Commutants and bicommutants of algebras of unbounded operators

Atsushi Inoue, Hideyuki Ueda, and Toshiyuki Yamauchi
Department of Applied Mathematics, Fukuoka University, 814 Fukuoka, Japan

(Received 24 April 1986; accepted for publication 17 September 1986)

The first purpose of this paper is to show that for each Op^*-algebra $(\mathscr{M},\mathscr{D})$ whose weak commutant $\mathscr{M}'_w$ is an algebra, there exists a closed Op^*-algebra $(\hat{\mathscr{M}},\hat{\mathscr{D}})$, which is the smallest extension of $(\mathscr{M},\mathscr{D})$ satisfying $\hat{\mathscr{M}}'_w = \mathscr{M}'_w$ and $\hat{\mathscr{M}}'_w \hat{\mathscr{D}} = \hat{\mathscr{D}}$. The second purpose is to characterize an unbounded bicommutant $\mathscr{M}''_{w\sigma}$ of an Op^*-algebra $\mathscr{M}$. The third purpose is to generalize the well-known Radon–Nikodym theorem for von Neumann algebras to Op^*-algebras $\mathscr{M}$ satisfying the von Neumann density type theorem $\bar{\mathscr{M}}^{t_s^*} = \mathscr{M}''_{w\sigma}$.

I. INTRODUCTION

In recent years algebras of unbounded operators have been studied by many mathematicians, both from the mathematical point of view and for applications in quantum physics.[1–6] In particular, Powers introduced the notions of closed, Hermitian, and self-adjoint Op^*-algebras in analogy with the notions of closed, Hermitian, and self-adjoint operators, respectively. The notation of self-adjointness has been indispensable in order to study Op^*-algebras in detail. For such a study the weak commutant $\mathscr{M}'_w$ of an Op^*-algebra $(\mathscr{M},\mathscr{D})$ plays an important role. Indeed, it is well known that if $\mathscr{M}$ is self-adjoint, then $\mathscr{M}'_w \mathscr{D} = \mathscr{D}$ and $\mathscr{M}'_w$ is a von Neumann algebra. Though the condition $\mathscr{M}'_w \mathscr{D} = \mathscr{D}$ does not necessarily imply the self-adjointness of $\mathscr{M}$, many results have been obtained under the assumption $\mathscr{M}'_w \mathscr{D} = \mathscr{D}$ without the one of self-adjointness of $\mathscr{M}$ (see Refs. 5 and 6). In Sec. III we shall show that though the condition "$\mathscr{M}'_w$ is an algebra" does not necessarily imply $\mathscr{M}'_w \mathscr{D} = \mathscr{D}$, there exists a closed Op^*-algebra $(\hat{\mathscr{M}},\hat{\mathscr{D}})$ which is the smallest extension of the Op^*-algebra $(\mathscr{M},\mathscr{D})$ satisfying $\hat{\mathscr{M}}'_w = \mathscr{M}'_w$ and $\hat{\mathscr{M}}'_w \hat{\mathscr{D}} = \hat{\mathscr{D}}$.

Mathot has investigated topological properties of an unbounded weak commutant $\mathscr{M}'_\sigma$ and unbounded bicommutants $\mathscr{M}''_{\sigma\sigma}$, $\mathscr{M}''_{w\sigma}$ of an Op^*-algebra $\mathscr{M}$ (see Ref. 7), and obtained that if $\mathscr{M}$ is an algebra of bounded operators (not necessarily leaving $\mathscr{D}$ invariant) then the strong*-closure $\bar{\mathscr{M}}^{t_s^*}$ of $\mathscr{M}$ equals the unbounded bicommutant $\mathscr{M}''_{w\sigma} = \mathscr{M}''_{\sigma\sigma}$. This result is a generalization of the von Neumann density theorem for bounded operator algebras to the unbounded case. In Sec. IV we shall characterize the unbounded bicommutant $\mathscr{M}''_{w\sigma}$ of $\mathscr{M}$ by using $(\hat{\mathscr{M}},\hat{\mathscr{D}})$ instead of $(\mathscr{M},\mathscr{D})$ and generalize the above Mathot result.

In Sec. V we shall study a Radon–Nikodym theorem for unbounded operator algebras. Such a study was begun by Gudder,[8] and hereafter was developed in Ref. 9. Here we generalize the well-known Radon–Nikodym theorem[10–12] for von Neumann algebras to Op^*-algebras satisfying the von Neumann density type theorem (that is, $\bar{\mathscr{M}}^{t_s^*} = \mathscr{M}''_{w\sigma}$), and investigate the Radon–Nikodym theorem obtained in Ref. 9 in more detail. Further, we apply the above results to the spatial theory for unbounded operator algebras.

II. PRELIMINARIES

Let $\mathscr{D}$ be a dense subspace in a Hilbert space $\mathscr{G}$. We denote by $C(\mathscr{D},\mathscr{G})$ the set of all linear operators X such that $\mathscr{D}(X)\cap\mathscr{D}(X^*)\supset\mathscr{D}$, and define a subset $\mathscr{L}^\dagger(\mathscr{D})$ of $C(\mathscr{D},\mathscr{G})$ by $\mathscr{L}^\dagger(\mathscr{D}) = \{X\in C(\mathscr{D},\mathscr{G});\ \mathscr{D}(X)=\mathscr{D},\ X\mathscr{D}\subset\mathscr{D}, X^*\mathscr{D}\subset\mathscr{D}\}$. Then $C(\mathscr{D},\mathscr{G})$ is a $*$-invariant vector space with the usual operations and the adjoint X^* and $\mathscr{L}^\dagger(\mathscr{D})$ is a $*$-algebra with involution $X^\dagger = X^*/\mathscr{D}$.

A $*$-subalgebra $\mathscr{M}$ of $\mathscr{L}^\dagger(\mathscr{D})$ with identity operator I is said to be an Op^*-algebra on $\mathscr{D}$. An Op^*-algebra $\mathscr{M}$ on $\mathscr{D}$ is also denoted by $(\mathscr{M},\mathscr{D})$.

Let $(\mathscr{M},\mathscr{D})$ by an Op^*-algebra. A locally convex topology on $\mathscr{D}$ defined by a family $\{\|\cdot\|_X; X\in\mathscr{M}\}$ of seminorms

$$\|\xi\|_X = \|X\xi\|,\quad \xi\in\mathscr{D}\,,$$

is said to be the induced topology on $\mathscr{D}$, which is denoted by $t_{\mathscr{M}}$. If $(\mathscr{D}, t_{\mathscr{M}})$ is complete, then $(\mathscr{M},\mathscr{D})$ is said to be closed. It follows from Ref. 5, Lemma 2.6, that for each Op^*-algebra $(\mathscr{M},\mathscr{D})$ there exists a closed Op^*-algebra $(\tilde{\mathscr{M}},\tilde{\mathscr{D}})$ that is the smallest closed extension of $(\mathscr{M},\mathscr{D})$, which is said to be the closure of $(\mathscr{M},\mathscr{D})$.

If $\mathscr{D} = \cap_{X\in\mathscr{M}}\mathscr{D}(X^*)$, then $(\mathscr{M},\mathscr{D})$ is said to be self-adjoint, and if $X^* = \overline{X^\dagger}$ for each $X\in\mathscr{M}$, then $(\mathscr{M},\mathscr{D})$ is said to be standard.

Let $\mathscr{A}$ be a $*$-algebra. A $*$-homomorphism π of $\mathscr{A}$ onto an Op^*-algebra on a dense subspace $\mathscr{D}(\pi)$ in a Hilbert space $\mathscr{G}_\pi$ is said to be a $*$-representation of $\mathscr{A}$ in $\mathscr{G}_\pi$ with domain $\mathscr{D}(\pi)$. A $*$-representation π of $\mathscr{A}$ is said to be closed (resp. self-adjoint, standard) if the Op^*-algebra $(\pi(\mathscr{A}),\mathscr{D}(\pi))$ is closed (resp. self-adjoint, standard).

Let π be a $*$-representation of $\mathscr{A}$. We let

$$\mathscr{D}(\tilde{\pi}) = \bigcap_{x\in\mathscr{A}}\mathscr{D}(\overline{\pi(x)}),$$

$$\tilde{\pi}(x)\xi = \overline{\pi(x)}\xi,\quad \text{for } x\in\mathscr{A} \text{ and } \xi\in\mathscr{D}(\tilde{\pi})\,.$$

Then $\tilde{\pi}$ is a closed $*$-representation of $\mathscr{A}$ that is the smallest closed extension of π, which is said to be the closure of π (see Ref. 5).

Let ϕ be a positive linear functional on a $*$-algebra $\mathscr{A}$. It is easily shown that $N_\phi \equiv \{x\in\mathscr{A}: \phi(x^*x)=0\}$ is a left ideal in $\mathscr{A}$. For each $x\in\mathscr{A}$ we denote by $\lambda_\phi(x)$ the coset of $\mathscr{A}/N_\phi$

 0022-2488/87/010001-07$02.50

that contains x, and define an inner product $(\ |\ )$ on $\lambda_\phi(\mathscr{A})$ by

$$(\lambda_\phi(x)|\lambda_\phi(y)) = \phi(y^*x), \quad x,y\in\mathscr{A}.$$

Let $\mathscr{G}_\phi$ be the Hilbert space that is completion of the pre-Hilbert space $\lambda_\phi(\mathscr{A})$. We denote by π_ϕ the closure of a $*$-representation π_ϕ^0 defined by

$$\pi_\phi^0(x)\lambda_\phi(y) = \lambda_\phi(xy), \quad x,y\in\mathscr{A}.$$

We call the triple $(\pi_\phi, \lambda_\phi, \mathscr{G}_\phi)$ the GNS construction for ϕ.

III. WEAK COMMUTANT

Let $\mathscr{M}$ be a closed Op^*-algebra on a dense subspace $\mathscr{D}$ in a Hilbert space $\mathscr{G}$. We define a weak commutant $\mathscr{M}'_w$ of $\mathscr{M}$ as follows:

$$\mathscr{M}'_w = \{C\in\mathscr{B}(\mathscr{G}):\ (CX\xi|\eta) = (C\xi|X^\dagger\eta) \text{ for all } X\in\mathscr{M},\ \xi,\eta\in\mathscr{D}\}.$$

Then $\mathscr{M}'_w$ is a $*$-invariant weakly closed subspace of $\mathscr{B}(\mathscr{G})$, but it is not necessarily an algebra.[5,13] For the relation between the self-adjointness of $\mathscr{M}$ and the weak commutant $\mathscr{M}'_w$ we have the following lemma.

Lemma 3.1: Let $(\mathscr{M},\mathscr{D})$ be a closed Op^*-algebra. Consider the following statements: (1) $\mathscr{M}$ is standard, (2) $\mathscr{M}$ is self-adjoint, (3) $\mathscr{M}'_w\mathscr{D} = \mathscr{D}$, (3′) $\bar{X}$ is affiliated with $\mathscr{M}''_{ww}$ for each $X\in\mathscr{M}$, (4) $\mathscr{M}'_w$ is a von Neumann algebra. Then the following implications hold:

$$(1)\Rightarrow(2)\Rightarrow\begin{array}{c}(3)\\ \Updownarrow\\ (3')\end{array}\Rightarrow(4).$$

Though the above implication $(4)\Rightarrow(3)$ does not necessarily hold, we have the following.

Theorem 3.2: Suppose $(\mathscr{M},\mathscr{D})$ is an Op^*-algebra such that $\mathscr{M}'_w$ is an algebra. Then there exists a closed Op^*-algebra $(\hat{\mathscr{M}},\hat{\mathscr{D}})$, which is the smallest extension of $(\mathscr{M},\mathscr{D})$ satisfying $\hat{\mathscr{M}}'_w = \mathscr{M}'_w$ and $\hat{\mathscr{M}}'_w\hat{\mathscr{D}} = \hat{\mathscr{D}}$.

Proof: For each $X\in\mathscr{M}$ we define an operator X_1, which is an extension of X as follows:

$$\mathscr{D}_1 = \left\{\sum_{i=1}^n C_i\xi_i:\ C_i\in\mathscr{M}'_w,\ \xi_i\in\mathscr{D},\ n=1,2,\ldots\right\},$$

$$X_1\left(\sum_{i=1}^n C_i\xi_i\right) = \sum_{i=1}^n C_iX\xi_i, \text{ for } \sum_{i=1}^n C_i\xi_i\in\mathscr{D}_1.$$

Since

$$\left(X_1\left(\sum_{i=1}^n C_i\xi_i\right)\middle|\eta\right) = \left(\sum_{i=1}^n C_iX\xi_i\middle|\eta\right) = \left(\sum_{i=1}^n C_i\xi_i\middle|X^\dagger\eta\right),$$

for all $\Sigma_{i=1}^n C_i\xi_i\in\mathscr{D}_1$ and $\eta\in\mathscr{D}$, it follows that X_1 is a well-defined linear operator on $\mathscr{D}_1$ that is an extension of X. Since $\mathscr{M}'_w$ is an algebra, it follows that

$$\left(X_1\left(\sum_{i=1}^n C_i\xi_i\right)\middle|\sum_{j=1}^m D_j\eta_j\right) = \sum_{i=1}^n\sum_{j=1}^m (D_j^*C_iX\xi_i|\eta_j)$$
$$= \sum_{i=1}^n\sum_{j=1}^m (D_j^*C_i\xi_i|X^\dagger\eta_j)$$
$$= \left(\sum_{i=1}^n C_i\xi_i\middle|\sum_{j=1}^m D_jX^\dagger\eta_j\right),$$

for all $\Sigma_{i=1}^n C_i\xi_i, \Sigma_{j=1}^m D_j\eta_j\in\mathscr{D}_1$, and $X\in\mathscr{M}$, which implies that $(\mathscr{M}_1,\mathscr{D}_1)$ is an Op^*-algebra satisfying $\mathscr{M}'_w = (\mathscr{M}_1)'_w$ and $\mathscr{M}'_w\mathscr{D}_1 = \mathscr{D}_1$. Hence it is easily shown that the closure $(\hat{\mathscr{M}},\hat{\mathscr{D}})$ of the Op^*-algebra $(\mathscr{M}_1,\mathscr{D}_1)$ satisfies our assertions. This completes the proof.

For polynomial algebras we have the following.

Corollary 3.3: Let $\mathscr{D}$ be a dense subspace in a Hilbert space $\mathscr{G}$. Then the following statements hold.

(1) Let X be an Hermitian operator in $\mathscr{L}^\dagger(\mathscr{D})$ and $\mathscr{P}(X)$ be the Op^*-algebra on $\mathscr{D}$ generated by X. Suppose X is essentially self-adjoint. Then there uniquely exists a closed Op^*-algebra $(\widehat{\mathscr{P}(X)},\hat{\mathscr{D}})$ which is an extension of $(\mathscr{P}(X),\mathscr{D})$ such that $\widehat{\mathscr{P}(X)}'_w = \mathscr{P}(X)'_w$ and $\widehat{\mathscr{P}(X)}'_w\hat{\mathscr{D}} = \hat{\mathscr{D}}$, which is a standard Op^*-algebra with $\hat{\mathscr{D}} = \cap_{n=1}^\infty\mathscr{D}(\bar{X}^n)$.

(2) Let X and Y be Hermitian operators in $\mathscr{L}^\dagger(\mathscr{D})$ such that $XY = YX$, and $\mathscr{P}(X,Y)$ be the commutative Op^*-algebra on $\mathscr{D}$ generated by X and Y. Then the following statements hold.

(a) Suppose X and Y are essentially self-adjoint. Then $\mathscr{P}(X,Y)'_w$ is an algebra.

(b) Suppose X^n *and* Y^n are essentially self-adjoint for $n = 1,2,\ldots$. Put

$$\mathscr{D}^*(\mathscr{P}(X,Y)) = \bigcap_{P(X,Y)\in\mathscr{P}(X,Y)}\mathscr{D}(P(X,Y)^*),$$
$$P^*(X,Y)\xi = \bar{P}(X,Y)^*\xi, \quad \xi\in\mathscr{D}^*(\mathscr{P}(X,Y)),$$

where $\bar{P}$ is the conjugate of a polynomial P. Then $\mathscr{P}^*(X,Y) = \{P^*(X,Y): P \text{ is a polynomial}\}$ is a self-adjoint Op^*-algebra on $\mathscr{D}^*(\mathscr{P}(X,Y))$.

(c) $\bar{X}$ and $\bar{Y}$ are self-adjoint operators with mutually commuting spectral projections if and only if $\mathscr{P}(X,Y)$ is standard. In this case

$$\mathscr{D} = \bigcap_{n=1}^\infty(\mathscr{D}(\bar{X}^n)\cap\mathscr{D}(\bar{Y}^n)).$$

Proof: (1) By Ref. 5, Lemma 3.2, $\mathscr{P}(X)'_w$ is an algebra. It hence follows from Theorem 3.2 that there exists a closed Op^*-algebra $(\widehat{\mathscr{P}(X)},\hat{\mathscr{D}})$ which is the smallest extension of $(\mathscr{P}(X),\mathscr{D})$ such that $\widehat{\mathscr{P}(X)}'_w = \mathscr{P}(X)'_w$ and $\widehat{\mathscr{P}(X)}'_w\hat{\mathscr{D}} = \hat{\mathscr{D}}$, which implies by Ref. 13, Theorem 2.1, that $(\widehat{\mathscr{P}(X)},\hat{\mathscr{D}})$ is standard and

$$\hat{\mathscr{D}} = \bigcap_{n=1}^\infty\mathscr{D}(\bar{X}^n).$$

(2) (a) Since $\mathscr{P}(X,Y)'_w = \mathscr{P}(X)'_w\cap\mathscr{P}(Y)'_w$, it follows from (Ref. 5, Lemma 3.2) that $\mathscr{P}(X,Y)'_w$ is an algebra.

(b) It is easily shown that $\mathscr{P}^*(X,Y)$ is an Op^*-algebra, which implies that $\mathscr{P}^*(X,Y)$ is self-adjoint.

(c) This follows from Ref. 13, Theorem 3.2.

For the study of Op^*-algebra $(\mathscr{M},\mathscr{D})$ many results have been obtained under the assumption that $(\mathscr{M},\mathscr{D})$ is self-adjoint.[2,5,6,9] By Theorem 3.2 we can obtain similar results under the weaker condition "$\mathscr{M}'_w$ is an algebra" than the self-adjointness of $\mathscr{M}$. For example, we have a slight extension of Powers results Ref. 5, Theorems 7.1 and 7.3.

Corollary 3.4: Suppose $(\mathscr{M},\mathscr{D})$ is a commutative Op^*-algebra such that $\mathscr{M}'_w$ is an algebra. Then $(\hat{\mathscr{M}},\hat{\mathscr{D}})$ is standard if and only if $\mathscr{M}''_{ww}$ is commutative.

Corollary 3.5: Let $\mathscr{A}$ be a commutatuve $*$-algebra with identity e and ϕ be a positive linear functional on $\mathscr{A}$. Suppose $\pi_\phi(\mathscr{A})'_w$ is an algebra. Then ϕ is strongly positive if and only if $\pi_\phi(\mathscr{A})$ is standard.

IV. UNBOUNDED BICOMMUTANTS

In Ref. 7 Mathot has introduced an unbounded commutant $\mathscr{M}'_\sigma$ and unbounded bicommutants $\mathscr{M}''_{w\sigma}$ and $\mathscr{M}''_{\sigma\sigma}$ of a $*$-invariant set $\mathscr{M}$ of unbounded operators, and investigated when $\mathscr{M}''_{w\sigma}$ or $\mathscr{M}''_{\sigma\sigma}$ equals the closure $\overline{\mathscr{M}}^{t_s^*}$ of $\mathscr{M}$ relative to the strong*-topology t_s^*. In particular, she has obtained the result that if $(\mathscr{M},\mathscr{D})$ is an algebra of bounded operators (not necessarily leaving $\mathscr{D}$ invariant), then $\overline{\mathscr{M}}^{t_s^*}=\mathscr{M}''_{\sigma\sigma}=\mathscr{M}''_{w\sigma}$.

In this section we characterize the unbounded bicommutant $\mathscr{M}''_{w\sigma}$ of a $*$-invariant set $\mathscr{M}$ of unbounded operators and generalize the above Mathot result.

We first define some locally convex topologies on $C(\mathscr{D},\mathscr{G})$. Locally convex topologies on $C(\mathscr{D},\mathscr{G})$ defined by systems $\{P_{\xi,\eta}(\cdot);\ \xi,\ \eta\in\mathscr{D}\}$, $\{P_\xi(\cdot);\ \xi\in\mathscr{D}\}$, and $\{P^*_\xi(\cdot);\xi\in\mathscr{D}\}$ of seminorms,

$$P_{\xi,\eta}(X)=(X\xi\,|\eta)\ ,$$

$$P_\xi(X)=\|X\xi\|,\quad P^*_\xi(X)=\|X\xi\|+\|X^*\xi\|\ ,$$

are said to be a weak topology, a strong topology, and a strong*-topology, which are denoted by t_w, t_s, and t_s^*, respectively.

We next define unbounded commutants $\mathscr{M}'_\sigma$, $\mathscr{M}'_c$ and unbounded bicommutants $\mathscr{M}''_{w\sigma}$, $\mathscr{M}''_{wc}$, $\mathscr{M}''_{\sigma\sigma}$, and $\mathscr{M}''_{cc}$ of a $*$-invariant subset $\mathscr{M}$ of $C(\mathscr{D},\mathscr{G})$ as follows:

$$\mathscr{M}'_\sigma=\{S\in C(\mathscr{D},\mathscr{G});\ (X\xi\,|S\eta)=(S^*\xi\,|X^*\eta),\ \text{for all } X\in\mathscr{M},\ \xi,\eta\in\mathscr{D}\}\ ,$$

$$\mathscr{M}'_c=\mathscr{M}'_\sigma\cap\mathscr{L}^+(\mathscr{D})\ ,$$

$$\mathscr{M}''_{w\sigma}=\{X\in C(\mathscr{D},\mathscr{G});\ (CX\xi\,|\eta)=(C\xi\,|X^*\eta),\ \text{for all } C\in\mathscr{M}'_w,\ \xi,\eta\in\mathscr{D}\}\ ,$$

$$\mathscr{M}''_{\sigma\sigma}=\{X\in C(\mathscr{D},\mathscr{G});\ (S\xi\,|X\eta)=(X^*\xi\,|S^*\eta),\ \text{for all } S\in\mathscr{M}'_\sigma,\ \xi,\eta\in\mathscr{D}\}\ ,$$

$$\mathscr{M}''_{cc}=\{X\in\mathscr{L}^+(\mathscr{D});\ XS=SX,\ \text{for all } S\in\mathscr{M}'_c\}\ .$$

Then we have the following.

Lemma 4.1: (See Ref. 7.) Let $\mathscr{M}$ be a $*$-invariant subset of $C(\mathscr{D},\mathscr{G})$. Then the following statements hold: (1) $\mathscr{M}'_\sigma$ is a strongly*-closed subspace of $C(\mathscr{D},\mathscr{G})$, (2) $\mathscr{M}'_c$ is an Op^*-algebra on $\mathscr{D}$, (3) $\mathscr{M}''_{w\sigma}$ is a strongly*-closed $*$-invariant subspace of $C(\mathscr{D},\mathscr{G})$ containing $\mathscr{M}\cup\mathscr{M}''_{ww}$, (4) $\mathscr{M}''_{\sigma\sigma}$ is a strongly*-closed $*$-invariant subspace of $C(\mathscr{D},\mathscr{G})$ containing $\mathscr{M}$, and (5) if $\mathscr{M}\mathscr{D}=\mathscr{D}$, then $\mathscr{M}''_{cc}$ is an Op^*-algebra on $\mathscr{D}$ containing $\mathscr{M}$.

We now investigate the relation between the unbounded commutant $\mathscr{M}''_{w\sigma}$ and the closure $\overline{\mathscr{M}''_{ww}}^{t_s^*}$ of $\mathscr{M}''_{ww}$ relative to the strong*-topology.

Theorem 4.2: Let $\mathscr{M}$ be a $*$-invariant subset of $C(\mathscr{D},\mathscr{G})$. Consider the following statements:

(1) $\mathscr{M}'_w$ is an algebra,

(2) $\overline{\mathscr{M}''_{ww}}^{t_s^*}=\mathscr{M}''_{w\sigma}$,

(3) $\overline{\mathscr{M}''_{ww}}^{t_s^*}\cap\mathscr{L}^+(\mathscr{D})=\mathscr{M}''_{wc}$.

Then the following implications hold:

$$\begin{array}{c}(1)\Longleftrightarrow(2)\\ \Downarrow\\ (3)\ .\end{array}$$

In particular if $\mathscr{M}$ is an Op^*-algebra on $\mathscr{D}$, then the statements (1)–(3) are equivalent.

Proof: (1)$\Rightarrow$(2). Since $\overline{\mathscr{M}''_{ww}}^{t_s^*}\subset\mathscr{M}''_{w\sigma}$ by Lemma 4.1(3), it is sufficient to show $\mathscr{M}''_{w\sigma}\subset\overline{\mathscr{M}''_{ww}}^{t_s^*}$. Take arbitrary $X\in\mathscr{M}''_{w\sigma}$. We let

$$\mathscr{D}_1=\left\{\sum_{k=1}^n C_k\xi_k:\ C_k\in\mathscr{M}'_w,\ \xi_k\in\mathscr{D}\right\},$$

$$X_1\left(\sum_{k=1}^n C_k\xi_k\right)=\sum_{k=1}^n C_kX\xi_k,\ \text{for}\quad\sum_{k=1}^n C_k\xi_k\in\mathscr{D}_1\ .$$

Since $\mathscr{M}'_w$ is an algebra, we have

$$\begin{aligned}\left(X_1\left(\sum_{i=1}^n C_i\xi_i\right)\Big|\sum_{j=1}^m D_j\eta_j\right)&=\sum_{i=1}^n\sum_{j=1}^m(X\xi_i|C_i^*D_j\eta_j)\\&=\sum_{i=1}^n\sum_{j=1}^m(D_j^*C_i\xi_i|X^*\eta_j)\\&=\left(\sum_{i=1}^n C_i\xi_i\Big|\sum_{j=1}^m D_jX^*\eta_j\right),\end{aligned}$$

for each $\Sigma_{i=1}^n C_i\xi_i,\ \Sigma_{j=1}^m D_j\eta_j\in\mathscr{D}_1$, which implies that $X_1\in C(\mathscr{D}_1,\mathscr{G})$ and $\overline{X}_1$ is affiliated with $\mathscr{M}''_{ww}$. Hence there exists a sequence $\{A_n\}$ in $\mathscr{M}''_{ww}$ such that $\lim_{n\to\infty}A_n\xi=\overline{X}_1\xi$ and $\lim_{n\to\infty}A_n^*\xi=X_1^*\xi$ for each $\xi\in\mathscr{D}_1$. Since $\mathscr{D}\subset\mathscr{D}_1$, it follows that $\lim_{n\to\infty}A_n\xi=X\xi$ and $\lim_{n\to\infty}A_n^*\xi=X^*\xi$ for each $\xi\in\mathscr{D}$. Hence $X\in\overline{\mathscr{M}''_{ww}}^{t_s^*}$.

(2)$\Rightarrow$(1). For each $C_1,C_2\in\mathscr{M}'_w$, $X\in\mathscr{M}$, and $\xi,\eta\in\mathscr{D}$ we have

$$\begin{aligned}(C_1C_2X\xi\,|\eta)&=\lim_\alpha(C_2A_\alpha\xi\,|C_1^*\eta)\\&=\lim_\alpha(C_2\xi\,|A_\alpha^*C_1^*\eta)\\&=\lim_\alpha(C_1C_2\xi\,|A_\alpha^*\eta)\\&=(C_1C_2\xi\,|X^*\eta)\ ,\end{aligned}$$

where $\{A_\alpha\}$ is a net in $\mathscr{M}''_{ww}$ which converges to X for the strong*-topology. Hence $C_1C_2\in\mathscr{M}'_w$.

(2)$\Rightarrow$(3). This is trivial.

Suppose $\mathscr{M}$ is an Op^*-algebra on $\mathscr{D}$. Since $\mathscr{M}\subset\mathscr{M}''_{wc}$, we can prove the implication (3)$\Rightarrow$(1) in the same way as in (2)$\Rightarrow$(1). This completes the proof.

Let $\mathscr{M}$ be a $*$-invariant subset of $C(\mathscr{D},\mathscr{G})$ such that $\mathscr{M}'_w$ is an algebra. Then

$$\begin{array}{c}\overline{\mathscr{M}}^{t_s^*}\subset\mathscr{M}''_{\sigma\sigma}\subset\mathscr{M}''_{w\sigma}=\overline{\mathscr{M}''_{ww}}^{t_s^*}\ ,\\ \cup\\ \mathscr{M}''_{cc}\subset\mathscr{M}''_{wc}=\overline{\mathscr{M}''_{ww}}^{t_s^*}\cap\mathscr{L}^\dagger(\mathscr{D})\ .\end{array}$$

By Theorem 4.2 we have the following Corollary 4.3, which is a slight generalization of Ref. 7, Proposition 9.

Corollary 4.3: Suppose $\mathscr{M}$ is a $*$-invariant subset of

$C(\mathscr{D},\mathscr{G})$ with identity operator such that $\mathscr{M}_b'' = \mathscr{M}_{ww}''$, where $\mathscr{M}_b = \{A\in\mathscr{M};\ \bar{A}$ is bounded$\}$. Then $\overline{\mathscr{M}}^{t_s^*} = \mathscr{M}_{\sigma\sigma}'' = \mathscr{M}_{w\sigma}''$.

V. RADON-NIKODYM THEOREMS

We first generalize the well-known following Radon–Nikodym theorem for von Neumann algebras to the unbounded case: every σ-weakly continuous positive linear functional ϕ on a von Neumann algebra $\mathscr{M}$ with a cyclic and separating vector ξ_0 is represented as

$$\phi(A) = (A\xi\,|\xi),\quad A\in\mathscr{M},$$

for a unique vector $\xi\in\mathscr{P}^{\#} \equiv \overline{\{AA^*\xi_0 : A\in\mathscr{M}\}}$: in particular,

$$\phi(A) = (AH\xi_0|H\xi_0),\quad A\in\mathscr{M},$$

for some positive self-adjoint operator H affiliated with $\mathscr{M}$.

Let $\mathscr{M}$ be an Op^*-algebra on a dense subspace $\mathscr{D}$ in a Hilbert space $\mathscr{G}$. We let

$$\mathscr{D}^\infty(\mathscr{M}) = \left\{\{\xi_k\}\subset\mathscr{D};\ \sum_{k=1}^{\infty}\|X\xi_k\|^2 < \infty,\ \text{for all } X\in\mathscr{M}\right\};$$

$$[X]\{\xi_k\} = \{X\xi_k\},\quad X\in\mathscr{M},\quad \{\xi_k\}\in\mathscr{D}^\infty(\mathscr{M});$$

$$[\mathscr{M}] = \{[X];\ X\in\mathscr{M}\}.$$

Then we have the following,

Lemma 5.1: (1) $([\mathscr{M}],\mathscr{D}^\infty(\mathscr{M}))$ is an Op^*-algebra.

(2) $(\mathscr{M},\mathscr{D})$ is closed if and only if $([\mathscr{M}],\mathscr{D}^\infty(\mathscr{M}))$ is closed.

(3) $\mathscr{M}_w'$ is an algebra if and only if $[\mathscr{M}]_w'$ is an algebra.

(4) $\mathscr{M}_w'\mathscr{D} = \mathscr{D}$ if and only if $[\mathscr{M}]_w'\mathscr{D}^\infty(\mathscr{M}) = \mathscr{D}^\infty(\mathscr{M})$.

(5) $(\mathscr{M},\mathscr{D})$ is self-adjoint if and only if $([\mathscr{M}],\mathscr{D}^\infty(\mathscr{M}))$ is self-adjoint.

Let $\mathscr{M}$ by an Op^*-algebra on a dense subspace $\mathscr{D}$ in a Hilbert space $\mathscr{G}$ and $\mathscr{G}^\infty$ be the direct sum of the Hilbert spaces $\mathscr{G}_n = \mathscr{G}$ for $n = 1,2,\ldots$. The weakest locally convex topology on $\mathscr{M}$ such that the map $X\to[X]$ of $\mathscr{M}$ into $(C(\mathscr{D}^\infty(\mathscr{M}),\mathscr{G}^\infty),\ t_w)$ [resp. $(C(\mathscr{D}^\infty(\mathscr{M}),\mathscr{G}^\infty),\ t_s)$, $(C(\mathscr{D}^\infty(\mathscr{M}),\mathscr{G}^\infty),\ t_s^*)$] is said to be a σ-weak (resp. σ-strong, σ-strong*) topology for $\mathscr{M}$, which is denoted by $t_{\sigma w}^{\mathscr{M}}$ (resp. $t_{\sigma s}^{\mathscr{M}}$, $t_{\sigma s}^{*\mathscr{M}}$).

Lemma 5.2: Let $\mathscr{M}$ be a closed Op^*-algebra on a dense subspace $\mathscr{D}$ in a Hilbert space $\mathscr{G}$ such that $\mathscr{M}_w'\mathscr{D} = \mathscr{D}$. Then the following statements hold.

(1) Suppose $\overline{\mathscr{M}}^{t_s^*} = \mathscr{M}_{w\sigma}''$. Then the positive linear functional ϕ on $\mathscr{M}$ is weakly continuous if and only if ϕ is represented as

$$\phi(X) = \sum_{k=1}^{n}(X\xi_k|\xi_k),\quad X\in\mathscr{M},$$

for some finite subset $\{\xi_k\}_{1\leqslant k\leqslant n}$ of $\mathscr{D}$.

(2) Suppose $\overline{[\mathscr{M}]}^{t_s^*} = [\mathscr{M}]_{w\sigma}''$. Then a positive linear functional ϕ on $\mathscr{M}$ is continuous with respect to the σ-weak topology for $\mathscr{M}$ if and only if ϕ is represented as

$$\phi(X) = \sum_{k=1}^{\infty}(X\xi_k|\xi_k),\quad X\in\mathscr{M},$$

for some $\{\xi_k\}\in\mathscr{D}^\infty(\mathscr{M})$.

In this case, if $\mathscr{M}_{ww}''$ possesses a separating vector, then

$$\phi(X) = (X\xi\,|\xi),\quad X\in\mathscr{M},$$

for some $\xi\in\mathscr{D}$.

Proof: We will first prove statement (2). Suppose ϕ is continuous relative to the σ-weak topology $t_{\sigma w}^{\mathscr{M}}$ for $\mathscr{M}$. Then there exists an element $\{\eta_k\}$ of $\mathscr{D}^\infty(\mathscr{M})$ such that

$$|\phi(X)|\leqslant\left|\sum_{k=1}^{\infty}(X\eta_k\,|\,\eta_k)\right|,$$

for all $X\in\mathscr{M}$. Hence it follows that

$$\|\lambda_\phi(X)\|\leqslant\|[X]\{\eta_k\}\|, \tag{5.1}$$

for all $X\in\mathscr{M}$. We let

$$C_0[X]\{\eta_k\} = \lambda_\phi(X),\quad X\in\mathscr{M}.$$

Then by (5.1), C_0 is extended to a continuous linear map $\overline{C}_0$ of $\overline{[\mathscr{M}]\{\eta_k\}}$ into $\mathscr{G}_\phi$. Since $[\mathscr{M}]_w'\mathscr{D}^\infty(\mathscr{M}) = \mathscr{D}^\infty(\mathscr{M})$ by Lemma 5.1 and $\overline{[\mathscr{M}]}^{t_s^*} = [\mathscr{M}]_{w\sigma}''$, we have

$$\overline{[\mathscr{M}]\{\eta_k\}} = \overline{[\mathscr{M}]_{ww}''\{\eta_k\}},$$

which implies that the projection P from $\mathscr{G}^\infty$ to $\overline{[\mathscr{M}]\{\eta_k\}}$ belongs to $[\mathscr{M}]_w'$. We now show

$$C\equiv(\overline{C}_0P)^*(\overline{C}_0P)\in[\mathscr{M}]_w'. \tag{5.2}$$

For each $A\in\mathscr{M}_{ww}''$ and $x, y\in\mathscr{G}^\infty$ there exist sequences $\{X_n\}$ and $\{Y_n\}$ in $\mathscr{M}$ such that $\lim_{n\to\infty}[X_n]\{\eta_k\} = Px$ and $\lim_{n\to\infty}[Y_n]\{\eta_k\} = Py$, and since

$$[\mathscr{M}_{ww}''] = [\mathscr{M}]_{ww}''\subset[\mathscr{M}]_{w\sigma}'' = \overline{[\mathscr{M}]}^{t_s^*},$$

there exists a net $\{X_\alpha\}$ in $\mathscr{M}$ such that

$$\lim_\alpha\sum_{k=1}^{\infty}\|X_\alpha\xi_k - A\xi_k\|^2 = \lim_\alpha\sum_{k=1}^{\infty}\|X_\alpha^+\xi_k - A^*\xi_k\|^2 = 0,$$

for each $\{\xi_k\}\in\mathscr{D}^\infty(\mathscr{M})$. Then we have

$$\begin{aligned}
(C[A]x|y) &= (\overline{C}_0P[A]x|\overline{C}_0Py)\\
&= \lim_{m,n\to\infty}(\overline{C}_0[A][X_n]\{\eta_k\}|\overline{C}_0[Y_m]\{\eta_k\})\\
&= \lim_{m,n\to\infty}\lim_\alpha(\overline{C}_0[X_\alpha X_n]\{\eta_k\}|\overline{C}_0[Y_m]\{\eta_k\})\\
&= \lim_{m,n\to\infty}\lim_\alpha(\lambda_\phi(X_\alpha X_n)|\lambda_\phi(Y_m))\\
&= \lim_{m,n\to\infty}\lim_\alpha(\lambda_\phi(X_n)|\lambda_\phi(X_\alpha^+Y_m))\\
&= \lim_{m,n\to\infty}\lim_\alpha(\overline{C}_0[X_n]\times\{\eta_k\}|\overline{C}_0[X_\alpha]^+[Y_m]\{\eta_k\})\\
&= \lim_{m,n\to\infty}(\overline{C}_0[X_n]\{\eta_k\}|\overline{C}_0[A]^*[Y_m]\{\eta_k\})\\
&= (\overline{C}_0Px|\overline{C}_0[A]^*Py) = (Cx|[A]^*y).
\end{aligned}$$

Hence $C\in[\mathscr{M}''_{ww}]'_w=[\mathscr{M}]'_w$. Since $[\mathscr{M}]'_w\mathscr{D}^\infty(\mathscr{M})=\mathscr{D}^\infty(\mathscr{M})$, it follows that $C^{1/2}\in[\mathscr{M}]'_w$ and $\{\xi_k\}\equiv C^{1/2}\{\eta_k\}\in\mathscr{D}^\infty(\mathscr{M})$, which implies that

$$\begin{aligned}\phi(X)=(\lambda_\phi(X)|\lambda_\phi(I))&=(C[X]\{\eta_k\}|\{\eta_k\})\\&=([X]C^{1/2}\{\eta_k\}|C^{1/2}\{\eta_k\})\\&=\sum_{k=1}^\infty(X\xi_k|\xi_k)\,,\end{aligned}$$

for all $X\in\mathscr{M}$.

The converse is trivial.

Suppose $\mathscr{M}''_{ww}$ possesses a separating vector η_0 and $\phi(X)=\Sigma_{k=1}^\infty(X\xi_k|\xi_k)$, $X\in\mathscr{M}$, for some $\{\xi_k\}\in\mathscr{D}^\infty(\mathscr{M})$. We let

$$E_1=\text{Proj}\ \overline{[\mathscr{M}]\{\xi_k\}},\quad E_2=\begin{pmatrix}1&0&\cdot&\cdots\\0&0&\cdot&\cdots\\\vdots&\vdots&&\ddots\end{pmatrix}.$$

It is proved in the same way as $P\in[\mathscr{M}]'_w$ that $E_1\in[\mathscr{M}]'_w$. It is clear that $E_2\in[\mathscr{M}]'_w$ and $Z(E_1)\leqslant Z(E_2)=I$, where $Z(E_i)$ is the central support of E_i $(i=1,2)$. Further we have

$$\{\xi_k\}\in E_1\mathscr{G}^\infty$$

and

$$\overline{[\mathscr{M}]''_{ww}E_1\{\xi_k\}}=\overline{[\mathscr{M}]\{\xi_k\}}=E_1\mathscr{G}^\infty\,,$$

and

$$\tilde{\eta}_0=\begin{pmatrix}\eta_0\\0\\\vdots\end{pmatrix}\in E_2\mathscr{G}^\infty$$

and

$$\overline{E_2[\mathscr{M}]'_wE_2\tilde{\eta}_0}=\begin{pmatrix}\overline{\mathscr{M}'_w\eta_0}\\0\\\vdots\end{pmatrix}=E_2\mathscr{G}^\infty\,,$$

since η_0 is a separating vector for $\mathscr{M}''_{ww}$.

It hence follows from Ref. 10, Part III, Chap. I, Lemma 4, that there exists an operator V in $[\mathscr{M}]'_w$ such that $V^*V=E_1$ and $VV^*\leqslant E_2$. Then we have

$$VV^*V\{\xi_k\}=VE_1\{\xi_k\}=V\{\xi_k\}\,,$$

and so $V\{\xi_k\}\in E_2\mathscr{G}^\infty$, which implies by $[\mathscr{M}]'_w\times\mathscr{D}^\infty(\mathscr{M})=\mathscr{D}^\infty(\mathscr{M})$, that

$$V\{\xi_k\}=\begin{pmatrix}\xi\\0\\\vdots\end{pmatrix}\text{ for some }\xi\in\mathscr{D}\,.$$

Thus we have

$$\begin{aligned}\phi(X)=([X]\{\xi_k\}|\{\xi_k\})&=(V^*V[X]\{\xi_k\}|\{\xi_k\})\\&=([X]V\{\xi_k\}|V\{\xi_k\})\\&=(X\xi|\xi)\,,\end{aligned}$$

for all $X\in\mathscr{M}$.

The proof of statement (1) is similar to (2).

Theorem 5.3: Let $(\mathscr{M},\mathscr{D})$ be a closed *Op**-algebra such that $\mathscr{M}'_w\mathscr{D}=\mathscr{D}$ and $\overline{[\mathscr{M}]}^{t^*_s}=[\mathscr{M}]''_{w\sigma}$ and a vector ξ_0 in $\mathscr{D}$ be strongly cyclic for $\mathscr{M}$ and separating for $\mathscr{M}''_{ww}$. Then, for every positive linear functional ϕ on $\mathscr{M}$, which is continuous with respect to the σ-weak topology of $\mathscr{M}$ there exists a unique element ξ of

$$\mathscr{P}^\#(\equiv\overline{\{AA^*\xi_0;\ A\in\mathscr{M}''_{ww}\}})\cap\mathscr{D}$$

such that

$$\phi(X)=(X\xi|\xi)\,,$$

for all $X\in\mathscr{M}$. In particular, there exists a positive self-adjoint operator H affiliated with $\mathscr{M}''_{ww}$ such that $\xi_0\in\mathscr{D}(H)$, $H\xi_0\in\mathscr{D}$, and

$$\phi(X)=(XH\xi_0|H\xi_0)\,,$$

for all $X\in\mathscr{M}$.

Proof: By Lemma 5.2 there exists a vector $\zeta_0\in\mathscr{D}$ such that

$$\phi(X)=(X\zeta_0|\zeta_0)\,,$$

for all $X\in\mathscr{M}$. We can prove this theorem in analogy with Ref. 12, Theorem 15.1 as follows. We define a σ-weakly continuous linear functional ω'_{ζ_0,ξ_0} on the von Neumann algebra $\mathscr{M}'_w$ by

$$\omega'_{\zeta_0,\xi_0}(C)=(C\zeta_0|\xi_0),\quad C\in\mathscr{M}'_w\,.$$

Let $\omega'_{\zeta_0,\xi_0}=R_{V'}\psi'$ be the polar decomposition of ω'_{ζ_0,ξ_0}, and put

$$\xi=V'^*\zeta_0\,.$$

Then since $\mathscr{M}'_w\mathscr{D}=\mathscr{D}$ and

$$\mathscr{P}^\#\cap\mathscr{D}=\{\zeta\in\mathscr{D};\ \omega'_{\zeta,\xi_0}\geqslant 0\}\,,$$

it follows that $\xi\in\mathscr{P}^\#\cap\mathscr{D}$ and $\zeta_0=V'\xi$, which implies

$$\begin{aligned}\phi(X)=(X\zeta_0|\zeta_0)&=(X\zeta_0|V'\xi)\\&=(XV'^*\zeta_0|\xi)\\&=(X\xi|\xi)\,,\end{aligned}$$

for all $X\in\mathscr{M}$. Thus the existence of ξ is shown. We next show the uniqueness of ξ. Suppose

$$\phi(X)=(X\xi_1|\xi_1)=(X\xi_2|\xi_2),\quad X\in\mathscr{M}\,,$$

for some $\xi_1,\xi_2\in\mathscr{P}^\#\cap\mathscr{D}$. Since

$$\mathscr{M}''_{ww}\subset\mathscr{M}''_{w\sigma}=\overline{\mathscr{M}}^{t^*_s}\,,$$

we have

$$(A\xi_1|\xi_1)=(A\xi_2|\xi_2)\,,$$

for all $A\in\mathscr{M}''_{ww}$, which implies by Ref. 12, Theorem 15.1, that $\xi_1=\xi_2$.

We now let

$$H_0C\xi_0=C\xi,\quad C\in\mathscr{M}'_w\,.$$

Since $\omega'_{\xi,\xi_0}\geqslant 0$, it follows that H_0 is a positive operator whose closure $\overline{H}_0$ is affiliated with $\mathscr{M}''_{ww}$. Friedrich's extension H of $\overline{H}_0$ fulfills our assertions. This completes the proof.

We next consider to extend Gudder's Radon–Nikodym theorem for *Op**-algebras.[8] Gudder introduced the following notions.

Definition 5.4: Let $(\mathscr{M},\mathscr{D})$ be a closed *Op**-algebra with a strongly cyclic vector ξ_0 and ϕ be a positive linear functional on $\mathscr{M}$. If $X\xi_0\rightarrow\lambda_\phi(X)$ is continuous (resp. closable), then ϕ is said to be ω_{ξ_0}-dominated (resp. strongly ω_{ξ_0}-absolutely continuous), where ω_{ξ_0} denotes a positive linear functional on $\mathscr{M}$ defined by $\omega_{\xi_0}(X) = (X\xi_0|\xi_0)$. If for each $X\in\mathscr{M}$ there exists a sequence $\{X_n\}$ in $\mathscr{M}$ such that $\lim_{n\to\infty} X_n\xi_0 = 0$ and $\lim_{n\to\infty}\lambda_\phi(X_n) = \lambda_\phi(X)$, then ϕ is said to be ω_{ξ_0}-singular.

We obtain the following result: for each strongly ω_{ξ_0}-absolutely continuous positive linear functional ϕ on $\mathscr{M}$ there exists a positive self-adjoint operator H in $\mathscr{G}$ such that

$$\phi(X) = (HX\xi_0|H\xi_0)\,,$$

for all $X\in\mathscr{M}$. However, as the relation between the above Gudder Radon–Nikodym derivative H and the Op^*-algebra $\mathscr{M}$ is vague, we obtained in Ref. 9 that under the assumption that $\pi_{\phi+\omega_{\xi_0}}(\mathscr{M})'_w$ is an algebra ϕ is strongly ω_{ξ_0}-absolutely continuous if and only if there exists a sequence $\{H'_n\}$ of positive operators in $\mathscr{M}'_w$ such that

$$\phi(X) = \lim_{n\to\infty}(H'_nX\xi_0|\xi_0)\,,$$

for all $X\in\mathscr{M}$. By investigating the above results in more details, we obtain the following lemma.

Lemma 5.5: Let $(\mathscr{M},\mathscr{D})$ be a closed Op^*-algebra with a strongly cyclic vector ξ_0 such that $\mathscr{M}'_w\mathscr{D} = \mathscr{D}$. Then the following statements hold.

(1) ϕ is a ω_{ξ_0}-dominated positive linear functional on $\mathscr{M}$ if and only if there exists a positive operator H' in $\mathscr{M}'_w$ such that

$$\phi(X) = (XH'\xi_0|H'\xi_0)\,,$$

for all $X\in\mathscr{M}$.

In the following (2) and (3), suppose ϕ is a positive linear functional on $\mathscr{M}$ such that $\pi_{\phi+\omega_{\xi_0}}(\mathscr{M})'_w$ is an algebra.

(2) The following statements are equivalent: (2′) ϕ is strongly ω_{ξ_0}-absolutely continuous, (2″) ϕ is represented as

$$\phi(X) = \lim_{n\to\infty}(H'_nX\xi_0|\xi_0),\quad X\in\mathscr{M}\,,$$

for some positive sequence $\{H'_n\}$ in $\mathscr{M}'_w$ such that $\lim_{n\to\infty} H'^{1/2}_n X\xi$ exists in $\mathscr{G}$ for each $X\in\mathscr{M}$, (2‴) ϕ is represented as

$$\phi(X) = (XH'\xi_0|H'\xi_0),\quad X\in\mathscr{M}\,,$$

for some positive self-adjoint operator H' affiliated with $\mathscr{M}'_w$ such that $\mathscr{D}(H')\supset\mathscr{M}\xi_0$.

(3) ϕ is decomposed into the sum

$$\phi = \phi_a + \phi_s\,,$$

where ϕ_a is a strongly ω_{ξ_0}-absolutely continuous positive linear functional on $\mathscr{M}$ and ϕ_s is a ω_{ξ_0}-singular positive linear functional on $\mathscr{M}$. If ϕ is strongly ω_{ξ_0}-absolutely continuous, then $\phi = \phi_a$; and if ϕ is ω_{ξ_0}-singular, then $\phi = \phi_s$.

Proof: Statements (1), (3), and the equivalence of (2′) and (2″) follow from Ref. 9, Theorem 3.2 and Theorem 3.3.

(2‴) ⇒ (2′). This is trivial.

(2″) ⇒ (2‴). We define an operator H'_0 in $\mathscr{G}$ as follows:

$$\mathscr{D}(H'_0) = \{\xi\in\mathscr{G};\ \lim_{n\to\infty} H'^{1/2}_n\xi \text{ exists in } \mathscr{G}\}\,,$$

$$H'_0\xi = \lim_{n\to\infty} H'^{1/2}_n\xi,\quad \text{for } \xi\in\mathscr{D}(H'_0)\,.$$

Then H'_0 is a positive operator in $\mathscr{G}$ such that $\mathscr{D}(H'_0)\supset\mathscr{M}\xi_0$. Further, since

$$\lim_{n\to\infty} H'^{1/2}_nA\xi = A\lim_{n\to\infty} H'^{1/2}_n\xi = AH'_0\xi\,,$$

for each $A\in\mathscr{M}''_{ww}$ and $\xi\in\mathscr{D}(H'_0)$, it follows that $\overline{H'_0}$ is affiliated with $\mathscr{M}'_w$. We denote by H' the Friedrichs self-adjoint extension of H'_0. Then H' is a positive self-adjoint operator in $\mathscr{G}$ affiliated with $\mathscr{M}'_w$ such that $\mathscr{D}(H')\supset\mathscr{M}\xi_0$ and

$$\phi(X) = (H'X\xi_0|H'\xi_0)\,,$$

for all $X\in\mathscr{M}$. Further, since $\{H'^{1/2}_nY\xi_0\}\subset\mathscr{D}$ and

$$\lim_{n\to\infty} XH'^{1/2}_nY\xi_0 = \lim_{n\to\infty} H'^{1/2}_nXY\xi_0 = H'XY\xi_0\,,$$

for each $X, Y\in\mathscr{M}$, so that

$$\begin{aligned}\phi(X) &= (H'X\xi_0|H'\xi_0)\\ &= (XH'\xi_0|H'\xi_0)\,,\end{aligned}$$

for all $X\in\mathscr{M}$. This completes the proof.

Theorem 5.6: Let $(\mathscr{M},\mathscr{D})$ be a closed Op^*-algebra such that $\mathscr{M}'_w\mathscr{D} = \mathscr{D}$ and a vector ξ_0 in $\mathscr{D}$ be strongly cyclic for $\mathscr{M}$ and separating for $\mathscr{M}''_{ww}$. Then the following statements hold.

(1) Every ω_{ξ_0}-dominated positive linear functional ϕ on $\mathscr{M}$ is represented as

$$\phi(X) = (XH\xi_0|H\xi_0),\quad X\in\mathscr{M}\,,$$

for a unique positive operator H in $\mathscr{M}''_{ww}$ such that $H\xi_0\in\mathscr{D}$.

(2) Every strongly ω_{ξ_0}-absolutely continuous positive linear functional ϕ on $\mathscr{M}$ such that $\pi_{\phi+\omega_{\xi_0}}(\mathscr{M})'_w$ is an algebra is represented as

$$\phi(X) = (X\xi|\xi),\quad X\in\mathscr{M}\,,$$

for a unique vector ξ in $\mathscr{P}^\#\cap\mathscr{D}$. In particular, there exists a positive self-adjoint operator H affiliated with $\mathscr{M}''_{ww}$ such that $\xi_0\in\mathscr{D}(H)$, $H\xi_0\in\mathscr{D}$ and

$$\phi(X) = (XH\xi_0|H\xi_0)\,,$$

for all $X\in\mathscr{M}$.

Proof: The existence of H in (1) and ξ, H in (2) follows from Lemma 5.5 and the proof of Theorem 5.3. We show the uniqueness of H in (1). Suppose

$$\phi(X) = (XH\xi_0|H\xi_0) = (XK\xi_0|K\xi_0),\quad X\in\mathscr{M}\,,$$

for some positive operators H and K in $\mathscr{M}''_{ww}$ such that $H\xi_0$, $K\xi_0\in\mathscr{D}$. Since ϕ is ω_{ξ_0}-dominated, it follows from Lemma 5.5 that there exist operators H' and K' in $\mathscr{M}'_w$ such that $H'X\xi_0 = XH\xi_0$ and $K'X\xi_0 = XK\xi_0$ for all $X\in\mathscr{M}$. Then we have

$$\begin{aligned}(H'^*H'X\xi_0|Y\xi_0) &= (XH\xi_0|YH\xi_0)\\ &= (XK\xi_0|YK\xi_0)\\ &= (K'^*K'X\xi_0|Y\xi_0)\,,\end{aligned}$$

for all $X, Y\in\mathscr{M}$, and hence $H'^*H' = K'^*K'$, which implies

$$\begin{aligned}(AH\xi_0|H\xi_0) = (AH'\xi_0|H'\xi_0) &= (AK'\xi_0|K'\xi_0)\\ &= (AK\xi_0|K\xi_0)\,,\end{aligned}$$

for all $A\in\mathscr{M}''_{ww}$. By the uniqueness of Sakai's Radon-Nikodym derivative (Ref. 12, Theorem 15.1), we have $H=K$. This completes the proof.

Remark 5.7: We may alter the condition "$\mathscr{M}'_w\mathscr{D}=\mathscr{D}$" in Lemma 5.2, Theorem 5.3, Lemma 5.5, and Theorem 5.6 to the weaker condition "$\mathscr{M}'_w$ is an algebra." For, by Theorem 3.2 there exists a closed Op^*-algebra $(\widehat{\mathscr{M}},\widehat{\mathscr{D}})$ which is an extension of $(\mathscr{M},\mathscr{D})$ satisfying $\widehat{\mathscr{M}}'_w=\mathscr{M}'_w$ and $\widehat{\mathscr{M}}'_w\widehat{\mathscr{D}}=\widehat{\mathscr{D}}$. Hence we have only to consider the Op^*-algebra $(\widehat{\mathscr{M}},\widehat{\mathscr{D}})$ instead of the Op^*-algebra $(\mathscr{M},\mathscr{D})$.

The spatial theory for Op^*-algebras was investigated in Refs. 14–16. In particular, Takesue[15] obtained many results for such a study. We have the following results by using Lemma 5.1, Theorem, 5.6, and Takesue's results (see Ref. 15, Theorem 3.1, Theorem 3.8).

Corollary 5.8: Let $(\mathscr{M},\mathscr{D})$ be a self-adjoint Op^*-algebra with a strongly cyclic vector ξ_0, η_0 be a separating vector for $\mathscr{M}''_{ww}$ and α be a $*$-automorphism of $\mathscr{M}$. Suppose $\overline{[\mathscr{M}]}'^{*}_{s}=[\mathscr{M}]''_{w\sigma}$. Then, if α and α^{-1} are continuous relative to the σ-weak topology for $\mathscr{M}$, then α is represented as

$$\alpha(X)=U^\dagger XU,\quad X\in\mathscr{M},$$

for some

$$U\in\mathscr{L}^\dagger(\mathscr{D})_u\equiv\{U\in\mathscr{L}^\dagger(\mathscr{D});\ \bar{U}\text{ is unitary}\}.$$

Corollary 5.9: Let $(\mathscr{M},\mathscr{D})$ be a self-adjoint Op^*-algebra, ξ_0 be a strongly cyclic vector for $\mathscr{M}$ and a separating vector for $\mathscr{M}''_{ww}$, and α be a $*$-automorphism of $\mathscr{M}$. Then the following statements hold.

(1) Suppose both the map $X\xi_0\to\alpha(X)\xi_0$ and $X\xi_0\to\alpha^{-1}(X)\xi_0$ are continuous. Then α is represented as

$$\alpha(X)=U^\dagger XU,\quad X\in\mathscr{M},$$

for some $U\in\mathscr{L}^\dagger(\mathscr{D})_u$.

(2) Suppose $\pi_{\omega_{\xi_0}+\omega_{\xi_0}\circ\alpha}(\mathscr{M})'_w$ is an algebra, and both the map $X\xi_0\to\alpha(X)\xi_0$ and $X\xi_0\to\alpha^{-1}(X)\xi_0$ are closable. Then α is represented as

$$\alpha(X)=U^\dagger XU,\quad X\in\mathscr{M},$$

for some $U\in\mathscr{L}^\dagger(\mathscr{D})_u$.

[1] H. J. Borchers, "Algebraic aspects of Wightman field theory," *Statistical Mechanics and Field Theory Lectures* (Haifa Summer School, New York, 1972), pp. 31–79.
[2] S. P. Gudder and W. Scruggs, Pac. J. Math. **70**, 369 (1977).
[3] S. P. Gudder and R. L. Hudson, Trans. Am. Math. Soc. **245**, 1 (1978).
[4] G. Lassner, Rep. Math. Phys. **3**, 279 (1972).
[5] R. T. Powers, Commun. Math. Phys. **21**, 85 (1971).
[6] R. T. Powers, Trans. Am. Math. Soc. **187**, 261 (1974).
[7] F. Mathot, J. Math. Phys. **26**, 1118 (1985).
[8] S. P. Gudder, Pac. J. Math. **80**, 141 (1979).
[9] A. Inoue, J. Operator Theory **10**, 77 (1983).
[10] J. Dixmier, *Von Neumann algèbras* (North-Holland, Amsterdam, 1981).
[11] S. Strătilă and L. Zsidó, *Lectures on von Neumann Algebras* (Abacus, Tunbridge Wells, 1979).
[12] M. Takesaki, "Tomita's theory of modular Hilbert algebras and its applications," *Lectures Notes in Mathematics*, Vol. 128 (Springer, Berlin, 1970).
[13] A. Inoue and K. Takesue, Trans. Am. Math. Soc. **280**, 393 (1983).
[14] A. Inoue and K. Takesue, Proc. Am. Math. Soc. **87**, 295 (1983).
[15] K. Takesue, Rep. Math. Phys. **21**, 347 (1985).
[16] A. Uhlmann, "Properties of the algebra $L^\dagger(D)$," JINR Comm. E 2-8149, Dubna, 1974.

On classical theory of moments: Finite-set-of-moments approach. I. Non-negative distribution: Its even moments and Hankel transform

Eugene S. Kryachko
Institute for Theoretical Physics, Kiev-130, 252130 USSR
Toshikatsu Koga
Muroran Institute of Technology, Muroran, Hokkaido, 050 Japan

(Received 16 May 1985; accepted for publication 10 September 1986)

For an unknown non-negative distribution $\Omega(z)$, the corresponding Hankel transform $F(k)$ is introduced. It is proposed to partition $F(k)$ in such a way that each component satisfies a linear differential relation whose solution gives an approximate Hankel transform in terms of a given finite set of even moments. As a result, for a known finite set of even moments, the non-negative distribution $\Omega(z)$ is obtained in the form of a finite sum of the definite differential and integral forms of the Gaussian distributions.

I. INTRODUCTION

One of the basic problems in the classical theory of moments[1–3] consists in constructing explicitly the unknown non-negative distribution or density $\omega(z)$ from its complete set of moments,

$$\langle z^n\rangle \equiv \int dz\, z^n\omega(z), \quad n=0,1,2,\ldots . \tag{1}$$

This problem is not purely mathematical, since the related problems arise also in theoretical physics and quantum chemistry. However, their peculiarity is related to the fact that only a few moments of the unknown non-negative distribution are given in contrast to the purely mathematical content of the classical problem of moments. Naturally, such empirical content of the classical problem of moments has a specificity, in particular, that the problem of constructing a non-negative distribution from a given finite set of its moments has no unique solution and, as emphasized by Corcoran and Langhoff,[4] is of continuing interest. The maximum entropy functional concept is one of the approaches to resolve such problems and is used widely in theoretical physics and quantum chemistry (see, for example, Ref. 5). This technique has been realized recently for momentum-space distributions of many-electron systems, in particular for the Compton profiles, and allows us to represent explicitly the momentum-space distributions in terms of their first few moments.[6,7]

The present series of papers gives an alternative solution of the finite-set-of-moments problem. In Paper I, we restrict ourselves to the particular case of this problem where a finite set of even moments $\langle z^{2m}\rangle$ is given. A general case of an incorporation of odd moments is studied in Paper II of this series. The structure of this paper is as follows. Some necessary definitions are given in the next section. The basic idea of the present approach is illustrated in Sec. III. A general solution of the problem for a given finite set of even moments is presented in Sec. IV. Some concluding remarks are made in Sec. V. Throughout this paper, the superscript $[n]$ denotes the nth order, while the superscript (n) does the nth derivative.

II. DEFINITIONS

In view of making bridges between the purely mathematical problem of the classical theory of moments and the related problems arising in theoretical physics and quantum chemistry, we may modify the definition of the moment $\langle z^n\rangle_\omega$ given by Eq. (1). For this purpose, we introduce a distribution $\Omega(z)$ and define its moment $\langle z^n\rangle_\Omega$ as follows:

$$\langle z^n\rangle \equiv 4\pi\int_0^\infty dz\, z^{n+2}\Omega(z), \quad n=-2,-1,0,1,\ldots . \tag{2}$$

For a given non-negative distribution $\Omega(z)$, let us define the function $F(k)$ as its Hankel transform,

$$F(k)\equiv 4\pi\int_0^\infty dz\, z^2\Omega(z)j_0(kz), \quad 0\leqslant k<\infty, \tag{3}$$

where $j_n(kz)$ is the spherical Bessel function. Inverting Eq. (3), one obtains

$$\Omega(z)=\frac{1}{2\pi^2}\int_0^\infty dk\, k^2F(k)\,j_0(kz), \tag{4}$$

that is, $F(k)$ and $\Omega(z)$ constitute a pair of Hankel transforms. From Definitions (2) and (3), it follows immediately that

$$F(0)=\langle z^0\rangle$$
and
$$F^{(2n)}(0)=(-1)^n\langle z^{2n}\rangle/(2n+1), \quad n\geqslant 1. \tag{5}$$

In the usual three-dimensional coordinate (r) space, $\Omega(r)$ plays the role of a spherically symmetric one-particle density or charge distribution $\rho(r)$ normalized to the number of particles N of a given many-particle system, i.e., $\langle r^0\rangle=N$. In momentum (p) space, $\Omega(p)$ can be a spherically symmetric momentum distribution $\rho(p)$ (see, for example, Ref. 7 as a review). In the particular case of closed-shell atoms and ions, $\rho(r)$ is one-electron density, and $F(k)$ is the so-called atomic scattering factor $f(k)$, which satisfies some well-known sum rules[8]

$$\langle r^{-1}\rangle=\frac{2}{\pi}\int_0^\infty dk\, f(k), \tag{6a}$$

$$\langle r^{-2}\rangle=\int_0^\infty dk\, kf(k). \tag{6b}$$

 0022-2488/87/010008-07$02.50

Equations (6a) and (6b) together with Eq. (5) in the form of $f(0) = \langle r^0 \rangle = N$ are the unique exact relations that connect the atomic scattering factor with the moments of the atomic charge distribution. There also exists an approximate relation of the same nature arising from the expansion of Eq. (3) at small values of k (Ref. 9):

$$f(k) = N - \langle r^2 \rangle k^2/3! + \langle r^4 \rangle k^4/5! + O(k^6). \qquad (7)$$

Equation (7) has been exploited recently to reconstruct $f(k)$ approximately,[10] but it is impossible to reproduce the corresponding $\rho(r)$ starting from Eq. (7) due to its divergence at large k.

III. BASIC IDEA: AN EXAMPLE OF FIRST FEW MOMENTS

To illustrate the basic idea of the present approach, let us differentiate Eq. (3) twice with respect to k. We obtain

$$F^{(1)}(k) = -4\pi \int_0^\infty dz\, z^3 \Omega(z) j_1(kz), \qquad (8)$$

$$F^{(2)}(k) = -\frac{4\pi}{3} \int_0^\infty dz\, z^4 \Omega(z) [j_0(kz) - 2j_2(kz)], \qquad (9)$$

since

$[j_n(x)]^{(1)} = (2n+1)^{-1}[nj_{n-1}(x) - (n+1)j_{n+1}(x)]$.

Introducing an operator function $U(k)$ in such a way that

$$U(k)F(k) \equiv 4\pi \int_0^\infty dz\, z^4 \Omega(z) j_0(kz), \qquad (10a)$$

or

$$U(k) \equiv \frac{\int_0^\infty dz\, z^4 \Omega(z) j_0(kz)}{\int_0^\infty dz\, z^2 \Omega(z) j_0(kz)}, \qquad (10b)$$

one can derive the following second-order differential relation for $F(k)$:

$$F^{(2)}(k) + (2/k)F^{(1)}(k) + U(k)F(k) = 0. \qquad (11)$$

As follows directly from Definition (10), $U(k)$ satisfies the constraint

$$U(0) = \langle z^2 \rangle / \langle z^0 \rangle = \alpha. \qquad (12)$$

Replacing $U(k)$ in Eq. (11) with its value at $k=0$, Eq. (12), one can interpret the resultant relation as a second-order linear differential equation for the zeroth-order approximate Hankel transform $F^{[0]}(k)$:

$$[F^{[0]}(k)]^{(2)} + (2/k)[F^{[0]}(k)]^{(1)} + \alpha F^{[0]}(k) = 0. \qquad (13)$$

Equation (13) is a special case of the differential equation for the spherical Bessel function, and its solution takes the following simple form:

$$F^{[0]}(k) = \langle z^0 \rangle j_0([\langle z^2 \rangle / \langle z^0 \rangle]^{1/2} k). \qquad (14)$$

When $F^{[0]}(k)$ is expanded for small k, the first two terms of $F^{[0]}(k)$ and $f(k)$ [Eq. (7)] coincide with one another.

Let us now evaluate $\Omega^{[0]}(z)$ corresponding to $F^{[0]}(k)$ via the rule given by Eq. (4). In contrast to the expansion (7), the Hankel transform of $F^{[0]}(k)$ does exist. The result is

$$\Omega^{[0]}(z) = \frac{\langle z^0 \rangle}{4\pi z} \left(\frac{\langle z^0 \rangle}{\langle z^2 \rangle} \right)^{1/2} \delta \left(z - \left[\frac{\langle z^2 \rangle}{\langle z^0 \rangle} \right]^{1/2} \right). \qquad (15)$$

Therefore, $\Omega^{[0]}(z)$ approximates the true non-negative distribution $\Omega(z)$ by means of a single delta-point "charge" (or "mass"). Evidently, this approximation is poor, and in order to improve it, let us return to Definition (10) of $U(k)$.

It is fairly easy to show that $U^{(1)}(0) = 0$ and

$$U^{(2)}(0) = \frac{1}{3} \left[\left(\frac{\langle z^2 \rangle}{\langle z^0 \rangle} \right)^2 - \frac{\langle z^4 \rangle}{\langle z^0 \rangle} \right] = -2\beta, \qquad (16)$$

if we remind the properties $j_0(0) = 1$ and $j_n(0) = 0$ $(n \geqslant 1)$ for Definition (10) and its differential forms. Then one can write the following second-order linear differential equation for the first-order approximation $F^{[1]}(k)$ to the Hankel transform $F(k)$:

$$[F^{[1]}(k)]^{(2)} + (2/k)[F^{[1]}(k)]^{(1)} + (\alpha - \beta k^2) F^{[1]}(k) = 0. \qquad (17)$$

The solution of Eq. (17) has the following explicit form [see Eq. (2.273.11) of Ref. 11]:

$$F^{[1]}(k) = k^{-3/2} \left[C_1 M_{\alpha/4\sqrt{\beta},1/4}(\sqrt{\beta} k^2) + C_2 M_{\alpha/4\sqrt{\beta},-1/4}(\sqrt{\beta} k^2) \right], \qquad (18)$$

where $M_{\kappa,\mu}(x)$ is the Whittaker function which is expressible in terms of the degenerate hypergeometric function as

$$M_{\kappa,\mu}(x) = x^{\mu+1/2} \exp(-x/2) {}_1F_1(\tfrac{1}{2} + \mu - \kappa, 2\mu + 1, x), \qquad (19)$$

and C_1 and C_2 are numerical constants. Since $F(k)$ is finite at $k = 0$, or in other words $\langle z^0 \rangle < \infty$ as follows from Eq. (5), we must put $C_2 = 0$. Taking Eq. (5) into account, one can obtain $C_1 = \langle z^0 \rangle \beta^{-3/8}$ and

$$F^{[1]}(k) = \langle z^0 \rangle \exp(-\gamma k^2) {}_1F_1(a,b,2\gamma k^2) = \langle z^0 \rangle \left[{}_1F_1 \left(a,b,-2\frac{d}{dx} \right) \exp(-\gamma x k^2) \right]_{x=1}, \qquad (20)$$

with $a = (3 - \alpha/\sqrt{\beta})/4$, $b = 3/2$, and $\gamma = \sqrt{\beta}/2$.

Expanding $F^{[1]}(k)$ in the Taylor series at $k = 0$ and taking into account the first three terms in this series expansion, one has

$$F^{[1]}(k) = \langle z^0 \rangle - \alpha \langle z^0 \rangle k^2/3! + (6\beta + \alpha^2) \langle z^0 \rangle k^4/5! + O(k^6), \qquad (21)$$

which coincides exactly up to the terms proportional to k^4 with the expansion (7) for the atomic scattering factor $f(k)$, if $\Omega(r) = \rho(r)$. It is clear that the form of $F^{[1]}(k)$ given by Eq. (20) is more applicable to problems in theoretical physics and quantum chemistry than the expansion (21) [or (7) for the atomic scattering factor], since in particular it is defined on the whole interval $0 \leqslant k < \infty$, converges as k approaches to an infinity, and its Hankel transform, the required distribution $\Omega^{[1]}(z)$, follows from a simple evaluation. In fact, bearing in mind the formula (11.4.28) of

Ref. 12, one obtains

$$\Omega^{[1]}(z) = \frac{2\langle z^0\rangle}{\pi^2}\beta^{-3/4}\sum_{n=0}^{\infty}\frac{(a)_n\Gamma(n+\frac{3}{2})}{2^{n+3/2}n!(b)_n} \times {}_1F_1(n+\tfrac{3}{2},\tfrac{3}{2},-z^2/4\gamma), \quad (22)$$

with $(a)_n \equiv \Gamma(a+n)/\Gamma(a)$. However, there exists an alternative way to express $\Omega^{[1]}(z)$ in a more convenient form. Taking Eq. (11.4.29) of Ref. 12 into account, we have

$$\int_0^\infty dk \exp(-\gamma k^2)k^{3/2}J_{1/2}(kz) = z^{1/2}(2\gamma)^{-3/2}\exp(-z^2/4\gamma), \quad (23)$$

$$\int_0^\infty dk \exp(-\gamma k^2)k^{2n+3/2}J_{1/2}(kz) = (-1)^n z^{1/2}2^{-3/2}\frac{d^n[\gamma^{-3/2}\exp(-z^2/4\gamma)]}{d\gamma^n}. \quad (24)$$

Therefore

$$\Omega^{[1]}(z) = \frac{\langle z^0\rangle}{8\pi^{3/2}}\sum_{n=0}^{\infty}\frac{(-1)^n(a)_n\beta^{n/2}}{n!(b)_n} \times \frac{d^n[\gamma^{-3/2}\exp(-z^2/4\gamma)]}{d\gamma^n}$$
$$= \frac{\langle z^0\rangle}{8\pi^{3/2}}\left[{}_1F_1\left(a,b,-2\frac{d}{dx}\right)(x\gamma)^{-3/2} \times \exp\left(\frac{-z^2}{4x\gamma}\right)\right]_{x=1}. \quad (25)$$

Hence the distribution $\Omega^{[1]}(z)$ takes a definite explicit form given by convergent degenerate hypergeometric series ${}_1F_1(a,b,-2d/dx)$ of the normal or Gaussian distribution. It should be emphasized here that in deriving $\Omega^{[1]}(z)$, we have assumed that α and β are positive to provide a nonnegativity of $\Omega^{[1]}(z)$. Notice that in the r space these conditions are automatically satisfied for all closed-shell atoms and ions whose moments, $\langle r^2\rangle$ and $\langle r^4\rangle$, are available. In particular, based on the known Z-dependence of $\langle r^2\rangle$ and $\langle r^4\rangle$ (see Ref. 10), one can demonstrate that in the limit $Z\to\infty$, α approaches $\frac{3}{4}$, and β is positive and behaves as $Z^{-0.95}$. Therefore $\rho^{[1]}(r) \equiv \Omega^{[1]}(r)$ is non-negative.

Obviously, the approximate distribution $\Omega^{[1]}(z)$ is correct in the sense that it gives the correct zeroth, second, and fourth moments as follows immediately from Eq. (21), i.e., $\langle z^0\rangle^{[1]} = \langle z^0\rangle$, $\langle z^2\rangle^{[1]} = \langle z^2\rangle$, and $\langle z^4\rangle^{[1]} = \langle z^4\rangle$, where

$$\langle z^{2n}\rangle^{[1]} \equiv 4\pi\int_0^\infty dz\, z^{2n+2}\Omega^{[1]}(z) = (-1)^n(2n+1)\left[\frac{d^{2n}F^{[1]}(k)}{dk^{2n}}\right]_{k=0}. \quad (26)$$

Moreover, $\Omega^{[1]}(z)$ provides the approximate expressions for higher moments, for example,

$$\langle z^{2m}\rangle^{[1]} = \langle z^0\rangle 2^m(2m+1)!!\sum_{n=0}^{\infty}\frac{(-1)^n(a)_n\beta^{n/2}}{n!(b)_n}\frac{d^n\gamma^m}{d\gamma^n}, \quad (27)$$

in particular

$$\langle z^6\rangle^{[1]} = \langle z^0\rangle\beta^{3/2}(105-212a+144a^2-64a^3), \quad (28a)$$

$$\langle z^8\rangle^{[1]} = \langle z^0\rangle\beta^2(945-1524a + 1952a^2-768a^3+256a^4), \quad (28b)$$

$$\langle z^{10}\rangle^{[1]} = \langle z^0\rangle\beta^{5/2}(10395-25236a+24480a^2 - 14720a^3+3840a^4-1024a^5). \quad (28c)$$

It is clear that these expressions, Eqs. (27) and (28a)–(28c), are useful in practice, for instance, to predict the unknown higher moments of the r- and p-space charge distributions $\rho(r)$ and $\rho(p)$.[13] In conclusion, it is worth noticing that $\Omega^{[1]}(z)$ and $F^{[1]}(k)$ constitute, by defintion, a pair of Hankel transforms similar to $\Omega(z)$ and $F(k)$. Hence, in r space the first-order approximate atomic scattering factor $f^{[1]}(k) \equiv F^{[1]}(k)$ satisfies the sum rules (6a) and (6b) where the moments $\langle r^{-1}\rangle$ and $\langle r^{-2}\rangle$ on the left-hand side of Eqs. (6a) and (6b) are replaced by $\langle r^{-1}\rangle^{[1]}$ and $\langle r^{-2}\rangle^{[1]}$ of the distribution $\rho^{[1]}(r) \equiv \Omega^{[1]}(r)$.

IV. GENERAL TREATMENT

In order to formulate analytically and generally the basic idea outlined in Sec. III, we first prove the following statement.

Proposition 1: Operator identities

$$Lj_0(kz) = -z^2 j_0(kz), \quad (29)$$

$$L^n j_0(kz) = (-1)^n z^{2n} j_0(kz), \quad n>1, \quad (30)$$

hold, where

$$L \equiv \frac{d^2}{dk^2} + \frac{2}{k}\frac{d}{dk} = \frac{1}{k^2}\frac{d}{dk}\left(k^2\frac{d}{dk}\right), \quad (31)$$

$$L^n = \left[\frac{d^2}{dk^2}+\frac{2}{k}\frac{d}{dk}\right]^n = \frac{d^{2n}}{dk^{2n}} + \frac{2n}{k}\frac{d^{2n-1}}{dk^{2n-1}}, \quad n>1, \quad (32)$$

Proof of Eq. (29) follows immediately from the definition of $j_0(kz)$. Repeating Eq. (29), one can obtain Eq. (30). The relation (32) can be proved via the method of mathematical induction.

Now expanding $j_0(kz)$ in the well-known series

$$j_0(kz) = 1 - \frac{(kz)^2}{3!} + \frac{(kz)^4}{5!} - + \cdots = \sum_{n=0}^{\infty} a_n, \quad (33)$$

with

$$a_n = (-1)^n(kz)^{2n}/(2n+1)!, \quad (34)$$

we have relations

$$L^n a_n = (-1)z^2L^{n-1}a_{n-1} = \cdots = (-1)^n z^{2n}a_0, \quad n\geqslant 1, \quad (35)$$

which can be rewritten in a formal way, by inverting, as follows:

$$a_n = (-1)^n z^{2n}L^{-n}(a_0). \quad (36)$$

Inserting expansion (33) into Eq. (3), one can formally partition $F(k)$ as

$$F(k) = \sum_{n=1}^{\infty} f^{[n]}(k), \quad (37)$$

where

$$f^{[1]}(k)\equiv 4\pi\int_0^\infty dz\, z^2\Omega(z)[a_0+a_1+a_2], \qquad (38a)$$

$$f^{[2]}(k)\equiv 4\pi\int_0^\infty dz\, z^2\Omega(z)[a_3+a_4+a_5], \qquad (38b)$$

etc., and generally

$$f^{[n]}(k)\equiv 4\pi\int_0^\infty dz\, z^2\Omega(z)[a_{3n-3}+a_{3n-2}+a_{3n-1}], \quad n>2. \qquad (38c)$$

As follows from Eqs. (5) and (34), the formal partitioning (37) possesses the following peculiarity: $f^{[n]}(k)$ contributes only to the moments $\langle z^{6n-6}\rangle$, $\langle z^{6n-4}\rangle$, and $\langle z^{6n-2}\rangle$, which are uniquely determined by

$$\langle z^{6n-2m}\rangle = (-1)^{3n-m}(6n-2m+1)\times\left[\frac{d^{6n-2m}f^{[n]}(k)}{dk^{6n-2m}}\right]_{k=0}, \quad m=3,2,1. \qquad (38d)$$

Let us consider in detail the leading term of the partitioning (37) of $F(k)$. As mentioned before, $f^{[1]}(k)$ was used in Refs. 9 and 10 to approximate the atomic scattering factor at small values of k, and taking into account Eq. (2), it can be represented as

$$f^{[1]}(k) = \langle z^0\rangle - \langle z^2\rangle k^2/3! + \langle z^4\rangle k^4/5!. \qquad (39)$$

Using Eq. (36), one may rewrite $f^{[1]}(k)$ as

$$f^{[1]}(k) = 4\pi\int_0^\infty dz\, z^2\Omega(z)[a_0 - z^2L^{-1}(a_0+a_1)]. \qquad (40)$$

Applying the operator L to both sides of Eq. (40), we have

$$Lf^{[1]}(k) = -4\pi\int_0^\infty dz\, z^4\Omega(z)(a_0+a_1) = -\langle z^2\rangle + \langle z^4\rangle k^2/3!. \qquad (41)$$

Based on Eq. (17), let us define

$$U^{[1]}(k)\equiv U_1(0)+\tfrac{1}{2}U_1^{(2)}(0)k^2 = \alpha^{[1]}-\beta^{[1]}k^2, \qquad (42)$$

then we can show that

$$U^{[1]}(k)f^{[1]}(k) = \langle z^2\rangle - \langle z^4\rangle k^2/3! + O(k^4), \qquad (43)$$

where we have used Eq. (39), and $\alpha^{[1]}\equiv\alpha$ and $\beta^{[1]}\equiv\beta$ are defined by Eqs. (12) and (16). Hence $f^{[1]}(k)$ satisfies the following second-order linear differential relation:

$$Lf^{[1]}(k) + U^{[1]}(k)f^{[1]}(k) = 0 + O(k^4), \qquad (44)$$

which is valid for small values of k with the condition $f^{[1]}(0) = \langle z^0\rangle$.

Similarly, for $f^{[n]}(k)$ ($n>1$) one can obtain the following relations:

$$\begin{aligned} f^{[n]}(k) &= (-1)^{3n-3}\left[\langle z^{6n-6}\rangle\frac{k^{6n-6}}{(6n-5)!} - \langle z^{6n-4}\rangle\frac{k^{6n-4}}{(6n-3)!} + \langle z^{6n-2}\rangle\frac{k^{6n-2}}{(6n-1)!}\right] \\ &= 4\pi(-1)^{3n-3}\int_0^\infty dz\, z^{6n-4}\Omega(z)L^{-(3n-3)} \\ &\quad\times[a_0 - z^2L^{-1}(a_0+a_1)], \end{aligned} \qquad (45)$$

and

$$Lg^{[n]}(k) = -4\pi\int_0^\infty dz\, z^{6n-2}\Omega(z)(a_0+a_1) = -\langle z^{6n-4}\rangle + \langle z^{6n-2}\rangle k^2/3!, \qquad (46)$$

where the function

$$\begin{aligned} g^{[n]}(k) &= 4\pi\int_0^\infty dz\, z^{6n-4}\Omega(z)[a_0 - z^2L^{-1}(a_0+a_1)] \\ &= (-1)^{3n-3}L^{3n-3}f^{[n]}(k) \\ &= \langle z^{6n-6}\rangle - \langle z^{6n-4}\rangle k^2/3! + \langle z^{6n-2}\rangle k^4/5! \end{aligned} \qquad (47)$$

is introduced. It is clear that $g^{[1]}(k) = f^{[1]}(k)$.

Now generalizing the function $U^{[1]}(k)$ to an arbitrary n (>1),

$$U^{[n]}(k) = U_n(0) + \tfrac{1}{2}U_n^{(2)}(0)k^2 = \alpha^{[n]} - \beta^{[n]}k^2, \qquad (48)$$

where

$$\alpha^{[n]} = \langle z^{6n-4}\rangle/\langle z^{6n-6}\rangle, \qquad (49a)$$

$$\beta^{[n]} = -\tfrac{1}{6}[(\langle z^{6n-4}\rangle/\langle z^{6n-6}\rangle)^2 - (\langle z^{6n-2}\rangle/\langle z^{6n-6}\rangle)], \qquad (49b)$$

one can prove fairly easily that $g^{[n]}(k)$ obeys the following second-order linear differential relation:

$$Lg^{[n]}(k) + U^{[n]}(k)g^{[n]}(k) = 0 + O(k^4), \qquad (50)$$

which is valid for small k under the constraint $g^{[n]}(0) = \langle z^{6n-6}\rangle$. Then

$$f^{[n]}(k) = (-1)^{3n-3}L^{-(3n-3)}g^{[n]}(k) \qquad (51)$$

with the conditions

$$[f^{[n]}(k)]_{k=0}^{(m)} = 0, \quad \text{for } 0\leqslant m\leqslant 6n-7, \quad n>1. \qquad (52)$$

Definition: The function $F^{[1]}(k)$ is defined as a solution of the second-order linear differential equation

$$LF^{[1]}(k) + U^{[1]}(k)F^{[1]}(k) = 0, \qquad (53)$$

where $U^{[1]}(k)$ is given by Eq. (42) with positive $\alpha^{[1]}$ and $\beta^{[1]}$.

Comparing Eq. (53) with the relation (44), one can conclude that $F^{[1]}(k)$ coincides exactly with $f^{[1]}(k)$ up to the term proportional to k^6 inclusively, since Eq. (53) coincides with Eq. (44) up to the term $\sim k^4$ as second-order differential relations [see also Eq. (21)]. The explicit form of $F^{[1]}(k)$ was obtained in Sec. III. One can define the first-order distribution $\Omega^{[1]}(z)$ as the Hankel transform of $F^{[1]}(k)$, viz.,

$$\Omega^{[1]}(z)\equiv\frac{1}{2\pi^2}\int_0^\infty dk\, k^2F^{[1]}(k)j_0(kz), \qquad (54)$$

which is non-negative due to the positiveness of $\alpha^{[1]}$ and $\beta^{[1]}$, and whose moments

$$\langle z^n\rangle^{[1]} \equiv 4\pi \int_0^\infty dz\, z^{n+2}\Omega^{[1]}(z) \tag{55}$$

with $n \geqslant 0$ are all nonvanishing in general [compare with Eqs. (27) and (28)]. The non-negativity of all the moments $\langle z^n\rangle^{[1]}$ for $n \geqslant 0$ follows directly from the expansion of $F^{[1]}(k)$, given by Eq. (20), similar to the formal expansion (37).

The first-order approximate distribution $\Omega^{[1]}$ generally predicts incorrect higher even moments $\langle z^{2m}\rangle^{[1]}$ with $m \geqslant 3$. The deviation of higher even moments from their true moments leads to the introduction of higher-order corrections $F^{[n]}(k)$'s and $\Omega^{[n]}(z)$'s, which are so defined as to correct the deviations in the even moments successively. Applying the similar arguments to $F^{[n]}(k)$'s and $\Omega^{[n]}(z)$'s, we can formulate the following statements.

Proposition 2: For a given finite set of even moments $\{\langle z^{2n}\rangle\}_{n=0}^{3N-1}$ the required distribution $\Omega(z)$, whose moments $\{\langle z^{2n}\rangle_\Omega\}_{n=0}^{3N-1}$ are just the given moments, and the corresponding Hankel transform $F(k)$ are determined by the following partitions:

$$\Omega(z) \equiv \sum_{n=1}^{N} \Omega^{[n]}(z), \tag{56}$$

$$F(k) \equiv \sum_{n=1}^{N} F^{[n]}(k). \tag{57}$$

The components or "pieces" of the expansions (56) and (57) of the same order, say n, $\Omega^{[n]}(z)$ and $F^{[n]}(k)$, constitute a pair of the Hankel transform, i.e.,

$$F^{[n]}(k) = 4\pi \int_0^\infty dz\, z^2 \Omega^{[n]}(z)\, j_0(kz), \tag{58a}$$

$$\Omega^{[n]}(z) = \frac{1}{2\pi^2} \int_0^\infty dk\, k^2 F^{[n]}(k)\, j_0(kz). \tag{58b}$$

The function $F^{[n]}(k)$ satisfies

$$F^{[n]}(k) = (-1)^{3n-3} L^{-(3n-3)} G^{[n]}(k) \tag{59}$$

with the conditions

$$[F^{[n]}(k)]_{k=0}^{(m)} = 0, \quad \text{for } 0 \leqslant m \leqslant 6n-7, \quad n > 1. \tag{60}$$

$$F^{[1]}(k) = G^{[1]}(k). \tag{61}$$

The function $G^{[n]}(k)$ obeys the second-order linear differential equation

$$LG^{[n]}(k) + U^{[n]}(k) G^{[n]}(k) = 0 \tag{62}$$

with

$$U^{[n]}(k) = \alpha^{[n]} - \beta^{[n]} k^2, \tag{63}$$

where

$$\alpha^{[n]} = \frac{\langle z^{6n-4}\rangle - {}^{[n-1]}\langle z^{6n-4}\rangle}{\langle z^{6n-6}\rangle - {}^{[n-1]}\langle z^{6n-6}\rangle}, \tag{64}$$

$$\beta^{[n]} = -\frac{1}{6}\left\{\left[\frac{\langle z^{6n-4}\rangle - {}^{[n-1]}\langle z^{6n-4}\rangle}{\langle z^{6n-6}\rangle - {}^{[n-1]}\langle z^{6n-6}\rangle}\right]^2 - \frac{\langle z^{6n-2}\rangle - {}^{[n-1]}\langle z^{6n-2}\rangle}{\langle z^{6n-6}\rangle - {}^{[n-1]}\langle z^{6n-6}\rangle}\right\}, \tag{65}$$

$${}^{[n]}\langle z^m\rangle \equiv 4\pi \int_0^\infty dz\, z^{m+2}\, {}^{[n]}\Omega(z) = \sum_{k=1}^{n} \langle z^m\rangle^{[k]}, \tag{66}$$

$${}^{[n]}\Omega(z) \equiv \sum_{k=1}^{n} \Omega^{[k]}(z), \quad \langle z^m\rangle^{[k]} \equiv 4\pi \int_0^\infty dz\, z^{m+2}\Omega^{[k]}(z). \tag{67}$$

Proposition 3: The function $G^{[n]}(k)$ as a solution of Eq. (62) under the conditions (63)–(67) takes the following explicit form:

$$\begin{aligned} G^{[n]}(k) &= (\langle z^{6n-6}\rangle - {}^{[n-1]}\langle z^{6n-6}\rangle) \\ &\quad \times \exp(-\gamma^{[n]}k^2)\, {}_1F_1(a^{[n]}, b, 2\gamma^{[n]}k^2) \\ &= (\langle z^{6n-6}\rangle - {}^{[n-1]}\langle z^{6n-6}\rangle) \\ &\quad \times \left[{}_1F_1\left(a^{[n]}, b, -2\frac{d}{dx}\right)\exp(-\gamma^{[n]}xk^2)\right]_{x=1}, \end{aligned} \tag{68}$$

where $b = \frac{3}{2}$ and

$$a^{[n]} = (3 - \alpha^{[n]}/\sqrt{\beta^{[n]}})/4, \quad \gamma^{[n]} = \sqrt{\beta^{[n]}}/2. \tag{69}$$

For a known $G^{[n]}(k)$ we now derive the expressions for $F^{[n]}(k)$ and $\Omega^{[n]}(z)$. As follows from Eq. (59),

$$L^{3n-3}F^{[n]}(k) = (-1)^{3n-3}G^{[n]}(k), \quad n > 1. \tag{70}$$

Using the identity (32), one can rewrite Eq. (70) as

$$\frac{d^{6n-6}F^{[n]}(k)}{dk^{6n-6}} + \frac{2(3n-3)}{k}\frac{d^{6n-7}F^{[n]}(k)}{dk^{6n-7}} = (-1)^{3n-3}G^{[n]}(k), \tag{71}$$

or introducing

$$H^{[n]}(k) \equiv \frac{d^{6n-7}F^{[n]}(k)}{dk^{6n-7}}, \tag{72}$$

in the following form:

$$\frac{dH^{[n]}(k)}{dk} + \frac{6n-6}{k}H^{[n]}(k) = (-1)^{3n-3}G^{[n]}(k), \tag{73}$$

with the subsidiary condition $H^{[n]}(0) = 0$ [see Eq. (60)]. Then, taking Eq. (68) into account, one obtains $H^{[n]}(k)$ as

$$\begin{aligned} H^{[n]}(k) &= (-1)^{3n-3}(\langle z^{6n-6}\rangle - {}^{[n-1]}\langle z^{6n-6}\rangle)k^{-(6n-6)}\left[{}_1F_1\left(a^{[n]}, b, -2\frac{d}{dx}\right)\int_0^k dt\, t^{6n-6}\exp(-\gamma^{[n]}xt^2)\right]_{x=1} \\ &= (-1)^{3n-3}\frac{\langle z^{6n-6}\rangle - {}^{[n-1]}\langle z^{6n-6}\rangle}{(6n-5)} \\ &\quad \times \left\{\left[{}_1F_1\left(a^{[n]}, b, -2\frac{d}{dx}\right){}_1F_1\left(1, 3n-\frac{3}{2}, -\frac{d}{dy}\right)k\exp(-\gamma^{[n]}xyk^2)\right]_{y=1}\right\}_{x=1}, \end{aligned} \tag{74}$$

where Eq. (6.5.12) of Ref. 12 has been applied.

Using the idea of fractional-order integral transform (see for instance, Ref. 14 and Chap. 13 of Ref. 15), one can obtain

$$F^{[n]}(k) = (-1)^{3n-3}\frac{\langle z^{6n-6}\rangle - {}^{[n-1]}\langle z^{6n-6}\rangle}{(6n-5)}\Big\{{}_1F_1\Big(a^{[n]},b,-2\frac{d}{dx}\Big)\Big[{}_1F_1\Big(1,3n-\frac{3}{2},-\frac{d}{dy}\Big)$$
$$\times[(6n-8)!]^{-1}\int_0^k dt\,(t-k)^{6n-8}t\exp(-\gamma^{[n]}xyt^2)\Big]_{y=1}\Big\}_{x=1}, \tag{75}$$

where the subsidiary conditions (60) for $F^{[n]}(k)$ are taken into account. Applying Eq. (13.1.15) of Ref. 15, one can evaluate the integral in Eq. (75):

$$[(6n-8)!]^{-1}\int_0^k dt\,(t-k)^{6n-8}t\exp(-\gamma^{[n]}xyt^2) = [(6n-5)!]^{-1}k^{6n-5}{}_2F_2(1,\tfrac{3}{2};3n-\tfrac{5}{2},3n-2;-\gamma^{[n]}xyk^2). \tag{76}$$

Therefore we finally have

$$F^{[n]}(k) = (-1)^{3n-3}(\langle z^{6n-6}\rangle - {}^{[n-1]}\langle z^{6n-6}\rangle)[(6n-5)(6n-6)!]^{-1}k^{6n-6}$$
$$\times\Big\{{}_1F_1\Big(a^{[n]},b,-2\frac{d}{dx}\Big)\Big[{}_1F_1\Big(1,3n-\frac{3}{2},-\frac{d}{dy}\Big){}_2F_2\Big(1,\frac{3}{2};3n-\frac{5}{2},3n-2;-\gamma^{[n]}xyk^2\Big)\Big]_{y=1}\Big\}_{x=1}. \tag{77}$$

Notice that we can express $F^{[n]}(k)$ in terms of the G-function by Meijer instead of ${}_2F_2$ via the application of Eq. (5.6.1) of Ref. 16.

Taking Eq. (58b) into account, we now obtain an explicit form of $\Omega^{[n]}(z)$. First, let us substitute Eq. (77) into Eq. (58b):

$$\Omega^{[n]}(z) = (-1)^{3n-3}(\langle z^{6n-6}\rangle - {}^{[n-1]}\langle z^{6n-6}\rangle)[2\pi^2(6n-5)(6n-6)!]^{-1}$$
$$\times\Big\{{}_1F_1\Big(a^{[n]},b,-2\frac{d}{dx}\Big)\Big[{}_1F_1\Big(1,3n-\frac{3}{2},-\frac{d}{dy}\Big)$$
$$\times\int_0^\infty dk\,k^{6n-4}{}_2F_2\Big(1,\frac{3}{2};3n-\frac{5}{2},3n-2;-\gamma^{[n]}xyk^2\Big)j_0(kz)\Big]_{y=1}\Big\}_{x=1}. \tag{78}$$

Similarly to Eq. (31), one can introduce in z space the operator L_z:

$$L_z \equiv \frac{d^2}{dz^2} + \frac{2}{z}\frac{d}{dz}, \tag{79}$$

and obtain the relation

$$L_z^{3n-2}j_0(kz) = (-1)^{3n-2}k^{6n-4}j_0(kz). \tag{80}$$

Then, taking Eq. (80) into account, we rewrite Eq. (78) as

$$\Omega^{[n]}(z) = -L_z^{3n-2}(\langle z^{6n-6}\rangle - {}^{[n-1]}\langle z^{6n-6}\rangle)[2\pi^2(6n-5)(6n-6)!]^{-1}$$
$$\times\Big(\frac{\pi}{2z}\Big)^{1/2}\Big\{{}_1F_1\Big(a^{[n]},b,-2\frac{d}{dx}\Big)\Big[{}_1F_1\Big(1,3n-\frac{3}{2},-\frac{d}{dy}\Big)$$
$$\times\int_0^\infty dk\,k^{-1/2}J_{1/2}(kz){}_2F_2\Big(1,\frac{3}{2};3n-\frac{5}{2},3n-2;-\gamma^{[n]}xyk^2\Big)\Big]_{y=1}\Big\}_{x=1}$$
$$= -L_z^{3n-2}(\langle z^{6n-6}\rangle - {}^{[n-1]}\langle z^{6n-6}\rangle)[2\pi^2(6n-5)(6n-6)!]^{-1}\Gamma(3n-\tfrac{5}{2})\,\Gamma(3n-2)z^{-1}$$
$$\times\Big[{}_1F_1\Big(a^{[n]},b,-2\frac{d}{dx}\Big){}_1F_1\Big(1,3n-\frac{3}{2},-\frac{d}{dy}\Big)G^{31}_{34}\Big(\frac{z^2}{4\gamma^{[n]}xy}\Big|\begin{matrix}1,3n-\frac{5}{2},3n-2\\ \frac{1}{2},1,\frac{3}{2},0\end{matrix}\Big)\Big]_{x=y=1}, \tag{81}$$

where Eq. (7.542.5) of Ref. 17 has been used.

According to Eq. (5) [see also Eq. (38d)] and due to the presence of the term k^{6n-6} in the expression for $F^{[n]}(k)$ [and similarly to L_z^{3n-2} in $\Omega^{[n]}(z)$], the nth piece of the Hankel transform $F^{[n]}(k)$ [or the nth piece of the distribution $\Omega^{[n]}(z)$, as proved via integration by parts] does not contribute to the moments $\langle z^m\rangle$ with $-2 \leqslant m \leqslant 6n-4$, i.e.,

$$\langle z^{2m}\rangle^{[n]} = 0, \quad -1 \leqslant m \leqslant 3n-2, \quad n > 1. \tag{82}$$

The first nonvanishing contribution of $F^{[n]}(k)$ is $\langle z^{6n-6}\rangle^{[n]}$:

$$\langle z^{6n-6}\rangle^{[n]} \equiv (-1)^{3n-3}(6n-5)\Big[\frac{d^{6n-6}F^{[n]}(k)}{dk^{6n-6}}\Big]_{k=0}$$
$$= \langle z^{6n-6}\rangle - {}^{[n-1]}\langle z^{6n-6}\rangle. \tag{83}$$

From Eqs. (69), (64), and (65), one obtains similarly that

$$\langle z^{6n-4}\rangle^{[n]} = \langle z^{6n-4}\rangle - {}^{[n-1]}\langle z^{6n-4}\rangle,$$
$$\langle z^{6n-2}\rangle^{[n]} = \langle z^{6n-2}\rangle - {}^{[n-1]}\langle z^{6n-2}\rangle, \tag{84}$$

and for $m \geqslant 0$

$$\langle z^{6n+2m} \rangle^{[n]}$$

$$\equiv (-1)^{3n+m}(6n+2m+1)\left[\frac{d^{\,6n+2m}F^{[n]}(k)}{dk^{\,6n+2m}}\right]_{k=0}$$

$$= \left[(6n+2m+1)(m+3)!\left(\frac{3}{2}\right)_{m+3}(\gamma^{[n]})^{m+3}\right]$$

$$\times \frac{\langle z^{6n-6}\rangle - {}^{[n-1]}\langle z^{6n-6}\rangle}{(6n-5)(3n-\frac{5}{2})_{m+3}(3n-2)_{m+3}}$$

$$\times \left[{}_1F_1\left(a^{[n]},b,-2\frac{d}{dx}\right){}_1F_1\left(1,3n-\frac{3}{2},-\frac{d}{dy}\right)\right.$$

$$\left.\times (xy)^{m+3}\right]_{x=y=1}. \tag{85}$$

Note that it has been assumed implicitly that for all $n \geqslant 1$ the quantities $\alpha^{[n]}$ and $\beta^{[n]}$, given by Eqs. (64) and (65), are positive for $\Omega^{[n]}(z)$ and the total distribution $\Omega(z)$, defined by Eq. (56), to be non-negative. However, one can prove (see Paper II) that this constraint is not so strong. Also in Paper II, it will be shown that the assumption made in Proposition 2 that the number of known moments is just $3N$ ($N \geqslant 1$) is not essential and can be removed.

V. SUMMARY

The present approach permits us to represent an unknown distribution in terms of a finite set of its known moments as a finite sum of the definite differential and integral forms of the Gaussian distribution [see Eq. (56)]. Naturally, one can consider such a representation as one of many ways (though only a few ways are known) in which a function chosen is expressed in terms of a sum of Gaussians and their differential forms. Nevertheless, it should be emphasized that the present representation is based on a given finite set of even moments. Restricting ourselves to the case where only even moments are known, we may assume that in fact $\Omega(z)$ is also a function of z^2, but its explicit representation in terms of Gaussians in z space is equivalent to its discrete Laplace representation in z^2 space.[18]

It is clear that such a representation of the unknown distribution is rather simple in a computational sense, and one can suggest that it is applicable and workable in processing experimental data on diverse symmetric spectroscopic contours, symmetric distributions of various signals, including radio signals and noises in technical problems, and also on charge distributions of molecules in r and p spaces, and in the related physical problems where only a finite number of moments is available from the experimental and computational data. The present approach may be applied to analytical approximation of one-electron densities in r and p spaces and scattering factors of atoms and to the prediction of unknown higher even moments via Eq. (85), which is of purely quantum-chemical interest and will be published elsewhere.[13] Notice that within the framework of the nth order approximation, the sum rules for the atomic scattering factor ${}^{[n]}f(k)$ as the Hankel transform of the distribution ${}^{[n]}\Omega(z)$ in r space, Eqs. (6a) and (6b), are satisfied exactly. Finally, it is worth noticing that the approach developed here can be generalized in a simple way to an arbitrary reference even moment $\langle z^{2m_0}\rangle$ instead of $\langle z^0\rangle$ as chosen in Sec. II.

ACKNOWLEDGMENTS

The authors acknowledge the referee for his valuable comments.

Part of this study has been supported by a Grant-in-Aid for Scientific Research from the Ministry of Education of Japan.

1. J. A. Shohat and J. D. Tamarkin, *The Problem of Moments* (Am. Math. Soc., New York, 1943).
2. N. I. Akhiezer, *The Classical Moment Problem* (Oliver and Boyd, London, 1965).
3. M. G. Krein and A. A. Nudel'man, *The Problem of Markov Moments and the Extremal Problems* (Nauka, Moscow, 1973).
4. C. T. Corcoran and P. W. Langhoff, J. Math. Phys. **18**, 651 (1977).
5. L. R. Mead and N. Papanicolaou, J. Math. Phys. **25**, 2404 (1984).
6. S. R. Gadre and S. B. Sears, J. Chem. Phys. **71**, 4321 (1976); T. Koga and M. Morita, J. Chem. Phys. **79**, 1933 (1983); T. Koga, Theor. Chim. Acta, **64**, 249 (1984).
7. E. S. Kryachko and T. Koga, Adv. Quantum Chem. **17**, 97 (1985).
8. J. N. Silverman and Y. Obata, J. Chem. Phys. **38**, 1254 (1963); R. Benesch and V. H. Smith, Jr., in *Wave Mechanics, The First Fifty Years*, edited by W. C. Price, S. C. Chissick, and T. Ravensdale (Buttersworths, London, 1973), p. 357; J. Oddershede and J. R. Sabin, Chem. Phys. **71**, 161 (1982); A. J. Thakkar and T. Koga, J. Chem. Phys. **83**, 747 (1985).
9. A. J. Thakkar and V. H. Smith, Jr., J. Phys. B **11**, 3803 (1978).
10. R. Pucci and N. H. March, J. Chem. Phys. **76**, 4089 (1982); N. H. March and R. Pucci, in *Local Density Approximations in Quantum Chemistry and Solid State Physics*, edited by J. P. Dahl and J. Avery (Plenum, New York, 1984), p. 53.
11. E. Kamke, *Differentialgleichungen. Lösungsmethoden and Lösungen* (Akademische Verlagsgesellschaft Geest and Portig, Leipzig, 1959).
12. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).
13. E. S. Kryachko and T. Koga, to be published.
14. I. I. Hirschmann and D. V. Widder, *The Convolution Transform* (Princeton U. P., Princeton, 1955).
15. M. Bateman and A. Erdelyi, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. II.
16. M. Bateman and A. Erdelyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953).
17. I. S. Gradsteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1980).
18. P. Henrici, *Applied and Computational Complex Analysis* (Wiley, New York, 1977), Vol. II, Sec. 10.10.

Approximate representations of SU(2) ordered exponentials in the adiabatic and stochastic limits

M. E. Brachet
Ecole Normale Supérieure, Paris, France

H. M. Fried[a)]
Physique Théorique, Université de Nice, 06034 Nice Cedex, France

(Received 24 June 1986, accepted for publication 10 September 1986)

Approximate representations for the SU(2) ordered exponential $U(t\,|E) = (\exp[i\int_0^t dt'\,\boldsymbol{\sigma}\cdot\mathbf{E}(t')])_+$, written as a functional of its input field $E(t)$, are derived in the adiabatic ($\rho \ll 1$) and stochastic ($\rho \gg 1$) limits, where $\rho \equiv |d\hat{E}/dt\,|/E$, $\hat{E} = \mathbf{E}/E$, $E = +(\mathbf{E}^2)^{1/2}$. An algorithm is set up for the adiabatic case, and fixed-point equations are obtained for situations of possible convergence. In the stochastic regime, "averaged" functions describing $U(t\,|E)$ are derived which reproduce its slowly varying dependence of large magnitude while missing, or approximating, rapid oscillations of small magnitude. Several functional integrals, analytic and machine are carried out over these approximate forms, and their results compared with the same functional integrals over the exact $U(t\,|E)$.

I. INTRODUCTION

Ordered exponentials are found in every branch of mathematical physics that deals with the causal time development of systems of more than one degree of freedom. Analytic treatments have typically been restricted to perturbative expansions, although computer calculations are now quite capable of dealing with any specific strong-coupling (SC) situation. However, when the variables in question are operators—numerical functions appearing in ordered exponentials and subsequently subjected to fluctuations as specified by an appropriate functional integral—the situation is much less clear. What would be most useful for such situations is a semianalytic approximation to the ordered exponential, which could then be inserted under the desired functional integral and its evaluation performed by some relevant approximation such as stationary phase. Functional integration aside, there are many instances when one would like to know the qualitative form of an ordered exponential as a functional of its input, without having to resort to a detailed numerical integration for each choice of input.

The purpose of this paper is to discuss and derive results, some of which have been previously quoted elsewhere,[1] for two classes of SC approximation to the ordered exponential solution of the differential equation

$$\frac{\partial U}{\partial t} = i\boldsymbol{\sigma}\cdot\mathbf{E}(t)\,U(t), \quad U(0) = 1, \tag{1.1}$$

where the σ_l denote 2×2 Pauli matrices, and the $E_l(t)$ are real, input functions. The unitary solution to (1.1) is

$$U(t) = \left(\exp\left[\int_0^t dt'\,\boldsymbol{\sigma}\cdot\mathbf{E}(t')\right]\right)_+$$
$$\equiv \sum_{n=0}^{\infty} \frac{i^n}{n!} \int_0^t dt_1 \cdots \int_0^t dt_n (\boldsymbol{\sigma}\cdot\mathbf{E}(t_1)\cdots\boldsymbol{\sigma}\cdot\mathbf{E}(t_n))_+, \tag{1.2}$$

where the symbol $(\ )_+$ denotes an ordering of the t_i-dependent factors, with those containing later times standing to the left.

Perhaps the most interesting applications are associated with the generalization to SU(N), obtained by replacing the σ_l of (1.1) by the $N\times N$ Hermitian matrices λ_l which form the defining representation of SU(N). In principle, the analysis of this paper could be extended from SU(2) to SU(N); however, the specific details appear quite complicated, and have not yet been carried through. Some work on the SU(2) SC adiabatic limit has already appeared in rather special contexts,[2,4] which is here generalized in a nontrivial way; to the best of the author's knowledge, the material presented for the SC stochastic limit is new. Generalizations of the adiabatic limit to SU(N) are not difficult, and have been used in quite different contexts, for Navier–Stokes fluid flow,[5] $N = 3$, and in one approach to QCD,[6] for arbitrary N.

The SC situation may be defined by the requirement $\int_0^t dt'\,E(t') > 1$, $E = +\sqrt{\mathbf{E}^2}$, in contrast to the weak-coupling, or perturbative regime for which one assumes the converse, $\int_0^t dt'E(t') < 1$; in the latter case it is simple to derive a valid representation for $\ln U$ in terms of an expansion in multiple integrals over ascending powers of $\mathbf{E}(t')$. For the SC case, two distinct limiting regions can be defined, one for which $|d\hat{E}/dt\,|$ is "small" (the adiabatic, or quasistatic limit), and the opposite ("stochastic") situation for which it is "large." Clearly, if $\hat{E}(t) \equiv \mathbf{E}(t)/E(t)$ did not depend on time, and were fixed in one direction, a choice of coordinate axes could be made so that only one of the σ_l need appear, and the ordered exponential would become an ordinary exponential involving that σ_l. When $\hat{E}(t)$ varies with time, however, the problem becomes nontrivial, and naturally divides into these two quite different limits. By "large" or "small" one must mean the magnitude of $|d\hat{E}/dt\,|$ with respect to the only other relevant quantity of like dimension, $E(t)$; and hence if one defines $\rho(t) \equiv |d\hat{E}/dt\,|/E$, the SC adiabatic and stochastic limits are defined by $\rho \ll 1$ and $\rho \gg 1$,

[a)] Permanent Address: Physics Department, Brown University, Providence, Rhode Island 02912.

 0022-2488/87/010015-13$02.50

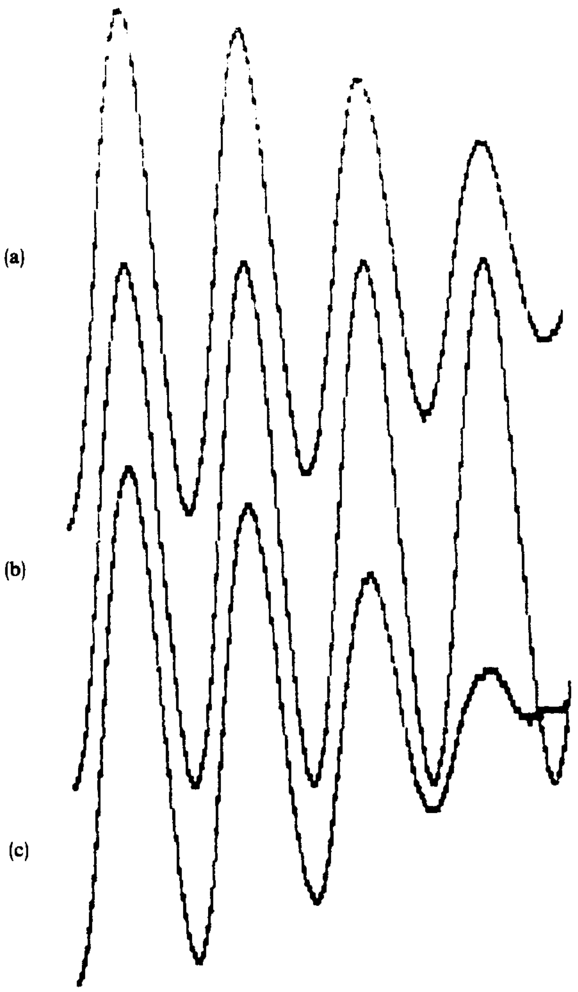

FIG. 1. Curves of (a) F_0, (b) $\cos Q_0$, and (c) $\cos Q_0 \cdot \cos q_1$; for the situation $\rho = 0.1$. (N.B. In all 18 figures, all curves plot the negative of every function indicated. Time increases from left to right.)

respectively. The word "stochastic" is appropriate because such behavior of ρ is expected in situations where a subsequent functional integration is performed with a "white-noise" Gaussian weighting; this will be fully discussed in Sec. IV.

To simplify the initial analyses as much as possible, and because it is always possible to reduce the problem to an equation of form (1.1) with a two-component $\mathbf{E}(t)$ vector, we first discuss both adiabatic and stochastic limits treating $\mathbf{E}$ as a vector in the (x, y) plane. The results of these investigations may be briefly summarized as follows, and will be described using the form of U which is convenient for numerical integration, $U(t) = F_0(t) + i\sigma\cdot\mathbf{F}(t)$.

A. The adiabatic limit

There exists here a sequence of corrections which can be written for (F_0, F_i), and which should approach the exact (that is, numerically integrated) solutions rapidly, if the first two approximations are at all representative. One can write an algorithm that can be used to generate successive approximations; and if (which we do not prove) convergence exists, then the solutions are given in terms of four simultaneous fixed-point equations. For brevity, we here only mention the form of solutions to $F_0 = \frac{1}{2}\mathrm{Tr}[U]$, with representations for $\mathbf{F}$ and all details reserved for the subsequent text.

In the adiabatic limit, $\rho = 0$, one finds

$$F_0(t) = \cos Q_0(t), \quad Q_0(t) = \int_0^t dt'\, E(t'). \tag{1.3}$$

For a constant magnitude E, F_0 varies as a simple cosine, with frequency $E/2\pi$. As ρ is increased from zero, but $\rho \ll 1$, this essential form remains but is modulated by a smaller competing frequency; e.g., if we suppose that $\omega = |d\hat{E}/dt|$ is also independent of time, and is chosen such that $\rho = \omega/E \simeq 0.1$, then the modulation will cause F_0 to shrink to zero after five cycles or so, then increase again, and repeat the same pattern. Variations of $E(t)$ and/or $\omega(t)$ will change the details but not the overall behavior, as long as $\rho \ll 1$.

Use of the algorithm discussed in the text leads to the first correction to (1.3) given by

$$F_0(t) = [\cos Q_0(t)]\left[\cos\left(\omega \int_0^t dt' |\sin Q_0(t')|\right)\right], \tag{1.4}$$

again assuming, for simplicity, that ω and E are constant. As pictured in Fig. 1, the numerically calculated F_0 is compared with the approximations of (1.3) and (1.4); and one can, in fact, see that (1.4) provides a bit too much modulation. If the procedure converges, the next approximation should modify that discrepancy, etc. We have not attempted further numerical work, and do not yet know whether the fixed-point equations written in Sec. II have solutions for certain $E(t)$.

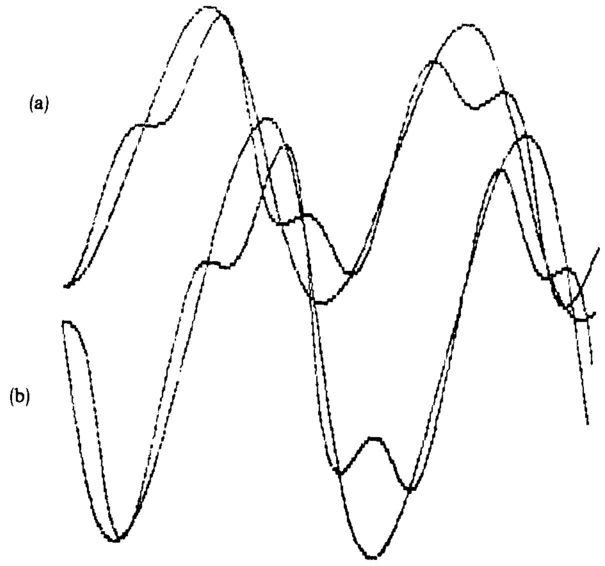

FIG. 2. Superpositions of (a) F_0 and $\overline{F}_0$, and of (b) F_3 and $\overline{F}_3$; for $\rho = 1$.

B. The stochastic limit

As ρ increases, the forms of the exact solutions change dramatically. For $\rho \sim 1$, with constant ω, E, the exact F_0 is displayed in Fig. 2, and bears no resemblance to its form in the adiabatic limit. As ρ is increased further, for $\rho \gg 1$ there is great simplification with F_0 taking the form of small, rapid ω-oscillations superimposed upon a cosine of larger magnitude and much slower frequency $\sim E^2/2\pi\omega$. When E and ω are themselves time-independent, the slowly varying behavior of F_0 can become considerably more complicated than a simple cosine.

Whatever physical properties are being described by these equations, it is surely the larger, slower oscillations—the "average" functional dependence—which should contain physically significant information, and not the smaller, faster oscillations riding on the "averaged" behavior. It is then a matter of some interest to be able to extract the "averaged" behavior of F_0 in any stochastic situation where $\rho \gg 1$; and such an "averaged" $F_0(t)$ can be crudely represented by $\bar{F}_0(t) = \cos(\int_0^t dt' E/\rho)$, a curve that misses all the rapid fluctuations of frequency $\omega/2\pi$ and order of magnitude $1/\rho$ in F_0, but reproduces its slower-frequency and large-magnitude behavior. (The phrase "of order" used in this paper means "of order relative to the slowly varying, averaged forms of $\bar{F}_{0,3}$," which are assumed to be correct when specifying the size of the small, rapidly fluctuating corrections. This is an operational definition of accuracy, not an absolute one.) Various forms of this "averaging," for the F_i as well as F_0, are illustrated in the associated figures. In fact, one can construct a simple argument, using unitarity, to include the rapid fluctuations correct to order $1/\rho$; but the main thrust of our discussion in Sec. III is to derive simple forms for $\bar{F}_0$, $\bar{F}_i$ in the stochastic limit. Slightly more complicated forms are derived in the text for use with smaller values of $\rho > 1$, and they even bear a certain resemblance to the exact forms for $\rho \lesssim 1$.

In Sec. IV we turn to the application of these "averaged" approximations in the stochastic limit generated by white-noise Gaussian (WNG) functional integration, by first comparing the known result of exact WNG integration over U with the result of WNG integration over $\bar{F}_0$, which can also be done exactly (and has an amusing form reminiscent of a Heisenberg, nearest-neighbor, spin–spin interaction). To within a spurious phase factor, which can be easily understood and "renormalized" away, both expressions agree. Other more general examples of functional integration over the adiabatic and stochastic approximations are also considered, and are compared to numerical results performed on the Saclay CRAY.

A generalization of the stochastic-limit approximations to a three-dimensional input $\mathbf{E}(t)$ is written in Sec. V, and is presented there along with the relevant, associated figures. Finally, in the next section, a "fine tuning" of the first stochastic averaged functions is performed, resulting in curves that reproduce the exact numerical integrations in an uncanny way, including the small rapid oscillations correct to order $1/\rho$. The last section is devoted to a very brief summary, and the posing of some relevant questions for future study.

II. AN ALGORITHM FOR THE ADIABATIC LIMIT

In the extreme adiabatic limit $\rho = 0$, corresponding to $d\hat{E}/dt = 0$, all the complexity of the problem disappears. For, as noted above one can choose an arbitrary spatial axis to lie along the direction of $\hat{E}$, and the ordered exponential becomes an ordinary exponential, so that $U(t) \Rightarrow \cos G + i\boldsymbol{\sigma}\cdot\hat{\mathbf{E}} \sin G$, with $G(t) = \int_0^t dt' E(t')$. The adiabatic algorithm which we now construct should involve forms close to this limiting case.

Suppose that $\hat{E}(t)$ is a slowly varying unit vector, in the sense of very small ρ; then it is reasonable to choose as a first guess for $U(t)$ the same limiting form

$$U_0(t) = \exp[i\boldsymbol{\sigma}\cdot\mathbf{Q}_0(t)], \tag{2.1}$$

where $\hat{Q}_0(t) = \hat{E}(t)$ and

$$Q_0(t) = |\mathbf{Q}_0(t)| = \int_0^t dt' E(t'), \quad E = +\sqrt{\mathbf{E}^2}.$$

This is not correct, but it is unitary, and its deviation from the exact U can be expressed by a unitary $V(t)$: if $U(t) = U_0(t)\cdot V(t)$ with U_0 given by (2.1), then V must satisfy the exact differential equation

$$\frac{\partial V}{\partial t} = i\boldsymbol{\sigma}\cdot\left(\mathbf{E} - \hat{Q}_0\frac{dQ_0}{dt}\right)V - iQ_0\int_0^1 d\mu\, e^{-i\mu\boldsymbol{\sigma}\cdot\mathbf{Q}_0}\boldsymbol{\sigma}\cdot\frac{d\hat{Q}_0}{dt}e^{+i\mu\boldsymbol{\sigma}\cdot\mathbf{Q}_0}\cdot V, \tag{2.2}$$

or

$$\frac{\partial V}{\partial t} = i\boldsymbol{\sigma}\cdot\mathbf{E}_1 V,$$

with

$$\mathbf{E}_1(t) = \mathbf{E} - \hat{Q}_0\frac{dQ_0}{dt} + \frac{1}{2}\left\{\sin(2Q_0)\frac{d\hat{Q}_0}{dt} - [1-\cos(2Q_0)]\hat{Q}_0\times\frac{d\hat{Q}_0}{dt}\right\}. \tag{2.3}$$

We write (2.3) in the form $\mathbf{E}_1 = \mathscr{E}(Q_0,\hat{Q}_0;\mathbf{E})$, and note that while the first two rhs terms of (2.3) will cancel for the specific choice of Q_0 and $\hat{Q}_0$, the functional form of (2.3) will be useful later on. Under the initial condition $V(0) = 1$, the exact solution to (2.2) is

$$V(t) = \left(\exp\left[i\int_0^t dt'\boldsymbol{\sigma}\cdot\mathbf{E}_1(t')\right]\right)_+. \tag{2.4}$$

But if, in the $\rho \ll 1$ regime, the U_0 of (2.1) is a reasonable first approximation to U, then a reasonable approximation to (2.4) should be given by

$$V_1(t) = e^{i\boldsymbol{\sigma}\cdot\mathbf{q}_1(t)}, \tag{2.5}$$

where

$$\hat{q}_1(t) = \hat{E}_1(t), \quad q_1(t) = |\mathbf{q}_1(t)| = \int_0^t dt'|\mathbf{E}_1(t')|. \tag{2.6}$$

With this approximation, we have an "improved" estimate of $U(t)$,

$$U_1(t) \equiv U_0V_1 = e^{i\boldsymbol{\sigma}\cdot\mathbf{Q}_0(t)}\cdot e^{i\boldsymbol{\sigma}\cdot\mathbf{q}_1(t)}. \tag{2.7}$$

But the combination of (2.7) is unitary, and can be rewritten in a manifestly unitary form as

$$U_1(t) = e^{i\boldsymbol{\sigma}\cdot\mathbf{Q}_1(t)}, \tag{2.8}$$

with

$$Q_1(t) \equiv |\mathbf{Q}_1(t)| = \arccos[\cos Q_0\cdot\cos q_1 - (\hat{Q}_0\cdot\hat{q}_1)\sin Q_0\cdot\sin q_1],$$

or

$$Q_1(t) = Q(Q_0,q_1;\hat{Q}_0\cdot\hat{q}_1); \tag{2.9}$$

and

$$\hat{Q}_1(t) = [\hat{Q}_0 \sin Q_0\cdot\cos q_1 + \hat{q}_1 \sin q_1\cdot\cos Q_0 + (\hat{q}_1\times\hat{Q}_0)\sin q_1\cdot\sin Q_0](\sin Q_1)^{-1},$$

or

$$\hat{Q}_1(t) = \hat{Q}(\hat{Q}_0,\hat{q}_1;Q_0,q_1), \tag{2.10}$$

where the quantities Q and $\hat{Q}$ are defined by the first lines of (2.9) and (2.10), respectively.

But the same process can be repeated: instead of the U_0 of (2.1) we now have the U_1 of (2.8), and can define a better approximation $U_2 = \exp[i\boldsymbol{\sigma}\cdot\mathbf{Q}_2]$, with

$$\mathbf{E}_2(t) = \mathscr{E}(Q_1,\hat{Q}_1;\mathbf{E}),$$

$$\hat{q}_2(t) = \hat{E}_2(t), \quad q_2(t) = \int_0^t dt'|\mathbf{E}_2(t')|,$$

$$Q_2(t) = Q(Q_1,q_2;\hat{Q}_1\cdot\hat{q}_2), \quad \hat{Q}_2(t) = \hat{Q}(\hat{Q}_1,\hat{q}_2;Q_1,q_2).$$

Clearly, the process can be repeated an infinite number of times; and if it converges, can be represented by the fixed-point equations

$$\begin{aligned} &Q^* = Q(Q^*,q^*;\hat{Q}^*\cdot\hat{q}^*),\\ &\hat{Q}^* = \hat{Q}(\hat{Q}^*,\hat{q}^*;Q^*,q^*),\\ &\hat{q}^* = \hat{\mathscr{E}}(Q^*,\hat{Q}^*;\mathbf{E}),\\ &q^* = \int_0^t dt'|\mathscr{E}(Q^*,\hat{Q}^*;\mathbf{E})|, \end{aligned} \tag{2.11}$$

where Q^*, $\hat{Q}^*$, $\hat{q}^*$, q^*, and $\mathbf{E}$ are functions of t, and the functional forms $\mathscr{E}$, Q, and $\hat{Q}$ are given by (2.3), (2.9), and (2.10).

For an arbitrary input $\mathbf{E}(t)$, there is probably little hope of finding or proving convergence; but for some suitably simple input this might be possible. For our purposes, we note that if $\mathbf{E}(t)$ is chosen to be a vector of constant magnitude E rotating in the (x, y) plane with a constant angular frequency ω, then for $\rho = \omega/E \sim 0.1$, U_1 is a better approximation to the exact U than is U_0, as illustrated in Fig. 1 where the first two approximations to $F_0(t)$ are compared with the exact, or numerically integrated result. In fact, U_1 provides somewhat too much modulation, which should be removed by U_2, etc.

Results equivalent to the U_1 correction to U_0 have been discussed, in special contexts, in Refs. 2 and 3. To our knowledge, the algorithm for general U_n, as well as the fixed-point equations (2.11), is new; however, these latter statements are probably too complicated to be of much practical use. Generalization to SU(N) is simple for U_0 (Refs. 5 and 6), and while the general algorithm can be defined for arbitrary N, the more complicated statement of unitarity there will make this task much more tedious.

III. THE STOCHASTIC LIMIT

For $\rho \gg 1$ we again choose for $U(t)$ the manifestly unitary form, $U(t) = \exp[i\boldsymbol{\sigma}\cdot\mathbf{G}(t)]$, with $\mathbf{G} = \hat{G}\cdot G$, $G = +\sqrt{\mathbf{G}^2}$, and substitute into (1.1) to obtain

$$\boldsymbol{\sigma}\cdot\mathbf{E}(t) = \int_0^1 d\mu\, e^{i\mu\boldsymbol{\sigma}\cdot\mathbf{G}}\boldsymbol{\sigma}\cdot\frac{d\,\mathbf{G}}{dt}e^{-i\mu\boldsymbol{\sigma}\cdot\mathbf{G}} \tag{3.1}$$

or

$$\mathbf{E}(t) = \hat{G}\frac{dG}{dt} - \frac{1}{2}[1-\cos(2G)]\left(\hat{G}\times\frac{d\hat{G}}{dt}\right) + \frac{1}{2}\sin(2G)\frac{d\hat{G}}{dt}, \tag{3.2}$$

which is equivalent to the pair of exact relations

$$\frac{dG}{dt} = \hat{G}(t)\cdot\mathbf{E}(t), \tag{3.3}$$

$$\frac{d\hat{G}}{dt} = E\,[\hat{\mathrm{G}}\times\hat{E} + \cot G(\hat{E} - \hat{G}(\hat{E}\cdot\hat{G}))]. \tag{3.4}$$

With the initial conditions $G(0) = 0$, $\hat{G}(0) = \hat{E}(0)$, the magnitude $G(t)$ is completely determined by $\hat{G}$. For simplicity, we suppose that E lies in the (x, y) plane; the three-dimensional generalization is treated in Sec. V.

Since $\hat{G}$ is a unit vector, it can be specified by two independent quantities which we choose in the following way. For $\rho \ll 1$, we know that $\hat{G}(t) \sim \hat{E}(\int_0^t dt'\,\omega)$, but as $\omega = |\hat{E}\times d\hat{E}/dt|$ increases from zero this cannot be retained; rather, we suppose that the argument of $\hat{E}$ can be specified by a phase change relative to $\int_0^t \omega dt'$: $\hat{E}(t) \to \hat{E}(\int_0^t \omega\, dt' - \delta(t))$. It will be convenient to use a dimensionless time variable, τ, given by $d\tau = E\,dt$, and so write this phase-shifted unit vector as $\hat{E}(\int_0^t d\tau' \rho - \delta(\tau))$. But since $\hat{E}$ lies in the (x, y) plane, and $\hat{G}$ will have a $\hat{z}$ component for arbitrary ρ while remaining a unit vector, we choose the ansatz

$$\hat{G}(\tau) = \cos\phi(\tau)\hat{E}\left(\int_0^\tau d\tau'\rho - \delta(\tau)\right) + \hat{z}\sin\phi(\tau), \tag{3.5}$$

with $\phi(\tau)$ and $\delta(\tau)$ the two independent functions needed to characterize $\hat{G}$. Substitution of (3.5) into (3.4) yields the two independent equations

$$\frac{d\delta}{d\tau} = \rho - \tan\phi\cdot\cos\delta - \frac{\sin\delta}{\cos\phi}\cdot\cot G, \tag{3.6}$$

$$\frac{d\phi}{d\tau} = \sin\delta - \sin\phi\cdot\cos\delta\cdot\cot G, \tag{3.7}$$

which, together with the initial conditions $\delta(0) = \phi(0) = 0$ and the relation

$$G(\tau) = \int_0^\tau d\tau' \cos\delta(\tau')\cdot\cos\phi(\tau'), \tag{3.8}$$

completely determine $\hat{G}$. Equation (3.8) is obtained using the assumed variation of $\hat{E}$, for arbitrary $\boldsymbol{\omega}(t)$, $E(t)$:

$$\frac{d\hat{E}}{d\tau}\left(\int_0^\tau \rho\, d\tau'\right) = \boldsymbol{\omega}\times\hat{E}\left(\int_0^\tau \rho\, d\tau'\right),$$

with

$$\hat{E}\left(\int_0^\tau \rho\, d\tau'\right) = \hat{\imath}\cos\left(\int_0^\tau \rho\, d\tau'\right) + \hat{\jmath}\sin\left(\int_0^\tau \rho\, d\tau'\right).$$

It is clear that Eqs. (3.6)–(3.8) are very nonlinear, and it is difficult to have any intuitive feeling about their solutions in the large ρ limit. In order to obtain this intuition, one may machine-integrate these equations—or, equivalently, those that follow by substituting the ansatz $U = F_0 + i\boldsymbol{\sigma}\cdot\mathbf{F}$ into (1.1), along with unitarity restriction $F_0^2 + \mathbf{F}^2 = 1$—and find for $\rho \gg 1$ a remarkable simplification. For simplicity, we again for the moment consider ω and E, and therefore ρ, as constants, and watch the exact solutions for $F_0 = \cos G$ change as ρ is increased through unity to large values. For $\rho \gg 1$, one finds that a very rapid oscillation, of frequency $\omega/2\pi$ and mangitude ρ^{-1}, is superimposed upon a relatively, slowly varying oscillation, of frequency $\sim E/2\pi\rho$ and magnitude unity. The rapid oscillations should be irrelevant to any physical property described by this system of equations, and it is therefore natural to phrase the question: is it possible to approximate $U(t) = U(t|E)$ as a function of E so that, in the large ρ limit, one reproduces only the "averaged," or slowly varying behavior, and not the rapid fluctuations? The answer to this question is, indeed, yes; it is the point of this section, and we now outline the derivation of such "averaged" functions, to be denoted by $\overline{F}_{0,i}$.

Just the "experimental" knowledge that, for $\rho \gg 1$, the output for, e.g., F_0 consists of rapid oscillations superimposed on a slowly varying function of form $\sim\cos(Et/\rho)$ is enough to suggest an argument that can be followed. For, from (3.8), this means that as far as the "averaged" behavior is concerned, the quantity $J \equiv \cos\delta(\tau)\cdot\cos\phi(\tau)$ can be treated as a constant. (This statement will be refined, in Sec. VI, when we discuss "fine tuning.") It will be useful to define the associated quantity $H \equiv \cos\phi\cdot\sin\delta$, so that $\cos^2\phi = J^2 + H^2$, and the exact equations (3.6)–(3.8) can be expressed as

$$J' = -\rho H + [1 - J^2]\cot G, \tag{3.9}$$

$$H' = -\sin\phi + \rho J - HJ\cot G, \tag{3.10}$$

and

$$G = \int_0^\tau d\tau'\, J(\tau'). \tag{3.11}$$

For the "averaged" behavior, $J \sim \text{const} \equiv \xi(\rho)$, (3.9) may be replaced by

$$H \simeq ((1 - \xi^2)/\rho)\cot G, \tag{3.12}$$

with $G \simeq \tau\xi$. Just as G depends on the slowly varying time dependence, so must the "averaged" H of (3.12). Substituting the latter into (3.10), with $G = \tau\xi$, yields an equation for an "averaged" $\sin\phi$,

$$\sin\phi = \rho\xi + \xi(1 - \xi^2)/\rho. \tag{3.13}$$

The form of (3.13) will be more complicated if ρ depends upon t, or τ, but for $\rho \gg 1$ this extra dependence need not be important; this will be discussed in Sec. V. For the remainder of this derivation we shall continue to assume that ρ is essentially constant; but we shall not hesitate to state our results for time-dependent ρ, where our final formulas continue to work in a satisfactory way.

If our analysis leading to (3.13) is correct, $\sin\phi$ should display an "averaged" behavior, with rapid oscillations superimposed on a constant background; and this is true, experimentally, as one can see in Fig. 3. One may note that there is a change of procedure used here, in the following sense. An exact (numerical) integration of (3.6)–(3.8) yields a value of G that never increases past π, while $\sin\phi$ and J are positive when the average G is increasing and negative when it decreases (so that $\hat{G}$ can cover all points on the unit sphere). In contrast, our "averaged" G will increase

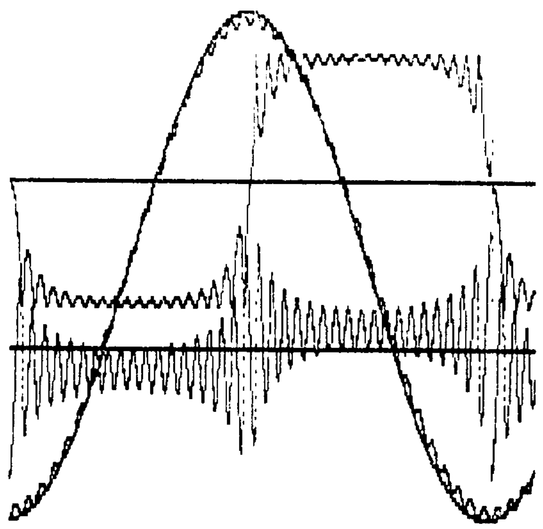

FIG. 3. Graphs of (a) $\sin\phi$, (b) $J = \cos\phi\cdot\cos\delta$, and (c) a superposition of F_0 and $\overline{F}_0$; for $\rho = 6$, $E = 10$.

without limit, so that $\sin G$ may become negative (just when the exact G was decreasing), while the averaged $\sin\phi$ and J are replaced by positive constants. In this way we are able to represent the correct signs of all the (F_0, F_i). This same feature of always positive $\sin\phi$ and J can occur in numerically integrated solutions of the exact equations (3.6)–(3.8) depending on the accuracy of the computation and the passage through the singular regions of $\cot G$. For our purposes, both $\sin\phi$ and J can be thought of as having an "averaged," constant value, even though in reality they oscillate about that value, and oscillate wildly near the regions $G \sim n\pi$. In contrast, a plot of $\sin\delta$ displays an almost uniform density of points spread over the same intervals.

We now use the "averaged" constancy of $\sin\phi$, or of $\cos^2\phi = J^2 + H^2$, to determine the dependence of ξ on ρ. For, if the "averaged" value of $(d/d\tau)$ $(\frac{1}{2}\cos^2\phi)$ is to vanish, from (3.9) and (3.10) one finds another expression for the "averaged" H,

$$0 \simeq -H\sin\phi + J[1 - (J^2 + H^2)]\cot G,$$

or

$$H = \xi\sin\phi\cot(\xi\tau). \tag{3.14}$$

Comparing with (3.12) we obtain

$$\xi\sin\phi = (1 - \xi^2)/\rho, \tag{3.15}$$

and finally, comparing (3.15) with (3.13) yields

$$(1 - \xi^2)/\rho = \rho\xi^2[1 + (1 - \xi^2)/\rho^2],$$

or

$$\xi(\rho) = \sqrt{1 + \rho^2/4} - \rho/2. \tag{3.16}$$

In obtaining (3.16) it has been supposed that $\xi > 0$ and $1 - \xi^2 > 0$. The slightly more complicated form of $\xi(\rho)$ used in Ref. 1 is exactly equivalent to (3.16). Limiting forms are

$$\xi(\rho)_{\rho \gg 1} \sim 1/\rho - 1/\rho^2 + \cdots$$

and

$$\xi(\rho)_{\rho \ll 1} \sim 1 - (\rho/2) + \cdots.$$

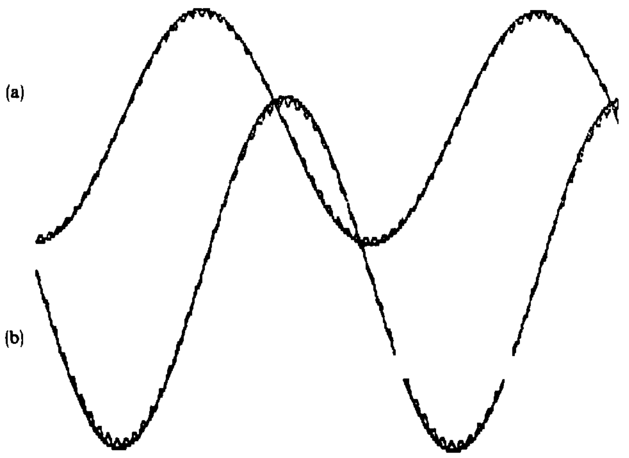

FIG. 4. Superpositions of (a) F_0 and $\overline{F}_0$, and (b) F_3 and $\overline{F}_3$; for $\rho = 6$, $E = 10$.

From (3.16) and (3.15), it follows that our "averaged" $\sin\phi$ is given by unity, which is certainly compatible with the curve of $\sin\phi$ illustrated in Fig. 3.

With these relations, our "averaged" solutions for F_0, F_3 are given by

$$\overline{F}_0 = \cos G, \tag{3.17}$$

$$\overline{F}_3 = \sin G, \tag{3.18}$$

where

$$G = \tau\xi(\rho) \to \int_0^\tau d\tau'\,\xi(\rho(\tau')) \tag{3.19}$$

is appropriate as a first generalization to time-dependent E and ω. The accuracy of these expressions is quite good for $\rho > 5$, where errors, or deviations from the numerically integrated F_0, F_3 are rarely worse than a few percent, and frequently much less. Even for $\rho \sim 1$, where the analysis is certainly not valid, one finds that these expressions for F_0 and F_3 do tend to average out the then, nonrapid fluctuations of the machine integrated F_0, F_3. Some typical outputs may be seen in Figs. 4–7, including several examples of t-dependent E and ω. One finds, generally, that even if ρ has some oscillation superimposed on a constant value$\gg 1$, the "averages"

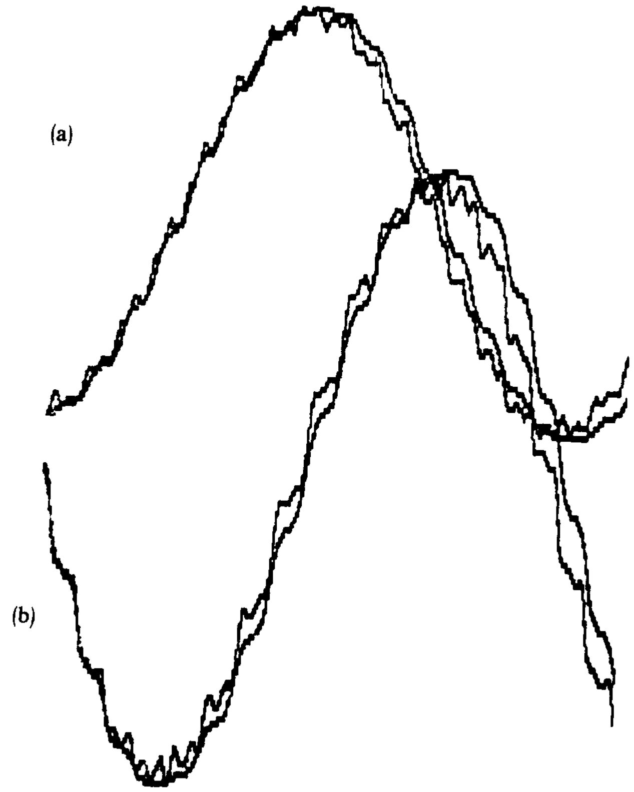

FIG. 6. Superpositions of (a) F_0 and $\overline{F}_0$, and (b) F_3 and $\overline{F}_3$; for $\omega = 60$, $E(t) = 10 + 5\sin(30t)$.

given by (3.17)–(3.19) continue to be reasonably accurate. One must, of course, be careful about the errors that accumulate in numerical integrations; a typical such effect will be the appearance of a phase lag between F_0 and $\overline{F}_0$, and between F_3 and $\overline{F}_3$, which is a "Δt-effect," and may be decreased by choosing a smaller integration step or a more accurate method of integration.

Analogous approximate expressions for $F_{1,2}$ are easily written. One has, exactly,

$$F_1 = \sin G\,[J\cos L + H\sin L\,], \tag{3.20}$$

$$F_1 = \sin G\,[J\sin L - H\cos L\,], \tag{3.21}$$

with $L = \int_0^t dt'\,\omega(t')$. Inserting the same "averaged" ap-

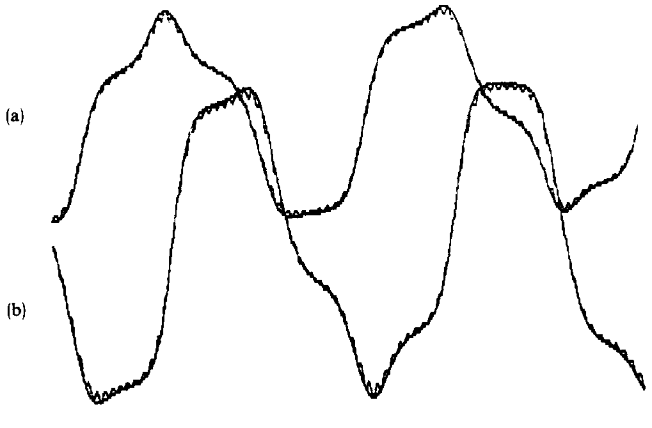

FIG. 5. Superpositions of (a) F_0 and $\overline{F}_0$, and (b) F_3 and $\overline{F}_3$; for $\omega = 60$, $E(t) = 10 + 5\sin(5t)$.

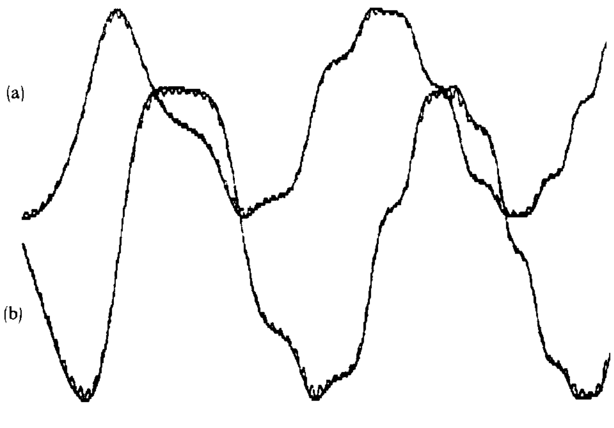

FIG. 7. Superpositions of (a) F_0 and $\overline{F}_0$, and (b) F_3 and $\overline{F}_3$; for $\omega = 60$, $E(t) = 10 + \cos(t^2)$.

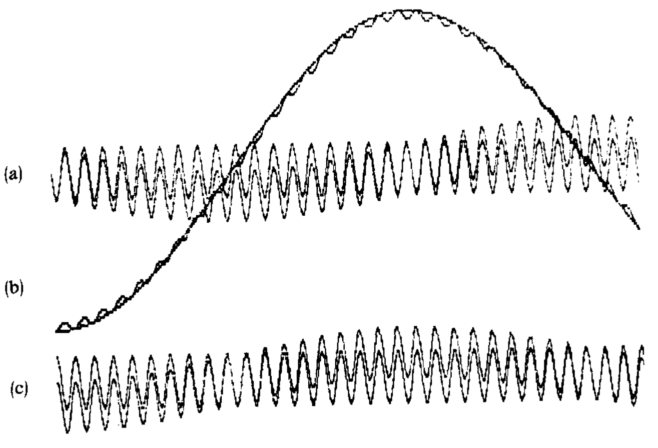

FIG. 8. Superpositions of (a) F_1 and $\overline{F}_1$, and (b) F_0 and $\overline{F}_0$, and (c) F_2 and $\overline{F}_2$; for $\rho = 6$, $E = 10$.

proximations for J, H, G as before, one finds

$$\overline{F}_1 = \xi \sin(G + L), \tag{3.22}$$

$$\overline{F}_2 = -\xi \cos(G + L), \tag{3.23}$$

where G is again given by (3.19). For large ρ, $\xi \sim 1/\rho$, and these $\overline{F}_{1,2} \sim O(1/\rho)$, in contrast to our $\overline{F}_{0,3} \sim O(1)$. These $\overline{F}_{1,2}$ are therefore small, and oscillate rapidly, and should have little physical importance in any specific problem. However, as seen in Fig. 8, Eqs. (3.22) and (3.23) do miss some of the slowly varying dependence of the exact $F_{1,2}$, even if the dependence is itself on the order of $1/\rho$. In Sec. VI we will give a simple argument to correct the $\overline{F}_{1,2}$ above, so that they will be correct to $O(1/\rho)$; and in the process, use the requirement of unitarity to "fine tune" our $\overline{F}_{0,3}$, giving them a rapid oscillation superimposed on their "averaged" values which is correct to $O(1/\rho^2)$. Unitarity is, of course, approximately satisfied by the $F_{0,1,2,3}$ above,

$$(\overline{F}_0^2 + \overline{F}_3^2) + (\overline{F}_1^2 + \overline{F}_2^2) = 1 + \xi^2 \simeq 1 + (1/\rho^2) + \cdots .$$

In Sec. VI we shall arrange to have this unitarity sum given by $1 + O(1/\rho^4)$, and hence infer that the "fine-tuned" $\overline{F}_{0,i}$ are themselves correct to at least $O(1/\rho)$.

There is one qualification to the remarks of this section that must be noted, and that will be relevant to some of the functional integrals performed over our stochastic averaged forms. If the input $\mathbf{E}(t)$ can be split into two nonparallel parts of radically different magnitudes, then it is the large-magnitude input that controls the final output of U. For example, motion corresponding to the input $\mathbf{E} = \hat{i}E_1 + \hat{j}E_2 \cos(\omega t)$, with $E_1 \gg E_2$, is essentially adiabatic, regardless of the value of ω.

More generally if $\mathbf{E}(t)$ is given as the sum of two nonparallel pieces, $\mathbf{E} = \mathbf{E}_1(t) + \mathbf{E}_2(t)$, with arbitrary time dependence but where the magnitude of one is much larger than the other, say $E_1 \gg E_2$, then the prudent way to set up the calculation is to separate all the E_1 dependence into a unitary V, leaving a rotated E_2 dependence, say E_2' in W: $U = VW$, where

$$V = \left(\exp\left[i \int_0^t dt' \, \boldsymbol{\sigma}\cdot\mathbf{E}_1(t') \right] \right)_+ .$$

We assume that all the components of E_1 are of the same order of magnitude; anything much smaller is put into E_2. Then

$$W = \left(\exp\left[i \int_0^t dt' \, \boldsymbol{\sigma}\cdot\mathbf{E}_2'(t') \right] \right)_+ ,$$

where

$$\boldsymbol{\sigma}\cdot\mathbf{E}_2' = V^{+} \boldsymbol{\sigma}\cdot\mathbf{E}_2 V, \quad \mathbf{E}_2'^2 = \mathbf{E}_2^2 .$$

When we calculate the ρ-value of $\mathbf{E}_2'$ we will find it of order (ω_2/E_2) and/or $(E_1 E_2)$; explicitly

$$\rho^2(\mathbf{E}_2') = \left[\left(\frac{d\,\mathbf{E}_2}{dt} + [\mathbf{E}_1 \times \mathbf{E}_2] \right)^2 - \left(\frac{dE_2}{dt} \right)^2 \right] (E_2^4)^{-1} .$$

As long as $\mathbf{E}_1 \times \mathbf{E}_2 \neq 0$—and this was assumed—there will be a piece of this ρ proportional to $(E_1/E_2) \gg 1$. It is then appropriate to use the stochastic form for W, leading to the contribution

$$G_W \sim \int_0^t dt' \, \frac{E_2}{\rho(E_2')} \sim \int_0^t dt' \, O\left(\frac{E_2^2}{E_1} \right),$$

which will be a small correction to the G_V if G_V is adiabatic, or a contribution of the same form as G_V if G_V itself is stochastic.

IV. STOCHASTIC FUNCTIONAL INTEGRATION

One very nice check on our approximations is their ability to reproduce the result of the one, nontrivial functional integration over an ordered exponential which can be performed analytically, that of WNG integration over the $U(t\,|E)$ of (1.2). Indeed, one type of application of our results should be to stochastic functional integration over weightings more complicated than Gaussian. In this section, we first show why the stochastic limit is appropriate to WNG integration, and then just how our approximate forms can reproduce the known, exact result

$$N \int d[E] \exp\left[-\frac{1}{2c} \int_0^t dt' \, \mathbf{E}^2(t') \right] U(t\,|\mathbf{E}) = e^{-tc}, \tag{4.1}$$

where N is a normalization constant defined by

$$N^{-1} = \int d[E] \exp\left[-\frac{1}{2c} \int_0^t dt' \, \mathbf{E}^2(t') \right].$$

In (4.1) we denote by c a real, positive constant; and continue to suppose that $\mathbf{E}$ lies in the (x, y) plane.

We first remaind the reader of the derivation of (4.1). Imagine the interval $(0,t)$ broken up into n subintervals, of width $\Delta t = t/n$ and labeled by an index i, so that the $\mathbf{E}(t')$ field in each subinterval is denoted by $\mathbf{E}_i$. Then, one definition of the functional integral is

$$\lim_{n \to \infty} \prod_{i=1}^{n} N_i \int d^2E_i \, e^{-\Delta t \, \mathbf{E}_i^2/2c} (e^{i\Delta t \boldsymbol{\sigma}\cdot\mathbf{E}_i}), \tag{4.2}$$

and the ordering of the brackets is such that terms with the larger value of i stand to the left. But each integral yields a result independent of i—that is, independent of $\boldsymbol{\sigma}$—by the following argument.

Because of the Gaussian weighting, each E_i scales as $(\Delta t)^{-1/2}$; that is, in (4.2) replace each $\mathbf{E}_i$ by $\mathbf{F}_i/\sqrt{\Delta t}$ (including the normalization, $N_i \to N_i'/\Delta t$), and for small Δt

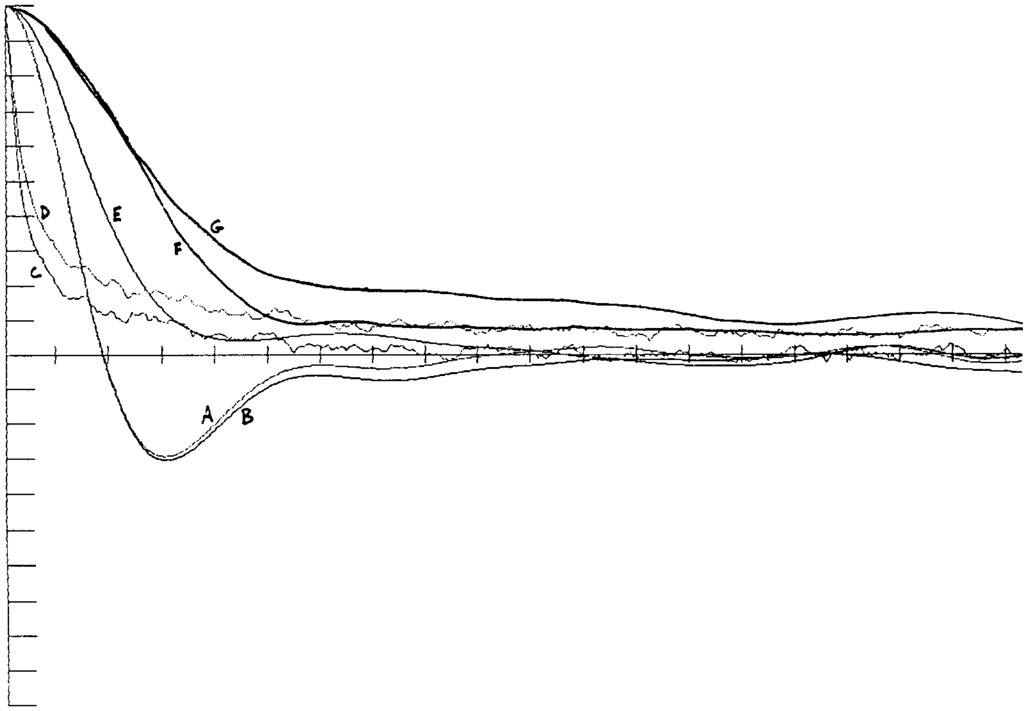

FIG. 9. A comparison of the result of functional integration over the exact $U(t\,|E)$ with several approximations, for $\tau^{-1} = 100$, $E_m = 10$, $\Delta t = 0.005$. The labeling of the curves is A = exact, numerical; B = renormalized $(1/\rho)$; $C = (1/\rho)$; D = full $\xi(\rho)$; E = renormalized, full $\xi(\rho)$; F = renormalized adiabatic; G = adiabatic.

expand each $\exp(i\sqrt{\Delta t}\,\boldsymbol{\sigma}\cdot\mathbf{F}_i)$ so that (4.2) becomes

$$\lim_{n\to\infty}\prod_{i=1}^{n} N_i' \int d^2F_i\, e^{-\mathbf{F}_i^2/2c}\Big(1 + i\sqrt{\Delta t}\,\boldsymbol{\sigma}\cdot\mathbf{F}_i - \frac{\Delta t}{2}\mathbf{F}_i^2 + \cdots\Big), \tag{4.3}$$

in which we retain only the leading, nonzero dependence proportional to Δt (the coefficient of $\sqrt{\Delta t}$ vanishes by symmetry). Each ith integral is the same; and it is trivial, yielding

$$\lim_{n\to\infty}(1 - c\Delta t)^n = e^{-ct}, \tag{4.4}$$

with $\Delta t = t/n$.

The essential part of this computation has been the observation that, for WNG integration, each $\mathbf{E}_i$ scales as $(\Delta t)^{-1/2}$. We now consider the same functional integration over our "averaged" forms. The first point to be settled is whether the stochastic limit is valid, and for this we must estimate the size of $\rho^2 = (d\hat{E}/dt)^2/E^2$. But, upon breaking up the interval $(0,t)$ into subintervals, any $\rho^2(t)$ would be replaced by

$$\rho_i^2 = (\hat{E}_i - \hat{E}_{i+1})^2/E_i^2(\Delta t)^2.$$

The $\hat{E}_i$ dependence is of $O(1)$; but because E_i scales as $(\Delta t)^{-1/2}$, $\rho_i^2 \sim O(1/\Delta t)$ and is *large*. Hence the stochastic limit most certainly is relevant, and we consider the functional integrals of our "averaged" forms in the limit of very large ρ, $\overline{U}\to\overline{F}_0 + i\sigma_3\overline{F}_3$, $\overline{F}_0 = \cos G$, $\overline{F}_3 = \sin G$. One then requires

$$N\int d[E]\exp\Big[-\frac{1}{2c}\int_0^t dt'\,\mathbf{E}^2(t')\Big]e^{\pm iG}, \tag{4.5}$$

which upon writing $G \simeq \int_0^t dt'(E/\rho)$ and breaking up the integration region into subintervals, generates

$$\lim_{n\to\infty}\prod_{i=1}^{n} N_i \int d^2E_i\, e^{-\Delta t\,\mathbf{E}_i^2/2c}\, e^{\pm i\Delta t E_i/\rho_i}, \tag{4.6}$$

where $\rho_i = +[(\hat{E}_i - \hat{E}_{i+1})^2]^{1/2}/E_i\Delta t$. Again rescaling $\mathbf{E}_i$, we now find in each subinterval both an integral over the magnitude F_i and a nontrivial angular dependence. Integration over each magnitude is trivial, leaving

$$\frac{1}{2\pi}\int_0^{2\pi} d\theta_i\Big(1 \mp \frac{ic\Delta t}{|\sin|(\theta_i/2)|}\Big)^{-1}. \tag{4.7}$$

The integral of (4.7) can be done exactly; with $q = \pm c\Delta t$, it is

$$1 + \frac{2iq}{\pi(1+q^2)^{1/2}}\ln\Big[\Big(\frac{1-(1+q^2)^{1/2}-iq}{1+(1+q^2)^{1/2}-iq}\Big)\times\Big(\frac{1+(1+q^2)^{1/2}}{1-(1+q^2)^{1/2}}\Big)\Big].$$

As $\Delta t\to 0$, the argument of the log becomes $\pm 2i/c\Delta t$, generating for the complete functional integral

$$\lim_{n\to\infty}\prod_{i=1}^{n}\Big(1 - c\Delta t \pm \frac{2ic\Delta t}{\pi}\ln\Big(\frac{2}{c\Delta t}\Big)\Big), \tag{4.8}$$

which can be written as

$$e^{-ct}e^{\pm(2ict/\pi)\ln(2/c\Delta t)}\big|_{\Delta t\to 0}. \tag{4.9}$$

Comparison with (4.1) shows that a spurious phase has appeared; but one that can be understood, and removed, by the following argument. In every subinterval's integration, our "averaged" forms have made a small error, which is (fortunately) imaginary, and which must be removed "by hand." Instead of calculating (4.5) as we have done, we must add the proviso that we keep only the real part of every subinterval's contribution; and in this way, by not retaining and compounding the small error generated by our "averaged" forms, we can reproduce (4.1). We expect this tendency towards a spurious phase factor to show up in more complicated functional integrals, or in functional integrals that are Gaussian but not precisely in the white-noise limit, and it will be necessary to remove such spurious dependence.

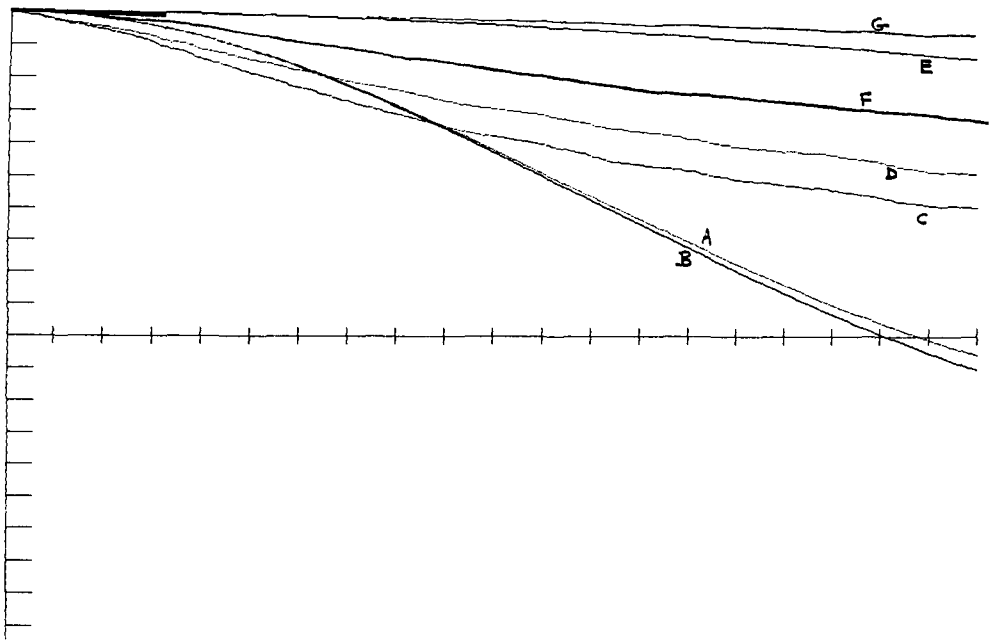

FIG. 10. The same comparisons, with the same labeling as in Fig. 9, using $\tau^{-1} = 100$, $E_m = 1$, $\Delta t = 0.005$.

This can be done most simply by replacing the functional integral over $e^{\pm iG}$, which we call $\langle e^{\pm iG}\rangle$, by the quantity $[|\langle e^{\pm iG}\rangle|^2]^{1/2}$, a computation we henceforth label "renormalized."

More general, non-WNG weightings may be treated by calculating Gaussian fluctuations with correlation function given by

$$\Delta_{ij}(t_1 - t_2) = \langle E_i(t_1)E_j(t_2)\rangle = \delta_{ij}(E_m/2\tau)e^{-|t_1 - t_2|/\tau},$$

where τ is a correlation time, and E_m an appropriate magnitude. The limit $\tau^{-1}\to\infty$ for $E_m = 1$ is the WNG case, $\Delta_{ij}\to\delta_{ij}\delta(t_1 - t_2)$, while the opposite limit, $\tau^{-1}\to 0$ is effectively the adiabatic limit. (This last remark would be strictly true if ρ were defined as $|d\,\hat{\mathbf{E}}/dt\,|/E^2$ rather than as $|d\hat{E}/dt\,|/E$; in practice there seems to be little difference.)

We illustrate in Figs. 9 and 10 calculations in the WNG region ($\tau^{-1} = 100$) over a variety of different possible approximations, and note that here the best agreement with the exact functional integration is given by first performing the large-ρ approximation of ξ, $\xi(\rho)\sim 1/\rho$, and then performing the functional integration. Why this is true—rather than using the exact $\xi(\rho)$ and letting the natural, large-ρ fluctuations automatically induce the effective large-ρ form of ξ—is a reflection of the comments made at the end of Sec. III. In the numerical computations there are many successive choices of E_i that correspond to large variations of $\hat{E}$, but of small magnitude, superimposed on a perpendicular component of large magnitude and slow variation; and these fluctuations are to be interpreted as adiabatic contributions of small, effective ρ. When the full $\xi(\rho)$ is used, such small-ρ contributions are incorrectly taken into account. However, with the large-ρ form of ξ, $\xi\sim 1/\rho$, the corresponding contributions to $\langle \mathrm{tr}\ U\rangle$ are small for small ρ, since such exponentiated terms are rapidly damped away. Using the large-ρ form of $\xi(\rho)$ suppresses such incorrect, effectively small-ρ behavior; and, as one can see from Figs. 9 and 10, provides fairly reasonable approximations to the exact result.

V. THREE-DIMENSIONAL INPUT

We here consider the generalization of the material of Sec. III to three-dimensional input $\mathbf{E}(t)$, which requires a generalization to time-dependent ρ. It will be appropriate to comment, firstly, on the derivation given in that section for a time-dependent ρ, and then to extend the analysis to three dimensions.

The passage from the exact equations (3.9)–(3.11) to our approximate, "averaged" forms was performed assuming a constant ρ, and using the "experimental" properties that J and $\sin\phi$ are given by rapid oscillations superimposed upon a constant background. If $\rho = \rho(t)$, one must first determine if the same properties of "averaged" constancy of J and $\sin\phi$ still exist, before an analysis of the same type can be given. The experimental answer, obtained for a variety of choices of the t dependence of ρ (but always insisting on $\rho \gg 1$) is that the angular integrations represented, e.g., by Fig. 11, are only slightly modified; experimentally, J and $\sin\phi$ may still be represented as constant quantities on which are superimposed rapid oscillations. This being the case, it does make sense to apply the same form of argument as was used to arrive at (3.12); but the form of (3.13) will now be complicated by the appearance of an extra term proportional to

$$\frac{d\rho}{dt}\left[\frac{(1-\xi^2)}{\rho^2} + \frac{2\xi}{\rho}\frac{\partial\xi}{\partial\rho}\right]\cot G.$$

The result is that (3.13) and (3.15) no longer yield an algebraic equation for $\xi(\rho)$, but rather, with specific input $d\rho/dt$, a differential equation for $\xi(\rho)$. The complication is decidedly nontrivial. Fortunately, if ξ still falls off as ρ increases, for $\rho \gg 1$ these terms should not have any important effect. More precisely, even if a time-dependent ρ (but, always, $\rho \gg 1$) adds small and rapid oscillations to our "averaged" forms, which need not agree with the small and rapid oscillations of the numerically integrated functions, the slowly varying behavior of the "averaged" forms still repro-

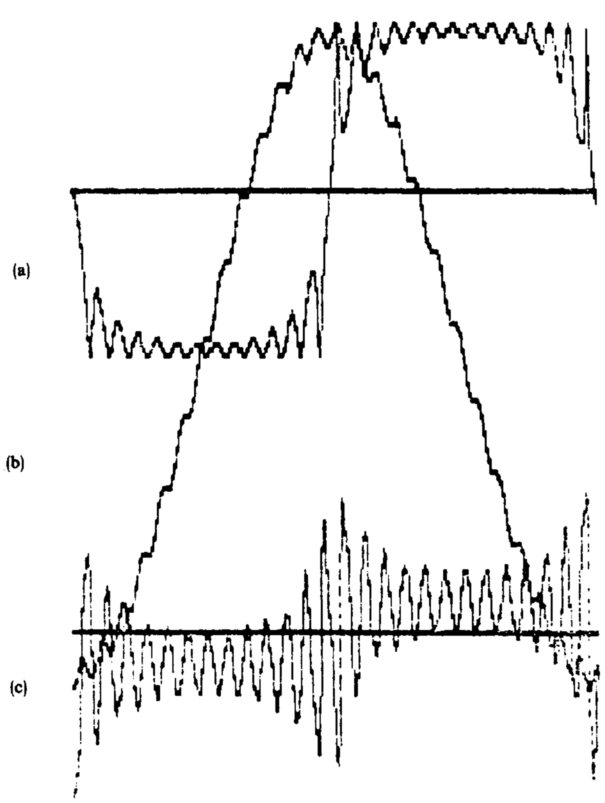

FIG. 11. Graphs of (a) $\sin\phi$, (b) F_0, and (c) $J=\cos\phi\cos\delta$; for $E=10$, $\rho(t)=30+30|\cos(60t)|$.

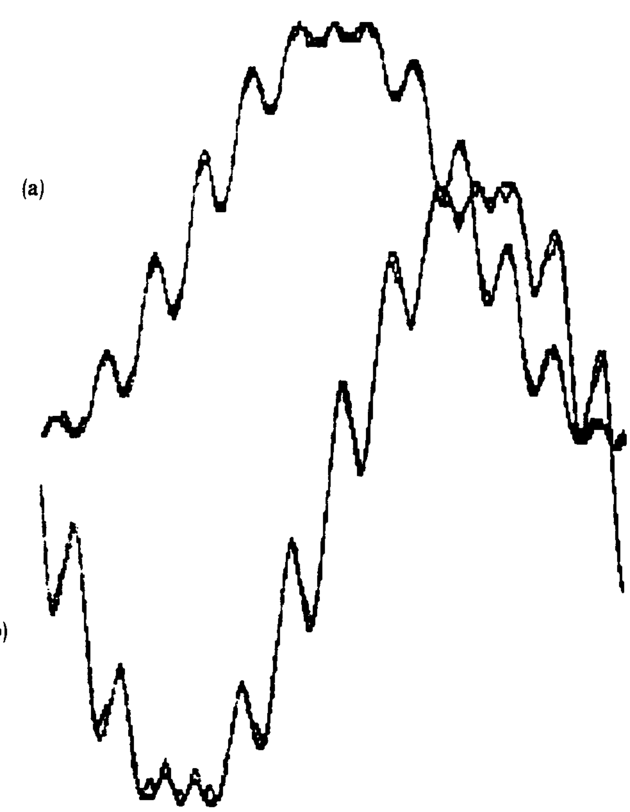

FIG. 12. Superpositions of (a) F_0 and $\overline{F}_0$, and (b) F_3 and $\overline{F}_3$; for $\omega=60$, $\nu=20$, $E_T=10$, and $E_L=5$.

duces that of the exact solutions. This can be clearly seen, even after the "fine-tuning" of Sec. VI, in Fig. 12, where an ω containing rapid oscillations inserted into our essentially constant (or slowly varyinbg) ω-analysis produces a curve whose small and rapid oscillations do not match the exact ones, but whose "averaged" shape continues to reproduce that of the exact curves.

We emphasize that we have not attempted a careful study of this quite complicated point; but we are convinced that, for $\rho\gg 1$, the specifically time-dependent effects of ρ are not important in developing the "averaged" forms in any way other than the elementary generalizations we have made,

$$\omega t\to\int_0^t dt'\,\omega(t'),\quad G=\tau\xi(\rho)\to\int_0^t dt'\,E(t')\xi(\rho(t')),$$

in writing our final formulas (3.17)–(3.24). To substantiate this claim, we point to the superimposed curves of F_0 and $\overline{F}_0$ of Figs. 5, 6, 7, and 12 made for a variety of choices of $\rho(t)$, and using only the $\overline{F}_0$ of (3.17).

In treating the problem of three-dimensional input $\mathbf{E}(t)$, it is always possible to perform a transformation on the basic equation (1.1) to yield a similar equation for a related quantity in which there appears a two-dimensional input $\mathscr{E}(t)$. For, if one defines another unitary quantity $V=e^{-(i/2)\theta(t)\sigma_3}\,U$, where $\theta(t)$ is a function to be determined, then the matrix V will satisfy

$$\frac{\partial V}{\partial t}=i\left(\sigma_1\mathscr{E}_1+\sigma_2\mathscr{E}_2+\sigma_3\left[E_3-\frac{\dot\theta}{2}\right]\right)V, \tag{5.1}$$

with $\theta(0)=0$, $V(0)=1$, and

$$\mathscr{E}_1\equiv E_1\cos\theta+E_2\sin\theta, \tag{5.2a}$$

$$\mathscr{E}_2\equiv E_2\cos\theta-E_1\sin\theta. \tag{5.2b}$$

If choose $\frac{1}{2}\theta(t)=\int_0^t dt'\,E_3(t')$, then the problem has been reduced to one of two-dimensional input.

Writing the exact solutions for V in the form $V=\mathscr{F}_0+i\boldsymbol{\sigma}\cdot\boldsymbol{\mathscr{F}}$, and comparing $U=F_0+i\boldsymbol{\sigma}\cdot\mathbf{F}$ with the solution obtained from $U=\exp[i\sigma_3(\theta/2)]\,V$, one has the exact statements

$$F_0=\mathscr{F}_0\cos(\theta/2)-\mathscr{F}_3\sin(\theta/2), \tag{5.3a}$$

$$F_3=\mathscr{F}_0\sin(\theta/2)+\mathscr{F}_3\cos(\theta/2), \tag{5.3b}$$

$$F_1=\mathscr{F}_1\cos(\theta/2)+\mathscr{F}_2\sin(\theta/2), \tag{5.3c}$$

$$F_2=\mathscr{F}_2\cos(\theta/2)-\mathscr{F}_1\sin(\theta/2). \tag{5.3d}$$

In order to write approximate, "averaged" expressions for the lhs of equations (5.3), we now apply the technique of Sec. III, writing, e.g.,

$$\overline{F}_0=\overline{\mathscr{F}}_0\cos(\theta/2)-\overline{\mathscr{F}}_3\sin(\theta/2), \tag{5.4}$$

and similarly for the other lines of (5.3). Here the $\overline{\mathscr{F}}$ are constructed in terms of a $\rho(\mathscr{E}_{1,2})$ of the two-dimensional problem. Clearly, $\mathscr{E}=(\mathscr{E}_1^2+\mathscr{E}_2^2)^{1/2}=(E_1^2+E_2^2)^{1/2}$, and $\hat{\mathscr{E}}_1=\mathscr{E}_1/\mathscr{E}$, $\hat{\mathscr{E}}_2=\mathscr{E}_2/\mathscr{E}$.

The description is simplest using cylindrical coordinates; if we choose $E_1=E_T\cos(\omega t)$, $E_2=E_T\sin(\omega t)$, $E_3=E_L\cos(\nu t)$, then $\mathscr{E}_1=E_T\cos(\omega t-\theta)$, $\mathscr{E}_2=E_T\sin(wt-\theta)$. For simplicity, suppose again that E_L and E_T, as well as ω and ν, are all constants; then one immediately calculates

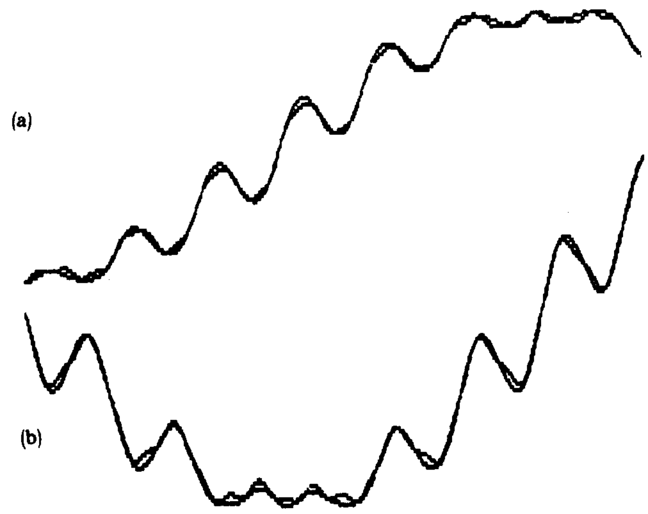

FIG. 13. Fine detail of the curves of Fig. 12, starting from $t = 0$.

$$\rho = |[\omega - 2E_L \cos(\nu t)]/E_T|, \tag{5.5}$$

exhibiting an explicitly time-dependent ρ, which will be used to calculate the $\mathscr{F}_{0,i}$. We again insist on the requirement

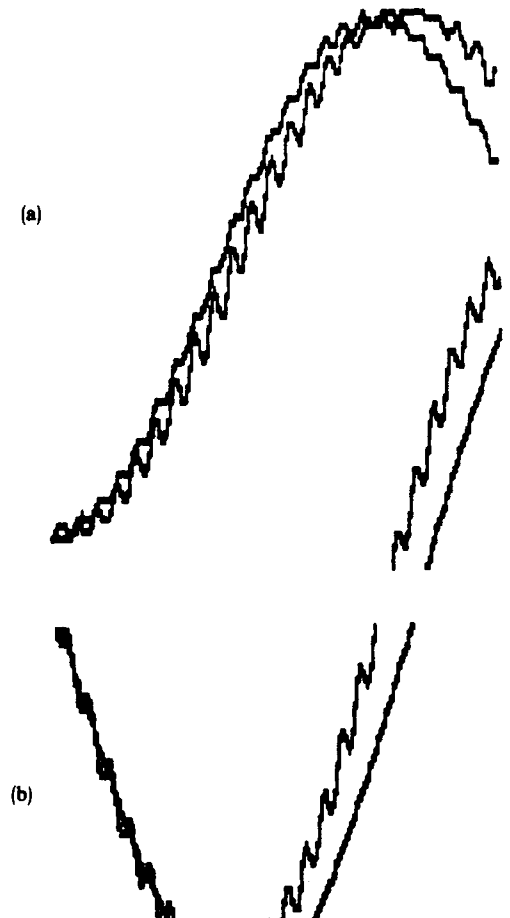

FIG. 14. Superpositions of (a) F_0 and $\overline{F}_0$, and (b) F_3 and $\overline{F}_3$; for $\omega = 60$, $\nu = 60$, $E_T = 10$, and $E_L = 5$.

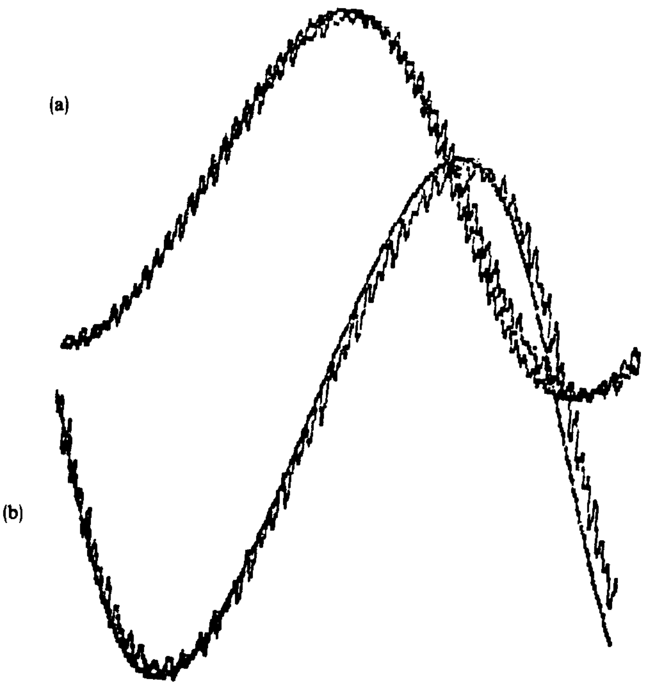

FIG. 15. Superpositions of (a) F_0 and $\overline{F}_0$, and (b) F_3 and $\overline{F}_3$; for $\omega = 60$, $\nu = 90$, $E_T = 10$, and $E_L = 5$.

$\rho \gg 1$, which condition governs the possible choices of ω, E_L, E_T.

Just how well the $\overline{F}_0$ and $\overline{F}_3$ reproduce the numerically integrated exact solutions can be seen from the examples of Figs. 12–15. For ω significantly larger or smaller than ν, the agreement is superb. For $\omega \sim \nu$ the agreement is less pleasing; but much of the discrepancy here seems to be tied up with the "in phase" errors made during the numerical computations. For example, a slowly forming phase lag gradually appearing between F_0 and $\overline{F}_0$, for $\omega = \nu$, is definitely diminished by using a finer time step in the numerical equations; however, we have not succeeded in completely removing this phase lag. This difficulty aside, which we believe is tied up with the deails of the numerical integration, it is difficult to be anything but enthusiatic over the quality of the results given by these "averaged" forms, using a three-dimensional input. Again, one finds that generalizations to time-dependent E_L, E_T, continue to be well represented by Eqs. (5.4), using in the computation of $\mathscr{F}_{0,i}$ the elementary generalizations of Sec. III for time-dependent E,ω.

VI. FINE TUNING

Of all the qualitative agreements between the exact solutions and our "averaged" functions, only the agreement between $F_{1,2}$ and $\overline{F}_{1,2}$ is less than satisfactory, because the $\overline{F}_{1,2}$ of Sec. III miss the low-frequency behavior clearly visible in the $F_{1,2}$; this is illustrated in Fig. 8. As a practical matter, it is not important because the $F_{1,2}$ are of order $1/\rho$ and are small; but as a matter of principle one would like to be able to extract all the correct, slowly varying behavior.

The trouble resides in our neglect of the small, rapid oscillations of J and H, in Sec. III, because those neglected fast oscillations could themselves be combined with similar

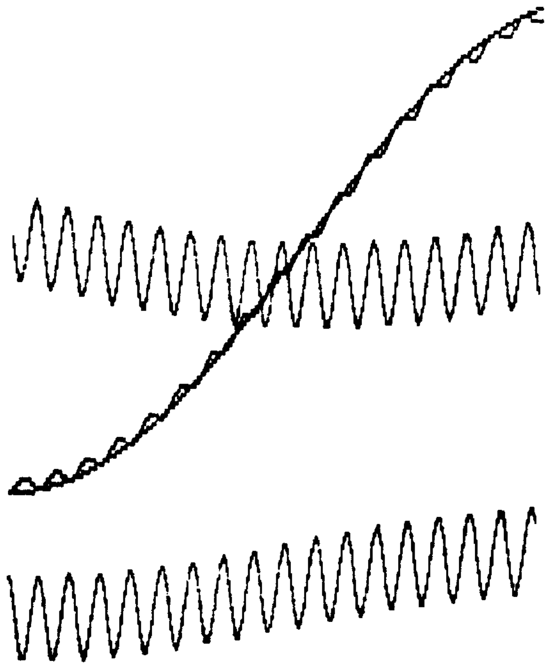

FIG. 16. Superpositions of (a) F_1 and $\overline{F}'_1$, and (b) F_0 and $\overline{F}_0$; (c) F_2 and $\overline{F}'_2$; for $\omega = 60$, $E = 10$.

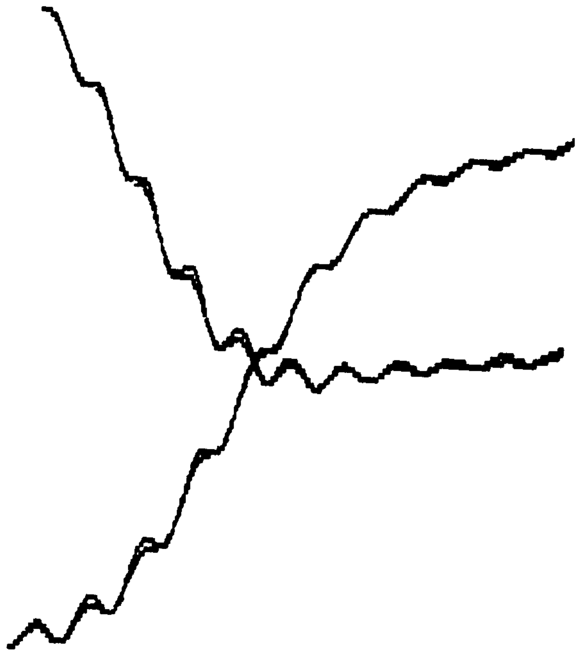

FIG. 18. Detail of the first shoulder for the superpositions of (a) F_0 and $\overline{F}'_0$, and (b) F_3 and $\overline{F}'_3$; for $\omega = 60$, $E(t) = 10 + 5\sin(5t)$.

oscillations appearing in the definition of the $F_{1,2}$, in (3.20) and (3.21), to generate terms independent of the rapid oscillations. To see this, denote by ξ and H_0 our previous choices of the constant and slowly varying J and H dependence, respectively; and then suppose that J and H shall each have a rapidly oscillating part of form

$$J = \xi + \xi\,[\cos L + \sin L \cot G\,], \tag{6.1a}$$

$$H = H_0 + \xi\,[\sin L - \cos L \cot G\,]. \tag{6.1b}$$

Imagine that there are constants, or slowly varying functions, α, β, γ, δ multiplying each of the $\sin L$, $\cos L$ terms in (6.1); and then imagine substituting (6.1) into the defining equations for $F_{1,2}$, to reproduce the $\overline{F}_{1,2}$ of (3.22)

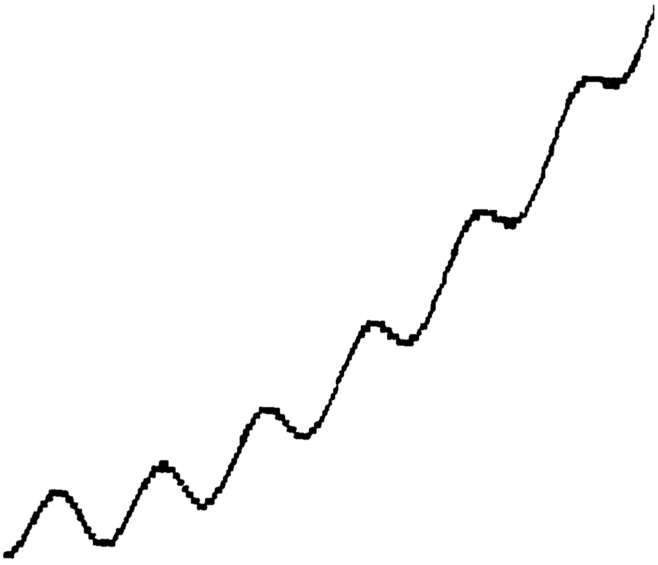

FIG. 17. Detail of the first shoulder of the superposition of F_0 and $\overline{F}'_0$; for $\omega = 60$, $E = 10$.

and (3.23) plus a part that has *only* a slowly varying time dependence. We denote by $\overline{F}'_{1,2}$ these new, improved functions, and find that we must choose $\alpha = \beta = \gamma = \delta = 1$, and then obtain

$$\overline{F}'_1 = \xi\,[\sin(G+L) + \sin G\,], \tag{6.2a}$$

$$\overline{F}'_2 = -\xi\,[\cos(G+L) - \cos G\,]. \tag{6.2b}$$

The agreement between (6.2) and the exact $F_{1,2}$ is so good that on the scale used in Fig. 16 there is no visible difference at all between them. Only when the scale is enlarged to show effects of order $1/\rho^2$ can one see superpositions of two curves.

These new values of J and H, given by (6.1), can now be used to define new $\overline{F}'_{0,3}$, which themselves are correct to order $1/\rho$. But it is much easier to use an argument suggested by unitarity, which requires

$$\overline{F}'^2_0 + \overline{F}'^2_3 + \overline{F}'^2_1 + \overline{F}'^2_2 = 1 + O(1/\rho^4) \tag{6.3}$$

if the new, "averaged" functions are to be correct to order $(1/\rho)$. For if we write

$$\overline{F}'_0 = \overline{F}_0 + \delta\overline{F}_0, \quad \overline{F}'_3 = \overline{F}_3 + \delta\overline{F}_3, \tag{6.4}$$

and substitute into (6.3), using the $\overline{F}'_{1,2}$ of (6.2) one obtains the relation

$$\delta\overline{F}_0 \cos G + \delta\overline{F}_3 \sin G + \xi^2[1 + \sin G \sin(G+L) - \cos G \cos(G+L)\,] = 0. \tag{6.5}$$

Rewriting the "1" coefficient of ξ, in (6.5) as $\sin^2 G + \cos^2 G$, and equating the coefficients of sin G and cos G, generates

$$\delta\overline{F}_0 = +\xi^2(\cos(G+L) - \cos G\,), \tag{6.6a}$$

$$\delta\overline{F}_3 = -\xi^2(\sin(G+L) + \sin G\,). \tag{6.6b}$$

The agreement between the $\overline{F}'_{0,3}$, given by (6.6) and (6.4) is extremely good, as displayed in Figs. 17 and 18.

From this construction we infer that these $\overline{F}'_{0,3}$ are correct to order $1/\rho^2$, while the $\overline{F}'_{1,2}$ are correct to order $1/\rho$. Again, this "fine tuning" is probably irrelevant in any given physical application, but it is pleasing to be able to improve the accuracy of our "averaged" curves in such a simple way.

VII. SUMMARY AND FURTHER QUESTIONS

In this paper we have suggested some methods for the approximate estimation of SU(2) ordered exponentials in the SC limits, adiabatic and stochastic, and have compared the results to exact or machine statements when certain functional integrals are carried out using our approximate forms. Our derivations have been mainly intuitive; but there can be no argument raised against the results which those derivations provide, which nicely match the numerically integrated functions representing the exact ordered exponential in both the adiabatic and stochastic limits. As such, we expect that these approximations will be immediately useful in a variety of physical problems, whose dynamical content can be expressed, approximated, or modeled in terms of SU(2) ordered exponentials.

There are three main areas in which the analyses of this paper raise questions that are surely deserving of futher attention.

(1) Are there possible choices of $\mathbf{E}(t)$ for which the fixed-point equations (3.1) have a nontrivial solution?

(2) A thorough analysis should be made of the generalization to time-dependent $\rho(t)$. Would the result of this investigation show that the $\overline{F}_{0,1}$ are insensitive to the time dependence of ρ, as suggested by all of our examples; or will there be certain situations, certain forms for $\rho(t) \gg 1$, for which our results are invalid?

(3) Can our results be extended to SU(N), $N > 2$?

It is not difficult to write the leading term of the adiabatic approximation for the case of SU(N), but its corrections will surely be more complicated because of the more cumbersome statement of unitarity.[5] In the stochastic limit, on the other hand, the situation seems less well-defined, and the methods of Sec. III would appear to be hazardous and uncertain. In principle, the same techniques can be used; in practice, the greater number of functions $F_{0,i}$, $1 \leqslant i \leqslant N^2 - 1$, makes for a certain amount of confusion. Surely, the much greater number of physical problems that involve SU(N), rather than SU(2), would make this a study of paramount interest.

ACKNOWLEDGMENTS

One of us (H.M.F) would like to express his warm thanks to U. Frisch and his colleagues, at the Observatoire de Nice where this work was begun, for useful criticism and the opportunity to use the necessary numerical facilities; and to colleagues at Physique Théorique, Université de Nice, where this paper was completed, for their very kind hospitality.

Much of this work was made possible by an NSF–CNRS Cooperative Research travel grant. H.M.F. would also like to acknowledge partial support by the U.S. Department of Energy under Contract No. DE-AC02-76A03130. A009—Task A.

[1] M. E. Brachet and H. M. Fried, Phys. Lett. A **103**, 309 (1984).
[2] L. Aspinall and J. C. Percival, Proc. R. Soc. **90**, 315 (1967); and references quoted therein.
[3] A. Dykhne, Sov. Phys. JETP **11**, 411 (1960); JETP Lett. **14**, 941 (1962).
[4] H. M. Fried, Phys. Rev. D **16**, 1916 (1977).
[5] H. M. Fried and J. Tessendorf, J. Math. Phys. **25**, 1144 (1984).
[6] H. M. Fried, "Strong coupling & the infrared structure of QCD," *Proceedings of the US–France Seminar "Theoretical Aspects of QCD,"* in Marseille, June, 1981.

Asymptotically flat space-times have no conformal Killing fields

David Garfinkle
Department of Physics and The McDonnell Center for the Space Sciences, Washington University, St. Louis, Missouri 63130

(Received 28 March 1986; accepted for publication 27 August 1986)

It is shown that asymptotically flat space-times with certain properties do not admit conformal Killing fields which are not Killing fields. The space-times must be vacuum, asymptotically Minkowskian, and have positive Bondi energy.

I. INTRODUCTION

Asymptotically flat space-times are thought to represent isolated gravitating systems. Thus there has been much interest in them. Research has been done on specific asymptotically flat space-times (e.g., the Schwarzschild, Kerr, and Weyl solutions of Einstein's equation[1]) and on general properties of asymptotically flat space-times.[2] One property of interest is the group of symmetries admitted by such a space-time. In particular, the possible isometries admitted by asymptotically flat space-times have been studied.[3,4]

However, conformal isometries have not been studied, largely because it has been thought that in some sense asymptotically flat space-times do not admit conformal isometries.[5] Results along these lines have been proven for space-times that are asymptotically flat at spatial infinity.[6,7] In this paper we present a result for space-times that are asymptotically flat at null infinity.

The conditions that the space-time must satisfy are as follows: (a) it must be vacuum, (b) it must be asymptotically Minkowskian, i.e., it must contain "all of $\mathscr{I}$," and (c) it must have positive Bondi energy. Under these conditions, the space-time does not admit a conformal Killing field that is not a Killing field.

The notation used will be as in Ref. 2. Section II contains a review of some of the properties of conformal Killing fields and of asymptotically flat space-times as well as a statement of the result. Section III contains a proof of the result.

II. NOTATION

Let $\widetilde{M}$ be a manifold, $\tilde{g}_{ab}$ a smooth (i.e., C^∞) metric, and ξ^a a vector field. Then ξ^a is said to be a conformal Killing field for the space-time $(\widetilde{M}, \tilde{g}_{ab})$ if

$$\mathfrak{L}_\xi \tilde{g}_{ab} = -2q\tilde{g}_{ab}, \tag{1}$$

for some scalar field q. (Note that the minus sign and factor of 2 are chosen for later convenience.) If $q=0$, then ξ^a is a Killing field. Let $\widetilde{\nabla}_a$, $\widetilde{R}_{abc}{}^d$, and $\widetilde{R}_{ab}$ be, respectively, the derivative operator and the Riemann and Ricci tensors associated with $\tilde{g}_{ab}$. Define the following tensor fields:

$$\tilde{\xi}_a \equiv \tilde{g}_{ab}\xi^b, \tag{2}$$

$$F_{ab} \equiv \widetilde{\nabla}_{[a}\tilde{\xi}_{b]}, \tag{3}$$

$$k_a \equiv -2\widetilde{\nabla}_a q, \tag{4}$$

$$L_{ab} \equiv \widetilde{R}_{ab} - \tfrac{1}{6}\tilde{g}^{cd}\widetilde{R}_{cd}\tilde{g}_{ab}. \tag{5}$$

Then the following equations hold[8]:

$$\widetilde{\nabla}_a\tilde{\xi}_b = F_{ab} - q\tilde{g}_{ab}, \tag{6}$$

$$\widetilde{\nabla}_a F_{bc} = \tilde{g}_{a[c}k_{b]} + \widetilde{R}_{cba}{}^d\tilde{\xi}_d, \tag{7}$$

$$\widetilde{\nabla}_a q = -\tfrac{1}{2}k_a, \tag{8}$$

$$\widetilde{\nabla}_a k_b = -\xi^c\widetilde{\nabla}_c L_{ab} + 2qL_{ab} - 2\tilde{g}^{cd}L_{c(a}F_{b)d}. \tag{9}$$

The values of the tensor fields $[\xi^a, F_{ab}, q, k_a]$ at a point p are called the conformal Killing data for ξ^a at point p. Given any conformal Killing data at a point p, Eqs. (6)–(9) can be integrated along any curve to give the value of ξ^a on that curve. If the conformal Killing data come from an actual conformal Killing field ξ^a, this integration procedure can be used to give the value of ξ^a everywhere. Since $\tilde{g}_{ab}$ is smooth, this procedure must yield a smooth ξ^a. Note that if the conformal Killing data vanish at any point, then $\xi^a = 0$ everywhere. Also note that if $(\widetilde{M}, \tilde{g}_{ab})$ is a vacuum space-time, then $L_{ab}=0$, and thus $\widetilde{\nabla}_a k_b = 0$.

The space-time $(\widetilde{M}, \tilde{g}_{ab})$ is said to be asymptotically flat at null infinity if there exists a manifold M with boundary $\mathscr{I}$ and smooth (i.e., C^∞) metric g_{ab}; a smooth function Ω on M and a mapping from $\widetilde{M}$ to M (by means of which we identify $\widetilde{M}$ with a subset of M) such that

(i) on $\widetilde{M}$, $\quad g_{ab} = \Omega^2\tilde{g}_{ab}$,

(ii) on $\mathscr{I}$, $\quad \Omega = 0, \quad \nabla_a\Omega \neq 0, \quad g^{ab}\nabla_a\Omega\nabla_b\Omega = 0.$

Here $\tilde{g}_{ab}$ is called the physical metric, g_{ab} the unphysical metric,[9] and (M, g_{ab}, Ω) an asymptote of $(\widetilde{M}, \tilde{g}_{ab})$, and $\mathscr{I}$, also called null infinity, is essentially a boundary added to the physical space-time at infinite distance in null directions. This $\mathscr{I}$ can be regarded as a manifold in its own right and tensor fields on the physical space-time give rise to tensor fields on $\mathscr{I}$. The tensor fields so produced are to be regarded in some sense as limits of the tensor fields on $(\widetilde{M}, \tilde{g}_{ab})$. This process has two parts: first, given a smooth tensor field defined on $\widetilde{M}$ we produce a smooth tensor field defined on M; second, given a smooth tensor field on M we produce a smooth tensor field on $\mathscr{I}$. Let $\tilde{\alpha}^{a\cdots b}{}_{c\cdots d}$ be a smooth tensor field defined on $\widetilde{M}$. If there exists a smooth tensor field $\alpha^{a\cdots b}{}_{c\cdots d}$ defined on all of M such that $\tilde{\alpha}^{a\cdots b}{}_{c\cdots d} = \alpha^{a\cdots b}{}_{c\cdots d}$ on $\widetilde{M}$, then $\tilde{\alpha}^{a\cdots b}{}_{c\cdots d}$ is said to be smoothly extendible to $\mathscr{I}$. Thus smoothly extendible tensor fields on $\widetilde{M}$ give rise to smooth tensor fields on M.

We now define a map ζ^* (also called pullback) from tensor fields on M to tensor fields on $\mathscr{I}$. This map acts on a subset $\mathfrak{C}$ of tensor fields on M. For a scalar field, $f, \zeta^*(f) = f|_{\Omega=0}$. For the exterior derivative of a scalar field, $\zeta^*(df) = d\zeta^*(f)$. Pullback of the sum (or outer

 0022-2488/87/010028-05$02.50

product) of tensor fields is the sum (or outer product) of their pullbacks. This defines ζ^* on all tensor fields of type $(0,p)$. For a general tensor field $\alpha^{a\cdots b}{}_{c\cdots d}$ consider the following equation:

$$\zeta^*(\alpha^{a\cdots b}{}_{c\cdots d}\beta_{a\cdots b}) = \zeta^*(\alpha^{a\cdots b}{}_{c\cdots d})\zeta^*(\beta_{a\cdots b}). \tag{10}$$

If the left-hand side of this equation vanishes whenever $\zeta^*(\beta_{a\cdots b}) = 0$, then Eq. (10) defines $\zeta^*(\alpha^{a\cdots b}{}_{c\cdots d})$; otherwise $\alpha^{a\cdots b}{}_{c\cdots d}$ is not in the set $\mathfrak{C}$ and cannot be pulled back.

Thus not all tensor fields on the physical space-time give rise to tensor fields on $\mathscr{I}$. First they must be smoothly extendible to $\mathscr{I}$, then their extensions must be in the set $\mathfrak{C}$. In particular, a smooth vector field defined on $\widetilde{M}$ can be pulled back if and only if it is smoothly extendible to $\mathscr{I}$ and its extension is tangent to $\mathscr{I}$. Here, ζ^* commutes with taking Lie derivatives, but not with contraction.

In what follows all indices will be raised and lowered with the unphysical metric g_{ab}. Define $n_a \equiv \nabla_a \Omega$. We obtain the following fields on $\mathscr{I}$:

$$\underline{g}_{ab} \equiv \zeta^*(g_{ab}), \tag{11}$$

$$\underline{n}^a \equiv \zeta^*(n^a), \tag{12}$$

$$\epsilon^{abc} \equiv \zeta^*(\epsilon^{abcd} n_d). \tag{13}$$

It is useful to define ϵ_{abc} as the unique antisymmetric field on $\mathscr{I}$ satisfying $\epsilon^{abc}\epsilon_{abc} = 6$. It follows from the conditions satisfied by Ω that $\underline{g}_{ab}\underline{n}^b = 0$ and that $\underline{g}_{ab}$ has signature $(0,+,+)$. Thus $\underline{g}_{ab}$ is degenerate and has no inverse. Nonetheless, it is useful to define the class of symmetric tensor fields $\underline{g}^{ab}$ satisfying

$$\underline{g}^{mn}\underline{g}_{am}\underline{g}_{bn} = \underline{g}_{ab}. \tag{14}$$

Members of this class differ by a tensor field of the form $v^{(a}\underline{n}^{b)}$.

The fields $\underline{g}_{ab}$ and $\underline{n}^a$ depend on the conformal factor Ω. In particular it is possible to choose Ω so that

$$\mathfrak{L}_{\underline{n}}\underline{g}_{ab} = 0. \tag{15}$$

Make this choice. The integral curves of $\underline{n}^a$ are called the generators of $\mathscr{I}$. The manifold of generators is called the base space B. A cross section of $\mathscr{I}$ is a two-dimensional submanifold which intersects each generator once. The physical space-time $(\widetilde{M},\tilde{g}_{ab})$ is said to be asymptotically Minkowskian if $\mathscr{I}$ is the manifold $S^2\times R$ and the generators are complete. From now on we will assume that $(\widetilde{M},\tilde{g}_{ab})$ is asymptotically Minkowskian. It is now possible to choose conformal factor Ω and coordinates u,θ, and ϕ on $\mathscr{I}$ such that

$$\underline{g}_{ab} = \partial_a\theta\,\partial_b\theta + \sin^2\theta\,\partial_a\phi\,\partial_b\phi, \tag{16}$$

$$\underline{n}^a = \left(\frac{\partial}{\partial u}\right)^a. \tag{17}$$

Make this choice. A tensor field $\Psi^{a\cdots b}{}_{c\cdots d}$ on $\mathscr{I}$ that satisfies $\mathfrak{L}_{\underline{n}}\Psi^{a\cdots b}{}_{c\cdots d} = 0$ and that yields zero when contracted on any index with $\underline{n}^a$ or $\partial_a u$ can be regarded as a tensor field on the two-sphere base space B. Thus, e.g., $\underline{g}_{ab}$ can be regarded as a positive definite metric on B.

Thus the fields defined so far on $\mathscr{I}$ are "kinematical"; they have been brought to a standard form by a choice of Ω. We now define the physical fields on $\mathscr{I}$. First assume that $(\widetilde{M},\tilde{g}_{ab})$ is a vacuum space-time. Then it can be shown (see Ref. 2) that the Weyl tensor C_{abcd} is zero on $\mathscr{I}$ and $\Omega^{-1}C_{abcd}$ is smooth. Define the following tensor fields on $\mathscr{I}$:

$$K^{ab} \equiv \epsilon^{amn}\epsilon^{bpq}\zeta^*(\Omega^{-1}C_{mnpq}), \tag{18}$$

$$*K^{ab} \equiv \epsilon^{amn}\epsilon^{bpq}\zeta^*(\tfrac{1}{2}\Omega^{-1}\epsilon_{mn}{}^{rs}C_{rspq}), \tag{19}$$

$$N_{ab} \equiv \underline{g}_{ac}\zeta^*(R^c{}_b - \tfrac{1}{6}R\delta^c{}_b) - \underline{g}_{ab}. \tag{20}$$

The tensor fields K^{ab}, $*K^{ab}$, and N_{ab} are symmetric. Define a derivative operator D_a on $\mathscr{I}$ by

$$D_a\zeta^*(\omega_b) = \zeta^*(\nabla_a\omega_b). \tag{21}$$

These are the physical fields on $\mathscr{I}$. They satisfy the following equations:

$$D_a\underline{n}^b = 0, \quad D_a\underline{g}_{bc} = 0, \tag{22}$$

$$N_{ab}\underline{n}^b = 0, \tag{23}$$

$$\underline{g}^{ab}N_{ab} = 0, \tag{24}$$

$$D_{[a}N_{b]c} = \tfrac{1}{4}\epsilon_{abm}*K^{mn}\underline{g}_{nc}, \tag{25}$$

$$D_aK^{ab} = 0, \tag{26}$$

$$K_a*K^{ab} = 0, \tag{27}$$

$$\underline{g}_{am}K^{mb} = -\epsilon_{amp}\underline{n}^p*K^{mb}, \tag{28}$$

$$\underline{g}_{am}*K^{mb} = \epsilon_{amp}\underline{n}^pK^{mb}. \tag{29}$$

Notice N_{ab} is determined by $*K^{ab}$ in the following sense: for a given $*K^{ab}$, N_{ab} is the unique symmetric tensor field satisfying Eqs. (23)–(25). Thus if $*K^{ab}\underline{g}_{bc} = 0$ then $N_{ab} = 0$.

Let C be a cross section of $\mathscr{I}$. Define the Bondi energy associated with C by

$$E(C) \equiv \frac{1}{8\pi}\int_C \epsilon_{abc}\left(\frac{1}{4}K^{am}l_m + (D_ml_n)\underline{g}^{np}N_{pq}\underline{g}^{q[m}\underline{n}^{a]}\right), \tag{30}$$

where l_a is any one form satisfying $\underline{n}^al_a = 1$. (Note that E does not depend on the choice of l_a or $\underline{g}^{ab}$.) Note that if $K^{ab} = *K^{ab} = 0$, then $E = 0$ for every cross section C. Let C_1 and C_2 be two cross sections of $\mathscr{I}$, where C_2 is to the future of C_1. Let A be the portion of $\mathscr{I}$ in between C_1 and C_2. Then

$$E(C_2) - E(C_1) = \frac{-1}{32\pi}\int_A (\underline{g}^{np}\underline{g}^{qm}N_{mn}N_{pq})\epsilon_{abc}. \tag{31}$$

We now state the result of this paper.

Theorem: Let $(\widetilde{M},\tilde{g}_{ab})$ be a space-time that is (i) asymptotically Minkowskian, (ii) vacuum, and (iii) has positive Bondi energy for every cross section of $\mathscr{I}$. Let ξ^a satisfy $\mathfrak{L}_\xi\,\tilde{g}_{ab} = -2q\,\tilde{g}_{ab}$. Then $q = 0$.

In other words, a space-time satisfying conditions (i)–(iii) admits no conformal Killing field that is not a Killing field.

III. PROOF

First we prove the following lemma.

Lemma 1: Let $(\widetilde{M},\tilde{g}_{ab})$ satisfy conditions (i)–(iii) and let ξ^a satisfy $\mathfrak{L}_\xi\tilde{g}_{ab} = -2q\tilde{g}_{ab}$. Then q is a constant.

Proof: Define $k_a \equiv -2\widetilde{\nabla}_aq$. Then Eq. (9) and condition (ii) imply

$$\widetilde{\nabla}_ak_b = 0. \tag{32}$$

Taking another derivative, antisymmetrizing, applying condition (ii), and using the conformal invariance of the Weyl tensor, we obtain

$$C_{abc}{}^{d}k_{d}=0, \tag{33}$$

where $C_{abc}{}^{d}$ is the Weyl tensor of g_{ab}. Condition (iii) implies that there is a point p at which $C_{abc}{}^{d}\neq 0$. It then follows from Eq. (33) that at point p, $\tilde{g}^{ab}k_{a}k_{b}=0$. Contracting Eq. (32) with $\tilde{g}^{bc}k_{c}$ we find that $\tilde{g}^{ab}k_{a}k_{b}$ is a constant. Thus

$$\tilde{g}^{ab}k_{a}k_{b}=0 \tag{34}$$

everywhere. Define $\tilde{k}^{a}\equiv\tilde{g}^{ab}k_{b}$. Then $\tilde{k}^{a}$ is a Killing field for $\tilde{g}_{ab}$. It then follows (see Ref. 2) that $\tilde{k}^{a}$ is smoothly extendible to $\mathscr{I}$ and is tangent to $\mathscr{I}$. Since $\tilde{k}^{a}$ is null, it follows that there is a smooth function γ such that

$$\lim_{\Omega\to 0}\tilde{k}^{a}=\gamma n^{a}. \tag{35}$$

Since $\Omega^{-1}C_{abcd}$ is smooth, it follows that

$$\lim_{\Omega\to 0}(\Omega^{-1}C_{abcd}\gamma n^{d})=0. \tag{36}$$

Thus at points of $\mathscr{I}$ where $\gamma\neq 0$,

$$\lim_{\Omega\to 0}(\Omega^{-1}C_{abcd}n^{d})=0. \tag{37}$$

It follows that at such points $K^{ab}={}^{*}K^{ab}=0$. Now assume $k_{a}\neq 0$. Then, since $\tilde{k}^{a}$ is a conformal Killing field for g_{ab}, the conformal Killing data for $\tilde{k}^{a}$ cannot vanish at a point. Thus $\tilde{k}^{a}$ cannot vanish in an open (three-dimensional) subset of a three-dimensional surface. Thus γ cannot vanish in any open (three-dimensional) subset of $\mathscr{I}$. Thus since K^{ab} and ${}^{*}K^{ab}$ are smooth, they must vanish everywhere on $\mathscr{I}$. Thus the Bondi energy is zero. But this contradicts condition (iii). Thus $k_{a}=0$, q is a constant.

We now show that ξ^{a} gives rise to a vector field on $\mathscr{I}$.

Lemma 2: Let $(\tilde{M},\tilde{g}_{ab})$ and ξ^{a} be as before. Then ξ^{a} has a smooth extension to $\mathscr{I}$ that is tangent to $\mathscr{I}$.

Proof: Since ξ^{a} is a conformal Killing field for $(\tilde{M},\tilde{g}_{ab})$, ξ^{a} is also a conformal Killing field for (M,g_{ab}), and ξ^{a} is thus smooth everywhere on M and thus automatically has a smooth extension to $\mathscr{I}$. Using Eq. (1) we obtain

$$\begin{aligned}-2q(\Omega^{-2}g_{ab})&\\ =\mathfrak{L}_{\xi}(\Omega^{-2}g_{ab})&\\ =\Omega^{-2}(\mathfrak{L}_{\xi}g_{ab}-2\Omega^{-1}g_{ab}\mathfrak{L}_{\xi}\Omega),&\end{aligned} \tag{38}$$

$$(2\Omega^{-1}\mathfrak{L}_{\xi}\Omega)g_{ab}=\mathfrak{L}_{\xi}g_{ab}+2qg_{ab}. \tag{39}$$

Since the right-hand side of Eq. (39) is smooth, $\Omega^{-1}\mathfrak{L}_{\xi}\Omega$ must be smooth. Thus $\mathfrak{L}_{\xi}\Omega=0$ on $\mathscr{I}$. Since $\nabla_{a}\Omega$ is the normal to $\mathscr{I}$, ξ^{a} is tangent to $\mathscr{I}$. This proves the lemma. Thus ξ^{a} satisfies the conditions necessary for a vector field to be pulled back to $\mathscr{I}$.

Define the following fields on $\mathscr{I}$:

$$\underline{\xi}^{a}\equiv\zeta^{*}(\xi^{a}), \tag{40}$$

$$\kappa\equiv\zeta^{*}(\Omega^{-1}\mathfrak{L}_{\xi}\Omega)-q. \tag{41}$$

Since ξ^{a} is a conformal Killing field we obtain

$$\mathfrak{L}_{\xi}C_{abc}{}^{d}=0. \tag{42}$$

Using ζ^{*} and Eqs. (1) and (42) we obtain

$$\mathfrak{L}_{\underline{\xi}}\underline{g}_{ab}=2\kappa\underline{g}_{ab}, \tag{43}$$

$$\mathfrak{L}_{\underline{\xi}}\underline{n}^{a}=(q-\kappa)\underline{n}^{a}, \tag{44}$$

$$\mathfrak{L}_{\underline{\xi}}K^{ab}=(q-5\kappa)K^{ab}. \tag{45}$$

$$\mathfrak{L}_{\underline{\xi}}{}^{*}K^{ab}=(q-5\kappa){}^{*}K^{ab}. \tag{46}$$

We assume that $q\neq 0$ and use Eqs. (43)–(46) to obtain a contradiction. Since both ξ^{a} and $-\xi^{a}$ are conformal Killing fields with opposite values of q, without loss of generality we now choose $q>0$.

First we examine the properties of the vector field $\underline{\xi}^{a}$ and its integral curves. Using our chosen coordinate system, we expand $\underline{\xi}^{a}$ in a coordinate basis

$$\underline{\xi}^{a}=A\left(\frac{\partial}{\partial u}\right)^{a}+B\left(\frac{\partial}{\partial\theta}\right)^{a}+C\left(\frac{\partial}{\partial\phi}\right)^{a}, \tag{47}$$

where A, B, and C are functions of u, θ, and ϕ. Using Eqs. (17) and (44) we obtain

$$A=(\kappa-q)u+f, \tag{48}$$

$$\frac{\partial B}{\partial u}=\frac{\partial C}{\partial u}=0, \tag{49}$$

where f is a function of θ and ϕ alone. Define the vector field η^{a} as

$$\eta^{a}\equiv\underline{\xi}^{a}-A\left(\frac{\partial}{\partial u}\right)^{a}, \tag{50}$$

then η^{a} is a vector field on the two-sphere. Equation (43) implies

$$\mathfrak{L}_{\eta}\underline{g}_{ab}=2\kappa\underline{g}_{ab}. \tag{51}$$

Since $\underline{g}_{ab}$ is, in our coordinates, the standard two-sphere metric, η^{a} is a conformal Killing field for the two-sphere. Define the scalar Z by

$$Z\equiv\underline{g}_{ab}\eta^{a}\eta^{b}. \tag{52}$$

Then Z is bounded, $Z\geqslant 0$, and $Z=0$ only at points where $\eta^{a}=0$. Contracting both sides of Eq. (51) with $\eta^{a}\eta^{b}$ we obtain

$$\mathfrak{L}_{\eta}Z=2\kappa Z. \tag{53}$$

Let $C(\lambda)$ be an integral curve of $\underline{\xi}^{a}$. Regard κ, u, f, and Z as functions of λ along the curve. Let $\tilde{C}(\lambda)$ be the corresponding curve on the two-sphere space of generators. Then $\tilde{C}(\lambda)$ is an integral curve of η^{a}.

Define the function $L(\lambda)$ by

$$L(\lambda)\equiv\int_{0}^{\lambda}\kappa(\lambda')d\lambda'. \tag{54}$$

The integral curves of η^{a} have the following property.

Lemma 3: Let $\tilde{p}$ be a point in the two-sphere and let $\tilde{C}(\lambda)$ be the integral curve of η^{a} for which $\tilde{C}(0)=\tilde{p}$. Then at least one of the following two conditions holds: (i) $L(\lambda)$ is bounded above, and/or (ii) $\tilde{C}(\lambda)=\tilde{p}$ for all λ. Furthermore, there are at most a finite number of points $\tilde{p}$ for which condition (i) fails to hold.

Proof: First treat the case where η^{a} is the zero vector field. Then $\kappa=0$, $L(\lambda)=0$ for all integral curves and the lemma is trivially true.

Next treat the case where η^{a} is not the zero vector field. First consider the case where $\eta^{a}\neq 0$ at point $\tilde{p}$. Then for all λ, the curve $\tilde{C}(\lambda)$ only passes through points where $\eta^{a}\neq 0$. Thus for all λ $Z(\lambda)>0$. Evaluating Eq. (53) on the curve $\tilde{C}(\lambda)$ and dividing by Z we obtain

$$\frac{d}{d\lambda}(\ln Z) = 2\kappa. \tag{55}$$

Integrating Eq. (55) we obtain

$$L(\lambda) = \tfrac{1}{2}\ln (Z(\lambda)/Z(0)). \tag{56}$$

Since $Z(\lambda)$ is bounded above, $L(\lambda)$ is bounded above.

Thus condition (i) can only fail to hold if $\eta^a = 0$ at point $\tilde{p}$. However, for a given two-sphere conformal Killing field η^a (other than the zero vector field), there are only a finite number of points where $\eta^a = 0$. Thus there are at most a finite number of points for which condition (i) fails to hold. If $\eta^a = 0$ at point $\tilde{p}$, then $\tilde{C}(\lambda) = \tilde{p}$ for all λ. This completes the proof of the lemma.

This property of integral curves of η^a allows us to derive the following property of integral curves of $\underline{\xi}^a$.

Lemma 4: Let p be a point in $\mathscr{I}$. Let G be the generator of $\mathscr{I}$ that contains p. Let $C(\lambda)$ be the integral curve of $\underline{\xi}^a$ for which $C(0) = p$. Then at least one of the following two conditions holds: (i) $L(\lambda)$ is bounded above and there exists a compact set $Q \subset \mathscr{I}$ such that for all $\lambda \geqslant 0$, $C(\lambda) \in Q$, and/or (ii) for all λ, $C(\lambda) \in G$. Furthermore, there are at most a finite number of generators G that contain points p at which condition (i) fails to hold.

Proof: Let $\tilde{C}(\lambda)$ be the two-sphere curve corresponding to $C(\lambda)$ and let $\tilde{p}$ be the two-sphere point corresponding to p. First treat the case where $L(\lambda)$ is bounded above. Applying Eqs. (47) and (48) to the curve $C(\lambda)$ we obtain

$$\frac{du}{d\lambda} = (\kappa - q)u + f. \tag{57}$$

The solution of Eq. (57) is

$$u(\lambda) = u(0)\exp[L(\lambda) - q\lambda] + \int_0^\lambda d\lambda' f(\lambda')\exp[J(\lambda,\lambda')]. \tag{58}$$

where $J(\lambda,\lambda')$ is given by

$$J(\lambda,\lambda') \equiv \int_{\lambda'}^{\lambda} d\lambda''(\kappa(\lambda'') - q). \tag{59}$$

Since $L(\lambda)$ is bounded above and f is a smooth function on the two-sphere, there exist positive constant M_1 and M_2 such that the following two inequalities are satisfied for $\lambda \geqslant 0$:

$$|u(0)|\exp[L(\lambda) - q\lambda] \leqslant M_1, \tag{60}$$

$$|f(\lambda)| \leqslant M_2. \tag{61}$$

It then follows that $u(\lambda)$ satisfies the following inequality for $\lambda \geqslant 0$:

$$|u(\lambda)| \leqslant M_1 + M_2\int_0^\lambda d\lambda' \exp[J(\lambda,\lambda')]. \tag{62}$$

We now show that there exist constants M_3 and α where $M_3 \geqslant 0$, $0 \leqslant \alpha < 1$, such that $J(\lambda,\lambda')$ satisfies the following inequality for $\lambda \geqslant 0$, $0 \leqslant \lambda' \leqslant \lambda$:

$$J(\lambda,\lambda') \leqslant (\alpha - 1)(q\lambda - q\lambda') + M_3. \tag{63}$$

First treat the case where $\kappa(\lambda'') \leqslant 0$ for all $\lambda'' \geqslant 0$. Then Eq. (63) is satisfied with $\alpha = M_3 = 0$. Next, treat the case where $\kappa(\lambda'') > 0$ for some $\lambda'' > 0$. Choose α such that $0 < \alpha < 1$ and the following two conditions are satisfied: (a) none of the points where $\eta^a = 0$ satisfy $\kappa = \alpha q$, and (b) for some $\lambda'' > 0$. $\kappa(\lambda'') > \alpha q$.

Let O be the region of the two-sphere where $\kappa \geqslant \alpha q$. Let λ_1 be a number satisfying $\lambda_1 > 0$ and $\tilde{C}(\lambda_1) \in O$. Then one of two conditions holds: either (a) for all λ'' satisfying $0 \leqslant \lambda'' \leqslant \lambda_1$, $\tilde{C}(\lambda'') \in O$; or (b) there exists λ_0 satisfying $0 < \lambda_0 \leqslant \lambda_1$ such that $\kappa(\lambda_0) = \alpha q$ and $\tilde{C}(\lambda'') \in O$ for all λ'' satisfying $\lambda_0 \leqslant \lambda'' \leqslant \lambda_1$. Note that $dL/d\lambda > 0$ in region O and $L(0) = 0$. Thus if condition (a) holds, then $L(\lambda_1) > 0$. If condition (b) holds, then $L(\lambda_1) \geqslant L(\lambda_0)$. Since none of the points where $\eta^a = 0$ satisfies $\kappa = \alpha q$, it follows that Z is bounded away from zero on the set of points in the two-sphere where $\kappa = \alpha q$. Thus $L(\lambda_0)$ is bounded below on the set of λ_0 satisfying $\kappa(\lambda_0) = \alpha q$. Thus $L(\lambda_1)$ is bounded below on the set of λ_1 satisfying $0 \leqslant \lambda_1$ and $\tilde{C}(\lambda_1) \in O$. Since $L(\lambda_1)$ is also bounded above, there exists a positive constant M_4 such that $|L(\lambda_1)| \leqslant M_4$ for all λ_1 satisfying $0 \leqslant \lambda_1$ and $\tilde{C}(\lambda_1) \in O$.

From Eq. (59) we obtain

$$J(\lambda,\lambda') = (\alpha - 1)q(\lambda - \lambda') + \int_{\lambda'}^{\lambda} d\lambda''(\kappa(\lambda'') - \alpha q). \tag{64}$$

For a given λ and λ' satisfying $\lambda \geqslant 0$, $0 \leqslant \lambda' \leqslant \lambda$; choose numbers λ_- and λ_+ as follows: if there is a λ'' satisfying $\lambda' \leqslant \lambda'' \leqslant \lambda$ and $\tilde{C}(\lambda'') \in O$; then let λ_- be the minimum such λ'' and let λ_+ be the maximum such λ''. If there is no such λ'', then choose $\lambda_- = \lambda_+ = \lambda$. In either case, note that the integrand in Eq. (64) is negative for all λ'' that satisfy $\lambda' \leqslant \lambda'' \leqslant \lambda$, but which do not satisfy $\lambda_- \leqslant \lambda'' \leqslant \lambda_+$. Thus using Eq. (64) we obtain the following inequality satisfied for all $\lambda \geqslant 0$, $0 \leqslant \lambda' \leqslant \lambda$:

$$\begin{aligned} J(\lambda,\lambda') &\leqslant (\alpha - 1)q(\lambda - \lambda') + \int_{\lambda_-}^{\lambda_+} d\lambda''(\kappa(\lambda'') - \alpha q) \\ &= (\alpha - 1)q(\lambda - \lambda') + L(\lambda_+) - L(\lambda_-) \\ &\quad - \alpha q(\lambda_+ - \lambda_-) \\ &\leqslant (\alpha - 1)q(\lambda - \lambda') + 2M_4. \end{aligned} \tag{65}$$

Thus Eq. (63) is satisfied with $M_3 = 2M_4$.

Using Eq. (63) in Eq. (62) we obtain for $\lambda \geqslant 0$

$$\begin{aligned} |u(\lambda)| &\leqslant M_1 + M_2\exp(M_3) \\ &\quad \times \int_0^\lambda d\lambda' \exp[(\alpha - 1)q(\lambda - \lambda')] \\ &= M_1 + \frac{M_2}{(1-\alpha)q}\exp(M_3) \\ &\quad \times [1 - \exp[(\alpha - 1)q\lambda]]. \end{aligned} \tag{66}$$

Define the constant M_5 by

$$M_5 \equiv M_1 + M_2 q^{-1}(1-\alpha)^{-1}\exp(M_3).$$

Then for $\lambda \geqslant 0$, $|u(\lambda)| \leqslant M_5$. Define the set Q by

$$Q \equiv \{\text{points } p \mid p \in \mathscr{I}, \quad |u(p)| \leqslant M_5\}, \tag{67}$$

then Q is compact and for all $\lambda \geqslant 0$, $C(\lambda) \in Q$. Thus if $L(\lambda)$ is bounded above then condition (i) of this lemma is satisfied.

If condition (i) of this lemma is not satisfied then the two-sphere point $\tilde{p}$ corresponding to the generator G does not satisfy condition (i) of Lemma 3. Thus $\tilde{C}(\lambda) = \tilde{p}$ for all λ. Thus $C(\lambda) \in G$ for all λ. Since there are at most a finite number of such two-sphere points $\tilde{p}$, there are at most a finite number of such generators G. This completes the proof of the lemma.

We are now ready to prove the theorem. We will use Eqs. (43)–(46) and Lemma 4 to show that for $q \neq 0$, the Bondi energy must vanish, thus establishing a contradiction with condition (iii). First define the scalar S by

$$S \equiv K^{ac} K^{bd} g_{ab} g_{cd}. \tag{68}$$

Then Eqs. (43) and (45) imply

$$\mathfrak{L}_{\xi} S = 2(q - 3\kappa) S. \tag{69}$$

Let p be a point in $\mathscr{I}$ for which condition (i) of Lemma 4 is satisfied. Let $C(\lambda)$ be the integral curve of ξ^a for which $C(0) = p$. Then the solution of Eq. (69) along the curve $C(\lambda)$ is

$$S(\lambda) = S(0)\exp[2q\lambda - 6L(\lambda)]. \tag{70}$$

For $\lambda \geqslant 0$, $C(\lambda)$ remains in a compact region. Thus, since S is smooth, $S(\lambda)$ is bounded for $\lambda \geqslant 0$. Since $L(\lambda)$ is bounded above, Eq. (70) then implies that $S(0) = 0$. Thus $S = 0$ at all points p for which condition (i) of Lemma 4 is satisfied. But since S is smooth, $S = 0$ everywhere.

Since $S = 0$, K^{ab} must be of the form

$$K^{ab} = 2\underline{n}^{(a} x^{b)}, \tag{71}$$

for some x^b. Applying the same proof to ${}^*K^{ab}$ we obtain

$${}^*K^{ab} = 2\underline{n}^{(a}\,{}^*x^{b)}, \tag{72}$$

for some ${}^*x^b$. Using Eq. (72) in Eq. (25) we obtain

$$\underline{n}^a D_{[a} N_{b]c} = 0. \tag{73}$$

Now applying Eq. (22) we obtain

$$\mathfrak{L}_{\underline{n}} N_{ab} = 0. \tag{74}$$

For each real number y let C_y be the cross section $u = y$ of $\mathscr{I}$; denote by $E(y)$ the Bondi energy associated with the cross section C_y and define $I(y)$ by

$$I(y) \equiv \frac{1}{8\pi} \int_{C_y} (g^{ac} g^{bd} N_{ab} N_{cd}) \underline{n}^e \epsilon_{efg}. \tag{75}$$

Then Eq. (74) implies that

$$\frac{dI}{dy} = 0. \tag{76}$$

Thus I is a constant. Equation (31) then implies

$$E(y_2) - E(y_1) = -\tfrac{1}{4} I(y_2 - y_1). \tag{77}$$

Thus if $I \neq 0$ then there exists a y such that $E(y) < 0$, i.e., there exists a cross section for which the Bondi energy is negative. But this contradicts condition (iii). Thus $I = 0$. Equation (23) then implies

$$N_{ab} = 0. \tag{78}$$

Equations (25) and (28) then imply

$${}^*K^{ab} = {}^*\beta \underline{n}^a \underline{n}^b, \tag{79}$$

$$K^{ab} = \beta \underline{n}^a \underline{n}^b, \tag{80}$$

for some scalar fields β and ${}^*\beta$. Equations (44) and (45) imply

$$\mathfrak{L}_{\xi} \beta = -(q + 3\kappa)\beta. \tag{81}$$

Equations (22) and (26) imply

$$\mathfrak{L}_{\underline{n}} \beta = 0. \tag{82}$$

Thus β is a function on the two-sphere and satisfies the equation

$$\mathfrak{L}_{\eta} \beta = -(q + 3\kappa)\beta. \tag{83}$$

Let $\tilde{p}$ be a point in the two-sphere for which condition (i) of Lemma 3 is satisfied. Let $\widetilde{C}(\lambda)$ be the integral curve of η^a for which $\widetilde{C}(0) = \tilde{p}$. Then the solution of Eq. (83) along the curve $\widetilde{C}(\lambda)$ is

$$\beta(\lambda) = \beta(0)\exp[-q\lambda - 3L(\lambda)]. \tag{84}$$

Since β is a smooth function on the two-sphere, $\beta(\lambda)$ is bounded for all λ. Since $L(\lambda)$ is bounded above, Eq. (84) then implies that $\beta(0) = 0$. Thus $\beta = 0$ at all points $\tilde{p}$ of the two-sphere for which condition (i) of Lemma 3 is satisfied. But since β is smooth, $\beta = 0$ everywhere. Similarly ${}^*\beta = 0$ everywhere. Thus $K^{ab} = {}^*K^{ab} = 0$. Thus the Bondi energy is zero. But this contradicts condition (iii) and thus proves the theorem. Space-times satisfying conditions (i)–(iii) of the theorem do not admit conformal Killing fields that are not Killing fields.

ACKNOWLEDGMENTS

I would like to thank Robert M. Wald, Robert Geroch, Carl Bender, and Doug Eardley for helpful discussions.

This research was supported in part by NSF Grant No. PHY 80-26043 to the University of Chicago and by NSF Grants No. PHY 83-13545 and No. PHY 85-13953 to Washington University.

[1]S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford U. P., New York, 1983).

[2]R. Geroch, in *Asymptotic Structure of Spacetime*, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977).

[3]A. Ashtekar and B. C. Xanthopoulos, J. Math. Phys. **19**, 2216 (1978).

[4]A. Ashtekar and B. G. Schmidt, J. Math. Phys. **21**, 862 (1980).

[5]J. Isenberg and V. Moncrief, unpublished.

[6]P. Yip, unpublished.

[7]D. Eardley, J. Isenberg, J. Marsden, and V. Moncrief, Commun. Math. Phys. **106**, 137 (1986).

[8]R. Geroch, Commun. Math. Phys. **13**, 180 (1969).

[9]It is an open question whether the requirement that the unphysical metric g_{ab} be C^∞ is "too strong"; i.e., whether it rules out a class of "physically reasonable isolated systems." See, e.g., J. Winicour, Found. Phys. **15**, 605 (1985). Nonetheless, for the purposes of this paper we assume that g_{ab} is C^∞.

Quantum theory on a regular tetrahedron

J. S. Dowker
Department of Theoretical Physics, The University of Manchester, Manchester, England

(Received 7 July 1986; accepted for publication 27 August 1986)

Quantum theory on the surface of a regular tetrahedron is discussed. The fact that the net of the tetrahedron tiles the plane allows the propagators to be constructed as image sums of standard ones. The effect of a constant magnetic field is discussed. The field theory Casimir energy is found.

I. INTRODUCTION

The study of the effects of the curvature and topology of space-time on quantum theories defined on that space-time is a well-established one.[1,2,3] The object of the present work is simply to present an investigation of a rather special and fixed situation but one that exemplifies certain general ideas and techniques.

First we assume that space-time is ultrastatic, i.e., that time is metrically separated from space. Second, we assume that space is d-dimensional and has a regular tetrahedron T as a section, i.e., that space takes the form $R^{\alpha-2} \times T$. We shall consider in detail only $d = 3$ and $d = 2$.

The tetrahedron is of interest because the curvature is concentrated at the vertices, which are conical singularities.[4] A single conical singularity, or wedge, has been considered previously[5,6,7] but the present example has the advantage of reducing to analysis on a space of finite volume. This makes the relevant eigenvalues discrete and allows integrated quantities to be discussed. The disadvantage is that the vertices all have a fixed deficit angle of π. It should be possible to discuss other polyhedra.

One motivation for studying quantum theories on (not in!) polyhedra is that such spaces arise in the $1 + 2$ theory of gravitation, or we could think of them as triangulation approximations to smooth manifolds in the manner of Regge's bones.

The particular fact that makes the regular tetrahedron so easy to deal with is that its net tiles the plane. Thus the Green's functions on T are image sums of plane Green's functions.

II. THE TETRAHEDRON

The net of a regular tetrahedron of edge length a is an equilateral triangle of side $2a$. The corresponding tiling of the plane is ancient and is illustrated in Fig. 1. The tetrahedron net is taken as the region OAB, the other triangles being equivalent copies. The unshaded triangles are obtained by discrete translations parallel to OA and OB through integer multiples of $2a$, while the shaded region can be obtained by first reflecting in the origin O and then translating. (Equivalent to reflection is a rotation through π.) The group Γ of this tiling is the non-Abelian, semidirect product of the cyclic group $C_2 = Z_2$ and the discrete translations $\mathbb{Z}^2$. Thus the images of a point (ξ,η) in OAB are $(\omega\xi + 2Ma, \omega\eta + 2Na)$, where $\omega = \pm 1$, $(M,N) \in \mathbb{Z}^2$, and ξ and η are oblique Cartesian coordinates. It is the non-Abelian, nonfree nature of the symmetry group that accounts for the vertex curvature (or vice versa), according to the Z_2 holonomy.

III. QUANTUM THEORY

Depending on whether we are considering quantum mechanics, statistical mechanics, or quantum field theory we are interested in propagators, partition functions, or the Green's functions.

There are two equivalent calculational methods—modes and image sums. Thus, for example, the propagator to go from the point (ξ,η) to the point (ξ',η') on the tetrahedron will be the sum of standard plane propagators from (ξ,η) to $(\omega\xi' + 2Ma, \omega\eta' + 2Na)$, $\forall\, M,N \in \mathbb{Z}$ and $\omega = \pm 1$.

By basic theory,[8] the distinct propagators we can obtain in this way are catalogued by $\mathrm{Hom}(\Gamma, U(1)) \sim Z_2$. [If Γ were just $\mathbb{Z}^2$, the propagators would be labeled by $2U(1)$, i.e., by two angles, but the presence of the reflection makes the Abelianized Γ equal to Z_2, and we know that $\mathrm{Hom}(Z_2, U(1)) \sim Z_2$. Whence the result just given.] The two propagators correspond to modes which are either periodic or antiperiodic ("twisted") under reflection.

Note that this approach does not produce the full freedom in the propagator allowed by the topology of the tetrahedron. The plane is not the universal covering space of T. In order to take the fundamental group T fully into account we should have to use not the standard plane propagator but the propagator in the presence of an infinite hexagonal lattice of Aharonov–Bohm flux lines through the points marked in

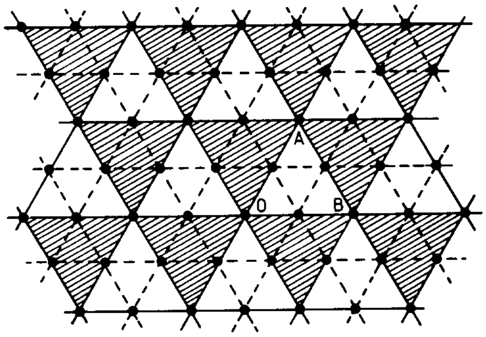

FIG. 1. The tiling of the plane by the net, OAB, of a regular tetrahedron. The lines, solid and dashed, are the edges and the dots are the vertices. Those at O, A, B, and their images, are to be identified. The shaded region is obtained from the fundamental zone OAB by first reflecting in the origin O and then translating. The unshaded region is obtained by just translating OAB.

 0022-2488/87/010033-06$02.50

Fig. 1 (the vertices of the tetrahedron). Some further remarks will be found in Sec. VII.

IV. MODES. ZETA FUNCTIONS

Whatever problem we are discussing we shall need the modes of the Laplace operator,

$$\Delta_2 = \frac{4}{3}\left(\frac{\partial^2}{\partial \xi^2} - \frac{\partial^2}{\partial \xi\, \partial \eta} + \frac{\partial^2}{\partial \eta^2}\right),$$

i.e., the solutions of

$$-\Delta_2 u = \lambda u .$$

By inspection, two sets of eigenfunctions are

$$u_+ = (2/\Omega)^{1/2}\cos(\pi/a)(m\xi + n\eta), \quad (m,n)\in \mathbb{Z}\times\mathbb{Z}^+ ,$$

and

$$u_- = (2/\Omega)^{1/2}\sin(\pi/a)(m\xi + n\eta) ,$$
$$(m,n)\in\mathbb{Z}\times\mathbb{Z}^+ \neq (0,0) .$$

with

$$\lambda = \lambda_{mn} = \tfrac{4}{3}(\pi^2/a^2)(m^2 - mn + n^2) .$$

The area of the tetrahedron is denoted by $\Omega = \sqrt{3}\, a^2$. The u_+ are untwisted and the u_- are twisted.

The integrated zeta functions are thus given by the Epstein function

$$\zeta_T(s) = \frac{1}{2}\left(\frac{3a^2}{4\pi^2}\right)^s \sum_{-\infty}^{\infty}{}' (m^2 - mn + n^2)^{-s} \qquad (1)$$

in both cases. This can be expressed in terms of "simpler" sums. Thus[9,10]

$$\zeta_T(s) = 3\left(\frac{3a^2}{4\pi^2}\right)^s \zeta_R(s) L_3(s) ,$$

where $\zeta_R(s)$ is the Riemann zeta function and $L_3(s)$ is the Dirichlet series

$$L_3(s) = 1 - 2^{-s} + 4^{-s} - 5^{-s} + 7^{-s} - 8^{-s} + \cdots .$$

It is amusing to note that $\zeta_T(s)$, for $s \geqslant 1.035$, is the minimum of those zeta functions defined in a similar fashion but with the quadratic form $am^2 + 2bmn + cn^2$, $ac - b^2 = \frac{3}{4}$, in place of $m^2 - mn + n^2$. This is related to closest packing and covering questions.[10]

For some purposes the nonlocal zeta functions are required. These are again Epstein,

$$\zeta_\pm(s;\xi',\eta'|\xi,\eta) = \frac{2}{\Omega}\sum_{-\infty}^{\infty}{}' \left.\begin{matrix}\cos\\ \sin\end{matrix}\right\}\frac{\pi}{a}(m\xi + n\eta) \left.\begin{matrix}\cos\\ \sin\end{matrix}\right\}\frac{\pi}{a}(m\xi' + n\eta')\lambda_{mn}^{-s} \qquad (2)$$

$$= \frac{2}{\Omega}\left[Z\left|\begin{matrix}0 & 0\\ (\xi-\xi')/2a & (\eta'-\eta)/2a\end{matrix}\right|(s)_\varphi \pm Z\left|\begin{matrix}0 & 0\\ (\xi+\xi')/2a & -(\eta'+\eta)/2a\end{matrix}\right|(s)_\varphi\right], \qquad (3)$$

in Epstein's notation[11] with $\varphi = m^2 - mn + n^2$.

Use of the transformation formula turns this eigenfunction expression into a classical paths form. Thus

$$\zeta_\pm(s;\xi'\eta'|\xi,\eta) = \frac{1}{3a^2\pi}\frac{\Gamma(1-s)}{\Gamma(s)}\left(\frac{3a^2}{4}\right)^s$$
$$\times\left[Z\left|\begin{matrix}(\xi-\xi')/2a & (\eta'-\eta)/2a\\ 0 & 0\end{matrix}\right|(1-s)_{\varphi^{-1}} \pm Z\left|\begin{matrix}(\xi+\xi')/2a & -(\eta+\eta')/2a\\ 0 & 0\end{matrix}\right|(1-s)_{\varphi^{-1}}\right], \qquad (4)$$

where φ^{-1} is the inverse form

$$\varphi^{-1} = 4(M^2 + MN + N^2)/3 ,$$

recognized as proportional to the distance between the origin and its (M,N)th image.

The second zeta functions on the right-hand sides of (3) and (4) are due to the reflection part of Γ. In the coincidence limit, $\xi' = \xi$, $\eta' = \eta$, they still give a position dependence indicating the inhomogeneous nature of the tetrahedron.

Finally in this section we note that the twisted modes vanish at the vertices.

V. QUANTUM MECHANICS ON THE TETRAHEDRON

The Hamiltonian for free particles is ($\hbar = 1$)

$$H = -(1/2\mu)\Delta_2 ,$$

and the propagator reads

$$K(t;\xi',\eta'|\xi,\eta) = \sum_i \psi_i(\xi',\eta')\psi_i^*(\xi,\eta)e^{-iE_it} ,$$

with $E_i = \lambda_{mn}/2\mu$ and the ψ_i equal to the u_+ or u_- modes.

Using Epstein's definition of the generalized theta function (see also Krazer[12]), the twisted and untwisted propagators can be written

$$K_\pm(t;\xi',\eta'|\xi,\eta) = \frac{1}{2\Omega}\left[\theta\left|\begin{matrix}0 & 0\\ (\xi-\xi')/2a & (\eta'-\eta)/2a\end{matrix}\right|\left(0,\frac{2\pi i t}{3\mu a^2}\right)_\varphi \pm \theta\left|\begin{matrix}0 & 0\\ (\xi+\xi')/2a & -(\eta+\eta')/2a\end{matrix}\right|\left(0,\frac{2\pi i t}{3\mu a^2}\right)_\varphi\right]. \qquad (5)$$

The theta function transformation formula yields expressions suitable for small t,

$$K_\pm(t;\xi',\eta'|\xi,\eta)=\frac{\mu}{2\pi it}\left[\theta\begin{vmatrix}(\xi-\xi')/2a & (\eta'-\eta)/2a\\ 0 & 0\end{vmatrix}\left(0,-\frac{3i\mu a^2}{2\pi t}\right)_{\varphi^{-1}}\right.$$
$$\left.\pm\theta\begin{vmatrix}(\xi+\xi')/2a & -(\eta+\eta')/2a\\ 0 & 0\end{vmatrix}\left(0,-\frac{3i\mu a^2}{2\pi t}\right)_{\varphi^{-1}}\right]. \tag{6}$$

This is the image sum of plane propagators referred to in Sec. III.

Equations (5) and (6) are related to (3) and (4) by Mellin transforms.

The integrated propagators follow directly from the eigenvalues of course, and are given by

$$K_\pm(t)=\frac{1}{2}\left[\theta\begin{vmatrix}0 & 0\\ 0 & 0\end{vmatrix}\left(0,\frac{2\pi it}{3\mu a^2}\right)_\varphi\pm 1\right] \tag{7}$$

$$=\frac{\Omega\mu}{2\pi it}\,\theta\begin{vmatrix}0 & 0\\ 0 & 0\end{vmatrix}\left(0,-\frac{3i\mu a^2}{2\pi t}\right)_{\varphi^{-1}}\pm\frac{1}{2}. \tag{8}$$

The small t expansion of (8) yields

$$K_\pm(t)\sim\frac{\mu}{2\pi it}\left(\Omega\pm\frac{\pi}{\mu}it\right)+O(e^{-1/t}), \tag{9}$$

in which we recognize the first term as the Weyl volume contribution. The second term agrees with the "corner terms"[13] (suitably adjusted for periodic boundary conditions).

The replacement $t\to -i/kT$ turns $K(t)$ into the quantum mechanical partition function and (9) then gives the high temperature expansion.

VI. MAGNETIC INTERACTIONS

The study of magnetic fields in two-dimensional systems is an important one and we wish to examine the case when a uniform magnetic field is applied perpendicularly to the plane of Fig. 1. For the tetrahedron this means that a uniform field H passes perpendicularly through each face. For consistency, as will be seen, we shall have to apply the Dirac quantization rule to the total flux, i.e.,

$$\Omega H=phc/e,\quad p\in\mathbb{Z}. \tag{10}$$

Thus we shall not have the luxury of being able to treat H as a continuous variable. We note that p labels the first Chern classes $H^2(T;\mathbb{Z})$ of the $U(1)$-bundles over T. Instead of beginning with the modes we shall construct an image sum analogous to (6).

The propagator on the plane in the presence of a uniform magnetic field is well known and we can simply transcribe the standard expression[14] to oblique coordinates to obtain

$$K_H(t;\xi'\eta'|\xi,\eta)=\frac{eH}{4\pi i\sin\frac{1}{2}\nu t}\exp\left[\frac{ieH}{4}\{((\xi'-\xi)^2+(\xi'-\xi)(\eta'-\eta)+(\eta'-\eta)^2)\cot\tfrac{1}{2}\nu t-\sqrt{3}(\xi'\eta-\xi\eta')\}\right], \tag{11}$$

with $\nu=eH/\mu$, $c=1$, and e is the charge on the particle.

The vector potential is the "symmetrical" one $\mathbf{A}=-\frac{1}{2}H(y,-x)$, in rectangular coordinates, and, unlike the magnetic field, does not possess the symmetry of the lattice. There are then well-known problems regarding the setting up of quantum mechanics.[15] Compensating gauge transformations are necessary.[16]

In order that the resulting "magnetic translations" (and, here, reflections) should induce a true representation in function space, rather than a ray representation, it is sufficient to impose the flux quantization (10). The general results of Ref. 8 are then immediately applicable.

If a point of the plane R^2 is denoted by q and the action of Γ by $q\gamma$ ($\gamma\in\Gamma$), then, instead of being periodic or antiperiodic, the wave function on R^2 satisfies

$$\psi(q\gamma)=a(\gamma)e^{i\Lambda_\gamma(q)}\psi(q), \tag{12}$$

where $\Lambda_\gamma(q)$ is a compensating gauge transformation and $a\in\mathrm{Hom}(\Gamma,U(1))$. In the present instance $a(\gamma)=\pm 1$.

Suitably interpreted (12) is valid for more general situations.

For a constant magnetic field

$$\Lambda_\gamma(q)=\tfrac{1}{2}\omega aeH\sqrt{3}(M\eta-N\xi),$$

where $q=(\xi,\eta)$ and $q\gamma=(\omega\xi+2Ma,\omega\eta+2Na)$. Introducing dimensionless coordinates $\bar\xi=\xi/a$ and $\bar\eta=\eta/a$ we find that when using (10),

$$\Lambda_\gamma(q)=\omega\pi p(M\bar\eta-N\bar\xi). \tag{13}$$

Iterating (12) produces no inconsistency and the magnetic group is isomorphic to the ordinary one.[15,17] This can be considered as another proof of the Dirac quantization rule—the so-called "integrality condition." Applying the general theory,[8] the propagator on the tetrahedron is

$$K_T(q',q)=\sum_\gamma a^{-1}(\gamma)e^{-i\Lambda_\gamma(q')}K_H(q'\gamma,q)$$
$$=\sum_\gamma K_H(q',q\gamma)a(\gamma)e^{i\Lambda_\gamma(q)}. \tag{14}$$

Written out explicitly in terms of theta functions, we find, from (11), (13), and (14),

$$K_T(t;\xi',\eta'|\xi,\eta)=\frac{p}{2\Omega i\sin\frac{1}{2}\nu t}\left[\exp\left\{\frac{1}{2}i\pi p(\xi\eta'-\xi'\eta)\right\}\right.$$
$$\times\theta\begin{vmatrix}\frac{1}{2}(\xi'-\xi) & -\frac{1}{2}(\eta'-\eta)\\ -\frac{1}{2}p(\eta+\eta') & \frac{1}{2}p(\xi+\xi')\end{vmatrix}\left(0,-\frac{\sqrt{3}ip}{2}\cot\tfrac{1}{2}\nu t\right)_{\varphi^{-1}}$$

$$\pm \exp\left\{-\frac{1}{2} i\pi p(\xi\eta' - \xi'\eta)\right\} \theta \begin{vmatrix} -\frac{1}{2}(\xi+\xi') & -\frac{1}{2}(\eta+\eta') \\ \frac{1}{2}p(\eta'-\eta) & \frac{1}{2}p(\xi-\xi') \end{vmatrix} \left(0, -\frac{\sqrt{3}ip}{2}\cot\tfrac{1}{2}\nu t\right)_{\varphi^{-1}} \Bigg], \tag{15}$$

where here, and from now on, we drop the bars over the coordinates and where $\nu = 2\pi p/\Omega\mu$ and $\varphi^{-1} = 4(M^2 + MN + N^2)/3$.

The theta function transformation formula can be used to rewrite (15) in eigenfunction form. However we restrict attention to the coincidence limit $K(t;\xi,\eta|\xi,\eta)$ since it is easier to handle. From (15),

$$K(t;\xi,\eta|\xi,\eta) = \frac{p}{2\Omega i \sin\frac{1}{2}\nu t}\left[\theta \begin{vmatrix} 0 & 0 \\ -p\eta & p\xi \end{vmatrix} \left(0, -\frac{\sqrt{3}ip}{2}\cot\frac{1}{2}\nu t\right)_{\varphi^{-1}} \pm \theta \begin{vmatrix} -\xi & -\eta \\ 0 & 0 \end{vmatrix} \left(0, -\frac{\sqrt{3}ip}{2}\cot\frac{1}{2}\nu t\right)_{\varphi^{-1}}\right]. \tag{16}$$

We further wish to integrate this quantity over the tetrahedron. Because of the symmetry of the integrand it is possible to extend the integration over the entire parallelogram $0 \leqslant \xi \leqslant 2$, $0 \leqslant \eta \leqslant 2$ with the penalty of a factor of 2. Then we note that this integration is most easily performed for those theta functions in which ξ and η occur on the bottom line so that it is calculationally advantageous to transform the second term in (16) to give

$$K(t;\xi,\eta|\xi,\eta) = \frac{1}{2\Omega}\left[\frac{p}{i\sin\frac{1}{2}\nu t}\theta \begin{vmatrix} 0 & 0 \\ -p\eta & p\xi \end{vmatrix} \left(0, -\frac{\sqrt{3}ip}{2}\cot\tfrac{1}{2}\nu t\right)_{\varphi^{-1}} \pm \frac{1}{\cos\frac{1}{2}\nu t}\theta \begin{vmatrix} 0 & 0 \\ \xi & \eta \end{vmatrix} \left(0, \frac{i2}{\sqrt{3}p}\tan\tfrac{1}{2}\nu t\right)_{\varphi}\right],$$

where $\varphi = m^2 - mn + n^2$. From the definition of the theta function, which we repeat here,

$$\theta \begin{vmatrix} 0 & 0 \\ \xi & \eta \end{vmatrix} (0,z)_\varphi = \sum_{-\infty}^{\infty} \exp[-\pi z(m^2 - mn + n^2) + 2\pi i(m\xi + n\eta)],$$

it can be seen that integration over ξ and η picks out the zero mode, i.e., the $m = n = 0$ term. Whence we find for the integrated propagator

$$K(t) = \tfrac{1}{2}(-pi\csc\tfrac{1}{2}\nu t \pm \sec\tfrac{1}{2}\nu t), \quad p \neq 0. \tag{17}$$

If $p = 0$ then we regain (8) instead.

It is a simple matter to obtain the eigenvalues of the Hamiltonian and their degeneracies from (17). We find the energy levels $E_k = (k + \frac{1}{2})\nu, k = 0,1,2,\ldots$, and the degeneracies $d_k = p \pm (-1)^k, p > 0$.

Without the flux quantization we would have found nonintegral degeneracies. This is another way of looking at the possible inconsistency in the quantum theory on a compact region.

This result is not unexpected because it is standard[18] that the area density of states in a given Landau level is $eH/2\pi$. With (10) this yields the above degeneracy for the tetrahedron region, apart from the reflection term $\pm(-1)^k$. Saying this again, the first term of (17) is obtained by integrating the coincidence limit of the plane propagator, (11), over a region Ω and then using (10). Thus, insofar as the integrated quantities go, the only novelty that the tetrahedron has to offer over any given region of the plane is the reflection contribution in (17). This will certainly have an effect on, say, the statistical mechanics, especially for small p. However, the main distinguishing property of the tetrahedron expressions is their position dependence the study of which is perhaps better appreciated in quantum field theories were local quantities are more freely available.

VII. QUANTUM FIELD THEORY. THE CASIMIR EFFECT

We approach this topic via the Green's functions of the free massless scalar field constructed as image sums. In order to have something to compute we shall concentrate on the Casimir effect and shall evaluate the vacuum average of the energy density $\langle \hat{T}_{00} \rangle$, a local quantity, and also the vacuum average of the total energy (the Hamiltonian) $E = \langle \hat{H} \rangle$.

Explicit expressions will be given for the cases of two- and three-dimensional space. The former for simplicity and the latter because we hope to make contact with the single wedge result mentioned earlier.

General theory[1,19] gives $\langle \hat{T}_{\mu\nu} \rangle$ as the coincidence limit of a differential operator acting on the Feynman Green's function D. In flat space-time (and rectangular coordinates)

$$\langle \hat{T}_{00}(\mathbf{x}) \rangle = i \lim_{\mathbf{x}' \to \mathbf{x}} [(2\epsilon + \tfrac{1}{2})\partial_0 \partial_{0'} + (2\epsilon - \tfrac{1}{2})\partial_i \partial_{i'}] D(\mathbf{x},\mathbf{x}'), \tag{18}$$

where ϵ is a parameter that, in order to give the improved energy-momentum tensor, equals $\frac{1}{6}$ in four-dimensional space-time and $\frac{1}{8}$ in three. These are the values we use.

In four-dimensional space-time D is given as the image sum[20] of standard Feynman Green's functions,

$$D(\mathbf{x},\mathbf{x}') = -\frac{i}{4\pi^2} \sum_{M,N=-\infty}^{\infty} \left(\frac{1}{\sigma^2_{MN}} \pm \frac{1}{(\sigma^R_{MN})^2}\right),$$

where

$$\sigma^2_{MN} = (t-t')^2 - [(\xi - \xi' - 2Ma)^2 + (\xi - \xi' - 2Ma)(\eta - \eta' - 2Na) + (\eta - \eta' - 2Na)^2] - (z-z')^2, \tag{19}$$

and the reflected interval is

$$(\sigma^R_{MN})^2 = (t-t')^2 - [(\xi + \xi' - 2Ma)^2 + (\xi + \xi' - 2Ma)(\eta + \eta' - 2Na)$$

$$+ (\eta + \eta' - 2Na)^2] - (z - z')^2 . \qquad (20)$$

Our notation is such that $\mathbf{x} = (\xi,\eta,z)$, $\mathbf{x}' = (\xi',\eta',z')$.

The differential operator in (18) is given by

$$\frac{i}{6}\left(5\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial z\,\partial z'} - \Delta_2^p\right),$$

where

$$\Delta_2^p = \frac{2}{3}\left(2\frac{\partial^2}{\partial\xi\,\partial\xi'} - \frac{\partial^2}{\partial\eta\,\partial\xi'} - \frac{\partial^2}{\partial\eta'\,\partial\xi} + 2\frac{\partial^2}{\partial\eta\,\partial\eta'}\right)$$

is the polarized Laplacian on the tetrahedron (or the plane).

Calculation produces the Epstein forms

$$\langle\hat{T}_{00}(\mathbf{x})\rangle = -\frac{1}{96\pi^2a^4}\left[3\sum_{M,N}{}'\frac{1}{(M^2+MN+N^2)^2} \pm \sum_{M,N}\frac{1}{[(\xi-M)^2+(\xi-M)(\eta-N)+(\eta-N)^2]^2}\right]. \qquad (21)$$

The $M = N = 0$ term in the first summation has been omitted as usual. It corresponds to the standard ultraviolet divergence on Minkowski space.[20] There is a divergence, which we can do nothing about, coming from the second term as $\mathbf{x}$ approaches a vertex. We easily find that as $r \to 0$ (r is the distance from the vertex)

$$\langle\hat{T}_{00}(\mathbf{x})\rangle \sim -1/96\pi^2r^4 ,$$

for untwisted fields, in agreement with the result[5,6] for a single wedge of angle π, again adjusted for periodic boundary conditions.

Obviously if $\langle\hat{T}_{00}(\mathbf{x})\rangle$ is integrated over all the tetrahedron to give the total energy, per unit z interval, the answer will be infinite due to the vertex effects. As described elsewhere,[5,21] an alternative procedure is to average the Hamiltonian, that is, to perform the integration first and then regularize/renormalize. Apart from any ultraviolet divergence this gives a finite result.

Standard theory[5,21] yields the expression for the total vacuum energy per unit z slice,

$$E = (1/8\pi)\lim_{s\to 1}\Gamma(s-2)\zeta_T(s-2) , \qquad (22)$$

where $\zeta_T(s)$ is the tetrahedron zeta function (1), which, in Epstein's notation, is

$$\zeta_T(s) = \frac{1}{2}\left(\frac{3a^2}{4\pi^2}\right)^s Z\begin{vmatrix}0 & 0\\ 0 & 0\end{vmatrix}(s)_\varphi .$$

For the numerical evaluation of E it is generally easiest to leave E as in (22) because the right hand side is given by Eq. (3) on p. 207 of Epstein[11] and this equation is ideally suited to rapid calculation in terms of the incomplete gamma function. We find

$$a^2E \doteqdot -0.0423 .$$

An application of the transformation formula to the expression for E yields

$$E = -\frac{1}{6\sqrt{3}a^2\pi^2}Z\begin{vmatrix}0 & 0\\ 0 & 0\end{vmatrix}(2)_{\varphi^{-1}} ,$$

which is identical to the integral of the first, constant term in (21) over the tetrahedron. Thus it appears that if we integrate before averaging, the zeta function method drops the *entire* reflected term. We have no feelings either way for the justification or significance of this procedure.

The analysis can be repeated for the three-dimensional space-time whose spatial section is the tetrahedron. The operator in (18) is now

$$\frac{i}{4}\left(3\frac{\partial^2}{\partial t^2} - \Delta_2^p\right),$$

but the Feynman Green's function is not so simply expressed. It is easiest to leave it as a proper-time integral,

$$D(\mathbf{x},\mathbf{x}') = \int_0^\infty \frac{i\,d\tau}{(-4\pi i\tau)^{3/2}}\sum_{M,N}\left[\exp\left(-\frac{i}{4\tau}\sigma_{MN}^2\right) \pm \exp\left(-\frac{i}{4\tau}(\sigma_{MN}^R)^2\right)\right],$$

where the space-time intervals are as in (19) and (20) except there are no $(z-z')^2$ terms. The integration over τ is left until the differentiations and coincidence limits have been taken. We find

$$\langle\hat{T}_{00}(\mathbf{x})\rangle = -\frac{1}{32\pi a^3}\left\{\sum_{M,N}{}'\frac{1}{(M^2+MN+N^2)^{3/2}} \pm \frac{1}{2}\sum_{M,N}\frac{1}{[(\xi-M)^2+(\xi-M)(\eta-N)+(\eta-N)^2]^{3/2}}\right\}. \qquad (23)$$

Near the vertex the behavior is again divergent,

$$\langle\hat{T}_{00}(\mathbf{x})\rangle \sim \mp 1/64\pi r^3 ,$$

and, as before, the total vacuum energy can be calculated by the zeta function method. This time

$$E = \frac{1}{2}\zeta_T(-\tfrac{1}{2}) = \frac{\pi}{2a\sqrt{3}}Z\begin{vmatrix}0 & 0\\ 0 & 0\end{vmatrix}\left(-\frac{1}{2}\right)_\varphi \doteqdot -\frac{0.19}{a} .$$

Again it is easy to show that this equals the integral of the first term in (23).

VIII. COMMENTS AND CONCLUSION

Although the particular numbers obtained in this paper may not have much practical use, the calculation does suggest a few interesting questions.

The two-sphere and the tetrahedron, including its vertices, are homeomorphic (they have the same Euler number) but of course are metrically different so that we would not expect them to have the same total Casimir energy. In fact for conformally invariant fields this energy is zero on S^2 (and on all even spheres).

We would not expect them, either, to have the same coefficient of *it* in the short-time expansion of the quantum mechanical propagator [cf. (9)]. Only for smooth manifolds (with or without boundaries) is this coefficient a topological invariant ($=2\pi\chi/3$, where χ is the Euler number). For manifolds whose boundaries contain corners there are nontopologically invariant terms.[13] This raises the question: What is the short-time expansion of the propagator on a general polyhedron? This would be of interest when discussing quantum field theory on those two-spaces whose curvature is concentrated on points.

It is easily proved using Carslaw's wedge results[7,22] that the contribution to the coefficient of *it* of a vertex of deficit angle $2\pi-\beta$ is $(4\pi^2-\beta^2)/6\beta$ (for $\mu=\frac{1}{2}$).

When the deficit angles α_i (i labels the vertices) become infinitesimally small,

$$\sum_i \frac{4\pi^2-\beta_i^2}{6\beta_i}=\sum_i \frac{1}{6}\alpha_i\left(\frac{4\pi-\alpha_i}{2\pi-\alpha_i}\right)$$

$$\to\frac{1}{3}\sum_i \alpha_i=\frac{4\pi}{3}=\frac{2\pi\chi}{3}=\frac{1}{6}\int_{\mathscr{M}} R(g)^{1/2}d^2\mathbf{x},$$

which is the standard coefficient for the smooth, limiting surface $\mathscr{M}$, diffeomorphic to the two-sphere.[23]

One might expect the higher terms in the expansion to vanish if only because they do so for the polygonal membrane.[24] The tetrahedron result (9) confirms this.

We might even extend this termination to simplicial approximations in d dimensions and speculatively say that only the first d terms can possibly appear.

Another point concerns the classification of propagators on the tetrahedron. General theory[3,25] says that the distinct propagators are in one to one correspondance with the elements of $\mathrm{Hom}(H_1,U(1))$ where H_1 is the first homology group of the configuration space. For the tetrahedron $H_1=3\mathbb{Z}$. Hence the propagators should be labeled by three angles.

In order to give these angles a "physical" meaning, imagine the particle to be charged and magnetic flux tubes to emerge from each vertex. The change of phase of the wave function when circling the tetrahedron will be $2\pi\Sigma_1^4 m_i\delta^i$, where n_i is the winding number, around the ith vertex, of the path and $\delta^i=e\Phi^i/hc$, where Φ^i is the flux through the ith vertex. Since the total flux is quantized $e\Sigma\Phi^i=phc$, $p\in\mathbb{Z}$, and only three δ parameters are independent. They can be taken to lie between 0 and $\frac{1}{2}$. Multiplied by 2π these are the three angles mentioned above.

The antiperiodic case discussed in Secs. III and IV corresponds to all the δ^i being $\frac{1}{2}$. Any other distribution of the δ^i, apart from all zero, requires the solution of Schrödinger's equation in an infinite lattice of flux tubes, an interesting problem in its own right.

Finally, we remark that it is not difficult to relax the regularity condition so long as the deficit angles remain equal to π.

[1] N. Birrell and P. C. W. Davies, *Quantum Fields on Curved Space* (Cambridge U. P., Cambridge, 1982).

[2] S. J. Avis and C. J. Isham, in *Recent Developments in Gravitation. Cargèse Lectures* (Plenum, New York, 1978); G. W. Gibbons, in *General Relativity*, edited by S. W. Hawking and W. Israel (Cambridge U. P., Cambridge, 1979).

[3] J. S. Dowker, "Selected topics in topology and quantum field theory," lectures Delivered at Center for Relativity, Austin, Texas, 1979.

[4] A. D. Aleksandrow, *Die innere Geometrie der Konvexen Flächen* (Akademie, Berlin, 1955); D. D. Sokolov and A. A. Starobinsky, Dok. Akad. Nauk. SSSR **22**, 312 (1977); T. Banchoff, Am. Math. Mon. **77**, 475 (1970); H. Gluck, K. Krigelmann and D. Singer, J. Diff. Geom. **9**, 601 (1974); J. C. Maxwell, Trans. Cambridge Philos. Soc. **IX**, pt. IV (1855).

[5] J. S. Dowker and G. Kennedy, J. Phys. A **11**, 895 (1978).

[6] D. Deutsch and P. Candelas, Phys. Rev. D **20**, 3063 (1979).

[7] J. S. Dowker, J. Phys. A **10**, 115 (1977); J. S. Dowker, J. Phys. A **18**, 3521 (1985).

[8] M. G. G. Laidlaw and C. DeWitt, Phys. Rev. D **3**, 1375 (1971); L. S. Schulman, J. Math. Phys. **12**, 304 (1971); J. S. Dowker, J. Phys. A **5**, 936 (1972).

[9] I. J. Zucker and M. M. Robertson, J. Phys. A **8**, 874 (1975).

[10] R. A. Rankin, Proc. Glasgow Math. Ass. **1**, 149 (1953).

[11] P. Epstein, Math. Ann. **63**, 205 (1907).

[12] A. Krazer, *Lehrbuch der Thetafunktionen* (Teubner, Leipzig, 1903).

[13] H. P. Baltes and E. R. Hilf, *Spectra of Finite Systems* (Bibliographisches Institut, Mannheim, 1976); S. A. Molchanov, Russ. Math. Surveys **30**, 1 (1975).

[14] E. H. Sondheimer and A. H. Wilson, Proc. R. Soc. London Ser. A **210**, 173 (1951); M. I. Glasser, Phys. Rev. **133**, B 831 (1964); L. M. Falicov and H. Stachowiak, Phys. Rev. **147**, 505 (1966).

[15] J. Zak, Phys. Rev. **134**, A 1602, A 1607 (1964); Phys. Rev. **136**, A 776 (1964); E Brown, Phys. Rev. **133**, A 1038 (1964).

[16] A. Janner and T. Janssen, Physica **53**, 1 (1971); L. Jansen and M. Boon, *Theory of Finite Groups* (North-Holland, Amsterdam, 1967); L. S. Schulman, Phys. Rev. **118**, 1139 (1969).

[17] W. Opechowski and W. G. Tam, Physica **42**, 529 (1969).

[18] R. E. Peierls, *Theory of Solids* (Clarendon, Oxford, 1955).

[19] B. S. DeWitt, Phys. Rep. C **19**, No. 6 (1975).

[20] J. S. Dowker and R. Critchley, J. Phys. A **9**, 535 (1976).

[21] J. S. Dowker and R. Banach, J. Phys. A **11**, 2255 (1978).

[22] H. S. Carslaw, Proc. London Math. Soc. **8**, 365 (1909). It is interesting historically to note that M. Kac [Am. Math. Mon. **73**, 1 (1969)] uses Carslaw's results in his derivation of the corner terms and is unaware of the work of B. V. Fedosov [Dok. Akad. Nauk. SSSR **151**, 786 (1963); **157**, 536 (1964)] who himself refers to a paper by G. Herglotz [Math. Ann. **124**, 219 (1952)] where many of Carslaw's formulas are rederived; J. Cheeger [J. Differential Geom. **18**, 575 (1983)] has given a formula for the constant term on the generalized cone $R^+\times N$, in terms of the spectral invariants on the Riemannian manifold N, that reduces to known results when N is a circle or an interval.

[23] Compare this with the discussion in A. N. Chaba and R. K. Pathria, Phys. Rev. A **8**, 3264 (1973) who use a particular approximation to the two-sphere.

[24] P. B. Bailey and F. H. Brownell, J. Math. Anal. Appl. **4**, 212 (1962). This termination means, somewhat surprisingly, that the expansion for a smooth boundary, like the circle, cannot be obtained as the limit of that for a polygonal boundary, except for the first three terms. Presumably the exponentially small terms conspire to produce the higher smooth terms.

[25] P. A. Horváthy, Lect. Notes Math. **905**, 197 (1982).

Spectral function methods for nonlinear diffusion equations

G. Delic[a)]

Nuclear Theory Research Group, Department of Physics and Schonland Research Center for Nuclear Sciences, University of the Witwatersrand, 1 Jan Smuts Avenue, Johannesburg 2001, South Africa

(Received 5 June 1986; accepted for publication 10 September 1986)

Two spectral function methods are developed for linear and nonlinear diffusion equations in one dimension where the nonlinearity is in the inhomogeneous term and occurs as a power of the solution. In the single spectral function method polynomial spectral functions in the spatial variable are introduced. The spectral resolution of the diffusion equation in the Hilbert space spanned by these functions yields a system of ordinary differential equations which is then integrated in discrete steps of the time variable. The double spectral method introduces polynomial spectral functions in both space and time variables and thereby eliminates the need for time integration through application of an iterative algorithm. Both methods are compared against analytical solutions for the linear cases and against the numerical solutions for the nonlinear cases. The second spectral function method was found to be more efficient than the first by a factor of 6 in the case of nonlinear problems.

I. INTRODUCTION

Traditionally the diffusion equation is encountered in statistical mechanics[1] where it occurs together with the Fokker–Planck[2] equation in the description of Brownian motion of microscopic particles. A comprehensive historical review of the diffusion equation and the relationship to stochastic processes is to be found in the essay by Montroll and Schlesinger.[3] An exposition on the diffusion equation, solution methods, and applications is given by Crank.[4] From the point of view of numerical techniques typical methods of solution are described in Refs. 5 and 6. The usual numerical approach to the solution of the diffusion equation is to discretize in either space or time degrees of freedom or in both. However, for large scale problems in several spatial dimensions such discretization methods become expensive in terms of computing resources. For this reason research into more efficient numerical methods continues and is described, for example, by Ghoniem and Sherman[7] or Rektorys.[8]

One alternative to discretization in space and time variables is the so-called spectral function method. Spectral function methods[9] provide concise and accurate approximations to operator equations characteristic of initial or boundary value problems. As an example, the method has enjoyed continued success in applications to fluid flow problems.[10] Interest in the method was revived by a series of key papers by Orszag and collaborators.[9,11] The spectral function method may be thought of as an L_2 projection of the operator equation of interest onto a finite dimensional Hilbert space spanned by a set of polynomial "coordinate"[12] or spectral functions. Given an L_2 norm it then becomes possible to apply the underlying powerful functional analytic theory as developed by Kantorovich and Akilov.[13]

A spectral function method was previously applied to linear Schrödinger operator equations characteristic of quantum scattering theory.[14,15] In the present study the method is developed for equations describing linear and nonlinear diffusion in one spatial dimension. Although only model cases are discussed in this communication, an application to a real-life problem in theoretical plasma physics is under study.[16]

Two spectral function methods are developed in the present work. In the first of these polynomial spectral functions in the spatial variable are introduced. This leads to a system of first-order ordinary differential equations, which is then integrated in discrete steps of the time variable. The second spectral method consists of polynomial spectral functions in both space and time variables and thereby eliminates the need for time integration. Both methods are compared against analytical solutions for the linear cases and against the numerical solutions of Rektorys[8] for the nonlinear cases.

This communication is divided into six sections. Section II introduces the two spectral function methods and the problems to which they are applied while Sec. III describes the extensions required for application to the nonlinear cases. Sections IV–VI give the results of numerical experiments using both methods for linear and nonlinear problems and Sec. VII summarizes conclusions.

II. TWO SPECTRAL FUNCTION METHODS

A. Statement of the problem and method

The diffusion equation studied here in one spatial dimension takes the form

$$\partial_t u = \partial_{rr} u + Q(r,t,u) , \tag{1}$$

where ∂_t denotes one derivative with respect to time t and ∂_{rr} denotes two derivatives with respect to one spatial degree of freedom. The domain of definition for the solution $u(r,t)$ is $r\in[0,\rho]$, $t\in[0,\tau]$ subject to the boundary conditions $u(0,t) = u(\rho,t) = 0$ and initial value $u(r,0) = u_0(r)$. The inhomogeneous term $Q(r,t,u)$ depends only on r, in the linear case, or on $u(r,t)$ in the nonlinear case, for the problems studied here. The coefficient functions of the derivative

[a)] Dedicated to Professor Dr. F. Beck (Institut für Kernphysik, Technische Hochschule, Darmstadt, West Germany) on the occasion of his sixtieth birthday.

 0022-2488/87/010039-21$02.50

terms have been chosen as unity. The extension of the method to cases where the coefficient functions depend on r, t, or $u(r,t)$ is not a problem in principle and the analysis is simplified with this assumption. However, a typical real-life problem from theoretical plasma physics does require this extension.[16]

The specific cases chosen for study were as follows for $t\in[0,1]$.

P1: $Q=0$, $r\in[0,\pi]$, and $u_0(r)=\sin(nr)$ for $n=1$.

P2: as in P1 for $n=2$.

P3: as in P1 for $n=4$.

P4: $Q=\sin(r)$, $r\in[0,\pi]$, and $u_0(r)=0$.

P5: $Q=-20u^3(r,t)$, $r\in[0,1]$, and

$$u_0(r)=\begin{cases} r, & r\in[0,\frac{1}{2}], \\ 1-r, & r\in[\frac{1}{2},1] . \end{cases}$$

P6: $Q=\int_0^t \{1-u^2(r,t')\}dt'$, $r\in[0,1]$, and $u_0(r)=0$.

Solutions for problems P1–P4 are known in closed form and for P1–P3 $u(r,t)=e^{-n^2t}\sin(nr)$ while for P4 $u(r,t)=(1-e^{-t})\sin(r)$. Solutions to P5 and P6 are available in numerical form and may be generated from tables of coefficients given by Rektorys using a time discretization method as described in Chaps. 5 and 6 of Ref. 8. Figure 1 shows perspective plots of the solutions for two of the linear problems, namely cases P1 and P3, while Fig. 2 shows the solutions of the nonlinear problems P5 and P6.

In essence the spectral function method has as its basis the method of separation of variables in that the solution has a representation as a sum of products of functions in the different degrees of freedom. The solution is then written as

$$u(r,t)=\sum_{n=1}^{\infty} p_n(t)w_n(x) , \tag{2}$$

with the change in variable $r=\rho x$. The set $\{w_n(x)\}_1^\infty$ is orthonormal on the interval $x\in[-1,+1]$ with respect to a weight function $\phi(x)>0$ on the interval. Conversely, given a Hilbert space $\mathbf{H}_r$ in which an inner product is defined as

$$(f,g)=\int_{-1}^{+1} f(x)g(x)\phi(x)dx , \tag{3}$$

then, for fixed t, Eq. (2) is simply the Fourier series[13] for $u(r,t)$. Therefore for a fixed t the solution $u(r,t)$ is fully specified by the set of bounded numbers $\mathbf{p}=\{p_n(t)\}_1^\infty\subset\mathbf{h}$ and the Hilbert spaces $\mathbf{h}$ and $\mathbf{H}_r$ are isomorphic. However, the exact Fourier coefficients are not known *a priori* and usually a Galerkin method is applied[11] to obtain approximate values for the exact Fourier coefficients for each value of t.

This is the usual approach[9] to solving partial differential equations by the spectral method and is referred to here as the single spectral function method (SSFM). The present study proposes an extension by introducing a second set of spectral functions for Eq. (2) as

$$p_n(t)=\sum_{m=1}^{\infty}{}' P_{nm}v_m(y) , \tag{4}$$

with the change in variable $t=\tau y$. The set $\{v_m(y)\}_1^\infty$ is orthonormal on the interval $y\in[0,1]$ and for some weight function $\psi(y)>0$ an inner product is defined by

$$(f,g)=\int_0^1 f(y)g(y)\psi(y)dy . \tag{5}$$

This extension will be referred to as the double spectral function method (DSFM). In the DSFM Eqs. (2) and (4) combined show that the solution $u(r,t)$ finds a representation in a tensor product space $\mathbf{H}=\mathbf{H}_r\otimes\mathbf{H}_t$ with $\{w_n(x)\}_1^\infty\in\mathbf{H}_r$ and $\{v_m(y)\}_1^\infty\in\mathbf{H}_t$.

B. Choice of spectral functions

1. Definition of the coordinate functions

The choice of basis for the spectral functions is usually limited to simple trigonometric or polynomial functions.[9] However, higher transcendental functions have also been used.[15] In the present application the boundary conditions suggest that the set $\{\sin(nr)\}_1^\infty$ is a natural choice in the trigonometric case. However, in more realistic problems the choice is not so obvious and for this reason a polynomial spectral method is also developed. For the reasons outlined previously[14] and because of their convenient numerical properties the Chebyshev polynomials are used. Chebyshev polynomials have previously been applied to the linear diffusion equation[17] and also to integral equations.[18,19]

In the following the notation will be a general one applicable to either the trigonometric or Chebyshev polynomial choice and also to problems beyond the scope of the present one. Since the approximation methods described below refer to finite and not infinite series this will be explicitly indicated in what follows.

Let $\mathbf{t}$ be the vector of $N+1$ components such that in the trigonometric case the transposed vector is

$$\mathbf{t}^T=\{\cos(0\pi x),\sin(1\pi x),\cos(1\pi x),\sin(2\pi x),\\ \cos(2\pi x),..., \sin(N\pi x/2),\cos(N\pi x/2)\} , \tag{6}$$

and in the polynomial case

$$\mathbf{t}^T=\{T_0(x),T_1(x),T_2(x),T_3(x),...,T_N(x)\} , \tag{7}$$

where $r=\rho x$ and $x\in[-1,+1]$ with N chosen as even. Then an approximation to the solution of Eq. (1) is

$$u=\mathbf{t}^T\mathbf{a} , \tag{8}$$

where $\mathbf{a}$ is a vector of $N+1$ coefficients which depend on time and

$$\mathbf{a}^T=\{\tfrac{1}{2}a_0,a_1,a_2,...,a_N\} . \tag{9}$$

Application of the two boundary conditions to Eq. (8) shows that two of the coefficients are linear combinations of the others. In the Chebyshev case these coefficients are $\frac{1}{2}a_0$ and a_1 while in the Fourier case they are $\frac{1}{2}a_0$ and a_2. On substituting the relevant linear combinations for this pair of coefficients in Eq. (8) and rearranging terms one obtains in the Fourier case

$$u(r,t)=\sum_{n=1}^{N/2} a_{2n-1}(t)e_n(x) \\ +\sum_{n=1}^{N/2-1} a_{2n+2}(t)e_{N/2+n}(x) , \tag{10}$$

where for $n = 1,\dots,N/2$,

$$e_n(x) = \sin(n\pi x)\,,$$

and for $n = 1,\dots,N/2 - 1$, (11)

$$e_{N/2+n}(x) = \begin{cases} \cos\{(n+1)\pi x\} - \cos(0\pi x)\,, & n \text{ odd}, \\ \cos\{(n+1)\pi x\} - \cos(1\pi x), & n \text{ even}. \end{cases} \tag{12}$$

In the Chebyshev case

$$u(r,t) = \sum_{n=1}^{N/2-1} a_{2n+1}(t) e_n(x) + \sum_{n=0}^{N/2-1} a_{2n+2}(t) e_{N/2+n}(x)\,, \tag{13}$$

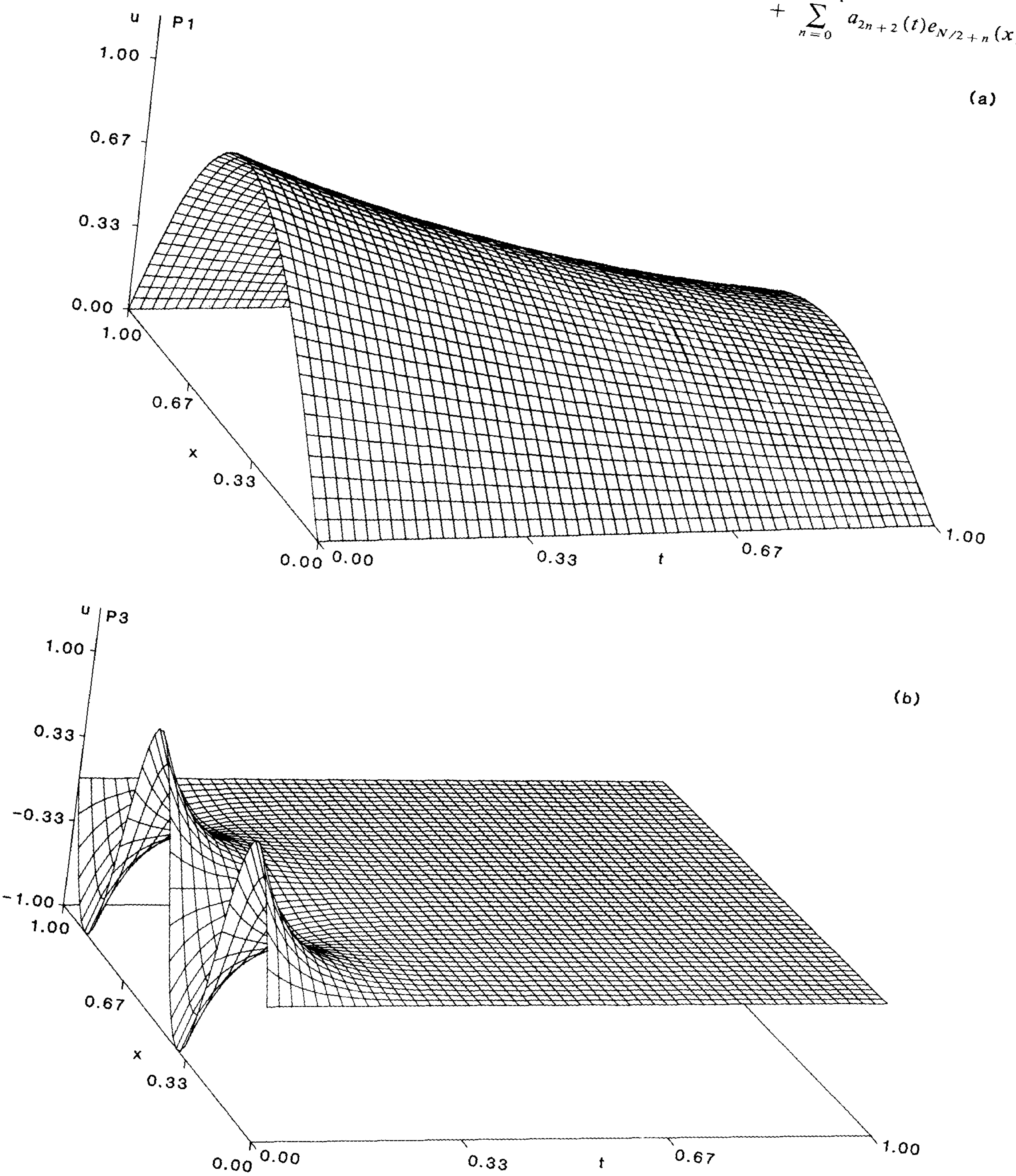

FIG. 1. Perspective plots of solutions to two of the linear problems, namely, (a) case P1 and (b) case P3. The figures were generated from the known analytical solutions as given in Sec. II A. The spatial variable is $x = r/\rho$, where ρ is the maximum value of r.

where for $n = 1,...,N/2-1$,

$$e_n(x) = T_{2n+3}(x) - T_1(x), \tag{14}$$

and for $n = 0,...,N/2-1$,

$$e_{N/2+n}(x) = \begin{cases} T_{2n+2}(x) + T_0(x) - 2T_1(x), & n \text{ even}, \\ T_{2n+2}(x) - T_0(x), & n \text{ odd}. \end{cases} \tag{15}$$

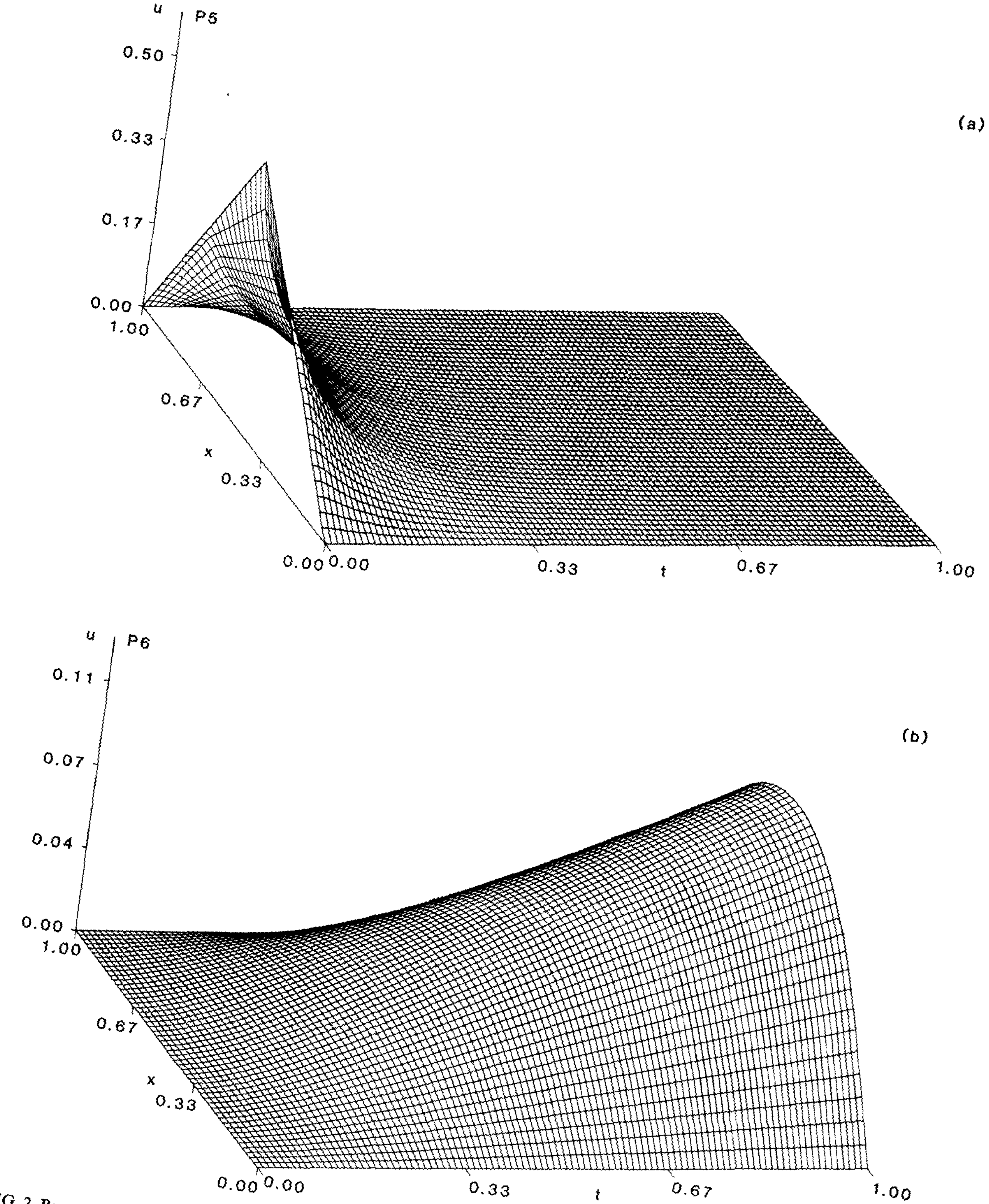

FIG. 2. Perspective plots of solutions to the two nonlinear problems, namely, (a) case P5 and (b) case P6. The spatial variable is $x = r/\rho$, where ρ is the maximum value of r. The plots were generated from the numerical solutions given by Rektorys.[8] In case P5 the solution was approximated with three finite elements at each discrete time point. Note the three distinct ridges parallel to the t axis corresponding to the knots at $r = 0.25$, 0.5, and 0.75. In case P6 the solution was approximated with a linear combination of two continuous functions, namely, $\sin(\pi x)$ and $\sin(3\pi x)$, at each discrete time point.

Each member of the linearly independent set $\{e_n\}_1^{N-1}$ satisfies the boundary conditions and a matrix $\mathbf{T}$ is defined by

$$\mathbf{e}=\mathbf{T}\mathbf{t}, \tag{16}$$

where $\mathbf{e}$ is the vector of $N-1$ coordinate functions $e_n(x)$ of Eqs. (10) and (13). Those functions which are odd in x are placed in the upper block of the vector $\mathbf{e}$ for reasons which become apparent in the next section. The forms of the vectors $\mathbf{t}$[Eqs. (6) or (7)] and $\mathbf{e}$[Eqs. (11) and (12) or (14) and (15)] define the matrix $\mathbf{T}$. This matrix is rectangular with $N-1$ rows and $N+1$ columns since $\mathbf{e}$ has two elements less than $\mathbf{t}$. The matrix $\mathbf{T}$ changes if either the boundary conditions or the basis $\mathbf{t}$ is changed.

2. Orthonormalization

Equation (1) can be solved directly with the expansion of Eqs. (10) or (13) similarly to the previous work.[14] However, while the upper block of the vector $\mathbf{e}$, namely the sine functions of Eq. (11) are orthonormal, the remaining elements e_n in Eqs. (12), (14), and (15) are not. From the point of view of the numerical application it is more efficient to use an orthonormalized set $\{w_n\}_1^{N-1}$ generated from the linearly independent set $\{e_n\}_1^{N-1}$. The Gram–Schmidt method generates an auxiliary set $\{d_n\}_1^{N-1}$ as follows:

$$\begin{aligned} &w_1=d_1/\|d_1\|, \quad d_1=e_1, \\ &w_n=d_n/\|d_n\|, \quad n=2,\ldots,N-1 \end{aligned} \tag{17}$$

with

$$d_n=e_n+\sum_{i=1}^{n-1}\alpha_{ni}w_i, \tag{18}$$

where

$$\alpha_{ni}=-(w_i,e_n). \tag{19}$$

The norm is $\|d_n\|=\sqrt{(d_n,d_n)}$ and the inner product is as defined in Eq. (3) with the weight function $\phi(x)=1$ in the Fourier case and $\phi(x)=1/\sqrt{1-x^2}$ in the Chebyshev case. The Gram–Schmidt method is used to determine the matrix elements S_{ij} of a matrix $\mathbf{S}$, where

$$\mathbf{w}=\mathbf{S}\mathbf{t} \tag{20}$$

using the matrix elements of $\mathbf{T}$ in Eq. (16).

If $\{t_i\}_1^{N+1}$ are elements of the vector $\mathbf{t}$ in Eqs. (6) and (7) then

$$e_n=\sum_{i=1}^{N+1}T_{ni}t_i \tag{16'}$$

and

$$w_n=\sum_{i=1}^{N+1}S_{ni}t_i\,. \tag{20'}$$

Initially, for $n=1$, $S_{1i}=T_{1i}/\|d_1\|$ and subsequently, for $n>1$, S_{ni} is calculated using the following results.

An expression for α_{ni} follows on substituting Eqs. (16′) and (20′) into (19) for $i=1,\ldots,n-1$. This gives for the Chebyshev case

$$\alpha_{ni}=-\tfrac{1}{2}\pi\left\{2S_{i1}T_{n1}+\sum_{j=2}^{N+1}S_{ij}T_{nj}\right\}. \tag{21}$$

Coefficients in the expression

$$d_n=\sum_{i=1}^{N+1}D_{ni}t_i \tag{22}$$

are obtained on substituting Eqs. (16′) and (20′) into (18)

$$D_{ni}=T_{ni}+\sum_{j=1}^{n-1}\alpha_{nj}S_{ji}, \tag{23}$$

whence, again for the Chebyshev case,

$$\|d_n\|=\left\{\tfrac{1}{2}\pi\left(2D_{n1}^2+\sum_{i=2}^{N+1}D_{ni}^2\right)\right\}^{1/2}. \tag{24}$$

It then follows from Eq. (17) that $S_{ni}=D_{ni}/\|d_n\|$. In the Fourier case the factors of $\frac{1}{2}\pi$ in Eqs. (21) and (24) are absent as they arise from the orthogonality relation of the Chebyshev polynomials. The scheme is required in the Fourier case to orthonormalize the set of functions in Eq. (12). The set of functions $\{w_n\}_1^{N-1}$ generated by this scheme were checked for orthogonality and no numerical difficulties were experienced for the applications discussed here.

3. Basis transformation

Comparison of Eqs. (8), (9), and (2) with the latter sum truncated at $N-1$ as required in (17) shows that the coefficients in the expansions are different. This is because there is a change in basis $\{t_i\}_1^{N+1}\to\{w_i\}_1^{N-1}$ with the latter set satisfying the boundary conditions of the problem. In the Chebyshev case it is more convenient from the numerical point of view to use the set of Chebyshev coefficients $\{a_i\}_0^N$ rather than the set $\{p_i\}_1^{N-1}$ because Chebyshev polynomial properties may then be exploited. Therefore two mappings are required

$$R:\ \{a_i\}_0^N\to\{p_i\}_1^{N-1}, \tag{25}$$

$$R^{-1}:\ \{p_i\}_1^{N-1}\to\{a_i\}_0^N. \tag{26}$$

Matrix representations of the mappings R and R^{-1} follow on comparing Eqs. (2), (8), and (9),

$$\sum_{n=1}^{N-1}p_n(t)w_n(x)=\sum_{m=1}^{N+1}{}'\,a_{m-1}(t)t_m(x), \tag{27a}$$

where the prime on the second summation indicates that the first term is halved. In vector notation

$$\mathbf{w}^T\mathbf{p}=\mathbf{t}^T\mathbf{a}. \tag{27b}$$

Left multiplication by the vector $\mathbf{w}$ or $\mathbf{t}$ followed by application of the inner product Eq. (3) gives, respectively,

$$\mathbf{p}=\mathbf{R}\mathbf{a} \tag{28}$$

and

$$\mathbf{R}^{-1}\mathbf{p}=\mathbf{a}. \tag{29}$$

It follows from the expression of Eq. (20) and the definition Eq. (9) of the vector $\mathbf{a}$ of coefficients that in the Chebyshev case $\mathbf{R}=(\mathbf{w},\mathbf{t}^T)$ is $\frac{1}{2}\pi$ times the matrix $\mathbf{S}$ with the first column multiplied by 2. In the Fourier case the factor of $\frac{1}{2}\pi$ is absent. Similarly $\mathbf{R}^{-1}=\mathbf{S}^T$ for both Fourier and Chebyshev cases. Note that the matrix $\mathbf{R}$ has $N-1$ rows and $N+1$ columns while the converse is the case for $\mathbf{R}^{-1}$.

C. The matrix problem

In the following it is assumed that the inhomogeneous term of Eq. (1) has an expansion analogous to Eq. (2), namely,

$$Q(r,t,u) = \sum_{n=1}^{N-1} q_n(t) w_n(x) . \tag{30}$$

This expansion is obtained by a basis transformation of an expansion similar to Eqs. (8) and (28)

$$\mathbf{w}^T\mathbf{q} = \mathbf{t}^T\mathbf{f} , \tag{31}$$

where

$$\mathbf{f}^T = \{\tfrac{1}{2}f_0, f_1, f_2, ..., f_N\} \tag{32}$$

is the vector of expansion coefficients of $Q(r,t,u)$ on the basis $\{t_i\}_1^{N+1}$. The procedure for obtaining these coefficients for nonlinear problems is described in Sec. III and here it is assumed that they are known.

Substitution of Eqs. (2) and (30) into Eq. (1) gives the equation

$$\mathbf{w}^T\left(\frac{d\mathbf{p}}{dt}\right) = \rho^{-2}\mathbf{w}''^T\mathbf{p} + \mathbf{w}^T\mathbf{q} , \tag{33}$$

where

$$\mathbf{p}^T = \{p_1(t), p_2(t), ..., p_{N-1}(t)\} \tag{34}$$

and similarly for $\mathbf{q}$. The double prime on $\mathbf{w}''$ indicates that the components are differentiated twice with respect to argument x where $r = \rho x$. Left multiplication of Eq. (33) by $\mathbf{w}$ and application of the inner product Eq. (3) yields the result

$$\frac{d\mathbf{p}}{dt} = \mathbf{H}\mathbf{p} + \mathbf{q} , \tag{35}$$

where, in the Fourier case, the matrix

$$\mathbf{H} = (\mathbf{w}, \mathbf{w}''^T)/\rho^2 \tag{36}$$

is diagonal. In the Chebyshev case $\mathbf{H}$ is neither diagonal nor symmetric but may be written in diagonal form as

$$\mathbf{H} = \mathbf{Y}\mathbf{\Lambda}\mathbf{Z}^T \tag{37}$$

where $\mathbf{\Lambda}$ is the diagonal matrix of (real) eigenvalues of $\mathbf{H}$ arranged in increasing magnitude. The matrices $\mathbf{Y}$ and $\mathbf{Z}$ have columns which are eigenvectors of $\mathbf{H}$ and $\mathbf{H}^T$, respectively, with normalization

$$\mathbf{Z}^T\mathbf{Y} = \mathbf{Y}\mathbf{Z}^T = \mathbf{1}. \tag{38}$$

Left multiplication of Eq. (35) by $\mathbf{Z}^T$ yields the vector equation

$$\frac{d\mathbf{p}^{(z)}}{dt} = \mathbf{\Lambda}\mathbf{p}^{(z)} + \mathbf{q}^{(z)} , \tag{39}$$

which is the spectral resolution of Eq. (1) with

$$\mathbf{p}^{(z)} = \mathbf{Z}^T\mathbf{p} \tag{40}$$

while from Eq. (38) it follows that

$$\mathbf{p} = \mathbf{Y}\mathbf{p}^{(z)} . \tag{41}$$

However, from Eqs. (28) and (29)

$$\mathbf{p}^{(z)} = \mathbf{Z}^T\mathbf{R}\mathbf{a} = \mathbf{F}\mathbf{a} \tag{42}$$

and

$$\mathbf{a} = \mathbf{R}^{-1}\mathbf{Y}\mathbf{p}^{(z)} = \mathbf{F}^{-1}\mathbf{p}^{(z)} \tag{43}$$

with similar results for $\mathbf{f}$, $\mathbf{q}$, and $\mathbf{q}^{(z)}$. From Eqs. (42) and (43) $\mathbf{F}$ is seen to be a matrix representation of a mapping F from a "physical" to a "spectral" space, namely,

$$F\colon \{a_n\}_0^N \to \{p_m^{(z)}\}_1^{N-1} \tag{44}$$

and F^{-1} is the inverse mapping. It suffices to consider the mapping as acting between the two sets of coefficients. This is because in either of the respective Hilbert spaces spanned by the orthogonal bases $\{t_n\}$ and $\{w_n\}$ the corresponding set of coefficients is a unique representation of an element in that space. This is a consequence of the Riesz–Fischer theorem.[13]

This section is concluded by showing how explicit expressions for the matrix elements $\mathbf{H}$ are obtained. The matrix elements are easily found once the $\{w''_n\}$ are expressed on the $\{t_m\}$ basis. The first step is to note that from Eq. (20′)

$$w''_n = \sum_{i=1}^{N+1} S_{ni} t''_i \tag{45}$$

and consequently the second step is to find an expression for t''_i on the $\{t_n\}$ basis.

At this point it is observed that the spectral basis of Eqs. (10) and (13) is grouped into two parts consisting, respectively, of odd and even powers in the argument x. This division has numerical advantages and is adhered to here.

For even powers of x in the Fourier case

$$t''_{2m+1} = \sum_{k=0}^{N/2} c_{mk} t_{2k+1} , \quad 1 \leqslant m \leqslant N/2 , \tag{46}$$

and from Eq. (6)

$$t_{2m+1}(x) = \cos(m\pi x), \quad 0 \leqslant m \leqslant N/2 , \tag{47}$$

therefore the only term which survives in Eq. (46) is $k = m$ with

$$c_{mm} = -m^2\pi^2 .$$

Similarly for odd powers of x

$$t''_{2m} = \sum_{k=1}^{N/2} c^{\circ}_{m-1,k-1} t_{2k} , \quad 1 \leqslant m \leqslant N/2 , \tag{48}$$

and

$$t_{2m}(x) = \sin(m\pi x), \quad 1 \leqslant m \leqslant N/2 , \tag{47'}$$

with the only nonzero term of Eq. (48) being

$$c^{\circ}_{m-1,m-1} = -m^2\pi^2 .$$

The corresponding expressions in the Chebyshev case follow on application of results given in Section 8.5.1 of Luke.[20] For even powers of x

$$t''_{2m+1} = \sum_{k=0}^{m-1} c_{mk} t_{2k+1} , \quad 1 \leqslant m \leqslant N/2 , \tag{49}$$

and from Eq. (7)

$$t_{2m+1}(x) = T_{2m}(x), \quad 0 \leqslant m \leqslant N/2 , \tag{50}$$

where

$$c_{m0} = 4m^3, \quad k = 0 ,$$

$$c_{mk} = 8m(m^2 - k^2), \quad 0 < k \leqslant m - 1 .$$

For odd powers of x

$$t''_{2m} = \sum_{k=1}^{m-1} c^{\circ}_{m-1,k-1} t_{2k} , \quad 2 \leqslant m \leqslant N/2 , \tag{51}$$

and

$$t_{2m}(x) = T_{2m-1}(x), \quad 1 \leqslant m \leqslant N/2 , \tag{50'}$$

where

$$c^{\circ}_{m-1,k-1} = 4(2m-1)(m-k)(m+k-1),$$
$$1 \leqslant k \leqslant m-1 .$$

Substitution of Eqs. (46) and (48) or Eqs. (49) and (51) into Eq. (45) yields for even powers of x

$$w''_n = \sum_{k=0}^{N3} K_{nk} t_{2k+1} , \tag{52}$$

where

$$K_{nk} = \sum_{m=k}^{M2} S_{n,2m+1} c_{mk} , \tag{53}$$

and for odd powers of x

$$w''_n = \sum_{k=1}^{N3} K^{\circ}_{nk}\, t_{2k} , \tag{52'}$$

with

$$K^{\circ}_{nk} = \sum_{m=k}^{M2} S_{n,2m} c^{\circ}_{m-1,k-1} , \tag{53'}$$

with summation indices in the Chebyshev case $N3 = N/2 - 1$, $M2 = N/2$, and in the Fourier case $N3 = N/2$, $M2 = k$. The matrix elements of $\mathbf{H}$ in Eq. (36) then follow from the inner product appropriate to the Chebyshev or Fourier cases with the resulting expression

$$\rho^{-2}(w_{n'},w''_n) = \frac{1}{2}\pi\rho^{-2}\left\{2S_{n'1}K_{n0} + \sum_{k=1}^{N3}\left[S_{n',2k+1}K_{nk} + S_{n',2k}K^{\circ}_{nk}\right]\right\}, \tag{54}$$

and the factor of $\frac{1}{2}\pi$ is absent in the Fourier case.

D. The single spectral function method (SSFM)

The SSFM consists of the transformation from the physical space of Eq. (1) into the spectral space of Eq. (39) as described above. This mapping is accomplished through the matrix representation $\mathbf{F}$ as defined in Eq. (42), namely, $\mathbf{p}^{(z)} = \mathbf{Fa}$. The vectors of coefficients $\mathbf{a}$ and $\mathbf{p}^{(z)}$ both depend on time and an initial value is specified for $\mathbf{p}^{(z)}$ by the vector $\mathbf{a}$ corresponding to $u_0(r) = u(r,0)$. If the inhomogeneous term Q of Eq. (1) is not zero and does not depend on time then $\mathbf{q}^{(z)}$ in Eq. (39) is obtained from $\mathbf{q}^{(z)} = \mathbf{Ff}$, where $\mathbf{f}$ is defined in Eq. (32). With these initial conditions the system of first-order ordinary differential equations (ODE's) in Eq. (39) can be integrated in time in the spectral space by either a predictor–corrector (PC) or Runge–Kutta (RK) method.[21] The solution at any time t in the physical space is obtained by the inverse mapping of Eq. (43), namely, $\mathbf{a} = \mathbf{F}^{-1}\mathbf{p}^{(z)}$. If Q does depend on time, as in the case for the nonlinear problems P5 and P6, then the spectral transforms of Eqs. (42) and (43) must be applied at each time step. The details of this case are described in Sec. III.

Although a large amount of literature on numerical solution of systems of ODE's exists, the text by Lapidus and Seinfeld[21] was found to be useful from the point of view of the practitioner. Since this reference documents and compares a substantial number of different integration formulas no detail is presented here. However, a full description of the ODE software package designed for the present application will be the subject of a separate report.[22] Therefore the description here is a brief one referencing only the formulas used and the important numerical parameters determining accuracy. Second to sixth order formulas for both RK and PC methods were taken from Lapidus and Seinfeld.[21] The formulas used come from tables in Eqs. (2), (3)–(7), (10), (14), (19), and (32) and the Adams Bashforth cases of Table 4.1 of that reference. Key parameters determining the numerical accuracy of the integrated solution were (a) the step size Δt, (b) the order of the formula k, (c) the number of iterations s, and (d) the number of components to be integrated in Eq. (39) N_s.

The iteration number s applies only to the PC formulas and its counts the number of times the predicted value is corrected. Thus $s = 0$ corresponds to prediction without correction, $s = 1$ one correction, etc. The procedure is defined in detail in Eq. (4.1-5) of Lapidus and Seinfeld.[21] In the present application the PC method is applied in preference to the RK method because for the choice of the above parameters used here it requires two to four times fewer function evaluations.

E. The double spectral function method (DSFM)

Clenshaw[23] observed that if an ODE has simple polynomial coefficients then the expansion coefficients of the solution in Chebyshev series may be generated by recurrence. The technique has previously been applied to the evaluation of higher transcendental functions[24] and in the present work it is applied to the system of ODE's of Eq. (39). In this equation each component of $\mathbf{p}^{(z)}$ is a function of time and from Eq. (42) it also follows that

$$\mathbf{p}^{(z)} = \mathbf{Fa} .$$

Therefore, for the components of $\mathbf{a}$ in Eqs. (8) and (9) introduce the expansions (written in vector form) as

$$\mathbf{a} = \mathbf{At^*} \tag{55a}$$

and similarly for the coefficients $\mathbf{f}$ of the inhomogeneous term defined in Eqs. (31) and (32)

$$\mathbf{f} = \mathbf{Et^*} . \tag{55b}$$

The vector $\mathbf{t^*}$ is

$$\mathbf{t}^{*T} = \{T^*_0(y), T^*_1(y), ..., T^*_M(y)\} , \tag{56}$$

where $T^*_m(y)$ is the shifted Chebyshev polynomial[20] and $t = \tau y$ with $t \in [0,\tau]$, $y \in [0,1]$. The first column of $\mathbf{A}$ ($\mathbf{E}$) has matrix elements $\frac{1}{2}A_{n0}$ ($\frac{1}{2}E_{n0}$) where A_{n0} (E_{n0}) is the first shifted Chebyshev coefficient of $a_n(t)$ $(f_n(t))$, $n = 0,1,...N$. In view of Eqs. (6), (7), and (56) matrices $\mathbf{A}$ and $\mathbf{E}$ have $M+1$ columns and $N+1$ rows.

Substituting Eq. (55a) into Eq. (42) yields

$$\mathbf{p}^{(z)} = \mathbf{Fa} = \mathbf{FAt^*} = \mathbf{Pt^*} \tag{57a}$$

and similarly

$$\mathbf{q}^{(z)} = \mathbf{Ff} = \mathbf{FEt^*} = \mathbf{Qt^*} , \tag{57b}$$

while the converse follows from Eq. (43)

$$\mathbf{a} = \mathbf{F}^{-1}\mathbf{p}^{(z)} = \mathbf{F}^{-1}\mathbf{Pt^*} = \mathbf{At^*} \tag{58a}$$

with

$$\mathbf{f} = \mathbf{F}^{-1}\mathbf{q}^{(z)} = \mathbf{F}^{-1}\mathbf{Qt^*} = \mathbf{Et^*} . \tag{58b}$$

In the DSFM matrix elements of **P** are obtained by recurrence in spectral space and the matrix **A** of Eq. (55a) in the physical space is obtained by the transform

$$\mathbf{A} = \mathbf{F}^{-1}\mathbf{P}, \tag{59a}$$

and similarly

$$\mathbf{E} = \mathbf{F}^{-1}\mathbf{Q}. \tag{59b}$$

In the case that the inhomogeneous term of Eq. (1) is a known function of r and t the matrix **E** of Eq. (55b) is known and $\mathbf{Q} = \mathbf{FE}$. However, if the inhomogeneous term depends on the solution $u(r,t)$, as is the case in the nonlinear problems P5, P6, then the DSFM described here has to be extended. This is done in Sec. III. This section is concluded with a description of how matrix elements of **P** are obtained by recurrence in the linear problems P1 to P4.

Consider the nth component of the vector Eq. (39) where the inhomogeneous term is considered to be zero,

$$\frac{dp_n^{(z)}}{dt} = \lambda_n p_n^{(z)}, \tag{60a}$$

with the change of variable $t = \tau y$ this becomes

$$\frac{dp_n^{(z)}}{dy} = \tau\lambda_n p_n^{(z)}, \tag{60b}$$

where λ_n is the nth (real) eigenvalue of the matrix **H** in Eq. (37). Introducing the expansion Eq. (57a)

$$p_n^{(z)}(y) = \sum_{m=0}^{M}{}' P_{nm} T_m^*(y), \tag{61a}$$

where the prime on the summation denotes that the first term contains the usual factor of $\frac{1}{2}$. For the derivative of $p_n^{(z)}(y)$ with respect to y let

$$\frac{dp_n^{(z)}}{dy} = \sum_{m=0}^{M}{}' P'_{nm} T_m^*(y) \tag{61b}$$

then, from the orthogonality property of the shifted Chebyshev polynomials, Eq. (60b) reduces to

$$P'_{nm} = \tau\lambda_n P_{nm}. \tag{62}$$

From Clenshaw[23]

$$4mP_{nm} = P'_{n,m-1} - P'_{n,m+1},$$

which, when combined with Eq. (62), gives the three-term recurrence

$$P_{n,m-1} = 4mP_{nm}/(\tau\lambda_n) + P_{n,m+1}. \tag{63}$$

In the case that the inhomogeneous term is not zero, then from Eq. (57b)

$$q_n^{(z)}(y) = \sum_{m=0}^{M}{}' Q_{nm} T_m^*(y), \tag{61a$'$}$$

and in place of Eq. (63) one has

$$P_{n,m-1} = 4mP_{nm}/(\tau\lambda_n) + P_{n,m+1} + (Q_{n,m+1} - Q_{n,m-1})/\lambda_n. \tag{64}$$

At time $t = 0$ it follows from $T_m^*(0) = (-1)^m$ that Eq. (61a) becomes

$$p_n^{(z)}(0) = \sum_{m=0}^{M}{}' (-1)^m P_{nm}. \tag{65}$$

The downward recurrence Eq. (64) is started with $P_{n,M+1} = 0$, $P_{nM} = 1$ and the resulting coefficients are normalized from Eq. (65). Numerical results are discussed in Sec. VI.

III. APPLICATION TO NONLINEAR PROBLEMS

A. Preliminaries

The type of nonlinearity discussed in this study is of a relatively simple type, namely, a (positive) integer power of the unknown solution. Furthermore the nonlinearity is assumed to occur only in the inhomogeneous term of the model problem given in Eq. (1). In the case of P5 (see Sec. II A) the nonlinearity is cubic in the solution $u(r,t)$ and in P6 it is a Volterra-like integral in time of a quadratic in $u(r,t)$. Because of the occurrence of only integer powers a simple property of both Chebyshev and Fourier series may be applied. The product of two series of $N+1$ terms can be written as a single series of $2(N+1)$ terms. In the vector notation of Eq. (8) this is the convolution product

$$u^2(r,t) = (\mathbf{t}^T\mathbf{a}) \otimes (\mathbf{t}^T\mathbf{a}) = \mathbf{t}^T\mathbf{b} \tag{66}$$

and the relevant expressions for **b** in the Chebyshev case are to be found at the end of Sec. 8.6.1 of Luke.[20] Similar expressions can be derived for the case of the product of two Fourier series. These results and the algorithm used to calculate the product of Eq. (66) for Chebyshev and Fourier cases are reported elsewhere.[25]

The remainder of this section discusses the application of Eq. (66) to problems P5 and P6 in both the SSFM and DSFM. In these applications the series which results from the convolution product of Eq. (66) is truncated at the same length as the component series namely $N+1$ terms. This approximation is a good one if the coefficients of the component series decrease sufficiently rapidly.

B. The single spectral function method

The SSFM performs a time integration of the vector Eq. (39) using either a RK or PC method. Therefore the time dependence of the inhomogeneous term must be computed at each time step. In the example of P5 the nonlinear inhomogeneous term is evaluated at each time step in three stages.

Algorithm 1:

(1) Back transform $\mathbf{a} = \mathbf{F}^{-1}\mathbf{p}^{(z)}$.

(2) Convolution product $-20(\mathbf{t}^T\mathbf{a}) \otimes [(\mathbf{t}^T\mathbf{a}) \otimes (\mathbf{t}^T\mathbf{a})] = \mathbf{t}^T\mathbf{f}$.

(3) Forward transform $\mathbf{Ff} = \mathbf{q}^{(z)}$,

where $\mathbf{F}$, $\mathbf{F}^{-1}$, $\mathbf{f}$, and $\mathbf{q}^{(z)}$ are as defined in Sec. II C. Stage 2 of the algorithm is a twofold product of the type in Eq. (66). Once $\mathbf{q}^{(z)}$ at the current time step has been calculated the PC or RK method can predict the vector $\mathbf{p}^{(z)}$ at the subsequent time step. For time $t = 0$ the algorithm commences at stage 2 using for **a** the coefficients of the initial distribution $u_0(r) = u(r,0)$.

In the case of problem P6 Algorithm 1 is modified in stage 2. After evaluation of one convolution product a quadrature is applied to determine the vector **f** at the current time step. Here the inhomogeneous term (Sec. II A) has the form

$$Q = \int_0^t \{1 - u^2(r,t')\} dt' . \quad (67)$$

If the convolution product for u^2 has been evaluated then

$$u^2(r,t') = \sum_{n=0}^{N}{}' b_n(t') t_{n+1}(x) , \quad (68)$$

and

$$Q = t - \sum_{n=0}^{N}{}' h_n(t) t_{n+1}(x) , \quad (69)$$

where

$$h_n(t) = \int_0^t b_n(t') dt' . \quad (70)$$

Therefore the coefficients **f** in the expression for the inhomogeneous term Eq. (31) are given by

$$f_n(t) = 2t\,\delta_{n,0} - h_n(t) , \quad (71)$$

where $\delta_{n,0}$ is the Kronecker delta function and the factor of 2 in the first term arises because of the (usual) factor of $\frac{1}{2}$ in the summation of Eqs. (31) and (32). In evaluating Eq. (70) numerically the simplest quadrature is the trapezoidal rule. For a time step Δt the coefficient $h_n(t)$ at the current time step t is computed from

$$h_n(t) = h_n(t - \Delta t) + \tfrac{1}{2}\Delta t\, [b_n(t - \Delta t) + b_n(t)] . \quad (72)$$

Therefore, for P6, the second stage of Algorithm 1 is as follows.

(2a) Convolution product $(\mathbf{t}^T\mathbf{a}) \otimes (\mathbf{t}^T\mathbf{a}) = \mathbf{t}^T\mathbf{b}$.

(2b) Quadrature of Eq. (72).

(2c) Form vector **f** from Eq. (71).

The initial condition for P6 is that $u_0(r) = 0$ and it follows from Eq. (67) that the inhomogeneous term is zero at $t = 0$. Consequently the modified form of Algorithm 1 is first applied at the first time step $t = \Delta t$.

C. The double spectral function method

This section describes the extension to the discussion of Sec. II E in applying the DSFM to nonlinear problems. The essence of the nonlinear DSFM is the solution for the coefficient matrix **A** of Eq. (55a) by an iterative technique. Schematically the procedure is set out in the following steps.

Algorithm 2:

(0) Compute matrix **P** by downward recurrence of Eq. (63).

(1) Back transform $\mathbf{A} = \mathbf{F}^{-1}\mathbf{P}$ of Eq. (59a).

(2) Compute coefficient matrix **E** of Eq. (55b).

(3) Forward transform $\mathbf{Q} = \mathbf{FE}$ of Eq. (57b).

(4) Compute new matrix **P** by downward recurrence of Eq. (64).

Step 0 is only used to provide an initial matrix **P** and the iteration consists of repetition of steps 1 to 4 until the matrix **A** has converged. Only step 2 is dependent on the specific form of the nonlinearity in the inhomogeneous term Q of Eq. (1), and the discussion here is specific to cases P5 and P6.

After back transformation **A** is known and **E** is computed as follows. From Eqs. (8), (9) and (55a), with $\epsilon_n = 1$, $n = 0$ and $\epsilon_n = 2$, $n > 0$.

$$u(r,t) = \sum_{n=0}^{N} \frac{1}{2}\epsilon_n \left[\sum_{m=0}^{M} \frac{1}{2}\epsilon_m A_{nm}\, T_m^*(y) \right] t_{n+1}(x) , \quad (73)$$

and, assuming that $t_{n+1}(x)$ is the Chebyshev case of Eq. (7), the convolution product of Eq. (66) becomes

$$u^2(r,t) = \sum_{n'n''}^{N} \frac{1}{2}\epsilon_{n'} \frac{1}{2}\epsilon_{n''} \left[\sum_{m'=0}^{M} \frac{1}{2}\epsilon_{m'} A_{n'm'}\, T_{m'}^*(y) \times \sum_{m''=0}^{M} \frac{1}{2}\epsilon_{m''} A_{n''m''}\, T_{m''}^*(y) \right] \times [t_{n'+1}(x) t_{n''+1}(x)] . \quad (74)$$

For each n', n'' the product of two series in the first parentheses is reduced to a single series

$$\sum_{m=0}^{M} \frac{1}{2}\epsilon_m C_{n'n''m}\, T_m^*(y)$$

similar to Eq. (66). For the second term in parentheses, from properties of Chebyshev polynomials,

$$t_{n'+1} t_{n''+1} = \tfrac{1}{2} t_{|n'-n''|+1} + \tfrac{1}{2} t_{n'+n''+1} .$$

Consequently the n', n'' summations may be performed to give the coefficients B_{nm} of

$$u^2(r,t) = \sum_{n=0}^{N} \frac{1}{2}\epsilon_n \left[\sum_{m=0}^{M} \frac{1}{2}\epsilon_m B_{nm}\, T_m^*(y) \right] t_{n+1}(x) , \quad (75)$$

where the term in the parentheses is the coefficient $b_n(t)$ of Eq. (68). For case P6 $h_n(t)$ of Eq. (70) then has the expansion

$$h_n(t) = \sum_{m=0}^{M} \frac{1}{2}\epsilon_m H_{nm}\, T_m^*(y) , \quad (76)$$

and Elliot[18] has given expressions relating Chebyshev expansion coefficients of the integral of a function to those for the function itself. Applying these results in the present case gives, for $m = 0$,

$$H_{n0} = \frac{1}{2}\tau \left[B_{n0} - \frac{1}{2} B_{n1} - 2 \sum_{k=2}^{M} \left\{ \frac{(-1)^k}{k^2 - 1} \right\} B_{nk} \right] , \quad (77a)$$

and for $m > 0$

$$H_{nm} = \tau [B_{nm-1} - B_{nm+1}]/(4m) . \quad (77b)$$

The first term of Eq. (69) has the first two expansion coefficients as 2τ and τ, respectively, and therefore, for each n, the matrix **E** of Eq. (55b) is defined by

$$\begin{aligned} E_{n0} &= 2\tau - H_{n0} , \\ E_{n1} &= \tau - H_{n1} , \\ E_{nm} &= -H_{nm}, \quad m > 1 . \end{aligned} \quad (78)$$

For case P5 the coefficients E_{nm} of Eq. (55b) follow directly from a second convolution of -20 times Eq. (73) with Eq. (75) and this is the exact analog of the reduction of Eq. (74) to Eq. (75).

IV. RESULTS FOR THE EIGENVALUE PROBLEM

In the trigonometric case, with the spectral basis of Eq. (11), the matrix **H** defined in Eq. (36) is diagonal. However, in the case of the Chebyshev spectral basis of Eqs. (14) and (15), **H** is not diagonal and is reduced to the diagonal form of Eq. (37) numerically by use of routines F01AKF, F01APF, and F02AQF of the NAGLIB scientific software package.[26] Since solutions of all the model problems P1–P6 are odd functions of the spatial variable, it suffices to consider only the Chebyshev basis of odd polynomials given in Eq. (14). This amounts to retaining only the first term of Eq. (13) and the first $N/2-1$ rows and columns of **H**. The diagonalization of **H** and $\mathbf{H}^T$ need only be performed once before application to any of the problems P1 to P6. However, in view of the transforms of Eqs. (40) and (41) and also Eq. (39), the accuracy of eigenvalues and eigenvectors determines the accuracy of subsequent operations. Therefore in this section typical results for eigenvalues are discussed and accuracy for eigenvectors is referred to in subsequent sections as appropriate.

The exact eigenvalues are known to be $-n^2$ for $n=1, 2,...,$ if $\rho^2=\pi^2$. However, convergence of approximate eigenvalues to the exact result in a numerical scheme has to be assessed. Two criteria have been proposed[14] to assess performance of a finite rank spectral approximation to an eigenvalue problem. The two criteria aim to answer the following two questions: (i) how many eigenvalues have converged to within a prescribed error for a given matrix order, and (ii) at which rate does a given eigenvalue converge to within a prescribed error as a function of increasing matrix size? The first criterion is determined by a plot of the magnitude of each eigenvalue as a function of the matrix order required for convergence to four significant figures. This type of plot shows the matrix size (i.e., the number of spectral functions) required to ensure convergence of all eigenvalues inside a bounded interval of the real line. For the present application this plot is shown in Fig. 3. Since eigenvalues are ordered in increasing magnitude it is seen that, typically, for a given matrix order, approximately half of the eigenvalues are accurate to within four significant figures or better. Note that the slope of this type of plot changes if the spectral basis is changed.[15] The second criterion of convergence is determined for each eigenvalue from a plot of the rate of convergence to a prescribed error of one digit in the Sth significant figure versus matrix order. For the present work typical results are shown in Fig. 4 and the rate of convergence is seen to be even more rapid than was the case for a simple Schrödinger operator as shown in Fig. 1 of Ref. 14. From Fig. 4 it follows that eigenvalues of smaller magnitude converge to more significant figures than do larger eigenvalues for a given matrix size.

Consequently it can be concluded that convergence of approximate eigenvalues to the exact ones is both rapid and stable in the present application of the Chebyshev spectral basis.

V. RESULTS FOR THE SINGLE SPECTRAL FUNCTION METHOD

The SSFM consists of time integration of the vector ODE of Eq. (39). If only the Fourier spectral basis of Eq. (11) is chosen then the matrix **H** of Eq. (36) is diagonal. In this case solution of the linear problems P1 to P4 is then trivial and reduces to time integration of a single component

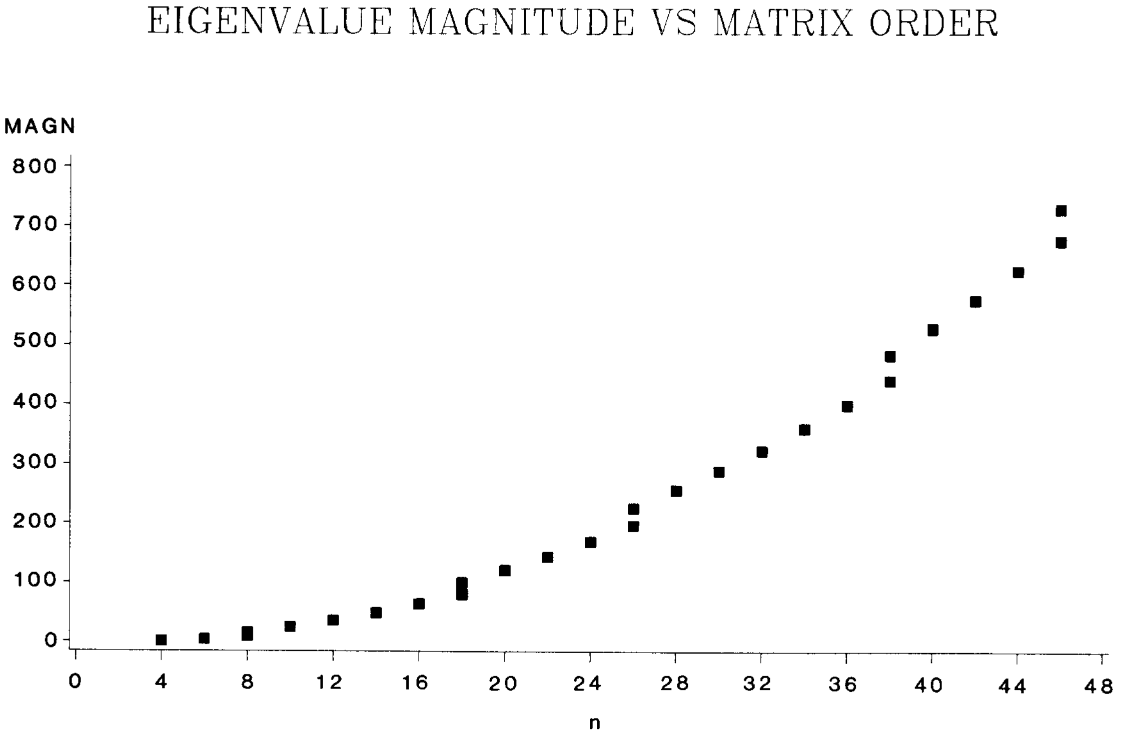

FIG. 3. The matrix order n required to produce convergence of one digit in the fourth figure in the Chebyshev case as a function of the magnitude of the eigenvalue of matrix **H**. The results are specific to the odd polynomial spectral function basis of the first term in Eq. (13) and consequently $n=N/2-1$ is the number of spectral functions used.

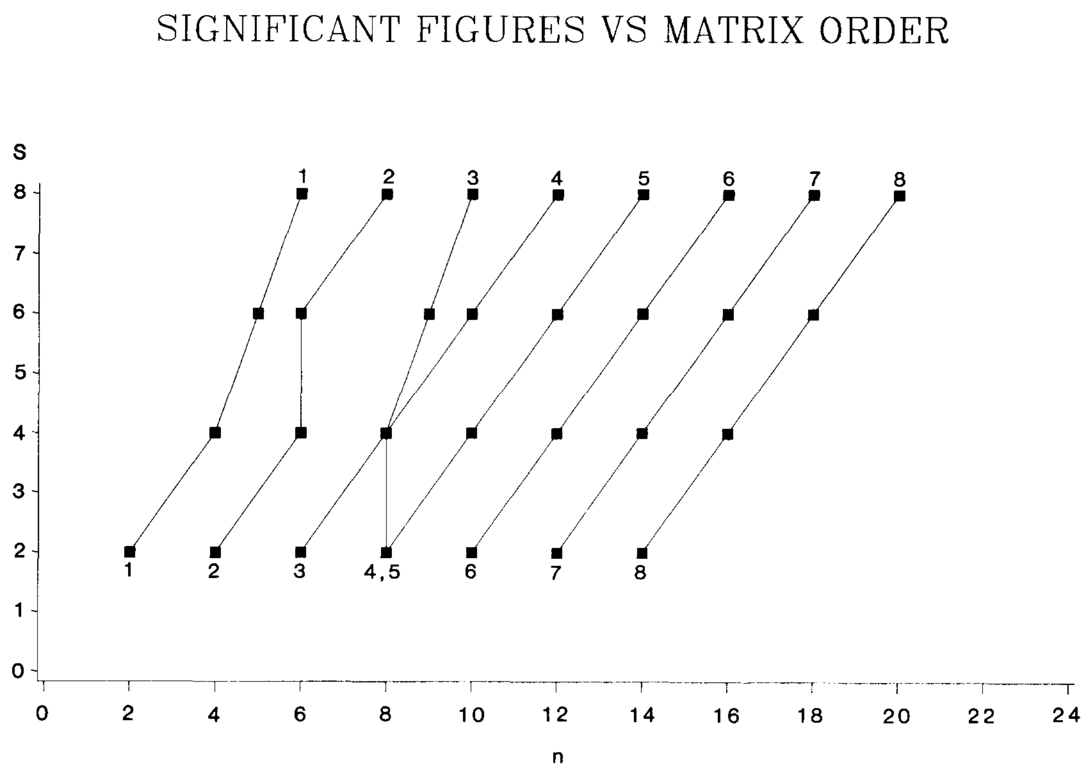

FIG. 4. The rate of convergence to one digit in the Sth decimal in the Chebyshev case as a function of the matrix order. The results are specific to the odd polynomial spectral function basis of the first term in Eq. (13) and consequently $n = N/2 - 1$ is the number of spectral functions used. The labels on the curves give the eigenvalue sequence when ordered in increasing magnitude.

of the vector ODE Eq. (39) in the appropriate coefficient of Eq. (10), namely $a_1(t)$, $a_3(t)$, and $a_7(t)$, respectively, corresponding to eigenvalues -1^2, -2^2, and -4^2. However, in the Chebyshev spectral basis the matrix $\mathbf{H}$ of Eq. (36) must first be diagonalized as discussed in Sec. IV. In terms of the eigenvectors the spectral transforms $\mathbf{F}$ and $\mathbf{F}^{-1}$ of Eqs. (42) and (43) are then defined. Again, in the case of the linear problems P1 to P4, the forward spectral transform Eq. (42) of the vector $\mathbf{a}$ of Eq. (9) gives only a single component in the vector Eq. (39). Thus the linear problems P1–P4 are characterized by a single eigenmode and each eigenmode corresponds to a different eigenvalue of the matrix $\mathbf{H}$. In the nonlinear problems P5 and P6 these eigenmodes are coupled together in physical space but the spectral transform of Eq. (44) in combination with the convolution product of Eq. (66) gives a system of uncoupled, simultaneous ODE's in spectral space. This coupling is manifest on inspection of the vector Eq. (40) in the nonlinear case. Several components are in evidence and the different components of $\mathbf{p}^{(z)}$ vary considerably in magnitude. Only the dominant components of $\mathbf{p}^{(z)}$ need be retained for time integration and the remainder are set to zero. The number of components retained, denoted by N_s, is determined by the accuracy required in the solution and how well eigenvalues and eigenvectors of the matrix $\mathbf{H}$ have converged.

Since the Fourier spectral basis has been investigated in detail by others,[9,11] the present numerical experiments concentrate mainly on the Chebyshev case. Although the linear problems are trivial they are still useful in assessing the resolving power of the Chebyshev spectral basis. Furthermore they also provide a benchmark against which performance for the nonlinear problems can be assessed. Figures 5–7 show results for the linear cases P1 to P3 obtained with the odd Chebyshev spectral basis of Eq. (14) which corresponds to retaining only the first term of Eq. (13). Therefore only the odd Chebyshev coefficients a_1, a_3, a_5,...,are present. The exact Chebyshev coefficients at any time t follow simply by multiplying the Chebyshev coefficients of the sine function in the spatial degree of freedom by the appropriate exponential function in time as required from the corresponding analytical solution given in Sec. II A. Figure 5 shows typical results for the logarithm of the difference between exact and calculated values of $a_{2n+1}(t)$ for cases P1–P3. The values of the PC parameters Δt, k, and s as defined in Sec. II D are shown in Fig. 5. The parameter N is defined in Eq. (7) and $n = N/2 - 1$ is the number of odd polynomial spectral functions defined in Eq. (14). The times shown in Fig. 5 correspond approximately to the points at half of the maximum magnitude of $u(r,t)$. Since the number of (half) oscillations in the spatial degree of freedom doubles from P1 to P2 and P2 to P3 (see Fig. 1) the number of spectral functions also doubles for a similar accuracy. These results suggest that four spectral functions per half oscillation in the spatial degree of freedom are sufficient to provide a resolving power of the order of 10^{-4} or better. This estimate is substantiated in Figs. 6 and 7 which show error curves in the time and spatial degrees of freedom, respectively. Figure 6 shows the error curves as a function of time for fixed values of x. The values of x chosen correspond to maxima in the oscillations of the spatial part of the solution. Figure 7 shows the error curves

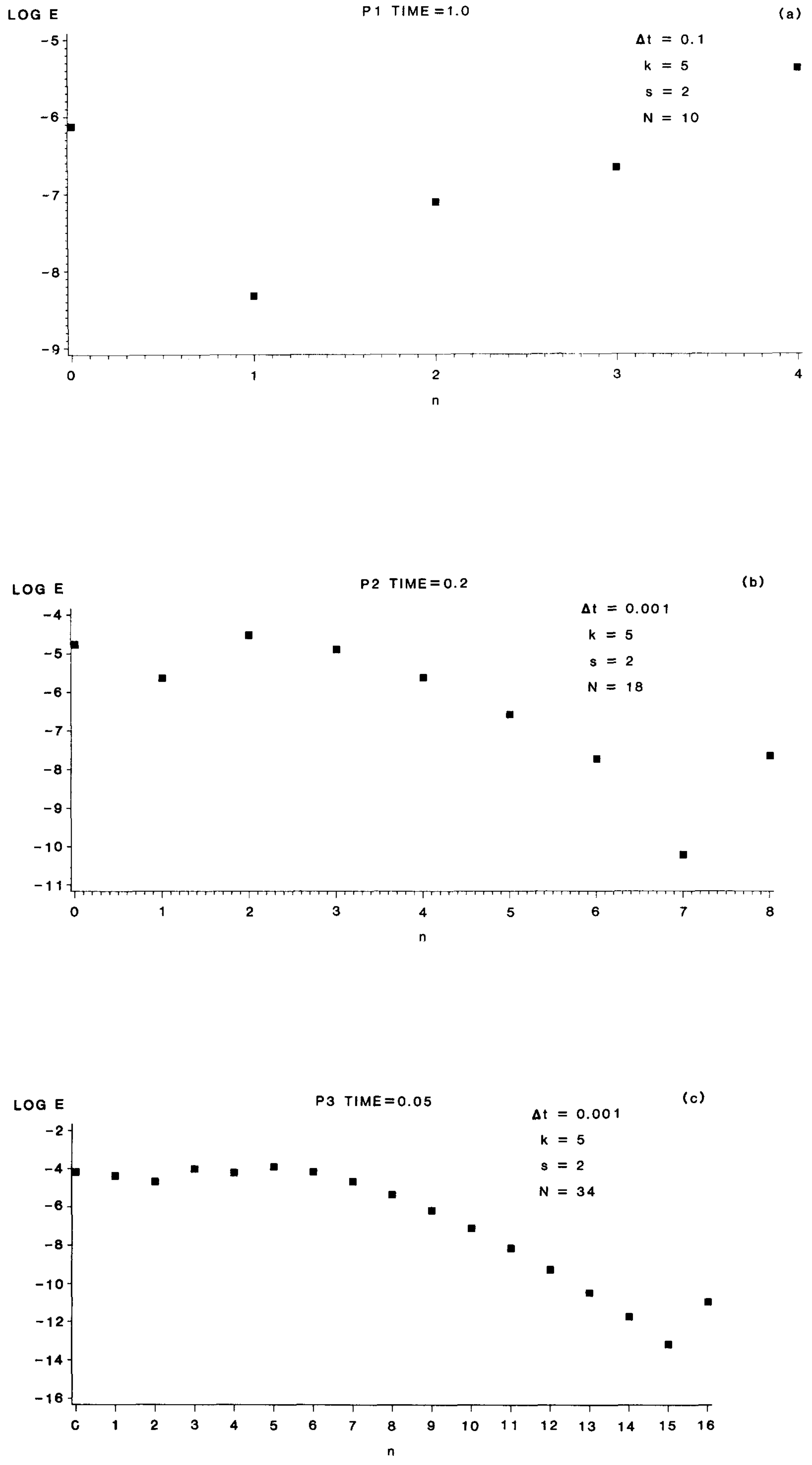

FIG. 5. The logarithm of the error in the calculated values of Chebyshev coefficients a_{2n+1} for the linear problems of (a) P1, (b) P2, and (c) P3, respectively, at the values of t shown. Only the odd Chebyshev spectral function basis of Eq. (14) was used and therefore $n = N/2 - 1$ is the number of spectral functions retained. For the SSFM the parameters of the PC method defined in Sec. II D have the values shown.

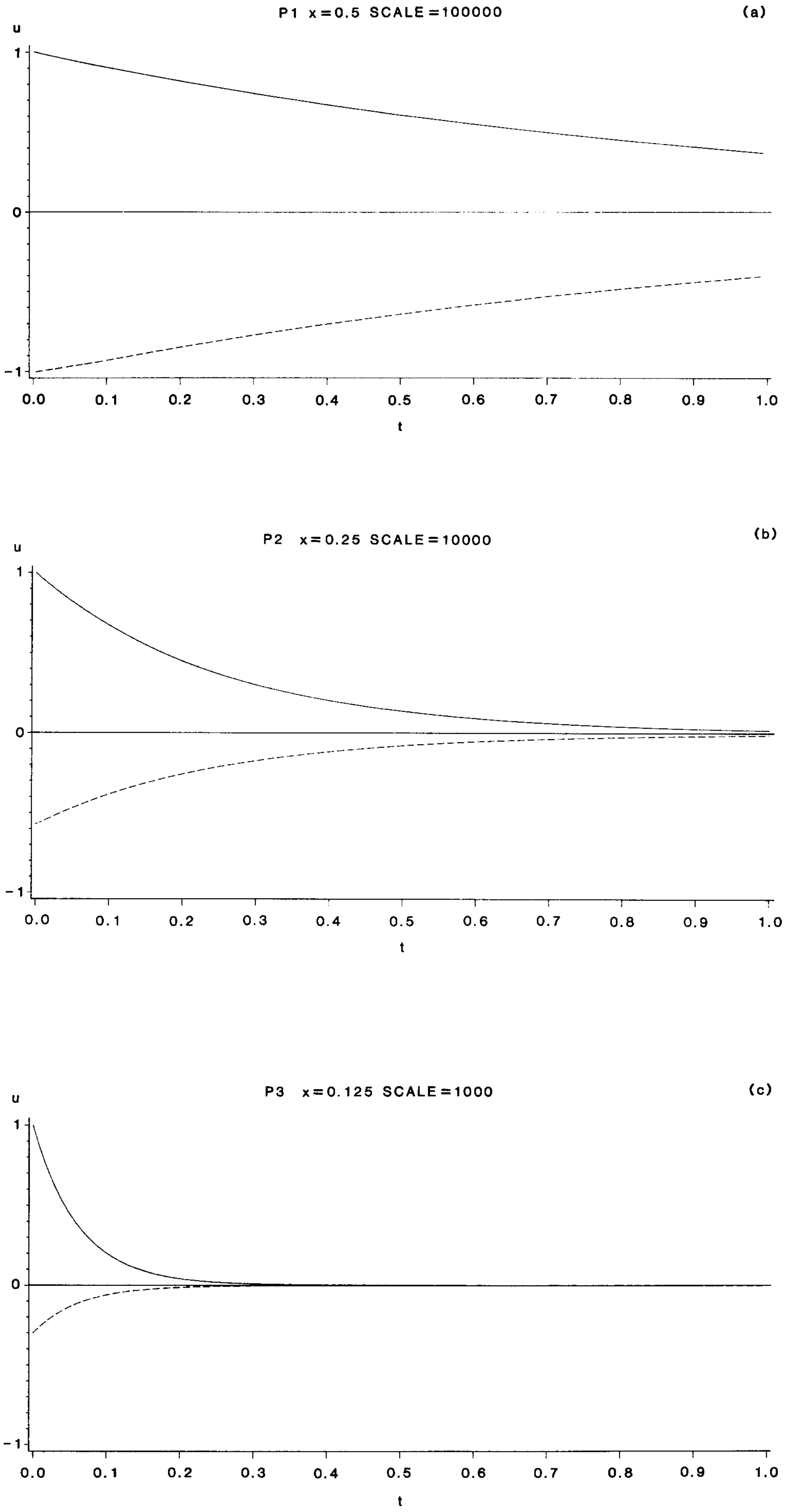

FIG. 6. Error curves as a function of time for the fixed values of x shown for the linear problems of cases (a) P1, (b) P2, and (c) P3, respectively. The unbroken curves are exact solutions and the broken curves are exact minus approximate solutions amplified by the factor SCALE. The approximate solutions were generated with the number of odd Chebyshev spectral functions and PC parameters as given in the corresponding parts of Fig. 5.

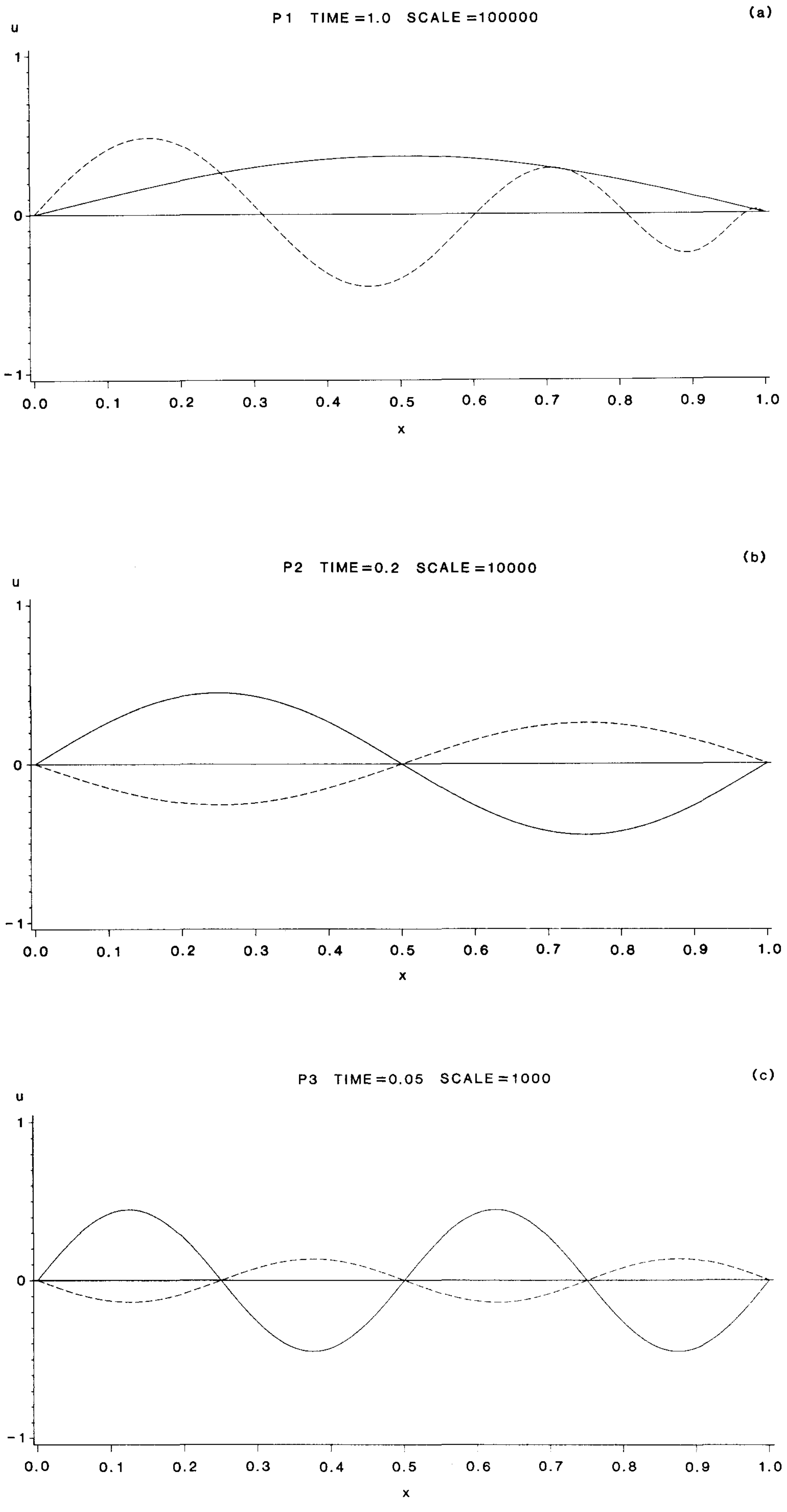

FIG. 7. Error curves as a function of x for the fixed values of time shown for the linear problems of cases (a) P1, (b) P2, and (c) P3, respectively. The unbroken curves are exact solutions and the broken curves are exact minus approximate solutions amplified by the factor SCALE. The approximate solutions were generated with the number of odd Chebyshev spectral functions and PC parameters as given in the corresponding parts of Fig. 5.

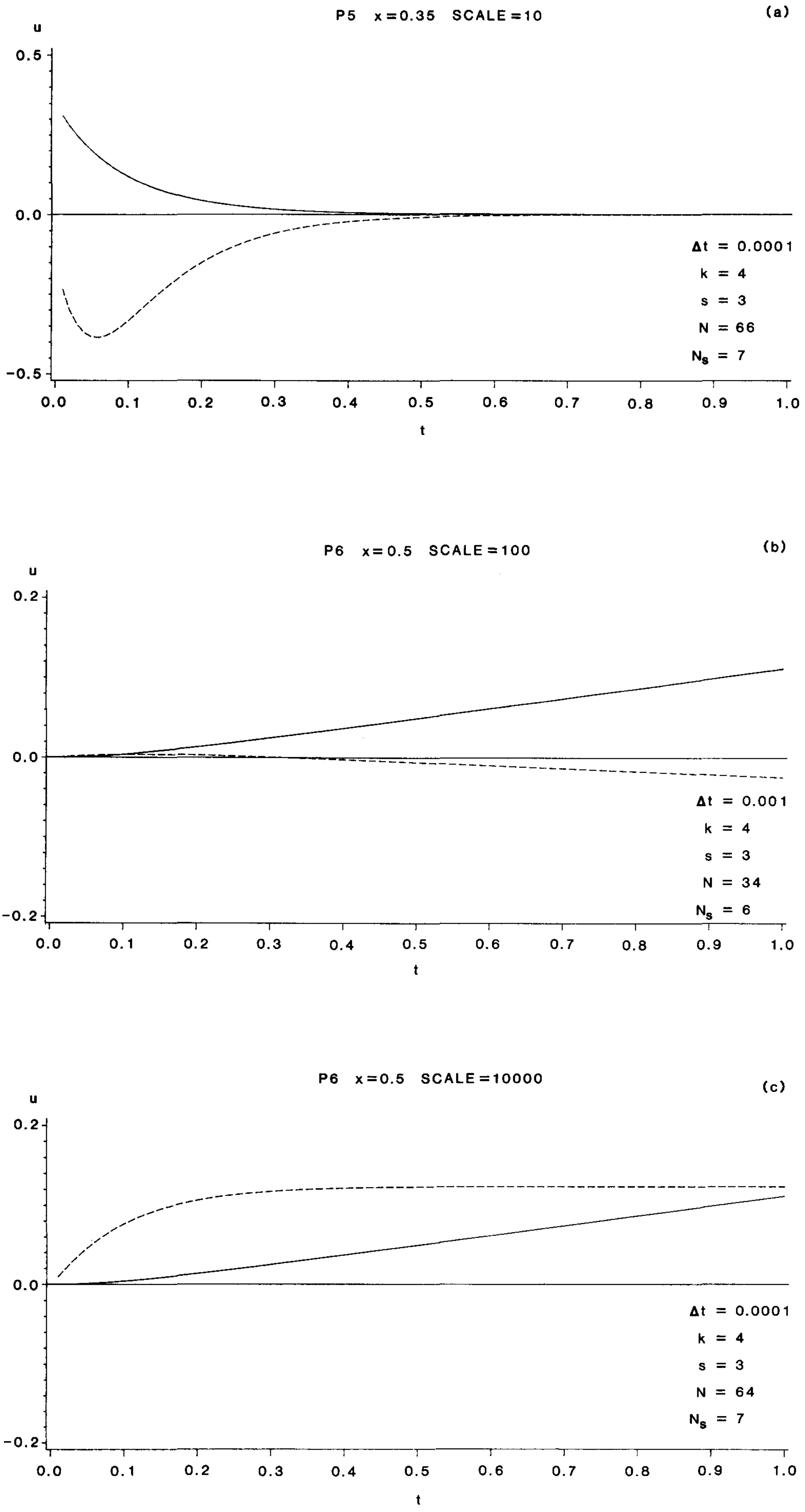

FIG. 8. Error curves as a function of time for the fixed values of x shown for the nonlinear problems of cases (a) P5, (b) P6, and (c) P6, respectively. The unbroken curves are exact solutions and the broken curves are exact minus approximate solutions amplified by the factor SCALE. The approximate solutions were generated from the PC parameters shown. For case P5 only odd Chebyshev spectral functions were used and the number retained was $n = 66/2 - 1$. For case P6 both even and odd Chebyshev spectral functions were used in Eq. (13) and the number retained was, respectively, (b) $n = 34 - 1$ and (c) $n = 64 - 1$. Here N_s is the number of components retained in the spectral space time integration of the vector $\mathbf{p}^{(z)}$ of Eq. (39). For case P5 the exact result was the finite element solution of Rektorys[8] while in case P6 the exact solution was computed from $N = 64$ and $N_s = 7$ and RK parameters of $\Delta t = 0.000\,001$, $k = 5$.

as a function of x for the fixed values of time t corresponding to Fig. 5. Note the correlation between the maximum error on the interval $x \in [0,1]$ as shown in Fig. 7 and the maximum error in the Chebyshev coefficients shown in Fig. 5.

For cases P5 and P6 solutions in closed form are not known. However, Rektorys[8] has computed approximations using a finite element method for case P5 and a time discretized variational method for case P6. These solutions of Rektorys will be treated as "exact" although they are not as accurate as solutions obtained in the present work using the SSFM and DSFM (see Sec. VI). Case P5 has an initial value $u_0(r)$ which is triangular with a discontinuity at $r = 0.5$ (see Sec. II A). The finite element solution as shown in Fig. 2(a) has two additional discontinuities at $r = 0.25$ and $r = 0.75$ which are not present in the initial value. The present work uses an approximation in continuous polynomial functions and it is to be expected that the convergence close to the discontinuity will be poor due to the Gibbs phenomenon at this point (see Gottlieb and Orszag[9]). As a consequence an expansion of the solution in polynomial (or trigonometric) functions is slowly convergent. Thus case P5 is a particularly challenging one in comparing approximation schemes.

In contrast to P5, case P6 is a continuous function which has a good approximation[8] as a linear combination of $\sin(\pi x)$ and $\sin(3\pi x)$ at each discrete time point. Therefore no difficulties are expected for a polynomial method in this case. Figures 8–11 show results of the SSFM applied to the nonlinear problems P5 and P6. For case P5 only the odd Chebyshev polynomial spectral basis is used. This corresponds to retaining only the first term of Eq. (13). Therefore only the odd Chebyshev coefficients a_1,a_3 and a_5,..., are present. For case P6 both odd and even polynomial spectral bases are used and this corresponds to retaining both terms of Eq. (13). The even polynomial basis elements are necessary because the inhomogeneous term of P6 has even polynomial terms.

Figure 8 shows the error curves as a function of time for fixed values of x. The values of x chosen correspond to the maximum in the solution of case P6 and a point between the discontinuities of the finite elements chosen by Rektorys for

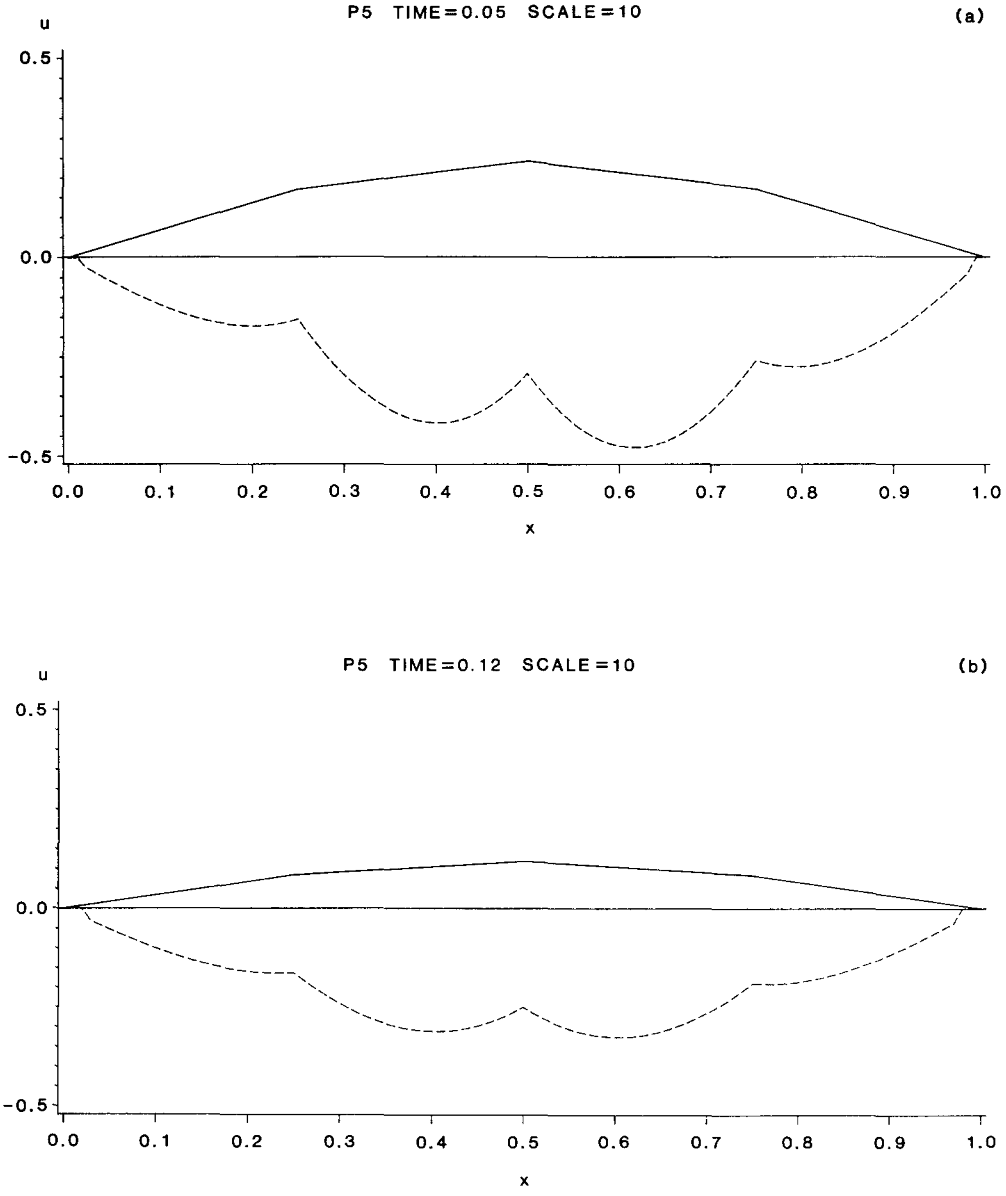

FIG. 9. Error curves as a function of x for the fixed values of time shown for the nonlinear problem of case P5. The unbroken curves are "exact" solutions and the broken curves are "exact" minus approximate solutions amplified by the factor SCALE. The approximate solutions were generated from the PC parameters of Fig. 8(a) and only $n = \frac{66}{2} - 1$ odd Chebyshev spectral functions were used. In this figure the "exact" result was the finite element solution of Rektorys.[8]

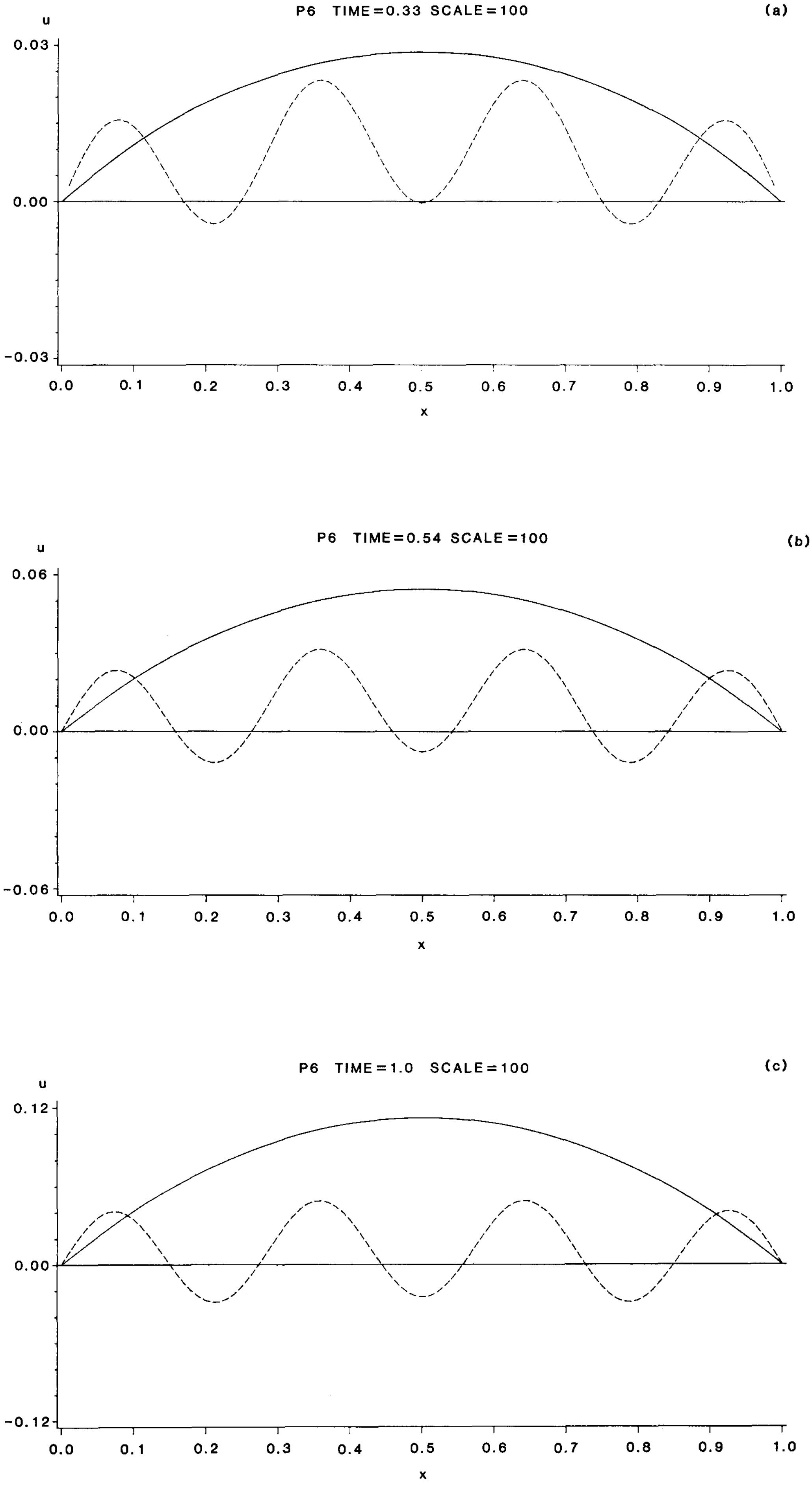

FIG. 10. Error curves as a function of x for the fixed values of time shown for the nonlinear problem of case P6. The unbroken curves are "exact" solutions and the broken curves are "exact" minus approximate solutions amplified by the factor SCALE. The approximate solutions were generated from the PC parameters given in Fig. 8(b). Both even and odd Chebyshev spectral functions were used in Eq. (13) with the number retained being $n = 34 - 1$. In this figure the "exact" solution was computed from $N = 64$, $N_s = 7$, and RK parameters of $\Delta t = 0.000\,001$, $k = 5$.

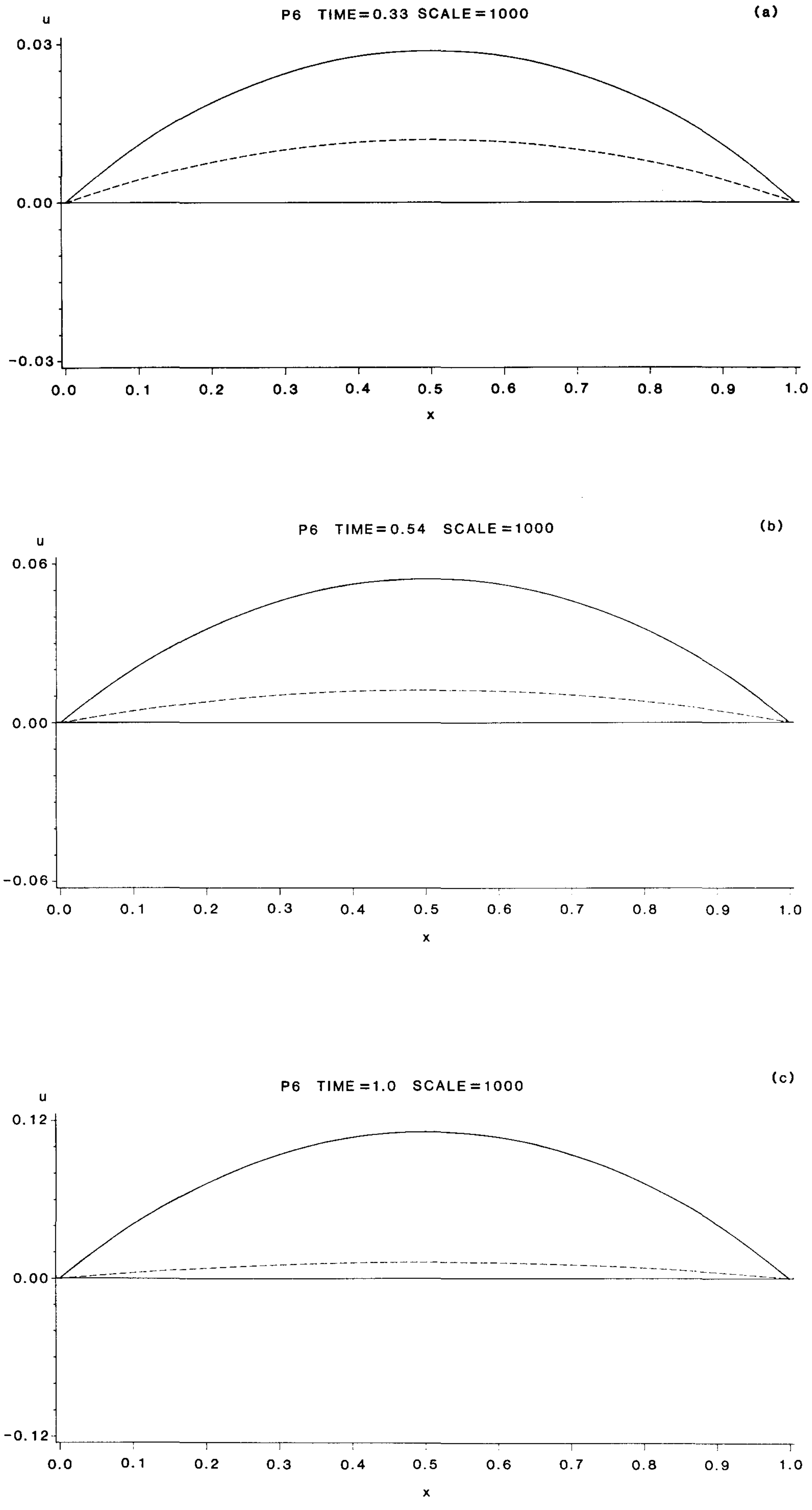

FIG. 11. Error curves as a function of x for the fixed values of time shown for the nonlinear problem of case P6. The unbroken curves are "exact" solutions and the broken curves are "exact" minus approximate solutions amplified by the factor SCALE. The approximate solutions were generated from the PC parameters given in Fig. 8(c). Both even and odd Chebyshev spectral functions were used in Eq. (13) and the number retained was $n = 64 - 1$. In this figure the "exact" solution was computed from $N = 64$, $N_s = 7$, and RK parameters of $\Delta t = 0.000\,001$, $k = 5$.

case P5. The values of the PC parameters Δt, k, and s as defined in Sec. II D are shown. The parameter N is defined in Eq. (7) and $n = N/2 - 1$ is the number of odd polynomial spectral functions defined in Eq. (14). In case P6 $N/2$ is the number of even polynomial spectral functions defined in Eq. (15). In case P6 the result of Fig. 8(b) is obtained with the same number of odd spectral functions as was used in the linear problem shown in Fig. 6(c) with $N/2$ even spectral functions added for the reasons given above. The error in this nonlinear problem is typically an order of magnitude larger than for the linear case of Fig. 6(c). Figure 8(c) demonstrates the precision possible by reduction of the step size and increase in the number of spectral functions. At the end point the improvement is a factor of approximately 20. For case P6 Figs. 10 and 11 show the error curves as a function of x for the fixed values of time t given. Figures 10 and 11 correspond, respectively, to the calculations of Figs. 8(b) and 8(c) and show that the relative error decreases as the solution grows with time. In Figs. 8–11 the exact solution was computed with $N = 64$ and $N_s = 7$ and RK parameters of $\Delta t = 0.000\,001$, $k = 5$. The solution so obtained is considerably more accurate than that given by Rektorys which would correspond to using only eigenmodes $N_s = 1$ and 3.

For case P5 Fig. 8(a) shows the error curve as a function of time for a fixed value of $x = 0.35$ and Fig. 9 shows the corresponding error curves as a function of x for the fixed values of time t given. The exact result in this case is generated by Rektorys[8] with three finite elements at each discrete time point. The three distinct ridges parallel to the t axis correspond to the knots at $r = 0.25$, 0.5, and 0.75. At these points the finite element method is more accurate and this is the reason for minima in the error curves of Fig. 9. The apparently larger error shown in Figs. 8(a) and 9 is more the result of the finite element approximation rather than the spectral method. However, the latter method is not without problems, as discussed above, when the solution is a discontinuous function. This is seen in the polynomial method on inspection of the Chebyshev expansion coefficients of the initial value $u_0(r)$ which is triangular with a discontinuity at $r = 0.5$. These coefficients were calculated with a high precision Newton–Cotes quadrature correct to the sixth decimal. Only odd polynomials are present and it was found that while $a_1(t = 0)$ is of the order of 0.2 the last coefficient retained in the SSFM calculation for case P5, namely, $a_{65}(t = 0)$ is of the order of 0.000 27. This is a very slow convergence for a Chebyshev method, where the coefficients usually decrease exponentially with the index $2n + 1$. This problem may lead to significant errors in the truncation of the convolution product discussed in Sec. III A. In view of the challenging nature of case P5 it was chosen to evaluate the DSFM as discussed in the next section.

VI. RESULTS FOR THE DOUBLE SPECTRAL FUNCTION METHOD

The DSFM as described in Sec. II E was applied to the linear cases P1 and P4. The simplest example is case P1 in the Fourier spectral basis of Eq. (11) where there is only one term, namely, $\sin(\pi x)$, corresponding to $n = 1$ with eigenvalue $\lambda = -1^2$. In this special case the mappings of Eqs. (42) and (43) have matrices with unit entries on the diagonal and zero elsewhere. Therefore the recurrence of Eq. (63) generates the expansion coefficients of e^{-t} in a series of shifted Chebyshev polynomials. It is a simple calculation to perform the recurrence from $M = 6$ and normalize the resulting coefficients according to Eq. (65) which in this case is e^{-t} evaluated at $t = 0$. Working to seven decimals this gives exact agreement with the values tabulated on p. 313 of Ref. 20.

A less trivial example is the linear case of P4 in the odd Chebyshev spectral basis of Eq. (14). Inspection of case P4 shows that the solution is simply the inhomogeneous term minus the solution of case P1. Therefore the DSFM consists of the recurrence Eq. (63), normalization as in Eq. (65), and then the recurrence of Eq. (64). This is all performed in spectral space and the negative sign is included at the normalization step. The matrix of coefficients in physical space defined in Eq. (55a) then follows from the back transform of Eq. (59a). The constant used in normalization is the forward transform, as given in Eq. (42), of the Chebyshev coefficients of the initial value in case P1. Similarly, the coefficients Q_{nm} are obtained by the forward transform of the physical space coefficients E_{nm} of the inhomogeneous term as defined by Eq. (57b). In case P4 the inhomogeneous term does not depend on time and the matrix $\mathbf{E}$ has only the first column as nonzero. This column is obtained from the vector $\mathbf{f}$ as described in the discussion following Eq. (56).

Table I compares the DSFM on the Chebyshev spectral basis with results of the SSFM using both Chebyshev and Fourier spectral bases for the PC parameters shown. The exact result is calculated from the closed form solution. The DSFM result is for the two iterations described above, namely, one application of Eq. (63) with $M = 4$, followed by one application of Eq. (64). All three methods used $N = 14$ corresponding to $n = 14/2 - 1$ spectral functions although only $n = 1$ is nonzero in the case of the Fourier SSFM. All three methods agree with the analytical result to within one or two digits in the fifth decimal.

While the matrix $\mathbf{A}$ is dense, the forward transformed matrix $\mathbf{P}$ has only one row for cases P1 and P4 because there is only one eigenmode. Therefore there is only one set of recurrences Eqs. (63) and (64) corresponding to $n = 1$. This is self-evident if the problem is solved in the Fourier sine spectral basis. However, in the case of nonlinear problems

TABLE I. Solution for case P4 at $x = 0.5$.

t	Exact	Chebyshev SSFM[a]	Fourier SSFM[a]	Chebyshev DSFM[b]
0.20	0.181 269	0.181 267	0.181 267	0.181 261
0.40	0.329 680	0.329 676	0.329 676	0.329 681
0.60	0.451 188	0.451 183	0.451 183	0.451 174
0.80	0.550 671	0.550 666	0.550 671	0.550 668
1.00	0.632 121	0.632 116	0.632 136	0.632 107
Time[c]		0.18	0.18	0.19

[a] With $N = 14$, $N_s = 1$, and PC parameters $\Delta t = 0.2$, $k = 4$, $s = 3$.
[b] With $N = 14$, $M = 4$, and two iterations.
[c] Execution time in seconds.

several eigenmodes $n = 1$ to N_s are present simultaneously in spectral space. Therefore each of the recurrences in Eqs. (63) and (64) has to be performed N_s times. The Chebyshev DSFM was applied to the nonlinear problem of case P5 using step 0 of Algorithm 2 (Sec. III C) for the initial iterate and steps 1–4 for subsequent iterates. The initial iterate consists of the recurrence Eq. (63) applied N_s times followed by a normalization for each of the N_s sets of coefficients. In the case of P5 the initial value of the solution is known as an expansion in odd Chebyshev coefficients as discussed in Sec. V. Therefore the set of N_s normalization constants in spectral space is obtained by the forward transform in Eq. (42). Subsequent iterations apply the same normalization requirement at each step. Inspection of the physical space coefficients defined by Eqs. (9) and (55a) at different times shows that nine iterations and $M = 13$ gives convergence to the accuracy of the initial value coefficients, namely, six decimals.

Table II compares the performance of the Chebyshev DSFM and SSFM. Values obtained from the finite element solution of Rektorys are also shown. The agreement between the SSFM and DSFM is better than with the finite element result. One source of error in the spectral function method is the truncation in the convolution product of Eq. (66). However, calculations with $N = 50$ and 42 for two iterations showed a maximum error of six digits in the fifth decimal. Thus the agreement between DSFM and SSFM results for times less than 0.15 suggests that most of the error shown in Figs. 8(a) and 9 is due to the finite element solution. A comparison of the computing times as given in Table II shows that in a nonlinear problem the DSFM is approximately six times faster than the SSFM.

VII. CONCLUSIONS

Two spectral function methods were applied to model studies of linear and nonlinear diffusion in one dimension. For the nonlinear cases the nonlinearity is in the inhomogeneous term and occurs as a (positive) integer power of the unknown solution. Although the linear problems are trivial they were used to assess the resolving power of the spectral functions and to provide a benchmark against which performance for the nonlinear problems was compared.

In the single spectral function method polynomial spectral functions in the spatial variable were introduced. Both Chebyshev polynomial and Fourier bases were used in the construction of spectral functions which were orthonormalized. Since the Fourier spectral basis has been investigated in detail by others,[9,11] the present numerical experiments concentrated mainly on the Chebyshev case. Because only integer powers of the solution occurred in the nonlinear term, a simple property of both Chebyshev (or Fourier) series was applied. This is the convolution product in which multiplication of two series is written as a single series. Projection of the diffusion equation onto the Hilbert space spanned by the orthonormalized set of spectral functions gave a matrix problem. Diagonalization of the matrix in the Chebyshev case showed that the rate of convergence of eigenvalues as a function of matrix size was more rapid than was the case for a simple Schrödinger operator studied previously.[14] A spectral resolution of the diffusion equation was obtained by a spectral transform which mapped the equation from physical to spectral space. In the case of linear problems the spectral resolution showed a single eigenmode and each eigenmode corresponded to a different eigenvalue of the matrix. In the nonlinear problems these eigenmodes were coupled together in physical space but the spectral transform in combination with the convolution product gave a system of uncoupled, simultaneous ODE's in spectral space. The spectral space mapping of the diffusion equation was integrated in time using predictor–corrector or Runge–Kutta integration methods and, for the nonlinear problems, this required computation of the inhomogeneous term at each time step.

The double spectral method introduced polynomial spectral functions in both space and time variables and thereby eliminated the need for time integration. A recurrence scheme was applied in spectral space and nonlinear terms were evaluated by convolution in physical space after application of the inverse spectral transform. The method was iterated until it converged, typically within a few cycles.

Both spectral function methods were compared against analytical solutions for the linear cases and against numerical solutions for the nonlinear cases. The DSFM was found to be faster than the SSFM by typically a factor of 6. Stable and accurate solutions were found in linear diffusion problems and also a nonlinear problem with a continuous initial value. Less accurate but stable solutions were found for a nonlinear problem with a discontinuous initial value.

A general formulation of both methods was given to facilitate the application to more complicated transport equations. In particular, work is in progress on a system of highly nonlinear coupled diffusion equations which arise in plasma theory.[16]

TABLE II. Solution for case P5 at $x = 0.35$.

t	Rektorys[a]	Chebyshev SSFM[b]	Chebyshev DSFM[c]	Chebyshev DSFM[d]
0.05	0.200 311	0.227 335	0.221 357	0.220 777
0.10	0.119 348	0.142 407	0.138 785	0.136 602
0.15	0.072 080	0.088 087	0.083 500	0.082 105
0.20	0.043 745	0.054 083	0.051 316	0.050 071
0.25	0.026 596	0.033 093	0.027 632	0.030 462
Time[e]		568	92	461

[a] Reference 8.
[b] With $N = 66$, $N_s = 7$, and PC parameters $\Delta t = 0.0001$, $k = 4$, $s = 3$.
[c] With $N = 66$, $M = 5$, and two iterations.
[d] With $N = 66$, $M = 13$, and five iterations.
[e] Execution time in seconds.

ACKNOWLEDGMENTS

This research was performed under contract to the Plasma Physics Division of the Atomic Energy Corporation of South Africa Limited. The author is grateful for the use of computing facilities at the University of the Witwatersrand Computer Center where computations were performed. He also thanks Mr. V. M. Gorringe and Mr. S. M. Malherbe for programming assistance.

[1]R. Balescu, *Equilibrium and Nonequilibrium Statistical Mechanics* (Wiley–Interscience, New York, 1975).
[2]H. Risken, *The Fokker–Planck Equation: Methods of Solution and Applications* (Springer, Berlin, 1984).
[3]E. W. Montroll and M. F. Schlesinger, in *Nonequilibrium Phenomena II: From Stochastics to Hydrodynamics*, edited by J. L. Lebowitz and E. W. Montroll (North-Holland, Amsterdam, 1984).
[4]J. Crank, *The Mathematics of Diffusion* (Oxford U. P., Oxford, 1983).
[5]R. D. Richtmyer and K. W. Morton, *Difference Methods for Initial-Value Problems*, 2nd ed. (Wiley–Interscience, New York, 1967).
[6]*A Survey of Numerical Methods for Partial Differential Equations*, edited by I. Gladwell and R. Wait (Oxford U. P., Oxford, 1979).
[7]A. F. Ghoniem and F. S. Sherman, J. Comput. Phys. **61**, 1 (1985).
[8]K. Rektorys, *The Method of Discretization in Time and Partial Differential Equations* (Reidel, Dordrecht, 1982).
[9]D. Gottlieb and S. A. Orszag, *Numerical Analysis of Spectral Methods: Theory and Application* (SIAM, Philadelphia, 1981).
[10]R. Peyret and T. D. Taylor, *Computational Methods for Fluid Flow* (Springer, New York, 1983).
[11]S. A. Orszag, Stud. Appl. Math. **L**, 293 (1971).
[12]S. G. Mikhlin and K. L. Smolitskiy, *Approximate Methods for Solution of Differential and Integral Equations* (American Elsevier, New York, 1967).
[13]L. V. Kantorovich and G. P. Akilov, *Functional Analysis*, 2nd ed. (Pergamon, New York, 1982).
[14]G. Delic and G. H. Rawitscher, J. Comput. Phys. **57**, 188 (1985).
[15]G. H. Rawitscher and G. Delic, J. Math. Phys. **27**, 816 (1986).
[16]G. Delic, J. D. Neethling, and D. F. B. Louw, Bull. Am. Phys. Soc. **31**, 1447 (1986).
[17]D. Elliot, Proc. Cambridge Philos. Soc. **57**, 823 (1961).
[18]D. Elliot, J. Austral. Math. Soc. **1**, 344 (1960).
[19]D. Elliott, Comput. J. **6**, 102 (1963).
[20]Y. L. Luke, *The Special Functions and Their Approximations* (Academic, New York, 1969).
[21]L. Lapidus and J. H. Seinfeld, *Numerical Solution of Ordinary Differential Equations* (Academic, New York, 1971).
[22]G. Delic and S. M. Malherbe, unpublished.
[23]C. W. Clenshaw, Proc. Cambridge Philos. Soc. **53**, 134 (1957).
[24]G. Delic, Comput. Phys. Commun. **18**, 73 (1979).
[25]G. Delic and S. M. Malherbe, unpublished.
[26]*Numerical Algorithms Group Library Manual* (Mark II) (Numerical Algorithms Group, Oxford, 1984).

Islands of stability and complex universality relations

R. Delbourgo and P. Hughes
Department of Physics, University of Tasmania, Hobart, Australia 7005

B. G. Kenny
Department of Physics, University of Western Australia, Perth, Western Australia 6009

(Received 10 April 1986; accepted for publication 10 September 1986)

For complex mappings of the type $z \to \lambda z(1-z)$, universality constants α and δ can be defined along islands of stability lying on filamentary sequences in the complex λ plane. As the end of the filament is approached, asymptotic values $\alpha_N \sim \lambda_\infty^{N-1}$, $\delta_N/\alpha_N^2 \sim 1$ are attained, where $\mu_\infty = \lambda_\infty(\lambda_\infty - 2)/4$, is associated with the limiting form of the universal function for that sequence, $g(z) = 1 - \mu_\infty z^2$. These results are complex generalizations of the real mapping case (applying to tangent bifurcations and windows of stability) where $\mu_\infty = 2$ and $\delta/\alpha^2 \to \frac{2}{3}$ correspond to the filament running along the real axis.

I. INTRODUCTION

In many branches of physics there is a real gain in understanding by extending the analysis of real dynamical variables into the complex plane. For dynamical nonlinear maps in one variable the advantages of such a generalization are much less obvious, but this has not hindered several researchers[1-3] from pursuing these studies. In this report we wish to point out the existence of complex universality relations between the analogs of the Feigenbaum constants α and δ as they pertain to various "kneading" sequences in the complex plane, thereby generalizing earlier work of ours[4] applying to the real case.

II. THEORY

We focus as usual on the Julia–Fatou mapping

$$z \to \lambda z(1-z), \tag{1a}$$

or

$$z \to 1 - \mu z^2, \tag{1b}$$

with $\mu = \lambda(\lambda - 2)/4$ providing the equivalence between (1a) and (1b). The detailed fractal nature of these maps have been highlighted by Mandelbrot[1] and Cvitanovic and Myrheim[5] and many fascinating properties have emerged. One of the most interesting of these[2] is the fact that "islands of stability" belonging to the Mandelbrot set are all connected to the main cactus by filaments. As a special instance, the real filament (with $\mathrm{Im}\,\lambda = \mathrm{Im}\,\mu = 0$) will connect the windows of stability that are seen in the counterpart real problem; that particular filament ends at $\lambda_\infty = 4$ or $\mu_\infty = 2$. If one follows the kneading sequences of the same type along that real filament then one can discover that as the end is approached, the universal constants associated with N-replication attain the asymptotic values,

$$\alpha \sim 4^{N-1}, \quad \delta/\alpha^2 \to \tfrac{2}{3}.$$

Cvitanovic and Myrheim[3,5] have shown that subcacti sprouting from the main Mandelbrot cactus allow one to define complex Feigenbaum constants which characterize their universal rates of shrinkage as one follows a particular Farey sequence. They have provided extensive lists of δ and α values connected with Farey sets m/n. In this note we will be examining other sequences which represent the analog of tangent bifurcations for real nonlinear maps. These sequences correspond to isolated islands of stability associated with N-replication that lie *outside* the main Mandelbrot cactus and its many leaves. It will be recalled that there exist as many as 2^N superstable N-cycles for quadratic maximum maps *some of which* touch the main cactus (and thus belong to the Farey sets). The islands we are referring to are the ones which do *not* touch the cactus, and they are generally quite small in size. In fact some are so miniscule that they are not at all easy to locate without some finesse. In Table I we have listed all of the island locations up to $N = 6$ for the readers' convenience, while in Table II we give the main islands for $N = 7$ and 8 that are relevant to our work, in that they lie on certain filaments. Also listed are some of the computed α and δ values for those N-plications as one follows a filament (see Figs. 1 and 2), which were determined by solving the complex versions of the functional equations.

III. ASYMPTOTIC RELATIONS

The values of α and δ are important only insofar as they relate to a particular kneading sequence. (The reader can convince himself that the asterisked sets do belong to the same sequences by tracing out the limit cycles in z.) There are of course an infinite number of such sequences or filaments in the complex plane. Our point is that for each of these sequences one can find an asymptotic δ–α relationship, showing that the two constants are not really independent of one another. In order to establish the α–δ relation, we turn to the renormalization group equations for them; they apply to complex just as well as real maps. The crucial observation is that as one approaches the end of a filament the renormalization group function $\overset{N}{g}(z)$ that determines α, namely

$$[\overset{N}{g}]^{N}(z) = -\overset{N}{g}(-\alpha z)/\alpha, \tag{2}$$

tends to the limiting quadratic form ($N \to \infty$),

$$\overset{\infty}{g}(z) = 1 - \mu_\infty z^2, \tag{3}$$

where μ_∞ labels the end point of the filament. This result is readily verified by solving Eq. (2) for successively large values of N and noticing how quickly the coefficients g^n in the

 0022-2488/87/010060-04$02.50

TABLE I. Superstable λ-values and cycle constants. A single asterisk refers to the rightmost real filamentary sequence, a double asterisk to the main complex filament, and a triple asterisk represents a subsidiary filament sequence. A # connotes a leaf on the main cactus. There exist other complex conjugate sequences and constants as well as reflected ones (where $\lambda \to 2 - \lambda$), which have been deliberately omitted.

Cyc. No.	Re λ	Im λ	Re α	Im α	Re δ	Im δ
2*	2.236 07	0	2.503	0	4.669	0
3*	3.831 87	0	9.277	0	55.26	0
3	2.552 65	0.959 46	2.097	− 2.358	4.600	8.981
4*	3.961 56	0	38.82	0	981.6	0
4#***	1.998 77	1.061 43	1.135	− 3.260	− 0.853	18.11
4**	2.741 22	1.185 66	10.56	− 5.375	100.4	− 69.34
5	3.743 91	0	20.13	0	255.5	0
5	3.905 71	0	45.80	0	1287.1	0
5*	3.990 27	0	160	0	16 931	0
5	1.667 62	0.988 55	0.380	− 3.554	− 9.520	26.37
5	2.841 45	0.611 22	3.874	− 2.181	18.97	14.56
5***	2.041 61	1.233 69	9.554	− 9.078	114.3	− 184.7
5	2.627 10	1.212 69	− 5.85	− 17.47	281.9	54.46
5**	2.808 89	1.216 51	40.43	− 2.526	1399	253
5	3.473 87	0.307 47	5.583	22.09	205.4	287.1
6	3.627 56	0	20.9	0	218.4	0
6	3.844 57	0	115.0	0	8508	0
6	3.937 54	0	207.6	0	28 020	0
6*	3.997 58	0	645.0	0	279 130	0
6	1.485 25	0.889 66	− 0.160	− 3.626	− 20.66	0
6	3.365 02	0.203 24				
6	2.610 95	1.068 40				
6	2.624 84	1.256 12	27.65	64.54	− 1458	4208
6	2.781 41	1.232 49	− 8.274	− 69.98	− 4270	462.9
6**	2.830 89	1.217 39	124.2	41.97	8769	12142
6	2.960 83	0.676 23	19.92	− 7.643	421.8	− 188.9
6	1.673 45	1.107 62	8.284	− 11.43	91.37	− 342.8
6***	2.080 40	1.267 61	37.22	− 8.569	1772	34.60
6	1.974 85	1.239 54	− 6.824	− 23.89	− 777.4	− 5.382

expansion,

$$\overset{N}{g}(z) = \sum_{n=0}^{\infty} \overset{N}{g}_n z^{2n} = 1 + \overset{N}{g}_1 z^2 + \cdots,$$

settle down to the limiting case (3). Of course the particular case of the rightmost cycle sequence for real maps, when $\mu_\infty = 2$, is now well known.[6] In much the same way, the renormalization group function $h(z)$, which determines δ,

$$\sum_{m=0}^{\infty} h([g]^{N-m})[g]^{m\prime}([g]^{N-m+1}) = -\delta/\alpha, \qquad (2')$$

also has coefficients h_n (in a power series in z^{2n}), which die off rapidly with n. The technique[4] for deriving a relation between α and δ makes great use of these facts.

One truncates the m"th iterate of g (here we drop the N superscript),

$$[g]^m(z) = -a_m + b_m z^2 + \cdots,$$

and uses the recurrence property $[g]^{m+1} = g([g]^m)$ to derive the recurrence relations of the truncated coefficients,

$$a_{m+1} = \mu_\infty a_m^2 - 1, \quad b_{m+1} = 2\mu_\infty a_m b_m. \qquad (4)$$

These formulas become more and more accurate for large N. Referring to (2) we also readily see that

$$a_N = 1/\alpha, \quad b_N = \mu_\infty \alpha \qquad (5)$$

offer separate ways of calculating α provided that the end point μ_∞ of the filament is known. Now, from the recur-

TABLE II. Superstable λ-values and some cycle constants for selected islands of stability corresponding to $N = 7,8,9$ replication.

Cycle	Re λ	Im λ	Re α	Im α	Re δ	Im δ
7*	3.999 40	0	2603	0	4.51E6	0
7**	2.837 39	1.215 05	310.8	269.8	− 2.2E4	1.5E5
7***	2.098 49	1.271 46	97.2	27.8	6436	10555
8*	3.999 85	0	10424	0	7.24E7	0
8**	2.839 03	1.213 56	567.1	1142	− 1.1E6	8.4E5
8***	2.105 37	1.269 36	179.7	180.2	− 35140	70041
9*	3.999 9	0	41715	0	1.16E9	0
9**	2.839 3	1.212 6				
9***	2.107	1.267				

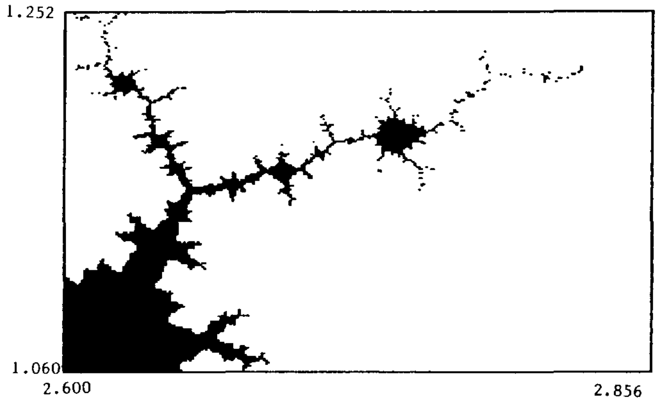

FIG. 1. Strings of islands in the range $2.600 < \mathrm{Re}\, z < 2.856$, $1.06 < \mathrm{Im}\, z < 1.252$. The ** filament extends to the right and terminates.

rence property (4), one easily derives

$$b_N = (2\mu_\infty)^{n-1}\mu_\infty \prod_{m=2}^{N-1} a_m$$

or (6)

$$\alpha = d(2\mu_\infty)^{N-1} \quad \text{with} \quad d = \prod_{m=2}^{N-1} a_m.$$

So it only remains to estimate the product d of the a-coefficients; this is possible by working downwards in m via (4), starting with (5):

$$a_{N-1} = \pm [(1 + 1/\alpha)/\mu]^{1/2}, \qquad a_{m-1} = \pm [(1 + a_m)/\mu]^{1/2}, \quad m < N. \tag{7}$$

The only *significant* ambiguity in (7) is the choice of sign for the root. This depends on the kneading sequence and completely characterizes it in fact. For instance the rightmost real filamentary sequence has the same root sign throughout.

Let us define the branch cut of $z^{1/2}$ to lie along the negative real axis (i.e., $|\arg z| < \pi$). Then for the ** complex filament sequence (see Fig. 1), the roots have to be taken as

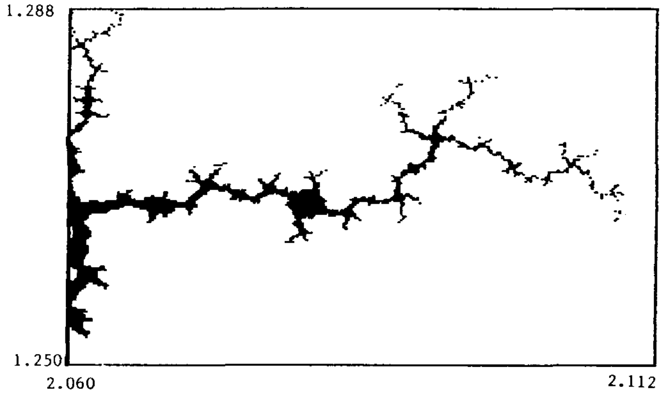

FIG. 2. Strings of islands in the range $2.060 < \mathrm{Re}\, z < 2.112$, $1.25 < \mathrm{Im}\, z < 1.288$. The *** filament branches to the right; another large filament straggles upwards.

$$a_m \leftrightarrow +, \quad m = N-1, N-2, \ldots, 3 \qquad a_2 \text{ and } a_1 \leftrightarrow -, \tag{8**}$$

whereas for the *** sequence the correct choice of signs is

$$a_m \leftrightarrow +, \quad m = N-1, \ldots, 4 \qquad a_3 \,\&\, a_2 \,\&\, a_1 \leftrightarrow -. \tag{8***}$$

[These choices can be verified directly by iterating (4) up in m, starting with $a_1 = -1$, which method gives no sign ambiguity.] The relevant point about this procedure is that for N asymptotic, we get to a good approximation,

$$a_{N-1} = -(1/\mu_\infty)^{1/2}, \ldots, \quad a_2 = \mu_\infty - 1, \quad a_1 = -1,$$

with *most* of the coefficients a_m (of a particular root sign) settling down to solutions of the algebraic equation,

$$a = \mu_\infty a^2 - 1$$

or

$$a = [(1 \pm (1 + 4\mu_\infty)^{1/2})/2\mu_\infty] = (\lambda_\infty / 2\mu_\infty)$$

or

$$(2 - \lambda_\infty)/2\mu_\infty.$$

Consequently a *crude* estimate of the asymptotic form of the product d is

$$d_{**} \simeq (\lambda/2\mu)^{N-3}\mu^{-1/2}, \quad \alpha_{**N} \simeq 4\mu^{3/2}\lambda^{N-3}, \tag{9**}$$

$$d_{***} \simeq (1-\mu)(\lambda/2\mu)^{N-4}\mu^{-1/2}, \qquad \alpha_{***N} \simeq (2\mu)^3(1-\mu)\lambda^{N-4}, \tag{9***}$$

each evaluated at the respective end points of the filaments. Clearly these estimates can be much improved numerically, but the salient prediction is that as $N \to \infty$

$$\lim \alpha_{N+1}/\alpha_N = \lambda_\infty,$$

which is well substantiated by the numerical facts; thus

$$\alpha_{**8}/\alpha_{**7} = 2.86 + 1.21i \quad (\text{cf. } \lambda_\infty = 2.84 + 1.20i),$$

$$\alpha_{***8}/\alpha_{***7} = 2.20 + 1.23i \quad (\text{cf. } \lambda_\infty = 2.11 + 1.27i),$$

although $N = 8$ is hardly an asymptotic number.

Turning next to the eigenvalue equation for δ, one finds that the coefficients h_n in the expansion,

$$h(z) = \sum_{h=0}^{\infty} h_n z^{2n} = 1 + h_1 z^2 + \cdots, \tag{10}$$

vanish very rapidly with $n (> = 1)$ for large N, ensuring that the lowest order approximation of (2′) has exactly the same form as in the real case,[4] namely

$$\sum_{m=0}^{N} [g]^{m\prime}([g]^{N-m+1}(0)) \simeq -\delta/\alpha. \tag{11}$$

Since $g(z) \simeq 1 - \mu_\infty z^2$ and $g'(z) \simeq -2\mu_\infty z$, one has

$$[g]^{n\prime}([g]^{N-n+1}(0)) = (2\mu_\infty)^n \prod_{j=N-n+1}^{N} a_j.$$

Remembering the definition d of (6), and its relation to α, we may rewrite (11) as

$$\frac{\delta}{\alpha^2}=\left[1-\frac{1}{2\mu_\infty}-\frac{1}{(2\mu_\infty)^2a_2}-\frac{1}{(2\mu_\infty)^3a_2a_3}-\cdots-\frac{1}{(2\mu_\infty)^{N-1}a_2a_3\cdots a_{N-1}}\right]\equiv c,$$

where c is completely fixed by μ_∞ and the kneading sequence of root signs for the a_m. Thus even in the complex case, we discover that δ is proportional to α^2. The facts bear out this prediction:

$$(\delta_8\alpha_8^{-2})_{***}=1.083+0.540\,i,\quad \text{while}\quad (\delta_7/\alpha_7^{-2})_{***}=1.081+0.544\,i,$$

and

$$(\delta_8\alpha_8^{-2})_{**}=0.837+0.245\,i,\quad \text{while}\quad (\delta_7\alpha_7^{-2})_{**}=0.835+0.248\,i.$$

Of course one can also estimate the coefficient c by approximating a with $(\lambda/2\mu)$ for the most part but the results are not especially reliable and we will therefore omit details at this point. There is one important point though that needs exposing. This has to do with the algebraic determination of the end point λ_∞. Consider the ** sequence. As $N\to\infty$, a_3 can be evaluated in one of two ways; either as $\mu(\mu-1)^2-1$ by going up in N for $N=2$, or as a solution $(\lambda/2\mu)$ of $\mu a^2-a=1$. This means that

$$2\mu^2(1-\mu)^2=\lambda+2\mu,$$

yielding a quartic in λ with roots $\lambda=0$, 4, $2.83929\pm1.21258i$, wherein we recognize the last pair as corresponding to the ** filament (see Table I). Similarly, the *** filament end point is obtained as a root of the equation

$$2\mu^2(1-\mu)^2=-\lambda+2\mu,$$

and occurs in the vicinity of $\lambda=2.107\pm1.267i$.

Clearly all the above considerations can be extended to other filamentary sequences. In all the cases we can anticipate that

$$\alpha_N\propto\lambda_\infty^{N-1}\quad\text{and}\quad \delta\propto\alpha_N^2,$$

with the proportionality constants depending on the end value μ_∞ of the filament and the kneading sequence. These results represent the complex generalization of the δ–α relation found in the real case.[4] We envisage that the same ideas will find application in circle maps.

[1]B. B. Mandelbrot, Ann. N.Y. Acad. Sci. **357**, 249 (1980).

[2]A. Douady and B. Hubbard, C. R. Acad. Sci. Paris **294**, 123 (1982).

[3]P. Cvitanovic and J. Myrheim, Phys. Lett. A **94**, 329 (1983).

[4]R. Delbourgo, W. Hart, and B. G. Kenny, Phys. Rev. A **31**, 514 (1985); Aust. J. Phys. **39**, 189 (1986); R. Delbourgo and B. G. Kenny, "Universality relations," Phys. Rev. A **33**, 3292 (1986).

[5]P. Cvitanovic and J. Myrheim, "Complex universality," Nordita preprint, 1984.

[6]J. P. Eckmann, H. Epstein, and P. Wittwer, Commun. Math. Phys. **93**, 495 (1984).

Classical mechanics with respect to an observer's past light cone

G. H. Derrick
School of Physics, University of Sydney, Sydney, New South Wales 2006, Australia

(Received 19 March 1986; accepted for publication 3 September 1986)

This work is motivated by the fact that it is impossible for an observer to know at time t_0 all the initial data of a system, if that data is specified in the conventional manner on the spacelike surface $t = t_0$. A Hamiltonian formulation for classical mechanics, first given by Dirac, is exploited, in which dynamical variables are specified by their values on an observer's past light cones. Starting from initial data given on the past light cone of an observer at some initial space-time point, the values of the variables on the observer's current past light cone are given by a canonical transform of the initial data. The method is illustrated for a spinless particle of mass m, which is either free or interacts with an external electromagnetic field. The remarkable result is obtained that the dynamics of this classical one-particle spin-0 system can be formulated in terms of a Dirac-like spinor, whose four components are formed from the generalized coordinates and momenta.

I. THE PROBLEM

In the conventional approach to both classical and quantum theory one predicts the future behavior of a system from initial data given at a certain time t_0, or more generally, given on a specified spacelike surface. Thus in a classical system with coordinates $q_1,q_2,\ldots$ and conjugate momenta $p_1,p_2,\ldots$ the values of these variables at time t_0 determine their values at all subsequent times t. The same is true of the corresponding quantum operators in the Heisenberg picture or of the quantum state vector in the Schrödinger picture.

However a difficulty arises because the finiteness of the velocity of light c rules out instantaneous information transfer. *At time t_0 no single observer can ever know the initial data belonging to that time.* While an infinite set of observers spread throughout the spacelike surface $t = t_0$ could know all this data collectively, no individual observer can make any use of what his fellow observers currently know. Suppose that at the present moment t_0 an observer wishes to make predictions relating to an experiment he intends to perform at some future time t at his spatial origin. The only part of space-time that can have any relevance to the outcome of the experiment will be within its past light cone (idealizing the experiment to be a point event). Thus in order to make a prediction from the initial data at time t_0 about the experiment to be performed at the later time t the observer needs the t_0 data within a sphere of radius $c(t - t_0)$. He cannot know this data at time t_0 but must wait for it to be transmitted to him by apparatuses or auxiliary observers spread throughout the sphere. If transmitted with the maximum signal velocity c, the data from the most distant regions will reach the central observer at time $t_0 + (t - t_0) = t$, just as the experiment is being carried out! Thus even with a comprehensive and reliable deterministic theory an observer can make no firm predictions about the future, but at best can merely have the satisfaction of verifying the consistency of the theory with the outcomes of different experiments *all performed within his past light cone.*

The observer can attempt to circumvent this dilemma by making the hypothesis that no event liable to influence the proposed experiment will occur outside some distance R, assumed to be less than $c(t - t_0)$. Typically R might have dimensions less than or comparable with those of the laboratory. The observer can then restrict his attention to the region inside a sphere of radius R and make his prediction at time $t_0 + R/c$, which is now earlier than the time t of the experiment. If his predictions turn out to be correct he may feel some reassurance concerning the validity of the theory and of his supplementary hypothesis. But what should be concluded if his prediction is wrong? Is the theory inadequate, or did something, contrary to hypothesis, indeed arrive from outside radius R? As a safeguard against unexpected incoming influences one might imagine covering the surface of the sphere of radius R with appropriate detectors. Unfortunately the information that one of these detectors has fired must always arrive too late to enable the observer to update his predictions in time.

The above discussion is on the basis of simple measurements idealized as point events. The situation is even worse for more complex measurements that extend over finite regions of space-time and hence involve auxiliary observers or apparatuses. In order to set up and effect such extended experiments the central observer must send instructions to his collaborators at other locations. Using signals of the optimum velocity c the most efficient extended measurements he can orchestrate lie on his future light cone. In summary, what an observer knows, and what he can have measured, lie, respectively, on or within his past and future light cones.

In a classic paper[1] Dirac derived a number of alternative Hamilton forms of relativistic dynamics, which differ from the conventional "instant" form in how initial data is prescribed. The present paper exploits a particular limiting case of Dirac's "point" form of dynamics in which the concept of initial data "at the present time t_0" is replaced by that of initial data on the observer's "current past light cone." The latter information, unlike the former, is available to the observer "here–now" without any need to wait for the future arrival of data transmitted from distant apparatuses or auxiliary observers. The price to pay is that initial data on the past

 0022-2488/87/010064-07$02.50

light cone is in general not sufficient to determine future behavior. This adds further uncertainty, over and above that inherent in quantum theory, to any predictions that an observer makes. He must always qualify his predictions by some phrase such as "on the basis of my present knowledge."

Accepting that past light cone data is the most that any observer can know, is there a "best" estimate he can make for the probability distributions of given dynamical variables in some future experiment? As a first step towards finding an answer to this question, the present paper develops a past light cone Hamiltonian approach to classical mechanics in preparation for quantum treatment in subsequent papers. The mode of treatment differs from that of Dirac in the choice of conjugate variables.

II. LIGHT CONE COORDINATES

Consider an observer whose (timelike) trajectory in four-dimensional Minkowski space is given in parametric form by

$$x^\lambda = z^\lambda(\tau)\,. \tag{1}$$

Here $z^\lambda(\tau)$, $\lambda = 0,1,2,3$, are four functions of the parameter τ, the latter being taken as the proper interval $\int (\eta_{\lambda\mu}\, dz^\lambda\, dz^\mu)^{1/2}$ measured along the trajectory from some arbitrary event.[2] Thus τ/c is the time elapsed since this initial event until the current "here–now" as experienced by the observer and recorded on an ideal clock carried by him. The four-velocity $v^\lambda = dz^\lambda/d\tau$ is required to be a future-pointing timelike unit vector, i.e., $\eta_{\lambda\mu} v^\lambda v^\mu = 1$, $v^0 \geqslant 1$, but otherwise the observer is allowed to move arbitrarily.

Let us focus attention on a particular value of τ, corresponding to the event $z^\lambda(\tau)$ being the observer's current "here–now." Then

$$x^\lambda = z^\lambda(\tau) + y^\lambda\,, \tag{2}$$

$$y^0 = -y\,, \tag{3}$$

where $y = |\mathbf{y}|$, is the equation of the past light cone with vertex at the observer. Here the past-pointing null vector y^λ serves to parametrize the light cone. Only three components of y^λ are independent on account of (3) and it is most convenient to choose as independent parameters the spatial components $\mathbf{y} = (y^1,y^2,y^3)$.

We may regard (2) and (3) as defining a change of coordinates from the original inertial coordinates x^λ to a new set (τ,y^1,y^2,y^3). These equations may be inverted and (τ,y^1,y^2,y^3) written as functions of x^λ. We first solve $z^0(\tau) - x^0 = |\mathbf{z}(\tau) - \mathbf{x}|$ to obtain $\tau = \tau(x^\lambda)$, the unique parameter value where the future light cone with vertex x^λ intersects the trajectory (1). Then $y^\kappa = x^\kappa - z^\kappa(\tau(x^\lambda))$ defines y^κ as a function of x^λ.

Writing $(x^0)' = \tau$, $(x^1)' = y^1$, $(x^2)' = y^2$, $(x^3)' = y^3$, the metric tensor in the primed system has components

$$(g_{\iota\kappa})' = \eta_{\lambda\mu} \frac{\partial x^\lambda}{\partial (x^\iota)'} \frac{\partial x^\mu}{\partial (x^\kappa)'}$$

and

$$(g^{\iota\kappa})' = \eta^{\lambda\mu} \frac{\partial (x^\iota)'}{\partial x^\lambda} \frac{\partial (x^\kappa)'}{\partial x^\mu}$$

given by[2]

$$\begin{aligned} &(g_{00})' = 1\,, \quad (g_{0j})' = -(v^0 y^j + v^j y)/y, \\ &(g_{jk})' = (y^j/y)(y^k/y) - \delta_{jk}, \quad (g^{00})' = 0, \\ &(g^{0j})' = y^j/(v_\lambda y^\lambda), \\ &(g^{jk})' = (y^j v^k + y^k v^j)/(-v_\lambda y^\lambda) - \delta^{jk}, \end{aligned} \tag{4}$$

where

$$v_\lambda y^\lambda = v^0 y^0 - \mathbf{v}\cdot\mathbf{y} = -(v^0 y + \mathbf{v}\cdot\mathbf{y})\,. \tag{5}$$

The volume element is

$$d^4x = (-v_\lambda y^\lambda) d\tau (d^3\mathbf{y}/y)\,. \tag{6}$$

Note that the first factor $(-v_\lambda y^\lambda)$ is necessarily positive, while the last factor $(d^3\mathbf{y}/y)$ is the standard Lorentz invariant measure on a light cone.[3]

What we now attempt is to express classical mechanics in a form where the observer's proper time τ/c replaces coordinate time $t = x^0/c$ as the evolution parameter. Thus initial data is to be specified on the past light cones $\tau = \text{const}$ rather than on the spacelike surfaces $t = \text{const}$. For the rest of this paper we shall restrict our attention to the specific problem of a single spinless particle of rest mass m, which is either free or suffers electromagnetic interactions. Section III first reviews the conventional treatment of such a system, and then the subsequent sections reformulate the theory in terms of past light cone data.

III. REVIEW OF CONVENTIONAL HAMILTONIAN THEORY FOR A CLASSICAL PARTICLE

The free motion of a classical spinless particle of mass m may be derived from the variation principle

$$\delta\left\{ -mc \int (\eta_{\lambda\mu}\, dx^\lambda\, dx^\mu)^{1/2} \right\} = 0\,. \tag{7}$$

Let $\mathbf{x} = \mathbf{q}(t)$ be the position vector of the particle at time t, and $\dot{\mathbf{q}} = d\mathbf{q}/dt$ its three-velocity vector. Then (7) assumes the form

$$\delta \int L\, dt = 0\,, \tag{8}$$

with

$$L = -mc^2(1 - \dot{\mathbf{q}}^2/c^2)^{1/2}\,. \tag{9}$$

Introducing the conjugate momentum vector $\mathbf{p} = \partial L/\partial \dot{\mathbf{q}}$ then leads to the Hamiltonian $H = \mathbf{p}\cdot\dot{\mathbf{q}} - L = c(\mathbf{p}^2 + m^2c^2)^{1/2}$ and the associated equations of motion

$$\frac{d\mathbf{q}}{dt} = \frac{\partial H}{\partial \mathbf{p}} = \frac{c^2\mathbf{p}}{H}\,, \quad \frac{d\mathbf{p}}{dt} = -\frac{\partial H}{\partial \mathbf{q}} = 0\,. \tag{10}$$

Here $H \equiv cp^0$ and $\mathbf{p}$ are interpreted as the energy and linear momentum, respectively, with the future-pointing timelike four-vector $p^\lambda \equiv (p^0,\mathbf{p})$ satisfying

$$\eta_{\lambda\mu} p^\lambda p^\mu = m^2c^2\,. \tag{11}$$

As a consequence of (10) any dynamical variable $f(\mathbf{q},\mathbf{p},t)$ evolves with time according to

$$\frac{df}{dt} = \{f,H\} + \frac{\partial f}{\partial t}\,, \tag{12}$$

where the Poisson bracket between any two functions f,g of $\mathbf{q},\mathbf{p}$ is defined by the convention

$$\{f,g\} = \left(\frac{\partial f}{\partial \mathbf{q}}\right)\cdot\left(\frac{\partial g}{\partial \mathbf{p}}\right) - \left(\frac{\partial f}{\partial \mathbf{p}}\right)\cdot\left(\frac{\partial g}{\partial \mathbf{q}}\right). \tag{13}$$

The angular momentum tensor $J^{\lambda\mu}$, by assumption, has no spin content, and is given by

$$(J^{23},J^{31},J^{12}) = \mathbf{q}\times\mathbf{p}\,, \quad (J^{01},J^{02},J^{03}) = ct\mathbf{p} - p^0\mathbf{q}\,. \tag{14}$$

Together p^λ and $J^{\lambda\mu}$ satisfy the Poisson brackets appropriate for the generators of the Poincaré group:

$$\{p^\lambda,p^\mu\} = 0\,, \tag{15}$$

$$\{J^{\iota\kappa},p^\lambda\} = \eta^{\kappa\lambda}p^\iota - \eta^{\iota\lambda}p^\kappa\,, \tag{16}$$

$$\{J^{\iota\kappa},J^{\lambda\mu}\} = \eta^{\iota\mu}J^{\kappa\lambda} + \eta^{\kappa\lambda}J^{\iota\mu} - \eta^{\iota\lambda}J^{\kappa\mu} - \eta^{\kappa\mu}J^{\iota\lambda}\,. \tag{17}$$

With the aid of the above brackets and the evolution equation (12) we find that both p^λ and $J^{\lambda\mu}$ are constant in time for a free particle.

When an external electromagnetic field derived from a vector potential $A^\lambda(x^\nu)$ is present, (7) must be changed to[4]

$$\delta\int\left\{-mc(\eta_{\lambda\mu}\,dx^\lambda\,dx^\mu)^{1/2} - \frac{e}{c}A_\lambda\,dx^\lambda\right\} = 0\,, \tag{18}$$

where e is the charge of the particle. the new Lagrangian leads to the Hamiltonian

$$H \equiv cp^0 = c[\{\mathbf{p} - (e/c)\mathbf{A}\}^2 + m^2c^2]^{1/2} + eA^0\,. \tag{19}$$

Instead of (11) one has

$$\eta_{\lambda\mu}\{p^\lambda - (e/c)A^\lambda\}\{p^\mu - (e/c)A^\mu\} = m^2c^2\,, \tag{20}$$

which corresponds to the standard prescription $p^\lambda \to p^\lambda - (e/c)A^\lambda$ for incorporating an electromagnetic field. However the Poincaré symmetry has been broken by the field, so that (15)–(17) are no longer true. To restore conservation of linear and angular momentum one would need to include the contributions from the electromagnetic field.

IV. CLASSICAL LIGHT CONE THEORY FOR A FREE PARTICLE

This section gives two alternative classical Hamiltonian formulations for the mechanics of a free particle of mass m. The first (Secs. IV A and IV B) uses $\mathbf{y}$ as past light cone coordinate, and the second (Secs. IV C and IV D) takes as coordinates spinor variables from which $\mathbf{y}$ may be constructed. The modifications necessary to accommodate interaction with an external electromagnetic field are considered in Sec. V.

A. Classical Hamiltonian formulation for coordinates $\mathbf{y}(\tau)$

We can define a particle trajectory in a four-dimensional space by specifying what functions any three of the coordinates are of the fourth coordinate for points lying on the path. In the conventional treatment of Sec. III we took inertial coordinates $(t,\mathbf{x})$ and wrote $\mathbf{x} = \mathbf{q}(t)$. Here, instead, we shall take as coordinates the light cone coordinates $(\tau,\mathbf{y})$ associated with an observer with the given world line (1), with the definitions of Sec. II. The particle trajectory is then specified by $\mathbf{y} = \mathbf{y}(\tau)$. Thus when the observer's clock reads τ/c he sees the particle at relative position $\mathbf{y}(\tau)$ by light which has just arrived from the particle. To $\mathbf{y}(\tau)$ we adjoin the fourth component $y^0(\tau) = -|\mathbf{y}(\tau)|$ to obtain the four-vector $y^\lambda(\tau)$.

As before the starting point is the variational principle (7). Transformation to our new variables with the aid of the metric tensor $(g_{\lambda\mu})'$ given in (4) yields

$$\delta\int L\,d\left(\frac{\tau}{c}\right) = 0\,, \tag{21}$$

where

$$L = -mc^2[1 - 2(v^0\hat{\mathbf{y}} + \mathbf{v})\cdot\mathbf{u} + (\hat{\mathbf{y}}\cdot\mathbf{u})^2 - \mathbf{u}^2]^{1/2}$$
$$= -mc^2[1 + 2v_\lambda u^\lambda + u_\lambda u^\lambda]^{1/2}\,. \tag{22}$$

In (22), $\hat{\mathbf{y}} = \mathbf{y}/|\mathbf{y}|$, $u^\lambda = dy^\lambda/d\tau$, and v^λ is the four-velocity of the observer, a given function of τ. On account of the identity $y_\lambda u^\lambda = 0$ we have $u^0 = -\hat{\mathbf{y}}\cdot\mathbf{u}$. The conjugate variable to $\mathbf{y}$ is

$$\boldsymbol{\pi} = c^{-1}\frac{\partial L}{\partial \mathbf{u}} = -m^2c^3L^{-1}[\mathbf{u} + \mathbf{v} + (u^0 + v^0)\hat{\mathbf{y}}], \tag{23}$$

with corresponding Hamiltonian

$$K = c\mathbf{u}\cdot\boldsymbol{\pi} - L = -m^2c^4L^{-1}(1 + v_\kappa u^\kappa)\,. \tag{24}$$

We must now solve (23) for $\mathbf{u}$ as a function of $\mathbf{y},\boldsymbol{\pi},\tau$ so that K may then be written in terms of these same variables. This yields

$$u^\lambda = -v^\lambda - v_\kappa y^\kappa(\mathbf{y}\cdot\boldsymbol{\pi})^{-1}p^\lambda\,, \tag{25}$$

$$K/c = v_\lambda p^\lambda\,, \tag{26}$$

where the four functions $p^\lambda \equiv p^\lambda(\mathbf{y},\boldsymbol{\pi})$ are

$$p^0 = \tfrac{1}{2}(\mathbf{y}\cdot\boldsymbol{\pi})^{-1}(\boldsymbol{\pi}^2 + m^2c^2)y\,,$$
$$\mathbf{p} = \boldsymbol{\pi} - \tfrac{1}{2}(\mathbf{y}\cdot\boldsymbol{\pi})^{-1}(\boldsymbol{\pi}^2 + m^2c^2)\mathbf{y} \equiv \boldsymbol{\pi} - p^0\hat{\mathbf{y}}\,, \tag{27}$$

and satisfy the identity

$$\eta_{\lambda\mu}p^\lambda p^\mu = m^2c^2\,. \tag{28}$$

The Possion brackets between any two dynamical variables $f(\mathbf{y},\boldsymbol{\pi},\tau)$, $g(\mathbf{y},\boldsymbol{\pi},\tau)$ is now defined in the standard way as[5]

$$\{f,g\} = \left(\frac{\partial f}{\partial \mathbf{y}}\right)\cdot\left(\frac{\partial g}{\partial \boldsymbol{\pi}}\right) - \left(\frac{\partial f}{\partial \boldsymbol{\pi}}\right)\cdot\left(\frac{\partial g}{\partial \mathbf{y}}\right). \tag{29}$$

The evolution of a dynamical variable $f(\mathbf{y},\boldsymbol{\pi},\tau)$ is then determined by its Poisson bracket with the Hamiltonian K of (26):

$$\frac{df}{d\tau} = v_\lambda\{f,p^\lambda\} + \frac{\partial f}{\partial \tau}\,, \tag{30}$$

with p^λ given by (27). By direct evaluation we find that

$$\{p^\lambda,p^\mu\} = 0\,, \tag{31}$$

which combined with (30) shows that p^λ is constant.

The form of (30) together with the identities (28) and (31) indicate that p^λ is the energy momentum four-vector, as anticipated by the notation.

We now seek an angular momentum tensor $j^{\lambda\mu} = j^{\lambda\mu}(\mathbf{y},\boldsymbol{\pi})$ to act as the generator of Lorentz transformations. In addition to satisfying the Poincaré group generator relations (16) and (17), the angular momentum must obey further constraints implied by the four-vector character of y^λ:

$$\{j^{\iota\kappa},y^\lambda\} = \eta^{\kappa\lambda}y^\iota - \eta^{\iota\lambda}y^\kappa .$$

A suitable definition is

$$(j^{23},j^{31},j^{12}) = \mathbf{y}\times\boldsymbol{\pi}, \quad (j^{01},j^{02},j^{03}) = -y\boldsymbol{\pi}. \tag{32}$$

Note the relation $j^{\lambda\mu} = y^\lambda p^\mu - y^\mu p^\lambda$, which follows from the definitions (27) and (32).

Applying (30) now shows that the quantity $J^{\lambda\mu} = j^{\lambda\mu} + z^\lambda(\tau)p^\mu - z^\mu(\tau)p^\lambda$ is a constant of motion. We interpret $J^{\lambda\mu}$ as the angular momentum referred to the fixed origin $x^\lambda = 0$, and $j^{\lambda\mu}$ that relative to the observer whose trajectory is defined by (1).

Equation (27) for the momentum p^λ is equivalent to the expression derived by Dirac[1] when allowance is made for his different choice of conjugate variables, and for his use of light cones that are future rather than past. On making the trivial change to a past light cone formulation, Dirac's result becomes

$$p^\lambda = \rho^\lambda - \tfrac{1}{2}(y^\kappa\rho_\kappa)^{-1}(\rho^\mu\rho_\mu - m^2c^2)y^\lambda . \tag{33}$$

In (33) there are *four* generalized coordinate-momenta pairs y^λ, ρ^λ satisfying $\{\rho^\lambda,y^\mu\} = g^{\lambda\mu}$, with $y^0 \approx -|\mathbf{y}|$ being a subsidiary condition imposed weakly. We can convert (33) to (27) by taking $y^0 = -|\mathbf{y}|$ as a strong identity and defining $\boldsymbol{\pi} = \boldsymbol{\rho} + \rho^0\hat{\mathbf{y}}$ as the variable conjugate to $\mathbf{y}$.

B. The canonical transformation $\mathbf{q},\mathbf{p}\rightarrow\mathbf{y},\boldsymbol{\pi}$

The change of phase space variables from the conventional set $\mathbf{q},\mathbf{p}$ to our light cone set $\mathbf{y},\boldsymbol{\pi}$ can be effected by a canonical transformation. Let us define the function $\tau[t]$ by $z^0(\tau[t]) = ct$. The position vector of the observer at time t is then $\mathbf{x} = \mathbf{z}(\tau[t])$, and that of the particle at the retarded time $t - y(\tau[t])/c$ is $\mathbf{x} = \mathbf{z}(\tau[t]) + \mathbf{y}(\tau[t])$. Since the particle has the constant three-velocity $c\mathbf{p}/p^0$ its position vector at time t is

$$\mathbf{x} = \mathbf{q}(t) = \mathbf{z}(\tau[t]) + \mathbf{y}(\tau[t]) + (\mathbf{p}/p^0)y(\tau[t]) . \tag{34}$$

Solving (34) for $\mathbf{y}(\tau[t])$ yields

$$\mathbf{y} = \mathbf{q} - \mathbf{z} - (mc)^{-2}[D - (\mathbf{q}-\mathbf{z})\cdot\mathbf{p}]\mathbf{p},$$

$$y = (mc)^{-2}[D - (\mathbf{q}-\mathbf{z})\cdot\mathbf{p}]p^0, \tag{35}$$

where

$$D = \{[(\mathbf{q}-\mathbf{z})\cdot\mathbf{p}]^2 + (\mathbf{q}-\mathbf{z})^2(mc)^2\}^{1/2} . \tag{36}$$

In (35) and subsequently in Sec. IV B the argument (t) for $\mathbf{q}$ and $(\tau[t])$ for $\mathbf{z}$, $\mathbf{y}$, and y is suppressed for brevity. The positive square root must be taken for D in (36) to ensure that $y \geqslant 0$. Substituting (35) into (27) then gives $\boldsymbol{\pi}$ as a function of $\mathbf{q},\mathbf{p}$:

$$\boldsymbol{\pi} = \{[D + (\mathbf{q}-\mathbf{z})\cdot\mathbf{p}]/(\mathbf{q}-\mathbf{z})^2\}(\mathbf{q}-\mathbf{z}) . \tag{37}$$

The inverse transformation found by solving (35) and (37) for $\mathbf{q},\mathbf{p}$ is

$$\mathbf{q} = \mathbf{z} + 2(\boldsymbol{\pi}^2 + m^2c^2)^{-1}D\boldsymbol{\pi}, \tag{38}$$

$$\mathbf{p} = \boldsymbol{\pi} - \tfrac{1}{2}(\boldsymbol{\pi}^2 + m^2c^2)D^{-1}\mathbf{y}, \tag{39}$$

where

$$D = \mathbf{y}\cdot\boldsymbol{\pi} . \tag{40}$$

The quantities D in (36) and (40) are identically equal, and are Lorentz invariant and non-negative on account of the identity $D = -y_\lambda p^\lambda$. An operator analog of D plays an important role in the corresponding quantum theory.[6]

It remains to establish that the transformation $\mathbf{q},\mathbf{p}\rightarrow\mathbf{y},\boldsymbol{\pi}$ defined by (35), (37) is canonical. A suitable generating function[5] is

$$F(\mathbf{p},\mathbf{y},t) = \mathbf{p}\cdot(\mathbf{y}+\mathbf{z}) + (\mathbf{p}^2 + m^2c^2)^{1/2}y ,$$

the t dependence coming from the term $\mathbf{z}\equiv\mathbf{z}(\tau[t])$. That $\mathbf{q} = \partial F/\partial\mathbf{p}$ and $\boldsymbol{\pi} = \partial F/\partial\mathbf{y}$ do indeed lead to (35) and (37) is seen most readily from Eqs. (27) and (34). The transformed Hamiltonian is

$$H' = cp^0 - \frac{\partial F}{\partial t} = cp^0 - \mathbf{p}\cdot\left(\frac{d\mathbf{z}}{dt}\right)$$

$$= cv_\lambda p^\lambda \frac{d\tau[t]}{dt},$$

which is consistent with the evolution equation (30) based on the Hamiltonian function K given by (26).

C. Classical Hamiltonian formulation in spinor coordinates

Any past-pointing null vector y^λ can be written in terms of a contravariant $D^{((1/2)0)}$ spinor[2] ξ^A, $A = 1,2$, and the standard Pauli spin matrices σ^λ. Writing $\xi = \left[\begin{smallmatrix}\xi^1\\ \xi^2\end{smallmatrix}\right]$, $\xi^\dagger = [(\xi^1)^*,(\xi^2)^*]$ the correspondence is

$$y^\lambda = -\xi^\dagger\sigma^\lambda\xi . \tag{41}$$

Let us make this substitution in (22) with $\xi^A = \xi^A(\tau)$ and develop a Hamiltonian formalism based on ξ^A as coordinates rather than $\mathbf{y}$. The real and imaginary parts of ξ^A give us four real coordinates, but since ξ may be multiplied by an overall phase factor without changing y^λ such a factor must be regarded as physically irrelevant. Writing $d\xi/d\tau = \zeta$ we have

$$\frac{dy^\lambda}{d\tau} \equiv u^\lambda = -\zeta^\dagger\sigma^\lambda\xi - \xi^\dagger\sigma^\lambda\zeta, \tag{42}$$

which, when substituted into (22), yields

$$L = -mc^2[1 - 2v_\lambda(\zeta^\dagger\sigma^\lambda\xi + \xi^\dagger\sigma^\lambda\zeta) + 4(\zeta^\dagger\xi\xi^\dagger\zeta - \zeta^\dagger\zeta\xi^\dagger\xi)]^{1/2}.$$

The term $u_\lambda u^\lambda$ has been simplified using the standard Pauli matrix identities[7] $a^\dagger\sigma_\lambda bc^\dagger\sigma^\lambda d = 2(a^\dagger bc^\dagger d - a^\dagger dc^\dagger b)$, which hold for any four spinors a,b,c,d. Now define the conjugate variable to ξ^A by[8] $\pi_A = c^{-1}\,\partial L/\partial\zeta^A$, noting that $(\pi_A)^*$ is then conjugate to $(\xi^A)^*$ on account of the reality of L. Under Lorentz transformations π_A is a covariant $D^{((1/2)0)}$ spinor. Writing $\pi = [\pi_1,\pi_2]$, $\pi^\dagger = \left[\begin{smallmatrix}(\pi_1)^*\\ (\pi_2)^*\end{smallmatrix}\right]$, we have

$$\pi = -m^2c^3L^{-1}[v_\lambda\xi^\dagger\sigma^\lambda + 2(\xi^\dagger\xi\zeta^\dagger - \zeta^\dagger\xi\xi^\dagger)]. \tag{43}$$

Whence

$$\pi\xi \equiv \pi_A\xi^A = -m^2c^3L^{-1}v_\lambda\xi^\dagger\sigma^\lambda\xi,$$

which, being real, implies the constraint equation

$$\pi\xi = (\pi\xi)^* \equiv \xi^\dagger\pi^\dagger . \tag{44}$$

Only three of the real and imaginary components of the momenta π_A and $(\pi_A)^*$ are independent owing to the constraint (44), which has its origin in the physical irrelevance of overall phase factors in ξ.

The Hamiltonian is

$$K = c(\pi\zeta + \zeta^\dagger\pi^\dagger) - L$$
$$= -m^2c^4L^{-1}v_\lambda[v^\lambda - (\xi^\dagger\sigma^\lambda\zeta + \zeta^\dagger\sigma^\lambda\xi)]\,. \qquad (45)$$

On account of the constraint (44), a term $\Lambda(\pi\xi - \xi^\dagger\pi^\dagger)$ with Λ arbitrary may be added to the above K without altering the physical content of the theory. For simplicity we choose $\Lambda = 0$.

While (43) cannot be solved unambiguously for ζ in terms of ξ and π, it is nevertheless possible to use (43) to eliminate ζ from (45) and express K as a function of ξ and π and their complex conjugates. To do this we need the following identity:

$$L(m^2c^2\xi^\dagger\sigma^\lambda\xi + \pi\bar\sigma^\lambda\pi^\dagger)$$
$$= -m^2c^3(\pi\xi + \xi^\dagger\pi^\dagger)[v^\lambda - (\xi^\dagger\sigma^\lambda\zeta + \zeta^\dagger\sigma^\lambda\xi)]. \qquad (46)$$

Here

$$\bar\sigma^\lambda = \epsilon(\sigma^\lambda)^*\epsilon^\dagger\,, \qquad (47)$$

with $\epsilon = \left(\begin{smallmatrix}0 & 1\\ -1 & 0\end{smallmatrix}\right)$ takes the values $\bar{\boldsymbol\sigma}^0 = \boldsymbol\sigma^0$, $\bar{\boldsymbol\sigma} = -\boldsymbol\sigma$. The relation (46) is most readily proved by invoking its manifest vector character. For any point along the observer's trajectory it is possible to find an inertial system in which $v^\lambda = (1,0,0,0)$ at that point. For v^λ of this simple form (46) is easily proved from (43) by direct evaluation of both sides, and hence, being a four-vector, the relation holds in general.

With the aid of (46) the Hamiltonian of (45) again assumes the form (26), but now with

$$p^\lambda = (m^2c^2\xi^\dagger\sigma^\lambda\xi + \pi\bar\sigma^\lambda\pi^\dagger)/(\pi\xi + \xi^\dagger\pi^\dagger)\,. \qquad (48)$$

Once again p^λ will be interpreted as the energy-momentum vector, the expression (48) actually being equivalent to (27). To see this, note that $\boldsymbol\pi$, the conjugate variable to $\mathbf{y}$, and π, the conjugate spinor to ξ, are related by the equation

$$\boldsymbol\pi = -\mathrm{Re}[(\pi\boldsymbol\sigma\xi)/(\xi^\dagger\xi)]\,, \qquad (49)$$

which follows from (23), (42), and (43) after some algebra. When this expression for $\boldsymbol\pi$ is substituted into (27) the form (48) is obtained. A like substitution into (33) yields the spinor form of the angular momentum tensor:

$$j^{\lambda\mu} = -\mathrm{Re}[i\pi\sigma^{\lambda\mu}\xi]\,, \qquad (50)$$

where the spin coefficients $\sigma^{\lambda\mu} = \frac{1}{2}i(\bar\sigma^\lambda\sigma^\mu - \bar\sigma^\mu\sigma^\lambda)$ are given by

$$(\sigma^{23},\sigma^{31},\sigma^{12}) = -i(\sigma^{01},\sigma^{02},\sigma^{03}) = \boldsymbol\sigma\,. \qquad (51)$$

Consider now the Poisson bracket structure with respect to the spinor coordinate ξ and its conjugate variable π. For this purpose we initially ignore the constraint (44) (that $\pi\xi$ is real) and regard the real and imaginary parts of the components of ξ and π as eight independent variables. The Poisson bracket between two functions f,g of these variables is defined as[8]

$$\{f,g\} = \frac{\partial f}{\partial\xi^A}\frac{\partial g}{\partial\pi_A} + \frac{\partial f}{\partial(\xi^A)^*}\frac{\partial g}{\partial(\pi_A)^*}$$
$$-\ (f,g \text{ interchanged}). \qquad (52)$$

The τ evolution is once again given by (30). Direct application of the definition (52) yields the following brackets:

$$\{p^\lambda, D - D^*\} = 0\,, \qquad (53)$$
$$\{p^\lambda,p^\mu\} = -4im^2c^2(D-D^*)(D+D^*)^{-3}\,\mathrm{Re}[\pi\sigma^{\lambda\mu}\xi]\,, \qquad (54)$$
$$\{j^{\iota\kappa},j^{\lambda\mu}\} = \eta^{\iota\mu}j^{\kappa\lambda} + \eta^{\kappa\lambda}j^{\iota\mu} - \eta^{\iota\lambda}j^{\kappa\mu} - \eta^{\kappa\mu}j^{\iota\lambda}\,, \qquad (55)$$
$$\{j^{\iota\kappa},p^\lambda\} = \eta^{\kappa\lambda}p^\iota - \eta^{\iota\lambda}p^\kappa\,, \qquad (56)$$
$$\{j^{\lambda\mu},\xi\} = \tfrac{1}{2}i\sigma^{\lambda\mu}\xi\,, \qquad (57)$$
$$\{j^{\lambda\mu},\pi\} = -\tfrac{1}{2}i\pi\sigma^{\lambda\mu}\,. \qquad (58)$$

Above we have written $\pi\xi = D$ and used the spinor expressions (48) and (50) for p^λ and $j^{\lambda\mu}$, respectively.

If we now impose the constraint (44), i.e., $D = D^*$, at some initial value of τ, then (53) and the evolution equation (30) ensure that this constraint is subsequently maintained. The rhs of (54) is then zero, as is necessary if p^λ is to be interpreted as the energy-momentum vector, i.e., the generator of space-time translations. The brackets for $j^{\iota\kappa}$ given by (55)–(58) support the contention that this tensor be regarded as the angular momentum. Equations (55) and (56) are necessary conditions for the Poincaré generators, while (57) and (58) are required in order that ξ and π transform, respectively, as contravariant and covariant $D^{((1/2)0)}$ spinors.

Finally let us note that when the constraint that $\pi\xi$ be real is satisfied, the quantity $D \equiv \pi\xi$ is identical to the D appearing in (36) and (40). This follows readily from (49).

D. Classical Hamiltonian formulation in Dirac spinor form

The content of the previous subsection can be written economically and elegantly in terms of a four-component column vector Ξ, whose elements Ξ^P, $P = 1,2,3,4$, are defined by

$$\begin{bmatrix}\Xi^1\\ \Xi^2\end{bmatrix} = \left(\frac{mc}{2}\right)^{1/2}\xi, \quad \begin{bmatrix}\Xi^3\\ \Xi^4\end{bmatrix} = -(2mc)^{-1/2}\pi^\dagger\,. \qquad (59)$$

Under an infinitesimal Lorentz transformation $(x^\kappa)' = x^\kappa + \eta^{\kappa\lambda}\omega_{\lambda\mu}x^\mu$, parametrized by the antisymmetric matrix $\omega_{\lambda\mu}$, the spinors transform according to

$$\xi' = [I - (i/4)\omega_{\lambda\mu}\sigma^{\lambda\mu}]\xi\,,$$
$$\pi' = \pi[I + (i/4)\omega_{\lambda\mu}\sigma^{\lambda\mu}]\,,$$

where I is the 2×2 unit matrix and $\sigma^{\lambda\mu}$ is given by (51). Correspondingly Ξ transforms as a Dirac spinor in the chiral representation[9]:

$$\Xi' = \Xi + (\omega_{\lambda\mu}\gamma^\lambda\gamma^\mu/4)\Xi\,, \qquad (60)$$

with the Dirac matrices given by

$$\gamma^0 \equiv \beta = \begin{pmatrix}0 & -I\\ -I & 0\end{pmatrix}, \quad \boldsymbol\gamma = \begin{pmatrix}0 & \underline{\sigma}\\ -\underline{\sigma} & 0\end{pmatrix}.$$

The adjoint spinor is

$$\bar\Xi = \Xi^\dagger\beta = [(2mc)^{-1/2}\pi, -(mc/2)^{1/2}\xi^\dagger]\,.$$

Denoting the components of $\bar\Xi$ by $\bar\Xi_Q$, $Q = 1,2,3,4$, (52) yields the Poisson brackets

$$\{\Xi^P,\bar\Xi_Q\} = \tfrac{1}{2}(\gamma^5)^P{}_Q, \quad \{\Xi^P,\Xi^Q\} = 0 = \{\bar\Xi_P,\bar\Xi_Q\}, \qquad (61)$$

where

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}.$$

The energy-momentum vector p^λ and the angular momentum $j^{\lambda\mu}$ given by (48) and (50), respectively, become

$$p^\lambda = mc\bar{\Xi}\gamma^\lambda\Xi/(\bar{\Xi}\Xi)\,, \tag{62}$$

$$j^{\lambda\mu} = \tfrac{1}{2}\bar{\Xi}\gamma^5(\gamma^\lambda\gamma^\mu - \gamma^\mu\gamma^\lambda)\Xi\,, \tag{63}$$

while the constraint that $\pi\xi$ be real is simply

$$\bar{\Xi}\gamma^5\Xi = 0\,. \tag{64}$$

In summary, we can reproduce the free motion of a classical spinless particle by the following scheme. The dynamics is described by a Dirac spinor Ξ whose physically realizable values are constrained by the relation (64). The evolution of Ξ as seen by an observer with trajectory $x^\lambda = z^\lambda(\tau)$ and four-velocity $v^\lambda = dz^\lambda/d\tau$ is governed by

$$\frac{d\Xi}{d\tau} = v_\lambda\{\Xi,p^\lambda\} = v_\lambda(\bar{\Xi}\Xi)^{-1}\gamma^5(mc\gamma^\lambda - p^\lambda)\Xi\,. \tag{65}$$

Here p^λ is the energy-momentum given by (62) and the Poisson bracket has been evaluated according to (61). When his proper time is τ/c the observer sees the particle at the point with Minkowski coordinates $x^\lambda = z^\lambda(\tau) + y^\lambda$, where

$$y^\lambda = -(mc)^{-1}\bar{\Xi}\gamma^\lambda(1+\gamma^5)\Xi\,. \tag{66}$$

It should be emphasized that we are dealing here with the *classical* motion of a particle with *zero spin*. Despite the appearance of the Dirac spinor Ξ with transformation law (60), the angular momentum $j^{\lambda\mu}$ given by (63) is the *orbital* angular momentum relative to the observer, viz. $y^\lambda p^\mu - y^\mu p^\lambda$, and as such has no internal content. A like paradoxical appearance of a four-spinor occurs in Dirac's quantum mechanical description of a spin-0 particle.[10] In that model one has a Dirac spinor that satisfies an equation superficially like the spin-$\frac{1}{2}$ Dirac equation, but nevertheless the spin is certainly zero. This raises the question of whether the present formalism is a classical analog of Dirac's quantum model. The answer appears to be no. The Dirac particle may be interpreted as a bound state of two particles interacting via harmonic oscillator potentials,[11] with the four-spinor being formed from two Hermitian harmonic oscillator coordinates and their conjugate momenta (also Hermitian). In contrast, the four-spinor Ξ of (59) is constructed from the two complex coordinates ξ and the complex conjugates of the associated generalized momenta π. The Poisson brackets given by (61) do not correspond to the commutator brackets of the Dirac model oscillator variables. Further, a simple prescription [see (72) below] exists for incorporating electromagnetic interactions into the present theory, whereas the Dirac model allows no such ansatz.

Finally it should be noted that Ξ is *not* a twistor in the sense of Penrose.[12] In the first place the present formalism applies to particles of nonzero rest mass, unlike twistors associated with massless particles. In the second place Ξ does not obey the twistor transformation law for translations. Indeed a translation of the origin of Minkowski coordinates leaves Ξ unchanged, while (65) shows that a translation of the observer's trajectory (1) changes Ξ by a *nonlinear* function of Ξ and $\bar{\Xi}$. In contrast twistors belong to a *linear* representation of the Poincaré group.[13]

V. ELECTROMAGNETIC INTERACTIONS

We consider here a Hamiltonian formulation in light cone coordinates for the classical motion of a particle of charge e and mass m in an external electromagnetic field derived from the vector potential $A^\lambda(x^\kappa)$. Inserting the light cone coordinates $\tau, y^\lambda(\tau)$ into the variation principle (18) yields a Lagrangian of the form (22) but with an additional term $-eA_\lambda(v^\lambda + u^\lambda)$. This adds a term $(e/c)(\mathbf{A} + A^0\hat{\mathbf{y}})$ to the expression (23) for $\boldsymbol{\pi}$. Thereafter the algebra follows that of Sec. IV A with

$$\boldsymbol{\Pi} = \boldsymbol{\pi} - (e/c)(\mathbf{A} + A^0\hat{\mathbf{y}}) \tag{67}$$

replacing $\boldsymbol{\pi}$. One finds the same form $K = cv_\lambda p^\lambda$ for the Hamiltonian as before, but with p^λ modified according to

$$\begin{aligned} p^0 - (e/c)A^0 &= \tfrac{1}{2}(\mathbf{y}\cdot\boldsymbol{\Pi})^{-1}(\boldsymbol{\Pi}^2 + m^2c^2)y\,,\\ \mathbf{p} - (e/c)\mathbf{A} &= \boldsymbol{\Pi} - \tfrac{1}{2}(\mathbf{y}\cdot\boldsymbol{\Pi})^{-1}(\boldsymbol{\Pi}^2 + m^2c^2)\mathbf{y}\,. \end{aligned} \tag{68}$$

Hence to the usual prescription $p^\lambda \to p^\lambda - (e/c)A^\lambda$ we add the ansatz $\boldsymbol{\pi} \to \boldsymbol{\Pi}$, with $\boldsymbol{\Pi}$ given by (67). Note that here A^λ means the function of $\tau,\mathbf{y}$ obtained by evaluating $A^\lambda(x^\kappa)$ at $x^\kappa = z^\kappa(\tau) + y^\kappa$. The identity (20) is satisfied by the expressions (68).

As in the noninteracting case, we may base an alternative Hamiltonian formalism on the $D^{((1/2)0)}$ spinor coordinate ξ, which is related to y^λ by (41). Following an argument parallel to that of Sec. IV C we obtain the Hamiltonian $K = cv_\lambda p^\lambda$ with

$$p^\lambda - (e/c)A^\lambda = (m^2c^2\xi^\dagger\sigma^\lambda\xi + \Pi\bar{\sigma}^\lambda\Pi^\dagger)/(\Pi\xi + \xi^\dagger\Pi^\dagger), \tag{69}$$

where

$$\Pi = \pi - (e/c)A_\lambda\xi^\dagger\sigma^\lambda\,. \tag{70}$$

As before $\pi \equiv (\pi_1,\pi_2)$ is the row vector of the conjugate variables to ξ^1,ξ^2, and satisfies the constraint that $\pi\xi$ be real. Hence we have the further rule that electromagnetic coupling requires the replacement $\pi \to \Pi$ with the definition (70).

Finally, in terms of the Dirac spinor Ξ defined by (59), (69) becomes

$$p^\lambda - (e/c)A^\lambda = mc\bar{\Lambda}\gamma^\lambda\Lambda/(\bar{\Lambda}\Lambda), \tag{71}$$

$$\Lambda = [I - e(2mc)^{-1}A_\lambda\gamma^\lambda(1+\gamma^5)]\Xi\,. \tag{72}$$

VI. DISCUSSION

It is interesting to ask what effect it would have on our formalism if a different choice were made for the space-time path (1) of the observer. The answer is that the conjugate pairs $(\mathbf{y},\boldsymbol{\pi})$ and (ξ,π) depend only on the observer's current here–now z^λ and on the trajectory of the particle. How the observer moved prior to arriving at z^λ is of no consequence. This is obviously true for $\mathbf{y}$, because this variable is defined by the point of intersection of the particle trajectory with the past light cone with vertex at z^λ. That the conjugate variable $\boldsymbol{\pi}$ also does not depend on the observer's past history can be seen by expressing $\boldsymbol{\pi}$ in terms of w^λ, the four-velocity of the particle at this point of intersection. We have the relations

$$w^\lambda = (v^\lambda + u^\lambda)/[(v_\alpha + u_\alpha)(v^\alpha + u^\alpha)]^{1/2},$$
$$\boldsymbol{\pi} = mc(\mathbf{w} + w^0\hat{\mathbf{y}}) + (e/c)(\mathbf{A} + A^0\hat{\mathbf{y}}),$$

the fields A^λ being evaluated at the point of intersection.

As a consequence of this independence on the past history of the observer, we can define functions $y^\lambda(z^\kappa),\boldsymbol{\pi}(z^\kappa)$ whose dependence on the four independent variables z^κ is determined by

$$\frac{\partial y^\lambda}{\partial z^\kappa} = \{y^\lambda,p_\kappa\}, \quad \frac{\partial \boldsymbol{\pi}}{\partial z^\kappa} = \{\boldsymbol{\pi},p_\kappa\}. \tag{73}$$

In a similar manner we find that the spinor variables ξ,π,Ξ do not depend on the route by which the observer reached z^κ, so that we may write equations analogous to (73) for these variables also.

At first sight it might appear that the one-particle system considered here is a counterexample to the assertion of Sec. I that knowledge of past light cone initial data is insufficient to enable prediction of future behavior. The point is that the observer cannot be sure that the system will continue to involve only one particle. There could well be, for example, incoming photons, for which there is no indication in the observer's initial past light cone data. This point was made by Dirac.[1]

[1]P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).

[2]*Alphabet conventions*: Early capital Latin letters $A,B,... = 1,2$. Late capital Latin letters $P,Q,... = 1,2,3,4$. Latin lowercase letters $= 1,2,3$. Greek lowercase letters $= 0,1,2,3$. The summation convention applies to repeated indices. *Minkowski metric tensor*: $\eta_{\lambda\mu} = \eta^{\lambda\mu} = \mathrm{diag}[1,-1,-1,-1]$. *Kronecker symbol*: $\delta_{jk} = \delta^{jk} = \mathrm{diag}[1,1,1]$. *Conjugation operations*: A superscript *, T, or † denotes, respectively, the complex conjugate, transpose, or Hermitian conjugate. *Pauli matrices*:

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

[3]S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper & Row, New York, 1962), pp. 48, 57, 117.

[4]L. Landau and E. Lifshitz, *The Classical Theory of Fields* (Pergamon, Oxford, 1975), p. 45.

[5]H. Goldstein, *Classical Mechanics* (Addison–Wesley, Cambridge, MA, 1953), Chap. 8.

[6]G. H. Derrick, "Quantum mechanics of a charged scalar boson with respect to an observer's past light cone," submitted to J. Math. Phys.

[7]E. M. Corson, *Introduction to Tensors, Spinors, and Relativistic Wave Equations* (Blackie, London, 1955), Chap. 2.

[8]If $w = u + iv$ is any complex variable, then the symbols $\partial/\partial w$ and ∂/w^* applied to any differentiable function of the real and imaginary parts u and v, mean, respectively, $\frac{1}{2}(\partial/\partial u - i\partial/\partial v)$ and $\frac{1}{2}(\partial/\partial u + i\partial/\partial v)$. No analyticity properties in the sense of complex variable function theory is implied.

[9]C. Itzykson and J-B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980), p. 694.

[10]P. A. M. Dirac, Proc. R. Soc. London Ser. A **322**, 435 (1971); **328**, 1 (1972).

[11]L. C. Biedenharn and H. van Dam, Phys. Rev. D **8**, 1735 (1973); N. Mukunda, H. van Dam, and L. C. Biedenharn, *ibid.* **22**, 1938 (1980).

[12]R. Penrose, in *Quantum Gravity*, edited by C. J. Isham, R. Penrose, and D. W. Sciama (Clarendon, Oxford, 1975), p. 268; R. Penrose and M. A. H. MacCallum, Phys. Rep. C **6**, 243 (1973).

[13]G. H. Derrick, Int. J. Theor. Phys. **23**, 359 (1984).

Maxwell's equations for a transverse field

William B. Gordon
Radar Division, Code 5311, Naval Research Laboratory, Washington, D. C. 20375

(Received 22 January 1986; accepted for publication 3 September 1986)

Hodge–deRham theory is applied to Maxwell's equations for a transverse electromagnetic wave with a given wave-front surface S. It is shown that the E and B fields, considered as tangent fields to S, are harmonic in the sense of Hodge theory. If S is a spheroid it is known that the space of harmonic fields on S has dimension zero, and hence transverse fields with spheroidal wave fronts do not exist. The same result holds, but for a different reason, if S is a noncircular cylinder or a surface of revolution, and it is conjectured that smooth, singularity-free, transverse solutions to Maxwell's equations exist only if S is a plane or a circular cylinder.

I. INTRODUCTION

A. Statement of the main results

We shall consider an electromagnetic wave propagating in free space along lines normal to a family of wave-front surfaces, each of which is given by $z=$ const, where z is a function that satisfies the eikonal equation $|\nabla z|^2 = 1$. Then for a general point P, $z(P)$ is equal to the signed perpendicular distance from P to the wave front S_0 given by $z=0$ (Ref. 1, p. 30). Let $u=(u^1,u^2)$ be a system of surface coordinates on S_0, and let $\boldsymbol{\eta}=\boldsymbol{\eta}(u)$ be the unit normal vector to S_0, which will be taken to point along the direction of propagation. If $\mathbf{X}$ denotes the position vector of a general point on S_0, then $\mathbf{X}=\mathbf{X}(u)$ for some vector-valued function $\mathbf{X}(u)$ with values in E^3 (= Euclidean three-space). The position vector $\mathbf{R}=\mathbf{R}(u,z)$ of a general point P in E^3 can then be written in the form

$$\mathbf{R}=\mathbf{X}(u)+z\boldsymbol{\eta}(u). \tag{1}$$

The correspondence that assigns to each point P the triple $(u,z)=(u^1,u^2,z)$ will be referred to as a "surface-normal" coordinate system, and the wave front obtained by setting $z=$ const in (1) will be denoted by S_z. Thus (1) has two uses, the first being a description of a surface-normal coordinate system for E^3, and the second, obtained by setting $z=$ const, being a parametric representation for the wave-front surface S_z.

The $\mathbf{E}$ and $\mathbf{B}$ fields will always be required to satisfy Maxwell's equations, viz.,

$$\begin{aligned} \nabla\times\mathbf{E} &= -\dot{\mathbf{B}}, & \nabla\cdot\mathbf{E}&=0,\\ \nabla\times\mathbf{B} &= (1/c^2)\dot{\mathbf{E}}, & \nabla\cdot\mathbf{B}&=0. \end{aligned} \tag{M}$$

In addition, it will be assumed that the fields are transverse, and it turns out that a necessary (but not sufficient) condition for (M) is that, considered as tangent fields to S_z, the fields $\mathbf{E}$ and $\mathbf{B}$ are harmonic in the sense of Hodge theory. This implies that

$$\tilde{\nabla}^2\mathbf{E}=0,\quad \tilde{\nabla}^2\mathbf{B}=0, \tag{2}$$

where $\tilde{\nabla}^2$ is the Hodge operator, a certain second-order differential operator on the tangent vector fields to S_z which is the analog to (but different from) the ordinary Laplace operator ∇^2 in three-space. These results are formalized in the following two propositions.

Proposition A: Suppose there exists an open interval of z values such that for each value of z in this interval both $\mathbf{E}(u,z,t)$ and $\mathbf{B}(u,z,t)$ are tangent to S_z over some interval of t values. Then (M) is satisfied only if $\mathbf{E}$ and $\mathbf{B}$ are harmonic on each of the surfaces S_z.

Proposition B: For monochromatic fields the tangency condition can be relaxed to require that it hold over some interval of z values, but only for a single value of t.

Suppose now that the wave-front surfaces are "closed" in the technical sense of modern differential geometry, i.e., they are compact (hence finitely extended) surfaces without boundary curves. Recall that the topological type of a closed surface is determined by its genus g (= number of holes); i.e., two surfaces with the same genus are "topologically equivalent" in the sense that one can be smoothly deformed into the other. A "spheroid" is any surface topologically equivalent to a sphere, and for spheroids we have $g=0$. For tori we have $g=1$, and for "double tori" we have $g=2$, etc. Moreover, for closed surfaces it is known from deRham cohomology theory that the space of harmonic tangent vector fields has dimension $2g$. In particular, this dimension is zero if the wave fronts are spheroids, and hence transverse fields with spheroidal wave fronts cannot satisfy Maxwell's equations (except for the trivial case when $\mathbf{E}$ and $\mathbf{B}$ are identically equal to 0).

The same result holds (but for a different reason) if the wave fronts are noncircular cylinders or surfaces of revolution, and we conjecture that transverse solutions to (M) exist only if the wave fronts are (flat) planes or circular cylinders. However, the "fields" in our discussion are always assumed to be smooth singularity-free functions of time and position, and the "wave fronts" are always assumed to be closed submanifolds of three-space. If either of these conditions is removed, then the solution set to (M) is enlarged and our results must be modified, as will be shown in Secs. I B and IV B.

B. General discussion

1. Comparison with Luneburg's work

As discussed by Luneburg in Ref. 1, a wave front is the furthest-on position of an expanding pulse of electromagnetic energy, and is taken to correspond to a "sudden discontinuity" in the field quantities. At any instant of time such a wave front occupies only a single surface position, and Luneburg shows that at this instant the **E** and **B** fields are tangent to this surface. The wave fronts discussed by Luneburg can be spheroidal, but there is no contradiction between Luneburg's results and ours since in our discussion the field quantities are assumed to be smooth functions of time and position, and the tangency condition is required to hold over an interval of z values. The wave fronts discussed in this paper are more akin to those discussed in geometrical optics (cf. Ref. 2), and perhaps might be described as the "ghosts" of a departed Luneburg front. The space in a neighborhood of an instantaneous wave-front surface position S will remain excited for some time after the Luneburg front passes through S, and if S is spheroidal at least one of the fields **E** or **B** will lose its tangency property immediately after the time of passage. In fact, in the examples of spheroidal waves given by Luneburg in Ref. 1, Sec. 13 one of the fields remains tangent to S while the other does not.

2. Comparison with geometrical optics

Geometrical optics provides a description of how a field attenuates as it recedes from its source. Assuming that the power density is proportional to $|\mathbf{E}|^2$ (or $|\mathbf{B}|^2$), and that energy is conserved along tubes of rays that span two surface patches on S_0 and S_z, one obtains the representation[2]

$$\mathbf{E}(u,z,t) = \frac{\mathbf{E}_0(u)\exp(ikz)}{[1 + 2\widetilde{H}(u)z + \widetilde{K}(u)z^2]^{1/2}}, \tag{3}$$

where $\mathbf{E}_0 = \mathbf{E}_0(u)$ are the values of **E** on S_0, k is the wave number, $\widetilde{H} = \widetilde{H}(u)$ is the mean curvature on S_0, and $\widetilde{K} = \widetilde{K}(u)$ is the total (or Gaussian) curvature on S_0. A similar expression holds for **B**, and we now ask whether these fields can be made to satisfy (M) by an appropriate choice of the boundary data $\mathbf{E}_0$ and $\mathbf{B}_0$. It turns out that the answer is "no." However, it will be shown that when transverse fields with a given wave front exist, (3) is essentially correct in the far field, i.e., for values of z that are large with respect to the radii of curvature and the wavelength.

3. Physical considerations

One does not expect to find strictly transverse fields generated by a physically realizable system of sources located in a finite region of space. Such fields generally contain longitudinal components whose amplitudes vary as $1/r^2$, whereas the transverse components vary as $1/r$ (where r is the range from source to observer). For example, considering this matter at its most "elementary" level, we note that Feynman's formula for the field generated by a moving charge contains two longitudinal components (one being the Coulomb field) (Ref. 3, Vol. I, Sec. 28-2; Vol. II, Sec. 21-1). However, theoretical fields generated by infinitely extended sources, such as Sommerfeld's solution to scattering from an infinite half-plane, can still provide insight into real physical problems. These classical solutions can also be used to estimate high-order scattering effects from finitely extended bodies, and for a recent account of this subject we refer the reader to Knott.[4]

II. HARMONICITY OF TRANSVERSE FIELDS

A. Vector operators in surface normal coordinates

We shall use the formalism of tensor analysis to obtain formulas for the calculation of vector operators in a general coordinate system, and then specialize these results to surface normal coordinates. In the discussion below we follow the notation and conventions given by Eisenhart in Refs. 5 and 6.

Let $u = (u^1,u^2,u^3)$ be a general coordinate system whose metric tensors and Christoffel symbols are denoted by g_{ij}, g^{ij}, and Γ^i_{jk}. Let $\mathbf{R} = \mathbf{R}(u)$ denote the position vector of a general point in three-space, set $\mathbf{R}_i = \partial\mathbf{R}/\partial u^i$, and let $\{\mathbf{R}^1,\mathbf{R}^2,\mathbf{R}^3\}$ be the basis dual to $\{\mathbf{R}_1,\mathbf{R}_2,\mathbf{R}_3\}$, so that

$$\mathbf{R}^i\cdot\mathbf{R}_j = \delta^i_j \quad (= \text{Kronecker delta}).$$

Then employing the tensor summation convention, for a general vector field $\mathbf{E} = \mathbf{E}(u)$ we have

$$\nabla\cdot\mathbf{E} = \mathbf{R}^i\cdot\frac{\partial\mathbf{E}}{\partial u^i}, \tag{4}$$

$$\nabla\times\mathbf{E} = \mathbf{R}^i\times\frac{\partial\mathbf{E}}{\partial u^i}, \tag{5}$$

$$\nabla^2\mathbf{E} = g^{ij}\left[\frac{\partial^2\mathbf{E}}{\partial u^i\,\partial u^j} - \Gamma^k_{ij}\frac{\partial\mathbf{E}}{\partial u^k}\right]. \tag{6}$$

We recall that a surface normal coordinate system is a triple (u^1,u^2,u^3), where (u^1,u^2) are surface coordinates on S_0 and $u^3 = z$. From (1) the pair (u^1,u^2) can also be used as surface coordinates on S_z, and the following conventions will sometimes be used to distinguish between surface quantities defined on S_z and space quantities defined on three-space: The former will be marked with a tilde (~), whereas the latter will be unmarked; Latin indices will have the range {1,2,3}, whereas Greek indices will have the range {1,2}. Thus the metric tensors, Christoffel symbols, and the coefficients of the second fundamental form of S_z will be denoted by $\tilde{g}_{\alpha\beta}$, $\tilde{g}^{\alpha\beta}$, $\widetilde{\Gamma}^\alpha_{\beta\gamma}$, and $\tilde{\omega}_{\alpha\beta}$. We also define a tensor $\widetilde{Q}^\alpha_\beta$ of mixed type according to

$$\widetilde{Q}^\alpha_\beta = -\tilde{g}^{\alpha\gamma}\tilde{\omega}_{\gamma\beta}.$$

The matrix $(\widetilde{Q}^\alpha_\beta)$ will also be called the "second fundamental form," and by definition the principal directions at a point on a surface are the unit eigenvectors of $\widetilde{Q}$, and the principal curvatures γ_1, γ_2 are the corresponding eigenvalues. The mean and total curvatures on S_z are defined according to[5]

$$2\widetilde{H} = \mathrm{Tr}(\widetilde{Q}) = \gamma_1 + \gamma_2,$$

$$\widetilde{K} = \mathrm{Det}(\widetilde{Q}) = \gamma_1\gamma_2.$$

As before we set

$$\mathbf{R}_\alpha = \frac{\partial\mathbf{R}}{\partial u^\alpha}, \quad \mathbf{R}_3 = \frac{\partial\mathbf{R}}{\partial z} = \eta.$$

The vectors $\mathbf{R}_1$, $\mathbf{R}_2$ are tangent to S_z, and $\mathbf{R}_3 = \boldsymbol{\eta}(u)$ is normal to S_z. It follows that in the dual basis the vectors $\mathbf{R}^1$,$\mathbf{R}^2$ are also tangent to S_z, and that $\mathbf{R}^3 = \boldsymbol{\eta}$. We recall that a vector field **E** has a covariant description (E_i) and a contravariant description (E^i) defined by

$$\mathbf{E} = E_i\mathbf{R}^i = E^i\mathbf{R}_i.$$

The tangential component of a vector field $\mathbf{E}$ will be denoted by $\widetilde{\mathbf{E}}$; i.e.,

$$\widetilde{\mathbf{E}} = E^{\alpha}\mathbf{R}_{\alpha} = E_{\alpha}\mathbf{R}^{\alpha}.$$

The space quantities g_{ij} and Γ^{i}_{jk} are defined according to

$$g_{ij} = \mathbf{R}_i \cdot \mathbf{R}_j, \quad \frac{\partial^2 \mathbf{R}}{\partial u^i \, \partial u^j} = \Gamma^k_{ij}\, \mathbf{R}_k,$$

whereas the corresponding surface quantities on S_z are defined by

$$\tilde{g}_{\alpha\beta} = \mathbf{R}_{\alpha} \cdot \mathbf{R}_{\beta}, \quad \frac{\partial^2 \mathbf{R}}{\partial u^{\alpha} \, \partial u^{\beta}} = \widetilde{\Gamma}^{\gamma}_{\alpha\beta}\, \mathbf{R}_{\gamma} + \widetilde{w}_{\alpha\beta}\boldsymbol{\eta}. \tag{7}$$

The space quantities and surface quantities can be related by comparing these two sets of results, and the surface quantities on S_z can be related to those on S_0 by differentiating (1) with respect to the surface coordinates u^{α} and using the well-known relation

$$\frac{\partial \boldsymbol{\eta}}{\partial u^{\alpha}} = \widetilde{Q}^{\gamma}_{\alpha}\, \mathbf{R}_{\gamma}.$$

Using all these relations, by straightforward calculations one can show that in surface normal coordinates (4)–(6) become

$$\nabla \cdot \mathbf{E} = \widetilde{E}^{\alpha}_{,\alpha} + (2\widetilde{H})E^3 + \frac{\partial E^3}{\partial z}, \tag{8}$$

$$\nabla \times \mathbf{E} = \frac{\widetilde{E}_{2,1} - \widetilde{E}_{1,2}}{\sqrt{[\det(\tilde{g}_{\alpha\beta})]}}\,\boldsymbol{\eta} + \left[\widetilde{\nabla} E^3 - \frac{\partial \widetilde{\mathbf{E}}}{\partial z} - Q\widetilde{\mathbf{E}}\right] \times \boldsymbol{\eta}, \tag{9}$$

$$\nabla^2 \mathbf{E} = \tilde{g}^{\alpha\beta}\left[\frac{\partial^2 \mathbf{E}}{\partial u^{\alpha}\, \partial u^{\beta}} - \Gamma^{\gamma}_{\alpha\beta}\, \frac{\partial \mathbf{E}}{\partial u^{\gamma}}\right] + 2\widetilde{H}\,\frac{\partial \mathbf{E}}{\partial z} + \frac{\partial^2 \mathbf{E}}{\partial z^2}, \tag{10}$$

where $\widetilde{E}^{\alpha}_{,\beta}$ and $\widetilde{E}_{\alpha,\beta}$ denote the surface covariant derivatives of the tangential component of E, $\widetilde{\nabla} E^3$ is the surface gradient of E^3 ($= E^3_{,\alpha}\, \mathbf{R}^{\alpha}$), and Q is the operator that sends tangent fields E into tangent fields F according to $F^{\alpha} = Q^{\alpha}_{\beta}\, E^{\beta}$.

B. Maxwell's equations for a transverse field

We now use (8) and (9) to express (M) in surface-normal coordinates. It will be assumed that the E and B fields are transverse; i.e., $E^3 = B^3 = 0$ over some intervals of z and t values, so that $\mathbf{E} = \widetilde{\mathbf{E}}$ and $\mathbf{B} = \widetilde{\mathbf{B}}$. Then from (8) the divergence equations in (M) yield the equations

$$\widetilde{E}^{\alpha}_{,\alpha} = \widetilde{B}^{\alpha}_{,\alpha} = 0.$$

In other words, the surface divergences of $\mathbf{E}$ and $\mathbf{B}$ vanish.

We now consider the curl equations in (M). Setting $E^3 = 0$ in (9), we get

$$-\dot{\mathbf{B}} = \frac{\widetilde{E}_{2,1} - \widetilde{E}_{1,2}}{\sqrt{[\det(\tilde{g}_{\alpha\beta})]}}\,\boldsymbol{\eta} - \left[\frac{\partial \widetilde{\mathbf{E}}}{\partial z} + Q\widetilde{\mathbf{E}}\right] \times \boldsymbol{\eta}.$$

By assumption the left-hand side of this equation is tangential, but the first term on the right-hand side is normal and the second term is tangential (since it is a cross product with η). Therefore the first term vanishes, and $-\dot{\mathbf{B}}$ is equal to the second term. Similar remarks apply to the other curl equation in (M), and putting everything together we see that Maxwell's equations for a transverse field reduce to

$$\begin{aligned} &\widetilde{E}^{\alpha}_{,\alpha} = \widetilde{B}^{\alpha}_{,\alpha} = 0, \\ &\widetilde{E}_{2,1} - \widetilde{E}_{1,2} = \widetilde{B}_{2,1} - \widetilde{B}_{1,2} = 0, \\ &\dot{\mathbf{B}} = \left[\frac{\partial \mathbf{E}}{\partial z} + Q\mathbf{E}\right] \times \boldsymbol{\eta}, \\ &-\left(\frac{1}{c^2}\right)\dot{\mathbf{E}} = \left[\frac{\partial \mathbf{B}}{\partial z} + Q\mathbf{B}\right] \times \boldsymbol{\eta}. \end{aligned} \tag{MT}$$

The tildes over $\mathbf{E}$ and $\mathbf{B}$ have been deleted in these last two equations since these fields are assumed to be tangential, but the tildes have been retained in the first two equations to emphasize that the covariant derivatives appearing there are surface quantities.

C. Hodge–deRham theory

The elements of Hodge and deRham theories are discussed in many texts in modern differential geometry, of which we shall only refer to Ref. 7 for definitions and notation. We recall that a tangent vector field $\widetilde{\mathbf{E}}$ on a surface S is said to be *closed* if it satisfies

$$\widetilde{E}_{1,2} - \widetilde{E}_{2,1} = 0,$$

and *coclosed* if its surface divergence vanishes; i.e.,

$$\widetilde{E}^{\alpha}_{,\alpha} = 0.$$

A tangent field E is *harmonic* if and only if it is both closed and coclosed. Every harmonic field satisfies (2), and the converse is true on closed surfaces but untrue on open surfaces. From the first two equations in (MT) it follows that transverse solutions to (M) are harmonic.

The structural properties of harmonic fields will be of importance in our subsequent analysis, and to discuss these we shall follow the customary practice of identifying the covariant description ($\widetilde{E}_{\alpha}$) of a tangent field $\mathbf{E}$ with the differential form E^{+} defined by

$$E^{+} = \widetilde{E}_{\alpha}\, du^{\alpha}.$$

The differential form E^{+} is said to be *exact* if it is the (exact) differential df of some function f on S; i.e., if $E_{\alpha} = \partial f / \partial u^{\alpha}$. Every exact form is closed, but the converse is not true unless S is simply connected. More specifically, a closed form E^{+} is exact if and only if its line integral around every closed circuit is zero, and using Green's theorem it is easy to show that this is necessarily the case if S is simply connected.

Every harmonic field E has the decomposition

$$E^{+} = df + \sum A_n \Omega_n,$$

where f is a harmonic function on S, the A_n are constants, and the Ω_n are certain closed but nonexact differential forms independent of $\mathbf{E}$ (i.e., are the same for all harmonic fields $\mathbf{E}$ on S), each having the form $d\theta$, where θ is a multivalued angular coordinate on S. We shall call df the "exact" part of $\mathbf{E}$, and the other terms will be called the "angular" part. If S is a closed surface, then the only harmonic functions on S are constants, and therefore the exact part of $\mathbf{E}$ vanishes. In this case there are $2g$ differential forms Ω_n, where g is the genus of S (cf. Sec. I A). If S is simply connected, then every closed form is exact, and therefore the angular part of $\mathbf{E}$ vanishes so that $\mathbf{E}$ is the gradient of a harmonic function. If S is both closed and simply connected, i.e., if S is a spheroid, then both parts vanish and $\mathbf{E}$ is identically zero.

D. Monochromatic fields

Monochromatic fields $\mathbf{E} = \mathbf{E}(u,z,t)$ have the form

$$\mathbf{E}(u,z,t) = \mathbf{p}(u,z)\cos(\omega t) + \mathbf{q}(u,z)\sin(\omega t).$$

We shall follow the customary practice of identifying $\mathbf{E}$ with its space-dependent part $\mathbf{E}(u,z)$ defined by the complex quantity

$$\mathbf{E}(u,z) = \mathbf{p}(u,z) + i\mathbf{q}(u,z).$$

Then in (M) and (MT) the time derivatives $\dot{\mathbf{B}}$ and $\dot{\mathbf{E}}$ are replaced with $-i\omega\mathbf{B}$ and $-i\omega\mathbf{E}$, respectively, and the curl equation in (M) relating $\dot{\mathbf{B}}$ to $\mathbf{E}$ becomes

$$i\omega\mathbf{B} = \nabla\times\mathbf{E}.$$

Hence if the fields are tangential at only a single instant of time, the left-hand side of this equation is tangential, and the same argument used before goes through to show that $\mathbf{E}$ and $\mathbf{B}$ are harmonic.

III. THE WAVE EQUATION FOR TRANSVERSE FIELDS

A. The equations (WT)

We have seen that the nonexistence of transverse solutions to (M) with spheroidal wave fronts is a simple consequence of the fact that a spheroid cannot support a nonzero harmonic field. However, there exist surfaces S which do support harmonic fields, but which still cannot be the wave fronts for transverse solutions. Results of this type will be obtained by analyzing a wave equation for transverse fields. This equation, denoted (WT), involves space derivatives in only the single space variable z, and (WT) will also be used to analyze the z variation of transverse solutions to (M) (when they exist).

We recall that a field $\mathbf{E} = \mathbf{E}(u^1,u^2,z,t)$ is said to be transverse if it is tangent to S_z over some intervals of z and t values (which by convention will be assumed to contain the value $z=0$), and that a transverse field is harmonic if it satisfies the first two sets of equations in (MT).

In standard field theory the solutions to (M) are shown to satisfy the wave equations

$$(1/c^2)\ddot{\mathbf{E}} = \nabla^2\mathbf{E}, \quad (1/c^2)\ddot{\mathbf{B}} = \nabla^2\mathbf{B}. \tag{W}$$

Conversely, every solution to (W) is a solution to (M), *provided* that one of the fields, say E, satisfies the divergence equation $\operatorname{div}(E) = 0$. Equation (W) can be derived by taking the time derivatives of the curl equations in (M) and making the appropriate substitutions. If one applies the same process to the last two equations in (MT), they one obtains a wave equation for transverse fields, viz.,

$$(1/c^2)\ddot{\mathbf{E}} = [2\widetilde{K}I - 2\widetilde{H}Q]\mathbf{E} + 2\widetilde{H}\frac{\partial\mathbf{E}}{\partial z} + \frac{\partial^2\mathbf{E}}{\partial z^2}, \tag{WT}$$

with a similar equation for $\mathbf{B}$. Here I denotes the identity operator, and Q is the second fundamental form defined in Sec. II A. This equation will be derived in the next section. The corresponding equation for the space-dependent parts of monochromatic fields is

$$-k^2\mathbf{E} = [2\widetilde{K}I - 2\widetilde{H}Q]\mathbf{E} + 2\widetilde{H}\frac{\partial\mathbf{E}}{\partial z} + \frac{\partial^2\mathbf{E}}{\partial z^2}. \tag{WTM}$$

Every transverse solution to (M) or (MT) is a solution to (WT), but the converse is not necessarily true. A solution to (WT) is a solution to (MT) if and only if it is harmonic, and we shall see that solutions to (WT) need not propagate as harmonic fields, even if harmonic initial data are prescribed on S_0.

B. Derivation of (WT)

Recall that the principal directions $\boldsymbol{\xi}_1,\boldsymbol{\xi}_2$ are unit vectors satisfying

$$Q\boldsymbol{\xi}_\alpha = \gamma_\alpha\boldsymbol{\xi}_\alpha \quad (\alpha = 1,2). \tag{11}$$

The principal directions are orthogonal, and we can therefore assume that $\boldsymbol{\xi}_1\times\boldsymbol{\xi}_2 = \boldsymbol{\eta}$, so that

$$\boldsymbol{\xi}_1\times\boldsymbol{\eta} = -\boldsymbol{\xi}_2, \quad \boldsymbol{\xi}_2\times\boldsymbol{\eta} = \boldsymbol{\xi}_1. \tag{12}$$

In our notation we shall ignore the dependence of the γ_α on the surface coordinates u, and we shall write $\gamma_1(0),\gamma_2(0)$ for the functional values of these quantities on S_0, and $\gamma_1(z),\gamma_2(z)$ for their corresponding values on S_z. Let ρ_1 and ρ_2 be the principal radii of curvature (the reciprocals of the principal curvatures). Then from simple geometrical arguments or otherwise, we have[2]

$$\rho(z) = \rho(0) + z, \tag{13}$$

and therefore

$$\gamma_\alpha(z) = \gamma_\alpha(0)/[1 + z\gamma_2(0)]. \tag{14}$$

In our derivations we shall resolve tangent fields E in two different ways:

$$\mathbf{E} = E_1\boldsymbol{\xi}_1 + E_2\boldsymbol{\xi}_2, \tag{15}$$

$$\mathbf{E} = \widetilde{E}_\alpha\mathbf{R}^\alpha = \widetilde{E}^\alpha\mathbf{R}_\alpha, \tag{16}$$

and we note that quantities $(\widetilde{E}_\alpha)$ and $(\widetilde{E}^\alpha)$ transform like tensors, whereas the components (E_1,E_2) appearing in (15) do not.

Our derivation of (WT) from (MT) will use the representation (15). Using (11), (12), and (15) to express the last two equations in (MT) in terms of components, we get the system

$$\begin{aligned} \dot{B}_1 &= (\gamma_2 E_2 + E_2'), \\ \dot{B}_2 &= -(\gamma_1 E_1 + E_1'), \\ (1/c^2)\dot{E}_1 &= -(\gamma_2 B_2 + B_2'), \\ (1/c^2)\dot{E}_2 &= (\gamma_1 B_1 + B_1'), \end{aligned} \tag{17}$$

where overdots are used to denote time derivatives and primes denote derivatives with respect to z.

Next we differentiate the last two equations in (17) with respect to time, and substitute the first two equations into the right-hand sides. In these manipulations we have to calculate $\gamma'(z)$, and from (14) we have

$$\gamma'(z) = -[\gamma(z)]^2. \tag{18}$$

Finally, putting everything together we get

$$(1/c^2)\ddot{E}_1 = (\gamma_1\gamma_2 - \gamma_1^2)E_1 + (\gamma_1 + \gamma_2)E_1' + E_1'',$$
$$(1/c^2)\ddot{E}_2 = (\gamma_1\gamma_2 - \gamma_2^2)E_2 + (\gamma_1 + \gamma_2)E_2' + E_2'', \tag{19}$$

with similar equations for B_1 and B_2. These equations are merely the equations (WT) written out in components, and hence the derivation of (WT) is complete. The corresponding equations for monochromatic fields are

$$-k^2E_1 = (\gamma_1\gamma_2 - \gamma_1^2)E_1 + (\gamma_1 + \gamma_2)E_1' + E_1'',$$
$$-k^2E_2 = (\gamma_1\gamma_2 - \gamma_2^2)E_2 + (\gamma_1 + \gamma_2)E_2' + E_2''. \tag{20}$$

C. Analysis of (WT)

We shall now hold the u variables fixed and use (20) to analyze how the monochromatic solutions to (WT) vary with z. There are three cases to consider, depending on whether both, one, or none of the principal curvatures vanish.

When both principal curvatures vanish, (20) reduces to the scalar wave equation in the variable z, and hence E propagates as a pure sinusoid in kz with no attenuation with increasing z.

Next we consider the case when $\gamma_1 = 0$ and $\gamma_2 \neq 0$. In this case (20) reduces to the system

$$E_1'' + (1/\rho_2)E_1' + k^2E_1 = 0,$$
$$\rho_2^2 E_2'' + \rho_2 E_2' + (k^2\rho_2^2 - 1)E_2 = 0.$$

Expressing E_1 and E_2 as functions of ρ_2, the general solutions are given by

$$E_1 = b_1J_0(k\rho_2) + c_1Y_0(k\rho_2),$$
$$E_2 = b_2J_1(k\rho_2) + c_2Y_1(k\rho_2), \tag{21}$$

where the b's and c's are independent of z and the J's and Y's are Bessel functions of the indicated order and type. From the asymptotic properties of the Bessel functions it follows that the components of $\mathbf{E}$ vary as $\exp(\pm ikz)/\sqrt{z}$ for large values of z.

Finally, although we suspect that transverse solutions to (M) do not exist when neither of the principal curvatures is identically zero, a discussion of this case is included for theoretical completeness. Referring to (14), we wish to avoid the singularities that occur when any of the curvatures becomes infinite for positive z. Therefore, setting

$$a_1 = \rho_1(0), \quad a_2 = \rho_2(0),$$

we assume

$$0 < a_1 \leqslant a_2.$$

We set

$$F_1 = \rho_1E_1, \quad F_2 = \rho_2E_2, \tag{22}$$

and substituting into (20) we eventually get

$$F_1'' - (2a/\rho_1\rho_2)F_1' + k^2F_1 = 0, \tag{23a}$$
$$F_2'' + (2a/\rho_1\rho_2)F_2' + k^2F_2 = 0, \tag{23b}$$

where

$$a = (a_2 - a_1)/2.$$

Let F be a solution to (23a), and let $\mathbf{Y}$ be the two-vector

$$\mathbf{Y} = \begin{bmatrix} kF \\ F' \end{bmatrix}.$$

Then (23a) can be written in the form

$$\frac{d\mathbf{Y}}{dz} = kJ\mathbf{Y} + \mathbf{V}, \tag{24}$$

where the matrix J and the vector $\mathbf{V}$ are defined by

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} 0 \\ 2aF'/\rho_1\rho_2 \end{bmatrix}.$$

Then using the method of variation of constants,[8] the general solution to (24) can shown to satisfy

$$|\mathbf{Y}(z) - e^{k(z-z_0)J}\mathbf{Y}(z_0)| \leqslant [2a|\mathbf{Y}(z_0)|/\rho_1(z_0)].$$

This result shows that $\mathbf{Y}(z)$ is sinusoidal in the far-field limit, since expanding $\exp(kzJ)$ as an infinite series using the relation $J^2 = -I$, we get

$$e^{kzJ} = \cos(kz)I + \sin(kz)J.$$

The same results hold when F is a solution to (23b), and from (22) it follows that in the far field the components of $\mathbf{E}$ have the form $\exp(\pm ikz)/z$.

IV. SOLUTIONS WITH PLANAR OR CYLINDRICAL WAVE FRONTS

A. Planar wave fronts

The discussion in the last section was concerned with the z variation of solutions for fixed values of the u coordinates, and we shall now allow u to vary and examine the global behavior of fields with planar or cylindrical wave fronts. Special solutions $\mathbf{E}$ to (W) in these cases are usually obtained by the method of separation of variables, and solutions to (M) are then obtained by imposing the additional condition, $\mathrm{div}(\mathbf{E}) = 0$. In our discussion equation (WT) will be combined with the structure theory for harmonic fields (discussed in Sec. II C) to establish rigorously that these are the only transverse solutions to (M) in these cases. We begin with the planar case.

Let (x^1,x^2,x^3) be rectangular coordinates, let $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$ be the corresponding unit basis vectors, and let S_0 be the plane $x^3 = 0$. Let $\mathbf{E} = \mathbf{E}(x^1,x^2,x^3)$ be a general solution to (M) whose boundary data on S_0 is given by $\mathbf{E}_0 = \mathbf{E}_0(x^1,x^2)$. From the discussion in Sec. II C we know that $\mathbf{E}_0$ must be the gradient of a harmonic function on S_0 (since the plane is simply connected), and from the last section we know that $\mathbf{E}$ must be sinusoidal in kz. In our present notation $z = x^3$, and it follows that the most general transverse solution to (M) must be of the form

$$\mathbf{E}(x^1,x^2,x^3) = e^{ikx^3}\,\mathrm{grad}(U) + e^{-ikx^3}\,\mathrm{grad}(V), \tag{25}$$

where $U = U(x^1,x^2)$ and $V = V(x^1,x^2)$ are harmonic functions on S_0, and grad denotes the surface gradient

$$\mathrm{grad}(U) = \frac{\partial U}{\partial x^1}\mathbf{e}_1 + \frac{\partial U}{\partial x^2}\mathbf{e}_2.$$

Conversely, it is easily seen that every vector function $\mathbf{E}$ of the form (25) is a solution to (W) and the divergence equation $\mathrm{div}(\mathbf{E}) = 0$. Hence (25) represents the most general transverse solution to (M) when the wave front is a plane.

B. Solutions with singularities

We illustrate by means of an example how the solution set to (M) must be enlarged, and our conclusions modified, if the fields are allowed to have singularities or if the wave fronts are not required to be closed submanifolds of three-space. Let S_0 now be an angular region in the x^1-x^2 plane that excludes the origin. Then the general structure theory for harmonic fields still holds, but in this case S_0 is not simply connected and every harmonic field $\mathbf{E}_0$ on S_0 has the covariant form $E_0^+ = df + a\, d\theta$, where f is a harmonic function, a is an arbitrary constant, and θ is the angular polar coordinate. Setting $r = [(x^1)^2 + (x^2)^2]^{1/2}$, we have

$$d\theta = (1/r^2)(x^1\, dx^2 - x^2\, dx^1),$$

and the contravariant field corresponding to $d\theta$ is therefore $(1/r^2)(x^1\mathbf{e}_2 - x^2\mathbf{e}_1)$. Hence, referring to (25), the most general solution to (M) in this case is now given by

$$\mathbf{E}(x^1,x^2,x^3) = e^{ikx^3}\,\mathrm{grad}(U) + e^{-ikx^3}\,\mathrm{grad}(V) + (a/r^2)(x^1 e_2 - x^2 e_1).$$

C. Cylindrical wave fronts

Let S_0 be a cylinder, i.e., the surface swept out by a generating line perpendicular to a plane as its intersection with the plane moves around a base curve. In the next section it will be shown that transverse solutions to (M) cannot exist when the curvature γ_0 of the base curve is nonconstant. We shall therefore assume that γ_0 is a nonzero constant. [The plane is a special case of a cylinder for which $\gamma_0 = 0$, and requires the separate treatment (given in Sec. IV A) from the nonzero case.] Then each of the wave fronts S_z is a circular cylinder, and from (14) the principal curvatures are given by $\gamma_1(z) = 0$ and $\gamma_2(z) = \gamma_0/(1 + z\gamma_0)$. The general solution to (M) must therefore have the form (21), where the coefficients must now be chosen to be functions of the surface coordinates that make $\mathbf{E}$ harmonic. It turns out that $\mathbf{E}$ is harmonic if and only if the coefficients are constants.

To prove this result let the base curve for S_0 lie in the x_1-x_2 plane, and let its parametric equations be given by $\{x^1 = x^1(s),\ x^2 = x^2(s)\}$, where the parameter s is arc length along the base curve. Let h denote the height of a general point on S_z above the plane, and take (h,s) to be the surface coordinates on S_z. Then (1) becomes

$$\mathbf{R}(h,s,z) = \mathbf{X}(h,s) + z\boldsymbol{\eta}(s),$$

and since $d\boldsymbol{\eta}/ds = \gamma_0\, \partial\mathbf{X}/\partial s$ along the base curve, we have

$$\mathbf{R}_1 = \mathbf{X}_1 = \frac{\partial\mathbf{X}}{\partial h},$$

$$\mathbf{R}_2 = (1 + z\gamma_0)\mathbf{X}_2 = (1 + z\gamma_0)\frac{\partial\mathbf{X}}{\partial s},$$

and therefore

$$\mathbf{R}^1 = \mathbf{X}_1, \qquad \mathbf{R}^2 = (1 + z\gamma_0)^{-1}\mathbf{X}_2. \tag{26}$$

We recall that the result (21) is based on the decomposition (15) of E into components along the principal directions, which in the present case are the unit vectors $\mathbf{X}_1$ and $\mathbf{X}_2$. Therefore the representations (15) and (16) become

$$\mathbf{E} = E_1\mathbf{X}_1 + E_2\mathbf{X}_2 = \widetilde{E}_1\mathbf{R}^1 + \widetilde{E}_2\mathbf{R}^2.$$

Taking the dot product of the second pair of equations with $\mathbf{R}_1$ and $\mathbf{R}_2$, from (21) we get

$$\begin{aligned} \widetilde{E}_1 &= E_1 = b_1 J_0(k\rho) + c_1 Y_0(k\rho), \\ \widetilde{E}_2 &= (1 + z\gamma_0)E_2 \\ &= (1 + z\gamma_0)[b_2 J_1(k\rho) + c_2 Y_1(k\rho)], \end{aligned} \tag{27}$$

where the b's and c's are functions of h and s, and $\rho = \rho(z)$ is the radius of curvature on S_z given by (13) with $\rho_0 = 1/\gamma_0$.

From the definitions in Sec. II C, $\mathbf{E}$ is closed if and only if

$$\frac{\partial\widetilde{E}_1}{\partial s} = \frac{\partial\widetilde{E}_2}{\partial h}. \tag{28a}$$

From (26) the components of the metric tensor are given by

$$\tilde{g}^{11} = 1, \quad \tilde{g}^{12} = \tilde{g}^{21} = 0, \quad \tilde{g}^{22} = (1 + z\gamma_0)^{-2},$$

and all the Christoffel symbols vanish since these components are constant on S_z. It follows that $\mathbf{E}$ is coclosed if and only if

$$0 = \tilde{g}^{11}\frac{\partial\widetilde{E}_1}{\partial h} + \tilde{g}^{22}\frac{\partial\widetilde{E}_2}{\partial s}. \tag{28b}$$

Applying Eqs. (28) to (27) we get certain linear relations between the Bessel functions whose coefficients are the partial derivatives of the b's and c's with respect to h and s. Hence all these partial derivatives must vanish, and therefore the b's and c's are constants.

V. OTHER CASES FOR WHICH TRANSVERSE FIELDS DO NOT EXIST

A. Alternate forms of (MT) and (WT)

We shall now give examples of surfaces S_0 that support harmonic tangent fields, but for which (WT) has no harmonic solution. From the discussion in Sec. III A we recall that such surfaces cannot be the wave fronts for transverse solutions to (M). The examples are noncircular cylinders and surfaces of revolution, and as previously mentioned, we suspect that singularity-free transverse solutions to (M) are only possible when the wave fronts are planes or circular cylinders. We shall confine our discussion to the monochromatic case.

In Sec. III we used the decomposition (15), and we now wish to express (MT) and (WT) in terms of the components given in the decomposition (16). However, for typographic ease we shall no longer use tildes to mark surface quantities. For example, $g_{\alpha\beta}$ will now denote the surface metric tensor defined by (7).

We begin by writing out a well-known representation for the "star" operator (*), which is the mapping that sends tangent fields $\mathbf{E}$ into $\mathbf{E}\times\boldsymbol{\eta}$. Setting $g = \det(g_{\alpha\beta})$, the result is[7]

$$*\mathbf{E} = (1/\sqrt{g})[(-g_{21}E_1 + g_{11}E_2)\mathbf{R}^1 + (-g_{22}E_1 + g_{12}E_2)\mathbf{R}^2]. \tag{29}$$

For a general tangent field $\mathbf{E}$, $**\mathbf{E} = -\mathbf{E}$ since $*$ is a rotation of 90 degrees. Using (29) one can also show that $*$ sends closed fields into coclosed fields, coclosed fields into closed fields, and hence $*$ sends harmonic fields into harmonic fields.[7]

Next we write out the following intermediate results, which will be proved at the end of this section, and in which we set $M = Q^2$:

$$\frac{\partial \mathbf{R}^\alpha}{\partial z} = -Q^\alpha_\beta \mathbf{R}^\beta, \tag{30a}$$

$$\frac{\partial Q}{\partial z} = -Q^2 = -M, \tag{30b}$$

$$\frac{\partial^2 \mathbf{R}}{\partial z^2} = 2M^\alpha_\beta \mathbf{R}^\beta, \tag{30c}$$

$$0 = Q^2 - 2HQ + KI. \tag{30d}$$

Then, applying $(\partial/\partial z)$ to (16) and using (30a) we get

$$\frac{\partial \mathbf{E}}{\partial z} = \frac{\partial E_\alpha}{\partial z}\mathbf{R}^\alpha + E_\alpha \frac{\partial \mathbf{R}^\alpha}{\partial z} = \frac{\partial E_\alpha}{\partial z}\mathbf{R}^\alpha - E_\alpha Q^\alpha_\beta \mathbf{R}^\beta.$$

Hence

$$\frac{\partial \mathbf{E}}{\partial z} + Q\mathbf{E} = \left(\frac{\partial E_\alpha}{\partial z}\right)\mathbf{R}^\alpha,$$

and using this result, the monochromatic form of the last two equations in (MT) can be as

$$\begin{aligned} -i\omega\mathbf{B} &= *\left[\left(\frac{\partial E_\alpha}{\partial z}\right)\mathbf{R}^\alpha\right], \\ \frac{i\omega}{c^2}\mathbf{E} &= *\left[\left(\frac{\partial B_\alpha}{\partial z}\right)\mathbf{R}^\alpha\right]. \end{aligned} \tag{31}$$

Similarly, using (30) to simplify the results, we find that (WTM) reduces to

$$-k^2 E_\alpha = \frac{\partial^2 E_\alpha}{\partial z^2} + 2H\frac{\partial E_\alpha}{\partial z} - 2Q^\beta_\alpha \frac{\partial E_\beta}{\partial z}. \tag{32}$$

Proofs of (30a)–(30d): Using the well-known relation[5]

$$\frac{\partial \boldsymbol{\eta}}{\partial u^\alpha} = -\omega_{\alpha\beta}\mathbf{R}^\beta,$$

from (1) we get

$$\frac{\partial \mathbf{R}_\alpha}{\partial z} = \frac{\partial^2 \mathbf{R}}{\partial z\,\partial u^\alpha} = \frac{\partial \boldsymbol{\eta}}{\partial u^\alpha} = -\omega_{\alpha\beta}\mathbf{R}^\beta.$$

Then differentiating the relation $\mathbf{R}^\alpha \cdot \mathbf{R}_\beta = \delta^\alpha_\beta$, we get

$$\frac{\partial \mathbf{R}^\alpha}{\partial z}\cdot\mathbf{R}_\beta + \mathbf{R}^\alpha\cdot\frac{\partial \mathbf{R}_\beta}{\partial z} = 0,$$

which combined with the above result gives (30a).

Equation (30b) is a consequence (18) and the fact that the principal directions ξ_1, ξ_2 are independent of z, i.e., do not vary along a ray normal to the surfaces S_z. Equation (30c) is obtained by differentiating (30a) and using (30b). Equation (30d) is the characteristic equation for Q. It is a consequence of the fact that Q is a symmetric operator (with respect to the g metric) with eigenvalues γ_1, γ_2. This relation can be used to reduce any rational expression in Q to an expression which is linear in Q.

B. Noncircular cylinders and surfaces of revolution

We now specialize (31) and (32) to the case when a surface coordinate system u exists for which the tangent basis vectors $\mathbf{R}_1$, $\mathbf{R}_2$ are scalar multiples of the principal directions. The principal directions are orthogonal, and from the relations $g_{\alpha\beta} = \mathbf{R}_\alpha \cdot \mathbf{R}_\beta$ and the definition of Q we get

$$\begin{aligned} &g_{12} = g_{21} = 0, \quad w_{12} = w_{21} = 0, \\ &Q^1_2 = Q^2_1 = 0, \quad Q^1_1 = \gamma_1, \quad Q^2_2 = \gamma_2. \end{aligned} \tag{33}$$

Using (1) it is easy to show that these properties propagate from S_0 to S_z. Applying (29) and (33) to (31), we reduce (MT) to the form

$$\begin{aligned} (i\omega/c^2)E_1 &= (g_{11}/g_{22})^{1/2}B'_2, \\ -(i\omega/c^2)E_2 &= (g_{22}/g_{11})^{1/2}B'_1, \\ -i\omega B_1 &= (g_{11}/g_{22})^{1/2}E'_2, \\ i\omega B_2 &= (g_{22}/g_{11})^{1/2}E'_1, \end{aligned} \tag{34}$$

where the primes again denote derivatives with respect to z. Similarly, (32) reduces to

$$E''_1 + (\gamma_2 - \gamma_1)E'_1 + k^2 E_1 = 0, \tag{35a}$$

$$E''_2 + (\gamma_1 - \gamma_2)E'_2 + k^2 E_2 = 0, \tag{35b}$$

with similar equations for the B's.

The conditions for (33) and therefore (34) and (35) apply to cylinders and surfaces of revolution endowed with the appropriate coordinate systems. For cylinders, we use the (h,s) coordinates discussed in Sec. IV. For a surface of revolution obtained by rotating a curve around the x^3 axis, we let (r,θ) be polar coordinates in the x^1-x^2 plane so that the parametric representation of the surface is given by

$$\{x^1 = r\cos\theta,\ x^2 = r\sin\theta,\ x^3 = -f(r)\},$$

where for simplicity it is assumed that $f = f(r)$ is a monotonically increasing function with $f(0) = 0$. Then by straightforward calculations one can show that

$$\gamma_1 = f''(r)/[F(r)]^{3/2}, \quad \gamma_2 = f'(r)/\left[r\sqrt{F(r)}\right],$$

where we set $F(r) = 1 + [f'(r)]^2$.

We now differentiate (35a) with respect to u^1, (35b) with respect to u^2, and subtract the results. Setting $d = \gamma_2 - \gamma_1$, we get

$$d_2 E'_1 + d_1 E'_2 = 0,$$

where $d_1 = \partial d/\partial u^1$, $d_2 = \partial d/\partial u^2$. But for the special coordinate systems described above, the principal curvatures and hence d depend on only one of the surface coordinates. Call this coordinate u^2. Then $d_1 = 0$, and from the above equation we get

$$d_2 E'_1 = 0.$$

For the cylinder, d is the curvature of the base curve, and is nonconstant for noncircular cylinders. Hence $d_2 \neq 0$ in this case, and it can be shown that the same result holds for surfaces of revolution. We therefore conclude that $E'_1 = 0$, from which it follows that $E_1 = 0$ since (WTM) has no nonzero solutions whose derivatives are identically zero in z. The same arguments apply to the $\mathbf{B}$ vector, and we conclude that $B'_1 = 0$ and $B_1 = 0$. But from (34) the vanishing of E'_1 implies the vanishing of B_2, and the vanishing of B'_1 implies the vanishing of E_2. Hence $E_1 = E_2 = B_1 = B_2 = 0$; in other words, for noncircular cylinders and surfaces of revolution we have shown that $\mathbf{E} = 0$ and $\mathbf{B} = 0$ are the only solutions to (MT).

[1]R. K. Luneburg, *Mathematical Theory of Optics* (Univ. California, Berkeley, 1966).
[2]M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1975), 5th ed.
[3]R. P. Feynman, R. B. Leighton, and M. Sands, *Lectures on Physics* (Addison–Wesley, Reading, MA, 1963).
[4]E. F. Knott, "A progression of high-frequency RCS prediction techniques," Proc. IEEE **73**, 252 (1985).
[5]L. P. Eisenhart, *An Introduction to Differential Geometry* (Princeton U. P., Princeton, 1947).
[6]L. P. Eisenhart, *Riemannian Geometry* (Princeton U.P., Princeton, 1949).
[7]S. I. Goldberg, *Curvature and Homology* (Dover, New York, 1982).
[8]P. Hartman, *Ordinary Differential Equations* (Wiley, New York, 1964).

On hyper-relativistic quantum systems

Awele Maduemezia
Department of Physics, University of Ibadan, Ibadan, Nigeria

(Received 5 March 1986; accepted for publication 6 August 1986)

A *hyper-relativistic system* is defined as one whose equation of motion is form invariant under coordinate transformations induced by a semisimple group whose algebra is contractible to the algebra of the Poincaré group. Such a system lies, categorically, in the domain between the special theory of relativity and the general theory, for whereas the former requires covariance under transformations between inertial systems, the latter imposes covariance with respect to arbitrary continuous transformations. In this paper, a new interpretation of a particular fiber-bundle structure constructed on the timelike homogeneous space $M = \mathrm{SO}(4,2)/\mathrm{SO}(4,1)$ is presented, and Minkowski space-time is realized as a subspace of the standard fiber of the tangent bundle over this hyperquadric. Through the process of group contraction, coupled with the commutation of the momentum vector fields with the principal bundle of linear frames with which the tangent bundle is associated, a hierarchy of "Heisenberg commutation relations," parametrized by the point spectrum of the center of the contracted algebra, is obtained. The classical Newtonian gravitational potential field enters as the fifth coordinate of an extended space-time manifold.

I. INTRODUCTION

In quantum mechanics, one usually represents physical observables as self-adjoint linear operators acting on a suitable inner product space, normally taken as Hilbert space. The quantization of classical mechanics is achieved through the postulation of the Heisenberg commutation relation $2\pi(xp - px) = ih$, between the (self-adjoint) operator representatives x, p, of the position and momentum observables, respectively. In spite of the undoubted successes of this construction, there have been serious questions of internal consistency raised in the literature. For instance, it is desirable to find a map F: $[\ ,\ ] \rightarrow \{\ ,\ \}$ from the Heisenberg commutator $[\ ,\ ]$ to the Poisson bracket $\{\ ,\ \}$ of classical mechanics. It would appear, however, that there is no such map.[1] There are other problems. These problems seem to derive from our conventional interpretation of the Heisenberg commutator as a Hilbert space operator equation. The operators that one would associate with x and p are both unbounded, and have no eigenvalues. It is thus necessary to make special arrangements to fit them into the measurement postulates of quantum mechanics.

Suppose $x = (x_1, x_2, \ldots, x_n)$ is a local coordinate system for a manifold M, and p is a vector field. Then the *derivation* properties of p are enough to yield the relation $2\pi(xp - px) = ih$. This is because a basis for the space of vector fields is the set of derivatives $\{\partial / \partial x_i\}$, x itself being regarded as a function of x.

Thus the Heisenberg postulate amounts to saying that p is a vector field and, in particular, a generator of translations in x. The *position operator* concept, by which we mean an operator whose effect on a wave function $f(x)$ is to multiply it by x, is not required at this stage, and one need not concern oneself with its spectral properties. Neither is it necessary to regard momentum as an operator in Hilbert space.

There has in the recent past been some effort to exploit the properties of homogeneous spaces of Lie groups in the formulation of quantum mechanics. Amongst many others, Doebner and Tolar[2] considered a homogeneous G-space M, where G is a Lie group. They studied the kinematics of quantum systems on M using Mackey's theory of imprimitivity systems.[3–5] They looked at particle kinematics on M as kinematics on a G-orbit equivalent to M in some Euclidean space R_n. Nevertheless, their approach involves the retention of the conventional characterization of all observables as Hilbert space operators over the base space M. The attitude one adopts in the present paper is that *position x* and *free particle momentum p* are properties of space-time, and are to be handled differently from other observables.

The complete geometrization of quantum mechanics proceeds most naturally within the framework of the theory of fiber bundles.[6,7] Within this framework, the problem of describing relativistic particle kinematics on a *flat* pseudo-Euclidean space, and of making this consistent with a quantization scheme based on the *curved* manifold M, does not pose a serious problem. Beginning with the conformal group of space-time $\mathrm{SO}(4,2)$, we systematically show how the structure of the tangent bundle on $\mathrm{SO}(4,2)/\mathrm{SO}(4,1)$ is rich enough to provide the required Heisenberg commutation relations on the one hand, and the relativistic frame-to-frame transformation equations on the other hand.

The general framework for this construction is given in Sec. II for an arbitrary dynamical group G. In Sec. III, we actualize this construction by choosing G to be the conformal group. This choice derives from the fact that the *local properties* of the conformal gorup directly provide the dynamical variables of quantum mechanics. On the other hand, its *global properties* in relation to its de Sitter subgroup $\mathrm{SO}(4,1)$ *ultimately* provide the space-time structure needed to relativize quantum mechanics.

II. THE GENERAL SCHEME

Let G be a noncompact Lie group whose Lie algebra g supplies all the dynamical variables needed for the description of any physical system, namely, angular momentum, energy momentum, world point variable or position four-vector, etc. Let G possess a maximal subgroup H (with Lie algebra h) such that the homogeneous space $M = G/H$ is of *unit rank*, that is, there is only one independent invariant of any pair of points of M with respect to the action of the group G on M, or equivalently,[8] there is only one independent invariant differential operator on M. Such an invariant differential operator is the *Laplace–Beltrami* operator[9]

$$D(M) = g^{-1/2}\frac{\partial}{\partial x_k}\left(m^{ik}g^{1/2}\frac{\partial}{\partial x_i}\right), \tag{2.1}$$

where g is the determinant of the metric tensor with components m_{ij}, and the m^{ij} are the components (elements) of the matrix inverse of m. Thus $m^{ij} = (m^{-1})_{ij}$, and the x_i are coordinate vectors defining the invariant elementary interval

$$ds^2 = m_{ij}\,dx_i\,dx_j\,. \tag{2.2}$$

The choice of a homogeneous space of unit rank is made for simplicity. The dimension of M is required to be small, typically 5, in the spirit of the Kaluza and Klein theories[10] which are currently being reexamined in the literature within the context of elementary particle physics.

We then construct the *principal H-bundle P* on M, and hence the tangent bundle over M, which is, in the terminology of Kobayashi and Nomizu,[7] *the fiber bundle associated with P, with structure group H, and standard fiber E(p,q)*, where $E(p,q)$ is a physically relevant pseudo-Euclidean space with metric $g_{ij} = (1,1,\ldots(p \text{ times}), -1, -1,\ldots(q \text{ times}))$, and $p+q=5$. Then the difference space $Q = g - h$ has five linearly independent *vector fields* as its basis. *It is the elements of the basis of Q that we commute with the momenta*, which are also vector fields belonging to g, in order to obtain the Heisenberg commutation relations in a suitable limit. This limit is provided by the classical Inönü–Wigner–Saletan contraction scheme.[11,12]

The standard fiber $E(p,q)$ of the tangent bundle turns out to consist of Minkowski space-time plus one extra coordinate intimately connected with the distribution of matter in space-time.

As is explained in Sec. III, there is a *duality* between the spaces $E(p,q)$ and Q provided by the quotient map

$$f\colon\; E(p,q)\times Q \to E(p,q)Q \tag{2.3}$$

that defines the *total space* of the tangent bundle. This *duality* replaces the usual quantum mechanical prescription of *replacing* position coordinates with *operators* in order to be able to commute them in a nontrivial way with the momenta. In this approach, we now have *vector fields associated* in a natural way with the position coordinates, *as part of the fiber bundle structure*, and these vector fields are commuted with the momenta, which already are vector fields, being elements of the Lie algebra g of the group G of the bundle.

III. EXPLICIT CONSTRUCTION OF THE GEOMETRIZATION SCHEME

Consider the pseudo-Euclidean space $E(4,2)$ with diagonal covariant metric tensor $m_{ij} = (1,1,1,-1,-1,1)$. Let $G = \mathrm{SO}(4,2)$, $H = \mathrm{SO}(4,1)$, and let $w = (w_1,w_2,w_3,w_4,w_5,w_6)$ be an arbitrary point of $E(4,2)$. Consider the homogeneous space $M = \mathrm{SO}(4,2)/\mathrm{SO}(4,1)$. Let z be an arbitrary point of M. Then M is topologically equivalent to the five-dimensional hyperquadric

$$m_{ij}w_iw_j = -1\,, \tag{3.1}$$

which we also denote by M. Thus M is suitably parametrized by[13,14]

$$\begin{aligned}
w_1 &= \cos\varphi_1 \sin\theta_2 \sinh\theta\,,\\
w_2 &= \sin\varphi_1 \sin\theta_2 \sinh\theta\,,\\
w_3 &= \cos\varphi_2 \cos\theta_2 \sinh\theta\,,\\
w_6 &= \sin\varphi_2 \cos\theta_2 \sinh\theta\,,\\
w_4 &= \cos\varphi_3 \cosh\theta\,,\\
w_5 &= \sin\varphi_3 \cosh\theta\,,
\end{aligned} \tag{3.2}$$

where

$$0\leqslant\theta<\infty\,,\quad 0\leqslant\theta_2\leqslant\pi/2\,,\quad 0\leqslant\varphi_1,\varphi_2,\varphi_3<2\pi\,. \tag{3.3}$$

The group H acts freely on G on the right and the canonical map

$$p\colon\; G\to M = G/H \tag{3.4}$$

is differentiable. The collection $\{G,p,M\}$ is the *principal H-bundle* over M. We denote it by P.

A *linear frame y* at a point z of M is an ordered basis (y_1,y_2,y_3,y_4,y_5) of the tangent space $T_z(M)$ at z. Thus with the parametrization given in (3.2), we could choose

$$(y_1,y_2,y_3,y_4,y_5) = \left(\frac{\partial}{\partial\theta},\frac{\partial}{\partial\theta_2},\frac{\partial}{\partial\varphi_1},\frac{\partial}{\partial\varphi_2},\frac{\partial}{\partial\varphi_3}\right). \tag{3.5}$$

However, anticipating the quantum mechanical applications we have in mind, and noting that the generators of $Q = g - h$ are also a possible linear frame, we take for the y_i, not the entries in (3.5), but rather such suitable $C(M)$-linear combinations of the entries in (3.5) as would in a suitable Lie algebra contraction limit provide a basis whose dual is indentifiable with space-time coordinates plus a fifth coordinate. Here $C(M)$ is the space of C^∞ — functions on M.

The group SO(4,2) has a Lie algebra g generated by the 15 left-invariant vector fields $M_{ba} = -M_{ab}, p_a, A_a, D$, $a,b = 1,2,3,4$, with the commutation relations[13]

$$\begin{aligned}
&[D,p_a] = p_a\,,\\
&[D,A_a] = -A_a\,,\\
&[M_{ab},p_c] = g_{ca}p_b - g_{cb}p_a\,,\\
&[M_{ab},A_c] = g_{ca}A_b - g_{cb}A_a\,,\\
&[M_{ab},M_{cd}] = g_{ac}M_{bd} + g_{da}M_{cb} + g_{cb}M_{da} + g_{db}M_{ac}\,,\\
&[A_b,p_a] = M_{ab} - g_{ab}D\,,\\
&[A_a,A_b] = [p_a,p_b] = [D,M_{ab}] = 0\,,
\end{aligned} \tag{3.6}$$

where g_{ab} is the metric $(1,1,1,-1)$.

Now D is the dilatation field, M_{ab} are angular momenta, p_a are linear momenta, and A_a are the so-called special conformal transformations. Let $r = 2^{1/2}$. We choose as a basis for the tangent space $T_z(M)$,

$$(y_1,y_2,y_3,y_4,y_5) = (rA_2,rA_1,rA_3,rA_4,rD)\,. \tag{3.7}$$

Let $L(M)$ be the set of all linear frames y at all points z of M. Thus a typical point of $L(M)$ is denoted by (z,y) or by y_z.

Also let

$$\pi: L(M) \to M \tag{3.8}$$

denote the projection

$$y_z = (z,y) \to z \tag{3.9}$$

onto the first entry. Any two linear frames y, y' are *de Sitter equivalent* if they are connected to each other by a de Sitter transformation. For two such frames,

$$y'_n = q_{nm} y_m \ , \tag{3.10}$$

where $q \in \mathrm{SO}(4,1)$. It follows that M is the quotient of $L(M)$ by the equivalence relation induced by this action of SO(4,1) on $L(M)$. The collection $\{L(M), \pi, M\}$ is a *principal de Sitter bundle*, which we denote by S.

Now let $E = E(4,1)$ be the pseudo-Euclidean space with metric $(1,1,1,1,-1)$. We construct "*the fiber bundle F, associated with S*," and with standard fiber[7] E. Consider the map

$$f: (L(M) \times E) \to F \tag{3.11}$$

defined by

$$(y_z, x) \to x y_z \ ; \tag{3.12}$$

$y_z = (y_1, y_2, y_3, y_4, y_5)$, written as a *column* matrix, and x is a *row* matrix $(x_1, x_2, x_3, x_4, x_5)$, where x_i is a real number. Thus

$$x y_z = x_i y_i \in T_z(M) \ . \tag{3.13}$$

Each y_z may, therefore, be regarded as a linear *isomorphism* of E onto $T_z(M)$. To get from one point of $T_z(M)$ to another, we keep y_z, that is, the basis, fixed, and vary x, that is, the components (or coordinates).

The group SO(4,1) acts on E covariantly from the right, and on $L(M)$ contravariantly from the left. For each q in SO(4,1), we have

$$q: \ x_i \to x_j q_{ji} \ , \quad q: \ y_i \to (q^{-1})_{ik} y_k \ .$$

This gives the induced action

$$q: \ x_i y_i \to x_j q_{ji} (q^{-1})_{ik} y_k = x_i y_i \tag{3.14}$$

on the quotient space F, *the space of all vector fields on M, with coefficients in E*. In order to bring out the quotient structure, F may be written as

$$F = (L(M) \times E) \bmod H \ . \tag{3.15}$$

Then F, taken together with the projection π onto M [Eqs. (3.8) and (3.9)], is the tangent bundle over M.

We now introduce physics into the system by identifying the standard fiber E with space-time, augmented with an as yet undefined fifth coordinate x_5. In the sequel, E will be referred to as *extended space-time*. The isomorphism

$$y_z: \ E \to T_z(M) \tag{3.16}$$

induced by each $y_z \in L(M)$ enables us set up a one-to-one correspondence between each position coordinate x and a vector $z_x = x_i y_i \in T_z(M)$. If we further identify the subset of generators $p = (p_1, p_2, p_3, p_4)$ of SO(4,2) with energy-momentum, we obtain the result that the dynamically meaningful commutation rules between momentum and position observables are given by the $C(M)$*-linear* commutators of the form $[p, z_x]$, between the vector fields. Due to this $C(M)$ linearity, this reduces to a commutation of the energy-momentum vector fields with the vector fields y, *dual* to extended space-time x with regard to the action of the group H on the *total space* F of the tangent bundle over M.

In order to obtain the Heisenberg commutation rules, we have to perform a suitable Lie algebra contraction that will collapse SO(4,2) into a doubly inhomogeneous Lorentz group. This is studied in the next section.

IV. HEISENBERG COMMUTATION RULES

Let e be a positive fractional parameter, and let $B(e)$: $g \to g'$ be an e-parameter-dependent linear transformation of the Lie algebra of g into its isomorph g', which transformation, however, becomes singular in the limit $e \to 0$. Let $[\ ,\ ]$ and $[\ ,\ ]_e$ denote the Lie products in g and g', respectively.

Explicitly,

$$[U,V]_e = B(e)^{-1}[B(e)U, B(E)V] \ , \quad U,V \in g \ . \tag{4.1}$$

We now use the symbol U to denote either A_a or p_a, $a = 1,2,3,4$, and choose for the *definition* of $B(E)$, for e different from zero,

$$\begin{aligned} &B(e)U = eU = U' \ , \quad B(e)^{-1}U' = e^{-1}U' \ , \\ &B(e)D = e^2 D = D' \ , \quad B(e)^{-1}D = e^{-2}D \ , \\ &B(e)M_{ab} = M_{ab} = B(e)^{-1}M_{ab} \ , \quad a,b = 1,2,3,4 \ . \end{aligned} \tag{4.2}$$

Using this in (3.6), then going to the limit $e \to 0$, and replacing $[\ ,\ ]_0$ with $[\ ,\ ]$, we obtain

$$\begin{aligned} &[y_a, p'_b] = -g_{ab} y_5 = -2^{1/2} g_{ab} D' \ , \\ &[y_a, y_5] = [p'_a, y_5] = [p'_a, p'_b] = 0 \ , \end{aligned} \tag{4.3}$$

$$[y_a, y_b] = [M_{ab}, y_5] = 0 \ . \tag{4.4}$$

All other commutators remain the same as in (3.6). This contracted algebra is that of the (doubly) inhomogeneous group $K \otimes Q$, where $K = \mathrm{SO}(3,1) \otimes T_4$ is the Poincaré group, and Q is the five-dimensional Abelian group generated by the contracted frame $y \in L(M)$. The T_4 are four-dimensional translations.

Equations (4.3) and (4.4) have important quantum mechanical implications. We note first that $y_5 = 2^{1/2} D'$ acts as a constant vector field in the sense that it commutes with all other vector fields. It belongs to, and in fact generates, the *center* of the contracted algebra L. This algebra L induces infinitesimal motions of $M = \mathrm{SO}(4,2)/\mathrm{SO}(4,1)$ into itself. Hence M is an *L-module*, and the natural action of L on M defines a representation R: $L \to \mathrm{gl}(M)$. By *Schur's Lemma*,[15] given such a representation R: $L \to \mathrm{gl}(M)$ of the algebra L, *if R is irreducible, then the only endomorphism of M commuting with all* $R(V)$ $(V \in L)$ *are the scalars*. Thus $y_5 = 2^{1/2} D'$ is a scalar multiple of the identity. We denote it by h_j, where $j = 1,2,3,\ldots,n$, $n \geqslant 2$.

Ordinarily, before contraction, the spectrum of D is continuous.[16] However, the contracted operator $y_5 = 2^{1/2} D'$, belonging as it does to the center of the algebra L, will have a discrete spectrum, which we have written h_j. That is, the process of contraction partitions the continuous spectrum into discrete groups, much as a magnetic field collects the energy levels of an erstwhile free electron into a countable set of *Landau levels* in solids. These eigenvalues h_j define the limits of uncertainty in measurement. For $j = 1$, we expect that $h_1 = 0$, and we would then have classical mechanics.

For $j=2$, we put $h_2 = -2^{-1/2} ih/(2\pi)$, where h is Planck's constant. Then

$$[y_a, p_b] = i g_{ab} h/(2\pi) , \tag{4.5}$$

which includes the Heisenberg commutation rule. For j greater than 2, we indeed have a *hierarchy of quantum mechanics.*

V. INTERPRETATION OF THE FIFTH COORDINATE

The special theory of relativity is based on the assumption that space-time is homogeneous, and has an existence independent of matter or field.[17] Then the Poincaré group, $P = SO(3,1) \otimes T_4$, is the invariance group of the laws of physics. "On the basis of the general theory of relativity, on the other hand, space, as opposed to what fills space, which is dependent on the coordinates, has no separate existence."[17] For *hyper-relativistic systems* such as we are now studying, a middle course has to be charted. First one must admit that uneven distribution of massive matter in the neighborhood of any point alters the properties of space-time at that point. At any given world point, it should not be enough to indicate the values of the four variables x_a that make up space-time. It would seem to be necessary also to indicate the value of the variable that allows for the effect at that point of a *nonuniform* and *nonstatic* distribution of matter in the universe. In general relativity, a pure gravitational field is described in terms of the metric tensor g_{ij} obtained from the solution of the Einstein field equations. For hyper-relativistic systems, it is enough to use the classical gravitational potential field g given by the Newtonian theory, subject to the so-called Einstein equivalence principle.[17] Thus g is locally equivalent to an acceleration. In addition to specifying x_a, we also have to indicate the mean gravitational potential field g at the world point due to ambient mass distributions. It is therefore reasonable to take as the fifth coordinate of the space $E(4,1)$, g, the classical (that is, Newtonian) gravitational potential field at that point due to all matter as they are distributed *at that instant of time.* We multiply this with a *physical constant* "a," so that the object ag has the dimension of length.

The standard fiber $E(4,1)$ is a five-dimensional metric space spanned by three ordinary coordinate variables (x_1, x_2, x_3) by one timelike variable $x_0 = ct$, and by one potential-like variable, $x_4 = ag$. The metric $(-1,1,1,1,1) = g_{ij}, i,j = 0,1,2,3,4$. Let $W = dg/dt$. W may or may not vanish, even though $\partial g/\partial t$ necessarily vanishes.

VI. de SITTER TRANSFORMATIONS OF INERTIAL FRAMES

Let us consider the action of SO(4,1) on E(4,1). It has been shown[18] that any element q of $SO(m,n)$ may, for $3 \leqslant m < \infty$, $0 < n < \infty$, $m \geqslant n$, be factorized in the form $q = A \cdot B(\theta) \cdot C$, where A,C belong to the group $SO(m,n-1) \times 1$, and $B(\theta)$ belongs to a one-parameter subgroup of $SO(m,n)$. With a slight variation of the arguments used in Ref. 18, one can show that an arbitrary element of the group SO(1,4) may be factorized in the form $q = A \cdot B(\theta) \cdot C$, where A,C belong to $SO(1,3) \times 1$, and $B(\theta)$ belongs to a one-parameter subgroup of SO(1,4). Here A and C are Lorentz transformations in the subspace spanned by the coordinates x_0, x_1, x_2, x_3. The result is that one can have a pure de Sitter transformation *without Lorentz boosts,* just as one can have a pure Lorentz transformation *without rotations.*

A more useful factorization is, however, obtained as a deduction from the factorization of the n-dimensional rotation group $SO(n)$ given by Murnaghan.[19] His construction can be extended to the pseudo-orthogonal groups $SO(m,n)$ by analytic continuation. Murnaghan's construction consists in factorizing any element of $SO(n)$ into a product of plane rotations through an angle φ in the half-open range $[-\pi,\pi)$. Considering SO(1,4), we note that due to the indefinite metric of $E(4,1)$, the rotations connecting x_0 with any one of x_1, x_2, x_3, x_4 is hyperbolic, rather than circular. Thus in going from the parametrization of SO(5) to that of SO(4,1), it suffices to make the replacement $\varphi \to i\varphi$, with respect to the angles of rotation for the planes (01), (02), (03), (04). There are thus two types of "rotation" matrices, R, T, the (mn)th element of which are defined by

$$\begin{aligned}(R_{ij}(\varphi))_{mn} &= (\cos\varphi - 1)(\delta_{mi}\delta_{ni} + \delta_{mj}\delta_{nj}) \\ &\quad + (\delta_{mj}\delta_{ni} - \delta_{mi}\delta_{nj})\sin\varphi + \delta_{mn} , \end{aligned}\tag{6.1}$$

$$\begin{aligned}(T_{ij}(\theta))_{mn} &= (\cosh\theta - 1)(\delta_{mi}\delta_{ni} + \delta_{mj}\delta_{nj}) \\ &\quad + (\delta_{mi}\delta_{nj} + \delta_{mj}\delta_{ni})\sinh\theta + \delta_{mn} , \end{aligned}\tag{6.2}$$

where

$$-\pi \leqslant \varphi < \pi, \quad 0 \leqslant \theta < \infty , \tag{6.3}$$

and the subscripts (ij) denote "the ij plane." In terms of these, an arbitrary element of SO(4,1) is factorized as

$$\begin{aligned} q &= R_{34}(\varphi_4) R_{23}(\varphi_1) R_{24}(\varphi_5) R_{12}(\varphi_2) R_{13}(\varphi_3) R_{14}(\varphi_6) S , \\ S &= T_{01}(\theta_1) T_{02}(\theta_2) T_{03}(\theta_3) T_{04}(\theta_4) . \end{aligned}\tag{6.4}$$

We impose the following simplifying restrictions: $\varphi_1 = \varphi_2 = \varphi_3 = 0$, which implies *no rotations in three-space,* and $\theta_4 = 0$, which implies *no* $(x_0 x_4)$ *rotations.* Also we consider an inertial frame S' moving with speed V in the z direction of the inertial frame S. This induces Lorentz boosts in the z direction only. The corresponding de Sitter boosts exists only in the z direction also. For this condition, we require to set $\theta_1 = \theta_2 = \varphi_5 = \varphi_6 = 0$. Hence we left with

$$q = R_{34}(\varphi_4) T_{03}(\theta_3) \tag{6.5}$$

$$= R(\varphi) T(\theta), \quad \text{for simplicity} ,$$

$$= \begin{pmatrix} \cosh\theta & 0 & 0 & \sinh\theta & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \cos\varphi \sinh\theta & 0 & 0 & \cos\varphi\cosh\theta & -\sin\varphi \\ \sin\varphi\sinh\theta & 0 & 0 & \sin\varphi\cosh\theta & \cos\varphi \end{pmatrix} . \tag{6.6}$$

We consider the motion with speed $\mathbf{v} = (0,0,V)$ in S of the origin o' of the inertial frame S'. Applying this to the transformation

$$dx_3' = q_{3j} dx_j = 0 , \tag{6.7}$$

we readily obtain

$$A_t = V - W A_g , \tag{6.8}$$

where

$$A_t = \left(\frac{\partial x_3}{\partial t}\right)_g, \quad A_g = \left(\frac{\partial x_3}{\partial g}\right)_t, \quad W = \frac{dg}{dt}. \tag{6.9}$$

If, as we suppose, g is the classical gravitational potential at a given world point,then $1/A_g$ is the z component of the gradient of g at any given time, while W is the total time rate of change of g. Both of these quantities are properties of extended space-time, and not of the particular transformation of coordinates induced by V. We write $1/A_g = K$. Then

$$\cos\varphi(c\sinh\theta + \cosh\theta A_t) + (\cos\varphi\cosh(\theta)/K - a\sin\varphi)W = 0. \tag{6.10}$$

For any given W, one solution is

$$\begin{aligned} &c\sinh\theta + \cosh\theta A_t = 0, \\ &\cosh\theta - a\,K\tan\varphi = 0, \end{aligned} \tag{6.11}$$

that is,

$$A_t = -c\tanh\theta, \quad a\,K\tan\varphi = \cosh\theta. \tag{6.12}$$

Now put

$$\begin{aligned} &cB_t = A_t, \quad B_g = A_g/a = 1/(aK), \\ &B = B_g(1 - B_g^2)^{-1/2}. \end{aligned} \tag{6.13}$$

Then

$$\tan\varphi = B, \quad \tanh\theta = -B_t. \tag{6.14}$$

We also define the three G variables

$$\begin{aligned} &G_t = (1 - B_t^2)^{-1/2}, \quad G_g = (1 - B_g^2)^{-1/2}, \\ &G = (1 + B^2)^{-1/2}. \end{aligned} \tag{6.15}$$

Hence

$$\begin{aligned} &\tan\varphi = G_t B_g = B, \\ &\cosh\theta = G_t, \quad \sinh\theta = -G_t B_t, \\ &\cos\varphi = G, \quad \sin\varphi = BG. \end{aligned} \tag{6.16}$$

Substituting all these in (6.6), and then in $dx' = q\,dx$, we obtain the de Sitter transformations in extended space-time:

$$\begin{aligned} &t' = G_t(t = x_3 B_t/c), \quad x_1' = x_1, \\ &x_2' = x_2, \quad x_3' = G(y_3 - Bag), \\ &g' = G(g + y_3 B/a), \end{aligned} \tag{6.17}$$

where $y_3 = G_t(x_3 - B_t ct)$.

The set of equations (6.17) is valid for any W. The special theory of relativity assumes that V is constant, but puts $W = 0$. Hence $V = A_t =$ const. For *hyper-relativistic systems*, as defined here, we admit of cases for which W does not vanish, and, in the next section, investigate its effect on the kinematics of free particles as mirrored in the energy-momentum mass relation.

VII. ENERGY-MOMENTUM RELATION

The invariant interval in $E(4,1)$ is

$$ds^2 = (dx_1)^2 + (dx_2)^2 + (dx_3)^2 - (c\,dt)^2 + (a\,dg)^2. \tag{7.1}$$

We put

$$\begin{aligned} &c^2\,dT^2 = -ds^2, \quad k = a/c, \\ &v_i = \frac{dx_i}{dt}, \quad w_i = \frac{dx_i}{dg}, \quad i = 1,2,3, \\ &G_v = (1 - v^2/c^2 - k^2W^2)^{-1/2}. \end{aligned} \tag{7.2}$$

Hence

$$c^2\,dT^2 = c^2\,dt^2/G_v^2. \tag{7.3}$$

Dividing $(dx_1, dx_2, dx_3, c\,dt, a\,dg)$ by $c\,dT$, we obtain a five-velocity

$$U = G_v(v_1, v_2, v_3, c, aW), \tag{7.4}$$

whose invariant square is

$$U^T m U = G_v^2(v_1^2 + v_2^2 + v_3^2 - c^2 + a^2W^2) \tag{7.5}$$

$$= -c^2, \tag{7.6}$$

the same result as the usual square of the four-velocity of special relativity. Here, m is the metric of extended space-time.

In order to make contact with special relativity, we introduce a number m_0 by rewriting (7.5) as

$$U^T m U = m_0^2\,G_v^2\,(\mathbf{v}\cdot\mathbf{v} - c^2 + a^2W^2)/m_0^2 \tag{7.7}$$

$$= (\mathbf{p}\cdot\mathbf{p} - E/c^2)m_0^{-2}, \tag{7.8}$$

where

$$\mathbf{p} = m_0 G_v \mathbf{v}, \tag{7.9}$$

$$E = m_0 G_v c^2(1 - a^2W^2)^{1/2}. \tag{7.10}$$

We identify $\mathbf{p}$ with momentum, and E with energy. There is a contribution to the energy from $W = dg/dt$. From (7.8), we see that the energy momentum relation remains

$$E^2 = c^2p^2 + m_0^2\,c^4. \tag{7.11}$$

However, the definition of E in terms of the *velocity* $\mathbf{v}$ now takes the form

$$\begin{aligned} E &= (m_0c^2(1 - a^2W^2)^{1/2}(1 - v^2/c^2 - a^2W^2)^{-1/2}) \\ &= m_0c^2(1 - Rv^2/c^2)^{-1/2}, \end{aligned} \tag{7.12}$$

$$R = 1/(1 - a^2W^2). \tag{7.13}$$

VIII. CONTRACTION OF THE STRUCTURE GROUP OF THE BUNDLE

The group SO(4,1) is the group of 5×5 matrices q, satisfying the relations

$$\det q = 1, \quad q^T Q q = Q, \tag{8.1}$$

where the matrix Q has elements

$$Q_{ij} = \delta_{ij} - c^2\delta_{0i}\delta_{j0} + a^2\delta_{4i}\delta_{j4}, \quad i,j = 0,1,2,3,4. \tag{8.2}$$

The associated invariant quadratic form in $E(4,1)$ is then $x^T Qx$, where $x = (t, x_1, x_2, x_3, g)$.

From (8.1) and (8.2), we have

$$\begin{aligned} 0 &= q_{mi}Q_{mn}q_{nk} - Q_{ik} \\ &= c^2(\delta_{0i}\delta_{k0} - q_{0i}q_{0k}) + a^2(q_{4i}q_{4k} - \delta_{4i}\delta_{k4}) \\ &\quad + q_{ni}q_{nk} - \delta_{ik}. \end{aligned} \tag{8.3}$$

If we now imagine "g" switched off, but at the same time letting "a" go to infinity in such a way that "ag" remains finite, then we have a local nongravitational situation in the sense of the free-falling test experiments of Will.[20]

In the limit a^2 going to infinity, there are two distinct possibilities: (a) both i and k less than 4,

$$q_{4i}q_{4k} = (\delta_{ik} - q_{ni}q_{nk} + c^2(q_{0i}q_{0k} - \delta_{0i}\delta_{k0}))/a^2$$

tends to zero; (b) $i = k = 4$,

$$q_{44}^2 - 1 = (1 + c^2 q_{04}^2 - q_{n4}q_{n4})/a^2,$$

and $q_{44}{}^2$ tends to 1. Thus we have the affine transformation

$$\begin{pmatrix} t' \\ x_1' \\ x_2' \\ x_3' \\ g' \end{pmatrix} = \left(\begin{array}{cccc|c} & & & & a_0 \\ & & & & a_1 \\ & & J_{ik} & & a_2 \\ & & & & a_3 \\ \hline 0 & 0 & 0 & 0 & 1 \end{array}\right) \cdot \begin{pmatrix} t \\ x_1 \\ x_2 \\ x_3 \\ g \end{pmatrix}, \tag{8.4}$$

where

$$a_i = q_{i4}, \quad J_{ik} = q_{ik}, \quad i,k = 0,1,2,3. \tag{8.5}$$

The 4×4 matrix J with elements J_{ik} belongs to SO(3,1). The contracted group is therefore the Poincaré group P, which as a subgroup of SO(4,2) acts on $M = \mathrm{SO}(4,2)/\mathrm{SO}(4,1)$, and can therefore be used to construct an affine tangent bundle over M. Our construction is akin to what Drechsler[21] has called "the contraction of the SO(4,1) de Sitter bundle over space-time to the affine tangent bundle over space-time." Under the contraction, the variable g becomes an *absolute* quantity, a *spectator variable*, in the same way that time is absolute in Galilean relativity. Gravity is thus essentially decoupled from special relativistic systems. Be that as it may, the intrinsic five-dimensional nature of the space of special relativity is clear from the above.

Now let us relate the limit $a^2 \to \infty$ to Eqs. (7.2) and (7.12), in which we note that it is the same $a = kc$ that is involved in the contraction process. Therefore, in order to have k^2W^2 vanish even under the limit $a^2 \to \infty$, we must have W^2/c^2 tending to zero sufficiently fast. This is achieved in the special theory of relativity by setting W identically zero at the outset. Then the term kW does not arise at all. For hyper-relativistic systems, however, *which include all quantum mechanical systems*, W need not be zero.

IX. CONCLUSION

We have derived the hyper-relativistic transformation laws for free particles on the SO(4,1) de Sitter bundle over SO(4,2)/SO(4,1), and shown how the Heisenberg commutation rules of quantum mechanics naturally arise from a particular contraction of the structure group of the bundle. An intrinsic five-dimensional flat space emerges as the configuration space of relativistic quantum mechanics. The first four coordinates are conventional space-time, but the fifth coordinate is identified with the classical Newtonian gravitational potential field g. In the special theory of relativity, this variable is an absolute variable. One special feature of this theory is that the coordinate variables x_i are not represented by operators in a Hilbert space, but rather by a set of vector fields dual to them in the sense explained in the text. There is also a suggestion that there might indeed be a *hierarchy* of Heisenberg commutation rules in nature.

ACKNOWLEDGMENT

Much of this paper was completed in 1983 when the author was a Senior Associate member of the International Centre for Theoretical Physics, Trieste, Italy. He is grateful to the Director of the Centre, Professor Abdus Salam, F. R. S., for hospitality at the Centre.

[1] R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin, Reading, MA, 1979).
[2] H. D. Doebner and J. Tolar, J. Math. Phys. **16**, 975 (1975).
[3] G. W. Mackey, Bull. Am. Math. Soc. **69**, 628 (1963).
[4] G. W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968).
[5] V. S. Varadarayan, *Geometry of Quantum Theory*, Vol. II, *Geometry of Covariant Systems* (Van Nostrand–Reinhold, New York, 1970).
[6] D. Husemoller, *Fibre Bundles* (McGraw-Hill, New York, 1966).
[7] S. Kobayashi and K. Monizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.
[8] I. M. Gelfand, Am. Math. Soc. Transl. Ser. 2 **37**, 31 (1964).
[9] S. Helgasson, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[10] H. T. Flint, *The Quantum Equation and the Theory of Fields* (Methuen, London, 1966).
[11] E. Inonu and E. P. Wigner, Proc. Natl. Acad. Sci. USA **39**, 510 (1953).
[12] E. J. Saletan, J. Math. Phys. **2**, 1 (1961).
[13] A. Maduemezia, Nuovo Cimento A **51**, 932 (1967).
[14] A. Maduemezia, Nuovo Cimento A **51**, 585 (1967).
[15] J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1972).
[16] A. O. Barut, *Dynamical Groups and Generalized Symmetries in Quantum Theory* (Univ. Canterbury, Christchurch, New Zealand, 1971).
[17] A. Einstein, *Relativity* (Methuen, London, 1954).
[18] A. Maduemezia, J. Math. Phys. **12**, 1681 (1971).
[19] F. D. Murnaghan, *The Unitary and Rotation Groups* (Spartan, Washington, D. C., 1962).
[20] C. M. Will, Phys. Rep. **113**, 345 (1984).
[21] W. Drechsler, J. Math. Phys. **18**, 1358 (1977).

Wiener measures for path integrals with affine kinematic variables

Ingrid Daubechies[a)]
Theoretical Physics, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium
John R. Klauder
A.T. & T. Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974
Thierry Paul
Centre de Physique Théorique, Centre National de Recherche Scientifique, Luminy Case 907, F-13288 Marseille Cedex, France

(Received 12 June 1986; accepted for publication 3 September 1986)

The results obtained earlier have been generalized to show that the path integral for the affine coherent state matrix element of a unitary evolution operator $\exp(-iTH)$ can be written as a well-defined Wiener integral, involving Wiener measure on the Lobachevsky half-plane, in the limit that the diffusion constant diverges. This approach works for a wide class of Hamiltonians, including, e.g., $-d^2/dx^2 + V(x)$ on $L^2(\mathbb{R}_+)$, with V sufficiently singular at $x = 0$.

I. INTRODUCTION

The observables that are the quantum kinematical operators are usually defined to have commutation relations analogous to the Poisson bracket structure of the associated classical kinematical variables. Examples are a single canonical pair and the Heisenberg commutation relation, or angular momentum variables and the Lie algebra of angular momentum operators. We shall say that p, q are classical affine variables if $q > 0$ (or $p > 0$), for example, with the other variable p (or q) being unrestricted. Since one variable is the generator of translations of the other, it follows that some conflict with the range restriction is possible, a situation that reflects itself in the quantum theory by the fact that the operators Q and P cannot both be observables (self-adjoint operators) satisfying the Heisenberg commutation relation if $Q > 0$ (or $P > 0$). An acceptable substitute for the nonobservable operator is the dilation operator $D = \frac{1}{2}(QP + PQ)$, which can always be chosen self-adjoint along with the positive operator. The Lie-algebra relation $[Q,D] = iQ$ with $Q > 0$ is just the quantum image of the Poisson-bracket relation $\{q,d\} = q$, $q > 0$, where $d = qp$. The generator D preserves the positivity of Q just as the classical counterpart d preserves the positivity of q. The indicated Lie algebra relation is that of the affine group, sometimes called the $(ax + b)$-group, which is the group of translations (b) and scale changes without reflection $(a > 0)$ of the real line into itself, $x \rightarrow x' = ax + b$. Thus we refer to Q (or P) and D as quantum affine kinematical variables, and in view of the simple relation between d, p, and q, we loosely refer to p, q with $q > 0$ (or $p > 0$) as classical affine kinematic variables as noted earlier.

Focusing on the $p > 0$ case for the moment, we may imagine a formal phase-space path integral quantization of such a system given by

$$\mathcal{N}^{-1} \int \exp\left\{ i \int [p\dot{q} - H(p,q)]\, dt \right\} \prod_t [dp_t dq_t]\,, \qquad (1.1)$$

where all paths satisfy the condition $p(t) > 0$. This expression is plagued by two problems. The first problem relates to what (1.1) could possibly represent since it *cannot* be the propagator expressed in the Q-representation for the simple reason that if $[Q,P] = i$ and $P > 0$ then no Q-representation is possible. A satisfactory answer to the first problem was given earlier[1] in which (1.1) was formally interpreted as the propagator expressed in the affine coherent-state representation (which makes fundamental use of the operators P and D rather than P and Q; see Refs. 2, 3). The second problem with (1.1) pertains to the formal nature of the path integral. In Ref. 1 meaning was given to (1.1) as the limit of a fairly standard lattice-space regularization. This approach made little direct contact with paths defined for continuous time as in the classical theory, and besides, it was relatively heuristic. On the other hand, in recent work[4] pertaining to the usual canonical case (and also for spin kinematical variables), it was shown how the appropriate coherent-state representation of the propagator can be defined as the limit of well-defined path integrals over pinned Brownian-motion measures as the diffusion constant diverges. The purpose of the present paper is to extend this alternative form of regularization and its associated rigorous definition of a path-integral representation to systems involving affine variables. To begin with, however, it is useful to give a brief description of the construction in Ref. 4 for the canonical case.

For a given Hamiltonian H, we defined[4] the path integrals

$$2\pi e^{\nu(t''-t')/2} \int \exp\left[\frac{i}{2} \int (p\, dq - q\, dp) - i \int h(p,q) dt \right] d\mu_W^\nu(p) d\mu_W^\nu(q)\,, \qquad (1.2)$$

where $d\mu_W^\nu(p)$ and $d\mu_W^\nu(q)$ are Wiener measures associated to two independent Brownian processes (one in p, one in q) with diffusion constant ν, and pinned at p',q' for $t = t'$, at p'',q'' for $t = t''$. The function h in (1.2) is the antinormal ordered symbol[2] of H. For finite ν, (1.2) is a perfectly well-defined path integral on phase space. It has been proved[4] that for a wide class of Hamiltonians, the limit for $\nu \rightarrow \infty$ of (1.2) gives the coherent state matrix element

[a)] "Bevoegdverklaard Navorser" at the National Foundation for Scientific Research, Belgium.

 0022-2488/87/010085-18$02.50

$\langle p'',q''|\exp[-i(t''-t')H]|p',q'\rangle$.

This procedure is not restricted to only the canonical kinematical variables. In Ref. 5 an outline is given of how the above construction can be extended to general semisimple Lie groups. One has then to use the corresponding generalized coherent states.[6] One can define a metric on the group manifold associated to these coherent states,[5] and use the corresponding Laplace–Beltrami operator to define a generalized Wiener measure. Examples of interest outlined in Ref. 3 are (i) the Weyl–Heisenberg group, (ii) the group SU(2), and (iii) the affine $(ax+b)$-group, corresponding to, respectively, canonical, spin, and affine kinematic variables. The first two were extensively discussed in Ref. 4. Here we present a more detailed study of the affine variable case. In particular, we derive explicit conditions characterizing the class of Hamiltonians that can be treated by our methods, and we give several examples as well.

This paper is organized as follows. In Sec. II we review the definition and some properties of the coherent states associated with the $(ax+b)$-group.[2,3] We shall adopt notation related to that in Ref. 3, which is different from the notation in Refs. 1 and 2. We shall also indicate how to pass from one notation to the other. It is convenient to break the construction into two parts. In Sec. III we study the path integral for zero Hamiltonian. We introduce the Brownian process on the half plane, use it to construct the path integral, and show that in the limit of diverging diffusion constant the path integral converges to the coherent state overlap function [as it should, since $\exp(-itH)=\mathbb{1}$ if $H=0$]. In Sec. IV we discuss the path integral with a nonzero Hamiltonian, and we derive sufficient conditions on the Hamiltonian so that the limit for diverging diffusion constant leads to the appropriate coherent-state matrix element of the evolution operator.

II. THE $(ax+b)$-GROUP AND THE AFFINE COHERENT STATES

Let us review the definition of the $(ax+b)$-group and the associated coherent states, and give some of their properties. Most of this discussion is analogous to what happens for the Weyl–Heisenberg group and its associated coherent states, the more familiar canonical coherent states. Both the affine and the canonical coherent states are examples of the construction of coherent states associated with general Lie-groups.[6]

A. The $(ax+b)$-group

The "$(ax+b)$-group" is the set $M_+ := \mathbb{R}^*_+ \times \mathbb{R}$, where $\mathbb{R}^*_+ = (0,\infty)$, with the group law

$$(a'',b'')(a',b') = (a''a',b''+a''b') .$$

This group has two (faithful) inequivalent irreducible unitary representations U_+ and U_-. We shall consider their following realizations on $L^2(\mathbb{R}_+)$. For $\psi \in L^2(\mathbb{R}_+)$, one defines

$$[U_\pm(a,b)\psi](x) = a^{1/2}e^{\pm ibx}\psi(ax) . \tag{2.1}$$

We shall mainly use U_+, except when specified otherwise. The subscript $+$ will often be dropped.

Both representations U_+ and U_- are square integrable. This means[7] that there exists an (unbounded) positive self-adjoint operator C on $L^2(\mathbb{R}_+)$ such that

$\forall \psi_1,\psi_2 \in D(C^{1/2})$, $\forall \phi_1,\phi_2 \in L^2(\mathbb{R}_+)$:

$$\int d\mu(a,b)\langle\phi_1,U_\pm(a,b)\psi_1\rangle\langle U_\pm(a,b)\psi_2,\phi_2\rangle = \langle C^{1/2}\psi_2,C^{1/2}\psi_1\rangle\langle\phi_1,\phi_2\rangle . \tag{2.2}$$

Here $d\mu(a,b) = (1/2\pi)a^{-2}\,da\,db$ is the left-invariant measure on the $(ax+b)$-group. The operator C is given by

$$(C\psi)(x) = x^{-1}\psi(x) . \tag{2.3}$$

In particular (2.2) implies that, for all $\psi \in D(C^{1/2})$, $\|\psi\|=1$,

$$\int d\mu(a,b)\,U_\pm(a,b)|\psi\rangle\langle\psi|U_\pm(a,b)^* = c(\psi)\mathbb{1} , \tag{2.4}$$

with

$$c(\psi) = |C^{-1/2}\psi|^2 = \int_0^\infty dx(1/x)|\psi(x)|^2 . \tag{2.5}$$

The closed spaces $\mathscr{H}_\pm$ spanned by the sets

$$\{\langle U_\pm(\cdot,\cdot)\psi,\phi\rangle;\ \ \psi\in D(C^{1/2}),\ \ \phi\in L^2(\mathbb{R}_+)\}$$

are mutually orthogonal subspaces of $L^2(M_+) := L^2(M_+;d\mu)$. Together, $\mathscr{H}_+$ and $\mathscr{H}_-$ span the whole space $L^2(M_+)$. This can easily be checked by explicit calculation.

All this enables us to build orthonormal bases of $L^2(M_+)$, starting from orthonormal bases in $L^2(\mathbb{R}_+)$. Let $\{\phi_j: j\in\mathbb{N}\}$, $\{\psi_j: j\in\mathbb{N}\}$ be two orthonormal bases in $L^2(\mathbb{R}_+)$ such that $\psi_j \in D(C^{-1/2})$ for all j. Define elements $f^\pm_{ij}$ of $L^2(M_+)$ by

$$f^\pm_{ij}(a,b) = \langle U_\pm(a,b)C^{-1/2}\psi_i,\phi_j\rangle . \tag{2.6}$$

It is clear that for all i, j, $f^\epsilon_{ij}\in\mathscr{H}_\epsilon$ ($\epsilon = +$ or $-$). On the other hand, both $\{f^+_{ij}; i,j\in\mathbb{N}\}$ and $\{f^-_{ij}; i,j\in\mathbb{N}\}$ are orthonormal sets, as a consequence of (2.2). One easily checks that, for $\epsilon = +$ or $-$, $\{f^\epsilon_{ij}; i,j\in\mathbb{N}\}$ constitutes a basis for $\mathscr{H}_\epsilon$. The set $\{f^\epsilon_{ij}; i,j\exists\mathbb{N}, \epsilon = +$ or $-\}$ is therefore an orthonormal basis for $L^2(M_+)$.

Let now B be a Hilbert–Schmidt operator on $L^2(\mathbb{R}_+)$ such that $C^{-1/2}B$ is trace class. Then

$$B = \sum_j \lambda_j|\psi_j\rangle\langle\phi_j| ,$$

where $\{\phi_j; j\in\mathbb{N}\}$, $\{\psi_j; j\in\mathbb{N}\}$ are orthonormal bases in $L^2(\mathbb{R}_+)$, with $\psi_j\in D(C^{-1/2})$ for all j, $\Sigma_j|\lambda_j|^2<\infty$. Since $C^{-1/2}B$ is trace class we can define

$$\begin{aligned}[F(B)](a,b) &= (1/\sqrt{2})\mathrm{Tr}[(U_+(a,b)+U_-(a,b))C^{-1/2}B] \\ &= (1/\sqrt{2})\sum_{j,\epsilon}\lambda_j\langle\phi_j,U_\epsilon(a,b)C^{-1/2}\psi_j\rangle .\end{aligned} \tag{2.7}$$

From the preceding paragraph it is clear that (2.7) can be considered as an expansion of $F(B)$ with respect to an orthonormal base in $L^2(M_+)$. Since the sequence of coefficients is square summable, $\Sigma_{j,\epsilon}|\lambda_j|^2 = 2\,\mathrm{Tr}(B^*B)$, we immediately see that $F(B)\in L^2(M_+)$, with

$$\int d\mu(a,b)\,|[F(B)](a,b)|^2=\frac{1}{2}\sum_{j,\epsilon}|\lambda_j|^2=\mathrm{Tr}(B^*B)\,. \tag{2.8}$$

The set of Hilbert–Schmidt operators B for which $C^{-1/2}B$ is trace class is dense in the space τ_2 of Hilbert–Schmidt operators. One can use this to extend the map $B\to F(B)$ to all of τ_2. This extension is a unitary map from τ_2 to $L^2(M_+)$. This is the $(ax+b)$-group analog of a well-known result for the Weyl–Heisenberg group.[8]

B. The affine coherent states

A special role in our path integral results below will be played by extremal-weight vectors for the unitary representation under consideration (see Ref. 5). In our case these are the vectors[1] (normalized to 1)

$$\psi_\beta(x)=2^\beta\Gamma(2\beta)^{-1/2}x^{\beta-1/2}e^{-x}\,. \tag{2.9a}$$

In order for $c_\beta\equiv c(\psi_\beta)$ to be finite, one has to impose $\beta>\frac{1}{2}$. One finds

$$c_\beta=(\beta-\tfrac{1}{2})^{-1}\,. \tag{2.9b}$$

We shall use these minimal weight vectors ψ_β as "fiducial vectors" [6] for the construction of the affine coherent states,

$$|a,b;\beta\rangle=U(a,b)\psi_\beta\,.$$

From (2.4) one now immediately has the affine coherent state resolution of the identity

$$c_\beta^{-1}\int d\mu(a,b)\,|a,b;\beta\rangle\langle a,b;\beta|=\mathbb{1}\,. \tag{2.10}$$

The "overlap function" of different coherent states (same value of β) is given by

$$\begin{aligned}\langle a'',b'';\beta|a',b';\beta\rangle&=\left[\frac{a''+a'+i(b''-b')}{2\sqrt{a''a'}}\right]^{-2\beta}\\&=\left[\frac{2}{1+\cosh d(a'',b'';a',b')}\right]^{\beta}\\&\quad\times\exp\left(-2\beta i\tan^{-1}\frac{b''-b'}{a''+a'}\right),\end{aligned} \tag{2.11}$$

where d denotes the metric distance[9] on the Lobachevsky half-plane M_+

$$\begin{aligned}d(a'',b'';a',b')&\\&=\cosh^{-1}\left[1+\frac{(a''-a')^2+(b''-b')^2}{2\,a''a'}\right].\end{aligned} \tag{2.12}$$

For every $\beta>\frac{1}{2}$ one can define the following map on $L^2(\mathbb{R}_+)$:

$$\begin{aligned}(U_\beta\phi)(a,b)&=c_\beta^{-1/2}\langle a,b;\beta|\phi\rangle\\&=c_\beta^{-1/2}2^\beta[\Gamma(2\beta)]^{-1/2}a^\beta\\&\quad\times\int_0^\infty dx\,x^{\beta-1/2}e^{-(a+ib)x}\phi(x)\,.\end{aligned} \tag{2.13}$$

It is clear from (2.10) that U_β is an isometry from $L^2(\mathbb{R}_+)$ to $L^2(M_+)$. These maps U_β are the analogs of the Bargmann transform for the Weyl–Heisenberg case.[10] The image $\mathscr{H}_\beta\equiv U_\beta L^2(\mathbb{R}_+)$ consists of exactly those elements f of $L^2(M_+)$ that can be written as

$$f(a,b)=a^\beta\phi(a+ib)\,,$$

where $\phi(z)$ is an entire analytic function on the half-plane $\mathrm{Re}\,z>0$.

The Hilbert space $\mathscr{H}_\beta$ is a reproducing kernel Hilbert space,[11] with reproducing kernel $c_\beta^{-1}\langle a'',b'';\beta|a',b';\beta\rangle$. In other words, for f in $\mathscr{H}_\beta$,

$$f(a,b)=c_\beta^{-1}\int d\mu(a',b')\langle a,b;\beta|a',b';\beta\rangle f(a',b')\,.$$

This means in particular that the orthogonal projection operator P_β mapping $L^2(M_+)$ onto $\mathscr{H}_\beta$ is an integral operator with integral kernel

$$P_\beta(a'',b'';a',b')=c_\beta^{-1}\langle a'',b'';\beta|a',b';\beta\rangle\,. \tag{2.14}$$

C. Correspondence with the *pq*-notation

We mentioned in the Introduction that our notation would not coincide with that in Ref. 1. To conclude this section we give the correspondence between our present notation and the pq-notation in Ref. 1.

For fixed β, define $p=\beta a^{-1}$, $q=-b$. We shall also rescale the measure; $d\tilde\mu(p,q)$ is the image of $c_\beta^{-1}d\mu(a,b)$, i.e.,

$$d\tilde\mu(p,q)=c_\beta^{-1}\beta^{-1}\frac{dp\,dq}{2\pi}=\frac{1-1/2\beta}{2\pi}\,dp\,dq\,.$$

With this change of notation, (2.13) becomes, for instance,

$$\begin{aligned}(\tilde U_\beta\psi)(p,q)&=(2\beta)^\beta[\Gamma(2\beta)]^{-1/2}p^{-\beta}\\&\quad\times\int_0^\infty dk\,k^\beta e^{-k(\beta p^{-1}-iq)}\psi(k)\,.\end{aligned} \tag{2.15}$$

This corresponds exactly with Eq. (24) in Ref. 1.

Using this correspondence every result we shall obtain here can be translated into the pq-notation used in Ref. 1, and vice versa. At the end of Sec. IV D we shall state our main result in pq-notation as well as in the ab-notation which will be used throughout this paper.

III. THE PATH INTEGRAL FOR ZERO HAMILTONIAN

In the ab-notation, with the correspondence rules of Sec. II C, (1.1) becomes

$$\mathscr{N}^{-1}\int\exp\left[-i\beta\int a^{-1}\,db-\int h(a,b)\,dt\right]\prod_t\frac{da_t\,db_t}{a_t^2}\,, \tag{3.1}$$

where $A>0$ throughout the integration domain. We shall give a sense to this expression by a regularization that leads to a Wiener measure, on the Lobachevsky half-plane, for diffusion constant ν. In the end we take the limit $\nu\to\infty$. For related ideas (regularization by extra factors that formally disappear in the limit as a diffusion constant diverges), see Ref. 12.

In this section we restrict ourselves to the case $h=0$. The general case $h\neq0$ will be handled in the next section.

Let us first define the Wiener measure on the Lobachevsky half-plane. The Laplace–Beltrami operator is given by

$$\Delta=a^2(\partial_a^2+\partial_b^2) \tag{3.2}$$

(in the pq-notation, $\Delta = \partial_p p^2 \partial_p + \beta^2 p^{-2} \partial_q^2$). This is a symmetric operator in $L^2(M_+)$, essentially self-adjoint on $C_0^\infty(M_+)$, the C^∞-functions on M_+ with compact support away from $a=0$ (this essential self-adjointness is most easily checked in the pq-notation).

The heat kernel for this Laplace–Beltrami operator is given by[9]

$$K_t(a'',b'';a',b') \equiv [\exp(t\Delta)](a'',b'';a',b') = \frac{e^{-t/4}}{2\sqrt{2\pi}t^{3/2}} \int_\delta^\infty \frac{x e^{-x^2/4t}}{\sqrt{\cosh x - \cosh\delta}}\, dx, \tag{3.3}$$

where $\delta = d(a'',b'';a'b')$ is the metric distance (2.12). We define the affine (pinned) Wiener measure with diffusion constant ν, denoted $d\mu^{\nu,T}_{W;a'',b'';a',b'}$, as the measure on path space, pinned at a',b' for $t=0$, at a'',b'' for $t=T$, such that

$$\int d\mu^{\nu,T}_{W;a'',b'';a',b'} = K_{\nu T}(a'',b'';a',b'). \tag{3.4}$$

Requiring (3.4) for all (a'',b''), $(a',b')\in M_+$, and all $T>0$ defines $d\mu^\nu_W$ unambiguously. We shall drop the super- and subscripts T, a'', b'', a', and b' in the sequel.

We use this measure to regularize (3.1) in the following way. We define

$$\mathscr{P}^0_\nu(a'',b'';a',b';T) = c_\beta e^{\nu T\beta} \int \exp\Big(-i\beta \int a^{-1}\, db\Big)\, d\mu^\nu_W(a,b). \tag{3.5}$$

The expression $\int a^{-1}\,db$ should be considered as a stochastic integral, to be calculated using the Stratonovich (midpoint rule) procedure. Formally (3.5) can be written as

$$\mathscr{P}^0_\nu(a'',b'';a',b';T) = \mathscr{N}\int \exp\Big[-i\beta\int a^{-1}\dot b\, dt - \frac{1}{\nu}\int a^{-2}(\dot a^2 + \dot b^2)\, dt\Big] \times \prod_t \frac{da_t\, db_t}{a_t^2},$$

where the factors c_β and $e^{\nu TB}$ have been absorbed in the (infinite) normalization constant $\mathscr{N}$. This formal expression shows how (3.5) can indeed be viewed as a regularization of (3.1) (for the case $h=0$). In the final step of our regularization procedure we take the limit for $\nu\to\infty$; in this limit the regularizing factor in the above formal expression vanishes.

It is our aim in this section to prove that

$$\lim_{\nu\to\infty} \mathscr{P}^0_\nu(a'',b'';a',b';T) = \langle a'',b'';\beta | a',b';\beta\rangle. \tag{3.6}$$

This is exactly what the general expression (3.1) or (1.1) should lead to[1] in the case $h=0$.

We start by studying $\mathscr{P}^0_\nu$ for finite ν.

Lemma 3.1: $c_\beta^{-1}\mathscr{P}^0_\nu$ is the integral kernel of a semigroup on $L^2(M_+)$:

$$\mathscr{P}^0_\nu(a'',b'';a',b';T) = c_\beta[\exp(-\nu TA)](a'',b'';a',b'). \tag{3.7}$$

The operator A is given by

$$A = -\beta - a^2[\partial_a^2 + (\partial_b + i\beta/a)^2] \tag{3.8a}$$

$$= a^2(\partial_a + \partial_b + i\beta/a)(i\partial_a - \partial_b - i\beta/a). \tag{3.8b}$$

In particular, A is a positive self-adjoint operator, with domain $D(-\Delta)$.

Proof: It is clear that the $c_\beta^{-1}\mathscr{P}^0_\nu$ satisfy a semigroup property, i.e.,

$$\int d\mu(a,b)\mathscr{P}^0_\nu(a'',b'';a,b;t_2)\mathscr{P}^0_\nu(a,b;a',b';t_1) = c_\beta\mathscr{P}^0_\nu(a'',b'';a',b';t_1+t_2).$$

On the other hand, we have

$$|c_\beta^{-1}\mathscr{P}^0_\nu(a,b;a',b';t)| \leqslant e^{\nu t\beta}K_{\nu t}(a,b,a',b').$$

This already implies that $c_\beta^{-1}\mathscr{P}^0_\nu$ is the integral kernel of a semigroup of operators, i.e., Eq. (3.7), with

$$A \geqslant -\beta + \tfrac{1}{4}. \tag{3.9}$$

Here we have used that $-\Delta\geqslant\frac{1}{4}$ on the Lobachevsky half-plane. Following the standard procedure, and using the midpoint rule for the stochastic integral $\int a^{-1}\,db$, one obtains the following differential equation for $\mathscr{P}^0_\nu$:

$$\partial_t\mathscr{P}^0_\nu(a,b;a',b';t) = \{-\beta - a^2[\partial_a^2 + (\partial_b + i\beta/a)^2]\}\mathscr{P}^0_\nu(a,b;a',b';t).$$

This implies that the infinitesimal generator A is given by (3.8). We have

$$A = -\Delta - \beta + \beta^2 - 2i\beta a\partial_b.$$

Since, for all $\psi\in D(-\Delta)$, and for all $\epsilon>0$,

$$\|a\partial_b\psi\|^2 = -\langle\psi, a^2\partial_b^2\psi\rangle \leqslant \langle\psi,(-\Delta)\psi\rangle \leqslant \epsilon\|-\Delta\psi\|^2 + (1/4\epsilon)\|\psi\|^2,$$

we see that $A-(-\Delta)$ is $(-\Delta)$-bounded with infinitesimally small bound. Hence A is self-adjoint, with domain $D(-\Delta)$. Finally it follows from (3.8b) that A is positive.

Note: It follows from the proof that every core for $-\Delta$ is a core for A. In particular, A is essentially self-adjoint on $C_0^\infty(M_+)$, the set of C^∞-functions on M_+ with compact support away from $a=0$.

We shall see below that we can do much better than Lemma 3.1. We shall see that A has an isolated eigenvalue at 0. If we denote by P_0 the projection onto the eigenspace of A for the eigenvalue 0, we then see that

$$e^{-\nu TA} \underset{\nu\to\infty}{\to} P_0 \quad (T>0).$$

This will then lead to statement (3.6).

To carry out this program, we have to determine the spectrum of A and the corresponding eigenspaces. We shall reduce this to a spectral problem on $L^2(\mathbb{R}_+)$ rather than on $L^2(M_+)$.

We first introduce the infinitesimal generators of $U_\pm(a,b)$. Both $V_\pm(b) = U_\pm(1,b)$ and $W(\alpha) = U(e^\alpha,0)$ are strongly continuous unitary one-parameter groups. Their generators are, respectively, Q and D, i.e.,

$$V_\pm(b) = e^{\pm ibQ}, \quad W(\alpha) = e^{i\alpha D},$$

where Q and D are defined by

$$(Q\psi)(x) = x\psi(x),$$

$$(D\psi)(x) = -ix\psi'(x) - (i/2)\psi(x).$$

One easily checks that these are indeed self-adjoint operators on $L^2(R_+)$. The set $C_0^\infty(\mathbb{R}_+)$ of all C^∞-functions with compact support away from 0 is a core for both D and Q. Then $U_\pm(a,b)$ can be written in terms of Q, D as follows:

$$U_\pm(a,b) = e^{\pm ibQ}e^{i(\log a)D} = e^{i(\log a)D}e^{\pm(ib/a)Q}. \tag{3.10}$$

Note that $C = Q^{-1}$. With the help of all this we prove the following lemma.

Lemma 3.2: On $L^2(\mathbb{R}_+)$ we define the operators $D^2 + Q^2 \mp 2\beta Q + (\beta - \frac{1}{2})^2$, with domain $C_0^\infty(\mathbb{R}_+)$. These are symmetric operators; we denote their closures by $H_\pm$. Then

(1) $H_\pm$ are self-adjoint,

(2) $\forall \psi,\phi \in C_0^\infty(R_+)$, $\langle U_\pm(\cdot,\cdot)C^{-1/2}\psi,\phi\rangle \in D(A)$,

and

$$A\langle U_\pm(a,b)C^{-1/2}\psi,\phi\rangle = \langle U_\pm(a,b)C^{-1/2}H_\pm\psi,\phi\rangle. \tag{3.11}$$

Proof: To prove the first statement it is convenient to make a unitary transformation from $L^2(\mathbb{R}_+)$ to $L^2(\mathbb{R})$. We define, for $\psi \in L^2(\mathbb{R}_+)$,

$$(U\psi)(s) = e^{s/2}\psi(e^s). \tag{3.12}$$

Accordingly $UC_0^\infty(\mathbb{R}_+) = C_0^\infty(\mathbb{R})$, the set of C^∞-functions with compact support. On the other hand,

$$U[D^2 + Q^2 \mp 2\beta Q + (\beta - \tfrac{1}{2})^2]U^{-1} = -\frac{d^2}{ds^2} + e^{2s} \mp 2\beta e^s + \left(\beta - \frac{1}{2}\right)^2. \tag{3.13}$$

Since the potential $V_\pm(s) = 2e^s \mp 2\beta e^s + (\beta - \frac{1}{2})^2$ is the sum of a bounded potential and a positive smooth potential, the operators (3.13) are essentially self-adjoint on $C_0^\infty(\mathbb{R})$ by Theorem X.29 in Ref. 13. This proves the first statement.

It is easy to check that for $\psi,\phi \in C_0^\infty(\mathbb{R}_+)$ the functions $f_{\psi,\phi}^\pm(a,b) = \langle U_\pm(a,b)C^{-1/2}\psi,\phi\rangle$ are well-defined C^∞-functions in a,b. Their support is contained in a set of the form $[c_1,c_2]\times\mathbb{R}$, with $c_1 > 0$; they decrease more rapidly in b than any inverse polynomial, and this uniformly in a. This is sufficient to ensure that $f_{\psi,\phi}^\pm \in D(-\Delta) = D(A)$, and also to justify the calculations below.

We have

$$\begin{aligned}
(Af_{\psi,\phi}^\pm)(a,b) &= a^2\langle(-i\partial_a + \partial_b - i\beta/a)(-i\partial_a - \partial_b + i\beta/a)U_\pm(a,b)Q^{1/2}\psi,,\phi\rangle \\
&= a^2\langle(-i\partial_a + \partial_b - i\beta/a)U_\pm(a,b)(1/a)(D \mp iQ + i\beta)Q^{1/2}\psi,\phi\rangle \\
&= a^2\langle U_\pm(a,b)[a^{-2}(D \pm iQ - i\beta)(D \mp iQ + i\beta) + ia^{-2}(D \mp iQ + i\beta)]Q^{1/2}\psi,\phi\rangle \\
&= \langle U_\pm(a,b)(D \pm iQ - i\beta + i)(D \mp iQ + i\beta)Q^{1/2}\psi,\phi\rangle \\
&= \langle U_\pm(a,b)Q^{1/2}(D \pm iQ - i\beta + i/2)(D \mp iQ + i\beta - i/2)\psi,\phi\rangle \\
&= \langle U_\pm(a,b)Q^{1/2}[D^2 + Q^2 \mp 2\beta Q + (\beta - \tfrac{1}{2})^2]\psi,\phi\rangle,
\end{aligned}$$

where we have repeatedly used that $[D,Q^\alpha] = -i\alpha Q^\alpha$. Hence (3.11) follows. As a consequence of (3.11) the subspaces $\mathscr{H}_\pm$ are invariant subspaces for A. Moreover the spectrum of $A|_{\mathscr{H}_\pm}$ is exactly the spectrum of $H_\pm$. □

Lemma 3.3: Let $A_\pm$ be the restrictions of A to $\mathscr{H}_\pm$, with domains $D(-\Delta)\cap\mathscr{H}_\pm$. Then $\sigma(A_\pm) = \sigma(H_\pm)$.

Proof: Let $P_\Omega^\pm$ be the family of spectral projection operators associated with $H_\pm$.

Let ψ_j be an orthonormal base in $L^2(\mathbb{R}_+)$, with $\psi_j \in D(H_\pm^2)$. This ensures that $\psi_j \in D(C^{-1/2})$ and $H_\pm\psi_j \in D(C^{-1/2})$. Define now $\mathbf{P}_\Omega^\pm$ on $\mathscr{H}_\pm$ by

$$\mathbf{P}_\Omega^\pm\left(\sum_{j,k} c_{jk}\langle U_\pm(\cdot,\cdot)C^{-1/2}\psi_j,\psi_k\rangle\right) = \sum_{j,k} c_{jk}\langle U_\pm(\cdot,\cdot)C^{-1/2}P_\Omega^\pm\psi_j,\psi_k\rangle. \tag{3.14}$$

Using (2.2) one finds $|\mathbf{P}_\Omega^\pm| \leqslant 1$ and $(\mathbf{P}_\Omega^\pm)^* = \mathbf{P}_\Omega^\pm$. On the other hand clearly $(\mathbf{P}_\Omega^\pm)^2 = \mathbf{P}_\Omega^\pm$, $\mathbf{P}_\mathbb{R}^\pm = 1_{\mathscr{H}_\pm}$, and $\mathbf{P}_{\Omega_1}^\pm\mathbf{P}_{\Omega_2}^\pm = \mathbf{P}_{\Omega_1\cap\Omega_2}^\pm$. This implies that the family $\{\mathbf{P}_\Omega^\pm;\ \Omega$ Borel set in $\mathbb{R}\}$ is the set of spectral projection operators for some self-adjoint operator on $\mathscr{H}_\pm$. It follows from (3.11) that this self-adjoint operator is exactly $A_\pm$. Since it is clear from (3.14) that the two projection-valued measures $P^\pm$ and $\mathbf{P}^\pm$ have the same support, $\sigma(A_\pm) = \sigma(H_\pm)$ follows immediately. □

Remark: Suppose that λ is an isolated eigenvalue of H_+ (we shall see below that H_- has only continuous spectrum) with eigenvector ϕ_λ (we assume the multiplicity of λ to be 1). Then λ is an isolated eigenvalue of A_+. It follows from the proof of Lemma 3.3 that the associated eigenspace E_λ of A_+ is given by

$$E_\lambda = \{\langle U(\cdot,\cdot)C^{-1/2}\phi_\lambda,\phi\rangle;\ \phi \in L^2(\mathbb{R}_+)\}, \tag{3.15}$$

E_λ is an infinite-dimensional closed subspace of $\mathscr{H}_+$, and every eigenvalue of A_+ is infinitely degenerate. This is completely analogous to what happens in the Weyl–Heisenberg case.[4] In order to find the spectrum of A and the associated eigenspaces we have thus only to determine the spectrum and eigenspaces for $H_\pm$. This turns out to be very easy, because $H_\pm$ are related to the exactly solvable Morse Schrödinger operator.[14]

Lemma 3.4: (1) H_- has only the continuous spectrum

$$\sigma(H_-) = [(\beta - \tfrac{1}{2})^2,\infty), \tag{3.16}$$

and (2) H_+ has the same continuous spectrum, and $\lfloor\beta + \frac{1}{2}\rfloor$ eigenvalues lying below it:

$$\sigma(H_+) = \{(\beta - \tfrac{1}{2})^2 - (\beta - n - \tfrac{1}{2})^2;\ n = 0,1,\ldots,\lfloor\beta - \tfrac{1}{2}\rfloor\} \cup [(\beta - \tfrac{1}{2})^2,\infty). \tag{3.17}$$

Note: Here we have used the notation $\lfloor x\rfloor$ for the largest integer strictly smaller than x:

$$|x| = \max\{n\in\mathbb{N}\,;\;\; n<x\}\,.$$

Proof: Again it will be convenient to consider $UH_{\pm}U^{-1}$ rather than $H_{\pm}$ itself, with U as defined by (3.12). We have [see (3.13)]

$$UH_{\pm}U^{-1} = -\frac{d^2}{ds^2} + V_{\pm}(s),$$

$$\text{with } V_{\pm}(s) = e^{2s} \mp 2\beta e^{s} + (\beta-\tfrac{1}{2})^2\,.$$

Since $\beta>0$, $V_-(s)$ is a continuous, monotonously increasing function of s, tending to $(\beta-\frac{1}{2})^2$ for $s\to-\infty$ and to ∞ for $s\to\infty$. It is clear therefore that $\sigma(H_-)\subset[(\beta-\frac{1}{2})^2,\infty)$. On the other hand wave functions with support in $[-2L,-L]$, with L very large, will "see" only the constant part $(\beta-\frac{1}{2})^2$ of the potential V_-. This means that the spectrum of H_- will at least contain

$$\sigma\Big(-\frac{d^2}{ds^2}\Big) + \Big(\beta-\frac{1}{2}\Big)^2 = \Big[\Big(\beta-\frac{1}{2}\Big)^2,\infty\Big).$$

Hence (3.16).

The operator $-d^2/ds^2 + v_+(s)$ is really the Morse operator.[14] Putting several constants equal to 1, one finds in Ref. 14 that the operator

$$-\frac{d^2}{dy^2} + D(e^{-2y} - e^{-y}) \quad \text{on } L^2(\mathbb{R}) \qquad (3.18)$$

has discrete spectrum

$$\{-[\sqrt{D} - (n+\tfrac{1}{2})]^2;\; n\in\mathbb{N},\; n<\sqrt{D}-\tfrac{1}{2}\}\,.$$

Its continuous spectrum is $[0,\infty)$. Putting $s = -y+\log\beta$, $D=\beta^2$, one finds that $-d^2/ds^2 + V_+(s) - (\beta-\frac{1}{2})^2$ reduces to (3.18). Hence

$$\begin{aligned}\sigma(H_+) &= \sigma\Big(-\frac{d^2}{ds^2} + V_+(s)\Big)\\ &= \{(\beta-\tfrac{1}{2})^2 - (\beta-\tfrac{1}{2}-n)^2;\; n\in\mathbb{N},\; n<\beta-\tfrac{1}{2}\}\\ &\quad\cup[(\beta-\tfrac{1}{2})^2,\infty)\,. \qquad\square\end{aligned}$$

Remark: Reference 14 also gives explicit formulas for the eigenvectors of $-d^2/dy^2 + D(e^{-2y}-e^{-y})$. We shall only need the ground state. This is given by

$$\phi_0(y) = [\Gamma(2\sqrt{D}-1)]^{-1/2}(2\sqrt{D}e^{-y})^{\sqrt{D}-1/2}e^{-\sqrt{D}e^{-y}}\,.$$

Substituting $y=-s-\log\beta$, and making the inverse transformation U^{-1}, we find the ground state ϕ_0 of H_+:

$$\phi_0(x) = [\Gamma(2\beta-1)]^{-1/2}2^{\beta-1/2}x^{\beta-1}e^{x}\,. \qquad (3.19)$$

If we bring together the results of Lemmas 3.2, 3.3, and 3.4 we see indeed that $A\geqslant 0$ and that 0 is an isolated eigenvalue of A. The associated eigenspace E_0 is given by [see (3.15)]

$$E_0 = \{\langle U(\cdot,\cdot)C^{-1/2}\phi_0,\;\phi\rangle;\;\phi\in L^2(\mathbb{R}_+)\}\,.$$

Here ϕ_0 is the ground state of H_+, as defined by (3.19). Note that

$$\begin{aligned}(C^{-1/2}\phi_0)(x) &= [\Gamma(2\beta-1)]^{-1/2}2^{\beta-1/2}x^{\beta-1/2}e^{-x}\\ &= \sqrt{\beta-\tfrac{1}{2}}[\Gamma(2\beta)]^{-1/2}2^{\beta}x^{\beta-1/2}e^{-x}\\ &= c_\beta^{-1/2}\psi_\beta(x)\,, \qquad (3.20)\end{aligned}$$

with c_β, ψ_β as defined by (2.9).

Hence, with the notations of Sec. II B,

$$\begin{aligned}E_0 &= \{\langle a,b;\beta\,|\phi\rangle;\;\phi\in L^2(\mathbb{R}_+)\}\\ &= U_\beta L^2(\mathbb{R}_+) = \mathscr{H}_\beta\,.\end{aligned}$$

This implies that the spectral projection operator $\mathbf{P}_0^+$ of A associated with the eigenvalue 0 is exactly P_β. Since $A\geqslant 0$, and since the eigenvalue 0 of A is isolated, we have therefore

$$\operatorname*{s-lim}_{\nu\to\infty} \exp(-\nu TA) = P_\beta \quad (T>0)\,.$$

This implies at least in a distributional sense, convergence of the corresponding integral kernels. In other words, and taking into account (2.14) and (3.7),

$$\mathscr{P}^0_\nu(a'',b'';a',b';T) \underset{\nu\to\infty}{\rightarrow} \langle a'',b'';\beta\,|a',b';\beta\rangle \quad (T>0)\,.$$

This is exactly what we set out to prove [see (3.6)].

We can do better, however, than only distributional convergence. In order to prove pointwise convergence of the $\mathscr{P}^0_\nu$, we first derive a formula relating the integral kernel of $\exp(-\nu AT)$ with $H_\pm$. This is done in the following two lemmas.

Lemma 3.5: For $t>0$, the operators $C^{-1/2}\times\exp[-tH_\pm]C^{-1/2}$ are trace class.

Lemma 3.6:

$$\begin{aligned}&[\exp(-At)](a'',b'';a',b')\\ &\quad= \sum_{\epsilon=+,-} \mathrm{Tr}[U_\epsilon(a'',b'')^{-1}U_\epsilon(a',b')C^{-1/2}\\ &\qquad\times\exp(-tH_\epsilon)C^{-1/2}]\,. \qquad (3.21)\end{aligned}$$

Proof of Lemma 3.6: We shall first derive (3.21), already assuming that $C^{-1/2}\exp[-tH_\pm]C^{-1/2}$ are trace class.

Let $\{\psi_j;\; j\in\mathbb{N}\}$ be an orthonormal base of $L^2(\mathbb{R}_+)$ such that $\psi_j\in D(C^{+1/2})\cap D(C^{-1/2})$ for all j. Define, as in (2.6),

$$f^{\pm}_{ij}(a,b) = \langle U_\pm(a,b)C^{-1/2}\psi_i,\psi_j\rangle\,.$$

The $f^{\pm}_{ij}$ constitute an orthonormal base of $L^2(M_+)$. Hence, at least in a distributional sense,

$$\begin{aligned}&[\exp(-At)](a'',b'';a',b')\\ &\quad= \sum_{i,j,\epsilon}\sum_{k,l,\epsilon'} f^{\epsilon}_{ij}(a'',b'')\langle f^{\epsilon}_{ij},\; e^{-At}f^{\epsilon'}_{kl}\rangle\,\overline{f^{\epsilon'}_{kl}(a',b')}\,.\end{aligned} \qquad (3.22)$$

It is clear from the proof of Lemma 3.3 that

$$(e^{-At}f^{\epsilon'}_{kl})(a,b) = \langle U_{\epsilon'}(a,b)C^{-1/2}e^{-H_{\epsilon'}t}\psi_k,\psi_l\rangle\,. \qquad (3.23)$$

Note that $C^{-1/2}e^{-H_{\epsilon'}t}\psi_k$ is well defined, since $\psi_k\in D(C^{1/2})$, hence $\psi_k = C^{-1/2}\phi_k$ for some ϕ_k, and since $C^{-1/2}\exp(-H_{\epsilon'}t)C^{-1/2}$ is a bounded (even trace-class) operator. From (3.23), (2.2), and the orthogonality of $\mathscr{H}_+$ and $\mathscr{H}_-$ we obtain

$$\langle f^{\epsilon}_{ij}, e^{-At}f^{\epsilon'}_{kl}\rangle = \delta_{\epsilon,\epsilon'}\delta_{jl}\langle e^{-H_\epsilon t}\psi_k,\psi_i\rangle\,.$$

Substituting this into (3.22) leads to

$$\begin{aligned}&[\exp(-At)](a'',b'';a',b')\\ &\quad= \sum_{\epsilon,i,j,k}\langle U_\epsilon(a'',b'')C^{-1/2}\psi_i,\psi_j\rangle\\ &\qquad\times\langle\psi_j,U_\epsilon(a',b')C^{-1/2}\psi_k\rangle\langle e^{-H_\epsilon t}\psi_k,\psi_i\rangle\\ &\quad= \sum_\epsilon \mathrm{Tr}[U_\epsilon(a'',b'')^{-1}U_\epsilon(a',b')C^{-1/2}e^{-H_\epsilon t}C^{-1/2}]\,.\end{aligned}$$

Since the final result of this calculation is clearly a continuous function in (a'',b''), (a',b'), we may conclude (3.21) pointwise, even though *a priori* (3.22) was true only in a distributional sense. □

We now turn to the proof of Lemma 3.5. In the course of the proof we shall not only prove that $C^{-1/2}\exp(-tH_\pm)C^{-1/2}$ are trace class, but also calculate an estimate of the trace. The method used in this estimation will be useful again in the next section, as well as the estimate itself.

Proof of Lemma 3.5: Again it is convenient to use the unitary transform (3.12). We have

$$UC^{-1/2}U^{-1} = (2\pi)^{-1/2}e^{s/2},$$

$$UH_\pm U^{-1} = -\frac{d^2}{ds^2} + V_\pm(s),$$

with $V_\pm(s) = e^{2s} \mp 2\beta e^s + (\beta - \frac{1}{2})^2$.

We thus have to study

$$e^{s/2}\exp[-T(-\Delta + V_\pm)]e^{s/2}$$

on $L^2(\mathbb{R})$. By the Feynman–Kac formula $\exp[-t(-\Delta + V_\pm)]$ and therefore also $e^{s/2}[-T(-\Delta + V_\pm)]e^{s/2}$ has a positive integral kernel. It is therefore trace class if and only if this integrable kernel is integrable, i.e., if

$$\int_{-\infty}^{\infty} ds\, e^{s/2}\{\exp[-T(-\Delta + V_\pm)]\}(s,s)e^{s/2} < \infty. \tag{3.24}$$

By the Feynman–Kac formula we have (see, e.g., Refs. 13 and 15)

$$\{\exp[-T(-\Delta + V_\pm)]\}(s,s) = \int d\rho_{W,T;s,s} \exp\left\{-\int_0^T dt\, V_\pm[\omega(t)]\right\}. \tag{3.25}$$

Here $d\rho_{W,T;s_1,s_2}$ is the familiar pinned Wiener measure. We have denoted it by ρ in order to distinguish it from our Wiener measure $d\mu_W^\nu$ on the Lobachevsky half-plane. The measure $d\rho_{W,T;s_1,s_2}$ is pinned at s_1 for $t=0$, at s_2 for $t=T$. It is a Gaussian measure with normalized connected convariance $(t_1 \leqslant t_2)$

$$\langle\omega(t_1)\omega(t_2)\rangle^c = \langle\omega(t_1)\omega(t_2)\rangle - \langle\omega(t_1)\rangle\langle\omega(t_2)\rangle = 2t_1(1 - t_2/T).$$

Substituting (3.25) into (3.24) gives

$$\int_{-\infty}^{\infty} ds\, e^s \int d\rho_{W,T;s,s} \exp\left\{-\int_0^T dt\, V_\pm[\omega(t)]\right\}$$

$$\leqslant \int_{-\infty}^{\infty} ds\, e^s \int d\rho_{W,T;s,s}\, T^{-1}\int_0^T dt \times \exp\{-TV_\pm[\omega(t)]\} \quad \text{(by Jensen's inequality)}$$

$$= \int_{-\infty}^{\infty} ds\, e^s \int d\rho_{W,T;0,0}\, T^{-1}\int_0^T dt \times \exp\{-TV_\pm[\omega(t)+s]\}$$

$$= \int d\rho_{W,T;0,0}\, T^{-1}\int_0^T dt \int_{-\infty}^{\infty} ds\, e^s \times \exp\{-TV_\pm[\omega(t)+s]\}$$

$$= \int d\rho_{W,T;0,0}\, T^{-1}\int_0^T dt \int_{-\infty}^{\infty} ds\, e^{s-\omega(t)} \times \exp[-TV_\pm(s)]$$

$$[\text{translates } s \to s + \omega(t) \text{ for every } t]. \tag{3.26}$$

(This technique, using first Jensen's inequality and then, after permuting the integrals over t and s, shifting s by $\omega(t)$, was used by Lieb[16] to derive bounds on the number of bound states for $-\Delta + V$; see also the discussion of the Lieb inequality in Ref. 15.)

One easily calculates

$$\int d\rho_{W,T;0,0}\, e^{-\omega(t)} = \frac{1}{\sqrt{\pi T}} e^{t(T-t)/T}, \tag{3.27}$$

and

$$T^{-1}\int_0^T dt\, e^{t(T-t)/T} = \int_0^1 dr\, e^{Tr(1-r)} \leqslant e^{T/4}. \tag{3.28}$$

On the other hand,

$$\int_{-\infty}^{\infty} ds\, e^s \exp[-TV_\pm(s)] = \int_0^\infty dx \exp\left\{-T\left[x^2 \mp 2\beta x + \left(\beta - \frac{1}{2}\right)^2\right]\right\} \leqslant \frac{\sqrt{\pi}}{\sqrt{T}} \exp\left[\left(\beta - \frac{1}{4}\right)T\right]. \tag{3.29}$$

Putting together (3.26), (3.27), (3.28), and (3.29) shows that condition (3.24) is fulfilled. This means that $C^{-1/2}\exp(-H_\pm T)C^{-1/2}$ is trace class, and

$$\mathrm{Tr}[C^{-1/2}\exp(-H_\pm T)C^{-1/2}] \leqslant (1/T)e^{\beta T}. \tag{3.30}$$

□

With the help of Lemmas 3.5 and 3.6, and of estimate (3.30), we can prove (3.6) pointwise.

Proposition 3.7: Let $\mathscr{P}_\nu^0$ be defined by (3.5). Then, for all $T > 0$, and for all (a'',b''), $(a',b') \in M_+$,

$$\lim_{\varphi\to\infty} \mathscr{P}_\nu^0(a'',b'';a',b';T) = \langle a'',b'';\beta|a',b';\beta\rangle.$$

Proof: By the definition of the affine coherent states in Sec. II B, and by (3.20), we have

$$\begin{aligned}\langle a'',b'';\beta|a',b';\beta\rangle &= \langle\psi_\beta|U_+(a'',b'')^{-1}U_+(a',b')\psi_\beta\rangle \\ &= c_\beta\langle C^{-1/2}\phi_0|U_+(a'',b'')^{-1}U_+(a',b')C^{-1/2}\phi_0\rangle \\ &= c_\beta\,\mathrm{Tr}[U_+(a'',b'')^{-1}U_+(a',b')C^{-1/2}P_0C^{-1/2}],\end{aligned} \tag{3.31}$$

where $P_0 = |\phi_0\rangle\langle\phi_0|$ is the zero-eigenvalue spectral projection operator of H_+.

Comparing (3.31) with Lemma 3.6, we find

$$\begin{aligned}&|c_\beta^{-1}[\mathscr{P}_\nu^0(a'',b'';a',b';T) - \langle a'',b'';\beta|a',b';\beta\rangle]| \\ &\leqslant |\mathrm{Tr}[U_-(a'',b'')^{-1}U_-(a',b')C^{-1/2}e^{-\nu H_- t}C^{-1/2}]| \\ &\quad + |\mathrm{Tr}[U_+(a'',b'')^{-1}U_+(a',b') \times C^{-1/2}(e^{-\nu H_+ T} - P_0)C^{-1/2}]| \\ &\leqslant |\mathrm{Tr}[C^{-1/2}e^{-\nu H_- T}C^{-1/2}]| \\ &\quad + |\mathrm{Tr}[C^{-1/2}(e^{-\nu H_+ T} - P_0)C^{-1/2}]|.\end{aligned} \tag{3.32}$$

The estimate (3.30) is not sufficient to conclude that this converges to 0 for $\nu\to\infty$. We can improve this estimate in the following way. For all $\lambda\in[0,1]$,

$$e^{-H_-t}\leqslant e^{H_-\lambda t}\|e^{-H_-(1-\lambda)t}\|$$
$$\leqslant e^{-H_-\lambda t}e^{-(1-\lambda)(\beta-1/2)^2t}.$$

Hence, for all $\lambda\in[0,1]$,

$$\mathrm{Tr}[C^{-1/2}e^{-H_-t}C^{-1/2}]$$
$$\leqslant e^{-(1-\lambda)(\beta-1/2)^2t}\,\mathrm{Tr}[C^{-1/2}e^{-\lambda H_-t}C^{-1/2}]$$
$$\leqslant e^{-(\beta-1/2)^2t}\frac{\exp\{[(\beta-1/2)^2+\beta]\lambda t\}}{2\lambda t}.$$

If $t\geqslant[\beta^2+1/4]^{-1}$, we can choose $\lambda=[t(\beta^2+1/4)]^{-1}\leqslant 1$, and we find

$$\mathrm{Tr}[C^{-1/2}e^{-H_-t}C^{-1/2}]\leqslant(\beta^2+\tfrac{1}{4})e^{1-(\beta-1/2)^2t}. \tag{3.33}$$

If $t\leqslant[\beta^2+\frac{1}{4}]^{-1}$, we take $\lambda=1$, and we find

$$\mathrm{Tr}[C^{-1/2}e^{-H_-t}C^{-1/2}]\leqslant(e/t)\,e^{-(\beta-1/2)^2t}. \tag{3.34}$$

Combining (3.33) and (3.34) we find that there exists a constant $¢$ such that, for all $t>0$,

$$\mathrm{Tr}[C^{-1/2}e^{-H_-t}C^{-1/2}]\leqslant ¢(1+t^{-1})e^{-(\beta-1/2)^2t}. \tag{3.35}$$

The same can be done for $e^{-H_+t}-P_0$. There, the basic inequality is

$$e^{-H_+t}-P_0\leqslant(e^{-H_+\lambda t}-P_0)\cdot\|e^{-H_+(1-\lambda)t}-P_0\|$$
$$\leqslant e^{-H_+\lambda t}\cdot\exp[-(1-\lambda)B(\beta)t],$$

with

$$B(\beta)=\begin{cases}(\beta-\frac{1}{2})^2, & \text{if } \beta<\frac{3}{2},\\ 2(\beta-1), & \text{if } \beta>\frac{3}{2}.\end{cases} \tag{3.36}$$

This distinction is due to the fact that H_+ has more than one bound state if $\beta>\frac{3}{2}$. In this case $2(\beta-1)$ is the energy difference between the ground state and the first excited state. The estimate for H_+, corresponding to the inequality (3.35) for H_-, is then

$$\mathrm{Tr}[C^{-1/2}(e^{-H_+t}-P_0)C^{-1/2}]\leqslant ¢(1+t^{-1})e^{-B(\beta)t}. \tag{3.37}$$

Substituting the estimates (3.37) and (3.35) into (3.32) leads to

$$|\mathscr{P}^0_\nu(a'',b'';a',b';T)-\langle a'',b'';\beta|a',b';\beta\rangle|$$
$$\leqslant ¢[1-(\nu T)^{-1}]\exp[-B(\beta)\nu T], \tag{3.38}$$

where $¢$ denotes a constant [not the same as in (3.37) or (3.35)] which depends on β, but not on ν or T. It is clear that $(3.38)\to 0$ for $\nu\to\infty$. This concludes our proof.

For zero Hamiltonian, we have thus achieved our aim. We have given a sense to the formal expression (3.1) by regularizing it by means of a Wiener measure with diffusion constant ν, and we have proved that we obtain the expected result for $\nu\to\infty$.

IV. THE PATH INTEGRAL FOR NONZERO HAMILTONIAN

For nonzero Hamiltonians our strategy will essentially be the same as for the zero-Hamiltonian case. We regularize (3.1) by means of a Wiener measure with diffusion constant ν, i.e., we define

$$\mathscr{P}^h_\nu(a'',b'';a',b';T)=c_\beta e^{\nu T\beta}\int\exp\Big[-i\beta\int a^{-1}\,db$$
$$-i\int h(a,b)dt\Big]\,d\mu^\nu_W(a,b). \tag{4.1}$$

Again the stochastic integral $\int a^{-1}\,db$ should be understood in the Stratonovich sense. We shall show that in the limit for ν tending to ∞, $\mathscr{P}^h_\nu$ tends to the affine coherent state matrix element $\langle a'',b'';\beta|\exp(-iTH)|a',b';\beta\rangle$, where

$$H=c_\beta^{-1}\int d\mu(a,b)|a,b;\beta\rangle h(a,b)\langle a,b;\beta|. \tag{4.2}$$

Our proof of this statement will run along the same lines as for the Weyl–Heisenberg case, in Ref. 2. We shall therefore not repeat the whole argument. We shall prove some basic estimates and show how, given these estimates, the proofs in Ref. 4 carry over to the affine path integrals studied here.

The proof, in Ref. 4 of the convergence, for $\nu\to\infty$, of the ν-dependent path integral $\mathscr{P}^h_\nu$ proceeded in essentially three steps. First it was shown that $\mathscr{P}^h_\nu$ was the integral kernel of a contraction semigroup. Then strong convergence, as $\nu\to\infty$, of these contraction operators was proved; this led to convergence of the $\mathscr{P}^h_\nu$ in a distributional sense. Finally, pointwise convergence of the $\mathscr{P}^h_\nu$ was proved. For these three steps, different conditions of a technical nature were imposed on the function h.

We shall distinguish these same three steps here. We start however with a subsection listing different conditions on h and estimates following from these conditions. These estimates will be needed in the following three subsections, outlining the proof of our main result.

A. Conditions on the function *h* and various estimates

The first estimate will ensure that $\mathscr{P}^h_\nu$ is a well-defined expression, i.e., that

$$\exp\Big\{-i\int_0^T dt\ h[a(t),b(t)]\Big\}$$

is integrable with respect to $d\mu^\nu_W$. For this it is sufficient that

$$\int d\mu^\nu_W\int_0^T dt\,|h[a(t),b(t)]|<\infty. \tag{4.3a}$$

This can be rewritten as

$$\int_0^T dt\int d\mu(a,b)K_{T-t}(a'',b'';a,b)$$
$$\times|h(a,b)|K_t(a,b;a',b')<\infty, \tag{4.3b}$$

with K_t as defined by (3.3).

The following lemma gives a sufficient condition on h for (4.3) to hold.

Lemma 4.1: Define

$$D(a,b) := d(a,b;1,0) = \cosh^{-1}\left[\frac{1+a^2+b^2}{2a}\right] \tag{4.4}$$

[see (2.12)]. If, for all $\alpha > 0$,

$$k_\alpha(h) = \int d\mu(a,b)\,|h(a,b)|^2 \exp[-\alpha D(a,b)^2] < \infty\,, \tag{4.5}$$

then, for all (a',b'), $(a'',b'')\in M_+$, and all $T>0$,

$$\int_0^T dt \int d\mu(a,b) K_{T-t}(a'',b'';a,b)\,|h(a,b)|\,K_t(a,b;a'b')$$
$$\leqslant \cent\,[k_{(16T)^{-1}}(h)]^{1/2}$$
$$\times \exp\{1/16T\,[D(a',b')^2 + D(a'',b'')^2]\}\,. \tag{4.6}$$

Note: We shall, throughout this section, denote all constants by $\cent$ without further identification. A constant $\cent$ may depend on β. Occasionally as in (4.6) the constant $\cent$ may also depend on T. In all the cases where the T dependence is important, however, we shall explicitly keep track of it.

Proof: By (3.3) we have

$$\int_0^T dt\, K_{T-t}(a'',b'';a,b)K_t(a,b;a',b')$$
$$\leqslant \cent \int_{\delta'}^\infty dx \frac{xe^{-x^2/8T}}{\sqrt{\cosh x - \cosh \delta'}}$$
$$\times \int_{\delta''}^\infty dy \frac{ye^{-y^2/8T}}{\sqrt{\cosh y - \cosh \delta''}} I(x,y)\,,$$

where $\delta' = d(a,b;a',b')$, $\delta'' = d(a,b;a'',b'')$, and with $I(x,y)$ given by

$$I(x,y) = \int_0^T dt\,[t(T-t)]^{-3/2} e^{-x^2/8t} e^{-y^2/8(T-t)}$$
$$\leqslant \left[\frac{T}{2}\right]^{-3/2} e^{-y^2/8T} \int_0^{T/2} dt\, t^{-3/2} e^{-x^2/8t}$$
$$+ \left[\frac{T}{2}\right]^{-3/2} e^{-x^2/8T} \int_0^{T/2} dt\, t^{-3/2} e^{-y^2/8t}$$
$$\leqslant \cent T^{-3/2}(x^{-1}+y^{-1}) \int_0^\infty ds\, s^{-3/2} e^{-1/8s}$$
$$\leqslant \cent T^{-3/2}(x^{-1}+y^{-1})\,.$$

On the other hand

$$\int_\delta^\infty dx \frac{e^{-\alpha x^2}}{\sqrt{\cosh x - \cosh \delta}}$$
$$\leqslant \frac{1}{\sqrt{\sinh \delta}} \int_0^\infty du \frac{e^{-\alpha(u+\delta)^2}}{\sqrt{u}}$$
$$\leqslant \cent \alpha^{-1/4} \frac{e^{-\alpha\delta^2}}{\sqrt{\sinh \delta}}$$

$$\int_\delta^\infty dx \frac{xe^{-\alpha x^2}}{\sqrt{\cosh x - \cosh \delta}}$$
$$= \int_0^\infty du \frac{(u+\delta)e^{-\alpha(u+\delta)^2}}{\sqrt{\cosh \delta(\cosh u - 1) + \sinh \delta \sinh u}}$$
$$\leqslant \frac{e^{-\alpha\delta^2}}{\sqrt{\cosh \delta}} \int_0^\infty du \frac{u}{\sqrt{\cosh u - 1}} + \frac{e^{-\alpha\delta^2}\delta}{\sqrt{\sinh \delta}} \cent \alpha^{-1/4}$$
$$\leqslant (1+\alpha^{-1/4}\delta^{1/2})e^{-\alpha\delta^2} \leqslant \cent(1+\alpha^{-1/2})e^{-\alpha\delta^2/2}\,.$$

Hence

$$\int_0^T dt\, K_{T-t}(a'',b'';a,b)K_t(a,b;a',b')$$
$$\leqslant \cent T^{-5/4}(1+T^{-1/2}) \left[\frac{1}{\sqrt{\sinh \delta'}} + \frac{1}{\sqrt{\sinh \delta''}}\right]$$
$$\times e^{-(\delta'^2+\delta''^2)/16T}\,. \tag{4.7}$$

This implies

$$\int_0^T dt \int d\mu(a,b) K_{T-t}(a'',b'';a,b)\,|h(a,b)|\,K_t(a,b;a',b')$$
$$\leqslant \cent T^{-5/4}(1+T^{-1/2})$$
$$\times \left\{\left[\int d\mu(a,b)\,|h(a,b)|^2 e^{-d(a,b;a',b')^2/8T}\right]^{1/2}\right.$$
$$\left.\times \left[\int d\mu(a,b) \frac{e^{-d(a,b;a'',b'')^2/8T}}{\sinh[d(a,b;a'',b'')]}\right]^{1/2}\right\}$$
$$+ \text{ idem with roles of } a',b' \text{ and } a'',b'' \text{ reversed}\,.$$

Since

$$D(a,b) \leqslant d(a,b;a'b') + D(a',b')$$

hence

$$-d(a,b;a',b')^2 \geqslant -\tfrac{1}{2}D(a,b)^2 + D(a',b')^2\,,$$

the first factor is finite by (4.5). We only need to prove still that, for all $\alpha > 0$,

$$\int d\mu(a,b) \frac{e^{-\alpha d(a,b;a',b')^2}}{\sinh[d(a,b;a',b')]} < \infty\,,$$

in order to conclude (4.6). Since both the measure $d\mu(a,b)$ and the metric distance d are (left) invariant, it suffices to prove, for all $\alpha > 0$,

$$\int d\mu(a,b) \frac{e^{-\alpha D(a,b)^2}}{\sinh D(a,b)} < \infty\,. \tag{4.8}$$

A careful analysis of the singularities of the integrand in (4.8), using the definition (4.4) of $D(a,b)$, shows that this integral is indeed finite.

Remark: We shall also need the following similar estimate. From (4.7) we obtain

$$\int_0^T dt \int d\mu(a,b) K_{T-t}(a'',b'';a,b)\,|h(a,b)|\,K_t(a.b;a',b')$$
$$\leqslant \cent T^{-5/4}(1+T^{-1/2}) \left(\int d\mu(a,b)\,|h(a,b)|^2\right.$$
$$\left.\times \exp\left\{\frac{1}{16T}[d(a,b;a',b')^2 + d(a,b;a'',b'')^2]\right\}\right)^{1/2}$$
$$\times \left\{\int d\mu(a,b)\,[\sinh D(a,b)]^{-1}\right.$$
$$\left.\times \exp\left[-\frac{1}{16T} D(a,b)^2\right]\right\}^{1/2}. \tag{4.9}$$

Using the triangle inequality for the metric d one finds that

$$d(a,b;a',b')^2 + d(a,b;a'',b'')^2$$
$$\geqslant \tfrac{1}{3}D(a,b)^2 + \tfrac{1}{3}D(a',b')^2 - D(a'',b'')^2\,.$$

Inserting this into (4.9) we find

$$\int_0^T dt \int d\mu(a,b) K_{T-t}(a'',b'';a,b)\,|h(a,b)|\,K_t(a,b;a',b')$$
$$\leqslant ¢\exp[(1/8T)D(a'',b'')^2 - (1/40T)D(a',b')^2]\,[k_{(80T)^{-1}}(h)]^{1/2}. \tag{4.10}$$

We shall impose conditions on the function h other than only (4.3). To formulate them, we first need the following definitions.

For $(a',b')\in M_+$, and $t>0$, we define the following functions on M_+:

$$\phi_{a',b';t}(a,b) = [\exp(-tA)](a,b;a',b'), \tag{4.11}$$
$$\phi_{a',b';\infty}(a,b) = c_\beta^{-1}\langle a,b;\beta\,|a',b';\beta\rangle = P_\beta(a,b;a',b'). \tag{4.12}$$

It is clear that

$$\overline{\phi_{a',b';t}(a,b)} = \phi_{a,b;t}(a',b'),$$
$$\overline{\phi_{a',b';\infty}(a,b)} = \phi_{a,b;\infty}(a',b').$$

Some of the calculations in Sec. III can be viewed as estimates on the L^2- and L^∞-norms of these vectors and their difference. We have

$$\|\phi_{a',b';\infty}\| = c_\beta^{-1/2} \quad [\text{by (2.10)}], \tag{4.13}$$
$$\|\phi_{a',b';t} - \phi_{a',b';\infty}\|^2 = \int d\mu(a,b)\,|[\exp(-tA) - P_\beta](a,b;a',b')|^2$$
$$= [\exp(-2tA) - P_\beta](a',b';a',b')$$
$$= \mathrm{Tr}\{C^{-1/2}[(e^{-2tH_+} - P_0) + e^{-2tH_-}]C^{-1/2}\} \quad [\text{by Lemma 3.6 and (3.31)}]$$
$$\leqslant ¢(1-t^{-1})e^{-2B(\beta)t} \quad [\text{by (3.35), (3.37)}],$$

where

$$B(\beta) = \begin{cases}(\beta-1/2)^2, & \text{if } \beta\leqslant\frac{3}{2},\\ 2(\beta-1), & \text{if } \beta\geqslant\frac{3}{2}.\end{cases} \tag{4.14}$$

Hence

$$\|\phi_{a',b';t} - \phi_{a',b';\infty}\| \leqslant ¢(1-t^{-1/2})e^{-B(\beta)t}, \tag{4.15}$$
$$\|\phi_{a',b';t}\| \leqslant c_\beta^{-1/2} + ¢(1+t^{-1/2})e^{-B(\beta)t}. \tag{4.16}$$

On the other hand, the estimate (3.38) can be rewritten as

$$\|\phi^{a',b';t} - \phi_{a',b';\infty}\|_\infty = \sup_{a,b\in M_+} |(\phi_{a',b';t} - \phi_{a',b';\infty})(a,b)| \leqslant ¢(1+t^{-1})\exp[-B(\beta)t]. \tag{4.17}$$

In the following three sections we shall consider the multiplication operator h on $L^2(M_+)$ defined by

$$(hf)(a,b) = h(a,b)f(a,b).$$

We shall restrict ourselves to real functions h. Then the multiplication operator is self-adjoint, with domain

$$D(h) = \{f\in L^2(M_+);\ hf\in L^2(M_+)\}.$$

In the remainder of this subsection we shall determine sufficient conditions on h ensuring that $\phi_{a',b';\infty}$ and $\phi_{a',b';t}$ are elements of $D(h)$, i.e.,

$$\int d\mu(a,b)\,|h(a,b)|^2|\phi_{a',b';\infty}(a,b)|^2 < \infty, \tag{4.18}$$
$$\int d\mu(a,b)\,|h(a,b)|^2|\phi_{a',b';t}(a,b)|^2 < \infty. \tag{4.19}$$

We shall also estimate $\|h(\phi_{a',b';t} - \phi_{a',b';\infty})\|$. We start with (4.18), the easiest one.

Lemma 4.2: If

$$\int d\mu(a,b)\,|h(a,b)|^2\left[\frac{2a}{1+a^2+b^2}\right]^{2\beta} < \infty, \tag{4.20}$$

then (4.18) is satisfied, and

$$\|h\phi_{a',b';\infty}\| \leqslant ¢\left[\frac{1+a'^2+b'^2}{2a'}\right]^\beta. \tag{4.21}$$

Proof: By (2.11), we find

$$\|h\phi_{a',b';\infty}\|^2 = \int d\mu(a,b)\,|h(a,b)|^2 c_\beta^{-2}|\langle a,b;\beta\,|a',b';\beta\rangle|^2$$
$$= c_\beta^{-2}\int d\mu(a,b)\,|h(a,b)|^2 \times\left[\frac{1+\cosh d(a,b;a',b')}{2}\right]^{-2\beta}$$
$$\leqslant ¢\int d\mu(a,b)\,|h(a,b)|^2\left[\frac{1+\cosh D(a,b)}{1+\cosh D(a',b')}\right]^{-2\beta}$$
$$[\text{use } D(a,b)\leqslant d(a,b;a',b') + D(a',b')]$$
$$\leqslant ¢\left(\frac{1+a'^2+b'^2}{2a'}\right)^{2\beta}\int d\mu(a,b)\,|h(a,b)|^2 \times\left[\frac{2a}{1+a^2+b^2}\right]^{2\beta} < \infty. \qquad \square$$

The other two estimates involve some additional calculation. We start by estimating weighted L^p-norms of $\phi_{a',b';t}$.

Lemma 4.3: For $\lambda>1$, $\mu>0$, one has

$$I_{\lambda,\mu}(t) \equiv \int d\mu(a,b)\,|[\exp(-tA)](a,b;1,0)|^\lambda \times\left[\frac{1+a^2+b^2}{2a}\right]^\mu \leqslant ¢\,(1/\sqrt{\lambda-1})t^{1-\lambda}\exp[\epsilon(\lambda;\mu)t], \tag{4.22}$$

with

$$\epsilon(\lambda,\mu) = \lambda\left(\beta-\frac{1}{4}\right) + \max\left[\frac{(1-\lambda)^2}{4} + \frac{(\mu+1/2)^2}{\lambda-1},\ M\left(\mu+\frac{1-\lambda}{2}\right) + \frac{1}{4(\lambda-1)}\right],$$

where, for all $\alpha\in\mathbb{R}$, $M(\alpha) = \max(\alpha,\alpha^2)$.

Proof: We first estimate $|[\exp(-tA)](a,b;1,0)|$, using the same technique as in the proof of Lemma 3.5. By Lemma 3.6

$$\overline{[\exp(-tA)](a,b;1,0)} = [\exp(-tA)](1,0;a,b) = \sum_\epsilon \mathrm{Tr}[U_\epsilon(a,b)C^{-1/2}e^{-H_\epsilon}C^{-1/2}].$$

Using again the unitary transform (3.12) we can rewrite this as (using the Feynman–Kac formula)

$$
\begin{aligned}
&[\exp(-TA)](1,0;a,b)\\
&\quad=\sum_\epsilon\int_{-\infty}^{\infty}ds\,e^{i\epsilon be^s}\frac{1}{2\pi}e^{s+(1/2)\ln a}\\
&\qquad\times\exp\Big[-T\Big(-\frac{d^2}{ds^2}+V_\epsilon\Big)\Big](s+\ln a,s)\\
&\quad=\not{c}\sum_\epsilon\int_{-\infty}^{\infty}ds\,e^{i\epsilon be^s}e^{s+(1/2)\ln a}\int d\rho_{W,T;0,\ln a}\\
&\qquad\times\exp\Big\{-\int_0^T dt\,V_\epsilon[s+\omega(t)]\Big\}\\
&\quad=\sum_\epsilon\int_0^\infty dx\,e^{i\epsilon bx}\sqrt{a}\int d\rho_{W,T;0,\ln a}\\
&\qquad\times\exp\Big\{-\int_0^T dt\,V_\epsilon[\omega(t)+\ln x]\Big\}\\
&\quad=\not{c}\int_{-\infty}^{\infty}dx\,e^{ibx}\sqrt{a}\int d\rho_{W,T;0,\ln a}\\
&\qquad\times\exp\Big\{-\int_0^T dt\,V_{\epsilon(x)}[\omega(t)+\ln|x|]\Big\},
\end{aligned}
\tag{4.23}
$$

where $\epsilon(x)=x/|x|$ for $x\neq 0$. Since

$$
\begin{aligned}
&V_{\epsilon(x)}[\omega(t)+\ln|x|]\\
&\quad=x^2e^{2\omega(t)}-2\beta\epsilon(x)|x|e^{\omega(t)}+(\beta-\tfrac12)^2\\
&\quad=x^2e^{2\omega(t)}-2\beta xe^{\omega(t)}+(\beta-\tfrac12)^2,
\end{aligned}
$$

we have

$$
\begin{aligned}
&[\exp(-TA)](1,0;a,b)\\
&\quad=\not{c}\sqrt{a}\,e^{(\beta-1/4)T}\int d\rho_{W,T;0,\ln a}\\
&\qquad\times\int_{-\infty}^{\infty}dx\,e^{ibx}\exp[-(a_0x^2-2a_1x+a_2)],
\end{aligned}
$$

where

$$
a_0=\int_0^T dt\,e^{2\omega(t)},\quad a_1=\beta\int_0^T dt\,e^{\omega(t)},
$$
$$
a_2=\beta^2T.
$$

Hence

$$
\begin{aligned}
&[\exp(-TA)](1,0;a,b)\\
&\quad=\not{c}\sqrt{a}\,e^{(\beta-1/4)T}\int d\rho_{W,T;0,\ln a}\,a_0^{-1/2}e^{iba_1/a_0}\\
&\qquad\times e^{-b^2/4a_0}e^{-a_2+a_1^2/a_0}.
\end{aligned}
\tag{4.24}
$$

From the Cauchy–Schwarz inequality

$$
a_1^2=\beta^2T^2\Big[\int_0^T\frac{dt}{T}e^{\omega(t)}\Big]^2\leqslant\beta^2T^2\int_0^T\frac{dT}{T}e^{2\omega(t)},
$$

hence $-a_2+a_1^2/a_0\leqslant 0$. Hence

$$
\begin{aligned}
&|[\exp(-TA)](a,b;1,0)|\\
&\quad\leqslant\not{c}\sqrt{a}\,e^{(\beta-1/4)T}\int d\rho_{W,T;0,\ln a}[a_0(\omega)]^{-1/2}\\
&\qquad\times\exp\Big[-\frac{b^2}{4a_0(\omega)}\Big],
\end{aligned}
$$

with

$$
a_0(\omega)=\int_0^T dt\,e^{2\omega(t)}.
$$

Since $\lambda>1$, we find (use either Jensen's or Young's inequality)

$$
\begin{aligned}
&\Big[\int d\rho_{W,T;0,\ln a}[a_0(\omega)]^{-1/2}e^{-b^2/4a_0(\omega)}\Big]^\lambda\\
&\quad\leqslant\not{c}\,T^{-(\lambda-1)/2}\exp\Big[-\frac{\lambda-1}{4T}(\ln a)^2\Big]\\
&\qquad\times\int d\rho_{W,T;0,\ln a}[a_0(\omega)]^{-\lambda/2}e^{-\lambda b^2/4a_0(\omega)}.
\end{aligned}
$$

Hence [see (4.22)],

$$
\begin{aligned}
I_{\lambda,\mu}(T)&\leqslant\not{c}\,e^{\lambda T(\beta-1/4)}T^{-(\lambda-1)/2}\int_0^\infty da\int_{-\infty}^{\infty}db\,a^{\lambda/2-2-\mu}\\
&\qquad\times(1+a^{2\mu}+b^{2\mu})\exp\Big[-\frac{\lambda-1}{4T}(\ln a)^2\Big]\int d\rho_{W,T;0,\ln a}[a_0(\omega)]^{-\lambda/2}e^{-\lambda b^2/4a_0(\omega)}\\
&\leqslant\not{c}\,e^{\lambda T(\beta-1/4)}T^{-(\lambda-1)/2}\int_0^\infty da\int d\rho_{W,T;0,\ln a}\,a^{\lambda/2-2-\mu}\\
&\qquad\times\exp\Big[-\frac{\lambda-1}{4T}(\ln a)^2\Big]\{(1+a^{2\mu})[a_0(\omega)]^{(1-\lambda)/2}+[a_0(\omega)]^{\mu+(1-\lambda)/2}\}.
\end{aligned}
\tag{4.25}
$$

Let us estimate

$$
J_{\delta,T}(x)\equiv\int d\rho_{W,T;0,x}\Big[\int_0^T dt\,e^{2\omega(t)}\Big]^\delta.
$$

If either $\delta\leqslant 0$ or $\delta>1$ we can apply Jensen's inequality and obtain

$$
\begin{aligned}
J_{\delta,T}(x)&\leqslant\int d\rho_{W,T;0,x}T^{\delta-1}\int_0^T dt\,e^{2\delta\omega(t)}\\
&\leqslant\not{c}\,T^{\delta-1}\int_0^T dt\frac{1}{\sqrt{t(T-t)}}\int_{-\infty}^{\infty}dy\exp\Big[-\frac{y^2}{4t}-\frac{(y-x)^2}{4(T-t)}+2\delta y\Big]\\
&\leqslant\not{c}\,T^{\delta-1}\int dy\exp\Big[-\frac{(y-x/2)^2}{T}+2\delta y\Big]\int_0^T dt\,[t(T-t)]^{-1/2}\leqslant\not{c}\,T^{\delta-1/2}e^{T\delta^2}e^{\delta x}.
\end{aligned}
\tag{4.26}
$$

If $0\leqslant\delta\leqslant 1$, then, by Young's inequality,

$$J_{\delta,T}(x)\leqslant\left[\int d\rho_{W,T;0,x}\int_0^T dt\, e^{2\omega(t)}\right]^{\delta}\left[\int d\rho_{W,T;0,x}\right]^{1-\delta}\leqslant ¢\, T^{\delta-1/2}e^{\delta T}e^{\delta x}. \tag{4.27}$$

Combining (4.26) with (4.27) we obtain, with $M(\delta)=\max(\delta,\delta^2)$,

$$J_{\delta,T}(x)\leqslant ¢\, T^{\delta-1/2}\, e^{\delta x}\, e^{M(\delta)T}.$$

Substituting this into (4.25) we find

$$I_{\lambda,\mu}(T)\leqslant ¢\, e^{\lambda T(\beta-1/4)}\, T^{-\lambda+1/2}\int_{-\infty}^{\infty}dx\,\exp\left[-\frac{\lambda-1}{4T}x^2+\frac{x}{2}\right]\left[e^{T(1-\lambda)^2/4}e^{-\mu x}(1+e^{2\mu x})+T^{\mu}e^{TM(\mu+(1-\lambda)/2)}\right]$$

$$\leqslant ¢\, T^{1-\lambda}e^{\lambda T(\beta-1/4)}\frac{1}{\sqrt{\lambda-1}}\left[\exp\left\{T\left[\frac{(1-\lambda)^2}{4}+\frac{(\mu+1/2)^2}{\lambda-1}\right]\right\}\right.$$

$$\left.+T^{\mu}\exp\left\{T\left[M\left(\mu+\frac{1-\lambda}{2}\right)+\frac{1}{4(\lambda-1)}\right]\right\}\right].$$

It is easy to see that this leads to (4.22). □

With the help of Lemma 4.3 we can now estimate

$$\|h(\phi_{a',b';t}-\phi_{a',b';\infty})\|.$$

Lemma 4.4: Let h be a function satisfying

$$C_{\mu,r}(h)\equiv\int d\mu(a,b)\,|h(a,b)|^{2+r}\left[\frac{2a}{1+a^2+b^2}\right]^{\mu}<\infty, \tag{4.28}$$

for positive parametes r,μ satisfying the following conditions:

$$\mu<r(\beta-\tfrac{1}{2})+2\beta, \tag{4.29a}$$

$$\sup_{\alpha\in(m,1)}\left[2(1-\alpha)B(\beta)-\frac{r}{r+2}\,\epsilon\left(\alpha\frac{2(\pi+2)}{r},\frac{2\mu}{r}\right)\right]>0, \tag{4.29b}$$

where

$$m=\frac{r}{2(\pi+2)}\max\left(1,\frac{1+2\mu/r}{\beta}\right).$$

Here

$$\epsilon(\lambda,\gamma)=\lambda\left(\beta-\frac{1}{4}\right)+\max\left[\frac{(1-\lambda)^2}{4}+\frac{(\gamma+1/2)^2}{\lambda-1},\ M\left(\gamma+\frac{1-\lambda}{2}\right)+\frac{1}{4(\lambda-1)}\right],$$

with $M(\delta)=\max(\delta,\delta^2)$, and $B(\beta)=(\beta-\frac{1}{2})^2$ if $\beta\leqslant\frac{3}{2}$, $B(\beta)=2(\beta-1)$ if $\beta\geqslant\frac{3}{2}$.

Then there exist constants $¢_1$, $¢_2>0$ such that

$$\|h(\phi_{a',b';t}-\phi_{a',b';\infty})\|\leqslant ¢_1[1+t^{-1+r/[2(r+2)]}]e^{-¢_2 t}\left(\frac{1+a'^2+b'^2}{2a'}\right)^{\mu/(r+2)}. \tag{4.30}$$

Proof:

$$\|h(\phi_{a',b';t}-\phi_{a',b';\infty})\|^2\leqslant\left[\int d\mu(a,b)\,|h(a,b)|^{2+r}\left(\frac{2a}{1+a^2+b^2}\right)^{\mu}\right]^{2/(r+2)}$$

$$\times\left[\int d\mu(a,b)\,|[\phi_{a',b';t}-\phi_{a',b';\infty}](a,b)|^{2(r+2)/r}\left(\frac{1+a^2+b^2}{2a}\right)^{2\mu/r}\right]^{r/(r+2)}$$

$$\leqslant[C_{\mu,r}(h)]^{2/(r+2)}\sup_{a,b}|[\phi_{a',b';t}-\phi_{a',b';\infty}](a,b)|^{2(1-\alpha)}$$

$$\times\left\{\int d\mu(a,b)\,|[\phi_{a',b';t}-\phi_{a',b';\infty}](a,b)|^{2\alpha(r+2)/r}[\cosh D(a',b')\cosh d(a,b;a',b')]^{2\mu/r}\right\}^{r/(r+2)}$$

$$\leqslant ¢[1+t^{-2(1-\alpha)}]\,e^{-2(1-\alpha)B(\beta)t}\left(\frac{1+a'^2+b'^2}{2a'}\right)^{2\mu/(r+2)}$$

$$\times\left\{\int d\mu(a,b)\left[|\phi_{1,0;t}(a,b)|^{2\alpha(r+2)/r}+|\phi_{1,0;\infty}(a,b)|^{2\alpha(r+2)/r}\right]\left(\frac{1+a^2+b^2}{2a}\right)^{2\mu/r}\right\}^{r/(r+2)}$$

[use (4.17) and the left invariance of the measure $d\mu$]. This holds for all $\alpha\in[0,1]$. If we choose α such that $\alpha>m$, with m as defined above, then $2\alpha\beta(r+2)/r-2\mu/r>1$, hence

$$\int d\mu |\phi_{1,0;\infty}(a,b)|^{2\alpha(r+2)/r}\left(\frac{1+a^2+b^2}{2a}\right)^{2\mu/r}$$
$$= \not c \int d\mu \left(\frac{2a}{1+a^2+b^2}\right)^{2\alpha\beta(r+2)/r - 2\mu/r} < \infty .$$

On the other hand $\alpha > m$ also implies $2\alpha(r+2)/r > 1$. Using Lemma 4.3 leads then to

$$\|h(\phi_{a',b';t} - \phi_{a',b';\infty})\|$$
$$\leqslant \not c [1+t^{-(1-\alpha)}][1+t^{r/[2(r+2)]-\alpha}]$$
$$\times e^{-2(1-\alpha)B(\beta)t}\left(\frac{1+a'^2+b'^2}{2a'}\right)^{2\mu/(r+2)}$$
$$\times [1+I_{2\alpha(r+2)/r,2\mu/r}]^{r/(r+2)}$$
$$\leqslant \not c [1+t^{-1+r/[2(r+2)]}]\left(\frac{1+a'^2+b'^2}{2a'}\right)^{2\mu/(r+2)}$$
$$\times \exp\left[-(1-\alpha)B(\beta)t + \frac{r}{2(r+2)}t\epsilon \times\left(\frac{2\alpha(r+2)}{r},\frac{2\mu}{r}\right)\right].$$

This holds for all $\alpha\in(m,1]$. It is clear from this that (4.30) follows if the conditions (4.29) are satisfied. ■

Remark: The conditions (4.29) are sufficient conditions on the pair (r,μ), given β, ensuring that (4.30) holds. The conditions (4.29) are however rather complicated, and may not be easy to check. It is possible, of course, to only consider one value for α, instead of the whole interval $(m,1)$. This considerably simplifies the condition on r,μ, but may be too restrictive. One possibility of choosing such a fixed value for α is, e.g., $\alpha = r/(r+2)$. It is then sufficient that

$$\mu < r(\beta - \tfrac{1}{2}),$$
$$r < \frac{4B(\beta)}{\epsilon(2,2\beta-1)},$$

to ensure that the conditions (4.29) are satisfied. This allows only a finite range for the parameter r, however, and is thus very restrictive. It turns out that it is easier to proceed in the inverse direction, i.e., to start from the pair (r,μ) and to determine for which values of β the conditions (4.29) are satisfied. One finds that the following conditions imply (4.29):

$$\beta \geqslant r(1+2\mu/r)/2(r+2)],$$
$$\beta > \tfrac{3}{2}, \tag{4.31a}$$
$$\beta > \frac{1}{2(2-3\alpha)}\left[4 - \tfrac{9}{2}\alpha + \frac{r}{r+2}\tilde{\epsilon}\left(2\alpha\frac{r+2}{r},\frac{2\mu}{r}\right)\right],$$

for some α satisfying

$$\frac{r}{2(r+2)} < \alpha < \tfrac{2}{3} \text{ and } \alpha < \frac{r(1+2\mu/r)}{2\beta(r+2)}. \tag{4.31b}$$

Here $\tilde{\epsilon}$ is defined by

$$\tilde{\epsilon}(\lambda,\gamma) = \epsilon(\lambda,\gamma) - \lambda(\beta - \tfrac{1}{4})$$
$$= \max\left[\frac{(1-\lambda)^2}{4} + \frac{(\gamma+1/2)^2}{\lambda-1}, M\left(\gamma + \frac{1-\lambda}{2}\right) + \frac{1}{4(\lambda-1)}\right], \tag{4.31c}$$

with $M(x) = \max(x,x^2)$.

Note that the second condition on α in (4.31b) is an implicit condition, since it contains β again, and β is bounded below by a function depending on α. In the explicit examples below (see Remark 2 at the end of Sec. IV) we shall first disregard this extra condition on α, compute a lower bound on β, and then verify that the condition is satisfied.

Our last estimate involves $\|h\phi_{a',b';t}\|$. From Lemma 4.2 and 4.4 one immediately has

$$|h\phi_{a',b';t}| < \not c\,[1+t^{-1+r/2(r+2)}]\left(\frac{1+a'^2+b'^2}{2a'}\right)^{2\mu/r} \tag{4.32}$$

if h satisfies (4.28), where μ, r, β fulfill either the conditions (4.29) or the conditions (4.31).

All in all we have three different technical conditions on h. The first one, (4.5), ensures that $\mathscr{P}_\nu^h$ is well defined. The second one, (4.20), ensures that $\phi_{a',b';\infty}\in D(h)$ for all $(a',b')\in M_+$. The third one, (4.28), ensures that $\phi_{a',b';t}\in D(h)$ for all $(a',b')\in M_+$, and all $t>0$. Note that $(4.28)\rightarrow(4.20)\rightarrow(4.5)$.

In what follows we shall always assume that (4.28) is satisfied.

B. The path as integral kernel of a contraction semigroup

Since h satisfied condition (4.28), hence condition (4.5), we know by Lemma 4.1 that $\mathscr{P}_\nu^h$ is well-defined. Copying the argument in Ref. 4 the following proposition can be proved.

Proposition 4.5: Let h be a real function satisfying condition (4.28). Then there exists a strongly continuous semigroup of contractions $E(\nu,h;t)$ on $L^2(M_+;d\mu)$ such that

$$[E(\nu,h;t)](a'',b'';a',b') = c_\beta^{-1}\,\mathscr{P}_\nu^h(a'',b'';a',b';t). \tag{4.33}$$

These contraction operators are related to $\exp(-\nu AT)$ by the integral equation

$$\langle f_2,E(\nu,h;T)f_1\rangle = \langle f_2,e^{-\nu AT}f_1\rangle - i\int_0^T dt \times\langle f_2,E(\nu,h;T-t)he^{-\nu AT}f_1\rangle. \tag{4.34}$$

This integral equation holds if $f_1, f_2\in C_0^\infty(M_+)$ or if $f_1\in D$ and $f_2\in C_0^\infty(M_+)\cup D$. Here D is the finite linear span of the vectors $\phi_{a,b;\infty}$ defined by (4.12).

Proof: This proposition is completely analogous to proposition 2.1 in Ref. 24, and the proof runs along exactly the same lines. We shall therefore only outline the main arguments, and fill in the technical details only where the present situation is different from that in Ref. 24.

Equation (4.33) is proved in three steps: for $h\in C_0^\infty(M_+)$, for $h\in L^\infty(M_+)$, and finally for all h satisfying (4.28).

For $h\in C_0^\infty(M_+)$ one uses the Trotter product formula to show that

$$\mathscr{P}_\nu^h(a'',b'',a',b';T)$$
$$= c_\beta\{\exp[-(\nu A+ih)T]\}(a'',b'';a',b'). \tag{4.35}$$

Since h is bounded, the operator $\nu A + ih$ is well defined, and generates a semigroup. Since $A\geqslant 0$, and h is a real function, this is a semigroup of contractions.

Using the dominated convergence theorem for $\mathscr{P}^h_\nu$, and strong resolvent convergence for $\exp[-(\nu A + ih)T]$, one can extend (4.35) to all bounded functions h.

In a next step one uses again dominated convergence arguments to show that, for all functions h satisfying (4.5), there exists a strongly continuous semigroup of contractions $E(\nu, h; t)$ satisfying (4.33). These operators are constructed as $\text{s-lim}_{n\to\infty} \exp[-(\nu A + ih_n)t]$, where $h_n(a,b) = h(a,b)$ if $|h(a,b)| < n$, $h_n(a,b) = n \operatorname{sgn} h(a,b)$ otherwise. (See Ref. 4; the arguments given there carry over without problems.)

To prove (4.31), we use the fact that the integral kernel of $E(\nu, h; t)$ is given by a path integral, i.e., (4.33). We have, for all (a', b'), $(a'', b'')\in M_+$ for all $T>0$ (see Ref. 4),

$$\begin{aligned}\mathscr{P}^h_\nu&(a'',b'';a',b';T)\\ &= \mathscr{P}^0_\nu(a'',b'';a',b';T) - i c_\beta^{-1}\int_0^T dt \int d\mu(a,b)\\ &\quad\times \mathscr{P}^h_\nu(a'',b'';a,b;T-t)h(a,b)\mathscr{P}^0_\nu(a,b;a',b';t)\,. \end{aligned}\tag{4.36}$$

Take now f_1, $f_2\in C_0^\infty(M_+)$. We multiply (4.35) by $\overline{f_2(a'',b'')}\ f_1(a',b')$ and integrate over $d\mu(a',b')\times d\mu(a'',b'')$. Using the upper bound (valid for all h [this follows from (4.1)])

$$|\mathscr{P}^h_\nu(a'',b'';a',b';t)|\leqslant c_\beta e^{\nu t\beta} K_{\nu t}(a'',b'';a',b')\,,\tag{4.37}$$

and the estimate (4.6), one sees that the resulting integral converges absolutely. This allows us to change the order of the integrations, and leads to (4.34), for all $f_1, f_2\in C_0^\infty(M_+)$.

We can extend this to the case where $f_1\in D$. To do this, we use (4.10). Take $f_1\in D, f_2\in C_0^\infty(M_+)$. Again we multiply (4.36) by $f_2(a'',b'')f_1(a',b')$ and integrate over $d\mu(a',b')\times d\mu(a'',b'')$. Since the resulting integral is absolutely convergent by (4.37) and (4.10), we may again reverse the order of the integrations. We thus obtain

$$\begin{aligned}\langle f_2,&E(\nu,h;T)f_1\rangle\\ &= \langle f_2, e^{-\nu A T} f_1\rangle - i c_\beta^{-2}\int_0^T dt\int d\mu(a'',b'')f_2(a'',b'')\\ &\quad\times\int d\mu(a,b)\mathscr{P}^h_\nu(a'',b'';a,b;T-t)h(a,b)\\ &\quad\times\int d\mu(a',b')\mathscr{P}^0_\nu(a,b;a',b';t)f_1(a',b').\end{aligned}\tag{4.38}$$

We know however that

$$f_1=\sum_{j=1}^N c_j\phi_{a_j,b_j;\infty}\in\mathscr{H}_\beta\,,$$

hence $e^{-\nu A t}f_1 = f_1$ for all t. This means in particular that $e^{-\nu A t}f_1\in D(h)$ for all t, so that we may rewrite (4.38) in the form (4.34).

Once (4.34) is obtained for $f_1\in D, f_2\in C_0^\infty(M_+)$, one uses a straightforward approximation argument, using again that $e^{-\nu A t}f_1 = f_1$, together with the fact that $C_0^\infty(M_+)$ is dense, to conclude (4.34) for $f_1, f_2\in D$.

Remark: By exactly the same arguments one can also prove that for all $f_1, f_2\in C_0^\infty(M_+)$

$$\begin{aligned}\langle f_2,&E(\nu,h;T)(1-P_\beta)f_1\rangle\\ &=\langle f_2, e^{-\nu A T}(1-P_\beta)f_1\rangle\\ &\quad + i\int_0^T dt\,\langle f_2, E(\nu,h;T-t)he^{-\nu A t}(1-P_\beta)f_1\rangle\,.\end{aligned}\tag{4.39}$$

C. Operator convergence of $E(\nu,h;T)$ for $\nu\to\infty$

The proof of the strong operator convergence of $E(\nu,h;T)$ hinges on Eq. (4.34). Again the proof in Ref. 4 can essentially be taken over, without major problems. The only difference is that we have to be a little more careful, because the operator A had a purely discrete spectrum in the Weyl–Heisenberg case, and we could therefore conveniently use an orthonormal basis consisting of eigenvectors of A. This is not possible here. We shall therefore, in our proof of Proposition 4.6 below (the analog of Proposition 2.2 in Ref. 4) pay attention only to those technical details where our argument differs from that in Ref. 4.

Proposition 4.6: Let h be a real function on M_+ satisfying (4.28). Define the operator $P_\beta h P_\beta$ on the domain $\{f; P_\beta f\in D(h)\}$. Clearly D, the finite linear span of the $\phi_{a,b;\infty}$, satisfies $D\subset D(P_\beta h P_\beta)$. If $P_\beta h P_\beta$ is essentially self-adjoint on $D\oplus\mathscr{H}_\beta^\perp$, then, for all $T>0$,

$$\underset{\nu\to\infty}{\text{s-lim}}\, E(\nu,h;T) = P_\beta \exp[-iP_\beta h P_\beta T]P_\beta\,.\tag{4.40}$$

Proof: To prove (4.40), the operator $E(\nu,h;T)$ is split into three parts,

$$\begin{aligned}E(\nu,h;T) &= E(\nu,h;T)(1-P_\beta) + P_\beta E(\nu,h;T)P_\beta\\ &\quad + (1-P_\beta)E(\nu,h;T)P_\beta\,.\end{aligned}$$

The treatment of the last two terms is completely analogous to the proof of Proposition 2.2 in Ref. 4. We shall therefore restrict ourselves here to a discussion of the first term and an estimate related to it.

From (4.39) we obtain, for all $f_1, f_2\in C_0^\infty(M_+)$,

$$|\langle f_2, E(\nu,h;T)(1-P_\beta)f_1\rangle|\leqslant\|f_2\|\cdot\|e^{-\nu A T}(1-P_\beta)\|\cdot\|f_1\| + \|f_2\|\cdot\int_0^T dt\,\|he^{-\nu A t}(1-P_\beta)f_1\|\,.\tag{4.41}$$

We have $\|e^{-\nu A T}(1-P_\beta)\|\leqslant e^{-\nu T B(\beta)}$, with $B(\beta)$ as defined by (4.14), and

$$\begin{aligned}\|he^{-\nu A t}(1-P_\beta)f_1\|^2 &= \|h(e^{-\nu A t}-P_\beta)f_1\|^2\leqslant\int d\mu(a',b')\int d\mu(a'',b'')|f_1((a',b')|\ |f_1(a'',b'')|\\ &\quad\times\left[\int d\mu(a,b)|h(a,b)|^2|(\phi_{a',b';\nu t}-\phi_{a',b';\infty})(a,b)|^2\right]^{1/2}\\ &\quad\times\left[\int d\mu(a,b)|h(a,b)|^2|\phi_{a'',b'';\nu t}-\phi_{a'',b'';\infty}(a,b)|^2\right]^{1/2}\\ &< \cancel{c}[1+(\nu t)^{-1+r/[2(r+2)]}]^2 e^{-2k\nu t}\left[\int d\mu(a,b)|f_1(a,b)|\left[\frac{1+a^2+b^2}{2a}\right]^\beta\right]^2,\end{aligned}$$

by Lemma 4.4. Substituting this into (4.41) leads to

$$\|E(\nu,h;T)(1-P_\beta)f_1\| \leqslant e^{-\nu TB(\beta)}\|f_1\| + \frac{k}{\nu}\left\| f_1\cdot\left[\frac{1+a^2+b^2}{2a}\right]^\beta\right\|_1 \times\int_0^\infty dt\,[1+t^{-1+r/2(r+2)}]e^{-kt}. \tag{4.42}$$

This holds for all $f_1 \in C_0^\infty(M_+)$. Since C_0^∞ is a dense subspace of $L^2(M_+;d\mu)$ and since the operators $E(\nu,h;T)$ are contractions, this implies, for all $T>0$,

$$\operatorname*{s-lim}_{\nu\to\infty} E(\nu,h;T)(\mathbb{1}-P_\beta)=0.$$

From (4.42) we can clearly also conclude that

$$\lim_{\nu\to\infty}\int_0^T dt\,|\langle f_2,E(\nu,h;T-t)(\mathbb{1}-P_\beta)f_1\rangle|=0,$$

for all $f_1,f_2 \in C_0^\infty(M_+)$, and hence (by the same density arguments as above) for all $f_1,f_2 \in L^2(M_+;d\mu)$. This estimate is needed in the discussion of $P_\beta E(\nu,h;T)P_\beta$ (see Ref. 4).

As already mentioned above, the remainder of the proof is a transcription of the proof of Proposition 2.2 in Ref. 4. □

Our ultimate goal is to link $\mathscr{P}_\nu^h$, at least in the limit for $\nu\to\infty$, to the unitary group $\exp(-iTH)$ generated by a Hamiltonian H on $L^2(\mathbb{R}_+)$. This is in fact achieved by Proposition 4.6. To see this, write the integral kernel of $P_\beta hP_\beta$,

$$(P_\beta hP_\beta)(a'',b'';a',b') = c_\beta^{-2}\int d\mu(a,b) \langle a'',b'';\beta|a,b;\beta\rangle h(a,b)\langle a,b;\beta|a',b';\beta\rangle.$$

One easily checks from (2.13) that this is exactly the integral kernel of $U_\beta HU_\beta^*$, with

$$H=c_\beta^{-1}\int d\mu(a,b)|a,b;\beta\rangle h(a,b)\langle a,b;\beta|.$$

Thus $P_\beta hP_\beta = U_\beta HU_\beta^*$. The condition that $P_\beta hP_\beta$ be essentially self-adjoint on $D\oplus\mathscr{H}_\beta$ is exactly equivalent to the condition that H be essentially self-adjoint on D_c, the finite linear span of the (affine) coherent states $|a,b;\beta\rangle$.

The conclusion (4.40) can now be rewritten in terms of H. One finds (see also Ref. 4)

$$[P_\beta\exp(-iP_\beta hP_\beta T)P_\beta](a'',b'';a',b') = c_\beta^{-1}\langle a''b'';\beta|\exp(-iHT)|a',b';\beta\rangle.$$

The strong convergence (4.40) implies, in particular, convergence of the corresponding integral kernels, in a distributional sense (i.e., when evaluated on test functions). We have therefore, at least in a distributional sense,

$$\lim_{\nu\to\infty}\mathscr{P}_\nu^h(a'',b'';a',b';T) = \langle a'',b'';\beta|e^{-iHT}|a'b';\beta\rangle. \tag{4.43}$$

This result will be sharpened to pointwise convergence in the next subsection.

D. Pointwise convergence of $\mathscr{P}_\nu^h$ for $\nu\to\infty$

To prove (4.43) for all points (a'',b''), $(a',b')\in M_+$, rather than in a distributional sense, we again use an integral equation relating $\mathscr{P}_\nu^h$ and $\mathscr{P}_\nu^0$, obtained by combining (4.36) with the complex conjugate version of (4.36) for $-h$.

$$\begin{aligned}\mathscr{P}_\nu^h(a'',b'';a',b';T) &= \mathscr{P}_\nu^0(a'',b'';a',b';T) - i\,c_\beta^{-1}\int_0^T dt\int d\mu(a,b)\,\mathscr{P}_\nu^0(a'',b'';a,b;T-t)h(a,b)\,\mathscr{P}_\nu^0(a,b;a',b';t)\\ &\quad - c_\beta^{-2}\int_0^T dt_2\int_0^{t_2}dt_1\int d\mu(a_1,b_1)\int d\mu(a_2,b_2)\,\mathscr{P}_\nu^0(a'',b'';a_2,b_2;T-t_2)h(a_2,b_2)\\ &\quad\times\mathscr{P}_\nu^h(a_2,b_2;a_1b_1;t_2-t_1)h(a_1,b_1)\,\mathscr{P}_\nu^0(a_1,b_1;a',b';t_1).\end{aligned}$$

Rewriting this in terms of $\phi_{a,b;t}$ and $\phi_{a,b;\infty}$, and combining it with an analogous integral equation for the coherent state matrix elements of $\exp(-iTH)$ leads to (see Ref. 4)

$$\begin{aligned}&c_\beta^{-1}[\mathscr{P}_\nu^h(a'',b'';a',b';T) - \langle a'',b'';\beta|e^{-iTH}|a',b';\beta\rangle]\\ &\quad=(\phi_{a'',b'';\nu T}-\phi_{a'',b'';\infty})(a',b')\\ &\qquad -i\int_0^T dt\,\langle\phi_{a'',b'';\nu(T-t)},h(\phi_{a',b';\nu t}-\phi_{a',b';\infty})\rangle - i\int_0^T dt\,\langle\phi_{a'',b'';\nu(T-t)}-\phi_{a'',b'';\infty},h\phi_{a',b';\infty}\rangle\\ &\qquad -\int_0^T dt_2\int_0^{t_2}dt_1\langle h\phi_{a'',b'';\nu(T-t_2)},E(\nu,h;t_2-t_1)h\,[\phi_{a',b';\nu t_1}-\phi_{a',b';\infty}]\rangle\\ &\qquad -\int_0^T dt_2\int_0^{t_2}dt_1\langle h\,[\phi_{a'',b'';\nu(T-t_2)}-\phi_{a'',b'';\infty}],E(\nu,h;t_2-t_1)h\phi_{a',b';\infty}\rangle\\ &\qquad -\int_0^T dt_2\int_0^{t_2}dt_1\langle h\phi_{a'',b'';\infty},[E(\nu,h;t_2-t_1)-P_\beta e^{-iP_\beta hP_\beta(t_2-t_1)}P_\beta]h\phi_{a',b';\infty}\rangle.\end{aligned} \tag{4.44}$$

Denote the six terms in the right-hand side of (4.45) by $\Delta_1,\ldots,\Delta_6$. We show that $\Delta_j \to_{\nu\to\infty} 0$ for $j=1,\ldots,6$.

The estimates (4.15) and (4.30) can be rewritten as

$$\|\phi_{a,b;t}-\phi_{a,b;\infty}\|\leqslant f(t)\,,$$

$$\|h(\phi_{a,b;t}-\phi_{a,b;\infty})\|\leqslant g(a,b;t)\,,$$

where the functions $f(\cdot)$ and $g(a,b;\cdot)$ $[(a,b)$ fixed$]$ are monotonically decreasing in t, and integrable,

$$\int_0^\infty dt\, f(t)\leqslant\infty\,,$$

$$\int_0^\infty dt\, g(a,b;t)\leqslant\infty\,.$$

On the other hand, (4.13) and Lemma 4.2 tell us that

$$\|\phi_{a,b;\infty}\| = c_\beta^{-1/2}\quad [\text{for all } (a,b)]\,,$$

and $[(a',b'),(a'',b'')$ fixed$]$

$$\|h\phi_{a'',b'';\infty}\|\leqslant \not{c},\quad \|h\phi_{a',b';\infty}\|\leqslant\not{c}\,.$$

We now discuss the terms $\Delta_1,\ldots,\Delta_6$ one by one.

Using (3.38) we have immediately

$$\Delta_1\leqslant\not{c}\,[1+(\nu T)^{-1}]e^{-B(\beta)\nu T}\underset{\nu\to\infty}{\to}0\,.$$

The next four terms can be estimated in terms of f, g,

$$\Delta_2\leqslant\not{c}\int_0^T dt\,[1+f(\nu(T-t))]g(a',b';\nu t)$$

$$\leqslant\not{c}\frac{1}{\nu}\int_0^\infty dt\, g(a',b';t)$$

$$+\not{c}f\Big(\frac{\nu T}{2}\Big)\frac{1}{\nu}\int_0^\infty dt\, g(a',b';t)$$

$$+\not{c}g\Big(a',b';\frac{\nu T}{2}\Big)\frac{1}{\nu}\int_0^\infty dt\, f(t)$$

$$\leqslant\Big(\frac{1}{\nu}\Big)\not{c}\underset{\nu\to\infty}{\to}0\,,$$

$$\Delta_3\leqslant\not{c}\int_0^T dt\, f(\nu(T-t))\leqslant\Big(\frac{1}{\nu}\Big)\not{c}\underset{\nu\to\infty}{\to}0\,,$$

$$\Delta_4\leqslant\not{c}\int_0^T dt_2\int_0^{t_2}dt_1[1+g(a'',b'';\nu(T-t_2))]g(a',b';\nu t_1)$$

$$\leqslant\not{c}\int_0^T dt_1\, g(a',b';\nu t_1)\cdot(T-t_1)$$

$$\times\frac{1}{\nu^2}\not{c}\Big[\int_0^\infty dt_2 g(a'',b'';t_2)\Big]\cdot\Big[\int_0^\infty dt_1 g(a',b';t_1)\Big]$$

$$\leqslant\not{c}\Big(\frac{1}{\nu^2}\Big)+\not{c}T\frac{1}{\nu}\int_0^\infty dt_1\, g(a',b';t_1)\underset{\nu\to\infty}{\to}0\,,$$

$$\Delta_5\leqslant\int_0^T dt_2\int_0^{t_2}dt_1\, g(a'',b'';\nu(T-t_2))$$

$$\leqslant T\frac{1}{\nu}\int_0^\infty dt\, g(a'',b'';t)\underset{\nu\to\infty}{\to}0\,.$$

Finally, $\Delta_6\underset{\nu\to\infty}{\to}0$ follows from Proposition 4.5 and the dominated convergence theorem. This completes the proof of our main result.

Theorem 4.6: Let h be a real function on M_+. Suppose that (1) h satisfies condition (4.28), (2) the operator

$$H=c_\beta^{-1}\int d\mu(a,b)\,|a,b;\beta\rangle h(a,b)\langle a,b;\beta|\qquad(4.45)$$

is essentially self-adjoint on D_c, the finite linear span of the affine coherent states. Then, for all (a',b'), $(a'',b'')\in M_+$ and for all $T>0$

$$\lim_{\nu\to\infty}c_\beta e^{\nu T\beta}\int\exp\Big[-i\beta\int a^{-1}db-i\int h(a,b)dt\Big]$$

$$\times d\mu_W^\nu(a,b)=\langle a'',b'';\beta|e^{-iTH}|a',b';\beta\rangle\,.$$

E. Remarks

1. The main result in the pq-notation

We define $\widetilde{\Delta}_\beta=\beta^{-1}\partial_p p^2\partial_p+\beta p^{-2}\partial_q^2$.

Let $\widetilde{K}_t$ be the associated heat kernel, in $L^2(M_+;[(1-1/2\beta)/2\pi]dp\,dq)$,

$$\widetilde{K}_t(p'',q'';\,p',q')\equiv[\exp(t\widetilde{\Delta}_\beta)](p'',q'';p',q')$$

$$=\frac{e^{-t/4\beta}\beta^{3/2}}{2\sqrt{2\pi}(\beta-\frac{1}{2})t^{3/2}}$$

$$\times\int_\delta^\infty\frac{xe^{-\beta x^2/4t}}{\sqrt{\cosh x-\cosh\delta}}dx,$$

where

$$\delta=d(p'',q'';p',q')$$

$$=\cosh^{-1}\Big\{1+\frac{p'p''}{2}[(p'^{-1}-p''^{-1})^2+\beta^2(q'-q'')^2]\Big\}.$$

Define $d\tilde{\mu}^{\nu,T}_{W;p'',q'';p',q'}$ to be the associated Wiener process with diffusion constant ν, pinned at p',q' for $t=0$, at p'',q'' for $t=T$. In particular $d\tilde{\mu}^\nu_W$ satisfies

$$\int d\tilde{\mu}^{\nu,T}_{W;p'',q'';p',q'}=\widetilde{K}_{\nu T}(p'',q'';p',q'),$$

$$\frac{1-1/2\beta}{2\pi}\int dp\,dq\;d\tilde{\mu}^{\nu,T-t}_{W;p'',q'';p,q}d\tilde{\mu}^{\nu,t}_{W;p,q;p',q'}$$

$$=d\tilde{\mu}^{\nu,T}_{W;p'',q'';p',q'}\,.$$

Let h be a function on M_+ satisfying

$$\int dp\,dq\,|h(p,q)|^{2+r}\Big[\frac{p}{1+p^2(q^2+1)}\Big]^\mu<\infty,\qquad(4.46)$$

for some μ,r satisfying condition (4.29).

Let H be the operator on $L^2(\mathbb{R}_+)$ defined by

$$H=\frac{1-1/2\beta}{2\pi}\int dp\,dq|p,q;\beta>h(p,q)<p,q;\beta|,$$

where, for $\psi\in L^2(\mathbb{R}_+)$,

$$\langle p,q;\beta|\psi\rangle=(2\beta)^\beta[\Gamma(2\beta)]^{-1/2}p^{-\beta}$$

$$\times\int_0^\infty dk\,k^\beta e^{-k(\beta p^{-1}-iq)}\psi(k)$$

[see (2.14)].

Define the path integral

$$\mathscr{P}^h_\nu(p'',q'';p',q';T)$$

$$=e^{\nu T/2}\int\exp\Big[i\int p\,dq-i\int h(p,q)dt\Big]d\tilde{\mu}^{\nu,T}_{W;p'',q'';p',q'}$$

[this differs by a factor c_β from (4.1); this factor is absorbed in the measure in the pq-notation].

Translated into the pq-notation, the main theorem now reads (1) if h satisfies (4.45), and (2) if H is essentially self-adjoint on $\widetilde{D}_c$, the finite linear span of the $|p,q;\beta\rangle$, then, for all (p'',q''), $(p',q')\in M_+$, for all $T>0$,

$$\lim_{\nu\to\infty} \widetilde{\mathscr{P}}^h_\nu(p'',q'';p',q';T)$$
$$= \langle p'',q'';\beta|\exp(-iTH)|p',q';\beta\rangle.$$

2. Examples

(a) The simplest example is, of course, provided by bounded functions $h(a,b)$,

$$|h(a,b)|\leqslant M.$$

In this case the operator H defined by (4.45) is also bounded by M; H is thus clearly essentially self-adjoint on D_c. Moreover the condition (4.28) is satisfied for arbitrary $r>0$ and for all $\mu>1$. Let us now determine from (4.29) or (4.31) the restrictions imposed on β by the condition $\mu>1$. Two possibilities have to be distinguished: $\frac{1}{2}<\beta\leqslant\frac{3}{2}$ or $\beta\geqslant\frac{3}{2}$. In the first case we have $B(\beta)=(\beta-\frac{1}{2})^2$ in (4.29b), leading to the condition

$$2(1-\alpha)\beta^2-2\beta+\tfrac{1}{2}>\frac{r}{r+2}\,\tilde{\epsilon}\Big(2\alpha\frac{r+2}{r},\frac{2\mu}{r}\Big), \tag{4.47}$$

with $\tilde{\epsilon}$ as defined by (4.31c). It turns out there is no set of values (α,r,μ) with $r/2(r+2)<\alpha<1$, and $\mu>1$, such that (4.47) is satisfied for $\beta\in(\frac{1}{2},\frac{3}{2}]$.

For $\beta\geqslant\frac{3}{2}$ we have to determine β satisfying the conditions (4.31). One has then to choose (α,r,μ) so as to produce the smallest possible lower bound on β consistent with the other conditions. For $\mu>1$, $r=\frac{1}{2}$, and $\alpha=\frac{1}{3}$ one finds that (4.31a) reduces to $\beta>2.06$, while all the other conditions are fulfilled also.

This means that Theorem 4.6 allows us to conclude that, for bounded Hamiltonians H associated to bounded functions $h(a,b)$,

$$\lim_{\nu\to\infty} c_\beta e^{\nu T\beta}\int\exp\Big[-i\beta\int a^{-1}\,db-i\int h(a,b)dt\Big]$$
$$\times d\mu^\nu_W(a,b)=\langle a'',b'';\beta|e^{iTH}|a',b';\beta\rangle, \tag{4.48}$$

for all $\beta>2.06$.

We believe that, for bounded functions h, (4.48) should hold for all $\beta>\frac{1}{2}$, since it holds for $h=\text{const}$ whenever $\beta>\frac{1}{2}$. The 2.06-bound found here is probably an artifact of our method of proof, which uses Young's and Jensen's inequalities several times (in the proof of Lemma 4.3).

(b) We next turn to examples of the form

$$H=-\frac{d^2}{dx^2}+V(x)$$

on $L^2(\mathbb{R}_+)$.

In order for this operator to be essentially self-adjoint on D_c, V must have a singularity at the origin. More precisely, H will be essentially self-adjoint on D_c (regardless of β), e.g., for $V(x)$ of the form

$$V(x)=C_1x^{-\alpha_1}+C_2x^{\alpha_2},$$

where either $\alpha_1>2$, $C_1>0$ or $\alpha_1=2$, $C_1\geqslant\frac{3}{4}$, and either $0\leqslant\alpha_2\leqslant2$, C_2 arbitrary, or $\alpha_2>2$, $C_2\geqslant0$. In all these cases V has a strong singularity at $x=0$; for $x\to\infty$, V may tend to ∞, a constant, or $-\infty$, depending on the values chosen for the different parameters.

Let us now construct the corresponding functions $h(a,b)$, and determine the values of β for which Theorem 4.6 applies. The function $h_0(a,b)$ corresponding to $-d^2/dx^2$ is given by

$$h_0(a,b)=b^2-(1/2\beta)a^2$$

[one easily checks that substitution of h_0 into (4.45) leads to $-d^2/dx^2$]. Similarly, the function $h_\alpha(a,b)$ associated with $x^{-\alpha}$ is given by

$$h_\alpha(a,b)=\frac{2^\alpha\Gamma(2\beta-1)}{\Gamma(2\beta+\alpha-1)}a^\alpha.$$

Hence the function $h(a,b)$ corresponding to the Hamiltonian $-(d^2/dx^2)+V$, with V as above, is given by

$$h(a,b)=b^2-\frac{1}{2\beta}a^2+C_1\frac{2^{\alpha_1}\Gamma(2\beta-1)}{\Gamma(2\beta+\alpha_1-1)}a^{\alpha_1}$$
$$+C_2\frac{2^{-\alpha_2}\Gamma(2\beta-1)}{\Gamma(2\beta-\alpha_2-1)}a^{-\alpha_2}.$$

If $C_2\neq0$ we have to impose the additional restriction $2\beta-\alpha_2-1\notin-\mathbb{N}$.

We shall restrict ourselves to one particular case now. We take $C_2=0$, $\alpha_1=2$, and $C_1\geqslant\frac{3}{4}$. The Hamiltonian H is essentially self-adjoint, and

$$h(a,b)=b^2+\frac{1}{\beta}\Big(\frac{C_1}{\beta-\frac{1}{2}}-\frac{1}{2}\Big)a^2.$$

The pairs (r,μ) for which this function satisfies the condition (4.28) are restricted by the condition $\mu>2(r+2)$. We have thus to find (r,α,μ) satisfying this condition as well as the conditions (4.31b); this then enables us, from (4.31a) to compute a β_0 such that Theorem 4.6 applies, for this Hamiltonian, for all $\beta>\beta_0$. For $\alpha=\frac{1}{3}$, $r=1$, and $\mu>6$, one finds that (4.31a) becomes $\beta>27.33$. It is easy to check that all the other conditions are satisfied as well. Hence Theorem 4.6 applies to $H=-d^2/dx^2+Cx^{-2}$, $C\geqslant\frac{3}{4}$, if $\beta>27.33$. Again we believe that this is not optimal. The true lower bound β_0 on β for which (4.48) would hold, whenever $\beta>\beta_0$, is probably much smaller than the here computed value 27.33, though possibly larger than $\frac{1}{2}$.

3. A formula giving the function h from the operator H

Formula (4.45) defines the operator H for a given function h. If we define the function $\mathrm{H}(a,b)$ to be the diagonal coherent state matrix elements of H,

$$\mathrm{H}(a,b)=\langle a,b;\beta|H|a,b;\beta\rangle,$$

then (4.45) leads to

$$\mathrm{H}(a,b)=c_\beta^{-1}\int\frac{d\mu(a',b')}{a'^2}h(a',b')|\langle a,b;\beta|a',b';\beta\rangle|^2$$
$$=c_\beta^{-1}\int\frac{d\mu(a',b')}{a'^2}h(a',b')$$
$$\times\Big[\frac{2}{1+\cosh d(a,b;a',b')}\Big]^{2\beta}.$$

This formula can be inverted. Using results in Ref. 17 one finds

$$h(a,b) = (\mathrm{TH})(a,b), \tag{4.49}$$

where the operator T, acting on the function H, is given by

$$\mathrm{T} = \prod_{n=0}^{\infty} \left[1 - \frac{\Delta}{(2\beta+n+1)(2\beta+n+2)} \right], \tag{4.50}$$

with $\Delta = a^2(\partial_a^2 + \partial_b^2)$, the Laplace–Beltrami operator on the Lobachevsky plane.

It turns out that this infinite product can be rewritten in terms of Γ-functions. One way to see this is to use the correspondence (4.49) for a family of special cases. For $H = x^{-\alpha}$ we know already that

$$h(a,b) = \frac{2^\alpha \Gamma(2\beta-1)}{\Gamma(2\beta+\alpha-1)} a^\alpha.$$

On the other hand, the corresponding function $\mathrm{H}(a,b)$ is

$$\mathrm{H}(a,b) = \langle a,b;\beta | x^{-\alpha} | a,b;\beta \rangle = \frac{2^\alpha \Gamma(2\beta-\alpha)}{\Gamma(2\beta)} a^\alpha.$$

This implies that

$$\prod_{n=0}^{\infty} \left[1 + \frac{-\alpha(\alpha-1)}{(2\beta+n+1)(2\beta+n+2)} \right] = \frac{\Gamma(2\beta)\Gamma(2\beta-1)}{\Gamma(2\beta-\alpha)\Gamma(2\beta+\alpha-1)}.$$

By analytic continuation one finds that, for all $t > 0$,

$$\prod_{n=0}^{\infty} \left[1 + \frac{t^2+1/4}{(2\beta+n+1)(2\beta+bn+2)} \right] = \frac{\Gamma(2\beta)\Gamma(2\beta-1)}{\Gamma(2\beta-it-\frac{1}{2})\Gamma(2\beta+it-\frac{1}{2})} = \frac{B(2\beta,2\beta-1)}{B(2\beta-it-\frac{1}{2},2\beta+it-\frac{1}{2})}. \tag{4.51}$$

Since the spectrum of $-\Delta = -a^2(\partial_a^2 + \partial_b^2)$ on the Lobachevsky plane is $[\frac{1}{4},\infty)$, (4.51) determines (4.50) completely. For particular values of β, (4.51) and hence (4.50) can be further simplified. For $\beta = 1$, e.g., we find

$$\frac{B(2,1)}{B(\frac{3}{2}-it,\frac{3}{2}+it)} = \frac{\pi}{(\frac{9}{4}+t^2)\cosh(\pi t)}.$$

This can then be used to give an integral representation for T. We have, e.g.,

$$\int_0^\pi dx \frac{\cos tx}{\cosh x/2} = \frac{\pi}{\cosh t\pi},$$

hence

$$\prod_{n=0}^{\infty} \left[1 + \frac{-\Delta}{(n+3)(n+4)} \right] = (-\Delta+2)^{-1} \int_0^\infty dt \frac{\cos[t\sqrt{-\Delta+\frac{1}{4}}]}{\cosh t/2},$$

with

$$\cos[t\sqrt{-\Delta+\tfrac{1}{4}}] = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} t^{2n} \left(-\Delta + \frac{1}{4} \right)^n.$$

ACKNOWLEDGMENTS

Two of us (I.D. and T.P.) would like to thank the Dipartimento di Matematica, Università di Bologna, for its hospitality during some of the early stages of this work. I.D. also gratefully acknowledges the hospitality of the BiBoS-project, Universität of Bielefeld, where this work was completed.

It is a pleasure for us to thank André Voros for many helpful remarks concerning the Lobachevsky plane.

[1]J. R. Klauder, "Path integrals for affine variables," in *Functional Integration Theory and Applications*, edited by J. P. Antoine and E. Tirapagui (Plenum, New York, 1980), p. 101.

[2]E. W. Aslaksen and J. R. Klauder, J. Math. Phys. **10**, 2267 (1969).

[3]T. Paul. J. Math. Phys. **25**, 3252 (1984).

[4]I. Daubechies and J. R. Klauder, J. Math. Phys. **26**, 2239 (1985).

[5]J. R. Klauder, "Coherent-state path integrals for unitary group representations," to be published in the *Proceedings of the 14th International Conference on Group Theoretical Methods in Physics*, Seoul, South Korea, August 1985.

[6]See, e.g., J. R. Klauder and B. S. Skagerstam, *Coherent States. Applications in Physics and Mathematical Physics* (World Scientific, Singapore, 1985).

[7]M. Duflo and C. C. Moore, J. Funct. Anal. **21**, 208 (1976); A. L. Carey, Bull. Austral. Math. Soc. **15**, 1 (1976); A. Grossmann, J. Morlet, and T. Paul, J. Math. Phys. **26**, 2473 (1985); S. A. Gaal, *Linear Analysis and Transformation Theory* (Springer, Berlin, 1973).

[8]J. E. Moyal, Proc. Cambridge Philos. Soc. **45**, 99 (1949); M. S. Bartlett and J. E. Moyal, *ibid.* **45**, 545 (1949); I. Segal, Math. Scand. **13**, 31 (1963); J. C. Pool, J. Math. Phys. **7**, 7 (1966).

[9]L. Robin, *Fonctions sphériques de Legendre et fonctions sphéroidales* (Gauthier–Villars, Paris, 1957).

[10]V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961).

[11]S. Bergman, *The Kernel Function and Conformal Mapping* [Am. Math. Soc., Providence, RI, 1950, 1970 (2nd ed.,)]; N. Aronszajn, Trans. Am. Math. Soc. **68**, 337 (1950).

[12]I. M. Gel'fand and A. M. Yaglom, J. Math. Phys. **1**, 47 (1960); K. Itô, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (California U.P., Berkeley, 1967), Vol. 2, Part 1, pp. 145–161; J. Tarski, Ann. Inst. H. Poincaré **17**, 313 (1972).

[13]M. Reed and B. Simon, *Methods of Modern Mathematical Physics. II Fourier Analysis, Self-adjointness* (Academic, New York, 1975).

[14]P. Morse, Phys. Rev. **34**, 57 (1929).

[15]B. Simon, *Functional Integration and Quantum Physics* (Academic, New York, 1979).

[16]E. Lieb, Bull. Am. Math. Soc. **82**, 751 (1976), More details can be found in E. Lieb, Proc. Am. Math. Soc. **36**, 241 (1980).

[17]F. A. Berezin, Commun. Math. Phys. **40**, 153 (1975).

Constructive representations of propagators for quantum systems with electromagnetic fields

T. A. Osborn, L. Papiez, and R. Corns
Department of Physics, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2

(Received 5 May 1986; accepted for publication 3 September 1986)

The quantum evolution of an N-body system of particles that mutually interact through scalar fields and couple to an arbitrary external electromagnetic field is rigorously described. Both operator- and kernel-valued solutions to the evolution problem are found. Based upon a particular realization of the Dyson expansion, a convergent series representation of the propagator (the kernel of the Schrödinger time evolution operator) is obtained. The basic approach is to embed the quantum evolution problem in the larger class of evolution problems that result if mass is allowed to be complex. Quantum evolution with real mass is considered to be the boundary value of the complex mass evolution problem. The constructive representation of the propagator is determined for the class of analytic scalar and vector fields that are given as Fourier transforms of time-dependent scalar- and vector-valued measures.

I. INTRODUCTION

The dynamical evolution of a quantum system is determined by the solutions of Schrödinger's time-dependent equation of motion

$$i\hbar \frac{\partial}{\partial t}\psi(x,t) = H(x,t)\psi(x,t), \tag{1.1}$$

where x denotes a generic point (in the d-dimensional Euclidean space $\mathbb{R}^d$) that specifies the position of all the particles in the system. The solution $\psi(\cdot,t)$ is a state vector that is an element of Hilbert space $\mathscr{H} = L^2(\mathbb{R}^d)$ for each value t of the time parameter. Within a given finite time interval $[0,T]$ the differential structure of the Hamiltonian is taken to be of the general form

$$H(x,t) = (1/2m)((\hbar/i)\nabla_x - a(x,t))^2 + v(x,t). \tag{1.2}$$

The function a above represents a time-dependent vector field mapping $\mathbb{R}^d\times[0,T]\to\mathbb{R}^d$, while v is a scalar potential from $\mathbb{R}^d\times[0,T]$ to $\mathbb{R}$.

The Schrödinger problem just outlined contains nonrelativistic N-body quantum dynamics. If each individual particle, having mass m, moves in three dimensions then $d=3N$. The presence of the vector potential in (1.2) means that the Hamiltonian describes a collection of particles that mutually interact through scalar fields and couple via the Lorentz force to a time-dependent, spatially inhomogeneous, external electromagnetic field.

This paper obtains both operator-valued and kernel-valued solutions of (1.1). In the ensuing analysis we place restrictions on a and v that ensure that the minimal operator $H(\cdot,t)$, defined on $C_0^\infty(\mathbb{R}^d)$, has a unique closed extension $H(t)$ and further that the (dense) domain of $H(t)$ is independent of time, i.e., $D((H(t)) = D_0$ for all $t\in[0,T]$. In these circumstances the Schrödinger evolution problem in $\mathscr{H}$, associated with (1.1), assumes the following form. Let T_Δ be the triangular region $\{(t,s)\in\mathbb{R}^2: 0\leqslant s\leqslant t\leqslant T\}$. Suppose ϕ is an arbitrary function chosen from $D_0\subset\mathscr{H}$ and s is the time at which the initial data condition is imposed, then the Schrödinger evolution problem in T_Δ consists of solving

$$i\hbar\frac{d}{dt}\psi(t) = H(t)\psi(t), \quad \psi(t)\in D_0, \tag{1.3}$$

with

$$\psi(s)=\phi. \tag{1.4}$$

One way of characterizing the solution $\psi(t)$ of Eqs. (1.3) and (1.4) is to determine a bounded operator family $U(t,s)$: $\mathscr{H}\to\mathscr{H}$ such that

$$\psi(t) = U(t,s)\phi, \quad t,s\in T_\Delta. \tag{1.5}$$

We will call $U(t,s)$ the Schrödinger evolution operator. In the time-independent case where $H(t)=H$ for all t and H is self-adjoint, $U(t,s)$ is only a function of the time displacement $\tau = t-s$ and may be written

$$U(\tau+s,s) = e^{-i\tau H/\hbar}. \tag{1.6}$$

The operators on the right-hand side of (1.6) form a one-parameter ($\tau\in\mathbb{R}$) unitary group. Perhaps the most basic effect of the t dependence of $H(t)$ is to cause $U(t,s)$ to depend separately on t and s.

Let us define the evolution operator more completely. Denote by the symbol $\mathscr{B}(\mathscr{H},\mathscr{H})$ the Banach space (with norm $\|\cdot\|$) of all bounded operators mapping $\mathscr{H}$ into $\mathscr{H}$. The identity operator on $\mathscr{H}$ is indicated by I.

Definition 1: A two-parameter operator family U: $T_\Delta\to\mathscr{B}(\mathscr{H},\mathscr{H})$ is said to be the *(Schrödinger) evolution operator* generated by $\{H(t): t\in[0,T]\}$ if

(1) $U(t,s)$: $D_0\to D_0$, $t,s\in T_\Delta$.

(2) U is uniformly bounded in T_Δ and for some finite $c>0$,

$$\|U(t,s)\|\leqslant e^{(t-s)c}, \quad t>s. \tag{1.7}$$

(3) U is strongly continuous in T_Δ.

(4) The following identities hold in $\mathscr{B}(\mathscr{H},\mathscr{H})$:

$$U(t,s) = U(t,\tau)U(\tau,s), \quad 0\leqslant s\leqslant\tau\leqslant t\leqslant T, \tag{1.8}$$

$$U(s,s) = I, \quad s\in[0,T]. \tag{1.9}$$

(5) On the domain D_0, U is strongly continuously differentiable relative to t and s. Furthermore, U satisfies the equations of motion ($t,s\in T_\Delta$),

 0022-2488/87/010103-21$02.50

$$i\hbar \frac{\partial}{\partial t} U(t,s) f = H(t) U(t,s) f, \quad f \in D_0, \qquad (1.10)$$

$$- i\hbar \frac{\partial}{\partial s} U(t,s) f = U(t,s) H(s) f, \quad f \in D_0. \qquad (1.11)$$

It is apparent that given the existence of a Schrödinger evolution operator properties (1), (4), and (5) ensure that $\psi(t)$ defined by (1.5) is an L^2 (strong) solution of (1.3) taking values in D_0.

For many physical problems (particular choices of a and v) it turns out that $U(t,s)$ is a type of integral operator. In this case we call its associated time evolution kernel the propagator of the system. Let $\langle \cdot,\cdot \rangle$ be the inner product in $\mathscr{H}$ (linear in the right element). We have, following Ref. (1),

Definition 2: A two-parameter family (in T_Δ, $t \neq s$) of functions $K(t,\cdot;s,\cdot)$: $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ that are measurable and locally integrable on $\mathbb{R}^d \times \mathbb{R}^d$ is called the *propagator* for evolution $\{U(t,s): t,s \in T_\Delta\}$ if for all L^∞ functions of compact support f,g

$$\langle f, U(t,s) g \rangle = \int_{\mathbb{R}^d \times \mathbb{R}^d} \overline{f(x)} K(t,x;s,y) g(y) dx\, dy. \qquad (1.12)$$

Technically Definition 2 asserts that $U(t,s)$ is a weak integral operator with kernel K. Examination of the exactly solvable free problem with $a(x,t) = 0$ and $v(x,t) = 0$ shows that one cannot expect the stronger statement that K be a Carleman kernel.[1] For example the kernel is not L^2 in y for almost all fixed x.

The principal goal of our study of the evolution problem (1.1) is to establish the existence of the propagator and to investigate its properties by finding a representation of K in terms of an explicit absolutely convergent series. We focus much of our attention on the propagator because this is arguably the most important object in quantum evolution theory.

Our basic approach is to embed the evolution problem (1.1) in a larger problem. The mass parameter is allowed to take all values in the upper half complex plane, and the original positive mass problem is considered as a boundary value of the complex mass evolution problem. It will be shown that evolution operator U is defined and continuous in the parameter m for $\operatorname{Im} m \geqslant 0$, $m \neq 0$ and similarly that the propagator K also has nice continuity properties in m. The advantage gained by this enlargement of the original problem is that computations involving the propagator K and its approximations are made simpler. The various multiple integrals appearing in the explicit construction of the propagator are all of a Gaussian type when $\operatorname{Im} m > 0$. Thus Fourier transforms behave well, and the several changes of limiting order that the analysis requires become manageable. Our use of mass as an analytic variable to assist in the determination of the propagator has some parallels with Nelson's program[2] of giving meaning to the Feynman path integral via the complex mass limit $\operatorname{Im} m \to 0$.

The constructive series representation of the propagator K is derived for the class of analytic vector and scalar fields that can be written as the Fourier image of certain time-dependent vector and scalar-valued measures. For example, if $\gamma(t)$ denotes for each $t \in [0,T]$ a complex vector-valued measure on the Borel subsets of $\mathbb{R}^d$, then

$$a(x,t) = \int e^{i\alpha \cdot x} d\gamma(t). \qquad (1.13)$$

Here α is the variable of integration (*not* displayed in the measure) with domain $\mathbb{R}^d$. The symbol $\alpha \cdot x$ is the scalar product. Furthermore it is assumed that $\gamma(t)$ has support on a compact set of $\mathbb{R}^d$. Thus (for each t) $a(x,t)$ is a real analytic function of x. Moreover sufficient continuity properties in t are imposed on the measures $\gamma(t)$ in order to ensure that $a(x,t)$ is continuously differentiable in t. In a similar fashion the scalar field is assumed to be the Fourier transform of a complex scalar-valued compactly supported measure $\nu(t)$.

$$v(x,t) = \int e^{i\alpha \cdot x} d\nu(t). \qquad (1.14)$$

The use of this class of potentials consisting of Fourier images of complex bounded measures was initiated by Ito in the study of the Feynman path integrals[3] and plays a central role in results Albeverio and Høegh-Krohn[4] obtain for the path integral. In the situation where no electromagnetic field is present and $v(x,t)$ is static, a constructive representation of the propagator K is found in Osborn and Fujiwara.[5] Potentials of the form (1.14) are suitable for modeling the total N-body scalar interaction energy because there is no assumption of decay as $|x| \to \infty$. Similar remarks apply to the vector field $a(x,t)$ obtained from (1.13). The Fourier image (1.13) does not necessarily imply any decrease in the vector field (or the associated electric and magnetic fields) as $|x| \to \infty$. For example, one may choose $\gamma(t)$ so that the Stark problem of a quantum system in a constant electric field is obtained.

A description of the measures γ and ν, together with the related behavior of the fields a and v is found in Sec. II. Properties of the family of Hamiltonians $H(t)$ such as closure, domain stability, and the existence of the strong derivative on D_0 are discussed in Sec. III. Our method of obtaining the existence and properties of the evolution operator $U(t,s)$ proceeds by adapting the theory of linear differential equations in Banach space[6] to the study of the Schrödinger evolution problem (1.3) and (1.4). Section IV is devoted to this topic.

Let $H(t)$ be interpreted as a perturbation of the free Laplacian operator $H_0 = -(\hbar^2/2m)\Delta_x$. In this spirit the full Hamiltonian is written

$$H(t) = H_0 + V(t). \qquad (1.15)$$

Using formal arguments it is evident that the abstract integral equation equivalent to the equation of motion (1.10) for $U(t,s)$ with boundary condition (1.9) is

$$U(t,s) = U_0(t,s) - \frac{i}{\hbar} \int_s^t d\tau\, U_0(t,\tau) V(\tau) U(\tau,s), \qquad (1.16)$$

where U_0 indicates the free evolution given by Eq. (1.6) with $H = H_0$.

In the physics literature a common approach to investigating $U(t,s)$ is to iterate (1.16) to obtain the familiar Dyson expansion[7]

$$U(t,s) = U_0(t,s) + \sum_{n=1}^{\infty} \left(-\frac{i}{\hbar}\right)^n \times \int_< dt_n \cdots dt_1\, U_0(t,t_n) V(t_n) \times U_0(t_n,t_{n-1}) \times \cdots \times V(t_1) U_0(t_1,s), \quad (1.17)$$

where the integral subscript $<$ denotes the time-ordered domain $t \geqslant t_n \geqslant \cdots \geqslant t_1 \geqslant s$. It is known (Ref. 8, Chap. IX), however, that this method is not generally successful in yielding rigorous results for $U(t,s)$. The presence of the singular derivative coupling terms in the Hamiltonian (1.2) means that the operator $V(t)$ is intrinsically unbounded. Thus, in general, one has the difficulty of showing that the product of the unbounded operators appearing on the right-hand side of (1.17) have the necessary domain and range consistency, that they define (operator-valued) integrable functions, and that the resulting integrals form a convergent series. Nevertheless, for the case considered in this paper, we shall be able to prove that series (1.17) determines the solution of the evolution problem (1.3) and (1.4). This is described in Sec. V.

The final section provides a construction of the evolution propagator K from formulas suggested by the perturbation series (1.17) and establishes that the real mass boundary value of the constructive series for K is actually the weak kernel of the quantum evolution operator $U(t,s)$.

II. MEASURES AND FIELDS

This section presents the precise definition of the scalar and vector fields that enter the Hamiltonian $H(t)$. These fields are Fourier images of certain measures. The basic mathematical properties of the relevant measure spaces are summarized.

Let the couple $(\mathbb{R}^d,B)$ specify the measurable space consisting of the set $\mathbb{R}^d$ and the smallest σ-algebra B of (Borel) subsets of $\mathbb{R}^d$ that contain all the open sets in $\mathbb{R}^d$. Indicate by $\mathbb{C}^r$ the complex r-dimensional Euclidean space. A complex $\mathbb{C}^r$-valued measured γ on $(\mathbb{R}^d,B)$ is a countably additive set function from B to $\mathbb{C}^r$ ($r=1$ for scalar measures). The associated total variation of γ is the non-negative set function $|\gamma|$ on $(\mathbb{R}^d,B)$ defined by

$$|\gamma|(e) = \sup_{\pi} \sum_{e_i \in \pi} |\gamma(e_i)| \quad (e_i \in B). \tag{2.1}$$

On the right-hand side of (2.1), $|\cdot|$ is the Euclidean norm for the space $\mathbb{C}^r$, and the supremum is taken over all countable partitions π of e allowed by B. The set function γ is said to be of bounded variation if $|\gamma|\,(\mathbb{R}^d) < \infty$.

Next we construct a Banach space of vector-valued measures γ. To this end, in the set of all $\mathbb{C}^r$-valued measures on $(\mathbb{R}^d,B)$ with bounded variation adjoin, in the standard way,[9] the operations of summation and multiplication by a complex scalar. It is evident, that so equipped, this set of measures is a vector space which we shall designate by the symbol $\mathfrak{M}(\mathbb{R}^d,\mathbb{C}^r)$. Further, one may use the total variation of the set $\mathbb{R}^d$ as the natural norm in $\mathfrak{M}(\mathbb{R}^d,\mathbb{C}^r)$, namely

$$\|\gamma\| = |\gamma|(\mathbb{R}^d), \quad \gamma \in \mathfrak{M}(\mathbb{R}^d,\mathbb{C}^r). \tag{2.2}$$

With this norm attached one can prove[9] that $\mathfrak{M}(\mathbb{R}^d,\mathbb{C}^r)$ is a Banach space.

Each vector-valued measure $\gamma \in \mathfrak{M}(\mathbb{R}^d,\mathbb{C}^r)$ generates a vector field a: $\mathbb{R}^d \to \mathbb{C}^r$ via the Fourier transform of the measure,

$$a(x) = \int e^{i\alpha \cdot x}\, d\gamma. \tag{2.3}$$

Clearly, for each $x \in \mathbb{R}^d$, the function $e^{i\alpha \cdot x}$ is Borel measurable and $L^1(\mathbb{R}^d, d\gamma)$. In fact, a can be shown to be both uniformly continuous and uniformly bounded throughout $\mathbb{R}^d$. In particular note

$$|a(x)| \leqslant \|\gamma\|, \quad x \in \mathbb{R}^d. \tag{2.4}$$

The statement of the evolution problem in Sec. I presupposes that the vector and scalar fields are real valued. This requirement imposes a restriction on the form of the measure γ. Let $\mathscr{F}^r$ be the Fourier image of $\mathfrak{M}(\mathbb{R}^d,\mathbb{C}^r)$, i.e., all possible a: $\mathbb{R}^d \to \mathbb{C}^r$ defined by (2.3) as γ varies throughout $\mathfrak{M}(\mathbb{R}^d,\mathbb{C}^r)$. If we define a norm for $\mathscr{F}^r$ by $\|a\| \equiv \|\gamma\|$, then $\mathscr{F}^r$ is also a Banach space. Furthermore, the Fourier transform mapping $\mathfrak{M}(\mathbb{R}^d,\mathbb{C}^r) \to \mathscr{F}^r$ establishes a one-to-one correspondence between $\mathscr{F}^r$ and $\mathfrak{M}(\mathbb{R}^d,\mathbb{C}^r)$. This is a consequence of the uniqueness[10] of the transform (2.3) which asserts that $a(x) = 0$ for all x if and only if $\gamma = 0$. A measure γ is said to satisfy the *reflection property* if

$$\gamma(e) = \overline{\gamma(-e)} \quad (e \in B),$$

where $-e$ is defined as $\{\alpha \in \mathbb{R}^d:\ -\alpha \in e\}$. If γ obeys the reflection property then the associated a is real. Conversely, if a is real for all x then (by the uniqueness of the Fourier transform) γ must satisfy the reflection property. We denote by $\mathfrak{M}^*(\mathbb{R}^d,\mathbb{C}^r)$ the set of all γ in $\mathfrak{M}(\mathbb{R}^d,\mathbb{C}^r)$ that satisfy the reflection property.

The polar decomposition of measure γ in terms of $|\gamma|$ is used extensively in the following. The polar decomposition of γ asserts that there is a Borel measurable function η: $\mathbb{R}^d \to \mathbb{C}^r$ such that $|\eta(\alpha)| = 1$ for all α and

$$\int_e d\gamma = \int_e \eta(\alpha)\, d\,|\gamma|, \quad e \in B. \tag{2.5}$$

Note, since $\gamma \ll |\gamma|$ the Radon–Nikodym theorem (Ref. 11, p. 64) establishes that there exists an $L^1(\mathbb{R}^d, d\,|\gamma|)$ function η which satisfies (2.5). The proof that $|\eta(\alpha)| = 1$ follows from a modification of the derivation given by Rudin (Ref. 12, Theorems 1.40 and 6.12) for the scalar case ($r=1$).

In the following we shall use products of complex scalar measures. Consider the case of two scalar measures: μ_1 over $(\mathbb{R}_1^d, B_1)$ and μ_2 over $(\mathbb{R}_2^d, B_2)$. It is known[9,12] that they uniquely determine a product measure, $\mu_1 \times \mu_2$, on the smallest σ-algebra $B_1 \times B_2$ (of Borel subsets) generated by the family of rectangles $A_1 \times A_2$, $A_1 \in B_1$, $A_2 \in B_2$, if it is required that for every rectangle $A_1 \times A_2$

$$(\mu_1 \times \mu_2)(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) \tag{2.6}$$

holds. It is evident that a Banach space $\mathfrak{M}(\mathbb{R}_1^d \times \mathbb{R}_2^d, \mathbb{C})$ may be constructed in this case exactly as before with a norm given now by the total variation of the set $\mathbb{R}_1^d \times \mathbb{R}_2^d$, i.e.,

$$\|\mu_1 \times \mu_2\| = |\mu_1 \times \mu_2|(\mathbb{R}_1^d \times \mathbb{R}_2^d). \tag{2.7}$$

It is worthwhile to notice that the following identity holds in $\mathfrak{M}(\mathbb{R}_1^d \times \mathbb{R}_2^d, \mathbb{C})$:

$$\mu_1 \times (\mu_2' + \mu_2'') = \mu_1 \times \mu_2' + \mu_1 \times \mu_2'', \tag{2.8}$$

where $\mu_1 \in \mathfrak{M}(\mathbb{R}_1^d, \mathbb{C})$ and $\mu_2', \mu_2'' \in \mathfrak{M}(\mathbb{R}_2^d, \mathbb{C})$. In addition we note (Ref. 9, p. 192)

$$\|\mu_1 \times \mu_2\| = \|\mu_1\| \, \|\mu_2\|. \tag{2.9}$$

The appearance of terms like $a(x,t) \cdot a(x,t)$ in the Hamiltonian (1.2) means that we need to understand the definition of the scalar convolution of measures in $\mathfrak{M}(\mathbb{R}^d, \mathbb{C}^r)$. Let γ, γ' be a pair of measures in $\mathfrak{M}(\mathbb{R}^d, \mathbb{C}^r)$ and a, a' the associated pair of vector fields in $\mathscr{F}^r$. The symbol $*$ denotes the map $\mathfrak{M}(\mathbb{R}^d, \mathbb{C}^r) \times \mathfrak{M}(\mathbb{R}^d, \mathbb{C}^r) \to \mathfrak{M}(\mathbb{R}^d, \mathbb{C})$ defined by

$$\gamma * \gamma'(e) = \int \chi_e(\alpha + \alpha') \eta(\alpha) \cdot \eta'(\alpha') d\, |\gamma| \times |\gamma'|. \tag{2.10}$$

Here η, η' are the unit length vectors occurring in the polar factorization of γ and γ'; $|\gamma| \times |\gamma'|$ is the product scalar measure on $\mathbb{R}^d \times \mathbb{R}^d$; and χ_e denotes the characteristic function on $\mathbb{R}^d$ for the set $e \in B$. The presence of $\eta \cdot \eta'$ in the integrand of (2.10) means the function $\gamma * \gamma'(e)$ takes values in $\mathbb{C}$. Furthermore, it is not difficult to establish that $\gamma * \gamma'$ is a measure in $\mathfrak{M}(\mathbb{R}^d, \mathbb{C})$ and that

$$\|\gamma * \gamma'\| \leqslant \|\gamma\| \, \|\gamma'\|. \tag{2.11}$$

The scalar product in $\mathscr{F}^r$ is related to the scalar convolution in $\mathfrak{M}(\mathbb{R}^d, \mathbb{C}^r)$ by the identity

$$a(x) \cdot a'(x) = \int e^{i\alpha \cdot x} d\gamma * \gamma'. \tag{2.12}$$

Consider subsets in $\mathfrak{M}(\mathbb{R}^d, \mathbb{C}^r)$ consisting of measures γ that have compact support. Let $S_k \subset \mathbb{R}^d$ be the closed ball of radius $k > 0$ and center at $\alpha = 0$. Here $\mathfrak{M}(S_k, \mathbb{C}^r)$ will denote the subset of measures in $\mathfrak{M}(\mathbb{R}^d, \mathbb{C}^r)$ that have their support contained by S_k. With respect to the norm $\|\cdot\|$, $\mathfrak{M}(S_k, \mathbb{C}^r)$ is a Banach subspace of $\mathfrak{M}(\mathbb{R}^d, \mathbb{C}^r)$. Similarly, $\mathfrak{M}^*(S_k, \mathbb{C}^r)$ is defined to be the set of measures in $\mathfrak{M}(S_k, \mathbb{C}^r)$ that satisfy the reflection property. This set turns out to be a Banach subspace of $\mathfrak{M}(S_k, \mathbb{C}^r)$ with respect to the norm (2.2).

We have not yet indicated the manner in which the vector field a acquires a time dependence. If the measure in $\mathfrak{M}(\mathbb{R}^d, \mathbb{C}^r)$ is not static but varies as t is altered then a will become time dependent. Suppose $[0,T]$ is the relevant time interval for the study of the evolution problem. Denote by $\gamma(t)$ the values of a map

$$\gamma(\cdot): \ [0,T] \to \mathfrak{M}(\mathbb{R}^d, \mathbb{C}^r).$$

Thus $\gamma(\cdot)$ is a Banach space-valued function of t. We say $\gamma(\cdot)$ is *continuous in* $[0,T]$ if

$$\|\gamma(t') - \gamma(t)\| \to 0 \quad \text{as } t' \to t \tag{2.13}$$

for all $t \in [0,T]$. In a similar fashion $\gamma(t)$ is said to be *continuously differentiable in* $[0,T]$ if there is a measure $\dot\gamma(t) \in \mathfrak{M}(\mathbb{R}^d, \mathbb{C}^r)$ for all $t \in [0,T]$ such that $\dot\gamma(t)$ is continuous in $[0,T]$ and

$$\left\| \frac{\gamma(t') - \gamma(t)}{t' - t} - \dot\gamma(t) \right\| \to 0 \quad \text{as } t' \to t. \tag{2.14}$$

Observe that for each t, $\dot\gamma(t)$ is uniquely defined.

With this terminology we may state the hypotheses on the vector and scalar fields used throughout the remainder of this paper.

Vector field hypothesis: The vector field a: $\mathbb{R}^d \times [0,T] \to \mathbb{R}^d$ is said to be in the class $\mathscr{V}_v(k)$ $[a \in \mathscr{V}_v(k)]$ if a is the Fourier image, Eq. (1.13), of the time-dependent family of measures $\gamma(t)$ satisfying

(1) $\gamma(t) \in \mathfrak{M}^*(S_{k/2}, \mathbb{C}^d)$, $t \in [0,T]$ for $k < \infty$.

(2) $\gamma(t)$ is continuously differentiable on $[0,T]$.

Scalar field hypothesis: The scalar field v: $\mathbb{R}^d \times [0,T] \to \mathbb{R}$ is said to be in the class $\mathscr{V}_s(k)$ $[v \in \mathscr{V}_s(k)]$ if v is the Fourier image, Eq. (1.14), of the time-dependent family of measures $\nu(t)$ satisfying

(1) $\nu(t) \in \mathfrak{M}^*(S_k, \mathbb{C})$, $t \in [0,T]$ for $k < \infty$.

(2) $\nu(t)$ is continuously differentiable on $[0,T]$.

For the time continuous measures appearing above one can extract two useful constants that will often appear in our estimates, namely,

$$\gamma_T = \sup \|\gamma(t)\|, \quad \nu_T = \sup \|\nu(t)\|, \tag{2.15}$$

where each supremum is taken over $t \in [0,T]$.

Our primary motivation for invoking these hypotheses is that they allow us to carry through efficiently the explicit construction of the propagator K. For some parts of our investigation these hypotheses are unnecessarily restrictive. For example, the proof of the existence of the evolution operator $U(t,s)$ using the theory of differential equations in Banach space will succeed for a much broader class of vector and potential fields. However, we keep the hypotheses $\mathscr{V}_v(k)$ and $\mathscr{V}_s(k)$ throughout all sections of our analysis since the main goal is to construct the kernel (propagator) of $U(t,s)$.

Let us illustrate some of the properties of the fields a and v implied by assumptions $\mathscr{V}_v(k)$ and $\mathscr{V}_s(k)$. Because the measures $\gamma(t)$ and $\nu(t)$ have compact support it follows that $a(\cdot,t)$ and $v(\cdot,t)$ are C^∞ functions of x for each t. Convenient Fourier transform formulas exist for the various terms entering the Hamiltonian. For example,

$$\frac{1}{2m} a(x,t)^2 + v(x,t) = \int e^{i\alpha \cdot x} d\mu(t), \tag{2.16}$$

where

$$\mu(t) = (1/2m)\gamma(t) * \gamma(t) + \nu(t) \in \mathfrak{M}^*(S_k, \mathbb{C}). \tag{2.17}$$

Note that the definition of scalar convolution has the possibility of increasing the support of the associated convolution measure from $S_{k/2}$ to S_k. Time derivatives of the fields are also simple,

$$\frac{\partial}{\partial t} a(x,t) = \int e^{i\alpha \cdot x} d\dot\gamma(t) \tag{2.18}$$

for $t \in [0,T]$. In order to verify this formula and to establish the nature of the limiting process associated with the t derivative of a, let $\dot a(x,t)$ be defined as the integral on the right-hand side of (2.18). Then it follows that

$$\frac{a(x,t')-a(x,t)}{t'-t}-\dot{a}(x,t)$$
$$=\int e^{i\alpha\cdot x}d\left[\frac{\gamma(t')-\gamma(t)}{t'-t}-\dot{\gamma}(t)\right]. \tag{2.19}$$

By hypothesis $\mathscr{V}_v(k)$, the measure in the square brackets is an element of $\mathfrak{M}^*(S_{k/2},\mathbb{C}^d)$. Statement (2.19) employs the additivity of integration with respect to different measures in $\mathfrak{M}(\mathbb{R}^d,\mathbb{C}^r)$. Inequality (2.4) applies to (2.19) and yields

$$\left\|\frac{a(\cdot,t')-a(\cdot,t)}{t'-t}-\dot{a}(\cdot,t)\right\|_\infty$$
$$\leqslant\left\|\frac{\gamma(t')-\gamma(t)}{t'-t}-\dot{\gamma}(t)\right\|, \tag{2.20}$$

where $\|\cdot\|_\infty$ is the usual supremum norm of $\mathbb{R}^r$-valued functions on $\mathbb{R}^d$. Since $\dot{\gamma}(t)$ is the derivative of $\gamma(t)$, the right-hand side of (2.20) vanishes as $t'\to t$, for all $t\in[0,T]$. Similarly the t continuity of $\dot{\gamma}(t)$ implies the t continuity of $\dot{a}(\cdot,t)$ in the $\|\cdot\|_\infty$ norm. Thus it is seen that $a(\cdot,t)$ is continuously t differentiable on $[0,T]$ in the $\|\cdot\|_\infty$ norm. A similar conclusion is valid for $v(\cdot,t)$ if $v\in\mathscr{V}_s(k)$.

The t continuity of scalar measures also implies a joint $t_1,\ldots,t_n$ continuity for the product measure $\mu_1(t_1)\times\mu_2(t_2)\times\cdots\times\mu_n(t_n)$. In the case of two measures (2.8) and (2.9) yield

$$\|\mu_1(t_1')\times\mu_2(t_2')-\mu_1(t_1)\times\mu_2(t_2)\|$$
$$\leqslant\|\mu_1(t_1')\|\,\|\mu_2(t_2')-\mu_2(t_2)\|$$
$$+\|\mu_2(t_2)\|\,\|\mu_1(t_1')-\mu_1(t_1)\|. \tag{2.21}$$

Since the continuity of $\mu_i(t_i)$ with respect to t_i $(i=1,2)$ gives uniform boundedness of $\|\mu_1(t_1')\|$ and $\|\mu_2(t_2)\|$ on $[0,T]$, the above relation establishes the joint continuity of $\mu_1(t_1)\times\mu_2(t_2)$ with respect to $(t_1,t_2)\in[0,T]\times[0,T]$. This result obviously extends to the product of n t_i-continuous scalar measures $\mu_i(t_i)$, $i=1\sim n$.

III. HAMILTONIAN PROPERTIES

The complex mass evolution problem is generated by Hamiltonians that form a two-parameter family (in m and t) of operators. The theory of linear differential equations in Banach space, used in the next section, requires that these Hamiltonian operators be closed, have time-independent domain D_0, and be strongly continuously differentiable on D_0. For fields $a\in\mathscr{V}_v(k)$ and $v\in\mathscr{V}_s(k)$ we establish these features of the associated Hamiltonian family.

The mass parameter m is assumed to take values in the upper half complex plane. The open (closed) upper half-plane is denoted by $\mathbb{C}_>$ ($\mathbb{C}_\geqslant$). The symbol $\mathbb{C}_+$ will represent the closed upper half-plane with the origin deleted, $\mathbb{C}_+=\mathbb{C}_\geqslant\setminus\{0\}$. The value of Planck's constant plays no role in the section, so locally we set $\hbar=1$. Let $\widehat{H}_0(m)$ be the minimal operator with domain $C_0^\infty(\mathbb{R}^d)$ associated with the Laplacian in $\mathbb{R}^d$

$$\widehat{H}_0(m)=-(1/2m)\Delta. \tag{3.1}$$

For $m>0$ it is well known (Ref. 8, Chap. V) that $\widehat{H}_0(m)$ acting in the space $L^2(\mathbb{R}^d)$ is essentially self-adjoint with a self-adjoint closure, $H_0(m)$. Furthermore, the spectrum of $H_0(m)$ is $[0,\infty)$. For complex mass $m\in\mathbb{C}_>$, writing $\widehat{H}_0(m)=m^{-1}\widehat{H}_0(1)$ shows that $\widehat{H}_0(m)$ is closable with closure $H_0(m)=m^{-1}H_0(1)$. Let $\hat{f}$ be the Fourier transform of an arbitrary element $f\in L^2(\mathbb{R}^d)$. The domain of $H_0(m)$, for all $m\in\mathbb{C}_+$, is

$$D_0=\{f\in L^2(\mathbb{R}^d):\ \alpha^2\hat{f}(\alpha)\in L^2(\mathbb{R}^d)\}. \tag{3.2}$$

In a similar fashion, for $t\in[0,T]$ and $m\in\mathbb{C}_+$ define the minimal Hamiltonian operator with vector field $a\in\mathscr{V}_v(k)$ by

$$\widehat{H}_1(t,m)=(1/2m)(i\nabla+a(\cdot,t))^2, \tag{3.3}$$

where ∇ is the gradient and the domain of $\widehat{H}_1$ is $C_0^\infty(\mathbb{R}^d)$. After expanding the square, $\widehat{H}_1$ may be written as the sum

$$\widehat{H}_1(t,m)=\widehat{H}_0(m)+\widehat{W}(t,m), \tag{3.4}$$

where the perturbing operator [again having domain $C_0^\infty(\mathbb{R}^d)$] is

$$\widehat{W}(t,m)=\frac{i}{m}a(\cdot,t)\cdot\nabla+\frac{i}{2m}(\nabla\cdot a)(\cdot,t)$$
$$+\frac{1}{2m}a(\cdot,t)^2. \tag{3.5}$$

The fact that the field a is real, together with an integration by parts shows that for $m>0$, $\widehat{W}(t,m)$ is symmetric and thus closable. Denote the closure of $\widehat{W}(t,1)$ by $W(t,1)$. Noting that $\widehat{W}(t,m)=m^{-1}\widehat{W}(t,1)$ shows $W(t,m)=m^{-1}W(t,1)$ is the closure of $\widehat{W}(t,m)$ for all $m\in\mathbb{C}_+$.

Let $Y(t)$ denote the bounded operator obtained by multiplication with $v(x,t)$ for $v\in\mathscr{V}_s(k)$. The minimal Hamiltonian, on $C_0^\infty(\mathbb{R}^d)$, with both scalar and vector fields present takes two possible forms,

$$\widehat{H}_2(t,m)=\widehat{H}_1(t,m)+Y(t), \tag{3.6}$$
$$\widehat{H}_2(t,m)=\widehat{H}_0(m)+\widehat{V}(t,m), \tag{3.7}$$

where

$$\widehat{V}(t,m)=\widehat{W}(t,m)+Y(t). \tag{3.8}$$

Since $\widehat{W}$ is closable and Y bounded, $\widehat{W}+Y$ is closable, and its closure will be indicated by the symbol $V(t,m)$.

The basic closure, perturbation, and domain properties of $\widehat{H}_1$ and $\widehat{H}_2$ are the following.

Lemma 1: Suppose $a\in\mathscr{V}_v(k)$ and $v\in\mathscr{V}_s(k)$. For all $(t,m)\in[0,T]\times\mathbb{C}_+$ the operators $\widehat{H}_1(t,m)$ and $\widehat{H}_2(t,m)$ have closures $H_1(t,m)$ and $H_2(t,m)$ obeying

(a) $D(H_1(t,m))=D(H_2(t,m))=D_0$; (3.9)

(b) $H_1(t,m)f=H_0(m)f+W(t,m)f$, $f\in D_0$, (3.10)

$H_2(t,m)f=H_0(m)f+V(t,m)f$, $f\in D_0$, (3.11)

where $D(W(t,m))\supseteq D_0$ and $D(V(t,m))\supseteq D_0$; and

(c) if $m>0$ then $H_1(t,m)$ and $H_2(t,m)$ are self-joint and bounded from below. Specifically,

$$H_1(t,m)\geqslant 0,\quad H_2(t,m)\geqslant -v_T. \tag{3.12}$$

Proof: We show first that $\widehat{W}(t,m)$ is Kato tiny relative to $\widehat{H}_0(m)$. Since multiplication by either $\nabla\cdot a(x,t)$ or $a(x,t)^2$ gives rise to a bounded operator, it suffices to consider just the term $a(x,t)\cdot\nabla$ in $\widehat{W}(t,m)$. Take $f\in C_0^\infty(\mathbb{R}^d)$, and consider

$$\|a(\cdot,t)\cdot\nabla f\|^2$$

$$=\int|a(x,t)\cdot\nabla f(x)|^2\,dx$$

$$\leqslant\gamma_T^2\int|\nabla f(x)|^2\,dx=-\gamma_T^2\int\overline{f(x)}\Delta f(x)dx.$$

The inequality follows from (2.4) and (2.15), and the subsequent equality by an integration by parts. There are no surface contributions since $f\in C_0^\infty(\mathbb{R}^d)$. For arbitrarily small $\delta>0$, let $E_\delta=\{x\in\mathbb{R}^d\colon\delta^{-1}|f(x)|\geqslant\delta|\Delta f(x)|\}$, then

$$\|a(\cdot,t)\cdot\nabla f\|^2$$

$$\leqslant\gamma_T^2\int_{E_\delta}|\delta^{-1}f(x)|^2\,dx+\gamma_T^2\int_{\mathbb{R}^d\setminus E_\delta}|\delta\Delta f(x)|^2\,dx$$

$$\leqslant\gamma_T^2\delta^{-2}\|f\|^2+|2m|^2\gamma_T^2\delta^2\|\widehat{H}_0(m)f\|^2.$$

Since the coefficient of the second term can be made as small as we like, by the choice of δ, it follows that $\widehat{W}(t,m)$ has $\widehat{H}_0(m)$-bound 0 (or equivalently that $\widehat{W}$ is Kato tiny relative to $\widehat{H}_0$).

Let $m>0$, then $\widehat{W}(t,m)$ is symmetric. We employ the perturbation theorem (Ref. 8, Theorem 4.4, p. 288) which asserts that if operator T is essentially self-adjoint and operator A is symmetric and furthermore T-bounded with T-bound smaller than 1, then $T+A$ is essentially self-adjoint and its closure $(T+A)^\sim$ is equal to $\widetilde{T}+\widetilde{A}$. Here T is $\widehat{H}_0(m)$ and the symmetric perturbation A is $\widehat{W}(t,m)$. Thus $H_1(t,m)$ is self-adjoint and (3.9) and (3.10) are established, if $m>0$.

In order to extend these conclusions to the complex mass case, note the simple behavior of H_1 with respect to its mass dependence. From $\widehat{H}_1(t,m)=m^{-1}\widehat{H}_1(t,1)$ and the definition of closure it follows that (for $m\neq0$)

$$H_1(t,m)=m^{-1}H_1(t,1).\tag{3.13}$$

Identity (3.13) and the fact that $H_1(t,1)$ satisfies (3.9) and (3.10) extends the validity of (3.9) and (3.10) to all $m\in\mathbb{C}_+$.

Finally let us verify the bound statement in (c) for $H_1(t,m)$. We must show, for $m>0$, that $\langle f,H_1(t,m)f\rangle$ is non-negative ($f\in D_0$). If follows from the definition of closure that it will suffice to establish the inequality above for f on a core of $H_1(t,m)$. On the core C_0^∞, $\widehat{H}_1=H_1$. Starting from the definition of the minimal operator $\widehat{H}_1$ and integrating by parts gives

$$\langle f,\widehat{H}_1(t,m)f\rangle$$

$$=(1/2m)\|[i\nabla+a(\cdot,t)]f\|^2\geqslant0,\quad f\in C_0^\infty(\mathbb{R}^d).$$

So $H_1(t,m)$ is non-negative.

Thus statements (a), (b), and (c) are proved for the operators $\widehat{H}_1(t,m)$ and $H_1(t,m)$. The extension of these results to $\widehat{H}_2(t,m)$ and $H_2(t,m)$ is trivial since perturbation $Y(t)$ is a bounded operator,

$$\|Y(t)\|\leqslant\nu_T,$$

and $Y(t)$ is symmetric since $v(\cdot,t)$ is real valued. □

The next goal is to establish that $H_2(t,m)$ is strongly continuously differentiable in t. Our first observation is that $H_2(t,m)$ has a convenient representation in terms of the momentum operator P acting in the space $L^2(\mathbb{R}^d,\mathbb{C})$. In order to define P, introduce two Hilbert spaces $\mathscr{H}_s$ and $\mathscr{H}_v$. The first space is $\mathscr{H}_s=L^2(\mathbb{R}^d,\mathbb{C})$; and the second is that of the vector-valued (in $\mathbb{C}^d$) square integrable functions, $\mathscr{H}_v=L^2(\mathbb{R}^d,\mathbb{C}^d)$. The standard norms of these two spaces are indicated by $\|\cdot\|_s$ and $\|\cdot\|_v$, respectively. Let F be the unitary Fourier integral transform mapping $\mathscr{H}_s\to\mathscr{H}_s$. Define $P\colon\mathscr{H}_s\to\mathscr{H}_v$ by specifying each of its vector components,

$$P_jf=F^{-1}(\alpha_j\hat{f}),\quad j=1,\dots,d,$$

where $\hat{f}=Ff$ and $\alpha=(\alpha_1,\dots,\alpha_d)$ is the vector argument of $\hat{f}$. Clearly, P_j has the interpretation of the generalized derivative $-i\partial/\partial x_j$. The domain of operator P is

$$D(P)=\{f\in\mathscr{H}_s\colon\alpha\hat{f}\in\mathscr{H}_v\}=\{f\in\mathscr{H}_s\colon|\alpha|\hat{f}\in\mathscr{H}_s\}.$$

In terms of P the free Hamiltonian is

$$H_0(m)=(1/2m)P\cdot P,\quad m\in\mathbb{C}_+.$$

In order to complete the construction of $H_2(t,m)$ in terms of P define the operator family $\mathring{V}(t,m)\colon\mathscr{H}_s\to\mathscr{H}_s$ for parameters $t\in[0,T]$ and $m\in\mathbb{C}_+$ by

$$\mathring{V}(t,m)=-\frac{1}{m}a(\cdot,t)\cdot P+\frac{i}{2m}(\nabla\cdot a)(\cdot,t)$$

$$+\frac{1}{2m}a(\cdot,t)^2+v(\cdot,t)$$

with domain $D(\mathring{V}(t,m))=D(P)$. From the estimate

$$\|a(\cdot,t)\cdot Pf\|_s\leqslant\gamma_T\|Pf\|_v$$

it follows without difficulty that $\mathring{V}(t,m)$ has $H_0(m)$-bound less than 1. This implies (Ref. 8, Theorem 1.1, p. 190) that $\mathring{H}_2(t,m)$ given by

$$\mathring{H}_2(t,m)=H_0(m)+\mathring{V}(t,m)\tag{3.14}$$

is closed, with domain $D(\mathring{H}_2(t,m))=D_0$. But on C_0^∞, $\mathring{H}_2(t,m)$ is equal to $\widehat{H}_2(t,m)$. Thus both $H_2(t,m)$ and $\mathring{H}_2(t,m)$ are closed extensions of $\widehat{H}_2(t,m)$. Since $H_2(t,m)$ is the smallest closed extension, $H_2(t,m)\subseteq\mathring{H}_2(t,m)$. In fact, since the domains of these two operators are the same we have

$$H_2(t,m)=\mathring{H}_2(t,m).\tag{3.15}$$

In the lemma to follow we show that the operator $\dot{H}_2(t,m)$ defined by

$$\dot{H}_2(t,m)=-(1/m)\dot{a}(\cdot,t)\cdot P+(i/2m)(\nabla\cdot\dot{a})(\cdot,t)$$

$$+(1/m)\dot{a}(\cdot,t)\cdot a(\cdot,t)+\dot{v}(\cdot,t)\tag{3.16}$$

with domain taken to be D_0, is the t derivative of $H_2(t,m)$. The functions $\dot{a}(x,t)$ and $\dot{v}(x,t)$ are the t-partial derivatives of $a(x,t)$ and $v(x,t)$ and are most conveniently represented by the Fourier transforms of the measures $\dot{\gamma}(t)$ and $\dot{\nu}(t)$, cf. (2.18).

Lemma 2: In the interval $[0,T]$, for each $m\in\mathbb{C}_+$, $H_2(t,m)$ is strongly continuously differentiable on the domain D_0, and

$$\frac{d}{dt}H_2(t,m)=\dot{H}_2(t,m).\tag{3.17}$$

Proof: Take $f\in D_0$, let $\delta=t'-t\neq0$ and set

$$Df=\delta^{-1}[H_2(t',m)-H_2(t,m)]f-\dot{H}_2(t,m)f$$

$$=\delta^{-1}[\mathring{V}(t',m)-\mathring{V}(t,m)]f-\dot{H}_2(t,m)f,$$

where the second equality results from using (3.14) and (3.15). Now, define

$$s(x,t) = (i/2m)(\nabla \cdot a)(x,t) + (1/2m)a(x,t)^2 + v(x,t),$$

and

$$\dot{s}(x,t) = (i/2m)(\nabla \cdot \dot{a})(x,t) + (1/m)\dot{a}(x,t) \cdot a(x,t) + \dot{v}(x,t).$$

With this notation, we have the estimate

$$\|Df\|_s \leqslant (1/|m|)\|\delta^{-1}[a(\cdot,t') - a(\cdot,t)] - \dot{a}(\cdot,t)\|_\infty \|Pf\|_v + \|\delta^{-1}[s(\cdot,t') - s(\cdot,t)] - \dot{s}(\cdot,t)\|_\infty \|f\|_s.$$

The fact that $\dot{\gamma}(t)$ is the derivative in the sense of (2.14) of the measure $\gamma(t)$, ensures via inequality (2.20) that the first term on the right-hand side vanishes as $\delta \to 0$. Similar reasoning shows that the $\|\cdot\|_\infty$ portion of the second term also goes to 0 as $\delta \to 0$. Thus $\dot{H}_2(t,m)$ is shown to be the strong derivative on D_0 of $H_2(t,m)$. The strong continuity of $\dot{H}_2(t,m)$ is a result of formula (3.16) plus the fact that $\dot{\gamma}(t)$ and $\dot{v}(t)$ are continuous. □

The results establishing closure, domain stability, and the strong differentiability of $H_2(t,m)$ have been easy to obtain because $a(x,t)$ and $v(x,t)$ have such nice differentiability and boundedness properties. However, even when $a(x,t)$ and $v(x,t)$ are allowed to be in classes of functions that tolerate local singularities, $H_2(t,m)$ (for $m > 0$) is known to remain essentially self-adjoint. For recent results in this direction see Leinfelder and Simader.[13]

We conclude our discussion of Hamiltonian properties by stating resolvent growth estimates for $H_1(t,m)$ and $H_2(t,m)$.

For an operator A: $D(A) \subset \mathscr{H} \to \mathscr{H}$ and $z \in \rho(A)$ (the resolvent set of A) we denote the associated resolvent by

$$R(A,z) = (A - z)^{-1}.$$

Lemma 3: Let $m \in \mathbb{C}_+$ and $t \in [0,T]$.

(a) For $\omega > 0$, $i\omega$ is in the resolvent set of $H_1(t,m)$ and

$$\|R(H_1(t,m),i\omega)\| = \omega^{-1}. \tag{3.18}$$

(b) For $\omega > \nu_T$, $i\omega$ is in the resolvent set of $H_2(t,m)$ and

$$\|R(H_2(t,m),i\omega)\| \leqslant (\omega - \nu_T)^{-1}. \tag{3.19}$$

Proof: Suppose $m \in \mathbb{C}_+$ has the polar representation $m = e^{i\phi}|m|$, $\phi \in [0,\pi]$. For positive mass argument, H_1 is self-adjoint with spectrum $\Sigma = [0,\infty)$, thus the norm of the associated resolvent satisfies, for Re $z \leqslant 0$,

$$\|R(H_1(t,|m|),z)\| = [\mathrm{dist}(z,\Sigma)]^{-1} = |z|^{-1}. \tag{3.20}$$

However by identity (3.13) we also have

$$e^{i\phi}R(H_1(t,|m|),e^{i\phi}i\omega) = R(H_1(t,m),i\omega). \tag{3.21}$$

Together (3.20) and (3.21) imply (3.18).

Consider (b). If $\omega > \nu_T$, then

$$\|Y(t)R(H_1(t,m),i\omega)\| \leqslant \|Y(t)\|\ \|R(H_1(t,m),i\omega)\| \leqslant \nu_T/\omega < 1. \tag{3.22}$$

Thus the operator $1 + Y(t)R(H_1(t,m),i\omega)$ has a bounded inverse; $i\omega$ is in the resolvent set of $H_2(t,m)$; and

$$R(H_2,i\omega) = R(H_1,i\omega)[1 + Y(t)R(H_1,i\omega)]^{-1}. \tag{3.23}$$

Taking the operator norm of (3.23) and upon using (a) and (3.22) yields

$$\|R(H_2,i\omega)\| \leqslant \frac{1}{\omega}\frac{1}{1-\nu_T/\omega} = \frac{1}{\omega - \nu_T}. \qquad \square$$

It is clear that the hypothesis $v \in \mathscr{V}_s$ is the unnecessarily restrictive in Lemma 3. In essence the proof only requires that $v(x,t)$ is (for each $t \in [0,T]$) Kato tiny relative to $H_0(m)$.

IV. EVOLUTION OPERATORS

This section establishes the existence and describes the properties of the Schrödinger evolution operator. One version[6] of the general theory of linear differential equations in Banach space with unbounded operator coefficients is summarized. This theory is then adapted to the study of the Schrödinger problem (1.3) and (1.4).

The evolution problem of interest here is the following. Suppose E is a Banach space and $A(t): D(A(t)) \subset E \to E$ is an unbounded operator for all $t \in [0,T]$. Operators $A(t)$ are always assumed to be closed and to have a stable common domain $D(A(t)) = D(A)$ that is dense in E. The differential equation of evolution for time interval $[0,T]$ is

$$\frac{df}{dt} = A(t)f, \quad 0 \leqslant t \leqslant T. \tag{4.1}$$

We say that $f(t)$ is a *solution of equation* (4.1) on the interval $[s,T]$ $(0 \leqslant s \leqslant T)$ if $f(t)$ has values in $D(A)$, possesses a strong derivative $\dot{f}(t)$, and satisfies (4.1) on the segment $[s,T]$. The *Cauchy problem in the triangle* T_Δ is the problem of finding for each fixed $s \in [0,T]$ a solution $f(t,s)$ of (4.1) on segment $[s,T]$ that satisfies the initial data condition

$$f(s,s) = f_0 \in D(A). \tag{4.2}$$

Definition 3: The Cauchy problem (4.1) and (4.2) is said to be *uniformly correct* if the following statements hold.

(1) For each $s \in [0,T]$ and any $f_0 \in D(A)$ there exists a unique solution $f(t,s)$ of (4.1) on the segment $[s,T]$ satisfying initial data condition (4.2).

(2) The function $f(t,s)$ and its t derivative $\dot{f}(t,s)$ are continuous for t,s in the triangle T_Δ.

(3) The solution depends continuously on the initial data in the sense that if the $f_{0,n} \in D(A)$ converge to zero then the corresponding solutions $f_n(t,s)$ converge to zero uniformly relative to t and s in T_Δ.

In the circumstance where the Cauchy problem is uniformly correct, the evolution operator $\mathscr{U}(t,s)$ is defined for each, $t,s \in T_\Delta$ as the linear map $f_0 \to f(t,s)$, or

$$f(t,s) = \mathscr{U}(t,s)f_0. \tag{4.3}$$

From (1) and (3) it follows that $\mathscr{U}(t,s)$ is bounded and since the domain $D(A)$ is dense we may extend (by continuity) $\mathscr{U}(t,s)$ to the entire space E. Henceforth $\mathscr{U}$ will denote this extension. Also from (1) it is apparent that $\mathscr{U}(t,s)$: $D(A) \to D(A)$. Statement (2) means that $\mathscr{U}(t,s)$ is strongly continuously differentiable with respect to $t \in [s,T]$ on the domain $D(A)$. Similarly, collecting these and other useful inferences of Definition 3 gives us the following.

Proposition 1: Suppose that the Cauchy problem in the triangle T_Δ is uniformly correct, then the evolution operators $\mathscr{U}(t,s)$ satisfy the following.

(1) $\mathscr{U}(t,s): D(A) \to D(A)$, $t,s \in T_\Delta$.

(2) The operator $\mathscr{U}(t,s)$ is uniformly bounded in T_Δ.

(3) The operator $\mathscr{U}(t,s)$ is strongly continuous in T_Δ.

(4) The following operator-valued identities hold in T_Δ:

$$\mathscr{U}(t,s) = \mathscr{U}(t,\tau)\mathscr{U}(\tau,s), \quad 0 \leqslant s \leqslant \tau \leqslant t \leqslant T, \tag{4.4}$$

$$\mathscr{U}(s,s) = I, \quad s \in [0,T]. \tag{4.5}$$

(5) On the region $D(A)$ the operator $\mathscr{U}(t,s)$ is strongly differentiable in $t \in [s,T]$. Furthermore $\partial \mathscr{U}(t,s)/\partial t$ is jointly continuous for $t,s \in T_\Delta$ and obeys (in T_Δ)

$$\frac{\partial \mathscr{U}(t,s)}{\partial t} f = A(t)\mathscr{U}(t,s) f, \quad f \in D(A). \tag{4.6}$$

Proof: For details see Ref. 6 (pp. 193–195). □

We notice that Proposition 1 has a natural converse. Namely if $\mathscr{U}(t,s)$ fulfills conditions (1)–(5) then the Cauchy problem is uniformly correct. Thus it is of interest to know under what conditions an evolution operator having properties (1)–(5) will uniquely exist. Sufficient conditions for the existence of $\mathscr{U}(t,s)$ together with several additional properties of $\mathscr{U}$ are given in the theorem below.

Theorem 1: Suppose that the operators $A(t)$, $t \in [0,T]$ are

(1) densely defined and closed with a t-invariant domain $D(A)$;

(2) strongly continuously differentiable on domain $D(A)$; and

(3) obey the resolvent estimate

$$\|R(A(t),\lambda)\| \leqslant (1+\lambda)^{-1}, \quad \lambda \geqslant 0. \tag{4.7}$$

Then

(a) the Cauchy problem in T_Δ is uniformly correct;

(b) the evolution operator $\mathscr{U}(t,s)$ associated with the uniformly correct Cauchy problem is strongly continuously differentiable in the variable $s \in [0,T]$ on the domain $D(A)$, and satisfies

$$\frac{\partial \mathscr{U}(t,s)}{\partial s} f = -\mathscr{U}(t,s)A(s) f, \quad t,s \in T_\Delta, \quad f \in D(A); \tag{4.8}$$

and

(c) the evolution operator $\mathscr{U}(t,s)$ has the uniform bound (in T_Δ)

$$\|\mathscr{U}(t,s)\| \leqslant 1. \tag{4.9}$$

These results are found in Chap. 3 of Krein's book[6] (cf. Theorem 3.11). The proof of Theorem 1 is based on the determination of $\mathscr{U}(t,s)$ via the limit of an approximating sequence of operators that have an explicit construction. We do not display these approximating operators (and the associated proofs) since their detailed form is not required in the subsequent analysis.

We turn next to the problem of determining the Schrödinger evolution operators. In this case the Banach space E for evolution (4.1) is the Hilbert space $L^2(\mathbb{R}^d)$. The obvious choice of setting $A(t,m) = H(t,m)$ leads to the difficulty that the resolvent inequality (4.7) is not fulfilled. However, if we let (for all $m \in \mathbb{C}_+$)

$$A(t,m) = -(i/\hbar)H(t,m) - (1 + \nu_T/\hbar), \tag{4.10}$$

it now turns out that (4.7) is satisfied. To see this observe that

$$R(A(t,m),\lambda) = i\hbar R(H(t,m), i[\nu_T + \hbar(1+\lambda)]).$$

From Lemma 3(b)

$$\|R(A(t,m),\lambda)\| \leqslant \frac{\hbar}{[\nu_T + \hbar(1+\lambda)] - \nu_T} = \frac{1}{1+\lambda}, \quad \lambda \geqslant 0. \tag{4.11}$$

As defined by (4.10) the operators $A(t,m)$ (for each $m \in \mathbb{C}_+$) are closed, densely defined, and have t-invariant domain D_0 (Lemma 1). Furthermore $A(t,m)$ are strongly continuously differentiable in $t \in [0,T]$ on the domain D_0 (Lemma 2). Employing Theorem 1, it follows that the evolution problem (4.1) defined by $A(t,m)$ in (4.10) is uniformly correct (for all $m \in \mathbb{C}_+$). Summarizing, we have the following conclusion.

Theorem 2: For each $m \in \mathbb{C}^+$, the Cauchy problem (in T_Δ) generated by $A(t,m)$ is uniformly correct. Let $\mathscr{U}(t,s;m)$ denote the associated evolution operators for the family of generators $\{A(t,m): t \in [0,T]\}$ then

$$U(t,s;m) = e^{(1+\nu_T/\hbar)(t-s)}\mathscr{U}(t,s;m) \tag{4.12}$$

is the Schrödinger evolution operator [satisfying (1)–(5) of Definition 1].

Proof: In view of the remarks prior to the theorem it only remains to show (4.12). Substituting the right-hand side of (4.12) into (4.6) immediately leads to statement (1.10) of Definition 1. Equation (1.11) is a direct consequence of identity (4.8). Properties (1)–(4) of Definition 1 follow from results (1)–(4) in Proposition 1 and Theorem 1(c). Constant c of (1.7) is $1 + \nu_T/\hbar$. □

The Schrödinger evolution operator defined by (4.12) is dependent on the complex mass m. For all $m \in \mathbb{C}_+$, including the real axis boundary, $U(t,s;m)$ is m-strongly continuous. Specifically we have the following.

Proposition 2: Let $U(t,s;m)$ be the Schrödinger evolution operator of Theorem 2. For each $t,s \in T_\Delta$, $U(t,s;m)$ is strongly continuous in $\mathbb{C}_+$.

Proof: Let $f \in D_0$, and $m,m_1 \in \mathbb{C}_+$ and $m \neq m_1$. Formula (4.12) shows that U and $\mathscr{U}$ have the same continuity properties in m. The function $\mathscr{U}(t,\tau;m_1)\mathscr{U}(\tau,s;m) f$ is strongly differentiable in τ with the result

$$\begin{aligned}\frac{\partial}{\partial \tau}[\mathscr{U}(t,\tau;m_1)\mathscr{U}(\tau,s;m)] f &= \mathscr{U}(t,\tau;m_1)\frac{\partial \mathscr{U}(\tau,s;m)}{\partial \tau} f + \frac{\partial \mathscr{U}(t,\tau;m_1)}{\partial \tau}\mathscr{U}(\tau,s,m) f \\ &= \mathscr{U}(t,\tau;m_1)[A(\tau,m) - A(\tau,m_1)]\mathscr{U}(\tau,s;m) f.\end{aligned} \tag{4.13}$$

The second equality follows upon using Eqs. (4.6) and (4.8). For fixed t, s, m, and m_1 the right-hand side of (4.13) is strongly continuous for $\tau \in [s,t]$, thus we may integrate (4.13) with respect to τ and obtain

$$\mathscr{U}(t,s;m) f - \mathscr{U}(t,s;m_1) f = \int_s^t d\tau\, \mathscr{U}(t,\tau;m_1)[A(\tau,m) - A(\tau,m_1)]\mathscr{U}(\tau,s;m) f. \tag{4.14}$$

The integral here is the strong Riemann integral. The evaluation of the left-hand side of (4.14) has employed the initial condition identity (4.5).

Put $\delta = (1 - m/m_1)$. From the definition (4.10) of $A(t,m)$ a little algebra shows

$$A(\tau,m) - A(\tau,m_1) = \delta\{A(\tau,m) + iY(t)/\hbar + 1 + \nu_T/\hbar\}. \tag{4.15}$$

Upon substituting (4.15) into (4.14) and taking the norm of both sides, one obtains after utilizing estimate (4.9)

$$\|\mathscr{U}(t,s;m)f - \mathscr{U}(t,s,m_1)f\| \leqslant |\delta| \int_s^t d\tau [\|A(\tau,m)\mathscr{U}(\tau,s;m)f\| + (2\nu_T/\hbar + 1)\|f\|].$$

The integral on the right-hand side is finite and independent of m_1. Since $\delta \to 0$ as $m_1 \to m \in \mathbb{C}_+$ the inequality above shows that $\mathscr{U}(t,s;m)$ is m-strongly continuous on the domain $D_0 \subset L^2(\mathbb{R}^d)$. The uniform boundedness property (4.9) of $\mathscr{U}(t,s;m)$ with respect to $m \in \mathbb{C}_+$ and the fact D_0 is dense in $L^2(\mathbb{R}^d)$ suffices to show that $\mathscr{U}(t,s;m)$ is strongly continuous in $m \in \mathbb{C}_+$ on the domain $L^2(\mathbb{R}^d)$. □

Let us mention here that other, more general, versions of the evolution theory (in the strong sense on a fixed Banach space) are available (see Tanabe[14]) which allow one to obtain similar existence theorems for $\mathscr{U}(t,s)$ but under significantly less restrictive assumptions on $A(t)$. However, for our considerations Krein's[6] approach is sufficient and has the advantage that the hypotheses are easily verified. The only possible artificial condition seems to be the demand of continuous t differentiability of $H(t,m)$. However, this requirement does not seriously restrict the physics one can describe in this problem. In particular, recall that the determination of the electric field from the scalar and vector potentials requires the existence of $(\partial/\partial t)a(x,t)$.

The Schrödinger operators $U(t,s;m)$ have several additional basic properties, which we do not describe here because they are not needed in the sequel. First, the mass continuity is stronger than that stated in Proposition 2. In fact, $U(t,s;m)$ is an analytic function of m in $\mathbb{C}_>$. Secondly, in the case of quantum mechanical evolution, i.e., when $m>0$, $U(t,s;m)$ is unitary. A discussion of unitarity in a context similar to the one here is found in Dollard and Friedman.[15]

The statements of Theorem 2 and Proposition 2 may be extended to include scalar perturbations $v(x,t)$ that generate an operator $Y(t)$, that is (for each $t \in [0,T]$) Kato tiny relative to $H_0(m)$, and remains strongly continuously differentiable with respect to $t \in [0,T]$, on the domain D_0.

V. CONVERGENT DYSON SERIES

It is established in this section that the Dyson series is strongly convergent for a certain class of initial data functions and that the series sum defines an L^2 solution of the Cauchy problem (1.3) and (1.4).

First we summarize a number of notational conventions that will make more economical the description of the multiple integrals that appear in the Dyson series (1.17). For each integer n, let $\mathbf{t}_n = (t_1,...,t_n)$. The allowed domain of $\mathbf{t}_n$ for each $t,s \in T_\Delta$ we denote by $\Delta_n(t,s) = \{\mathbf{t}_n \colon s \leqslant t_1 \leqslant \cdots \leqslant t_n \leqslant t\}$. The nth order iterated time integration of series (1.17) will be abbreviated by

$$\int_< d\mathbf{t}_n = \int_s^t dt_n \int_s^{t_n} dt_{n-1} \cdots \int_s^{t_2} dt_1.$$

Several combinations of the measures $\gamma(t)$ and $\nu(t)$ occur repeatedly in representations of the kernels associated with the Dyson series. Suppose β is a fixed vector in $\mathbb{R}^d$ and $a \in \mathscr{V}_v(k)$. Define the set function [on the measure space $(\mathbb{R}^d, B)$] by

$$\mu_0(t,\beta)(e) = \int_e \beta \cdot \eta(t,\alpha) d|\gamma|(t) \quad (e \in B). \tag{5.1}$$

The measure μ_0 depends parametrically on t and β; $\alpha \in e$ is the variable of integration and $\eta(t,\alpha)$ is the vector-valued function defined by the polar decomposition (2.5) of $\gamma(t)$. Notice that $\mu_0(t,\beta)$ is a continuous $\mathfrak{M}(S_{k/2},\mathbb{C})$ valued function of the two variables β and t.

Two other measures that appear frequently we denote by μ_l^n and $\hat{\mu}_l^n$. Let $\boldsymbol{\alpha}_n = (\alpha_1,...,\alpha_n)$ be an n tuple of vectors in S_k. For each positive index $l \leqslant n$ and parameters t, $\boldsymbol{\alpha}_{l-1}$ define the set function

$$\mu_l^n(t,\alpha_1,...,\alpha_{l-1})(e) = \int_e \left(\frac{1}{2}\alpha + \sum_{j=1}^{l-1} \alpha_j\right) \cdot \eta(t,\alpha) d|\gamma|(t) \quad (e \in B).$$

Again, for each allowed $\boldsymbol{\alpha}_{l-1}$, $\mu_l^n(t,\alpha_1,...,\alpha_{l-1})$ is a t-continuous Banach space-valued function in $\mathfrak{M}(S_{k/2},\mathbb{C})$. A related measure $\hat{\mu}_l^n$ is given by

$$\hat{\mu}_l^n(t,\alpha_{l+1},...,\alpha_n)(e) = \int_e \left(\frac{1}{2}\alpha + \sum_{j=l+1}^{n} \alpha_j\right) \cdot \eta(t,\alpha) d|\gamma|(t) \quad (e \in B).$$

If $l=1$ the sum is absent in the expression for μ_l^n; and if $l=n$ the sum is missing in the integral for $\hat{\mu}_l^n$. As a consequence

$$\mu_1^n(t) = \mu_1^1(t) = \hat{\mu}_1^1(t) = \hat{\mu}_n^n(t). \tag{5.2}$$

In terms of the measures μ_l^n and $\hat{\mu}_l^n$ we define additional composite measures (for $l = 1,...,n$) by

$$\lambda_l^n(t) = \mu(t) - (\hbar/m)[\mu_0(t,\alpha_0) + \mu_l^n(t,\alpha_1,...,\alpha_{l-1})], \tag{5.3}$$

$$\hat{\lambda}_l^n(t) = \mu(t) - (\hbar/m)[\mu_0(t,\alpha) - \hat{\mu}_l^n(t,\alpha_{l+1},...,\alpha_n)], \tag{5.4}$$

where $\mu(t)$ is given by (2.17). Measure $\lambda_l^n(t)$ depends implicitly on the parameters $\alpha_0,\alpha_1,...,\alpha_{l-1}$, $\hbar$, and m as well as on the three explicit parameters n, l, and t. The measure $\hat{\lambda}_l^n(t)$ has implicit dependence on $\alpha,\alpha_{l+1},...,\alpha_n$, $\hbar$, and m. Note that we have changed, for reasons of subsequent utility, the vector parameter in measure μ_0 from α_0 to α as we go from (5.3) to (5.4).

Henceforth, in order to keep track of the different vector arguments appearing in the iterated integrals involving measures $\lambda_1^1,...,\lambda_n^n$ it is advantageous to introduce the variable of integration in the measure symbol. For measures $\lambda_l^n(t)$ and $\hat{\lambda}_l^n(t)$ the integration variable is most often $\alpha_l \in \mathbb{R}^d$. So if g is any integrable function on $\mathbb{R}^d$, we now write

$$\int g\, d\lambda_l^n(t) \quad \text{as} \quad \int g(\alpha_l)\, d\lambda_l^n(t;\alpha_l).$$

The family of measures λ_l^n and $\hat{\lambda}_l^n$ have simple norm bounds. Since it is assumed that $\alpha_j \in S_k$, $j = 1 \sim n$, it is evident that

$$\|\lambda_l^n(t)\| \leqslant \|\mu(t)\| + (\hbar/|m|)(|\alpha_0| + nk)\|\gamma(t)\| \quad (5.5)$$

and

$$\|\hat{\lambda}_l^n(t)\| \leqslant \|\mu(t)\| + (\hbar/|m|)(|\alpha| + nk)\|\gamma(t)\|. \quad (5.6)$$

Obviously, these bounds are uniform with respect to the parameters $\boldsymbol{\alpha}_n$ and the index l.

Finally, $\mathscr{S}(\mathbb{R}^d)$ shall indicate that Schwartz space of C^∞ functions of rapid decrease in $\mathbb{R}^d$. The normalization convention for the Fourier transform is taken to be

$$\hat{g}(\alpha) = \frac{1}{(2\pi)^{d/2}} \int e^{-ix\cdot\alpha} g(x)\,dx, \quad g \in \mathscr{S}(\mathbb{R}^d).$$

Lemma 4: Let $\phi \in \mathscr{S}(\mathbb{R}^d)$. Suppose $t,s \in T_\Delta$, $m \in \mathbb{C}_+$, and n is any positive integer.

(a) If $\tau_1,\tau_2 \geqslant 0$, the operator $\exp[i\tau_2 H_0(m)/\hbar]\, V(\tau_1,m)$ maps $\mathscr{S} \to \mathscr{S}$. Thus for each $\mathbf{t}_n \in \Delta_n(t,s)$, a set of functions $\{\psi_n\}$ in $\mathscr{S}$ is defined by the recurrence relation

$$\psi_n(t,s;\mathbf{t}_n) = \exp[-i(t - t_n)H_0(m)/\hbar] \times V(t_n,m)\psi_{n-1}(t_n,s;\mathbf{t}_{n-1}), \quad (5.7)$$

where

$$\psi_0(t,s) = \exp[-i(t-s)H_0(m)/\hbar]\phi. \quad (5.8)$$

(b) The value of $\psi_n(t,s;\mathbf{t}_n)$ at $x \in \mathbb{R}^d$ is given by the iterated integral

$$\psi_n(t,s;\mathbf{t}_n)(x) = \frac{1}{(2\pi)^{d/2}} \int d\alpha_0\, \hat{\phi}(\alpha_0) \times \left\{ \int d\lambda_1^n(t_1;\alpha_1) \cdots \int d\lambda_n^n(t_n;\alpha_n) \times \exp\left[ix \cdot \sum_{j=0}^n \alpha_j - \frac{i\hbar}{2m} \sum_{l,j=0}^n (t - t_l \vee t_j)\alpha_l \cdot \alpha_j \right]\right\}, \quad (5.9)$$

where $t_l \vee t_j = \mathrm{Max}(t_l,t_j)$ and $t_0 \equiv s$. If $n = 0$ the multiple integral symbol is replaced with 1.

(c) The Fourier transform of $\psi_n(t,s;\mathbf{t}_n)$ is the function $\hat{\psi}_n(t,s;\mathbf{t}_n) \in \mathscr{S}$, defined pointwise by

$$\hat{\psi}_n(t,s;\mathbf{t}_n)(\alpha) = \int d\hat{\lambda}_n^n(t_n;\alpha_n) \cdots \int d\hat{\lambda}_1^n(t_1;\alpha_1)\hat{\phi}\left(\alpha - \sum_{j=1}^n \alpha_j\right) \times \exp\left\{ -\frac{i\hbar}{2m}\left[(t - t_n)\alpha^2 + (t_n - t_{n-1}) \times (\alpha - \alpha_n)^2 + \cdots + (t_1 - s)\left(\alpha - \sum_{j=1}^n \alpha_j\right)^2\right]\right\}. \quad (5.10)$$

Proof: Since the action of the operators $V(t,m)$ [and $W(t,m)$] are the same on both C_0^∞ and $\mathscr{S}$, the differential form of V on $\mathscr{S}$ is given by (3.5) and (3.8). The fact that $a(\cdot,t)$ and $v(\cdot,t)$ are C^∞ functions of x, imply that $V(t,m)$ maps $\mathscr{S}$ into $\mathscr{S}$. Combining this with the fact that operator $\exp[-i\tau_2 H_0(m)/\hbar]$ is unitarily equivalent to multiplication by $\exp[-i\hbar\tau_2\alpha^2/2m]$ establishes (a).

Consider (b) and (c). First observe that the expression (5.9) for ψ_n is just the inverse Fourier transform of $\hat{\psi}_n$ in (5.10) followed by a change of variable $\alpha \to \alpha_0 = \alpha - \alpha_1 - \cdots - \alpha_n$ and use of the identity

$$(t - t_n)(\alpha_n + \dots + \alpha_0)^2 + \cdots + (t_1 - s)(\alpha_0)^2 = \sum_{l,j=0}^n (t - t_l \vee t_j)\alpha_l \cdot \alpha_j. \quad (5.11)$$

Thus it only remains to verify (5.10) for all $n \geqslant 0$. If $n = 0$ the right-hand side of (5.10) is just

$$\exp[-i\hbar(t-s)\alpha^2/2m]\hat{\phi}(\alpha)$$

and (5.9) results as the inverse Fourier transform. For $n \geqslant 1$, let $\tau_2,\tau_1 \geqslant 0$ and calculate the effect of $\exp[-i\tau_2 H_0(m)/\hbar]\, V(\tau_1,m)$ on an arbitrary element f of $\mathscr{S}$. The action of $V(\tau_1,m)$ on f is given by the sum of terms that appear in (3.5) and (3.8). The use of the integral representation (1.13) for $a(x,t)$ leads to

$$\left[\frac{i\hbar}{m} a(x,\tau_1) \cdot \nabla f\right](x) = \frac{-1}{(2\pi)^{d/2}} \frac{\hbar}{m} \int d\alpha_0 \hat{f}(\alpha_0) \times \int d|\gamma|(\tau_1;\alpha') e^{ix\cdot(\alpha_0+\alpha')}\alpha_0 \cdot \eta(\tau_1,\alpha').$$

Likewise one has

$$\left[\frac{i\hbar}{2m}(\nabla \cdot a(x,\tau_1))f\right](x) = \frac{-1}{(2\pi)^{d/2}} \frac{\hbar}{2m} \int d\alpha_0 \hat{f}(\alpha_0) \times \int d|\gamma|(\tau_1;\alpha') e^{ix\cdot(\alpha_0+\alpha')}\alpha' \cdot \eta(\tau_1,\alpha').$$

Summing this pair of formulas and adding the similar identity that describes the operator multiplication by $(2m)^{-1}a(x,\tau_1)^2 + v(x,\tau_1)$ [see Eq. (2.16)] gives

$$[V(\tau_1,m)f](x) = \frac{1}{(2\pi)^{d/2}} \int d\alpha_0 \hat{f}(\alpha_0) \left\{ \int d\lambda_1^1(\tau_1;\alpha') e^{ix\cdot(\alpha_0+\alpha')} \right\},$$

where λ_1^1 depends also on the parameter α_0, as in (5.3). A change of variables with $\alpha_0 = \alpha - \alpha'$ and reading off the Fourier transform from the last identity yields

$$[V(\tau_1,m)f]^\wedge(\alpha) = \int d\hat{\lambda}_1^1(\tau_1;\alpha')\hat{f}(\alpha - \alpha'),$$

where $\hat{\lambda}_1^1$ depends now on the parameter α, as in (5.4). Finally multiply by the operator $\exp[-i\tau_2 H_0(m)/\hbar]$ to obtain

$$\{\exp[-i\tau_2 H_0(m)/\hbar]\, V(\tau_1,m) f\}^\wedge(\alpha) = \int d\hat{\lambda}_1^1(\tau_1;\alpha')\exp\left[-\frac{i\hbar}{2m}\tau_2\alpha^2\right]\hat{f}(\alpha - \alpha'). \quad (5.12)$$

In representation (5.12) we are free to replace measure $\hat{\lambda}_1^1$ with $\hat{\lambda}_n^n$, since (5.2) shows that these two measures are identical. Note that measures $\hat{\lambda}_1^1$ and $\hat{\lambda}_n^n$ depend parametrically on α.

In order to obtain expression (5.10), just apply (5.12) iteratively, together with an appropriate induction argument. First, set $\tau_2 = t - t_1$, $\tau_1 = t_1$, $\alpha' = \alpha_1$, and $\hat{f}(\alpha) = \hat{\psi}_0(t_1,s)(\alpha)$ and observe that (5.12) is the result (5.10) for $n = 1$. Now assume (5.10) is valid for some $n-1 \geqslant 0$, put $\tau_2 = t - t_n$, $\tau_1 = t_n$, $\alpha' = \alpha_n$, $\hat{\lambda}^1_1 = \hat{\lambda}^n_n$ and further suppose

$$\hat{f}(\alpha) = \hat{\psi}_{n-1}(t_n,s;\mathbf{t}_{n-1})(\alpha)$$

in (5.12). Then after some careful algebra, it follows that (5.10) holds for n. $\square$

If the evolution problem is specialized, by setting $a(x,t) = 0$ or equivalently $\gamma(t) = 0$ then statements (b) and (c) of Lemma 4 simplify substantially. Measure definitions (5.3) and (5.4) reduce to

$$\lambda^n_l(t) = \hat{\lambda}^n_l(t) = \nu(t),$$

and formulas (5.9) and (5.10) are closely related to prior known results for the static scalar field problem [Ref. 4, Eq. (3.12); Ref. 5, Lemma 2].

The continuity properties of ψ_n and $\hat{\psi}_n$ follow from their defining integral expressions. It proves useful to label the region of allowed t, s, and $\mathbf{t}_n$ variation. Put $\Delta_n(T) = \{(t,s;\mathbf{t}_n): t,s\in T_\Delta, \mathbf{t}_n\in\Delta_n(t,s)\}$. Since $\psi_n\in\mathscr{S}$, it is in the domain of $H_0(m)$ and thereby $H_0(m)\psi_n$ is a well defined element of L^2. A pointwise construction of $H_0(m)\psi_n$ is available from the Fourier transform. The action of $H_0(m)$ on $\hat{\psi}_n$ is multiplication by $(\hbar^2/2m)\alpha^2$, or

$$[H_0(m)\psi_n(t,s;\mathbf{t}_n)]^\wedge(\alpha) = \hat{\Phi}_n(t,s;\mathbf{t}_n)(\alpha)$$
$$\equiv (\hbar^2/2m)\alpha^2\hat{\psi}_n(t,s;\mathbf{t}_n)(\alpha). \tag{5.13}$$

Having defined $\hat{\Phi}_n$ pointwise for each $\alpha\in\mathbb{R}^d$, it is evident that $\hat{\Phi}_n(t,s,\mathbf{t}_n)\in\mathscr{S}$ and also that $H_0(m)\psi_n(t,s;\mathbf{t}_n) = \Phi_n(t,s;\mathbf{t}_n)\in\mathscr{S}$.

Lemma 5: Let $\phi\in\mathscr{S}(\mathbb{R}^d)$, $m\in\mathbb{C}_+$ and $n\geqslant 1$.

(a) The $\mathscr{S}$-valued functions $\hat{\psi}_n(\cdot)$ and $\hat{\Phi}_n(\cdot)$ mapping $\Delta_n(T)\to L^p$ are continuous throughout (the compact domain) $\Delta_n(T)$ in the norm $\|\cdot\|_p$ for all $1\leqslant p\leqslant\infty$.

(b) The $\mathscr{S}$-valued functions $\psi_n(\cdot)$ and $\Phi_n(\cdot)$ mapping $\Delta_n(T)\to L^q$ are continuous throughout $\Delta_n(T)$ in the norm $\|\cdot\|_q$ for all $2\leqslant q\leqslant\infty$.

Proof: Observe first that statement (b) is an immediate consequence of (a). The Hausdorff–Young theorem for Fourier transforms asserts (Ref. 16, p. 11)

$$\|\psi\|_q \leqslant (2\pi)^{d(2^{-1}-q^{-1})}\|\hat{\psi}\|_p,$$

where $p^{-1}+q^{-1} = 1$ and $1\leqslant p\leqslant 2$. In our context this means that if $\hat{\psi}_n(\cdot)$: $\Delta_n(T)\to L^p$ is continuous in $\|\cdot\|_p$ then so is its inverse Fourier image $\psi_n(\cdot)$: $\Delta_n(T)\to L^q$ in norm $\|\cdot\|_q$ where $2\leqslant q\leqslant\infty$. The same remark applies to $\hat{\Phi}_n(\cdot)$.

The proof of (a) is based on the formula (5.10) for $\hat{\psi}_n$. In order to carry out the required estimates it is useful to revise (5.10) somewhat. Let us scale the measure $\hat{\lambda}^n_l$ by dividing by $(1+|\alpha|+nk)$. So we define

$$\tilde{\lambda}^n_l(t_l) \equiv (1+|\alpha|+nk)^{-1}\hat{\lambda}^n_l(t_l), \quad l = 1\sim n. \tag{5.14}$$

From (5.14), (5.6), (2.15), and (2.17) a bound for the total variation of $\tilde{\lambda}^n_l$ is obtained. If we set

$$\mu_T \equiv (1/2|m|)\gamma_T^2 + \nu_T,$$

then

$$\|\tilde{\lambda}^n_l(t_l)\| \leqslant \mu_T + (\hbar/|m|)\gamma_T. \tag{5.15}$$

This bound is clearly uniform in α, $\alpha_{l+1},\ldots,\alpha_n$, and t_l. Furthermore, note that the convergence of $\tilde{\lambda}^n_l(t'_l)\to\tilde{\lambda}^n_l(t_l)$ as $t'_l\to t_l$ also has an α, $\alpha_{l+1},\ldots,\alpha_n$ uniform character, since

$$\|\tilde{\lambda}^n_l(t'_l) - \tilde{\lambda}^n_l(t_l)\|$$
$$\leqslant \|\mu(t'_l) - \mu(t_l)\| + (\hbar/|m|)\|\gamma(t'_l) - \gamma(t_l)\|. \tag{5.16}$$

Consider the proof of (a) for $\hat{\Phi}_n$. Choose two different points from $\Delta_n(T)$, say t', s', $\mathbf{t}'_n$ and t, s, $\mathbf{t}_n$, then put

$$D\hat{\Phi}_n = \hat{\Phi}_n(t',s';\mathbf{t}'_n) - \hat{\Phi}_n(t,s;\mathbf{t}_n).$$

It proves convenient to let $f(t,s;\mathbf{t}_n)$ be the exponent factor in (5.10), i.e.,

$$f(t,s;\mathbf{t}_n) = \exp\left[-\frac{i\hbar}{2m}\sum_{l,j=0}^{n}(t - t_l\vee t_j)\alpha_l\cdot\alpha_j\right], \tag{5.17}$$

where $\alpha_0 = \alpha - \alpha_1 - \cdots - \alpha_n$. Formulas (5.13) and (5.10) allow us to write $D\hat{\Phi}_n$ as the sum of two parts

$$D\hat{\Phi}_n = D_1\hat{\Phi}_n + D_2\hat{\Phi}_n,$$

$$D_1\hat{\Phi}_n(\alpha) = \left[\int d\tilde{\lambda}^n_n(t'_n;\alpha_n)\cdots\int d\tilde{\lambda}^n_1(t'_1;\alpha_1)\right.$$
$$\left.-\int d\tilde{\lambda}^n_n(t_n;\alpha_n)\cdots\int d\tilde{\lambda}^n_1(t_1;\alpha_1)\right](\hbar^2/2m)$$
$$\times f(t',s';\mathbf{t}'_n)\alpha^2(1+|\alpha|+nk)^n\hat{\phi}(\alpha_0),$$

$$D_2\hat{\Phi}_n(\alpha) = \int d\tilde{\lambda}^n_n(t_n;\alpha_n)\cdots\int d\tilde{\lambda}^n_1(t_1;\alpha_1)$$
$$\times[f(t',s';\mathbf{t}'_n) - f(t,s;\mathbf{t}_n)]$$
$$\times(\hbar^2/2m)\alpha^2(1+|\alpha|+nk)^n\hat{\phi}(\alpha_0).$$

Investigate $D_2\hat{\Phi}_n$ first. The allowed argument of the exponential f always falls in the right-half complex plane. If $z_1,z_2\in\mathbb{C}$ with $\operatorname{Re} z_1\geqslant 0$, $\operatorname{Re} z_2\geqslant 0$, recall the standard bound

$$|e^{-z_1} - e^{-z_2}| \leqslant |z_1 - z_2|.$$

Applying this inequality to the difference of two f's gives the estimate

$$|f(t',s';\mathbf{t}'_n) - f(t,s;\mathbf{t}_n)|$$
$$\leqslant |m|^{-1}\hbar(|\alpha|+nk)^2$$
$$\times[|t'-t| + |t'_n - t_n| + \cdots + |s'-s|]$$

for all $\alpha\in\mathbb{R}^d$ and $\alpha_j\in S_k$.

The function $\hat{\phi}\in\mathscr{S}$, thus there is a constant $C_\phi<\infty$ such that

$$(1+|\alpha|+nk)^{n+4}\left|\hat{\phi}\left(\alpha - \sum_{j=1}^{n}\alpha_j\right)\right| < C_\phi(1+|\alpha|)^{-d-1} \tag{5.18}$$

for all $\alpha\in\mathbb{R}^d$, $\alpha_j\in S_k$. Combining the two previous inequalities with (5.15) and the definition of $D_2\hat{\Phi}_n$ leads to the estimate

$$|D_2\hat{\Phi}_n(\alpha)| < \frac{1}{2}\hbar^3|m|^{-2}C_\phi\left(\mu_T + \frac{\hbar}{m}\gamma_T\right)^n(1+|\alpha|)^{-d-1}$$
$$\times\left[|t'-t| + |s'-s| + \sum_{j=1}^{n}|t'_j - t_j|\right].$$

The function $(1+|\alpha|)^{-d-1}$ is in L^p for all $1\leqslant p\leqslant\infty$, thereby we obtain that $D_2\hat{\Phi}_n\to 0$ in the $\|\cdot\|_p$ norm as $t',s',\mathbf{t}'_n\to t,s,\mathbf{t}_n$.

To complete the study of $D\hat{\Phi}_n$, turn to $D_1\hat{\Phi}_n$. In this term the difference in the values of $\hat{\Phi}_n$ at t', s', $\mathbf{t}'_n$ and t, s, $\mathbf{t}_n$ arise solely from the different measures in the integrals entering $D_1\hat{\Phi}_n$. The formula (5.10) for $\hat{\psi}_n$ is defined as an n-fold iterated integral. However, this iterated integral may be replaced by a multiple integral on the measure space $(\mathbb{R}^d_n\times\cdots\times\mathbb{R}^d_1, B\times\cdots\times B)$ with respect to the scalar-valued product measure $\tilde{\lambda}^n_n(t_n)\times\cdots\times\tilde{\lambda}^n_1(t_1)$. In defining this product measure one should recall that the conventional construction of the product measures from the component measures, $\tilde{\lambda}^n_i(t_i)$, does not allow the variable, say α_i, of one measure to be a parameter in another measure as it is here for $\tilde{\lambda}^n_l(t_l)$ for $l<i$. This difficulty is easily circumvented by expanding out the measures $\tilde{\lambda}^n_i$ according to their definition in (5.4) and (5.14). After this is done $\tilde{\lambda}^n_n(t_n)\times\cdots\times\tilde{\lambda}^n_1(t_1)$ may be expressed as a sum of product measures that takes the form of a string whose elements are either $|\gamma|(\cdot)$ or $\mu(\cdot)$, e.g., $|\gamma|(t_n)\times|\gamma|(t_{n-1})\times\mu(t_{n-2})\times\cdots\times|\gamma|(t_2)\times\mu(t_1)$. This product string is such that the parameters are only the $\mathbf{t}_n$ and so the variables of each of the constituent measures are independent. From (2.21) it is evident that the product measure is continuous for $\mathbf{t}_n\in\Delta_n(t,s)$ in the total variation norm.

An addition and subtraction of cross terms of the form $\tilde{\lambda}^n_n(t_n)\times\cdots\times\tilde{\lambda}^n_j(t_j)\times\tilde{\lambda}^n_{j-1}(t'_{j-1})\times\cdots\times\tilde{\lambda}^n_1(t'_1)$ together with Eq. (2.9) and estimates (5.15) and (5.16) shows that

$$\begin{aligned}\|\tilde{\lambda}^n_n(t'_n)\times\cdots\times\tilde{\lambda}^n_1(t'_1)-\tilde{\lambda}^n_n(t_n)\times\cdots\times\tilde{\lambda}^n_1(t_1)\|\\ \leqslant\Big(\mu_T+\frac{\hbar}{|m|}\gamma_T\Big)^{n-1}\sum_{j=1}^{n}\Big[\|\mu(t'_j)-\mu(t_j)\|\\ +\frac{\hbar}{|m|}\|\gamma(t'_j)-\gamma(t_j)\|\Big].\end{aligned}\tag{5.19}$$

Using inequality (5.18) and the bound $|f(t',s,\mathbf{t}'_n)|\leqslant 1$ gives us the pointwise estimate

$$\begin{aligned}|D_1\hat{\Phi}_n(\alpha)|\leqslant\frac{\hbar^2}{2|m|}\frac{C_\phi}{(1+|\alpha|)^{d+1}}\Big(\mu_T+\frac{\hbar}{|m|}\gamma_T\Big)^{n-1}\\ \times\sum_{j=1}^{n}\Big[\|\mu(t'_j)-\mu(t_j)\|\\ +\frac{\hbar}{|m|}\|\gamma(t'_j)-\gamma(t_j)\|\Big].\end{aligned}$$

Since $(1+|\alpha|)^{-d-1}$ is in $L^p(d\alpha)$ for all $1\leqslant p\leqslant\infty$ it is seen that $\|D_1\hat{\Phi}_n\|_p\to 0$ as $t',s',\mathbf{t}'_n\to t,s,\mathbf{t}_n$.

Furthermore note that $\hat{\psi}_n$ differs from $\hat{\Phi}_n$ only by the multiplicative factor $(\hbar^2/2m)\alpha^2$. Thus the argument above is easily modified to show the continuity of $\hat{\psi}_n$ in $\|\cdot\|_p$. □

The next task is to give a precise meaning to the time ordered multiple integrals that appear in the Dyson expansion (1.17). This may be done either in an abstract manner on the space L^q or in a pointwise sense. The lemma below establishes the equivalence of these two approaches.

Part (b) of Lemma 5 asserts the strong continuity in $\mathbf{t}_n$ of the function $\psi_n(t,s;\cdot):\Delta_n(t,s)\to L^q$, $2\leqslant q\leqslant\infty$. This property makes it possible to define a family (for $t,s,m\in T_\Delta\times\mathbb{C}_+$) of linear operators $D_n(t,s;m):\mathscr{S}\to L^q$ by the abstract L^q integral,

$$D_n(t,s;m)\phi=\Big(\frac{-i}{\hbar}\Big)^n\int_< d\mathbf{t}_n\ \psi_n(t,s;\mathbf{t}_n).\tag{5.20}$$

The integral above is the Riemann integral on the n-dimensional domain $\Delta_n(t,s)$ which converges in the $\|\cdot\|_q$ topology. The variability of q means that $D_n(t,s;m):\mathscr{S}\to L^2\cap L^\infty$.

Lemma 6: Let $\phi\in\mathscr{S}(\mathbb{R}^d)$, $m\in\mathbb{C}_+$, and $n\geqslant 1$. Suppose $t,s\in T_\Delta$.

(a) For each $x\in\mathbb{R}^d$, $\psi_n(t,s;\cdot)(x):\Delta_n(t,s)\to\mathbb{C}$ is Riemann integrable on the domain $\Delta_n(t,s)$ and for almost all x

$$[D_n(t,s;m)\phi](x)=\Big(\frac{-i}{\hbar}\Big)^n\int_< d\mathbf{t}_n\ \psi_n(t,s;\mathbf{t}_n)(x).\tag{5.21}$$

(b) For each $\alpha\in\mathbb{R}^d$, $\hat{\psi}_n(t,s;\cdot)(\alpha):\Delta_n(t,s)\to\mathbb{C}$ is Riemann integrable on the domain $\Delta_n(t,s)$ and for almost all α

$$[D_n(t,s;m)\phi]^\wedge(\alpha)=\Big(\frac{-i}{\hbar}\Big)^n\int_< d\mathbf{t}_n\ \hat{\psi}_n(t,s;\mathbf{t}_n)(\alpha).\tag{5.22}$$

Proof: Relation (5.21) asserts that the abstract Riemann integral in (5.20) and the numerically valued Riemann integral on the right-hand side of (5.21) construct the same L^q element. For fixed n, t, s, and m the function $\psi_n(t,s;\cdot)(x)$ is [by Lemma 5(b), with $q=\infty$] known to be continuous for $\mathbf{t}_n\in\Delta_n(t,s)$ uniformly with respect to $x\in\mathbb{R}^d$. Thus $\psi_n(t,s;\cdot)(x)$ is Riemann integrable on the n-dimensional domain $\Delta_n(t,s)$, and thereby one has that the numerically-valued Riemann integral on the right-hand side of (5.21) is well defined for each x. The almost everywhere equality (5.21) is obtained as a direct consequence of the defining convergence criteria for vector-valued and the numerically valued Riemann integrals. For a more general characterization of this type of equivalence see Ref. 17 (p. 69).

Finally observe that (b) is the result of the Fourier transform (in the Plancherel sense) of (a). Specifically,

$$\begin{aligned}[D_n(t,s;m)\phi]^\wedge(\alpha)=\lim_{L\to\infty}\frac{1}{(2\pi)^{d/2}}\int_{|x|<L}dx\ e^{i\alpha\cdot x}\\ \times\Big\{\Big(\frac{-i}{\hbar}\Big)^n\int_< d\mathbf{t}_n\ \psi_n(t,s;\mathbf{t}_n)(x)\Big\}.\end{aligned}$$

Fubini's theorem applies and justifies interchanging the $dx\,d\mathbf{t}_n$ order of integration. Dominated convergence together with Lemma 5(a) and Lemma 4(c) then gives result (5.22). □

In preparation for showing that the sum over n of $D_n(t,s;m)\phi$ constructs a solution of the complex mass Schrödinger evolution equation we require a statement of the recurrence relation that the operators D_n obey. In the following we need the standard theorem justifying the interchange of operator order and strong-Riemann integration. Let A be a closed operator on $L^2(\mathbb{R}^d)$. Suppose $f(\boldsymbol{\tau})$ is an L^2-valued function on the compact region $\Delta\subset\mathbb{R}^n$. If both $f(\boldsymbol{\tau})$ and $[Af](\boldsymbol{\tau})$ are strongly continuous in Δ, then

$$A\int_\Delta f(\boldsymbol{\tau})d\boldsymbol{\tau}=\int_\Delta[Af](\boldsymbol{\tau})d\boldsymbol{\tau}.\tag{5.23}$$

Both integrals here are the strong-Riemann integral. The identity is a straightforward consequence of the closedness of A and the definition of the strong-Riemann integral.

Lemma 7: Let $\phi\in\mathscr{S}(\mathbb{R}^d)$, $m\in\mathbb{C}_+$, $s\in[0,T]$, and $n\geqslant 0$, then for all $t\in[s,T]$.

(a) $D_n(t,s;m)\phi\in D_0$.

(b) $D_n(t,s;m)\phi$ is strongly continuously differentiable with respect to $t\in[s,T]$ with derivative denoted by $\dot{D}_n(t,s;m)\phi$.

(c) The recurrence relation (where $D_{n-1}=0$, if $n=0$)

$$i\hbar\dot{D}_n(t,s;m)\phi=H_0(m)D_n(t,s;m)\phi+V(t,m)D_{n-1}(t,s;m)\phi \tag{5.24}$$

holds.

Proof: Property (a) is satisfied if $\alpha^2[D_n(t,s;m)]^\wedge(\alpha)$ as in L^2. But from (5.13) and (5.22) one has the pointwise representation

$$\frac{\hbar^2}{2m}\alpha^2[D_n(t,s;m)\phi]^\wedge(\alpha)=\left(\frac{-i}{\hbar}\right)^n\int_< d\mathbf{t}_n\,\hat{\Phi}_n(t,s;\mathbf{t}_n)(\alpha). \tag{5.25}$$

Recall that $\hat{\Phi}_n(t,s;\cdot):\Delta_n(t,s)\to L^2$ is $\|\cdot\|_2$ continuous for all $\mathbf{t}_n\in\Delta_n(t,s)$. Similarly the pointwise function $\hat{\Phi}_n(t,s;\cdot)(\alpha)$: $\Delta_n(t,s)\to\mathbb{C}$ is continuous in absolute value norm $|\cdot|$ for all $\mathbf{t}_n\in\Delta_n(t,s)$ and each $\alpha\in\mathbb{R}^d$. [Lemma 5, part (a) with $p=2$ and $p=\infty$, respectively.] Thus the L^2 and the numerical Riemann integrals of $\hat{\Phi}_n$ are equivalent for almost all α, i.e.,

$$\int_< d\mathbf{t}_n\,\hat{\Phi}_n(t,s;\mathbf{t}_n)(\alpha)=\left[\int_< d\mathbf{t}_n\,\hat{\Phi}_n(t,s;\mathbf{t}_n)\right](\alpha).$$

So the right-hand side integral in (5.25) defines an element of L^2. Furthermore because of the $\|\cdot\|_2$ continuity of $\hat{\Phi}_n$ in $\Delta_n(T)$, it is seen that (5.25) defines a $\|\cdot\|_2$ continuous function of $t\in[s,T]$.

Consider (b) and (c) together. We introduce an L^2-valued auxiliary function of $\delta>0$ and $\mathbf{t}_n\in\Delta_n(t,s)$ (for some fixed s and t),

$$R(t,s;\delta;\mathbf{t}_n)\equiv\psi_n(t+\delta,s;\mathbf{t}_n)-\psi_n(t,s;\mathbf{t}_n)+(i/\hbar)\delta H_0(m)\psi_n(t,s;\mathbf{t}_n).$$

Obviously, from the known properties of ψ_n and $H_0\psi_n$ this is an L^2-continuous function of $\mathbf{t}_n\in\Delta_n(t,s)$. So we may write

$$\Delta_+(\delta)\phi\equiv(1/\delta)[D_n(t+\delta,s;m)-D_n(t,s;m)]\phi,$$

where (for all $\delta>0$)

$$\begin{aligned}\Delta_+(\delta)\phi=\frac{1}{\delta}\Bigg\{&\delta\left(\frac{-i}{\hbar}\right)^{n+1}\int_{\Delta_n(t,s)}d\mathbf{t}_n\,H_0(m)\psi_n(t,s;\mathbf{t}_n)\\&+\left(\frac{-i}{\hbar}\right)^n\int_{\Delta_n(t+\delta,s)}d\mathbf{t}_n\,\psi_n(t+\delta,s;\mathbf{t}_n)\\&-\left(\frac{-i}{\hbar}\right)^n\int_{\Delta_n(t,s)}d\mathbf{t}_n\,\psi_n(t+\delta,s;\mathbf{t}_n)\\&+\left(\frac{-i}{\hbar}\right)^n\int_{\Delta_n(t,s)}d\mathbf{t}_n\,R(t,s;\delta;\mathbf{t}_n)\Bigg\}.\end{aligned} \tag{5.26}$$

Since $H_0(m)$ is closed, identity (5.23) means that the first term on the right-hand side of (5.26) is equal to

$$\left(\frac{-i}{\hbar}\right)^{n+1}H_0(m)\int_{\Delta_n(t,s)}d\mathbf{t}_n\,\psi_n(t,s;\mathbf{t}_n)=\frac{-i}{\hbar}H_0(m)D_n(t,s;m)\phi. \tag{5.27}$$

The contribution of the R integral is also easy to evaluate. The operators $H_0(m)$, $\exp[-i\delta H_0(m)/\hbar]$, and 1 are all closed. Equations (5.7) and (5.23) then yield

$$\delta^{-1}\left(\frac{-i}{\hbar}\right)^n\int_{\Delta(t,s)}d\mathbf{t}_n\,R(t,s;\delta;\mathbf{t}_n)=\delta^{-1}[e^{-i\delta H_0(m)/\hbar}-1+i\delta H_0(m)/\hbar]D_n(t,s;m)\phi. \tag{5.28}$$

By part (a), $D_n(t,s;m)\phi\in D_0$. Since $e^{-i\delta H_0(m)/\hbar}$ is strongly differentiable with respect to δ on D_0 with right derivative $-iH_0(m)/\hbar$ at $\delta=0$, the $\|\cdot\|_2$ norm of the right-hand side of (5.28) vanishes as $\delta\to 0^+$.

It remains to consider the middle two terms on the right-hand side of (5.26). The difference in value of these two terms arises from the different domains of integration. Note that

$$\int_{\Delta_n(t+\delta,s)\setminus\Delta_n(t,s)}d\mathbf{t}_n=\int_t^{t+\delta}dt_n\int_{\Delta_{n-1}(t_n,s)}d\mathbf{t}_{n-1}.$$

Defining,

$$g(t_n)\equiv\int_{\Delta_{n-1}(t_n,s)}d\mathbf{t}_{n-1}\,\psi_n(t,s;\mathbf{t}_n) \tag{5.29}$$

we have upon multiplication by $(i\hbar)^n$

$$\begin{aligned}\delta^{-1}\int_t^{t+\delta}dt_n&\int_{\Delta_{n-1}(t_n,s)}d\mathbf{t}_{n-1}\,\psi_n(t+\delta,s;\mathbf{t}_n)\\&=\delta^{-1}\int_t^{t+\delta}dt_n\,g(t_n)+\delta^{-1}\int_t^{t+\delta}dt_n\\&\quad\times\left\{\int_{\Delta_{n-1}(t_n,s)}d\mathbf{t}_{n-1}[\psi_n(t+\delta,s;\mathbf{t}_n)-\psi_n(t,s;\mathbf{t}_n)]\right\}.\end{aligned} \tag{5.30}$$

Lemma 5 demonstrated that ψ: $\Delta_n(T)\to L^2$ is continuous throughout the compact domain $\Delta_n(T)$. Thus ψ is uniformly continuous in this domain and in particular for any $\epsilon>0$ there is a δ_0 independent of t, s, $\mathbf{t}_n$ such that for all $\delta<\delta_0$

$$\|\psi(t+\delta,s;\mathbf{t}_n)-\psi(t,s;\mathbf{t}_n)\|_2<\epsilon.$$

This suffices to show that the $\|\cdot\|_2$ norm of the right-most integral in (5.30) vanishes as $\delta\to 0^+$.

The integral over $g(t_n)$ may be evaluated as follows. Application of the mean-value theorem for the strong-Riemann integral (Ref. 18, Proposition 2.4.7) states that

$$\frac{1}{\delta}\int_t^{t+\delta}dt_n\,g(t_n)=\bar{g}_{H(t,t+\delta)},$$

where $\bar{g}$ is some element of the closed convex hull of the set of values of $g(t_n)$ on the interval $[t,t+\delta]$. In our context g is defined by integral (5.29) and is a strongly continuous function of $t_n\in[s,t]$. Thus as $\delta\to 0^+$ one has

$$\text{s-lim}\,\frac{1}{\delta}\int_t^{t+\delta}dt_n\,g(t_n)=\text{s-lim}\,\bar{g}_{H(t,t+\delta)}=g(t).$$

Using 5.7 and upon noting that $V(t,m)$ is closed it is found that

$$\left(\frac{-i}{\hbar}\right)^n g(t) = \left(\frac{-i}{\hbar}\right)^n \int_{\Delta_{n-1}(t,s)} d\,\mathbf{t}_{n-1}\, \psi_n(t,s;\mathbf{t}_{n-1},t)$$
$$= \frac{-i}{\hbar} V(t,m) D_{n-1}(t,s;m)\phi$$

is the $\delta \to 0^+$ L^2-limiting value of the right-hand side of (5.30).

Combining these conclusions establishes that $D_n(t,s;m)\phi$ is right differentiable in t and obeys

$$i\hbar \frac{\partial^+}{\partial t} D_n(t,s;m)\phi$$
$$= H_0(m) D_n(t,s,m)\phi + V(t,m) D_{n-1}(t,s;m)\phi.$$

In order to obtain the statement (c) of the lemma, recall that [in the analysis of (a)] $H_0(m)D_n(t,s;m)\phi$ was shown to be a continuous L^2-valued function of t. Furthermore, the function $g(t)$ or equivalently $V(t,m)D_{n-1}(t,s;m)\phi$ is also known to be a continuous L^2-valued function of t. The fact that $D_n(t,s;m)\phi$ is right continuously differentiable in t for every point of $[s,T]$ suffices (Ref. 6, p. 4) to show that $D_n(t,s;m)\phi$ is continuously differentiable in t with respect to norm $\|\cdot\|_2$. □

We conclude this section by establishing that the sum over n of the functions $D_n(t,s;m)\phi$ given by (5.20) constitutes a solution of the Schrödinger evolution problem (1.3) and (1.4). Showing convergence of these sums compels us to restrict the class of initial data functions to $\hat{\phi}\in C_0^\infty(\mathbb{R}^d)$—the compactly supported C^∞ functions in momentum space. In the next section we will show how this class of functions may be enlarged.

Proposition 3: Assume $a\in\mathscr{V}_v(k)$ and $v\in\mathscr{V}_s(k)$. Let $\hat{\phi}\in C_0^\infty(\mathbb{R}^d)$. For $t,s\in T_\Delta$, $m\in\mathbb{C}_+$ define

$$t^* = s + |m|/ek\gamma_T \tag{5.31}$$

and

$$\Psi_N(t,s;m) = \sum_{n=0}^{N} D_n(t,s;m)\phi. \tag{5.32}$$

(a) If $t\in[s,t^*)$ the sum over $n = 1 \sim \infty$ of $\|D_n(t,s;m)\phi\|$, $\|H_0(m)D_n(t,s;m)\phi\|$, $\|V(t,m)D_n(t,s;m)\phi\|$, and $\|\dot{D}_n(t,s;m)\phi\|$ are all finite.

(b) If $t\in[s,t^*)$ the $L^2(\mathbb{R}^d)$-valued function of t, s, and m given by $\Psi(t,s;m) = \lim_{N\to\infty}\Psi_N(t,s;m)$ has range D_0; is t-strongly continuously differentiable; obeys the Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\Psi(t,s;m) = H(t,m)\Psi(t,s;m); \tag{5.33}$$

and satisfies the initial condition

$$\Psi(s,s;m) = \phi. \tag{5.34}$$

Proof: Consider the sum of $\|H_0(m)D_n(t,s;m)\phi\|$. From (5.10), (5.17), and (5.22) we have the Fourier representation

$$I_n(\alpha) = [H_0(m)D_n(t,s;m)\phi]^\wedge(\alpha),$$

where

$$I_n(\alpha) = \frac{\hbar^2\alpha^2}{2m}\left(\frac{-i}{\hbar}\right)^n \frac{1}{(2\pi)^{d/2}}$$
$$\times\int_< d\,\mathbf{t}_n \int d\hat{\lambda}_n^n(t_n;\alpha_n)\cdots$$
$$\times\int d\hat{\lambda}_1^n(t_1;\alpha_1) f(t,s;\mathbf{t}_n)\hat{\phi}(\alpha_0).$$

Since $\hat{\phi}$ has compact support in $\mathbb{R}^d$ there is a positive $b<\infty$ such that this support is contained within the ball S_{bk}. Recall $\alpha_0 = \alpha - \Sigma_{j=1}^n \alpha_j$; thus if $|\alpha| \geqslant (n+b)k$ the support criteria for $\hat{\phi}$ implies $|\hat{\phi}(\alpha_0)| = 0$ for all $\alpha_j\in S_k$, $j = 1\sim n$. As a consequence I_n has its support inside the ball $S_{(n+b)k}$.

A pointwise bound for $I_n(\alpha)$ results from using $|f|\leqslant 1$ and the inequality

$$\int d\,|\hat{\lambda}_j^n(t_j;\alpha_j)| \leqslant \mu_T + \frac{\hbar}{|m|}(n+b)k\gamma_T.$$

One readily finds

$$|I_n(\alpha)| \leqslant \frac{[\hbar k(n+b)]^2}{2|m|(2\pi)^{d/2}}\|\hat{\phi}\|_\infty \frac{1}{n!}$$
$$\times\left[\left(\frac{t-s}{\hbar}\right)\left(\mu_T + \frac{\hbar}{|m|}(n+b)k\gamma_T\right)\right]^n.$$

In view of the compact support of I_n on $S_{(n+b)k}$, this estimate implies that I_n has finite L^2 norm. For some constant finite C, independent of n,

$$\|I_n\| \leqslant C\frac{[(n+b)k]^{2+d/2}}{n!}$$
$$\times\left[\left(\frac{t-s}{\hbar}\right)\left(\mu_T + \frac{\hbar}{|m|}(n+b)k\gamma_T\right)\right]^n. \tag{5.35}$$

An application of the ratio test to the sum over $n = 1\sim\infty$ of the terms on the right-hand side of (5.35) shows that the series converges if $t - s < |m|(ek\gamma_T)^{-1}$. This implies, under the same restriction on t, that

$$\sum_{n=0}^{\infty}\|H_0(m)D_n(t,s;m)\phi\| < \infty.$$

A similar argument shows the sum over n of $\|D_n(t,s;m)\phi\|$ is finite.

The operator $V(t,m)$ is $H_0(m)$-bounded, so the convergences established above imply that the sum over n of $\|V(t,m)D_n(t,s;m)\phi\|$ is also finite. Finally, because of identity (5.24), the sum of $\|\dot{D}_n(t,s;m)\phi\|$ is seen also to converge. Lastly, note that all these series of norms are uniformly convergent for t in compact subsets of $[s,t^*)$.

Consider (b). First recall the standard interchange of limit theorem for derivatives. Let $\{h_N(t)\}$ be a sequence of L^2-valued strongly continuously differentiable functions on $[a,b]$, which converge in norm $\|\cdot\|$ to $h(t)$ as $N\to\infty$. Suppose further that $\{\partial h_N(t)/\partial t\}$ is a sequence which on $[a,b]$ uniformly converges in norm $\|\cdot\|$ to $g(t)$. Then $h(t)$ is continuously differentiable in norm $\|\cdot\|$ and

$$\frac{\partial}{\partial t}h(t) = g(t), \quad t\in[a,b]. \tag{5.36}$$

Here h_N corresponds to $\Sigma_{n=0}^N D_n(t,s;m)\phi$, and $\partial h_N(t)/\partial t$ is $\Sigma_{n=0}^N \dot{D}_n(t,s;m)\phi$. Thus we have

$$\frac{\partial}{\partial t}\Psi(t,s;m) = \sum_{n=0}^{\infty} \dot{D}_n(t,s;m)\phi, \quad t\in[s,t^*-\delta].$$

To continue note that $H_0(m)$ is closed, and that both $\Sigma_{n=0}^N D_n(t,s;m)\phi$ and $\Sigma_{n=0}^N H_0(m)D_n(t,s;m)\phi$ are Cauchy sequences in L^2. Thus the definition of closure gives

$$H_0(m)\Psi(t,s;m) = \sum_{n=0}^{\infty} H_0(m)D_n(t,s;m)\phi.$$

The operator $V(t,m)$ is also closed and is $H_0(m)$ bounded. Arguing as above shows

$$V(t,m)\Psi(t,s;m) = \sum_{n=0}^{\infty} V(t,m)D_{n-1}(t,s;m)\phi.$$

Combining the last three identities with recurrence relation (5.24) proves (5.33).

The initial condition (5.34) follows from the uniform strong convergence in t of the series for $\Psi(t,s;m)$ together with the fact that as $t\to s$, $D_n(t,s;m)\phi$ goes strongly to zero while $D_0(t,s;m)\phi$ goes strongly to ϕ. □

VI. COMPLEX MASS PROPAGATOR AND ITS BOUNDARY VALUE

The goal in this section is to prove that, for masses with a positive imaginary part, $\Sigma_{n=0}^{\infty} D_n(t,s;m)$ is a bounded integral operator. Growth estimates for the kernels of $D_n(t,s;m)$ are found which allow one to sum over the index n in order to obtain a candidate kernel for the evolution operator $U(t,s;m)$ of Theorem 2. The uniqueness property of the uniformly correct Cauchy problem is then used to establish that this candidate kernel is the integral kernel of $U(t,s;m)$. Finally the continuity in mass is utilized to extend this kernel representation to the real mass axis.

We begin by defining the measures that enter the ensuing kernel representations. In analogy with Eq. (5.3) set

$$\rho_l(t) = \mu(t) - (\hbar/m)\mu_l^n(t,\alpha_1,...,\alpha_{l-1}), \tag{6.1}$$

where the measure of $\rho_l\in\mathfrak{M}(S_k,\mathbb{C})$ and depends parametrically on $t,\alpha_1,...,\alpha_{l-1},\hbar$, and m. Notice that the parameter α_0 found in λ_l^n is now absent. This is natural for the representations given below since the α_0 dependence is removed by a Fourier transform. The total variation norm of $\rho_l(t)$ has the bound

$$\|\rho_l(t)\|\leqslant\mu_T + (\hbar/|m|)nk\gamma_T \quad (l\leqslant n),$$

which is uniform with respect to all the parameters not appearing on the right-hand side of the estimate.

Let n be the order of the Dyson iterate. For each positive index $r\leqslant n$, put $\mathbf{j}_r = (j_1,j_2,...,j_r)$ and define the ordered index set

$$J_{n,r} = \{\mathbf{j}_r\text{:}\ 1\leqslant j_1 < j_2 < \cdots < j_r \leqslant n\}.$$

There are $\binom{n}{r}$ elements $\mathbf{j}_r$ in $J_{n,r}$. To each $\mathbf{j}_r$ we associate a measure in the product space $(\mathbb{R}_1^d\times\cdots\times\mathbb{R}_n^d, B\times\cdots\times B)$,

$$\Lambda^n(\mathbf{j}_r,\mathbf{t}_n) = \rho_1(t_1)\times\cdots\times|\gamma|(t_{j_1})\times\cdots\times|\gamma|(t_{j_r})\times\cdots\times\rho_n(t_n).$$

The right-hand side of this equality is to be understood in the following sense. If $r=0$, the measure involves only the product of $\rho_i(t_i)$ for $i=1\sim n$. In the case where $r>0$ and $\mathbf{j}_r = (j_1,...,j_r)$ then the j_i term of the product for the $r=0$ case has the element $\rho_{j_i}(t_{j_i})$ replaced with $|\gamma|(t_{j_i})$, $i=1\sim r$. It is evident that Λ^n has the uniform bound [for $\mathbf{j}_r\in J_{n,r}$ and $\mathbf{t}_n\in\Delta_n(0,T)$]

$$\|\Lambda^n(\mathbf{j}_r;\mathbf{t}_n)\|\leqslant\Big(\mu_T + \frac{\hbar}{|m|}nk\gamma_T\Big)^{n-r}\gamma_T^r. \tag{6.2}$$

In order to prepare for Lemma 9, we require a certain product derivative formula. The greatest integer less than or equal to $r/2$ is denoted by $[r/2]$.

Lemma 8: Suppose $\{\eta_i\}_1^r$ is a set of r vectors in $\mathbb{R}^d$. If $x\in\mathbb{R}^d$ and $z\in\mathbb{C}$ the formula

$$\begin{aligned}(\eta_1\cdot\nabla)\cdots(\eta_r\cdot\nabla)e^{z|x|^2} &= e^{z|x|^2}\sum_{l=0}^{[r/2]}(2z)^{r-l}\sum_{r,l}{}'(x\cdot\eta_{i_1})\cdots(x\cdot\eta_{i_{r-2l}})\\ &\quad\times(\eta_{i_{r-2l+1}}\cdot\eta_{i_{r-2l+2}})\cdots(\eta_{i_{r-1}}\cdot\eta_{i_r})\end{aligned} \tag{6.3}$$

holds. Here $\Sigma'_{r,l}$ represents the sum over all divisions of the elements $\{\eta_i\}_1^r$ into particular subsets. For a given l, a particular division is obtained when $r-2l$ vectors from $\{\eta_i\}_1^r$ are chosen (indexed by $i_1\sim i_{r-2l}$ these vectors enter the $r-2l$ scalar products with x), followed by the selection of l pairs from the remaining $2l$ elements [indexed by $i_{r-2l+1}\sim i_r$, each pair forms one of the scalar products of η_i on the right-hand side of (6.3)]. For a given r and l, $\Sigma'_{r,l}$ is the sum over all distinct choices of this type. There are $r![2^l(r-2l)!l!]^{-1}$ terms in the sum $\Sigma'_{r,l}$.

The one-dimensional case ($d=1$) of formula (6.3) is in Ref. 19 (p. 20, 0.432.2). An inductive proof of the general case ($d>1$) which is combinatorial in nature may be found.

Lemma 9: Assume $a\in\mathscr{V}_v(k)$ and $v\in\mathscr{V}_s(k)$. Let $t,s\in T_\Delta$, $(t\neq s)$, and $n\geqslant 0$.

(a) If $m\in\mathbb{C}_>$ then the operator $D_n(t,s;m)$: $\mathscr{S}\to L^2\cap L^\infty$ is an integral operator with a jointly continuous Carleman kernel $d_n(t,\cdot;s\cdot;m)$: $\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{C}$ such that for almost all x

$$[D_n(t,s;m)\phi](x) = \int d_n(t,x;s,y;m)\phi(y)dy, \quad \phi\in\mathscr{S}. \tag{6.4}$$

(b) For $m\in\mathbb{C}_>$ the pointwise value of d_n is given by

$$\begin{aligned}&d_n(t,x;s,y;m)\\ &= \Big(\frac{-i}{\hbar}\Big)^n\Big(\frac{m}{2\pi i\hbar(t-s)}\Big)^{d/2}\\ &\quad\times\sum_{r=0}^{n}\sum_{J_{n,r}}\Big(\frac{-i\hbar}{m}\Big)^r\int_< d\mathbf{t}_n\int d\Lambda^n(\mathbf{j}_r,\mathbf{t}_n)\\ &\quad\times\exp\Big[ix\cdot\sum_{p=1}^{n}\alpha_p - \frac{i\hbar}{2m}\sum_{l,p=1}^{n}(t-t_l\vee t_p)\alpha_l\cdot\alpha_p\Big]\\ &\quad\times\prod_{i=1}^{r}[\eta(t_{j_i},\alpha_{j_i})\cdot\nabla_y]\exp\Big[\frac{im}{2\hbar(t-s)}(X_n-y)^2\Big],\end{aligned} \tag{6.5}$$

where $X_n\in\mathbb{C}^d$ and

$$X_n = x - \frac{\hbar}{m}\sum_{p=1}^{n}(t-t_p)\alpha_p. \tag{6.6}$$

The integral in (6.5) remains well defined if $m\in\mathbb{R}\setminus\{0\}$. In this way the definition of d_n is extended to $\mathbb{C}_+$.

(c) For $m \in \mathbb{C}_+$, d_n has the estimate (for all $x,y \in \mathbb{R}^d \times \mathbb{R}^d$)

$$|d_n(t,x;s,y;m)| < \left(\frac{|m|}{2\pi\hbar(t-s)}\right)^{d/2} \left[\frac{2ek(t-s)}{|m|}\gamma_T\right]^n \times \exp\frac{1}{2\hbar}\left[\frac{-\operatorname{Im} m}{t-s}(x-y)^2 + c_1|x-y| + c_2\right], \tag{6.7}$$

where

$$c_1 = \frac{|m|}{k(t-s)}, \quad c_2 = \frac{|m|}{2k^2(t-s)} + \frac{|m|\mu_T}{k\gamma_T}.$$

Proof: Taken together Lemmas 6(a) and 4(b) provide (for $m \in \mathbb{C}_+$) the pointwise representation

$$[D_n(t,s;m)\phi](x) = \left(\frac{-i}{\hbar}\right)^n \int_< d\,\mathbf{t}_n \left\{\frac{1}{(2\pi)^{d/2}}\int d\alpha_0\,\hat{\phi}(\alpha_0) \times \int d\lambda_1^n(t_1;\alpha_1)\cdots\int d\lambda_n^n(t_n;\alpha_n) \times f(t,s;\mathbf{t}_n)\exp\left[ix\cdot\sum_{j=0}^n \alpha_j\right]\right\},$$

where f is the exponential function given by (5.17). The region of $\mathbf{t}_n$ integration has finite volume; $|f| \leqslant 1$, $\hat{\phi} \in \mathscr{S}$ and in view of the inequality (5.5) it follows the multiple integral above is absolutely convergent. An application of Fubini's theorem justifies the interchange of the $d\alpha_0$ and the $d\,\mathbf{t}_n$ order of integration. Thus for almost all x

$$[D_n(t,s;m)\phi](x) = \int \hat{d}_n(t,x;s,\alpha_0;m)\hat{\phi}(\alpha_0)d\alpha_0, \tag{6.8}$$

where (for $n \geqslant 1$)

$$\hat{d}_n(t,x;s,\alpha_0;m) = \left(\frac{-i}{\hbar}\right)^n \frac{1}{(2\pi)^{d/2}}\int_< d\,\mathbf{t}_n \times \int d\lambda_1^n(t_1;\alpha_1)\cdots\int d\lambda_n^n(t_n;\alpha_n) \times f(t,s;\mathbf{t}_n)\exp\left[ix\cdot\sum_{j=0}^n \alpha_j\right]. \tag{6.9}$$

If $n = 0$, the multiple integral is absent and instead one has

$$\hat{d}_0(t,x;s,\alpha_0;m) = \frac{1}{(2\pi)^{d/2}}\exp\left[\frac{-i\hbar}{2m}(t-s)\alpha_0^2 + ix\cdot\alpha_0\right].$$

Expression (6.9) implies a pointwise bound for $|\hat{d}_n|$. Since $|\alpha_i| \leqslant k$, $i = 1 \sim n$, the inequality

$$\sum_{l,j=1}^n (t - t_l \vee t_j)\alpha_l\cdot\alpha_j \geqslant 0 \tag{6.10}$$

valid for all $\mathbf{t}_n \in \Delta_n(t,s)$ and all $\boldsymbol{\alpha}_n$ (Ref. 5, Lemma 5), gives the estimate

$$|f(t,s;\mathbf{t}_n)| \leqslant h_n(\alpha_0) \equiv \exp\left\{\operatorname{Im}\left(\frac{\hbar}{2m}\right)[\alpha_0^2(t-s) - 2nk|\alpha_0|(t-s)]\right\}.$$

Inserting this inequality into (6.9), it is found that

$$|\hat{d}_n(t,x;s,\alpha_0;m)| \leqslant \frac{1}{(2\pi)^{d/2}}\left(\frac{(t-s)}{\hbar}\right)^n \frac{h_n(\alpha_0)}{n!} \times \left[\mu_T + \frac{\hbar}{|m|}(nk + |\alpha_0|)\gamma_T\right]^n. \tag{6.11}$$

At this point we restrict m and insist that $\operatorname{Im} m > 0$. In this case it is seen that the bound (6.11) is in the form of an nth order polynomial in $|\alpha_0|$ times a Gaussian function. Since $t - s > 0$ and $\operatorname{Im}(1/m) < 0$, this Gaussian is decaying. Thus the bound function on the right-hand side of (6.11) belongs to $L^1(d\alpha_0) \cap L^\infty(d\alpha_0)$. Furthermore, note that $\hat{\phi} \in L^2(d\alpha_0)$, thereby the integral of (6.8) may be interpreted as the inner product of two $L^2(d\alpha_0)$ functions. Parseval's relation for L^2 inner products lets one write (6.8) in the form of (6.4) where the function d_n is defined as the Fourier transform of $\hat{d}_n$. Specifically

$$d_n(t,x;s,y;m) = \frac{1}{(2\pi)^{d/2}}\int e^{-iy\cdot\alpha_0}\hat{d}_n(t,x;s,\alpha_0;m)d\alpha_0. \tag{6.12}$$

Because $\hat{d}_n(t,x;s,\cdot\,;m) \in L^1(d\alpha_0)$, this Fourier integral exists absolutely. In particular the Hausdorff–Young theorem for Fourier transforms (Ref. 16, p. 11), shows that $d_n(t,x;s,\cdot\,;m) \in L^2(dy) \cap L^\infty(dy)$. This establishes that the function d_n is a Carleman kernel. So part (a) of the lemma (with exception of the x,y joint continuity of d_n) is demonstrated.

Statement (b) is the consequence of an explicit evaluation of the integral (6.12). In representation (6.9) we use the expansion

$$\lambda_l^n(t_l) = \rho_l(t_l) - (\hbar/m)\mu_0(t_l,\alpha_0), \tag{6.13}$$

which follows from (5.3) and (6.1). Inserting (6.9) into (6.12) and employing (6.13) yields

$$d_n = \frac{1}{(2\pi)^d}\left(\frac{-i}{\hbar}\right)^n \sum_{r=0}^n \left(\frac{-\hbar}{m}\right)^r \sum_{\mathbf{j}_r \in J_{r,n}} \int d\alpha_0 \int_< d\,\mathbf{t}_n \times \int d\Lambda^n(\mathbf{j}_r,\mathbf{t}_n)f(t,s;\mathbf{t}_n)\left[\prod_{i=1}^r \alpha_0\cdot\eta(t_{j_i},\alpha_{j_i})\right] \times \exp\left[i(x-y)\cdot\alpha_0 + ix\cdot\sum_{p=1}^n \alpha_p\right].$$

Absolute convergence of the multiple integral holds if $\operatorname{Im} m > 0$. After changing orders of integration the d_n integral is

$$d_n = \left(\frac{-i}{\hbar}\right)^n \sum_{r=0}^n \left(\frac{-\hbar}{m}\right)^r \sum_{\mathbf{j}_r \in J_{r,n}} \int_< d\,\mathbf{t}_n \int d\Lambda^n(\mathbf{j}_r,\mathbf{t}_n) \times f_1(t,s;\mathbf{t}_n)\exp\left[ix\cdot\sum_{p=1}^n \alpha_p\right]\mathscr{F}(\mathbf{j}_r,\mathbf{t}_n)(\boldsymbol{\alpha}_n), \tag{6.14}$$

where

$$f_1(t,s;\mathbf{t}_n) = \exp\left[\frac{-i\hbar}{2m}\sum_{i,p=1}^n (t - t_i \vee t_p)\alpha_i\cdot\alpha_p\right]$$

and

$$\mathscr{F}(\mathbf{j}_r,\mathbf{t}_n)(\boldsymbol{\alpha}_n) = \frac{1}{(2\pi)^d}\int d\alpha_0\Big[\prod_{i=1}^{r}\alpha_0\cdot\eta(t_{j_i},\alpha_{j_i})\Big] \times\exp\Big[\frac{-i\hbar}{2m}(t-s)\alpha_0^2 - i\alpha_0\cdot(y-X_n)\Big]. \quad (6.15)$$

X_n is the complex vector (6.6). If $r=0$, then the product of $\alpha_0\cdot\eta$ factors in (6.15) is replaced by 1.

All the integrals in (6.15) are $L^1(d\alpha_0)$, $(\operatorname{Im} m>0)$, and may be evaluated in closed form. For $r=0$, the integral (6.15) is of the Fresnel type. It is evaluated by completing the square in the exponential. In this fashion one obtains

$$\mathscr{F}(\mathbf{j}_0,\mathbf{t}_n)(\boldsymbol{\alpha}_n) = \Big(\frac{m}{2\pi i\hbar(t-s)}\Big)^{d/2}\exp\Big(\frac{im}{2\hbar(t-s)}(X_n-y)^2\Big).$$

If X_n is replaced by x the right-hand side is $d_0(x,t;y,s;m)$. The square root here is the one that maps the cut plane $\mathbb{C}\setminus(-\infty,0]$ onto the right half complex plane. For $r>0$, the integrand has a polynomial in the components of α_0, and again can be found in closed form,

$$\mathscr{F}(\mathbf{j}_r,\mathbf{t}_n)(\boldsymbol{\alpha}_n) = \Big(\frac{m}{2\pi i\hbar(t-s)}\Big)^{d/2}\prod_{i=1}^{r}[\eta(t_{j_i},\alpha_{j_i})\cdot i\nabla_y] \times\exp\Big(\frac{im}{2\hbar(t-s)}(X_n-y)^2\Big). \quad (6.16)$$

Combining (6.16) and (6.14) gives expression (6.5) for d_n in the case where $\operatorname{Im} m>0$.

For each fixed $x,y,t>s$ formula (6.5) defines a holomorphic function of m in the half-plane $\mathbb{C}_>$. For m in any compact subset of $\mathbb{C}_>$, expression (6.5) is a uniformly convergent integral of a holomorphic function of m. If the range of m is enlarged to the region $\mathbb{C}_+$, then (for fixed $x,y,t>s$) the integral (6.5) defines a continuous function of m. So $d_n(t,x;s,y;m)$ is an analytic function in $\mathbb{C}_>$ and continuous on $\mathbb{C}_+$. The joint x,y continuity of d_n arises similarly. For x,y in any compact subset of $\mathbb{R}^d\times\mathbb{R}^d$ the integral (6.5) is uniformly convergent (in x,y) and has an integrand that is x,y jointly continuous in this compact region. This conclusion holds for all $m\in\mathbb{C}_+$.

Consider (c). A detailed understanding of the behavior of the exponential functions in (6.5) is necessary for an optimal estimate. First, note that

$$f_1(t,s;\mathbf{t}_n)\exp\Big[\frac{im}{2\hbar(t-s)}(X_n-y)^2\Big] = f_2(t,s;\mathbf{t}_n)\exp\Big[\frac{im(x-y)^2}{2\hbar(t-s)} - i\frac{(x-y)}{t-s}\cdot\sum_{p=1}^{n}(t-t_p)\alpha_p\Big], \quad (6.17)$$

where

$$f_2(t,s;\mathbf{t}_n) = \exp\Big\{\frac{i\hbar}{2m}\Big[\frac{1}{t-s}\Big(\sum_{l=1}^{n}(t-t_l)\alpha_l\Big)^2 - \sum_{l,p=1}^{n}(t-t_l\vee t_p)\alpha_l\cdot\alpha_p\Big]\Big\} = \exp\Big[\frac{-i\hbar}{2m}\sum_{l,p=1}^{n}\theta(t_p,t_l)\alpha_p\cdot\alpha_l\Big]. \quad (6.18)$$

The function θ is the Green's function for the one-dimensional operator $-d^2/d\tau^2$ on the interval $[s,t]$ that vanishes at both end points, namely

$$\theta(\tau,\tau') = (t-\tau_>)(\tau_<-s)/(t-s),$$

where $\tau_< = \min(\tau,\tau')$ and $\tau_> = \max(\tau,\tau')$. In addition a useful fact is that the summation in f_2 is non-negative [Ref. 5, (2.37)], i.e., for all $\mathbf{t}_n\in\Delta_n(t,s)$ $(t_0=s)$

$$\sum_{i,p=1}^{n}\theta(t_p,t_i)\alpha_p\cdot\alpha_i = \sum_{p=1}^{n}\Big[\frac{1}{t-t_p}-\frac{1}{t-t_{p-1}}\Big]\Big(\sum_{i=p}^{n}(t-t_i)\alpha_i\Big)^2 \geqslant 0. \quad (6.19)$$

This implies that $|f_2(\cdot)|\leqslant 1$ for $m\in\mathbb{C}_+$, $t>s$.

Proceed further by employing Lemma 8 in order to evaluate the gradient operators in (6.5). In the notation developed above

$$d_n = \Big(\frac{-i}{\hbar}\Big)^n\Big(\frac{m}{2\pi i\hbar(t-s)}\Big)^{d/2}\sum_{r=0}^{n}\Big(\frac{-i\hbar}{m}\Big)^r\sum_{J_{n,r}}\sum_{l=0}^{[r/2]}\sum_{r,l}{}'\Big(\frac{im}{\hbar(t-s)}\Big)^{r-l}\int_< d\mathbf{t}_n\int d\Lambda^n(\mathbf{j}_r,\mathbf{t}_n)$$
$$\times f_2(t,s;\mathbf{t}_n)\exp\Big[ix\cdot\sum_{p=1}^{n}\alpha_p + \frac{im(x-y)^2}{2\hbar(t-s)} - i\frac{(x-y)}{t-s}\cdot\sum_{p=1}^{n}(t-t_p)\alpha_p\Big]\eta(t_{q_1},\alpha_{q_1})\cdot(y-X_n)$$
$$\times\cdots\times\eta(t_{q_{r-2l}},\alpha_{q_{r-2l}})\cdot(y-X_n)\eta(t_{q_{r-2l+1}},\alpha_{q_{r-2l+1}})\cdot\eta(t_{q_{r-2l+2}},\alpha_{q_{r-2l+2}})\times\cdots\times\eta(t_{q_{r-1}},\alpha_{q_{r-1}})\cdot\eta(t_{q_r},\alpha_{q_r}), \quad (6.20)$$

where for each $\mathbf{j}_r\in J_{r,n}$ the indices $q_1\sim q_r$ enumerate the possible divisions of the set $\{j_1,j_2,\ldots,j_r\}$ required by the summation convention $\Sigma'_{r,l}$ of Lemma 8. Although (6.20) is a complicated and lengthy expression it can be bounded in a convenient form. Take the absolute value of both sides of (6.20). Recall that $|\eta|=1$ and $|f_2|\leqslant 1$ for all allowed arguments and note that

$$|y-X_n|\leqslant Z_n \equiv |x-y| + (n\hbar k/|m|)(t-s). \quad (6.21)$$

Thus the integrand of (6.20) is bounded in absolute value by $(Z_n)^{r-2l}\exp[-\operatorname{Im} m(x-y)^2/2\hbar(t-s)]$. With this estimate one finds

$$|d_n|\leqslant\Big(\frac{|m|}{2\pi\hbar(t-s)}\Big)^{d/2}\Big(\frac{t-s}{\hbar}\Big)^n\frac{1}{n!} \times\exp\Big(\frac{-\operatorname{Im} m}{2\hbar(t-s)}(x-y)^2\Big) \times\sum_{r=0}^{n}\binom{n}{r}\Big(\mu_T+\frac{\hbar}{|m|}nk\gamma_T\Big)^{n-r}\Big(\frac{\hbar\gamma_T}{|m|}\Big)^r B_r, \quad (6.22)$$

where

$$B_r = \sum_{l=0}^{[r/2]}\Big(\frac{|m|}{\hbar(t-s)}\Big)^{r-l}\frac{r!}{2^l(r-2l)!l!}(Z_n)^{r-2l}. \quad (6.23)$$

The result (6.22) uses bound (6.2) for the measure $\Lambda^n(\mathbf{j}_r,\mathbf{t}_n)$ as well as the $\mathbf{t}_n$-volume formula

$$\int_{\Delta_n(t,s)} d\,\mathbf{t}_n = (t-s)^n/n!.$$

Continue by obtaining a simple upper bound for B_r. In (6.23) replace $[(r-2l)!]^{-1}$ with the larger $n^l[(r-l)!]^{-1}$ and then extend the sum over l to $l=0\sim r$. In this way one finds

$$B_r < \left(\frac{|m|Z_n}{\hbar(t-s)} + \frac{n}{2Z_n}\right)^r. \tag{6.24}$$

Placing (6.24) into (6.22) brings the sum over $r=0\sim n$ into the form of the binomial expansion. After utilizing $(n!)^{-1} < (e/n)^n$, we have

$$|d_n| < \left(\frac{|m|}{2\pi\hbar(t-s)}\right)^{d/2}\left(\frac{e(t-s)}{n\hbar}\right)^n \times\left\{\mu_T + \left[\frac{n\hbar}{|m|}\left(k+\frac{1}{2Z_n}\right)+\frac{Z_n}{t-s}\right]\gamma_T\right\}^n \times\exp\left(\frac{-\operatorname{Im} m}{2\hbar(t-s)}(x-y)^2\right). \tag{6.25}$$

The term in the curly brackets proportional to Z_n^{-1} is bounded by

$$\frac{n\hbar}{2|m|Z_n} = \frac{n\hbar}{2|m|}\left[|x-y| + \frac{nk\hbar}{|m|}(t-s)\right]^{-1} \leqslant \frac{1}{2k(t-s)}$$

and so

$$\{\ \}^n \leqslant \left[\frac{2nk\hbar}{|m|}\gamma_T\right]^n\left[1+\frac{1}{n}\frac{|m|}{2k\hbar}\times\left(\frac{\mu_T}{\gamma_T}+\frac{|x-y|}{t-s}+\frac{1}{2k}\frac{1}{t-s}\right)\right]^n$$
$$\leqslant \left[\frac{2nk\hbar}{|m|}\gamma_T\right]^n \times\exp\left[\frac{|m|}{2k\hbar}\left(\frac{|x-y|}{t-s}+\frac{\mu_T}{\gamma_T}+\frac{1}{2k}\frac{1}{t-s}\right)\right].$$

This last inequality, when combined with (6.25) establishes bound (6.7) for $|d_n|$. The estimate (6.7) holds for $n\geqslant 1$ and is valid for all $m\in\mathbb{C}_+$. □

Lemma 9 establishes that d_n is the Carleman kernel for the Dyson operator D_n provided that the mass has a positive imaginary part. However, it is worthwhile to keep in mind that d_n is also well defined for real masses, $m\neq 0$. We investigate now the sum over n of the kernels d_n. The radius of convergence condition we denote by $0<\theta<1$ where

$$\theta = 2ek(t-s)|m|^{-1}\gamma_T. \tag{6.26}$$

Except for the factor of 2 this is the same convergence radius as found in Proposition 3. Another useful notation is to set ($t,s\in T_\Delta$, $t\neq s$, $m\in\mathbb{C}_+$)

$$g(x-y;t,s,m) \equiv \left(\frac{|m|}{2\pi\hbar(t-s)}\right)^{d/2} \times\exp\left\{\frac{1}{2\hbar}\left[\frac{-\operatorname{Im} m}{t-s}(x-y)^2 + c_1|x-y| + c_2\right]\right\}. \tag{6.27}$$

Clearly, if $\operatorname{Im} m>0$ then $g(\cdot;t,s,m)\in L^1\cap L^2$.

Lemma 10: Let $t,s\in T_\Delta$, $t\neq s$.

(a) If $0<\theta<1$, then for each $(x,y,m)\in\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{C}_+$ the sum over $n=0\sim\infty$ of $d_n(t,x;s,y;m)$ is absolutely convergent. The pointwise value of this series is defined to be

$$K(t,x;s,y;m) = \sum_{n=0}^{\infty} d_n(t,x;s,y;m). \tag{6.28}$$

(b) Suppose R_2 and M are arbitrary compact subsets of $\mathbb{R}^d\times\mathbb{R}^d$ and $\mathbb{C}_+$, respectively, and let m_- be the smallest value of $|m|$ in M. If

$$0 < (2ek\gamma_T/m_-)(t-s) < 1 \tag{6.29}$$

then series (6.28) is uniformly convergent in $R_2\times M$. Thus for $m\in\mathbb{C}_+$, $K(t,\cdot;s,\cdot;m)$ is jointly continuous in $\mathbb{R}^d\times\mathbb{R}^d$. Furthermore for $x,y\in\mathbb{R}^d\times\mathbb{R}^d$ and $t,s\in T_\Delta$, $K(t,x;s,y,\cdot)$ is continuous in M.

(c) If $0<\theta<1$, then $K(t,x;s,y;m)$ satisfies the estimate

$$|K(t,x;s,y;m)| \leqslant (1-\theta)^{-1} g(x-y;t,s,m) \tag{6.30}$$

for all $(x,y,m)\in\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{C}_+$.

Proof: Consider (a) and (b) together. Lemma 9(c) gives the inequality

$$|d_n(t,x;s,y;m)| \leqslant \theta^n g(x-y;t,s,m), \tag{6.31}$$

whereby

$$\sum_{n=0}^{\infty} |d_n(t,x;s,y;m)| \leqslant \frac{1}{1-\theta} g(x-y;t,s,m) < \infty.$$

This proves (a) and (c).

Take D to be the largest value of $|x-y|$ in R_2 and m_+ the largest value of $|m|$ in M. Estimate (6.7) and the explicit expression for c_1 and c_2 let us write

$$|d_n| \leqslant \left(\frac{m_+}{2\pi\hbar(t-s)}\right)^{d/2}\left[\frac{2ek(t-s)}{m_-}\gamma_T\right]^n \times\exp\left\{\frac{1}{2\hbar}\left[\frac{m_+D}{k(t-s)} + \frac{m_+}{2k^2(t-s)} + \frac{m_+\mu_T}{k\gamma_T}\right]\right\}.$$

The function on the right-hand side of the inequality is independent of $(x,y,m)\in R_2\times M$. Furthermore, the sum over n is finite if (6.29) is fulfilled. Thus the uniform convergence of the series (6.28) on $R_2\times M$ is established. The continuity properties in x, y, and m are a consequence of the termwise continuity of d_n in these variables plus the uniform convergence. □

The next stage of our analysis is to interpret for positive $\operatorname{Im} m$ the function $K(t,x;s,y;m)$ as an integral kernel and investigate the nature of the operator it defines. Inequality (6.30) means that the integral

$$h(x) = \int K(t,x;s,y;m) f(y)\,dy \tag{6.32}$$

is well defined for a large class of functions f. Let us make a few preliminary observations concerning (6.32). Assume $f\in L^2$, then

$$|h(x)| \leqslant \frac{1}{1-\theta}\int g(x-y;t,s,m)\,|f(y)|\,dy. \tag{6.33}$$

The right-hand side of (6.33) is in the form of a convolution. Applying Young's inequality (Ref. 16, p. 29) shows that

$g(\cdot;t,s;m)\in L^1$ implies $h(\cdot)\in L^2$. Thus (6.32) defines a family of bounded operators on L^2, specifically $K(t,s;m): L^2\to L^2$ for each $t,s,m\in T_\Delta\times\mathbb{C}_>$ where

$$h = K(t,s;m)f. \tag{6.34}$$

Similar reasoning applies to the understanding of representation (6.4) of the operator D_n. Estimate (6.7) provides the kernel d_n with an L^1-convolution bound. This means that if $\operatorname{Im} m>0$ then D_n has a bounded extension, which we also denote by D_n.

Proposition 4: Let $t,s\in T_\Delta$ and $m\in\mathbb{C}_>$. Assume $0<\theta<1$.

(a) $$\|K(t,s;m)\|\leqslant\frac{1}{1-\theta}\|g(\cdot;t,s,m)\|_1. \tag{6.35}$$

(b) The operator-valued series defined by the sum over $n=0\sim\infty$ of $D_n(t,s;m)$ is convergent in the operator norm topology to $K(t,s;m)$, i.e.,

$$K(t,s;m) = \sum_{n=0}^{\infty} D_n\,(t,s;m). \tag{6.36}$$

(c) Let $U(t,s;m)$ be the Schrödinger evolution given in Theorem 2, then

$$K(t,s;m) = U(t,s;m). \tag{6.37}$$

Proof: (a) is a consequence of the Young inequality for convolutions, namely

$$\|h\|_2\leqslant(1-\theta)^{-1}\|g(\cdot;t,s;m)\|_1\|f\|_2.$$

In order to demonstrate (b) recall that both $K(t,s;m)$ and $\Sigma_{n=0}^{N} D_n(t,s;m)$ (for finite N) are integral operators. By (6.4) and (6.28) one finds

$$\left[K(t,s;m)f-\sum_{n=0}^{N} D_n(t,s;m)f\right](x)$$
$$\equiv h_N(x)$$
$$=\int\left[K(t,x;s,y;m)-\sum_{n=0}^{N} d_n(t,x;s,y;m)\right]f(y)dy.$$

Inequality (6.31) then yields the bound

$$|h_N(x)|\leqslant\frac{\theta^{N+1}}{1-\theta}\int g(x-y;t,s,m)|f(y)|dy$$

and an application of Young's convolution inequality leads to

$$\left\|K(t,s;m)-\sum_{n=0}^{N} D_n(t,s;m)\right\|\leqslant\frac{\theta^{N+1}}{1-\theta}\|g(\cdot;t,s,m)\|_1.$$

The right-hand side here vanishes as $N\to\infty$ and so (b) is established.

Consider (c). Here we bring together the conclusions of Proposition 3 and Theorems 1 and 2. Let the initial data function $\phi\in\hat{C}_0^\infty$, where $\hat{C}_0^\infty$ denotes the Fourier image of $C_0^\infty(\mathbb{R}^d)$. Proposition 3 states that, if $0<\theta<1$,

$$\Psi(t,s;m) = \sum_{n=0}^{\infty} D_n(t,s;m)\phi \tag{6.38}$$

is a t-strongly continuously differentiable solution of the Schrödinger equation (5.33) that satisfies the initial condition (5.34). Theorem 1 asserts that if the fields a and v obey hypotheses $\mathscr{V}_v(k)$ and $\mathscr{V}_s(k)$, then the Cauchy problem in T_Δ is uniformly correct. In particular, the strongly continuously differentiable L^2 solutions of the Cauchy problem are unique and may always be represented in terms of the bounded Schrödinger evolution operator $U(t,s;m)$ given by (4.12). This means that $\Psi(t,s;m)$ has a representation

$$\Psi(t,s;m) = U(t,s;m)\phi. \tag{6.39}$$

Note that if the conclusion of part (b) is used in (6.38) then also

$$\Psi(t,s;m) = K(t,s;m)\phi. \tag{6.40}$$

Since $\hat{C}_0^\infty$ is dense in L^2 and both U and K are bounded operators Eqs. (6.39) and (6.40) show that these two operators are identical for all parameters t, s, and m falling in the domain of their joint definition. Thus (c) is proved. □

Several remarks about the results of Proposition 4 are in order. First of all, from (c), (4.12), and (4.9) we have a bound

$$\|K(t,s;m)\|\leqslant\exp[(1+\nu_T/\hbar)(t-s)],$$

which is a stronger result than (6.35). (It is uniform in $m\in\mathbb{C}_+$!) However, to get this bound we must rely on the specific methods of abstract evolution theory.[6] This is not the case with (6.35) and, at least for $m\in\mathbb{C}_>$, it guarantees, together with (6.36) and Proposition 3, that the Dyson series construction alone provides a unique propagator type solution for the evolution problem considered in this paper. So it seems worthwhile to state (a) independently.

The evolution operator $U(t,s;m)$ is well defined for all $(t,s,m)\in T_\Delta\times\mathbb{C}_+$. However, equality (6.37) is demonstrated for a more restricted set of time displacements, i.e., those that satisfy $\theta<1$.

Estimate (6.7) of Lemma 9 provides the critical control of the analysis that lets us prove that $K(t,s;m)$ and $D_n(t,s;m)$ are bounded operators if $\operatorname{Im} m>0$. Note $\|g(\cdot;t,s,m)\|_1\to\infty$ as $\operatorname{Im} m\to 0$ so that inequality (6.35) says little about the real axis boundary value of the operator $K(t,s;m)$.

If in Lemma 9 part (a) one tries to extend the statement of (6.4) so that it applies when $\operatorname{Im} m=0$ then Fourier transform (6.12) is no longer the transform of an L^1 function and furthermore the Fresnel integrals (6.15) will no longer converge absolutely and at best then require some sort of distributional interpretation. The complex mass method avoids all these difficulties.

Via Eqs. (6.37), (6.34), and (6.32), Proposition 4 shows that the function $K(t,\cdot;s,\cdot;m)$ is the (Carleman) integral kernel of the evolution operator $U(t,s;m)$, namely for almost all x

$$[U(t,s;m)\phi](x) = \int K(t,x;s,y;m)\phi(y)dy \quad (m\in\mathbb{C}_>) \tag{6.41}$$

for all $\phi\in L^2$, provided of course that $\operatorname{Im} m>0$. The final task is to extend the meaning of (6.41) to the positive mass axis. We denote by $L_0^p(\mathbb{R}^d)$ the L^p functions on $\mathbb{R}^d$ with compact support.

Theorem 3: Assume $a\in\mathscr{V}_v(k)$ and $v\in\mathscr{V}_s(k)$. Let $\phi\in L_0^2(\mathbb{R}^d)$, $m>0$, and $0<\theta<1$. If $U(t,s;m)$ is the Schrödinger evolution operator of Theorem 1 and if $K(t,\cdot;s,\cdot;m)$:

$\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{C}$ is the continuous function constructed in Lemma 10, then for almost all $x\in\mathbb{R}^d$

$$[U(t,s;m)\phi](x)=\int K(t,x;s,y;m)\phi(y)dy\quad(m\in\mathbb{R}^+). \tag{6.42}$$

Proof: If $\phi\in L_0^2(\mathbb{R}^d)$ then also $\phi\in L_0^1(\mathbb{R}^d)$. Since $K(t,\cdot;s,\cdot;m)$ is a jointly continuous function on compact subsets of $\mathbb{R}^d\times\mathbb{R}^d$, the integral in (6.41) defines in $\mathbb{R}^d$ a continuous function of x for each $m\in\mathbb{C}_+$. Thus, upon setting $m=m_1+i\delta$ $(m_1>0,\delta>0)$, we can consider the $\delta\to 0$ limit of the integral in (6.41) for any point $x\in\mathbb{R}^d$. To this end notice that for every fixed $x\in\mathbb{R}^d$ Lemma 10(b) implies that there exists a finite constant $C_{x,t,s}$ such that

$$|K(t,x;s,y;m_1+i\delta)\phi(y)|\leqslant C_{x,t,s}|\phi(y)|, \tag{6.43}$$

provided that t, $s\in T_\Delta$, $t\neq s$, and

$$0<(2ek/m_1)\gamma_T(t-s)<1. \tag{6.44}$$

Estimate (6.43) and the fact that $K(t,x;s,y;m_1+i\delta)$ has a limiting value $K(t,x;s,y;m_1)$ allows the application of the dominated convergence theorem to obtain

$$\lim_{\delta\to 0}\int K(t,x;s,y;m_1+i\delta)\phi(y)dy=\int K(t,x;s,y;m_1)\phi(y)dy \tag{6.45}$$

where, by Lemma 10(b) the right-hand side of (6.45) is also a continuous function of x.

Put $\delta=n^{-1}$, $n=1\sim\infty$. Proposition 2 states that

$$\operatorname*{s-lim}_{n\to\infty} U(t,s;m_1+i/n)\phi=U(t,s;m_1)\phi.$$

A standard result (Ref. 12, Theorem 3.12, p. 70) asserts that there exists a subsequence $\{n_j\}$ such that for almost all $x\in\mathbb{R}^d$

$$\lim_{j\to\infty}[U(t,s;m_1+i/n_j)\phi](x)=[U(t,s;m_1)\phi](x). \tag{6.46}$$

Identity (6.41) and the existence of the two limits (6.45) and (6.46) establish (6.42). □

Some comments on the results of Theorem 3 are appropriate. Theorem 3 shows that the function $K(t,x;s,y;m)$, $m>0$, is the propagator (the weak integral kernel in the sense of Definition 2) of the Schrödinger evolution operator if the time displacement $t-s$ satisfies condition (6.44). It is worthwhile to note that the function $K(t,x;s,y;m)$ of (6.28) is in fact a propagator in a stronger sense. The boundedness of $U(t,s;m)$ allows the relation (6.42) to be extended as follows. For any $f\in L^2(\mathbb{R}^d)$, $m>0$

$$[U(t,s;m)f](x)=\lim_{L\to\infty}\int_{|y|<L}K(t,x;s,y;m)f(y)dy.$$

This statement is structurally identical to the meaning traditionally assigned to the free propagator,[2,8,16] and as such indicates that the result obtained in Theorem 3 is close to optimal.

In the case of evolution for complex mass ($\operatorname{Im} m>0$) one may establish without difficulty that the local solutions of Theorem 3 may be pieced together to obtain a solution that is valid for all $t,s\in T_\Delta$. In this case the radius of convergence condition $\theta<1$ for the Dyson series is eliminated. In the following corollary to Theorem 3 we show how to define an extended kernel $K(t,x;s,y;m)$ that remains valid even if $\theta\geqslant 1$.

Corollary 1: Assume $a\in\mathscr{V}_v(k)$, $v\in\mathscr{V}_s(k)$, and $m\in\mathbb{C}_>$. For all $t,s\in T_\Delta$, $t\neq s$.

(1) The complex mass evolution operator $U(t,s;m)$ is an integral operator with a Carleman kernel $K(t,\cdot;s,\cdot;m)\in C(\mathbb{R}^d\times\mathbb{R}^d)$.

(2) The kernel K is Gaussian bounded, i.e., there exists $0<A$, $B<\infty$ (dependent upon m, t, and s) such that

$$|K(t,x;s,y;m)|\leqslant Ae^{-B(x-y)^2}. \tag{6.47}$$

(3) The composition rule

$$K(t,x;s,y;m)=\int K(t,x;\tau,z;m)K(\tau,z;s,y;m)dz \tag{6.48}$$

holds for all s, τ, and t such that $0\leqslant s<\tau<t\leqslant T$.

Proof: Suppose $t-\tau$ and $\tau-s$ are such that $\theta<1$ for both of these time displacements. Then representation (6.41) holds for $U(t,\tau;m)$ and $U(\tau,s;m)$ with the appropriate kernel functions defined by the constructive series (6.28). From the composition property (1.8)

$$U(t,s;m)=U(t,\tau,m)U(\tau,s;m)$$

and (6.41) it follows that, for $\phi\in L^2$,

$$\begin{aligned}&[U(t,s;m)\phi](x)\\&\quad=\int K(t,x;\tau,z;m)[U(\tau,s;m)\phi](z)dz\\&\quad=\int K(t,x;\tau,z;m)\left[\int K(\tau,z;s,y;m)\phi(y)dy\right]dz.\end{aligned}$$

Estimate (6.30) suffices to show that the $dy\,dz$ integration is absolutely convergent. Applying the Fubini theorem gives Eq. (6.41) wherein the kernel function $K(t,x;s,y;m)$ is now defined by the absolutely convergent integral (6.48).

The convolution bound estimate (6.47) follows from an application of the free heat kernel composition identity, i.e., $\beta_1,\beta_2>0$

$$\begin{aligned}&[4\pi(\beta_1+\beta_2)]^{-d/2}\exp\left\{-\frac{(x-y)^2}{4(\beta_1+\beta_2)}\right\}\\&\quad=[(4\pi\beta_1)(4\pi\beta_2)]^{-d/2}\int\exp\left\{-\frac{(x-z)^2}{4\beta_1}\right\}\\&\qquad\times\exp\left\{-\frac{(z-y)^2}{4\beta_2}\right\}dz,\end{aligned}$$

and the fact that in inequality (6.30) g may be replaced by

$$(1-\theta)^{-1}g(x-y;t,s,m)<A'e^{-B'(x-y)^2}$$

for suitable $0<A'$, $B'<\infty$.

Estimate (6.47) shows that the extended kernel K constructed by (6.48) is also a Carleman kernel. In addition bound (6.47) suffices to show that the integral (6.48) is uniformly convergent with respect to x and y in compact subsets of $\mathbb{R}^d\times\mathbb{R}^d$. The continuity of the integrand in x,y together with the uniform convergence establishes that the left-hand side of (6.48) is jointly continuous in x and y. One may also verify without difficulty that the value of the integral (6.48) is independent of the particular choice of the intermediate time τ. Thereby properties (1)–(3) are verified if $t-\tau$ and $\tau-s$ both have $\theta<1$, or equivalently (after an appropriate choice of τ) for $t-s$ with $\theta<2$. If this argument (with suitable adjustments) is repeated n times one obtains via composition rule (6.48) the kernel K for $t-s$ with $\theta<2^n$. □

ACKNOWLEDGMENT

This work has been supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.

[1]B. Simon, Bull. Am. Math. Soc. **7**, 447 (1982).
[2]E. Nelson, J. Math. Phys. **5**, 332 (1964).
[3]K. Ito, in *Proceedings of the Fourth Berkeley Symposium on Mathematics Statistics and Probability* (Univ. California, Berkeley, 1961), Vol. II, p. 227; K. Ito, in *Proceedings of the Fifth Berkeley Symposium on Mathematics Statistics and Probability* (Univ. California, Berkeley, 1967), Vol. 2, part I, p. 145.
[4]S. A. Albeverio and R. J. Høegh-Krohn, *Mathematical Theory of Feynman Path Integrals* (Springer, Berlin, 1976).
[5]T. A. Osborn and Y. Fujiwara, J. Math. Phys. **24**, 1093 (1983).
[6]S. G. Krein, *Linear Differential Equations in Banach Space* (American Mathematical Society, Providence, 1971).
[7]F. J. Dyson, Phys. Rev. **75**, 486 (1949); **75**, 1736 (1949).
[8]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966).
[9]N. Dunford and J. T. Schwartz, *Linear Operators Part I* (Interscience, New York, 1967).
[10]W. Rudin, *Fourier Analysis on Groups* (Interscience, New York, 1961), Chap. 1.
[11]J. Diestel and J. J. Uhl, *Vector Measures* (American Mathematical Society, Providence, 1977).
[12]W. Rudin, *Real and Complex Analysis* (McGraw-Hill, New York, 1974).
[13]H. Leinfelder and C. G. Simader, Math. Z. **176**, 1 (1981).
[14]H. Tanabe, *Equations of Evolution* (Pitman, London, 1979), Chap. 4.
[15]J. D. Dollard and C. N. Friedman, *Product Integration* (Addison–Wesley, Reading, 1979), Chap. 3.
[16]M. Reed and B. Simon, *Methods of Modern Mathematical Physics II, Fourier Analysis, Self-Adjointness* (Academic, New York, 1975).
[17]E. Hille and R. S. Phillips, *Functional Analaysis and Semi-Groups* (American Mathematical Society, Providence, 1957).
[18]R. Abraham, J. E. Marsden, and T. Ratiu, *Manifolds, Tensor Analysis, and Applications* (Addison–Wesley, Reading, 1983).
[19]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1980).

Nontrivial zeros of the Wigner (3*j*) and Racah (6*j*) coefficients. II. Some nonlinear solutions

Simcha Brudno
Chicago College of Osteopathic Medicine, Chicago, Illinois 60615

(Received 5 May 1986; accepted for publication 3 September 1986)

Additional formulas for nontrivial zeros in the $3j$ and $6j$ symbols have been found for some higher-order cases, i.e., where $k=2$, 3, and 4 (numerical examples only).

I. INTRODUCTION

In a previous publication,[1] a set of formulas was presented that finds nontrivial zeros of the $3j$ and $6j$ symbols for the linear case where $k=1$. (The nontrivial zeros are zeros of the "polynomial part" of Wigner or Racah coefficients.) Subsequently, Brudno and Louck[2] demonstrated this general formula for the "linear" case to be complete. In this publication a relationship was also demonstrated between these zeros and two well-known Diophantine equations,

$$X^2+Y^2=U^2+V^2 \tag{1.1}$$

and

$$X^3+Y^3+Z^3=U^3+V^3+W^3, \tag{1.2}$$

$$X+Y+Z=U+V+W. \tag{1.3}$$

Equation (1.1) was solved by Pasternak; recent work by Bremner and Brudno[3] provides the total solution for the equation set [(1.2), (1.3)].

Progress also has been made in finding similar formula sets for the higher-order cases. Beyer, Louck, and Stein[4] published an exhaustive review article on the quadratic zeros that includes discussions of Pell equations, the solutions of which "give parametric families of nontrivial zeros of Racah coefficients." A key element in the formulation of the quadratic problem is presented in Sec. II of this manuscript for two purposes: (1) to orient the reader for a parallel development in the cubic case, and (2) to illuminate a feature of the quadratic formulation that may be of value in physics but that is not readily apparent in the development of this problem as presented by Beyer, Louck, and Stein.[4] In these more complex cases "total" solutions remain elusive; however, enough headway has been made to warrant publication of current results in the hope that these bits and pieces may enlighten the paths of others seeking to understand the complex interrelationships operant within the Wigner and Racah symbols.

II. QUADRATIC ZEROS

For the quadratic case, i.e., where $k=2$, it can be assumed without loss of generality that the Racah coefficient has the form

$$\begin{Bmatrix} a & b & a+b-2 \\ d & c & f \end{Bmatrix} \equiv 0, \tag{2.1}$$

where the following equality is implied:

$$\begin{aligned} &A(A-1)(F)(F+1)E(E+1) \\ &\quad -2(A-1)(F+1)(E+1)BCD \\ &\quad +B(B-1)C(C-1)D(D-1)\equiv 0, \end{aligned} \tag{2.2}$$

and where

$$A=a+b+c+d+1, \tag{2.3}$$

$$B=a+c-f, \tag{2.4}$$

$$C=b+d-f, \tag{2.5}$$

$$D=-a-b+c+d+2, \tag{2.6}$$

$$E=b-d+f-1, \tag{2.7}$$

$$F=a-c+f-1. \tag{2.8}$$

Equations (2.2)–(2.8), with slight changes in notation, are given in (2.2) of Beyer, Louck, and Stein.[4]

For further analysis, the above equations are limited to the particular branch where $b=2$ and $d=f$, with a, b, and c yet to be determined. The more particular case where c also equals d,

$$\begin{Bmatrix} a & 2 & a \\ d & d & d \end{Bmatrix} \equiv 0, \tag{2.9}$$

has been solved by Brudno and Louck.[5] For the particular case where $a=j_1=j_3, j_2=2, d=l_1=l_3, c=l_2$, namely,

$$\begin{Bmatrix} a & 2 & a \\ d & c & d \end{Bmatrix} \equiv 0, \tag{2.10}$$

there exists a "hyperbolic ladder" that may have meaning in the physical realm. The first 13 examples of this ladder are given in Table I. Every $3n$th term is a hinge of this ladder

TABLE I. The first 13 examples of the hyperbolic ladder.

Index	a	d	c
0	0	−0.5	−0.5
1	1	0.5	0.5
2	2	1.5	1.5
3	3	0.5	2.5
4	18	15.5	15.5
5	33	28.5	28.5
6	48	13.5	41.5
7	257	222.5	222.5
8	466	403.5	403.5
9	675	194.5	584.5
10	3586	3105.5	3105.5
11	6497	5626.5	5626.5
12	9408	2715.5	8147.5

 0022-2488/87/010124-04$02.50

TABLE II. Parametrical solutions for Eq. (3.10).

a	b	$2c+1$	
$6s^2+6s+3$	$36s^4+72s^3+57s^2+24s+5$	$72s^5+180s^4+198s^3+123s^2+43s+7$	(i)
$6s^2+6s+5$	$12s^4+24s^3+23s^2+12s+3$	$24s^5+60s^4+82s^3+65s^2+31s+7$	(ii)
$6s^2+7s+3$	$36s^3+42s^2+21s+5$	$72s^4+132s^3+104s^2+41s+7$	(iii)
$6s^2+8s+3$	$18s^3+33s^2+18s+5$	$36s^4+96s^3+91s^2+39s+7$	(iv)
$6s^2+9s+5$	$12s^3+18s^2+11s+3$	$24s^4+60s^3+64s^2+33s+7$	(v)
$6s^2+12s+5$	$6s^3+15s^2+10s+3$	$12s^4+48s^3+65s^2+35s+7$	(vi)
$8s^2+9s+3$	$24s^2+15s+5$	$64s^3+80s^2+37s+7$	(vii)
$12s^2+9s+3$	$12s^2+15s+5$	$48s^3+72s^2+37s+7$	(viii)

where all of the a terms lie on the hyperbola, $4y^2-3x^2$ = constant, and where the remaining a's represent rungs located equidistant from the hinges. The explicit formulas for the hinges are

$$a_{3n}=\cosh 2n(\alpha)/2-\tfrac{1}{2} \text{ or } [\sinh n(\alpha)]^2, \tag{2.11}$$

$$d_{3n}=\sinh 2n(\alpha)\sinh(\alpha)/12-\tfrac{1}{2}, \tag{2.12}$$

$$c_{3n}=3(d+\tfrac{1}{2})-\tfrac{1}{2}, \tag{2.13}$$

where $\cosh(\alpha)\equiv 2$. The explicit formulas for the two interstitial rungs, $(3n+1)$ and $(3n+2)$, are, respectively,

$$a_{(3n+1)}=[\cosh 2(n+1)\alpha+\cosh 2n\alpha]/6-\tfrac{1}{2}, \tag{2.14}$$

$$a_{(3n+2)}=[2\cosh 2(n+1)\alpha+2\cosh 2n\alpha]/6-\tfrac{1}{2}, \tag{2.15}$$

$$d=c=[a\tanh\alpha]+\tfrac{1}{2}, \tag{2.16}$$

where $[a\tanh\alpha]$ is the integer part of $a\tanh\alpha$. Each number P in Table I obeys the following recurrence relation:

$$P_n=15(P_{n-3}-P_{n-6})+P_{n-9}, \tag{2.17}$$

where, whenever P is a negative number, $P=-P-1$. [Because all the quantum numbers appear in the $3nj$ coefficients through the form $j\,(j+1)$ and powers thereof, in every case, whenever a solution j_i satisfies a particular equation, there exists also the parallel solution $j_2=-j_1-1$, so that the expression $j(j+1)$ has the same value. This permits the exchange of every $-j$ with its concomitant $|j|-1$.]

III. CUBIC ZEROS

Considering cubic zeros, i.e., zeros of order 3,

$$\begin{Bmatrix} a & b & a+b-3 \\ d & c & f \end{Bmatrix}\equiv 0, \tag{3.1}$$

where, without loss of generality, the polynomial to be zeroed is

$$\begin{aligned}&1A(A-1)(A-2)E(E+1)(E+2)F(F+1)(F+2)\\&\quad-3(A-1)(A-2)(E+1)\\&\quad\times(E+2)(F+1)(F+2)BCD\\&\quad+3(A-2)(E+2)(F+2)\\&\quad\times B(B-1)C(C-1)D(D-1)\\&\quad-1B(B-1)(B-2)\\&\quad\times C(C-1)(C-2)D(D-1)(D-2)\end{aligned} \tag{3.2}$$

and where

$$A=a+b+c+d+l, \tag{3.3}$$

$$B=a+c-f, \tag{3.4}$$

$$C=b+d-f, \tag{3.5}$$

$$D=-a-b+c+d+3, \tag{3.6}$$

$$E=b-d+f-2, \tag{3.7}$$

$$F=a-c+f-2. \tag{3.8}$$

Zeros of (3.2) have been found. For the particular branch where $c=d=f$, the polynomial expression (3.2) reduces to

$$4(a+b-1)(a+b-2)\{3(2c+1)^2-[2ab(a+b)-a^2-5ab-b^2+2(a+b)]\}, \tag{3.9}$$

and from here

$$3(2c+1)^2=2ab(a+b)-a^2-5ab-b^2+2(a+b). \tag{3.10}$$

A particular set of parametrical solutions for (3.10) is listed in Table II. A two-parameter solution for (3.10) that also includes (i)–(viii) in Table II is given by

$$a=6s^2+6s+1+n, \tag{3.11}$$

$$b=6s^2+6s+1+(36s^4+72s^3+51s^2+18s+3)/n, \tag{3.12}$$

TABLE III. Specific numerical examples of (3.11)–(3.13), where $s=3$.

Index	a	b	$2c+1$
1	75	5450	38665
2	76	2762	19856
3	77	1866	13591
4	78	1418	10462
5	80	970	7340
6	81	842	6451
7	82	746	5786
8	86	522	4246
9	88	458	3812
10	90	410	3490
11	95	330	2965
12	98	298	2762
13	102	266	2566
14	106	242	2426
15	116	202	2216
16	122	186	2146
17	130	170	2090
18	138	158	2062

TABLE IV. Parametrical solutions for (3.10).

a	b	$2c+1$	
$6s^2+6s+3$	$35s^4+72s^3+57s^2+18s+2$	$72s^5+180s^4+198s^3+111s^2+31s+5$	(i)
$6s^2+6s+5$	$12s^4+24s^3+23s^2+10s+2$	$24s^5+60s^4+82s^3+61s^2+27s+5$	(ii)
$6s^2+7s+2$	$36s^3+78s^2+57s+14$	$72s^4+204s^3+212s^2+93s+14$	(iii)
$6s^2+8s+3$	$18s^3+33s^2+18s+2$	$36s^4+96s^3+91s^2+33s+3$	(iv)
$6s^2+9s+2$	$12s^3+30s^2+23s+6$	$24s^4+84s^3+100s^2+45s+6$	(v)
$6s^2+12s+5$	$6s^3+15s^2+10s+2$	$12s^4+48s^3+65s^2+33s+5$	(vi)
$8s^2+7s+2$	$24s^2+33s+14$	$64s^3+112s^2+69s+14$	(vii)
$12s^2+9s+2$	$12s^2+15s+6$	$48s^3+72s^2+37s+6$	(viii)

$$2c+1=(2s+1)(a+b)-3s-1, \tag{3.13}$$

where n assumes, not only the algebraical factors of $3(s+1)\times(2s+1)(6s^2+3s+1)$, but the numerical factors as well, and where s is a rational parameter such that a and b are integers. Specific numerical examples (for $s=3$) are given in Table III.

Another parametrical solution set for (3.10) that is also associated with (3.11)–(3.13) is given in Table IV.

The corresponding two-parameter solution that includes (i)–(viii) in Table IV is given by

$$a=6s^2+6s+2+n, \tag{3.14}$$

$$b=6s^2+6s+2+3(6s^2+9s+4)(2s+1)s/n, \tag{3.15}$$

$$2c+1=(2s+1)(a+b)-(3s+2), \tag{3.16}$$

where, again, n assumes, not only the algebraical factors of $3s(6s^2+9s+4)(2s+1)$, but the numerical factors as well, and where, again, s is a rational parameter such that a and b are integers.

For any numerical example of (3.10), an s can be found that generates that particular numerical example, together with the other members of the set generated by that particular s. For example, calculating the numerical value of the quotient

$$(2c+a+b)/[2c-(a+b)+3] \tag{3.17}$$

from the solution

$$\begin{Bmatrix} 9 & 15 & 21 \\ 21 & 21 & 21 \end{Bmatrix} \equiv 0, \tag{3.18}$$

which is a specific case of (3.10),

$$\begin{Bmatrix} a & b & a+b-3 \\ c & c & c \end{Bmatrix} \equiv 0, \tag{3.19}$$

a value of $\frac{66}{21}$ is obtained:

$$n/m=\tfrac{22}{7}, \tag{3.20}$$

$$s=m/(n-m)=\tfrac{7}{15}. \tag{3.21}$$

Another analysis of (3.10) can be achieved by rewriting (3.10) in the following form:

$$3(2c+1)^2-(2b-1)(a+b/2-1)^2 = (b-2)[(2b^2-b+2)/4]. \tag{3.22}$$

With b held constant, Eq. (3.22) becomes a Pell equation; therefore, with every b there exists an infinity of a's and c's. An orbit of b is obtained by multiplying a particular solution, a_0, b_0, by the algebraical units given by

$$3y^2-(2b-1)x^2=1. \tag{3.23}$$

Analysis of the orbits for the different b's currently is in progress; however, a specific example, where $b=17$, is given in Table V.

IV. QUARTIC ZEROS

Considering quartic zeros, i.e., zeros of order 4,

$$\begin{Bmatrix} a & b & a+b-4 \\ d & c & f \end{Bmatrix} \equiv 0, \tag{4.1}$$

where, without loss of generality, the polynomial to be zeroed is

$$\begin{aligned} &1A(A-1)(A-2)(A-3)F(F+1)(F+2)(F+3)E(E+1)(E+2)(E+3) \\ &\quad -4(A-1)(A-2)(A-3)(F+1)(F+2)(F+3)(E+1)(E+2)(E+3)BCD \\ &\quad +6(A+2)(A+3)(F+2)(F+3)(E+2)(E+3)B(B-1)C(C-1)D(D-1) \\ &\quad -4(A+3)(F+3)(E+3)B(B-1)(B-2)C(C-1)(C-2)D(D-1)(D-2) \\ &\quad +1B(B-1)(B-2)(B-3)C(C-1)(C-2)(C-3)D(D-1)(D-2)(D-3) \end{aligned} \tag{4.2}$$

and where

$$A=a+b+c+d+1, \tag{4.3}$$

$$B=a+c-f, \tag{4.4}$$

$$C=b+d-f, \tag{4.5}$$

$$D=-a-b+c+d+4, \tag{4.6}$$

$$E=b-d+f-3, \tag{4.7}$$

$$F=a-c+f-3. \tag{4.8}$$

Presented below are ten examples of zeros of (4.2):

TABLE V. A specific example of (3.22), where $b = 17$.

Index	a	b	$2c+1$
1	2	17	17
2	71	17	259
3	74	17	269
4	863	17	2887
5	2903	17	9653
6	31154	17	103351
7	32351	17	107321

$$\begin{Bmatrix} 20.5 & 13.5 & 30.0 \\ 35.0 & 29.0 & 27.5 \end{Bmatrix} \equiv 0, \tag{4.9}$$

$$\begin{Bmatrix} 28.5 & 11.5 & 36.0 \\ 40.5 & 35.5 & 34.0 \end{Bmatrix} \equiv 0, \tag{4.10}$$

$$\begin{Bmatrix} 17.5 & 9.5 & 23 \\ 34 & 39 & 32.5 \end{Bmatrix} \equiv 0, \tag{4.11}$$

$$\begin{Bmatrix} 12.5 & 8.0 & 16.5 \\ 27.0 & 26.5 & 29.0 \end{Bmatrix} \equiv 0, \tag{4.12}$$

$$\begin{Bmatrix} 16.5 & 7.0 & 19.5 \\ 32.0 & 30.5 & 29.0 \end{Bmatrix} \equiv 0, \tag{4.13}$$

$$\begin{Bmatrix} 15.0 & 6.0 & 17.0 \\ 25.0 & 30.0 & 24.0 \end{Bmatrix} \equiv 0, \tag{4.14}$$

$$\begin{Bmatrix} 29.5 & 16.0 & 41.5 \\ 44.0 & 41.5 & 33.0 \end{Bmatrix} \equiv 0, \tag{4.15}$$

$$\begin{Bmatrix} 33.0 & 10.5 & 39.5 \\ 43.5 & 38.0 & 40.0 \end{Bmatrix} \equiv 0, \tag{4.16}$$

$$\begin{Bmatrix} 18.0 & 5.5 & 19.5 \\ 28.5 & 36.0 & 30.0 \end{Bmatrix} \equiv 0, \tag{4.17}$$

$$\begin{Bmatrix} 14.0 & 18.5 & 28.5 \\ 58.0 & 32.5 & 43.5 \end{Bmatrix} \equiv 0. \tag{4.18}$$

Examples (4.9) and (4.10) are related by the equation

$$A = 3(D-1), \tag{4.19}$$

examples (4.11)–(4.14) by the equation

$$A = 2D + 1, \tag{4.20}$$

and examples (4.15) and (4.16) by the equation

$$A = 3D. \tag{4.21}$$

Examples (4.17) and (4.18) appear unrelated, each of them, respectively, obeying the equations

$$A = 2D - 1, \tag{4.22}$$

$$A = 2D. \tag{4.23}$$

In Eq. (4.19)–(4.23), A and D are defined by Eqs. (4.3) and (4.6), respectively.

ACKNOWLEDGMENTS

The author thanks John Conway for providing an essential insight, Dennis Bartley for assistance, and Thomas Glonek for critical review.

[1]S. Brudno, J. Math. Phys. **26**, 434 (1985).
[2]S. Brudno and J. D. Louck, J. Math. Phys. **26**, 2092 (1985).
[3]A. Bremner and S. Brudno, J. Math. Phys. **27**, 2613 (1986).
[4]W. A. Beyer, J. D. Louck, and P. R. Stein, "Zeros of Racah coefficients and the Pell equation. I," submitted to J. Math. Phys.
[5]S. Brudno and J. D. Louck, J. Math. Phys. **26**, 1125 (1985).

The sum rule for spectroscopic factors in the seniority scheme of identical particles

Masao Nomura
Institute of Physics, College of Arts and Sciences, University of Tokyo, Tokyo, 153, Japan

(Received 27 August 1985; accepted for publication 10 September 1986)

Spectroscopic factors partitioning an identical-particle state into a couple of substates with definite seniorities are summed into a simple form. This sum rule is an extension of the reduction relation for the fixed-seniority average of a many-body operator. The seniority projection operator is utilized throughout the present formulation.

I. INTRODUCTION

It is known[1] that the average interaction energy in the n-body states with a definite seniority is describable as a linear combination of the corresponding average energies in the two-body states. Extension of this theorem has been done in the study of symmetry and statistical properties inherent to nuclear spectroscopy.[2–10]

In the seniority scheme of identical particles,[11–13] which the present paper concerns, the reduction relation for the average two-body interaction energy has been extended to that for the average of a general n_1-body operator $O(n_1)$ as

$$\sum_\alpha \frac{\langle nv\alpha|O(n_1)|nv\alpha\rangle}{d(v)} = \sum_{v_1} Z(nv,n_1v_1) \sum_{\alpha_1} \frac{\langle n_1v_1\alpha_1|O(n_1)|n_1v_1\alpha_1\rangle}{d(v_1)}, \qquad (1)$$

where $|nv\alpha\rangle$ stands for the orthonormalized n-body state labeled by the seniority quantum number v, defined for the finite set of degenerate orbits,[14] together with the set of the other quantum numbers α. Here, the symbol $d(v)$ denotes the dimensionality of orthogonal states with the definite v. The coefficient Z has properties characteristic to the propagator, and is called the propagation coefficient.[9] The explicit form of Z was given in Ref. 6.

We express $O(n_1)$ as a linear combination of operators $A^+(n_1v_1\alpha_1)A(n_1v_1'\alpha_1')$, where $A^+(n_1v_1\alpha_1)$ stands[3,15–18] for the polynomial of single-particle creation operators (C^+'s) of rank n fulfilling the condition $A^+(n_1v_1\alpha_1)|0\rangle = |n_1v_1\alpha_1\rangle$. Ingredients of (1) are written in terms of A^+ and A as

$$\sum_\alpha \frac{\langle nv\alpha|A^+(n_1v_1\alpha_1)A(n_1v_1'\alpha_1')|nv\alpha\rangle}{d(v)} = \delta(\alpha_1,\alpha_1')\delta(v_1,v_1')\,\frac{Z(nv,n_1v_1)}{d(v_1)}. \qquad (2)$$

A characteristic feature of (2) is nonexistence of an α_1-(α_1'-)dependent term on the right-hand side (rhs) except $\delta(\alpha_1,\alpha_1')$. The matrix element of A^+ is related to the coefficient of fractional parentage (cfp) as[17]

$$\langle nv\alpha|A^+(n_1v_1\alpha_1)|n_2v_2\alpha_2\rangle = \binom{n}{n_1}^{1/2}\langle n_2v_2\alpha_2 + n_1v_1\alpha_1|nv\alpha\rangle, \qquad (3)$$

where $n_1+n_2=n$. Square of the matrix element of A^+ is called the spectroscopic factor.[19]

In the present paper, we extend (2) to the form

$$\sum_\alpha \frac{\langle nv\alpha|A^+(n_1v_1\alpha_1)P(n_2,v_2)A(n_1v_1'\alpha_1')|nv\alpha\rangle}{d(v)} = \frac{\delta(\alpha_1,\alpha_1')\delta(v_1,v_1')}{d(v_1)} \times F(\equiv \text{a factor independent of } \alpha_1), \qquad (4)$$

where $P(n_2,v_2)$ stands for the projection operator onto the n_2-body space with the definite seniority v_2, i.e.,

$$P(n_2,v_2) = \sum_{\alpha_2} |n_2v_2\alpha_2\rangle\langle n_2v_2\alpha_2|. \qquad (5)$$

We show that the factor F, a function in n, n_1, v, v_1, and v_2, is represented by a single quotient of products of binomial coefficients (or factorials). This result is hardly suggested by the involved form of Z. The present formulation relies on reexpressions of $P(n_2,v_2)$ in different forms.

The expression (4) in the case of $\alpha_1=\alpha_1'$ and $v_1=v_1'$ represents the sum rule for the spectroscopic factors. We can rewrite it as

$$\begin{aligned} F(nv,n_1v_1,n_2v_2) &= \sum_{\alpha\alpha_2} |\langle nv\alpha|A^+(n_1v_1\alpha_1)|n_2v_2\alpha_2\rangle|^2 \frac{d(v_1)}{d(v)} \qquad (6)\\ &= \sum_{\alpha_1\alpha_2} |\langle nv\alpha|A^+(n_1v_1\alpha_1)|n_2v_2\alpha_2\rangle|^2 \qquad (7)\\ &= F(nv,n_2v_2,n_1v_1), \qquad (8)\end{aligned}$$

where in the second step the particle–hole symmetry is used, see Sec. IV. To clarify the physical implication of F, let us assign $|nv\alpha\rangle$ to the fixed-nv component of the initial state of the target (or projectile) nucleus in the heavy-ion reaction, where the term "initial" means the stage when the transfer process begins to occur. The factor F given by (7) represents the sum of spectroscopic factors partitioning $|nv\alpha\rangle$ into a couple of states $|n_1v_1\alpha_1\rangle$ and $|n_2v_2\alpha_2\rangle$ with α_1 and α_2 being varied. The value of F does not rely on α, the specification of the fixed-nv component of the initial state. It is evident that F is symmetric with respect to the interchange of $|n_1v_1\alpha_1\rangle$ and $|n_2v_2\alpha_2\rangle$.

We expect that the relation (7), or generally (4), with F being written as a closed and compact form, will serve global analyses of multiparticle transfer or nuclear fragmentation in heavy-ion reactions.[20]

The significance of (4) viewed from mathematical

 0022-2488/87/010128-06$02.50

physics lies in its interrelation with (2). The sum over v_2 on both sides of (4) produces (2), identifying

$$Z(nv,n_1v_1) = \sum_{v_2} F(nv,n_1v_1,n_2v_2). \tag{9}$$

We show in turn that the factor F can be expanded as a sum of products of a v_2-dependent factor and Z. These relations between Z and F as well as the simple form of F suggests that the relation (4) is more fundamental than (2).

In Sec. II, relations for the seniority projection operator $P(n_2,v_2)$ are deduced. In Sec. III, we verify (4) by representing its left-hand side (lhs) in terms of that of (2). At the same time, we get the expansion of F as a weighted sum of Z. Section IV deals with some properties of F. In Sec. V, a recurrence relation for F is shown. As its solution, we obtain the explicit F written in a compact form. In Sec. VI, asymptotic forms of F, of Z, and of the interrelation between F and Z are given for the dilute system where the Pauli principle is negligible. In Sec. VII, the factor F is described by means of the quasispin formalism. Short remarks are given in Sec. VIII. In Appendices A and B, we supply discussions for those in Secs. III and V, respectively.

II. RELATIONS FOR THE SENIORITY PROJECTION OPERATOR

Here, we deduce relations [(13), (15) and (26)] concerning the seniority projection operator.

A characteristic feature of the seniority scheme of identical particles lies in the simple (de)composition rule for seniority-0 components of operators and of wave functions. The essence, from which almost all of the present results stem, is represented as

$$A^+(n_1v\alpha)A^+(n_20) = \langle nv\alpha|A^+(n_1v\alpha)|n_20\rangle A^+(nv\alpha), \tag{10}$$

where the index α_2 of $A^+(n_20\alpha_2)$ is abbreviated. The matrix element on the rhs is evaluated by virtue of the Wigner–Eckart theorem in quasispin space[13] as

$$\begin{aligned}\langle nv\alpha|A^+(n_1v\alpha)|n_20\rangle &= \left(\frac{\Omega}{2}\,\frac{n_1}{2}\,\frac{n_2-\Omega}{2}\,\frac{n_1}{2}\,\Big|\,\frac{\Omega-v}{2}\,\frac{n-\Omega}{2}\right)\\ &\quad\times\left(\frac{\Omega}{2}\,\frac{n_1}{2}\,\frac{-\Omega}{2}\,\frac{n_1}{2}\,\Big|\,\frac{\Omega-v}{2}\,\frac{n_1-\Omega}{2}\right)^{-1} &(11)\\ &= \binom{(n-v)/2}{n_2/2}^{1/2}\binom{\Omega-(n_1+v)/2}{n_2/2}^{1/2}\binom{\Omega}{n_2/2}^{-1/2}, &(12)\end{aligned}$$

where Ω stands for half the number of single-particle states. In the following we dispense with the index α that appears on the lhs, since it is not involved on the rhs. By virtue of (10), the projection operator $P(n,v)$ is related to $P(n_1v)$, where $v\leqslant n_1\leqslant n$, by

$$P(n,v) = \frac{A^+(n-n_1\,0)P(n_1,v)A(n-n_1\,0)}{|\langle nv|A^+(n_1v)|(n-n_1)0\rangle|^2}. \tag{13}$$

The number projection operator $P(n)$ onto the n-body space can be expressed in terms of $P(n,v)$ as

$$P(n) = \sum_{v'} P(n,v'), \tag{14}$$

where v' runs over n, $n-2$, ..., 0 or 1.

Let us show a type of inversion of (14).

Proposition:

$$P(n,v) = P(n)\sum_l f(n,v,l)A^+(l0)A(l0), \tag{15}$$

with

$$\begin{aligned}f(n,v,l) &= \frac{\Omega+1-v}{(\Omega+1-(n+v)/2)}\\ &\quad\times\binom{l/2}{(n-v)/2}\binom{2\Omega+1-(n+v)/2}{\Omega-l/2}\\ &\quad\times\binom{2\Omega+1-(n+v)/2}{\Omega}(-1)^{(l-n+v)/2}. &(16)\end{aligned}$$

Here, l in (15) runs over $n-v$, $n-v+2$, ..., and $2[n/2]$, where [ ] stands for the largest integer contained in it.

Proof: Let us show that the rhs of (15) is transformed into the lhs. Substituting (14) into the rhs of (15), we transform the resultant operator product as

$$\begin{aligned}P(n,v)A^+(l0)A(l0) &= A^+(l0)P(n-l,v)A(l0) &(17)\\ &= |\langle nv|A^+((n-l)v)|l0\rangle|^2P(n,v), &(18)\end{aligned}$$

where we use (13) in the last step. Let us sum over l, using

$$\sum_q(-1)^q\binom{m}{q}\binom{x+q}{r+q} = (-1)^m\binom{x}{m+r}, \tag{19}$$

where $q=(l-n+v)/2$. Then, we see that all the terms with $v'\neq v$ vanish. The sum of terms with $v'=v$ yields $P(n,v)$, the lhs of (15).

Let us show a recurrence relation for $P(n,v)$. We start from

$$\begin{aligned}vP(v,v) &= \sum_\beta C_\beta^+C_\beta P(v,v) &(20)\\ &= \sum_\beta C_\beta^+P(v-1,v-1)C_\beta\{1-P(v,v-2)\}, &(21)\end{aligned}$$

where it is postulated that $P(n,v)=0$ if $v<0$ and $P(0,0)=1$. Using (13), we express $P(v,v-2)$ on the rhs of (21) in terms of $P(v-2,v-2)$. Subsequently, the relation (13) of Ref. 18 is utilized to commute C_β and the resultant $A^+(20)$. It then follows that

$$\begin{aligned}vP(v,v) &= \sum_\beta C_\beta^+P(v-1,v-1)C_\beta + \sum_{\beta\beta'}C_\beta^+\langle 1\beta|C_{\beta'}|20\rangle\\ &\quad\times\frac{P(v-1,v-1)C_{\beta'}^+A(20)}{|\langle vv-2|A^+(v-2v-2)|20\rangle|^2}. &(22)\end{aligned}$$

We transform $P(v-1,v-1)C_{\beta'}^+$ seen on the rhs as

$$\begin{aligned}P(v-1,v-1)C_{\beta'}^+ &= \{1-P(v-1,v-3)\}C_{\beta'}^+P(v-2,v-2) &(23)\\ &= C_{\beta'}^+P(v-2,v-2)\\ &\quad+A^+(20)\sum_{\beta''}C_{\beta''}P(v-2,v-2)\\ &\quad\times\frac{\langle 20|C_{\beta''}^+|1\beta'\rangle}{|\langle v-1v-3|A^+(20)|v-3v-3\rangle|^2}. &(24)\end{aligned}$$

After a slight rearrangement of the resultant rhs of (22), we get

$$vP(v,v) = \sum_\beta C_\beta^+ P(v-1,v-1)C_\beta - \left\{ \frac{2\Omega+4-v}{\Omega+3-v} \right\} P(v,v-2). \tag{25}$$

Using (13), we extend the last expression to the form

$$\left\{ \frac{v(2\Omega+2-n-v)}{\Omega+1-v} \right\} P(n,v) = 2\sum_\beta C_\beta^+ P(n-1,v-1)C_\beta - \left\{ \frac{(n-v+2)(2\Omega+4-v)}{(\Omega+3-v)} \right\} P(n,v-2), \tag{26}$$

a reduction relation for $P(n,v)$.

The operators $P(n)$ and $P(n,v)$ can be represented completely in terms of A^+ and A by using

$$|0\rangle\langle 0| = \sum_n (-1)^n \binom{\mathbf{n}}{n} = \sum_n (-1)^n \sum_{v\alpha} A^+(nv\alpha)A(nv\alpha), \tag{27}$$

where $\mathbf{n}$ stands for the number operator.

III. THE SUM OF SPECTROSCOPIC FACTORS AS A SUM OF PROPAGATION COEFFICIENTS

We transform the lhs of (4) using (10) and (15) as

$$\sum_\alpha \langle nv\alpha|A^+(n_1v_1\alpha_1)P(n_2,v_2)A(n_1v_1'\alpha_1')|nv\alpha\rangle = \sum_l f(n_2,v_2,l)\langle l'v_1|A^+(n_1v_1)l0\rangle\langle l0|A(n_1v_1')|l'v_1'\alpha_1'\rangle \times \sum_\alpha \langle nv\alpha|A^+(l'v_1\alpha_1)A(l'v_1'\alpha_1')|nv\alpha\rangle, \tag{28}$$

where $l' = n_1 + l$, and l runs over $n_2 - v_2, n_2 - v_2 + 2,\ldots,$ and $2[(n_2 - |v - v_1|)/2]$. Combining (28) with (2), we verify (4) with F being represented in terms of Z as

$$F(nv,n_1v_1,n_2v_2) = \sum_l (-1)^{(l-n_2+v_2)/2} Z(nv,n_1+lv_1)\frac{\Omega+1-v_2}{\Omega+1+(l-n_2-v_2)/2} \times \binom{l/2}{(n_2-v_2)/2}\binom{(n_1+l-v_1)/2}{l/2}\binom{\Omega-(n_1+v_1)/2}{l/2}\binom{\Omega+(l-n_2-v_2)/2}{l/2}^{-1}, \tag{29}$$

where both (12) and (16) are explicitly used. Notice that the argument v_2 on the rhs is not involved in Z.

The explicit form of Z was deduced in Ref. 6 as

$$Z(nv,n_1v_1) = Z(n-n_1+v_1v,v_1v_1)|\langle nv|A^+(n-n_1+v_1v)|n_1-v_10\rangle|^2/|\langle n_1v_1|A^+(v_1v_1)|n_1-v_10\rangle|^2 \tag{30}$$

$$= \frac{((n-v)/2)!(\Omega-(n_1+v_1)/2)!(\Omega+1-v_1)}{(n-n_1)!((n_1-v_1)/2)!(\Omega-(n+v)/2)!} \times \sum_m \frac{(-1)^m(n-n_1+v_1-2m)!(\Omega+m-(n+v-n_1+v_1)/2)!}{m!(v_1-2m)!(\Omega+1-v_1+m)!((n-n_1+v_1-v)/2-m)!}. \tag{31}$$

A reexpression of (31) is given in Appendix A.

Substituting (31) into (29) surely yields the explicit form of F. However, the resultant expression is too involved to treat. A compact form of F is deduced in Sec. V by means of another device.

IV. SOME PROPERTIES OF THE SUM OF SPECTROSCOPIC FACTORS

Here, some properties of F are summarized as a preparation for later discussions.

The particle–hole symmetry in the fermion system gives

$$\langle nv\alpha|A^+(n_1v_1\alpha_1)|n_2v_2\alpha_2\rangle = \langle(2\Omega-n_2)v_2\alpha_2|A^+(n_1v_1\alpha_1)|(2\Omega-n)v\alpha\rangle. \tag{32}$$

Using this and (4), we see that

$$\sum_{\alpha_1\alpha_2} |\langle nv\alpha|A^+(n_1v_1\alpha_1)|n_2v_2\alpha_2\rangle|^2 = \text{independent of } \alpha. \tag{33}$$

Let us sum over α on both sides and, subsequently, apply (4) to the resulting lhs. Then, we obtain (7). The expression (32) gives the symmetry relation for F as

$$F(nv,n_1v_1,n_2v_2) = F((2\Omega-n_2)v_2,n_1v_1,(2\Omega-n)v)\frac{d(v_2)}{d(v)}. \tag{34}$$

The relation (4) is reduced to (2) in the cases of $v_1 = 0$, $n_1 = n-1$, $v = n - n_1 - v_1$, or 0, since in these cases $P(n_2,v_2)$ in (4) can be put to unity. In the case of $v_1 = 0$, it follows that

$$F(nv,n_10,n_2v) = Z(nv,n_10) = \text{square of the rhs of (12) with } n_1 \text{ being interchanged by } n_2. \tag{35}$$

In the case of $n_1 = n - 1$, we get

$$F(nv,n-1v_1,11) = Z(nv,n-1v_1) = \begin{cases} v(2\Omega+2-n-v)/(2\Omega+2-2v), & \text{if } v_1 = v-1 \\ (n-v)(2\Omega+2-v)/(2\Omega+2-2v), & \text{if } v_1 = v+1, \end{cases} \tag{36}$$

a well-known expression as the special orthogonality relation for $n \to (n-1)+1$ cfp's.[12] The dimensionality $d(v)$ is describable[9] in terms of Z and F as

$$d(v) = Z(2\Omega 0,nv) = F(2\Omega 0,nv,(2\Omega - n)v). \tag{37}$$

It is explicitly given by

$$d(v) = \frac{\Omega+1-v}{\Omega+1}\binom{2\Omega+2}{v}. \tag{38}$$

The factor $F(nv,n_1v_1,n_2v_2)$ is a polynomial in the quasispin S^2, i.e.,

$$(n-v)(2\Omega+2-n-v)/2. \tag{39}$$

We see it from (29) combined with the fact that the coefficient Z is a polynomial[6] in S^2.

V. THE SUM RULE FOR SPECTROSCOPIC FACTORS

In this section, we deduce a compact form of F as a solution of a recurrence relation for F.

Substituting (26) into (4), we get

$$\left\{\frac{v_2(2\Omega+2-n_2-v_2)}{(\Omega+1-v_2)}\right\}F(nv,n_1v_1,n_2v_2) = 2\sum_{v'} F(nv,(n_1+1)v',(n_2-1)(v_2-1))F((n_1+1)v',n_1v_1,11)$$
$$-\left\{\frac{(n_2-v_2+2)(2\Omega+4-v_2)}{(\Omega+3-v_2)}\right\}F(nv,n_1v_1,n_2(v_2-2)). \tag{40}$$

Here, the factor $F(...,11)$ seen on the rhs is explicitly given by (36). To solve (40), we start from the case of $v_2 = 1$. Using (35), we get the explicit form of $F(nv,n_1v_1,n_21)$. Next, let us put $v_2 = 2$ into (40). The factor $F(nv,n_1v_1,n_22)$ is then represented in terms of $F(nv,(n_1+1)v',(n_2-1)1)$ and $F(nv,n_1v_1,n_20)$ whose explicit forms have been deduced. Repeating this way, we obtain

$$\begin{aligned} F(nv,n_1v_1,n_2v_2) &\equiv \binom{n}{n_1}\sum_{\alpha_1\alpha_2} |\langle n_2v_2\alpha_2 + n_1v_1\alpha_1|nv\alpha\rangle|^2 \\ &= \left\{\frac{(\Omega+1-v_1)(\Omega+1-v_2)}{(\Omega+1)(v+1)}\binom{\Omega+1}{(v+v_1-v_2)/2}\binom{\Omega+1}{(v+v_2-v_1)/2}\binom{2\Omega+2-v}{(v_1+v_2-v)/2}\right. \\ &\quad\times\binom{2\Omega-(v+v_1+v_2)/2}{\Omega}\binom{\Omega+1-(v_1+v_2-v)/2}{v+1}^{-1}\Bigg\}\Bigg\{\binom{\Omega-(v+v_1+v_2)/2}{\Omega-(n+v)/2} \\ &\quad\times\left.\binom{(n-v_1-v_2)/2}{(n_1-v_1)/2}\left[\binom{2\Omega-(v+v_1+v_2)/2}{(n-v)/2}\binom{2\Omega-(n+v_1+v_2)/2}{\Omega-(n_1+v_1)/2}\right]^{-1}\right\}, \end{aligned} \tag{41}$$

where the first equality is the summary of (7) combined with (3). The rhs of (41) is factorized, though not in a unique way, into two parts, one depending on any of n_1, n_2, and n, and the other written only in terms of v_1, v_2, v, and Ω.

The expression (41) has a remarkably compact form in comparison with the expression of Z, i.e., (31): we see the convenience of (41) by checking (37). Substitution of (31) into (29) would agree with (41), though it is difficult to check. Substitution of (41) into (9) yields a new type of expression of Z.

We show in Appendix B that the relation (40), from which (41) is obtained, is derivable without recourse to the seniority projection operator. We actually find, using (B5), that expression (41) is valid also for the seniority scheme of identical bosons,[21] if we have only to replace Ω here by $-\Omega$ and to take the absolute value of the resultant expression.[22]

It is not easy to give a physical interpretation for (41). The following two sections are devoted to extract some properties embedded in (41).

VI. ASYMPTOTIC FORMS IN THE DILUTE SYSTEM

We present here asymptotic forms of various expressions in the dilute system, i.e., the system with $\Omega \gg n$, where the correction due to Pauli principle disappears.

From (41), it follows that

$$\lim_{\Omega\to\infty} F(nv,n_1v_1,n_2v_2) = 2^{(v_1+v_2-v)/2}\binom{v}{(v+v_2-v_1)/2}\times\binom{(n-v_1-v_2)/2}{(n_1-v_1)/2}\binom{(n-v)/2}{(v_1+v_2-v)/2}. \tag{42}$$

In the case $v = v_1 + v_2$, the rhs gives the number of ways to pick up v_1 and $n_1 - v_1$ particles, respectively, from v unpaired and $n - v$ paired particles. If $n \gg v_1$, v_2, v, and n_1, which is often realized in practical cases, the rhs with v_2 being varied takes the maximum value at $v_2 = v_1 + v$. The relations (29) and (31) have asymptotic forms

$$\lim_{\Omega\to\infty} F(nv,n_1v_1,n_2v_2) = \binom{(n-v_1-v_2)/2}{(n_2-v_2)/2} \sum_l \binom{(n_1+l-v_1)/2}{(n-v_1-v_2)/2} \times(-1)^{(l-n_2+v_2)/2} \lim_{\Omega\to\infty} Z(nv,n_1+lv_1) \tag{43}$$

and

$$\lim_{\Omega\to\infty} Z(nv,n_1v_1) = \binom{(n-v)/2}{(n_1-v_1)/2} \sum_m (-1)^m \binom{(n-n_1+v_1-v)/2}{m} \times\binom{n-n_1+v_1-2m}{n-n_1}, \tag{44}$$

respectively. Expressions (42) and (44) are associated with each other by (9), as is shown by using the identity

$$\sum_m 2^m \binom{p}{m}\binom{q}{m+r} = \sum_m (-1)^m \binom{p}{m}\binom{2p+q-2m}{2p+r}. \tag{45}$$

To see the consistency of (42)–(44), we substitute (44) into (43) and, subsequently, sum over l and m by using (45), etc. Then, we get (42).

VII. DESCRIPTION BY MEANS OF THE QUASISPIN FORMALISM

We point out here that the relation (41) contains a new relation that is not easy to deduce by means of the quasispin formalism.[13]

The factor F is easily transformed, by using the Wigner–Eckart theorem in quasispin space[13] as

$$\begin{aligned} F(nv,n_1v_1,n_2v_2) &= \left(\frac{\Omega-v_1}{2}\frac{n_2}{2}\frac{n_1-\Omega}{2}\frac{n_2}{2}\,\Big|\,\frac{\Omega-v}{2}\frac{n-\Omega}{2}\right)^2 \\ &\times\left(\frac{\Omega-v_2}{2}\frac{v_1}{2}\frac{n_2-\Omega}{2}\frac{v_1}{2}\,\Big|\,\frac{\Omega-v}{2}\frac{n_2+v_1-\Omega}{2}\right)^2 \\ &\times\left\{(\Omega-v+1)\left(\frac{\Omega-v_1}{2}\frac{n_2}{2}\frac{v_1-\Omega}{2}\frac{n_2}{2}\,\Big|\,\frac{\Omega-v}{2}\frac{n_2+v_1-\Omega}{2}\right)^2\right\}^{-1} \\ &\times\sum_{\alpha\alpha_2}\left|{}_s\!\left\langle\frac{(\Omega-v)}{2}\alpha\right|\left|T\left(\frac{v_1}{2},v_1\alpha_1\right)\right|\left|\frac{(\Omega-v_2)}{2}\alpha_2\right\rangle_s\right|^2\frac{d(v_1)}{d(v)}. \end{aligned} \tag{46}$$

Here, the operator $T(SS_z,v_1\alpha_1)$ denotes the quasispin tensor of the rank S with its z component S_z that belongs to the same multiplet as $A^+(n=2S,v_1\alpha_1)$ does: for example, the quasispin S^+ is represented as $\sqrt{\Omega}\,T(11,0)$. The state $|nv\alpha\rangle$ is specified in the quasispin space as $|(\Omega-v)/2\;(n-\Omega)/2\;\alpha\rangle_s$.

The dependence of F on n_1, n_2 and n is surely obtained from (46), and the result agrees with the relevant part of (41). However, it is difficult to get the explicit dependence of F on v etc. by means of the quasispin formalism.

Let us compare (46) with (41). Then, we obtain

$$\begin{aligned} &\sum_{\alpha\alpha_2}\left|{}_s\!\left\langle\frac{(\Omega-v)}{2}\alpha\right|\left|T\left(\frac{v_1}{2},v_1\alpha_1\right)\right|\left|\frac{(\Omega-v_2)}{2}\alpha_2\right\rangle_s\right|^2 \\ &= \frac{(\Omega+1-v_2)(\Omega+1-v)}{\Omega+1-v_1}\binom{2\Omega+2-v_1}{(v+v_2-v_1)/2} \\ &\times\binom{2\Omega+2-v_1}{\Omega+1-(v+v_1-v_2)/2}\binom{2\Omega+2-v_1}{\Omega+1}^{-1}, \end{aligned} \tag{47}$$

a relation that has not been discussed before.

The multiplet of the operator T is describable in terms of normal products as

$$\begin{aligned} T(v/2(v/2-k),v\alpha) &= \sum_{\alpha'\alpha''}(-1)^{k(k-1)/2}\langle kk\alpha''|A(v'v'\alpha')|vv\alpha\rangle \\ &\times A^+(v'v'\alpha')\tilde{A}(kk\alpha'')\binom{v}{k}^{-1/2}, \end{aligned} \tag{48}$$

where $v' = v - k$ and, in the case of $\alpha = (\beta JM)$,

$$\tilde{A}(nv\beta JM) = (-1)^{J-M}A(nv\beta J-M). \tag{49}$$

We can verify (48), applying S_- by k times to $A^+(vv\alpha)$. The expression (48) makes it possible to modify (47) further.

VIII. REMARKS

Expression (41) shows that the spectroscopic factors are summed into the closed algebraic form including no summation. The result looks still involved, which, however, is due to several arguments in it and to the effect of the Pauli principle. The asymptotic form (42), where the effect of the Pauli principle disappears, clarifies a specific form of (41).

In the analysis of the multinucleon transfer in the heavy reaction, it will reasonably be assumed that the initial state to be fragmented is statistically distributed with probability proportional to $\exp(-E/T)$, where E stands for the excitation energy, roughly proportional to the seniority[12] v, and T the temperature relying on the incident beam energy. Relation (41), which is used for the spectral analysis of projectile (or target) residues, will then be accompanied by multiplication of the statistical factor and the sum over v. Seniority schemes for protons and for neutrons are to be treated separately. Study on projectile residues along this line will be given in a future paper.

APPENDIX A: ANOTHER REEXPRESSION OF Z

The seniority projection operator $P(n,v)$, rewritten in form (15), makes it possible to transform the fixed-seniority average into a normal average as

$$\sum_\alpha \langle nv\alpha|O|nv\alpha\rangle = \sum_\gamma \langle n\gamma|OP(n,v)|n\gamma\rangle, \tag{A1}$$

where O stands for an operator or an operator product that is not always arranged in normal products. The index γ specifies the complete set of the n-body states. For evaluation of the average, we have only to apply (14) of Ref. 8 to the rhs into which (15) is substituted.

As a simple example, we consider the case when $O = A^+(v_1v_1\alpha_1)A(v_1v_1\alpha_1)$. It is easy to obtain

$$Z(nv,v_1v_1)$$

$$\equiv \sum_\alpha \langle nv\alpha | A^+(v_1v_1\alpha_1)A(v_1v_1\alpha_1)|nv\alpha\rangle \frac{d(v_1)}{d(v)} \quad (A2)$$

$$= \sum_{m=0}^{[v/2]} (-1)^m \frac{\Omega+1-v_1}{\Omega+1-v+m} \binom{(n-v)/2+m}{m} \times \binom{2\Omega+2}{v_1} \left\{ \binom{\Omega-v+m}{(n-v)/2+m} \binom{2\Omega+2}{v} \right\}^{-1} \times \sum_l \binom{v_1}{l} \binom{\Omega-l}{\Omega-(n-v)/2-m} \times \binom{2\Omega-n+v-v_1-2m}{2\Omega-n-l}. \quad (A3)$$

This gives another expression of Z different from (31) and from (9) linked with (41), though these are algebraically identical. In the dilute system, the $m \neq 0$ term on the rhs of (A3) vanishes and we get

$$\lim_{\Omega\to\infty} Z(nv,v_1v_1) = \sum_l 2^l \binom{(n-v)/2}{l} \binom{v}{k-l}, \quad (A4)$$

which agrees with (44), with $n_1 = v_1$ due to the identity (45).

APPENDIX B: THE REDUCTION RELATION FOR F

Here, we deduce (40) without recourse to the seniority projection operator. This derivation is applicable to a general fixed-symmetry average.

Let us consider the quantity

$$Q = \sum_{\alpha\alpha_0\alpha_1} \langle nv\alpha | A^+(n_1v_1\alpha_1)A^+(n_0v_0\alpha_0) \times P(n-n_0-n_1,v_m)A(n_0v_0\alpha_0)A(n_1v_1\alpha_1)|nv\alpha\rangle. \quad (B1)$$

We transform it as

$$Q = \sum_{v'\alpha'v''\alpha''} \Big(\sum_{\alpha_0} \langle (n-n_1)v'\alpha' | A^+(n_0v_0\alpha_0) \times P(n-n_0-n_1,v_m)A(n_0v_0\alpha_0)|(n-n_1)v''\alpha''\rangle \Big) \times \Big(\sum_\alpha \langle nv\alpha | A^+((n-n_1)v'\alpha') \times P(n_1,v_1)A((n-n_1)v''\alpha'')|nv\alpha\rangle \Big). \quad (B2)$$

Using (4), verified in Sec. III, we represent the rhs in terms of F as

$$Q = d(v) \sum_{v'} F((n-n_1)v',n_0v_0,(n-n_0-n_1)v_m) \times F(nv,(n-n_1)v',n_1v_1), \quad (B3)$$

where v' runs over the range

$$|v_m - v_0| \leqslant v' \leqslant v_m + v_0 \text{ and } |v - v_1| \leqslant v' \leqslant v + v_1, \quad (B4)$$

according to the selection rule for seniority. The rhs of (B3) is to be symmetric with respect to interchange of the set $(n_0v_0\alpha_0)$ and the set $(n_1v_1\alpha_1)$, which implies, in the case of $n_0 = 1$,

$$F(n_2(v_m-1),(n_2-1)v_m,11)F(nv,n_2(v_m-1),n_1v_1) + F(n_2(v_m+1),(n_2-1)v_m,11) \times F(nv,n_2(v_m+1),n_1v_1) = F(nv,(n-1)(v-1),11) \times F((n-1)(v-1),(n_2-1)v_m,n_1v_1) + F(nv,(n-1)(v+1),11) \times F((n-1)(v+1),(n_2-1)v_m,n_1v_1), \quad (B5)$$

where $n_2 = n - n_1$. The first term on the lhs (rhs) is to be excluded in the case of $v_m = 0$ $(v = 0)$.

Let us substitute (34) into each of $F(\ldots,n_1v_1)$ on the rhs of (B5) and, subsequently, change notations as $n_1 \to 2\Omega - n$, $n_2 \to n_2 + 1$, $n \to 2\Omega + 1 - n_1$, $v_1 \to v$, $v_m \to v_2$, and $v \to v_1 - 1$. Then, we obtain (40).

The relation (B5) can be regarded as a reduction relation for F with n_1 and v_1 being fixed. The same type of relations in various coupling schemes other than the seniority scheme are deducible for a fermion and a boson systems alike. Summing over v_m on both sides of (B5) by the use of (9) and (36), we get the Chapman–Kolmogorov equation for Z, which was given by (12′) of Ref. 9.

[1]G. Racah, in *Farkas Memorial Volume* (Research Council of Israel, Jerusalem, 1952), p. 294.
[2]S. A. Moszkowski, Prog. Theor. Phys. **28**, 1 (1962).
[3]J. B. French, in *Nuclear Structure*, edited by A. Hossain, Harum-ar-Raschid, and M. Islam (North-Holland, Amsterdam, 1967), p. 85.
[4]M. Ichimura, Nucl. Phys. A **131**, 398 (1969).
[5]M. Nomura, Prog. Theor. Phys. **47**, 1858 (1972).
[6]M. Nomura, Prog. Theor. Phys. **51**, 489 (1974).
[7]C. Quesne and S. Sptiz, Ann. Phys. (NY) **112**, 304 (1978).
[8]M. Nomura, J. Math. Phys. **26**, 738 (1985).
[9]M. Nomura, J. Math. Phys. **26**, 965 (1985).
[10]*Theory and Applications of Moment Methods in Many-Fermion Systems*, edited by B. J. Dalton, S. M. Grimes, J. P. Vary, and S. A. Williams (Plenum, New York, 1980).
[11]G. Racah, Phys. Rev. **63**, 367 (1943).
[12]A. de-Shalit and I. Talmi, *Nuclear Shell Theory* (Academic, New York, 1963).
[13]R. D. Lawson and M. H. Macfarlane, Nucl. Phys. **66**, 80 (1965).
[14]A. Arima and H. Kawarada, J. Phys. Soc. Jpn. **19**, 1768 (1964).
[15]M. Moshinsky and C. Quesne, J. Math. Phys. **11**, 1631 (1970).
[16]P. J. Brussaard and P. W. M. Glaudemans, *Shell-Model Applications in Nuclear Spectroscopy* (North-Holland, Amsterdam, 1977).
[17]M. Nomura, Prog. Theor. Phys. **59**, 451 (1978).
[18]M. Nomura, J. Math. Phys. **26**, 732 (1985).
[19]M. H. Macfarlane and J. B. French, Rev. Mod. Phys. **32**, 567 (1960).
[20]X. Campi, J. Desbois, and E. Lipparini, Nucl. Phys. A **428**, 327 (1984).
[21]M. Nomura, Nucl. Phys. A **150**, 193 (1970).
[22]M. Nomura, Ann. Phys. (NY) **91**, 83 (1975).

The quasiperiodic solutions to the discrete nonlinear Schrödinger equation

Siraj Ahmad and A. Roy Chowdhury
High Energy Physics Division, Department of Physics, Jadavpur University, Calcutta-700 032, India

(Received 6 January 1986; accepted for publication 28 July 1986)

The quasiperiodic solutions to the discrete nonlinear Schrödinger equation are obtained by a variant of a method due to Date and Tanaka [E. Date and S. Tanaka, Suppl. Prog. Theor. Phys. **59**, 107 (1976)]. It is shown explicitly that the nonlinear field variable at the different lattice points can be determined in a recursive fashion in terms of combinations of Reimann's functions depending on lattice position and time.

I. INTRODUCTION

Periodic solutions to nonlinear partial differential equations have attracted the attention of researchers over the last decade. At present there exist at least three different approaches for obtaining such solutions. The most general method being that of Krichever and Dubrovin[1] based on the technique of algebrogeometry. Next is the age-old technique[2] of analyzing the periodic spectrum of the associated linear operator as in the case of the full line IST problem. Third is the method put forward by Date and Tanaka[3] and elaborated by Forest and McLaughlin.[4] Though in the paper of Date and Tanaka both the cases of continuous and discrete nonlinear equations are considered, here we show that a variation of this method can lead to an elegant formulation of the periodic-inverse problem for a discrete nonlinear equation. We present the methodology by an analysis of the discrete nonlinear Schrödinger equation (DNLSE).

II. THE DNLSE AND THE INVERSE PROBLEM

The equation under consideration is

$$iq_{n_t} = [1/(\Delta x)^2](q_{n+1} + q_{n-1} - 2q_n) \pm q_n q_n^*(q_{n+1} + q_{n-1}). \tag{1}$$

The inverse problem associated with Eq. (1) is written as

$$\vartheta_{n+1} = F_n(z)\vartheta_n, \quad \frac{\partial \vartheta_n}{\partial t} = G_n(z)\vartheta_n, \tag{2}$$

where $F_n(z), G_n(z)$ are the matrices

$$F_n(z) = \begin{pmatrix} z & q_n \Delta x \\ \mp q_n^* \Delta x & \bar{z}^1 \end{pmatrix}, \tag{3}$$

$$G_n(z) = \frac{i}{(\Delta x)^2} \begin{pmatrix} 1 - z^2 + (\Delta x)^2 q_n q_{n-1}^* & \Delta x(-q_n z + z^{-1} q_{n-1}) \\ \pm \Delta x(q_{n-1}^* z - \bar{z}^1 q_n^*) & -(1 - \bar{z}^2 \mp (\Delta x)^2 q_{n-1} q_n^*) \end{pmatrix}, \tag{4}$$

and ϑ_n is a two-component vector $(\vartheta_{1n}, \vartheta_{2n})$. The consistency between (3) and (4) is equivalent to (1) and is written as

$$\frac{\partial F_n}{\partial t} = G_{n+1} F_n - F_n G_n. \tag{5}$$

We now assume that the nonlinear field q_n obeys the periodic boundary condition $q_{n+N+1} = q_n$ with period $N+1$, where N is an arbitrary non-negative integer. From the first equation of (2) we observe that F_n can be interpreted as the transfer matrix over the single lattice site, so that we can define the translation operator $H_n(z)$ by

$$\vartheta_{n+N+1} = H_n(z)\vartheta_n. \tag{6}$$

Then it is easy to ascertain that

$$H_n(z) = \overset{\curvearrowleft}{\prod_{j=1}^{N}} F_{j+1}(z), \tag{7}$$

where the curved arrow indicates that the order of increase of the indices in the product.

The monodromy matrix $H_n(z)$ satisfies the equations

$$H_{n+1} F_n = F_n H_n, \quad \frac{\partial H_n}{\partial t} = G_n H_n - H_n G_n. \tag{8}$$

The matrix H_n is such that

$$\det H_n = \det H_m, \quad S\beta H_n = S\beta H_m,$$
$$\frac{\partial}{\partial t}\det H_n = \frac{\partial}{\partial t}\det H_m = 0, \tag{9}$$
$$\frac{\partial}{\partial t} S\beta H_n = 0, \quad n,m = 0,1,2,\ldots .$$

Let us consider the matrix H_n as

$$H_n = \begin{pmatrix} \varphi_n + f_n & -g_n \\ h_n & \varphi_n - f_n \end{pmatrix}. \tag{10}$$

Using these conditions Eq. (8) becomes

$$\begin{aligned} z(f_{n+1} - f_n) &= \mp g_{n+1} q_n^* \Delta x + h_n q_n \Delta x, \\ \bar{z}^1(f_{n+1} - f_n) &= h_{n+1} q_n \Delta x - g_n q_n^* \Delta x, \\ -z g_n + \bar{z}^1 g_{n+1} &= q_n \Delta x(f_{n+1} + f_n), \\ -\bar{z}^1 h_n + z h_{n+1} &= \mp q_n^* \Delta x(f_{n+1} + f_n). \end{aligned} \tag{11}$$

The corresponding time evolutions are governed by

$$\frac{\partial f_n}{\partial t} = [i/(\Delta x)^2][\Delta x(-q_n z + \bar{z}^1 q_{n-1})h_n \pm \Delta x(q^*_{n-1} z - \bar{z}^1 q^*_n)g_n],$$

$$\begin{aligned}\frac{\partial g_n}{\partial t} &= [i/(\Delta x)^2[2f_n \Delta x(-q_n z + \bar{z}^1 q_{n-1}) \\ &\quad + g_n\{(2 - z^2 - \bar{z}^2) \\ &\quad \mp (\Delta x)^2(q_n q^*_{n-1} + q_{n-1} q^*_n)\}],\end{aligned} \tag{12}$$

$$\begin{aligned}\frac{\partial h_n}{\partial t} &= [i/(\Delta x)^2][\pm 2f_n \Delta x(q^*_{n-1} z - \bar{z}^1 q^*_n) \\ &\quad + h_n\{(-2 + z^2 + \bar{z}^2) \\ &\quad \pm (\Delta x)^2(q_n q^*_{n-1} + q^*_n q_{n-1})\}].\end{aligned}$$

III. SOLUTION OF THE PERIODIC PROBLEM

It is now interesting to observe that Eqs. (11) and (12) are analogs of those equations satisfied by the square eigenfunctions f^2, g^2, and fg, as deduced by Date and Tanaka and Forest and McLaughlin.[4] Indeed we can deduce that the quantity

$$f_n^2 - g_n h_n = P(z^2) \tag{13}$$

is independent of h and t, so that it is an absolute constant. With these comments in mind we now represent the function $P(z^2)$ as

$$P(z^2) = \sum_{k=0}^{2N+2} P_k z^{2k} = P_{2N+2} \prod_{j=1}^{2N+2} (z^2 - E_j), \tag{14}$$

so that the E_j are the zeros of $P(z^2)$. Now we try a polynomial solution for f_n, g_n, and h_n in the form

$$\begin{aligned} f_n(z,t) &= \sum_{k=0}^{N+1} f_n^{(k)}(t) z^{2k}, \\ h_n(z,t) &= \sum_{k=0}^{N} h_n^{(k)}(t) z^{2k+1}, \\ g_n(z,t) &= \sum_{k=0}^{n} g_n^{(k)}(t) z^{2k+1}. \end{aligned} \tag{15}$$

Substituting these in Eqs. (11) and (12) and comparing various powers of z, we obtain

$$f_{n+1}^{(N+1)} = f_n^{(N+1)}, \tag{16}$$

$$f_{n+1}^{(0)} - f_n^{(0)} = \Delta x[\mp q^*_n q^{(0)}_{n+1} + q_n h_n^{(0)}]. \tag{17}$$

From the third equation of the set (11) we deduce

$$(f_{n+1}^{(N+1)} + f_n^{(N+1)}) q_n \Delta x = -q_n^{(N)} \text{ or } q_n = -\frac{q_n^{(N)}}{2\Delta x f_n^{(N+1)}}. \tag{18a}$$

Also we get

$$g_{n+1}^{(0)} = \Delta x q_n [f_{n+1}^{(0)} + f_n^{(0)}], \tag{18b}$$

and from the same set equating term involving z^{2N+2} we get

$$q^*_{n-1} = \mp h_n^{(N)}/2\Delta x f_n^{(N+1)}. \tag{19a}$$

Then

$$-h_n^{(0)} = \mp q^*_n \Delta x(f_{n+1}^{(0)} + f_n^{(0)}). \tag{19b}$$

From (17), (18b), and (19b) we deduce

$$f_{n+1}^{(0)} = f_n^{(0)}. \tag{20}$$

In addition we also take notice of the following important relations:

$$\begin{aligned} f_{n(z,t)} &= z^{2(N+1)} f_n(\bar{z}^1, t), \\ q_n h_n(z,t) &= \mp z^{2(N+1)} q^*_n g_n(\bar{z}^1, t), \\ q^*_n g_n(z,t) &= -z^{2(N+1)} q_n h_n(\bar{z}^1, t). \end{aligned} \tag{21}$$

Using expansion (15) in (21) we obtain

$$\begin{aligned} f_n^{(k)}(t) &= f_n^{(N+1-k)}(t), \\ q_n h_n^{(N)}(t) &= \mp q^*_n g_n^{(0)}(t), \\ q^*_n g_n^{(N)}(t) &= -q_n h_n^{(0)}(t). \end{aligned} \tag{22}$$

From these equations follow two important relations;

$$\begin{aligned} q_n h_n^{(N)} &= q^*_n (f_n^{(0)}/f_n^{(N+1)*}) h_n^{(N)*}, \\ q^*_n g_n^{(N)} &= \pm q_n (f_n^{(0)}/f_n^{(N+1)*}) g_n^{(N)*}. \end{aligned} \tag{23}$$

As in (14) we now postulate the existence of the zeros of the function $g_n(z,t)$ and write it in the form

$$g_n(z,t) = z g_n^{(N)} \prod_{j=1}^{N} (z^2 - \mu_j(n,t)). \tag{24}$$

When this is used in conjunction with (21), we deduce

$$\frac{h_n^{(N)*}}{g_n^{(N)}} = \pm \frac{f_n^{(N+1)*}}{f_n^{(0)}} (-1)^{N+1} \prod_{j=1}^{N} \mu_j(n,t).$$

Also, from (18a) and (19a),

$$\frac{q_{n-1}}{q_n} = \pm \frac{h_n^{(N)*}}{g_n^{(N)}} \frac{f_n^{(N+1)}}{f_n^{(N+1)*}},$$

which finally leads to

$$q_{n-1}(t) = q_n(t)(-1)^{N+1} \prod_{j=1}^{N} \mu_j(n,t). \tag{25}$$

Thus the lattice field $q_n(t)$ obeys a recursion relation, given via $\mu_j(n,t)$, indicating the fact that they can be explicitly determined if we can have a determination of $\mu_j(n,t)$.

To deduce such an equation we equate coefficients of z^{2N+1} from both sides of the second equation of (12):

$$\begin{aligned}\frac{\partial g_n^{(N)}}{\partial t} &= [i/(\Delta x)^2][2\Delta x q_{n-1} f_n^{(N+1)} - 2\Delta x q_n f_n^{(N)} + g_n^{(N)} \\ &\quad \times \{1 \mp (\Delta x)^2(q_n q^*_{n-1} + q_{n-1} q^*_{n-1})\} - g_n^{N-1}].\end{aligned} \tag{26}$$

We then use

$$\begin{aligned} g_n^{(N)} &= -2f_n^{(N+1)} q_n \Delta x, \\ g_n^{(N-1)} &= -g_n^{(N)} \sum_{j=1}^{N} \mu_j = 2\Delta x f_n^{(N+1)} q_n \sum_{j=1}^{N} \mu_j, \\ (f_n^{(N+1)})^2 &= P_{2N+2}, \\ 2f_n^{(N+1)} f_n^{(N)} - g_n^{(N)} h_n^{(N)} &= P_{2N+1}, \\ \frac{f_n^{(N)}}{f_n^{(N+1)}} &= \mp 2q_n q^*_{n-1}(\Delta x)^2 + \frac{P_{2N+1}}{2P_{2N+2}}. \end{aligned} \tag{27}$$

This immediately leads to

$$
\begin{aligned}
&-2f_n^{(N+1)}\frac{\partial q_n}{\partial t}\Delta x \\
&=\frac{i}{(\Delta x)^2}\Big[2\Delta x q_{n-1}f_n^{(N+1)}-2\Delta x q_n f_n^{(N)} \\
&\quad -2f_n^{(N+1)}q_n\Delta x\{!\mp(\Delta x)^2(q_nq_{n-1}^*+q_{n-1}q_n^*)\} \\
&\quad -2\Delta x f_n^{(N+1)}q_n\sum\mu_j\Big],
\end{aligned}
$$

which finally, by use of (25), can be converted to

$$
\begin{aligned}
\frac{\partial}{\partial t}\ln q_n=\frac{i}{(\Delta x)^2}\Big[&1\pm 3|q_n|^2(\Delta x)^2(-1)^N\prod_{j=1}^N\mu_j^*(n,t) \\
&+\frac{P_{2N+1}}{2P_{2N+2}}+(-1)^N\prod_{j=1}^N\mu_j(n,t)+\sum_{j=1}^N\mu_j(n,t) \\
&\pm|q_n|^2(\Delta x)^2(-1)^N\prod_{j=1}^N\mu_j(n,t)\Big]. \qquad (28)
\end{aligned}
$$

We now use the original nonlinear difference equation shifted by 1:

$$
i\frac{\partial q_{n-1}}{\partial t}=\frac{q_n+q_{n-2}-2q_{n-1}}{(\Delta x)^2}\pm q_{n-1}q_{n-1}^*(q_n+q_{n-2}).
$$

Now equating the two expressions of $\partial q_{n-1}/\partial t$, we arrive at

$$
\begin{aligned}
&q_n+q_{n-1}\sum_{j=1}^N\mu_j(n,t) \\
&=\pm q_{n-1}\Big[3q_{n-1}q_{n-2}^*(\Delta x)^2-q_{n-1}^*q_n(\Delta x)^2 \\
&\quad \mp\frac{P_{2N+1}}{2P_{2N+2}}\pm 1\Big]. \qquad (29)
\end{aligned}
$$

Equations (25), (28), and (29) will be of utmost importance in the computations that will follow.

IV. MOTION OF THE ZEROS $\mu_j(n,t)$

Since we have assumed $g_n(t,z)$ to have the zeros $\mu_j(n,t)$,

$$
g_n(t,z)=zg_n^{(N)}(t)\prod_{j=1}^N(z^2-\mu_j(n,t)).
$$

Differentiating this equation with respect to t, we get

$$
\begin{aligned}
&z\frac{\partial g_n^{(N)}}{\partial t}\prod_{j=1}^N(z^2-\mu_j)+zg_n^{(N)}\prod_{j=1}^N(z^2-\mu_j)\Big(\frac{\partial\mu_j}{\partial t}\Big) \\
&=\frac{i}{(\Delta x)^2}\Big[2\Delta x f_n(-q_nz+\bar z^1q_{n-1})+zg_n^{(N)} \\
&\quad\times\prod_{j=1}^N(z^2\mu_j)\{1-z^2-\bar z^2\mp(q_nq_{n-1}^*+q_{n-1}q_n^*)\}\Big].
\end{aligned}
$$

Setting $z^2=\mu_j$, we obtain

$$
2f_n^{(N+1)}q_n\Delta x\frac{\partial\mu_j}{\partial t}=\frac{2if_n}{\Delta x}\frac{1}{\Pi_{l\neq j}(\mu_j-\mu_l)}\Big(-q_n+\frac{1}{\mu_j}q_{n-1}\Big),
$$

which finally leads to

$$
\begin{aligned}
\frac{\partial\mu_j}{\partial t}=\frac{i}{(\Delta x)^2}&\frac{\sqrt{P(\mu_j)}}{f_n^{(N+1)}\Pi_{l\neq j}(\mu_j-\mu_l)} \\
&\times\Big[-1+(-1)^{N+1}\prod_{l\neq j}\mu_l\Big]. \qquad (30)
\end{aligned}
$$

V. EXPLICIT SOLUTION FOR $\mu_j(n,t)$

Equation (30) shows that the variables $\mu_j(n,t)$ satisfy a complicated system of dynamical equations. This flow can be straightened out by exploiting the fact that each μ_j resides on the Reimann surface of

$$
R^2(E)=\prod_{j=1}^{2N+2}(E-E_j). \qquad (31)
$$

We define N Abelian differentials of the first kind on it by

$$
du_\nu=\frac{c_{\nu_1}E^{N-1}+\cdots+c_{\nu N}}{R(E)}dE,\quad \nu=1,2,\dots,N.
$$

The matrix of the constants $c_{\nu\mu}$ are fixed in terms of $\{E_j\}$ and the normalization conditions

$$
\oint_a du_\nu=\delta_{\nu\mu},
$$

while the "b" cycles define the period matrix

$$
B_{\mu\nu}=\int_b du_\nu.
$$

On the Reimann surface designated in this manner, Eq. (30) can be written as

$$
\begin{aligned}
\frac{\partial\mu_j}{\partial t}=\frac{i}{(\Delta x)^2}&\Big[-1+(-1)^{N+1}\prod_{l\neq j}\mu_l\Big] \\
&\times\frac{(\Pi_{k=1}^{2N+2}(\mu_j-E_k)^{1/2})}{\Pi_{l\neq j}(\mu_j-\mu_l)}. \qquad (32)
\end{aligned}
$$

We now define some functionals of μ_j on this Reimann surface through

$$
l_j(\mu)=-\sum_{k=1}^N\int^{\mu_k}du_j=-\sum_{l=1}^N c_{jl}\sum_{k=1}^N\int_0^{\mu_k}\frac{E^{l-1}}{R(E)}dE, \qquad (33)
$$

from which we deduce

$$
\begin{aligned}
&\frac{dl_j(\mu)}{dt} \\
&=-\frac{i}{(\Delta x)^2}\sum_{l=1}^N c_{jl}\sum_{k=1}^N\Big[-1+(-1)^{N+1}\prod_{m\neq k}\mu_m\Big] \\
&\quad\times\frac{\mu_k^{l-1}}{\Pi_{n\neq k}(\mu_k-\mu_n)}. \qquad (34)
\end{aligned}
$$

With the help of standard Lagrange interpolation formulas it can seen that these summations over the μ_c's are really very simple. Instead of the general case we indicate here the result for μ_1 and μ_2. In this case we get

$$\frac{\partial l_j}{\partial t}(\mu_1,\mu_2)$$

$$= \frac{i}{(\Delta x)^2} c_{j1}\left[(1+\mu_2)\frac{\mu_1}{\mu_1-\mu_2} + (1+\mu_1)\frac{\mu_2}{\mu_2-\mu_1}\right]$$
$$+ \frac{i}{(\Delta x)^2} c_{j2}\left[(1+\mu_2)\frac{1}{\mu_1-\mu_2}\right.$$
$$\left. + (1+\mu_1)\frac{1}{\mu_2-\mu_1}\right]\frac{i}{(\Delta x)^2}[c_{j1}-c_{j2}].$$

Hence we have a linear flow given as

$$l_j(\mu_1,\mu_2) = [i/(\Delta x)^2][c_{j1}-c_{j2}]t + l_j^0(\mu_1,\mu_2). \quad (35)$$

Similar considerations hold also for $l_j(\mu_1,\mu_2,...,\mu_N)$ and we have

$$l_j(\mu_1,...,\mu_N) = [i/(\Delta x)^2][c_{j1} - (-1)^N c_{jN}]t + l_j^0. \quad (36)$$

So now the solution for the μ_j's can be obtained by the famous Jacobi inversion problem.

VI. EXPLICIT FORM OF THE SOLUTION

From our previous discussions we observe that the periodic nonlinear field at the nth lattice point, defined to be q_n, satisfies the elegant recursion relation (25) with $\mu_j(n,t)$ satisfying (32). Also there is another algebraic connection between $\mu_j(n,t)q_n$, and q_{n-1}, given by Eq. (29). All these equations involve some symmetric functions of the zeros $\mu_j(m,t)$. It is interesting to note that these symmetric functions of $\mu_j(n,t)$ can be expressed in terms of combinations of Reimann θ functions. Since this result is now quite standard[5] we will quote the result and use it in the sequel. It is known that

$$\prod_{j=1}^{N}\mu_j(n,t) = \ln\frac{\theta(\alpha t+\gamma+\delta)\theta(\alpha t+\gamma-\delta)}{\theta(\alpha t+\gamma+\delta)\theta(\alpha t+\gamma-\delta)} + \sum_{j=1}^{N}\oint_{a_j}\ln z\, du_j(z), \quad (37)$$

$$\sum_{j=1}^{N}\mu_j(n,t) = \frac{\partial}{\partial t}\left|\ln\frac{\theta(\alpha t+\gamma+\sigma+\tau\beta)}{\theta(\alpha t+\gamma-\sigma+\tau\beta)}\right|_{\tau=0} + \sum_{j=1}^{N}\oint z\, du_j(z), \quad (38)$$

where the vectors, α, γ, δ, σ, etc. are defined as below:

$$\delta_j = u_j(0), \quad \gamma_j = \sum_{i=1}^{N} u_j(\mu_i,0) + \frac{1}{2}\sum_{i=1}^{N} B_{i,j}{}^{-1/2},$$

$$\sigma_j = u_j(\alpha), \quad u_j = \int du_j\, dz,$$

$$B_{ij} = \oint_{b_j} du_j(z), \quad \alpha_j = \frac{c_{j,0}}{f_{n0}}.$$

We now set

$$\sum u_j = \chi, \quad \prod u_j = \psi,$$

for the right-hand sides of (37) and (38). Using these with Eqs. (25) and (29), we can eliminate q_{n-1} and solve for $|q_n|^2$ as

$$|q_n|^2 = \left[\frac{\psi^{-1}+\chi+\beta}{2(\Delta\chi)^2\psi^*}\right], \quad \beta = \frac{P_{2N+1}}{P_{2N+2}}.$$

It is useful to remember that these formulas are valid for N-phase periodic waves in general.

VII. DISCUSSION

In the previous sections we have developed an analog of the methods of Date and Tanaka and of Forest and McLaughlin to investigate the quasiperiodic solutions of the discrete nonlinear Schrödinger equation. It is important to remember that the first periodic inverse problem for the discrete nonlinear equation—the Toda lattice—was solved by Kac[6] using the spectrum of the Jacobi matrix. Also the algebrogeometric method exists, and it has been applied by Dubrovin and Krichever,[7] but our approach, in the line of Refs. 3 and 4, is perhaps simplest and does have some computational convenience.

[1] I. M. Krichever, Funct. Anal. Appl. **9**, 105 (1975); B. A. Dubrovin, *ibid.* **9**, 215 (1975).

[2] P. Lac, Lect. Appl. Math. **15**, 85 (1974); Y. C. Ma and M. J. Ablowitz, Stud. Appl. Math. **65**, 113 (1981).

[3] E. Date and S. Tanaka, Suppl. Prog. Theor. Phys. **59**, 107 (1976).

[4] G. Forest and D. W. McLaughlin, J. Math. Phys. **23**, 1248 (1982).

[5] B. A. Dubrovin, V. B. Matveev, and S. P. Novokov, Russ. Math. Surv. **31** (1976).

[6] M. Kac and P. Van Moerbeke, Natl. Acad. Sci. (USA) **72**, 1627 (1975).

[7] B. A. Dubrovin and I. M. Krichever, "Integrable systems," in *London Mathematical Society Lecture Note Series*, Vol. 60 (Cambridge U. P., London, 1981).

Towards the proof of the cosmic censorship hypothesis in cosmological space-times

Andrzej Królak[a)]
Department of Applied Mathematics and Astronomy, University College, Cardiff CFI IXL, United Kingdom

(Received 6 May 1986; accepted for publication 9 September 1986)

A theorem supporting the view that the cosmic censorship hypothesis proved recently by Królak [A. Królak, Gen. Relativ. Gravit. **15**, 99 (1983); J. Class. Quantum Grav. **3**, 267 (1986)] for asymptotically flat space-times, is true in general, is generalized so that it is applicable to cosmological situations.

I. INTRODUCTION

An idea originally put forward by Penrose[1] to circumvent the problem of space-time singularities is that there exists a "cosmic censor" that forbids the appearance of naked singularities (i.e., singularities visible from infinity) in generic collapse that starts off from a perfectly nonsingular initial state, clothing each one in an event horizon.

It is important to know whether or not the cosmic censorship is true for the following three reasons. First, if it is true then the singularities resulting from gravitational collapse cannot appear arbitrarily in space-time but they are always hidden behind an event horizon. Thus the indeterminism introduced by the presence of singularities is considerably reduced. Second, many important results of relativity, the black hole area principle, the black hole uniqueness theorem, and the positivity of gravitational energy, all rely on the assumption that Penrose's hypothesis is true. Third, the validity of the cosmic censorship hypothesis is important for astrophysics. If it is valid then the final stages of evolution of stars of mass above certain critical value must be black holes and not something as ill defined as a naked singularity.

The author has recently proved a theorem that supports the validity of the cosmic censorship hypothesis in asymptotically flat space-times.[2,3] The idea of his approach and the outline of the proof of the theorem is given in an essay submitted to the Gravity Research Foundation in the year 1982.[2] The full account is given in Ref. 3. The aim of this paper is to generalize the censorship theorem given in Ref. 3 to cosmological situations. The paper is based on a part of an essay submitted by this author to the Gravity Research Foundation in the year 1983.

In Sec. II we shall give basic notions and definitions including the statement of the cosmic censorship, which we shall prove in the following section. In that section we shall also formulate our proposals to characterize the singularities arising in physically realistic situations. These proposals constitute the basic assumptions of the censorship theorem given in Sec. III. In Sec. III we shall state and prove our censorship theorem.

[a)] On leave of absence from the Mathematical Institute, Polish Academy of Sciences, Warsaw, Poland.

II. BASIC NOTIONS AND DEFINITIONS

By space-time we shall mean a pair (M,g), where M is a connected orientable four-dimensional Hausdorff C^∞ manifold and g is a C^∞ Lorentz metric on M.

The cosmic censorship hypothesis asserts that no space-time singularity visible from infinity can arise from generic gravitational collapse that starts off from nonsingular initial surface. In this section we shall define the concepts contained in the above rather rough statement of Penrose's hypothesis in precise mathematical terms.

We shall denote by $\mathscr{E}$ a connected component of the set of all past end points of the future-complete null geodesics. We shall call the set $\mathscr{E}$ an *external region*. The external region $\mathscr{E}$ is interpreted as a region of space-time visible from infinity. In definition of $\mathscr{E}$ we demanded merely the existence of the future-complete null geodesics without further restrictions on the behavior of space-time at infinity (cf. the definition of future-null infinity $\mathscr{I}^+$ in Ref. 4 for asymptotically flat space-times). Thus our definition of region visible from infinity is very general and is applicable to cosmological situations with exception to closed cosmological models in which no complete null geodesics exist.

We shall say that a future-endless nonspacelike curve λ *terminates in a naked singularity* if $I^-(\lambda)$, the chronological past of λ, is contained in the chronological past of some point p of the external region $\mathscr{E}$.

In the following definition we shall describe the concept of nonsingular initial surface.

Definition 1: A space-time (M,g) is said to be generalized partially future asymptotically predictable from a partial Cauchy surface S with respect to a nonempty subset $\mathscr{E}_0$ of $\mathscr{E}$ if the following conditions hold: (1) $\mathscr{E}_0 \cap I^+(S) \subset D^+(S)$; (2) when $\mathscr{E}_0 \neq \mathscr{E}$, then $H^+(S)$ contains a future-complete null geodesic generator γ such that $I^-(\gamma) \cap I^+(S) \subset \mathscr{E}_0$; and (3) for every past-endless nonspacelike curve α such that $I^+(\alpha) \subset I^+(S)$ there exists a point s in $I^+(S)$ such that $I^+(\alpha) \subset I^+(s)$.

Remarks: Condition (1) in the above definition means that space-time is predictable in some neighborhood of the initial surface S.

Condition (2) ensures that the predictable region extends to infinity.

The additional condition (3) ensures that violations of

 0022-2488/87/010138-04$02.50

predictability are not associated with the initial surface S. For example, consider a Cauchy surface S in Minkowski space-time. Suppose that a point of the surface S is removed. Let the resulting space-time be denoted by (M',η), where η is the flat metric and let the surface S with the point removed be denoted by S'. It is clear that the external region $\mathscr{E}$ is the whole of M' and that M' contains a naked singularity [any nonspacelike curve in Minkowski space-time with future end point p is a future-endless curve terminating in a naked singularity in space-time (M',η)]. In (M',η) conditions (1) and (2) hold but not condition (3). The naked singularity in this example can be thought of as located on the surface S whereas by cosmic censorship one may only expect that no naked singularities will arise providing that the initial surface is nonsingular. The role of condition (3) is to eliminate initial singularities such as described in the above example.

We shall say that the cosmic censorship holds in space-time (M,g) if it is generalized future asymptotically predictable as defined below.

Definition 2: A space-time generalized partially future asymptotically predictable from a partial Cauchy surface S with respect to a nonempty subset $\mathscr{E}_0$ of $\mathscr{E}$ is said to be generalized future asymptotically predictable from S if the region $\mathscr{E}_0$ coincides with the external region $\mathscr{E}$.

We shall prove that the above definition is equivalent to space-time being free of naked singularities.

Lemma 1: A generalized partially future asymptotically predictable space-time from a partial Cauchy surface S with respect to a nonempty subset $\mathscr{E}_0$ of $\mathscr{E}$ is generalized future asymptotically predictable from S if and only if there is no nonspacelike curve η terminating in a naked singularity such that $I^-(\eta)\cap I^+(S)\neq\varnothing$ and $I^-(\eta)\cap I^+(S)\subset\mathscr{E}_0$.

Proof: Suppose that a generalized partially future asymptotically predictable space-time from a partial Cauchy surface S with respect to a nonempty subset $\mathscr{E}_0$ of $\mathscr{E}$ is not generalized future asymptotically predictable from S. Then $\mathscr{E}_0\neq\mathscr{E}$. Let γ be the null geodesic generator given in condition (2) of Definition 1. Clearly $I^+(\gamma)\subset I^+(S)$. Since S has no edge, γ is past-endless (Ref. 4, Prop. 6.5.3). Thus there exists a point s in $I^+(S)$ such that $I^+(\gamma)\subset I^+(s)$. Let p be any point on γ. The closure of the set $I^-(p)\cap I^+(s)$ is not compact since γ is past-endless. Hence $s\notin \operatorname{int} D^-(I^-(p))$ whereas $s\in I^-(p)$. Thus there exists a future-endless nonspacelike curve η with past end point s such that $I^-(\eta)\subset I^-(p)$. Thus by definition η terminates in a naked singularity. By construction $I^-(\eta)\cap I^+(S)\neq\varnothing$ and $I^-(\eta)\cap I^+(S)\subset\mathscr{E}_0$.

Suppose that there is a nonspacelike curve η terminating in a naked singularity such that $I^-(\eta)\cap I^+(S)\neq\varnothing$. Thus there is a point p in $\mathscr{E}\cap I^+(S)$ such that $I^-(\eta)\subset I^-(p)$. Let (q_i) be a sequence of points on η in $I^+(S)$ such that $q_{i+1}\subset I^+(q_i)$ for all i and with no limit point in M. Let (μ_i) be a sequence of nonspacelike curves from the point q_1 to p, where μ_i consists of the segment of η from q_1 to q_i and a timelike curve from q_i to p. The curve μ_1 is simply a timelike curve μ from q_1 to p. If the sequence (μ_i) had a limit nonspacelike curve μ from q_1 to p then $I^-(\eta)=I^-(\mu')$, where $\mu'=\mu\cap I^-(\eta)$, and where μ' is not future-endless, being continued as μ to the future end point p. We shall show that in such a case the set $\mathscr{E}\cap I^+(S)$ cannot be globally hyperbolic. If the strong causality fails in $\mathscr{E}\cap I^+(S)$, then this set cannot be globally hyperbolic by definition (see Ref. 4, p. 206). If the strong causality holds in $\mathscr{E}\cap I^+(S)$, then we can use the causal boundary construction of Geroch *et al.*[5] Thus the set $I^-(\eta)$ is a terminal indecomposable past set (TIP) and therefore any nonspacelike curve β such that $I^-(\beta)=I^-(\eta)$ must be future-endless (see Ref. 5, Theorem 2.3). This contradicts the existence of the curve μ' above. Thus the sequence (μ_i) has no limit curve. Hence the set $I^-(p)\cap I^+(q_1)$ does not have a compact closure and therefore the set $\mathscr{E}\cap I^+(S)$ cannot be globally hyperbolic (see Ref. 4, p. 206). Consequently $\mathscr{E}\cap I^+(S)$ cannot be in $\operatorname{int} D^+(S)$ as this latter set is globally hyperbolic. Thus $\mathscr{E}\neq\mathscr{E}_0$ and by Definition 2 space-time cannot be generalized future asymptotically predictable from S. □

We now come to the main problem of our description of generic gravitational collapse. We shall propose two conditions that in our opinion fully characterize the generic gravitational collapse. These two conditions constitute the main assumptions of the censorship theorem proved in the following section. It has been agreed that a necessary condition for space-time to be singular is that it is timelike or null geodesically incomplete.[6] Geodesic incompleteness is exhibited in the singularity theorems of Penrose, Hawking, and Geroch. However, the existence of incomplete geodesics in space-time does not mean that it contains singularities important from a physical point of view. For example, a mere removal of a point from a geodesically complete space-time results in a space-time that is singular. Examples such as the one above can be comparatively easily dealt with, for example, by assuming that space-time is maximal and hole-free.[7] Unfortunately there exist examples of singularities like, for example, the shell-crossing singularities of Yodzis *et al.*[8,9] and the shell-focusing singularities of Eardley and Smarr[10] at which even the curvature invariants blow up but which are commonly considered as unphysical. Recently the shell-focusing singularities were thoroughly investigated by Christodoulou.[11] His claim is that the collapse situations in which these singularities appear may be a serious alternative to the standard picture of collapse given by Oppenheimer and Snyder. In our opinion the most promising way of distinguishing the class of singularities important from the physical point of view is to observe how they affect physical objects. This approach has been pioneered by Ellis and Schmidt.[12] They introduced a class of what they called strong curvature singularities which have the property that all objects are crushed to zero volume at such a singularity no matter what the physical properties of the object. It was noticed by Seifert[13] that tidal forces at the shell-crossing singularity are finite. This means that a sufficiently strong detector will not be destroyed at such a singularity. Thus shell-crossing singularities are not of strong curvature type. On the other hand, the Schwarzschild singularity and the singularities in Friedmann cosmologies, i.e., singularities that are regarded as archetypes for real singularities, are of strong curvature type. This has led Tipler *et al.*[14] to suggest that "in any physically realistic space-time, all incomplete causal geodesics terminate in strong curvature singularities." The author of this paper has independently put forward a similar conjecture stating that "singularities in all reasonable physical cases are of the strong curvature type."[15] The concept of the strong curvature singularity was precisely defined in geometrical terms by Tipler.[16] Here we shall adopt the following definition which is a modification of Tipler's original idea.

Definition 3: A future-endless nonspacelike geodesic λ is said to terminate in a strong curvature singularity in the future, if for every point p on λ the expansion Θ of the future-directed nonrotating congruence of geodesics from p containing λ becomes negative somewhere on λ.

In our opinion the above definition is a precise geometrical description of the singularity arising as a result of reaching by the gravitational field the point of no return beyond which only its further, unbounded increase in strength is possible. All the other singularities are in a certain sense artificial originating from some special conditions imposed on space-time, e.g., symmetry, Petrov type, special initial conditions.

In the above definition, not only timelike but also null geodesics are included. The strong curvature singularity is identified by the property that it focuses all the congruences of the null geodesics approaching it.

We shall say that the *strong curvature condition* holds if every future endless nonspacelike geodesics terminating in a naked singularity terminates in a strong curvature singularity in the future.

Our second condition is based on the idea that only those singularities that are forced on us by singularity theorems are generic. This idea has been recently expressed by Seifert[17] and Hawking.[18] We shall say that the *trapped surface condition* holds to the future of a partial Cauchy surface S if for any nonspacelike curve η terminating in a naked singularity, there is a trapped surface $\mathscr{T}$ such that the intersection $\mathscr{T} \cap I^-(\eta) \cap I^+(S)$ is not empty. The support for its validity has been provided by the proof by Schoen and Yau[19] of the trapped surface conjecture put forward by Seifert.[13] The now established conjecture says that "any mass that is concentrated within a region of a sufficiently small diameter can be surrounded by a trapped surface." To end the discussion of the trapped surface condition recall that Penrose was led to the formulation of the cosmic censorship hypothesis[1] by the following remark: in the case of the Kerr–Newman black hole the singularities that are not naked are accompanied by trapped surfaces and therefore there exist theorems (singularity theorems) that say that such singularities will not disappear under generic perturbations.

III. CENSORSHIP THEOREM FOR COSMOLOGICAL SPACE-TIMES

Before we can state our censorship theorem we need the following definition.

Definition 4: A future-endless null geodesic of space-time (M,g) that forms an achronal set and is future complete is called an outgoing null geodesic. A null geodesic that is a limit curve of outgoing null geodesics is called a marginally outgoing null geodesic.

Theorem 1: Let (M,g) be a generalized partially future asymptotically predictable space-time from a partial Cauchy surface S with respect to a nonempty subset $\mathscr{E}_0$ of the external region $\mathscr{E}$, and let γ be the null geodesic generator of $H^+(S)$ given in Definition 1.

If (1) $R_{ab}k^ak^b \geq 0$ for all null vectors k^a, and either (2a) (i) for any compact set $\mathscr{C}$ such that $\mathscr{C} \subset I^-(\gamma) \cap I^+(S)$, the boundary $\dot{I}^+(\mathscr{C})$ contains an outgoing null geodesic contained in $D^+(S)$, (ii) the trapped surface condition holds to the future of S, or (2b) (i) for any future-endless nonspacelike curve η, such that $\dot{I}^-(\eta) \cap I^+(S) \neq \varnothing$ and $I^-(\eta) \cap I^+(S) \subset \mathscr{E}_0$, the boundary $\dot{I}^-(\eta) \cap I^+(S)$ contains a marginally outgoing null geodesic λ, (ii) the strong curvature condition holds; then (M,g) is a generalized future asymptotically predictable space-time from the partial Cauchy surface S.

Remarks: It has been shown by this author that in an asymptotically flat regular partially predictable space-time, condition (2a) (i) is satisfied and condition (2b) (i) is fulfilled if, at the Cauchy horizon, global hyperbolicity is violated in such a way that the condition of causal simplicity does not hold also (see Ref. 3 for details). This last fact means that the censorship theorem proved in Ref. 3 does not deal with completely general naked singularities. There may exist naked singularities such that everywhere on the Cauchy horizon, which forms as a result of the naked singularity, causal simplicity holds.

We assume that the conditions discussed above also hold in cosmological situations. Thus our censorship theorem proved below is subject to at least the same shortcomings as the censorship theorem proved for asymptotically flat space-times in Ref. 3.

Proof of Theorem 1: If the generalized partially future asymptotically predictable space-time (M,g) is not generalized future asymptotically predictable, then by Lemma 1 there exists a future-endless nonspacelike curve η terminating in a naked singularity such that $I^-(\eta) \cap I^+(S) \neq \varnothing$ and $I^-(\eta) \cap I^+(S) \subset \mathscr{E}_0$.

If condition (2a) holds, then by condition (2a) (ii) there exists a trapped surface $\mathscr{T} \subset I^-(\eta) \cap S$. Since $I^-(\eta) \subset I^-(\gamma)$ there exists a future-complete null geodesic generator α of $\dot{J}^+(\mathscr{T})$ by condition (2a) (i). The null geodesic α has the past end point on $\mathscr{T}$ since α is contained in int $D^+(S)$ and the set int $D^+(S)$ is globally hyperbolic and therefore causally simple. Thus we have a contradiction, since by future completeness of α, the Raychaudhuri equation, and condition (1) there must exist a point s conjugate to $\mathscr{T}$ on α. Consequently, points of $\dot{J}^+(\mathscr{T})$ to the future of s can be joined to $\mathscr{T}$ by timelike curves. This is impossible as the set $\dot{J}^+(\mathscr{T})$ is achronal.

If condition (2b) holds, then by (2b) (i) the boundary $\dot{I}^-(\eta)$ contains a marginally outgoing null geodesic λ. Consider the sequence (λ_i) of outgoing null geodesics such that λ is its limit curve. Choose a sequence of points (p_i), such that for all i p_i is a point on λ_i, and the sequence (p_i) converges to some point p on λ. Consider the expansion Θ_i of a nonrotating congruence of null geodesics from p_i containing λ_i: Θ_i cannot become negative anywhere on λ_i to the future of p_i, since otherwise, by completeness of λ_i, the Raychaudhuri equation, and condition (1), there would be a point conjugate to p_i on λ_i. This is impossible since by definition, each λ_i generates an achronal set. Thus in the limit, the expansion Θ of a nonrotating congruence of null geodesics from p and containing λ cannot become negative. However by condition (2b) (ii), λ terminates in a strong curvature singularity. Therefore by Definition 3, Θ must become negative somewhere on λ. This is a contradiction. □

ACKNOWLEDGMENTS

I would like to thank Dr. St. Bażański and Mr. W. Rudnicki for helpful discussions.

[1]R. Penrose, Riv. Nuovo Cimento, **1**, 252 (1969).
[2]A. Królak, Gen. Relativ. Gravit. **15**, 99 (1983).
[3]A. Królak, J. Class. Quantum Grav. **3**, 267 (1986).
[4]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1973).
[5]R. Geroch, E. H. Kronheimer, and R. Penrose, Proc. R. Soc. London Ser. A **327**, 545 (1972).
[6]R. Geroch, Ann. Phys. (NY) **48**, 526 (1968); and Ref. 4, § 8.1.
[7]C. J. S. Clarke, Commun. Math. Phys. **49**, 17 (1976).
[8]P. Yodzis, H. J. Seifert, and H. Müller zum Hagen, Commun. Math. Phys. **37**, 29 (1973).
[9]P. Yodzis, H. J. Seifert, and H. Müller zum Hagen, Commun. Math. Phys. **37**, 135 (1973).
[10]D. M. Eardley and L. Smarr, Phys. Rev. D **19**, 2239 (1979).
[11]D. Christodoulou, Commun. Math. Phys. **93**, 171 (1984).
[12]G. F. R. Ellis and B. Schmidt, Gen. Relativ. Gravit. **8**, 915 (1977).
[13]H. J. Seifert, Gen. Relativ. Gravit. **10**, 1065 (1979).
[14]F. J. Tipler, C. J. S. Clarke, and G. F. R. Ellis, *In General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980).
[15]A. Królak, Master's thesis, Warsaw University, 1978 (unpublished); A. Królak, Acta Phys. Pol. B **12**, 643 (1981).
[16]F. J. Tipler, Phys. Lett. **64**, A, 8 (1977).
[17]H. J. Seifert, in *Differential Geometrical Methods in Mathematical Physics, Lecture Notes in Mathematics*, Vol. 570, edited by A. Dold and B. Eckman (Springer, Berlin, 1978).
[18]S. W. Hawking, Gen. Relativ. Gravit. **10**, 1047 (1979).
[19]R. Schoen and S.-T. Yau, Commun. Math. Phys. **90**, 545 (1983).

On singularity theorems and curvature growth

L. B. Szabados
Central Research Institute for Physics, P.O. Box 49, H-1525 Budapest 114, Hungary

(Received 29 April 1986; accepted for publication 27 August 1986)

It is shown that the proofs of a series of classical singularity theorems of general relativity can be modified such that these theorems also state the maximality of the incomplete nonspacelike geodesics. Since along maximal incomplete nonspacelike geodesics with affine parameter u certain parts of the tidal curvature cannot blow up faster than $(\bar{u}-u)^{-2}$, where $\bar{u}$ is the parameter value until which the geodesics cannot be extended, the classical singularity theorems do restrict the behavior of the curvature.

I. INTRODUCTION

In the general theory of relativity the singularity theorems of Hawking, Penrose, and others state the existence of incomplete nonspacelike geodesics under very general conditions; i.e., space-time is singular.[1,2] However, these theorems do not show which geodesics are incomplete and in many cases they do not tell us explicitly if these geodesics are future or past incomplete.

In other parts of classical physics the notion of singularities is connected with the misbehavior of a characteristic quantity; e.g., field strength. Thus, based on this intuitive notion of singularities, certain misbehavior of the curvature tensor is expected along the incomplete nonspacelike geodesics. Unfortunately, such a misbehavior does not follow from incompleteness, furthermore the singularity theorems say nothing about the curvature's behavior.[3]

Surprisingly, while we are expecting a lower bound to the rate of growth of some parts of the curvature along incomplete nonspacelike geodesics approaching a "true, physical" singularity, in certain cases the existence of an upper bound can be proved, which is due to the causal structure.[4–7] The basic idea[1,2] is the fact that a nonspacelike curve is maximal iff it is geodesic without any pair of conjugate points. But the occurrence of conjugate points is a very general phenomenon: if a nonspacelike geodesic gathers enough curvature in some sense, then conjugate points must occur. Null geodesics are maximal if they lie in an achronal set, furthermore there exists a maximal nonspacelike geodesic between causally related points in a globally hyperbolic set. Thus maximality, in certain situations, follows from causality. Consequently, if γ is a maximal nonspacelike geodesic, then γ cannot gather arbitrarily large curvature; and in particular if γ is a maximal incomplete nonspacelike geodesic (and so it hits a singularity), we obtain an upper bound to the rate of the blowing up of certain parts of the curvature. This argument was used to obtain an upper bound to the rate of the divergence of the Ricci and Weyl part of the tidal force along incomplete null geodesics lying in an achronal set by Tipler[5] and the author,[6] respectively: they cannot blow up faster than $(\bar{u}-u)^{-2}$, where u is the affine parameter and $\bar{u}$ is the parameter value until which the geodesic cannot be extended.

Unfortunately, incomplete nonspacelike geodesics are not necessarily maximal, so we do not have any upper bound in general. For example, if γ is a future incomplete null geodesic [so the TIP $P:=I^{-}\gamma$ represents a singular point of the causal boundary ∂_c (see Ref. 3)], then γ is not necessarily in an achronal set, therefore γ is not necessarily maximal. On the other hand, though the boundary ∂P is achronal and is generated by future endless null geodesics, these geodesics are not necessarily incomplete, even if γ is. Therefore one may ask the question: Is there any "physically realistic" situation in which the incomplete nonspacelike geodesics are maximal; i.e., in which the existence of an upper bound can be proved?

In this paper we show that the proofs of a number of classical singularity theorems can be modified such that they state not the existence of incomplete nonspacelike geodesics only, but that these geodesics must be maximal, therefore these theorems do restrict the curvature's behavior. We conclude that in "physically realistic" situations, defined by the conditions of the classical singularity theorems, certain parts of the curvature cannot blow up faster than $(\bar{u}-u)^{-2}$ along the geodesics approaching the singularity. Furthermore, in many cases they show where the incomplete geodesics lie.

This paper consists of four parts. In the first one we review what restrictions on the growth of the curvature can be obtained from the maximality of incomplete nonspacelike geodesics. The second part contains the auxiliary statements we need for the new proofs. Four of the most important classical singularity theorems with modified proofs are contained in the third part. Finally, in the fourth part some remarks are given.

This paper is based on the matter given in Refs. 1 and 2, so the continuous references to well-known statements are omitted. Our conventions and notations are the same as those of the book of Hawking and Ellis[1] except that the chronological, causal, etc. future of the set K is denoted by $I^{+}K, J^{+}K$, etc., respectively.

II. MAXIMAL NONSPACELIKE CURVES

A nonspacelike curve γ from p to q is said to be maximal[1] if p and q cannot be joined by any nonspacelike curve, obtained from γ by small deformation, longer than γ. Since, by definition, incomplete nonspacelike curves are inextendable, we are interested in nonspacelike curves without end points. A future inextendable nonspacelike curve γ starting at p is said to be maximal if for every $q\in\gamma$, the segment of γ between p and q is maximal. Based on the statements of Chap. 4 of Ref. 1, one can state that a future inextendable nonspacelike curve γ from p is maximal iff it is geodesic and

 0022-2488/87/010142-04$02.50

contains no point conjugate to p along γ.

Ruling out the possibility of the occurrence of conjugate points along future directed incomplete nonspacelike geodesics, Tipler has a restriction on the growth of the $R_{ab}K^aK^b$ part of the curvature, where K^a is the tangent to the geodesics.[4-7]

Proposition 2.1: Let γ: $[0,\bar{u})\to M$ be a maximal incomplete nonspacelike geodesic with affine parameter u and tangent K^a. If the convergence condition $R_{ab}K^aK^b\geqslant 0$ holds, then, for $\forall\alpha>0$, the inequality

$$\liminf_{\Delta u\to 0}\left(-\Delta u\int_{\bar{u}-\Delta u}^{\bar{u}-(1+\alpha)\Delta u}R_{ab}K^aK^b\,du\right)\leqslant 2n$$

must be satisfied, where $n=2$ for null and $n=3$ for timelike geodesics.

This proposition implies that the energy-density-like expression $R_{ab}K^aK^b$ cannot blow up faster than $(\bar{u}-u)^{-2}$ as we approach the singularity at $\bar{u}$, provided the metric is C^2. For timelike geodesics, the equations describing the Jacobi fields (and so the conjugate points) are very complicated, therefore obtaining restrictions on further parts of the curvature seems to be almost hopeless.

Along maximal incomplete null geodesics, however, a restriction on the eigenvalue ϵ of $C_{manb}K^aK^b$ can be obtained in certain situations. Before considering the statement giving us this restriction, we have to examine the possible behaviors of $C_{manb}K^aK^b$ as $u\to\bar{u}$. Since $C_{manb}K^aK^b$ is a symmetric and traceless 2×2 matrix, it has the form

$$\epsilon\begin{pmatrix}\cos 2\chi & \sin 2\chi\\ \sin 2\chi & -\cos 2\chi\end{pmatrix}.$$

Thus it is completely characterized by the functions $\epsilon(u)$ and $\chi(u)$. Consequently, the behavior of $C_{manb}K^aK^b$ is determined by those of ϵ and χ (see Ref. 7). From physical points of view the most important case is that in which the components of $C_{manb}K^aK^b$ diverge; i.e., there is a definite limiting eigenframe of $C_{manb}K^aK^b$ [$\lim_{u\to\bar{u}}\chi(u)$ exists] and ϵ tends to infinity. Introducing the notation

$$\begin{aligned}W_{u_1,u-u_1} &:= \frac{1}{2}\int_{u_1}^{u}\left(\left(\int_{u_1}^{u'}C_{manb}K^aK^b\,du''\right)\right.\\ &\quad\times\left.\left(\int_{u_1}^{u'}C_{menf}K^eK^f\,du''\right)\right)du'\\ &=\int_{u_1}^{u}\left[\left(\int_{u_1}^{u'}\epsilon\cos 2\chi\,du''\right)^2\right.\\ &\quad+\left.\left(\int_{u_1}^{u'}\epsilon\sin 2\chi\,du''\right)^2\right]du',\end{aligned}$$

the next proposition gives us a restriction on the diverging $C_{manb}K^aK^b$.[6]

Proposition 2.2: Let γ: $[0,\bar{u})\to M$ be an incomplete null geodesic lying in an achronal set, u be its affine parameter, and K^a be its tangent. If the convergence condition $R_{ab}K^aK^b\geqslant 0$ holds, $\lim_{u\to\bar{u}}\chi(u)$ exists and $\exists\theta>0$ such that $\epsilon(u)$ does not change sign on $(\bar{u}-\theta,\bar{u})$, then for $\forall\alpha>0$, the inequality

$$\liminf_{\Delta u\to 0}\left(-\Delta u W_{(\bar{u}-\Delta u)-\Delta u,\,-\alpha\Delta u}\right)\leqslant(2+\alpha)(1+\alpha)^3$$

must be satisfied.

This statement implies that ϵ cannot blow up faster than $(\bar{u}-u)^{-2}$ along maximal incomplete null geodesics, provided the metric is C^2 and $\lim\chi(u)$ exists. Otherwise, assuming for example that $\epsilon(u)=b(\bar{u}-u)^{-2-\nu}$ for some $b\neq 0$ and $\nu>0$, we would obtain that

$$-\Delta u W_{(\bar{u}-2\Delta u),\,-\alpha\Delta u}=b^2G(\alpha,\nu)(\Delta u)^{-2},$$

$\forall\alpha>0$, where $G(\alpha,\nu)$ is a nowhere zero expression of α and ν; which would tend to infinity as $\Delta u\to 0$.

III. AUXILIARY STATEMENTS

In this section we state four lemmas that will be used in the modification of the proofs of the classical singularity theorems.

Lemma 3.1: If (1) the null convergence condition holds on M, (2) the null generic condition holds on M (i.e., on each inextendable null geodesic with tangent K^a there is a point where $K_{[a}R_{b]ef[c}K_{d]}K^eK^f\neq 0$), and (3) the chronology condition holds on M, then the strong causality condition holds on M or there is a point p where the strong causality condition is violated and there is an incomplete null geodesic through p lying in $E^-\{p\}\cup E^+\{p\}$.

This lemma is a modified form of Proposition 6.4.6 of Ref. 1 and its proof is almost the same.

The following lemmas state compactness of certain subsets of space-time, provided the maximal nonspacelike geodesics generating them, in some sense, leave these sets. The proofs are based on the standard matter given in Chap. 6 of Ref. 1. The first one has been published yet,[8] so its proof will be omitted here.

Lemma 3.2: Let K be a nonempty set. If each future directed null geodesic γ generating E^+K-K leaves E^+K in the future direction (i.e., each γ has a point r such that the points of γ following r do not belong to E^+K and the points of γ preceding r belong to E^+K), then E^+K is compact.

Lemma 3.3: If there is no past directed past endless nonspacelike geodesic γ from p such that its segment $\gamma-\{p\}$ is maximal, or each such a nonspacelike geodesic leaves $D^-E^-\{p\}$ in the past direction, then $\overline{D^-E^-\{p\}}$ is compact.

Sketch of proof: Using the technique developed in Refs. 1 and 2, one can show that for each point q of $\overline{D^-E^-\{p\}}$ there exists a past directed nonspacelike geodesic γ from p through q, such that the segment (q,p) of γ is maximal.

Let $\{q_n\}$ be an infinite sequence of points of $\overline{D^-E^-\{p\}}$. Because of Lemma 3.2, $E^-\{p\}$ is compact. Thus without loss of generality one may assume that p is not limit point of $\{q_n\}$ and no point of this sequence belongs to $E^-\{p\}$. Let γ_n be the past directed nonspacelike geodesic from p through q_n such that its segment (q_n,p) is maximal. $\{\gamma_n\}$ has a limit curve γ from p which is geodesic. Furthermore, the segment of γ consisting of those points that are limit points of the maximal segments (q_n,p) of the γ_n's is maximal. Consequently, there is a point $q\in\gamma-\{p\}$ such that this maximal segment is $[q,p)$ or (q,p). Of course, $q\in\overline{D^-E^-\{p\}}$, and q is a limit point of $\{q_n\}$, i.e., $\overline{D^-E^-\{p\}}$ is compact.

In a similar way one can prove our fourth lemma.

Lemma 3.4: Let S be a compact C^2 partial Cauchy surface. If there is no future inextendable maximal timelike geodesic orthogonal to S, or each such a timelike geodesic leaves D^+S in the future direction, then $\overline{D^+S}$ is compact.

IV. CLASSICAL SINGULARITY THEOREMS MODIFIED

In this section the four classical singularity theorems contained in Ref. 1 will be reexamined and it will be clear that the original conditions of these theorems guarantee the maximality of incomplete nonspacelike geodesics.

The first one is Penrose's theorem[9] and, since its modified proof has been published elsewhere,[8] its new proof, which is based on Lemma 3.2, will not be repeated here.

Theorem 4.1: If (1) $R_{ab}K^aK^b \geqslant 0$ for every null vector K^a; (2) (M,g) admits a noncompact Cauchy surface; and (3) there exists at least one of the following: (a) a closed trapped surface T, (b) a point t such that along each future directed null geodesic from t the expansion $\hat{\theta}$ becomes negative (t may be called a future trapped point); then there exists a future incomplete null geodesic lying in ∂J^+T or in $\partial J^+\{t\}$, respectively.

Possibility (b) in condition (3) is due to Tipler[10] and we note that this concept of trapped point differs from the trapped point of Królak.[11,12]

The second theorem is that of Hawking and Penrose.[13]

Theorem 4.2: If (1) $R_{ab}K^aK^b \geqslant 0$ for every nonspacelike vector K^a; (2) the chronology condition holds on M; (3) the generic condition holds on M; and (4) there exists at least one of the following: (a) a compact achronal set S without edge, (b) a closed trapped surface T, (c) a future trapped point t; then at least one of the following statements holds: (α) there exists a compact set $C \neq 0$ and an incomplete null geodesic lying in $E^+C \cup E^-C$, and/or (β) there exists an open globally hyperbolic set D and an incomplete maximal nonspacelike geodesic in D.

Proof: Let K be S or T or $\{t\}$ in case (a), or (b), or (c), respectively. In cases (b) and (c), $E^+K - K$ is generated by future directed null geodesics with past end points on K. If each null geodesic generator of $E^+K - K$ leaves E^+K, then E^+K is compact (Lemma 3.2). Thus, if E^+K is not compact then there must be a null geodesic generator γ which does not leave E^+K in the future direction. However, γ cannot be future complete, as otherwise a point conjugate to K would occur on γ, thus statement (α) holds with $C = T$ or $\{t\}$. Since S is achronal without edge, one has $E^\pm S = S$ and so, because of the compactness of S, E^+S is compact. Hence one may assume that E^+K is compact (i.e., K is a future trapped set).

Since K is a future trapped set in a strongly causal space-time, there exists a future inextendable timelike curve μ in $\operatorname{int} D^+E^+K$. The set $F := E^+K \cap \overline{J^-\mu}$ is compact and achronal. Furthermore, $E^-F = F \cup G$, where G is a connected subset of $\partial J^-\mu$. Thus through each point of G there is a future inextendable null geodesic. Here E^-F may be compact or noncompact. If E^-F is not compact, then by Lemma 3.2 there is a null geodesic generator γ of $E^-F - F$ that does not leave E^-F in the past direction; i.e., γ is an inextendable null geodesic in $\partial J^-\mu$ through $p := \gamma \cap F$. But γ must be incomplete, as otherwise conditions (1) and (2) would imply the existence of a pair of conjugate points along γ. Thus with $C = \{p\}$, statement (α) holds, therefore one can assume that E^-F is compact.

F is a past trapped set in a strongly causal space-time, thus there exists a past inextendable timelike curve λ in $\operatorname{int} D^-E^-F$. From this point the proof is the same one given in Ref. 1: one can show that there is a point $q \in E^+K$ and a maximal inextendable nonspacelike geodesic γ through q in $D := \operatorname{int} DE^-F$. This geodesic must be incomplete, because if it were complete then a pair of conjugate points would occur, which would contradict its maximality.

The original version of the following theorems were published by Hawking.[14,1]

Theorem 4.3: If (1) $R_{ab}K^aK^b \geqslant 0$ for every nonspacelike vector K^a; (2) the strong causality condition holds on M; and (3) there is a point p, a past directed unit timelike vector W at p and a positive number b such that the expansion $\theta = V^a{}_{;a}$ of the past directed timelike geodesics from p with unit tangent V^a becomes less than $-3|W^aV_a|b^{-1}$ within parameter distance $\Delta t = |W^aV_a|^{-1}b$; then at least one of the following statements holds: (α) there is a past directed incomplete null geodesic from p in $E^-\{p\}$, and/or (β) there is a past directed maximal incomplete nonspacelike geodesic from p in $D^-E^-\{p\}$.

Sketch of proof: If $E^-\{p\}$ is not compact, then by Lemma 3.2 there is a null geodesic from p in $E^-\{p\}$. If $E^-\{p\}$ is compact, then $\overline{D^-E^-\{p\}}$ must be noncompact, as otherwise there would be a past imprisioned timelike curve in $\overline{D^-E^-\{p\}}$. Thus, because of Lemma 3.3, there is a past endless maximal nonspacelike geodesic in $D^-E^-\{p\}$. However, these curves must be incomplete, according to conditions (1) and (2).

Theorem 4.4: If (1) $R_{ab}K^aK^b \geqslant 0$ for every timelike vector K^a, (2) there is a compact C^2 partial Cauchy surface, (3) the unit normals to S are everywhere converging then there exists a maximal future incomplete timelike geodesic orthogonal to S in D^+S.

Based on Lemma 3.4, a similar argument can be used to prove this statement too.

V. DISCUSSION AND FINAL REMARKS

Since we summarized the results in the Abstract and the Introduction, we do not repeat them, but we have some final remarks.

The second condition of Theorem 4 of Ref. 1 is weaker than that of Theorem 4.4: for the proof of the existence of incomplete timelike geodesic, only a compact spacelike hypersurface without any edge is needed. Of course, in Hawking's covering space $M_{\rm H}$, the maximality of the incomplete timelike geodesic γ orthogonal to an S-homeomorphic preimage $S_{\rm H}$ of S (Ref. 15) can be proved. Thus, it would be interesting to see whether or not the geodesic might lose its maximality under the projection $\pi: M_{\rm H} \to M$. If not then, of course, the second condition of Theorem 4.4 can be weakened to that of Theorem 4 of Ref. 1 (see *Note added in proof*).

There is another class of singularity theorems[16,17] where the maximality of the incomplete nonspacelike geodesics can be proved in certain covering spaces. However, it is not clear whether or not their maximality is preserved under the covering projection (see *Note added in proof*).

Among the oldest singularity theorems,[18,19] there is a great variety of statements that predict maximal incomplete timelike geodesics. For example, one can show easily the next theorem.

Theorem 5.1: If (1) $R_{ab} V^a V^b \geqslant 0$ for every timelike vector V^a; (2) (M,g) admits a compact C^2 Cauchy surface S; and (3) the unit normals to S are everywhere converging; then every future inextendible timelike curve has finite total length measured from one of its points, furthermore for every TIP P there exists a maximal incomplete timelike geodesic γ orthogonal to S such that $I^- \gamma \subseteq P$.

Condition (3) can be replaced by one of the series of conditions that guarantee the expansion of the timelike geodesics becoming negative.[18,19,20] If the future causal boundary of space-time has no null part,[3,21] then Theorem 4.1 states that each point of the singular future boundary can be reached by maximal timelike geodesics.

Finally, it is worth noting that the incomplete null geodesic in Theorem 1 of Ref. 22 (which states that, roughly speaking, chronology violation creates incomplete null geodesics in an asymptotically flat space-time) is also maximal.

Note added in proof: Since the covering projection is a local diffeomorphism, the maximality of the inextendable nonspacelike geodesics is preserved: if there were a nontrivial Jacobi field along $\pi \circ \gamma$ with zeros then a Jacobi field describing conjugate points could be given along γ. (I am indebted to C. J. S. Clarke for suggesting this idea of the proof.)

ACKNOWLEDGMENTS

I am indebted to Dr. Á. Sebestyén and the whole relativity group of CRIP for the continuous discussions. Special thanks to Miss M. Vida for the clarification of the properties of $G(\alpha,\nu)$.

[1] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U. P., Cambridge, 1973).
[2] R. Penrose, *Techniques of Differential Topology in Relativity* (SIAM, Philadelphia, 1972).
[3] R. Penrose; in *Theoretical Principles in Astrophysics and Relativity*, edited by N. R. Lebovitz, W. H. Reid, and P. O. Vandervort (Chicago U. P., Chicago, IL, 1978).
[4] F. J. Tipler, C. J. S. Clarke, and G. F. R. Ellis, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980).
[5] F. J. Tipler, Phys. Rev. D **15**, 942 (1977).
[6] L. B. Szabados, Gen. Relativ. Gravit. **14**, 891 (1982).
[7] L. B. Szabados, Gen. Relativ. Gravit. **15**, 187 (1983).
[8] L. B. Szabados, in *Proceedings of the Fourth Marcel Grossmann Meeting*, edited by R. Ruffini (to be published).
[9] R. Penrose, Phys. Rev. Lett. **14**, 57 (1965).
[10] F. J. Tipler, Phys. Rev. D. **17**, 2721 (1978).
[11] A. Królak, Acta. Phys. Pol. B. **12**, 643 (1981).
[12] A. Królak, Gen. Relativ. Gravit. **14**, 793 (1982).
[13] S. W. Hawking and R. Penrose, Proc. R. Soc. London Ser. A **314**, 529 (1970).
[14] S. W. Hawking, Proc. R. Soc. London Ser. A **300**, 187 (1967).
[15] B. C. Haggman, G. W. Horndeski, and G. Mess, J. Math. Phys. **21**, 2412 (1980).
[16] D. Gannon, Gen. Relativ. Gravit. **7**, 219 (1976).
[17] D. Gannon, J. Math. Phys. **16**, 2364 (1975).
[18] S. W. Hawking, Proc. R. Soc. London Ser. A **294**, 511 (1966).
[19] S. W. Hawking, Proc. R. Soc. London Ser. A **295**, 490 (1966).
[20] F. J. Tipler, Astrophys. J. **209**, 12 (1976).
[21] R. M. Wald and P. Yip, J. Math. Phys. **22**, 2659 (1981).
[22] F. J. Tipler, Phys. Rev. Lett. **37**, 879 (1976).

Hamiltonian dynamics of higher-order theories of gravity

Victor Szczyrba
Department of Physics, Yale University, 217 Prospect Street, New Haven, Connecticut 06511

(Submitted 20 June 1986; accepted for publication 17 September 1986)

The Hamiltonian dynamics of gravitational theories with general Lagrangians quadratic in components of the curvature tensor is investigated. It is shown that the Noether generator corresponding to the action of the diffeomorphism group of space-time naturally defines the energy-momentum function E. The analysis of the differential of E gives rise to a formula for the symplectic two-form Ω and thereby defines symplectic (canonical) variables of the system. This construction is four-covariant, that is, independent of a chosen slicing of space-time. A more thorough analysis is performed in the $(3+1)$ decomposition. In this scheme the canonical classification of quadratic Lagrangians is presented. It singles out the class of canonically regular Lagrangians as well as four classes of degenerate ones. The Cauchy–Kowalewska problem for all these classes is formulated.

I. INTRODUCTION

The dynamics of classical fields is governed by Euler–Lagrange equations of corresponding variational problems. In mechanics, where the least action principle is related to one-dimensional variational integrals, the correspondence between Lagrangian and Hamiltonian formulations was already discovered in the 19th Century. Early attempts to generalize that construction for multidimensional integrals, that is, for the field theory, cannot be considered satisfactory mainly because of the lack of a single-out time coordinate for relativistic Lagrangian field theories.[1]

Even if these apparent difficulties can be overcome by introducing families of spacelike surfaces in space-time to parametrize the evolution picture ("many-fingered time"),[2] the next problem, namely, the canonical degeneracy of physically interesting Lagrangians and the related noninvertibility of corresponding Legendre transformations, seems to be the main obstacle for the construction of an adequate Hamiltonian dynamics.

In the late 60s and early 70s in order to construct a consistent Hamiltonian fomulation of field theories several groups of mathematical physicists started investigations on the geometric structure of the multidimensional variational problems. It turned out that the necessary mathematical structure had already been presented in the literature in 1953 by Dedecker.[3] His paper gave a geometrized and updated version of the results that had earlier been obtained by Caratheodory, Weyl, Lepage, and De Donder. Written in a rather difficult language of modern differential geometry, Dedecker's paper was not known in the physical community for several years. In this scheme the geometric formulation of first-order variational problems is based on the Hamilton–Cartan four-form θ_{HC} constructed from the Lagrangian of the system under consideration. θ_{HC} is not simply a differential form on space-time M but a four-form on some bundle P over M, whose fibers are parametrized by four-momenta (or four-velocities) of variational potentials. The variational Euler–Lagrange equations for the Lagrangian L can be rewritten in an elegant geometric form using the exterior derivative $d\theta_{\mathrm{HC}}$ of the Hamilton–Cartan four-form. Moreover the five-form $d\theta_{\mathrm{HC}}$ transvected with two vector fields on P, which represent vectors tangent to the variety of all geometric configurations of the system, and integrated over a three-dimensional initial surface σ gives us the numerical value of the symplectic two-form Ω on the variety of all geometric configurations $\mathbf{G}$.

The action of the diffeomorphism group of space-time in the bundle P induces a mapping $Z \to X_Z$ from the space of vector fields on space-time to the space of vector fields on P. Therefore we can construct the energy-momentum three-form ν_Z on P transvecting θ_{HC} with X_Z. Taking the pullback of ν_Z onto M and integrating it over a three-dimensional initial surface σ we obtain the energy-momentum function E on $\mathbf{G}$.

Now, the functional Hamilton equation reads

$$dE \cdot V = -\Omega(Y \wedge V), \tag{1.1}$$

where V is an arbitrary (sample) vector tangent to $\mathbf{G}$, and Y is the vector of evolution in $\mathbf{G}$ generated by the vector field Z. In local coordinates Y is represented by the Lie derivatives of corresponding geometric variables taken in the direction of Z.

For first-order variational principles the geometric multisymplectic formulation of the calculus of variations with applications to physical problems has been developed in papers by Garciá-Pérez-Rendon, Gawedzki, Kijowski, Goldschmidt–Sternberg, Szczyrba, Kijowski–Tulczyjew, Kondracki, and Aldaya–Azcarraga[4]; see also a review paper by Kastrup.[5]

The general method of how to pass from the multisymplectic formulation of field theories to the Hamiltonian dynamics, as it has been outlined above, was presented in the author's previous papers[6] and then applied to gravitational theories with general Lagrangians.[7,8]

These papers reveal relations between the symplectic Hamiltonian dynamics and Dirac's theory of constrained Hamiltonian systems.[9] Recently analyzed cases of gravitational theories with quadratic Lagrangians[10] show, however, that some results effectively obtained in the symplectic approach would be very difficult to get in a rather formal Dirac scheme.

 0022-2488/87/010146-13$02.50

So far, the symplectic dynamics is known for theories based on first-order variational principles. In theories of gravity these cases correspond to the Einstein–Palatini method of variation, where the components of a metric $g_{\mu\nu}$ and the coefficients of a connection $\Gamma_{\mu}{}^{\lambda}{}_{\nu}$ are varied independently.

For higher-order variational problems a consistent multisymplectic formulation has not yet been found. In spite of essential efforts in that field[11] the problem seems to be more elusive than ever. Up to now nobody has succeeded in the finding of appropriate symplectic positions and momenta for general higher-order variational systems, and it could even seem that for such theories the symplectic Hamiltonian formulation does not exist.

In the present paper we show, however, how to formulate the Hamiltonian dynamics for second-order gravitational theories based on Einstein–Hilbert variational principle with ten variational potentials $g_{\mu\nu}$ and the connecting coefficients expressed by the Christoffel symbols $\{_{\mu}{}^{\lambda}{}_{\nu}\}$. Here the primary entity is the Lagrangian L depending on the metric, its first and second derivatives. The invariance properties of the Lagrangian with respect to the action of the diffeomorphism group of space-time enable us to define a conserved Noether current E^{λ}. The integral of the corresponding dual three-form over a three-dimensional surface σ defines the energy-momentum function E on the space of geometric configurations **G**.

Of course, for first-order variational problems such a method is equivalent to the geometric construction of E described previously. For higher-order systems, however, we do not know how to construct the symplectic two-form Ω. In order to overcome this difficulty we compute the differential dE and show that the integrand in the explicit formula for $dE\cdot V$ can be rearranged in such a way that dE satisfies the equation of type (1.1) with the vector of evolution Y given by the Lie derivatives of corresponding field variables taken in the direction of the field Z. The proper formula for the symplectic two-form Ω can be naturally inferred from these considerations.

This construction is entirely four-covariant and does not depend on a particular choice of a surface σ in space-time. It enables us to find elegant formulas for symplectic variables without $(3+1)$ decomposition.[8]

However, four-covariant symplectic variables derived in such a way are not independent. In order to determine kinematically independent symplectic variables on the initial surface σ we have to decompose the field variables in their σ-tangential and σ-normal parts. This decomposition is performed by means of the bar operation (see Ref. 6 and Appendix A). As the result we get 12 symplectic positions

$$\bar{g}_{ij},\quad z_{ij}, \tag{1.2a}$$

and 12 symplectic momenta

$$\Xi^{ij},\quad \zeta^{ij}, \tag{1.2b}$$

on the initial surface σ.

Here $\bar{g}_{ij}$ is a Riemannian metric on σ and z_{ij} is proportional to the second fundamental form K_{ij} of the embedding

$$\sigma \to M \quad (z_{ij} = -2K_{ij}).$$

The formulas for the momenta Ξ^{ij}, ζ^{ij} are given in Sec. II.

Gravitational Lagrangians divide in two classes. For regular Lagrangians 24 symplectic momenta (1.2b) are kinematically independent quantities on the initial surface σ. For singular Lagrangians there are kinematical (primary) constraints among the quantities (1.2). In such cases we have to find kinematically independent symplectic variables and express the symplectic two-form in terms of them. In Sec. III we present a classification of gravitational Lagrangians quadratic in the curvature tensor and determine their kinematically independent symplectic variables.

The field equations

$$(Eq)^{\mu\nu} = \frac{\delta L}{\delta g_{\mu\nu}} = 0 \tag{1.3}$$

written in terms of symplectic variables divide in two subsets,

$$\overline{(Eq)}^{0\nu} = 0, \tag{1.4a}$$

and

$$\overline{(Eq)}^{mn} = 0. \tag{1.4b}$$

We show that Eqs. (1.4a) are the symplectic constraints. They can be expressed in terms of symplectic variables and their spatial derivatives. Equations (1.4b) determine the dynamics of the system. The dynamics preserves the constraints.

Of course, as in any theory of gravity, we have four gauge variables N, N^k [ADM's (Arnowitt–Deser–Misner) lapse and shift] whose dynamics is not governed by the field equations and they have to be fixed arbitrarily on space-time.

The analysis presented in this paper enables us to give a canonical classification of gravitational Lagrangians. If we restrict ourselves to Lagrangians at most quadratic in the curvature tensor then such a general Lagrangian can be written as

$$L = A\sqrt{-g}R + 1/2B\sqrt{-g}R^2 + 1/4C\sqrt{-g}R^{\mu\nu}R_{\mu\nu}$$
$$+ 1/4D\sqrt{-g}R^{\alpha\beta\mu\nu}R_{\alpha\beta\mu\nu}. \tag{1.5}$$

By virtue of the Bach–Lanczos identity [cf. (3.2)] we may neglect the D term. Let $D=0$. Then Lagrangians (1.5) are canonically regular if and only if $C\neq 0$ and $C\neq -6B$.

Besides class I of canonically regular Lagrangians there are four classes of singular Lagrangians:

II. $C\neq 0,\quad C=-6B,\quad A\neq 0$
III. $C\neq 0,\quad C=-6B,\quad A=0,$
IV. $C=0,\quad B\neq 0,$
V. $C=0,\quad B=0,\quad A\neq 0.$

In case III we have additional gauge invariance with respect to conformal (scale) transformation $g_{\mu\nu}\to \tau g_{\mu\nu}$ and $\operatorname{tr} z$ is the gauge variable related to this gauge transformation.

The canonical analysis of quadratic theories of gravity has recently been discussed by Boulware.[12] His paper follows the routine procedure of the $(3+1)$ decomposition of the action integral as a method of finding the canonical variables. He separates the conformal case (III in our terminology) but does not make clear distinction between classes II, IV, and V. The classical (standard) method employed in that paper requires tedious calculations of Poisson brackets to prove that the constraints (2.31) are first class. In our approach this property is obvious because the explicit formula for the energy momentum vector density (2.12′) shows that the constraints are generators of the action of the diffeomorphism group of space-time.

The symplectic-Hamiltonian picture helps us to formulate the dynamical initial value problem. It is not the only useful application of this technique. In the case of Einstein's gravity geometric-Hamiltonian methods were used by Fischer, Marsden, Moncrief, Arms, and Anderson[13] to investigate singularities of the space of solutions. In a subsequent paper we will show how the symplectic methods enable us to extend these results for general gravitational Lagrangians. In the present paper we deal only with second-order gravitational theories whose Lagrangians are constructed from the metric and Riemann tensor. It is clear, however, that the idea and methods presented here can also be applied to higher-order gravitational theories with Lagrangians constructed from the metric, the Riemann tensor, and its covariant derivatives up to a certain order. Moreover, we may couple a tensor matter field to gravity with a Lagrangian depending on matter potentials and their partial derivatives up to a fixed order.

The Hamiltonian fomulation of the gravitational theories as presented in Secs. II and III requires the boundary integrals to be neglected. It is justifiable if initial surfaces are compact manifolds without boundary, that is to say, for spatially closed space-times. If initial surfaces are noncompact manifolds then the functional Hamilton equation (1.1) is not valid. In such cases the value of the symplectic form depends essentially on the choice of the initial surface and instead of (1.1) we get

$$dE = -\Omega(Y\wedge\cdot) + \Sigma, \tag{1.6}$$

where Σ is a one-form on **G**, which is responsible for nonconservation of the symplectic structure. In general, Σ is not closed and it cannot be expressed as the differential of a function on **G**. Therefore for spatially open space-times the dynamics is not Hamiltonian in sense of Eq. (1.1).

However, if we restrict the space of geometric configuration **G** to metrics asymptotically Euclidean at spatial infinity on the initial surface σ, then we can preserve the Hamiltonian dynamics (1.1) with a modified energy function $E_{\text{tot}} = E + E_\infty$ and with the same symplectic two-form Ω.

Then the Hamilton equation reads

$$dE_{\text{tot}} = -\Omega(Y\wedge\cdot). \tag{1.7}$$

The additional term E_∞ is expressed by the surface integral over the two-dimensional boundary of σ. We prove that for the general Lagrangian (1.5) only its Einstein term contributes to E_∞. This result was earlier observed by Boulware[12] and Strominger.[14]

Because the energy formula for the general quadratic Lagrangian coincides with the standard ADM expression,[15] it is natural to ask whether the positivity arguments by Shoen–Yau, Witten, Horowitz–Tod, Reula, and Parker–Taubes[16] can be applied for a general case. Boulware[12] argues that for higher-order Lagrangians the positivity of energy does not hold in general. However, Strominger[14] proved the positivity of energy for Lagrangians of class IV.

The notation in this paper follows that of Ref. 6–8. In particular, Greek indices run from 0 through 3; Latin indices run from 1 through 3. Three-covariant ("barred") geometric quantities are defined by means of the $(3+1)$ decomposition of corresponding four-covariant ("unbarred") quantities according to rules presented in Appendix A.

II. THE SYMPLECTIC DYNAMICS OF GENERAL GRAVITATIONAL LAGRANGIANS

A general gravitational Lagrangian L is a scalar density constructed from the metric $g_{\mu\nu}$ and its first- and second-order partial derivatives. Natural transformation properties of L with respect to the action of the diffeomorphism group of space-time imply that the Lagrangian can be rewritten as a function of the metric $g_{\mu\nu}$ and components of the curvature tensor $R_{\alpha\beta\mu\nu}$. We take this fact for granted; the idea of a proof is presented in Refs. 8, 17, and 18. Thus,

$$L = L(g_{\mu\nu}, R_{\alpha\beta\mu\nu}). \tag{2.1}$$

The diffeomorphism invariance of L can be expressed as the following relation:

$$\pounds_Z L = \frac{\partial L}{\partial g_{\mu\nu}} \pounds_Z g_{\mu\nu} + \frac{\partial L}{\partial R_{\alpha\beta\mu\nu}} \pounds_Z R_{\alpha\beta\mu\nu}, \tag{2.2}$$

where Z is an arbitrary vector field on space-time and $\pounds_Z$ denotes the Lie derivative in the direction of Z.

We define the gravitational momenta

$$P^{\mu\nu\alpha\beta} = 2\frac{\partial L}{\partial R_{\alpha\beta\mu\nu}}. \tag{2.3}$$

The properties of the curvature tensor in Riemannian space-times give rise to the following relations:

$$\begin{gathered} P^{\mu\nu\alpha\beta} = P^{\alpha\beta\mu\nu}, \quad P^{\mu\nu\alpha\beta} = -P^{\nu\mu\alpha\beta} = P^{\nu\mu\beta\alpha}, \\ P^{\mu\nu\alpha\beta} + P^{\mu\alpha\beta\nu} + P^{\mu\beta\nu\alpha} = 0. \end{gathered} \tag{2.4}$$

Taking into account the formulas

$$\begin{aligned} &\pounds_Z L = \partial_\tau (Z^\tau L), \quad \pounds_Z g_{\mu\nu} = \nabla_\mu Z_\nu + \nabla_\nu Z_\mu, \\ &\pounds_Z R_{\alpha\beta\mu\nu} = Z^\tau \nabla_\tau R_{\alpha\beta\mu\nu} + \nabla_\alpha Z^\epsilon R_{\epsilon\beta\mu\nu} + \nabla_\beta Z^\epsilon R_{\alpha\epsilon\mu\nu} \\ &\qquad + \nabla_\mu Z^\epsilon R_{\alpha\beta\epsilon\nu} + \nabla_\nu Z^\epsilon R_{\alpha\beta\mu\epsilon}, \end{aligned} \tag{2.5}$$

we get from (2.2) the following relations:

$$\left(Lg^{\mu\nu} - \frac{2\partial L}{\partial g_{\mu\nu}} - 2P^{\mu\tau\omega\xi}R^{\nu}{}_{\tau\omega\xi}\right)\nabla_\mu Z_\nu = 0, \tag{2.6}$$

$$(\nabla_\tau L - 1/2P^{\alpha\beta\mu\nu}\nabla_\tau R_{\alpha\beta\mu\nu})Z^\tau = 0. \tag{2.7}$$

It follows from (2.6) that

$$P^{\mu\tau\omega\xi}R^{\nu}{}_{\tau\omega\xi} = P^{\nu\tau\omega\xi}R^{\mu}{}_{\tau\omega\xi} \tag{2.8a}$$

and

$$Lg^{\mu\nu} - \frac{2\partial L}{\partial g_{\mu\nu}} - 2P^{\mu\tau\omega\xi}R^{\nu}{}_{\tau\omega\xi} = 0. \tag{2.8b}$$

From (2.7) we get

$$\nabla_\tau L = 1/2P^{\alpha\beta\mu\nu}\nabla_\tau R_{\alpha\beta\mu\nu}. \tag{2.9}$$

The variational Euler–Lagrange equations for L read

$$\frac{\delta L}{\delta g_{\mu\nu}} = (Eq)^{\mu\nu}$$
$$= \tfrac{1}{2}\,[Lg^{\mu\nu} - P^{\mu\omega\alpha\beta}R^{\ \nu}_{\ \ \omega\alpha\beta} - \nabla_\alpha\nabla_\beta(P^{\alpha\mu\beta\nu} + P^{\alpha\nu\beta\mu})] = 0. \tag{2.10}$$

Now we return to formula (2.2). Making use of (2.5) and (3.8) we rewrite it in the Noether form

$$\partial_\lambda E^\lambda + (Eq)^{\mu\nu}Ł_Z g_{\mu\nu} = 0. \tag{2.11}$$

The explicit expression for the energy vector density E^λ reads

$$E^\lambda = -[L\delta^\lambda_{\ \tau} - P^{\lambda\omega\alpha\beta}R_{\tau\omega\alpha\beta} - \nabla_\alpha\nabla_\beta(P^{\alpha\lambda\beta}_{\ \ \ \ \tau} + P^{\alpha}_{\ \tau}{}^{\beta\lambda})]Z^\tau - \nabla_\tau B^{\tau\lambda}, \tag{2.12}$$

where

$$B^{\tau\lambda} = 2\nabla_\omega P^{\tau\lambda\omega\nu}Z_\nu - P^{\tau\lambda\omega\mu}\nabla_\omega Z_\mu \tag{2.13}$$

is a skew-symmetric tensor density on M.

We observe that

$$E^\lambda = -2(Eq)^{\tau\lambda}Z_\tau + \text{divergence}. \tag{2.12$'$}$$

This result could be expected because in theories of gravity the energy-momentum generators are expressed by the left-hand sides of field equations.[15,19,20,21,6–8] It follows from (2.11) that

$$\partial_\lambda E^\lambda = 0 \tag{2.14}$$

provided that the field equations (2.10) are satisfied. Therefore the vector density E^λ defines a conserved integral quantity. We take the dual three-form on space-time

$$\nu = \sum_{\lambda=0}^{3} (-1)^\lambda E^\lambda\, dx^0 \wedge \cdots_\lambda \cdots \wedge dx^3 \tag{2.15}$$

and integrate it over any three-dimensional surface σ,

$$E = \int_\sigma \nu. \tag{2.16}$$

The convergence of the integral in (2.16) can be achieved by imposing appropriate boundary conditions at the infinity on σ or by assuming that σ is a compact three-manifold without boundary.

For a fixed vector field Z on space-time the formula (2.16) gives us a function E on the space of all Lorentz metrics on space-time. In general, E depends on the choice of a surface of integration σ. Only if we restrict ourselves to metrics satisfying the field equations (2.10) do we get a function independent of σ. In this case, however, E is the trivial zero function.

Nonetheless the function E plays the role of the Hamiltonian of an infinite-dimensional dynamical system. The idea is to consider E as a function in the space of all Lorentz metrics and to compute its differential dE. We will show that the differential can be expressed as the sum of two integrals. The first of them vanishes if the field equations (2.10) hold, the latter should be expressed by means of the symplectic two-form Ω that is to be defined.

Let us observe that the integral of the divergence in (2.12′) can be transformed into an integral over the boundary of σ and therefore neglected. To calculate dE we take variations of the first summand in (2.12′) and integrate the result over σ. After performing some algebraic-differential calculations we are able to write

$$\begin{aligned} dE\cdot V = &-\int_\sigma \sum_{\lambda=0}^{3} (-1)^\lambda Z^\lambda (Eq)^{\mu\nu}\delta g_{\mu\nu}\, dx^0\wedge\cdots_\lambda\cdots\wedge dx^3 \\ &-\int_\sigma \sum_{\tau=0}^{3} (-1)^\tau [Ł_Z(\nabla^\omega P^{\tau(\mu}{}_{\omega}{}^{\nu)})\delta g_{\mu\nu} - Ł_Z g_{\mu\nu}\delta(\nabla^\omega P^{\tau(\mu}{}_{\omega}{}^{\nu)}) \\ &+ Ł_Z P^{\tau(\mu}{}_{\lambda}{}^{\nu)}\,\delta\Gamma_\mu{}^\lambda{}_\nu - Ł_Z\Gamma_\mu{}^\lambda{}_\nu\,\delta P^{\tau(\mu}{}_{\lambda}{}^{\nu)}]dx^0\wedge\cdots_\tau\cdots\wedge dx^3 \\ &+\int_{\partial\sigma} \sum_{\lambda<\tau} (-1)^{\lambda+\tau} A^{\lambda\tau}\, dx^0\wedge\cdots_\lambda\cdots_\tau\cdots\wedge dx^3. \end{aligned} \tag{2.17}$$

Here V is a (sample) vector tangent to the space of all Lorentz metrices, which is represented by variations $\delta g_{\mu\nu}$; $A^{\lambda\tau}$ is a skew-symmetric tensor density [see (B1)].

Formulas for the Lie derivatives are these of (2.5) and

$$Ł_Z\Gamma_\mu{}^\lambda{}_\nu = \nabla_\mu\nabla_\nu Z^\lambda + R^\lambda{}_{\nu\sigma\mu}Z^\sigma,$$
$$Ł_Z P^{\tau\mu}{}_\lambda{}^\nu = \nabla_\sigma(Z^\sigma P^{\tau\mu}{}_\lambda{}^\nu) + \nabla_\lambda Z^\epsilon P^{\tau\mu}{}_\epsilon{}^\nu - \nabla_\sigma Z^\tau P^{\sigma\mu}{}_\lambda{}^\nu - \nabla_\sigma Z^\mu P^{\tau\sigma}{}_\lambda{}^\nu - \nabla_\sigma Z^\nu P^{\tau\mu}{}_\lambda{}^\sigma, \tag{2.18}$$

and an analogous formula for $Ł_Z(\nabla^\omega P^{\tau\mu}{}_\omega{}^\nu)$. In (2.17)

$$\delta\Gamma_\mu{}^\lambda{}_\nu = \tfrac{1}{2}g^{\lambda\sigma}(\nabla_\mu\delta g_{\nu\sigma} + \nabla_\nu\,\delta g_{\mu\sigma} - \nabla_\sigma\,\delta g_{\mu\nu}), \tag{2.19}$$

the variations $\delta P^{\tau\mu}{}_\lambda{}^\nu$ can be computed by means of (2.3) and the following relations:

$$\delta R^\tau{}_{\beta\mu\nu} = \nabla_\mu\,\delta\Gamma_\nu{}^\tau{}_\beta - \nabla_\nu\,\delta\Gamma_\mu{}^\tau{}_\beta. \tag{2.20}$$

The first term in (2.17) vanishes if the field equations (2.10) are satisfied, the third one vanishes by virtue of the boundary conditions on $\partial\sigma$. The second term can be considered as expressed by a skew-symmetric two-form Ω constructed from the variables

$$g_{\mu\nu},\quad \Gamma_\mu{}^\lambda{}_\nu \tag{2.21a}$$

and their conjugate momenta

$$\nabla^\omega P^{\tau(\mu}{}_\omega{}^{\nu)},\quad P^{\tau(\mu}{}_\lambda{}^{\nu)}. \tag{2.21b}$$

Now we postpone the discussion of the formula (2.17) and define the following symplectic two-form on the space of all Lorentz metrics on space-time [Lor(M)]. For two vectors V_1 and V_2 tangent to Lor(M) and represented by symmetric tensor fields $\delta_1 g_{\mu\nu}$ and $\delta_2 g_{\mu\nu}$ we set

$$
\begin{aligned}
\Omega(V_1,V_2) &= \int_\sigma \sum_{\tau=0}^{3} (-1)^\tau [\delta_1(\nabla^\omega P^{\tau(\mu}{}_\omega{}^{\nu)}) \wedge \delta_2 g_{\mu\nu} \\
&\quad + \delta_1 P^{\tau(\mu}{}_\lambda{}^{\nu)} \wedge \delta_2 \Gamma_\mu{}^\lambda{}_\nu] dx^0 \wedge \cdots{}_\tau \cdots \wedge dx^3. \qquad (2.22)
\end{aligned}
$$

To compute the integral (2.22) formulas (2.19), (2.20), and the linearized version of (2.3) should be used. If appropriate boundary conditions on $\partial\sigma$ are satisfied then (2.22) defines a differential two-form on $\mathrm{Lor}(M)$. Here Ω is a closed two-form because the integrand in (2.22) is expressed in terms of complete variations.

If vectors V_1, V_2 are tangent to $\mathrm{Lor}(M)$, then the value of the integral essentially depends on the choice of a surface σ. If, however, we take a background metric $g_{\mu\nu}$ satisfying the field equations (2.10) and vectors V_1, V_2 satisfying the linearized version of field equations, then the value of Ω does not depend on σ. This fact is proven in Appendix B.

Therefore the formula (2.22) restricted to the space of solutions of field equations (Sol) gives us a natural symplectic structure on this space. In our further considerations we treat Ω as a two-form on $\mathrm{Lor}(M)$ but remember that it is an extension of a geometric object internally related to the space of solutions.

Let us return to formula (2.17). From now on, we assume that the vector field Z is transversal to σ. Therefore the vector Y with components

$$
\delta_Y g_{\mu\nu} = \pounds_Z g_{\mu\nu}, \qquad (2.23)
$$

and, respectively,

$$
\begin{aligned}
&\delta_Y \Gamma_\mu{}^\lambda{}_\nu = \pounds_Z \Gamma_\mu{}^\lambda{}_\nu, \quad \delta_Y(\nabla^\omega P^{\tau(\mu}{}_\omega{}^{\nu)}) = \pounds_Z(\nabla^\omega P^{\tau(\mu}{}_\omega{}^{\nu)}), \\
&\delta_Y P^{\tau(\mu}{}_\lambda{}^{\nu)} = \pounds_Z P^{\tau(\mu}{}_\lambda{}^{\nu)} \qquad (2.23')
\end{aligned}
$$

determines infinitesimal transformations of the symplectic variables induced by the evolution of their initial values from the surface σ into the direction of the vector field Z. Here Y is called the vector of evolution.

Now we formulate the following.

Theorem: The variational Euler–Lagrange equations (2.10) are equivalent to the functional Hamilton equation

$$
dE\, V = -\Omega(Y \wedge V), \qquad (2.24)
$$

where V is an arbitrary vector tangent to $\mathrm{Lor}(M)$ and Y is the vector of evolution.

The four-covariant Hamiltonian formulation presented above is very elegant and simple. It is not very useful, however, for the dynamical analysis of particular Lagrangians. As a matter of fact, on the initial surface σ only ten of 40 quantities $\Gamma_\nu{}^\lambda{}_\nu$ are independent. The remaining 30 express by means of σ tangential derivatives of $g_{\nu\mu}$. Similar problems may arise with the momenta (2.21b).

Therefore in order to determine truly independent symplectic variables we have to perform the (3 + 1) decomposition of four-covariant symplectic variables (2.21). By means of the bar operation described in Appendix A we obtain

$$
\begin{aligned}
&\delta_1(\nabla^\omega P^{0(\mu}{}_\omega{}^{\nu)}) \wedge \delta_2 g_{\mu\nu} + \delta_1 P^{0(\mu}{}_\lambda{}^{\nu)} \wedge \delta_2 \Gamma_\mu{}^\lambda{}_\nu \\
&\quad = \delta_1 \Xi^{ij} \wedge \delta_2 \bar g_{ij} + \delta_1 \zeta^{ij} \wedge \delta_2 z_{ij} \\
&\qquad + \partial_p [\delta_1(1/N\overline{P}^{0p}{}_{0u}) \wedge \delta_2 N^u - \delta_1 \overline{P}^{0ipj} \wedge \delta_2 \bar g_{ij}]. \qquad (2.25)
\end{aligned}
$$

Here

$$
\begin{aligned}
&z_{ij} = 2\overline{\Gamma}_i{}^0{}_j, \quad \zeta^{ij} = \overline{P}^{0(i}{}_0{}^{j)}, \\
&\Xi^{ij} = (\overline{\nabla^\omega P^{0(i}{}_\omega{}^{j)}}) + \overline{\nabla}^v \overline{P}^{0(i}{}_v{}^{j)} \\
&\qquad + \tfrac{1}{2}(\overline{P}^{0u0i}\overline{\Gamma}_u{}^j{}_0 + \overline{P}^{0u0j}\overline{\Gamma}_u{}^i{}_0). \qquad (2.26)
\end{aligned}
$$

Therefore the quantities

$$
\bar g_{ij}, \quad z_{ij}, \qquad (2.27a)
$$

$$
\Xi^{ij}, \quad \zeta^{ij}, \qquad (2.27b)
$$

are three-covariant symplectic variables on the initial surface σ.

The symplectic two-form Ω can be rewritten in terms of these variables,

$$
\begin{aligned}
\Omega(V_1,V_2) &= \int_\sigma [\delta_1 \Xi^{ij} \wedge \delta_2 \bar g_{ij} \\
&\quad + \delta_1 \zeta^{ij} \wedge \delta_2 z_{ij}] dx^1 \wedge dx^2 \wedge dx^3, \qquad (2.28)
\end{aligned}
$$

where the boundary terms have been neglected.

Let us observe that the quantities z_{ij} are in a one-to-one correspondence with $\bar\partial_0 \bar g_{ij}$, that is,

$$
z_{ij} = \overline{\nabla}_0 \bar g_{ij} \quad [\text{cf. (A13)}], \qquad (2.29)
$$

and therefore $\bar g_{ij}$ and z_{ij} are kinematically independent symplectic variables. Their conjugate momenta Ξ^{ij} and ζ^{ij} are related to third and second time derivatives of $\bar g_{ij}$. For canonically regular gravitational Lagrangians the quantities ζ^{ij} are in a one-to-one correspondence with $\bar\partial_0^2 \bar g_{ij}$ and Ξ^{ij} are in such relations with $\bar\partial_0^3 \bar g_{ij}$. These problems are discussed in Sec. III.

Now we are able to write the field equations (2.10) in the dynamical formulation

$$
\begin{aligned}
(\overline{Eq})^{mn} = &-\overline{\nabla}_0 \Xi^{mn} + \tfrac{1}{2}\bar g^{mn}\overline{L} - \tfrac{1}{4}(\overline{P}^{m\tau\omega\epsilon}\overline{R}^n{}_{\tau\omega\epsilon} + \overline{P}^{n\tau\omega\epsilon}\overline{R}^m{}_{\tau\omega\epsilon}) - \tfrac{1}{2}\overline{\nabla}_0 \overline{\Gamma}_r{}^n{}_0 \zeta^{mr} - \tfrac{1}{2}\overline{\nabla}_0 \overline{\Gamma}_r{}^m{}_0 \zeta^{nr} \\
&- \tfrac{1}{2}[-\lambda_a{}^n{}_p \overline{P}^{0pam} + (\overline{\nabla}_a + \partial_a \ln N)(\overline{\nabla}_b + \partial_b \ln N)\overline{P}^{ambn} \\
&+ (\overline{\nabla}_a + \partial_a \ln N)(-\overline{\nabla}^a \ln N \zeta^{mn} - \overline{\Gamma}_w{}^m{}_0 \overline{P}^{0awn} + \overline{\Gamma}_w{}^a{}_0 \overline{P}^{0nwm} + \overline{\nabla}^m \ln N \zeta^{an} + \overline{\Gamma}_w{}^a{}_0 \overline{P}^{0mwn} + \overline{\Gamma}_w{}^a{}_0 \overline{P}^{0nwm}) \\
&+ \overline{\nabla}^n \ln N(\overline{\nabla}_w \zeta^{mw} + \overline{\Gamma}_w{}^0{}_v \overline{P}^{0wmv}) + \overline{\Gamma}_r{}^n{}_0(-\overline{\Gamma}_v{}^m{}_0 \zeta^{vr} + (\overline{\nabla}_v + \partial_v \ln N)\overline{P}^{0mvr} + \partial_v \ln N \overline{P}^{0rvm} + \overline{\Gamma}_v{}^0{}_w \overline{P}^{wmvr}) \\
&+ \overline{\Gamma}_r{}^n{}_0(\overline{\nabla}_s \overline{P}^{0smr} + \overline{\Gamma}_s{}^r{}_0 \zeta^{ms} - \overline{\Gamma}_s{}^m{}_0 \zeta^{rs})] - \tfrac{1}{2}[m \leftrightarrow n] = 0. \qquad (2.30)
\end{aligned}
$$

Here the Christofel-like tensors $\lambda_a{}^n{}_p$ are defined in (A18).

$$
\begin{aligned}
(\overline{Eq})^{00} &= \tfrac{1}{2}[z_{ij}\Xi^{ij} - \overline{L} - 2\zeta^{sb}\overline{R}_{0s0b} \\
&\quad + 2\overline{\nabla}_p \overline{\nabla}_q \zeta^{pq} + \tfrac{1}{2} z_{ij} z^j{}_p \zeta^{ip}] = 0, \qquad (2.31a)
\end{aligned}
$$

$$
(\overline{Eq})^{0n} = \overline{\nabla}_p \Xi^{pn} + \overline{\nabla}_p (z_q{}^n \zeta^{qp}) - \tfrac{1}{2}\zeta^{pq}\overline{\nabla}^n z_{pq} = 0. \qquad (2.31b)
$$

Equations (2.30) determine the dynamics of Ξ^{ij}, Eqs. (2.31) are constraints. It is easy to see that the left-hand sides of (2.31b) are functions of the symplectic variables (2.27) and their spatial derivatives. The same is true for (2.31a) but to prove it we have to consider the linearized version of this equation and observe that it depends only on variations of the symplectic variables [cf. (B4)].

The constraints (2.31) are conserved in the process of evolution. This fact can be proved either by a direct computation or making use of the contracted Bianchi identities

$$\nabla_\mu (Eq)^{\mu\nu} = 0. \tag{2.32}$$

Those, in turn, we get by a direct differentiation of (2.10).

Equations (2.30) and (2.31) in the form presented above give us the complete dynamical picture only for regular gravitational Lagrangians. In these cases we have 24 kinematically independent symplectic variables (2.27) and their initial values have to satisfy four symplectic constraints (2.31). The constraints reduce four degrees of freedom. Moreover, the action of the diffeomorphism group, Diff M, reduces four further degrees of freedom. Thus for canonically regular gravitational Lagrangians of second order we have $24 - 2\cdot 4 = 16$ independent degrees of freedom in the phase space.

If we write the system (2.30) and (2.31) in terms of $\bar{g}_{ij}$ and gauge variables N, N^k then we have six fourth-order dynamical equations and four constraints. For degenerate Lagrangians we have kinematical relations among variables (2.27) and to accomplish the dynamical analysis we have at first to find kinematically independent symplectic variables. In Sec. III we deal with several such examples.

III. THE CANONICAL CLASSIFICATION OF QUADRATIC GRAVITATIONAL LAGRANGIANS

A general gravitational Lagrangian (at most) quadratic in components of Riemann tensor is

$$L = A\sqrt{-g}R + \tfrac{1}{2}B\sqrt{-g}R^2 + \tfrac{1}{4}C\sqrt{-g}R^{\mu\nu}R_{\mu\nu} + \tfrac{1}{4}D\sqrt{-g}R^{\alpha\beta\mu\nu}R_{\alpha\beta\mu\nu}. \tag{3.1}$$

It turns out, however, that the last term in this expression can be replaced with a linear combination of other terms. The Bach–Lanczos identity

$$\tfrac{1}{4}\sqrt{-g}\,\epsilon_{\alpha\beta\lambda\tau}\epsilon^{\mu\nu\rho\sigma}R^{\alpha\beta}{}_{\mu\nu}R^{\lambda\tau}{}_{\rho\sigma} = \sqrt{-g}(R^2 - 4R^{\alpha\beta}R_{\alpha\beta} + R^{\alpha\beta\mu\nu}R_{\alpha\beta\mu\nu}) \tag{3.2}$$

and triviality of the variational equations for the left-hand side Lagrangian in (3.2) enable us to eliminate the D term in (3.1). Therefore without loss of generality we may assume that $D=0$.

The variational principle for the Lagrangian (3.1) leads to the Euler–Lagrange equations (2.10) which are expected to form a fourth-order system for the components of a metric on space-time. It is known, however, that for the Einstein theory (i.e., $B=C=0, A\neq 0$) we have a second-order system for the components of the metric. Therefore we expect that different combinations of the coefficients A, B, and C in (3.1) give rise to qualitatively different systems of field equations.

First of all, we specify conditions for the general Lagrangian (3.1) to be canonically regular. This means that the Euler–Lagrange equations in the $(3+1)$ picture split into the set of six independent fourth-order equations for six components of the spatial metric $\bar{g}_{ij}$,

$$(\overline{Eq})^{ij} = 0, \tag{3.3}$$

and four Hamiltonian constraints

$$(\overline{Eq})^{0\mu} = 0. \tag{3.4}$$

In such a case we expect a one-to-one correspondence between the values of the metric components $\bar{g}_{ij}$, its first, second, and third derivatives on an initial surface σ, and the values of the canonical variables $\Xi^{ij}, \bar{g}_{ij}, \zeta^{ij}, z_{ij}$ on σ. We start with some formulas valid for a general Lagrangian (3.1)

$$\begin{aligned}P^{\alpha\beta\mu\nu} = {} & (A+BR)\sqrt{-g}(g^{\alpha\mu}g^{\beta\nu} - g^{\alpha\nu}g^{\beta\mu}) \\ & + \tfrac{1}{4}C\sqrt{-g}(g^{\alpha\mu}R^{\beta\nu} - g^{\alpha\nu}R^{\beta\mu} \\ & - g^{\beta\mu}R^{\alpha\nu} + g^{\beta\nu}R^{\alpha\mu}),\end{aligned} \tag{3.5}$$

$$\begin{aligned}\nabla_\omega P^{\alpha\beta\omega\nu} = {} & B\sqrt{-g}(\nabla^\alpha R g^{\beta\nu} - \nabla^\beta R g^{\alpha\nu}) \\ & + \tfrac{1}{4}C\sqrt{-g}(\nabla^\alpha R^{\beta\nu} - \nabla^\beta R^{\alpha\nu}) \\ & + \tfrac{1}{8}C\sqrt{-g}(-g^{\alpha\nu}\nabla^\beta R + g^{\beta\nu}\nabla^\alpha R).\end{aligned} \tag{3.6}$$

To get (3.6) we make use of the contracted Bianchi identities

$$\nabla_\mu (R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R) = 0. \tag{3.7}$$

In the $(3+1)$ picture

$$\overline{P}^{0b0n} = -\tfrac{1}{4}C\sqrt{\bar{g}}\,\overline{R}^{bn} + \sqrt{\bar{g}}(-A - B\overline{R} + \tfrac{1}{4}C\overline{R}^{00})\bar{g}^{bn}, \tag{3.8}$$

$$\overline{P}^{0bmn} = \tfrac{1}{4}C\sqrt{\bar{g}}(-\bar{g}^{bm}\overline{R}^{0n} + \bar{g}^{bn}\overline{R}^{0m}), \tag{3.9}$$

$$\begin{aligned}\overline{\nabla_\omega P^{0b\omega n}} = {} & (B + \tfrac{1}{8}C)\sqrt{\bar{g}}\bar{g}^{bn}\,\overline{\nabla^0 R} \\ & + \tfrac{1}{4}C\sqrt{\bar{g}}(\overline{\nabla^0 R^{bn}} - \overline{\nabla^b R^{0n}}).\end{aligned} \tag{3.10}$$

Taking into account formulas of Appendix A we have

$$\begin{aligned}\overline{P}^{0b0n} = {} & -\tfrac{1}{4}C\,[\sqrt{\bar{g}}\,{}^3\overline{R}^{bn} + \overline{\nabla}_0(\sqrt{\bar{g}}\,\overline{\Gamma}^{b0n}) + 2\sqrt{\bar{g}}\,\overline{\Gamma}^{b0u}\overline{\Gamma}^{n0}{}_u - \sqrt{\bar{g}}(\overline{\nabla}^b + \overline{\nabla}^b \ln N)\overline{\nabla}^n \ln N\,] \\ & + \tfrac{1}{4}C\bar{g}^{bn}[\sqrt{\bar{g}}(\overline{\nabla}_s + \overline{\nabla}_s \ln N)\overline{\nabla}^s \ln N - \overline{\nabla}_0(\sqrt{\bar{g}}\,\overline{\Gamma}_s{}^s{}_0) + \sqrt{\bar{g}}((\overline{\Gamma}_s{}^s{}_0)^2 - \overline{\Gamma}_r{}^0{}_s\overline{\Gamma}^{r0s})] \\ & - B\bar{g}^{bn}[-2\sqrt{\bar{g}}(\overline{\nabla}_s + \overline{\nabla}_s \ln N)\overline{\nabla}^s \ln N + 2\overline{\nabla}_0(\sqrt{\bar{g}}\,\overline{\Gamma}_s{}^s{}_0) + \sqrt{\bar{g}}\,{}^3\overline{R} + \sqrt{\bar{g}}(\overline{\Gamma}_p{}^0{}_q\overline{\Gamma}^{p0q} - (\overline{\Gamma}_s{}^s{}_0)^2)] - A\sqrt{\bar{g}}\bar{g}^{bn};\end{aligned} \tag{3.11}$$

$$\overline{P}^{0bmn} = \tfrac{1}{4}C\sqrt{\bar{g}}[\bar{g}^{bm}(\overline{\nabla}_s\overline{\Gamma}^{s0n} - \overline{\nabla}^n\overline{\Gamma}_s{}^s{}_0) - \bar{g}^{bn}(\overline{\nabla}_s\overline{\Gamma}^{s0m} - \overline{\nabla}^m\overline{\Gamma}_s{}^s{}_0)]; \tag{3.12}$$

$$\overline{\nabla_\omega P^{0b\omega n}} = -(B+\tfrac{1}{8}C)\bar g^{bn}(\overline\nabla_0 - \overline\Gamma_s{}^s{}_0)\big[2\overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma_s{}^s{}_0) - 2\sqrt{\bar g}(\overline\nabla_s + \overline\nabla_s \ln N)\overline\nabla^s \ln N$$
$$+\sqrt{\bar g}\,{}^3\overline R + \sqrt{\bar g}(-(\overline\Gamma_s{}^s{}_0)^2 + \overline\Gamma_p{}^0{}_q\overline\Gamma^{p0q})\big] - \tfrac14 C(\overline\nabla_0 - \overline\Gamma_s{}^s{}_0)\big[\sqrt{\bar g}\,{}^3\overline R{}^{bn} + \overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma^{b0n}) + 2\sqrt{\bar g}\,\overline\Gamma^{b0}{}_v\overline\Gamma^{n0v}$$
$$-\sqrt{\bar g}(\overline\nabla^b + \overline\nabla^b \ln N)\overline\nabla^n \ln N\big] - \tfrac14 C(\overline\nabla^b + \overline\nabla^b \ln N)\big[\sqrt{\bar g}(-\overline\nabla_v\overline\Gamma^{v0n} + \overline\nabla^n\overline\Gamma_s{}^s{}_0)\big]$$
$$-\tfrac14 C\sqrt{\bar g}\,\overline\nabla^n \ln N(-\overline\nabla_v\overline\Gamma^{b0v} + \overline\nabla^b\overline\Gamma_s{}^s{}_0) - \tfrac14 C(\overline\Gamma_r{}^b{}_0 + \overline\Gamma^{b0}{}_r)\big[\sqrt{\bar g}\,{}^3\overline R{}^{rn} + \overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma^{r0n}) + 2\sqrt{\bar g}\,\overline\Gamma^{r0}{}_v\overline\Gamma^{n0v}$$
$$-\sqrt{\bar g}(\overline\nabla^r + \overline\nabla^r \ln N)\overline\nabla^n \ln N\big] - \tfrac14 C\overline\Gamma_r{}^n{}_0\big[\sqrt{\bar g}\,{}^3\overline R{}^{br} + \overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma^{b0r}) + 2\sqrt{\bar g}\,\overline\Gamma^{b0}{}_v\overline\Gamma^{r0v}$$
$$-\sqrt{\bar g}(\overline\nabla^b + \overline\nabla^b \ln N)\overline\nabla^r \ln N\big] - \tfrac14 C\overline\Gamma^{b0n}\big[-\overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma_s{}^s{}_0) - \sqrt{\bar g}\,\overline\Gamma_p{}^0{}_q\overline\Gamma^{p0q} + \sqrt{\bar g}(\overline\Gamma_s{}^s{}_0)^2$$
$$+\sqrt{\bar g}(\overline\nabla_s + \overline\nabla_s \ln N)\overline\nabla^s \ln N\big]. \tag{3.13}$$

It follows from (3.11) that if $C \neq 0$ then there is a chance to solve these equations for $\overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma^{b0n})$ and obtain these quantities as functions of symplectic variables, the gauge variable N, and spatial derivatives of these variables. Similarly, we may try to solve (3.13) for $\overline\nabla_0\overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma^{b0n})$. [The terms $\overline\nabla_0{}^3\overline R{}^{ij}$ and ${}^3\overline R$ in (3.13) do not cause any problems because we can transform them making use of (A24) and (A25).]

Let us observe that even if $C \neq 0$ some difficulties may arise in determining the traces $\overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma^{i0j})\bar g_{ij}$ and $\overline\nabla_0\overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma^{i0j})\bar g_{ij}$ from (3.11) and (3.13). We have

$$\overline P^{0b0n}\bar g_{bn} = (-3B - \tfrac14\cdot C)\sqrt{\bar g}\,{}^3\overline R + (\tfrac34\cdot C + 3B)\sqrt{\bar g}((\overline\Gamma_s{}^s{}_0)^2 - \overline\Gamma_p{}^0{}_q\overline\Gamma^{p0q}) - 3A\sqrt{\bar g}$$
$$+(C+6B)(\overline\nabla_s + \overline\nabla_s \ln N)\overline\nabla^s \ln N - (C+6B)\overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma_s{}^s{}_0), \tag{3.14}$$

$$\overline{\nabla_\omega P^{0b\omega n}}\bar g_{bn} = -3(B+\tfrac18 C)(\overline\nabla_0 - \overline\Gamma_s{}^s{}_0)\big[\sqrt{\bar g}\,{}^3\overline R + \sqrt{\bar g}(\overline\Gamma_p{}^0{}_q\overline\Gamma^{p0q} - (\overline\Gamma_s{}^s{}_0)^2)\big] - \tfrac14 C(\overline\nabla_0 - \overline\Gamma_s{}^s{}_0)(\sqrt{\bar g}\,{}^3\overline R)$$
$$+\tfrac14 C\overline\Gamma_s{}^s{}_0\big[2\overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma_s{}^s{}_0) - 2\sqrt{\bar g}(\overline\nabla^s + \overline\nabla^s \ln N)\overline\nabla_s \ln N - \sqrt{\bar g}(\overline\Gamma_s{}^s{}_0)^2 + \sqrt{\bar g}\,\overline\Gamma_p{}^0{}_q\overline\Gamma^{p0q}\big]$$
$$-\tfrac12 C\sqrt{\bar g}\,\overline\nabla_b \ln N(-\overline\nabla_v\overline\Gamma^{b0v} + \overline\nabla^b\overline\Gamma_s{}^s{}_0) - \tfrac14 C\sqrt{\bar g}\overline\nabla_n(-\overline\nabla_v\overline\Gamma^{n0v} + \overline\nabla^n\overline\Gamma_s{}^s{}_0)$$
$$-\tfrac14 C\overline\Gamma_n{}^0{}_s\big[\sqrt{\bar g}\,{}^3\overline R{}^{ns} + \overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma^{n0s}) + 2\sqrt{\bar g}\,\overline\Gamma^{n0v}\overline\Gamma^{s0}{}_v - \sqrt{\bar g}(\overline\nabla^n + \overline\nabla^n \ln N)\overline\nabla^s \ln N)\big]$$
$$-(6B+C)\big[\overline\nabla_0\overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma_s{}^s{}_0) - \overline\nabla_0(\sqrt{\bar g}(\overline\nabla_s + \overline\nabla_s \ln N)\overline\nabla^s \ln N)\big]. \tag{3.15}$$

We infer from (3.14) and (3.15) that if $C \neq -6B$ then $\overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma_s{}^s{}_0)$ and $\overline\nabla_0\overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma_s{}^s{}_0)$ can be determined as functions of the symplectic variables, their spatial derivatives, the lapse N, and its spatial and time derivatives. Therefore we can compute $\overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma^{i0j})$ and $\overline\nabla_0\overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma^{i0j})$ from (3.11) and (3.13). Bearing in mind that $\overline\nabla_0\bar g_{ij} = 2\overline\Gamma_i{}^0{}_j$ [cf. (A13)] we conclude that if $C \neq 0$ and $C \neq -6B$ then we have a regular fourth-order system for the spatial metric $\bar g_{ij}$. In such a case we have a one-to-one correspondence between the momenta ζ^{ij} and the time derivatives of z_{ij} as well as a one-to-one correspondence between the momenta Ξ^{ij} and the time derivatives of ζ^{ij}. That is,

$$\begin{aligned}\zeta^{ij} &= \zeta^{ij}(\overline\nabla_0 z_{pq}, z_{pq}, \bar g_{pq}, N,\\ &\qquad \text{spatial derivatives of } (z_{pq}, \bar g_{pq}, N)),\\ \Xi^{ij} &= \Xi^{ij}(\overline\nabla_0\zeta^{pq}, \zeta^{pq}, z_{pq}, \bar g_{pq}, N,\\ &\qquad \text{spatial derivatives of } (\zeta^{pq}, z_{pq}, \bar g_{pq}, N)).\end{aligned} \tag{3.16}$$

Remark: In (3.16) first- and second-order spatial derivatives of corresponding variables may appear.

It follows from (3.11) and (3.13) that for regular Lagrangians the expressions

$$\begin{aligned}&\overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma^{i0j}) - \sqrt{\bar g}(\overline\nabla^i + \overline\nabla^i \ln N)\overline\nabla^j \ln N,\\ &\overline\nabla_0\big[\overline\nabla_0(\sqrt{\bar g}\,\overline\Gamma^{i0j}) - \sqrt{\bar g}(\overline\nabla^i + \overline\nabla^i \ln N)\overline\nabla^j \ln N\big]\end{aligned} \tag{3.17}$$

are functions of ζ^{pq}, z_{pq}, $\bar g_{pq}$, ${}^3\overline R_{pq}$ and of Ξ^{pq}, ζ^{pq}, z_{pq}, $\bar g_{pq}$, ${}^3\overline R_{pq}$, $\overline\nabla_r z_{pq}$, $\overline\nabla_s\overline\nabla_r z_{pq}$, $\overline\nabla_r \ln N$, respectively.

Taking into account relations (3.16) and (3.17) and the field equations (2.30) we have the following first-order dynamical system:

$$\begin{aligned}\overline\nabla_0\bar g_{ij} &= z_{ij},\\ \overline\nabla_0 z_{ij} &= \overline\nabla_0 z_{ij}(\zeta^{pq}, z_{pq}, \bar g_{pq}, {}^3\overline R_{pq}, \overline\nabla_r \ln N, \overline\nabla_s\overline\nabla_r \ln N),\\ \overline\nabla_0\zeta^{ij} &= \overline\nabla_0\zeta^{ij}(\Xi^{pq}, \zeta^{pq}, z_{pq}, \bar g_{pq}, \overline\nabla_r z_{pq}, \overline\nabla_s\overline\nabla_r z_{pq}, {}^3\overline R_{pq},\\ &\qquad \overline\nabla_r \ln N, \overline\nabla_s\overline\nabla_r \ln N),\\ \overline\nabla_0\Xi^{ij} &= \overline\nabla_0\Xi^{ij}(\Xi^{pq}, \zeta^{pq}, z_{pq}, \bar g_{pq}, \overline\nabla_r\zeta^{pq}, \overline\nabla z_{pq}\overline\nabla_s\overline\nabla_r z_{pq}, {}^3\overline R_{pq},\\ &\qquad \overline\nabla_r \ln N, \overline\nabla_s\overline\nabla_r \ln N).\end{aligned} \tag{3.18}$$

If $C \neq 0$ and $C = -6B$ then the terms $\bar g^{ij}\overline\nabla_0 z_{ij}$ and $\bar g_{ij}\overline\nabla_0\zeta^{ij}$ cannot be determined from (3.11) and (4.13). In such cases Eqs. (3.14) and (3.15) become symplectic constraints, which in the explicit form read

$$\operatorname{tr}\zeta = \tfrac32 B\sqrt{\bar g}\,[{}^3\overline R + \tfrac14((\operatorname{tr} z)^2 - z_{pq}z^{pq})] + 3A\sqrt{\bar g}, \tag{3.19}$$

$$\operatorname{tr}\Xi = -\tfrac12\zeta^{pq}z_{pq} + \tfrac32 B\sqrt{\bar g}(\overline\nabla_p\overline\nabla_p z^{pq} - \overline\nabla^p\overline\nabla_p \operatorname{tr} z) - {}^t\zeta^{pq}\,{}^t z_{pq}. \tag{3.20}$$

We have a one-to-one correspondence between ${}^t\zeta^{pq}$ and $\overline\nabla_0{}^t z_{ij}$ as well as between ${}^t\Xi^{pq}$ and $\overline\nabla_0{}^t\zeta^{pq}$ but from the kinematical relations (3.11) and (3.13) nothing can be said about $\overline\nabla_0 \operatorname{tr} z$ and $\overline\nabla_0 \operatorname{tr}\zeta$. (Here ${}^t\zeta^{pq}$, ${}^t z_{pq}$, ${}^t\Xi^{pq}$ are the traceless parts of corresponding entities, e.g., ${}^t z_{pq} = z_{pq} - \tfrac13\bar g_{pq} \operatorname{tr} z$.]) In order to understand better the dynamics of the system we rewrite the formula for the symplectic two-form (2.28) in terms of new symplectic variables. After some differential-algebraic operations the integrand in (2.28) can be transformed to the following form:

$$\begin{aligned}
&\delta_1\Xi^{ij}\wedge\delta_2\bar{g}_{ij}+\delta_1\zeta^{ij}\wedge\delta_2 z_{ij}\\
&\quad=\delta_1\Upsilon^{ij}\wedge\delta_2(\bar{g}^{-1/3}\bar{g}_{ij})+\delta_1(\bar{g}^{-1/3}\,{}^t\zeta^{ij})\wedge\delta_2(\bar{g}^{1/3}\,{}^tz_{ij})\\
&\qquad+\delta_1(\bar{g}^{-1/3}\operatorname{tr}\zeta)\wedge\delta_2(\bar{g}^{1/3}(\tfrac{1}{3})\operatorname{tr}z)\\
&\qquad+\delta_1(\bar{g}^{-1/3}(\operatorname{tr}\Xi-{}^t\zeta^{ij}\,{}^tz_{ij})\wedge\delta_2(\bar{g}^{1/3}). \qquad (3.21)
\end{aligned}$$

The quantities

$$\bar{g}^{-1/3}\bar{g}_{ij},\quad \bar{g}^{1/3}\,{}^tz_{ij} \qquad (3.22a)$$

are the dynamical symplectic positions with the conjugate momenta

$$\Upsilon^{ij}=\bar{g}^{1/3}({}^t\Xi^{ij}+\tfrac{1}{3}({}^tz^{ij}\operatorname{tr}\zeta+{}^t\zeta^{ij}\operatorname{tr}z)),\quad \bar{g}^{-1/3}\,{}^t\zeta^{ij}. \qquad (3.22b)$$

These 20 quantities are kinematically independent on the initial surface.

It follows from (3.11), (3.13), and the equations

$$(\overline{Eq})^{pq}-\tfrac{1}{3}\bar{g}^{pq}(\overline{Eq})^{mn}\bar{g}_{mn}=0 \qquad (3.23)$$

that the dynamics of the quantities

$$\bar{g}^{-1/3}\bar{g}_{ij},\quad {}^tz_{ij},\quad {}^t\zeta^{ij},\quad {}^t\Xi^{ij}$$

is determined by the following first-order system:

$$\begin{aligned}
&\overline{\nabla}_0(\bar{g}^{-1/3}\bar{g}_{ij})=\bar{g}^{-1/3}\,{}^tz_{ij},\\
&\overline{\nabla}_0{}^tz_{ij}=\overline{\nabla}_0{}^tz_{ij}({}^t\zeta^{pq},{}^tz_{pq},\operatorname{tr}z,\bar{g}^{-1/3}\bar{g}_{pq},\bar{g},{}^3\overline{R}_{pq},\\
&\qquad\overline{\nabla}_r\ln N,\overline{\nabla}_s\overline{\nabla}_r\ln N),\\
&\overline{\nabla}_0{}^t\zeta^{ij}=\overline{\nabla}_0{}^t\zeta^{ij}({}^t\Xi^{pq},{}^t\zeta^{pq},{}^tz_{pq},\operatorname{tr}z,\bar{g}^{-1/3}\bar{g}_{pq},\bar{g},\overline{\nabla}_r{}^tz_{pq},\overline{\nabla}_s\overline{\nabla}_r{}^tz_{pq},\\
&\overline{\nabla}_r\operatorname{tr}z,\overline{\nabla}_s\overline{\nabla}_r\operatorname{tr}z,{}^3\overline{R}_{pq},\overline{\nabla}_r\ln N,\overline{\nabla}_r\overline{\nabla}_s\ln N), \qquad (3.24)\\
&\overline{\nabla}_0{}^t\Xi^{ij}=\overline{\nabla}_0{}^t\Xi^{ij}({}^t\Xi^{pq},{}^t\zeta^{pq},{}^tz_{pq},\operatorname{tr}z,\bar{g}^{-1/3}\bar{g}_{pq},\bar{g},\\
&\qquad\overline{\nabla}_r{}^t\zeta^{pq},\overline{\nabla}_r{}^tz_{pq},\overline{\nabla}_s\overline{\nabla}_r{}^tz_{pq},\overline{\nabla}_r\operatorname{tr}z,\\
&\qquad\overline{\nabla}_s\overline{\nabla}_r\operatorname{tr}z,{}^3\overline{R}_{pq},\overline{\nabla}_r\ln N,\overline{\nabla}_s\overline{\nabla}_r\ln N).
\end{aligned}$$

The quantities $\operatorname{tr}\Xi$ and $\operatorname{tr}\zeta$ do not appear at the right-hand sides of (3.24) for they have been eliminated by means of (3.19) and (3.20). Of course, we remember that the dynamics of $\bar{g}$ is given by

$$\overline{\nabla}_0\bar{g}=\bar{g}\operatorname{tr}z. \qquad (3.25)$$

Therefore the only quantity whose dynamics is not determined yet is $\operatorname{tr} z$. To clarify the situation we recall that so far we have made use only of five dynamical Hamiltonian equations (3.23). The question is, what does the equation

$$(\overline{Eq})^{mn}\bar{g}_{mn}=0 \qquad (3.26)$$

bring about? The following lemma helps us to answer this problem.

Lemma 1: If in the general formula (3.1) $C=-6B$ then

$$(Eq)^{\mu\nu}g_{\mu\nu}=AR\sqrt{-g}. \qquad (3.27)$$

Let $A\neq 0$ then $(\overline{Eq})^{mn}\bar{g}_{mn}-(\overline{Eq})^{00}=A\sqrt{\bar{g}}\overline{R}$.

Taking into account (3.26) and the Hamiltonian constraint (2.31a) we get

$$\begin{aligned}
\sqrt{\bar{g}}\,\overline{R}\;=&-2\sqrt{\bar{g}}(\overline{\nabla}^s+\overline{\nabla}^s\ln N)\overline{\nabla}_s\ln N+\overline{\nabla}_0(\sqrt{\bar{g}}\operatorname{tr}z)\\
&+\sqrt{\bar{g}}\,{}^3\overline{R}+\tfrac{1}{4}\sqrt{\bar{g}}({}^tz_{pq}\,{}^tz_{pq}-\tfrac{2}{3}(\operatorname{tr}z)^2)=0. \qquad (3.28)
\end{aligned}$$

Equations (3.26) and (3.28) enable us to determine the dynamics of $\operatorname{tr} z$. Let us observe that the first-order dynamical system (3.24), (3.25), and (3.28) can be written as a system of five fourth-order equations for (five independent) quantities $\bar{g}^{-1/3}\bar{g}_{ij}$ and one second-order equation for $\bar{g}^{1/3}$.

Now we discuss the case $C\neq 0$, $C=-6B$, $A=0$. We observe that Eq. (3.26) is equivalent to the Hamiltonian constraint (2.31a). Therefore we have no equation for the evolution of $\operatorname{tr} z$. This quantity seems to be a gauge-type variable. As a matter of fact, the following lemma describes the corresponding gauge transformation.

Lemma 2: If $C=-6B$, $A=0$ then the Lagrangian (3.1) is invariant (up to a total divergence) under the conformal (scale) transformation

$$g_{\mu\nu}\rightarrow\tau g_{\mu\nu}, \qquad (3.29)$$

where $\tau=\tau(x^\lambda)$ is a positive function on space-time. For details see Appendix C.

Therefore the theory is scale invariant and

$$\operatorname{tr}z=\bar{g}^{-1}\overline{\nabla}_0\bar{g} \qquad (3.30)$$

is the gauge variable for the transformation (3.29). In such a case the system (3.24), (3.25) gives rise only to five fourth-order equations for five independent components of the unimodular, symmetric 3×3 matrix $[\bar{g}^{-1/3}\bar{g}_{ij}]$.

Now we consider $R+R^2$ gravity, i.e., $C=0, B\neq 0$. We have from (2.10),

$$\begin{aligned}
(Eq)^{\mu\nu}=&B\sqrt{-g}(\nabla^\mu\nabla^\nu R-g^{\mu\nu}\nabla^\sigma\nabla_\sigma R)\\
&+B\sqrt{-g}R(\tfrac{1}{4}g^{\mu\nu}R-R^{\mu\nu})\\
&+A\sqrt{-g}(\tfrac{1}{2}g^{\mu\nu}-R^{\mu\nu})=0. \qquad (3.31)
\end{aligned}$$

Let us observe that

$$(Eq)^{\alpha\beta}g_{\alpha\beta}=-3B\sqrt{-g}\nabla^\sigma\nabla_\sigma R+A\sqrt{-g}R \qquad (3.32)$$

and

$$\begin{aligned}
&(Eq)^{\mu\nu}-\tfrac{1}{3}g^{\mu\nu}(Eq)^{\alpha\beta}g_{\alpha\beta}\\
&\quad=B\sqrt{-g}\nabla^\mu\nabla^\nu R+B\sqrt{-g}R(\tfrac{1}{4}g^{\mu\nu}R-R^{\mu\nu})\\
&\qquad+A\sqrt{-g}(\tfrac{1}{6}g^{\mu\nu}R-R^{\mu\nu}). \qquad (3.33)
\end{aligned}$$

In the symplectic picture the symplectic variables (2.27) are not independent. It follows from (3.8)–(3.10) that

$$\begin{aligned}
&\zeta^{ij}=\sqrt{\bar{g}}(A+B\overline{R})\bar{g}^{ij},\\
&\Xi^{ij}=-B\sqrt{\bar{g}}\bar{g}^{ij}\overline{\nabla}_0\overline{R}-(A+B\overline{R})\sqrt{\bar{g}}\,\overline{\Gamma}^{i0j}.
\end{aligned} \qquad (3.34)$$

The symplectic integrand in (2.28) can be expressed as follows:

$$\begin{aligned}
&\delta_1\Xi^{ij}\wedge\delta_2\bar{g}_{ij}+\delta_1\zeta^{ij}\wedge\delta_2 z_{ij}\\
&\quad=\delta_1\Lambda^{ij}\wedge\delta_2(\bar{g}^{-1/3}\bar{g}_{ij})\\
&\qquad+\delta_1\Psi\wedge\delta_2(\bar{g}^{1/3})+\delta_1(\operatorname{tr}\pi)\wedge\delta_2 B\overline{R}), \qquad (3.35)
\end{aligned}$$

where

$$\pi^{ij}=\sqrt{\bar{g}}(\overline{\Gamma}^{i0j}-\bar{g}^{ij}\overline{\Gamma}_p{}^0{}_q\bar{g}^{pq}) \qquad (3.36a)$$

are the standard ADM momenta,

$$\begin{aligned}
&\Lambda^{ij}=\bar{g}^{1/3}(A+B\overline{R})\,{}^t\pi^{ij},\\
&\Psi=\bar{g}^{-1/3}(-3B\sqrt{\bar{g}}\overline{\nabla}_0\overline{R}+(A+B\overline{R})\operatorname{tr}\pi).
\end{aligned} \qquad (3.36b)$$

The dynamics can be written in terms of 14 dynamical symplectic variables,

$$\Lambda^{ij}, \quad \bar{g}^{-1/3}\bar{g}_{ij}, \quad \Psi, \quad \bar{g}^{1/3}, \quad \operatorname{tr}\pi, \quad B\overline{R}. \tag{3.37}$$

These quantities are kinematically independent.

Let us observe that if $\overline{R} \neq -A/B$ then the set (3.37) is in a one-to-one correspondence with the following system of variables:

$$\bar{g}_{ij}, \quad z_{ij}, \quad \overline{R}, \quad \overline{\nabla}_0 \overline{R}. \tag{3.38}$$

In the $(3+1)$ picture we write

$$\begin{aligned}(\overline{Eq})^{\alpha\beta}\bar{g}_{\alpha\beta} = & -3B\sqrt{\bar{g}}[-\overline{\nabla}_0\overline{\nabla}_0\overline{R} + (\overline{\nabla}^p + \overline{\nabla}^p \ln N)\overline{\nabla}_p\overline{R} \\ & -\bar{g}^{pq}\overline{\Gamma}_p{}^0{}_q\overline{\nabla}_0\overline{R}\,] + A\sqrt{\bar{g}}\,\overline{R} = 0, \end{aligned} \tag{3.32'}$$

$$\begin{aligned}&(\overline{Eq})^{pq} - \tfrac{1}{3}\bar{g}^{pq}(\overline{Eq})^{\alpha\beta}\bar{g}_{\alpha\beta} \\ &= -\overline{\nabla}_0[(A + B\overline{R})\,{}^t\pi^{pq}] + B\sqrt{\bar{g}}\overline{\nabla}^p\overline{\nabla}^q\overline{R} \\ &\quad -\tfrac{1}{3}B\sqrt{\bar{g}}\bar{g}^{pq}\overline{\nabla}_0\overline{R}\,\overline{\Gamma}_v{}^v{}_0 + \tfrac{2}{3}(A + B\overline{R})\sqrt{\bar{g}}\,\overline{\Gamma}^{p0q}\overline{\Gamma}_v{}^v{}_0 \\ &\quad -(A + B\overline{R})\big[\sqrt{\bar{g}}\,{}^3\overline{R}^{\,pq} + 2\sqrt{\bar{g}}\overline{\Gamma}^{p0u}\overline{\Gamma}^{q0}{}_u \\ &\quad -\sqrt{\bar{g}}(\overline{\nabla}^p + \overline{\nabla}^p \ln N)\overline{\nabla}^q \ln N\,\big] \\ &\quad +\tfrac{1}{6}(A + B\overline{R})\bar{g}^{pq}\big[-2\sqrt{\bar{g}}(\overline{\nabla}_s + \overline{\nabla}_s \ln N)\overline{\nabla}^s \ln N \\ &\quad +\sqrt{\bar{g}}\,{}^3\overline{\mathrm{R}} + \sqrt{\bar{g}}(-(\overline{\Gamma}_s{}^s{}_0)^2 + \overline{\Gamma}_u{}^0{}_v\overline{\Gamma}^{u0v})\big] \\ &\quad +\tfrac{1}{12}B\sqrt{\bar{g}}\bar{g}^{pq}\overline{R}^{\,2} = 0. \end{aligned} \tag{3.33'}$$

Taking into account relations (A22) we see that Eq. (3.32′) is a fourth-order dynamical equation for $\bar{g}$ and the system (3.33′) gives us five second-order dynamical equations for the metric density $\bar{g}^{-1/3}\bar{g}_{ij}$.

The last case that should be considered is the Einstein theory, i.e., $C = B = 0$, $A \neq 0$. Then

$$\begin{aligned}\Xi^{ij} &= -\tfrac{1}{2}\sqrt{\bar{g}}(\bar{g}^{ir}\overline{\Gamma}_r{}^j{}_0 + \bar{g}^{jr}\overline{\Gamma}_r{}^i{}_0), \\ \zeta^{ij} &= \overline{P}^{0i}{}_0{}^j = \sqrt{\bar{g}}\bar{g}_{ij}. \end{aligned} \tag{3.39}$$

After some algebraic-differential transformations the symplectic integrand in (2.28) reads

$$\delta_1\Xi^{ij}\wedge\delta_2\bar{g}_{ij} + \delta_1\zeta^{ij}\wedge\delta_2 z_{ij} = \delta_1\pi^{ij}\wedge\delta_2\bar{g}_{ij}. \tag{3.40}$$

This result was already known to ADM[15] (see Ref. 22 for a geometric derivation). For the Einstein theory Eqs. (2.30) govern the dynamics of π^{ij}. When expressed in terms of $\bar{g}_{ij}$ they form a second-order dynamical system for components of the metric on slices. We do not discuss this system because such an analysis has already been presented in the literature.[15,19,21–23]

IV. INDEPENDENT DEGREES OF FREEDOM IN GRAVITATIONAL THEORIES WITH QUADRATIC LAGRANGIANS

The gravitational theories discussed in the previous section are theories with constraints. For all the cases considered in Sec. III the diffeomorphism group of space-time is contained in the full gauge group and ADM's lapse and shift N, N^k are the diffeomorphism gauge variables.

It is obvious that in theories with constraints we have reduction of the number of independent degrees of freedom. q symplectic constraints reduce q primary (kinematically independent) geometric degrees of freedom. Moreover, the symplectic analysis shows us that the m-parameter action of a gauge group reduces m further degrees of freedom. More precisely, such an action induces a distribution W of vectors tangent to the subvariety C of symplectic variables satisfying constraints. W is the gauge distribution of the symplectic two-form Ω and the quotient space TC/W represents directions of independent degrees of freedom. Therefore if p is the number of kinematically independent (geometric) symplectic variables, then we have

$$f = p - (q + m)$$

independent degrees of freedom in the phase space.

For the Einstein theory this problem was profoundly investigated in Refs. 21–24. The symplectic approach advocated in those papers can be considered as an alternative to the Dirac theory of constrained Hamiltonian systems.[9] In our opinion, for theories of gravity with quadratic Lagrangians the symplectic approach seems to be more effective than Dirac's original one. It can be concluded from the results of the present paper as well as from the analysis of SO(3,1)-gauge theories of gravity.[7,8,10]

Let us briefly summarize the results of Sec. III. We have five different cases of quadratic gravitational Lagrangians.

(I) Canonically regular cases $C \neq 0$, $C \neq -6B$. We have 24 dynamical variables,

$$\Xi^{ij}, \quad \bar{g}_{ij}, \quad \zeta^{ij}, \quad z_{ij},$$

whose initial values are subject to four Hamiltonian constraints (2.31). The full gauge group is Diff M:

$$p = 24, \quad q = 4, \quad m = 4, \quad f = 16.$$

(II) $C \neq = 0$, $C = -6B$, $A \neq 0$. We have 22 kinematically independent dynamical variables [cf. (3.22)],

$$\Upsilon^{ij}, \quad \bar{g}^{-1/3}\bar{g}_{ij}, \quad \bar{g}^{-1/3}\,{}^t\zeta^{ij},$$

$$\bar{g}^{1/3}\,{}^t z_{ij}, \quad \bar{g}^{1/3}, \quad \bar{g}^{1/3}\operatorname{tr} z,$$

$$p = 22, \quad q = 4, \quad m = 4, \quad f = 14.$$

(III) $C \neq = 0, C = -6B, A = 0$. We have 21 kinematically independent dynamical variables,

$$\Upsilon^{ij}, \quad \bar{g}^{-1/3}\bar{g}_{ij}, \quad \bar{g}^{-1/3}\,{}^t\zeta^{ij}, \quad \bar{g}^{1/3}\,{}^t z_{ij}, \quad \bar{g}^{1/3}$$

and an additional one-parameter gauge transformation $g_{\mu\nu} \to \tau g_{\mu\nu}$ with the corresponding gauge variable tr z,

$$p = 21, \quad q = 4, \quad m = 5, \quad f = 12.$$

(IV) $R^2 + R$ − gravity, $C = 0, B \neq 0$. We have 14 independent dynamical variables [cf. (3.37)].

$$\Lambda^{ij}, \quad \bar{g}^{-1/3}\bar{g}_{ij}, \quad \Psi, \quad \bar{g}^{1/3}, \quad \operatorname{tr}\pi, \quad B\overline{R},$$

$$p = 14, \quad q = 4, \quad m = 4, \quad f = 6.$$

(V) Einstein gravity $C = 0$, $B = 0$, $A \neq 0$. We have 12 kinematically independent dynamical variables,

$$\pi^{ij}, \quad \bar{g}_{ij}, \quad p = 12, \quad q = 4, \quad m = 4, \quad f = 4.$$

V. BOUNDARY INTEGRALS IN THE HAMILTONIAN FORMULATION AND THE DEFINITION OF ENERGY

The functional Hamiltonian equation (2.24) has been obtained from simple geometric principles. The energy-momentum vector density E^λ is the Noether current for the action of the diffeomorphism group. The symplectic two-form Ω is given by a natural expression built from the field potentials $g_{\mu\nu}$ and their momenta $P^{\lambda\tau\mu\nu}$. Let us observe, however, that the elegant dynamical formulation as described in Sec. II is relevant only for spatially closed spacetimes. In order to get the equivalence of the Euler–Lagrange equations with the Hamilton equation (2.24) we have to neglect the following boundary integrals over two-dimensional surfaces: (i) in formula (2.16), which defines the energy-momentum function E; (ii) in formula (2.17), which gives us the differential of E; and (iii) in formula (2.25), when we pass to the $(3+1)$ representation of the symplectic two-form Ω.

Now we are going to discuss these boundary terms more thoroughly. It is clear that a successful discussion can be carried on only if specific asymptotic falloffs of field variables at the spatial infinity are assumed. Then we can preserve the Hamiltonian form of the dynamics with a modified energy-momentum function.

The idea is to analyze the asymptotic behavior of boundary integrals, discard those of them that vanish asymptotically when the two-dimensional boundaries of compact domains on the initial surface σ tend to infinity, and to reformulate the remaining ones as contributions to the energy function. This is a classical approach due to ADM,[15] De Witt,[19] and Regge–Teitelboim.[20] In order to give a precise definition of asymptotic conditions we assume that the initial surface σ is decomposed into a compact set K and its complement CK that is diffeomorphic to the complement of a contractible compact set in R^3 (see Parker–Taubes[16] for a more general case). Thus, outside K we have a Cartesian coordinate system (x^k) on σ and the limit at spatial infinity is well defined.

We assume the following asymptotic conditions:

$$\begin{aligned} &\bar{g}_{ij}=\delta_{ij}+O(1/r),\quad \bar{\Gamma}_i{}^0{}_j=O(1/r^2),\\ &N=1+O(1/r),\quad N^k=O(1/r),\\ &\partial_0N=O(1/r^2),\quad \partial_0N^k=O(1/r^2), \end{aligned}\tag{5.1}$$

where $r=(\Sigma_{i=1}^3(x^i)^2)^{1/2}$ is the Euclidean radius. First of all, we assume that

$$Z^0=1,\quad Z^k=0\tag{5.2}$$

and start the discussion of the boundary integral in (2.16). It is given by the following formula [cf. (2.12)]:

$$-\int_{\partial\sigma}\sum_{\tau<\lambda}(-1)^{\tau+\lambda}B^{\tau\lambda}\,dx^0\wedge\cdots_{\hat{\tau}}\cdots_{\hat{\lambda}}\cdots\wedge dx^3.$$

Taking into account that $x^0=\text{const}$ on σ we get

$$\int_{\partial\sigma}\sum_{k=1}^3(-1)^{k+1}\bar{B}^{0k}\,dx^1\wedge\cdots_{\hat{k}}\cdots\wedge dx^3.\tag{5.3}$$

From (2.13) we get

$$\begin{aligned}\bar{B}^{0k}=2(&-\overline{\nabla_\omega P^{0k\omega0}}N+\overline{\nabla_\omega P^{0k\omega s}}N_s)\\ &+\bar{P}^{0k0a}(-2\partial_aN-\bar{\partial}_0N_a-1/N\partial_aN^sN_s).\end{aligned}\tag{5.4}$$

It follows from (A15), (3.8), and (3.10) that

$$\overline{\nabla_\omega P^{0k\omega0}}=O(1/r^4),\quad \overline{\nabla_\omega P^{0k\omega s}}=O(1/r^4).\tag{5.5}$$

Taking into account (5.5) we see that the only $O(1/r^2)$ asymptotic term in (5.4) is

$$\bar{B}^{0k}_\infty=A(2\partial^kN+\partial_0N^k).\tag{5.6}$$

The integral of $\bar{B}^{0k}_\infty$ over the two-dimensional boundary of σ depends on the asymptotic values of ∂^kN and ∂_0N^k on $\partial\sigma$. Such a contribution destroys the Hamiltonian form of the field equations because the coefficients at variations δN and δN^k in $dE\cdot V$ have to be proportional to the Hamiltonian constraints (2.31) and this property is ensured by other terms. We conclude that the divergence term in (2.12) must be discarded *a priori*. Therefore we define

$$E=-\int_\sigma\sum_{\lambda=0}^3(-1)^\lambda 2(Eq)^{\tau\lambda}Z_\tau\,dx^0\wedge\cdots_{\hat{\lambda}}\cdots\wedge dx^3.\tag{5.7}$$

Computing the differential of E we obtain the integral of the divergence $\partial_\tau A^{\tau\lambda}$ as well as the symplectic two-form term. The integral of the divergence will be analyzed later. Now we observe that the symplectic term while transformed to the $(3+1)$ picture gives rise to the following surface integral [cf. (2.25)]

$$\begin{aligned}\int_{\partial\sigma}\sum_{k=1}^3(-1)^k[&\delta_Y(1/N\bar{P}^{0k}{}_{ou})\wedge\delta_VN^u\\ &-\delta_Y\bar{P}^{0ikj}\wedge\delta_V\bar{g}_{ij}]dx^1\wedge\cdots_{\hat{k}}\cdots\wedge dx^3.\end{aligned}\tag{5.8}$$

Here δ_Y are variations corresponding to the vector of evolution Y and δ_V are variations corresponding to a sample vector V. We know from Refs. 6 and 22 that

$$\begin{aligned}\delta_YN&=N(\bar{\nabla}_0\bar{Z}^0+\bar{Z}^k\partial_k\ln N)\\ &=N(\bar{\partial}_0N+N^k\partial_k\ln N)=\partial_0N,\\ \delta_YN^k&=N(\bar{\nabla}_0\bar{Z}^k+\bar{Z}^0\partial^k\ln N-\partial^k\bar{Z}^0)\\ &=N(\bar{\partial}_0N^k+1/NN^p\partial_pN^k)=\partial_0N^k,\\ \delta_Y\bar{g}_{ij}&=(\bar{\nabla}_i\bar{Z}_j+\bar{\nabla}_j\bar{Z}_i)+2/\sqrt{\bar{g}}(\pi_{ij}-\tfrac{1}{2}\bar{g}_{ij}\operatorname{tr}\pi)\bar{Z}^0\\ &=(\bar{\nabla}_iN_j+\bar{\nabla}_jN_i)+2N/\sqrt{\bar{g}}(\pi_{ij}-\tfrac{1}{2}\bar{g}_{ij}\operatorname{tr}\pi).\end{aligned}\tag{5.9}$$

We see that all these quantities have $O(1/r^2)$ asymptotic behavior. It follows from (3.8) and (3.9) that

$$\bar{P}^{0p0q}=-A\delta^{pq}+O(1/r),\quad \bar{P}^{p0ij}=O(1/r^3).\tag{5.10}$$

The corresponding variations of these quantities by no means decrease slower and the integrand in (5.8) has $O(1/r^3)$ asymptotic behavior. Therefore in the limit $r\to\infty$ the integral (5.8) vanishes. The boundary integral in (2.17) reads

$$-\int_{\partial\sigma}\sum_{k=1}^3(-1)^{k+1}\bar{A}^{0k}\,dx^1\wedge\cdots_{\hat{k}}\cdots\wedge dx^3.\tag{5.11}$$

It follows from (3.5), (5.2), and (B1) that only the term

$$\bar{A}^{0k}_1=N(-\bar{P}^{k\lambda\mu\nu}\,\overline{\nabla_\mu\delta g_{\nu\lambda}}-\bar{P}^{0k\omega\lambda}\,\overline{\nabla_\omega\delta g_{0\lambda}})\tag{5.12}$$

may have $O(1/r^2)$ asymptotics. Other terms fall off more rapidly. Expanding (5.12) and taking into account (5.10) we observe that the unique term with $O(1/r^2)$ behavior is

$$\bar{A}^{0k}_2=-N\bar{P}^{kamb}\bar{\nabla}_m\delta\bar{g}_{ab}.\tag{5.13}$$

Finally, making use of (3.5) we get the following formula:

$$\bar{A}^{0k}_\infty=NA\sqrt{\bar{g}}(\bar{g}^{kb}\bar{g}^{ma}-\bar{g}^{km}\bar{g}^{ab})\bar{\nabla}_m\delta\bar{g}_{ab}.\tag{5.14}$$

If we substitute (5.14) into (5.11) and change the sign [the energy term should be on the left-hand side of (2.17)], then we get

$$E_\infty = \int_{\partial\sigma_\infty} \sum_{k=1}^{3} (-1)^{k+1} A(\delta^{kb}\delta^{ma} - \delta^{km}\delta^{ab})\partial_m \bar{g}_{ab}\, dx^1 \wedge \cdots \hat{k} \cdots \wedge dx^3, \quad (5.15)$$

where $\partial\sigma_\infty$ is a two-dimensional surface representing the boundary of σ at infinity.

The formula (5.15) coincides with the classical ADM expression for the energy of asymptotically flat gravitational fields. We observe that only the Einstein part of Lagrangian (3.1) contributes to the final result. This fact was earlier observed by Boulware[12] and Strominger.[14]

ACKNOWLEDGMENTS

The author would like to thank Professor V. Moncrief and Professor L. Smolin for their interest in this work, stimulating discussion, and critical remarks.

This work was supported by the NSF Grant No. PHY 8503072 to the Yale University.

APPENDIX A: THE (3+1) DECOMPOSITION OF GEOMETRIC OBJECTS

Let $(\sigma)_{t\in R}$ be a slicing of space-time M in a family of three-dimensional surfaces. Let (x^μ) be local coordinates consistent with the slicing, that is,

$$\sigma_t = \{x \in M : x^0 = t\}.$$

We consider only such Lorentz metrics $g_{\mu\nu}$ on M with signature $(-+++)$ that all the surfaces σ_t are spacelike for these metrics. Let

$$N = (-g^{00})^{-1/2}, \quad N_k = g_{0k} \quad (A1)$$

be ADM's lapse and shift. For any tensor density $F^{\alpha_1 \cdots \alpha_k}{}_{\beta_1 \cdots \beta_s}$ of weight r on M we define its bar components

$$\bar{F}^{\alpha_1 \cdots \alpha_k}{}_{\beta_1 \cdots \beta_s} = N^{-r} A^{\bar{\alpha}_1}{}_{\mu_1} \cdots A^{\bar{\alpha}_k}{}_{\mu_k} A^{-1\nu_1}{}_{\bar{\beta}_1} \cdots A^{-1\nu_s}{}_{\bar{\beta}_s} F^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_s}, \quad (A2)$$

where

$$A^{\bar{0}}{}_0 = N, \quad A^{\bar{k}}{}_0 = N^k, \quad A^{\bar{\mu}}{}_s = \delta^\mu{}_s,$$
$$A^{-1 0}{}_{\bar{0}} = 1/N, \quad A^{-1k}{}_{\bar{0}} = -N^k/N, \quad A^{-1\nu}{}_{\bar{s}} = \delta^\nu{}_s. \quad (A3)$$

In particular for the metric tensor we have

$$\bar{g}_{00} = -1, \quad \bar{g}_{0k} = 0, \quad \bar{g}_{ij} = g_{ij},$$
$$\bar{g}^{00} = -1, \quad \bar{g}^{0k} = 0, \quad \bar{g}^{ij}\bar{g}_{kj} = \delta^i{}_k, \quad (A4)$$
$$g = \det[g_{\mu\nu}] = -N^2\bar{g} = -N^2\det[\bar{g}_{ij}],$$
$$N^k = \bar{g}^{ks}N_s. \quad (A5)$$

For connection coefficients we define

$$\bar{\Gamma}_\mu{}^\lambda{}_\nu = A^{\bar{\lambda}}{}_\tau A^{-1\alpha}{}_{\bar{\mu}} A^{-1\beta}{}_{\bar{\nu}} \Gamma_\alpha{}^\tau{}_\beta + A^{\bar{\lambda}}{}_\tau \bar{\partial}_\mu A^{-1\tau}{}_{\bar{\nu}}, \quad (A6)$$

where

$$\bar{\partial}_\mu = A^{-1\tau}{}_{\bar{\mu}}\partial_\tau. \quad (A7)$$

For the Riemannian connection $\Gamma_\mu{}^\lambda{}_\nu = \{{}_\mu{}^\lambda{}_\nu\}$ we have

$$\begin{aligned} &\bar{\Gamma}_0{}^0{}_0 = 0, && \bar{\Gamma}_k{}^0{}_0 = 0 \\ &\bar{\Gamma}_0{}^p{}_0 = \bar{\nabla}^p \ln N, && \bar{\Gamma}_k{}^s{}_0 = N\bar{g}^{su}\Gamma_k{}^0{}_u \\ &\bar{\Gamma}_0{}^0{}_k = \partial_k \ln N, && \bar{\Gamma}_k{}^0{}_s = N\Gamma_k{}^0{}_s, \\ &\bar{\Gamma}_0{}^s{}_k = N\bar{g}^{su}\Gamma_k{}^0{}_u + 1/N\,\partial_k N^s, && \bar{\Gamma}_k{}^r{}_s = \Gamma_k{}^r{}_s + N^r\Gamma_k{}^0{}_s. \end{aligned} \quad (A8)$$

Here $\bar{\Gamma}_k{}^r{}_s$ are the Christoffels of $\bar{g}_{ij}$; $\bar{\Gamma}_k{}^0{}_s$ represent the second fundamental form of the embedding $\sigma \to M$.[6,25] For a tensor density $\bar{F}^a{}_b$ of weight r on σ we define covariant derivatives

$$\bar{\nabla}_k \bar{F}^a{}_b = \bar{\partial}_k \bar{F}^a{}_b - r\bar{\Gamma}_k{}^p{}_p \bar{F}^a{}_b + \bar{\Gamma}_k{}^a{}_p \bar{F}^p{}_b - \bar{\Gamma}_k{}^q{}_b \bar{F}^a{}_q, \quad (A9a)$$

$$\bar{\nabla}_0 \bar{F}^a{}_b = \bar{\partial}_0 \bar{F}^a{}_b - r\cdot\sigma^u{}_u \bar{F}^a{}_b + \sigma^a{}_p \bar{F}^p{}_b - \sigma^q{}_b \bar{F}^a{}_q, \quad (A9b)$$

where

$$\sigma^q{}_p = (1/N)\partial_p N^q \quad (A10)$$

are $\bar{\partial}_0$-connection coefficients. Formula (A10) assures proper transformation properties of (A9) under reparametrizations of the slicing

$$x^0 \to x^{0'}(x^0), \quad x^k \to x^{k'}(x^0, x^s). \quad (A11)$$

It was observed in Ref. 7 that the lapse N is an x^0 density of weight 1 under transformations (A11). Nonetheless $\partial_s \ln N$ is a covector field on σ. The shift N^k is not a vector field on slices. Its transformation rules are explained in Ref. 7.

The covariant derivative of a tensor density (of weight r) $F^\alpha{}_\beta$ is a tensor density of valence (1,2). Therefore $\overline{\nabla_\lambda F^\alpha{}_\beta}$ is defined by means of (A2). It turns out, however, that the following very useful formula holds[6]:

$$\overline{\nabla_\lambda F^\alpha{}_\beta} = \bar{\partial}_\lambda \bar{F}^\alpha{}_\beta - r\bar{\Gamma}_\lambda{}^\tau{}_\tau \bar{F}^\alpha{}_\beta + \bar{\Gamma}_\lambda{}^\alpha{}_\epsilon \bar{F}^\epsilon{}_\beta - \bar{\Gamma}_\lambda{}^\sigma{}_\beta \bar{F}^\alpha{}_\sigma. \quad (A12)$$

Analogous formulas hold for tensor densities of any valence and they enable us to express the left-hand side of (A12) by three-covariant derivatives. Some examples are the following:

$$0 = \overline{\nabla_0 g_{ij}} = \bar{\nabla}_0 \bar{g}_{ij} - 2\bar{\Gamma}_i{}^0{}_j, \text{ i.e., } \bar{\nabla}_0 \bar{g}_{ij} = 2\bar{\Gamma}_i{}^0{}_j; \quad (A13)$$

$$\overline{\nabla_\omega P^{0a\omega b}} = \bar{\nabla}_0 \bar{P}^{0a0b} + (\bar{\nabla}_r + \partial_r \ln N)\bar{P}^{0arb} + \bar{\Gamma}_s{}^a{}_0 \bar{P}^{0s0b} + \partial_r \ln N \bar{P}^{ra0b} + \bar{\Gamma}_r{}^0{}_s \bar{P}^{sarb};$$

$$\overline{\nabla_\omega P^{0a\omega 0}} = \bar{\nabla}_r \bar{P}^{0ar0} + \bar{\Gamma}_r{}^0{}_s \bar{P}^{sar0}; \quad (A14)$$

$$\begin{aligned} \overline{\nabla_0 R^{bn}} &= \bar{\nabla}_0 \bar{R}^{bn} + \bar{\nabla}^b \ln N \bar{R}^{0n} + \bar{\nabla}^n \ln N \bar{R}^{b0} + \bar{\Gamma}_r{}^b{}_0 \bar{R}^{rn} + \bar{\Gamma}_r{}^n{}_0 \bar{R}^{br}; \\ \overline{\nabla_k R^{0n}} &= \bar{\nabla}_k \bar{R}^{0n} + \bar{\Gamma}_k{}^0{}_s(\bar{R}^{sn} + \bar{g}^{sn}\bar{R}^{00}); \end{aligned} \quad (A15)$$

$$\begin{aligned} \overline{\nabla^\tau \nabla_\tau R} &= -\bar{\nabla}_0 \bar{\nabla}_0 \bar{R} + (\bar{\nabla}^p + \bar{\nabla}^p \ln N)\bar{\nabla}_p \bar{R} - \bar{g}^{pq}\bar{\Gamma}_p{}^0{}_q \bar{\nabla}_0 \bar{R}; \\ \overline{\nabla_p \nabla_q R} &= \bar{\nabla}_p \bar{\nabla}_q \bar{R} - \bar{\Gamma}_p{}^0{}_q \bar{\nabla}_0 \bar{R}, \end{aligned} \quad (A16)$$

the commutation relations for $\bar{\nabla}_\lambda$.

Let $\bar{F}^m{}_n$ be a tensor density of weight r on σ, then

$$[\overline{\nabla}_i,\overline{\nabla}_j]\overline{F}^m{}_n = {}^3\overline{R}^m{}_{sij}\overline{F}^s{}_n - {}^3\overline{R}^u{}_{nij}\overline{F}^m{}_u, \tag{A17a}$$

$$[\overline{\nabla}_0,\overline{\nabla}_j]\overline{F}^m{}_n = \partial_j \ln N\overline{\nabla}_0\overline{F}^m{}_n - r\cdot\lambda_j{}^p{}_p\overline{F}^m{}_n + \lambda_j{}^m{}_p\overline{F}^p{}_n - \lambda_j{}^u{}_n\overline{F}^m{}_u, \tag{A17b}$$

where the Christoffel-like tensor $\lambda_j{}^u{}_p$ is given by

$$\lambda_j{}^u{}_p = \tfrac{1}{2}\overline{g}^{us}[(\overline{\nabla}_j + \partial_j \ln N)\overline{\nabla}_0\overline{g}_{ps} + (\overline{\nabla}_p + \partial_p \ln N)\overline{\nabla}_0\overline{g}_{js} - (\overline{\nabla}_s + \partial_s \ln N)\overline{\nabla}_0\overline{g}_{jp}], \tag{A18}$$

where $\overline{\nabla}_0\overline{g}_{pq}$ should be expressed by means of (A13). For the curvature tensor its bar components $\overline{R}^\alpha{}_{\beta\mu\nu}$ are defined by means of (A2). We have the following relations as well:

$$\overline{R}^\alpha{}_{\beta\mu\nu} = \overline{\partial}_\mu\overline{\Gamma}_\nu{}^\alpha{}_\beta - \overline{\partial}_\nu\overline{\Gamma}_\mu{}^\alpha{}_\beta + \overline{\Gamma}_\mu{}^\alpha{}_\epsilon\overline{\Gamma}_\nu{}^\epsilon{}_\beta - \overline{\Gamma}_\nu{}^\alpha{}_\epsilon\overline{\Gamma}_\mu{}^\epsilon{}_\beta - (\overline{\partial}_\mu A^{-1\epsilon}{}_{\overline{\nu}} - \overline{\partial}_\nu A^{-1\epsilon}{}_{\overline{\mu}})A^{\overline{\omega}}{}_\epsilon\overline{\Gamma}_\omega{}^\alpha{}_\beta. \tag{A19}$$

For the curvature tensor of a Riemannian connection we get from (A19) and (A8)

$$\overline{R}_{i0mn} = \overline{\nabla}_m\overline{\Gamma}_n{}^0{}_i - \overline{\nabla}_n\overline{\Gamma}_m{}^0{}_i,$$
$$\overline{R}_{ijmn} = {}^3\overline{R}_{ijmn} + (\overline{\Gamma}_m{}^0{}_i\overline{\Gamma}_n{}^0{}_j - \overline{\Gamma}_n{}^0{}_i\overline{\Gamma}_m{}^0{}_j), \tag{A20}$$
$$\overline{R}_{i0m0} = \overline{\nabla}_m\overline{\nabla}_i \ln N + \overline{\nabla}_i \ln N\overline{\nabla}_m \ln N - \overline{\nabla}_0\overline{\Gamma}_m{}^0{}_i + \overline{\Gamma}_i{}^0{}_a\overline{\Gamma}_m{}^a{}_0.$$

Here ${}^3\overline{R}_{ijmn}$ is the Riemann tensor of $\overline{g}_{ij}$. From (A20) we get

$$\sqrt{g}\,\overline{R}^{ij} = \sqrt{g}\,{}^3\overline{R}_{ij} + \overline{\nabla}_0(\sqrt{g}\overline{\Gamma}^{i0j}) + 2\sqrt{g}\overline{\Gamma}^{i0u}\overline{\Gamma}^{j0}{}_u - \sqrt{g}(\overline{\nabla}^j + \overline{\nabla}^j \ln N)\overline{\nabla}^i \ln N, \tag{A21}$$

$$\sqrt{g}\,\overline{R} = -2\sqrt{g}(\overline{\nabla}_s + \overline{\nabla}_s \ln N)\overline{\nabla}^s \ln N + 2\overline{\nabla}_0(\sqrt{g}\,\overline{\Gamma}_a{}^0{}_b\overline{g}^{ab}) + \sqrt{g}\,{}^3\overline{R} + \sqrt{g}(-(\overline{\Gamma}_s{}^s{}_0)^2 + \overline{\Gamma}_p{}^0{}_q\overline{\Gamma}^{p0q}). \tag{A22}$$

We also have

$$\sqrt{g}\,\overline{R}^{00} = \sqrt{g}(\overline{\nabla}_s + \overline{\nabla}_s \ln N)\overline{\nabla}^s \ln N - \overline{\nabla}_0(\sqrt{g}\overline{\Gamma}_s{}^s{}_0) + \sqrt{g}((\overline{\Gamma}_s{}^s{}_0)^2 - \overline{\Gamma}_r{}^0{}_s\overline{\Gamma}^{r0s}),$$
$$\overline{R}_{0k} = \overline{\nabla}_s\overline{\Gamma}_k{}^s{}_0 - \overline{\nabla}_k\overline{\Gamma}_s{}^s{}_0, \tag{A23}$$
$$\overline{\nabla}_0{}^3\overline{R}^i{}_{jpq} = (\overline{\nabla}_p + \partial_p \ln N)\lambda_q{}^i{}_j - (\overline{\nabla}_q + \partial_q \ln N)\lambda_p{}^i{}_j \quad [\text{cf. (A18)}]. \tag{A24}$$

In particular,

$$\overline{\nabla}_0{}^3\overline{R} = (\overline{\nabla}^u + \overline{\nabla}^u \ln N)(\overline{\nabla}^k + \overline{\nabla}^k \ln N)\overline{\nabla}_0\overline{g}_{uk} - (\overline{\nabla}^u + \overline{\nabla}^u \ln N)(\overline{\nabla}_u + \overline{\nabla}_u \ln N)(\overline{\nabla}_0\overline{g}_{ks}\overline{g}^{ks}) + {}^3\overline{R}_{pq}\overline{\nabla}_0\overline{g}^{pq}. \tag{A25}$$

APPENDIX B: TIME CONSERVATION OF THE SYMPLECTIC TWO-FORM

The skew-symmetric tensor density in (2.17) is given by the following formula:

$$A^{\tau\alpha} = (-P^{\alpha\lambda\mu\nu}Z^\tau + P^{\tau\lambda\mu\nu}Z^\alpha)\nabla_\mu\delta g_{\nu\lambda} - Z^\sigma P^{\tau\alpha\omega\xi}\nabla_\omega\delta g_{\sigma\xi} - 2\delta(\nabla_\omega P^{\alpha\tau\omega\mu})Z_\mu - Z^\sigma\nabla_\omega P^{\alpha\tau\omega\epsilon}\delta g_{\sigma\epsilon} + (\nabla_\omega P^{\alpha\nu\omega\lambda}Z^\tau - \nabla_\omega P^{\tau\nu\omega\lambda}Z^\alpha)\delta g_{\nu\lambda} + Z^\sigma(\nabla_\omega P^{\tau\omega\alpha\xi} - \nabla_\omega P^{\alpha\omega\tau\xi})\delta g_{\sigma\xi} - \delta P^{\tau\alpha\mu\nu}\nabla_\mu Z_\nu + P^{\tau\alpha\nu\sigma}\nabla_\sigma Z^\mu\delta g_{\nu\mu}. \tag{B1}$$

Direct calculations show that

$$\partial_\tau[\delta_1(\nabla^\omega P^{\tau(\mu}{}_\omega{}^{\nu)})\wedge\delta_2 g_{\mu\nu} + \delta_1(P^{\tau(\mu}{}_\lambda{}^{\nu)})\wedge\delta_2\Gamma_\mu{}^\lambda{}_\nu] = -\delta_1(Eq)^{\mu\nu}\wedge\delta_2 g_{\mu\nu}. \tag{B2}$$

It follows from (B2) that if vectors V_1, V_2 represented by tensor fields $\delta_1 g_{\mu\nu}$, $\delta_2 g_{\mu\nu}$ satisfy the linearized field equations

$$\delta(Eq)^{\mu\nu} = 0, \tag{B3}$$

then the integrand in (2.22) has vanishing divergence. Therefore the value of the integral (2.22) is independent of the choice of σ.

The variation of the Lagrangian gives rise to the following formula:

$$\delta\overline{L} = (\tfrac{1}{2}\overline{L}\overline{g}^{ab} - \overline{P}^{a\tau\omega\xi}\overline{R}^b{}_{\tau\omega\xi})\delta\overline{g}_{ab} + \tfrac{1}{2}\overline{P}^{\alpha\beta\mu\nu}\delta\overline{R}_{\alpha\beta\mu\nu}. \tag{B4}$$

Therefore

$$\delta(\overline{L} + 2\zeta^{ab}\overline{R}_{0a0b}) = (\tfrac{1}{2}\overline{L}\overline{g}^{ab} - \overline{P}^{a\tau\omega\xi}\overline{R}^b{}_{\tau\omega\xi})\delta\overline{g}_{ab} + 2\delta\zeta^{ab}\overline{R}_{0a0b} + 2\overline{P}^{0bmn}\delta\overline{R}_{0bmn} + \tfrac{1}{2}\overline{P}^{abmn}\delta\overline{R}_{abmn}. \tag{B5}$$

It follows from (A20) and (B5) that the linearized constraints (2.31a) can be expressed by variations of the symplectic variables (2.27).

APPENDIX C: CONFORMAL GAUGE TRANSFORM

If the scale transformation

$$g_{\mu\nu} \to = {}'g_{\mu\nu} = \tau g_{\mu\nu} \tag{C1}$$

is performed, where τ is an arbitrary positive function on M then

$${}'\Gamma_\lambda{}^\alpha{}_\beta = \Gamma_\lambda{}^\alpha{}_\beta + \tfrac{1}{2}(\delta^\alpha{}_\beta\partial_\lambda \ln \tau + \delta^\alpha{}_\lambda\partial_\beta \ln \tau - g_{\lambda\beta}\partial^\alpha \ln \tau); \tag{C2}$$

$${}'R_{\beta\nu} = R_{\beta\nu} - \nabla_\nu\nabla_\beta \ln \tau - \tfrac{1}{2}g_{\nu\beta}\nabla^\sigma\nabla_\sigma \ln \tau + \tfrac{1}{2}(\nabla_\beta \ln \tau\nabla_\nu \ln \tau - g_{\nu\beta}\nabla^\epsilon \ln \tau\nabla_\epsilon \ln \tau), \tag{C3}$$

$${}'R = (1/\tau)R + 1/\tau[-3\nabla^\sigma\nabla_\sigma \ln \tau - \tfrac{3}{2}\nabla^\sigma \ln \tau\nabla_\sigma \ln \tau].$$

Taking into account these relations we get for the Lagrangian (3.1) with $C = -6B$ and $A = 0$,

$${}'L = L + \tfrac{1}{2}B\,\partial_\mu[\sqrt{-g}(-3\nabla_\nu \ln \tau\nabla^\mu\nabla^\nu \ln \tau + 3\nabla^\mu \ln \tau\nabla_\sigma\nabla^\sigma \ln \tau + \tfrac{3}{2}\nabla^\mu \ln \tau\nabla^\nu \ln \tau\nabla_\nu \ln \tau)]. \tag{C4}$$

[1]P. Weiss, Proc. R. Soc. London **156**, 192 (1936); **169**, 102, 119 (1939).
[2]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[3]P. Dedecker, in *Colloque International de Géometrie Différentielle*, Strasbourg, 1953 (CNRS, Paris, 1953).
[4]P. L. García and A. Perez-Rendon, Commun. Math. Phys. **13**, 24 (1969); Arch. Ration. Mech. Appl. **43**, 101 (1971); K. Gawedzki, Rep. Math. Phys. **3**, 307 (1972); J. Kijowski, Commun. Math. Phys. **30**, 99 (1973); H. Goldschmidt and S. Sternberg, Ann. Inst. Fourier (Grenoble) **23**, 203 (1973); J. Kijowski and W. Szczyrba, Commun. Math. Phys. **46**, 183

(1976); W. Szczyrba, Ann. Polon. Math. **32**, 145 (1976); W. Kondracki, Rep. Math. Phys. **16**, 9 (1976); J. Kijowski and W. Tulczyjew, "Symplectic framework for field theories," in *Lecture Notes in Physics*, Vol. 107 (Springer, Berlin, 1979); V. Aldaya and J. A. Azcarraga, Rev. Nuovo Cimento **3** (10) (1980).
[5]H. A. Kastrup, Phys. Rep. **101**, 1 (1983).
[6]W. Szczyrba, J. Math. Phys. **22**, 1926 (1981); Phys. Rev. D **25**, 2548 (1982).
[7]W. Szczyrba, Ann. Phys. (NY) **158**, 320 (1984).
[8]M. Antonowicz and W. Szczyrba, Phys. Rev. D **31**, 3104 (1985); J. Math. Phys. **26**, 1711 (1985).
[9]P. A. M. Dirac, *Lectures on Quantum Mechanics* (Yeshiva University, New York, 1964); A. Hanson, T. Regge, and C. Teitelboim, *Constrained Hamiltonian Systems* (Academia Nazionale dei Lincei, Rome, 1976); K. Sundermeyer, "Constrained dynamics," in *Lecture Notes in Physics*, Vol. 169 (Springer, Berlin, 1982).
[10]W. Szczyrba, "Stephenson–Kilmister–Yang theory of gravity and its dynamics," Yale University preprint (1986).
[11]V. Aldaya and J. M. Azcarraga, J. Math. Phys. **19**, 1869 (1978); W. F. Shadwick, Lett. Math. Phys. **6**, 409 (1982); M. Francaviglia and D. Krupka, Ann. Inst. H. Poincaré **37**, 295 (1982); Commentaire de la Redaction, *ibid.* **42**, 213 (1985); P. Dedecker, C. R. Acad. Sci. Paris Ser. 1 **298**, 397 (1984); **299**, 363 (1984).
[12]D. G. Boulware, in *Quantum Theory of Gravity*, edited by S. M. Christensen (Hilger, Bristol, 1984).
[13]A. Fischer, J. Marsden, and V. Moncrief, Ann. Inst. H. Poincare, Sec. A **33**, 147 (1981); J. Arms, J. Marsden, and V. Moncrief, Ann. Phys. (NY) **144**, 81 (1982); J. Arms and I. Anderson, *ibid.* **167**, 354 (1986).
[14]A. Strominger, Phys. Rev. D **30**, 2257 (1984).
[15]R. Arnowitt, S. Deser, and C. Misner, in *Gravitation—an Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).
[16]R. Schoen and S. T. Yau, Commun. Math. Phys. **65**, 45 (1979); **79**, 45 (1981); E. Witten, *ibid.* **80**, 381 (1981); O. Reula, J. Math. Phys. **23**, 810 (1982); T. Parker and C. Taubes, Commun. Math. Phys. **84**, 223 (1982); G. Horowitz and P. Tod, *ibid.* **85**, 429 (1982).
[17]W. Szczyrba, in *Cosmology and Gravitation*, edited by P. Bergman and V. De Sabbata (Plenum, New York, 1980).
[18]F. W. Hehl, J. Nitsch, and P. von der Heyde, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980).
[19]B. De Witt, Phys. Rev. **113**, 1113 (1967).
[20]T. Regge and C. Teitelboim, Ann. Phys. (NY) **88**, 286 (1974).
[21]A. Fischer and J. Marsden, in *General Relativity—An Einstein Centenary Survey*, edited by S. Hawking and W. Israel (Cambridge U. P., Cambridge, 1979).
[22]W. Szczyrba, Commun. Math. Phys. **51**, 163 (1976).
[23]J. York, Jr., in *Sources of Gravitational Radiation, Proceedings of the Batelle Seattle Workshop*, edited by L. Smarr (Cambridge U. P., Cambridge, 1979).
[24]V. Moncrief, J. Math. Phys. **16**, 493 (1975).
[25]W. Szczyrba, Commun. Math. Phys. **60**, 215 (1978).

On the physical properties of a nonquadratic solution for the McVittie metric

Henning Knutsen
Stavanger College of Engineering, Box 2527, Ullandhaug, 4001 Stavanger, Norway

(Received 17 March 1986; accepted for publication 27 August 1986)

A particular class of McVittie's new nonquadratic solutions is examined with respect to its physical characteristics. It is found that center regularity is not compatible with negative pressure and density gradients. These solutions also have the strange geometric feature that the physical radius is a decreasing function of comoving radial coordinate. Models for nonstatic "gaseous" spheres (i.e., the density ρ vanishes at the outer surface of the perfect fluid sphere together with the pressure p) have been constructed. The global motion is studied, and it has been found that pulsations are not possible. It is shown, however, that the density and pressure gradients are both *negative*. The pressure and the density are thus *positive* within the boundary of these "gaseous" spheres, and it is also seen that the density is increasing for contracting models. For the layers close to the outer boundary, it is shown that the pressure is increasing when the sphere is contracting. The speed of sound is thus real, and it is further seen that this speed is less than the speed of light. It is found that there exist bouncing models where the rate of change of circumference as measured by an observer riding on the boundary of the "gaseous" sphere is zero just at the moment when the sphere starts its reexpansion.

I. INTRODUCTION

It is common opinion today that the problem of gravitational collapse of massive objects is very important. However, the exact nature of the collapse process is generally not understood, partially because we do not know which "equation of state" matter obeys during the last stages of gravitational collapse. Even worse, it is expected that general relativity itself breaks down and cannot describe the physics for these moments. Hence, we shall suppose that the reader is aware of the importance of having at hand some easily surveyable exact models related to these strange phenomena.

McVittie[1] has given a very important procedure for obtaining exact solutions of Einstein's field equations without having to specify a particular equation of state. McVittie demonstrated that for a special form of the metric and for certain symmetry conditions, the field equations break up into three differential equations that may be solved independently. McVittie dealt with two of these equations in his fundamental paper, but the last equation was only solved for the special case where the derivative of the unknown function y is a quadratic function of y. Later, McVittie[2] has given several new nonquadratic solutions.

However, McVittie's approach does not necessarily lead to models that are physically plausible. That is our main motivation for investigating in some detail the physical properties of one of McVittie's new nonquadratic solutions. In particular, we find that there must exist an irregularity at the origin at the fluid sphere if the pressure and the density have negative gradients. Closely connected to this result is the following strange geometric feature of the model; the physical radius is a decreasing function of the radial coordinate.

For McVittie's class of solutions the matter density does not generally vanish at the outer boundary of the perfect fluid sphere. However, the heavenly bodies are generally gaseous spheres where the density drops to zero at surface together with the pressure. We have previously[3–5] shown that for McVittie's class of quadratic solutions, oscillatory motion is forbidden for physically reasonable gaseous spheres.

The global motion for the subclass of gaseous models is also studied in this paper, and we show that pulsations are forbidden.

Previously we have also put forward[6–9] several different models for gaseous spheres and examined their physical properties. This kind of analysis is also carried out in this paper. Particularly, we have constructed models where the density and the pressure p are positive, and their respective gradients are negative. The density is seen to be increasing for contracting spheres, and it is decreasing for expanding models.

Moreover, we find that even more physical conditions are valid for the outermost layers of the nonstatic gaseous sphere: (i) the pressure is increasing for contracting spheres, and it is decreasing for expanding spheres, (ii) the adiabatic speed of sound is less than the speed of light, and (iii) the energy condition $\rho > 3p$ holds.

The total mass is also found to be negative, and we point out that there exist models where the rate of change of circumference as measured by an observer riding on top of the sphere is zero just at the moment when the sphere changes its motion from collapse and starts expanding.

II. THE BASIC EQUATIONS

McVittie's line element reads

$$ds^2 = y^2\, dt^2 - e^{\eta} S^2\, [dr^2 + f^2(d\theta^2 + \sin^2\theta\, d\varphi^2)]\,. \tag{1}$$

Here is S the so-called scale function and is a function of t alone, and f is a function of comoving radial coordinate r alone, and η and y are functions of a variable z defined by

$$e^z = Q/S\,, \tag{2}$$

 0022-2488/87/010159-06$02.50

where Q is another function of r alone.

For a perfect fluid the following equation is now obtained from $T_1^{\ 4} = 0$ ($T_\mu^{\ \nu}$ here denotes the energy-momentum tensor),

$$y = 1 - \tfrac{1}{2}\eta_z . \tag{3}$$

Here the suffix means differentiation with respect to z.

Assuming stress isotropy, Walker's[10] equation yields the following three differential equations:

$$\frac{Q''}{Q} - \frac{f'Q'}{fQ} = a\left(\frac{Q'}{Q}\right)^2, \tag{4}$$

$$\frac{f''}{f} + \frac{1-f'^2}{f^2} = b\left(\frac{Q'}{Q}\right)^2, \tag{5}$$

$$y_{zz} + (a-3+y)\,y_z + [a+b-2-(a-3)y-y^2]\,y = 0, \tag{6}$$

where the prime denotes differentiation with respect to r.

Moreover, McVittie[1] obtains the following expressions for the density ρ and the pressure p:

$$8\pi G\rho = 3\left(\frac{\dot S}{S}\right)^2 + \frac{e^{-\eta}}{S^2}\Big\{3\,\frac{1-f'^2}{f^2} - 6(1-y)\frac{f'Q'}{fQ} - [2b - 2y_z + (1-y)(2a-1-y)]\left(\frac{Q'}{Q}\right)^2\Big\}, \tag{7}$$

$$8\pi Gp = \frac{1}{y}\Big\{-2\frac{\ddot S}{S} - (3y-2)\left(\frac{\dot S}{S}\right)^2 - \frac{e^{-\eta}}{S^2}\Big[y\,\frac{1-f'^2}{f^2} + 2(y^2-y-y_z)\frac{f'Q'}{fQ} + (1-y)(y^2-y-2y_z)\left(\frac{Q'}{Q}\right)^2\Big]\Big\}, \tag{8}$$

where an overdot means differentiation with respect to time.

III. THE NONQUADRATIC SOLUTION

In this paper we shall examine one of McVittie's[2] new nonquadratic solutions of Eq. (6), i.e., we shall investigate his solution

$$y^2 = 9\beta^2\frac{(\cosh\beta z + 1)}{(\cosh\beta z - 1)(\cosh\beta z + 2)^2}, \tag{9}$$

$$e^\eta = e^{\epsilon + 2z}\left(\frac{\cosh\beta z - 1}{\cosh\beta z + 2}\right)^2, \tag{10}$$

where

$$a = 3, \tag{11}$$

$$b = -1-\beta^2, \tag{12}$$

and ϵ is an arbitrary integration constant.

Inserting into Eq. (6) we have in fact found that the only valid solution is

$$y = -3\beta\frac{1}{(X+2)}\sqrt{\frac{X+1}{X-1}}, \tag{13}$$

with

$$y_z = 3\beta^2\frac{(X+\frac{1}{2})^2 + \frac{3}{4}}{(X-1)(X+2)^2}. \tag{14}$$

Here we have written

$$X = \cosh\beta z . \tag{15}$$

IV. IRREGULARITY AT THE ORIGIN

The pressure gradient is most easily found from the equation which represents conservation of linear momentum, i.e., $T_1^{\ \mu}{}_{;\mu} = 0$ (a semicolon denotes covariant differentiation), which yields

$$p' = -(y_zQ'/yQ)(\rho + p). \tag{16}$$

The general expression for the density gradient ρ' for the McVittie metric has been given in a previous paper.[4] For our particular model we now obtain

$$8\pi G\rho' = \frac{10e^{-\eta}}{S^2}\frac{Q'}{Q}(y_z + y^2 + \beta^2)\left(\frac{f'Q'}{fQ} + \frac{Q'^2}{Q^2}\right). \tag{17}$$

From Eq. (14) we observe that y_z is positive.

Restricting the pressure and the density to be non-negative, and their respective gradients to be nonpositive, Eqs. (16) and (17) immediately yield

$$Q'/y > 0, \tag{18}$$

and

$$Q'(f'Q'/fQ + Q'^2/Q^2) < 0. \tag{19}$$

However, to have regularity at the origin it must be the case that the "physical radius"

$$R = \pm e^{\eta/2}Sf, \tag{20}$$

vanishes at the center,[11] i.e., $f(0) = 0$.

Hence, we have

$$f'/f > 0. \tag{21}$$

It is thus seen that $Q' > 0$ is not compatible with condition (19).

Differentiating Eq. (20) we now find

$$R' = R\left[\left(\frac{f'}{f} + \frac{Q'}{Q}\right) - y\frac{Q'}{Q}\right]. \tag{22}$$

Using conditions (18) and (19) with a negative Q' it is immediately seen that we have $R' < 0$. This is not, however, compatible with vanishing of the non-negative function R at the origin. Hence, it is the case that *positive pressure and density and negative gradients imply center irregularity.*

V. ON THE POSSIBILITY OF OSCILLATIONS

To fit the internal solution to an external vacuum Schwarzschild solution it is necessary and sufficient to put the pressure equal to zero at the boundary:

$$p_b \equiv 0. \tag{23}$$

Henceforth the suffix b denotes boundary values. We have previously[3] shown that this gives an ordinary first-order differential equation for $\dot S^2$. Solving that equation we obtain for the present case (after some calculations)

$$\dot S^2 = e^{-\epsilon}S^2\left(\frac{Y+2}{Y-1}\right)^3\Big[-\frac{A}{Y+2} + B\frac{(Y-1)}{(Y+2)^2}\sqrt{\frac{Y+1}{Y-1}} - C\frac{Y+1}{(Y+2)^3} + D\Big], \tag{24}$$

where

$$A = 3\left(\frac{f'^2-1}{f^2} + 2\frac{f'Q'}{fQ} + \frac{Q'^2}{Q^2}\right)_b, \tag{25}$$

$$B = 6\beta \left(\frac{f'Q'}{fQ} + \frac{Q'^2}{Q^2}\right)_b , \tag{26}$$

$$C = -9\beta^2 \left(\frac{Q'}{Q}\right)_b^2 . \tag{27}$$

Here D is an arbitrary integration constant, and we have written

$$Y = X_b . \tag{28}$$

Here we have also followed McVittie and Stabell[12] and without loss of generality put $Q_b = 1$. From Eq. (24) we may study the global motion of the perfect fluid sphere. Moreover, this equation also gives a consistency relation which must be fulfilled, i.e., $\dot{S}^2 \geqslant 0$.

From Eqs. (13), (16), and (17) it is seen that we should demand

$$\beta \left(\frac{f'Q'}{fQ} + \frac{Q'^2}{Q^2}\right) > 0 , \tag{29}$$

to obtain a physically meaningful model.

Hence, we take the constant B to be positive. Necessary oscillatory conditions are then that the function within square brackets on the right-hand side of Eq. (24), call it $V(Y)$, has at least two positive roots, and that $V(Y) > 0$ between these two roots.

It is thus immediately seen from Eq. (24) that pulsations are not possible for models with $A < 0$ and $D > 0$, since the fluid in that case never comes to rest.

However, the necessary oscillatory conditions *are* fulfilled if we choose A, B, C, on D such that we have $V(1) < 0$, $D = \lim_{Y \to \infty} V(Y) < 0$ and observe that we can find $Y > 1$ such that $V(Y) > 0$.

The first restriction is equivalent to

$$A > -\tfrac{4}{3} C + 3D , \tag{30}$$

and the last restriction may be written

$$A < \frac{Y+1}{(Y+2)^2} C + (Y+2)D + (Y+2) . \tag{31}$$

Hence, the following inequality must be fulfilled:

$$-\frac{2Y+1}{9(Y+2)} C - D < \frac{B}{Y+2} \sqrt{\frac{Y+1}{Y-1}} . \tag{32}$$

We thus choose the constants in the following way: first we fix $D < 0$, $B > 0$, and $C < 0$. Then we observe that for $Y \gtrsim 1$ condition (32) is always fulfilled. Thereafter A is easily chosen such that restrictions (30) and (31) are also fulfilled. We thus arrive at an oscillating model. We emphasize, however, that these conditions are not sufficient to have a *physically acceptable* pulsating model.

VI. GASEOUS SPHERES

From now on we restrict the sphere to be gaseous, i.e., we demand

$$p_b \equiv 0 . \tag{33}$$

From Eq. (7) the following differential equation is obtained for the scale function:

$$\dot{S}^2 = e^{-\epsilon} \left(\frac{Y+2}{Y-1}\right)^2 S^2 \left[\frac{1}{3} A + \frac{B}{Y+2} \sqrt{\frac{Y+1}{Y-1}} + \frac{1}{27} C \frac{2Y^2 + 14Y + 11}{(Y+2)^2}\right] . \tag{34}$$

Comparing Eqs. (24) and (34) it is seen that the nonstatic sphere is gaseous if and only if the arbitrary integration constant D takes the value

$$D = \frac{1}{3} A + \frac{2}{27} C = \left[\frac{f'^2 - 1}{f^2} + 2\frac{f'Q'}{fQ} + \left(1 - \frac{2}{3}\beta^2\right)\left(\frac{Q'}{Q}\right)^2\right]_b . \tag{35}$$

The junction condition $p_b \equiv 0$ is thus fulfilled for our gaseous model. This result is also an immediate consequence of the equation which represents conservation of energy, i.e., $T_4{}^{\mu}{}_{;\mu} = 0$, since this equation reads

$$\dot{\rho} = -3y(\dot{S}/S)(\rho + p) . \tag{36}$$

VII. NONEXISTENCE OF PULSATING GAS SPHERES

To prove that oscillatory motion is forbidden for these gaseous models, we investigate the function

$$W(Y) = \frac{1}{3} A + \frac{B}{Y+2} \sqrt{\frac{Y+1}{Y-1}} + \frac{1}{27} C \frac{2Y^2 + 14Y + 11}{(Y+2)^2} . \tag{37}$$

We first observe

$$\lim_{Y \to 1} W(Y) = \infty . \tag{38}$$

From Rolle's theorem it now follows that for pulsating spheres it must be the case that the equation

$$\frac{dW}{dY} = 0 \tag{39}$$

has at least two solutions larger than unity.

We are now going to prove, however, that Eq. (39) can have at most one solution.

Differentiating (37) we find that Eq. (39) is fulfilled if and only if

$$U(Y) = \frac{(Y^2 + Y + 1)(Y+2)}{(Y-1)^{5/2} \sqrt{Y-1}} = -\frac{2}{9} \frac{C}{B} > 0 . \tag{40}$$

Differentiating once more, we find, however,

$$\frac{dU}{dY} = (Y-1)^{-7/2} (Y+1)^{3/2} \times (-5Y^3 - 15Y^2 - 18Y - 7) . \tag{41}$$

Hence, $U(Y)$ is a strictly decreasing function of Y, and Eq. (40) thus can have at most one solution. We conclude that *oscillatory motion is not possible for these gaseous spheres*.

VIII. INTEGRATION OF THE ISOTROPY EQUATIONS, NEGATIVE DENSITY GRADIENT, AND POSITIVE DENSITY

The first of the two remaining isotropy equations (4) and (5) is immediately integrated to give

$$f = A_1 Q'/Q^3 . \tag{42}$$

Here A_1 is an arbitrary integration constant. To integrate the second equation we follow McVittie[1] and introduce a new radial coordinate q defined by

$$q = -\tfrac{1}{2} A_1 Q^{-2} , \tag{43}$$

which yields

$$\frac{dq}{dr} = f. \tag{44}$$

Equation (5) now reads

$$f_{qq} + \frac{1}{f^3} = \frac{-\beta^2 - 1}{4} \frac{f}{q^2} . \tag{45}$$

Equation (45) is solved by the double substitution[1]

$$q = e^w, \quad f = e^{w/2} v(w) . \tag{46}$$

This double substitution yields, when inserted into Eq. (45),

$$\frac{d^2 v}{dw^2} = -\frac{\beta^2}{4} v - \frac{1}{v^3} . \tag{47}$$

To avoid complications we put an arbitrary integration constant equal to zero, and the solution of Eq. (47) reads

$$f^2 = (2q/\beta) \sin(\beta \ln q) . \tag{48}$$

We also have

$$Q = \sqrt{q_b / q} . \tag{49}$$

We further find

$$\frac{1 - f'^2}{f^2} - 2\frac{f'Q'}{fQ} - \frac{Q'^2}{Q^2} = \frac{\beta^2 f^2}{4q^2} , \tag{50}$$

and

$$\frac{f'Q'}{fQ} + \left(\frac{Q'}{Q}\right)^2 = -\frac{1}{2q} \cos(\beta \ln q) . \tag{51}$$

We now obtain two different models.

(i) Models with the following characteristics:

$$Q' > 0, \quad f < 0, \quad y > 0, \quad \beta < 0, \quad \frac{f'Q'}{fQ} + \left(\frac{Q'}{Q}\right)^2 < 0 . \tag{52}$$

From Eq. (17) it is now seen that *the density gradient is negative, and the density is thus positive for these gaseous models.*

To have $f^2 > 0$, it is also seen that we must have

$$1 < q < e^{-\pi/2\beta} , \tag{53}$$

i.e., we must have

$$q_{\text{center}} < e^{-\pi/2\beta}, \quad q_b > 1 . \tag{54}$$

(ii) Models with the following characteristics:

$$Q' < 0, \quad f > 0, \quad y < 0, \quad \beta > 0, \quad \frac{f'Q'}{fQ} + \left(\frac{Q'}{Q}\right)^2 > 0 . \tag{55}$$

From Eq. (17) it is again seen that *the density gradient is negative, and the density is thus positive for these gaseous models.*

To have $f^2 > 0$, it is now seen that we must take

$$e^{\pi/2\beta} < q < e^{\pi/\beta} , \tag{56}$$

i.e., we must choose

$$q_b > e^{\pi/2\beta}, \quad q_{\text{center}} < e^{\pi/\beta} . \tag{57}$$

IX. POSITIVE PRESSURE AND NEGATIVE PRESSURE GRADIENT

For this gaseous model we find that the density is given by

$$8\pi G\, e^\epsilon \rho = H - H_b , \tag{58}$$

where

$$\begin{aligned} H = {} & \frac{3}{Q^2}\left(\frac{1 - f'^2}{f^2} - 2\frac{f'Q'}{fQ} - \frac{Q'^2}{Q^2}\right)\left(\frac{X+2}{X-1}\right)^2 \\ & - 18\frac{\beta}{Q^2}\left(\frac{f'Q'}{fQ} + \frac{Q'^2}{Q^2}\right)\frac{X+2}{(X-1)^2} \\ & \times \sqrt{\frac{X+1}{X-1}} + \frac{\beta^2}{Q^2}\frac{Q'^2}{Q^2}\frac{2X^2 + 14X + 11}{(X-1)^2} . \end{aligned} \tag{59}$$

Further, we find (after some calculations) that the pressure p is given by

$$\begin{aligned} & -8\pi G\, e^\epsilon \beta \sqrt{\frac{X+1}{X-1}}\,\frac{1}{X+2}\,p \\ & = \Bigg\{\left[\frac{1 - f'^2}{f^2} - 2\frac{f'Q'}{fQ} - \frac{Q'^2}{Q^2}\right]_b \left[2\frac{Y+2}{(Y-1)^2}\sqrt{\frac{Y+1}{Y-1}} - 3\frac{(Y+2)^2}{(X-2)\,(X-1)^2}\sqrt{\frac{X+1}{X-1}}\right] \\ & \quad + \frac{1}{Q^2}\left[\frac{1 - f'^2}{f^2} - 2\frac{f'Q'}{fQ} - \frac{Q'^2}{Q^2}\right]\frac{X+2}{(X-1)^2}\sqrt{\frac{X+1}{X-1}}\Bigg\} + 2\beta\left\{\left[\frac{f'Q'}{fQ} + \frac{Q'^2}{Q^2}\right]_b\right. \\ & \quad \times \left[-\frac{Y^2 + 7Y + 7}{(Y-1)^3} + 9\frac{Y+2}{(X+2)(Y-1)^2}\sqrt{\frac{(X+1)(Y+1)}{(X-1)(Y-1)}}\right] + \left.\frac{1}{Q^2}\left(\frac{f'Q'}{fQ} + \frac{Q'^2}{Q^2}\right)\frac{X^2 - 2X - 2}{(X-1)^3}\right\} \\ & \quad + \beta^2\left\{\left(\frac{Q'}{Q}\right)_b^2 \left[2\frac{Y+2}{(Y-1)^2}\sqrt{\frac{Y+1}{Y-1}} - \frac{2Y^2 + 14Y + 11}{(Y-1)^2(X+2)}\sqrt{\frac{X+1}{X-1}}\right]\right. \\ & \quad + \left.\frac{3}{Q^2}\left(\frac{Q'}{Q}\right)^2 \frac{2X+1}{(X-1)^2(X+1)}\sqrt{\frac{X+1}{X-1}}\right\} . \end{aligned} \tag{60}$$

From this equation it is easily seen that $p_b \equiv 0$.

Remembering Eq. (16) we now find that the pressure gradient may be written

$$8\pi G e^{\epsilon} p' = -2\frac{y_z Q'}{yQ}(X+2)\sqrt{\frac{X-1}{X+1}}\,(I-I_b), \tag{61}$$

where I reads

$$\frac{1}{Q^2}\left\{\left[\frac{1-f'^2}{f^2}-2\frac{f'Q'}{fQ}-\frac{Q'^2}{Q^2}\right]\frac{X+2}{(X-1)^2}\sqrt{\frac{X+1}{X-1}} -\beta\left[\frac{f'Q'}{fQ}+\frac{Q'^2}{Q^2}\right]\frac{X^2+7X+7}{(X-1)^3} +\beta^2\left(\frac{Q'}{Q}\right)^2\frac{X+2}{(X-1)^2}\sqrt{\frac{X+1}{X-1}}\right\}. \tag{62}$$

To be assured that the pressure gradient is negative, we like to prove $(I-I_b)<0$.

However, we find

$$\frac{\partial I}{\partial r}=\frac{5\beta^2 f}{4qq_b}\cos(\beta\ln q)\,\frac{(X+2)^2}{(X-1)^4}\sqrt{X^2-1}, \tag{63}$$

and from Eqs. (52)–(57) it is thus seen that we have

$$\frac{\partial I}{\partial r}<0. \tag{64}$$

Hence, we have proved that *the pressure gradient is negative, and the pressure is thus positive.*

X. THE LAYERS CLOSE TO THE BOUNDARY

We shall now consider the outer layers of these nonstatic "gaseous" spheres.

Differentiating Eq. (16) we find

$$p''_b = -(y_z Q'/yQ)_b\,\rho'_b. \tag{65}$$

The pressure is thus a convex function for the surface layers. Since we also have $p'_b=0$ and $\rho'_b<0$, it must also be the case that the energy condition $\rho>3p$ is fulfilled close to the boundary.

We have previously given the formula for the mass function.[8] Now we find that the total mass M is given by

$$M=-\frac{20}{3}\pi\rho'_b R_b^3\left(\frac{f'}{f}+\frac{Q'}{Q}\right)_b^{-1}. \tag{66}$$

It is thus seen that the total mass is negative.

XI. CHANGE OF DENSITY AND PRESSURE WITH RESPECT TO TIME

From Eq. (36) it is immediately seen that the density is increasing for contracting spheres, and it is decreasing when the sphere expands.

To show that we have $\dot p(r\approx r_b)<0$ for expanding models, it is enough to show that we have $\dot p''_b<0$. Differentiating Eq. (16) we obtain

$$\dot p''_b=\left[-\frac{\partial}{\partial t}\left(\frac{y_z}{y}\right)+3y_z\frac{\dot S}{S}\right]_b\left(\frac{Q'}{Q}\right)_b\rho'_b =\left(\frac{Q'}{Q}\right)_b\beta^2\frac{\dot S}{S}\,\frac{8Y^3+21\,Y^2+24\,Y+10}{(Y^2-1)\,(Y+2)^2}\,\rho'_b. \tag{67}$$

From this equation it is easily seen that we have $\dot p''_b<0$.

XII. SPEED OF SOUND VERSUS SPEED OF LIGHT

The adiabatic speed of sound V_s is given by[6]

$$V_s^2=\dot p/\dot\rho.$$

It is thus seen that this speed is real for the layers close to the boundary.

To show that this speed is less than the speed of light, i.e.,

$$(V_s^2)_b<1, \tag{68}$$

it is enough to prove that the following relation holds for expanding spheres:

$$\dot p'_b-\dot\rho'_b<0. \tag{69}$$

We obtain

$$(\dot p'-\dot\rho')_b=3y_b(\dot S/S)\rho'_b. \tag{70}$$

We also have

$$\dot R=\pm e^{\eta/2}fy\dot S. \tag{71}$$

Hence, we conclude that the adiabatic speed of sound is less than the speed of light for the surface layers.

XIII. RATE OF CHANGE OF CIRCUMFERENCE

McVittie[1] and Nariai[13] takes V_M where

$$V_M^2=e^{\eta}f^2\dot S^2 \tag{72}$$

to be the matter velocity. However, this "velocity" is in fact the change of $(1/2\pi)\times$circumference as measured by an observer riding in a shell of matter. Remembering Eq. (34) we obtain

$$\frac{d}{dt}(V_{M,b}^2)=-\beta\sqrt{Y^2-1}\,\frac{\dot S}{S}f_b^2\left\{-6\left[\frac{f'Q'}{fQ}+\frac{Q'^2}{Q^2}\right]_b \times\frac{(Y^2+Y+1)\sqrt{Y^2-1}}{(Y-1)^2\,(Y+2)^2(Y+1)} +2\beta^2\left(\frac{Q'}{Q}\right)_b^2\frac{Y-1}{(Y+2)^3}\right\}. \tag{73}$$

Taking $Y\approx1$, it is seen that there exists a certain time interval such that we have

$$\frac{d}{dt}(V_{M,b}^2)<0, \tag{74}$$

for expanding models.

Using Eq. (40) it found that the equation

$$\frac{d}{dt}(V_{M,b}^2)=0 \tag{75}$$

has a solution if and only if we have

$$\beta\left(\frac{f'Q'}{fQ}+\frac{Q'^2}{Q^2}\right)_b<\frac{1}{3}\beta^2\left(\frac{Q'}{Q}\right)_b^2. \tag{76}$$

However, remembering Eqs. (48)–(57) it is easily found that condition (76) is fulfilled.

Equation (75) may in fact be written

$$-3\cot(\beta\ln q_b)=\frac{\sqrt{Y^2-1}\,(Y-1)^2}{(Y+2)(Y^2+Y+1)}. \tag{77}$$

We also find that we have $\dot S=0$ if and only if

$$-3\cot(\beta\ln q_b)=\frac{54\sqrt{Y+1}\,(Y+2)}{\sqrt{Y-1}\,(5Y^2+26Y+23)}. \qquad (78)$$

From Eqs. (77) and (78) we now obtain the following equation:

$$5Y^5=43Y^4+310Y^3+482Y^2+389Y+239, \qquad (79)$$

which is seen to have a solution $Y>1$. Hence, it is possible to have models where $\dot{S}=0$ and $d/dt(V^2_{M,b})=0$ at the same moment.

XIV. CONCLUSION

A particular class of McVittie's nonquadratic solutions of one of the isotropy equations has been discussed. This class contains models for gaseous spheres where the density and the pressure are positive within the outer boundary of the nonstatic sphere, and their respective gradients are negative. But we have shown that there is a center irregularity, and this fact is connected with the total mass being negative and the physical radius being a decreasing function of radial coordinate. However, we certainly expect a singularity to develop during the last stages of gravitational collapse if general relativity is able to tell what will happen. As long as a satisfactory theory for quantum gravity does not exist, we will therefore not discard these models as being without astrophysical interest.

[1]G. C. McVittie, Ann. Inst. H. Poincaré **6**, 1 (1967).
[2]G. C. McVitte, Ann. Inst. H. Poincaré **40**, 235 (1984).
[3]H. Knutsen and R. Stabell, Ann. Inst. Henri Poincaré **31**, 339 (1979).
[4]H. Knutsen, Ann. Inst. H. Poincaré **39**, 101 (1983).
[5]H. Knutsen, Phys. Scr. **28**, 129 (1983).
[6]H. Knutsen, Gen. Relativ. Gravit. **16**, 777 (1984).
[7]H. Knutsen, Phys. Scr. **31**, 305 (1985).
[8]H. Knutsen, Gen. Relativ. Gravit. **17**, 1121 (1985).
[9]H. Knutsen, Phys. Scr. **32**, 568 (1985).
[10]A. G. Walker, Q. J. Math. **6**, 81 (1935).
[11]C. W. Misner and D. H. Sharp, Phys. Rev. **136**, B571 (1964).
[12]G. C. McVittie and R. Stabell, Ann. Inst. H. Poincaré **9**, 371 (1968).
[13]H. Nariai, Prog. Theor. Phys. **38**, 92 (1967).

Exact self-gravitating disks and rings: A solitonic approach

P. S. Letelier and S. R. Oliveira
Departamento de Física, Universidade de Brasília, 70.910 Brasília-D. F., Brazil

(Received 2 January 1986; accepted for publication 27 August 1986)

The Belinsky–Zakharov version of the inverse scattering method is used to generate a large class of solutions to the vacuum Einstein equations representing uniformly accelerating and rotating disks and rings. The solutions studied are generated from a simple class of static disks and rings that can be expressed in a simple form using suitable complex functions of the usual cylindrical coordinates.

I. INTRODUCTION

Solutions representing disklike configurations of matter have been extensively studied in the context of general relativity.[1–3] Less studied are the solutions representing ringlike configurations.[4–5] Much of what is known about the solutions to the Einstein equation describing rotating disks represents the study of approximate solutions using numerical techniques.[3]

The purpose of this paper is to use the inverse scattering method (ISM) to generate exact solutions to the vacuum Einstein equations representing disks and rings. The method used is the Belinsky–Zakharov[6] version of the ISM that we recently employed to produce a large class of metrics of physical interest.[7]

The application of the ISM requires the knowledge of static solutions which are simple enough in order to explicitly solve the system of differential equations[7] for the functions F, the key functions of the ISM for a Weyl solution.

In Sec. II we study a family of static disks and a family of static rings with the above-mentioned characteristics. The first two members of the family of disks represent solutions that are already known.[2] Both families can be generated from the knowledge of a single member belonging to any one of them and can be expressed in a simple form using complex functions of the usual cylindrical coordinates.

In Sec. III we present a summary of the one- and the two-soliton solutions generated using the ISM with a general Weyl seed solution. In Sec. IV we explicitly compute the functions F associated with different disks and rings, and study the one- and the two-soliton solutions. In Sec. V we discuss some of the found results, we especially point out that the interpretation of the studied solutions as representing true disks and rings may not be totally correct. Finally, in the Appendix we compute the metric function ν for one disk and two rings.

II. STATIC DISKS AND RINGS

The vacuum Einstein equations for the static axially symmetric space-time

$$ds^2 = e^{\sigma_0(r,z)}\,(dr^2+dz^2) + r^2\,e^{-\phi(r,z)}\,d\theta^2 - e^{\phi(r,z)}\,dt^2 \tag{2.1}$$

are equivalent to

$$\phi_{,rr} + \phi_{,r}/r + \phi_{,zz} = 0\,, \tag{2.2}$$

$$\sigma_0[\phi] = -\phi + \nu[\phi]\,, \tag{2.3}$$

$$\nu[\phi] \equiv \frac{1}{2}\int r\,[\,(\phi_{,r}^2 - \phi_{,z}^2)\,dr + 2\phi_{,r}\phi_{,z}\,dz\,]\,, \tag{2.4}$$

where $(\ )_{,r} \equiv \partial_r$ and $(\ )_{,z} \equiv \partial_t$.

A physical image of the solutions of (2.2) can be obtained considering that the metric function ϕ can be related to the Newtonian potential U by[8]

$$U = \phi/2\,. \tag{2.5}$$

A. Disks

The Newtonian potential for the exterior of a disk with axially symmetric matter density can be written as[9]

$$U = -\sum_{n=0}^{\infty} C_{2n}\,P_{2n}(\eta)q_{2n}(\xi)\,, \tag{2.6}$$

where C_{2n} are constants, and $P_{2n}(\eta)$ and $q_{2n}(\xi) = i^{2n+1}Q_{2n}(i\xi)$ are the usual Legendre polynomials and the Legendre functions of second kind, respectively. The variables η and ξ are the oblate ellipsoidal coordinates that are related to the cylindrical coordinates by

$$r^2 = a^2(1+\xi^2)\,(1-\eta^2)\,, \tag{2.7a}$$

$$z = a\xi\eta \quad (-1\leqslant\eta\leqslant 1,\quad 0\leqslant\xi<\infty)\,. \tag{2.7b}$$

The Newtonian matter density associated to (2.7) is[9]

$$\rho = S(r)\,\delta(z)\,, \tag{2.8a}$$

$$S(r) = \{2\pi a[1-(r/a)^2]^{1/2-1}\} \times \sum_{n=0}^{\infty}(2n+1)\,C_{2n}q_{2n}(0)P_{2n}([1-(r/a)^2]^{1/2})\,. \tag{2.8b}$$

We shall study the specialization obtained by choosing the constants C_{2n} as

$$C_{2n} = (-1)^{n+1}(4n+1)\,(2m+1)!\Gamma(m-n+\tfrac{1}{2}) \times[\,(2n+1)2^{2n+1}(2m-2n)!Q_{2n+1}(0) \times\Gamma(m+n+\tfrac{3}{2})\,]^{-1}\,(M/a)\,, \tag{2.9}$$

for $n\leqslant m$ and $C_{2n}=0$, and for $n>m$, where m is a fixed positive integer. With this particular choice of C_{2n}, the potential (2.6) defines a family of disks with Newtonian densities

$$S^{Dm} = \frac{(2m+1)M}{2\pi a^2}\left[1-\left(\frac{r}{a}\right)^2\right]^{m-1/2}, \quad r\leqslant a\,. \tag{2.10}$$

Note that the Newtonian mass of each disk is M. The first three members of the family have the associated functions ϕ,

 0022-2488/87/010165-06$02.50

$$\phi^{D_0} = -(2M/a)\cot^{-1}\xi, \tag{2.11}$$

$$\phi^{D_1} = -(2M/a)\{\cot^{-1}\xi + \tfrac{1}{4}[(3\xi^2+1)\cot^{-1}\xi - 3\xi](3\eta^2-1)\}, \tag{2.12}$$

$$\phi^{D_2} = -(2M/a)[\cot^{-1}\xi + (2\eta^2-1)K + (35\eta^4 - 30\eta^2+3)L], \tag{2.13a}$$

$$K \equiv \tfrac{5}{14}[(3\xi^2+1)\cot^{-1}\xi - 3\xi], \tag{2.13b}$$

$$L \equiv \tfrac{3}{448}[(35\xi^4+30\xi^2+3)\cot^{-1}\xi - 35\xi^2 + \tfrac{55}{3}\xi]. \tag{2.13c}$$

The associated Newtonian densities are

$$S^{D_0} = (M/2\pi a)(a^2-r^2)^{-1/2}, \tag{2.14}$$

$$S^{D_1} = (3M/2\pi a^3)(a^2-r^2)^{1/2}, \tag{2.15}$$

$$S^{D_2} = (5M/2\pi a^5)(a^2-r^2)^{3/2}. \tag{2.16}$$

The first member of the family—the monopole term—is the well-known "integrable" disk studied in electrostatics that has a singular charge (matter) density on the rim. The second member—monopole + quadrupole term—has a well-behaved matter density, maximum on the disk center, and zero on the rim.[10] The third and following members have also a density with the same characteristics as D_1.

In cylindrical coordinates, the metric potentials (2.12) and (2.13) can be written as[11]

$$\phi^{D_0} = -(2M/a)\,\mathrm{Im}\ln\mu, \tag{2.17}$$

$$\phi^{D_1} = -(3M/a^3)\,\mathrm{Im}[(a^2+z^2-r^2/2)\ln\mu + \tfrac{1}{2}(3z+ia)R], \tag{2.18}$$

$$\phi^{D_2} = -(15M/2a^5)\,\mathrm{Im}(A\ln\mu - BR - \tfrac{1}{4}CR^3), \tag{2.19}$$

where

$$A \equiv \tfrac{1}{2}(a^4 - r^4/8 - z^4 + z^2r^2) + (z^2 - r^2/2)(a^2+z^2-r^2/2), \tag{2.20a}$$

$$B \equiv \tfrac{1}{8}(ia-z)^3 + (ia-z)(3z^2 - r^2/16) - zr^2, \tag{2.20b}$$

$$C = ia + 25z/3, \tag{2.20c}$$

and

$$R \equiv +[(ia-z)^2 + r^2]^{1/2}, \tag{2.21}$$

$$\mu \equiv ia - z + R. \tag{2.22}$$

The relations

$$a\xi = \mathrm{Re}\,R, \tag{2.23a}$$

$$a\eta = \mathrm{Im}\,R \tag{2.23b}$$

are particularly useful in computing[11] (2.17)–(2.19).

The associated functions ν for the first disk of the family can be found in Ref. 1. This function can be expressed in a very simple way in terms of the complex function μ,

$$\nu[\phi^{D_0}] = -\frac{2M^2}{a^2}\ln\frac{r^2+|\mu|^2}{|r^2+\mu^2|} \tag{2.24}$$

(see the Appendix). The function $\nu[\phi^{D_1}]$ is computed in Ref. 2 (see also Refs. 12 and 13).

The disk D_n corresponds to the sum of the first 2^{2n}-pole terms in (2.6) with C_{2n} given by (2.9). We can pass to the next D_{n+1} disk by adding the corresponding 2^{2n+2}-pole term. The following algorithm involving a parametric integration does automatically the indicated operation:

$$\phi^{D_{n+1}} = \frac{2n+3}{a^{2n+3}}\int a^{2n+2}\phi^{D_n}\,da, \tag{2.25}$$

since from (2.11) we have

$$S^{D_{n+1}} = \frac{2n+3}{a^{2n+3}}\int a^{2n+2}S^{D_n}\,da. \tag{2.26}$$

B. Rings

The simplest solution of the Laplace equation in toroidal coordinates that represents a ring with radius a, i.e., it does not depend on the cylindrical angle θ, is[14]

$$U = -(M/\sqrt{2})(\cos\xi - \cos\eta)^{1/2}\cos(\eta/2), \tag{2.27}$$

where now

$$z + ir = a\cot[(\eta + i\xi)/2] \quad (0\leqslant\xi<\infty,\quad 0\leqslant\eta<2\pi). \tag{2.28}$$

The associated function ϕ in cylindrical coordinates reads[15]

$$\phi^{R_0} = -2M\,\mathrm{Re}(1/R). \tag{2.29}$$

The ring R_0 is related to the disk D_0 by

$$\phi^{R_0} = \frac{\partial}{\partial a}(a\phi^{D_0}), \tag{2.30a}$$

$$\phi^{D_0} = \frac{1}{a}\int\phi^{R_0}\,da. \tag{2.30b}$$

These relations tell us that the disk D_0 is formed by rings like R_0 of different radii. From R_0 we can also generate a family of rings as follows:

$$\phi^{R_{n+1}} = \frac{\partial}{\partial a}(a\phi^{R_n}). \tag{2.31}$$

The functions ϕ^{R_1} and ϕ^{R_2} are

$$\phi^{R_1} = -2M\,\mathrm{Re}\left[\frac{1}{R} + \frac{a(a+iz)}{R^3}\right], \tag{2.32}$$

$$\phi^{R_2} = -2M\,\mathrm{Re}\left[\frac{1}{R} + \frac{a(4a+3iz)}{R^3} + \frac{3a^2(a+iz)^2}{R^5}\right]. \tag{2.33}$$

The expression (2.29) coincides up to the quadrupole moment with the corresponding function ϕ for a uniform ring of radius[16] $b = a\sqrt{2}$. The associated function ν is in this case

$$\nu[\phi^{R_0}] = -2M^2r^2\left[\frac{|\mu|^2}{R^2}\frac{1}{(r^2+|\mu|^2)^2} + \mathrm{Re}\left(\frac{1}{4R^4}\right)\right]. \tag{2.34}$$

The functions $\nu[\phi^{R_1}]$ is given in the Appendix.

The functions ϕ^{D_0} and ϕ^{R_0} are equivalent to the functions ϕ associated with a bar of complex "length" $2ia$, i.e., with a complexified Weyl δ-metric,[17] and with two equal point masses located on the z axis at $z = \pm ia$, i.e., with a complexified two center Chazy–Curzon metric,[17] respectively.

Note that the Newtonian potentials associated with the family of disks of densities (2.10) as well as the family of rings can be obtained from the potential associated with any one of their members as Eqs. (2.5), (2.25), (2.30), and (2.31) indicate.

III. ONE- AND TWO-SOLITON SOLUTIONS

The vacuum Einstein equations for the stationary axially symmetric space-time

$$ds^2 = e^{\sigma(r,z)}\,(dr^2 + dz^2) + \gamma_{ab}(r,z)dx^a dx^b\,, \tag{3.1}$$

with $a,b = 3,4$, $(\theta,t) = (x^3,x^4)$, and $\det(\gamma_{ab}) = -r^2$, are equivalent to

$$(r\gamma_{,r}\gamma^{-1})_{,r} + (r\gamma_{,z}\gamma^{-1})_{,z} = 0\,, \tag{3.2}$$

$$\sigma = -\ln r - \frac{1}{4}\int r\,[\mathrm{tr}(\gamma_{,r}\gamma_{,r}^{-1} - \gamma_{,z}\gamma_{,z}^{-1})\,dr + 2\mathrm{tr}(\gamma_{,r}\gamma_{,z}^{-1})dz]\,, \tag{3.3}$$

where $\gamma \equiv (\gamma_{ab})$ and $\gamma_{,r}^{-1} \equiv (\gamma^{-1})_{,r}$, etc.

The one- and the two-soliton solutions of (3.2) are defined as those solutions obtained using the inverse scattering method with a "scattering matrix" with one and two single poles, respectively. When the initial solution (seed solution) is taken as the Weyl solution (2.2)–(2.4), we find for the one-soliton case[7,18]

$$\gamma_{33} = -\frac{p_1^2(r/\mu_1)Y_1^2 - q_1^2(\mu_1/r)Y_1^{-2}}{p_1^2Y_1^2 + q_1^2Y_1^{-2}}\,(r^2e^{-\phi})\,, \tag{3.4a}$$

$$\gamma_{34} = -\frac{p_1q_1r(r/\mu_1 + \mu_1/r)}{p_1^2Y_1^2 + q_1^2Y_1^{-2}}\,, \tag{3.4b}$$

$$\gamma_{44} = -\frac{p_1^2(\mu_1/r)Y_1^2 - q_1^2(r/\mu_1)Y_1^{-2}}{p_1^2Y_1^2 + q_1^2Y_1^{-2}}\,(-e^{\phi})\,, \tag{3.4c}$$

$$\sigma_1 = \sigma_0 + \ln\left\{r^{1/2}\frac{p_1^2Y_1^2 + q_1^2Y_1^{-2}}{R_1}\right\}, \tag{3.5}$$

where

$$Y_k \equiv (r/\mu_k)^{1/2}\exp(F_k - \phi/2)\,, \tag{3.6}$$

$$\mu_k \equiv \alpha_k - z + \epsilon_k R_k\,, \tag{3.7}$$

$$R_k \equiv +[(\alpha_k - z)^2 + r^2]^{1/2}\,. \tag{3.8}$$

The p_k, q_k, and α_k are arbitrary constants and $\epsilon_k = \pm 1$. The function $F = F(r,z;\lambda)$ is the solution of the system of differential equations

$$(r\partial_r - \lambda\partial_z + 2\lambda\partial_\lambda)F = r\phi_{,r}\,, \tag{3.9a}$$

$$(r\partial_z + \lambda\partial_r)F = r\phi_{,z}\,, \tag{3.9b}$$

$$F(r,z;0) = \phi\,. \tag{3.10}$$

The variable λ is a spectral parameter defined on the complex field, and

$$F_k \equiv F(r,z;\mu_k)\,. \tag{3.11}$$

Along the poles' trajectories, $\lambda = \mu_k$, the system of equations (3.9) and (3.10) admits the solution

$$\begin{aligned} F_k[\phi] &\equiv F[\phi;\mu_k]\,, \\ &= \frac{1}{2}\int \frac{r}{\mu_k}\,[(\mu_{k,r}\phi_{,r} - \mu_{k,z}\phi_{,z})dr \\ &\quad + (\mu_{k,r}\phi_{,z} + \mu_{k,z}\phi_{,r})dz]\,. \end{aligned} \tag{3.12}$$

And for the two-soliton case we get[7,19]

$$\gamma_{33} = \frac{[r(\mu_2-\mu_1)P_1]^2 - [(r^2+\mu_1\mu_2)P_2]^2}{[r(\mu_2-\mu_1)S_1]^2 + [(r^2+\mu_1\mu_2)S_2]^2}\,(r^2e^{-\phi})\,, \tag{3.13a}$$

$$\gamma_{34} = 4(\alpha_1 - \alpha_2)r\mu_1\mu_2 \times \frac{\epsilon_2R_2T_1 - \epsilon_1R_1T_2}{[r(\mu_2-\mu_1)S_1]^2 + [(r^2+\mu_1\mu_2)S_2]^2}\,, \tag{3.13b}$$

$$\gamma_{44} = \frac{[r(\mu_2-\mu_1)Q_1]^2 - [(r^2+\mu_1\mu_2)Q_2]^2}{[r(\mu_2-\mu_1)S_1]^2 - [(r^2+\mu_1\mu_2)S_1]^2}\,(-e^{\phi})\,, \tag{3.13c}$$

$$\sigma_2 = \sigma_0 + \ln\left\{\frac{[r(\mu_2-\mu_1)S_1]^2 + [(r^2+\mu_1\mu_2)S_2]^2}{\mu_1\mu_2R_1R_2}\right\}, \tag{3.14}$$

where

$$S_1 \equiv p_1p_2Y_1Y_2 - q_1q_2(Y_1Y_2)^{-1}\,, \tag{3.15a}$$

$$S_2 \equiv p_1q_2(Y_1/Y_2) - q_1p_2(Y_2/Y_1)\,, \tag{3.15b}$$

$$P_1 \equiv p_1p_2(r^2/\mu_1\mu_2)^{1/2}Y_1Y_2 + q_1q_2(\mu_1\mu_2/r^2)^{1/2}(Y_1Y_2)^{-1}\,, \tag{3.16a}$$

$$P_2 \equiv p_1q_2(\mu_2/\mu_1)^{1/2}(Y_1/Y_2) - q_1p_2(\mu_1/\mu_2)^{1/2}(Y_2/Y_1)\,, \tag{3.16b}$$

$$T_1 \equiv (p_1Y_1)^2 - (q_1Y_1^{-1})^2\,, \tag{3.17a}$$

$$T_2 \equiv (p_2Y_2)^2 - (q_2Y_2^{-1})^2\,, \tag{3.17b}$$

$$Q_1 \equiv p_1p_2(\mu_1\mu_2/r^2)^{1/2}\,Y_1Y_2 + q_1q_2(r^2/\mu_1\mu_2)^{1/2}\,(Y_1Y_2)^{-1}\,, \tag{3.18a}$$

$$Q_2 \equiv p_1q_2(\mu_1/\mu_2)^{1/2}\,(Y_1/Y_2) - q_1p_2(\mu_2/\mu_1)^{1/2}\,(Y_2/Y_1)\,. \tag{3.18b}$$

An important property of the two-soliton transform is that it maps diagonal asymptotically flat solutions into asymptotically flat solutions.[7]

Note that the soliton solutions associated to a given Weyl seed solution characterized by ϕ, i.e., a solution to the usual Laplace equation in cylindrical coordinates, are completely determined by the set of constants[20] α_k, p_k, and q_k, the known functions μ_k, and the function $F[\phi;\lambda]$ solution to the system of differential equations (3.9) and (3.10). Note that the function σ_0 is related to the seed solution that is assumed to be known.

IV. ACCELERATING AND ROTATING DISKS AND RINGS

In this section we first compute the functions F associated to the family of disks and rings presented in Sec. II. Then we study the corresponding one- and two-soliton solutions.

Using (3.9)–(3.12) one can demonstrate the following theorems.

Theorem 1: Let ϕ_1 and ϕ_2 be solutions of Eq. (2.3) and α an arbitrary constant, then

$$F[\phi_1+\phi_2;\lambda]=F[\phi_1;\lambda]+F[\phi_2;\lambda], \quad (4.1a)$$

$$F[\alpha\phi_1;\lambda]=\alpha F[\phi;\lambda]. \quad (4.1b)$$

Theorem 2: Let $\phi(a)\equiv\phi(r,z,a)$ be a solution of Eq. (2.9) depending on a constant parameter a, then

$$\frac{\partial}{\partial a}F[\phi(a);\lambda]=F\left[\frac{\partial\phi}{\partial a};\lambda\right], \quad (4.2a)$$

$$\int F[\phi(a);\lambda]da=F\left[\int\phi(a)da;\lambda\right]. \quad (4.2b)$$

From (2.25), (2.30), (2.31), (4.1), and (4.2) we get

$$F[\phi^{D_{n+1}};\lambda]=\frac{2n+3}{a^{2n+3}}\int a^{2n+2}F[\phi^{D_n};\lambda]\,da, \quad (4.3)$$

$$F[\phi^{R_0};\lambda]=\frac{\partial}{\partial a}\{aF[\phi^{D_0};\lambda]\}, \quad (4.4)$$

$$F[\phi^{R_{n+1}};\lambda]=\frac{\partial}{\partial a}\{aF[\phi^{R_n};\lambda]\}. \quad (4.5)$$

Thus, the class of functions F associated with both families of disks and rings can be obtained from the knowledge of the F associated with a single member of any one of the families.

A simple verification shows that[21]

$$F[\ln\mu;\lambda]=\ln(\mu-\lambda) \quad (4.6)$$

satisfies Eqs. (3.9) and (3.10). From (2.17), (4.1), and (4.5), for the first disk, we get

$$F[\phi^{D_0};\lambda]=-\frac{M}{ia}\ln\frac{\mu-\lambda}{\bar\mu-\lambda}, \quad (4.7)$$

where the bar denotes complex conjugation, and from (4.3) and (4.6), for the second, we have

$$\begin{aligned}F[\phi^{D_1};\lambda]&=-\frac{3M}{4ia^3}\Big\{2\Big(a^2+z^2-\frac{r^2}{2}+\lambda z+\frac{\lambda^2}{4}\Big)\ln\frac{\mu-\lambda}{\bar\mu-\lambda}\\&\quad+\tfrac14[\mu(\mu+8z+2\lambda)-\bar\mu(\bar\mu+8z+2\lambda)]\\&\quad-\frac{2zr^2}{\lambda}\ln\frac{1-\lambda/\mu}{1-\lambda/\bar\mu}\\&\quad+\frac{r^2}{2\lambda}\Big(\frac1\mu-\frac1{\bar\mu}+\frac1\lambda\ln\frac{1-\lambda/\mu}{1-\lambda/\bar\mu}\Big)\Big\}. \quad (4.8)\end{aligned}$$

The expressions

$$\lim_{\lambda\to0}\frac1\lambda\ln\Big(1-\frac{\lambda}{\mu}\Big)=-\frac1\mu, \quad (4.9a)$$

$$\lim_{\lambda\to0}\left|\frac{1}{\lambda\mu}+\frac{1}{\lambda^2}\ln\Big(1-\frac\lambda\mu\Big)\right|=-\frac{1}{2\mu^2}, \quad (4.9b)$$

tell us that the Eq. (3.10) is verified.

The functions F associated with the first two rings are[22]

$$F[\phi^{R_0};\lambda]=\phi^{R_0}-\lambda M\left[\frac{1}{R(\mu-\lambda)}+\frac{1}{\bar R(\bar\mu-\lambda)}\right], \quad (4.10)$$

$$\begin{aligned}F[\phi^{R_1};\lambda]=\phi^{R_1}-\lambda M&\left[\frac{a+iz}{R^3(\mu-\lambda)}+\frac{a-iz}{\bar R^3(\bar\mu-\lambda)}\right.\\&\left.-\frac{i\mu}{R^2(\mu-\lambda)^2}+\frac{i\bar\mu}{\bar R^2(\bar\mu-\lambda)^2}\right]. \quad (4.11)\end{aligned}$$

From (2.4), (4.1), (3.9), (3.6), (3.7), and (3.11) one can get the following relations:

$$\nu[\phi+\beta\ln r]=\nu[\phi]+\beta\phi+\tfrac12\beta^2\ln r, \quad (4.12)$$

$$F[\phi+\beta\ln r;\lambda]=F[\phi;\lambda]+(\beta/2)\ln(r^2-2\lambda z-\lambda^2), \quad (4.13)$$

$$Y_k[\phi+\beta\ln r]=(2\alpha_k\mu_k/r)^{\beta/2}Y_k[\phi], \quad (4.14)$$

that will be useful to study the one-soliton solutions.

Let us recall that the functions ϕ associated to the Newtonian potentials of (i) an infinite wire of density δ located on the z aixs, (ii) as semi-infinite wire of density δ laying on the z axis and located along $[\alpha_k,+\infty]$, and (iii) a similar semi-infinite wire located along $[-\infty,\alpha_k]$, are, respectively,

$$\phi=4\delta\ln r, \quad (4.15)$$

$$\phi=2\delta\ln\mu_k^+, \quad (4.16)$$

$$\phi=2\delta\ln\mu_k^-, \quad (4.17)$$

where we have introduced the notation

$$\mu_k^\pm=\mu_k|_{\epsilon_k=\pm1}. \quad (4.18)$$

The static limit of the one-soliton solution is obtained by putting either $p_1=0$ or $q_1=0$ in (3.4) and (3.5). We get

$$\ln\gamma_{44}=\phi+\ln(r/\mu_1)\quad(p_1=0) \quad (4.19)$$

$$\sigma_1=\sigma_0+\ln\Big(\frac{r^{1/2}Y_1^{-2}}{R_1}\Big)+2\ln q_1, \quad (4.20)$$

$$\ln\gamma_{44}=\phi+\ln(\mu_1/r)\quad(q_1=0) \quad (4.21)$$

$$\sigma_1=\sigma_0+\ln\Big(\frac{r^{1/2}Y_1^2}{R_1}\Big)+2\ln p_1, \quad (4.22)$$

and for both cases[23]

$$\gamma_{33}=-r^2/\gamma_{44}. \quad (4.23)$$

Thus, if we take ϕ as representing a disk or a ring we have that the one-soliton transform adds two semi-infinite wires along the z axis that have opposite densities starting from the point $z=\alpha_1$. To get a more physical solution we can eliminate one of the semi-infinite wires by adding to the seed solution the potential of an infinite wire of density $\delta=\frac14$. In this case we have

$$\ln\gamma'_{44}=\phi+\ln(r^2/\mu_1)\quad(p_1=0), \quad (4.24)$$

$$\ln\gamma'_{44}=\phi+\ln\mu_1\quad(q_1=0). \quad (4.25)$$

Note that by virtue of the identity

$$\mu_k^+\mu_k^-=-r^2, \quad (4.26)$$

both solutions (4.24) and (4.25) represent a disk or a ring depending on the meaning of ϕ and a semi-infinite wire. To be more precise let me choose $\alpha_1>0$ and $\epsilon_1=-1$, then (3.23) represents a disk or a ring and a semi-infinite wire along $[\alpha_1,+\infty]$. (See Fig. 1.) But, the Weyl solution corresponding to a semi-infinite wire along ($\phi=0$) with Newtonian density $\delta=\frac12$ represents a uniformly accelerated spacetime.[7] Then the one-soliton solutions constructed with the seed solutions

$$\phi^{D_m}+\ln r, \quad (4.27)$$

$$\phi^{R_m}+\ln r, \quad (4.28)$$

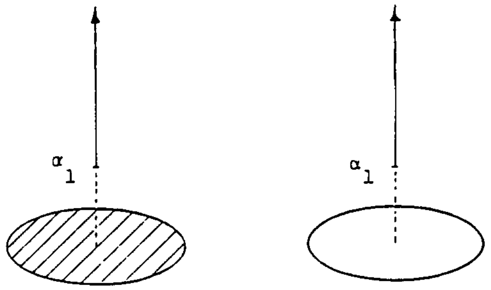

FIG. 1. Disk and ring with a semi-infinite wire.

represent a rotating and accelerating disk and ring, respectively.

Note that in this case

$$\sigma_0[\phi + \ln r] = \nu[\phi] + \tfrac{1}{2}\ln r, \tag{4.29}$$

$$Y_k\ [\phi + \ln r] = (2\alpha_k\mu_k/r)^{1/2}\ Y_k[\phi]. \tag{4.30}$$

The static solutions associated to the seed solutions (4.28) and (4.29) with $m=0$ are characterized by

$$\gamma_{44}^{D_0} = \mu_1(\mu/\bar{\mu})^{M/ia}, \tag{4.31}$$

$$\gamma_{44}^{R_0} = \mu_1 \exp[-M(1/R + 1/R)]. \tag{4.32}$$

Thus, the one-soliton solution constructed with (4.27) where $m=0$ can be considered as a rotating complexified accelerating Weyl δ-metric, i.e., to a complexified generalization of the Kinnersley metric.[17] And the one soliton constructed with (4.28) with $m=0$ corresponds to a rotating complexified Bonnor–Swaminarayan metric.[17]

The static limit of the two-soliton solution is obtained by putting $q_1 = q_2 = 0$ in (3.13)–(3.18). We find the Weyl solution characterized by

$$\ln\gamma_{44} = \ln(\mu_1\mu_2/r^2) + \phi, \tag{4.33}$$

$$\sigma_2 = \sigma_0 + \ln\left\{\frac{|r(\mu_2-\mu)S_1|^2}{\mu_1\mu_2R_1R_2}\right\}. \tag{4.34}$$

The other possible choices of the constants p_1, p_2, q_1, and q_2 that make $\gamma_{34}=0$ yield essentially the same solution (4.33) and (4.34) as a consequence of (4.26). Taking $\alpha_1 = -\alpha_2 = \alpha$ we have that (4.33) represents the superposition of a bar of length 2α, and either a disk or a ring depending on the meaning of ϕ, i.e., the superposition of a Schwarzschild solution[8,17] with mass equal to α located on the origin of the coordinates and either a disk or a ring depending whether ϕ is taken as ϕ^{Dm} or ϕ^{Rm}, respectively. (See Fig. 2.)

The full solution (3.13) and (3.14) with ϕ taken as either ϕ^{Dm} or ϕ^{Rm} represents the superposition of a Kerr metric with either a disk or a ring, respectively.[17] The Kerr metric is obtained in the limit[1,2] $\phi = F_k = 0$.

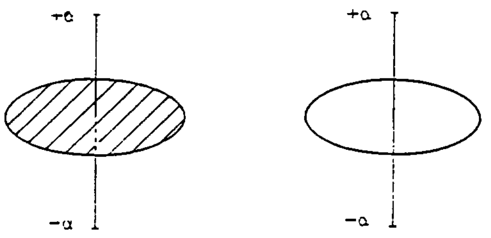

FIG. 2. Disk and ring with a bar of length 2α.

V. DISCUSSION

The expressions (4.33) and (2.17) suggest that the disk D_0 itself can be interpreted as a soliton solution since

$$\phi^{D_0} = -(M/ia)\ln(\mu^+\bar{\mu}^-/r^2), \tag{5.1}$$

where

$$\mu^{\pm} \equiv ia - z \pm R. \tag{5.2}$$

The factor M/ia can be thought as arising from the coalescence of M/ia two-soliton solution. Note that the appearance of such a factor suggests that (5.1) may be obtained from the inverse scattering method with a scattering matrix with a more general singular structure.[24]

In the present case, the utilization of complex functions together with the use of cylindrical coordinates simplified the computations to the point that the functions $\nu[\phi]$ and $F[\phi,\lambda]$ are found in an explicit way at least for ϕ^{D_0}, ϕ^{D_1}, ϕ^{R_0}, and ϕ^{R_1}. We want to point out that the generalization of the previous results to the case of d coaxial disks and r coaxial rings can also be computed easily in an explicit way, in the case that the rings and disks belong to the class D_0, R_0, and R_1. The most interesting case appears to be the case of r rings on a plane. Work along this line will soon be reported.

One can also consider the solution of Laplace equations in the context of the van Stockum solution and its solitonic generalizations.[25] The functions needed to solve the inverse scattering method are exactly the same functions $F[\phi,\lambda]$ considered in the present paper. Unhappily to get a physical interpretation or a physical image of the solutions generated is not an easy task.

Finally, we want to indicate that the interpretation of seed solutions (Weyl solutions) with $\phi = \phi^{D_m}$, ϕ^{R_m} as representing disks and rings is not without pitfalls since, as we indicated, the spherically symmetric Scharzschild solution is represented as a bar. By this reason we choose to call the interpretation of the solutions representing disks, rings, etc., a physical image rather than a physical interpretation. The study of the curvature singularities associated with the different solutions presented in Secs. II and III will be the subject matter of another paper.

ACKNOWLEDGMENT

The authors were partially supported by CNPq (a Brazilian government agency).

APPENDIX: THE COMPUTATION OF $\nu[\Phi]$

By using the notation of Sec. IV and the definitions (2.4) and

$$\nu[\phi_1,\phi_2] \equiv \frac{1}{2}\int r[(\phi_{1,r}\phi_{2,r} - \phi_{1,z}\phi_{2,z})dr + (\phi_{1,r}\phi_{2,z} + \phi_{1,z}\phi_{2,r})dz], \tag{A1}$$

it is not difficult to show the useful relations

$$\nu[\phi_1+\phi_2] = \nu[\phi_1] + \nu[\phi_2] + 2\nu[\phi_1,\phi_2], \tag{A2}$$

$$\nu[\alpha\phi] = \alpha^2\nu[\phi], \tag{A3}$$

$$\nu[\phi(a_1)] = \lim_{a_2\to a_1}\nu[\phi(a_2),\phi(a_1)], \tag{A4}$$

where a_1 and a_2 are parameters. Also

$$\nu[\phi_1,\phi_2] = \nu[\phi_2,\phi_1] , \tag{A5}$$

$$\bar{\nu}[\phi_1,\phi_2] = \nu[\bar{\phi}_2,\bar{\phi}_1] , \tag{A6}$$

$$\frac{\partial}{\partial a} \nu[\phi(a),\phi_1] = \nu\left[\frac{\partial\phi(a)}{\partial a},\phi_1\right] . \tag{A7}$$

From (A1) and (3.12) we have

$$\nu[\phi,\ln \mu_k] = F_k[\phi] . \tag{A8}$$

Now we shall compute $\nu[\phi^{D_0}]$, from (2.18) and (A1)–(A3). We get

$$\nu[\phi^{D_0}] = -(2M^2/a^2)(\operatorname{Re}\nu[\ln \mu] - \nu[\ln \mu,\ln \bar{\mu}]) . \tag{A9}$$

Thus, to obtain $\nu[\phi^{D_0}]$ we only need to compute $\nu[\ln \mu_1,\ln \mu_2]$ as (A4) indicates. From (4.6) and (A8) we have that

$$\nu[\ln \mu_1,\ln \mu_2] = \ln(\mu_1 - \mu_2) . \tag{A10}$$

To take the limit $\mu_1 \to \mu_2$, the identity

$$(r_2 + \mu_1\mu_2)(\mu_2 - \mu_1) = 2(\alpha_2 - \alpha_1)\mu_1\mu_2 , \tag{A11}$$

which is a direct consequence of the definition of μ_k, will be used. Thus

$$\nu[\ln \mu] = \lim_{\mu' \to \mu} \nu[\ln \mu,\ln \mu'] , \tag{A12}$$

$$= \ln[\mu^2/(\mu^2 + r^2)] . \tag{A13}$$

The expressions (A12) and (A13) are equal modulo a constant of integration that we set equal to zero, a practice that we shall follow in the sequel. From (A9)–(A12) we obtain (2.25).

For $\nu[\phi^{R_0}]$ we have

$$\nu[\phi^{R_0}] = 2M^2\left(\operatorname{Re}\nu\left[\frac{1}{R}\right] + \nu\left[\frac{1}{R},\frac{1}{\bar{R}}\right]\right) . \tag{A14}$$

Equations (A5), (A7), and (3.7) tell us that

$$\nu\left[\frac{1}{R_1},\frac{1}{R_2}\right] = \frac{\partial^2}{\partial\alpha_1\,\partial\alpha_2}\nu[\ln \mu_1, \ln \mu_2] , \tag{A15}$$

$$= -r^2\frac{\mu_1\mu_2}{R_1R_2}\frac{1}{(r^2 + \mu_1\mu_2)^2} . \tag{A16}$$

The expressions (A14)–(A16) give us Eq. (2.37).

The coefficient $\nu[\phi^{R_1}]$ can be written as

$$\nu[\phi^{R_1}] = \nu[\phi^{R_0}] - 2M^2a^2\left\{\nu\left[\frac{ia-z}{R^3},\frac{ia+z}{\bar{R}^3}\right] + \operatorname{Re}\nu\left[\frac{ia-z}{R^3}\right]\right\} - 4M^2a\left\{\operatorname{Im}\nu\left[\frac{1}{R},\frac{ia-z}{R^3}\right] + \operatorname{Im}\nu\left|\frac{1}{R},\frac{ia+z}{\bar{R}^3}\right|\right\} . \tag{A17}$$

By using (A16) and (A5)–(A7) we get

$$\nu\left[\frac{ia-z}{R^3},\frac{ia+z}{\bar{R}^3}\right] = \frac{r^2|\mu|^2|R - ia + z|^2}{|R|^6(r^2 + |\mu|^2)^2} - \frac{6r^4|\mu|^4}{|R|^4(r^2 + |\mu|^2)^4} + \frac{2r^2|\mu|^4}{|R|^4(r^2 + |\mu|^2)^3}\operatorname{Re}\left(\frac{ia-z}{R}\right) , \tag{A18}$$

$$\nu\left[\frac{ia-z}{R^3}\right] = -\frac{r^2(R - ia + z)^2}{4R^8} + \frac{3r^4}{8R^8} - \frac{r^2\mu(ia-z)}{4R^8} , \tag{A19}$$

$$\nu\left[\frac{1}{R},\frac{ia-z}{R^3}\right] = -\frac{r^2(R - ia + z)}{4R^6} + \frac{r^2\mu}{4R^6} , \tag{A20}$$

$$\nu\left[\frac{1}{R},\frac{ia+z}{\bar{R}^3}\right] = -\frac{r^2|\mu|^2(R - ia + z)}{R^2|R|^2(r^2 + |\mu|^2)^2} + \frac{2r^2|\mu|^4}{R|R|^2(r^2 + |\mu|^2)^3} . \tag{A21}$$

Note that the direct integration of (2.4) for ϕ^{R_1} is quite involved,[26] and that the knowledge of (A10) and (A5)–(A7) allows a direct computation of $\nu[\phi^{R_1}]$. The integral (A10)—the expression used to compute $\nu[\phi^{D_0}]$, $\nu[\phi^{R_0}]$, and $\nu[\phi^{R_1}]$—is a by-product of the inverse scattering method. In principle, the coefficients $\nu[\phi^{R_2}]$, $\nu[\phi^{R_3}]$,..., can be computed in a similar way.

[1]P. Y. Chou, Am. J. Math. **53**, 289 (1931).
[2]T. Morgan and L. Morgan, Phys. Rev. **183**, 1097 (1969).
[3]P. S. Wesson and A. Lermann, Astrophys. Space Sci. **57**, 203 (1978). This paper contains abundant references on general relativistic disks.
[4]R. Bach and H. Weyl, Math. Z. **13**, 134 (1922).
[5]K. S. Thorne, J. Math. Phys. **16**, 1860 (1975), and references therein.
[6]V. A. Belinsky and V. E. Zakarov Zh. Eksp. Teor. Fiz. **75**, 1955 (1978); **77**, 3 (1979) [Sov. Phys. JETP **48**, 985 (1978); **50**, 1 (1979)].
[7]P. S. Letelier, J. Math. Phys. **26**, 467 (1985).
[8]See, for instance, H. P. Robertson and T. W. Noonan, *Relativity and Cosmology* (Saunders, Philadelphia, 1969), Sec. 10.4.
[9]H. Bateman, *Partial Differential Equations of Mathematical Physics* (Cambridge U.P, Cambridge, 1932), p. 447.
[10]This particular case is studied in Ref. 2.
[11]A different approach to the Weyl solution using complex coordinates can be found in P. Szekeres, Phys. Rev. **176**, 1446 (1968).
[12]J. H. Young and C. A. Coulter, Phys. Rev. **184**, 1313 (1969).
[13]J. Winicour, A. Janis, and E. Newman, Phys. Rev. **176**, 1507 (1968).
[14]See, for instance, Ref. 9, p. 461.
[15]The interpretation of (2.30) as a ring can be found in P. Appell, Ann. Math. (Leipzig) **30**, 155 (1887), see also E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U.P, Cambridge, 1950), p. 400.
[16]N. Santos, Lett. Nuovo Cimento **14**, 327 (1975).
[17]See, for instance, Ref. 7.
[18]The components (3.4) are the physical components. Also, (3.5) differs from the corresponding equation of Ref. 7 in that we have corrected a misprint ($r^{1/2}$) and we have set a constant of integration equal to zero—a procedure that we shall follow without further warning.
[19]Note that the Eq. (3.13b) and (3.14) differ from the corresponding equation of Ref. 7 in that we have used the identity (A11).
[20]When p_k, q_k, and α_k are complex the expressions (3.13) and (3.14) can be cast in a more appealing form, see Ref. 7.
[21]P. S. Letelier, J. Math. Phys. **25**, 2675.
[22]P. S. Letelier, Rev. Bras. Fis. **14**, 371 (1984).
[23]Note that we can have $\gamma_{44} > 0$ and $\gamma_{33} < 0$. But, we can always add an integration constant $i\pi$ either to σ_1 or to σ_0 in order to have t as a time variable.
[24]The ISM with a scattering matrix with order two poles can be found in M. Francaviglia and C. Sgarra, Nuovo Cimento B **80**, 223 (1984).
[25]P. S. Letelier, J. Math. Phys. **27**, 564 (1986).
[26]See, for instance, B. Hoffman, *Les Théories Relativistes de la Gravitation* (CNRS, Paris, 1962).

The group-theoretical classification of the 11-dimensional classical homogeneous Kaluza–Klein cosmologies

M. Demiański
Institute of Theoretical Physics, Warsaw University, Hoża 69, 00-681 Warsaw, Poland
Z. Golda
N. Copernicus Astronomical Centre, Orla 171, 30-244 Krakow, Poland
L. M. Sokołowski and M. Szydłowski
Astronomical Observatory, Jagellonian University, Orla 171, 30-244 Krakow, Poland
P. Turkowski
Academy of Agriculture, Mickiewicza 21, 31-120 Krakow, Poland

(Received 17 January 1986; accepted for publication 3 September 1986)

In the context of the classical Kaluza–Klein cosmology the generalized Bianchi models in 11 dimensions are considered. These are space-times whose spacelike ten-dimensional sections are the hypersurfaces of transitivity for a ten-dimensional isometry group of the total space-time. Such a space-time is a trivial principal fiber bundle $P(M,G_7)$, where M is a four-dimensional physical space-time with an isometry group G_3 (of a Bianchi type) and G_7 is a compact isometry group of the compact internal space. The isometry group of P is $G_{10} = G_3 \otimes G_7$, hence all the generalized Bianchi models are classified by enumerating the relevant groups G_7. Due to the compactness of G_7 the result is astonishingly simple: there are three distinct homogeneous internal spaces in addition to the 11 ordinary Bianchi types for M.

I. INTRODUCTION

There is little doubt that irrespective of the present troubles of the supersymmetric quantum Kaluza–Klein theories, the very idea of generating the fundamental interactions by means of the dimensional reduction will remain as a significant achievement of theoretical physics; therefore studying the geometrical aspects of the higher-dimensional theories is well founded. Usually the classical Kaluza–Klein (KK) theory is understood as the "ground state" of the full quantum version of the KK theory. Initially it was naively assumed that the "ground state" of the theory is a product space $P = \mathfrak{M}_4 \times B$, P being the total $(4 + d)$-dimensional space-time, $\mathfrak{M}_4$ the Minkowski space, and B the microscopic (i.e., "internal") compact, d-dimensional Riemannian space of the internal degrees of freedom, the higher dimensions being treated as physically real. Such a definition is, however, incorrect for two reasons.

(a) Even in the absence of the fermionic matter and the Yang–Mills fields, the empty macroscopic space-time is curved in general.

(b) For this choice of the "ground state" (with the additional assumption that the metric g_{AB} of P, $A,B = 0,\ldots,3 + d$, is the direct sum of the macroscopic metric $g_{\mu\nu}(x^\alpha) = \eta_{\mu\nu}$, $\mu,\nu,\alpha = 0,\ldots,3$, and the microscopic metric $g_{ab}(x^c)$ on B, $a,b,c = 4,\ldots,d + 3$), the vacuum Einstein equations $R_{AB} = 0$ reduce to $R_{ab}(g_{cd}) = 0$. A theorem has for a long time been known to mathematicians: a compact, Riemannian, Ricci-flat manifold admits only covariantly constant (if any) Killing vector fields (cf. Appendix B). Thus such a space cannot induce non-Abelian gauge fields.

We therefore discard the assumption that the macroscopic space-time is flat, instead we assume that it is a certain Lorentzian four-dimensional manifold M. Every space-time $P = M \times B$ satisfying the vacuum Einstein equations will be referred to as a classical cosmological solution of the Kaluza–Klein theory.

It is interesting to study the cosmological (time-dependent) solutions for the KK theory, since one knows from the ordinary cosmology that our universe was much smaller in its early stages than it is today. This raises the question of whether the effective number of space-time dimensions has always been equal to 4. Indeed, the present four-dimensional stage of the universe could have been preceded by a higher-dimensional stage, which at "later times" becomes effectively four-dimensional in the sense that the microscopic dimensions become unobservably small due to dynamic contraction.

Higher-dimensional cosmologies have been studied in a number of papers (see, e.g., Refs. 1 and 2 and references therein), usually in the context of a supersymmetric KK theory. Most authors have assumed that the macroscopic space M is a Robertson–Walker space-time. There are, however, plausible arguments, both theoretical and observational,[3] that M is anisotropic spatially homogeneous, i.e., its metric structure is described by one of the Bianchi types (we do not take into account, for the time being, the Kantowski–Sachs models). Some authors[2,4] have also considered homogeneous models, in which M is of Bianchi type I and B is a homogeneous space with an Abelian simply transitive isometry group; without, however, attempting to answer the question of how large is that class in the whole set of the spatially homogeneous (in $d + 3$ dimensions) cosmological solutions. The purpose of the present paper is to give a gen-

 0022-2488/87/010171-03$02.50

eral classification of the 11-dimensional homogeneous cosmological space-times, or more exactly, of the so-called generalized Bianchi types, i.e., those space-times whose ten-dimensional spatial sections possess simply transitive isometry groups.

We restrict our analysis to the 11-dimensional space-times since this dimension is distinguished by the realistic supersymmetric versions of the KK theory demanding that the symmetry group of the internal space B be at least $SU(3) \otimes SU(2) \otimes U(1)$. It should be stressed, however, that our approach excludes the most interesting (for the supersymmetric theory) case when B is a "round" or "squashed" seven-sphere. The seven-sphere is a coset space rather than a group manifold, hence the spacelike hypersurfaces of transitivity of the form $G_3 \times S^7$, where G_3 is a three-dimensional group space, do not admit a subgroup of the full isometry group of the space-time acting simply transitively on the hypersurfaces. This exclusion is not a disadvantage of the generalized Bianchi cosmology since the supersymmetric theory assumes that there are (at least bosonic) matter fields in the macroscopic sector of the space-time (the "Freund–Rubin ansatz") and furthermore the SO(8) isometry group of S^7 does not contain as a subgroup the physically interesting phenomenological group $SU(3) \otimes SU(2) \otimes U(1)$.

Spatially homogeneous space-times of dimension less than or not much larger than 11 can be treated in a very similar way (see Appendix A).

II. GEOMETRIC FOUNDATIONS

The microscopic space of internal degrees of freedom is, by definition, a compact Riemannian space, hence the following theorem holds.

Theorem 1[5]: The group $I(B)$ of isometries of a Riemannian manifold B is a Lie transformation group with respect to the compact-open topology in B. If B is compact, $I(B)$ is also compact.

We assume that the isometry group $I(B) \equiv G_7$ acts simply transitively on B, hence there exists a distance-preserving diffeomorphism of G_7 onto B, $\dim G_7 = \dim B = 7$. The macroscopic space-time M admits a three-dimensional isometry group G_3 (of a Bianchi type) acting simply transitively on its spacelike sections.

A generic $(4+d)$-dimensional KK space-time has the structure of an associated or principal fiber bundle.[6] As we consider only simply transitive isometry groups we do not need to use coset spaces, therefore the generalized Bianchi space-times are the principal bundles. From the physical considerations one knows the base space M and the structure group G_7. To uniquely determine the bundle space P one additionally needs to know the transition functions connecting different trivializations of the bundle for a given open covering of M. Physics, however, does not determine these functions, hence it is quite sufficient to consider the simplest case, i.e., a trivial principal bundle $P(M,G_7)$ with $P = M \times G_7$.

To correctly perform the dimensional reduction yielding the effective four-dimensional field theory one applies the "Kaluza–Klein ansatz"[7] expressing the metric g_{AB} of P in terms of the metric tensors $g_{\mu\nu}$ on M and g_{ab} on B, the Killing vectors of B, and the gauge fields on the space-time M. The "ground state" of the KK theory is then defined as a principal fiber bundle with the vanishing gauge fields on M. It means that the metric tensor g_{AB} is block diagonal, or more exactly, the metric g_P on the bundle space is the direct sum of the metric tensors for M and B, $g_P = g_M \oplus g_B$. In other words, the bundle space P is a product space $M \times G_7$ not only in the topological but also in the metric sense.

III. THE ALGEBRAIC CLASSIFICATION OF THE 11-DIMENSIONAL BIANCHI MODELS

As usual one classifies the homogeneous spaces in terms of distinct Lie algebras for appropriate Lie groups. From the very notion of the trivial principal fiber bundle it follows that the ten-dimensional isometry group G_{10} of the bundle space P decomposes into the direct product $G_{10} = G_3 \otimes G_7$. This implies that the Lie algebra L_{10} of the Lie group G_{10} is the direct sum of the Lie algebras of the Lie groups G_3 and G_7, $L_{10} = L_3 \oplus L_7$. This is an important result allowing one to reduce the classification problem for ten-dimensional Lie algebras to that for seven-dimensional ones.

To enumerate all distinct real seven-dimensional Lie algebras is a hopelessly uphill task. Four-dimensional real Lie algebras were enumerated long ago (see Ref. 8 and references therein); five-dimensional real Lie algebras were classified by Mubarakzyanov,[9] who divided them into six nilpotent, 33 solvable, and a number of decomposable algebras. In six dimensions only nilpotent Lie algebras were studied. Fortunately, one needs not to enumerate all distinct real seven-dimensional Lie algebras, most of which are noncompact, for as is well known, the Lie algebra of a compact Lie group is also compact. For compact Lie algebras one applies the following theorem.

Theorem2[10]: A compact Lie algebra L is a direct sum $L = N \oplus S_1 \oplus \cdots \oplus S_n$, where N is the center of L and the S_i are simple algebras.

There are only three distinct real simple Lie algebras of dimension $\leqslant 7$ (for the classification method see Ref. 10); two of them are three dimensional—these are the Bianchi types $L_3(\mathrm{VIII})$ and $L_3(\mathrm{IX})$, the third one is six dimensional and, according to Ref. 10, is denoted by $\mathrm{sl}(2,\mathbb{C})^R$. Among these only the algebra $L_3(\mathrm{IX})$ is compact. One thus infers that there exist exactly three distinct real seven-dimensional compact Lie algebras, namely

$$\bigoplus_{i=1}^{7} L_1 \quad \text{(the Abelian algebra)},$$

$$\bigoplus_{i=1}^{4} L_1 \oplus L_3(\mathrm{IX}),$$

$$L_1 \oplus L_3(\mathrm{IX}) \oplus L_3(\mathrm{IX}).$$

One sees that the compactness condition is crucial for the classification problem—it implies that there are no more than three distinct homogeneous seven-dimensional internal spaces B. Each of them can be joined to any of the 11 Bianchi types for the macroscopic space-time M, hence one arrives at a surprisingly simple result: there are 33 generalized Bianchi types for 11 dimensional spatially homogeneous Kaluza–Klein space-times P.

Solutions for the vacuum Einstein field equations in the physically most interesting cases (the macroscopic space-time M is of Bianchi types I, V, and IX) and their interpretation will be presented in subsequent papers.

APPENDIX A: THE $d=6$ CASE

Although in the light of the preceding discussion it may seem rather trivial, we also give here, for the sake of completeness, the algebraic classification of the ten-dimensional generalized Bianchi space-times, as $D=10$ is one of the dimensions distinguished by the superstring theories. Applying Theorem 2 and the classification method for real simple Lie algebras of dimension $\leqslant 6$ one finds that there are three distinct real six-dimensional compact Lie algebras:

$$\bigoplus_{i=1}^{6} L_1 \quad \text{(the Abelian algebra)},$$

$$\bigoplus_{i=1}^{3} L_1 \oplus L_3(\mathrm{IX}),$$

$$L_3(\mathrm{IX}) \oplus L_3(\mathrm{IX}) \quad \text{(a semisimple Lie algebra)}.$$

APPENDIX B: RICCI-FLAT MANIFOLDS ADMIT ONLY COMMUTING KILLING VECTORS

Here we give the exact formulation and proof of the theorem mentioned in the Introduction. Its content is well known to the active researchers in the field but, to the best of our knowledge, its proof has never been published in the physical literature. The theorem traces back to Bochner.[11]

Theorem: Let B be a compact connected Riemannian manifold without boundary. Then (a) if the Ricci tensor field is negative definite everywhere on B, then the isometry group $I(B)$ is finite, i.e., every Killing vector field is zero; and (b) if B is Ricci flat, $R_{ab}=0$, then every Killing vector field is parallel, therefore any two Killing vectors commute.

The original proof is rather long. We give a short proof adapted from Ref. 12. One can assume that B is orientable, otherwise one has only to consider the orientable twofold covering space of B. Then for any vector field A on B one has, from the Gauss theorem,

$$\int_B \operatorname{div} A \, dV = 0.$$

Let K be any Killing vector field on B. One has for it

$$\operatorname{div}(\nabla_K K) \equiv \nabla_a (K^b \nabla_b K^a) = -K^{a;b}K_{a;b} + R_{ab}K^aK^b.$$

For a Riemannian metric one gets $-K^{a;b}K_{a;b} \leqslant 0$. Assuming $R_{ab}K^aK^b \leqslant 0$ one obtains $\operatorname{div}(\nabla_K K) \leqslant 0$. On the other hand, $\int_B \operatorname{div}(\nabla_K K)\,dV = 0$, hence

$$\operatorname{div}(\nabla_K K) = 0 = K^{a;b}K_{a;b} = R_{ab}K^aK^b.$$

(a) For the negative definite Ricci tensor, $R_{ab}K^aK^b = 0$ iff $K=0$ everywhere on B.

(b) For $R_{ab}=0$, $K^{a;b}K_{a;b}=0$ implies $K_{a;b}=0$—every Killing field is parallel on B. If K and L are any two Killing vector fields on B, then one easily checks by a short calculation that $[K, L] = 0$ on B.

[1] A. Chodos and S. Detweiler, Phys. Rev. D **21**, 2167 (1980); P. G. O. Freund, Nucl. Phys. B **209**, 146 (1982); S. Randjbar-Daemi, A. Salam, and J. Strathdee, Phys. Lett. B **135**, 388 (1984); Y. Okada, Tokyo preprint UT-448, February, 1985; D. Lorenz-Petzold, Phys. Lett. B **158**, 110 (1985); M. Gleiser, S. Rajpoot, and J. G. Taylor, Ann. Phys. (NY) **160**, 299 (1985).

[2] J. Demaret, J. Hanquin, M. Henneaux, and P. Spindel, Nucl. Phys. B **252**, 538 (1985).

[3] M. A. H. MacCallum, "The mathematics of anisotropic spatially homogeneous cosmologies," in *Physics of the Expanding Universe, Lecture Notes in Physics*, Vol. 109, edited by M. Demiański (Springer, Berlin, 1979).

[4] D. Lorenz-Petzold, Phys. Rev. D **31**, 929 (1985); Phys. Lett. B **149**, 79 (1984); **153**, 134 (1985); R. Bergamini and C. Orzalesi, *ibid.* **135**, 38 (1984).

[5] S. B. Myers and N. Steenrod, Ann. Math. **40**, 400 (1939).

[6] Y. M. Cho, J. Math. Phys. **16**, 2029 (1975); R. Percacci and S. Randjbar-Daemi, *ibid.* **24**, 807 (1983).

[7] M. J. Duff, B. Nilsson, C. Pope, and N. Warner, Phys. Lett. B **149**, 90 (1984).

[8] J. Patera, R. Sharp, and P. Winternitz, J. Math. Phys. **17**, 986 (1976); M. A. H. MacCallum, "On the classification of the real four-dimensional Lie algebras," Queen Mary College (London) preprint, undated.

[9] G. M. Mubarakzyanov, Izv. Vyssh. Uchebn. Zaved. Mat. **3**(34), 99 (1963).

[10] A. O. Barut and R. Rączka, *Theory of Group Representations and Applications* (Polish Scientific, Warszawa, 1980), 2nd ed.

[11] S. Bochner, Bull. Am. Math. Soc. **52**, 776 (1946).

[12] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.

Canonical commutation relations on the interval

G. Lassner and G. A. Lassner
Karl-Marx-Universitat Leipzig, Sektion Mathematik and Sektion Physik, 7010 Leipzig, German Democratic Republic
C. Trapani
Istituto di Fisica dell Universita di Palermo, I-90123 Palermo, Italy

(Received 11 July 1986; accepted for publication 10 September 1986)

On the Hilbert space $\mathscr{H} = \mathscr{L}^2(0,L)$, where $(0,L)$ is a bounded interval of R^1, the domain for the canonical commutation relation (CCR) and the CCR quasi-*-algebra is constructed. It is shown that the Bogolubov inequality for a Bose gas (in a box) is fulfilled.

I. INTRODUCTION

The basic ingredient in the algebraic approach to statistical physics is the set $\mathfrak{A}$ of observables of the system. This set $\mathfrak{A}$ is assumed to be a linear space. The dynamics is then described by a one-parameter family τ_t of linear transformations of $\mathfrak{A}$.

When one tries to perform the thermodynamical limit of the local Heisenberg dynamics some difficulties arise, namely this limit does not belong to $\mathfrak{A}$. This problem leads directly to consideration in the set of observables of the structure of a quasi-*-algebra.[2]

In this paper we consider the quasi-*-algebra generated by the operators P and Q (the usual position and momentum operators) acting in the Hilbert space $\mathscr{L}^2(Q,L)$, where $(0,L)$ is a bounded interval of R.[1]

In Secs. II and III we give the domain of definition for a canonical commutation relation (CCR) and construct the CCR quasi-*-algebra. This quasi-*-algebra $\mathscr{A}$ consists of elements of the form $\Sigma F_k P^k$, where the F_k's are distributions. Thus $\mathscr{A}$ may also contain very singular objects.

In Section IV we consider the question posed by the Bogolubov inequality. As is known, this inequality for a Bose gas leads to a contradiction. The reason for this lies in the fact that the operators involved are unbounded and the CCR is usally taken as $[Q,P] = i$. Applying the results of the previous sections we show that the boundary condition leads, in this case, to $[Q,P] = i - iL\delta(x)$. By means of this no contradiction arises from the Bogolubov inequality.

As a result, the quasi-*-algebras approach allows us not only to perform thermodynamical limits, but also helps us to overcome some other difficulties of the theory such as the one discussed above.

II. DOMAIN FOR CCR

Let us consider the operators $P = -i d/dx$ and $Q = x$, acting on the Hilbert space $\mathscr{H} = \mathscr{L}^2(0,L)$, where $(0,L)$ is a bounded interval of R^1. For technical simplicity we take 0 as the left boundary point.

As is known, the operator Q is bounded in $\mathscr{L}^2(0,L)$, whereas the operator P is unbounded and it is self-adjoint on the domain

$$D(P) = \{f \in \mathscr{L}^2(0,L)\colon f \text{ is absolutely continuous and } f(L) = \alpha f(0),\quad |\alpha| = 1\}.$$

We restrict ourself to the case $\alpha = 1$. The operator P has a simple spectrum with eigenvectors

$$\phi_n(x) = \phi_p(x) = 1/\sqrt{L}\, e^{ipx},\qquad p = 2\pi n/L,\quad n = 0, \pm 1, \pm 2,\ldots. \tag{2.1}$$

Let us now consider the following linear manifold of $\mathscr{H}$:

$$\mathscr{D} = \{\phi = \textstyle\sum \xi_n \phi_n \in \mathscr{H}\colon \|\phi\|_n < \infty\},$$

where

$$\|\phi\|_k^2 = \sum |\xi_n|^2 \left(1 + \left(\frac{2\pi n}{L}\right)^2\right)^{2k} = \|(1+p^2)^k \phi\|^2 < \infty,\quad k = 0,1,\ldots. \tag{2.2}$$

The seminorms $\phi \to \|\phi\|_k$ define a topology t on $\mathscr{D}$ which is Fréchet and reflexive with respect to t.

Lemma 2.1: The domain $\mathscr{D}$ is the following space of functions

$$\mathscr{D} = \{\phi(x) \in C^\infty[0,L]\colon \phi^{(k)}(0) = \phi^{(k)}(L),\ k = 0,1,\ldots\}. \tag{2.3}$$

With respect to the usual multiplication of functions and the involution $\phi^+(x) = \overline{\phi(x)}$, $\mathscr{D}$ becomes a topological *-algebra, and for $\phi, \chi \in \mathscr{D}$, one has

$$\|\chi\phi\|_k \leqslant c_k \|\chi\|_k \|\phi\|_k, \tag{2.4}$$

where C_k are certain constants.

Proof: The proof of (2.3) is straightforward, and (2.4) follows from the following estimations:

$$\|\chi\phi\|_k^2 = \int_0^L |(1+p^2)^k(\chi\phi)|^2\, dx \leqslant \int_0^L \left(\sum_{\upsilon=0}^{k} \binom{k}{\upsilon} |p^{2\upsilon}(\chi\phi)|\right)^2 dx \leqslant a_k \sum_{i=0}^{2k} \int_0^L |p^i\chi|^2\, dx \sum_{j=0}^{k} \int_0^L |p^j\phi|^2\, dx,$$

where the a_k are certain constants. Since every integral on the right-hand side can be estimated by $\|\chi\|_k$ (resp. $\|\phi\|_k$). (2.4) is proved.

Let $\mathscr{D}_k$ be the completion of $\mathscr{D}$ with respect to the Hilbert norm $\|\cdot\|_k$. Then we get the sequence

$$\mathscr{D} = \cap \mathscr{D}_k \subset \cdots \subset \mathscr{D}_{k-1} \subset \mathscr{D}_k \subset \cdots \subset \mathscr{D}_0 = \mathscr{H}.$$

By $\mathscr{D}_{-k}$ we denote the dual space of $\mathscr{D}_k$ with the norm

 0022-2488/87/010174-04$02.50

$\|\cdot\|_{-k}$. Then $\|\phi\|_{-k} = \|(1+p^2)^{-k}\phi\|$ and we get that $\mathscr{D}' = \cup \mathscr{D}_k$ is the dual space of $\mathscr{D}$. So we have the scale of Hilbert spaces

$$\lim_{n\to\infty} \operatorname{proj} \mathscr{D}_k = \mathscr{D}[t] \subset \mathscr{D}_0 = \mathscr{H} \subset \mathscr{D}'[t'] = \lim_{n\to\infty} \operatorname{ind} \mathscr{D}_k. \tag{2.5}$$

By interpolation (Ref. 3, I.4) we get $\mathscr{D}_s$ for every real s. The scalar product $\langle F,\phi\rangle = \langle \phi^+,F^+\rangle = \langle \overline{\phi,F}\rangle$ is defined for

$$\phi\in\mathscr{D}_k,\quad F\in\mathscr{D}_{-k},\quad k\in Z.$$

The space $\mathscr{D}'$ carries the strong dual topology t' defined by the seminorms

$$F\in\mathscr{D}' \to \sup_{\chi\in\mathscr{M}} |\langle F,\chi\rangle|, \tag{2.6}$$

where $\mathscr{M}$ runs over all bounded subsets of $\mathscr{D}[t]$.

The topology t can also be described by the seminorms

$$(a_n):\ \|F\|_{(a_n)} = \left(\sum |x_n|^2 a_n\right)^{1/2}$$

$$F = \sum_{n\in Z} x_n \phi_n, \tag{2.7}$$

where (a_n) runs over all positive sequences with $\Sigma_{n\in Z} a_n n^{2k} < \infty$ for all k.

Lemma 2.2: If $F\in\mathscr{D}'$, $g\in\mathscr{D}$, then $gF = Fg\in\mathscr{D}'$ is defined by

$$\langle gF,\varphi\rangle = \langle F,g^+\varphi\rangle,$$

and the map $F\to gF$ is continuous in $\mathscr{D}'$.

Proof: We need only to prove the continuity. This is a consequence of the following estimates. Let K be a bounded set in $\mathscr{D}[t]$,

$$\sup_{\chi\in K} |\langle Fg,\chi\rangle| = \sup_{\chi\in K} |\langle F,g^+\chi\rangle| = \sup_{\chi\in g^+K} |\langle F,\chi\rangle|.$$

As $\mathscr{D}$ is a topological *-algebra, the set g^+K is also bounded in $\mathscr{D}[t]$.

The foregoing lemmas prove the following theorem.

Theorem 2.3: $(\mathscr{D}'[t'],\mathscr{D})$ is a topological quasi-*-algebra,[2] i.e., (i) $\mathscr{D}'[t']$ is a linear topological space with continuous involution and a dense distinguished linear subspace $\mathscr{D}$, (ii) for $g\in\mathscr{D}$ the multiplication $F\to gF = Fg$ is defined and continuous, $F\in\mathscr{D}'$, and (iii) $\mathscr{D}$ is a *-algebra. The involution is $F^+ = \bar{F}$. For a topological quasi-*-algebra $(\mathfrak{A}[\xi],\mathfrak{A}_0)$ in[4] the left and the right strong topologies ${}^{\mathfrak{A}}\beta, \beta^{\mathfrak{A}}$ on $\mathfrak{A}_0$ have been introduced. If $(\mathfrak{A},\mathfrak{A}_0)$ is a commutative quasi-*-algebra, then both topologies coincide, ${}^{\mathfrak{A}}\beta = \beta^{\mathfrak{A}}$, and they are defined by the system of seminorms

$$\beta^{\mathfrak{A}}:\ \|A\|^{N,P} = \sup_{B\in N} p(AB), \tag{2.8}$$

where p runs over all seminorms of the topology ξ and N runs over all bounded subsets $N\subset\mathfrak{A}[\xi]$.

Now we get the following lemma.

Lemma 2.4: For the topological quasi-*-algebra $(\mathscr{D}'[t'],\mathscr{D})$ the strong topology $\beta^{\mathscr{D}'}$ on $\mathscr{D}$ coincides with the Fréchet topology t of $\mathscr{D}$.

Proof: First we prove that t is stronger than $\beta^{\mathscr{D}'}$. In fact by (2.6) and (2.8) the seminorms $\|\cdot\|^{N,p}$ are given by

$$\|A\|^{N,p} = \sup_{B\in N} \sup_{\chi\in M} |\langle AB,\chi\rangle$$

$$= \sup_{B\in N,\chi\in M} |F_{B^+\chi}(A)|,$$

where $F_{B^+\chi}(A) = \langle B^+\chi,A\rangle$ is a continuous linear functional. The set $\mathscr{F} = \{F_{B^+\chi};\ B\in N,\ \chi\in M\}$ of continuous functionals are pointwise bounded. By the generalized Banach–Steinhaus theorem Ref. 5, 15.13 $\mathscr{F}$ is equicontinuous and therefore $\|A\|^{N,p}$ is continuous in t.

Now we prove that t is weaker than $\beta^{\mathscr{D}'}$. For this we show that every norm $\|\cdot\|_k$ of t can be estimated by a norm $\|A\|^{N,p}$. Now we have $\|A\|_k = \sup_{F\in B} |\langle F,A\rangle|$ with a bounded set B of $\mathscr{D}'$. On $\mathscr{D}'$ we take the very simple seminorm $p(F) = |\langle F,1\rangle|$, $1\in\mathscr{D}$ and $N = B^+$. Then we get

$$\|A\|^{N,p} = \sup_{B\in B^+} |\langle AB,1\rangle| = \sup_{F\in B} |\langle F,A\rangle| = \|A\|_k.$$

Therefore the proof is complete.

III. THE CCR QUASI-*-ALGEBRA

Let $\mathscr{D}\subset\mathscr{H}\subset\mathscr{D}'$ be the rigged Hilbert space (2.5). Then $\mathscr{L}(\mathscr{D},\mathscr{D}')$ denotes the space of all continuous linear operators of $\mathscr{D}$ into $\mathscr{D}'$ and τ_∞ the topology of uniformly bounded convergence on $\mathscr{L}(\mathscr{D},\mathscr{D}')$ defined by the seminorms

$$\|A\|_{\mathscr{M}} = \sup_{\phi,\chi\in\mathscr{M}} |\langle A\phi,\chi\rangle|,\quad \mathscr{M} \text{ bounded in } \mathscr{D}[t]. \tag{3.1}$$

In $\mathscr{L}(\mathscr{D},\mathscr{D}')$ the involution $A\to A^+$ is defined by $\langle A\phi,\chi\rangle = \langle\phi,A^+\chi\rangle$, $\mathscr{L}^+(\mathscr{D})$ is the *-algebra of all operators $A\in\mathscr{L}(\mathscr{D},\mathscr{D}')$ with $A,A^+\mathscr{D}\subset\mathscr{D}$, $(\mathscr{L}(\mathscr{D},\mathscr{D}')\ [\tau_\infty], \mathscr{L}^+(\mathscr{D}))$ is a topological quasi-*-algebra. For this and other properties of $\mathscr{L}(\mathscr{D},\mathscr{D}')$ see Refs. 1 and 2.

Especially, every operator $A\in\mathscr{L}^+(\mathscr{D})$ has a continuation to an operator $A: \mathscr{D}'\to\mathscr{D}'$ defined by $\langle AF,\varphi\rangle = \langle F,A^+\varphi\rangle, \varphi\in\mathscr{D}$. We call (iP) the derivation also on $\mathscr{D}'$ and write for $F\in\mathscr{D}'$

$$(iP)^k F = F^{(k)},\quad k = 1,2,\ldots. \tag{3.2}$$

For $F\in\mathscr{D}'$ we denote by $\hat{F}$ the multiplication operator of $\mathscr{D}$ into $\mathscr{D}'$ (see Lemma 2.2)

$$\hat{F}\phi = F\phi,\quad \hat{F}\in\mathscr{L}(\mathscr{D},\mathscr{D}'), \tag{3.2}$$

$F\to\hat{F}$ defines a injection of $\mathscr{D}'$ into $\mathscr{L}(\mathscr{D},\mathscr{D}')$. With this notations it is straightforward to prove the commutation relations

$$P\hat{F} - \hat{F}P = -iF^{(1)}. \tag{3.4}$$

Definition 3.1: By $\mathscr{A}$ we denote the linear space of all polynomials B in a variable p with coefficients in $\mathscr{D}'$.

$$f = \sum_{k=0}^{N} F_k p^k,\quad F_k\in\mathscr{D}'. \tag{3.5}$$

This $\mathscr{A}$ is isomorphic to the direct sum $\overset{\infty}{\underset{1}{\oplus}}\mathscr{D}'$. The topology of the direct sum on $\mathscr{A}$ will be denoted by t' also. By $\mathscr{A}_0$ we denote the subspace of $\mathscr{A}$ of all f with $F_K\in\mathscr{D}$, $\mathscr{A}_0 = \overset{\infty}{\oplus}\mathscr{D}$. By t we denote the direct sum topology on $\mathscr{A}_0$.

In what follows we shall equip $\mathscr{A}$ with a partial multiplication (see also Ref. 6) such that $(\mathscr{A},\mathscr{A}_0)$ becomes a topological quasi-*-algebra. We start with the following representation π_0 of $\mathscr{A}$ in $\mathscr{L}(\mathscr{D},\mathscr{D}')$:

$$\pi_0(f)=\sum \widehat{F}_k p^k \in \mathscr{L}(\mathscr{D},\mathscr{D}') . \tag{3.6}$$

We put $\mathfrak{A}=\pi_0(\mathscr{A})$, $\mathfrak{A}_0=\pi_0(\mathscr{A}_0)$. The following lemma shows that $(\mathfrak{A},\mathfrak{A}_0)$ is a quasi-*-subalgebra of $(\mathscr{L}(\mathscr{D},\mathscr{D}'),\ \mathscr{L}^+(\mathscr{D}))$.

Lemma 3.2: (i) π_0 is a bijection of $\mathscr{A}$ onto $\mathfrak{A}$; (ii) $\mathfrak{A}_0=\pi_0(\mathscr{A}_0)\subset\mathscr{L}^+(\mathscr{D})$ is an Op*-algebra; (iii) For $A\in\mathfrak{A}$ and $B\in\mathfrak{A}_0$ we have AB, $BA\in\mathfrak{A}$.

Proof: (i) Let $f=\Sigma F_k p^k\neq 0$. Show $\pi_0(f)\neq 0$ we prove the existence of two elements ϕ_r,ϕ_s of the basis (2.1) such that $\langle\pi_0(f)\phi_r,\phi_s\rangle\neq 0$. We have

$$\langle\pi\eta_0(f)\phi_n,\phi_m\rangle=\sum_{k=0}^{N}((2\pi/L)n)^k\langle F_K\phi_n,\phi_m\rangle$$
$$=\sqrt{L}\sum_{K=0}^{N}((2\pi/L)n)^k\langle F_K,\phi_{m-n}\rangle . \tag{3.7}$$

The vector $(\langle F_1,\phi_t\rangle,...,\langle F_N,\phi_t\rangle)$ is different from zero for a certain t. Furthermore, it is not orthogonal to every vector $\langle 1,(2\pi/L)n,...,((2\pi/L)n)^N\rangle$, $n\in Z$. Therefore for one r we have

$$\sqrt{L}\sum_{k=0}^{\infty}((2\pi/L)r)^k\langle F_K|\phi_t\rangle\neq 0 .$$

This is equal to $\langle\pi_0(f)\phi_r,\phi_s\rangle$ for $s=t+r$. Therefore (i) is proved.

(ii) Let $f_1=\Sigma a_k(Q)p^k$, $f_2=\Sigma b_l(Q)p^l$ be elements of $\mathscr{A}_0$, i.e., $a_k(Q),b_l(Q)\in\mathscr{D}$, then making use of the commutation relations between P and Q on $\mathscr{D}$, we have

$$\pi_0(f_1):\quad \pi_0(f_2)=\pi_0(f_3) , \tag{3.8}$$

where

$$f_3=\sum_{K,l}\sum_{s=0}^{k}\binom{K}{S}(-i)^{k-s}a_k\, b_l^{(k-s)}p^{s+l} .$$

(iii) This follows from calculations similar to the previous ones taking into account the commutation relations (3.4).

Lemma 3:3: $(\mathfrak{A},\mathfrak{A}_0)$ is a topological quasi-*-algebra with respect to the topology τ_∞.

Proof: We need only to prove that $\mathfrak{A}_0$ is dense in $\mathfrak{A}$ with respect to $\tau_{\mathscr{D}}$. We will make use of the following facts: (a) $\mathscr{D}$ is dense in $\mathscr{D}'[t']$, and (b) since $\mathscr{D}[t]$ is a topological *-algebra with jointly continuous multiplication, if $\mathscr{M}$ and $\mathscr{N}$ are bounded in $\mathscr{D}$ so also $\mathscr{M}\mathscr{N}$ is.

Now let $\sum_{k=1}^{N}\widehat{F}_k p^k\in\mathfrak{A}$, for each $F_k\in\mathscr{D}'$ there is a net F_k^α in $\mathscr{D}$ which converges to F_k with respect to t'. We have therefore for a bounded subset $\mathscr{M}$ of $\mathscr{D}\ [t]$

$$\sup_{\phi,\psi\in\mathscr{M}}\left|\sum_k\langle(\widehat{F}_k-\widehat{F}_k^\alpha)p^k\phi,\chi\rangle\right|$$
$$\leqslant\sum_k\sup_{\phi,\chi\in\mathscr{M}}|\langle(\widehat{F}_k-\widehat{F}_k^\alpha)p^k\phi,\chi\rangle|$$
$$\leqslant\sum_k\sup_{\phi,\chi\in\mathscr{M}}|\langle F_k-F_k^\alpha,\ \overline{p^k\phi\chi}\rangle|$$
$$\leqslant\sum_k\sup_{\chi\in\mathscr{M}\mathscr{N}}|\langle F_k-F_k^\alpha,\chi\rangle|\to 0 ,$$

where $\mathscr{N}=\bigcup_{k=1}^{N}\overline{p^k\mathscr{M}}$. This proves the statement [see (3.1)].

IV. BOGOLUBOV INEQUALITY

In this section we are going to describe the dynamics and the equilibrium state of the free Bose system[7] in the interval (box) $[0,L]$ by a derivation

$$\delta(A)=i[H,A] , \tag{4.1}$$

and the Gibbs state

$$\langle A\rangle=(1/Z)\mathrm{tr}\, e^{-\beta H}A \tag{4.2}$$

on the quasi-*-algebra $(\mathfrak{A},\mathfrak{A}_0)$, where $H=p^2$ is the free Hamiltonian and β the inverse temperature.

At first sight the definition of (4.1) and (4.2) might look trivial but formal calculations with the Bogolubov inequality

$$\tfrac{1}{2}\beta\langle AA^++A^+A\rangle\langle[[C,H],C^+]\rangle\geqslant|\langle[C,A]\rangle|^2 \tag{4.3}$$

and the commutation relation $[Q,P]=i$ leads to the contradiction $1\leqslant 0(!)$ for $C=P$ and $A=Q$. In Ref. 8 this contradiction has been solved by a new definition of states on unbounded operators as sesquilinear forms and a corresponding generalization of Bogolubov inequality (see Ref. 8, Lemma 2).

In fact the derivation (4.1) and the state (4.2) are well-defined on the quasi-*-algebra $(\mathfrak{A},\mathfrak{A}_0)$ and also the Bogolubov inequality (3.3) is valid. There cannot be a contradiction. The problem will be solved by the fact that the commutator of Q and P, which is well-defined on $\mathfrak{A}$ is different from i.

Lemma 4.1: Let $F(x)$ be continuous differentiable on the closed interval $[0,L]$. Then $F\in\mathscr{D}'$ and we get [see (2.2)]

$$F^{(1)}=F'(x)-(F(L)-F(0))\delta(x) , \tag{4.4}$$

where $F'(x)$ is the usual derivative of $F(x)$.

Proof: (4.4) follows from the relation

$$\langle iPF,\varphi\rangle=-i\langle F,P\varphi\rangle$$
$$=-\int_0^L\overline{F}\varphi'\,dx=-\overline{F}(x)\varphi(x)\,|_0^L+\int_0^L\overline{F}'\varphi\,dx$$
$$=\int_0^L\overline{F}'\varphi(x)\,dx-(\overline{F}(L)-\overline{F}(0))\varphi(0) ,$$

for all $\varphi\in\mathscr{D}$, because $\varphi(0)=\varphi(L)$.

If $Q=x$ is the position then we get, from (4.4) and (3.4), the commutation relation

$$[Q,P]=QP-PQ=i-iL\hat{\delta}(x) . \tag{4.5}$$

The products QP,PQ are well defined in $\mathfrak{A}$, since $P\in\mathfrak{A}$. Furthermore, all products in (4.2) and (4.3) are well defined.

Lemma 4.2: (i) $\tau_t(A)=e^{iHt}Ae^{-iHt}$ defines a one-parameter group of linear transformations on $\mathscr{L}(\mathscr{D},\mathscr{D}')$.

(ii) The state $\langle A\rangle=(1/Z)\mathrm{tr}\, e^{-\beta H}A$, $Z=\mathrm{tr}\, e^{-\beta H}$, is well defined on $\mathscr{L}(\mathscr{D},\mathscr{D}')$.

Proof: (i) This follows from the fact that $e^{iHt} \in \mathscr{L}^+(\mathscr{D})$. It is even an operator of order zero,[7] so that

$$\|e^{iHt}\phi\|_k = \|(1+p^2)^k e^{iHt}\phi\| \leqslant \|\phi\|_k\,, \quad \text{for all } k.$$

(ii) For every $A \in \mathscr{L}(\mathscr{D},\mathscr{D}')$ the operator $e^{-\beta H}{}_A$ is nuclear[7] and we have

$$\operatorname{tr} e^{-\beta H}A = \sum_{n\in Z} e^{-\beta\epsilon_n}\langle \phi_n, A\phi_n\rangle\,, \tag{4.6}$$

where $\epsilon_n = (4\pi^2/L^2)n^2$ are the eigenvalues of H.

Let us yet remark that the dynamics $\tau_t(A)$ does not leave $\mathfrak{A}$ invariant, whereas its derivative (4.1) leaves $\mathfrak{A}$ invariant.

Since $\langle \phi_n, \hat{\delta}\phi_n\rangle = 1/L$ for all n we get $\langle\hat{\delta}\rangle = 1/L$. Hence

$$\langle [P,Q]\rangle = 0\,. \tag{4.7}$$

Therefore there is no contradiction in applying the Bogolubov inequality (4.3) to $C = P, A = Q$, because we get zero on both sides.

Let us finally remark that the considerations of this section are valid also for different Bose gases with interaction in a finite box.[8] The essential condition is the nuclearity of the operator $e^{-\beta H}A$ for $A \in \mathscr{L}(\mathscr{D},\mathscr{D}')$, which is satisfied in such cases.

[1]G. Lassner, "Topological algebras and their application in quantum statistics," Wiss. Z. Karl-Marx-Univ. Leipzig Math.-Natur. Reihe **30**, 572 (1981).

[2]G. Lassner, "Algebras of unbounded operators and quantum dynamics," Physica A 471 (1984).

[3]M. E. Taylor, *Pseudodifferential Operators* (Princeton U. P., Princeton, NJ, 1981).

[4]G. Lassner, G. A. Lassner "Twistprodukt und quasi-*-algebren," Z. Anal. Chem. **4**(4), 363 (1985).

[5]G. Köthe, *Topologische Lineare Räum* (Springer, Berlin, 1966), Bd. I.

[6]J. P. Antoine, and W. Karwowski, "Partial *-algebras of closed operators in Hilbert space," Publ. Res. Inst. Math. Sci., Kyoto University **21**, 205 (1985).

[7]G. A. Lassner, "Continuity of state functions of the Bose gas," Rep. Math. Phys. **20**, 65 (1984).

[8]M. Bouziane, and P. A. Martin, "Bogolubov inequality for unbounded operators and the Bose gas," J. Math. Phys. **17**, 1848 (1976).

A singular integral equation formulation and solution for transport in semi-infinite ducts

Fausto Malvagi and G. C. Pomraning
School of Engineering and Applied Science, University of California, Los Angeles, California 90024

(Received 3 June 1986; accepted for publication 17 September 1986)

An unusual two-group transport equation, associated with particle transport in a duct, and possessing a full "removal matrix" is considered. An approximate solution for the exiting distribution is obtained by an application of the facile (F_N) method. An integral transform technique is used to transform the problem into a singular integral equation, which serves as a basis for the F_N approximation. The set of basis functions is derived from the singular eigenfunctions of the original problem. Comparisons of the reflection probability with numerical solutions show the accuracy and the remarkable efficiency of the F_N method.

I. INTRODUCTION

In recent papers a new transport equation has arisen, presenting some peculiar and interesting features. The physical problem is that of treating neutral particle transport in an evacuated duct, with partial isotropic scattering at the wall. Pomraning and Prinja[1] derived a simple model for the monoenergetic, time-independent problem, which reduces the number of independent variables from 5 (3 in space, 2 in angle) to 2 (1 in space, 1 in angle). This was shown[2,3] to be the lowest-order approximation in a hierarchy of approximations derived by a weighted residual procedure. The next-order approximation, which was then derived,[3] leads to two coupled planar geometry transport equations. They can be written in vector form as

$$\mu\frac{\partial}{\partial x}\mathbf{\Phi}(x,\mu) + (1-\mu^2)^{1/2}\mathsf{S}\mathbf{\Phi}(x,\mu) = \frac{2c}{\pi}(1-\mu^2)^{1/2}\mathsf{B}\int_{-1}^{1} d\mu'(1-\mu'^2)^{1/2}\mathbf{\Phi}(x,\mu'), \tag{1.1}$$

with $0<x<\infty$, and with the boundary conditions

$$\mathbf{\Phi}(0,\mu) = \mathbf{f}(\mu), \quad 0\leqslant\mu\leqslant 1, \tag{1.2}$$

$$\lim_{x\to\infty} \mathbf{\Phi}(x,\mu) = 0, \tag{1.3}$$

where the two components of $\mathbf{\Phi}$ are weighted averages of the particle flux, $\mathbf{f}(\mu)$ is the prescribed incident flux, and c is the scattering probability at the wall. The elements of the two matrices S and B have been defined for arbitrary cross sectional geometry of the duct.[3] We will focus our attention on the case of a circular duct of semi-infinite length. The numerical values of the elements of S and B are then given by[3]

$$s_{11} = b_{11} = 2/\pi, \tag{1.4}$$

$$s_{12} = b_{12} = (3\pi - 16/\pi)(9\pi^2-64)^{-1/2}, \tag{1.5}$$

$$s_{21} = b_{21} = -2(3\pi - 16/\pi)(9\pi^2-64)^{1/2}(128-9\pi^2)^{-1}, \tag{1.6}$$

$$s_{22} = -(16/\pi)b_{22} = 16(3\pi-16/\pi)(128-9\pi^2)^{-1}. \tag{1.7}$$

Equation (1.1) represents a two-group transport equation with anisotropic cross sections. The peculiarity of this problem is that both the "removal matrix" and the "scattering matrix" are full, while in ordinary two-group transport equations the "removal matrix" is always diagonal. An additional unusual feature is the occurrence of a continuous spectrum which is complex and involves lines of infinite length.

In this paper we are interested in an application of the F_N method to the problem described by Eq. (1.1). The F_N method, initially introduced in the context of neutron transport theory by Siewert and Benoist[4] and Grandjean and Siewert,[5] has also proved to be particularly efficient in solving basic transport problems in the field of radiative transfer and rarefied gas dynamics (an interesting review of the applications of the F_N method was recently given by Garcia[6]). The method, though approximate, can yield very accurate numerical results with modest computational efforts. The method has already been applied to an infinite spectrum,[7] and to a two-group problem,[8] but never to a problem with a complex spectrum, nor one involving a full removal matrix.

We first use an integral transform technique suggested by Siewert[8,9] to derive a singular integral equation for the exiting flux. The exiting flux itself is then approximated by a finite expansion in terms of a set of basis functions. Those functions are chosen, after a short discussion, to be the eigenfunctions of the problem. The coefficients of the expansion are found by requiring that the integral equation be satisfied at certain (collocation) points. Once our approximate solution is established, we check its accuracy and efficiency by evaluating the reflection probability (albedo) at the open end $x=0$, defined by

$$\alpha = \frac{\int_0^1 d\mu\, \mu\Phi_1(0,-\mu)}{\int_0^1 d\mu\, \mu f_1(\mu)}, \tag{1.8}$$

for different choices of the incoming flux f and the wall scattering probability c. The results are then compared to the ones given in Ref. 3, obtained from a direct numerical solution of Eq. (1.1).

II. DERIVATION OF THE SINGULAR INTEGRAL EQUATION

We start with a change of variable from μ to ξ according to

$$\xi = \mu(1-\mu^2)^{-1/2}, \tag{2.1}$$

$$\boldsymbol{\psi}(x,\xi) = \mathbf{\Phi}(x,\mu), \tag{2.2}$$

 0022-2488/87/010178-06$02.50

$$\mathbf{F}(\xi) = \mathbf{f}(\mu). \tag{2.3}$$

Equations (1.1) through (1.3) then become

$$\xi \frac{\partial}{\partial x} \boldsymbol{\psi}(x,\xi) + \mathsf{S}\boldsymbol{\psi}(x,\xi) = \mathsf{C} \int_{-\infty}^{\infty} d\xi'(\xi'^2 + 1)^{-2} \boldsymbol{\psi}(x,\xi'), \tag{2.4}$$

$$\boldsymbol{\psi}(0,\xi) = \mathbf{F}(\xi), \quad 0 \leqslant \xi < \infty, \tag{2.5}$$

$$\lim_{x\to\infty} \boldsymbol{\psi}(x,\xi) = 0, \tag{2.6}$$

where $\mathsf{C} = (2c/\pi)\mathsf{B}$.

Now we follow the same procedure reported in Refs. 8 and 9, and change ξ to $-\xi$, multiply the resulting equation by $e^{-x/s}$ and integrate with respect to x from 0 to ∞. Upon integration by parts and multiplication by s we obtain

$$(s\mathsf{S} - \xi\mathsf{I}) \int_0^{\infty} dx\, e^{-x/s} \boldsymbol{\psi}(x, -\xi) = s\mathsf{C}\boldsymbol{\rho}^*(s) - s\xi\boldsymbol{\psi}(0, -\xi), \tag{2.7}$$

where

$$\boldsymbol{\rho}^*(s) = \int_0^{\infty} dx\, e^{-x/s} \int_{-\infty}^{\infty} d\xi(\xi^2 + 1)^{-2} \boldsymbol{\psi}(x, -\xi), \tag{2.8}$$

and I is the identity matrix of rank 2. We must also consider only $\operatorname{Re} s > 0$.

Disregarding, for the time being, the occurrence of the singularities of $(s\mathsf{S} - \xi\mathsf{I})$, we formally multiply Eq. (2.7) by $(\xi^2 + 1)^{-2}(\xi\mathsf{I} - s\mathsf{S})^{-1}$ and integrate over ξ from $-\infty$ to ∞. Making use of the boundary condition (2.5), we write the result as

$$\begin{aligned}\boldsymbol{\Lambda}(s)\boldsymbol{\rho}^*(s) = s \int_0^{\infty} d\xi\, \xi(\xi^2 + 1)^{-2}(\xi\mathsf{I} - s\mathsf{S})^{-1} \\ \times \boldsymbol{\psi}(0, -\xi) + s\boldsymbol{\Gamma}(s),\end{aligned} \tag{2.9}$$

where

$$\boldsymbol{\Lambda}(s) = \mathsf{I} + s \int_{-\infty}^{\infty} d\xi(\xi^2 + 1)^{-2}(\xi\mathsf{I} - s\mathsf{S})^{-1}\mathsf{C}, \tag{2.10}$$

$$\boldsymbol{\Gamma}(s) = \int_0^{\infty} d\xi\, \xi(\xi^2 + 1)^{-2}(\xi\mathsf{I} + s\mathsf{S})^{-1}\mathbf{F}(\xi). \tag{2.11}$$

Since the singularities of Eq. (2.9) occur when the determinant of the matrix $(s\mathsf{S} - \xi\mathsf{I})$ vanishes, we look at

$$\begin{aligned}\det(s\mathsf{S} - \xi\mathsf{I}) &= (\xi - g_+ s)(\xi - g_- s) \\ &= \det(\mathsf{S})(s - f_+\xi)(s - f_-\xi),\end{aligned} \tag{2.12}$$

where

$$g_\pm = (\tfrac{1}{2})(s_{11} + s_{22}) \pm \tfrac{1}{2}((s_{11} + s_{22})^2 - 4\det(\mathsf{S}))^{1/2}, \tag{2.13}$$

$$f_\pm = (g_\mp)^{-1}. \tag{2.14}$$

In our case $(s_{11} + s_{22})^2 < 4\det(\mathsf{S})$, so that the singular spectrum consists of the two conjugate linear branches in the complex half-plane $\operatorname{Re} s > 0$ defined by $s = f_\pm \nu$, with ν real and positive. Those are the continuum eigenvalues of the problem, as we will also see later.

If we now let s approach, say, $f_+\nu$ (upper branch) in Eq. (2.9), we find according to the Plemelj formulas

$$\begin{aligned}\boldsymbol{\Lambda}_+^\pm(\nu)\boldsymbol{\rho}_+^*(\nu) = f_+\nu P \int_0^{\infty} d\xi\, \xi\, [(\xi^2 + 1)^2(\xi - \nu)(\xi - \kappa_+\nu)]^{-1}(\xi\mathsf{I} - f_+\nu\mathsf{S}^\dagger)\boldsymbol{\psi}(0, -\xi) \\ \pm \pi i f_+\nu^2[(\nu^2 + 1)^2(1 - \kappa_+)]^{-1}(\mathsf{I} - f_+\mathsf{S}^\dagger)\boldsymbol{\psi}(0, -\nu) + f_+\nu\boldsymbol{\Gamma}_+(\nu),\end{aligned} \tag{2.15}$$

where

$$\begin{aligned}\boldsymbol{\Lambda}_+^\pm(\nu) = \mathsf{I} + f_+\nu P \int_{-\infty}^{\infty} d\xi\, [(\xi^2 + 1)^2(\xi - \nu)(\xi - \kappa_+\nu)]^{-1} \\ \times(\xi\mathsf{I} - f_+\nu\mathsf{S}^\dagger)\mathsf{C} \pm \pi i f_+\nu[(\nu^2 + 1)^2(1 - \kappa_+)]^{-1}(\mathsf{I} - f_+\mathsf{S}^\dagger)\mathsf{C},\end{aligned} \tag{2.16}$$

$$\kappa_+ = g_+(g_-)^{-1}. \tag{2.17}$$

Here P indicates the principal value of the integral, $\mathsf{S}^\dagger$ is the adjoint of S, namely $\det(\mathsf{S})$ times S^{-1}, the subscript $+$ indicates that we are referring to the upper branch, and the $\pm$ refers to approaching the branch from above or from below, respectively. We have also rewritten $\boldsymbol{\rho}^*(f_+\nu)$ and $\boldsymbol{\Gamma}(f_+\nu)$ as $\boldsymbol{\rho}_+^*(\nu)$ and $\boldsymbol{\Gamma}_+(\nu)$, respectively. Adding and subtracting the equations that correspond to the two ways of approach we find

$$\begin{aligned}\boldsymbol{\Delta}_+(\nu)\boldsymbol{\rho}_+^*(\nu) = f_+\nu P \int_0^{\infty} d\xi\, \xi\, [(\xi^2 + 1)^2(\xi - \nu) \\ \times(\xi - \kappa_+\nu)]^{-1}(\xi\mathsf{I} - f_+\nu\mathsf{S}^\dagger)\boldsymbol{\psi}(0, -\xi) \\ + f_+\nu\boldsymbol{\Gamma}_+(\nu),\end{aligned} \tag{2.18}$$

$$(\mathsf{I} - f_+\mathsf{S}^\dagger)\boldsymbol{\psi}(0, -\nu) = (1/\nu)(\mathsf{I} - f_+\mathsf{S}^\dagger)\mathsf{C}\boldsymbol{\rho}_+^*(\nu), \tag{2.19}$$

where

$$\begin{aligned}\boldsymbol{\Delta}_+(\nu) = \mathsf{I} - f_+\nu P \int_{-\infty}^{\infty} d\xi\, [(\xi^2 + 1)^2(\xi - \nu) \\ \times(\xi - \kappa_+\nu)]^{-1}(f_+\nu\mathsf{S}^\dagger - \xi\mathsf{I})\mathsf{C}.\end{aligned} \tag{2.20}$$

When we let s approach the lower branch, the equations we obtain are simply the complex conjugates of Eqs. (2.18)–(2.20).

From Eq. (2.18) and the corresponding one related to the lower branch, we can readily find expressions for $\boldsymbol{\rho}_+^*$ and $\boldsymbol{\rho}_-^*$, but the use of Eq. (2.19) and its complex conjugate is somewhat less straightforward, since the two matrices $(\mathsf{I} - f_\pm\mathsf{S}^\dagger)$ are singular. Recalling that the flux $\boldsymbol{\psi}$ is a real quantity, it can be checked after some algebra that the real and imaginary parts of Eq. (2.19) are linearly dependent, so that we can use either one of them to obtain

$$\psi(0,-\nu)=P\int_0^\infty d\xi\,\mathsf{H}(\nu,\xi)\psi(0,-\xi)+\mathbf{K}(\nu), \tag{2.21}$$

where

$$\mathsf{H}(\nu,\xi)=\{\mathrm{Re}(\mathsf{I}-f_+\mathsf{S}^\dagger)\}^{-1}\,\mathrm{Re}\{(\mathsf{I}-f_+\mathsf{S}^\dagger)\mathsf{C}\boldsymbol{\Delta}_+^{-1}(\nu)\times f_+\xi\,[(\xi^2+1)^2(\xi-\nu)\times(\xi-\kappa_+\nu)]^{-1}(\xi\mathsf{I}-f_+\nu\mathsf{S}^\dagger)\}, \tag{2.22}$$

$$\mathbf{K}(\nu)=\{\mathrm{Re}(\mathsf{I}-f_+\mathsf{S}^\dagger)\}^{-1}\times\mathrm{Re}\{(\mathsf{I}-f_+\mathsf{S}^\dagger)\mathsf{C}\boldsymbol{\Delta}_+^{-1}(\nu)f_+\boldsymbol{\Gamma}_+(\nu)\}. \tag{2.23}$$

Since we can straightforwardly compute H and $\mathbf{K}$ (some sketches of this algebra are shown in the appendix), Eq. (2.21) is the integral equation for the exiting flux which constitutes the basic equation to which the F_N approximation is applied.

III. EIGENFUNCTIONS AND THE F_N METHOD

The F_N method consists of approximating the exiting flux in Eq. (2.21) as a linear combination of a finite number of basis functions. Any complete set of functions could be used, but the choice of this set is very crucial to the rate of convergence of the expansion. Initially the expansion

$$\psi(0,-\xi)=\sum_{n=1}^{N}(\xi+\xi_n)^{-1}\mathbf{A}_n,\quad \xi,\ \xi_n>0, \tag{3.1}$$

was assumed, in which scalar basis functions are used, with vectorial expansion coefficients. This was suggested by the solution of the one group equation.[1] However, the convergence of the numerical results proved to be relatively slow, as will be shown in Sec. IV. Thus the eigenfunctions of the problem were derived, in order to infer from them a more natural choice for the basis functions.

In order to find these eigenfunctions the classic procedure is followed.[10] We start with the ansatz

$$\psi(x,\xi)=e^{-x/s}\mathbf{F}_s(\xi), \tag{3.2}$$

where, because of the condition of finiteness at infinity, we restrict ourselves to $\mathrm{Re}\,s>0$. After substitution into Eq. (2.4) we obtain

$$(\mathsf{S}s-\xi\mathsf{I})\mathbf{F}_s(\xi)=s\mathsf{C}\mathbf{N}_s, \tag{3.3}$$

where

$$\mathbf{N}_s=\int_{-\infty}^{\infty}d\xi(\xi^2+1)^{-2}\mathbf{F}_s(\xi). \tag{3.4}$$

The singularities of the matrix in the left-hand side of Eq. (3.3) are the eigenvalues of our problem. They lie on the two conjugate branches of the complex half-plane defined by $s=f_\pm\nu, 0<\nu<\infty$. The singular eigenfunctions appear then as complex conjugate pairs, and can be written (say for the upper branch) as

$$\mathbf{F}_{+,\nu}(\xi)=Pf_+\nu[(\xi-\nu)(\xi-\kappa_+\nu)]^{-1}\times(f_+\nu\mathsf{S}^\dagger-\xi\mathsf{I})\mathsf{C}\mathbf{N}_{+,\nu}+\mathbf{D}_+\lambda_+(\nu)\delta(\xi-\nu),\quad \xi>0, \tag{3.5}$$

$$\mathbf{F}_{+,\nu}(-\xi)=Pf_+\nu[(\xi+\nu)(\xi+\kappa_+\nu)]^{-1}\times(f_+\nu\mathsf{S}^\dagger+\xi\mathsf{I})\mathsf{C}\mathbf{N}_{+,\nu},\quad \xi>0, \tag{3.6}$$

where $\mathbf{D}_+$ is the vector

$$\mathbf{D}_+=\begin{bmatrix} f_+s_{12}/(1-f_+s_{11})-1\\ f_+s_{21}/(1-f_+s_{22})-1\end{bmatrix}, \tag{3.7}$$

and P, as usual, indicates that integrals must be interpreted as principal values, and the subscript $+$ refers to the upper branch. Again the eigenfunctions corresponding to the lower branch are the complex conjugate of the ones corresponding to the upper branch. Substitution of Eqs. (3.5) and (3.6) into Eq. (3.4) gives

$$\boldsymbol{\Delta}_+(\nu)\mathbf{N}_{+,\nu}=(\nu^2+1)^{-2}\mathbf{D}_+\lambda_+(\nu), \tag{3.8}$$

where $\boldsymbol{\Delta}_+(\nu)$ is the same matrix defined by Eq. (2.20). Now, choosing the value of the normalization for the first component of $\mathbf{N}_{+,\nu}$, which physically corresponds to particle conservation, we obtain from Eq. (3.8)

$$N_{+,\nu,1}=1, \tag{3.9}$$

$$N_{+,\nu,2}=[D_{+,2}\Delta_{+,11}(\nu)-D_{+,1}\Delta_{+,21}(\nu)]\times[D_{+,1}\Delta_{+,22}(\nu)-D_{+,2}\Delta_{+,12}(\nu)]^{-1}, \tag{3.10}$$

$$\lambda_+(\nu)=\det(\boldsymbol{\Delta}_+)[D_{+,1}\Delta_{+,22}(\nu)-D_{+,2}\Delta_{+,12}(\nu)]^{-1}. \tag{3.11}$$

Assuming that the eigenfunctions (3.5) and (3.6) form a complete set, we can then write the exiting flux as a superposition of the eigenfunctions themselves, i.e.,

$$\psi(0,-\xi)=\int_0^\infty d\nu[A_+(\nu)\mathbf{F}_{+,\nu}(-\xi)+A_-(\nu)\mathbf{F}_{-,\nu}(-\xi)], \tag{3.12}$$

with $0<\xi<\infty$. Since $A_\pm(\nu)$ and $\mathbf{F}_{\pm,\nu}$ are complex, but the solution ψ must be real, Eq. (3.12) can also be written as

$$\psi(0,-\xi)=\int_0^\infty d\nu\{a(\nu)\mathrm{Re}[\mathbf{F}_{+,\nu}(-\xi)]+b(\nu)\mathrm{Im}[\mathbf{F}_{+,\nu}(-\xi)]\}. \tag{3.13}$$

It therefore appears natural to expand the exiting flux as

$$\psi(0,-\xi)=\sum_{n=1}^{N}[a_n\mathbf{R}_n(\xi)+b_n\mathbf{I}_n(\xi)], \tag{3.14}$$

where

$$\mathbf{R}_n(\xi)=\mathrm{Re}[\mathbf{F}_{+,\nu_n}(-\xi)], \tag{3.15}$$

$$\mathbf{I}_n(\xi)=\mathrm{Im}[\mathbf{F}_{+,\nu_n}(-\xi)]. \tag{3.16}$$

Equations (3.15) and (3.16) represent our choice for the basis functions, and Eq. (3.14) is the related F_N approximation. Comparing Eqs. (3.14) and (3.1) we see that now the basis functions are vectors, and the coefficients of the expansion are scalars. Substituting Eq. (3.14) into Eq. (2.21) we obtain

$$\sum_{n=1}^{N}\left\{a_n\left[\mathbf{R}_n(\nu)-P\int_0^\infty d\xi\,\mathsf{H}(\nu,\xi)\mathbf{R}_n(\xi)\right]+b_n\left[\mathbf{I}_n(\nu)-P\int_0^\infty d\xi\,\mathsf{H}(\nu,\xi)\mathbf{I}_n(\xi)\right]\right\}=\mathbf{K}(\nu). \tag{3.17}$$

If we now evaluate Eq. (3.17) at N collocation points ν_m, $m = 1,...,N$, we are led to a linear system of $2N$ scalar equations for the $2N$ unknowns a_n and b_n. This system can be indicated as

$$\mathbf{MA} = \mathbf{L}, \tag{3.18}$$

where

$$M_{2m+i-2,2n-1} = R_{n,i}(\nu_m) - P\int_0^\infty d\xi\,[H(\nu_m,\xi)R_n(\xi)]_i,$$

$$i = 1,2, \tag{3.19}$$

$$M_{2m+i-2,2n} = I_{n,i}(\nu_m) - P\int_0^\infty d\xi\,[H(\nu_m,\xi)I_n(\xi)]_i,$$

$$i = 1,2, \tag{3.20}$$

$$A_{2m-1} = a_m, \tag{3.21}$$

$$A_{2m} = b_m, \tag{3.22}$$

$$L_{2m-1} = K_1(\nu_m), \tag{3.23}$$

$$L_{2m} = K_2(\nu_m), \tag{3.24}$$

with $m,n = 1,2,...,N$.

Once the system (3.18) is solved, Eq. (3.14) gives our approximate solution for the exiting flux.

IV. NUMERICAL RESULTS

In this section we are interested in testing the accuracy and efficiency of our approximate solution. Therefore we compare the results for the albedo α, defined, after the change of variable (2.1), by

$$\alpha = \frac{\int_0^\infty d\xi\,\xi(\xi^2+1)^{-2}\psi_1(0,-\xi)}{\int_0^\infty d\xi\,\xi(\xi^2+1)^{-2}F_1(\xi)}, \tag{4.1}$$

with those reported in Ref. 3 from a completely numerical solution of Eq. (1.1). Substituting into Eq. (4.1) the approximate solution given by Eq. (3.14), the reflection probability is computed as

$$\alpha = \frac{1}{Q}\sum_{n=1}^{N}\left\{a_n\int_0^\infty d\xi\,\xi(\xi^2+1)^{-2}R_{n,1}(\xi) + b_n\int_0^\infty d\xi\,\xi(\xi^2+1)^{-2}I_{n,1}(\xi)\right\}, \tag{4.2}$$

where

$$Q = \int_0^\infty d\xi\,\xi(\xi^2+1)^{-2}F_1(\xi). \tag{4.3}$$

As for the choice of the collocation points in Eq. (3.17), we initially used a set of points uniformly spaced in the variable μ, i.e.,

$$\xi_i = \nu_i = \mu_i(1-\mu_i^2)^{-1/2}, \quad 1 \leqslant i \leqslant N, \tag{4.4}$$

where

$$\mu_i = i/(N+1), \quad 1 \leqslant i \leqslant N. \tag{4.5}$$

However, the choice

$$\mu_i = \cos(i/(2N+1)), \quad 1 \leqslant i \leqslant N, \tag{4.6}$$

proved to give better results. The latter set of points is based upon the Chebyshev polynomials of the second kind,[11] which are the natural (Gaussian) quadrature points for the integral in Eq. (1.1).

We then consider two different kinds of boundary conditions. The first case

$$\mathbf{F}(\xi) = \begin{bmatrix}1\\0\end{bmatrix} \tag{4.7}$$

corresponds to an isotropic distribution of the incoming flux. Another physically interesting case is that of a delta function incidence

$$\mathbf{F}(\xi) = \begin{bmatrix}\delta(\xi-\xi_0)\\0\end{bmatrix}. \tag{4.8}$$

Table I shows the results for the albedo with an isotropic incidence, computed for various values of the wall scattering probability c. The numerical results perfectly agree up to $c = 0.7$. The slight difference in the range from $c = 0.7$ to $c = 0.99$ is believed to be related to a noncomplete convergence of the results given in Ref. 3.

Tables II and III refer to a delta function incidence for two different values of c. Again our results agree remarkably well with those given by the numerical solution.

TABLE I. Reflection probability with isotropic incidence.

	Reflection probability	
c	F_N	Numerical
0.1	0.0256	0.0256
0.2	0.0540	0.0540
0.3	0.0861	0.0861
0.4	0.1227	0.1227
0.5	0.1654	0.1654
0.6	0.2167	0.2167
0.7	0.2810	0.2814
0.8	0.3673	0.3677
0.9	0.5010	0.5014
0.95	0.6147	0.6152
0.99	0.8019	0.8027

TABLE II. Reflection probability with delta function incidence ($c = 0.6$).

	Reflection probability	
μ_0	F_N	Numerical
0.1	0.356 5	0.356 4
0.2	0.335 9	0.336 0
0.4	0.293 9	0.293 9
0.6	0.249 6	0.249 5
0.8	0.192 9	0.192 9
0.99	0.054 57	0.054 05

TABLE III. Reflection probability with delta function incidence ($c = 0.95$).

	Reflection probability	
μ_0	F_N	Numerical
0.1	0.7662	0.767
0.2	0.7502	0.751
0.4	0.7150	0.716
0.6	0.6720	0.672
0.8	0.6039	0.603
0.99	0.2887	0.280

TABLE IV. Convergence pattern with isotropic incidence, $c = 0.6$.

Number of collocation points	Reflection probability Chebyshev quadrature; eigenfunction basis	Equal spacing; eigenfunction basis	Equal spacing; scalar basis
2	0.214 24	0.210 97	0.228 26
4	0.216 70	0.216 48	0.217 60
6	0.216 72	0.216 70	0.216 87
8	0.216 72	0.216 72	0.216 75
10	0.216 72	0.216 72	0.216 73
12	0.216 72	0.216 72	0.216 72

In Tables IV and V we compare the convergence patterns for the two choices of the basis functions we considered, Eqs. (3.1) and (3.14), and for the two choices of the collocation points, Eqs. (4.5) and (4.6). The basis functions derived from the eigenfunctions of the problem and the Chebyshev collocation points consistently provide the best approximation. Moreover we see that a very small number of collocation points, as small as 6, is enough to reach convergence up to the fifth decimal place.

As far as the efficiency is concerned, the computations reported in Ref. 3 typically required a Cray-1 time of the order of 1 minute. Our computations required less than one second on the much slower IBM—3090 machine to reach convergence.

V. CONCLUDING REMARKS

The F_N approximate solution we developed for the two-group transport problem proved to be accurate and remarkably efficient. From this point we can see two different directions in which to proceed. From a theoretical point of view, it would be interesting to show the orthogonality and completeness of the eigenfunctions we derived [Eqs. (3.5) and (3.6)]. While proving the orthogonality should be, if not simple, feasable, proving the completeness of the set could be prohibitively complex, if possible at all. From a more applied point of view, it would be of interest to extend our analysis to finite ducts. In this case Eqs. (1.1) and (1.2) would hold in a region $0<x<L$, but Eq. (1.3) would be replaced by a prescribed distribution incident at the open end $x = L$. The derivation of the singular integral equation would be modified, using a finite Laplace transform. Moreover, one should obtain two coupled equations for the exiting distributions at the two open ends.

TABLE V. Convergence pattern with delta incidence $\mu_0 = 0.6$, $c = 0.6$.

Number of collocation points	Reflection probability Chebyshev quadrature; eigenfunction basis	Equal spacing; eigenfunction basis	Equal spacing; scalar basis
2	0.247 89	0.246 92	0.265 87
4	0.249 53	0.249 47	0.248 94
6	0.249 55	0.249 54	0.249 74
8	0.249 55	0.249 55	0.249 54
10	0.249 55	0.249 55	0.249 55

ACKNOWLEDGMENTS

The authors are grateful to C. E. Siewert for several very helpful discussions.

The work of the first author (F.M.) was supported by Science Applications International Corporation and that of the second author (G.C.P.) partially by the National Science Foundation.

APPENDIX: PRINCIPAL VALUE INTEGRALS

Full details concerning the way the system (3.18) was solved are given in Ref. 12. Here we only want to point out the way in which the principal value integrals were treated.

First, in order to know the matrix H, an explicit expression for $\mathbf{\Delta}_+(\nu)$, defined by (2.20) as

$$\mathbf{\Delta}_+(\nu) = \mathsf{I} - f_+\nu P\int_{-\infty}^{\infty} d\xi\,[(\xi^2+1)^2(\xi-\nu)\times(\xi-\kappa_+\nu)]^{-1}(f_+\nu\mathsf{S}^\dagger - \xi\mathsf{I})\mathsf{C}, \tag{A1}$$

has to be found. The function to be integrated is an analytic function everywhere in the complex ξ plane, except for four poles. Two of them, at $\xi = \pm i$, are second order, while the other two, one at $\xi = \kappa_+\nu$, the other on the real line at $\xi = \nu$, are simple. We can then think of closing the contour of integration in the lower half of the complex plane, where only the second-order pole at $\xi = -i$ is extant. Applying Cauchy's theorem we have that the integral in Eq. (A1) must equal the residue at $\xi = -i$ plus one half the residue at $\xi = \nu$, both taken with a negative sign because we are closing in the lower plane. Then we have

$$\mathbf{\Delta}_+(\nu) = \mathsf{I} - \pi i(\mathsf{R}_\nu + 2\mathsf{R}_{-i})\mathsf{C}, \tag{A2}$$

where, omitting the algebra, it is found that

$$\mathsf{R}_\nu = (\nu^2+1)^{-2}(1-\kappa_+)^{-1}(\mathsf{I} - f_+\nu\mathsf{S}^\dagger), \tag{A3}$$

$$\mathsf{R}_{-i} = \tfrac{1}{8}(i+\nu)^{-2}(i+\kappa_+\nu)^{-2}\{[4i\nu(\kappa_++1)-8]\mathsf{I} + f_+\nu[2\nu(\kappa_++1) + i(\kappa_++\nu+1)]\mathsf{S}^\dagger\}. \tag{A4}$$

Once H is known, it can be seen[12] that the integrations required in Eq. (3.17) are of the type

$$\mathrm{Int} = P\int_0^\infty d\xi(\alpha_1\xi + \alpha_2\xi^2 + \alpha_3\xi^3 + \alpha_4\xi^4 + \alpha_5\xi^5)\times\{(\xi^2+1)^2[(\xi-\kappa_1\nu_m)^2 + \kappa_2^2\nu_m^2]\times[(\xi+\kappa_1\nu_n)^2 + \kappa_2^2\nu_n^2](\xi-\nu_m)(\xi+\nu_n)\}^{-1}, \tag{A5}$$

where

$$\kappa_1 = \mathrm{Re}\,\kappa_+, \quad \kappa_2 = \mathrm{Im}\,\kappa_+, \tag{A6}$$

and the α_i's are just constants. Decomposing the integrand into simple fractions, we obtain

$$\mathrm{Int} = \lim_{r\to\infty}\Bigg\{\int_0^r d\xi(\beta_1\xi^3 + \beta_2\xi^2 + \beta_3\xi + \beta_4)(\xi^2+1)^{-2} + \int_0^r d\xi(\beta_5\xi + \beta_6)[(\xi-\kappa_1\nu_m)^2 + \kappa_2^2\nu_m^2]^{-1} + \int_0^r d\xi(\beta_7\xi + \beta_8)[(\xi+\kappa_1\nu_n)^2 + \kappa_2^2\nu_n^2]^{-1}$$

$$+\beta_9 \int_0^r d\xi(\xi+\nu_n)^{-1}+\beta_{10} P \int_0^r d\xi(\xi-\nu_m)^{-1}\Big\}, \tag{A7}$$

where the β_i's can be obtained from the α_i's from a 10×10 linear system, whose solution was obtained numerically. All the integrations can now be performed analytically, yielding

$$\begin{aligned}\mathrm{Int} = {} & (\pi/4)(\beta_2+\beta_4)+\tfrac{1}{2}(\beta_3-\beta_1)+(\pi/2\kappa_2)\\ & \times[(\beta_5-\beta_7)\kappa_1+(\beta_6/\nu_m+\beta_8/\nu_n)]+(1/\kappa_2)\\ & \times\tan^{-1}(\kappa_1/\kappa_2)[(\beta_5+\beta_7)\kappa_1+(\beta_6/\nu_m-\beta_8/\nu_n)]\\ & -\tfrac{1}{2}(\beta_5+\beta_7)\ln(\kappa_1^2+\kappa_2^2)-(\beta_7+\beta_9)\ln(\nu_n)\\ & -(\beta_5+\beta_{10})\ln(\nu_m).\end{aligned} \tag{A8}$$

[1]A. K. Prinja and G. C. Pomraning, Transp. Theory Stat. Phys. **13**, 567 (1984).
[2]E. W. Larsen, Transp. Theory Stat. Phys. **13**, 599 (1984).
[3]E. W. Larsen, F. Malvagi, and G. C. Pomraning, Nucl. Sci. Eng. **93**, 13 (1986).
[4]C. E. Siewert and P. Benoist, Nucl. Sci. Eng. **69**, 156 (1979).
[5]P. Grandjean and C. E. Siewert, Nucl. Sci. Eng. **69**, 161 (1979).
[6]R. D. M. Garcia, Transp. Theory Stat. Phys. **14**, 391 (1985).
[7]R. D. M. Garcia, G. C. Pomraning, and C. E. Siewert, Plasma Phys. **24**, 903 (1982).
[8]C. E. Siewert and J. R. Thomas, Jr., Nucl. Sci. Eng. **94**, 264 (1986).
[9]C. E. Siewert, Z. Ange. Math. Phys. **35**, 144 (1984).
[10]K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison–Wesley, Reading, MA, 1967).
[11]M. Abramowitz and I. Stegun, *Handbook of Mathematical Tables* (Natl. Bur. Stand., Washington, DC, 1964).
[12]F. Malvagi, Masters thesis, submitted to the University of California, Los Angeles, 1986.

Vacuum polarization in curved backgrounds deduced from Hadamard kernels

Mario A. Castagnino
Instituto de Física de Rosario, Av. Pellegrini 250, 2000 Rosario, Argentina and Instituto de Astronomía y Física del Espacio, Casilla de Correo 67-Sucursal 28, 1428 Buenos Aires, Argentina

Diego D. Harari and Carmen A. Núñez
Instituto de Astronomía y Física del Espacio, Casilla de Correo 67-Sucursal 28, 1428 Buenos Aires, Argentina

(Received 26 March 1985; accepted for publication 28 July 1986)

The point-splitting method of regularization is applied to alternative Hadamard kernels, for scalar fields in curved backgrounds, in order to study the polarization term in the vacuum expectation value (VEV) of the energy-momentum tensor. An ansatz to define this term is introduced and verified in several cases. The renormalized VEV of the energy-momentum tensor of a scalar massive conformally coupled field in a Robertson–Walker universe is explicitly computed. Two terms appear, the polarization term, which coincides with the one computed using the above-mentioned ansatz, and a creation of particle term.

I. INTRODUCTION

The aim of this paper is to study the different terms of the renormalized energy-momentum tensor, $\langle T_{\mu\nu}\rangle^{\rm ren}$, for a scalar quantum field in a curved background,[1] and the relation of these terms with the Hadamard structure of the Green's function $G(x,x') = \langle 0|\{\phi(x), \phi(x')\}|0\rangle$ ($\{\ ,\ \}$ is the anticommutator). There are several causes for a nonvanishing vacuum expectation value of the energy-momentum tensor, namely the presence of boundaries or nontrivial topologies,[2] the presence of horizons,[3] vacuum definition related to noninertial motion,[4] particle creation due to the evolution of the universe,[5] etc. In these, and other examples that can be found in the literature, a "vacuum-polarization" term appears. However, a precise and general definition of this concept is lacking, and there is only a unanimous agreement on the "local" nature of such a term. Also a "particle-creation" contribution is sometimes computed, and is said to be of nonlocal origin. Although the physical meaning of these terms is clear in each example examined, we lack a general definition. In a general situation, $\langle T_{\mu\nu}\rangle^{\rm ren}$ will be nonlocal. Thus, a division of $\langle T_{\mu\nu}\rangle^{\rm ren}$ into local and nonlocal terms (without a precise definition of these terms) is of course arbitrary, since a nonlocal plus a local term is still nonlocal.

To solve this problem and to single out the vacuum-polarization term, we will make a natural ansatz: we shall assume that the (local) polarization term of $\langle T_{\mu\nu}\rangle^{\rm ren}$ can be obtained from a propagator with the structure of a symmetric Hadamard solution, constructed entirely in terms of geometric quantities of the background at which $\langle T_{\mu\nu}\rangle^{\rm ren}$ is being evaluated. In fact, using the point-splitting regularization method,[6–9] $\langle T_{\mu\nu}\rangle$ can be evaluated as the coincidence limit of $\mathscr{D}_{\mu\nu'} G(x,x')$, where $\mathscr{D}_{\mu\nu'}$ is a bivector differential operator. The best and more general candidate for computing it from the polarization term of $\langle T_{\mu\nu}\rangle^{\rm ren}$ is a Hadamard kernel.[9–19] Indeed, any Hadamard solution of the Klein–Gordon equation is endowed, in a covariant way, with the same divergent structure as the flat-space kernel. Moreover, the divergences of any Hadamard kernel are of a geometric nature, and they can be absorbed by the bare constants of a generalized Hilbert–Einstein action for gravity (including terms quadratic in the curvature). Thus, a symmetric Hadamard kernel containing only geometric objects of the background might account perfectly well for the polarization term of $\langle T_{\mu\nu}\rangle^{\rm ren}$, i.e., $P_{\mu\nu} = \langle T_{\mu\nu}\rangle^{\rm ren}_{\rm Had}$.

We cannot prove that this ansatz works at any possible background geometry, but we will show in a few cases in which a trustable vacuum definition is available, at least in some regions of the background, that the true vacuum polarization and the result of our ansatz coincide. If, as we suspect, the same result can be found in more examples, then the method might have some physical relevance.

The paper is organized as follows: in Sec. II we briefly review some known facts about Hadamard elementary solutions and we present the formulas necessary to evaluate $\langle T_{\mu\nu}\rangle$ with the point-splitting method applied to the difference of two Hadamard elementary solutions. In fact, $\langle T_{\mu\nu}\rangle$ will be renormalized, as usual,[1,6,19,20] by subtraction of the infinities that appear when a particular Hadamard kernel, the de Witt–Schwinger expansion, is subjected to the point-splitting formula. This section partly duplicates some results of our previous paper,[15] to which we refer for more details, but that was only concerned with the trace of $\langle T_{\mu\nu}\rangle^{\rm ren}$.

In Secs. III–V we will construct the polarization term from symmetric Hadamard kernels built out at particular background geometries. In Sec. III we shall see that when Hadamard kernels constructed entirely in terms of the Riemann tensor and its derivatives are required to be symmetric only within conformally flat metrics, the resulting $P_{\mu\nu} = \langle T_{\mu\nu}\rangle^{\rm ren}_{\rm Had}$ is entirely determined, up to a term $m^2 G_{\mu\nu}$ (where $G_{\mu\nu}$ is the Einstein tensor and m the mass of the field). The result, in the massless case, is equivalent to that of Ref. 8, where similar requirements were imposed directly over $\langle T_{\mu\nu}\rangle^{\rm ren}$ instead of over the kernel. It leads, of course, to the usual trace anomaly. In Sec. IV we restrict ourselves to spatially flat Robertson–Walker universes, and we show that even if we try to enlarge the family of Hadamard symmetric

 0022-2488/87/010184-09$02.50

kernels by introducing quantities related to the comoving coordinate system, no change for $P_{\mu\nu}$ is obtained from the results of Sec. III. In Sec. IV B we then explicitly compute $\langle 0|T_{\mu\nu}|0\rangle^{\rm ren}$ in the vacuum states that instantaneously minimize the energy, for massive conformally coupled fields, with the result

$$\langle 0|T_{\mu\nu}(x)|0\rangle^{\rm ren} = P_{\mu\nu} + C_{\mu\nu}.$$

Here $P_{\mu\nu}$ is the same polarization term of Sec. III ($P_{\mu\nu} = \langle T_{\mu\nu}\rangle^{\rm ren}_{\rm Had}$) and $C_{\mu\nu}$ vanishes at the time when the energy is minimized (it can be taken as the particle-creation term). Thus Hadamard kernels that are symmetric at conformally flat metrics lead to the correct vacuum polarization in this simple case, in which the vacuum definition is well known.

In Sec. V we show that in a Schwarzschild background no symmetric Hadamard kernel can be constructed entirely in terms of the Riemann tensor. However, if that "purely geometric" condition is relaxed, we can construct a family of $\langle T_{\mu\nu}\rangle^{\rm ren}_{\rm Had}$ that contains the correct vacuum polarization in the Hartle–Hawking vacuum at the horizon. Some final conclusions are drawn in Sec. VI.

II. POINT-SPLITTING METHOD APPLIED TO HADAMARD KERNELS

A real scalar field arbitrarily coupled to the curvature of the background geometry is described by the action

$$S[\phi] = -\frac{1}{2}\int d^4x\sqrt{-g}\{\phi_{,\mu}\phi,^{\mu} + m^2\phi^2 + \xi R\phi^2\},$$

which is invariant under conformal transformations of the metric only if $m=0$ and $\xi = \frac{1}{6}$ (conformal coupling).[21] A variational principle leads to the Klein–Gordon equation

$$(\Box - m^2 - \xi R)\phi(x) = 0. \qquad (1)$$

The energy-momentum tensor operator is defined as the response of the field system with respect to metric variations

$$\begin{aligned} T_{\mu\nu} &\equiv (2/\sqrt{-g})(\delta S/\delta g^{\mu\nu}) \\ &= \tfrac{1}{2}(1-2\xi)\{\phi_{;\mu;\nu}\} + \tfrac{1}{2}(2\xi - \tfrac{1}{2})\{\phi_{;\rho}\phi_{;}{}^{\rho}\}g_{\mu\nu} \\ &\quad - \xi\{\phi_{;\mu\nu},\phi\} + \xi\{\Box\phi,\phi\}g_{\mu\nu} \\ &\quad + [\tfrac{1}{2}\xi(R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu}) - \tfrac{1}{4}m^2 g_{\mu\nu}]\{\phi,\phi\}, \end{aligned} \qquad (2)$$

where $\{\phi,\psi\} = \phi\psi + \psi\phi$ is the anticommutator. This tensor is traceless for massless conformally coupled fields.

The point-splitting method allows one to write the quantum vacuum-expectation value of $T_{\mu\nu}$ as the coincidence limit ($x\to x'$) of a differential operator applied to the biscalar kernel $G(x,x')$, defined as

$$G_1(x,x') = \langle 0|\{\phi(x),\phi(x')\}|0\rangle. \qquad (3)$$

Alternative definitions of the vacuum state (which is not a trivial matter in a curved background) are related to different boundary conditions for G.[22] The point-splitting expression for $\langle T_{\mu\nu}\rangle$ is obtained from Eq. (2) by changing in a symmetrized way one of the points from x to x' in each product of field operators at x, writing the result in terms of G and its x and x' derivatives, and finally by taking the coincidence limit. The formal (divergent) final expression is

$$\begin{aligned} \langle T_{\mu\nu}\rangle &= -\tfrac{1}{2}[G_{1;\mu\nu}] - \tfrac{3}{4}(\xi - \tfrac{1}{3})[\Box G_1]g_{\mu\nu} \\ &\quad + \tfrac{1}{2}(\tfrac{1}{2} - \xi)[G_1]_{;\mu\nu} + \tfrac{1}{2}(\xi - \tfrac{1}{4})\Box[G_1]g_{\mu\nu} \\ &\quad + \{\tfrac{3}{4}(\xi - \tfrac{1}{3})m^2 + \xi R)g_{\mu\nu} + \tfrac{1}{2}\xi R_{\mu\nu}\}[G_1], \end{aligned} \qquad (4)$$

where we have abbreviated the coincidence limit as

$$[G_1] = \lim_{x\to x'} G_1(x,x').$$

Equation (4) differs from its usual version in the literature because we have changed every x' derivative into x derivatives using Synge's theorem,[6] which, applied to a symmetric biscalar, leads to $[G_{1;\mu}] = \frac{1}{2}[G_1]_{;\mu}$, $[G_{1;\mu\nu}] = -[G_{1;\mu\nu}] + \frac{1}{2}[G_1]_{;\mu\nu}$, and $[G_{1;\mu'\nu'}] = [G_{1;\mu\nu}]$.

Now we shall assume that G is a Hadamard-type elementary solution of the Klein–Gordon equation (1)

$$G_1(x,x') = [\Delta^{1/2}(x,x')/8\pi^2]\{2/\sigma + v\ln\mu^2\sigma + w\}. \qquad (5)$$

In fact such $G_1(x,x')$ has the same kind of divergences as the $\Delta_1(x-x')$ of flat space-time, and therefore we can make the ansatz that the $G_1(x,x')$ which yields the local term of the energy-momentum tensor must be found in the family of Hadamard elementary solutions. Here σ is half the square of the geodesic distance between x and x', μ is an arbitrary mass scale (μ^2 has the same dimension as σ^{-1} if $h=c=1$), $\Delta(x,x')$ is the Van Vleck determinant, while $v(x,x')$ and $w(x,x')$ are given by

$$v(x,x') = \sum_{n=0}^{\infty} v_n(x,x')\sigma^n; \quad w(x,x') = \sum_{n=0}^{\omega} w_n(x,x')\sigma^n,$$

$$v_0 + v_{0;}{}^{\mu}\sigma_{;\mu} = V - \Delta^{-1/2}\Box(\Delta^{1/2}), \quad (V\equiv m^2 + \xi R), \qquad (6a)$$

$$v_n + \frac{1}{n+1}v_{n;}{}^{\mu}\sigma_{;\mu} = \frac{1}{2n(n+1)}\{Vv_{n-1} - \Delta^{-1/2}\Box(\Delta^{1/2}v_{n-1})\} \quad (n\geqslant 1), \qquad (6b)$$

$$\begin{aligned} w_n + \frac{1}{n+1}w_{n;}{}^{\mu}\sigma_{;\mu} &= \frac{1}{2n(n+1)}\{Vw_{n-1} - \Delta^{-1/2}\Box(\Delta^{1/2}w_{n-1})\} \\ &\quad - \frac{2n+1}{n(n+1)}v_n - \frac{1}{n(n+1)}v_{n;}{}^{\mu}\sigma_{;\mu} \quad (n\geqslant 1), \end{aligned} \qquad (6c)$$

where $\Delta(x,x')$ and $v(x,x')$ are univocally determined by the background geometry (their coincidence limits and those of some of their derivatives can be seen in Refs. 6, 11, and 15). Every election of the function $w_0(x,x')$ determines the complete $w(x,x')$ through Eq. (6c). For instance, a flat-space kernel $\Delta_1(x,x')$,

$$\Delta_1(x-x') = \frac{m^2}{4\pi}\,\mathrm{Im}\left\{\frac{H_1^{(1)}(\sqrt{2m^2\sigma})}{\sqrt{2m^2\sigma}}\right\},$$

can be shown to be the Hadamard solution characterized by

$$w_0{}^M = m^2(2\gamma - \ln 2 - 1), \quad \mu^2 = m^2,$$

where we use a superscript M to denote a "Minkowskian" value. Here γ is the Euler constant. The v_n^M and w_n^M satisfying the recurrence relations (6) are

$$v_n^M = \frac{2(m^2/2)^{n+1}}{n!(n+1)!},$$
$$w_n^M = -2\frac{(m^2/2)^{n+1}}{n!(n+1)!} \times\{\ln 2 + \psi(n+2) + \psi(n+0)\},$$

where $\psi(n)$ is the derivative of the Γ function's logarithm. The mass scale is, naturally, $\mu^2 = m^2$, and no problem arises in the massless limit, because v vanishes identically (there is no logarithmic term in the massless kernel in flat spacetime).

All the arbitrariness in any Hadamard solution is completely contained in $w_0(x,x')$ (except for changes in the mass scale μ). In other words, $w(x,x')$ (except for changes in the mass scale μ). In other words, $w(x,x')$ is the unique part of G_1 depending on boundary conditions. Define $G_1^W \equiv (\Delta^{1/2}/8\pi^2)\, w(x,x')$ as the "boundary" part of G_1, and apply the point-splitting expression (4) only to G_1^W instead of the complete G_1. The result is

$$\begin{aligned}16\pi^2\langle T_{\mu\nu}\rangle^W &= -\{[w_{0;\mu\nu}] - \tfrac{1}{4}g_{\mu\nu}[\Box w_0]\} + \tfrac{1}{3}\{[w_0]_{;\mu\nu} \\ &\quad - \tfrac{1}{4}g_{\mu\nu}\Box[w_0]\} - (\xi - \tfrac{1}{6})\{[w_0]_{;\mu\nu} - g_{\mu\nu}\Box[w_0]\} \\ &\quad + (\xi - \tfrac{1}{6})(R_{\mu\nu} - \tfrac{1}{4}g_{\mu\nu}R)[w_0] - \tfrac{1}{4}m^2[\omega_0]g_{\mu\nu} \\ &\quad + 9(\xi - \tfrac{1}{6})[v_1]g_{\mu\nu}. \end{aligned} \tag{7}$$

This formula does not give the complete $\langle T_{\mu\nu}\rangle$, but only the contribution that changes when the point-splitting method is applied to alternative Hadamard-type symmetric kernels. Contrary to Eq. (4), Eq. (7) gives finite results. Moreover, we will not need to know the complete $\langle T_{\mu\nu}\rangle$ in order to evaluate the renormalized stress tensor, because the renormalization recipe will consist in the subtraction of the infinities appearing when a particular Hadamard kernel is adopted [the de Witt–Schwinger representation, $G_1^{\rm DS}$ (Refs. 19, 20, and 23)]. Then, with $G_1^{\rm DS}$ being a Hadamard kernel,[15,20] we have

$$\begin{aligned}\langle T_{\mu\nu}\rangle^{\rm ren}_{\rm Had} &= \langle T_{\mu\nu}\rangle_{\rm Had} - {}^{(4)}\langle T_{\mu\nu}\rangle_{\rm DS} \\ &= \langle T_{\mu\nu}\rangle^W - {}^{(4)}\langle T_{\mu\nu}\rangle^W_{\rm DS}.\end{aligned} \tag{8}$$

Here ${}^{(4)}\langle T_{\mu\nu}\rangle_{\rm DS}$ denotes the expectation value of T evaluated from the de Witt–Schwinger kernel up to fourth order in the metric derivatives. The divergences appear only up to fourth order.

Here $G_1(x,x')$ must be, by definition, a symmetric kernel. Moreover, if the point-splitting expression were applied to a nonsymmetric kernel, the resulting $\langle T_{\mu\nu}\rangle$ would have nonzero covariant divergence.[13] For $w(x,x')$ to be a symmetric biscalar it can be shown, using the recurrence relations (6) and Synge's theorem, that $w(x,x')$ is restricted to verify[15]

$$\begin{aligned}[w_{0;\mu\nu}]_;{}^{\nu} - \tfrac{1}{4}[\Box w_0]_{;\mu} &= \tfrac{1}{4}\Box[w_0]_{,\mu} + \tfrac{1}{12}R_{\mu\nu}[w_0]^{,\nu} \\ &\quad + \tfrac{1}{4}(\xi - \tfrac{1}{6})R_{,\mu}[w_0] - \tfrac{1}{4}\{m^2 + (\xi - \tfrac{1}{6})R\} \\ &\quad \times[w_0]_{,\mu} + \tfrac{1}{2}[v_1]_{,\mu},\end{aligned} \tag{9}$$

with

$$\begin{aligned}[v_1] &= m^4/4 + (m^2/2)(\xi - \tfrac{1}{6})R + \tfrac{1}{4}(\xi - \tfrac{1}{6})R^2 \\ &\quad - \tfrac{1}{12}(\xi - \tfrac{1}{5})\Box R + \tfrac{1}{360}(R_{\theta\rho\tau\epsilon}R^{\theta\rho\tau\epsilon} - R_{\theta\rho}R^{\theta\rho} + \Box R).\end{aligned}$$

Any Hadamard elementary solution defined by a w satisfying condition (9), when inserted into the point-splitting expression (4), leads to a $\langle T_{\mu\nu}\rangle$ with zero covariant divergence. However, $\langle T_{\mu\nu}\rangle^W$ given by (7), which is not the complete $\langle T_{\mu\nu}\rangle$, will not necessarily have zero divergence. But the renormalized expression obtained through (8) will be covariantly conserved, because it will be the difference of two conserved expressions, $\langle T_{\mu\nu}\rangle$ and $\langle T_{\mu\nu}\rangle_{\rm DS}$. The de Witt–Schwinger kernel is such that, up to fourth order,[15]

$$\begin{aligned}[w_0^{\rm DS}] &= w_0{}^M + (\xi - \tfrac{1}{6})(2\gamma - \ln 2)R \\ &\quad + (1/m^2)\{T - \tfrac{1}{60}(2\gamma - \ln 2)\Box R\}, \\ [w_{0,\mu\nu}^{\rm DS}] &= (28 - \ln 2)\{[\tfrac{1}{3}(\xi - \tfrac{1}{6}) - \tfrac{1}{180}]R_{;\mu\nu} - \tfrac{1}{90}RR_{\mu\nu} \\ &\quad + \tfrac{1}{30}(R^{\theta\rho}R_{\mu\theta\nu\rho} + \tfrac{1}{2}\Box R_{\mu\nu})\},\end{aligned} \tag{10}$$

with

$$\begin{aligned}T &= \tfrac{1}{180}(R_{\theta\rho\tau\epsilon}R^{\theta\rho\tau\epsilon} - R_{\theta\rho}R^{\theta\rho}) \\ &\quad + \tfrac{1}{2}(\xi - \tfrac{1}{6})^2R^2 - \tfrac{1}{6}(\xi - \tfrac{1}{5})\Box R.\end{aligned} \tag{11}$$

Thus, inserting (10) in (7), we get

$$\begin{aligned}16\pi^2\langle T_{\mu\nu}\rangle^W_{\rm DS} &= -m^2(\xi - \tfrac{1}{6})(R_{\mu\nu} - \tfrac{1}{4}Rg_{\mu\nu}) + 9(\xi - \tfrac{1}{6})[v_1]g_{\mu\nu} \\ &\quad - \tfrac{1}{4}g_{\mu\nu}T + (2\gamma - \ln 2)[|\xi - \tfrac{1}{6}| \\ &\quad \times G_{\mu\nu} - \tfrac{1}{2}(\xi - \tfrac{1}{6})^2\,{}^{(2)}H - \tfrac{1}{180}({}^{(1)}H_{\mu\nu} - 3{}^{(2)}H_{\mu\nu})],\end{aligned} \tag{12}$$

where $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$ is the Einstein tensor and ${}^{(1)}H_{\mu\nu}$ and ${}^{(2)}H_{\mu\nu}$ are defined in the Appendix as the metric variations of the integrals of R^2 and $R_{\theta\rho}R^{\theta\rho}$, respectively. The quantity given by (12) will be the one subtracted in (8) in order to renormalize any $\langle T_{\mu\nu}\rangle_{\rm Had}$.

The kernel $G_1^{\rm DS}$ has the natural mass scale $\mu^2 = m^2$. But what happens if we subtract $G_1^{\rm DS}$ from a Hadamard kernel with a different mass scale? As quoted from Wald for the massless conformally coupled case,[13] it is the finite part of $\langle T_{\mu\nu}\rangle$ that changes when the mass scale is modified, and the change is proportional to the combination $3{}^{(2)}H_{\mu\nu} - {}^{(1)}H_{\mu\nu}$. We can deduce the change in $\langle T_{\mu\nu}\rangle$ in the general case, when the change $\mu \to \mu'$ is made, by application of the point-splitting expression (4) to the difference $G_1' - G_1 = -2(\Delta^{1/2}/8\pi^2)v(x,x')\ln(\mu'/\mu)$. The result is

$$\begin{aligned}16\pi^2\langle T'_{\mu\nu}\rangle^W &= 16\pi^2\langle T_{\mu\nu}\rangle^W - \ln(\mu'/\mu)[\tfrac{1}{180}(3{}^{(2)}{\rm H}_{\mu\nu} - {}^{(1)}{\rm H}_{\mu\nu}] \\ &\quad + \tfrac{1}{2}(\xi - \tfrac{1}{6})^2\,{}^{(1)}H_{\mu\nu} - (\xi - \tfrac{1}{6})m^2G_{\mu\nu} + \tfrac{1}{4}m^4g_{\mu\nu}].\end{aligned} \tag{13}$$

The ambiguity in $\langle T_{\mu\nu}\rangle$ due to mass-scale changes is proportional to $g_{\mu\nu}$, $G_{\mu\nu}$, ${}^{(1)}H_{\mu\nu}$, and ${}^{(2)}H_{\mu\nu}$. Thus, a mass-scale change only affects the finite renormalization of the bare constants multiplying λ, R, R^2, and $R_{\theta\rho}R^{\theta\rho}$ in the gravitational Lagrangian. Moreover, the scale dependence disappears in conformally flat metrics[9,13] where ${}^{(1)}H_{\mu\nu} = 3{}^{(2)}H_{\mu\nu}$ and in vacuum solutions of Einstein equations where $G_{\mu\nu} = {}^{(1)}H_{\mu\nu} = {}^{(2)}H_{\mu\nu} = 0$.

III. CONFORMALLY FLAT METRICS

We have shown in Ref. 15 that it is possible to construct Hadamard kernels for massless particles in terms of local geometric quantities that verify the symmetry condition at any conformally flat metric, although the same construction would not lead to a symmetric kernel in an arbitrary background geometry.

Here we shall compute the resulting $\langle T_{\mu\nu}\rangle^{\text{ren}}_{\text{Had}}$ applying (8) and (12) to such Hadamard kernels. The assumptions for the construction of w_0 are the following: (i) in the Minkowskian limit it should reduce to w_0^M; (ii) the massless limit should be well behaved, i.e., there should be no m in the denominator as in the de Witt–Schwinger expansion (10); and (iii) it should be built from the Riemann tensor, its derivatives, or contractions. We can then postulate the following expression:

$$[w_0] = w_0^M + AR,$$
$$[w_{0;\mu\nu}] = m^2BR_{\mu\nu} + C_1RR_{\mu\nu} + C_2R_{;\mu\nu} + C_3R_{\mu\theta}R^{\theta}{}_{\nu}. \tag{14}$$

We have also omitted a term in $\Box R_{\mu\nu}$ because it can be written in terms of $RR_{\mu\nu}$, $R_{;\mu\nu}$, and $R_{\mu\theta}R^{\mu\theta}$, in addition to irrelevant terms proportional to $g_{\mu\nu}$, as a consequence of the Gauss–Bonnet theorem (see the Appendix). Terms proportional to $g_{\mu\nu}$ are irrelevant because they belong in fact to w_1: when a covariant Taylor expansion of w_0 is made, a term in $g_{\mu\nu}$ would combine to $g_{\mu\nu}\sigma,{}^{\mu}\sigma,{}^{\nu} = 2\sigma$ (in Ref. 15 such terms were unnecessarily included, without any consequence to the final results).

When the symmetry condition (9) is imposed to (14) it reduces to a linear combination of the four independent "geometric variables" X_1–X_4 listed in the Appendix, which must vanish. Thus the coefficients at (14) turn to be restricted by

$$B = -A, \quad C_1 = -C_3 = -\tfrac{1}{180}, \quad C_2 = \tfrac{1}{540} + \tfrac{1}{3}A.$$

Only A remains as an arbitrary coefficient. The introduction of this result in Eq. (7) gives

$$\begin{aligned}2880\pi^2\langle T_{\mu\nu}\rangle^W &= 180Am^2G_{\mu\nu} + RR_{\mu\nu} - \tfrac{1}{3}R_{;\mu\nu}\\ &\quad - R_{\mu\theta}R^{\mu\theta} + \tfrac{1}{48}g_{\mu\nu}(R_{\theta\rho}R^{\theta\rho} - R^2 + \tfrac{1}{3}\Box R).\end{aligned} \tag{15}$$

A unique local term is obtained for the massless case, independently of the value of A, and the ambiguity for the massive case is again proportional to a tensor also appearing at the left-hand side (lhs) of Einstein equations.

We can now renormalize expression (15), using (8) and (13). No ambiguity due to mass-scale changes exists for conformal coupling and conformally flat metrics (conformal triviality). Then

$$\begin{aligned}P_{\mu\nu} &= \langle T_{\mu\nu}\rangle^{\text{ren}}_{\text{Had}}\\ &= (1/2880\pi^2)[180Am^2G_{\mu\nu} + RR_{\mu\nu} - \tfrac{1}{3}R_{;\mu\nu}\\ &\quad - R_{\mu\theta}R^{\theta}{}_{\nu} + \tfrac{1}{48}g_{\mu\nu}(2R_{\theta\rho}R^{\theta\rho} + \tfrac{4}{3}\Box R - \tfrac{4}{3}R^2)],\end{aligned} \tag{16}$$

and A is still arbitrary. The trace is, of course, the usual trace anomaly in the massless case[24]

$$\begin{aligned}P_{\mu}{}^{\mu} &= (1/2880\pi^2)\\ &\quad\times[-180m^2AR + R_{\theta\rho}R^{\theta\rho} + \Box R - \tfrac{1}{3}R^2].\end{aligned} \tag{17}$$

Thus, the restriction to conformal triviality (in the massless case) determines a unique polarization term in the energy-momentum tensor, according with our ansatz.

IV. COMOVING OBSERVER IN ROBERTSON–WALKER UNIVERSES

A. Construction of $P_{\mu\nu}$ in the massless case

Now we shall consider another kind of possible construction of $\langle T_{\mu\nu}\rangle^{\text{ren}}_{\text{Had}}$: we shall do it in a particular coordinate system, introducing a priviliged decomposition in spacelike and timelike directions. Consider the comoving coordinate system in Robertson–Walker universes, where the arc length reads

$$\begin{aligned}dS^2 &= -dt^2 + a^2(t)(dx^2 + dy^2 + dz^2)\\ &= a^2(\eta)(-d\eta^2 + dx^2 + dy^2 + dZ^2).\end{aligned}$$

The metric is conformally flat in the conformal-time variable defined by $d\eta = dt/a(t)$. We have shown in Ref. 15 that a family of Hadamard symmetric kernels bigger than the one obtained for any conformally flat background geometry can be constructed by Robertson–Walker metrics for massless conformally coupled particles if the comoving coordinate system is privileged. It is given by

$$[w_0] = A_1\alpha_1 + A_2\alpha_2,$$
$$[w_{0;0}{}^{0}] = \sum_{n=1}^{5} T_n\beta_n; \quad [w_{0;ij}] = g_{ij}\sum_{n=1}^{5} S_n\beta_n,$$

where

$$\begin{aligned}&\alpha_1 = H^2, \quad \alpha_2 = \tfrac{1}{6}R, \quad \beta_1 = \alpha_1^2; \quad \beta_2 = \alpha_2^2;\\ &\beta_3 = \alpha_1\alpha_2; \quad \beta_4 = H\dot\alpha_2; \quad \beta_5 = \ddot\alpha_2,\end{aligned} \tag{18}$$

where H is the Hubble constant, $H = \dot a/a$ (the dot means time derivative). The scalar curvature is $R = 6(\dot H + 2H^2)$. Latin indices $(i,j = 1,2,3)$ denote spatial components. Here A_1 and A_2 may be chosen at will, and the differences $D_n = T_n - S_n$ $(n = 1,2,...,5)$ must verify

$$\begin{aligned}&D_1 = \tfrac{4}{90} - \tfrac{28}{3}A_1, \quad D_2 = -\tfrac{2}{90} - \tfrac{2}{3}A_1, \quad D_3 = \tfrac{2}{90} + 6A_1,\\ &D_4 = \tfrac{1}{90} - \tfrac{4}{3}A_1 + \tfrac{1}{3}A_2, \quad D_5 = -\tfrac{1}{90} - \tfrac{1}{3}A_2,\end{aligned}$$

in order to satisfy the symmetry condition (9). Equation (7) gives in this case, for the (0,0) component of $\langle T_{\mu\nu}\rangle^W$,

$$\begin{aligned}16\pi^2\langle T_{00}\rangle^W &= \frac{3}{4}D_n\beta_n + \frac{1}{4}\left\{\frac{d^2}{dt^2}[w_0] - H\frac{d}{dt}[w_0]\right\}\\ &= \tfrac{1}{180}(6\beta_1 - 3\beta_2 + 3\beta_3 + \tfrac{3}{2}\beta_4 - \tfrac{3}{2}\beta_5).\end{aligned}$$

Taking into account that $\langle T_{\mu}{}^{\mu}\rangle^W = 0$, then $\langle T_{ij}\rangle^W = g_{ij}\langle T_{00}\rangle^W$. Thus, in spite of all the arbitrariness in A_1, A_2, and T_n (for given S_n), the resulting $\langle T_{\mu\nu}\rangle^B$ is unique. It is the same as the one given by (15), which is valid for any conformally flat metric, in particular for R-W universes. This can be verified by explicit evaluation of the (0,0) component of (15), using the values listed in the Appendix for quantities such as RR_{00}, etc. Renormalization, obtained by subtraction of

$$16\pi^2\, {}^{(4)}\langle T_{\mu\nu}\rangle^W_{\rm DS} = -\tfrac{1}{4} g_{\mu\nu}\tfrac{1}{180}\{12\beta_1 - 12\beta_3 - 18\beta_4 - 6\beta_5\},$$

gives, of course, the same result as (16). It is given, in components, by

$$P_{00} = \langle T_{00}\rangle^{\rm ren}_{\rm Had} = (1/2880\pi^2)[3\beta_1 - 3\beta_2 + 6\beta_3 + 6\beta_4],$$

$$P_{ij} = \langle T_{ij}\rangle^{\rm ren}_{\rm Had} = (1/2880\pi^2)\tfrac{1}{3}g_{ij}[15\beta_1 - 3\beta_2 - 6\beta_3 - 12\beta_4 - 6\beta_5]. \tag{19}$$

Thus, the potential dependence of $\langle T_{\mu\nu}\rangle^{\rm ren}_{\rm Had}$ on quantities like H related to the comoving reference system does not change nor does it enlarge the purely geometric result obtained at Sec. III for any conformally flat metric. Morevoer, the result (19) is the true vacuum polarization in the conformal vacuum,[8] a very reliable definition in the massless, conformally coupled case.

B. Hamiltonian diagonalization

It is interesting to make a connection between the result given at (19), which was obtained via the point-splitting method applied to Hadamard propagators, with the evaluation of $\langle T_{\mu\nu}\rangle$ via a definition of the vacuum state through a normal-mode decomposition of the field. Renormalization will be made through adiabatic regularization,[25] a method which allows cancellation of the infinites before integrating the normal modes. Moreover, we shall see terms representing massive particle creation appear in $\langle T_{\mu\nu}\rangle^{\rm ren}$.

The real scalar field can be decomposed in normal modes as follows

$$\phi(x) = \int d^3\mathbf{k}\,\frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{(2\pi a)^{3/2}}\{a_{\mathbf{k}} f_k(t) + a^+_{-\mathbf{k}} f^*_k(t)\}, \tag{20}$$

where f_k and f^*_k must verify the Wronskian condition

$$\dot f^*_k f_k - f^*_k \dot f_k = i,$$

which can be automatically satisfied, for any real $\Omega_k(t)$, writing[26]

$$f_k(t) = \frac{\exp\{i\int^t \Omega_k(t')dt'\}}{[2\Omega_k(t)]^{1/2}}. \tag{21}$$

In order to make a solution of Klein–Gordon, Eq. (1), Ω_k must satisfy

$$\frac{1}{2}\frac{\ddot\Omega_k}{\Omega_k} - \frac{3}{4}\left(\frac{\dot\Omega_k}{\Omega_k}\right)^2 + \Omega_k^2 = \omega_k^2 + \left(\xi - \frac{1}{4}\right)R + \frac{3}{4}H^2, \tag{22}$$

where $w_k^2 = k^2/a^2 + m^2$. A vacuum state can be defined by

$$a_{\mathbf{k}}|0\rangle = 0, \quad \forall\mathbf{k}.$$

Obviously, this definition depends upon the normal modes f_k used in (20). Taking the vacuum expectation value of the classical expression (2) for $T_{\mu\nu}$ gives

$$\begin{aligned}
\langle T_{00}\rangle &= \frac{1}{2}\int\frac{d^3\mathbf{k}}{(2\pi a)^3}\Big\{|fk|^2\Big[w_k^2 + \Big(\frac{9}{4} - 12\xi\Big)H^2\Big] + |\dot f_k|^2 + 4\Big(3\xi - \frac{3}{4}\Big)H\,\mathrm{Re}\{\dot f_k f^*_k\}\Big\}\\
&= \int\frac{d^3\mathbf{k}}{(2\pi a)^3}\frac{1}{4\Omega_k}\Big\{\Omega_k^2 + \omega_k^2 + \Big(\frac{9}{4} - 12\xi\Big)\\
&\quad\times H^2 - \Big(6\xi - \frac{3}{2}\Big)H\frac{\dot\Omega_k}{\Omega_k} + \frac{1}{4}\Big(\frac{\dot\Omega_k}{\Omega_k}\Big)^2\Big\},\\
\langle T_{ij}\rangle &= g_{ij}\int\frac{d^3\mathbf{k}}{(2\pi a)^3}\Big\{|f_k|^2\Big[\frac{1}{3}\frac{k^2}{a^2} + \Big(2\xi - \frac{1}{2}\Big)\omega_k^2 + 2\xi\Big(\xi - \frac{1}{6}\Big)R + \Big(\frac{9}{8} - \frac{13}{2}\xi\Big)H^2\Big]\\
&\quad - \Big(2\xi - \frac{1}{2}\Big)|\dot f_k|^2 + \Big(8\xi - \frac{3}{2}\Big)H\,\mathrm{Re}\{\dot f_k f^*_k\}\Big\}\\
&= \frac{1}{3}g_{ij}\int\frac{d^3\mathbf{k}}{(2\pi a)^3}\frac{1}{2\Omega_k}\Big\{(\omega_k^2 - m^2) + \Big(6\xi - \frac{3}{2}\Big)\\
&\quad\times(\omega_k^2 - \Omega_k^2) + \xi(6\xi - 1)R + 3\Big(\frac{9}{8} - \frac{13}{2}\xi\Big)\\
&\quad\times H^2 - 3\Big(4\xi - \frac{3}{4}\Big)H\frac{\dot\Omega_k}{\Omega_k} - \frac{3}{4}\Big(2\xi - \frac{1}{2}\Big)\Big(\frac{\dot\Omega_k}{\Omega_k}\Big)^2\Big\}.
\end{aligned} \tag{23}$$

The dependence of Eq. (23) upon f_k (or Ω_k) reveals the dependence upon the quantum state in which the expectation value is evaluated. A possible way to define the quantum state is through boundary conditions for the function Ω_k at a particular time τ, i.e., giving as data $\Omega_k(\tau)$ and $\dot\Omega_k(\tau)$. A physical criterium, which has often been suggested, is Hamiltonian diagonalization, which coincides with Hamiltonian minimization between alternative normal modes.[17,18,27–30] The vacuum state $|0\rangle_\tau$, which makes the Hamiltonian a minimum at $t = \tau$, leads, through Eq. (3), to a propagator $G_1^{(\tau)}(x,x')$ which is not a Hadamard-type solution.[17,30,31] However, it leads to the same divergences in the energy-momentum tensor as any Hadamard solution for conformal coupling (and only for this kind of coupling).[31] We shall verify this fact explicitly. (In Ref. 17 it is shown that the $G_1(x,x')$ from Hamiltonian diagonalization is not a Hadamard kernel, and the term that originates the difference is singled out [Eq. (2.34)]. The authors conclude that the renormalized VEV of the energy-momentum tensor in the vacuum of Hamiltonian diagonalization will be divergent. However, as can be seen from Eq. (78) of Ref. 9, the term that originates the difference does not contribute with divergences to the energy-momentum tensor. Thus, the author's conclusion is not proved. In fact, to have a Hadamard structure is a sufficient condition for a propagator to yield a finite energy-momentum tensor when renormalized by subtraction of $G_1^{\rm DS}$, but it is not a necessary one, as the example of this section shows.) Thus, we shall restrict ourselves from now on to conformally coupled particles. The Hamiltonian

$$H(t) = \int d^3\mathbf{x}\sqrt{-g}\langle T_{00}(t,\mathbf{x})\rangle,$$

is minimized with respect to alternative Cauchy data $\Omega_k(\tau)$ and $\dot\Omega_k(\tau)$ in Eq. (23) if

$$\Omega_k(\tau)=\omega_k(\tau),\quad \dot{\Omega}_k(\tau)/\Omega_k(\tau)=-H(\tau)\quad (\xi=\tfrac{1}{6}). \tag{24}$$

In the massless conformally coupled case these conditions are satisfied, at any t, by $\Omega_k=k/a$, which is a solution of Eq. (22). The corresponding normal modes are the natural generalization of the positive frequencies of flat space, written in terms of the conformal time. Positive and negative frequencies so defined do not mix during the universe evolution, and then particle creation does not occur.[5] Massive particle creation does occur. We can evaluate the contribution of the created particles to the energy-momentum tensor by writing the normal modes which define the vacuum of minimum energy at $t=\tau$ as

$$\begin{aligned} f_k^{(\tau)}(t)&=\alpha_k(\tau,t)\,\frac{\exp\{i\int^t\omega_k(t')dt'\}}{\sqrt{2\omega_k}}\\ &\quad+\beta_k(\tau,t)\,\frac{\exp\{i\int^t\omega_k(t')dt'\}}{\sqrt{2\omega_k}},\\ \dot{f}_k^{(\tau)}(t)&=i\omega_k\bigg[\alpha_k(\tau,t)\,\frac{\exp\{i\int^t\omega_k(t')dt'\}}{\sqrt{2\omega_k}}\\ &\quad+\beta_k(\tau,t)\,\frac{\exp\{-i\int^t\omega_k(t')dt'\}}{\sqrt{2\omega_k}}\bigg]\\ &\quad+\tfrac{1}{2}Hf_k^{(\tau)}(t), \end{aligned} \tag{25}$$

with

$$\alpha_k(\tau,\tau)=1,\quad \beta_k(\tau,\tau)=0.$$

These boundary conditions on α_k and β_k are adequate to make the boundary conditions (24) hold. The Wronskian condition implies

$$|\alpha_k(\tau,t)|^2-|\beta_k(\tau,t)|^2=1,$$

and the Klein–Gordon equation (1) is satisfied if α_k and β_k obey the following coupled first-order equations[23]:

$$\dot{\alpha}_k=(m^2H/2\omega_k^2)\beta_k\exp\Big(-2i\int^t\omega_k(t')dt'\Big),$$

$$\dot{\beta}_k=(m^2H/2\omega_k^2)\alpha_k\exp\Big(2i\int^t\omega_k(t')dt'\Big).$$

Let us introduce (25) in Eq. (23) for $\langle T_{\mu\nu}\rangle$

$$\begin{aligned} {}_\tau\langle T_{00}(t)\rangle_\tau&=\int\frac{d^3\mathbf{k}}{(2\pi a)^3}\frac{1}{2}\omega_k\{1+2|\beta_k(\tau,t)|^2\},\\ {}_\tau\langle T_{ij}(t)\rangle_\tau&=\frac{1}{3}g_{ij}\int\frac{d^3\mathbf{k}}{(2\pi a)^3}\frac{1}{2}\omega_k\bigg\{\Big(1-\frac{m^2}{\omega_k^2}\Big)\\ &\quad\times(1+2|\beta_k(\tau,t)|^2)\\ &\quad-\frac{2}{H}\frac{d}{dt}(|\beta_k(\tau,t)|^2)\bigg\}. \end{aligned} \tag{26}$$

This expression leads, after integration, to an infinite result. We can get finite answers by adiabatic regularization,[25] which means to fix, as the (time-dependent) zero-energy point, the energy of the vacuum state defined by an adiabatic expansion of the normal modes, i.e., by the normal modes given by the solution of Eq. (22) for Ω_k obtained by an iterative procedure. Moreover, it has been shown that the adiabatic expansion of normal modes leads through (3) to a propagator that, up to fourth order in the metric derivatives, coincides with the de Witt–Schwinger representation $G_1^{\rm DS}$.[32,33] Then adiabatic regularization must coincide with the subtraction prescription (8). In this case we cannot use Eq. (8) to renormalize $\langle T_{\mu\nu}\rangle$ because the $G_1(x,x')$ of the Hamiltonian diagonalization is not a Hadamard structure. The adiabatic solution of Eq. (22) neglecting derivatives of the metric higher than the fourth is

$$\begin{aligned} \Omega_k(t)&=\omega_k(t)\bigg\{1+\frac{1}{\omega_k^2}\Big[\frac{1}{2}(6\xi-1)\alpha_2\Big]\\ &\quad-\frac{1}{4\omega_k^4}\Big[m^2(\alpha_1+\alpha_2)+\frac{1}{2}(6\xi-1)^2\beta_2\\ &\quad+16\xi-1)(\beta_2+\beta_3+\frac{5}{2}\beta_4+\frac{1}{2}\beta_5)\Big]\\ &\quad+\frac{m^2}{\omega_k^6}\Big[\frac{5}{8}m^2\alpha_1+\frac{1}{4}\Big\{9\xi\beta_2-(2-39\xi)\beta_3\\ &\quad+\Big(15\xi+\frac{1}{4}\Big)\beta_4+\frac{1}{4}\beta_5\Big\}\Big]-\frac{1}{4}\frac{m^4}{\omega_4^8}\\ &\quad\times\Big[\frac{19}{8}(\beta_1+\beta_2)+\frac{1}{2}(75\xi+39)\beta_3+\frac{7}{2}\beta_4\Big]\\ &\quad+\frac{221}{32}\frac{m^6}{\omega_k^{10}}(\beta_1+\beta_3)-\frac{1105}{128}\frac{m^8}{\omega_k^{12}}\beta_1\bigg\},\\ \frac{\dot{\Omega}_k}{\Omega_k}&=-H+\frac{1}{\omega_k^2}\Big[m^2H+(6\xi-1)(H\alpha_2+\frac{1}{2}\dot{\alpha}_2)\Big]\\ &\quad-\frac{m^2}{\omega_k^4}\Big[\Big(6\xi+\frac{1}{2}\Big)H\alpha_2+\frac{1}{4}\dot{\alpha}_2\Big]\\ &\quad+\frac{m^4}{\omega_k^6}\frac{9}{4}H(\alpha_1+\alpha_2)-\frac{15}{4}\frac{m^6}{\omega_k^8}H\alpha_1. \end{aligned} \tag{27}$$

Inserting (27) in Eq. (23) and computing all the integrals that give finite results, we obtain the "adiabatic vacuum expectation value" of $T_{\mu\nu}$, up to fourth order (and for $\xi=\frac{1}{6}$)

$$\langle T_{00}\rangle_{\rm Ad}^{(4)}=\int\frac{d^3\mathbf{k}}{(2\pi a)^3}\frac{1}{2}\omega_k-P_{00},$$

$$\langle T_{ij}\rangle_{\rm Ad}^{(4)}=\frac{1}{3}g_{ij}\int\frac{d^3\mathbf{k}}{(2\pi a)^3}\frac{1}{2}\omega_k\Big(1-\frac{m^2}{\omega_k^2}\Big)-P_{ij},$$

where

$$\begin{aligned} 2880\pi^2P_{00}&=-30m^2\alpha_1+3\beta_1-3\beta_2+6\beta_3+6\beta_4,\\ 2880\pi^2P_{ij}&=\tfrac{1}{3}g_{ij}[-30m^2(\alpha_1-2\alpha_2)+15\beta_1\\ &\quad-3\beta_2-6\beta_3-12\beta_4-6\beta_5]. \end{aligned} \tag{28}$$

Now the renormalization ${}_\tau\langle T_{\mu\nu}\rangle_\tau^{\rm ren}={}_\tau\langle T_{\mu\nu}\rangle_\tau-\langle T_{\mu\nu}\rangle_{\rm Ad}^{(4)}$ is made:

$$\begin{aligned} {}_\tau\langle T_{00}(t)\rangle_\tau^{\rm ren}&=P_{00}(t)+\int\frac{d^3\mathbf{k}}{(2\pi a)^3}|\beta_k(\tau,t)|^2\omega_k,\\ {}_\tau\langle T_{ij}(t)\rangle_\tau^{\rm ren}&=P_{ij}(t)+\frac{1}{3}g_{ij}\int\frac{d^3\mathbf{k}}{(2\pi a)^3}\omega_k\bigg\{\Big(1-\frac{m^2}{\omega_k^2}\Big)\\ &\quad\times|\beta_k(\tau,t)|^2-\frac{1}{H}\frac{d}{dt}|\beta_k(\tau,t)|^2\bigg\}. \end{aligned} \tag{29}$$

Here two different contributions are clearly identified. A vacuum-polarization term, $P_{\mu\nu}$, which depends only upon the local geometry at the instant t, and the particle creation term $C_{\mu\nu}$ [the remainder of Eq. (29)], which depends upon boundary conditions at $t = \tau$, and also upon the evolution of the radius of the universe between τ and t,

$$C_{00}(t) = \int \frac{d^3\mathbf{k}}{(2\pi a)^3} |\beta_k(\tau,t)|^2 \omega_k = \rho(t),$$

$$C_{ij}(t) = \frac{1}{3} g_{ij} \int \frac{d^3\mathbf{k}}{(2\pi a)^3} \omega_k \left\{ \left(1 - \frac{m^2}{\omega_k^2}\right) |\beta_k(\tau,t)|^2 - \frac{1}{H}\frac{d}{dt} |\beta_k(z,t)|^2 \right\}.$$

It can be easily verified that both contributions, $P_{\mu\nu}$ and $C_{\mu\nu}$, are covariantly conserved, i.e., $P_{\mu\nu;}{}^{\nu} = C_{\mu\nu;}{}^{\nu} = 0$.

Here $P_{\mu\nu}$ is the vacuum polarization in the sense that if we take $t = \tau$ in Eq. (29) we obtain $P_{\mu\nu} = {}_\tau\langle 0|T_{\mu\nu}(\tau)|0\rangle_\tau^{\text{ren}}$. Note that the vacuum-polarization term obtained through adiabatic regularization of the normal modes which produces an instantaneous minimization of the energy is the same as the one arrived at through the point-splitting method applied to the most general Hadamard geometric kernel in the conformally-flat case [compare Eqs. (28), (16), and (19)]. That is to say,

$$P_{\mu\nu} = \langle T_{\mu\nu} \rangle_{\text{Had}}^{\text{ren}}.$$

The massive term at (16), originated from Hadamard kernels, is not univocally determined, and the coincidence of (16) with (28) is obtained taking $A = -\frac{1}{18}$. In other words, the polarization term of the vacuum states of minimum instantaneous energy coincide with the local term obtained from the Hadamard kernel having $[w_0] = w_0^M - R/18$. In Ref. 15 we have shown that the conformal kernel for massless particles (the one which originated as a conformal transform of the flat-space kernel) is precisely characterized by $[w_0] = -R/18$ ($w_0^M = 0$ for massless fields).

V. SCHWARZSCHILD BACKGROUND

Now consider a Schwarzschild background geometry, where the arc length reads

$$dS^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \frac{dr^2}{1 - 2M/r} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2).$$

No polarization term exists for the energy-momentum tensor of a massless conformally coupled field that can be obtained from Hadamard kernels determined by the background geometry exclusively. Indeed, taking into account that the Schwarzschild solution is a vacuum solution of Einstein equations ($R = R_{\mu\nu} = 0$), the unique fourth-order purely geometric quantity to be included in $[w_{0;\mu\nu}]$ is $R_{\mu\theta\rho\tau} R^{\mu\theta\rho\tau} = \frac{1}{4} g_{\mu\nu} C_{\theta\rho\tau\epsilon} C^{\theta\rho\tau\epsilon}$. However, terms proportional to $g_{\mu\nu}$ are irrelevant. So $[w_0] = [w_{0;\mu\nu}] = 0$ is the most general purely geometric construction, but it does not verify the symmetry condition (9) because

$$360[v_1] = C_{\theta\rho\tau\epsilon} C^{\theta\rho\tau\epsilon} \equiv C^2 = 48(M^2/r^6),$$

and then $[v_1]_{,r} \neq 0$. Another kind of construction of the Hadamard kernels must be done in order to obtain possible polarization terms for the energy-momentum tensor. We shall show that symmetric Hadamard kernels exist in a Schwarzschild background if the purely geometric restriction is given up. We shall construct them in a particular reference frame, i.e., (t,r,θ,φ). Any two-point function, sharing the symmetries of the Schwarzschild geometry, satisfies

$$[\omega_{0;t}{}^{t}] = T(r), \quad [\omega_{0;r}{}^{r}] = R(r),$$
$$[\omega_{0,\phi}{}^{\phi}] = [\omega_{0;\theta}{}^{\theta}] = A(r). \tag{30}$$

The renormalized energy-momentum tensor computed as

$$\langle T_{\mu\nu}\rangle_{\text{Had}}^{\text{ren}} = \langle T_{\mu\nu}\rangle^{W} - \langle T_{\mu\nu}\rangle_{\text{DS}}^{W} = \langle T_{\mu\nu}\rangle^{W} + \tfrac{1}{4} g_{\mu\nu}(C^2/2880\pi^2)$$

is then given by

$$P_t{}^{t} = \langle T_t{}^{t}\rangle_{\text{Had}}^{\text{ren}} = -\frac{1}{32\pi^2}(I + J) + \frac{1}{4}\frac{c^2}{2880\pi^2},$$

$$P_t{}^{t} = \langle T_r{}^{r}\rangle_{\text{Had}}^{\text{ren}} = -\frac{1}{32\pi^2}(I - J) + \frac{1}{4}\frac{c^2}{2880\pi^2},$$

$$P_\theta{}^{\theta} = P_\phi{}^{\phi} = \langle T_\theta{}^{\theta}\rangle_{\text{Had}}^{\text{ren}} = \langle T_\phi{}^{\phi}\rangle_{\text{Had}}^{\text{ren}} = \frac{1}{32\pi^2} I + \frac{1}{4}\frac{c^2}{2880\pi^2},$$

where

$$I \equiv \tfrac{1}{2}(R + T) - A, \quad J = T - R.$$

The symmetry condition (9) is equivalent to the following differential condition on I and J:

$$\frac{1}{2}(\dot I - \dot J) + \frac{2}{r}\left(I - \frac{1}{2}J\right) - \frac{M}{r^2(1 - 2M/r)} J = -\frac{2}{5}\frac{M^2}{r^7} \quad \left(\cdot = \frac{d}{dr}\right).$$

A possible solution (which gives $P_t{}^{t} = P_r{}^{r}$) is

$$I = (2M^2/5r^6)(1 - \alpha/r^4), \quad J = 0,$$

with α an arbitrary constant.

There are many possible polarization terms constructed from Hadamard kernels with Schwarzschild symmetries (30) in a Schwarzschild background. It is possible that some of them can be written in a geometrized way, using perhaps some other geometric quantities of the Schwarzschild geometry, such as Killing vectors, in the construction of ω_0. For the moment, we have just proved the possibility of constructing symmetric Hadamard kernels for massless fields in Schwarzschild geometry. We can also mention that one member of the family (the one with $\alpha = -\frac{43}{720}$) is such that $P_{\mu\nu}$ evaluated at the horizon ($r = 2M$) coincides with the value of ${}_{\text{HH}}\langle 0|T_{\mu\nu}|0\rangle_{\text{HH}}$ at the horizon computed in Ref. 35, where $|0\rangle_{\text{HH}}$ denotes the Hartle–Hawking vacuum state. The Hartle–Hawking state can be considered a good vacuum definition at the horizon (but not necessarily at any other region), so its value at $r = 2M$ can be trusted as the vacuum polarization there.

VI. CONCLUSIONS

We have made an explicit construction at some particular backgrounds of the polarization term of the quantum expectation values of the stress-energy tensor, which originates in the application of the point-splitting method to Hadamard-type kernels. Nonlocal contributions, such as those appearing when particle creation occurs, are not included in this kind of calculation.

When a normal-mode decomposition of the field is made, the vacuum state is generally defined through boundary conditions for the modes (for example, requiring Hamiltonian diagonalization over certain hypersurfaces, or requiring the modes to be of "positive-frequency type" with respect to a particular timelike direction). We have considered an example: the quantum states for scalar conformally coupled fields in Robertson–Walker universes defined through instantaneous Hamiltonian diagonalization. Equation (29) suggests the separation of ${}_\tau\langle T_{\mu\nu}\rangle_\tau^{\rm ren}$ into two clearly different contributions. One is defined in terms of purely geometric quantities: the polarization $P_{\mu\nu} = \langle 0|T_{\mu\nu}(\tau)|0\rangle$. The other one, $C_{\mu\nu}$, is nonlocal and depends on the particular time chosen to diagonalize the Hamiltonian. The interesting result is that $P_{\mu\nu}$ can also be computed as a polarization term originated from symmetric and geometric Hadamard kernels in conformally flat metrics, which is a completely different approach to P than the normal modes identification of the quantum states. As another example we have studied the polarization tensor at the horizon in a Schwarzschild background. We have shown that it is possible to construct Hadamard kernels in that geometry. However, there is a difference in the prescription used to construct the functions $[w_0]$ and $[w_{0;\mu\nu}]$ in the Robertson–Walker and Schwarzschild metrics. In the former case we used functionals of the metric and its derivatives, whereas in the latter we considered general functions, sharing only the symmetries with the metric. Although we realize that it would be interesting to have a unique prescription, we believe we have somehow clarified the nature of the polarization term. In fact, until now this term defined only as a "local" term had no precise meaning. We have constructed reasonable polarization terms that coincide with the ones obtained by other techniques. We have proved that these terms can also be computed with our Hadamard ansatz using local objects related to a particular coordinate system or a particular property of the metric. We believe that these reference systems have a privileged role because they could be considered as the observers' system. We begin to study this approach in Refs. 35 and 36 for the vacuum definition. The study of the polarization terms, under this view, will be the subject of forthcoming papers.

ACKNOWLEDGMENTS

Two of the authors (M.A.C.) and (D.D.H.) were fellows of The Consejo Nacional de Investigaciones Científicas y Técnicas.

C.A.N. was supported by a Scholarship of the Consejo Nacional de Investigaciones Científicas y Técnicas.

APPENDIX: SOME GEOMETRIC RELATIONS

Our conventions are signature $(-,+++)$,

$$R^{\mu}{}_{\nu\theta\rho} = \Gamma^{\mu}_{\nu\rho,\theta} - \Gamma^{\mu}_{\nu\theta,\rho} + \Gamma^{\mu}_{\theta\tau}\Gamma^{\tau}_{\nu\rho} - \Gamma^{\mu}_{\tau\rho}\Gamma^{\tau}_{\nu\theta},$$

$$R_{\mu\nu} = R^{\theta}{}_{\mu\theta\nu},$$

$$C_{\theta\rho\tau\epsilon} = R_{\theta\rho\tau\epsilon} - \tfrac{1}{2}(g_{\theta\tau}R_{\rho\epsilon} - g_{\theta\epsilon}R_{\rho\tau} - g_{\rho\tau}R_{\theta\epsilon} + g_{\rho\epsilon}R_{\theta\tau}) - \tfrac{1}{6}R(g_{\theta\epsilon}g_{\rho\tau} - g_{\theta\tau}g_{\rho\epsilon}).$$

Gauss–Bonnet theorem: In four dimensions the Gauss–Bonnet density G is a topological invariant[29]

$$G = \int d^4x\sqrt{-g}\,(R_{\theta\rho\tau\epsilon}R^{\theta\rho\tau\epsilon} + R^2 - 4R_{\theta\rho}R^{\theta\rho}).$$

Then a metric functional variation of G vanishes, which implies

$$ {}^{(3)}H_{\mu\nu} = 4\,{}^{(2)}H_{\mu\nu} - {}^{(1)}H_{\mu\nu}, \qquad \text{(A1)}$$

where

$$\begin{aligned}
{}^{(1)}H_{\mu\nu} &= \frac{1}{\sqrt{-g}}\frac{\delta}{\delta g^{\mu\nu}}\int d^4x\sqrt{-g}R^2 \\
&= 2R_{;\mu\nu} - 2RR_{\mu\nu} + \tfrac{1}{2}g_{\mu\nu}(R^2 - 4\Box R), \\
{}^{(2)}H_{\mu\nu} &= \frac{1}{\sqrt{-g}}\frac{\delta}{\delta g_{\mu\nu}}\int d^4x\sqrt{-g}R_{\theta\rho}R^{\theta\rho} \\
&= R_{;\mu\nu} - \Box R_{\mu\nu} - 2R^{\theta\rho}R_{\mu\theta\nu\rho} \\
&\quad + \tfrac{1}{2}g_{\mu\nu}(R_{\theta\rho}R^{\theta\rho} - \Box R), \\
{}^{(3)}H_{\mu\nu} &= \frac{1}{\sqrt{-g}}\frac{\delta}{\delta g_{\mu\nu}}\int d^4x\sqrt{-g}R_{\theta\rho\tau\epsilon}R^{\theta\rho\tau\epsilon} \\
&= 4R_{\mu\theta}R^{\theta}{}_{\nu} - 2R_{;\mu\nu} - 4R^{\theta\rho}R_{\mu\theta\nu\rho} \\
&\quad - 4\Box R_{\mu\nu} - 2R_{\mu\theta\rho\tau}R_{\nu}{}^{\theta\rho\tau} + \tfrac{1}{2}g_{\mu\nu}R_{\theta\rho\tau\epsilon}R^{\theta\rho\tau\epsilon}.
\end{aligned}$$

In conformally flat metrics, where the Weyl tensor vanishes, one has

$$R_{\theta\rho\tau\epsilon}R^{\theta\rho\tau\epsilon} = 2R_{\theta\rho}R^{\theta\rho} - \tfrac{1}{3}R^2,$$

and then a variation of G leads to

$$ {}^{(1)}H_{\mu\nu} = 3\,{}^{(2)}H_{\mu\nu} \quad \text{if } C_{\theta\rho\tau\epsilon} = 0. \qquad \text{(A2)}$$

Fifth-order geometric variables: All the terms appearing when the symmetry condition (9) is imposed on the expression (14) can be reduced, in conformally flat metrics, to the following fifth-order independent geometric variables:

$$\begin{aligned}
X_1 &= R_{\mu\nu}R_{,}{}^{\nu}, \quad X_2 = R_{;\theta}{}^{\theta}{}_{\mu}, \\
X_3 &= RR_{,\mu}, \quad X_4 = (R_{\theta\rho}R^{\theta\rho})_{,\mu}.
\end{aligned} \qquad \text{(A3)}$$

They are independent in the sense that it is impossible to find a relation between them that is valid for any possible conformally flat geometry, although particular relations exist at particular background geometries.

Robertson–Walker geometric quantities: Following is a list of the spatial and temporal components of the geometric quantities appearing through this paper when evaluated at the comoving coordinate system:

$$\begin{aligned}
&R_{;00} = 6\beta_5, \quad R_{;ij} = -6\beta_4 g_{ij}, \\
&RR_{00} = 18(-\beta_2 + \beta_3), \quad RR_{ij} = 6(\beta_2 + \beta_3)g_{ij}, \\
&\Box R_{00} = -3(4\beta_1 + 2\beta_2 - 6\beta_3 - \beta_4 - \beta_5),
\end{aligned}$$

where $\beta_1,\ldots,\beta_5$ are defined in Eq. (18) in the main text.

[1] For a recent review on the subject see N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge U. P., London, 1982).

[2]B. S. de Witt, in *General Relativity, An Einstein Centennary Survey*, edited by S. Hawking and W. Israel (Cambridge U. P., London, 1979).
[3]G. Gibbons and S. W. Hawking, Phys. Rev. D **15**, 2738 (1977).
[4]N. Sanchez, Phys. Rev. D **24**, 2100 (1981).
[5]L. Parker, in *Asymptotic Structure of Space Time*, edited by F. Esposito and L. Witten (Plenum, New York, 1977).
[6]S. M. Christensen, Phys. Rev. D **14**, 2490 (1976).
[7]S. M. Christensen, Phys. Rev. D **17**, 946 (1978).
[8]P. C. W. Davies, S. A. Fulling, S. M. Christensen, and T. S. Bunch, Ann. Phys. (NY) **109**, 108 (1977).
[9]S. L. Adler, J. Lieberman, and Y. J. Ng, Ann. Phys. (NY) **106**, 279 (1977).
[10]J. Hadamard, *Lectures on Cauchy Problem in Linear Partial Differential Equations* (Dover, New York, 1952).
[11]B. S. de Witt and R. W. Brehme, Ann. Phys. (NY) **9**, 220 (1960).
[12]R. M. Wald, Commun. Math. Phys. **54**, 1 (1977).
[13]R. M. Wald, Phys. Rev. D **17**, 1477 (1978).
[14]M. R. Brown and A. C. Ottewill, Proc. R. Soc. London Ser. A **389**, 379 (1983).
[15]M. Castagnino and D. Harari, Ann. Phys. (NY) **152**, 85 (1984).
[16]M. R. Brown, J. Math. Phys. **25**, 136 (1984).
[17]A. M. Najmi and A. C. Ottewill, Phys. Rev. D **30**, 1733 (1984).
[18]A. M. Najmi and A. C. Ottewill, Phys. Rev. D **30**, 2573 (1984).
[19]B. S. de Witt, Phys. Rep. C **19**, 295 (1975).
[20]L. Parker, in *Recent Developments in Gravitation*, edited by S. Deser and M. Levy (Plenum, New York, 1979).
[21]R. Penrose, in *Relativity, Groups and Topology*, edited by C. De Witt and B. De Witt (Gordon and Breach, New York, 1964).
[22]M. Castagnino, Gen. Relativ. Gravit. **9**, 101 (1978).
[23]L. S. Brown, Phys. Rev. D **15**, 1469 (1977).
[24]Originally derived by D. M. Capper and M. J. Duff, Nuovo Cimento A **2**, 173 (1974). See also Ref. 1, Chap. 6 and references therein.
[25]L. Parker and S. A. Fulling, Phys. Rev. D **9**, 341 (1974).
[26]M. Castagnino, D. Harari, and L. Chimento, Phys. Rev. D **24**, 290 (1981).
[27]S. G. Mamaev, V. M. Mostepanenko, and A. A. Starobinsky, Sov. Phys. JETP **43**, 823 (1976).
[28]S. A. Fulling, Gen. Relativ. Gravit. **10**, 807 (1979).
[29]M. Castagnino, Gen. Relativ. Gravit. **15**, 1149 (1983).
[30]A. M. Najmi, J. Math. Phys. **25**, 623 (1984).
[31]D. Harari, Ph. D. thesis, Universidad de Buenos Aires, Argentina, 1983.
[32]M. Castagnino, D. Harari, and C. Nunez, in *Proceedings of the Europhysics Study Conference: Unification of Fundamental Interactions II*, edited by S. Ferrara and J. Ellis (Plenum, New York, 1982).
[33]T. S. Bunch and L. Parker, Phys. Rev. D **20**, 2499 (1979).
[34]K. W. Howard and P. Candelas, Phys. Rev. Lett. **53**, 403 (1984).
[35]C. Lanczos, Ann. Math. **39**, 842 (1938).
[36]M. A. Castagnino and R. Ferraro, Phys. Rev. D **34**, 497 (1986).

An unconventional canonical quantization of local scalar fields over quantum space-time

Miklós Banai[a)]
Institute of Theoretical Physics, Eötvös University, H-1088 Budapest, Puskin U.5-7, Hungary

(Received 4 October 1983; accepted for publication 17 September 1986)

The central issue addressed in this paper is the following: Is the conventional procedure the only one for extending the canonical quantization method for local field theories or does another way exist? Here an unconventional extension of the canonical quantization method is presented for a classical local field theory consisting of N real scalar fields. This approach is essentially a reconsideration of the conventional procedure in an alternative way offered by a recent new approach of classical local field theories. The proposed canonical commutation relations have a solution in the A-valued Hilbert space $\mathscr{H}_A = \mathscr{H} \otimes A$, unique up to A-unitary equivalence, where A is the algebra of the bounded operators of the Hilbert space $L^2(\mathbb{R}^3)$. The canonical equations as operator equations are equivalent in form to the classical field equations, and are *a priori* well-defined for interacting systems, too. This model of quantized fields lacks some of the difficulties of the conventional approach, e.g., the rigorous application of the interaction picture is not stemmed by Haag's theorem and the ultraviolet catastrophe (there are no ultraviolet divergences in the S matrices of the model). Examples of the model satisfying the asymptotic condition provide examples for the axioms of Haag–Kastler while they satisfy the axioms of Wightman only partially. The consistent interpretation of the model requires a new concept of space-time, a quantum space-time. The "local" state space $\mathscr{H}_A$ of the model is constructed over this quantum space-time.

I. INTRODUCTION

If quantum mechanics is considered as a standard for a sound theory then, from this point of view, quantum field theory lies not entirely in this standard. The experimentally most successful conventional quantum field theory appended by the renormalization prescription has serious internal inconsistencies.[1–3] In fact, the detailed examinations had revealed the following difficulties of this theory.

Problem 1: The canonical commutation relations have infinitely many unitarily inequivalent representations (unlike the quantum mechanical case).[3,4]

Problem 2: If the basic hypotheses of the theory hold true then the interaction picture is not applicable to describe nontrivial interactions.[3] This is stemmed by two difficulties: (a) in the framework of the theory, by means of the interaction picture, one can derive only the trivial S matrix (Haag's theorem),[3,4] and (b) the interaction Hamiltonian consisting of higher powers than quadratic of the field does not possess a definite mathematical meaning in the Fock space of the free field (ultraviolet catastrophe).[4]

Problem 3: The quantum fields defined at the points of Minkowski space $\mathbb{M}^4$ do not exist as operators in a Hilbert space.[2]

The renormalization theory resolves only Problem 2(b) in such a way that it formally defines, in the basic dynamical equation, the term describing the interaction and which is *a priori* not defined.[5,6]

There are many approaches (see Ref. 6 for a brief overview) to remedy the problems of the conventional theory. Thus, e.g., the axiomatic approach of Streater and Wightman[7] and of Haag and Kastler[8] and others.[6] But none of these approaches is complete in the sense that one could apply them to quantize an arbitrary classical field theory to obtain numerical predictions for comparison with experimental results. Therefore, considering this situation, one may ask if the conventional procedure is the only one for extending the canonical quantization method of quantum mechanics to the cases of local field theories or does there exist another way?

The physical origin of the difficulties of the conventional theory was already pointed out by Schwinger in the 1950's when he had shown by physical arguments that a well-defined and convergent theory could not be imagined in the present (classical) space-time conceptions.[9] Together with many others (e.g., cf. Refs. 10 and 11), the present author also shares this opinion (cf. Ref. 12). In Ref. 13 the quantum logical approach of quantum mechanics was generalized to the cases of local field theories. This approach offers (a) a new class of representation spaces for representing the kinematical properties of quantum local field theories[14] and (b) an axiomatic approach to general space-time models.[15] In the latter approach the resulted quantum space-time models can be formulated, in the framework of a quantum relativity theory,[16] in terms of Davis' quantum relativity principle.[17] A specific quantum relativistic model of space-time (the quantum relativistic substitute of the Minkowski space-time) can be consistently defined[16] by exploiting the results of the classically relativistic quantum theory of Prugovečki.[10] The central problem of quantum mechanics of a quantum relativistic particle living in this quantum space-time is a mass eigenvalue problem[16] that predicts in the first approximation phenomenologically well-established classically relativistic

[a)] This paper is a revised version of the author's work, "The Canonical quantization of local scalar fields over quantum space-time," report KFKI-1983-51.

 0022-2488/87/010193-22$02.50

quantum particle ("elementary particle") spectra.[18]

Here, answering the above question in the affirmative,[19,12] we formulate a canonical quantization method in the "local state space" $\mathscr{H}_A$ constructed over quantum space-time of event space $L^2(\mathbb{R}^3)$,[15,16] for the illustrative case of N real classically relativistic scalar fields of Lagrangian density,

$$\mathscr{L}(t,\mathbf{x}) = \Big[\frac{1}{2}\sum_{\alpha=1}^{N} (\partial_\mu \phi_\alpha \partial^\mu \phi_\alpha - m_\alpha^2 \phi_\alpha^2) - V(\phi_1,\ldots,\phi_N)\Big](t,\mathbf{x}), \quad (t,\mathbf{x})\in\mathbb{M}^4. \tag{1.1}$$

Our guiding principle is the locality: all information obtainable by the system can be obtained by measuring the system at the points of the pertinent space-time model. In classical local field theory this principle is formulated in a natural way, i.e., the classical system is described by local fields and local observables, and the global observables are generated by local ones usually by integrating up local observables over spacelike hypersurfaces in classical space-time $\mathbb{M}^4$. Quantum local field theory should have a similar structure. Namely, the system should be described by local observables and local states. The global observables and global states should be generated by local ones in such a way that they are integrated up on a "spacelike hypersurface" of the space-time model corresponding to the system [on a maximal set of the causally disconnected points (events) of the space-time model].[15,16]

From the quantization algorithm itself the following is expected. Classical local field theory is built up from an infinite collection of identical classical mechanical systems of finitely many degrees of freedom, connected in space. The quantization should preserve this structure in such a way that it inserts in the place of the classical mechanical systems their quantum mechanical refinements obtained by canonical quantization. Consequently the quantized system should also be built up from an infinite collection of identical quantum mechanical systems connected in space.

For example, the classical local field theory of Lagrangian (1.1) consists of an infinite collection of identical classical anharmonic oscillators of N degrees of freedom connected in space. Then the corresponding quantum local field theory should consist of an infinite collection of identical quantum anharmonic oscillators of N degrees of freedom connected in space.

In our approach the Hilbert realization of the system of local propositions[13] of the studied quantum local field theory will be determined by means of a (trivial) Hilbert A module (or A-valued Hilbert space[20]) $\mathscr{H}_A$, where A is the algebra of bounded operators of the Hilbert space $L^2(\mathbb{R}^3)$. Thus all information obtained by local measurements of the quantum system are contained by $\mathscr{H}_A$: the local states can be represented by the rays of the $\mathscr{H}_A$ and the local bounded observables[13] by self-adjoint bounded operators (A-module homomorphisms) in $\mathscr{H}_A$. The expectation value of a local bounded observable F in the local state Ψ in $\mathscr{H}_A$ can be given by the formula

$$\overline{F} = EpF = \langle\Psi|F|\Psi\rangle_A \in A \tag{1.2}$$

using the A-valued Hermitian inner product of $\mathscr{H}_A$.[12,14,16] (Of course, here the locality means the locality in quantum space-time.[12,16])

In this way we construct a model of quantized fields which applies a new (quantum) conception of space-time and, in accordance with Schwinger's observation,[9] it lacks the main difficulties (Problems 1, 2, and 3 above) of the conventional theory.[12] Of course, the correct theory of quantized fields must reproduce the main physical result of the conventional theory, the renormalized S-matrix elements referring to scattering processes. Our model is very "close" in a certain sense[12] to the conventional theory because it is only a reconsideration, in an alternative way, of the extension of the canonical quantization method to local field theories (cf. Fig. 1). However, the question of whether the model reproduces the renormalized S-matrix elements of the conventional theory has not yet been answered. Nevertheless, an intuitive line of thought referring to an answer of the question in the affirmative can be found in Sec. IX.

Finally to this point we note that the formalism we present in this paper is an *extension* of the conventional Hilbert space formulation of quantum mechanics, thus this formulation is consistent as far as the former is consistent. To achieve this unconventional extension the relatively new methods, Boolean-valued models,[21] quantum set theory of Takeuti,[22] and Hilbert A modules,[20] of extension theory in mathematics is applied.[14] Another point is that we call our model quantum local field theory to distinguish it from the conventional theory usually named as local quantum field theory.

This paper is presented as follows. Section II deals with the physical and mathematical preliminaries. Here we summarize the physical motivations of our model and the applied mathematical tools. Section III is devoted to the formulation of the canonical quantization, and the representation of the canonical commutation relations in terms of the A-valued Hilbert space $\mathscr{H}_A$. After formulating the kinematical description of the quantized system in the local state space $\mathscr{H}_A$ in Sec. III, we treat the global description of the quantum system in Sec. IV. The dynamical description of the quantized system is formulated in Sec. V. These three sections (Secs. III–V) contain the unconventional extension of the canonical quantization method of quantum mechanics to local field theories consisting of real scalar fields, thus they carry the main contribution of this paper. Section VI is devoted to the extension of the interaction picture method of quantum mechanical perturbation theory, in this new framework, to quantum local field theories considered. The Lorentz invariance and the classical limit of our model of quantized fields are discussed in Sec. VII. The discussion of the axioms of Wightman and Haag–Kastler can be found in Sec. VIII. In Sec. IX some concluding remarks close this paper. In Appendices A and B, the basic notions of Hilbert A modules and of the Boolean-valued models of set theory and of quantum set theory will be briefly summarized, and in Appendix C Feynman's graph rules will also be briefly summarized.

Finally we collect the abbreviations we use in what follows:

c = classical,

q = quantum,

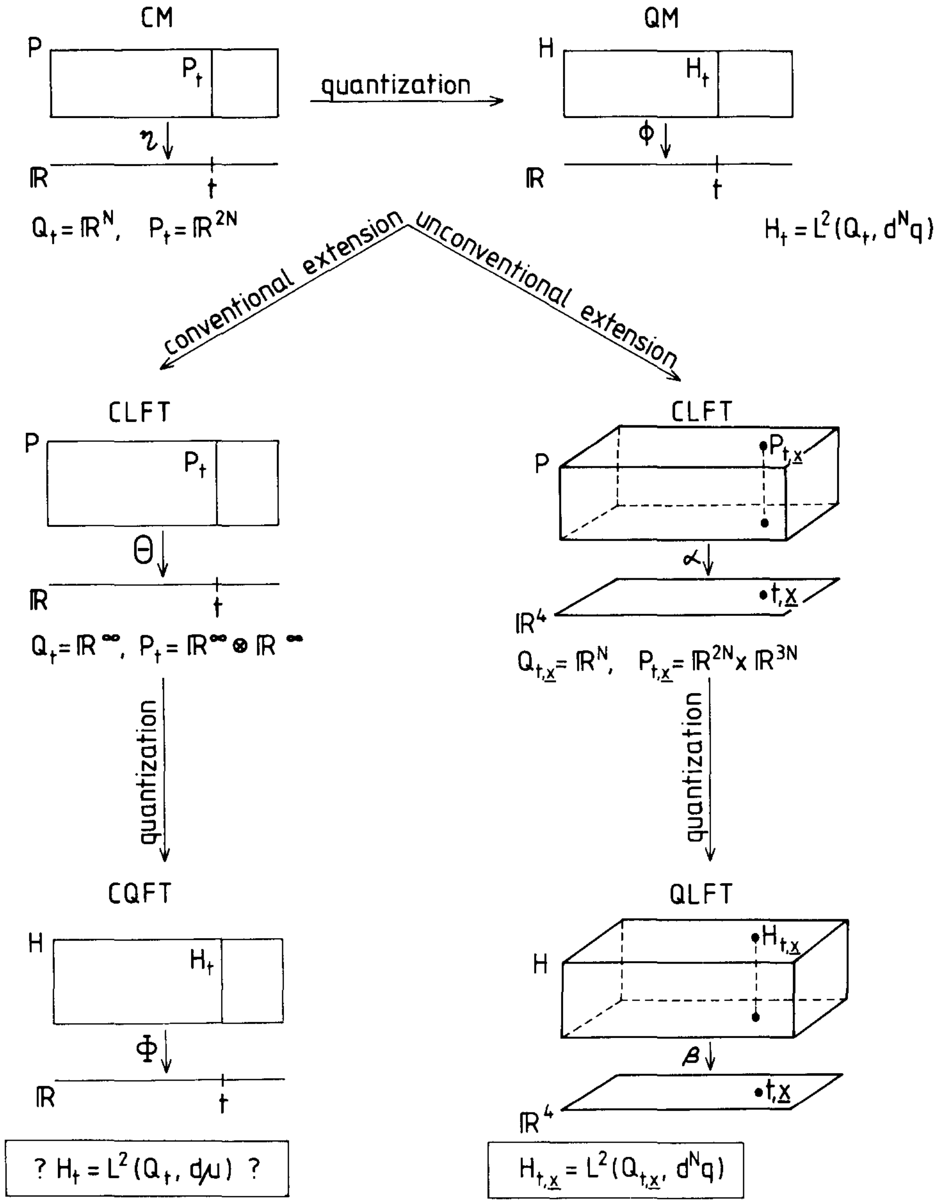

FIG. 1. An illustration of the basic idea of the unconventional extension of the quantization to CLFT. The column corresponding to the conventional extension knocks against the difficulties of the analysis in function space.[6] The column corresponding to the unconventional extension suggests a "passable road."

KG = Klein–Gordon,
CM = classical mechanics,
QM = quantum mechanics,
CCR = canonical commutation relation,
CLFT = classical local field theory,
CQFT = conventional quantum field theory,
QLFT = quantum local field theory,
IDPS = incomplete direct product space.

II. PHYSICAL AND MATHEMATICAL PRELIMINARIES

A. Physical motivations

(a) The basic idea of our unconventional extension of the canonical quantization method of CLFT under consideration is illustrated in Fig. 1, by comparing it with the conventional extension. According to the approach in Ref. 23, the fundamental geometrical space of CM is the configuration bundle $\xi: Q \to \mathbb{R}$ over the time line. Here Q is a fiber bundle with fiber $\xi^{-1}(t) = Q_t$ at time t. For example, for a system consisting of N one-dimensional anharmonic oscillators, $Q_t = \mathbb{R}^N$. The other fundamental object of CM is the phase bundle $\eta: P \to \mathbb{R}$. Here P is a fiber bundle with fiber $\eta^{-1}(t) = P_t$ at time t. In our example $P_t = \mathbb{R}^{2N}$. The canonical quantization substitutes P with a Hilbert bundle $\phi: H \to \mathbb{R}$ over the time line. H is a fiber bundle with fiber $\phi^{-1}(t) = H_t$ at time t. By von Neumann's uniqueness theorem one can identify the state space H_t with the Hilbert space $L^2(Q_t)$ up to unitary equivalence. In our example, $H_t = L^2(Q_t, d\mu) = L^2(\mathbb{R}^N, d^Nq)$.

The conventional approach extends this scheme in the following way. Now one can also attach to CLFT a configuration bundle $\xi: Q \to \mathbb{R}$ over the time line.[23] Q is a fiber bundle with fiber $\xi^{-1}(t) = Q_t$ at time t. Here Q_t is the function space of the Dirichlet data of CLFT at time t, and Q_t is infinite dimensional. For example, if the field $\phi(t,\mathbf{x})$ of CLFT consisting of a single real scalar field is square integrable in $\mathbf{x}$ then one can identify Q_t with the real Hilbert space l_2, so $Q_t = \mathbb{R}^\infty = l_2$. A phase bundle $\theta: P \to \mathbb{R}$ also corresponds to the system. The fiber $\theta^{-1}(t) = P_t$ is the phase space consisting of the Cauchy data of the system at time t. In our example we can write $P_t = \mathbb{R}^\infty \otimes \mathbb{R}^\infty = l_2 \otimes l_2$. The conventional quantization should also like to substitute P with a Hilbert bundle $\Phi: H \to \mathbb{R}$ of fiber $\Phi^{-1}(t) = H_t$ at time t. But now, from Problem 1, one cannot identify H_t, up to unitary equivalence, with a Hilbert space $L^2(Q_t, d\mu)$ because there are infinitely many unitarily inequivalent measures μ (corresponding to the representations of the CCR's) in the infinite dimensional function space Q_t.[24] In this direction the quantized system cannot be formulated, in terms of Hilbert spaces, in the same unique way as QM is formulated.

The recent formulation[23] of CLFT in terms of fiber bundles over c space-time $\mathbb{M}^4$ offers a new possibility to approach the quantization of CLFT over $\mathbb{M}^4$. According to this symplectic approach the fundamental geometrical space of CLFT is a configuration bundle ξ: $Q \to \mathbb{M}^4$ over the four-dimensional space-time manifold. Q is a fiber bundle with fiber $\xi^{-1}(t,\mathbf{x}) = Q_{t,\mathbf{x}}$ at the space-time point $(t,\mathbf{x})$. In the case of CLFT of Lagrangian (1.1) we have $Q_{t,\mathbf{x}} = \mathbb{R}^N$. The other fundamental object of CLFT is the phase bundle α: $P \to \mathbb{M}^4$. Here P is also a fiber bundle with fiber $\alpha^{-1}(t,\mathbf{x}) = P_{t,\mathbf{x}}$ at the space-time point $(t,\mathbf{x})$. In our example $P_{t,\mathbf{x}} = \mathbb{R}^{2N} \times \mathbb{R}^{3N}$. The factor $\mathbb{R}^{3N}$ is provided by the spatial "generalized velocities" $p^{\alpha}_{\mathbf{x}}(t) = \nabla_{\mathbf{x}} \phi_{\alpha}(t,\mathbf{x})$ in the Lagrangian of (1.1).[23] Then the quantization should substitute this geometrical picture with the following picture. P is substituted with a Hilbert bundle β: $\mathscr{H} \to \mathbb{R}^4$. Here $\mathscr{H}$ is a fiber bundle with fiber $\beta^{-1}(t,\mathbf{x}) = H_{t,\mathbf{x}}$, at the point $(t,\mathbf{x})$, and $H_{t,\mathbf{x}}$ is a complex separable Hilbert space. Because at the point $\mathbf{x}$ there is a mechanical system of *finitely many degrees of freedom* and of configuration space $Q_{t,\mathbf{x}}$, von Neumann's theorem holds true *locally*. Therefore we can identify $H_{t,\mathbf{x}}$ with the Hilbert space $L^2(Q_{t,\mathbf{x}})$ up to unitary equivalence. In our example $H_{t,\mathbf{x}} = L^2(Q_{t,\mathbf{x}}, d\mu) = L^2(\mathbb{R}^N, d^N q)$.

This approach grasps the fact that local field theory (both c and q) consists of an infinite collection of *mechanical systems* of *finitely many degrees of freedom* (sitting at the points of space $\mathbb{R}^3$ at a given instance t), more naturally than the conventional approach.

(b) We can establish a canonical formalism for CLFT in accordance with our guiding principle, the locality, i.e., we establish this formalism for the local fields and local observables.[19,12] For simplicity considering now the $N=1$ case in our example of Lagrangian (1.1), Poisson brackets of the local physical quantities of the form $F(t,\mathbf{x}) = F(\phi, \pi, \nabla\phi, \nabla\pi)(t,\mathbf{x})$ can be defined as follows:

$$\{F_1, F_2\} = \frac{\delta F_1}{\delta \phi}\frac{\delta F_2}{\delta \pi} - \frac{\delta F_2}{\delta \phi}\frac{\delta F_1}{\delta \pi}, \tag{2.1}$$

where $\pi = \partial \mathscr{L}/\partial\dot{\phi}$ and $\delta F/\delta\phi = \partial F/\partial\phi - \nabla(\partial F/\partial \nabla \phi)$ or $\delta F/\delta \pi = \partial F/\partial \pi - \nabla(\partial F/2\nabla \pi)$. Poisson's brackets of the basic local variables ϕ and π and their gradients $\nabla\phi$ and $\nabla\pi$ are

$$\{\pi,\pi\} = \{\phi, \phi\} = 0, \quad \{\pi, \phi\} = -1, \tag{2.2a}$$

$$\begin{aligned} &\{\nabla\pi, \pi\} = \{\nabla\phi, \phi\} = 0, \\ &\{\nabla\pi, \phi\} = \{\pi, \nabla\phi\} = (-\nabla)(-1), \end{aligned} \tag{2.2b}$$

$$\begin{aligned} &\{\nabla\pi, \nabla\pi\} = \{\nabla\phi, \nabla\phi\} = 0, \\ &\{\nabla\pi, \nabla\phi\} = (-\nabla)^2(-1). \end{aligned} \tag{2.2c}$$

Then the canonical quantization of the c theory means that we replace the local quantities $(\phi, \nabla\phi, \pi, \nabla\pi)$ with local observables $(\hat{\phi}, \widehat{\nabla\phi}, \hat{\pi}, \widehat{\nabla\pi})$, i.e., with self-adjoint elements of a noncommutative algebra, and Poisson's brackets (2.2a)–(2.2c) are replaced by commutator brackets (cf. Ref. 12),

$$[\hat{\pi}, \hat{\pi}] = [\hat{\phi}, \hat{\phi}] = 0, \quad [\hat{\pi}, \hat{\phi}] = -i\mathbb{1}, \tag{2.3a}$$

$$\begin{aligned} &[\widehat{\nabla\pi}, \hat{\pi}] = [\widehat{\nabla\phi}, \hat{\phi}] = 0, \\ &[\widehat{\nabla\pi}, \hat{\phi}] = [\hat{\pi}, \widehat{\nabla\phi}] = (-i\nabla)(-i\mathbb{1}), \end{aligned} \tag{2.3b}$$

$$\begin{aligned} &[\widehat{\nabla\pi}, \widehat{\nabla\pi}] = [\widehat{\nabla\phi}, \widehat{\nabla\phi}] = 0, \\ &[\widehat{\nabla\pi}, \widehat{\nabla\phi}] = (-i\nabla)^2(-i\mathbb{1}) \end{aligned} \tag{2.3c}$$

(we use the units $\hbar = c = 1$ in this paper). A solution of these CCR's with linear self-adjoint operators is

$$\hat{\phi}\Psi = \phi\cdot\Psi(\phi, \mathbf{x}), \quad \hat{\pi}\Psi = -i\frac{\partial\Psi(\phi, \mathbf{x})}{\partial\phi}, \tag{2.4a}$$

$$\begin{aligned} &\widehat{\nabla\phi}\Psi = -i\nabla_{\mathbf{x}}(\phi\cdot\Psi(\phi, \mathbf{x})), \\ &\widehat{\nabla\pi}\Psi = -i\nabla_{\mathbf{x}}\left(-i\frac{\partial\Psi(\phi, \mathbf{x})}{\partial\phi}\right), \end{aligned} \tag{2.4b}$$

where ϕ does not depend explicitly on $\mathbf{x}$, i.e., $[-i\nabla_{\mathbf{x}}, \hat{\phi}] = 0$, and Ψ is an (appropriate) element of the space of L^2 sections of the Hilbert bundle β_t: $\mathscr{H}_t \to \mathbb{R}$, where $\mathscr{H}_t = L^2(\mathbb{R}) \times \mathbb{R}^3$ is a subbundle of β: $\mathscr{H} \to \mathbb{R}^4$ [i.e., $\Psi \in L^2(\mathbb{R} \times \mathbb{R}^3)$]. By von Neumann's theorem the representation (2.4a) of the CCR's (2.3a) is unique at each point $\mathbf{x} \in \mathbb{R}^3$, up to unitary equivalence. Thus, as we expected, this quantization substitutes the trivial phase bundle $P = \mathbb{R}^2 \times \mathbb{R}^3 \times (\mathbb{R}^3 \times \mathbb{R})$ with the trivial Hilbert bundle $\mathscr{H} = L^2(\mathbb{R}) \times (\mathbb{R}^3 \times \mathbb{R})$; the fiber $P_{t,\mathbf{x}} = \mathbb{R}^2 \times \mathbb{R}^3$ is replaced by the fiber $H_{t,\mathbf{x}} = L^2(Q_{t,\mathbf{x}}) = L^2(\mathbb{R})$. The fiber $H_{t,\mathbf{x}}$ carries an irreducible solution of the CCR's (2.3a). In this way this quantization algorithm satisfies our expectation in point (a) above.[12,19]

From the irreducibility requirement for the system $(\hat{\phi}, \hat{\pi})$ at each point $\mathbf{x}$ and with the notation $\hat{\mathbf{p}} = -i\nabla_{\mathbf{x}}$ we have the general relations[12]

$$\widehat{\nabla\phi} = \hat{\mathbf{p}}\hat{\phi} = \hat{\phi}\hat{\mathbf{p}}, \quad \widehat{\nabla\pi} = \hat{\mathbf{p}}\hat{\pi} = \hat{\pi}\hat{\mathbf{p}}. \tag{2.5}$$

Thus $\widehat{\nabla\phi}$ has the same role as the spatial generalized velocity $p_{\mathbf{x}}(t) = \nabla_{\mathbf{x}}\phi(t,\mathbf{x})$ in the c theory. Namely it connects the "neighboring" anharmonic oscillators in space.

Then, the commutators of (2.3b) and (2.3c) follow from (2.3a) and (2.5) and conversely, the CCR's (2.3a)–(2.3c) with the irreducibility requirement for $(\hat{\phi}, \hat{\pi})$ at each point $\mathbf{x} \in \mathbb{R}^3$ imply the relations of (2.5).

The task is to determine the general abstract Hilbert representations of the CCR's (2.3). By considering the commutators in (2.3b) and (2.3c) one can think of $\hat{\pi}$, $\hat{\phi}$, $\widehat{\nabla\phi} = \hat{\phi}\hat{\mathbf{p}}$, and $\widehat{\nabla\pi} = \hat{\pi}\hat{\mathbf{p}}$ such that they act as operators in such a space $\mathscr{H}_{\hat{A}}$ in which the product among the elements of the space and the elements of an operator algebra $\hat{A}$ containing $\hat{\mathbf{p}}$ is defined and $\mathscr{H}_{\hat{A}}$ is equipped with an $\hat{A}$-valued scalar product, i.e., $\mathscr{H}_{\hat{A}}$ is a Hilbert $\hat{A}$ module and $\hat{\pi}$, $\hat{\phi}$, $\widehat{\nabla\phi}$, and $\widehat{\nabla\pi}$ are $\hat{A}$-module homomorphisms in $\mathscr{H}_{\hat{A}}$.

(c) In CQFT one encounters the following types of formal equations[1] (considering again the $N=1$ case in our example):

$$[\pi(t,\mathbf{x}), \phi(t,\mathbf{x}')] = -i\delta^3(\mathbf{x}-\mathbf{x}'), \tag{2.6}$$

$$[\nabla_{\mathbf{x}}\pi(t,\mathbf{x}), \phi(t,\mathbf{x}')] = -i\nabla_{\mathbf{x}}\delta^3(\mathbf{x}-\mathbf{x}'), \tag{2.7}$$

and for the free field

$$[\phi(t_1,\mathbf{x}_1),\phi(t_2,\mathbf{x}_2)] = -i\Delta(t_1-t_2,\mathbf{x}_1-\mathbf{x}_2), \tag{2.8}$$

$$\Delta(t_1-t_2,\mathbf{x}_1-\mathbf{x}_2) = (2\pi)^{-4}\int_{\mathbb{R}^3}\oint_C \frac{e^{ik_0(t_1-t_2)}}{m^2-k_0^2+\mathbf{k}^2} \times e^{-i\mathbf{k}(\mathbf{x}_1-\mathbf{x}_2)}\,dk_0\,d^3\mathbf{k}, \tag{2.9}$$

$$\underbrace{\phi(t_1,\mathbf{x}_1)\,\phi(t_2,\mathbf{x}_2)} = \langle 0|T(\phi(t_1,\mathbf{x}_1)\,\phi(t_2,\mathbf{x}_2))|0\rangle = i\Delta_F(t_1-t_2,\mathbf{x}_1-\mathbf{x}_2), \tag{2.10}$$

$$\Delta_F(t_1-t_2,\mathbf{x}_1-\mathbf{x}_2) = \lim_{\eta\to 0}(2\pi)^{-4}\int_{\mathbb{R}^4}\frac{e^{ik_0(t_1-t_2)}}{m^2-k_0^2+\mathbf{k}^2-i\eta} \times e^{-i\mathbf{k}(\mathbf{x}_1-\mathbf{x}_2)}\,dk_0\,d^3\mathbf{k}, \tag{2.11}$$

etc. The singular "functions" on the right-hand side of these types of equations make sense mathematically as the integral kernel representations of the corresponding linear operators in the rigged Hilbert space $S(\mathbb{R}^3)\subset L^2(\mathbb{R}^3)\subset S^*(\mathbb{R}^3)$.[2] Thus $\delta^3(\mathbf{x}-\mathbf{x}')$ is the "integral kernel" of the unit operator [restricting its domain to $S(\mathbb{R}^3)$], $-i\nabla_{\mathbf{x}}\delta^3(\mathbf{x}-\mathbf{x}')$ is the integral kernel of the three-momentum operator

$$\hat{\mathbf{p}} = \int_{\mathbb{R}^3}\mathbf{k}\,dP(\mathbf{k}) \tag{2.12}$$

($P(\mathbf{k})$ is the spectral decomposition of $\hat{\mathbf{p}}$ and its integral kernel is $P(\mathbf{k},\mathbf{x},\mathbf{x}') = (2\pi)^{-3}\exp[-i\mathbf{k}(\mathbf{x}-\mathbf{x}')]$)[25] and $\Delta(t_1-t_2,\mathbf{x}_1-\mathbf{x}_2)$ and $\Delta_F(t_1-t_2,\mathbf{x}_1-\mathbf{x}_2)$ are the integral kernels of the following functions of $\hat{\mathbf{p}}$:

$$\hat{\Delta}(t_1-t_2) = f(t_1-t_2,\hat{\mathbf{p}}) = \int_{\mathbb{R}^3}dP(\mathbf{k})\oint_C\frac{1}{2\pi} \times\frac{e^{ik_0(t_1-t_2)}}{m^2-k_0^2+\mathbf{k}^2}dk_0, \tag{2.13}$$

$$\hat{\Delta}_F(t_1-t_2) = f_F(t_1-t_2,\hat{\mathbf{p}}) = \int_{\mathbb{R}^3}dP(\mathbf{k})\lim_{\eta\to 0}\frac{1}{2\pi} \times\int_{-\infty}^{+\infty}\frac{e^{ik_0(t_1-t_2)}}{m^2-k_0^2+\mathbf{k}^2-i\eta}dk_0, \tag{2.14}$$

respectively. Thus the so-called "c-numbers" of CQFT, $\delta^3(\mathbf{x}-\mathbf{x}')$, $\Delta(t_1-t_2,\mathbf{x}_1-\mathbf{x}_2)$, $\Delta_F(t_1-t_2,\mathbf{x}_1-\mathbf{x}_2)$, etc., are in fact (commuting) elements of an operator algebra $\hat{A}$ in $L^2(\mathbb{R}^3)$ which contains $\hat{\mathbf{p}}$ and its functions. Therefore the equations of the types (2.6)–(2.8) are meaningful if π, ϕ, $\nabla\pi$, and $\nabla\phi$ act in such a space $\mathscr{H}_{\hat{A}}$ in which the product among the elements of the space and the elements of $\hat{A}$ is defined and, because of (2.10), it is equipped with an $\hat{A}$-valued scalar product, i.e., $\mathscr{H}_{\hat{A}}$ is again a Hilbert $\hat{A}$ module.

B. Mathematical tools

We apply the relatively new methods (Boolean-valued models of set theory,[21] q set theory of Takeuti,[22] Hilbert modules over operator algebras[20,26,27]) of modern extension theory of mathematics to establish the unconventional extension of the canonical quantization method to CLFT under consideration. We had presented these methods in Ref. 14 and a brief summary of them can be found in Appendices A and B.

Let us introduce the following notations. Let $\mathscr{H}$ be a separable Hilbert space and $(\hat{p},\hat{q})$ denote an irreducible system of self-adjoint operators in $\mathscr{H}$. Let $\mathscr{P}(\mathscr{H})$ be the lattice of orthogonal projectors of $\mathscr{H}$ and let $\mathscr{B}$ denote a complete Boolean sublattice of $\mathscr{P}(\mathscr{H})$. $B(\mathscr{H})$ denotes the algebra of bounded operators of $\mathscr{H}$, B is the Abelian von Neumann algebra generated by $\mathscr{B}$, and $\hat{B}$ denotes the unbounded extension of B, i.e., $\hat{B}$ is the *-algebra of operators having spectral projectors exclusively from $\mathscr{B}$ (cf. Refs. 28 and 14). Let $\hat{B}(\mathscr{H})$ be the operator *-algebra generated by the identity of $\mathscr{H}$ and by the members of $(\hat{p},\hat{q})$. If $\mathscr{H}=L^2(\mathbb{R}^N)$ then the domain of $\hat{B}(\mathscr{H})$ is the function space $S(\mathbb{R}^N)$.[2,29] We call $\hat{B}(\mathscr{H})$ the unbounded extension of $B(\mathscr{H})$. Especially let $\mathscr{P}=\mathscr{P}(L^2(\mathbb{R}^3))$, $A=B(L^2(\mathbb{R}^3))$, and $\hat{A}=\hat{B}(L^2(\mathbb{R}^3))$. Let $\mathscr{H}_A$ ($\mathscr{H}_B$) denote the trivial Hilbert A (B) module [or A (B)-valued Hilbert space] $\mathscr{H}\otimes A$ ($\mathscr{H}\otimes B$).[20] $B(\mathscr{H}_A)$ ($B(\mathscr{H}_B)$) denotes the algebra of the bounded operators [$A(B)$-module homomorphisms][27] of $\mathscr{H}_A$ ($\mathscr{H}_B$).

Lemma 2.1: $B(\mathscr{H}_A)=B(\mathscr{H})\otimes A$.

Proof: From a theorem of Kasparov (cf. Ref. 20, p. 137) we have $B(\mathscr{H}_A)=B(\mathscr{H}\otimes A)=M(K\otimes A)$, where $K=K(\mathscr{H})$ is the set of compact operators in $\mathscr{H}$, and $M(K\otimes A)$ denotes the multiplier algebra of $K\otimes A$.[20] For a C^*-algebra $\mathscr{A}$, $M(\mathscr{A})$ is defined as follows: $M(\mathscr{A}) := \{x|x\in\mathscr{A}^{dd}, x\mathscr{A}\subseteq\mathscr{A}\supseteq\mathscr{A}x\}$, where $\mathscr{A}^{dd}$ is the bidual of $\mathscr{A}$. In our case $\mathscr{A}=K\otimes A$ and $\mathscr{A}^{dd}=(K\otimes A)^{dd}=B(\mathscr{H})\otimes A$ (cf. Ref. 3, p. 111). The compact operators have the following property (see Ref. 30, Theorem 8.1 on p. 369): If $a\in K$ and $b\in B(\mathscr{H})$ then $ab\in K$ and $ba\in K$. Then this implies that

$$M(K\otimes A) = \{x|x\in B(\mathscr{H})\otimes A,\ x(K\otimes A)\subseteq K\otimes A\supseteq(K\otimes A)x\} = B(\mathscr{H})\otimes A. \qquad \text{Q.E.D.}$$

Furthermore we have $B(\mathscr{H}_B)=B(\mathscr{H})\otimes B$.[26]

V denotes the ordinary universe of set theory, while $V^{(\mathscr{B})}$ is the universe of the $\mathscr{B}$-valued model of set theory[21] and $V^{(\mathscr{P})}$ is the universe of the $\mathscr{P}$-valued model of Takeuti.[22] Then the *-algebra of complex numbers $\mathbb{C}^{(\mathscr{B})}$ in $V^{(\mathscr{B})}$ can be identified with $\hat{B}$, i.e., $\mathbb{C}^{(\mathscr{B})}=\hat{B}$.[31] The bounded part of $\mathbb{C}^{(\mathscr{B})}$ is equal to B.[32] We identify the *-algebra of complex numbers $\mathbb{C}^{(\mathscr{P})}$ in $V^{(\mathscr{P})}$ with $\hat{A}$. The bounded part of $\mathbb{C}^{(\mathscr{P})}=\hat{A}$ is A. A complex Hilbert space $\mathscr{H}^{(\mathscr{B})}$ in $V^{(\mathscr{B})}$ can be identified with $\mathscr{H}_{\hat{B}}=\mathscr{H}\otimes\hat{B}$,[28] while we identify a complex Hilbert space $\mathscr{H}^{(\mathscr{P})}$ in $V^{(\mathscr{P})}$ with $\mathscr{H}_{\hat{A}}=\mathscr{H}\otimes\hat{A}$, i.e., $\mathscr{H}^{(\mathscr{B})}=\mathscr{H}\otimes\hat{B}$, $\mathscr{H}^{(\mathscr{P})}=\mathscr{H}\otimes\hat{A}$. The bounded part of $\mathscr{H}^{(\mathscr{B})}$ ($\mathscr{H}^{(\mathscr{P})}$) is $\mathscr{H}\otimes B$ ($\mathscr{H}\otimes A$).[32] Furthermore we can write:

$$B(\mathscr{H}^{(\mathscr{B})}) = B(\mathscr{H}_{\hat{B}}) = B(\mathscr{H}_B) = B(\mathscr{H})\otimes B$$

(see Ref. 28),

$$\hat{B}(\mathscr{H}^{(\mathscr{B})}) = \hat{B}(\mathscr{H})\otimes\hat{B},\ \hat{B}(\mathscr{H}^{(\mathscr{P})}) = \hat{B}(\mathscr{H})\otimes\hat{A}.$$

If otherwise is not stated, in what follows $\mathscr{B}$ will denote the maximal Boolean sublattice of $\mathscr{P}$ generated by the spectral projections of $\hat{\mathbf{p}}$ in $\hat{A}$. We call $\mathscr{H}_{\hat{A}}$ ($\mathscr{H}_{\hat{B}}$) the unbounded extension of $\mathscr{H}_A$ ($\mathscr{H}_B$) and the operator *-algebra $\hat{B}(\mathscr{H}_{\hat{A}})=\hat{B}(\mathscr{H})\otimes\hat{A}$ [$\hat{B}(\mathscr{H}_{\hat{B}})=\hat{B}(\mathscr{H})\otimes\hat{B}$] with domain $\mathscr{D}=S(\mathbb{R}^N)\otimes S(\mathbb{R}^3)$ [if $\mathscr{H}=L^2(\mathbb{R}^N)$] are called the unbounded extension of $B(\mathscr{H}_A)$ [$B(\mathscr{H}_B)$].

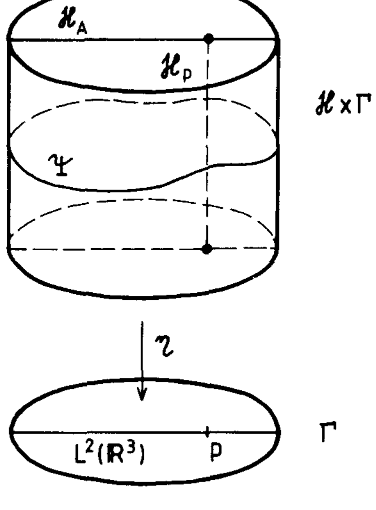

FIG. 2. An illustration of the geometrical structure of the local state space $\mathscr{H}_A$. One might think of this structure as a trivial, "noncommutative" Hilbert bundle over q space-time of event space $L^2(\mathbb{R}^3)$. The points of q space-time are represented by the rays f of $L^2(\mathbb{R}^3)$ or by the corresponding projectors $p(f) = |f\rangle\langle f|, f \in f$.

Notes: (1) $\mathscr{H}^{(\mathscr{B})}$ is complete in the $\hat{B}$-valued norm $\|\phi\|^{(\mathscr{B})} = \langle\phi|\phi\rangle_{\hat{B}}^{1/2}$, $\phi \in \mathscr{H}^{(\mathscr{B})}$,[28] and $\mathscr{H}^{(\mathscr{P})}$ is complete in the $\hat{A}$-valued norm $\|\phi\|^{(\mathscr{P})} = \langle\phi|\phi\rangle_{\hat{A}}^{1/2}$, $\phi \in \mathscr{H}^{(\mathscr{P})}$.

(2) $\mathscr{H}_B = \mathscr{H} \otimes B$ is isomorphic to the B module of sections of the trivial Hilbert bundle $\mathscr{H} \times \Gamma$ where $\Gamma = \mathrm{Sp}\,\mathscr{B}$ (the spectrum space of $\mathscr{B}$).[33] In this sense one might think of $\mathscr{H}_A = \mathscr{H} \otimes A$ as the A module of sections Ψ of the trivial "noncommutative" Hilbert bundle η: $\mathscr{H}_A \to L^2(\mathbb{R}^3)$ over q space-time of event space $L^2(\mathbb{R}^3)$ (Ref. 15) with fiber $\mathscr{H}_p$ at the point p (Fig. 2).[16]

(3) $(\mathscr{P}(\mathscr{H} \otimes A), \mathscr{P}, V)$ is an irreducible CROC-valued propositional system (representing in this way a pure QLFT),[13] where $\mathscr{P}(\mathscr{H} \otimes A)$ is the set of closed sub-Hilbert A modules of $\mathscr{H} \otimes A$.[14]

(4) $\mathscr{H}_B$ is a closed subspace of $\mathscr{H}_A$ and $\mathscr{H}_{\hat{B}}$ is a subspace of $\mathscr{H}_{\hat{A}}$.

(5) Mathematics ($\mathscr{B}$-valued analysis) in $V^{(\mathscr{B})}$ has already partly been developed (cf. Refs. 28, 31, and 32) but the development of mathematics ($\mathscr{P}$-valued analysis) in $V^{(\mathscr{P})}$ (in q set theory) is essentially still waiting for further research (cf. Ref. 22). It belongs to the aims of the present paper to call attention on this new branch of mathematics by applying its methods to solve physical problems.

III. THE CANONICAL QUANTIZATION IN TERMS OF AN A-VALUED HILBERT SPACE $\mathscr{H}_A = \mathscr{H} \otimes A$

Now we formulate the canonical quantization of CLFT of Lagrangian (1.1) in terms of $\mathscr{H}_A$ and its unbounded extension $\mathscr{H}_{\hat{A}}$. One postulates that the variables ϕ_α, $\pi_\alpha = \dot{\phi}_\alpha$, $\nabla\phi_\alpha$, $\nabla\pi_\alpha$, $\alpha = 1,...,N$, of CLFT are local observables and thus represented by self-adjoint operators in $\hat{B}(\mathscr{H}_{\hat{A}})$ such that the local fields ϕ_α and momentums π_α satisfy the equal time CCR's

$$[\hat{\phi}_\alpha, \hat{\phi}_\beta] = [\hat{\pi}_\alpha, \hat{\pi}_\beta] = 0, \quad \alpha, \beta = 1,...,N, \tag{3.1}$$
$$[\hat{\pi}_\alpha, \hat{\phi}_\beta] = -i\delta_{\alpha\beta}\mathbb{1}\cdot 1,$$

where $\mathbb{1} = 1_{\mathscr{H}} \otimes 1$, $1 \in A$, and equalities are of course understood on the common domain of both sides. An example for operators satisfying (3.1) is the following. Let $\{\varphi_n\}_{n=1}^\infty$ be a basis in $L^2(\mathbb{R}^3)$, then

$$\hat{\phi}_\alpha = \sum_n \hat{q}_\alpha \otimes c_n^\alpha |\varphi_n\rangle\langle\varphi_n|, \qquad \hat{\pi}_\alpha = \sum_n \hat{p}_\alpha \otimes \frac{1}{c_n^\alpha}|\varphi_n\rangle\langle\varphi_n|, \tag{3.2a}$$

where

$$[\hat{p}_\alpha, \hat{p}_\beta] = [\hat{q}_\alpha, \hat{q}_\beta] = 0, \quad [\hat{p}_\alpha, \hat{q}_\beta] = -i\delta_{\alpha\beta}1_{\mathscr{H}} \tag{3.2b}$$

and $c_n^\alpha \in \mathbb{R}$, $\hat{p}_\alpha, \hat{q}_\beta \in \hat{B}(\mathscr{H})$, and $\alpha, \beta = 1,...,N$.

From the relations (2.5) and

$$\begin{aligned}[\hat{p}_1 \otimes \hat{a}_1, \hat{p}_2 \otimes \hat{a}_2] &= \hat{p}_1\hat{p}_2 \otimes [\hat{a}_1, \hat{a}_2] + [\hat{p}_1, \hat{p}_2] \otimes \hat{a}_2\hat{a}_1 \\ &= \hat{p}_2\hat{p}_1 \otimes [\hat{a}_1, \hat{a}_2] + [\hat{p}_1, \hat{p}_2] \otimes \hat{a}_1\hat{a}_2, \\ &\hat{a}_1, \hat{a}_2 \in \hat{A}, \quad \hat{p}_1, \hat{p}_2 \in \hat{B}(\mathscr{H}),\end{aligned}$$

we obtain the general form of $\hat{\phi}_\alpha$ and $\hat{\pi}_\alpha$. They are

$$\hat{\phi}_\alpha = \hat{\phi}_\alpha(\hat{\mathbf{p}}) = \hat{q}_\alpha \otimes \hat{c}_\alpha, \qquad \hat{\pi}_\alpha = \hat{\pi}_\alpha(\hat{\mathbf{p}}) = \hat{p}_\alpha \otimes \hat{c}_\alpha^{-1}, \quad \hat{c}_\alpha = \hat{c}_\alpha^*, \tag{3.3}$$

where $\hat{c}_\alpha$'s are invertible elements of $\hat{B}$ ($\subset \hat{A}$) and $(\hat{p}, \hat{q}) = (\hat{p}_1,..., \hat{p}_N, \hat{q}_1,..., \hat{q}_N)$ is an irreducible system of self-adjoint operators in $\mathscr{H}$ satisfying the CCR's (3.2b). A well-known solution for $(\hat{p}, \hat{q})$ is

$$\mathscr{H} = L^2(\mathbb{R}^N), \quad \hat{q}_\alpha = q_\alpha\cdot, \qquad \hat{p}_\alpha = -i\frac{\partial}{\partial q_\alpha}, \quad \alpha = 1,..., N. \tag{3.4}$$

Let us cast the CCR's (3.1) into Weyl's form. We define the unitary operators

$$U(\alpha) = \exp\left\{i\sum_{n=1}^N \alpha_n \hat{\phi}_n\right\}, \quad V(\alpha) = \exp\left\{i\sum_{n=1}^N \alpha_n \hat{\pi}_n\right\}$$

in $\mathscr{H}_A$, where $\alpha = (\alpha_1,...,\alpha_N) \in \mathbb{R}^N$. We assume that U and V are continuous with respect to the parameters α and satisfy the relations

$$U(\alpha)U(\beta) = U(\alpha + \beta), \quad V(\alpha)V(\beta) = V(\alpha + \beta),$$
$$U(\alpha)V(\beta) = \exp\left\{i\sum_{n=1}^N \alpha_n\beta_n\right\}V(\beta)U(\alpha). \tag{3.5}$$

Definition 3.1: A system (U,V) of bounded operators in the A-valued Hilbert space $\mathscr{H}_A$ is called a $\mathscr{B}$-irreducible system if the set of bounded operators in $\mathscr{H}_A$ commuting with all the members of the system (U,V) is equal to the Abelian von Neumann algebra B.

Note that this irreducibility notion is the natural extension of the irreducibility notion formulated in the $\mathscr{B}$-valued universe $V^{(\mathscr{B})}$ to the $\mathscr{P}$-valued universe $V^{(\mathscr{P})}$. The extension of the irreducibility notion in the ordinary universe V to the irreducibility notion in $V^{(\mathscr{B})}$ is as follows: A system of bounded operators in the Hilbert space $\mathscr{H}^{(\mathscr{B})}$ in $V^{(\mathscr{B})}$ is irreducible if its center is equal to the bounded part of the complex numbers $\mathbb{C}^{(\mathscr{B})}$ in $V^{(\mathscr{B})}$, i.e., equal to B (cf. Ref. 28).

Then we have the following extension of von Neumann's theorem.

Proposition 3.2: A $\mathscr{B}$-irreducible set of unitary operators $U(\alpha)$ and $V(\alpha)$ in the A-valued Hilbert space $\mathscr{H}_A$ satisfying the CCR's (3.5) is uniquely determined up to A-unitary equivalence, i.e., if $(U'(\alpha),V'(\alpha))$ is another set of $\mathscr{B}$-irreducible unitary operators in $\mathscr{H}'_A$ also satisfying (3.5), then an A-unitary operator S: $\mathscr{H}'_A \to \mathscr{H}_A$ exists such that $U(\alpha) = SU'(\alpha)S^{-1}$ and $V(\alpha) = SV'(\alpha)S^{-1}$.

Proof: We apply the methods of $\mathscr{B}$- and $\mathscr{P}$-valued models for the proof of the proposition. Here $\mathscr{H}_A = \mathscr{H} \otimes A$ is the bounded part of the Hilbert space $\mathscr{H}^{(\mathscr{P})} = \mathscr{H} \otimes \hat{A}$ in $V^{(\mathscr{P})}$ and $\mathscr{H}'_A = \mathscr{H}' \otimes A$ is the bounded part of the Hilbert space $\mathscr{H}'^{(\mathscr{P})} = \mathscr{H}' \otimes \hat{A}$ in $V^{(\mathscr{P})}$. Furthermore, it follows from the $\mathscr{B}$-irreducibility property of the two sets $(U(\alpha),V(\alpha))$ and $(U'(\alpha),V'(\alpha))$ that $U(\alpha)$, $V(\alpha) \in B(\mathscr{H}_B) = B(\mathscr{H}^{(\mathscr{B})}) \subset B(\mathscr{H}_A)$ and $U'(\alpha)$, $V'(\alpha) \in B(\mathscr{H}'_B) = B(\mathscr{H}'^{(\mathscr{B})}) \subset B(\mathscr{H}'_A)$ [$U(\alpha)$ and $V(\alpha)$ map the subspace $\mathscr{H}_B = \mathscr{H} \otimes B$ of $\mathscr{H}_A$ onto itself and $U'(\alpha)$ and $V'(\alpha)$ map the subspace $\mathscr{H}'_B = \mathscr{H}' \otimes B$ of $\mathscr{H}'_A$ onto itself]. In this way the system $(U(\alpha),V(\alpha))$ is an irreducible system of unitary operators in the Hilbert space $\mathscr{H}^{(\mathscr{B})}$ in $V^{(\mathscr{B})}$ and the system $(U'(\alpha),V'(\alpha))$ is an irreducible system in the Hilbert space $\mathscr{H}'^{(\mathscr{B})}$ in $V^{(\mathscr{B})}$. Now, because $V^{(\mathscr{B})}$ satisfies the axioms ZFC (Zermelo–Fraenkel plus the axiom of choice)[14,21] we can apply theorem 30 in Ref. 21 (pp. 55–57) to von Neumann's theorem in the ordinary universe V which yields that von Neumann's theorem also holds true in $V^{(\mathscr{B})}$ [i.e., $(U(\alpha),V(\alpha))$ in $\mathscr{H}^{(\mathscr{B})}$ and $(U'(\alpha),V'(\alpha))$ in $\mathscr{H}'^{(\mathscr{B})}$ are unitarily equivalent in $V^{(\mathscr{B})}$]. Then the canonical extension of the corresponding unitary operator in between $\mathscr{H}^{(\mathscr{B})}$ and $\mathscr{H}'^{(\mathscr{B})}$ provides the A-unitary operator S in between $\mathscr{H}_A$ and $\mathscr{H}'_A$. Q.E.D.

This extension of von Neumann's theorem offers the possibility that we formulate QLFT in terms of the A-valued Hilbert spaces in the same unique way, up to A-unitary equivalence (avoiding Problem 1 of CQFT in this way) as QM is formulated in terms of complex Hilbert spaces up to unitary equivalence. As we pointed out in Ref. 16 one can consider the representation space $\mathscr{H}_A$ as the local state space of QLFT constructed over q space-time of event space $L^2(\mathbb{R}^3)$ (cf. Note 2 in Sec. II B) ("local" here means the locality in q space-time[15,16]). We can straightforwardly adapt the "rules" of the Hilbert space formulation of QM[25] to QLFT.

(A1) The *local state space* of QLFT is an A-valued separable Hilbert space $\mathscr{H}_A = \mathscr{H} \otimes A$; the local states are represented by rays $\mathbf{\Psi}$, $\langle\Psi|\Psi\rangle_A = 1$, $\Psi \in \mathbf{\Psi}$, of $\mathscr{H}_A$, where $\langle|\rangle_A$ is the A-valued scalar product of $\mathscr{H}_A$.

(A2) The local bounded observables[13] are represented by self-adjoint bounded operators in $\mathscr{H}_A$ (the spectral values of these operators are self-adjoint elements of the operator algebra A).

(A3) The expectation value of a local bounded observable F in the local state $\mathbf{\Psi}$ is

$$\bar{F} = \langle\Psi|F|\Psi\rangle_A \in A_*, \quad \Psi \in \mathbf{\Psi}, \tag{$*$}$$

where A_* denotes the set of the self-adjoint elements of A. If we deal with unbounded quantities then we use the unbounded extension $\mathscr{H}_{\hat{A}} = \mathscr{H} \otimes \hat{A}$ of $\mathscr{H}_A$ instead of $\mathscr{H}_A$.

Generally we can say that $\mathscr{H}_A$ carries all the information one gets by local measurements in q space-time.[13,16] Furthermore, in QLFT the q local fields $\hat{\phi}_\alpha$, at a given instant t, is operationally defined at the points of q space-time, i.e., the q local fields are operator-valued functions $\mathbf{f} \to \hat{\phi}_\alpha(\mathbf{f}) = p(f)\hat{\phi}_\alpha\, p(f) \in \hat{B}(\mathscr{H}_{\hat{A}})$, $\alpha = 1,\ldots, N$, where $\mathbf{f}$ is a ray of $L^2(\mathbb{R}^3)$ representing a point (event) of q space-time and $p(f) = |f\rangle\langle f|$ is the one-dimensional projector corresponding to $\mathbf{f}$.[16]

Notes: (1) Two elements of a ray $\mathbf{\Psi}$ differ from each other by a unitary operator, i.e., $\Psi_1 = \Psi_2 u$, $\Psi_1,\Psi_2 \in \mathbf{\Psi}$. For $\langle\Psi_1|\Psi_1\rangle_A = 1$, $\langle\Psi_2|\Psi_2\rangle_A = 1$, and $\langle\Psi_2 u|\Psi_2 u\rangle_A = u^*\langle\Psi_2|\Psi_2\rangle_A u = u^*u = 1$. The ray $\mathbf{\Psi}$ can be identified with a "one-dimensional" projector $P(\Psi) = |\Psi\rangle_A\langle\Psi|$ projecting on the one-dimensional A-valued subspace $\{c\Psi a|\ c\in\mathbb{C},\ a\in A\}$ (the A-valued subspace is itself an A module over A in $\mathscr{H}_A$, too).

(2) If p is an atom in $\mathscr{P}$ [i.e., an event (point) in q space-time], then the ray of the form $\Phi \otimes p$, $\langle\Phi|\Phi\rangle = 1$, in $\mathscr{H} \otimes A$ corresponds to a *pure local state* of the q system (according to Ref. 13) and conversely, every pure local state is represented in this way in $\mathscr{H}_A$. If we prepare the q system on a spacelike hypersurface in q space-time generated by a maximal set of causally disconnected events,[15]

$$\Omega = \{p_\alpha|\ \alpha\in\mathbb{N},\ p_\alpha = p(\varphi_\alpha),\ \varphi_\alpha \in L^2(\mathbb{R}^3),$$
$$\langle\varphi_\alpha|\varphi_\beta\rangle = \delta_{\alpha\beta},\ \Sigma_\alpha |\varphi_\alpha\rangle\langle\varphi_\alpha| = 1\},$$

then this means the specification of a set of pure local states, $\{\mathbf{\Phi}_\alpha \otimes p_\alpha|\ p_\alpha \in \Omega,\ \Phi_\alpha \in \mathscr{H}\}$. Generally, one can think of a local state $\mathbf{\Psi}$ as a section of norm 1 in the noncommutative Hilbert bundle η: $\mathscr{H}_A \to L^2(\mathbb{R}^3)$ over q space-time of event space $L^2(\mathbb{R}^3)$ (according to the sense of Note 2 in Sec. II B).

(3) The L^2-sections $\Phi(\phi_1,\ldots,\phi_N,\mathbf{x},t)$ of the Hilbert bundle $\mathscr{H} = L^2(\mathbb{R}^N) \times \mathbb{R}^4$ over $\mathbb{R}^4$ can be generated by the elements $\hat{\Psi}$ of $\mathscr{H}_A = L^2(\mathbb{R}^N) \otimes A$ as follows:

$$\Phi(\phi_1,\ldots,\phi_N,\mathbf{x},t) = \Psi(\phi_1,\ldots,\phi_N,\hat{\mathbf{p}},\hat{\mathbf{x}},t)\chi(\mathbf{x}),$$

where $\chi(\mathbf{x}) \in L^2(\mathbb{R}^3)$.[19] The A-valued scalar product in $\mathscr{H}_A$ in this case is of the form

$$\langle\hat{\Psi}_1|\hat{\Psi}_2\rangle_A = \int_{\mathbb{R}^N} d^N\phi\, \Psi_1^*(\phi_1,\ldots,\phi_N,\hat{\mathbf{p}},\hat{\mathbf{x}}) \times \Psi_2(\phi_1,\ldots,\phi_N,\hat{\mathbf{p}},\hat{\mathbf{x}}). \tag{3.6}$$

We might say that a unique (in the mentioned sense) representation space for describing QLFT is obtained, by means of canonical quantization, in the way that one constructs the A-valued Hilbert space $\mathscr{H}_A = \mathscr{H} \otimes A$ over the sub-Hilbert bundle β_t: $\mathscr{H}_t \to \mathbb{R}^3$ [$\mathscr{H}_t = L^2(\mathbb{R}^N) \times \mathbb{R}^3$] at the given instant t.

(4) Let $\{\hat{\Phi}_n\}$ be a complete orthonormal system in $\mathscr{H}_{\hat{A}}$, i.e.,

$$\langle\hat{\Phi}_n|\hat{\Phi}_m\rangle_{\hat{A}} = \delta_{nm}\cdot 1, \quad \sum_{n=1}^{\infty} |\hat{\Phi}_n\rangle_{\hat{A}}\langle\hat{\Phi}_n| = \mathbb{1}, \tag{3.7}$$

such that every $\hat{\Phi}_n$ is in $\mathscr{H}_{\hat{B}}$ ($\subset \mathscr{H}_{\hat{A}}$). If $\mathscr{H} = L^2(\mathbb{R}^N)$ and $\{\Phi_n\} = \{\Phi_n(\phi_1,\ldots,\phi_N)\}$ is a complete orthonormal system in $L^2(\mathbb{R}^N)$, then the form of $\hat{\Phi}_n$ is as follows:

$$\hat{\Phi}_n = \Phi_n(\phi_1(\hat{\mathbf{p}}),\ldots,\phi_N(\hat{\mathbf{p}})) = \int dP(\mathbf{k})\Phi_n(\phi_1(\mathbf{k}),\ldots,\phi_N(\mathbf{k})),$$

where $P(\mathbf{k})$ is the spectral decomposition of $\hat{\mathbf{p}}$. Then the precise meaning of (3.7) is as follows:

$$\langle\hat{\Phi}_n|\hat{\Phi}_m\rangle_{\hat{A}} = \int d^N\phi(\hat{\mathbf{p}})\Phi_n^*(\phi_1(\hat{\mathbf{p}}),..., \phi_N(\hat{\mathbf{p}}))$$
$$\times\Phi_m(\phi_1(\hat{\mathbf{p}}),...,\phi_N(\hat{\mathbf{p}}))$$
$$= \int_{\mathbf{R}^3} dP(\mathbf{k}) \int_{\mathbf{R}^N} d^N\phi(\mathbf{k})\Phi_n^*(\phi_1(\mathbf{k}),..., \phi_N(\mathbf{k}))$$
$$\times\Phi_m(\phi_1(\mathbf{k}),..., \phi_N(\mathbf{k}))$$
$$= \int_{\mathbf{R}^3} dP(\mathbf{k})\delta_{nm} = \delta_{nm}\ 1\,, \tag{3.8}$$

and

$$\sum_{n=1}^{\infty} |\hat{\Phi}_n\rangle_{\hat{A}}\langle\hat{\Phi}_n| = \sum_{n=1}^{\infty}\int_{\mathbf{R}^3} dP(\mathbf{k})|\Phi_n(\mathbf{k})\rangle\langle\Phi_n(\mathbf{k})|$$
$$= \sum_{n=1}^{\infty}\int_{\mathbf{R}^3} dP(\mathbf{k})\Phi_n^*(\phi_1'(\mathbf{k}),..., \phi_N'(\mathbf{k}))$$
$$\times\ \Phi_n(\phi_1(\mathbf{k}),..., \phi_N(\mathbf{k}))$$
$$= \int_{\mathbf{R}^3} dP(\mathbf{k})\delta^N(\phi'(\mathbf{k}) - \phi(\mathbf{k}))$$
$$= \delta^N(\phi' - \phi)\cdot 1\,, \tag{3.9}$$

where $\phi = (\phi_1,..., \phi_N)$ and $\delta^N(\phi' - \phi)$ is the integral kernel of the unity of $L^2(\mathbb{R}^N)$.[2] By the aid of this orthonormal system we can write the elements of $\mathscr{H}_{\hat{A}} = L^2(\mathbb{R}^N)\otimes\hat{A}$ in the following form:

$$\hat{\Phi} = \Phi(\phi_1(\hat{\mathbf{p}}),..., \phi_N(\hat{\mathbf{p}}),\hat{\mathbf{p}},\hat{\mathbf{x}})$$
$$= \sum_{n=1}^{\infty} c_n\ \hat{\Phi}_n\hat{a}_n$$
$$= \sum_{n=1}^{\infty} c_n\ \Phi_n(\phi_1(\hat{\mathbf{p}}),..., \phi_N(\hat{\mathbf{p}}))\cdot a_n(\hat{\mathbf{p}},\hat{\mathbf{x}})\,, \tag{3.10}$$

where $c_n\in\mathbb{C}$, $\hat{a}_n\in\hat{A}$. The $\hat{A}$-valued scalar product of two elements $\hat{\Phi}_1$, $\hat{\Phi}_2\in\mathscr{H}_{\hat{A}}$ is as follows:

$$\langle\hat{\Phi}_1|\hat{\Phi}_2\rangle_{\hat{A}} = \sum_{n,m=1}^{\infty}\bar{c}_n^1 c_m^2\hat{a}_{1n}^*\langle\hat{\Phi}_n|\hat{\Phi}_m\rangle_{\hat{A}}\hat{a}_{2m}$$
$$= \sum_{n=1}^{\infty}\bar{c}_n^1 c_n^2\hat{a}_{1n}^*\hat{a}_{2n}\,. \tag{3.11}$$

Let the local observable $\hat{F}$ be the function of $(\hat{\phi},\hat{\pi},\widehat{\nabla\phi},\widehat{\nabla\pi})$, i.e., $\hat{F} = F(\hat{\phi},\hat{\pi},\widehat{\nabla\phi},\widehat{\nabla\pi})$. [In what follows we will use the abbreviations $\hat{\phi} = (\hat{\phi}_1,..., \hat{\phi}_N)$, $\hat{\pi} = (\hat{\pi}_1,...,\hat{\pi}_N)$, $\widehat{\nabla\phi} = (\widehat{\nabla\phi}_1,..., \widehat{\nabla\phi}_N)$, and $\widehat{\nabla\pi} = (\widehat{\nabla\pi}_1,...,\widehat{\nabla\pi}_N)$, too.] Then, because of (2.5), we have

$$\hat{F} = F(\hat{\phi}(\hat{\mathbf{p}}), \hat{\pi}(\hat{\mathbf{p}}),\hat{\mathbf{p}})\in\hat{B}(\mathscr{H}_{\hat{B}})$$
$$= \hat{B}(\mathscr{H})\otimes\hat{B}\subset\hat{B}(\mathscr{H}_{\hat{A}})\,.$$

Since in the universe $V^{(\mathscr{B})}$ the spectral theorem holds true,[21,14] we can easily derive the spectral form of $\hat{F}$. Let $\hat{F}$ be of discrete (nondegenerate) spectrum, i.e.,

$$\hat{F}\hat{\Phi}_n = \hat{\Phi}_n\hat{f}_n,\quad \langle\hat{\Phi}_n|\hat{\Phi}_m\rangle_{\hat{A}} = \delta_{nm}\cdot 1\,,$$
$$\sum_{n=1}^{\infty}|\hat{\Phi}_n\rangle_{\hat{A}}\langle\hat{\Phi}_n| = \mathbb{1}\,, \tag{3.12}$$

where $\hat{f}_n\in\hat{B}_*$ ($\hat{B}_*$ is the set of self-adjoint elements of $\hat{B}$, the set of real numbers $\mathbb{R}^{(\mathscr{B})}$ in $V^{(\mathscr{B})}$). Then the spectral form of $\hat{F}$ is

$$\hat{F} = \sum_{n=1}^{\infty}\hat{E}_n\hat{f}_n = \sum_{n=1}^{\infty}|\hat{\Phi}_n\rangle_{\hat{A}}\langle\hat{\Phi}_n|\hat{f}_n$$
$$= \sum_{n=1}^{\infty}\int_{\mathbf{R}^3} dP(\mathbf{k})|\Phi_n(\mathbf{k})\rangle\langle\Phi_n(\mathbf{k})|f_n(\mathbf{k})\,. \tag{3.13}$$

If $\hat{F}$ is of continuous spectrum then its spectral form is

$$\hat{F} = \int d\hat{E}(\lambda)\hat{f}(\lambda) = \int_{\mathbf{R}^3} dP(\mathbf{k})\int_{-\infty}^{+\infty} dE(\mathbf{k},\lambda)f(\mathbf{k},\lambda)\,, \tag{3.14}$$

where $\hat{E}(\lambda) = E(\hat{\mathbf{p}},\lambda)$ is the spectral family in $\mathscr{H}^{(\mathscr{B})} = \mathscr{H}_{\hat{B}}\subset\mathscr{H}_{\hat{A}}$, belonging to $\hat{F}$, and $\hat{f}(\lambda) = f(\hat{\mathbf{p}},\lambda)\in\hat{B}_*$. If $\hat{F}$ has discrete as well as continuous spectra then its spectral form from (3.13) and (3.14) is

$$\hat{F} = \sum_n\hat{E}_n\hat{f}_n + \int d\hat{E}(\lambda)\hat{f}(\lambda)$$
$$= \sum_n\int_{\mathbf{R}^3} dP(\mathbf{k})E_n(\mathbf{k})f_n(\mathbf{k})$$
$$+ \int_{\mathbf{R}^3} dP(\mathbf{k})\int_{-\infty}^{+\infty} dE(\mathbf{k},\lambda)f(\mathbf{k},\lambda)\,. \tag{3.15}$$

Since $\mathscr{H}_{\hat{B}}$ is isomorphic to the $\hat{B}$-module of sections of the trivial Hilbert bundle $\mathscr{H}\times\Gamma$, where $\Gamma = \mathrm{Sp}\ \mathscr{B}$ (cf. Note 2 in Sec. II B), thus we can write the maps $n(\hat{\mathbf{p}}):\Gamma\to\mathbb{N}$ instead of the natural numbers $n\in\mathbb{N}$ in the above formulas, i.e., the natural numbers in $V^{(\mathscr{B})}$, $n(\hat{\mathbf{p}})\in\mathbb{N}^{(\mathscr{B})}$. The map $n(\hat{\mathbf{p}}) = \int_{\mathbf{R}^3}dP(\mathbf{k})n(\mathbf{k})$ associates natural numbers with the spectrum points $\mathbf{k}\in\Gamma$ (cf. Sec. VI A).

(5) In contrast with the irreducibility hypothesis of CQFT, the system $(\hat{\phi},\hat{\pi}) = (\hat{\phi}_1,...,\hat{\phi}_N, \hat{\pi}_1,...,\hat{\pi}_N)$ of self-adjoint operators is only a $\mathscr{B}$-irreducible system [in a straightforward sense according to Definition 3.1, i.e., the set of operators in $\hat{B}(\mathscr{H}_{\hat{A}})$ commuting with all the numbers of $(\hat{\phi},\hat{\pi})$ equals $\hat{B}$] in QLFT. This irreducibility notion is clearly weaker than the usual irreducibility notion and this is the main difference between CQFT and our model named QLFT. However, the C^*-algebra $B(\mathscr{H}_A) = B(\mathscr{H})\otimes A$ associated with the local bounded observables of QLFT is a factor, thus the local state space $\mathscr{H}_A$ describes the quantized system *coherently*; the unbounded extension $\hat{B}(\mathscr{H}_{\hat{A}}) = \hat{B}(\mathscr{H})\otimes\hat{A}$ of $B(\mathscr{H}_A)$ is generated not only by the local observables $(\hat{\phi}(\hat{\mathbf{p}}),\hat{\pi}(\hat{\mathbf{p}}),\widehat{\nabla\phi}(\hat{\mathbf{p}}),\widehat{\nabla\pi}(\hat{\mathbf{p}}))$ but by the three-position operator $\hat{\mathbf{x}}$, too. The physical meaning of $\hat{\mathbf{x}}$ is to measure the positions in three-space of the q mechanical systems constituting QLFT, at the instant t of the preparation of the q system. This is in accordance with the c theory, where, in the canonical formalism, the three-position vector $\mathbf{x}$ also occurs explicitly in the local observables of the system (e.g., in the angular momentum density $\mathscr{M}_{\mu\nu\rho} = T_{\mu\rho}x_\nu - T_{\mu\nu}x_\rho$, where $T_{\mu\nu}$ is the energy-momentum density) and $\mathbf{x}$ is not a function of the basic local dynamical variables (ϕ,π).

IV. THE GLOBAL DESCRIPTION OF THE QUANTUM LOCAL SYSTEM

In keeping with our strategy (cf. Sec. I and Refs. 12 and 16) the q system globally is described via integrations over the sets of information obtained by local measurements in q space-time, i.e., over the local state space $\mathscr{H}_A$ over q space-time. Measures in q space-time are determined by Gleason's theorem.[15,34] We consider those measures on $\mathscr{P}$ that correspond to positive semidefinite self-adjoint operators $\rho \in A$ of the form

$$\rho = 1 \text{ or } \rho = \sum_{n=1}^{\infty} c_n p_n, \quad 0 < c_n < 1,$$
$$\sum_n c_n = 1, \quad p_n \perp p_m, \quad n \neq m,$$
$$\dim p_n = 1, \quad \sum_{n=1}^{\infty} p_n = 1 . \tag{4.1}$$

In the second case the ρ's are probability measures on $\mathscr{P}$. The global state spaces are obtained by averaging, with the aid of the above measures, over the local state space $\mathscr{H}_A$ as follows:

$$H^{\rho} = \operatorname{Tr} \rho \mathscr{H}_A := \{ \phi | \phi \in \mathscr{H} \otimes A, \ \operatorname{Tr} \rho \langle \phi | \phi \rangle_A < \infty \} \tag{4.2}$$

and by equipping them with the $\mathbb{C}$-valued scalar product

$$\langle \phi_1 | \phi_2 \rangle_\rho := \operatorname{Tr} \rho \langle \phi_1 | \phi_2 \rangle_A = \sum_k \bar{c}_k^1 c_k^2 \times \langle f_k^2 | f_k^2 \rangle \operatorname{Tr} \rho a_{k1}^* a_{k2} , \tag{4.3}$$

where $c_k^1, c_k^2 \in \mathbb{C}$, $f_k^1, f_k^2 \in \mathscr{H}$, and $a_{k1}, a_{k2} \in A$.

Proposition 4.1: H^{ρ} with the scalar product $\langle \cdot | \cdot \rangle_\rho$ is a complex separable Hilbert space. Furthermore, if $\rho \neq 1$ then the map u: $H^{\rho} \to H^1$; $\phi \to \phi \sqrt{\rho}$ is a unitary map.

Proof: (1) Let $\operatorname{Tr} A := \{a | a \in A, \ \operatorname{Tr} a^* a < \infty\}$. Then $\operatorname{Tr} A$ with the scalar product $\langle a_1 | a_2 \rangle_1 := \operatorname{Tr} a_1^* a_2$ is the separable Hilbert space of the Hilbert–Schmidt operators in A.[35]

(2) Let $\operatorname{Tr} \rho A := \{a | a \in A, \operatorname{Tr} \rho a^* a < \infty\}$ and $\langle a_1 | a_2 \rangle_\rho := \operatorname{Tr} \rho a_1^* a_2$. Then, $a \in \operatorname{Tr} \rho A$ iff $a\sqrt{\rho} \in \operatorname{Tr} A$. For, if $a \in \operatorname{Tr} \rho A$ then $\operatorname{Tr} \rho a^* a = \operatorname{Tr}(a\sqrt{\rho})^*(a\sqrt{\rho}) < \infty$ then $a\sqrt{\rho} \in \operatorname{Tr} A$ and if $a\sqrt{\rho} \in \operatorname{Tr} A$ then $\operatorname{Tr}(a\sqrt{\rho})^*(a\sqrt{\rho}) = \operatorname{Tr} \rho a^* a < \infty$ then $a \in \operatorname{Tr} \rho A$. Taking this into account, it can easily be checked that $\operatorname{Tr} \rho A$ is a complex vector space and $\langle | \rangle_\rho$ is a complex scalar product on it. To this latter we only note that $\{a | a \in \operatorname{Tr} \rho A, \operatorname{Tr} \rho a^* a = 0\} \equiv \{0\}$. For, $\operatorname{Tr} \rho a^* a = 0$ iff $a\sqrt{\rho} = 0$. Beside the form of ρ's in (4.1) this equation has the only solution $a = 0$, which can be seen from the equation

$$\langle \phi | (a\sqrt{\rho})^*(a\sqrt{\rho}) | \phi \rangle = \sum_n c_n \langle a p_n \phi | a p_n \phi \rangle = 0, \ \forall \phi \in L^2(\mathbb{R}^3) .$$

Finally $\operatorname{Tr} \rho A$ is complete in the norm $\|a\|_\rho = (\operatorname{Tr} \rho a^* a)^{1/2}$. For, let (a_i) be a Cauchy sequence in $\operatorname{Tr} A$. Then

$$\lim_{i,j \to \infty} \|a_i - a_j\|_\rho = 0 \quad \text{iff} \quad \lim_{i,j \to \infty} \|a_i \sqrt{\rho} - a_j \sqrt{\rho}\|_1 = 0$$

then

$$\exists a' = a\sqrt{\rho} \in \operatorname{Tr} A, \ \lim_{i \to \infty} \|a_i \sqrt{\rho} - a\sqrt{\rho}\|_1 = 0$$

$$\text{iff} \ \lim_{i \to \infty} \|a_i - a\|_\rho = 0 ,$$

then $a \in \operatorname{Tr} \rho A$.

(3) Let us take the tensor product of $\mathscr{H}$ and $\operatorname{Tr} \rho A$, then from (4.2) and (4.3) we have $\mathscr{H} \otimes \operatorname{Tr} \rho A = \operatorname{Tr} \rho(\mathscr{H} \otimes A) = H^{\rho}$. The second part of the proposition is clear from (4.3). Q.E.D.

We observe the following. (a) H^{ρ} is an invariant structure over q space-time because $\langle \phi_1 | \phi_2 \rangle_\rho$ is invariant under symmetry transformations $\rho' = u \rho u^{-1}$, $\langle \phi_1 | \phi_2 \rangle'_A = u \langle \phi_1 | \phi_2 \rangle_A u^{-1}$ in q space-time, where u is a unitary operator in A.[15,16]

(b) For the bounded linear operators in H^{ρ} we have $B(H^{\rho}) = B(\mathscr{H}) \otimes B(\operatorname{Tr} \rho A) \cong B(\mathscr{H}) \otimes A$. Then a bounded global observable (represented by a bounded self-adjoint operator in H^{ρ}) corresponds to every local bounded observable and conversely. For the unbounded extension of $B(H^{\rho})$ we have

$$\widehat{B}(H^{\rho}) = \widehat{B}(\mathscr{H}) \otimes \widehat{B}(\operatorname{Tr} \rho A) \cong \widehat{B}(\mathscr{H}) \otimes \widehat{A}$$

with the domain $\mathscr{D} = S(\mathbb{R}^N) \otimes d$, where

$$d := \overline{\left\{ \phi | \ \phi \in \operatorname{Tr} \rho A, \ \phi = \sum_k c_k | f_k \rangle \langle g_k |, \quad c_k \in \mathbb{C}, \quad f_k, g_k \in S(\mathbb{R}^3) \right\}}$$

[the summation for k means a finite sum and the closure is understood in the product topology of $S(\mathbb{R}^3) \otimes S(\mathbb{R}^3)^2$] (cf. Note 1 below).

(c) Let

$$\operatorname{Tr} \mathscr{H}_B := \{ \phi | \phi \in \mathscr{H} \otimes B, \quad \operatorname{Tr} \langle \phi | \phi \rangle_A < \infty \} .$$

Then, by Ref. 28,

$$\operatorname{Tr} \mathscr{H}_B \cong \int_\Gamma \oplus \mathscr{H}(\mathbf{k}) d^3\mathbf{k} = H ,$$

where $\Gamma = \operatorname{Sp} \mathscr{B}$ $(= \mathbb{R}^3)$. Furthermore

$$B(H) = B(\operatorname{Tr} \mathscr{H}_B) = B(\mathscr{H}) \otimes B$$

and

$$\widehat{B}(H) \cong \widehat{B}(\operatorname{Tr} \mathscr{H}_B) \cong \widehat{B}(\mathscr{H}) \otimes \widehat{B}.$$

It is clear that $\operatorname{Tr} \mathscr{H}_B$ is a closed subspace of $\operatorname{Tr} \mathscr{H}_A$.

In the sense of Proposition 4.1 the global state space of QLFT of local state space $\mathscr{H}_A$ can generally be identified with the separable Hilbert space $H^1 = \operatorname{Tr} \mathscr{H}_A$. However, if we know that the joint distribution in three-position and three-momentum space of the infinite collection of the connected q mechanical systems constituting the whole system is characterized by the statistical operator ρ of the type (4.1) during our study of the system, then the appropriate global state space is the separable Hilbert space $H^{\rho} = \operatorname{Tr} \rho \mathscr{H}_A$. We can apply such an H^{ρ} when the apparatuses measuring the system are all in the q statistical state characterized by ρ (these apparatuses measure the system locally in q space-time[13,15]). This means that the collections of local measuring apparatuses (sitting at the points of q space-time[15]) defining the measuring apparatuses[13] are prepared in such a way that they be in the state described by ρ concerning their data in three-space and three-momentum space. These apparatuses measure the q local fields and the local observables generated by them, in a given instant, and not the operators $\hat{\mathbf{x}}$ and $\hat{\mathbf{p}}$ (see Ref. 16 and Note 5 below).

We see, by defining the *global states* by the rays of H^{ρ}, that a local state $\mathbf{\Phi}$ in $\mathscr{H}_A$ generates a global state in H^{ρ} if $\rho \neq 1$. For, $\langle \Phi|\Phi\rangle_A = 1$, $\Phi \in \mathbf{\Phi}$ and $\operatorname{Tr}\rho\langle\Phi|\Phi\rangle_A = \operatorname{Tr}\rho \cdot 1 = 1$, i.e., $\Phi \in H^{\rho}$ and $\langle\Phi|\Phi\rangle_{\rho} = 1$. But, if $\rho = 1$ then the local state $\mathbf{\Phi}$ does not generate a global state in H^1 since in that case $\langle\Phi|\Phi\rangle_1 = \operatorname{Tr}\langle\Phi|\Phi\rangle_A = \operatorname{Tr} 1 = \infty$. In the formation of the local state each q mechanical system (the collection of them forms the whole system) participates with unit statistical weight, the "norm" of $\mathbf{\Phi}$ is the unity, 1, in A.[16] However, in the formation of the global state of the q system these q mechanical systems participate with a statistical weight lying in between 0 and 1 and this is described by a statistical operator ρ of the type (4.1). Then we obtain the global state of the q system from the local one in such a way that we "renormalize" $\mathbf{\Phi}$ with ρ in the following way. Let $\rho = a^*a$, $a \in \operatorname{Tr} A$, and $\operatorname{Tr} a^*a = 1$. Then $\Psi = \Phi a$, $\Phi \in \mathbf{\Phi}$, already generates a global state in H^1 since $\langle\Psi|\Psi\rangle_1 = \operatorname{Tr}\langle\Psi|\Psi\rangle_A = \operatorname{Tr} a^*\langle\Phi|\Phi\rangle_A a = \operatorname{Tr} a^*a = 1$. Thus we can write for the global state $\mathbf{\Psi}$ that $\mathbf{\Psi} = \mathbf{\Phi}a$, i.e., two elements of $\mathbf{\Psi}$ differ also from each other by a unitary operator $u \in A$. The decomposition $\rho = a^*a$ is also unique only up to a unitary operator, i.e., ua provides the same ρ as a since $(ua)^*(ua) = a^*u^{-1}ua = a^*a = \rho$. We can melt this ambiguity into $\mathbf{\Phi}$ and for the sake of unambiguity we can choose $\sqrt{\rho}$ for a, i.e., $a = \sqrt{\rho}$. Consequently, a local state $\mathbf{\Phi}$ determines a global state of the q system together with the specification of a statistical operator ρ. Here $\rho = \rho(\hat{\mathbf{x}},\hat{\mathbf{p}})$ can be considered as the q mechanical distribution function of the q mechanical systems forming the q system, in the phase space $\mathbb{R}^6$; ρ describes the joint distribution of these q mechanical systems in three-space and three-momentum space.

Thus, in conclusion, as to the interpretation of the global state space H^{ρ} we note again that QLFT consists of a "trivial noncommutative bundle" of infinitely many q mechanical systems of state space $\mathscr{H}$. So, when $\rho = \Sigma\, c_n\, p_n$, we can say that these infinitely many q mechanical systems are described globally by H^{ρ} by the aid of q statistical mechanics. Here ρ describes the q statistical state of measuring apparatuses (measuring the system locally in q space-time),[13,16] i.e., a common preparation of these aparatuses on a spacelike hypersurface Γ, the set of atoms of which is $\Omega = \{p_n, n \in \mathbb{R}\}$.[15] The statistical weight of the q mechanical system at the point $p_n \in \Omega$ is c_n. Thus a global state describes the q system when the system is measured in the q statistical state ρ. In this case every local state $\mathbf{\Phi}$ in $\mathscr{H}_A$ generates a global state in H^{ρ}. Finally let us collect the rules of the global description of QLFT (cf. Ref. 12).

(a1) *The global state space* of QLFT of local state space $\mathscr{H}_A$ in the measuring procedures characterized by the statistical operator $\rho \in A$ of the type (4.1) is the complex separable Hilbert space $H^{\rho} = \operatorname{Tr}\rho\mathscr{H}_A$; its *global states* are described by the rays of H^{ρ}.

(a2) The *global observables* are represented by self-adjoint operators in H^{ρ}.

(a3) The *expectation value* of the global observable $\hat{f}$ generated by the local one, $\widehat{F}$, in the global state $\boldsymbol{\phi}$ generated by the local one, $\mathbf{\Phi}$, is

$$\bar{f}_{\rho} = \langle \phi|\hat{f}|\phi\rangle_{\rho} = \operatorname{Tr}\rho\langle \Phi|\widehat{F}|\Phi\rangle_A, \quad \Phi \in \mathbf{\Phi}. \qquad (**)$$

As we said, H^{ρ} carries all the information on the infinite collection of connected q mechanical systems constituting QLFT, which are obtained in the common q statistical state ρ of the local measuring apparatuses[13] in q space-time.[15,16]

Notes: (1) The elements of A and $\hat{A}$ define, in a natural way, linear operators in $\operatorname{Tr} A$ with the correspondence $\hat{f}(a) = f \cdot a$, where $a \in \operatorname{Tr} A$ and $f \in A$ or $f \in \hat{A}$. It is clear from the definition of $\operatorname{Tr} A$ that the system of self-adjoint operators, $(\hat{f}_{\hat{\mathbf{x}}}, \hat{f}_{\hat{\mathbf{p}}}) = (\hat{\mathbf{x}}\cdot, \hat{\mathbf{p}}\cdot)$, corresponding to the system $(\hat{\mathbf{x}}, \hat{\mathbf{p}})$, is irreducible in $\operatorname{Tr} A$.

(2) Let $\{\Phi_n\}$ be a complete orthonormal system in $\mathscr{H}_A$, i.e., this determines a complete orthogonal family of local states in $\mathscr{H}_A$. Now, if we specify the same ρ to every member of this family then the appropriate global state space of the q system is provided by the Hilbert space $H^{\rho} = \operatorname{Tr}\rho\mathscr{H}_A$; on the other hand, if we specify a ρ_n to each Φ_n then the appropriate global state space is $H^1 = \operatorname{Tr}\mathscr{H}_A$. In this case the family $\{\Psi_n\} = \{\Phi_n\sqrt{\rho_n}\}$ gives a complete orthonormal system in H^1, i.e.,

$$\langle\Psi_n|\Psi_m\rangle_1 = \operatorname{Tr}\sqrt{\rho_n}\langle\Phi_n|\Phi_m\rangle_A\sqrt{\rho_m} = \delta_{nm}\,,$$

$$\sum_n |\Psi_n\rangle\langle\Psi_n| = 1 \qquad (4.4)$$

[of course, the completeness is only satisfied if every ρ_n is of the form (4.1)]. Thus $\{\Psi_n\}$ provides a complete orthogonal family of global states in H^1. Then we can write a vector Ψ of unit norm in H^1 in the following form:

$$\Psi = \sum_n c_n\, \Phi_n\, \sqrt{\rho_n}, \quad c_n \in \mathbb{C}, \quad \sum_n |c_n|^2 = 1, \qquad (4.5)$$

where

$$c_n = \langle\Psi_n|\Psi\rangle_1, \quad \Psi_n = \Phi_n\sqrt{\rho_n} \qquad (4.6)$$

is the complex amplitude of the transition probability from the global state Ψ to the global state Ψ_n (cf. below).

(3) Let $\Psi_1 = \Phi_1\sqrt{\rho_1}$ and $\Psi_2 = \Phi_2\sqrt{\rho_2}$ be two rays in H^1. We obtain the transition amplitude from Ψ_1 to Ψ_2 in the following way:

$$\langle\Psi_1|\Psi_2\rangle_1 = \operatorname{Tr}\langle\Psi_1|\Psi_2\rangle_A = \operatorname{Tr}\sqrt{\rho_1}\langle\Phi_1|\Phi_2\rangle_A\sqrt{\rho_2}$$
$$= \operatorname{Tr}\sqrt{\rho_2\rho_1}\cdot\langle\Phi_1|\Phi_2\rangle_A\,.$$

But this expression is not unambiguous. Let $\Phi_1' = \Phi_1 u_1 \in \mathbf{\Phi}_1$ and $\Phi_2' = \Phi_2 u_2 \in \mathbf{\Phi}_2$, then

$$\langle\Psi_1'|\Psi_2'\rangle_1 = \operatorname{Tr}\sqrt{\rho_1}\cdot u_1^{-1}\langle\Phi_1|\Phi_2\rangle_A u_2\sqrt{\rho_2}$$
$$= \operatorname{Tr} u_2\sqrt{\rho_2\rho_1}\, u_1^{-1}\langle\Phi_1|\Phi_2\rangle_A\,,$$

and this is only equal to the former expression if $\sqrt{\rho_2\rho_1} = u_2\sqrt{\rho_2\rho_1}u_1^{-1}$. This problem occurs in (4.6) in such a way that, if we choose the element $\Psi_n' = \Phi_n u^{-1}\rho_n$ from $\mathbf{\Psi}_n$, then we obtain $\langle\Psi_n'|c_n\,\Psi_n\rangle_1 = c_n \operatorname{Tr}\rho_n u$ and $\operatorname{Tr}\rho_n u$ in general does not equal 1 or $e^{-i\varphi}$, $\varphi \in \mathbb{R}$, and $|c_n \operatorname{Tr}\rho_n u|^2 \neq |c_n|^2$. This phenomenon follows from the fact that the transition probability in between two vectors of the global state $\mathbf{\Psi} = \mathbf{\Phi}\sqrt{\rho}$ is not 1 as in QM but $|\langle\Psi_1|\Psi_2\rangle_1|^2 = |\operatorname{Tr}\rho u|^2$, $\Psi_2 = \Phi_1 u\sqrt{\rho}$, and that gives, in general, nontrivial transition probability from Ψ_1 to Ψ_2.

(4) Let us denote the self-adjoint operator in H^{ρ} generated by the local observable $\hat{F}$ in the following way:

$$\hat{f}_\rho := \operatorname{Tr} \rho \hat{F}, \quad \hat{f} := \hat{f}_1 = \operatorname{Tr} \hat{F}, \tag{4.7}$$

the meaning of which is as follows:

$$\langle \Phi_1 | \hat{f}_\rho | \Phi_2 \rangle_\rho = \operatorname{Tr} \rho \langle \Phi_1 | \hat{F} | \Phi_2 \rangle_A, \quad \Phi_1, \Phi_2 \in H^{\rho}$$

(of course Φ_2 is in the domain of $\hat{F}$). With this notation an $\hat{f}_\rho$ corresponding to an $\hat{F}$ of the form (3.15) is of the following form:

$$\hat{f}_\rho = \operatorname{Tr} \rho \hat{F} = \sum_n \operatorname{Tr} \rho \int_{\mathbb{R}^3} dP(\mathbf{k}) E_n(\mathbf{k}) f_n(\mathbf{k}) + \operatorname{Tr} \rho \times \int_{\mathbb{R}^3} dP(\mathbf{k}) \int_{-\infty}^{+\infty} dE(\mathbf{k},\lambda) f(\mathbf{k},\lambda). \tag{4.8}$$

We can write these operators, in the case $\rho = 1$, in the subspace $H = \int_{\mathbb{R}^3} \oplus \mathscr{H}(\mathbf{k}) d^3\mathbf{k}$ of $H^1 = \operatorname{Tr} \mathscr{H}_A$ in the following form[36]:

$$\hat{f} = \hat{f}_1|_H = \operatorname{Tr}_H \hat{F} = \int_{\mathbb{R}^3}^{\oplus} \hat{F}(\mathbf{k}) d^3\mathbf{k} = \sum_n \int_{\mathbb{R}^3}^{\oplus} E_n(\mathbf{k}) f_n(\mathbf{k}) d^3\mathbf{k} + \int_{\mathbb{R}^3}^{\oplus} \int_{-\infty}^{+\infty} dE(\mathbf{k},\lambda) f(\lambda,\mathbf{k}) d^3\mathbf{k}. \tag{4.9}$$

(5) As we said in Sec. III (and in Ref. 16), the q local fields are operator-valued functions $\mathbf{f} \to \hat{\phi}_\alpha(\mathbf{f}) = p(f) \times \hat{\phi}_\alpha p(f)$ of the points $\mathbf{f}$ of q space-time. By the relations (3.3) one can write

$$\hat{\phi}_\alpha(\mathbf{f}) = \hat{q}_\alpha \otimes p(f) \hat{c}_\alpha p(f) = \hat{q}_\alpha \otimes c_{\mathbf{f}}^{\alpha} p(f),$$

where $c_{\mathbf{f}}^{\alpha} = \langle f | \hat{c}_\alpha | f \rangle \in \mathbb{C}$.

[Clearly

$$\hat{\phi}_\alpha(\mathbf{f}) = \hat{\phi}_\alpha(p) \in \hat{B}(\eta^{-1}(p(f))) = \hat{B}(\mathscr{H}) \otimes p(f),$$

where $\eta^{-1}(p(f)) = \mathscr{H}_p = \mathscr{H} \otimes p(f)$ is the fiber at the point $p = p(f)$ of the "noncommutative Hilbert bundle η: $\mathscr{H}_A \to L^2(\mathbb{R}^3)$" over q space-time (see Note 2 in Sec. II).] Thus a member of an apparatus[13] measuring the q local field $\hat{\phi}_\alpha$, which sits at the q space-time point $p = p(f)$, measures the operator $\hat{q}_\alpha$.

V. THE LOCAL DESCRIPTION OF THE DYNAMICS

We have given the local and global kinematical description of QLFT in the foregoing two sections. Now we turn to the question of the dynamics. To specify the dynamics of the q local system in accordance with our guiding principle, the locality (cf. Sec. I), we use the following motivation from CLFT.

Let us implement a Legendre transformation on the Lagrangian density $\mathscr{L}$ by applying the c field equations. We get

$$d\left(\sum_{\alpha=1}^{N} \pi_\alpha \dot{\phi}_\alpha - \mathscr{L}\right) = \sum_{\alpha=1}^{N} \left[-\left(\dot{\pi}_\alpha - \nabla \frac{\partial \mathscr{L}}{\partial \nabla \phi_\alpha}\right) d\phi_\alpha + \dot{\phi}_\alpha \, d\pi_\alpha - \frac{\partial \mathscr{L}}{\partial \nabla \phi_\alpha} d\nabla \phi_\alpha \right], \tag{5.1}$$

where $\nabla = \nabla_{\mathbf{x}}$, hence we obtain the canonical equations[16]

$$-\dot{\pi}_\alpha = \frac{\partial \mathscr{H}}{\partial \phi_\alpha} - \nabla \frac{\partial \mathscr{H}}{\partial \nabla \phi_\alpha}, \quad \dot{\phi}_\alpha = \frac{\partial \mathscr{H}}{\partial \pi_\alpha}, \quad \alpha = 1,\ldots,N, \tag{5.2}$$

where $\mathscr{H} = \mathscr{H}(\phi,\pi,\nabla\phi) = \Sigma_{\alpha=1}^{N} \pi_\alpha \dot{\phi}_\alpha - \mathscr{L}$ is the *Hamiltonian density* [cf. these equations with Eqs. (17.7) on p. 118 in Ref. 23]. Equations (5.2) are equivalent to the c field equations. By applying a Poisson bracket of the form (2.1) we can write (5.2) as follows:

$$\dot{\pi}_\alpha = \{\mathscr{H}, \pi_\alpha\}, \quad \dot{\phi}_\alpha = \{\mathscr{H}, \phi_\alpha\}, \quad \alpha = 1,\ldots, N, \tag{5.3}$$

and, in general, the equation of motion of a local quantity $F = F(\phi,\pi,\nabla\phi,\nabla\pi,t)$ can be expressed as

$$\frac{dF}{dt} = \frac{\partial F}{\partial t} + \{\mathscr{H}, F\}, \tag{5.4}$$

which, of course, gives back (5.3) for π and ϕ. Then the condition that F be a constant of motion (a conserved local observable) is

$$\frac{\partial F}{\partial t} + \{\mathscr{H}, F\} = 0. \tag{5.5}$$

If F does not depend explicitly on t then F is conserved if $\{\mathscr{H}, F\} = 0$. For example, $\mathscr{H}$ itself is such a conserved local observable, but the momentum density $T_{ok} = \Sigma_\alpha \pi_\alpha \partial_k \phi_\alpha$, $k = 1,2,3$, belonging to the total three-momentum $\mathbf{P}$ is locally not conserved since $\{\mathscr{H}, T_{ok}\} \neq 0$.

By taking into account this canonical formulation[19] of the dynamics of CLFT we define the local Schrödinger time evolution in the local state space $\mathscr{H}_A$ by the one-parameter unitary group $t \to \exp(-i\hat{\mathscr{H}}t)$ in $\mathscr{H}_A$ (cf. Note 2 below) as follows:

$$\Phi(t) = \exp(-i\hat{\mathscr{H}}t)\Phi(0), \quad \Phi \in \mathscr{H}_A, \tag{5.6}$$

where $\hat{\mathscr{H}}$ is the *local Hamiltonian* obtained from the c Hamiltonian density by applying the quantization algorithm:

$$\hat{\mathscr{H}} = \mathscr{H}(\hat{\phi},\hat{\pi},\widehat{\nabla\phi}) = \sum_{\alpha=1}^{N} \left[\frac{1}{2}\hat{\pi}_\alpha^2 + \frac{1}{2}(\widehat{\nabla\phi}_\alpha)^2 + \frac{1}{2} m_\alpha^2 \hat{\phi}_\alpha^2\right] + V(\hat{\phi}_1,\ldots,\hat{\phi}_N) = \frac{1}{2}\sum_{\alpha=1}^{N} [\hat{\pi}_\alpha^2 + \hat{\phi}_\alpha^2 \hat{\mathbf{p}}^2 + m_\alpha^2 \hat{\phi}_\alpha^2] + V(\hat{\phi}_1,\ldots,\hat{\phi}_N), \tag{5.7}$$

where Eq. (2.5) has been used. In differential form this prescription provides the local Schrödinger equation

$$i\frac{\partial \Phi(t)}{\partial t} = \hat{\mathscr{H}}\Phi(t), \quad \Phi \in \mathscr{H}_{\hat{A}}, \tag{5.8a}$$

$$i\frac{\partial \Phi(t)}{\partial t} = \left\{\frac{1}{2}\sum_{\alpha=1}^{N} [\hat{\pi}_\alpha^2 + \hat{\phi}_\alpha^2 \hat{\mathbf{p}}^2 + m_\alpha^2 \hat{\phi}_\alpha^2] + V(\hat{\phi}_1,\ldots,\hat{\phi}_N)\right\}\Phi(t). \tag{5.8b}$$

This specification of the dynamics is confirmed by the following facts[19]: Eq. (5.8b) implies (1) a continuity equation for the (appropriate) L^2-sections of the trivial Hilbert bundle $\mathscr{H}$ over $\mathbb{R}^4$ (cf. Fig. 1), (2) an Ehrenfest theorem (cf.

Sec. VII), and (3) the c limit of (5.8b) can be derived in an analogous way as the c limit of the q mechanical Schrödinger equation (cf. Note 3 in Sec. III).

The formula $\hat{F}(t) = \exp(-i\hat{\mathscr{H}}t)\hat{F}(0)\exp(i\hat{\mathscr{H}}t)$ provides the time evolution of the local observables $\hat{F} = F(\hat{\phi},\hat{\pi},\nabla\hat{\phi},\nabla\hat{\pi},t)$. In differential form it has the form

$$\frac{d\hat{F}}{dt} = \frac{\partial\hat{F}}{\partial t} + i[\hat{\mathscr{H}},\hat{F}]\,. \tag{5.9}$$

The condition that $\hat{F}$ be a conserved local observable is

$$\frac{\partial\hat{F}}{\partial t} + i[\hat{\mathscr{H}},\hat{F}] = 0\,. \tag{5.10}$$

If $\hat{F}$ does not depend explicitly on t then it is conserved when $[\hat{\mathscr{H}},\hat{F}] = 0$.

The local Heisenberg equation of motion (5.9), for $\hat{\phi}$ and $\hat{\pi}$, gives

$$\dot{\hat{\phi}}_\alpha = i[\hat{\mathscr{H}},\hat{\phi}_\alpha] = \frac{\partial\hat{\mathscr{H}}}{\partial\hat{\pi}_\alpha},$$
$$-\dot{\hat{\pi}} = i[\hat{\pi}_\alpha,\hat{\mathscr{H}}] = \frac{\partial\hat{\mathscr{H}}}{\partial\hat{\phi}_\alpha} - i\nabla\left(\frac{\partial\hat{\mathscr{H}}}{\partial\nabla\hat{\phi}_\alpha}\right), \tag{5.11}$$

taking into account Eqs. (2.5) and (3.1). These equations as operator equations have the same form as the c canonical equations (5.2). Using the specific form of $\hat{\mathscr{H}}$ in (5.7) and Eqs. (2.5) and (3.1) we get

$$\dot{\hat{\phi}}_\alpha = \hat{\pi}_\alpha, \quad \dot{\hat{\pi}}_\alpha = -\hat{\phi}_\alpha\,\hat{\mathbf{p}}^2 - m_\alpha^2\,\hat{\phi}_\alpha - \frac{\partial\hat{V}}{\partial\hat{\phi}_\alpha},$$
$$\alpha = 1,\ldots,N, \tag{5.12}$$

or

$$\ddot{\hat{\phi}}_\alpha(t) + \hat{\mathbf{p}}^2\hat{\phi}_\alpha(t) + m_\alpha^2\hat{\phi}_\alpha(t) = -\frac{\partial\hat{V}}{\partial\hat{\phi}_\alpha}(t),$$
$$\alpha = 1,\ldots,N. \tag{5.12'}$$

By taking into account that in position representation $\mathbf{p}^2 = -\Delta$, we have obtained that the c field equations belonging to the Lagrangian (1.1) recover in the quantized theory as operator equations [cf. Sec. VII and Eqs. (7.12)]. In contrast with the corresponding dynamical equations in CQFT[5,6] these equations are well-defined operator equations. In fact, not only the left-hand side of Eqs. (5.12) is well defined as in CQFT but their right-hand side, too. For, if V is an analytical function of ϕ_α then $\partial V/\partial\phi_\alpha$ is, too, and thus by the relation (2.5) we see that $\partial\hat{V}/\partial\hat{\phi}_\alpha$ is a well-defined element of the operator *-algebra $\hat{B}(\mathscr{H})\otimes\hat{B}$, which is a subalgebra of $\hat{B}(\mathscr{H}_{\hat{A}}) = \hat{B}(\mathscr{H})\otimes\hat{A}$.

In order to know completely the dynamics of the quantized system, one should solve the local Schrödinger equation (5.8) or, equivalently, the local Heisenberg equations (5.11) or concretely Eqs. (5.12). For local states of the form $\Psi(t) = \Phi\exp(-i\hat{e}t)$ we get the local eigenvalue equation

$$\hat{\mathscr{H}}\Phi = \Phi\hat{e}, \quad \hat{e}\in\hat{B}\subset\hat{A}, \quad \Phi\in\mathscr{H}_{\hat{\mathscr{B}}}\subset\mathscr{H}_{\hat{A}}\,,$$
$$\left[\frac{1}{2}\sum_{\alpha=1}^{N}(\hat{\pi}_\alpha^2 + \hat{\phi}_\alpha^2\,\hat{p}_{0\alpha}^2) + V(\hat{\phi}_1,\ldots,\hat{\phi}_N)\right]\Phi = \Phi\hat{e}, \tag{5.13}$$

where $\hat{p}_{0\alpha}^2 = \hat{\mathbf{p}}^2 + m_\alpha^2$. This eigenvalue equation formulated in terms of $\mathscr{H}_{\hat{A}}$ is the natural extension of the q mechanical eigenvalue equation

$$\left[\frac{1}{2}\sum_{\alpha=1}^{N}(\hat{p}_\alpha^2 + \hat{q}_\alpha^2\omega_\alpha^2) + V(\hat{q}_1,\ldots,\hat{q}_N)\right]\Phi = e\Phi \tag{5.13'}$$

formulated in terms of a complex Hilbert space. Thus we can adapt, in a straightforward way, the well-developed methods for solving Eq. (5.13') to Eq. (5.13). One can think of Eq. (5.13) as the *natural extension* of the q mechanical eigenvalue [Eq. (5.13')] given in terms of the universe V into the universe $V^{(\mathscr{P})}$ of q set theory.[22]

Notes: (1) Equations (5.7) and (3.3) show that the local Schrödinger evolution operator $U(t) = \exp(-i\hat{\mathscr{H}}t)$ maps $\mathscr{H}^{(\mathscr{B})} = \mathscr{H}_{\hat{B}}$ in $\mathscr{H}^{(\mathscr{P})} = \mathscr{H}_{\hat{A}}$ onto itself, hence all the spectral values $\hat{e}$ of $\hat{\mathscr{H}}$ in Eq. (5.13) lie in $\hat{B}\subset\hat{A}$.

(2) In the sense of the basic theorem of $\mathscr{B}$-valued analysis[31,14] (see Theorem B1 in Appendix B), Stone's theorem[37] holds true in the subuniverse $V^{(\mathscr{B})}$ of the universe $V^{(\mathscr{P})}$. Thus the local Hamiltonian $\hat{\mathscr{H}}$ in (5.7) ($\hat{\mathscr{H}}\in\hat{B}(\mathscr{H})\otimes\hat{B}\subset\hat{B}(\mathscr{H}_{\hat{A}})$) and the one-parameter unitary group $t\to U(t) = \exp(-i\hat{\mathscr{H}}t)$ mutually determine each other in the subspace $\mathscr{H}_B$ of $\mathscr{H}_A$.

VI. PERTURBATION THEORY USING THE INTERACTION PICTURE

One of the best developed methods for solving the Schrödinger equation corresponding to (5.13') is the perturbation theory in the interaction picture. We now extend this method formulated in V to the local Schrödinger equation (5.8) formulated in $V^{(\mathscr{P})}$.

The interaction picture is appropriate to use it for the solution of the dynamical problem when the local Hamiltonian is of the form $\hat{\mathscr{H}} = \hat{\mathscr{H}}_0 + \hat{\mathscr{H}}_I$, where $\hat{\mathscr{H}}_0$ leads to exactly solvable problem ("free fields") while $\hat{\mathscr{H}}_I$ (the interaction term) is "appropriately small" with respect to $\hat{\mathscr{H}}_0$ (cf. Refs. 4 and 38). In our example

$$\hat{\mathscr{H}}_0 = \frac{1}{2}\sum_{\alpha=1}^{N}(\hat{\pi}_\alpha^2 + \hat{\phi}_\alpha^2\,\hat{\mathbf{p}}^2 + m_\alpha^2\,\hat{\phi}_\alpha^2),$$
$$\hat{\mathscr{H}}_I = V(\hat{\phi}_1,\ldots,\hat{\phi}_N)\,. \tag{6.1}$$

Denoting the quantities in the Heisenberg picture by the prefix H, we have the following basic relations. The Heisenberg picture and the interaction picture is connected by the unitary map $U(t,t_0)$ in $\mathscr{H}_A$ and contracted at the instant $t = t_0$ as follows:

$$\Psi(t) = U(t,t_0)\Psi^H, \quad \Psi^H\in\mathscr{H}_A, \quad \Psi(t_0) = \Psi^H,$$
$$\hat{F}(t) = U(t,t_0)\hat{F}^H(t_0)U^{-1}(t,t_0), \quad \hat{F}(t_0) = \hat{F}^H(t_0)\,. \tag{6.2}$$

The local states $\Psi(t)$ satisfy the Tomonaga–Schwinger equation

$$i\frac{\partial\Psi(t)}{\partial t} = \hat{\mathscr{H}}_I(t)\Psi(t)\,, \tag{6.3}$$

while the local observables $\hat{F}(t)$ are governed by the equation

$$\frac{\partial\hat{F}(t)}{\partial t} = i[\hat{\mathscr{H}}_0(t),\hat{F}(t)]\,. \tag{6.4}$$

The evolution operator $U(t,t_0)$ satisfies the equation

$$i\frac{\partial U(t,t_0)}{\partial t} = \hat{\mathscr{H}}_I(t)U(t,t_0), \quad U(t_0,t_0) = \mathbb{1}. \tag{6.5}$$

Then we observe that first, from (6.4), the local fields $\hat{\phi}_\alpha(t)$ satisfy free field dynamical equations and second, from (6.1) and (2.5), $U(t,t_0)$ lies in $B(\mathscr{H}_B)$ $[=B(\mathscr{H})\otimes B \subset B(\mathscr{H}_A)]$ and maps $\mathscr{H}_B$ onto itself in $\mathscr{H}_A$. The formal solution of Eq. (6.5) is given by the von Neumann–Liouville series

$$U(t,t_0)=\mathbb{1}+\sum_{n=1}^{\infty}(-i)^n\frac{1}{n!}\times\int_{t_0}^{t}dt^1\cdots dt^n\,T(\hat{\mathscr{H}}_I(t')\cdots\hat{\mathscr{H}}_I(t^n))$$
$$=T\sum_{n=0}^{\infty}\frac{1}{n!}\left[-i\int_{t_0}^{t}\hat{\mathscr{H}}_I(t')dt'\right]^n, \qquad (6.6)$$

where T is Dyson's T product. However, this $U(t,t_0)$ is not well-defined, in the same way as its q mechanical counterpart, on account of the formal definition of Dyson's T product (cf. Ref. 4, pp. 271–273).

A. Free fields

By Eqs. (5.12), (6.1), and (6.4) we have the operator KG equations

$$\ddot{\hat{\phi}}_\alpha(t)+\hat{\phi}_\alpha(t)\hat{\mathbf{p}}^2+m_\alpha^2\hat{\phi}_\alpha(t)=0,\quad \alpha=1,\ldots,N, \qquad (6.7)$$

for the fields $\hat{\phi}_\alpha$. The solution of these equations is

$$\hat{\phi}_\alpha(t)=\hat{\phi}_\alpha^-(t)+\hat{\phi}_\alpha^+(t)=\hat{a}_\alpha\otimes\hat{c}_\alpha(t)+\hat{a}_\alpha^+\otimes\hat{c}_\alpha^*(t), \qquad (6.8)$$

where

$$\hat{c}_\alpha(t)=(2\hat{p}_0^\alpha)^{-1/2}e^{-i\hat{p}_0^\alpha t}=\int_{\mathbf{R}^3}\frac{1}{\sqrt{2k_0^\alpha}}e^{-ik_0^\alpha t}\,dP(\mathbf{k}), \qquad (6.9)$$

where $\hat{p}_0^\alpha=(\hat{\mathbf{p}}^2+m_\alpha^2)^{-1/2}$, $k_0^\alpha=\sqrt{\mathbf{k}^2+m_\alpha^2}$, and

$$[\hat{a}_\alpha,\hat{a}_\beta]=[\hat{a}_\alpha^+,\hat{a}_\beta^+]=0,\quad [\hat{a}_\alpha,\hat{a}_\beta^+]=\delta_{\alpha\beta}\mathbb{1} \qquad (6.10)$$

($\hat{a}_\alpha^+$ denotes the adjoint of $\hat{a}_\alpha$). The operators $(\hat{a},\hat{a}^+)=(\hat{a}_1,\ldots,\hat{a}_N,\hat{a}_1^+,\ldots,\hat{a}_N^+)$ of course constitute a $\mathscr{B}$-irreducible system in $\mathscr{H}_{\hat{A}}$ and thus they are functions of $\hat{\mathbf{p}}$, i.e., $\hat{a}_\alpha=\hat{a}_\alpha(\hat{\mathbf{p}})\in\hat{B}(\mathscr{H})\otimes\hat{B}$, $\alpha=1,\ldots,N$ (cf. Note 5 in Sec. III and see Note 1 below). By a formal calculation $\hat{\mathscr{H}}_0$ gets the form

$$\hat{\mathscr{H}}_0=\sum_{\alpha=1}^{N}\left(\hat{N}_\alpha+\frac{1}{2}\right)\hat{p}_0^\alpha,\quad \hat{N}_\alpha=\hat{a}_\alpha^+\hat{a}_\alpha. \qquad (6.11)$$

In the Fock representation of (6.10), $\mathscr{H}_{\hat{A}}$ is spanned by the Fock basis,

$$\Phi_{\hat{n}_1,\ldots,\hat{n}_N}=(\hat{n}_1!\ldots\hat{n}_N!)^{-1/2}[\hat{a}_1^+(\hat{\mathbf{p}})]^{\hat{n}_1}\cdots[\hat{a}_N^+(\hat{\mathbf{p}})]^{\hat{n}_N}\Phi_0(\hat{\mathbf{p}})$$
$$=\int_{\mathbf{R}^3}dP(\mathbf{k})[n_1(\mathbf{k})!\cdots n_N(\mathbf{k})!]^{-1/2}\times[\hat{a}_1^+(\mathbf{k})]^{n_1(\mathbf{k})}\cdots[\hat{a}_N^+(\mathbf{k})]^{n_N(\mathbf{k})}\Phi_0(\mathbf{k}), \qquad (6.12)$$

where $\hat{n}_\alpha$ is a non-negative integer-valued function of $\hat{\mathbf{p}}$ [i.e., $\hat{n}_\alpha:\Gamma\to\mathrm{N}\cup\{0\}$, $\Gamma=\mathrm{Sp}\,\mathscr{B}$ (see Note 4 in Sec. III)] and $\Phi_0(\hat{\mathbf{p}})$ is the unique element of $\mathscr{H}_{\hat{A}}$ with the property

$$\hat{a}_\alpha(\hat{\mathbf{p}})\Phi_0(\hat{\mathbf{p}})=\int_{\mathbf{R}^3}dP(\mathbf{k})\hat{a}_\alpha(\mathbf{k})\Phi_0(\mathbf{k})=0,\quad\forall\alpha.$$

Here $\hat{a}_\alpha$, $\hat{a}_\alpha^+$, and $\hat{N}_\alpha$ act on this basis in the usual manner.[2] Since $\hat{N}_\alpha$ is diagonal on this basis, thus $\hat{\mathscr{H}}_0$ is, too, i.e., the eigenvectors of $\hat{\mathscr{H}}_0$ are the vectors of the form (6.12) in $\mathscr{H}_{\hat{A}}$. The complete orthogonal system (6.12) in $\mathscr{H}_{\hat{A}}$ determines a complete orthogonal family of local states of the q system. One can say that the KG equations (6.7) describe a "trivial fiber bundle" of harmonic oscillators of species N over q space-time, i.e., there are N one-dimensional harmonic oscillators of frequencies $k_0^\alpha=\sqrt{\mathbf{k}^2+m_\alpha^2}$, $\alpha=1,\ldots,N$, at each point $\mathbf{k}$ of $\Gamma=\mathrm{Sp}\,\mathscr{B}$ (of the "spacelike hypersurface" Γ in q space-time[15]). Here $\Phi_0(\mathbf{k})$ is the lowest energy state (vacuum) of the q mechanical system consisting of N one-dimensional harmonic oscillators and sitting at the spectrum point $\mathbf{k}$ of $\hat{\mathbf{p}}$, and $\hat{N}_\alpha(\mathbf{k})$ counts the quanta of energy $\sqrt{\mathbf{k}^2+m_\alpha^2}$ of the harmonic oscillator characterized by the frequency $\omega_\alpha=\sqrt{\mathbf{k}^2+m_\alpha^2}$, thus $\hat{N}_\alpha$ is the *local quantum number operator* and $\hat{N}=\Sigma_{\alpha=1}^N\hat{N}_\alpha$ is the total local quantum number operator. One can say that in the state $\Phi_{n_1(\mathbf{k})\cdots n_N(\mathbf{k})}(\mathbf{k})$ there are $n_1(\mathbf{k}),\ldots,n_N(\mathbf{k})$ pieces of quanta of energy $(\mathbf{k}^2+m_1^2)^{1/2},\ldots,(\mathbf{k}^2+m_N^2)^{1/2}$, respectively (plus the zero-point energy of magnitude $\Sigma_{\alpha=1}^N\frac{1}{2}\sqrt{\mathbf{k}^2+m_\alpha^2}$).

Now let us introduce the local observable

$$\hat{\mathscr{P}}=\sum_{\alpha=1}^{N}\left(\hat{N}_\alpha+\frac{1}{2}\right)\hat{\mathbf{p}}. \qquad (6.13)$$

It is clear that $[\hat{\mathscr{H}}_0,\hat{\mathscr{P}}]=0$, i.e., $\hat{\mathscr{P}}$ is a conserved local observable in the sense of (5.10). We identify the self-adjoint operator $\hat{\mathscr{P}}$ with the *local three-momentum observable* of the q system. The local states belonging to the vectors in (6.12) are also eigenstates of $\hat{\mathscr{P}}$. By calling the ray $\mathbf{\Phi}_0$ the *local vacuum state* of the free q fields $\hat{\phi}_\alpha$, we can give another way of interpretation (beside the above) for the system of free q fields. According to this, the creation operator $\hat{a}_\alpha^+$ creates a KG particle of four-momentum observable $(\hat{p}_0^\alpha,\hat{\mathbf{p}})$ from the local vacuum $\mathbf{\Phi}_0$. Thus in the local state $\mathbf{\Phi}_{n_1,\ldots,n_N}$ there are $n_1,\ldots,n_N$ KG particles of four-momentum observables $(\hat{p}_0^1,\hat{\mathbf{p}}),\ldots$ and $(\hat{p}_0^N,\hat{\mathbf{p}})$, respectively. So the "quanta" of the free q fields $\hat{\phi}_\alpha$ are free KG particles. In this sense the free field description in QLFT is *physically* equivalent with the free field theory of CQFT (cf. still Note 2 below).

With Eqs. (6.8)–(6.10) we obtain the following nonequal time commutators:

$$[\hat{\phi}_\alpha^+(t_1),\hat{\phi}_\beta^+(t_2)]=[\hat{\phi}_\alpha^-(t_1),\hat{\phi}_\beta^-(t_2)]=0,$$
$$[\hat{\phi}_\alpha^+(t_1),\hat{\phi}_\beta^-(t_2)]=-i\hat{\Delta}_\alpha^+(t_1-t_2)\delta_{\alpha\beta}\mathbb{1}=i\hat{\Delta}_\alpha^-(t_2-t_1)\delta_{\alpha\beta}\mathbb{1},$$
$$[\hat{\phi}_\alpha^-(t_1),\hat{\phi}_\beta^+(t_2)]=-i\hat{\Delta}_\alpha^-(t_1-t_2)\delta_{\alpha\beta}\mathbb{1}=i\hat{\Delta}_\alpha^+(t_2-t_1)\delta_{\alpha\beta}\mathbb{1},$$
$$[\hat{\phi}_\alpha(t_1),\hat{\phi}_\beta(t_2)]=-i\hat{\Delta}_\alpha(t_1-t_2)\delta_{\alpha\beta}\mathbb{1}, \qquad (6.14)$$

where $\hat{\Delta}_\alpha(t_1-t_2)$ is given by (2.13) with mass m_α and $\hat{\Delta}_\alpha^-$ and $\hat{\Delta}_\alpha^+$ are the negative and positive frequency parts of $\hat{\Delta}_\alpha$, respectively. The integral kernel representation of the operator $\hat{\Delta}_\alpha(t_1-t_2)$ $(\in\hat{B})$ in $L^2(\mathbf{R}^3)$ is provided by the formula (2.9) with mass m_α. For the normal and ordinary product we have the relation

$$N(\hat{\phi}_\alpha(t_1)\hat{\phi}_\beta(t_2)) = \hat{\phi}_\alpha(t_1)\hat{\phi}_\beta(t_2) - i\hat{\Delta}_\alpha(t_1-t_2)\delta_{\alpha\beta}\mathbb{1}. \tag{6.15}$$

The relation between the T and ordinary product is

$$T(\hat{\phi}_\alpha(t_1)\hat{\phi}_\beta(t_2)) = \hat{\phi}_\alpha(t_1)\hat{\phi}_\beta(t_2) + i\Theta(t_2-t_1)\hat{\Delta}_\alpha(t_1-t_2)\delta_{\alpha\beta}\mathbb{1}. \tag{6.16}$$

The local vacuum expectation values of the ordinary, normal, and T product in the local vacuum Φ_0 are

$$\langle\Phi_0|\hat{\phi}_\alpha(t_1)\hat{\phi}_\beta(t_2)|\Phi_0\rangle_{\hat{A}} = -i\hat{\Delta}^-_\alpha(t_1-t_2)\delta_{\alpha\beta} = i\hat{\Delta}^+_\alpha(t_2-t_1)\delta_{\alpha\beta}, \tag{6.17}$$

$$\langle\Phi_0|N(\hat{\phi}_\alpha(t_1)\hat{\phi}_\beta(t_2))|\Phi_0\rangle_{\hat{A}} = 0,$$

$$\langle\Phi_0|T(\hat{\phi}_\alpha(t_1)\hat{\phi}_\beta(t_2))|\Phi_0\rangle_{\hat{A}} = i\hat{\Delta}^\alpha_F(t_1-t_2)\delta_{\alpha\beta}, \tag{6.18}$$

respectively, where the *propagator* of the free field $\hat{\phi}_\alpha$, $\hat{\Delta}^\alpha_F(t_1-t_2)$, is given by Eq. (2.14) with mass m_α. The integral kernel representation, in $L^2(\mathbb{R}^3)$, of the operator $\hat{\Delta}^\alpha_F \in \hat{B}$ is provided by the formula (2.11) with mass m_α. For the pairing of $\hat{\phi}_\alpha(t_1)$ and $\hat{\phi}_\beta(t_2)$ we have

$$\underbrace{\hat{\phi}_\alpha(t_1)\hat{\phi}_\beta(t_2)} \equiv T(\hat{\phi}_\alpha(t_1)\hat{\phi}_\beta(t_2)) - N(\hat{\phi}_\alpha(t_1)\hat{\phi}_\beta(t_2)) = \langle\Phi_0|T(\hat{\phi}_\alpha(t_1)\hat{\phi}_\beta(t_2))|\Phi_0\rangle_{\hat{A}}\cdot\mathbb{1} = i\hat{\Delta}^\alpha_F(t_1-t_2)\delta_{\alpha\beta}\mathbb{1}. \tag{6.19}$$

We see, as we have expected in Sec. II C, that these relations recover in QLFT as mathematically well-defined equations (up to the ambiguity in the definition of the T product) in contrast with their counterparts in CQFT[1] (some of them are listed in Sec. II C [Eqs. (2.8) and (2.10)]). In the present approach, the linear operators $\hat{\Delta}_\alpha(t_1-t_2)$ and $\hat{\Delta}^\alpha_F(t_1-t_2)$ in $\hat{B}$ occur instead of the "c numbers" $\Delta_\alpha(t_1-t_2, \mathbf{x}_1-\mathbf{x}_2)$ and $\Delta^\alpha_F(t_1-t_2, \mathbf{x}_1-\mathbf{x}_2)$ of CQFT [the latters are the integral kernels in $L^2(\mathbb{R}^3)$ of the formers].

Notes: (1) The relation between the operator families $(\hat{a},\hat{a}^+)$ and $(\hat{\phi},\hat{\pi})$ is as follows:

$$\hat{a}_\alpha = \sqrt{(\hat{p}^\alpha_0/2)}\,[\hat{\phi}_\alpha + i(\hat{p}^\alpha_0)^{-1}\hat{\pi}_\alpha],$$
$$\hat{a}^+_\alpha = \sqrt{(\hat{p}^\alpha_0/2)}\,[\hat{\phi}_\alpha - i(\hat{p}^\alpha_0)^{-1}\hat{\pi}_\alpha]. \tag{6.20}$$

Furthermore, if $\mathscr{H} = L^2(\mathbb{R}^N)$ in $\mathscr{H}_{\hat{A}} = \mathscr{H}\otimes\hat{A}$ then the basis (6.12) is represented by the vectors

$$\Phi_{\hat{n}_1,\ldots,\hat{n}_N}(\phi_1,\ldots,\phi_N) = H_{\hat{n}_1,\ldots,\hat{n}_N}(\sqrt{\hat{p}^1_0}\phi_1,\ldots,\sqrt{\hat{p}^N_0}\phi_N) \times\exp\{-(\hat{p}^1_0\phi_1^2+\cdots+\hat{p}^N_0\phi_N^2)\}, \tag{6.21}$$

in $L^2(\mathbb{R}^N)\otimes\hat{A}$, where $H_{n_1,\ldots,n_N}$ denote the Hermite polynomials in $L^2(\mathbb{R}^N)$.[2] We obtain this basis directly if we solve the Schrödinger equation (5.8) of the q system instead of the operator equations (6.7).

(2) On the basis of Sec. IV the complete orthonormal system (6.12) [or (6.21)] provides a complete orthogonal family $\{\Psi_{\hat{n}}\} = \{\Phi_{\hat{n}}\sqrt{\rho_{\hat{n}}}\}$ of global states in the global state space $H^1 = \mathrm{Tr}\,\mathscr{H}_A$, where $\hat{n} = (\hat{n}_1,\ldots,\hat{n}_N)$. Let $\rho_{\hat{n}}$ be of the form $\Sigma_m c^{\hat{n}}_m p_m$ where $p_m = |\phi_m\rangle\langle\phi_m|$ and $\{\phi_m\}$ is a complete orthonormal family in $L^2(\mathbb{R}^3)$ such that the support of ϕ_m is a cube V_m in $\mathbb{R}^3$ (and these cubes have the properties that $V_m\cap V_{m'} = 0, m\neq m'$, and $\cup_m V_m = \mathbb{R}^3$). Let the edge of the cube V_m be ϵ_m and $\bar{\mathbf{x}}_m$ denote the center of the cube in $\mathbb{R}^3$. Then we can choose ϕ_m in such a way that $\langle\phi_m|\hat{\mathbf{x}}|\phi_m\rangle = \bar{\mathbf{x}}_m$ and the dispersion[25] $\Delta_m\mathbf{x} = \delta(\phi_m,\hat{\mathbf{x}})$ take the value ϵ_m. Furthermore, if $\langle\phi_m|\hat{\mathbf{p}}|\phi_m\rangle = \bar{\mathbf{k}}_m$ then the dispersions $\Delta_m\mathbf{x}$ and $\Delta_m\mathbf{p}$ do satisfy the relation $\Delta_m\mathbf{p}\Delta_m\mathbf{x}\approx\hbar$, i.e., $\Delta_m\mathbf{p}\approx\hbar/\epsilon_m$. Then, by (6.11) and (6.13), we can write

$$\mathrm{Tr}\,p_m\langle\Phi_{\hat{n}}|\hat{\mathscr{P}}_\mu|\Phi_{\hat{n}}\rangle_{\hat{A}}p_m = \mathrm{Tr}\,p_m\Big[\sum_{\alpha=1}^N\Big(n_\alpha(\hat{\mathbf{p}})+\frac{1}{2}\Big)\hat{p}^\alpha_\mu\Big] = \sum_{\alpha=1}^N\langle\phi_m|n_\alpha(\hat{\mathbf{p}})\hat{p}^\alpha_\mu|\phi_m\rangle + \frac{1}{2}\sum_{\alpha=1}^N\langle\phi_m|\hat{p}^\alpha_\mu|\phi_m\rangle \approx \sum_{\alpha=1}^N n_\alpha(\bar{\mathbf{k}}_m)\bar{\mathbf{k}}^\alpha_{\mu,m} + \frac{1}{2}\sum_{\alpha=1}^N\bar{k}^\alpha_{\mu,m}, \tag{6.22}$$

where $\hat{\mathscr{P}}_0 = \hat{\mathscr{H}}_0$ and $\bar{k}^\alpha_m = (\bar{k}^\alpha_{0,m},\bar{\mathbf{k}}_m)$, $\bar{k}^\alpha_{0,m} = \sqrt{\bar{\mathbf{k}}^2_m + m^2_\alpha}$. Consequently, in the global state $\Phi_{\hat{n}}\sqrt{\rho_{\hat{n}}}$ in the cube V_m there are $(n_1(\bar{\mathbf{k}}_m),\ldots,n_N(\bar{\mathbf{k}}_m))$ pieces of quanta ("KG particles") of four-momentum $(\bar{k}^1_{\mu,m},\ldots,\bar{k}^N_{\mu,m})$, respectively, with statistical weight $c^{\hat{n}}_m$ (plus the zero-point contribution $\frac{1}{2}\Sigma^N_{\alpha=1}\bar{k}^\alpha_{\mu,m}$). In such a global state the expectation value of the global four-momentum $\hat{P}_\mu = \mathrm{Tr}\,\hat{\mathscr{P}}_\mu$ of the system is

$$\bar{P}_\mu = \mathrm{Tr}\langle\Psi_{\hat{n}}|\hat{\mathscr{P}}_\mu|\Psi_{\hat{n}}\rangle_{\hat{A}} = \mathrm{Tr}\,\rho_{\hat{n}}\Big[\sum_{\alpha=1}^N\Big(n_\alpha(\hat{\mathbf{p}})+\frac{1}{2}\Big)\hat{p}^\alpha_\mu\Big] \approx\sum_m c^{\hat{n}}_m\sum_{\alpha=1}^N n_\alpha(\bar{\mathbf{k}}_m)\bar{k}^\alpha_{\mu,m} + \sum_m c^{\hat{n}}_m\sum_{\alpha=1}^N\frac{1}{2}\bar{k}^\alpha_{\mu,m}. \tag{6.23}$$

[In this brief consideration the domain question in Eqs. (6.22) and (6.23) is slightly put away.] Since $\hat{\mathscr{P}}_\mu$ is of discrete spectrum, thus by (3.13) we can write

$$\hat{\mathscr{P}}_\mu = \sum_{\alpha=1}^N\sum_{\hat{n}_\alpha=0}^\infty\hat{E}_{\hat{n}}\Big[n_\alpha(\hat{\mathbf{p}})+\frac{1}{2}\Big]\hat{p}^\alpha_\mu = \int_{\mathbb{R}^3}dP(\mathbf{k})\sum_{\alpha=1}^N\sum_{n_\alpha(\mathbf{k})=0}^\infty E_{n(\mathbf{k})}(\mathbf{k})\Big[n_\alpha(\mathbf{k})+\frac{1}{2}\Big]k^\alpha_\mu, \tag{6.24}$$

where $\hat{E}_{\hat{n}} = |\hat{\Phi}_{\hat{n}}\rangle_{\hat{A}}\langle\hat{\Phi}_{\hat{n}}|$. For the global observable generated by $\hat{\mathscr{P}}_\mu$ we get, by (4.8),

$$\hat{P}_\mu = \hat{P}^1_\mu = \mathrm{Tr}\int_{\mathbb{R}^3}dP(\mathbf{k})\sum_{\alpha=1}^N\sum_{n_\alpha(\mathbf{k})=0}^\infty E_{n(\mathbf{k})}(\mathbf{k}) \times\Big[n_\alpha(\mathbf{k})+\frac{1}{2}\Big]k^\alpha_\mu. \tag{6.25}$$

This, by (4.9), gets the form

$$\hat{P}^1_\mu|_H = \int^\oplus_{\mathbb{R}^3}d^3\mathbf{k}\sum_{\alpha=1}^N\Big[\hat{N}_\alpha(\mathbf{k})+\frac{1}{2}\Big]k^\alpha_\mu = \int^\oplus_{\mathbb{R}^3}d^3\mathbf{k}\sum_{\alpha=1}^N\sum_{n_\alpha(\mathbf{k})=0}^\infty E_{n(\mathbf{k})}(\mathbf{k})\Big[n_\alpha(\mathbf{k})+\frac{1}{2}\Big]k^\alpha_\mu \tag{6.26}$$

in the subspace $H = \int^\oplus_{\mathbb{R}^3}\mathscr{H}(\mathbf{k})d^3\mathbf{k}$. This expression shows a complete formal coincidence with its counterpart in CQFT.[1] However, in the present approach it is not required to introduce the normal ordering on account of the presence of the statistical weights $c_m^{\hat{n}}$ in (6.23).

B. Interacting fields, *S* matrix

In the interaction picture the local states are governed by the evolution operator $U(t,t_0)$ given by the formal expansion (6.6) and where $\hat{\phi}_\alpha(t)$ is inserted from (6.8) using (6.1). For describing a local scattering problem we obtain the local S matrix from $U(t,t_0)$, along the usual line of thoughts, by the adiabatic hypothesis,

$$S = \mathbb{1} + \sum_{n=1}^{\infty} (-i)^n \frac{1}{n!} \times \int_{-\infty}^{+\infty} dt^1 \cdots dt^n \, T(\hat{\mathscr{H}}_I(t^1) \cdots \hat{\mathscr{H}}_I(t^n)). \quad (6.27)$$

Clearly, S maps $\mathscr{H}_{\hat{B}}$ in $\mathscr{H}_{\hat{A}}$ into itself (cf. Notes 1 and 2 below). The relevant matrix element of S referring to a transition between specified initial and final local states $\mathbf{\Phi}_i$ and $\mathbf{\Phi}_f$ in $\mathscr{H}_A$ are given as

$$S_{if} = \langle \Phi_f | S | \Phi_i \rangle_A \in A, \quad (6.28)$$

when $\hat{\Phi}_i$ and $\hat{\Phi}_f$ are eigenvectors of $\hat{\mathscr{H}}_0$ then $\hat{\Phi}_i, \hat{\Phi}_f \in \mathscr{H}_{\hat{B}}$ and $S_{if} \in \hat{B}$. The transition amplitude between the global states $\Psi_i = \Phi_i \sqrt{\rho_i}$ and $\Psi_f = \Phi_f \sqrt{\rho_f}$ in $H^1 = \mathrm{Tr}\, \mathscr{H}_A$, belonging to the local ones, Φ_i and Φ_f, is provided by the expression

$$S_{if} = \langle \Psi_f | S_1 | \Psi_i \rangle_1 = \mathrm{Tr}\, \sqrt{\rho_i \rho_f} \langle \Phi_f | S | \Phi_i \rangle_A. \quad (6.29)$$

Assume that the interaction local Hamiltonian $\hat{\mathscr{H}}_I = V(\hat{\phi})$ is a polynomial of the fields $\hat{\phi}_\alpha$, i.e.,

$$V(\hat{\phi}_1,...,\hat{\phi}_N) = \sum_{n=0}^{M} g_n \hat{\phi}_1^{n_1} \cdots \hat{\phi}_N^{n_N}, \quad n = (n_1,...,n_N), \quad g_n \in \mathbb{R}. \quad (6.30)$$

[Of course, the "smallness" condition for $\hat{\mathscr{H}}_I$ with respect to $\hat{\mathscr{H}}_0$—which is necessary that $U(t,t_0)$ in (6.6) be applicable, i.e., the interaction picture be applicable—imposes restrictions for $V(\hat{\phi})$ in (6.30).[38] Intuitively this means that the coupling constants g_n should be small dimensionless numbers. For examples, in the case of a single scalar field ($N = 1$) this implies that $M \leqslant 4$.] Then the terms of the local S matrix in (6.27) can be expanded as a sum of normal products, using the pairings defined in (6.19), by Wick's first theorem and the corresponding terms can be represented by Feynman's graph techniques[1] (see Appendix C).

Proposition 6.1: Let $V(\hat{\phi})$ be as given in (6.30), then the terms of the series (6.27) are well defined, up to the ambiguity following from the T product, i.e., they do not contain infinite factors (ultraviolet divergences) when the matrix elements of them are taken between any two local states from the domain of them in $\mathscr{H}_{\hat{A}}$. Consequently, the matrix elements computed in between the corresponding global states also do not contain such divergences.

Proof: In fact, when we expand $S^{(n)}$ in (6.27) by the first Wick theorem then by the expressions (6.8), (6.9), (6.20), (2.13), (6.15), (2.14), and (6.19) we take products and sums of well-defined elements of the operator *-algebras $\hat{B}(\mathscr{H}) \otimes \hat{B}$ and $\hat{B}$—up to the ambiguity given in the T product by the "function" $\Theta(t - t')$—which manipulations are legitimate in between the elements of the algebra $\hat{B}(\mathscr{H}) \otimes \hat{B}$ and $\hat{B}$ since $\hat{B}(\mathscr{H}) \otimes \hat{B}$ is a module over $\hat{B}$. The scalar product $\langle \Phi_1 | S^{(n)} | \Phi_2 \rangle_A$, $\Phi_1, \Phi_2 \in \mathscr{D}(S^{(n)}) \subseteq \mathscr{H}_A$ [$\mathscr{D}(S^{(n)})$ is the domain of $S^{(n)}$ in $\mathscr{H}_A$], yields also a well-defined element of A, up to the mentioned ambiguity. The same can be said about the complex amplitude

$$\langle \Psi_1 | S_1^{(n)} | \Psi_2 \rangle_1 = \mathrm{Tr} \sqrt{\rho_2 \rho_1} \langle \Phi_1 | S^{(n)} | \Phi_2 \rangle_A,$$

too, where

$$\Psi_1, \Psi_2 \in \mathscr{D}(S_1^{(n)}) \subseteq H^1 = \mathrm{Tr}\, \mathscr{H}_A.$$

Q.E.D.

In CQFT the total interaction Hamiltonian

$$\hat{H}_I(t) = \int_{\mathbb{R}^3} \hat{\mathscr{H}}_I(\mathbf{x},t) d^3\mathbf{x} = \int_{\mathbb{R}^3} V(\hat{\phi}(\mathbf{x},t)) d^3\mathbf{x}$$

for a nontrivial interaction does not exist as a well-defined operator in the Fock space of the free fields [Problem 2(b) in Sec. I] and this leads to ultraviolet divergences.[5] Now in QLFT the *local interaction Hamiltonian* $\hat{\mathscr{H}}_I(t)$ is well-defined in the local state space $\mathscr{H}_{\hat{A}}$ of the free fields $\hat{\phi}_\alpha(t)$ [and the term $-\partial\hat{V}/\partial\hat{\phi}$ in Eq. (5.12) describing the interaction now is *a priori defined* in contrast with CQFT[5,6]]: by (6.8) and (6.9) $\hat{\mathscr{H}}_I(t)$ given by (6.30) is a well-defined element of $\hat{B}(\mathscr{H}) \otimes \hat{B}$. Consequently, the global interaction Hamiltonian $\hat{H}_I^1(t) = \mathrm{Tr}\, \hat{\mathscr{H}}_I(t)$ generated by $\hat{\mathscr{H}}_I(t)$ is also a well-defined operator in the global state space $H^1 = \mathrm{Tr}\, \mathscr{H}_A$. Thus the ultraviolet catastrophe of CQFT does not occur in QLFT. (Physically we can say that the local Hamiltonian does not contain contributions of far away fluctuations.[19]) The Haag theorem is also avoided in this approach because the system of operators $(\hat{\pi},\hat{\phi})$ constitutes only a $\mathscr{B}$-irreducible system in $\mathscr{H}_{\hat{A}}$ (cf. Note 5 in Sec. III). In summary, Problem 2 of CQFT (see Sec. I) is avoided in QLFT and so nontrivial interactions can be described by the aid of the interaction picture in this model of quantized fields.

Notes: (1) The rigorous proof of the unitarity of the S matrix (6.27) formally derived from $U(t,t_0)$ in (6.6) would require a proof of the asymptotic completeness. However, such a proof lacks even in the q mechanical N-body scattering theory for the general case when $N \geqslant 4$. New results along this line of research have been obtained by Sigal and Soffer[39] who prove the asymptotic completeness for short-range q mechanical systems consisting of arbitrary number of particles. Short range means that the pair potentials vanish at ∞ faster than r^{-1}.

(2) As to the convergence of the S matrix in (6.27) we mention that it is convergent as much as its q mechanical counterpart [belonging to the Schrödinger equation, the eigenvalue equation of which is given by Eq. (5.13')] is. For the former is the extension of the latter formulated in the universe V, via the universe $V^{(\mathscr{B})}$ to the universe $V^{(\mathscr{P})}$. Our general example is the extension of a q mechanical anharmonic oscillator system of state space $\mathscr{H}$ to an infinite collection of such systems, the members of which are connected in space. Consequently the S matrix (6.27) is convergent in $\mathscr{H}_A = \mathscr{H} \otimes A$ if the corresponding q mechanical S matrix is convergent in $\mathscr{H}$.

VII. THE LORENTZ INVARIANCE AND THE CLASSICAL LIMIT

Let us see first the local quantities $T_{\mu\nu}$ and $\mathscr{M}_{\mu\nu\rho}$ belonging to the generators $(P_\mu, M_{\nu\rho})$ of the proper Poincaré group $P^{\uparrow}_{+}$ in CLFT, where

$$T_{\mu\nu} = \sum_{\alpha=1}^{N} \frac{\partial \mathscr{L}}{\partial\, \partial_\mu \phi_\alpha} \partial_\nu \phi_\alpha - \mathscr{L} g_{\mu\nu}$$

and

$$\mathscr{M}_{\mu\nu\rho} = T_{\mu\rho} x_\nu - T_{\mu\nu} X_\rho .$$

Since $T_{\mu\nu} = T_{\mu\nu}(\phi,\pi,\nabla\phi)$ and $\mathscr{M}_{\mu\nu\rho} = \mathscr{M}_{\mu\nu\rho}(\phi,\pi,\nabla\phi,x)$, i.e., they are functions of the local observables $(\phi,\pi,\nabla\phi)$ and the space-time points $x = (t,\mathbf{x})$, thus, applying the canonical quantization algorithm, we can associate operators in $\mathscr{H}_{\hat{A}}$ with them by the substitution $\hat{T}_{\mu\nu} = T_{\mu\nu}(\hat{\phi},\hat{\pi},\hat{\nabla}\phi)$ and $\hat{\mathscr{M}}_{\mu\nu\rho} = \mathscr{M}_{\mu\nu\rho}(\hat{\phi},\hat{\pi},\hat{\nabla}\phi,t,\hat{\mathbf{x}})$. However, by formal computation with the aid of the CCR's of the form (2.3) we can check that $[\hat{T}_{0\mu},\hat{\phi}_\alpha] = -i\partial_\mu \hat{\phi}_\alpha$ but $[\hat{T}_{0\mu},\hat{\pi}_\alpha] \neq -i\partial_\mu \hat{\pi}_\alpha$ for $\mu = k$ $(k = 1,2,3)$, because $[\hat{T}_{0k},\hat{\pi}_\alpha] = i\partial_k \hat{\pi}_\alpha$. Furthermore $[\hat{T}_{0k},\hat{x}_j] \neq -i\delta_{jk}\mathbb{1}$ and $[\hat{T}_{0k},\hat{T}_{00}] \neq 0$. Thus $\hat{T}_{0k}$ is not a conserved local observable and does not generate translations in $\mathbb{R}^3$ in the k direction and consequently the operators $(\hat{T}_{0\mu},\hat{\mathscr{M}}_{0\nu\rho})$ do not satisfy the commutation rules of the generators of $P^{\uparrow}_{+}$ and so do not generate a representation of $P^{\uparrow}_{+}$ in $\mathscr{H}_A$. We have to look for another way.

First we discuss the question of Poincaré covariance and invariance in the case of free q fields. In this case we identified the local three-momentum observable of the system with the self-adjoint operator $\hat{\mathscr{P}}$ in (6.13). Then by means of this operator we can generate the three-space translations of the system and the four space-time translations (or evolution) of the system can be described by the four-parameter unitary group $U(a) = \exp(-i\hat{\mathscr{P}}_\mu a^\mu)$ in $\mathscr{H}_A$ where $\hat{\mathscr{P}} = (\hat{\mathscr{H}}_0,\hat{\mathscr{P}})$ and $[\hat{\mathscr{P}}_\mu,\hat{\mathscr{P}}_\nu] = 0$ and $a = (x - x_0) = (t - t_0, \mathbf{x} - \mathbf{x}_0)$. Then the local states and local observables evolve in four space-time as follows:

$$\Phi_x = U(x)\Phi_0 = \exp(-i\hat{\mathscr{P}}_\mu x^\mu)\Phi_0,$$

$$\hat{F}(x) = \exp(-i\hat{\mathscr{P}}_\mu x^\mu)\,\hat{F}_0 \exp(i\hat{\mathscr{P}}_\mu x^\mu)$$

(with the choice $x_0 = 0$) or in differential form

$$i\frac{\partial \Phi_x}{\partial x^\mu} = \hat{\mathscr{P}}_\mu \Phi_x, \tag{7.1}$$

$$\frac{\partial \hat{F}(x)}{\partial x^\mu} = i[\hat{\mathscr{P}}_\mu,\hat{F}(x)] . \tag{7.2}$$

Then for the field $\hat{\phi}_\alpha$ in the decomposition (6.8) we have, with the aid of (6.10), (6.11), and (6.13),

$$\begin{aligned}\hat{\phi}_\alpha(x) &= e^{-i\hat{\mathscr{P}}_\mu x^\mu}\hat{\phi}_\alpha(\hat{\mathbf{p}})e^{i\hat{\mathscr{P}}_\mu x^\mu} \\ &= (2\hat{p}_0^\alpha)^{-1/2}\big[\hat{a}_\alpha^+(\hat{\mathbf{p}})e^{i\hat{p}_\mu^\alpha x^\mu} \\ &\quad + \hat{a}_\alpha(\hat{\mathbf{p}})e^{-i\hat{p}_\mu^\alpha x^\mu}\big] \\ &= \int_{\mathbb{R}^3} dP(\mathbf{k})\,\frac{1}{\sqrt{2k_0^\alpha}} \\ &\quad \times\big[\hat{a}_\alpha^+(\mathbf{k})e^{ik_\mu^\alpha x^\mu} + \hat{a}_\alpha(\mathbf{k})e^{-ik_\mu^\alpha x^\mu}\big] ,\end{aligned} \tag{7.3}$$

where $\hat{p}^\alpha = (\hat{p}_0^\alpha, \hat{\mathbf{p}})$ and $k^\alpha = (k_0^\alpha,\mathbf{k})$. If we take into account the integral kernel representation of $P(\mathbf{k})$ then we see that this expression is very similar to the form of the free field operator in CQFT.[1] However, the latter operator is not well-defined in the Fock space of CQFT, according to the Problem 3 of CQFT (see Sec. I). Now $\hat{\phi}_\alpha(x)$ in (7.3) is a well-defined operator in $\mathscr{H}_{\hat{A}}$ for any value of the parameter x. Among the hypotheses of the negative results leading to Problem 3 (see Theorems 10.6 and 10.7 on p. 283 in Ref. 2) that one does not satisfy, in the present approach, which requires the irreducibility of the system $(\hat{\pi},\hat{\phi})$ of operators. Now this system $(\hat{\pi},\hat{\phi})$ is only a $\mathscr{B}$-irreducible system in $\mathscr{H}_{\hat{A}}$ (cf. Note 5 in Sec. III).

Equation (7.2) for $\hat{\phi}_\alpha(x)$ and $\mu = 0$ gives $\dot{\hat{\phi}}_\alpha(x) = i[\hat{\mathscr{H}}_0,\hat{\phi}_\alpha(x)] = \hat{\pi}_\alpha(x)$, while for $\mu = 1,2,3$ it yields

$$\begin{aligned}\nabla_{\mathbf{x}}\hat{\phi}_\alpha(t,\mathbf{x}) &= i[\hat{\mathscr{P}},\hat{\phi}_\alpha(t,\mathbf{x})] \\ &= i\hat{\mathbf{p}}(2\hat{p}_0^\alpha)^{-1/2}\big[\hat{a}_\alpha^+(\hat{\mathbf{p}})e^{i\hat{p}_\mu^\alpha x^\mu} - \hat{a}_\alpha(\hat{\mathbf{p}})e^{-i\hat{p}_\mu^\alpha x^\mu}\big] .\end{aligned} \tag{7.4}$$

For the derivative $\nabla_{\mathbf{x}}(\nabla_{\mathbf{x}}\hat{\phi}_\alpha(x)) = \Delta_{\mathbf{x}}\hat{\phi}_\alpha(x)$ we obtain

$$\begin{aligned}\Delta_{\mathbf{x}}\phi_\alpha(x) &= i[\hat{\mathscr{P}},i[\hat{\mathscr{P}},\hat{\phi}_\alpha(x)]] = -\hat{\mathbf{p}}^2(2\hat{p}_0^\alpha)^{-1/2} \\ &\quad\times\big[\hat{a}_\alpha^+(\hat{\mathbf{p}})e^{i\hat{p}_\mu^\alpha x^\mu} + \hat{a}_\alpha(\hat{\mathbf{p}})e^{-i\hat{p}_\mu^\alpha x^\mu}\big] \\ &= -\hat{\mathbf{p}}^2\hat{\phi}_\alpha(x).\end{aligned} \tag{7.5}$$

Comparing this equation with Eq. (6.7) we see that $\hat{\phi}_\alpha(x)$ satisfies the relativistically invariant KG equation

$$(\Box_x + m_\alpha^2)\hat{\phi}_\alpha(x) = 0, \quad \alpha = 1,\ldots,N, \tag{7.6}$$

where $\Box_x = \partial_t^2 - \Delta_{\mathbf{x}}$ acts on the parameters $x = (t,\mathbf{x})\in\mathbb{R}^4$. Because $\hat{\phi}_\alpha(x)$ is a well-defined operator at every parameter point $x\in\mathbb{R}^4$, thus (7.6) is a well-defined operator equation in contrast with its conventional counterpart,[1] which is only of formal content because of Problem 3.[2]

Now let us see the Lorentz rotations. By the aid of the operators $\hat{\mathscr{P}}_\mu$ the generators of these rotations can be easily identified as follows:

$$\hat{\mathscr{M}}_{\mu\nu} = x_\nu\hat{\mathscr{P}}_\mu - x_\mu\hat{\mathscr{P}}_\nu . \tag{7.7}$$

For on the set of the local states satisfying Eq. (7.1) we formally have $[\hat{\mathscr{P}}_\mu,x_\nu]\Phi_x = ig_{\mu\nu}\Phi_x$. From this we formally get that the family of operators $(\hat{\mathscr{P}}_\mu,\hat{\mathscr{M}}_{\rho\nu})$ satisfies the commutation relations of the Lie algebra of $P^{\uparrow}_{+}$. Thus the ten-parameter unitary group

$$U(a,\Lambda) = \exp\big[i(a_\mu\hat{\mathscr{P}}^\mu + \tfrac{1}{2}\Lambda_{\mu\nu}\hat{\mathscr{M}}^{\mu\nu})\big]$$

in $\mathscr{H}_A$ realizes formally a unitary representation of $P^{\uparrow}_{+}$ in $\mathscr{H}_A$. The action of $U(a,\Lambda)$ for the local states is $U(a,\Lambda)\Phi_x(\phi,\hat{\mathbf{p}},\hat{\mathbf{x}}) = \Phi_{\Lambda^{-1}(x-a)}(\phi',\hat{\mathbf{p}},\hat{\mathbf{x}})$ where ϕ' denotes the action on Φ of the operators (in $\hat{\mathscr{P}}_\mu$ and $\hat{\mathscr{M}}_{\nu\rho}$) acting on ϕ. The action of $U(a,\Lambda)$ for the local observables $\hat{F}$ is $\hat{F}' = U(a,\Lambda)\hat{F}U^{-1}(a,\Lambda)$. The differential form of this relation is partly in Eq. (7.2) and is the following:

$$(x_\nu\partial_\mu - x_\mu\partial_\nu)\hat{F}(x) = i[\hat{\mathscr{M}}_{\mu\nu},\hat{F}(x)] \tag{7.8}$$

(cf. Note 4 below).

Let us consider the Poincaré covariance and invariance in the case of the interacting fields. Let us integrate the operator equations (5.12) with the boundary condition that fields $\hat{\phi}_\alpha^{\rm in}(t)$ and $\hat{\phi}_\alpha^{\rm out}(t)$ exist in such a way that they satisfy the KG equations (6.7) (i.e., free q fields) and

$$\lim_{t\to\pm\infty} (\hat{\phi}_\alpha(t) - \hat{\phi}_\alpha^{\rm out/in}(t)) = 0, \quad \alpha = 1,...,N. \tag{7.9}$$

Then we obtain the following Yang–Feldman equations:

$$\hat{\phi}_\alpha(t) = \hat{\phi}_\alpha^{\rm in}(t) + \int_{-\infty}^{+\infty} \hat{\Delta}_R^\alpha(t-t')\frac{\partial\hat{V}}{\partial\hat{\phi}_\alpha}(t')dt', \tag{7.10a}$$

$$\hat{\phi}_\alpha(t) = \hat{\phi}_\alpha^{\rm out}(t) + \int_{-\infty}^{+\infty} \hat{\Delta}_A^\alpha(t-t')\frac{\partial\hat{V}}{\partial\hat{\phi}_\alpha}(t')dt', \tag{7.10b}$$

$\alpha = 1,...,N,$

where $\hat{\Delta}_R^\alpha(t) = -\theta(t)\hat{\Delta}^\alpha(t)$ and $\hat{\Delta}_A^\alpha(t) = \theta(t)\hat{\Delta}^\alpha(t)$ and $\hat{\Delta}^\alpha(t)$ is given by (2.13) with mass m_α.[40] By inserting Eqs. (7.10) into Eqs. (5.12) we can see that the asymptotic fields $\hat{\phi}_\alpha^{\rm out/in}(t)$ indeed satisfy the KG equations (6.7); furthermore (7.10) fulfills formally the asymptotic condition (7.9), too. Since $\hat{\phi}_\alpha^{\rm in}(t)$ and $\hat{\phi}_\alpha^{\rm out}(t)$ are free q fields, they are connected by a unitary operator S in the local state space $\mathscr{H}_{\hat{A}}$ of the free q fields in the way $\hat{\phi}_\alpha^{\rm out}(t) = S^{-1}\hat{\phi}_\alpha^{\rm in}(t)S$, and this S is formally identical with the local S matrix in (6.27) when one can apply perturbation theory using the interaction picture (cf. Ref. 40).

The local Hamiltonian $\hat{\mathscr{H}}$ of the interacting q fields in (7.10) (from its time independence) equals the local Hamiltonian $\hat{\mathscr{H}}^{\rm in/out}$ of the free q fields $\hat{\phi}_\alpha^{\rm in/out}(t)$ (cf. Ref. 41), i.e., formally we have

$$\lim_{t\to-\infty}\left\{\frac{1}{2}\sum_{\alpha=1}^{N}[\hat{\phi}_\alpha(t) + (\hat{p}_0^\alpha)^2\hat{\phi}_\alpha^2(t)] + V(\hat{\phi}(t))\right\}$$
$$= \frac{1}{2}\sum_{\alpha=1}^{N}[(\hat{\phi}_\alpha^{\rm in}(t))^2 + (\hat{p}_0^\alpha)^2(\hat{\phi}_\alpha^{\rm in}(t))^2]. \tag{7.11}$$

Because for the free q fields we have found the generators of $P_+^\uparrow$, thus for the interacting q fields in (7.10) satisfying (7.9) the generators of $P_+^\uparrow$ are provided by the operator family $(\hat{\mathscr{P}}_\mu^{\rm in}, \hat{\mathscr{M}}_{\rho\nu}^{\rm in})$ or $(\hat{\mathscr{P}}_\mu^{\rm out}, \hat{\mathscr{M}}_{\nu\rho}^{\rm out})$, where $\hat{\mathscr{P}}^{\rm in/out} = (\hat{\mathscr{H}}_0^{\rm in/out}, \hat{\mathscr{P}}^{\rm in/out})$ [the form of $\hat{\mathscr{P}}^{\rm in/out}$ is given by (6.13) (cf. Note 2 below)] and $\hat{\mathscr{M}}_{\mu\nu}^{\rm in/out} = x_\nu\hat{\mathscr{P}}_\mu^{\rm in/out} - x_\mu\hat{\mathscr{P}}_\nu^{\rm in/out}$ according to (7.7). Consequently the ten-parameter unitary group $U(a,\Lambda) = \exp[i(a^\mu\hat{\mathscr{P}}_\mu^{\rm in/out} + \frac{1}{2}\Lambda^{\mu\nu}\hat{\mathscr{M}}_{\mu\nu}^{\rm in/out})]$ in $\mathscr{H}_A$ gives formally a unitary representation of $P_+^\uparrow$ in $\mathscr{H}_A$.

Since, in the case of the solutions of Eqs. (5.12) satisfying (7.9), the translations of the q system in three-space are generated by a local three-momentum observable of the form (6.13) (cf. Notes 2 and 3 below), thus the operator

$$\hat{\phi}_\alpha(t,\mathbf{x}) = \exp(i\hat{\mathscr{P}}\mathbf{x})\hat{\phi}_\alpha(t)\exp(-i\hat{\mathscr{P}}\mathbf{x})$$

satisfies, by Eqs. (7.5) and (5.12), the operator equation

$$\Box_x\hat{\phi}_\alpha(x) + m_\alpha^2\hat{\phi}_\alpha(x) = -\frac{\partial\hat{V}}{\partial\hat{\phi}_\alpha}(x), \quad \alpha = 1,...,N. \tag{7.12}$$

These equations, relativistically invariant in form, are well-defined at every parameter point $x\in\mathbb{R}^4$ as a consequence of the well-definedness of $\hat{\phi}_\alpha(x)$, in contrast with their conventional counterparts in CQFT. Consequently the field equations of the c theory recover in the quantized theory as well-defined operator equations of the *same form* if the asymptotic condition (7.9) is fulfilled. In this model of quantized fields Problem 3 of CQFT is avoided.

In this way for the free q fields and the interacting q fields satisfying the asymptotic condition (7.9) (for the most important practical cases) we have cleared up the question of Lorentz invariance (see still Note 3 below).

Now let us see the c limit of the model. By taking the expectation value of Eqs. (7.12) in any local or global state (from their domains) we get the Ehrenfest theorem (cf. Ref. 19)

$$\Box_x\bar{\phi}_\alpha(x) + m_\alpha^2\bar{\phi}_\alpha(x) = -\overline{\frac{\partial V}{\partial\phi_\alpha}}(x), \quad \alpha = 1,...,N, \tag{7.13}$$

where, e.g.,

$$\bar{\phi}_\alpha(x) = \langle\Phi|\hat{\phi}_\alpha(x)|\Phi\rangle_{\hat{A}}, \quad \Phi\in\mathscr{H}_{\hat{A}},$$

or

$$\bar{\phi}_\alpha(x) = {\rm Tr}\,\rho\langle\Phi|\hat{\phi}_\alpha(x)|\Phi\rangle_A, \quad \Phi\in H^\rho,$$

and similarly for $\overline{(\partial V/\partial\phi_\alpha)}(x)$, too. The field equations of CLFT hold true in expectation value in the quantized theory. If we remind that $\hat{\phi}_\alpha(x)$ and $\hat{\pi}_\alpha(x)$ are functions of the operator $\hat{\mathbf{p}}$ then it easily follows from Heisenberg's uncertainty principle that the c limit ($\hbar\to 0$) of QLFT is provided by such (global) states in which the dispersions $\Delta\hat{\pi}_\alpha$ and $\Delta\hat{\phi}_\alpha$ referring to the canonical pair $(\hat{\pi},\hat{\phi})$ *and* the dispersions $\Delta\mathbf{p}$ and $\Delta\mathbf{x}$ referring to the canonical pair $(\hat{\mathbf{p}},\hat{\mathbf{x}})$ are simultaneously small. In these states the basic measuring hypothesis referring to both the c fields $\phi_\alpha(x)$ and the points (events) of c Minkowski space-time are essentially satisfied. The points $x = (t,\mathbf{x})\in\mathbb{R}^4$ in the arguments of the q fields $\hat{\phi}_\alpha(x)$ are only *parameters* in QLFT, they *do not have* the *operational meaning* of the *points* (events) of the c Minkowski space-time (cf. Ref. 10). In QLFT we *cannot retain* the *physical* (operational) *meaning* of *classical space-time*, we need the quantum space-time for the consistent interpretation of QLFT [the local state space $\mathscr{H}_A$ of QLFT carrying all the physical information obtained of the q system by local measurements is constructed over q space-time (cf. Secs. III and V)][16] (cf. still Refs. 9 and 10).

Notes: (1) By Eqs. (7.2)–(7.4) and the substitution $t = 0$ we get the operator

$$\hat{\mathbf{P}}(\mathbf{x}) := \frac{1}{2}\sum_{\alpha=1}^{N}[\hat{\pi}_\alpha(\mathbf{x})\nabla_\mathbf{x}\hat{\phi}_\alpha(\mathbf{x}) + (\nabla_\mathbf{x}\hat{\phi}_\alpha(\mathbf{x}))\hat{\pi}_\alpha(\mathbf{x})]$$
$$= \hat{\mathscr{P}} - \frac{1}{2}\sum_{\alpha=1}^{N}[\hat{a}_\alpha^{+2}(\hat{\mathbf{p}})e^{-2i\hat{\mathbf{p}}\mathbf{x}} + \hat{a}_\alpha^2(\hat{\mathbf{p}})e^{2i\hat{\mathbf{p}}\mathbf{x}}]. \tag{7.14}$$

The c counterpart of this operator is the three-momentum density $T_{0k} = \Sigma_{\alpha=1}^N\pi_\alpha\,\partial_k\phi_\alpha$, but neither T_{0k} nor $\hat{\mathbf{P}}(\mathbf{x})$ is a conserved local quantity in the sense of Eqs. (5.5) and (5.10) ($[\hat{\mathscr{H}}_0,\hat{\mathbf{P}}]\neq 0$) (cf. the following Notes 2 and 3).

(2) By the aid of the relations (6.20) the local three-momentum $\hat{\mathscr{P}}$ introduced in (6.13) can be expressed by $(\hat{\phi},\hat{\pi})$ in the following "nondiagonal" form:

$$\hat{\mathscr{P}} = \sum_{\alpha=1}^{N} [\hat{\phi}_\alpha^2 \hat{\mathbf{p}}\hat{p}_0^\alpha + \hat{\pi}_\alpha^2 \hat{\mathbf{p}}(\hat{p}_0^\alpha)^{-1}]$$
$$= \sum_{\alpha=1}^{N} [\hat{\phi}'_\alpha \widehat{\nabla} \hat{\phi}'_\alpha + \hat{\pi}'_\alpha \widehat{\nabla} \hat{\pi}'_\alpha],$$

where $\hat{\phi}'_\alpha = \hat{\phi}_\alpha \sqrt{\hat{p}_0^\alpha}$ and $\hat{\pi}'_\alpha = \hat{\pi}_\alpha(\sqrt{\hat{p}_0^\alpha})^{-1}$ and in the sense of Proposition 3.2 the transformation $(\hat{\phi},\hat{\pi}) \to (\hat{\phi}',\hat{\pi}')$ is a canonical transformation. The c local quantity $\mathscr{P} = \Sigma_{\alpha=1}^{N} [\phi_\alpha \nabla \phi_\alpha + \pi_\alpha \nabla \pi_\alpha]$ corresponding to the above $\hat{\mathscr{P}}$ is locally conserved. For

$$\{\mathscr{H},\mathscr{P}\} = \sum_{\alpha=1}^{N} \left[\frac{\delta \mathscr{H}}{\delta \pi_\alpha}\frac{\delta \mathscr{P}}{\delta \phi_\alpha} - \frac{\delta \mathscr{P}}{\delta \pi_\alpha}\frac{\delta \mathscr{H}}{\delta \phi_\alpha}\right] = 0,$$

since

$$\frac{\delta \mathscr{P}}{\delta \phi_\alpha} = (\nabla \phi_\alpha - \nabla \phi_\alpha) = 0$$

and

$$\frac{\delta \mathscr{P}}{\delta \pi_\alpha} = (\nabla \pi_\alpha - \nabla \pi_\alpha) = 0.$$

(3) In CLFT, when derivative coupling does not occur in the Lagrangian, the total three-momentum is identical with the three-momentum belonging to the free fields.[1] Essentially this fact survives after the quantization, too. For the global three-momentum belonging to the local one in (6.13), by (6.25) and (6.26), respectively, formally coincides with the three-momentum given in CQFT by the expression

$$\hat{\mathscr{P}}_k = \int_{\mathbb{R}^3} d^3\mathbf{x} \hat{T}_{0k}(\mathbf{x}) = \int_{\mathbb{R}^3} d^3\mathbf{x} \sum_{\alpha=1}^{N} (\hat{\pi}_0^\alpha(\mathbf{x}) \partial_k \hat{\phi}_\alpha(\mathbf{x})),$$

where $\pi_0^\alpha = \partial \mathscr{L}_0 / \partial \dot{\phi}_\alpha$ is the momentum density obtained from the free Lagrangian $\mathscr{L}_0$. "Integrating up" the operator in (7.14) in the parameter $\mathbf{x}$ [i.e., forming its trace appropriately in an orthonormal function system from $L^2(\mathbb{R}^3)$], the term differing from $\hat{\mathscr{P}}$ vanishes, consequently $\hat{\mathbf{P}}(\mathbf{x})$ and $\hat{\mathscr{P}}$ provide essentially the same global (total) three-momentum.

(4) In the present approach the three-position parameter $\mathbf{x}$ is treated on an equal footing with the time parameter t. This physically reflects essentially the fact that, in the absence of explicit four-position dependence in the Lagrangian (1.1), the system in four space-time is homogeneous and isotropic.

VIII. DISCUSSION OF THE AXIOMS OF WIGHTMAN AND HAAG–KASTLER

Now we discuss the axioms of Wightman and Haag–Kastler in the specific models of QLFT satisfying the asymptotic condition (7.9). We use the form of these axioms that can be found in Ref. 5, pp. 96–98. First we see Wightman's axioms.

W1: The Hilbert space of the q theoretical states is provided by the global state space $H^1 = \mathrm{Tr}\,\mathscr{H}_A$, which, in Fock representation, is spanned by the basis $\{\Psi_{\hat{n}}\} = \{\Phi_{\hat{n}}\sqrt{\rho_{\hat{n}}}\}$ corresponding to the Fock basis in (6.12). The continuous unitary group

$$U(a,\Lambda) = \exp[-i(a^\mu :\hat{\mathscr{P}}:_\mu^{\text{in}} + \tfrac{1}{2}\Lambda^{\mu\nu} :\hat{\mathscr{M}}:_{\mu\nu}^{\text{in}})]$$

in $\mathscr{H}_A$ generates a continuous unitary representation of $P_+^\uparrow$ in H^1, where : : denotes that the generators $\hat{\mathscr{P}}_\mu^{\text{in}}$ are renormalized in such a way that the ground state of $:\hat{\mathscr{P}}:_0^{\text{in}}$ be 0, i.e., $:\hat{\mathscr{P}}:_\mu^{\text{in}} = \Sigma_{\alpha=1}^{N} \hat{N}_\alpha \hat{p}_\mu^\alpha$. (One can achieve this, e.g., by the normal product in $\mathscr{H}_0$). Then the generators $(:\hat{\mathscr{P}}:_0^{\text{in}}, :\hat{\mathscr{P}}:^{\text{in}})$ of the translation subgroup have spectrum only in the forward cone. For

$$:\hat{\mathscr{P}}:_0^{\text{in}} \Phi_{\hat{n}} = \left(\sum_{\alpha=1}^{N} \hat{n}_\alpha \hat{p}_0^\alpha\right)\Phi_{\hat{n}}, \quad :\hat{\mathscr{P}}:^{\text{in}} \Phi_{\hat{n}} = \left(\sum_{\alpha=1}^{N} \hat{n}_\alpha \hat{\mathbf{p}}\right)\Phi_{\hat{n}}$$

and

$$(:\hat{\mathscr{P}}_0:^{\text{in}\,2} - :\hat{\mathscr{P}}:^{\text{in}\,2})\Phi_{\hat{n}} = (E_{\hat{n}}^2 - \mathbf{P}_{\hat{n}}^2)\Phi_{\hat{n}},$$

where

$$\begin{aligned} E_{\hat{n}}^2 - \mathbf{P}_{\hat{n}}^2 &= \left(\sum_{\alpha=1}^{N} \hat{p}_0^\alpha \hat{n}_\alpha\right)^2 - \left(\sum_{\alpha=1}^{N} \hat{n}_\alpha\right)^2 \hat{\mathbf{p}}^2 \\ &\geqslant \left(\sum_{\alpha=1}^{N} \hat{n}_\alpha\right)^2 (\hat{p}_{0,\min}^2 - \hat{\mathbf{p}}^2) \\ &= \int_{\mathbb{R}^3} dP(\mathbf{k}) \left(\sum_{\alpha=1}^{N} n_\alpha(\mathbf{k})\right)^2 \\ &\quad \times (k_{0,\min}^2 - \mathbf{k}^2) \geqslant 0, \quad \forall \mathbf{k}, \end{aligned}$$

where $\hat{p}_{0,\min} = \min_{m_\alpha} \hat{p}_0^\alpha$. The vacuum vector $\Psi_0 = \Phi_0 \sqrt{\rho_0}$ is invariant under the action of $U(a,\Lambda)$, because

$$\exp[-i(a^\mu :\hat{\mathscr{P}}:_\mu^{\text{in}} + \tfrac{1}{2}\Lambda^{\mu\nu} :\hat{\mathscr{M}}:_{\mu\nu}^{\text{in}})]\Phi_0 = \Phi_0.$$

W2: Since we deal with canonical formalism, we consider an appropriately modified version of this axiom. Namely we consider the canonical pairs $\{\hat{\pi}(f), \hat{\phi}(f);\ f \in S(\mathbb{R}^3)\}$ "smeared" only in $\mathbb{R}^3$ instead of the operators $\hat{\phi}(f)$, $f \in S(\mathbb{R}^4)$, "smeared" in $\mathbb{R}^4$ (Ref. 2). We denote this modified axiom by W′2. We define (cf. Note 1 below)

$$\begin{aligned} &\hat{\phi}_\alpha(f) := P(f)\hat{\phi}_\alpha(\hat{\mathbf{p}})P(f), \\ &\hat{\pi}_\alpha(f) := P(f)\hat{\pi}_\alpha(\hat{\mathbf{p}})P(f), \\ &P(f) = |f\rangle\langle f|, f \in S(\mathbb{R}^3), \quad \alpha = 1,...,N. \end{aligned} \tag{8.1}$$

Then the operators $\{\hat{\phi}(f), \hat{\pi}(f); f \in S(\mathbb{R}^3)\}$ are densely defined in $H^1 = \mathrm{Tr}\,\mathscr{H}_A = \mathscr{H} \otimes \mathrm{Tr}\,A$. The vector $\Omega = \Psi_0 = \Phi_0 \sqrt{\rho_0}$ is in the domain of any polynomial of the operators $(\hat{\phi}(f), \hat{\pi}(f))$ and the subspace $\mathscr{D}$ spanned algebraically by the vectors

$$\{\hat{\phi}_{i_1}(f_1)\hat{\pi}_{i_2}(f_2)\cdots\hat{\phi}_{i_{2N-1}}(f_{2n-1})\hat{\pi}_{i_{2N}}(f_{2n})\Phi_0\sqrt{\rho_0}: n \geqslant 0,\ f_j \in S(\mathbb{R}^3),\ j = 1,...,2n,\ i \in \mathrm{Per}(1,...,2N)\}$$

is dense in H^1. This follows from the facts that the family $(\hat{\phi},\hat{\pi})$ is $\mathscr{B}$ irreducible in $\mathscr{H}_{\hat{A}}$ [see Sec. III and (3.3)] and the elements $\{|f_i\rangle\langle f_k|: f_i, f_k \in S(\mathbb{R}^3)\}$ generate a dense subspace in $\mathrm{Tr}\,A$.[35] The field $\hat{\phi}_\alpha(f)$ is covariant under the action of $P_+^\uparrow$ in H^1 since

$$\begin{aligned} &U(a,\Lambda)\hat{\phi}_\alpha(f)U^{-1}(a,\Lambda) \\ &\quad = U(a,\Lambda)P(f)U^{-1}(a,\Lambda)U(a,\Lambda)\hat{\phi}_\alpha U^{-1}(a,\Lambda) \\ &\quad U(a,\Lambda)P(f)U^{-1}(a,\Lambda) = P((a,\Lambda)f)\hat{\phi}_\alpha P((a,\Lambda)f). \end{aligned}$$

However, $\hat{\phi}(f)$ and $\hat{\pi}(f)$ are not linear in f because, e.g.,

$$\begin{aligned} \hat{\phi}_\alpha(f_1 + f_2) &= P(f_1 + f_2)\hat{\phi}_\alpha P(f_1 + f_2) \\ &\neq P(f_1)\hat{\phi}_\alpha P(f_1) + P(f_2)\hat{\phi}_\alpha P(f_2) \\ &= \hat{\phi}_\alpha(f_1) + \hat{\phi}_\alpha(f_2). \end{aligned}$$

W3: If the supports of f and h are disjoint in $\mathbb{R}^3$ then

$$\hat{\phi}_\alpha(f)\hat{\phi}_\beta(h)=\hat{\pi}_\alpha(f)\hat{\pi}_\beta(h)=\hat{\pi}_\alpha(f)\hat{\phi}_\beta(h)=0$$

because $\langle f|h\rangle=0$ and, e.g.,

$$\hat{\phi}_\alpha(f)\hat{\phi}_\beta(h)=P(f)\hat{\phi}_\alpha P(f)P(h)\hat{\phi}_\beta P(h)$$
$$=P(f)\hat{\phi}_\alpha|f\rangle\langle f|h\rangle\langle h|\hat{\phi}_\beta P(h).$$

Consequently $\hat{\phi}_\alpha(f)$ and $\hat{\pi}_\alpha(f)$, respectively, commutes with $(\hat{\phi}(h),\hat{\pi}(h))$, but from the CCR's (3.1) and Eqs. (3.3)

$$[\hat{\pi}_\alpha(f),\hat{\phi}_\beta(f)]$$
$$=-i\delta_{\alpha\beta}c_{\alpha,f}^{-1}c_{\beta,f}\mathbb{1},\quad c_{\alpha,f}=\langle f|\hat{c}_\alpha|f\rangle,$$
$$c_{\alpha,f}^{-1}=\langle f|\hat{c}_\alpha^{-1}|f\rangle$$

and

$$[\hat{\pi}_\alpha(f),\hat{\pi}_\beta(f)]=[\hat{\phi}_\alpha(f),\hat{\phi}_\beta(f)]=0.$$

W4: The vector $\Omega=\Psi_0$ is the unique vector in H^1, up to the multiplication (from right) by an element $a\in A$ with the property $\operatorname{Tr}\rho_0 aa^*<\infty$, which is invariant under time translations. For

$$\exp(-i:\hat{\mathscr{P}}_0^{\text{in}}:)\Phi_0\sqrt{\rho_0}a=\Phi_0\sqrt{\rho_0}a$$

and Φ_0 is unique in $\mathscr{H}_{\hat{A}}$. Consequently, Ω is unique up to the operator $a\in A$ of the indicated property and not up to a complex number.

In summary we have the following.

Proposition 8.1: The q fields

$$\{\hat{\phi}(f),\hat{\pi}(f):f\in S(\mathbb{R}^3)\}$$

defined in (8.1) in the Hilbert space $H^1=\operatorname{Tr}\mathscr{H}_A$, with the vacuum vector $\Omega=\Psi_0=\Phi_0\sqrt{\rho_0}$, satisfy the axioms W1 and W3, and the axiom W'2 except for the linearity in f, and the axiom W4 in such a way that Ω is unique under time translations not only up to the multiplication by a complex number but up to the multiplication (from right) by an operator a of A with the property $\operatorname{Tr}\rho_0 aa^*<\infty$.

Now let us consider the Haag-Kastler axioms.

HK1: Since we deal with canonical formalism we give the correspondence $B\to\mathscr{A}(B)$ with the aid of an arbitrarily chosen time instant t. Let B_t be a bounded open set of the $t=$ constant hyperplane, i.e., $B_t\subset\mathbb{R}^3$. Here $\hat{B}_t$ denotes the causal shadow of B_t.[2] Let

$$B_t\to\mathscr{A}(B_t)=P(B_t)B(\mathscr{H}_A)P(B_t), \tag{8.2}$$

where $P(B_t)$ is the characteristic function of B_t, i.e., $P(B_t)\in\mathscr{P}$. Then, from Lemma 2.1,

$$\mathscr{A}(B_t)=P(B_t)(B(\mathscr{H})\otimes A)P(B_t)$$
$$=B(\mathscr{H})\otimes P(B_t)AP(B_t).$$

Thus $\mathscr{A}(B_t)$ is a C^*-algebra with identity and it contains all the local bounded observables of the system in B_t as self-adjoint elements in $\mathscr{A}(B_t)$. For the open bounded region $\hat{B}_t$ of $\mathbb{R}^4$ let the correspondence be

$$\hat{B}_t\to\mathscr{A}(\hat{B}_t):=\mathscr{A}(B_t) \tag{8.3}$$

(cf. Ref. 2). Because the sets of the form $\hat{B}_t$ provide the open sets of the Alexandrov topology of the Minkowski space $\mathbb{M}^4$, in this topology we have given the correspondence $B\to\mathscr{A}(B)$ between the open bounded sets of $\mathbb{M}^4$ and the C^*-algebras with identity. Let

$$\mathscr{A}=\bigcup_{\hat{B}_t}\mathscr{A}(\hat{B}_t)=\bigcup_{B_t}\mathscr{A}(B_t)$$
$$=\bigcup_{B_t}P(B_t)B(\mathscr{H}_A)P(B_t)=B(\mathscr{H}_A).$$

Since the $\hat{B}_t$'s at a fixed t cover $\mathbb{M}^4$, this $\mathscr{A}$ provides the C^*-algebra of the (quasilocal) bounded observables. [In fact, $\mathscr{A}=B(\mathscr{H}_A)=B(\mathscr{H})\otimes A$, according to Lemma 2.1, is a C^*-algebra with identity, moreover it is a factor of type I.[3]] $\mathscr{A}=B(\mathscr{H}_A)=B(\mathscr{H})\otimes A$ has an irreducible faithful representation because it is isomorphic to $B(H)$ where $H=L^2(\mathscr{H},\mathbb{R}^3)=L^2(\mathbb{R}^N\times\mathbb{R}^3)$.

HK2: Let $\hat{B}_t^1\subset\hat{B}_t^2$ then $B_t^1\subset B_t^2$ and $P(B_t^1)\subset P(B_t^2)$, so

$$\mathscr{A}(\hat{B}_t^1)=B(\mathscr{H})\otimes P(B_t^1)AP(B_t^1)$$
$$\subset B(\mathscr{H})\otimes P(B_t^2)AP(B_t^2)=\mathscr{A}(\hat{B}_t^2).$$

HK3: Let $\hat{B}_t^1$ be spacelike separated from $\hat{B}_t^2$ then $B_t^1\cap B_t^2=\varnothing$ and $P(B_t^1)P(B_t^2)=0$, thus

$$\mathscr{A}(\hat{B}_t^1)\mathscr{A}(\hat{B}_t^2)$$
$$=B(\mathscr{H})\otimes P(B_t^1)AP(B_t^1)P(B_t^2)AP(B_t^2)=0.$$

HK4: Let $\{a,\Lambda\}\in P_+^{\uparrow}$ then the map $\sigma_{\{a,\Lambda\}}F=U(a,\Lambda)FU^{-1}(a,\Lambda)$ is a *-automorphism of $\mathscr{A}$ where $F\in\mathscr{A}$ and $U(a,\Lambda)=\exp[-i(a^\mu\hat{\mathscr{P}}_\mu^{\text{in}}+\frac{1}{2}\Lambda^{\mu\nu}\hat{\mathscr{M}}_{\mu\nu}^{\text{in}})]$. Furthermore we have

$$\sigma_{\{a,\Lambda\}}(\mathscr{A}(\hat{B}_t))=\sigma_{\{a,\Lambda\}}(\mathscr{A}(B_t))=\sigma_{\{a,\Lambda\}}(P(B_t)B(\mathscr{H}_A)P(B_t))$$
$$=U(a,\Lambda)P(B_t)U^{-1}(a,\Lambda)U(a,\Lambda)B(\mathscr{H}_A)U^{-1}(a,\Lambda)U(a,\Lambda)P(B_t)U^{-1}(a,\Lambda)$$
$$=P(\{a,\Lambda\}B_t)B(\mathscr{H}_A)P(\{a,\Lambda\}B_t)=\mathscr{A}(\{a,\Lambda\}B_t)=\mathscr{A}(\{a,\Lambda\}\hat{B}_t)$$

and the map $\{a,\Lambda\}\to\sigma_{\{a,\Lambda\}}$ is a representation of $P_+^{\uparrow}$ in $\mathscr{A}=B(\mathscr{H}_A)$.

The axioms HK2 and HK3 have been verified for a fixed t; now, by applying HK4, we can verify them for arbitrary spacelike hypersurfaces in $\mathbb{M}^4$.

Finally, we have the following.

Proposition 8.2: The quasilocal algebra $\mathscr{A}=B(\mathscr{H}_A)$ corresponding to the bounded local observables of the models of QLFT satisfying the asymptotic condition (7.9), with the correspondence $\hat{B}_t\to\mathscr{A}(\hat{B}_t)$ given in (8.2) and (8.3), satisfies the axioms HK1–HK4.

In this way nontrivial models are constructed which satisfy the axioms of Haag–Kastler, while Wightman's axioms hold true only partially for them.

Notes: (1) It could seem natural, according to (3.3), to consider the operators $\hat{\phi}_\alpha(f)=\hat{q}_\alpha\hat{c}_\alpha(f)$ and $\hat{\pi}_\alpha(f)=\hat{p}_\alpha\hat{c}_\alpha^{-1}(f)$, $f\in S(\mathbb{R}^3)$, as the operators $\hat{\phi}(f)$ and $\hat{\pi}(f)$ appearing in the axiom W'2. But $\hat{c}_\alpha(f)$, $\hat{c}_\alpha^{-1}(f)\in L(\mathbb{R}^3)$, thus $\hat{\phi}_\alpha(f)$ and $\hat{\pi}_\alpha(f)$ do not define operators that are everywhere dense in $\mathscr{H}\otimes\hat{A}$ and also in H^1; the common domain of $\hat{q}_\alpha\hat{c}_\alpha(f)$ and $\hat{p}_\alpha\hat{c}_\alpha^{-1}(f)$ is $S(\mathbb{R}^N)$, i.e., they define linear operators only in $\mathscr{H}$ and not in $\mathscr{H}\otimes\hat{A}$. Consequently this choice does not satisfy W'2.

(2) It is clear that the collection $\mathscr{F} = \{\mathscr{A}(B_t) | B_t \subset \mathbb{R}^3, B_t$ is a bounded open set of $\mathbb{R}^3\}$ of C^* subalgebras of $\mathscr{A} = B(\mathscr{H}_A)$ is a net in the sense of Ref. 42.

IX. DISCUSSION

(a) It is clear that this canonical quantization method has a general status. Its application for more general CLFT including spin-$\frac{1}{2}$ and spin-1 fields, too, will be discussed in a subsequent paper, where QED and its numerical results will be also studied.

(b) Let $\mathscr{B}$ be a maximal atomic Boolean sublattice in $\mathscr{P}$, Ω be the dense set of isolated points of $\Gamma = \mathrm{Sp}\, \mathscr{B}$ (Ω is the set of atoms of $\mathscr{B}$). Then $\mathscr{H} \otimes \Gamma$ is the trivial Hilbert bundle corresponding to $\mathscr{H} \otimes B$. We can select out, in infinitely many different ways, from $\mathscr{H} \otimes \Omega$ an infinite sequence of copies of $\mathscr{H}$, then construct from these sequences the corresponding IDPS.[3] Every IDPS supports an irreducible representation of the usual ("decomposed" [cf. Eqs. (3.1) and (3.2)]) form of CCR's.[3] In this set of IDPS's there are infinitely many which support unitarily inequivalent representations of the usual form of the CCR.[3] In the present approach, thus unitarily inequivalent representations are "glued" together in a unique mathematical object $\mathscr{H}_A = \mathscr{H} \otimes A$, such that $\mathscr{H}_A$ supports a $\mathscr{B}$-irreducible representation of the CCR of the form (3.5), unique up to A-unitary equivalence.

(c) Intuitively it is clear that the q fields $\hat{\phi}_\alpha(x)$ satisfying (7.2) and (7.9) should correspond to the renormalized q fields, the asymptotic counterparts, $\hat{\phi}_\alpha^{\text{in/out}}(x)$, of which correspond to the physical (not bare) particles (the quanta of which are physical free q particles).[1] This is supported by the following intuitive line of thoughts (concerning a single scalar field). Following from the CCR (3.1), the commutator of $\hat{\phi}(\mathbf{x})$ and $\hat{\pi}(\mathbf{x}) = \dot{\hat{\phi}}(\mathbf{x})$ at identical point $\mathbf{x}$ equals $-i\mathbb{1}$ ($[\hat{\pi}(\mathbf{x}),\hat{\phi}(\mathbf{x})] = -i\mathbb{1}$) in contrast with the corresponding commutator of the bare field $\phi_0(\mathbf{x})$ of CQFT, for which

$$\lim_{\mathbf{x}'\to\mathbf{x}} [\dot{\phi}_0(\mathbf{x}'),\phi_0(\mathbf{x})] = \lim_{\mathbf{x}'\to\mathbf{x}} -i\delta^3(\mathbf{x}'-\mathbf{x}) = \infty.$$

In contrast with the latter commutator, the renormalized field $\phi(\mathbf{x}) = Z^{-1/2}\phi_0(\mathbf{x})$ commutes as follows: $[\dot{\phi}(\mathbf{x}'),\phi(\mathbf{x})] = -iZ^{-1}\delta^3(\mathbf{x}'-\mathbf{x})$.[1] Since Z is infinite (e.g., by applying a spatial lattice regularization and then letting the lattice spacing to vanish), thus Z normalizes $\delta^3(\mathbf{x}'-\mathbf{x})$, in an appropriate way, at $\mathbf{x}' = \mathbf{x}$, to the unity (in the mentioned regularization process), i.e., the renormalized field $\phi(\mathbf{x})$ also commutes at identical space points to $-i\mathbb{1}$. Consequently $\phi(\mathbf{x})$ satisfies the same commutation relation as $\hat{\phi}(\mathbf{x})$ and its basic dynamical equation is of the same form as Eq. (7.12).[40] It follows from this, essentially, that $\hat{\phi}(\mathbf{x})$ and $\phi(\mathbf{x})$ carry the same physical content, i.e., the elements of the S matrix belonging to $\hat{\phi}(\mathbf{x})$ should carry the same physical information as the (renormalized) S-matrix elements belonging to $\phi(\mathbf{x})$. [It is known that the renormalization makes definite the nonlinear term of the basic dynamical equation of the bare field $\phi_0(x)$, which term is *a priori* not defined.[5] Thus the dynamical equation of the renormalized field $\phi(x)$ has already a (formally) defined nonlinear term.[40] Then we could say that our formalism refers essentially directly to the renormalized fields and it derives the basic dynamical equations (5.12) and (7.12), respectively, for these fields without referring to the bare fields.] However, this intuitive line of thought requires a rigorous proof in the further research.

(d) In conclusion, by applying Hilbert A-module techniques and the $\mathscr{B}$-valued and $\mathscr{P}$-valued models of the mathematical extension theory, a nonconventional extension of the canonical quantization method of q mechanics to LFT is presented, which is a reconsideration of the conventional procedure in an alternative way offered by a recent new approach of CLFT.[23] We have shown that in the present approach Problems 1, 2, and 3 of CQFT (see Sec. I) do not occur. Furthermore, we have seen that this formalism is very "close" to the conventional theory: many important equations of the latter theory which do not have exact mathematical meaning recover, in the present formulation, as mathematically well-defined equations. Thus the *a priori* not defined basic dynamical equation of the conventional theory occurs in the present approach as well-defined equations. Furthermore, in this approach, the interaction picture is well-defined for nontrivial scattering matrices, too. The local S matrix is well defined, at least to the extent of the q mechanical S matrix. Feynman's graph technique can be applied with a straightforward modification because the S matrix of the model has the same structure as its conventional counterpart. The local and global scattering amplitudes of the local states (and the corresponding global states) are determined by the products of Feynman's propagators of the free q fields. However, that our model of quantized fields be physically realistic, one has still to show that the matrix element (6.29) in between initial and final global states, $\Psi_i = \Phi_i\sqrt{\rho_i}$ and $\Psi_f = \Phi_f\sqrt{\rho_f}$, respectively, which correspond to the many-particle states of the conventional theory, is, within an error bound, arbitrarily small, equal to the corresponding renormalized matrix element of the conventional theory.

In this paper we have demonstrated on a model theory that, in accordance with Schwinger's observation,[9] a theory of quantized fields based on a new, quantum conception of space-time[15,16] reveals a much higher mathematical regularity than the conventional theory based on the classical conception of space-time.

ACKNOWLEDGMENTS

I profited considerably from discussions with Professor A. S. Wightman in Erice in 1985 and I should like to thank Professor T. Matolcsi and Professor M. Huszár for valuable and stimulating discussions.

This work was supported by the Hungarian Academy of Sciences.

APPENDIX A: HILBERT A MODULES

The Hilbert A modules can be considered as the direct extension of complex Hilbert spaces in the following sense.

Definition A1: $\mathscr{H}_A$ is a Hilbert A module if $\mathscr{H}_A$ is a complex linear vector space and it is a right module over the C^* algebra A and it is equipped with an A-valued inner product $\langle\ |\ \rangle_A : \mathscr{H}_A \times \mathscr{H}_A \to A$ which has the following properties: for all $c \in \mathbb{C}$, $a \in A$, and $\phi_1, \phi_2, \phi \in \mathscr{H}_A$,

$$(1)\ \langle\phi|\phi_1+\phi_2\rangle_A = \langle\phi|\phi_1\rangle_A + \langle\phi|\phi_2\rangle_A,$$
$$\langle\phi_1|c\phi_2\rangle_A = c\langle\phi_1|\phi_2\rangle_A;$$
$$(2)\ \langle\phi_1|\phi_2 a\rangle_A = \langle\phi_1|\phi_2\rangle_A a;$$
$$(3)\ \langle\phi_1|\phi_2\rangle_A^* = \langle\phi_2|\phi_1\rangle_A;$$
$$(4)\ \langle\phi|\phi\rangle_A \geqslant 0, \text{ if } \langle\phi|\phi\rangle_A = 0 \text{ then } \phi = 0.$$

Finally $\mathscr{H}_A$ is complete in the norm $\|\cdot\|$ where $\|\phi\| = \|(\langle\phi|\phi\rangle_A)^{1/2}\|$ and $\|\ \|$ denotes the norm of A.[20,27]

Clearly, if A is commutative then this definition gives back the notion of Kaplansky's C^* modules.[26] We see from the definition that the class of Hilbert A modules is wider than the class of complex Hilbert spaces, while this wider class belongs of course to the class of Banach spaces.

Let us consider the following two examples of $\mathscr{H}_A$.

(1) $\mathscr{H}_A = A$, $A = B(H)$, $\langle\phi_1|\phi_2\rangle_A = a_1^* a_2$, a_1, $a_2 \in A$. We can easily verify Definition A1. (Note that $\langle a|a\rangle_A \geqslant 0$ iff $\langle\phi|a^*a|\phi\rangle = \langle a\phi|a\phi\rangle \geqslant 0$ where $\phi \in H$!)

(2) $\mathscr{H}_A = \mathscr{H} \otimes A$ where $\mathscr{H}$ is a complex separable Hilbert space and $\otimes$ denotes the algebraic tensor product, and the A-valued inner product is of the form

$$\langle\phi_1|\phi_2\rangle_A = \sum_k \bar{c}_k^1 c_k^2 \langle f_k^1 | f_k^2\rangle a_{k1}^* a_{k2}, \quad \phi_1, \phi_2 \in \mathscr{H}_A, \tag{A1}$$

where $c_k^1, c_k^2 \in \mathbb{C}$, $f_k^1, f_k^2 \in \mathscr{H}$, and $a_{k1}, a_{k2} \in A$. One can again check with direct verification that Definition A1 holds true ($\mathscr{H} \otimes A$ denotes the completion of the algebraic tensor product in the norm $\|(\langle\phi|\phi\rangle_A)^{1/2}\|$).

We note that the countably infinite direct sum $\mathscr{H}_A = \Sigma_1^\infty \oplus A$ of the Hilbert A module $\mathscr{H}_A = A$ of the first example is isomorphic to $\mathscr{H} \otimes A$ and the Hilbert A modules of this type is called *trivial* Hilbert A-modules[43] (or simply this type of Hilbert A modules is called as Hilbert spaces over A [20]). Another note is that one can define the unbounded extension $\mathscr{H}_{\hat{A}} = \mathscr{H} \otimes \hat{A}$ of $\mathscr{H} \otimes A$ as follows. In the definition of $\mathscr{H}_A$ we replace A by $\hat{A}$ and in the properties 1–4 we only require equalities on the set $\mathscr{D} = S(\mathbb{R}^3)$ everywhere dense in $H = L^2(\mathbb{R}^3)$, and the completeness is required in the $\hat{A}$-valued norm $\|\phi\|_{\hat{A}} = \sqrt{\langle\phi|\phi\rangle_{\hat{A}}}$, the equality is again restricted to the everywhere dense set $\mathscr{D}$.

Let us consider the operators in Hilbert A modules.

Definition A2: For Hilbert A modules $\mathscr{H}_A^1$ and $\mathscr{H}_A^2$ we denote by $B(\mathscr{H}_A^1, \mathscr{H}_A^2)$ the set of such maps $T: \mathscr{H}_A^1 \to \mathscr{H}_A^2$ that there exists $T^*: \mathscr{H}_A^2 \to \mathscr{H}_A^1$ satisfying the condition $\langle Tx|y\rangle_A = \langle x|T^*y\rangle_A$, $\forall x \in \mathscr{H}_A^1$, $\forall y \in \mathscr{H}_A^2$.

Lemma A3: Every map $T \in B(\mathscr{H}_A^1, \mathscr{H}_A^2)$ is a bounded linear A-module map. For $T \in B(\mathscr{H}_A^1, \mathscr{H}_A^2)$ the operator T^* is uniquely defined and belongs to $B(\mathscr{H}_A^2, \mathscr{H}_A^1)$. With the norm induced from the space of bounded linear operators on $\mathscr{H}_A$, $B(\mathscr{H}_A, \mathscr{H}_A) = B(\mathscr{H}_A)$ is a C^* algebra.[27]

Theorem A4 (Kasparov[20]): Let $\mathscr{H}_A$ be a Hilbert space over A. Then $B(\mathscr{H}_A) \cong M(K \otimes A)$ where $K = K(\mathscr{H})$, the C^* algebra of compact operators in $\mathscr{H}$ and $M(\mathscr{A})$ denotes the multiplier algebra of the C^* algebra $\mathscr{A}$ (cf. Sec. II B).

APPENDIX B: $\mathscr{B}$-VALUED AND $\mathscr{P}$-VALUED MODELS OF SET THEORY

Let $\mathscr{P}$ be the lattice of the orthogonal projectors of a complex separable Hilbert space H. Here $\mathscr{B}$ denotes a complete Boolean sublattice of $\mathscr{P}$.

It is a well-known fact of set theory that all statements and theorems of modern mathematics can be derived from the system of axioms of Zermelo and Fraenkel plus the axiom of choice (denoted this system by ZFC).[21] If V denotes the ordinary universe of set theory then we know trivially that V satisfies ZFC, i.e., a model of ZFC. Cohen and later Scott and Solovay had shown that there are nontrivial models in set theory which also satisfy ZFC. These models are called Boolean-valued (or $\mathscr{B}$-valued) models and their universes are denoted by $V^{(\mathscr{B})}$. These models differ from V in that while a sentence ϕ formulated in V takes its truth values $[\![\phi]\!]$ in the two elements Boolean algebra $\{0,1\}$, a sentence ϕ formulated in $V^{(\mathscr{B})}$ takes its truth values $[\![\phi]\!]$ in a complete Boolean algebra $\mathscr{B}$. We say that ϕ is *true* if $[\![\phi]\!] = 1$ (1 is the maximal element of $\mathscr{B}$), *false* if $[\![\phi]\!] = 0$ (0 is the minimal element of $\mathscr{B}$) and *undecidable* (or *true–false*) if $[\![\phi]\!] = b$ ($b \in \mathscr{B}$, $b \neq 1,0$) ("true as much as b" and "false as much as $1-b$"). If the usual sets, the elements of V, are represented by the corresponding character functions, then one can think of the elements of $V^{(\mathscr{B})}$ as "generalized sets," the character functions of which can take their values in $\mathscr{B}$.[21]

The following theorem constitutes the basis of Takeuti's $\mathscr{B}$-valued analysis (see Theorem 30 in Ref. 21 on pp. 55–57).

Theorem B1: Let ϕ be a theorem logically derivable from the axioms ZFC. Then $[\![\phi]\!] = 1$ in $V^{(\mathscr{B})}$, too.

This theorem provides us with a procedure to produce a new theorem in $V^{(\mathscr{B})}$ from an old theorem ϕ in V.

Let us see an example of $\mathscr{B}$-valued analysis. Since $V^{(\mathscr{B})}$ satisfies ZFC, thus we can construct real numbers in $V^{(\mathscr{B})}$ by Dedekind's cuts. Then the real numbers in $V^{(\mathscr{B})}$ can be interpreted as the self-adjoint operators in $\hat{B}$ where $\hat{B}$ consists of the linear operators in H having spectral projections exclusively from $\mathscr{B}$. In this case the procedure in Theorem B1 provides us with a machinery to transform theorems on real numbers into theorems on self-adjoint operators in $\hat{B}$.[31,17]

Now let us see the notion of Takeuti's q set theory.[22] Takeuti extended the $\mathscr{B}$-valued models in such a way that he replaced the Boolean algebra $\mathscr{B}$ of c logic with the Hilbert lattice $\mathscr{P}$ of q logic.[44] In this way he extended the c set theory based on c logic to a set theory based on q logic and called this set theory *q set theory*. In the universe $V^{(\mathscr{P})}$ of Takeuti's $\mathscr{P}$-valued models, a statement ϕ takes its truth values, $[\![\phi]\!]$, in $\mathscr{P}$ and the elements of $V^{(\mathscr{P})}$ can be thought of as generalized sets, the character functions of which take their values in $\mathscr{P}$. Takeuti had shown that a natural generalization of the ZFC axiom system holds true in $V^{(\mathscr{P})}$, thus a reasonable mathematics can be developed based on q logic.[22]

As an example for a mathematical object in $V^{(\mathscr{P})}$ we can see again the real numbers in $V^{(\mathscr{P})}$ defined by Dedekind's cuts. These real numbers in $V^{(\mathscr{P})}$ correspond to self-adjoint operators in H. Thus the theory of real numbers in

TABLE I. Feynman's graph rules.

Name	Graph element	Mathematical equivalent	Physical interpretation
outgoing ϕ line	t	$\hat{\phi}^{+}(t)$	ϕ emitted
incoming ϕ line	t	$\hat{\phi}^{-}(t)$	ϕ absorbed
Internal ϕ line	$t^1 \quad t^2$	$i\hat{\Delta}_F(t^1 - t^2)$	virtual ϕ
Vertex	t	g and $\int_{-\infty}^{+\infty} dt$	interaction
loop	t	$i\Delta^{+}(0)$	self-interaction at t

$V^{(\mathscr{P})}$ corresponds to the theory of self-adjoint operators in H.[22]

APPENDIX C: FEYNMAN'S GRAPH RULES

Feynman's graph rules for a single self-interacting scalar field $\hat{\phi}$ of mass m and of $\hat{\mathscr{H}}_I = g\hat{\phi}^n$, $n \geqslant 3$, are collected in Table I. The loop representing a "self-interaction" at the instant t appears because there is no normal ordering in $\hat{\mathscr{H}}_I$ [in this approach there is no need for normal ordering of fields $\hat{\phi}_\alpha$, the local observables are well-defined operators without normal ordering, too (see Sec. VI)]. Thus, according to the first Wick theorem, pairing of the field $\hat{\phi}$ with itself at the instant t^n also appears in the expansion of $S^{(n)}$. This is, with Eqs. (6.15), (6.16), (6.19), and (2.13), the following:

$$\hat{\phi}(t)\hat{\phi}(t) = i\hat{\Delta}^{+}(t-t) = \int_{\mathbb{R}^3} \frac{1}{2k_0}\, dP(\mathbf{k}). \tag{C1}$$

As to the interpretation of the graph elements we note that $\hat{\phi}(t)$ satisfies the operator KG equation, thus $\hat{\phi}(t)$ describes a field of KG particle pure states of mass m. Keeping in mind this we can say by the usual nomenclature that ϕ emitted (or ϕ absorbed) for $\hat{\phi}^{+}(t^1)$ [or $\hat{\phi}^{-}(t^2)$] at the time t^1, and now this means that at time t^1 the field of KG particle pure states was emitted on the "spacelike hypersurface" Sp $\mathscr{B}$ in q space-time[15] [and, respectively, for $\hat{\phi}^{-}(t^2)$]. The internal ϕ line also means the exchange of the field of virtual particle pure states on Sp $\mathscr{B}$. As a further note we mention that the integrals over the time variables in the integrand of $S^{(n)}$ can be trivially performed and one obtains that with every vertex an energy conservation factor will be associated. Every "elementary" local (i.e., at time t on a spacelike hypersurface Sp $\mathscr{B}$) interaction respects energy conservation. The local three-momentum conservation holds true trivially because S lies in $\hat{B}(\mathscr{H}) \otimes \hat{B}$ and is a function of $\hat{\mathbf{p}}$. Thus S connects only such local states which have the same amount of three-momentum [more precisely, the spectral projections $P(\mathbf{k})$ guarantee that the outgoing momenta will be equal to the incoming momenta, the local transition amplitude only for those parts of local states does not vanish which have the same amounts of momenta].

Feynman's graph rules can be obtained in the k_0 space for computing the A-valued S-matrix elements between special initial and final local states of the form $\hat{\Phi}_n = (1/\sqrt{n!})(\hat{a}^{+})^n\hat{\Phi}_0$, in a straightforward way from the rules of Table I.

[1] P. Roman, *Introduction to Quantum Field Theory* (Wiley, New York, 1969).
[2] N. N. Bogolubov, A. A. Logunov, and I. T. Todorov, *Introduction to Axiomatic Quantum Field Theory* (Benjamin, London, 1975).
[3] G. G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley–Interscience, New York, 1972).
[4] A. S. Wightman, in *High Energy Electromagnetic Interactions and Field Theory, Cargèse Lectures in Theoretical Physics* (Gordon and Breach, New York, 1967), pp. 171–279.
[5] J. Glimm and A. Jaffe, *Quantum Physics* (Springer, New York, 1981).
[6] I. E. Segal, Phys. Scr. **24**, 827 (1981).
[7] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964).
[8] R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964).
[9] J. Schwinger, *Selected Papers on Quantum Electrodynamics* (Dover, New York, 1958), p. xvi.
[10] E. Prugovečki, *Stochastic Quantum Mechanics and Quantum Space-time* (Reidel, Dordrecht, 1984).
[11] R. Penrose, in *Quantum Gravity—An Oxford Symposium*, edited by C. F. Isham, R. Penrose, and D. W. Schiama (Claredon, Oxford, 1975).
[12] M. Banai, in *Quantum Field Theory*, edited by F. Mancini (North-Holland, Amsterdam, 1986), pp. 97–118.
[13] M. Banai, Int. J. Theor. Phys. **20**, 147 (1981).
[14] M. Banai, Hadronic J. **5**, 1812 (1982); erratum **5**, 2155 (1982).
[15] M. Banai, Found. Phys. **15**, 1203 (1985).
[16] M. Banai, Int. J. Theor. Phys. **23**, 1043 (1984).
[17] M. Davis, Int. J. Theor. Phys. **16**, 867 (1977).
[18] M. Banai and B. Lukács, Lett. Nuovo Cimento **36**, 533 (1983).
[19] M. Banai and B. Lukács, Lett. Nuovo Cimento **42**, 235 (1985); erratum **43**, 160 (1985).
[20] G. G. Kasparov, J. Operator Theory **4**, 133 (1980).
[21] T. Jech, *Lectures in Set Theory* (Springer, Berlin, 1971).
[22] G. Takeuti, in *Current Issues in Quantum Logic*, edited by E. Beltrametti and B. van Fraassen (Plenum, New York, 1981), pp. 303–322.
[23] J. Kijowski and W. M. Tulczyjew, *A Symplectic Framework for Field Theories* (Springer, Berlin, 1979).
[24] R. Haag, in *Lectures in Theoretical Physics*, edited by W. E. Brittin, B. W. Downs, and J. Downs (Interscience, New York, 1961), Vol. III, pp. 353–381.
[25] J. von Neumann, *Mathematical Foundation of Quantum Mechanics* (Princeton U.P., Princeton, 1955).
[26] J. Kaplansky, Am. J. Math. **75**, 839 (1953).
[27] W. L. Paschke, Trans. Am. Math. Soc. **182**, 443 (1973).
[28] G. Takeuti, J. Math. Soc. Jpn. **35**, 1 (1983).
[29] K. Schmüdgen, Publ. RIMS Kyoto Univ. **19**, 601 (1983).
[30] E. Prugovečki, *Quantum Mechanics in Hilbert Space* (Academic, New York, 1981).
[31] G. Takeuti, J. Symbolic Logic **44**, 417 (1979).
[32] G. Takeuti, "C*-algebras and Boolean valued analysis," preprint, 1982.
[33] J. Dixmier and A. Douady, Bull. Soc. Math. France **91**, 227 (1963).
[34] A. M. Gleason, J. Math. Mech. **6**, 885 (1957).
[35] T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966).
[36] M. Guenin, J. Math. Phys. **7**, 271 (1966).
[37] M. H. Stone, Ann. Math. **33**, 643 (1932).
[38] T. Kato, Trans. Am. Math. Soc. **70**, 195 (1951); E. Wienholtz, Math. Ann. **435**, 50 (1958).
[39] I. M. Sigal and A. Soffer, "N-particle scattering problem: Asymptotic completeness for short-range systems," preprint (CALTECH), 1985.
[40] C. N. Yang and D. Feldman, Phys. Rev. **79**, 972 (1950).
[41] T. Bałaban and R. Rączka, in *Mathematical Physics and Physical Mathematics*, edited by K. Maurin and R. Rașzka (Reidel, Dordrecht, 1976), pp. 203–228.
[42] R. Haag, R. V. Kadison, and D. Kastler, Commun. Math. Phys. **16**, 81 (1970).
[43] P. Fillmore and M. J. Dupré, in *Operator Theory: Advanced Application* (Birkhauser, Basel, 1981), Vol. 2, pp. 71–79.
[44] G. Birkhoff and J. von Neumann, Ann. Math. **37**, 823 (1936).

Linear de Sitter gravity

C. Fronsdal
Department of Physics, University of California, Los Angeles, California 90024
W. F. Heidenreich
Institut für Theoretische Physik Technische Universität Clausthal, D-3392 Clausthal–Zellerfeld, West Germany

(Received 6 June 1986; accepted for publication 20 August 1986)

It is shown that there are two equivalent field potentials for massless spin-2 fields in (3,2)–de Sitter space; a symmetric two-tensor and a three-tensor of mixed symmetry. Each theory carries two inequivalent Gupta–Bleuler triplets, which correspond to the two helicities. They differ in their behavior at spatial infinity. All four fields appear when writing linear conformal gravity in de Sitter space. Two of them describe the unitary conformal gravitons, and two describe part of the conformal ghost.

I. INTRODUCTION

Einstein's theory of gravitation (with a nonzero cosmological constant) can be interpreted as a theory of a metric field; that is, a symmetric tensor field of rank 2, on a fixed de Sitter background.

Four-dimensional conformal gravity[1] cannot be expressed entirely in terms of a metric field, even in the linear approximation. The only natural vehicle for such a theory includes a tensor field of rank 3 and mixed symmetry. Therefore, to learn something from Einstein's theory of gravity one should begin by reformulating it in terms of a mixed symmetry tensor field. Here we study this problem in the linear approximation.

When Einstein's theory is approximated by expanding the metric field about the fixed de Sitter background, and discarding all but the leading terms, then one obtains a linear field theory that will be referred to as the metric form of linear de Sitter gravity.[2] It turns out that linear de Sitter gravity has an alternative formulation, in terms of a tensor field of rank 3 and mixed symmetry. In this paper we study linear de Sitter gravity in both of its alternative formulations.

The physical propagating states of linear de Sitter gravity are described by two copies of the massless representation $D(3,2)$. This is the unitary irreducible representation that is defined up to equivalence as follows. The spectrum of the energy operator is bounded below by $E_0 = 3$, and in the lowest energy eigenspace the angular momentum is $s = 2$. In the limit to zero curvature these physical representations turn into the massless helicity ± 2 representations of the Poincaré group. They can be extended to the massless helicity ± 2 representations of the conformal group SO(4,2) without enlarging the modules. The associated gauge fields transform by the Weyl-equivalent[3] representations $D(4,1)$, $D(0,2)$, and $D(-1,1)$; see Fig. 1.

We will, in Sec. II, get a first overview on the tensors that can describe massless spin-2 theories. In Sec. III we will show that the minimal tensors in linear de Sitter gravity are a symmetric two-tensor or a mixed three-tensor. We obtain a group-theoretical description of the solution spaces in the form of Gupta-Bleuler triplets, which exhibit the gauge freedom of the field potentials. A mapping between the two formulations is given and their properties in the limit to spatial infinity are discussed.

In Sec. IV we rewrite linear conformal gravity in de Sitter fields and discuss it in Sec. V using the results of Secs. II and III.

II. TENSOR STRUCTURE

In $\mathbb{R}^5$, coordinated by (u_μ), $\mu = 0, 1, 2, 3, 5$, consider the set of rays

$$u^2 = u_0^2 - u_1^2 - u_2^2 - u_3^2 + u_5^2 > 0, \quad u_\mu \triangleq \lambda u_\mu, \quad \lambda \neq 0. \tag{2.1}$$

In this paper, what is called de Sitter space is just this four-dimensional manifold. A more conventional formulation is obtained by using coordinates

$$\bar{u}_\mu = u_\mu/(\rho u^2)^{1/2}, \quad \bar{u}_R = u^2, \quad \rho > 0.$$

Then de Sitter space (2.1) becomes

$$\bar{u}_0^2 - \bar{u}_1^2 - \bar{u}_2^2 - \bar{u}_3^2 + \bar{u}_5^2 = \rho^{-1},$$

the auxiliary coordinate $\bar{u}_R$ distinguishes points along the rays. Employing the homogeneous coordinates u_μ here is mostly a matter of convenience. The limit to spatial infinity, $\bar{u}_\mu \to \infty$, is the limit $u^2 \to 0$. Spatial infinity itself is the three-dimensional manifold $u^2 = 0$.

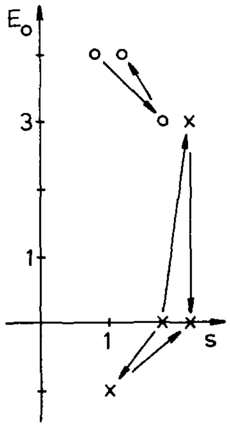

FIG. 1. The Gutpa–Bleuler triplets of linear de Sitter gravity. Shown are the (absolute and relative) lowest weights and the leaks.

 0022-2488/87/010215-06$02.50

It will be assumed that the fields of linear de Sitter gravity are tensor fields; our first job is to find out what are the possible choices of rank and symmetry. A homogeneous tensor field of rank n and degree N on $\mathbb{R}^5$ is a n-linear function of n vector variables,

$$\begin{aligned} &\psi(u,z_1,\ldots,z_n) = \psi_{\mu\nu\cdots}(u) z_1^\mu z_2^\nu \cdots, \\ &\hat{N}\psi = N\psi, \quad \hat{N} \equiv u\cdot\partial_u. \end{aligned} \tag{2.2}$$

The degree of homogeneity N can be changed arbitrarily by multiplying with $(u^2)^d$, $d\in\mathbb{R}$, on de Sitter space; but not on spatial infinity $u^2 = 0$. Fields $\psi(\bar{u}_\mu)$ in coordinates $\bar{u}_\mu$ are obtained by

$$\psi(u_\mu) = \psi((\rho\bar{u}_R)^{1/2}\bar{u}_\mu) = (\rho u_R)^{N/2}\psi(\bar{u}_\mu).$$

We consider in this section tensors of specific symmetry that are transverse, traceless, and divergenceless; that is,

$$u_\mu \psi_{\mu\nu\cdots} = 0, \quad \psi_{\mu\mu\nu\cdots} = 0, \quad \partial_\mu \psi_{\mu\nu\cdots} = 0. \tag{2.3}$$

The action of a basis $L_{\mu\nu} = -L_{\nu\mu}$ of the Lie algebra so(3.2) on tensor fields is given by

$$L_{\mu\nu}\psi = (M_{\mu\nu} + S_{\mu\nu})\psi, \quad M_{\mu\nu} = i(u_\mu\,\partial_\nu - u_\nu\,\partial_\mu),$$

$$S_{\mu\nu} = i\sum_k \left(z_{k\mu}\frac{\partial}{\partial z_k^\nu} - z_{k\nu}\frac{\partial}{\partial z_k^\mu}\right).$$

The second-order Casimir operator is

$$Q = \tfrac{1}{2}L_{\mu\nu}L_{\mu\nu}.$$

For tensors satisfying Eqs. (2.3), one finds

$$Q = \hat{N}(\hat{N}+3) - u^2\,\partial^2 + n_1(n_1+1) + n_2(n_2-1),$$

where $n_1 \geqslant n_2 \geqslant 0$ are integers that label the symmetry type according to the lengths of the rows of the Young diagrams. We do not have to consider three or more rows, as such tensors can be transformed to two-row tensors by multiplying with the fully antisymmetric ϵ-tensor.

For the positive energy representation $D(E_0,s)$, Q takes the value $E_0(E_0-3) + s(s+1)$; for gravitons $D(3,2)$, $Q = 6$. We choose the degree of homogeneity N so that the wave equation $(Q-6)\psi = 0$ takes the simple form

$$u^2\,\partial^2\psi = 0. \tag{2.4}$$

This is convenient, especially for the passage to the limit to spatial infinity, $u^2 \to 0$. Thus N should satisfy

$$N(N+3) + n_1(n_1+1) + n_2(n_2-1) = 6.$$

Here is a list of the possible integer degrees and associated symmetry types (n_1,n_2):

$$\begin{array}{lccccccccccc} \text{degree} & +1 & & 0 & & -1 & & -2 & & -3 & & -4 \\ \text{symmetry} & (1,0) & \longrightarrow & (2,0) & \multicolumn{3}{c}{\longrightarrow} & (2,2) & \longleftarrow & (2,1) & \longleftarrow & (1,1) \\ & (1,1) & \longrightarrow & (2,1) & \longrightarrow & (2,2) & \multicolumn{3}{c}{\longleftarrow} & (2,0) & \longleftarrow & (1,0). \end{array} \tag{2.5}$$

The arrows will be explained next.

Coboundary operator: The spaces of tensor fields defined by Eqs. (2.3), (2,4), and (2.5) are not irreducible. Certain invariant subspaces can be defined in terms of a coboundary operator d. This operator acts within a space defined by Eqs. (2.3)–(2.5). Each arrow in (2.5) represents an action of d. It acts on tensor fields that satisfy (2.3) and (2.4) to give tensor fields that satisfy the same equations. To express the action of d we first introduce the operator δ_μ defined by

$$\delta_\mu\psi = \begin{cases} \partial_\mu\psi, & \text{for degree } (\psi) \geqslant -\tfrac{3}{2}, \\ u^2\,\partial_\mu - u_\mu(2\hat{N}+3), & \text{for degree } (\psi) \leqslant -\tfrac{3}{2}. \end{cases} \tag{2.6}$$

The degree must be an integer in de Sitter space (2.1); it could be real in the universal covering space.

Now d is defined by

$$(d\phi)_{\mu\nu} = (2,0)\delta_\mu\Phi_\nu \equiv \delta_\mu\Phi_\nu + \delta_\nu\Phi_\mu, \tag{2.7}$$

$$(d\Lambda)_{\mu\nu\lambda} = (2,1)\delta_\lambda\Lambda_{\mu\nu} \equiv 2\delta_\lambda\Lambda_{\mu\nu} - \delta_\mu\Lambda_{\nu\lambda} - \delta_\nu\Lambda_{\lambda\mu}, \tag{2.8}$$

$$\begin{aligned} (dh)_{\mu\nu\lambda\rho} &= (2,2)\delta_\lambda\delta_\rho h_{\mu\nu} \\ &\equiv \delta_\nu\delta_\rho h_{\mu\lambda} - \delta_\lambda\delta_\nu h_{\rho\mu} \\ &\quad - \delta_\rho\delta_\mu h_{\nu\lambda} + \delta_\mu\delta_\lambda h_{\nu\rho}, \end{aligned} \tag{2.9}$$

$$\begin{aligned} (d\psi)_{\mu\nu\lambda\rho} &= (2,2)\delta_\rho\psi_{\mu\nu\lambda} \\ &\equiv \delta_\rho\psi_{\mu\nu\lambda} - \delta_\lambda\psi_{\mu\nu\rho} + \delta_\nu\psi_{\lambda\rho\mu} - \delta_\mu\psi_{\lambda\rho\nu}. \end{aligned} \tag{2.10}$$

As before (n_1,n_2) stands for the corresponding Young symmetrizer; Λ is antisymmetric, h is symmetric, and ψ is of mixed symmetry,

$$\begin{aligned} &\Lambda_{\mu\nu} = -\Lambda_{\nu\mu}, \quad h_{\mu\nu} = h_{\nu\mu}, \\ &\psi_{\mu\nu\lambda} = -\psi_{\nu\mu\lambda}, \quad \psi_{\mu\nu\lambda} + \psi_{\nu\lambda\mu} + \psi_{\lambda\mu\nu} = 0. \end{aligned} \tag{2.11}$$

One easily checks that

$$\delta_\mu\delta_\nu = \delta_\nu\delta_\mu,$$

and that $d\cdot d = 0$. Furthermore, every sequence of the form

$$\psi \xrightarrow{d} \psi' \xrightarrow{d} \psi''$$

is exact.

III. FIELD EQUATIONS AND SOLUTION SPACES

A. Tensor products

Fixing the symmetry of our traceless tensors means that the constant tensor carries a finite representation $D_f = D(-n_1,n_2)$ of SO(3.2). Fixing the degree of homogeneity to N' means that the scalar field equation (2.4) has positive energy solution spaces that carry $D(-N',0)$ and $D(3+N',0)$. The two solution spaces differ in their transformations under $Ru = -u$: $D(E_0,0)$ transforms like[4]

$$\psi(-u) = (-1)^{E_0}\psi(u).$$

They also behave differently at spatial infinity: states in $D(-N',0)$ are finite on $u^2 = 0$, while those in $D(3+N',0)$

are of the form $(u^2)^{3/2-N'}\Phi$, with Φ finite on $u^2=0$.

We have two options: either we use one field with degree N' (or 3-N'), which describes two positive energy Hilbert spaces carried by functions symmetric or antisymmetric under R; or we use two fields with $N=N'$ and $N=3-N'$, and fix the symmetry. We choose the second possibility with

$$\psi(-u)=(-1)^N\psi(u). \tag{3.1}$$

Then the scalar Eq. (2.4) carries $D(-N,0)$ and all fields are finite at spatial infinity.

The tensor field which satisfies (2.4) and (3.1) carries the tensor product

$$D_f\otimes D(-N,0), \tag{3.2}$$

of a finite and an infinite representation. Comparing weight diagrams all the tensor products for the cases of Eq. (2.5) can be calculated[5] up to the possibility of indecomposable representations with Weyl-equivalent subquotients. For our present purposes it is sufficient to consider only representations that are Weyl equivalent to the graviton $D(3,2)$, that is $D(-1,1)$, $D(2,0)$, $D(3,2)$, and $D(4,1)$.

In the cases of (1,0)- and (1,1)-symmetry we find that the tensor product (3.2) contains

$$\begin{aligned} &D(-1,1) \text{ and } D(0,2) \text{ for } N=1,\\ &\qquad D(4,1) \text{ for } N=2; \end{aligned} \tag{3.3}$$

in the cases of (2,0)- and (2,1)-symmetry it contains

$$\begin{aligned} &D(3,2) \text{ and two times } D(0,2) \text{ and } D(-1,1) \text{ for } N=0,\\ &D(3,2) \text{ and two times } D(4,1) \text{ for } N=-3. \end{aligned} \tag{3.4}$$

Finally in the cases with (2,2)-symmetry it contains

$$\begin{aligned} &D(0,2) \text{ and } D(4,1) \text{ and two times } D(3,2) \text{ for both}\\ &\quad N=-1,-2. \end{aligned} \tag{3.5}$$

The minimal tensors that describe gravitons have (2,0)- or (2,1)- symmetry with $N=0$ or $N=-3$. The two possible degrees will replace in de Sitter space the two helicities; the two possible symmetries will become connected by a kind of ϵ-duality transformation.

B. Field equations and Gupta–Bleuler triplets

Field potentials describe massless particles not irreducibly but they have gauge freedom. Here we will give field equations for the full theories, generalized Lorentz conditions, and the form of pure gauge fields. It turned out that the states in the reduction (3.4) are precisely those necessary for Gupta–Bleuler triplets: The potentials h and ψ with $N=0$ carry

$$D(0,2)\rightarrow(D(3,2)\oplus D(-1,1))\rightarrow D(0,2), \tag{3.6}$$

with "scalar" and gauge modes $D(0,2)$, physical modes $D(3,2)$, and a ten-dimensional finite $D(-1,1)$. The potentials with $N=-3$ carry

$$D(4,1)\rightarrow D(3,2)\rightarrow D(4,1). \tag{3.7}$$

We will give justification for these forms in the sequel. They can be rigorously proved by calculation of all absolute and relative ground states and leakages between them, as will be demonstrated for pure gauge fields in the next section.

C. Pure gauge fields: vectors and (1,1)-tensors

The lowest weights of the vector ϕ_μ are for $N=1$

$$(-1,1)_0=u_+z_i-u_iz_+,$$
$$(0,2)_0=u_iz_j+u_jz_i-2u_+^{-1}(u_iu_j-\tfrac{1}{3}u_ku_k\delta_{ij})z_+,$$

and for $N=-4$,

$$(4,1)_0=u_+^{-5}(u_+z_i-u_iz_+),$$
$$\text{with } u_+=u_0+iu_5,\quad z_+=z_0+iz_5.$$

Acting with an energy-lowering operator $M_i^-=M_{i0}-iM_{i5}$ on the states $(0,2)_0$ reveals a leak to the states $(-1,1)_0$. So the solution space of (2.4) carries—among other things—a $D(0,2)\rightarrow D(-1,1)$. A direct calculation shows that these states and therefore the full representations fulfill additional field equations

$$u\cdot\partial_z\phi=0,\quad \delta_u\cdot\delta_z\phi=0.$$

These together with Eq. (2.4) fix the second-order Casimir operator to $Q=6$. So they project out those terms in the tensor products (3.2) which are not Weyl equivalent to $D(3.2)$. For the states $(-1,1)_0$ only holds $z\cdot\delta_u\phi=0$, i.e., $d\phi=0$ projects on the invariant subspace $D(-1,1)$. The symmetric tensor $d\phi$ carries $D(0,2)$ [resp. $D(4,1)$] only. Similarly we find for the (1,1)-tensor $\Lambda_{\mu\nu}$ that

$$u\cdot\delta_z\Lambda=0,\quad \delta_u\cdot\delta_z\Lambda=0$$

project on $D(0,2)\rightarrow D(-1,1)$ for $n=1$ and $D(4,1)$ for $N=-4$. As before $d\Lambda=0$ projects on $D(-1,1)$; the mixed tensor $d\Lambda$ carries $D(0,2)$ [resp. $D(4,1)$] only.

D. The potentials: (2,0)- and (2,1)-tensors

The "gradients" $d\phi$ and $d\Lambda$ are nonvanishing for $D(0,2)$ and $D(4,1)$; they are pure gauge states of the $h_{\mu\nu}$ and $\psi_{\mu\nu\rho}$ tensors. The expressions $u_\mu h_{\mu\nu}$ and $u_\mu\psi_{\mu[\nu\rho]}$ for $N=0$ and $u^{-2}u_\mu h_{\mu\nu}$, $u^{-2}\,u_\mu\psi_{\mu[\nu\rho]}$ for $N=-3$ fulfill all equations of the ϕ and Λ gauge fields. So they too carry the representations $D(0,2)\rightarrow D(-1,1)$ [resp. $D(4,1)$]. We get the following field equations for the potentials h: the full Gupta–Bleuler-triplets (3.6) and (3.7) fulfill

$$\begin{aligned} &(u\cdot\partial)h^{(0)}=0\quad\text{or}\quad(u\cdot\partial)h^{(-3)}=-3h^{(-3)},\\ &h_{\mu\nu}=h_{\nu\mu},\quad h_{\mu\mu}=0,\\ &u_\mu u_\nu h_{\mu\nu}=0,\quad \delta_\mu h_{\mu\nu}=0,\quad \partial_u^2h=0. \end{aligned} \tag{3.8}$$

The Lorentz conditions, which map on physical and gauge modes, are

$$u_\mu h_{\mu\nu}=0. \tag{3.9}$$

For $N=0$, the finite modes satisfy $d(u\cdot h)=0$. The pure gauge fields are of the form

$$h=d\phi. \tag{3.10}$$

The corresponding equations for the ψ field are

$$\begin{aligned} &(u\cdot\partial)\psi^{(0)}=0,\text{ or }(u\cdot\partial)\psi^{(-3)}=-3\psi^{(-3)},\\ &\psi_{\mu\nu\rho}=-\psi_{\nu\mu\rho},\quad \psi_{\mu\nu\nu}=0,\\ &\psi_{\mu\nu\rho}+\psi_{\nu\rho\mu}+\psi_{\rho\mu\nu}=0,\\ &u_\mu\psi_{\mu(\nu\rho)}=0,\quad \delta_\mu\psi_{\mu\nu\rho}=0, \end{aligned} \tag{3.11}$$

for the full triplets. The Lorentz conditions are

$$u_\mu \psi_{\mu[\nu\rho]} = 0, \tag{3.12}$$

pure gauge fields have the form

$$\psi = d\Lambda. \tag{3.13}$$

With these field equations, the physical and gauge modes fulfill $(Q-6)h = (Q-6)\psi = 0$, while for the scalar modes we have $(Q-6)h =$ gauge and $(Q-6)\psi =$ gauge.

E. Field strengths: the (2,2)-tensors

Finally we investigate the (2,2)-tensors C. Specifically we want to show that the proportionalities

$$dh^{(0)} \propto d\psi^{(-3)} \quad \text{and} \quad dh^{(-3)} \propto d\psi^{(0)} \tag{3.14}$$

hold. We have to exclude the *a priori* possibility that the two copies of $D(3,2)$ to each degree in the tensor products (3.5) correspond to the two different potentials ψ and h.

The tensors dh and $d\psi$ fulfill the field equations (2.4) and all subsidiary conditions of C,

$$C_{\mu\nu\rho\sigma} = -C_{\nu\mu\rho\sigma} = C_{\rho\sigma\mu\nu}, \quad C_{\mu\nu\mu\sigma} = 0,$$
$$u_\mu C_{\nu\mu\rho\sigma} = 0, \quad \delta_\mu C_{\mu\nu\rho\sigma} = 0. \tag{3.15}$$

A straightforward calculation shows that $u_\mu C^{(-1)}_{\mu\nu\rho\sigma}$ fulfills all Eqs. (3.11) of the ψ-tensor and its Lorentz condition (3.12), if $\delta C^{(-1)} = 0$ and $u_\mu u_\rho \, C^{(-1)}_{\mu\nu\rho\sigma} = 0$. So it carries a $D(3,2) \to D(2,0)$. Similarly $u_\mu u_\rho \, C^{(-2)}_{\mu\nu\rho\sigma}$ fulfills all Eqs. (3.8) and (3.9) of $h^{(0)}$ if $\delta \cdot C^{(-2)} = 0$, and also carries a $D(3,2) \to D(2,0)$.

Therefore one of the two copies of $D(3,2)$ for each degree in the tensor products (3.5) is carried by $u \cdot C^{(-1)}$ [resp. $uu \cdot C^{(-2)}$]. We conclude that the conditions $u \cdot C = \delta \cdot C = 0$ project on maximally one $D(3,2)$ to each degree. This strongly suggests Eqs. (3.14), that is, $C^{(-1)}$ has two potentials $\psi^{(0)}$ and $h^{(-3)}$, $C^{(-2)}$ has two potentials $\psi^{(-3)}$ and $h^{(0)}$.

It will be shown in Sec. V, that the conformal spin-2 field satisfies equations of the type (3.14).

We have two field strength tensors $C^{(-1)}$ and $C^{(-2)}$ corresponding to the two helicities ± 2. The two formulations of spin-2 theories in de Sitter space, with h and ψ tensors, describe the same Gupta–Bleuler triplets, that is, equivalent free theories. There is a mapping between them.

F. Isomorphy of the h and ψ tensor potentials

Let us define two tensors

$$\bar{\psi}_{\mu\nu\lambda} = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma\tau} u_\rho \partial_\sigma h_{\tau\lambda}, \quad \bar{h}_{\mu\lambda} = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma\tau} u_\nu \partial_\rho \psi_{\sigma\tau\lambda}. \tag{3.16}$$

Explicit calculation shows that $\epsilon_{\mu\nu\rho\sigma\tau}\bar{h}_{\sigma\tau} = 0$ if ψ satisfies the field equations (3.11) and the Lorentz condition (3.12). So $\bar{h}_{\mu\nu}$ is symmetric. Similarly $\epsilon_{\mu\nu\rho\sigma\tau}\bar{\psi}_{\rho\sigma\tau} = 0$ if h satisfies (3.8) and (3.9). So $\bar{\psi}$ is antisymmetric in the first two indices, but not fully antisymmetric: it has the same symmetry as ψ. Finally we can show under the same conditions that equating $\bar{h} = h$ gives $\bar{\psi} = \psi$ and the reverse. So Eqs. (3.15) describe an isomorphy between the two potentials ψ and h, if they both satisfy their field equations and their Lorentz conditions.

The mapping of the scalar modes can be accomplished by adding gauge terms of form $\epsilon\, \partial(u_\mu h_{\mu\nu})$, $\epsilon\, \partial(u_\mu \psi_{\mu\nu\rho})$.

G. The de Sitter parity, helicity, and limit to spatial infinity

The de Sitter parity transformation distinguishes between the two degrees of the h and the ψ fields:

$$h^{(0)}(-u) = +h^{(0)}(u), \quad \psi^{(0)}(-u) = +\psi^{(0)}(u),$$
$$h^{(-3)}(-u) = -h^{(-3)}(u), \quad \psi^{(-3)}(-u) = -\psi^{(-3)}(u). \tag{3.17}$$

This reflection replaces the concept of helicity in de Sitter space. In bipolar coordinates $\bar{u}_i = \rho^{-1}\sinh(r)\ (\vartheta,\varphi)$; $i = 1, 2, 3$, $u_0 + iu_5 = \rho^{-1}\cosh(r)\, e^{i\tau}$ it reflects space coordinates at the origin and maps time τ to $\tau + \pi$. Initial data given on $\tau = 0$ completely determine the fields on $\tau = \pi$. No information can come in from spatial infinity.[6]

In the limit $u^2 \to 0$ or $r \to \infty$ to spatial infinity the two degrees behave differently: $h^{(0)}$ and $\psi^{(0)}$ approach ∞ like 1, $h^{(-3)}$ and $\psi^{(-3)}$ like $\cosh^{-3}(r)$.

IV. CONFORMAL LINEAR GRAVITY ON DE SITTER SPACE

Conformal linear gravity[1] was written down in Dirac's six-cone formalism, then translated to Minkowski notation. Here we interpret it as field theory on de Sitter space.

Six-space coordinates (y_α) are related to de Sitter coordinates by

$$u_\mu = y_\mu, \quad u_B = y^2, \quad y_4 = (u^2 - u_B)^{1/2}.$$

The range of indices, once and for all, is

$$\mu,\nu,\ldots = 0,1,2,3,5; \quad \alpha,\beta,\ldots = 0,1,2,3,4,5;$$
$$a,b,\ldots = 0,1,2,3,5,B.$$

Six-tensors $\tilde{\psi}(y)$ are related to complexes of five-tensors by

$$\tilde{\psi}_{\alpha\beta\cdots}(y) = u^a_\alpha u^b_\beta \cdots \psi_{ab\cdots}(u), \quad u^a_\alpha = u^a/\partial y^\alpha.$$

This works out to

$$\tilde{\psi}_\mu(y) = \psi_\mu(u) + 2u_\mu \psi_B(u), \quad \tilde{\psi}_4(y) = 2y_4\psi_B(u), \tag{4.1}$$

and similarly for tensors of higher rank.

The intrinsic gradient operator, on tensor fields of degree N, is

$$\text{grad}_\alpha = (N+1)\partial_\alpha - \tfrac{1}{2}y_\alpha\, \partial^2.$$

It gives rise to a "covariant derivative" ∇_a:

$$\text{grad}_\alpha \tilde{\psi}_{\beta\ldots}(y) = u^a_\alpha u^b_\beta \cdots \nabla_a \psi_{b,\ldots}(u). \tag{4.2}$$

After some calculation one finds that

$$\nabla_\mu \psi_{\nu\lambda\cdots} = (N+1)\{\partial_\mu + nu_\mu/u^2)\psi_{\nu\lambda\cdots} + 2\Gamma_{\mu\nu}\psi_{B\lambda\cdots} + \cdots\},$$
$$\nabla_B \psi_{\nu\lambda} = -\tfrac{1}{4}\{\partial^2 + n(2N+3-n)/u^2\}\psi_{\nu\lambda\cdots} - \{\partial_\nu + (n-1-N)u_\nu/u^2\}\psi_{B\lambda\cdots} - \cdots - 2\Gamma_{\nu\lambda}\psi_{BB\cdots} - \cdots. \tag{4.3}$$

Here $\partial^2 = (\partial/\partial u^\mu)^2$, n is the number of B-indices, the connection coefficients are

$$\Gamma_{\mu\nu} = \delta_{\mu\nu} - u_\mu u_\nu/u^2,$$

and every pair of Greek indices gives rise to a term like the last one.

The equations

$$\tilde{\psi}_{\alpha\beta\beta}=0, \quad y_\alpha \tilde{\psi}_{\alpha\beta}=0, \quad \mathrm{grad}_\alpha\, \tilde{\psi}_{\alpha\beta}=0,$$

take the following form in terms of ψ:

$$\psi_{\mu\mu c\cdots}+2u_\mu(\psi_{B\mu c\cdots}+\psi_{\mu Bc\cdots})=0, \quad u_\mu \psi_{\mu b\cdots}=0, \tag{4.4}$$

$$\nabla_\mu \psi_{\mu b\cdots}+2u_\mu(\nabla_B \psi_{\mu b\cdots}+\nabla_\mu \psi_{Bb\cdots})=0. \tag{4.5}$$

The tensor field $\widetilde{\Psi}$ of linear conformal gravity is a traceless tensor of rank 3, mixed symmetry (and antisymmetric in the first two indices), and degree zero. The corresponding complex (Ψ_{abc}) contains the following four tensors:

$$\bar{\gamma}_{\mu\nu\lambda}=\Psi_{\mu\nu\lambda}, \quad \bar{h}_{\mu\nu}=\bar{h}_{\nu\mu}=\Psi_{B\mu\nu}+\Psi_{B\nu\mu},$$
$$\bar{f}_{\mu\nu}=-\bar{f}_{\nu\mu}=\Psi_{\mu\nu B}, \quad \bar{a}_\mu=\Psi_{\mu BB}. \tag{4.6}$$

Tracelessness, the first of Eqs. (3.4), is expressed by

$$\bar{\gamma}_{\mu\nu\nu}+(3\bar{f}_{\mu\nu}-\bar{h}_{\mu\nu})u_\nu=0, \quad \bar{h}_{\mu\mu}-4u_\mu \bar{a}_\mu=0. \tag{4.7}$$

The first subsidiary condition, the second of Eqs. (4.4), becomes

$$u_\mu \bar{\gamma}_{\mu\nu\lambda}=u_\mu \bar{h}_{\mu\nu}=u_\mu \bar{f}_{\mu\nu}=u_\mu \bar{a}_\mu=0. \tag{4.8}$$

The second subsidiary condition, Eq. (4.5), reads

$$\partial_\mu \bar{\gamma}_{\mu\nu\lambda}+2\bar{h}_{\nu\lambda}-4\bar{f}_{\nu\lambda}=0, \tag{4.9}$$

$$\partial_\mu \bar{h}_{\mu\nu}=2\partial_\mu \bar{f}_{\mu\nu}=8\bar{a}_\nu, \tag{4.10}$$

if (4.8) holds.

The main dynamical equation of linear conformal gravity is

$$(2,2)\mathrm{grad}_\delta\, \widetilde{\Psi}_{\alpha\beta\gamma}=\widetilde{S}_{\alpha\beta\gamma\delta}, \quad \text{i.e.} \quad (2,2)\nabla_d \Psi_{abc}=S_{abcd}. \tag{4.11}$$

The degrees of $\widetilde{\Psi}$ and $\widetilde{S}$ are 0 and -1; that is

$$\begin{aligned} &u_\mu \nabla_\mu \Psi=0, \quad (u_\mu \nabla_\mu+1)S=0, \\ &\hat{N}\bar{\gamma}=(\hat{N}+1)\bar{h}=(\hat{N}+1)\bar{f}=(\hat{N}+2)\bar{a}=0. \end{aligned} \tag{4.12}$$

The source S is traceless, transverse, and divergenceless:

$$\begin{aligned} &S_{\mu b\mu d}=0, \quad u_\mu S_{\mu bcd}=0, \\ &\partial_\mu S_{\mu bcd}+(4-2n)S_{Bbcd}=0. \end{aligned} \tag{4.13}$$

As before, n denotes the number of B-indices.

Writing out Eq. (4.11) explicitly we find

$$S_{\mu\nu\lambda\rho}=(2,2)(\partial_\rho \bar{\gamma}_{\mu\nu\lambda}+\Gamma_{\mu\rho}\bar{h}_{\nu\lambda}-\Gamma_{\nu\rho}\bar{h}_{\mu\lambda}), \tag{4.14}$$

$$\begin{aligned} S_{\mu\nu B\rho}=&\tfrac{1}{4}\partial^2\bar{\gamma}_{\mu\nu\rho}+(2,1)\partial_\rho \bar{f}_{\mu\nu}+(u_\nu \bar{h}_{\mu\rho}-u_\mu \bar{h}_{\nu\rho})/u^2 \\ &+4(\Gamma_{\nu\rho}\bar{a}_\mu-\Gamma_{\mu\rho}\bar{a}_\nu), \end{aligned} \tag{4.15}$$

$$\begin{aligned} S_{B\nu B\rho}=&\tfrac{1}{4}(\partial^2+2/u^2)\bar{h}_{\nu\rho}-2(\partial_\rho+u_\rho/u^2)\bar{a}_\nu \\ &-2(\partial_\nu+u_\nu/u^2)\bar{a}_\rho. \end{aligned} \tag{4.16}$$

It is a good check to use these equations to verify (4.13).

Finally, the wave equation is

$$(\partial^2)^2\widetilde{\Psi}=\tilde{j}, \quad \text{i.e.,} \quad 16\nabla_B^2 \Psi_{abc}=j_{abc}. \tag{4.17}$$

It follows that

$$\begin{aligned} &(2,2)y_\delta(\partial^2)^2\widetilde{\Psi}_{\alpha\beta\gamma} \\ &\quad=-2\partial^2(d\widetilde{\Psi})_{\alpha\beta\gamma\delta}=-2\,\partial^2\widetilde{S}_{\alpha\beta\gamma\delta}; \end{aligned} \tag{4.18}$$

that is,

$$(2,2)\delta_d^B j_{abc}=-2\,\partial^2(d\Psi)_{abcd}=-2(\partial^2 S)_{abcd}. \tag{4.19}$$

This determines $j_{\mu\nu\lambda}$ and $j_{B\mu\nu}+j_{B\nu\mu}$, but not $j_{\mu\nu B}$ or $j_{\mu BB}$. Hence we have further constraints on S,

$$\begin{aligned} &(\partial^2 S)_{\mu\nu\lambda\rho}=0, \quad 2(\partial^2 S)_{B\nu\lambda\rho}=j_{\lambda\rho\nu}, \\ &2(\partial^2 S)_{B\nu B\rho}=j_{B\nu\rho}+j_{B\rho\nu}, \end{aligned} \tag{4.20}$$

and additional equations for $\bar{f}_{\mu\nu}$ and $\bar{a}_\mu$:

$$\begin{aligned} &j_{\mu\nu B}=(\partial^2-2/u^2)(\partial^2+2/u^2)\bar{f}_{\mu\nu}+(8/u^2)(d\bar{a})_{\mu\nu}, \\ &j_{\mu BB}=(\partial^2-6/u^2)(\partial^2+2/u^2)\bar{a}_\mu. \end{aligned} \tag{4.21}$$

Free conformal gravity has a gauge freedom

$$\psi_{\alpha\beta\gamma}\to\psi_{\alpha\beta\gamma}+2\,\mathrm{grad}_\gamma\, \Lambda_{\alpha\beta}-\mathrm{grad}_\alpha\, \Lambda_{\beta\gamma}+\mathrm{grad}_\beta\, \Lambda_{\alpha\gamma}, \tag{4.22}$$

which, in de Sitter space, gives a gauge freedom of the $\bar{\gamma}$ and $\bar{h}$ fields and

$$\begin{aligned} &\bar{f}_{\mu\nu}\to\bar{f}_{\mu\nu}+\tfrac{1}{4}\partial^2\Lambda_{\mu\nu}-3u_\mu(u^2)^{-1}\Lambda_{B\nu}+3u_\nu(u^2)^{-1}\Lambda_{B\mu}, \\ &\bar{a}_\mu\to\bar{a}_\mu-\tfrac{3}{8}(\partial^2+(4/u^2))\Lambda_{B\mu}. \end{aligned} \tag{4.23}$$

So it is possible to impose a de Sitter (not conformal) invariant gauge fixing $\bar{f}_{\mu\nu}=0$, $\bar{a}_\mu=0$. Then the equations of free conformal gravity with "Lorentz condition" (4.4), (4.5), and (4.17) give, in de Sitter space, (4.7)–(4.10) and

$$\partial^2\partial^2\bar{\gamma}_{\mu\nu\lambda}+4\,\partial^2\partial_\nu \bar{h}_{\mu\lambda}-4\,\partial^2\partial_\mu \bar{h}_{\nu\lambda}=0. \tag{4.24}$$

The solution space of these equations carries the physical, the ghost, and part of the gauge modes of conformal gravity. Next we want to analyze which fields in de Sitter space describe the physical and the ghost modes.

V. CONFORMAL SPIN-2 FIELDS DECOMPOSED INTO DE SITTER SPIN-2 FIELDS

The Gupta–Bleuler triplets of conformal gravity give

$$\left.\begin{matrix} D(0,2)\oplus D(4,1) \\ \oplus\ \text{more} \end{matrix}\right\}\to\left\{\begin{matrix} \mathrm{D}(3,2) \\ \oplus\,\mathrm{D}(1,2)\oplus\mathrm{D}(2,2)\oplus\mathrm{D}(3,2) \\ \oplus\ \text{finite} \end{matrix}\right\}\to\left\{\begin{matrix} D(0,2)\oplus D(4,1) \\ \oplus\ \text{more} \end{matrix}\right., \tag{5.1}$$

when decomposed into de Sitter representations. The physical modes remain irreducible, while the ghost modes not only contain nonunitary representations $D(1,2)\oplus D(2,2)$, but also the unitary massless $D(3,2)$; "more" in the gauge and scalar sectors means representations that are not Weyl-equivalent to $D(3,2)$ and therefore cannot appear in the de Sitter Gupta–Bleuler triplets. This structure (5.1) appears twice, one time from each helicity.

We found fields that describe the ghost modes. The nonunitary ghosts are carried by

$$H_{\mu\nu}=-(u^2\partial^2+2)\bar{h}_{\mu\nu}, \tag{5.2}$$

$$\Gamma_{\mu\nu\lambda} = \partial^2\gamma_{\mu\nu\lambda} - 4u_\mu (u^2)^{-1}\bar{h}_{\nu\lambda} + 4u_\nu (u^2)^{-1}\bar{h}_{\mu\lambda'}, \tag{5.3}$$

the unitary ghost by

$$C_{\mu\nu\lambda\rho} = (2,2)\,[\partial_\rho \bar{\gamma}_{\mu\nu\lambda} + (\delta_{\mu\rho} - u_\mu u_\rho (u^2)^{-1})\bar{h}_{\nu\lambda} - (\delta_{\nu\rho} - u_\nu u_\rho (u^2)^{-1})\bar{h}_{\mu\lambda}\,]. \tag{5.4}$$

The physical modes satisfy $C = 0$, which implies $\Gamma = 0$ and $H = 0$.

To show all this we employ Eq. (4.24), which gives $(-u^2\partial^2 + \hat{N}(\hat{N}+3))H = 0$. In addition H is traceless, divergenceless, and transverse due to Eqs. (4.7)–(4.10). So the second-order Casimir operator has eigenvalue $QH = 4H$; H can only describe nonunitary ghosts. If we put $H = 0$, the same is true for Γ.

The remaining solutions of Eq. (4.24) have to satisfy $H = 0$ and $\Gamma = 0$, i.e.,

$$\partial^2 h_{\mu\nu} = 0, \quad \partial^2\gamma_{\mu\nu\lambda} = 0, \tag{5.5}$$

where

$$h_{\mu\nu} \equiv (u^2)^{1/2}\bar{h}_{\mu\nu}, \quad \gamma_{\mu\nu\lambda} \equiv \bar{\gamma}_{\mu\nu\lambda} - u^2\partial_\mu \bar{h}_{\nu\lambda} + u^2\partial_\nu \bar{h}_{\mu\lambda}. \tag{5.6}$$

The fields h and γ have degree of homogeneity 0 and satisfy the subsidiary conditions (3.8), (3.9), (3.11), and (3.12); we decomposed the tensor $\bar{\gamma}$ into a divergenceless tensor γ and the divergence $\bar{h}$. Now γ and h have $Q = 6$ and therefore can only carry the $D(3,2)$ representations and pure gauge.

The field equation $C = 0$ projects on physical and gauge modes only. So C carries the $D(3,2)$ ghost. It has degree -1 and fulfills the equation $\partial^2 C = 0$ and the subsidiary conditions (3.15), and therefore also $QC = 6C$. It can be brought into the form

$$C = d\gamma - 2(u^2)^{1/2}dh. \tag{5.7}$$

To simplify comparison with Eq. (3.14) we decompose both fields h and γ into parts which remain finite at $u^2 = 0$:

$$h = h^{(0)} + (u^2)^{3/2}h^{(-3)}, \tag{5.8}$$

$$\gamma = \gamma^{(0)} + (u^2)^{3/2}\gamma^{(-3)}. \tag{5.9}$$

Then the equation $C = 0$ for the physical and gauge modes becomes

$$d\gamma^{(0)} = 2dh^{(-3)}, \quad d\gamma^{(-3)} = 2dh^{(0)}. \tag{5.10}$$

So the physical conformal gravitons satisfy equations of the form (3.14).

Comparison of Eqs. (4.14)–(4.16) with Eqs. (5.2)–(5.4) shows that our conformal coupling to an external current S is simply

$$S_{\mu\nu\lambda\rho} = C_{\mu\nu\lambda\rho}, \quad S_{\mu\nu B\lambda} = \tfrac{1}{4}\Gamma_{\mu\nu\lambda}, \quad S_{B\mu B\nu} = -(4u^2)^{-1}H_{\mu\nu}. \tag{5.11}$$

All the ghost modes C, Γ, and H vanish in empty space $S = 0$.

ACKNOWLEDGMENTS

We thank Moshe Flato for many stimulating discussions.

W. H. thanks Professor Doebner for encouragement and the DAAD (German Academic Exchange Service) for support during the early stages of this work.

[1]B. Binegar, C. Fronsdal, and W. Heidenreich, "*Linear conformal quantum gravity*," Phys. Rev. D **10**, 2249 (1983).

[2]J. Fang and C. Fronsdal, "*Elementary particles in curved space V: Massive and massless spin-2 fields*," Lett. Math. Phys. **2**, 391 (1978); L. F. Abbot and S. Deser, "*Stability of gravity with a cosmological constant*," Nucl. Phys. B **195**, 76 (1982).

[3]B. Binegar, C. Fronsdal, and W. Heidenreich, "*Conformal QED*," J. Math. Phys. **24**, 2828 (1983).

[4]L. Castell, "*Goldstone Particles in de Sitter space*," Nuovo Cimento A **61**, 585 (1969).

[5]W. Heidenreich, "*Quantum theory of spin-½ fields with gauge freedom*," Nuovo Cimento A **80**, 220 (1984).

[6]B. Binegar, C. Fronsdal, and W. Heidenreich, "*de Sitter QED*," Ann. Phys. (NY) **149**, 254 (1983).

On new substitutions connecting different Feynman diagrams

Till B. Anders and Wolfgang Jachmann
Sektion Physik, Universität München, Munich, West Germany

(Received 2 October 1985; accepted for publication 3 September 1986)

Feynman diagrams for electroweak reactions representing partial helicity amplitudes are considered. Substitutions connect all diagrams that differ only in the direction of fermion currents passing through the reaction and in the property of whether a particle is incoming or outgoing. The substitutions act on suitably introduced signatures of the particles and of the special currents. A special substitution that reverses the direction of a current is elaborated. It is used to construct a reversal of the sequence of the interactions in a current. The crossing symmetries, which are derived from the "substitution rule," are different from these. The application of all these substitutions helps to reduce considerably the number of Feynman diagrams that have to be calculated explicitly. The method is illustrated by a treatment of the reactions $ee \to ee\gamma\gamma$ and $e\gamma \to eZ_0$. Thereby calculational aids are given that refine the methods of an earlier paper for the evaluation of helicity amplitudes and cross sections for polarized particles in the standard model, including massive spin-1 bosons. A replacement of Dirac factors like $\not p = p_\mu \gamma^\mu$ by dyadic products of four-component helicity spinors helps to separate complicated fermion currents into sums of simple spinor scalars for which the optimal forms are given.

I. INTRODUCTION

The reaction $ee \to ee\gamma Z_0$ ($ee \to ee\gamma\gamma$) has 40 partial amplitudes or Feynman diagrams in the lowest order of the perturbation expansion. Most of them are connected by the exchange of the final electrons (or photons). Ten (four) pairs of the remaining 20 (ten) diagrams are connected by a crossing symmetry of the initial and final electron states. The remaining ten (six) diagrams can be reduced to four (three) diagrams which generate all others if a substitution can be constructed that reverses the order of the interactions in a current.

It is shown that such a reversal of a current can be constructed if the original Feynman diagrams are equipped with certain dummy sign factors or signatures on which the reversal operation acts.

We show that this substitution goes beyond the frame of the discrete spinor transformations. It also has different roots than the operations of crossing symmetry which are based on the "substitution rule."[1]

If the original Feynman diagram represents a partial helicity amplitude, the crossing exchanges and the current reversals are formulated in such a way that they can be applied also if the amplitude is given in an explicit form in the sense of an earlier paper,[2] quoted here as I.

We show in Sec. II that the polarized photons are treated very similarly to the polarized electrons or spin-$\frac{1}{2}$ particles as explicitly shown in paper I. The massive spin-1 particles are similarly treated but need a 3×3 matrix for spin-1 density.

Appendix A contains useful proposals showing how to shorten the evaluation of helicity amplitudes by the use of dyadic products of four-component helicity spinors.

The symmetry relations and the explicit forms of the helicity amplitudes for the reactions $e\gamma \to eZ_0$ and $ee \to ee\gamma\gamma$—including the Z_0 exchange—are shown in Secs. VI and VII and Appendix D. This illustrates the general concepts of this paper.

II. CROSS SECTIONS IN TERMS OF HELICITY AMPLITUDES

The amplitude of a reaction involving spin-$\frac{1}{2}$ and spin-1 particles, for instance, of the inelastic scattering $ee \to ee\gamma Z_0$,

$$\mathfrak{M} = (\bar u^{Q_1}(p_1,\hat s_1)\bar u^{Q_3}(p_3,\hat s_3)O_{\mu\nu}u^{Q_2}(p_2,\hat s_2)u^{Q_4}(p_4,\hat s_4)) \times \epsilon^{*\mu}(k_5,\hat s_5)\epsilon^{*\nu}(p_6,t_6), \tag{2.1}$$

contains normalized spinors[2] u^+ for fermions and u^- for antifermions

$$\bar u^Q u^Q = Q, \tag{2.2}$$

describing the states of the incoming and outgoing spin-$\frac{1}{2}$ particles.

The general four-component spinors may be expanded in terms of helicity spinors[2]

$$\begin{aligned} u^Q(p,\hat s) &= \sum_{N=\pm1} \alpha_N^Q(\hat p,\hat s)u_N^Q(p),\\ \bar u^Q(p,\hat s) &= \sum_{N=\pm1} \alpha_N^{*Q}(\hat p,\hat s)\bar u_N^Q(p). \end{aligned} \tag{2.3}$$

Here $\hat s$ means the spin vector at rest and p the momentum. The helicity spinors

$$u_N^Q(p) = \frac{1}{\sqrt2}\begin{pmatrix} \eta_M & |M,\hat p\rangle \\ Q\eta_{-M} & |M,\hat p\rangle \end{pmatrix}, \quad M\equiv QN, \tag{2.4}$$

$$\bar u_N^Q(p) = \frac{1}{\sqrt2}(Q\eta_{-M}\langle M,\hat p|,\eta_M\langle M,\hat p|), \tag{2.5}$$

consist of energy factors

$$\eta_\pm \equiv \eta_\pm(p) = [(E\pm p)/m]^{1/2}, \tag{2.6}$$

and two-component spinors

 0022-2488/87/010221-13$02.50

$$|+,\hat p\rangle = \begin{pmatrix} C \\ S \end{pmatrix}, \quad |-,\hat p\rangle = \begin{pmatrix} -S^* \\ C \end{pmatrix}, \tag{2.7}$$

with

$$C = \cos(\theta/2), \quad S = \sin(\theta/2)e^{i\phi}. \tag{2.8}$$

They are connected with the direction of the momentum

$$\hat p = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \tag{2.9}$$

by the relations

$$\langle N,\hat p|\sigma^\mu|N,\hat p\rangle = (1, N\hat p). \tag{2.10}$$

The states of massless left-handed neutrinos show $Q = +1$, $M = -1$, while right-handed neutrinos show $Q = -1, M = +1$. The energy E and the absolute value p of the momentum are equal. Only two components of the spinor (2.4) are not vanishing. In this case the quantity m in the formula (2.6) is treated as a constant and arbitrary unit of the energy.

The coefficients $\alpha_N^Q(\hat p,\hat s)$ of the expansion (2.3) are given by

$$\alpha_N^Q(\hat p,\hat s) = \langle QN,\hat p|Q,\hat s\rangle = N\alpha_N^{*-Q}(\hat p,\hat s). \tag{2.11}$$

The amplitude (2.1) also contains a normalized polarization vector $\epsilon^{*\mu}(\hat k,\hat s)$, which describes the state of an outgoing photon with momentum k and a polarization represented by the Stokes vector $\hat s$.[1] An incoming photon would require the complex conjugate vector $\epsilon^\mu(\hat k,\hat s)$. The general polarization vectors can be expanded in terms of two helicity vectors[3]

$$\epsilon^\mu(\hat k,\hat s) = \sum_{N=\pm 1} \alpha_N^+(\hat k,\hat s)\epsilon_N^\mu(\hat k). \tag{2.12}$$

In this expansion of a photon state, the same coefficients as shown for a fermion state (with $Q = +1$) can be used. This shows a formal analogy between the Stokes vector of a photon and the spin vector of a fermion.[4] However, the Stokes "vector" is not a vector—as is already known[5] and demonstrated later. For instance, if the space rotates around the axis of the photon propagation with angle α, the Stokes "vector" rotates with 2α. More specifically, if the real polarization vector for linearly polarized photons turns from the positive x axis to the positive y axis, the Stokes "vector" rotates from the negative to the positive x axis.

The polarization vectors for incoming photons in helicity states are given in Coulomb gauge by

$$\epsilon_+^\mu(\hat k) = (1/\sqrt2)(0, -\vec g(\hat k)), \quad \epsilon_-^\mu(\hat k) = (1/\sqrt2)(0, \vec g^*(\hat k)). \tag{2.13}$$

This can also be written as

$$\begin{aligned} \epsilon_+^\mu(\hat k) &= -(1/\sqrt2)\langle -,\hat k|\sigma^\mu|+,\hat k\rangle, \\ \epsilon_-^\mu(\hat k) &= (1/\sqrt2)\langle +,\hat k|\sigma^\mu|-,\hat k\rangle, \end{aligned} \tag{2.14}$$

because $\vec g(\hat k)$ is the complex vector

$$\vec g(\hat k) = \langle -,\hat k|\vec\sigma|+,\hat k\rangle, \tag{2.15}$$

which has the following properties:

$$\vec g\vec g = 0, \quad \vec g\vec g^* = 2, \quad \vec g\vec k = 0. \tag{2.16}$$

The vector $\vec g$ has been given by components in Eq. (A27) of paper I.

The amplitude (2.1) has further a normalized polarization vector $\epsilon^{*\nu}(p,t)$ describing the state of an outgoing massive spin-1 particle (a Z_0 in our case), with momentum p and five polarization parameters t.

Generally the state $\xi^{(1)}(p,\theta\phi)$ of a massive spin-1 particle moving in the direction $\theta\phi$ can be expanded in terms of three helicity states,

$$\begin{aligned} \xi^{(1)}(p,\theta\phi) &= \beta_+\xi_+^{(1)}(p,\theta\phi) + \beta_-\xi_-^{(1)}(p,\theta\phi) \\ &\quad + \beta_0\xi_0^{(1)}(p,\theta\phi). \end{aligned} \tag{2.17}$$

In the rest frame the helicity states $\xi_N^{(1)}$ are represented by columns of three complex numbers, the spin-1 spinors, which are given in the standard representation (Appendix C) by

$$\begin{aligned} &[\xi_+^{(1)}, \xi_-^{(1)}, \xi_0^{(1)}](0,\theta\phi) \\ &= \begin{pmatrix} C^2 & S^{*2} & -\sqrt2 CS^* \\ S^2 & C^2 & \sqrt2 CS \\ \sqrt2 CS & -\sqrt2 CS^* & C^2 - |S|^2 \end{pmatrix}, \end{aligned} \tag{2.18}$$

or in the vector representation (Appendix C) by

$$\begin{aligned} &[\vec\xi_+^{(1)}, \vec\xi_-^{(1)}, \vec\xi_0^{(1)}](0,\theta\phi) \\ &= [-(1/\sqrt2)\vec g(\theta\phi), (1/\sqrt2)\vec g^*(\theta\phi), \hat p(\theta\phi)]. \end{aligned} \tag{2.19}$$

Both representations are connected by a unitary transformation

$$\begin{aligned} \vec\xi_N^{(1)} &= T\xi_N^{(1)}, \\ T &= [\vec\xi_+^{(1)}, \vec\xi_-^{(1)}, \vec\xi_0^{(1)}](0,00) \\ &= \frac{1}{\sqrt2}\begin{pmatrix} -1 & 1 & \cdot \\ -i & -i & \cdot \\ \cdot & \cdot & \sqrt2 \end{pmatrix}. \end{aligned} \tag{2.20}$$

The polarization vectors of an incoming spin-1 particle with mass m are constructed with the components of the vector representations $\vec\xi^{(1)}$ by a Lorentz boost $L_{\theta\phi}$ in the direction $\theta\phi$,

$$\epsilon_N^\mu(p,\theta\phi) = L_{\theta\phi}(0, \vec\xi_N^{(1)}). \tag{2.21}$$

Therefore the vectors $\epsilon_\pm^\mu$ are identical with (2.13) but the third vector is given by

$$\epsilon_0^\mu(p,\theta\phi) = [p/m, (E/m)\vec\xi_0^{(1)}(0,\theta\phi)], \quad E^2 - p^2 = m^2. \tag{2.22}$$

The expansions in terms of helicity states (2.3), (2.12), and (2.17) convert the amplitude (2.1) to a sum of helicity amplitudes $\mathrm{HA}(N_i)$:

$$\mathfrak{M} = \sum_{N_1}\cdots\sum_{N_6} \alpha_{N_1}^{*Q_1}\alpha_{N_2}^{Q_2}\alpha_{N_3}^{*Q_3}\alpha_{N_4}^{Q_4}\alpha_{N_5}^{*+}\beta_{N_6}^* \,\mathrm{HA}(N_1,\ldots,N_6). \tag{2.23}$$

The cross section of a reaction is proportional to the transition probability W. For reactions with incompletely polarized particles, the quantity W is obtained from $|\mathfrak{M}|^2$ if the products $\alpha_N^{*Q}\alpha_{N'}^Q$ and $\beta_N^*\beta_{N'}$ are substituted by the components of the spin density matrix.[2,3,5] For spin-$\frac12$ particles and photons we have the replacements

$$\alpha_N^Q\alpha_{N'}^{*Q} \to R_{NN'}^Q, \tag{2.24}$$

with

$$R^{Q}_{NN'} = \frac{1}{2}\begin{pmatrix} 1+s_{\|} & Qs_{\perp}e^{-iQ\psi} \\ Qs_{\perp}e^{iQ\psi} & 1-s_{\|} \end{pmatrix}. \tag{2.25}$$

Here $s_{\|}$ and $s_{\perp}$ represent the components of the spin vector of a spin-$\frac{1}{2}$ particle or of the Stokes "vector" of a photon parallel and perpendicular to the momentum p,

$$\vec{s} = s_{\|}\hat{p} + \vec{s}_{\perp}, \quad |\vec{s}|^2 \leqslant 1. \tag{2.26}$$

The phase factor $\exp(i\psi)$ is defined by the relation

$$e^{i\psi} = \hat{s}\hat{g}(p)[1-(\hat{p}\hat{s})^2]^{-1/2}. \tag{2.27}$$

The angle ψ is the azimuth of $\hat{s}$ with regard to the helicity frame of the particle as introduced in paper I.

The dyadic matrix for the products of the coefficients β_N for massive spin-1 states,

$$B = \begin{pmatrix} \beta_+\beta_+^* & \beta_+\beta_-^* & \beta_+\beta_0^* \\ \beta_-\beta_+^* & \beta_-\beta_-^* & \beta_-\beta_0^* \\ \beta_0\beta_+^* & \beta_0\beta_-^* & \beta_0\beta_0^* \end{pmatrix}, \tag{2.28}$$

has to be replaced by the density matrix ρ for massive spin-1 beams:

$$B \to \rho = \rho^{\dagger}, \quad \mathrm{tr}[\rho] = 1. \tag{2.29}$$

The spin-density matrix is usually parametrized with Cartesian spin-1 operators given in the standard representation and relative to the helicity frame,[6] by

$$\rho = \begin{pmatrix} \frac{1}{3} - \frac{1}{6}(p_{xx}+p_{yy}) + \frac{1}{2}p_z & \frac{1}{6}(p_{xx}-p_{yy}) - \frac{i}{3}p_{xy} & \frac{1}{2\sqrt{2}}(p_x - ip_y) + \frac{1}{3\sqrt{2}}(p_{xz} - ip_{yz}) \\ \frac{1}{6}(p_{xx}-p_{yy}) + \frac{i}{3}p_{xy} & \frac{1}{3} - \frac{1}{6}(p_{xx}+p_{yy}) - \frac{1}{2}p_z & \frac{1}{2\sqrt{2}}(p_x + ip_y) - \frac{1}{3\sqrt{2}}(p_{xz} + ip_{yz}) \\ \frac{1}{2\sqrt{2}}(p_x + ip_y) + \frac{1}{3\sqrt{2}}(p_{xz} + ip_{yz}) & \frac{1}{2\sqrt{2}}(p_x - ip_y) - \frac{1}{3\sqrt{2}}(p_{xz} - ip_{yz}) & \frac{1}{3} + \frac{1}{3}(p_{xx}+p_{yy}) \end{pmatrix}, \tag{2.30}$$

$$p_{xx} + p_{yy} + p_{zz} = 0 \tag{2.31}$$

or with spherical spin-1 operators given in the standard representation and for the helicity frame of the particle (the z axis is the direction of motion),[7,8] by

$$\rho = \frac{1}{3}\begin{pmatrix} 1 + (1/\sqrt{2})t_0^{(2)} + \sqrt{\tfrac{3}{2}}t_0^{(1)} & \sqrt{3}t_{-2}^{(2)} & -\sqrt{\tfrac{3}{2}}(t_{-1}^{(1)} + t_{-1}^{(2)}) \\ \sqrt{3}t_{+2}^{(2)} & 1 + (1/\sqrt{2})t_0^{(2)} - \sqrt{\tfrac{3}{2}}t_0^{(1)} & \sqrt{\tfrac{3}{2}}(t_{+1}^{(1)} - t_{+1}^{(2)}) \\ \sqrt{\tfrac{3}{2}}(t_{+1}^{(1)} + t_{+1}^{(2)}) & -\sqrt{\tfrac{3}{2}}(t_{-1}^{(1)} - t_{-1}^{(2)}) & 1 - \sqrt{2}t_0^{(2)} \end{pmatrix}, \tag{2.32}$$

$$t_{-m}^{(l)} = (-1)^m t_m^{*(l)}. \tag{2.33}$$

The degree of polarization is defined by

$$D = \left[\frac{3}{4}(p_x^2 + p_y^2 + p_z^2) + \frac{1}{6}\sum_{ik} p_{ik}^2\right]^{1/2} = \left[\frac{1}{2}\sum_{l=1,2}\sum_{k=-l}^{+l}|t_k^{(l)}|^2\right]^{1/2}. \tag{2.34}$$

A comparison of the general matrix ρ, (2.30) for the spin-1 density and the special matrix R^{+}, and (2.25) for the spin density of a photon moving in the z direction yields

$$s_x = \tfrac{1}{3}(p_{xx} - p_{yy}), \quad s_y = \tfrac{2}{3}p_{xy}, \quad s_z = p_z, \tag{2.35}$$

and

$$p_{xx} + p_{yy} = -p_{zz} = -1, \quad p_{xz} = p_{yz} = p_x = p_y = 0. \tag{2.36}$$

The result (2.35) shows that the transversal components of the Stokes "vector" are really tensor components that behave under rotations as described before.

Now the transition probability W is given as the sum

$$W = \sum_{N_i}\cdots\sum\sum_{N_i'}\cdots\sum R^{Q_1}_{N_1N_1'}R^{Q_2}_{N_2'N_2}R^{Q_3}_{N_3N_3'}R^{Q_4}_{N_4'N_4} \times R^{+}_{N_5N_5'}\rho_{N_6N_6'}\mathrm{HA}(N_1'\cdots N_6')\mathrm{HA}^*(N_1\cdots N_6). \tag{2.37}$$

This expression can easily be realized analytically or by a computer if the helicity amplitudes are known.

III. HELICITY CURRENTS

The amplitude of a reaction is represented by a sum of Feynman diagrams. Each Feynman diagram is composed of currents passing through the reaction. The helicity currents including products of even or odd numbers of Dirac γ-matrices can easily be elaborated according to the methods of paper I. We obtain, with $\kappa = \pm 1$,

$$\begin{aligned} \bar{u}^{Q_e}_{N_e}(p_e)\prod_{i=1}^{2n}\gamma^{\nu_i}\gamma_5^{(1-\kappa)/2}u^{Q_s}_{N_s}(p_s) &= (Q_e/2)\eta_{-M_e}\eta_{M_s}\langle M_e,\hat{p}_e|\sigma_{\nu_1}\sigma^{\nu_2}\cdots\sigma^{\nu_{2n}}|M_s,\hat{p}_s\rangle \\ &\quad + \kappa(Q_s/2)\eta_{M_e}\eta_{-M_s}\langle M_e,\hat{p}_e|\sigma^{\nu_1}\sigma_{\nu_2}\cdots\sigma_{\nu_{2n}}|M_s,\hat{p}_s\rangle, \end{aligned} \tag{3.1}$$

$$\begin{aligned} \bar{u}^{Q_e}_{N_e}(p_e)\prod_{i=1}^{2n+1}\gamma^{\nu_i}\gamma_5^{(1-\kappa)/2}u^{Q_s}_{N_s}(p_s) &= \tfrac{1}{2}\eta_{M_e}\eta_{M_s}\langle M_e,\hat{p}_e|\sigma^{\nu_1}\sigma_{\nu_2}\cdots\sigma^{\nu_{2n+1}}|M_s,\hat{p}_s\rangle \\ &\quad + \kappa(Q_eQ_s/2)\eta_{-M_e}\eta_{-M_s} \\ &\quad \times\langle M_e,\hat{p}_e|\sigma_{\nu_1}\sigma^{\nu_2}\cdots\sigma_{\nu_{2n+1}}|M_s,\hat{p}_s\rangle. \end{aligned} \tag{3.2}$$

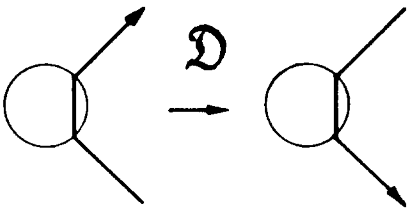

FIG. 1. Visualization of the reversal 𝔇 of a current.

The current operators of Feynman diagrams consist of vertex operators separated by propagators. The vertices have the most general form (a factor i omitted)

$$V = \rho_u \sum_\mu v_\mu \gamma^\mu + v_4 1 + \rho_5 v_5 \gamma_5 + \rho_a \rho_5 \sum_{\mu<\nu} v'_{\mu\nu} \gamma^\mu \gamma^\nu \gamma_5 + \rho_a \sum_{\mu<\nu} v_{\mu\nu} \gamma^\mu \gamma^\nu + \rho_u \rho_a \rho_5 \sum_\mu v_{5\mu} \gamma_5 \gamma^\mu. \quad (3.3)$$

The fermion propagators have the form (factor i omitted)

$$S(\tau p) = (\tau \underline{p} - m + i0)^{-1}. \quad (3.4)$$

Here we connected the vertex and propagator with dummy sign factors ρ_a, ρ_u, ρ_5, and τ, which form sets belonging to specific currents of a partial amplitude. They are used as indices for the action of certain substitutions as explained later. The idea is that these signs are intended to be finally $+1$, but a substitution can require a change of some of these signs. If a reversal of the sequence of the factors γ^μ in the vertex V is required, the sign ρ_α is changed. A transformation $\gamma^\mu \to -\gamma^\mu$ is equivalent to $\rho_u \to -\rho_u$. A transformation $\gamma_5 \to -\gamma_5$ is described by $\rho_5 \to -\rho_5$.

We introduce for each (real or virtual) particle a running signature[9] (in/out character)

$$\varepsilon_i = \begin{cases} -1, & \text{for an incoming particle } i, \\ +1, & \text{for an outgoing particle } i. \end{cases} \quad (3.5)$$

The particle momenta with regard to a special current

$$P_i = \varepsilon_i p_i, \quad K_l = \varepsilon_l k_l \quad (3.6)$$

constitute the energy momentum balance

$$P_s + P_e + \sum_{l=1}^n K_l = 0. \quad (3.7)$$

We observe that the product of charge and running signature is -1 for a particle state "s" starting the current, and $+1$ for a particle state "e" ending the current

$$D_s \equiv \varepsilon_s Q_s = -1, \quad D_e \equiv \varepsilon_e Q_e = +1. \quad (3.8)$$

The "starting" states and the "ending" states of currents should be distinguished from the "initial" and "final" states, which indicate the order of time.

A current describing n interaction points of a fermion has the form

$$J\left[{}^{Q_e}_{N_e}P_e, 1 \cdots n, {}^{Q_s}_{N_s}P_s\right] = \bar{u}^{Q_e}_{N_e}(p_e)\Gamma_{P_e,1\cdots n,P_s} u^{Q_s}_{N_s}(p_s), \quad (3.9)$$

with the current operator

$$\begin{aligned} \Gamma_{P_e,1\cdots n,P_s} &= V_1 S(\tau P_e + \tau K_1) V_2 \cdots S(\tau P_e + \tau K_1 + \cdots + \tau K_{n-1}) V_n \\ &= V_1 S(-\tau P_s - \tau K_n - \cdots - \tau K_2) V_2 \cdots S(-\tau P_s - \tau K_n) V_n. \end{aligned} \quad (3.10)$$

The current similar to the original one but with reversed direction is

$$J\left[{}^{-Q_s}_{N_s}P_s, n \cdots 1, {}^{-Q_e}_{N_e}P_e\right] = \bar{u}^{-Q_s}_{N_s}(p_s)\Gamma_{P_s,n\cdots 1,P_e} u^{-Q_e}_{N_e}(p_e), \quad (3.11)$$

with the current operator

$$\begin{aligned} \Gamma_{P_s,n\cdots 1,P_e} &= V_n S(\tau P_s + \tau K_n) \cdots V_2 S(\tau P_s + \tau K_n + \cdots + \tau K_2) V_1 \\ &= V_n S(-\tau P_e - \tau K_1 - \cdots - \tau K_{n-1}) \cdots V_2 S(-\tau P_e - \tau K_1) V_1. \end{aligned} \quad (3.12)$$

The reversal of the direction of a current may be visualized by reversing the arrow of a particle line in a Feynman diagram. This implies that the running signatures ε_s and ε_e are conserved, but the charges Q_s and Q_e are changed. As a consequence, the state "s" now ends the current and the state "e" now starts the reversed current. (See Fig. 1.)

The comparison of the original current operator (3.10) and the operator for the reversed current (3.12) yields

$$\Gamma_{P_s,n\cdots 1,P_e}(\tau,\rho_a,\rho_u,\rho_5) = \{\Gamma_{P_e,1\cdots n,P_s}(-\tau,\rho_a,\rho_u,\rho_5)\}_{V,S\text{-rev}}. \quad (3.13)$$

The current operator with reversed direction is the original one after reversing the sequence of all vertices V_i and propagators $S(q)$ and a change of the signs for all momenta in the arguments of the propagators, which is described by the change of the dummy sign factor τ. The reversal of the factors V_i and S can be expressed by reversing the sequence of all Dirac γ-matrices in Γ if first the sequence of the γ's in V_i is reversed. The reversal of the γ's in V_i is equivalent to the change of the dummy sign factor ρ_a in the expression (3.3)

$$\Gamma_{P_s,n\cdots 1,P_e}(\tau,\rho_a,\rho_u,\rho_5) = \{\Gamma_{P_e,1\cdots n,P_s}(-\tau,-\rho_a,\rho_u,\rho_5)\}_{\gamma\text{-rev}}. \quad (3.14)$$

At this place we should note that the scalar and pseudoscalar couplings in the general vertex (3.3) may also contain derivative couplings of the form

$$\rho_a v'^\mu R^{(i)}_\mu, \quad \rho_a \rho_5 v''^\mu R^{(i)}_\mu \gamma_5. \quad (3.15)$$

Here $R^{(i)}$ means the sum of momenta to the right and to the left of the vertex i

$$\begin{aligned} R^{(i)} &= 2\left(P_e + \sum_{r=1}^{i-1} K_r\right) + K_i \\ &= -2\left(P_s + \sum_{r=i+1}^{n} K_r\right) - K_i. \end{aligned} \quad (3.16)$$

The comparison of (3.12) and (3.10) shows that the sign of $R^{(i)}$ must be changed in a current of reversed order. And this is the reason why the "Gordon vertex" (3.15) is multiplied with the sign factor ρ_a. In the succeeding sections we construct representations of an operator $\mathfrak{D}$ that reverses the direction of a current, and all these representations as given in Eqs. (5.3)–(5.6) changes ρ_a into $-\rho_a$.

IV. SYMMETRIES OF HELICITY CURRENTS

We observe that the reversal of the sequence of γ-matrices in a current operator Γ can be realized with the help of some symmetry relations for helicity currents.

We use the relations for "Pauli" currents I(4.1) and I(4.2)[10]

$$\langle M_e,\hat{p}_e|\sigma^\mu\sigma_\nu\cdots\sigma^\tau|M_s,\hat{p}_s\rangle^* = \langle M_s,\hat{p}_s|\sigma^\tau\cdots\sigma_\nu\sigma^\mu|M_e,\hat{p}_e\rangle, \tag{4.1}$$

$$\langle M_e,\hat{p}_e|\sigma^\mu\sigma_\nu\cdots\sigma^\tau|M_s,\hat{p}_s\rangle = M_eM_s\langle -M_s,\hat{p}_s|\sigma_\tau\cdots\sigma^\nu\sigma_\mu|-M_e,\hat{p}_e\rangle. \tag{4.2}$$

The following symmetries for helicity currents including products of Dirac γ-matrices can be derived from the formulas (3.1) and (3.2) with (4.1) and (4.2):

$$\bar{u}^{Q_e}_{N_e}(p_e)\left\{\prod^n\gamma^{\nu_i}\right\}_{\gamma\text{-rev}}u^{Q_s}_{N_s}(p_s)$$
$$= \left(\bar{u}^{Q_s}_{N_s}(p_s)\left\{\prod^n\gamma^{\nu_i}\right\}u^{Q_e}_{N_e}(p_e)\right)^* \tag{4.3}$$
$$= (-1)^{n+1}N_eN_s\bar{u}_{N_s}^{-Q_s}(p_s)\left\{\prod^n\gamma^{\nu_i}\right\}u_{N_e}^{-Q_e}(p_e) \tag{4.4}$$
$$= N_eN_s\bar{u}^{Q_s}_{-N_s}(p_s)\left\{\prod^n\gamma^{\nu_i}\right\}u^{Q_e}_{-N_e}(p_e). \tag{4.5}$$

Let Γ be a sum of products of Dirac γ-matrices. We introduce some transformations of Γ:

$$\Gamma^{(5)} = \gamma_5\Gamma\gamma_5, \tag{4.6}$$
$$\Gamma^{(*)} = -\gamma_5\gamma^2\Gamma^*\gamma^2\gamma_5, \tag{4.7}$$
$$\bar{\Gamma} = \gamma^0\Gamma^+\gamma^0. \tag{4.8}$$

The transformation $^{(5)}$ supplies each factor γ^μ in Γ with a negative sign. The transformation $^{(*)}$ modifies all coefficients of the products of γ^μ ($\mu = 0,1,2,3$; not 5) into the complex conjugate values.

The symmetry relations (4.3)–(4.5) can now be generalized to currents including the more general operator Γ,

$$\bar{u}^{Q_e}_{N_e}(p_e)\{\Gamma\}_{\gamma\text{-rev}}u^{Q_s}_{N_s}(p_s)$$
$$= \left[\bar{u}^{Q_s}_{N_s}(p_s)\Gamma^{(*)}u^{Q_e}_{N_e}(p_e)\right]^* \tag{4.9}$$
$$= -N_eN_s\bar{u}_{N_s}^{-Q_s}(p_s)\Gamma^{(5)}u_{N_e}^{-Q_e}(p_e) \tag{4.10}$$
$$= N_eN_s\bar{u}^{Q_s}_{-N_s}(p_s)\Gamma u^{Q_e}_{-N_e}(p_e). \tag{4.11}$$

Equation (4.9) is equivalent to the well-known relation

$$(\bar{u}_a\Gamma u_b)^* = \bar{u}_b\bar{\Gamma}u_a, \tag{4.12}$$

because

$$\{\Gamma\}_{\gamma\text{-rev}} = \bar{\Gamma}^{(*)}. \tag{4.13}$$

Two of the relations (4.9)–(4.11) can be combined to new relations that show no reversal of the sequence of γ-matrices,

$$\bar{u}^{Q_e}_{N_e}(p_e)\Gamma u^{Q_s}_{N_s}(p_s)$$
$$= -N_eN_s\left[\bar{u}_{N_e}^{-Q_e}(p_e)\Gamma^{(*,5)}u_{N_s}^{-Q_s}(p_s)\right]^* \tag{4.14}$$
$$= N_eN_s\left[\bar{u}^{Q_e}_{-N_e}(p_e)\Gamma^{(*)}u^{Q_s}_{-N_s}(p_s)\right]^* \tag{4.15}$$
$$= -\bar{u}^{-Q_e}_{-N_e}(p_e)\Gamma^{(5)}u^{-Q_s}_{-N_s}(p_s). \tag{4.16}$$

These identities[11] can also be derived with the help of the spinor operations[12]

charge conjugation $\underrightarrow{\mathbf{C}} = i\gamma^2|^*/$, $\underleftarrow{\mathbf{C}} = \backslash^*|i\gamma^2$, (4.17)

helicity change $\underrightarrow{\mathbf{H}} = i\gamma_5\gamma^2|^*/$, $\underleftarrow{\mathbf{H}} = \backslash^*|i\gamma^2\gamma_5$, (4.18)

gamma 5 $\gamma_5 = \underrightarrow{\mathbf{C}}\underrightarrow{\mathbf{H}} = \underleftarrow{\mathbf{C}}\underleftarrow{\mathbf{H}}$. (4.19)

The symbols

$$|^*/,\backslash^*| \text{ and } \backslash^*/ = \backslash^*||^*/ \tag{4.20}$$

here mean complex conjugation to the right side, left side, and to both sides, i.e., to the whole expression

$$\begin{aligned} &|^*/^2 = \underrightarrow{\mathbf{C}}^2 = \underrightarrow{\mathbf{H}}^2 = \gamma_5^2 = 1,\\ &\backslash^*|^2 = \underleftarrow{\mathbf{C}}^2 = \underleftarrow{\mathbf{H}}^2 = 1,\\ &\backslash^*/ = \underleftarrow{\mathbf{C}}\underrightarrow{\mathbf{C}} = \underleftarrow{\mathbf{H}}\underrightarrow{\mathbf{H}}. \end{aligned} \tag{4.21}$$

These spinor operations act on helicity spinors and current operators Γ according to[13]

$$\underrightarrow{\mathbf{C}}u^Q_N = Nu_N^{-Q},\quad \bar{u}^Q_N\underleftarrow{\mathbf{C}} = -N\bar{u}_N^{-Q},\quad \underleftarrow{\mathbf{C}}\Gamma\underrightarrow{\mathbf{C}} = \backslash^*/\Gamma^{(*,5)}, \tag{4.22}$$

$$\underrightarrow{\mathbf{H}}u^Q_N = Nu^Q_{-N},\quad \bar{u}^Q_N\underleftarrow{\mathbf{H}} = N\bar{u}^Q_{-N},\quad \underleftarrow{\mathbf{H}}\Gamma\underrightarrow{\mathbf{H}} = -\backslash^*/\Gamma^{(*)}, \tag{4.23}$$

$$\gamma_5u^Q_N = u^{-Q}_{-N},\quad \bar{u}^Q_N\gamma_5 = -\bar{u}^{-Q}_{-N},\quad \gamma_5\Gamma\gamma_5 = \Gamma^{(5)}. \tag{4.24}$$

The identity (4.14) now can be derived by inserting the unit matrices $\underleftarrow{\mathbf{C}}^2$ and $\underrightarrow{\mathbf{C}}^2$, and by the use of (4.20)

$$\bar{u}^{Q_e}_{N_e}\Gamma u^{Q_s}_{N_s} = \bar{u}^{Q_e}_{N_e}\underleftarrow{\mathbf{C}}\underleftarrow{\mathbf{C}}\Gamma\underrightarrow{\mathbf{C}}\underrightarrow{\mathbf{C}}u^{Q_s}_{N_s} = -N_eN_s\backslash^*/\bar{u}_{N_e}^{-Q_e}\Gamma^{(*,5)}u_{N_s}^{-Q_s}. \tag{4.25}$$

In a similar way the proof of the identities (4.15) and (4.16) works with the insertion of $\underleftarrow{\mathbf{H}}^2$, $\underrightarrow{\mathbf{H}}^2$, and γ_5^2.

V. THE REVERSAL OF HELICITY CURRENTS

We introduce a special symbol for the change of signatures. This symbol includes also, if required, an overall complex conjugation as defined in (4.20)

$\backslash\tau/ \equiv \{\tau\to-\tau\}$, $\backslash^*/ \equiv$ overall complex conjugation,[14]

$\backslash\tau/\backslash\lambda/ \equiv \backslash\tau,\lambda/$, $\backslash\tau,\lambda/\backslash^*/ \equiv \backslash\tau,\lambda,^*/$, etc. (5.1)

The operation $\mathfrak{D}$ reverses the direction of a helicity current, $\mathfrak{D}$ transforms the current (3.9) into the reversed current (3.11)

$$\mathfrak{D}J\left[{}^{Q_e}_{N_e}P_e,1\cdots n,{}^{Q_s}_{N_s}P_s\right] \equiv J\left[{}^{-Q_s}_{N_s}P_s,n\cdots 1,{}^{-Q_e}_{N_e}P_e\right]. \tag{5.2}$$

In a Feynman diagram $\mathfrak{D}$ may be visualized simply by reversing the arrow of a current line passing through the reaction. (See Fig. 1.)

The operator $\mathfrak{D}$ can be realized by a change of some signatures. Eventually an additional multiplication with a phase factor is required. We use the symmetry relations

(4.9), (4.10), or (4.11) to realize the reversal of the sequence of γ-matrices in the current operators Γ required by (3.14). The following four realizations of $\mathfrak{D}$ are found:

$$\mathfrak{D}_{\mathrm{C}} = -N_eN_s\backslash\rho_a,\rho_u/, \tag{5.3}$$

$$\mathfrak{D}_{\mathrm{H}} = N_eN_s\backslash\tau,\rho_a,N_e,N_s,Q_e,Q_s/, \tag{5.4}$$

$$\mathfrak{D}_{*} = \backslash\tau,\rho_u,\rho_5,Q_e,Q_s,{*}/ \quad (v \text{ real}), \tag{5.5}$$

$$\mathfrak{D}_{\gamma_5} = -\backslash\rho_a,\rho_u,\rho_5,N_e,N_s,{*}/ \quad (v \text{ real}). \tag{5.6}$$

According to (4.10), the γ-reversal of a current operator between fermion states is equivalent to a charge conjugation of the interchanged "starting" and "ending" states and an additional transformation ${}^{(5)}$ as defined in (4.6). The charge conjugation requires the change of Q_e and Q_s and a multiplication with the phase factor $-N_eN_s$.

The transformation ${}^{(5)}$ requires that the matrix γ_5 is shifted through the current operator Γ with factors defined in (3.3) and (3.4). The result is an additional change of ρ_u and τ. The changes of Q_e, Q_s, and τ due to (4.10) are canceled by the requirements of (3.14) and (5.2). Thus the realization $\mathfrak{D}_{\mathrm{C}}$ (5.3) is obtained.

By similar arguments with the use of (4.11) the realization $\mathfrak{D}_{\mathrm{H}}$ is found.

The realization $\mathfrak{D}_{*}$ uses the symmetry (4.9). The transformation ${}^{(*)}$ is equivalent to the change of ρ_5 if all coefficients v in the vertices (3.3) are real.

The fourth realization $\mathfrak{D}_{\gamma_5}$ is simply the product of the realizations just mentioned,

$$\mathfrak{D}_{\gamma_5} = \mathfrak{D}_{\mathrm{C}}\mathfrak{D}_{\mathrm{H}}\mathbf{D}_{*}. \tag{5.7}$$

We notice that an identity substitution

$$\mathbf{I}\ J\left[{}^{Q_e}_{N_e}P_e,1\cdots n,{}^{Q_s}_{N_s}P_s\right] = J\left[{}^{Q_e}_{N_e}P_e,1\cdots n,{}^{Q_s}_{N_s}P_s\right], \tag{5.8}$$

is obtained by products of two different realizations of $\mathfrak{D}$. The identity operations

$$\mathbf{I}_{\mathrm{C}*} = \mathfrak{D}_{\mathrm{C}}\mathfrak{D}_{*} = -N_eN_s\backslash\tau,\rho_u,\rho_5,Q_e,Q_s,{*}/ \quad (v \text{ real}), \tag{5.9}$$

$$\mathbf{I}_{\mathrm{H}*} = \mathfrak{D}_{\mathrm{H}}\mathfrak{D}_{*} = N_eN_s\backslash\rho_5,N_e,N_s,{*}/ \quad (v \text{ real}), \tag{5.10}$$

$$\mathbf{I}_{\mathrm{CH}} = \mathfrak{D}_{\mathrm{C}}\mathfrak{D}_{\mathrm{H}} = -\backslash\tau,\rho_u,N_e,N_s,Q_e,Q_s/, \tag{5.11}$$

allow to reproduce the identity relations (4.14), (4.15), and (4.16). The transformation ${}^{(*)}$ in (4.14) and (4.15) requires a change of ρ_5 (if the coefficients v of the vertices are real). The transformation ${}^{(5)}$ in (4.14) and (4.16) requires a change of ρ_u and τ.

We ask for spinor transformations that possibly could be equivalent to these operations. In Sec. IV we have derived the different realizations of $\mathbf{I}$ with the help of the spinor transformations like the change of charge and helicity. This is not so for $\mathfrak{D}$. It is evident that a change of ρ_a and ρ_u, whereas τ remains constant—and this is just required by $\mathfrak{D}_{\mathrm{C}}$—cannot be canceled by a matrix shifted through the current operator Γ. Similar remarks hold for the other realizations of $\mathfrak{D}$. We see that the realizations of $\mathfrak{D}$ cannot be formulated without the introduction of special dummy signs like ρ_a, ρ_u, ρ_5, and τ into the propagators and vertices of the currents. Therefore, they exceed the frame of the spinor operations.

For the QED that shows only vector couplings and possibly additional anomalous magnetic moment couplings, the reversal of a current in the realization $\mathfrak{D}_{\mathrm{C}}$ is a pure phase transformation. In this special case $\mathfrak{D}_{\mathrm{C}}$ requires only a multiplication with the sign $-N_eN_s$ and with factors (-1) for each vertex. The fact that in this case $\mathfrak{D}_{\mathrm{C}}$ is different from the identity only by a sign factor may explain why this operation does not seem to have been considered so far explicitly in the literature.

VI. PARTICLE STATE EXCHANGES, CROSSING OPERATIONS, AND REVERSAL OF INTERACTIONS

Consider a reaction with several spin-$\frac{1}{2}$ particles of the same type. The exchange of two particle states i and j

$$\mathfrak{P}_{ij} \equiv \{p_i \leftrightarrow p_j,\ Q_i \leftrightarrow Q_j,\ N_i \leftrightarrow N_j,\ \varepsilon_i \leftrightarrow \varepsilon_j\}, \tag{6.1}$$

leads to another allowed physical reaction if the arrow indicators $D_i = \varepsilon_i Q_i$ and $D_j = \varepsilon_j Q_j$ of these states are equal, i.e., if both external particle lines are directed into or both away from the interaction region. On the other side, if $D_i = -D_j$, the exchange of the two particle states has to be completed by a change of the charges Q_i and Q_j

$$\mathfrak{X}_{ij} \equiv \begin{cases} \mathfrak{X}^{(+)}_{ij} = \mathfrak{P}_{ij}, & \text{for } D_i = D_j, \\ \mathfrak{X}^{(-)}_{ij} = \backslash Q_i,Q_j/\mathfrak{P}_{ij}, & \text{for } D_i = -D_j. \end{cases} \tag{6.2}$$

The action of the crossing exchange $\mathfrak{X}_{ij}$ on a Feynman diagram may be visualized simply by deflecting the external line of the state i into the line of the state j and vice versa without changing the arrows connected with these lines. (See Fig. 2.) The arrow indicators D_i and D_j are conserved under $\mathfrak{X}_{ij}$,

$$D_j' \equiv \mathfrak{X}_{ij}D_j = D_j. \tag{6.3}$$

Let one of the states i or j belong to an incoming state ($\varepsilon_i = -1$) and the other to an outgoing state ($\varepsilon_j = +1$). In this case $\mathfrak{X}_{ij}$ corresponds essentially (apart from a possible helicity flip) to a double application of the "substitution rule," which tells us how to transform an incoming particle into an outgoing antiparticle. The "substitution rule" is usually formulated[1] by a change of spinors and particle momenta

$$\begin{aligned} u^+(p) &\rightleftarrows u^-(q'), \quad p \rightleftarrows -q', \\ \bar{u}^-(q) &\rightleftarrows \bar{u}^+(p'), \quad q \rightleftarrows -p'. \end{aligned} \tag{6.4}$$

The transformation of momenta on the positive mass shell to momenta on the negative mass shell as required by the "substitution law" can be realized by an analytical continuation of the amplitude.[15,16] It is difficult to make this procedure unique, moreover it can be avoided.

The formulation of the "substitution rule" is more transparent with the help of the running signatures ε_p of the particles as introduced in (3.5). Instead of (6.4) we have

$$Q \to Q' = -Q, \quad p \to p', \quad \varepsilon \to \varepsilon' = -\varepsilon. \tag{6.5}$$

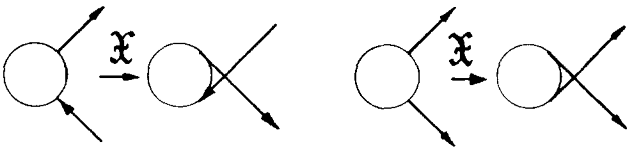

FIG. 2. Visualization of the crossing $\mathfrak{X}$ of two peripherical states.

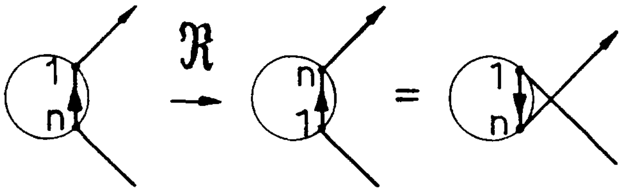

FIG. 3. Visualization of the reversal $\mathfrak{R}$ of the sequence of the interactions in a current.

If an amplitude is calculated so that the arguments of all resulting functions (like logarithm, square root, or Spence function having cuts) show how they are put infinitesimally above or below the real axis, the substitutions as required by the "substitution law" or by $\mathfrak{X}_{ij}$ can be realized uniquely.

We notice that an exchange $\mathfrak{X}_{ac}$ of two particle states a and c, which both are outgoing (incoming), can be realized by two operations $\mathfrak{X}_{ab}$ and $\mathfrak{X}_{bc}$, which exchange an incoming state (for instance, b) and an outgoing state. The relation

$$\mathfrak{X}_{ac} = \mathfrak{X}_{ab}\mathfrak{X}_{bc}\mathfrak{X}_{ab} \tag{6.6a}$$

can be written in more detail by

$$\mathfrak{X}_{ac}^{(D_aD_c)} = \mathfrak{X}_{ab}^{(D_cD_b)}\mathfrak{X}_{bc}^{(D_aD_c)}\mathfrak{X}_{ab}^{(D_aD_b)}. \tag{6.6b}$$

This relation confirms that the crossing exchanges $\mathfrak{X}_{ij}$ of any particle states are finally based on the "substitution rule."[17]

We point out that the crossing exchanges $\mathfrak{X}_{ij}$ and the current reversal $\mathfrak{D}$ have quite different features: $\mathfrak{X}_{ij}$ acts on the external lines, whereas $\mathfrak{D}$ changes also the interior of a Feynman diagram; $\mathfrak{X}_{ij}$ conserves the arrow indicators D of the states, whereas $\mathfrak{D}$ changes them.

Let the states e and s be connected by a current, so that $D_e = -D_s$. In this case the product of the crossing exchange $\mathfrak{X}_{es} = \mathfrak{X}_{es}^{(-)}$ and the current reversal causes a reversal of the sequence of the interactions this current meets. (See Fig. 3.)

The operation

$$\mathfrak{R} = \backslash Q_e, Q_s/\mathfrak{P}_{es}\mathfrak{D} \tag{6.7}$$

has the following effect on a current (3.9):

$$\mathfrak{R}\; J\left[{}^{Q_e}_{N_e}P_e, 1\cdots n, {}^{Q_s}_{N_s}P_s\right] = J\left[{}^{Q_e}_{N_e}P_e, n\cdots 1, {}^{Q_s}_{N_s}P_s\right]. \tag{6.8}$$

The realizations of $\mathfrak{R}$ are

$$\mathfrak{R}_{\mathrm{C}} = -N_eN_s\backslash \rho_a, \rho_u, Q_e, Q_s/\mathfrak{P}_{es}, \tag{6.9}$$

$$\mathfrak{R}_{\mathrm{H}} = N_eN_s\backslash \tau, \rho_a, N_e, N_s/\mathfrak{P}_{es}, \tag{6.10}$$

$$\mathfrak{R}_{*} = \backslash \tau, \rho_a, \rho_5, {*}/\mathfrak{P}_{es} \quad (v \text{ real}), \tag{6.11}$$

$$\mathfrak{R}_{\gamma_5} = -\backslash \rho_a, \rho_u, \rho_5, N_e, N_s, Q_e, Q_s, {*}/\mathfrak{P}_{es} \quad (v \text{ real}). \tag{6.12}$$

For applications in pure QED, the reversal $\mathfrak{R}$ was not unknown so far,[18] but has not been discussed in detail. In this special case $\mathfrak{D}$ causes only a certain sign as mentioned before, and therefore, $\mathfrak{R}$ is essentially equivalent to a crossing substitution $\mathfrak{X}_{es}^{(-)}$ as Eq. (6.7) shows.

We notice that the realizations

$$\mathfrak{D}_*, \mathfrak{D}_{\gamma_5}, \mathbf{I}_{\mathrm{C}*}, \mathbf{I}_{\mathrm{H}*}, \mathfrak{R}_*, \mathfrak{R}_{\gamma_5}, \tag{6.13}$$

are not very practical. They require in principle that all coefficients v of the vertices should be real. Moreover, all these realizations imply the application of a complex conjugation to a special current. Mostly an amplitude contains several currents. The constituents of these currents are twisted if the helicity amplitude is worked out, contractions are done, and the explicit form is found according to the calculational aids given in Appendix A. Therefore it is not possible to formulate a complex conjugation of one specific current, whereas the other currents are unchanged. But at least overall complex conjugation[14] may be obtained if $\mathfrak{R}_*$,... is applied to one current and the identities $\mathbf{I}_{\mathrm{C}*}$ or $\mathbf{I}_{\mathrm{H}*}$ to all other currents.

As mentioned in Sec. II, the polarization vector for a helicity state of a spin-1 particle depends upon whether the particle is incoming ($\varepsilon = -1$) or outgoing ($\varepsilon = +1$),

$$\epsilon_N^\mu(p,\varepsilon) = (-\varepsilon)^N\epsilon_M^\mu(p), \quad M \equiv -\varepsilon N. \tag{6.14}$$

Moreover, it is complex. The effect of a complex conjugation on this polarization vector is described according to Eq. (2.13) by

$$\backslash{*}/\epsilon_N^\mu(p,\varepsilon) = (-1)^N\epsilon_{-N}^\mu(p,\varepsilon). \tag{6.15}$$

This is an unwanted effect of $\backslash{*}/$ in the realizations (6.13) of the substitutions $\mathfrak{D}$, $\mathbf{I}$, and $\mathfrak{R}$. It can be avoided if the complex conjugations in these realizations are completed by a change of helicity N_j and a factor $(-1)^{N_j}$ for each affected vertex of a real spin-1 particle j:

$$\backslash{*}/\rightarrow(-1)^{N_j+\cdots}\backslash{*}, N_j, \ldots/. \tag{6.16}$$

We observe that all the realizations of $\mathfrak{D}$, $\mathbf{I}$, and $\mathfrak{R}$ keep the relation

$$\tau = -\varepsilon_sQ_s = \varepsilon_eQ_e \tag{6.17}$$

invariant. The meaning of this identification of τ is that a change of τ is included in $\backslash Q_e, Q_s/$, or in $\mathfrak{P}_{es}$. The relation (6.17) allows us to eliminate τ, especially in the following combinations appearing in the arguments of the propagators:

$$\tau P_s = -Q_sP_s, \quad \tau P_e = Q_eP_e. \tag{6.18}$$

As a simple example for the verification of the symmetries $\mathfrak{R}$ and $\mathfrak{X}$, let us calculate the partial amplitudes $\mathfrak{M}_1$ and $\mathfrak{M}_2$ of the reaction $e_{(1)}\gamma_{(3)}\rightarrow e_{(2)}Z_{0(4)}$ in the lowest order of the perturbation expansion. The amplitudes are written and elaborated according to the methods of Appendix A:

$$\begin{aligned}\mathfrak{M}_1 = -\mathfrak{F}_\kappa &\frac{(\rho_a\rho_5)^c\varepsilon_3(-\varepsilon_4)^{N_4}}{m(P_1K_3)}\frac{2^{-|M_4|/2}}{2\sqrt{2}m_{z_0}}\\ &\times(M_4+M_4')\langle 2|-A\,4'\rangle\langle -3|1\rangle\\ &\times[M_3N_1 2p_1\zeta_{12,A\,4',4'\|-\kappa Q_{12}}\langle -4'|1\rangle\langle 1|3\rangle\\ &-\tau\varepsilon_3\zeta_{1233,A\,4',4'\|-\kappa Q_{12}}\langle -4'|3\rangle],\end{aligned} \tag{6.19}$$

$$\begin{aligned}\mathfrak{M}_2 = -\mathfrak{F}_\kappa &\frac{(-\rho_a\rho_5)^c\varepsilon_3(-\varepsilon_4)^{N_4}}{m(P_2K_3)}\frac{2^{-|M_4|/2}}{2\sqrt{2}m_{z_0}}\\ &\times(M_4+M_4')\langle -1|-A\,4'\rangle\langle -3|-2\rangle(-N_1N_2)\\ &\times[M_3N_2 2p_2\zeta_{-1-2,A\,4',4'\|-\kappa Q_{12}}\langle -4'|-2\rangle\langle -2|3\rangle\\ &-\tau\varepsilon_3\zeta_{-1-233,A\,4',4'\|-\kappa Q_{12}}\langle -4'|3\rangle].\end{aligned} \tag{6.20}$$

The second partial amplitude is related to the first one by a reversal of the two interactions included in the current. Therefore for all realizations of $\mathfrak{R}$, the following relation

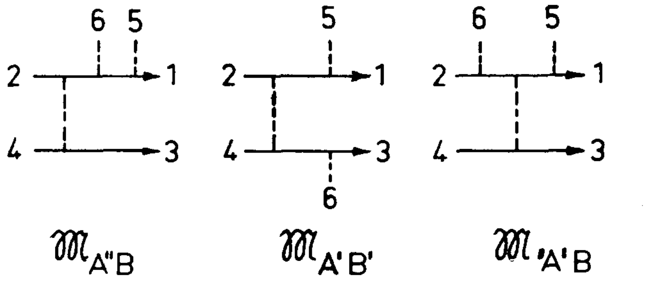

FIG. 4. The three generating Feynman diagrams for the reaction $ee \to ee\gamma\gamma$. The indices A and B refer to the currents A: $2\to 1$ and B: $4 \to 3$.

holds:

$$\mathfrak{M}_2 = \mathfrak{R}\mathfrak{M}_1. \tag{6.21}$$

VII. SUBSTITUTIONS APPLIED TO THE REACTION $ee \to ee\gamma\gamma$

Figure 4 shows three partial amplitudes contributing to the reaction $ee \to ee\gamma\gamma$. These amplitudes are given by

$$\begin{aligned}\mathfrak{M}_{A''B} &= \bar{u}_1(-\varepsilon_5)\underline{\epsilon_5}\rho_u^A S(\tau^A,P_1+K_5)(-\varepsilon_6)\underline{\epsilon_6}\rho_u^A\\ &\quad\times S(\tau^A,P_1+K_5+K_6)\gamma^\mu\rho_u^A(a+b\rho_a^A\rho_5^A\gamma_5)u_2\\ &\quad\times \bar{u}_3\gamma_\mu\rho_u^B(a+b\rho_a^B\rho_5^B\gamma_5)u_4 D(P_3+P_4),\end{aligned} \tag{7.1}$$

$$\begin{aligned}\mathfrak{M}_{A'B'} &= \bar{u}_1(-\varepsilon_5)\underline{\epsilon_5}\rho_u^A S(\tau^A,P_1+K_5)\\ &\quad\times \gamma^\mu\rho_u^A(a+b\rho_a^A\rho_5^A\gamma_5)u_2\\ &\quad\times \bar{u}_3(-\varepsilon_6)\underline{\epsilon_6}\rho_u^B S(\tau^B,P_3+K_6)\gamma_\mu\rho_u^B\\ &\quad\times (a+b\rho_a^B\rho_5^B\gamma_5)u_4 D(P_3+P_4+K_6),\end{aligned} \tag{7.2}$$

$$\begin{aligned}\mathfrak{M}_{'A'B} &= \bar{u}_1(-\varepsilon_5)\underline{\epsilon_5}\rho_u^A S(\tau^A,P_1+K_5)\gamma^\mu\rho_u^A(a+b\rho_a^A\rho_5^A\gamma_5)\\ &\quad\times S(\tau^A,-P_2-K_6)(-\varepsilon_6)\underline{\epsilon_6}\rho_u^A u_2\\ &\quad\times \bar{u}_3\gamma_\mu\rho_u^B(a+b\rho_a^B\rho_5^B\gamma_5)u_4 D(P_3+P_4).\end{aligned} \tag{7.3}$$

The vertex coefficients and the propagator $D(q)$ are for photons,

$$a=e, \quad b=0, \quad D(q)=q^{-2}, \tag{7.4}$$

and for Z_0-bosons,

$$a=eg_V, \quad b=-eg_A, \quad D(q)=(q^2-m_{z_0}^2)^{-1}. \tag{7.5}$$

The reversal of the interaction points of the current A: $2\to 1$ and B: $4\to 3$ is defined according to (6.7) as

$$\mathfrak{R}_A \equiv \backslash Q_1,Q_2/\mathfrak{P}_{12}\mathfrak{D}^A, \tag{7.6}$$

$$\mathfrak{R}_B \equiv \backslash Q_3,Q_4/\mathfrak{P}_{34}\mathfrak{D}^B. \tag{7.7}$$

The substitution $\mathfrak{D}^A$ acts on the signatures $\tau^A,\rho_a^A,\rho_u^A,\rho_5^A,Q_1,Q_2,N_1,N_2$ and $\mathfrak{D}^B$ on $\tau^B,\rho_a^B,\cdots,N_3,N_4$. The polarization vectors $\epsilon_i \equiv \epsilon_{N_i}$ of the photons with defined helicity N_i are complex valued. Therefore, we avoid the representations $\mathfrak{D}_*$ (5.5) and $\mathfrak{D}_{\gamma_5}$ (5.6), which are applicable only if all coefficients of the vertices are real or if the precautions (6.16) are taken.

The permutation of the currents A and B is defined by

$$\mathfrak{P}_{AB} \equiv \mathfrak{P}_{13}\mathfrak{P}_{24}\{\tau^A,\rho^A,\ldots \leftrightarrow \tau^B,\rho^B,\ldots\}. \tag{7.8}$$

The amplitudes (7.1)–(7.3) show internal symmetries

$$\mathfrak{M}_{A''B} = \mathfrak{R}_B\mathfrak{M}_{A''B}, \tag{7.9}$$

$$\mathfrak{M}_{A'B'} = \mathfrak{P}_{56}\mathfrak{P}_{AB}\mathfrak{M}_{A'B'}, \tag{7.10}$$

$$\mathfrak{M}_{'A'B} = \mathfrak{R}_B\mathfrak{M}_{'A'B}, \tag{7.11}$$

$$= \mathfrak{P}_{56}\mathfrak{R}_A\mathfrak{M}_{'A'B}. \tag{7.12}$$

Four partial amplitudes arise from $\mathfrak{M}_{A''B}$ by substitutions, all having two succeeding external photons:

$$(1+\mathfrak{P}_{AB})(1+\mathfrak{R}_A)\mathfrak{M}_{A''B}. \tag{7.13}$$

Similarly, $\mathfrak{M}_{A'B'}$ is the source of four amplitudes having two external photons at different currents:

$$(1+\mathfrak{R}_B)(1+\mathfrak{R}_A)\mathfrak{M}_{A'B'}. \tag{7.14}$$

Two partial amplitudes with two photons at the same current derived from

$$(1+\mathfrak{P}_{AB})\mathfrak{M}_{'A'B}. \tag{7.15}$$

These ten amplitudes are completed to 40 by the addition of all contributions with reversed photons and with final fermions exchanged, if the following operator is applied:

$$(1-\mathfrak{P}_{13})(1+\mathfrak{P}_{56}). \tag{7.16}$$

The internal symmetries (7.9)–(7.15) allow us to formulate the sum of all 40 partial amplitudes by

$$\begin{aligned}\mathfrak{M} &= (1-\mathfrak{P}_{13})(1+\mathfrak{P}_{56})(1+\mathfrak{P}_{AB})(1+\mathfrak{R}_A)(1+\mathfrak{R}_B)\\ &\quad\times[\tfrac{1}{2}\mathfrak{M}_{A''B}+\tfrac{1}{2}\mathfrak{M}_{A'B'}+\tfrac{1}{4}\mathfrak{M}_{'A'B}].\end{aligned} \tag{7.17}$$

This result shows that only three partial amplitudes of the reaction $ee\to ee\gamma\gamma$ have to be calculated explicitly whereas the remaining 37 amplitudes are obtained from these by various substitutions. The three generating helicity amplitudes have been worked out in Appendix D according to the methods described in the earlier paper I (Ref. 2) and Appendix A.

VIII. CONCLUDING REMARKS

The example of the preceding section illustrates that all the Feynman diagrams contributing to a physical reaction in the same order of the perturbation expansion can be classified in groups of topologically equivalent members. And all the members are connected by three types of transformations: (a) the permutation of two external particles of the same type—like $\mathfrak{P}_{12}$ for two electrons 1 and 2 or $\mathfrak{P}_{56}$ for two photons 5 and 6 in the preceding example; (b) the permutation of two currents of equal fermions passing through the reaction—like $\mathfrak{P}_{AB}$ which exchanges the current A: $2\to 1$ and B: $4\to 3$; and (c) the reversal of the order of the interaction points of a specific current—like $\mathfrak{R}_A$ or $\mathfrak{R}_B$. The amplitudes of two reactions that are different by a change of an incoming particle into an outgoing antiparticle are connected by a further type of transformation: (d) which applies the "substitution rule" and the crossing exchange $\mathfrak{X}$—as discussed in Sec. VI. Some of these transformations combine two symmetry operations, which reproduce each member of a group of topologically equivalent diagrams.

We showed in this paper that all these transformations can be formulated as simple substitutions that act on the particle indices and their natural signatures like charge, helicity, or in/out characters (Secs. II and III) as well as on some artificial dummy signs that have to be attached to the fermion propagators and the vertex operators in the different currents of a diagram (Secs. III and VI).

For each group of topologically equivalent diagrams there exists a representative expression that shows the particle signatures and dummy signs as general symbols. The representing expressions are expanded in explicit forms of functions of particle energies and momenta according to the methods described in Appendix A. The particle signatures and dummy signs are likewise apparent in the explicit and in the original form of the representative amplitude. Therefore the explicit contributions of the different Feynman diagrams can be generated from the explicit form of the representative amplitude by the application of simple substitutions. We notice that the symmetry operations can be used to prove the explicit forms of the generating expressions. If crossing operations are required— and the analytic form of the generating expressions are calculated—no difficult analytical continuations but only simple substitutions have to be performed.

In this way, the use of the substitutions connecting topologically equivalent Feynman diagrams reduces the calculational efforts to evaluate the helicity amplitudes.

ACKNOWLEDGMENTS

The authors thank Professor H. Salecker and Professor A. O. Barut for many constructive remarks. Critical discussions with W. Bauer, R. v. Mellenthin, and F. Ch. Simm are truly appreciated.

APPENDIX A: CALCULATIONAL AIDS FOR THE EVALUATION OF HELICITY AMPLITUDES

It is useful to formulate the helicity amplitudes not with the helicity spinors u_N^Q, Eq. (2.4)–(2.8), but with the spinors $w_N^Q = \sqrt{m}\, u_N^Q$ having the physical dimension $p^{1/2}$,

$$w_N^Q(p) = \frac{1}{\sqrt{2}}\begin{pmatrix} \zeta_M |M,\hat{p}\rangle \\ Q\zeta_{-M}|M,\hat{p}\rangle \end{pmatrix}, \quad M = QN,$$
$$\bar{w}_N^Q(p) = \frac{1}{\sqrt{2}}\,(Q\zeta_{-M}\langle M,\hat{p}|, \zeta_M\langle M,\hat{p}|), \tag{A1}$$

with

$$\zeta_M = (E + Mp)^{1/2}. \tag{A2}$$

These spinors are normalized according to

$$\bar{w}_N^Q(p)\, w_{\tilde{N}}^{\tilde{Q}}(p) = Qm\,\delta_{Q\tilde{Q}}\,\delta_{N\tilde{N}}. \tag{A3}$$

Sometimes we abbreviate these four-component spinors for different particles i by setting

$$w_i \equiv w_i^i \equiv w_{Q_iM_i}^{Q_i}(p_i), \quad w_i^{-i} \equiv w_{Q_iM_i}^{-Q_i}(p_i), \quad \text{etc.} \tag{A4}$$

Here we recall that according to (A1) $M = QN$ for spin-$\frac{1}{2}$ particles, but according to (6.14) $M = -\varepsilon N$ for spin-1 particles. Similarly, we abbreviate

$$|i\rangle \equiv |M_i,p_i\rangle, \quad \langle -i|j\rangle \equiv \langle -M_i,\hat{p}_i|M_j,\hat{p}_j\rangle,$$
$$Q_{12\cdots} \equiv Q_1Q_2\cdots, \quad \underline{\epsilon}_i \equiv \underline{\epsilon}_{N_i}(k_i), \ \text{etc.} \tag{A5}$$

The spinors of type w_N^Q can also be introduced for massless particles. In this case we have

$$w_N^Q(k) = QNw_{-N}^{-Q}(k) = QNw_Q^N(k), \quad \text{for } k^2 = 0. \tag{A6}$$

The factor γ_5 acts on a four-component spinor according to (4.24),

$$\gamma_5 w_N^Q(p) = w_{-N}^{-Q}(p), \quad \bar{w}_N^Q(p)\gamma_5 = -\bar{w}_{-N}^{-Q}(p). \tag{A7}$$

The numerators of Feynman amplitudes show several Dirac factors $\underline{a} = a_\mu\gamma^\mu$. In many cases it is useful to replace these factors by dyadic products of helicity spinors.[19] We give some examples, using the convention that the two values ± 1 of the primed indices should be summed

$$\underline{p} + Qm = 2w_{N'}^Q(p)\bar{w}_{N'}^Q(p), \quad \text{for } p^2 = m^2, \tag{A8}$$

$$\underline{p} = w_{N'}^{Q'}(p)\bar{w}_{N'}^{Q'}(p), \tag{A9}$$

$$\underline{k} = 2w_{N'}^Q(k)\bar{w}_{N'}^Q(k), \quad \text{for } k^2 = 0, \tag{A10}$$

$$\underline{\epsilon}_N(p) = -(1/m)2^{-|N|/2}(N - Q'N')w_{AN'}^{Q'}(p)\bar{w}_{-N'}^{-Q'}(p),$$
$$\text{for } N = 0, \pm 1, \quad A = (-1)^N, \ AN = -N, \tag{A11}$$

$$\underline{\epsilon}_N(p) = -N(\sqrt{2}/m)w_{NQ'}^{Q'}(p)\bar{w}_{NQ'}^{-Q'}(p), \quad \text{for } N = \pm 1, \tag{A12}$$

$$\underline{p}\underline{\epsilon}_N(p) = -\underline{\epsilon}_N(p)\underline{p}$$
$$= 2^{-|N|/2}(NQ' - N')w_{-AN'}^{-Q'}(p)\bar{w}_{N'}^{Q'}(p),$$
$$\text{for } N = 0, \pm 1, \quad A = (-1)^N, \tag{A13}$$

$$\underline{k}\underline{\epsilon}_N(k) = -\underline{\epsilon}_N(k)\underline{k} = 2\sqrt{2}Qw_{QN}^Q(k)\bar{w}_{-QN}^Q(k),$$
$$\text{for } N = \pm 1, \quad k^2 = 0, \tag{A14}$$

$$1 = (1/m)Q'w_{N'}^{Q'}(p)\bar{w}_{N'}^{Q'}(p). \tag{A15}$$

In Eqs. (A10) and (A14) the charge $Q = \pm 1$ is really fictitious and has no affect on physical results.

From the well-known formula (A8) one derives (A9), (A10), and (A15). The relation (A11) shows how to treat the vertex for photons or other (massive) spin-1 particles. Equation (A11) is derived separately in Appendix B. Equation (A12) specializes (A11). Equation (A13) is a consequence of (A9) and (A11) with regard to the normalization (A3). Equation (A14) is obtained by a specialization of (A13) with the help of (A6).

If two or more successive factors γ_μ appear in a Dirac current, our method separates these by the dyadic form (A15) of the unit operator.

If all the Dirac factors included in the helicity currents are replaced by their dyadic forms (A8)–(A15), the numerator of a Feynman helicity amplitude is resolved into a sum of products with four-spinor scalars like $\bar{w}_i w_j$ or contracted helicity vector currents like $\bar{w}_1\gamma_\mu w_2\bar{w}_3\gamma^\mu w_4$. Moreover, the last expression can be transformed into four-spinor scalars because of the following formula[20]:

$$\bar{w}_1^1\gamma_\mu w_2^2\bar{w}_3^3\gamma^\mu w_4^4 = \bar{w}_1^1 w_4^4\bar{w}_3^3 w_2^2 - \bar{w}_1^1 w_{-4}^{-4}\bar{w}_3^3 w_{-2}^{-2}$$
$$- N_2N_3\bar{w}_1^1 w_{-3}^3\bar{w}_{-2}^2 w_4^4$$
$$+ N_2N_3\bar{w}_1^1 w_3^{-3}\bar{w}_{-2}^2 w_{-4}^{-4}. \tag{A16}$$

If we proceed in this way we obtain many factors $\bar{w}_i w_j$ that can be expressed according to (A19) by scalars of two-component spinors like $\langle M_i,\hat{p}_i|M_j,\hat{p}_j\rangle$ and energy functions of the form

$$\zeta_{M_1M_2\cdots M_n\|Q}$$
$$\equiv \zeta_{M_1}(p_1)\cdots\zeta_{M_n}(p_n) + Q\zeta_{-M_1}(p_1)\cdots\zeta_{-M_n}(p_n). \tag{A17}$$

Sometimes we write this energy function in the abbreviated notation

$$\zeta_{12\cdots n\|Q} \equiv \zeta_{M_1M_2\cdots M_n\|Q}. \tag{A18}$$

However, it is not useful to replace schematically all the Dirac factors by dyadic products. In this way the number of summations over the intermediate indices Q_i and N_i would increase too much. Therefore we propose to maintain several forms of currents and current products that can be expressed in a manner as compact as $\bar{w}_i w_j$.

We gather here some useful compact expression of several currents and current products:

$$\bar{w}_1 w_2 = \tfrac{1}{2} Q_2 \zeta_{1-2\|Q_{12}} \langle 1|2\rangle = \tfrac{1}{2} Q_1 \zeta_{-12\|Q_{12}} \langle 1|2\rangle, \tag{A19}$$

$$\bar{w}_1 \gamma_\mu w_2 \bar{w}_3 \gamma^\mu w_4$$
$$= \tfrac{1}{2} M_2 M_3 [Q_{34} \zeta_{12-3-4\|Q_{1234}} \langle 1|4\rangle\langle -2|-3\rangle$$
$$- \zeta_{1234\|Q_{1234}} \langle 1|-3\rangle\langle -2|4\rangle], \tag{A20}$$

$$\bar{w}_1 \underline{\epsilon}_{N_2}(p_2) w_3$$
$$= (1/2m_2) 2^{-|N_2|/2} (N_2 + N_2') \zeta_{M_1, AN_2', N_2', M_3\| -Q_{13}}$$
$$\times \langle M_1| -AN_2'\rangle\langle -N_2'|M_3\rangle,$$
$$\text{for } N_2 = 0, \pm 1, \quad A = (-1)^{N_2}, \quad AN_2 = -N_2. \tag{A21}$$

This formula is specialized for $N_2 = 0$,

$$\bar{w}_1 \underline{\epsilon}_0(p_2) w_3$$
$$= (1/2m_2) N_2' \zeta_{M_1 N_2' N_2' M_3\| -Q_{13}}$$
$$\times \langle M_1| -N_2'\rangle\langle -N_2'|M_3\rangle, \tag{A21$'$}$$

and for $N_2 = \pm 1$,

$$\bar{w}_1 \underline{\epsilon}_{N_2}(p_2) w_3 = (1/\sqrt{2}) N_2 \zeta_{13\| -Q_{13}} \langle 1|2\rangle\langle -2|3\rangle, \tag{A21$''$}$$

$$\bar{w}_1 \underline{\epsilon}_{N_2} \underline{\epsilon}_{N_3} w_4 = -(1/2m_2 m_3) 2^{-(|N_2| + |N_3|)/2}$$
$$\times Q_4 (N_2 + N_2')(N_3 - N_3')$$
$$\times \zeta_{M_1, A_2 N_2', N_2', A_3 N_3', N_3', -M_4\|Q_{14}}$$
$$\times \langle M_1| -A_2 N_2'\rangle\langle -N_2'|A_3 N_3'\rangle\langle N_3'|M_4\rangle,$$
$$\text{for } N_i = 0, \pm 1, \quad A_i = (-1)^{N_i}. \tag{A22}$$

Especially[21] for $N_2 = \pm 1$, $N_3 = \pm 1$,

$$\bar{w}_1 \underline{\epsilon}_{N_2} \underline{\epsilon}_{N_3} w_4 = -Q_4 N_2 N_3 \zeta_{1-4\|Q_{14}} \langle 1|2\rangle\langle -2|3\rangle\langle -3|4\rangle, \tag{A22$'$}$$

$$\bar{w}_1 \underline{p}_2 w_3 = \tfrac{1}{2} \zeta_{12'2'3\|Q_{13}} \langle 1| -2'\rangle\langle -2'|3\rangle, \tag{A23}$$

$$\bar{w}_1 \underline{p}_2 \underline{p}_3 w_4 = \tfrac{1}{2} Q_4 \zeta_{12'2'3'3'-4\|Q_{14}} \langle 1| -2'\rangle\langle -2'|3'\rangle\langle 3'|4\rangle. \tag{A24}$$

In paper I we derived some formulas that expressed products of currents with several contractions like $\bar{w}_1 \gamma_\mu \gamma_\nu w_2 \bar{w}_3 \gamma^\nu \gamma^\mu w_4$, etc. These formulas have been derived from the very general formula I(3.11). Our present method, namely to separate the factors by the dyadic form (A15) of the unit operator and to apply (A20) repeatedly, works just as effectively. In some cases several summations over helicities N_i or "charges" Q_i—denoted by primed indices—appear in the results. These summations can be done in many cases. For this purpose we found the following formulas useful:

$$\zeta_{M_1\cdots M_m\|AQ'} \zeta_{M_{m+1}\cdots M_n\|BQ'} = 2\zeta_{M_1\cdots M_n\|AB}, \tag{A25}$$

$$Q' \zeta_{M_1\cdots M_m\|AQ'} \zeta_{M_{m+1}\cdots M_n\|BQ'}$$
$$= 2A\zeta_{-M_1\cdots -M_m M_{m+1}\cdots M_n\|AB}$$
$$= 2B\zeta_{M_1\cdots M_m - M_{m+1}\cdots - M_n\|AB}, \tag{A26}$$

$$\zeta_{N_k N_k N_k' N_k'\cdots\|Q} |N_k'\rangle\langle N_k'| = 2k\zeta_{N_k N_k\cdots\|Q} |N_k\rangle\langle N_k|,$$
$$\text{for } k^2 = 0, \tag{A27}$$

$$\zeta_{M\tilde{M}\|+1} = 2E\delta_{M\tilde{M}} + 2m\delta_{M,-\tilde{M}}, \tag{A28}$$

$$\zeta_{M\tilde{M}\|-1} = 2Mp\delta_{M\tilde{M}}, \tag{A29}$$

$$\zeta_M \zeta_{-M} = m, \tag{A30}$$

$$\zeta_{M_1\cdots M_n M_r - M_r\|Q} = m_r \zeta_{M_1\cdots M_n\|Q}, \tag{A31}$$

$$\zeta_{M_1\cdots\|Q} = Q\zeta_{-M_1 - \cdots\|Q}, \tag{A32}$$

$$\zeta_{M_k M_k\cdots\|Q} = M_k \zeta_{M_k M_k\cdots\| -Q}, \quad \text{for } k^2 = 0, \tag{A33}$$

$$\zeta_{M_i M_j\cdots\|Q_1} \zeta_{M_i M_r\cdots\|Q_2} = \zeta_{M_i M_i M_j\cdots M_r\cdots\|Q_1 Q_2}$$
$$+ m_i Q_2 \zeta_{M_j\cdots -M_r -\cdots\|Q_1 Q_2}, \tag{A34}$$

$$\langle M,\hat{p}|\tilde{M},\hat{p}\rangle = \delta_{M\tilde{M}}, \tag{A35}$$

$$|M',\hat{p}\rangle\langle M',\hat{p}| = 1, \tag{A36}$$

$$\langle -M_1,\hat{p}_1| -M_2,\hat{p}_2\rangle = M_1 M_2 \langle M_2,\hat{p}_2|M_1,\hat{p}_1\rangle. \tag{A37}$$

The relations (A25)–(A34) are easily derived from the definitions (A2) and (A17); (A35) is the normalization relation of the two-component spinors, (A20) is found in paper I as I(B12), and (A37) is a special case of I(A35). The completeness relation (A36) is contained in I(A32) for $i = k$, while (A19) is directly derived from (A1) with (A17), and (A16) is obtained by a comparison of (A19) and (A20). Also, (A21) and (A22) are derived from (A11) with (A19) and (A25). We notice that (A22) can also be reduced to (A21) if the dyadic form (A15) of the unit operator is inserted between $\underline{\epsilon}_2$ and $\underline{\epsilon}_3$, and (A26) is applied. The specializations (A21$''$) and (A22$'$) use (A30). The relation (A9) with (A19) and (A25) yields (A23) and (A24).

If a vertex in a current shows a factor $v(a + b\gamma_5)$ we replace this factor by

$$v\mathfrak{F}_\kappa (\rho_a \rho_5 \gamma_5)^c, \quad \text{if} \quad v = \gamma_\mu, \tag{A38}$$

$$v\mathfrak{F}_\kappa (\rho_5 \gamma_5)^c, \quad \text{if} \quad v = 1, \tag{A39}$$

$$v\mathfrak{F}_\kappa \gamma_5^c, \quad \text{if} \quad v = \gamma_\mu \gamma_\nu, \quad \mu \neq \nu. \tag{A40}$$

Here the exponent of γ_5 is a function of a dummy sign $\kappa = \pm 1$,

$$c(\kappa) = \tfrac{1}{2}(1 - \kappa) = c, \tag{A41}$$

and $\mathfrak{F}_\kappa$ represents a substitution operator

$$\mathfrak{F}_\kappa \equiv a + b \setminus \kappa / \tag{A42}$$

in front of the current. The meaning is that κ, like all other dummy sign factors introduced in this paper, gets the value $+1$ after the substitutions have been worked.

The surrogate factor γ_5^c appears only in the neighborhood of a spinor or of a numerator of a propagator replaced by a dyadic product according to (A8)–(A15). Therefore the formulas (A7) or more directly

$$\gamma_5^{\zeta(\kappa)} w_N^Q = w_{\kappa N}^{\kappa Q}, \quad \bar{w}_N^Q \gamma_5^{\zeta(\kappa)} = \kappa \bar{w}_{\kappa N}^{\kappa Q}, \tag{A43}$$

can always be applied to eliminate γ_5^ζ.

If the propagators of the bosons $a,b,\ldots$ belong to photons or massive vector bosons, we introduce dummy masses $\lambda_a, \lambda_b, \ldots$ in these propagators and specialize at the end of the calculation. Thereby we use the symbol

$$|_{\lambda = m} \equiv \{\lambda = m \text{ in the succeeding expression}\} \tag{A44}$$

to assign a special mass to the dummy λ.

APPENDIX B: THE DYADIC FORM OF THE VERTEX $\underline{\epsilon}_N$ OF A MASSIVE SPIN-1 PARTICLE OF DEFINITE HELICITY N

For the "transversal" helicities $N = \pm 1$ we had in paper I, Eq. (3.10), the formula

$$\epsilon_N^\mu(p)\sigma_\mu = -\epsilon_N^\mu(p)\sigma^\mu = N\sqrt{2}|N,\hat{p}\rangle\langle -N,\hat{p}|. \tag{B1}$$

For the polarization vector (2.22) of "longitudinal" helicity $N = 0$, i.e.,

$$\epsilon_0^\mu(p) = \left(\frac{p}{m}, \frac{E}{m}\hat{p}\right) = \frac{E}{mp}p^\mu - \frac{m}{p}(1,\vec{0}), \tag{B2}$$

we use (A36) and from paper I, Eq. I(3.8), to derive[22] the formulas

$$\epsilon_0^\mu(p)\sigma_\mu = \frac{1}{m}\sum_{N' = \pm 1} N'\zeta_{N'}^2(p)| -N',\hat{p}\rangle\langle -N',\hat{p}|, \tag{B3}$$

$$\epsilon_0^\mu(p)\sigma^\mu = \frac{1}{m}\sum_{N' = \pm 1} N'\zeta_{N'}^2(p)|N',\hat{p}\rangle\langle N',\hat{p}|. \tag{B4}$$

Equations (B3) and (B1) as well as (B4) and (B1) can be combined using (A30), $A = (-1)^N$, and $N = 0, \pm 1$ to give

$$\epsilon_N^\mu(p)\sigma_\mu = \frac{1}{m}2^{-|N|/2}\sum_{N' = \pm 1}(N+N')\zeta_{N'}(p)\zeta_{AN'}(p) \times | -AN',\hat{p}\rangle\langle -N',\hat{p}|, \tag{B5}$$

$$\epsilon_N^\mu(p)\sigma^\mu = -\frac{1}{m}2^{-|N|/2}\sum_{N' = \pm 1}(N+N')\zeta_{-N'}(p)\zeta_{-AN'}(p) \times | -AN',\hat{p}\rangle\langle -N',\hat{p}|. \tag{B6}$$

Using (A1) we construct the dyadic product ($\widetilde{M} = Q\widetilde{N}$)

$$w_{-A\widetilde{N}}^{-Q}(p)\bar{w}_{\widetilde{N}}^Q(p) = \frac{1}{2}\begin{pmatrix} Q\zeta_{A\widetilde{M}}\zeta_{-\widetilde{M}} & \zeta_{A\widetilde{M}}\zeta_{\widetilde{M}} \\ -\zeta_{-A\widetilde{M}}\zeta_{-\widetilde{M}} & -Q\zeta_{-A\widetilde{M}}\zeta_{\widetilde{M}} \end{pmatrix}|A\widetilde{M},\hat{p}\rangle\langle\widetilde{M},\hat{p}|. \tag{B7}$$

The multiplication with $N - Q\widetilde{N}$ and summation over Q and $\widetilde{N}$ yields

$$\sum_{Q,\widetilde{N} = \pm 1}(N - Q\widetilde{N})w_{-A\widetilde{N}}^{-Q}(p)\bar{w}_{\widetilde{N}}^Q(p) = \sum_{N' = \pm 1}(N+N')\begin{pmatrix} 0 & \zeta_{-AN'}\zeta_{-N'} \\ -\zeta_{AN'}\zeta_{N'} & 0 \end{pmatrix} \times | -AN',\hat{p}\rangle\langle -N',\hat{p}|. \tag{B8}$$

Multiply now with $-(1/m)2^{-|N|/2}$. The result is equal to

$$\underline{\epsilon}_N = \begin{pmatrix} 0 & \epsilon_N^\mu\sigma^\mu \\ \epsilon_N^\mu\sigma_\mu & 0 \end{pmatrix}, \tag{B9}$$

where the elements are given by (B5) and (B6). Hence, formula (A11) is correct.

APPENDIX C: SPIN-1 REPRESENTATIONS

The standard representation of spin-1 states is obtained from the Kronecker product of two Weyl spinors $u^{\dot{}}$ and $v.$ by

$$\xi = U\begin{pmatrix} u^1 v_1 \\ u^1 v_2 \\ u^2 v_1 \\ u^2 v_2 \end{pmatrix}, \quad U = \begin{pmatrix} 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 \\ \cdot & 1/\sqrt{2} & 1/\sqrt{2} & \cdot \\ \cdot & 1/\sqrt{2} & -1/\sqrt{2} & \cdot \end{pmatrix}. \tag{C1}$$

In the rest system of the particle, the unitary transformation U separates the spin-1 space (components 1,2,3) and the spin-0 space (component 4). A further unitary transformation T leads to the vector representation

$$\overset{\triangle}{\xi} = T\xi, \quad T = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 & 1 & \cdot & \cdot \\ -i & -i & \cdot & \cdot \\ \cdot & \cdot & \sqrt{2} & \cdot \\ \cdot & \cdot & \cdot & \sqrt{2} \end{pmatrix}. \tag{C2}$$

We find

$$\overset{\triangle}{\xi} = \frac{1}{\sqrt{2}}(u_{\dot{}}^T\sigma^{\dot{}\cdot}v_{.}) = (\vec{\xi},\xi^0). \tag{C3}$$

Here

$$\sigma^{\dot{}\cdot} = (-\sigma_x, -\sigma_y, -\sigma_z, 1) \tag{C4}$$

are the Pauli matrices, and

$$u_{\dot{}} = \chi u^{\dot{}}, \quad \chi = \begin{pmatrix} \cdot & 1 \\ -1 & \cdot \end{pmatrix}. \tag{C5}$$

This result helps to understand Eqs. (2.14) and (2.19).

APPENDIX D: THE BASIC HELICITY AMPLITUDES OF THE REACTION $ee \to ee\gamma\gamma$

The following reactions $e^\pm e^\pm \to e^\pm e^\pm\gamma\gamma$, $e^+e^- \to e^+e^-\gamma\gamma$, and $\gamma e^\pm \to e^\pm e^+e^-\gamma$ are described by the same helicity amplitude for various configurations: two spin-$\frac{1}{2}$ particles in the "starting" states 2 and 4 are scattered into the "ending" states 1 and 3. Thereby two photons 5 and 6 are emitted.

For abbreviation we use particle indices instead of particle signatures if the meaning of the indexed expression is unique. These abbreviations are explained in (A4), (A5), and (A18). According to Eqs. (2.4), (3.6), (6.14), and (6.17), the following relations are used in the calculation (fermion indices $j = 1,2,3,4$ photon indices $l = 5,6$):

$$M_j = Q_j N_j, \quad M_l = -\varepsilon_l N_l, \quad P_j = \varepsilon_j p_j, \quad K_l = \varepsilon_l k_l, \tag{D1}$$

$$Q_1 = \tau^A \varepsilon_1, \quad Q_2 = -\tau^A \varepsilon_2, \quad Q_3 = \tau^B \varepsilon_3, \quad Q_4 = -\tau^B \varepsilon_4. \tag{D2}$$

According to the convention introduced in Appendix A, primed signatures or indices should be summed.

The Feynman diagrams of the basic helicity amplitudes are shown in Fig. 4. These diagrams contain two vertices and one boson propagator describing the γ-exchange as well as the Z_0-exchange. The two vertices are replaced by $(\rho_a^A \rho_5^A)^{C^A} \gamma^\mu \gamma_5^{C^A}$ with $C^A = \frac{1}{2}(1-\kappa^A)$, and $(\rho_a^B \rho_5^B)^{C^B} \gamma_\mu \gamma_5^{C^B}$ with $C^B = \frac{1}{2}(1-\kappa^B)$. The boson propagator has a dummy mass λ, which is zero for a γ-exchange, and $m_{z_0} + i\Gamma_{z_0}$ for a Z_0-exchange. The amplitudes are therefore dependent on κ^A, κ^B, and λ.

As an example for the method of calculation described in Appendix A, consider the partial amplitude (7.2)

$$\begin{aligned}\mathfrak{M}_{A'B'}^{\kappa^A,\kappa^B,\lambda} &= (1/m^2) Q_1 Q_3 \varepsilon_5 \varepsilon_6 (\rho_a^A \rho_5^A)^{C^A} (\rho_a^B \rho_5^B)^{C^B} \{2(P_1K_5) 2(P_3K_6)[(P_3+P_4+K_6)^2 - \lambda^2]\}^{-1} \\ &\times [\bar{w}_1^1 \underline{\epsilon}_5 (\underline{p}_1 + Q_1 m)\gamma^\mu \gamma_5^{C^A} w_2^2 \bar{w}_3^3 \underline{\epsilon}_6 (\underline{p}_3 + Q_3 m)\gamma_\mu \gamma_5^{C^B} w_4^4 + \varepsilon_1 \varepsilon_5 \bar{w}_1^1 \underline{\epsilon}_5 \underline{k}_5 \gamma^\mu \gamma_5^{C^A} w_2^2 \bar{w}_3^3 \underline{\epsilon}_6 (\underline{p}_3 + Q_3 m)\gamma_\mu \gamma_5^{C^B} w_4^4 \\ &+ \varepsilon_3 \varepsilon_6 \bar{w}_1^1 \underline{\epsilon}_5 (\underline{p}_1 + Q_1 m)\gamma^\mu \gamma_5^{C^A} w_2^2 \bar{w}_3^3 \underline{\epsilon}_6 \underline{k}_6 \gamma_\mu \gamma_5^{C^B} w_4^4 + \varepsilon_1 \varepsilon_3 \varepsilon_5 \varepsilon_6 \bar{w}_1^1 \underline{\epsilon}_5 \underline{k}_5 \gamma^\mu \gamma_5^{C^A} w_2^2 \bar{w}_3^3 \underline{\epsilon}_6 \underline{k}_6 \gamma_\mu \gamma_5^{C^B} w_4^4] .\end{aligned} \tag{D3}$$

The relations (A8) and (A14) allow us to substitute $\underline{p} + Qm$ and $\underline{\epsilon}_l \underline{k}_l$ by (sums of) dyadic spinor products

$$\begin{aligned}\mathfrak{M}_{A'B'}^{\kappa^A,\kappa^B,\lambda} &= \frac{1}{m^2} Q_1 Q_3 \varepsilon_5 \varepsilon_6 (\rho_a^A \rho_5^A)^{C^A} (\rho_a^B \rho_5^B)^{C^B} \{2(P_1K_5) 2(P_3K_6)[(P_3+P_4+K_6)^2 - \lambda^2]\}^{-1} \\ &\times [4\bar{w}_1^1 \underline{\epsilon}_5 w_{1'}^1 \bar{w}_{1'}^1 \gamma^\mu \gamma_5^{C^A} w_2^2 \bar{w}_3^3 \underline{\epsilon}_6 w_{3'}^3 \bar{w}_{3'}^3 \gamma_\mu \gamma_5^{C^B} w_4^4 - \varepsilon_1 \varepsilon_5 Q_5 4\sqrt{2} \bar{w}_1^1 w_5^5 \bar{w}_{-5}^5 \gamma^\mu \gamma_5^{C^A} w_2^2 \bar{w}_3^3 \underline{\epsilon}_6 w_{3'}^3 \bar{w}_{3'}^3 \gamma_\mu \gamma_5^{C^B} w_4^4 \\ &- \varepsilon_3 \varepsilon_6 Q_6 4\sqrt{2} \bar{w}_1^1 \underline{\epsilon}_5 w_{1'}^1 \bar{w}_{1'}^1 \gamma^\mu \gamma_5^{C^A} w_2^2 \bar{w}_3^3 w_6^6 \bar{w}_{-6}^6 \gamma_\mu \gamma_5^{C^B} w_4^4 + \varepsilon_1 \varepsilon_3 \varepsilon_5 \varepsilon_6 Q_5 Q_6 8 \bar{w}_1^1 w_5^5 \bar{w}_{-5}^5 \gamma^\mu \gamma_5^{C^A} w_2^2 \bar{w}_3^3 w_6^6 \bar{w}_{-6}^6 \gamma_\mu \gamma_5^{C^B} w_4^4].\end{aligned} \tag{D4}$$

The spinor expressions $\bar{w}_a \gamma^\mu \gamma_5^{C^A} w_b \bar{w}_c \gamma_\mu \gamma_5^{C^B} w_d$, $\bar{w}_a \underline{\epsilon} w_b$, and $\bar{w}_a w_b$, are now evaluated using the relations (A43), (A20), (A21″), and (A19). Helicities with primed indices are summed. The relation (A29) helps to perform these summations. The fictitious dependence upon the arbitrary "charges of the photons" Q_5 and Q_6 will disappear. This result is obtained with the help of (A32) and (A34),

$$\begin{aligned}\mathfrak{M}_{A'B'}^{\kappa^A,\kappa^B,\lambda} &= (1/m^2)(\rho_a^A \rho_5^A)^{C^A} (\rho_a^B \rho_5^B)^{C^B} Q_{13} M_2 N_6 \{2(P_1K_5) 2(P_3K_6)[(P_3+P_4+K_6)^2 - \lambda^2]\}^{-1} \langle 1|5\rangle \langle 3|6\rangle \\ &\times \{N_5 M_1 4 p_1 p_3 \langle -5|1\rangle \langle -6|3\rangle [\kappa^B Q_{34} \zeta_{12-3-4\|Q\kappa^A\kappa^B} \langle 1|4\rangle \langle -2|-3\rangle - \zeta_{1234\|Q\kappa^A\kappa^B} \langle 1|-3\rangle \langle -2|4\rangle] \\ &+ \varepsilon_1 2 p_3 \langle -6|3\rangle [\kappa^B Q_{34} \zeta_{12-3-4-5-5\|Q\kappa^A\kappa^B} \langle -5|4\rangle \langle -2|-3\rangle - \zeta_{1234-5-5\|Q\kappa^A\kappa^B} \langle -5|-3\rangle \langle -2|4\rangle] \\ &+ \varepsilon_3 \varepsilon_6 N_5 M_1 2 p_1 \langle -5|1\rangle [\kappa^B Q_{34} \zeta_{12-3-466\|Q\kappa^A\kappa^B} \langle 1|4\rangle \langle -2|6\rangle - \zeta_{1234-6-6\|Q\kappa^A\kappa^B} \langle 1|6\rangle \langle -2|4\rangle] \\ &+ \varepsilon_1 \varepsilon_3 \varepsilon_6 [\kappa^B Q_{34} \zeta_{12-3-4-5-566\|Q\kappa^A\kappa^B} \langle -5|4\rangle \langle -2|6\rangle - \zeta_{1234-5-5-6-6\|Q\kappa^A\kappa^B} \langle -5|6\rangle \langle -2|4\rangle]\}.\end{aligned} \tag{D5}$$

Similarly we calculate the remaining two Feynman amplitudes

$$\begin{aligned}\mathfrak{M}_{A'B}^{\kappa^A,\kappa^B,\lambda} &= (1/m^2) \rho_u^A \rho_u^B (\rho_a^A \rho_5^A)^{C^A} (\rho_a^B \rho_5^B)^{C^B} M_2 M_3 \{2(P_1K_5)[(P_1+K_5+K_6)^2 - m^2][(P_3+P_4)^2 - \lambda^2]\}^{-1} \langle 1|5\rangle \\ &\times \{N_5 N_6 4 p_1^2 \langle -5|1\rangle \langle 1|6\rangle \langle -6|1\rangle [\kappa^B Q_{34} \zeta_{12-3-4\|Q\kappa^A\kappa^B} \langle 1|4\rangle \langle -2|-3\rangle - \zeta_{1234\|Q\kappa^A\kappa^B} \langle 1|-3\rangle \langle -2|4\rangle] \\ &+ \varepsilon_1 \varepsilon_5 N_5 N_6 M_1 2 p_1 \langle -5|1\rangle \langle 1|6\rangle \langle -6|5'\rangle [\kappa^B Q_{34} \zeta_{12-3-45'5'\|-Q\kappa^A\kappa^B} \langle 5'|4\rangle \langle -2|-3\rangle \\ &- \zeta_{12345'5'\|-Q\kappa^A\kappa^B} \langle 5'|-3\rangle \langle -2|4\rangle] + \varepsilon_1 N_5 M_1 2 p_1 \langle -5|1\rangle \langle 1|6\rangle \\ &\times [\kappa^B Q_{34} \zeta_{12-3-4-6-6\|Q\kappa^A\kappa^B} \langle -6|4\rangle \langle -2|-3\rangle - \zeta_{1234-6-6\|Q\kappa^A\kappa^B} \langle -6|-3\rangle \langle -2|4\rangle] \\ &+ \epsilon_1 N_6 \langle -5|6\rangle \langle -6|1'\rangle \zeta_{11'-5-5\|-} [\kappa^B Q_{34} \zeta_{1'2-3-4\|Q\kappa^A\kappa^B} \langle 1'|4\rangle \langle -2|-3\rangle - \zeta_{1'234\|Q\kappa^A\kappa^B} \langle 1'|-3\rangle \langle -2|4\rangle] \\ &+ N_5 N_6 2 k_5 \langle -5|6\rangle \langle -6|-5\rangle [\kappa^B Q_{34} \zeta_{12-3-4-5-5\|Q\kappa^A\kappa^B} \langle -5|4\rangle \langle -2|-3\rangle \\ &- \zeta_{1234-5-5\|Q\kappa^A\kappa^B} \langle -5|-3\rangle \langle -2|4\rangle] + \langle -5|6\rangle [\kappa^B Q_{34} \zeta_{12-3-4-5-5-6-6\|Q\kappa^A\kappa^B} \langle -6|4\rangle \langle -2|-3\rangle \\ &- \zeta_{1234-5-5-6-6\|Q\kappa^A\kappa^B} \langle -6|-3\rangle \langle -2|4\rangle]\},\end{aligned} \tag{D6}$$

$$\begin{aligned}\mathfrak{M}_{A'B}^{\kappa^A,\kappa^B,\lambda} &= (1/m^2) \rho_u^A \rho_u^B (\rho_a^A \rho_5^A)^{C^A} (\rho_a^B \rho_5^B)^{C^B} Q_{12} M_3 N_6 \{2(P_1K_5) 2(P_2K_6)[(P_3+P_4)^2 - \lambda^2]\}^{-1} \langle 1|5\rangle \langle -6|2\rangle \\ &\times \{N_5 M_1 4 p_1 p_2 \langle -5|1\rangle \langle 2|6\rangle [\kappa^B Q_{34} \zeta_{12-3-4\|Q\kappa^A\kappa^B} \langle 1|4\rangle \langle -2|-3\rangle - \zeta_{1234\|Q\kappa^A\kappa^B} \langle 1|-3\rangle \langle -2|4\rangle] \\ &+ \varepsilon_1 2 p_2 \langle 2|6\rangle [\kappa^B Q_{34} \zeta_{12-3-4-5-5\|Q\kappa^A\kappa^B} \langle -5|4\rangle \langle -2|-3\rangle - \zeta_{1234-5-5\|Q\kappa^A\kappa^B} \langle -5|-3\rangle \langle -2|4\rangle] \\ &+ \varepsilon_2 \varepsilon_6 N_5 M_1 2 p_1 \langle -5|1\rangle [\kappa^B Q_{34} \zeta_{12-3-466\|Q\kappa^A\kappa^B} \langle 1|4\rangle \langle -6|-3\rangle - \zeta_{123466\|Q\kappa^A\kappa^B} \langle 1|-3\rangle \langle -6|4\rangle] \\ &+ \varepsilon_1 \varepsilon_2 \varepsilon_6 [\kappa^B Q_{34} \zeta_{12-3-4-5-566\|Q\kappa^A\kappa^B} \langle -5|4\rangle \langle -6|-3\rangle - \zeta_{1234-5-566\|Q\kappa^A\kappa^B} \langle -5|-3\rangle \langle -6|4\rangle]\}.\end{aligned} \tag{D7}$$

These expressions satisfy the symmetry relations (7.9)–(7.12) as required.

The sum of amplitudes for γ- and Z_0-exchanges is obtained by the application of an operator for the correct permutation of κ^A,κ^B, and the assignment of the mass λ according to (A42) and (A44). This operator is evidently

$$\{e|_{\lambda=0} + (a + b \backslash \kappa^A/)(a + b \backslash \kappa^B/)|_{\lambda=m_{Z_0}}\}. \qquad \text{(D8)}$$

[1]J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Addison–Wesley, Reading, MA, 1959), 2nd ed.

[2]T. B. Anders and W. Jachmann, J. Math. Phys. **24**, 2847 (1983).

[3]W. Jachmann, internal report, University of Munich, 1980.

[4]According to our definition, for linearly polarized photons moving in the z'-direction of the local helicity frame[2] and having a real polarization vector in the positive x'-direction, the Stokes vector points in the negative x'-direction. This convention is useful to show the analogy of Stokes vectors with the spin vectors of spin-$\frac{1}{2}$ particles.

[5]H. A. Tolhoek, Rev. Mod. Phys. **28**, 227 (1956).

[6]G. G. Ohlsen, Rep. Prog. Phys. **35**, 717 (1972).

[7]B. A. Robson, *The Theory of Polarization Phenomena* (Clarendon, Oxford, 1974).

[8]A. Lindner, *Drehimpulse in der QM* (Teubner, Stuttgart, 1984).

[9]D. R. Yennie, S. C. Frautschi, and H. Suura, Ann. Phys. **13**, 379 (1961); K. E. Eriksson, Nuovo Cimento **19**, 1010 (1961).

[10]Equation I (4.2) contains a misprint. The correct relation is

$$\langle N_1,p_1|\sigma^{\nu_1}{}_{\nu_2}\cdots|N_2,p_2\rangle = N_1N_2\langle -N_2,p_2|\sigma\cdots{}^{\nu_2}{}_{\nu_1}|-N_1p_1\rangle\,.$$

We notice that the Eq. I (1.9) should show the factor $1/4m$, not $Q/4m$. Moreover, in Eq. I (3.2) each first factor $\alpha^{Q'}_{\pm}$ should be corrected into $\alpha''^{Q'}_{\pm}$.

[11]Equation (4.15) can also be found in paper I as I (4.3).

[12]With the usual matrix $C = i\gamma^2\gamma^0$ for the charge conjugation we find $u_c = C(\bar{u})^T = (i\gamma^2\gamma^0)(u^+\gamma^0)^T = i\gamma^2 u^* = i\gamma^2|{*}/u$. The usual operator for the time reversal is $\mathbf{T} = i\gamma^1\gamma^3|{*}/ = \gamma^0\gamma_5\gamma^2|{*}/$, for the parity transformation $\mathbf{P} = \gamma^0$, and for the helicity change $\mathbf{H} = i\mathbf{PT}$.

[13]We find $\mathbf{C}u^Q = u^{-Q}$. The complex conjugation $|{*}/$ in $\mathbf{C}$ causes additional phases N in the coefficients α^Q_N of the expansion (2.3) according to (2.11). These phases cancel the factors N in the charge conjugated helicity spinors $\mathbf{C}\,u^Q_N = Nu^{-Q}_N$.

[14]The overall complex conjugation $\backslash{*}/$ applied in the present context does not change possible imaginary terms of the propagator denominators.

[15] A. D. Martin and T. D. Spearman, *Elementary Particle Theory* (North–Holland, Amsterdam, 1970), p. 319ff.

[16]A. O. Barut, *The Theory of the Scattering Matrix* (Macmillan, New York, 1967), p. 79.

[17]S. Gasiorowicz, *Elementary Particle Physics* (Wiley, New York, 1966), p. 342.

[18]H. Salecker (private communication).

[19]W. Jachmann, "Calculations of helicity amplitudes in QED," internal report, University of Munich, 1982.

[20]We notice a further interesting formula for the contraction of two general vector currents for arbitrary spin states $u^{Q_j} = u^{Q_j}(p_j,s_j)$:

$$\bar{u}^{Q_1}\gamma_\mu u^{Q_2}\bar{u}^{Q_3}\gamma^\mu u^{Q_4} = \bar{u}^{Q_1}u^{Q_4}\bar{u}^{Q_3}u^{Q_2} - \bar{u}^{Q_1}\gamma_5 u^{Q_4}\bar{u}^{Q_3}\gamma_5 u^{Q_2}$$
$$- \bar{u}^{Q_1}u^{-Q_3}\bar{u}^{-Q_2}u^{Q_4} + \bar{u}^{Q_1}\gamma_5 u^{-Q_3}\bar{u}^{-Q_2}\gamma_5 u^{Q_4}.$$

This formula is derived from (A16) with the relations (2.3), (2.11), and (4.24).

[21]The formula (A22') was derived first by F. Ch. Simm, "QED tests with polarized Compton scattering," Diplomarbeit, University of Munich, December 1984, Eq. (B17).

[22]R. v. Mellenthin (private communication).

The O(3,1) symmetry problem of the charge–monopole interaction

L. Gy. Fehér
Bolyai Institute, H-6720 Szeged, Hungary

(Received 5 May 1986; accepted for publication 6 August 1986)

The question of whether there exists a smooth, time-independent conserved vector observable in the classical mechanics of the charge–monopole interaction that spans an O(3,1) Lie algebra together with the angular momentum or not is examined. It turns out that any candidate for such dynamical symmetry algebra has hard singularities at that part of the phase space that corresponds to charge–monopole collisions. In the course of the investigation we use the transformation of Boulware *et al.* [D. G. Boulware, L. S. Brown, R. N. Cahn, S. D. Ellis, and C. Lee, Phys. Rev. D **14**, 2708 (1976)] relating the charge–monopole system to a point mass moving in an inverse square potential. This transformation is shown to be a complete isomorphism between the scattering parts of the related Hamiltonian systems; its global behavior is described in terms of an U(1) principal fiber bundle of nontrivial topology. Several remarks on the symmetries of various monopole problems are made, e.g., the most general O(4) symmetry algebra is given for a special form of the charge–dyon interaction.

I. INTRODUCTION

There has been a renewed interest[1–8] in the symmetries of the electric-charge–Dirac-monopole interaction. Jackiw[1] treated the quantum mechanics of this problem on the basis of a time-dependent $O(2,1)\times SO(3)$ symmetry and he asked about the possible existence of a manifestly O(3,1) invariant formalism that would facilitate the computations and deepen the analogy to the Coulomb–Kepler scattering.[9,10] The existence of an O(3,1) symmetry in the charge–monopole quantum mechanics, which is properly analogous to that of the Coulomb–Kepler problem, would imply that there should already exist a time-independent O(3,1) invariance algebra in the classical mechanical version of the problem. The investigation of this latter existence question is the main subject of this paper and it will be answered in the negative in Sec. III. Golo[4] announced an O(3,1) algebra of conserved observables. We will see that this, and in fact any time-independent O(3,1) completion of the rotational symmetry algebra, has hard singularities at that part of the phase space that corresponds to charge–monopole collisions. In our investigation we use a method[11] originally developed for looking for dynamical symmetries of spherically symmetric potentials. This is possible because there is a transformation[12,13] relating the charge–monpole interaction to a point mass moving in an inverse square potential. For further use and for its own interest, in Sec. II we analyze the properties of this transformation in detail. It is an isomorphism between some restricted Hamiltonian systems and its nontrivial global topology will be unfolded, too. In the concluding final section, among other things, we give the most general O(4) symmetry algebra for a special form of the charge–dyon interaction.[14,15]

II. THE RELATION BETWEEN THE CHARGE–MONOPOLE SYSTEM AND A POINT MASS IN AN INVERSE SQUARE POTENTIAL

The classical mechanics of a nonrelativistic point charge in the field of a Dirac monopole is described here by the Hamiltonian system (M_D, W_D, H_D). The phase space for the particle, M_D, is identified with the set of pairs $(\mathbf{r},\mathbf{p}) = (r_i,p_i)$ with $r\neq 0$. The phase space M_D carries the symplectic form

$$W_D = -d\theta_D - e\mathscr{F}, \tag{2.1}$$

where $\theta_D = p_i\, dr^i$ is the Cartan form and

$$\mathscr{F} = \tfrac{1}{2}\mathscr{F}_{ij}\, dr^i\wedge dr^j, \quad \mathscr{F}_{ij} = g\epsilon_{ijk}r^k/r^3 \tag{2.2}$$

is the field strength tensor of a monopole of magnetic charge g. Our test particle, whose electric charge is e and mass is m, moves along the integral curves of the Hamiltonian vector field of

$$H_D = (1/2m)\|\mathbf{p}\|^2 . \tag{2.3}$$

The Poisson bracket $\{\ ,\ \}_D$ of two functions on the phase space is defined by W_D (see, e.g., Refs. 16 and 17): $\{f,g\}_D = W_D(X_f,X_g)$, with the Hamiltonian vector fields X_f, X_g. For the coordinate functions, for example,

$$\{r_i,r_j\}_D = 0, \quad \{r_i,p_j\}_D = \delta_{ij}, \quad \{p_i,p_j\}_D = e\mathscr{F}_{ij}.$$

There exists an interesting relation[12,13] between the charge–monopole system and a particle in an attractive inverse square potential. The phase space of this latter, denoted by M_p, is again identified with $(\mathbb{R}^3\setminus\{0\})\times\mathbb{R}^3$, whose points we call $(\mathbf{R},\mathbf{P})$ in this case. Its symplectic form is $W_p = -d\theta_p = -d(P_j\, dR^j)$ and the Hamiltonian is

$$H_p = \frac{1}{2m}\|\mathbf{P}\|^2 - \frac{e^2g^2}{2m}\frac{1}{R^2}. \tag{2.4}$$

To W_p is associated the Poisson bracket $\{\ ,\ \}_p$ giving for the coordinate functions $\{R_i,R_j\}_p = 0$, $\{P_i,P_j\}_p = 0$, $\{R_i,P_j\}_p = \delta_{ij}$. Both systems are spherically symmetric; $\mathbf{J} = \mathbf{r}\times\mathbf{p} - eg\mathbf{r}/r$ and $\mathbf{L} = \mathbf{R}\times\mathbf{P}$ are the corresponding angular momenta. Here $\mathbf{J}$ and $\mathbf{L}$ span an SO(3) Lie algebra with respect to $\{\ ,\ \}_D$ and $\{\ ,\ \}_p$, respectively. Observe, that the two systems (M_D,W_D,H_D) and (M_p,W_p,H_p) cannot be globally equivalent since W_p is an exact two-form while W_D is not exact because $\mathscr{F}$ describes a magnetic monopole.

 0022-2488/87/010234-06$02.50

Let us define

$$M_{\mathrm{D}}^{0}=\{(\mathbf{r},\mathbf{p})\in M_{\mathrm{D}}\mid \|\mathbf{J}(\mathbf{r},\mathbf{p})\|^{2}>e^{2}g^{2}\} \tag{2.5a}$$

and

$$M_{p}^{0}=\{(\mathbf{R},\mathbf{P})\in M_{p}\mid \|\mathbf{L}(\mathbf{R},\mathbf{P})\|^{2}>e^{2}g^{2}\}, \tag{2.5b}$$

where M_{D}^{0} is a smooth open submanifold in M_{D} that is invariant with respect to the flow of H_{D}. Therefore we can introduce the restricted Hamiltonian system $(M_{\mathrm{D}}^{0}, W_{\mathrm{D}}^{0}, H_{\mathrm{D}}^{0})$, where $W_{\mathrm{D}}^{0}=W_{\mathrm{D}|M_{\mathrm{D}}^{0}}$ and $H_{\mathrm{D}}^{0}=H_{\mathrm{D}|M_{\mathrm{D}}^{0}}$. Analogously, we will consider the system $(M_{p}^{0}, W_{p}^{0}, H_{p}^{0})$. Note that M_{D}^{0} and M_{p}^{0}, respectively, contain the scattering trajectories. Now we proceed to show that the above two "restricted Hamiltonian systems" can be canonically identified. To see this define the map λ: $M_{\mathrm{D}}^{0}\to M_{p}^{0}$ by

$$R_{i}\circ\lambda(\mathbf{r},\mathbf{p})=\frac{J}{l}\left[r_{i}+\frac{eg}{J^{2}}rJ_{i}\right], \tag{2.6a}$$

$$P_{i}\circ\lambda(\mathbf{r},\mathbf{p})=\frac{J}{l}\left[p_{i}+\frac{eg}{J^{2}}\frac{\langle\mathbf{r},\mathbf{p}\rangle}{r}J_{i}\right], \tag{2.6b}$$

where $J=\|\mathbf{J}\|$, $l=\|\mathbf{l}\|=\|\mathbf{r}\times\mathbf{p}\|$. The relation $\langle\mathbf{J}, \mathbf{r}/r\rangle = -eg$ implies that, for fixed $\mathbf{J}$, the trajectories of the particle interacting with the monopole lie on a cone whose axis is $\mathbf{J}$. Geometrically, (2.6a) means that one rotates $\mathbf{r}$ into the plane perpendicular to $\mathbf{J}$. Hereby the trajectories are carried into those of the potential $-e^{2}g^{2}/(2mR^{2})$, this is why Boulware *et al.*[12] introduced the transformation. Anyway, the background potential seems simpler than the field of the monopole. It was also observed in Ref. 12 that (2.6a) used together with the relations $\mathbf{p}(t)=m\dot{\mathbf{r}}(t)$, $\mathbf{P}(t)=m\dot{\mathbf{R}}(t)$ converts H_{D} and $\mathbf{J}$ into H_{p} and $\mathbf{L}$.

As a matter of fact (2.6) gives us the only transformation completing (2.6a) in a way that

$$H_{p}(\lambda(\mathbf{r},\mathbf{p}))=H_{\mathrm{D}}(\mathbf{r},\mathbf{p}),\quad \mathbf{L}(\lambda(\mathbf{r},\mathbf{p}))=\mathbf{J}(\mathbf{r},\mathbf{p}) \tag{2.7}$$

hold on M_{D}^{0}. Here we are interested in the symplectic nature and global properties of the transformation (2.6). First, λ is a global diffeomorphism from M_{D}^{0} to M_{p}^{0}, its inverse σ: $M_{p}^{0}\to M_{\mathrm{D}}^{0}$ is described in our coordinates by the equations

$$r_{i}\circ\sigma(\mathbf{R},\mathbf{P})=\left[1-\frac{e^{2}g^{2}}{L^{2}}\right]^{1/2}R_{i}-\frac{eg}{L^{2}}RL_{i}, \tag{2.8a}$$

$$p_{i}\circ\sigma(\mathbf{R},\mathbf{P})=\left[1-\frac{e^{2}g^{2}}{L^{2}}\right]^{1/2}P_{i}-\frac{eg}{L^{2}}\frac{\langle\mathbf{R},\mathbf{P}\rangle}{R}L_{i}, \tag{2.8b}$$

with $L=\|\mathbf{R}\times\mathbf{P}\|$. The point is that actually λ is a symplectomorphism, $\lambda^{*}(W_{p}^{0})=W_{\mathrm{D}}^{0}$. This is equivalent to

$$\{f\circ\lambda,h\circ\lambda\}_{\mathrm{D}}=\{f,h\}_{p}\circ\lambda \tag{2.9}$$

for any f, $h\in C^{\infty}(M_{p}^{0}$.) It is enough to prove (2.9) for the coordinate functions, then

$$\begin{aligned}&\{R_{i}\circ\lambda,R_{j}\circ\lambda\}_{\mathrm{D}}=\{R_{i},R_{j}\}_{p}\circ\lambda=0,\\&\{R_{i}\circ\lambda,P_{j}\circ\lambda\}_{\mathrm{D}}=\{R_{i},P_{j}\}_{p}\circ\lambda=\delta_{ij},\\&\{P_{i}\circ\lambda,P_{j}\circ\lambda\}_{\mathrm{D}}=\{P_{i},P_{j}\}_{p}\circ\lambda=0\end{aligned} \tag{2.10}$$

are verified by explicit calculations. Taking into account (2.7) we summarize the above results in the following theorem.

Thereom 1: λ is a diffeomorphism from M_{D}^{0} to M_{p}^{0} that carries the Hamiltonian system $(M_{\mathrm{D}}^{0}, W_{\mathrm{D}}^{0}, H_{\mathrm{D}}^{0})$ into $(M_{p}^{0}, W_{p}^{0}, H_{p}^{0})$ and commutes with the action of the rotational symmetry group SO(3) of the related systems. (Of course, the inverse transformation $\sigma=\lambda^{-1}$ is the same in the reverse direction.)

As a second step, now we investigate the behavior of our maps on the boundaries of the open submanifolds M_{D}^{0} and M_{p}^{0}. The closure of M_{D}^{0} is M_{D}, which can be decomposed as $M_{\mathrm{D}}^{0}\cup B_{\mathrm{D}}$, where

$$B_{\mathrm{D}}=\{(\mathbf{r},\mathbf{p})\in M_{\mathrm{D}}\mid\mathbf{r}\times\mathbf{p}=0\}. \tag{2.11}$$

The "boundary" B_{D} is a line bundle over $\mathbb{R}^{3}\setminus\{0\}$ because (2.11) tells us that for any fixed $\mathbf{r}$, $\mathbf{p}$ varies on the line parallel to $\mathbf{r}$. On the other hand, the closure of M_{p}^{0} in M_{p} can be decomposed as $\overline{M}_{p}^{0}=M_{p}^{0}\cup B_{p}$ with the boundary B_{p} given as

$$B_{p}=\{(\mathbf{R},\mathbf{P})\in M_{p}\mid\|\mathbf{R}\times\mathbf{P}\|^{2}=e^{2}g^{2}\}. \tag{2.12}$$

Here B_{p} is a "cylinder bundle" over $\mathbb{R}^{3}\setminus\{0\}$, for any fixed $\mathbf{R}$, $\mathbf{P}$ can vary on the surface of an infinite rotation cylinder with axis $\mathbf{R}$ and radius $|eg|/R$. From these we see immediately that B_{p} cannot be diffeomorphic to B_{D} since $\dim B_{p}=5$ while $\dim B_{\mathrm{D}}=4$. Notice, however, that the formula (2.8) is well defined also on the boundary, so we obtain a unique extension denoted also by σ, which maps $M_{p}^{0}\cup B_{p}$ onto $M_{\mathrm{D}}^{0}\cup B_{\mathrm{D}}$. The relations $H_{p}=H_{\mathrm{D}}\circ\sigma$, $\mathbf{L}=\mathbf{J}\circ\sigma$ remain valid for the extended map. Explicitly, on the boundary $(\mathbf{r}, \mathbf{p})=\sigma(\mathbf{R}, \mathbf{P})$ is given by

$$\mathbf{r}=-(R/eg)\mathbf{L},\quad \mathbf{p}=-(\langle\mathbf{R},\mathbf{P}\rangle/eqR)\mathbf{L}, \tag{2.13}$$

because of (2.8) and (2.12). Our first result in the characterization of σ is the following theorem.

Theorem 2: $B_{p}\xrightarrow{(\sigma)}B_{\mathrm{D}}$ is a U(1) principal fiber bundle with projection σ described in (2.13).

From their pictures as cylinder and line bundles over $\mathbb{R}^{3}\setminus\{0\}$ we see that both B_{p} and B_{D} are smooth manifolds, and closed submanifolds in M_{p} and M_{D}, respectively. On account of (2.13), σ: $B_{p}\to B_{\mathrm{D}}$ is a smooth surjective map. There is a natural action of U(1) on B_{p}, which makes it a principal bundle with projection σ. Let $(\mathbf{R}, \mathbf{P})\in B_{p}$ and $g(\alpha)=e^{i\alpha}\in\mathrm{U}(1)$. The vectors $\mathbf{R}$ and $\mathbf{P}$ lie in the plane Σ, which is perpendicular to $\mathbf{L}=\mathbf{R}\times\mathbf{P}$ and contains the origin. Now denote by $\mathbf{R}_{\alpha}$ and $\mathbf{P}_{\alpha}$ the vectors in Σ obtained by rotating, respectively, $\mathbf{R}$ and $\mathbf{P}$ around $\mathbf{L}$ with angle α in a counterclockwise direction. The formula

$$R_{g(\alpha)}[(\mathbf{R},\mathbf{P})]=(\mathbf{R}_{\alpha},\mathbf{P}_{\alpha}) \tag{2.14}$$

provides us with a free right action of U(1) on B_{p} for which only the identity $g(\alpha)=e^{i\alpha}=1$ has any fixed point. It follows easily from (2.13) and (2.14) that $\sigma\circ R_{g(\alpha)}=\sigma$, and that each $(\mathbf{R}, \mathbf{P})$ with fixed image under σ belongs to a unique orbit of U(1) in B_{p}. These facts together prove Theorem 2.

Now we clarify the structure of the bundle $B_{p}\xrightarrow{(\sigma)}B_{\mathrm{D}}$. Let B_{p}^{0} be the part of B_{p} in which $\langle\mathbf{R}, \mathbf{P}\rangle=0$. Analogously, we denote by B_{D}^{0} the $\langle\mathbf{r}, \mathbf{p}\rangle=0$ (and then $\mathbf{p}=0$) part of B_{D}. It is convenient to consider first the $B_{p}^{0}\to B_{\mathrm{D}}^{0}$ subbundle of $B_{p}\to B_{\mathrm{D}}$. We remark that the integral curves of the $1/R^{2}$

potential in B_p^0 describe uniform circular motions and σ projects any trajectory of circular motion onto a certain phase point in $B_{\rm D}^0$ representing the charge at rest in the monopole's field. Using the obvious identification of $B_{\rm D}^0$ with $\mathbb{R}^3\setminus\{0\}$ and that of M_p with pairs $(\mathbf{R}, \mathbf{P})$, B_p^0 can be regarded as

$$B_p^0 = \left\{(\mathbf{r};\mathbf{R},\mathbf{P})\,\middle|\,\begin{matrix}\|\mathbf{r}\|\neq 0,\ \|\mathbf{R}\| = \|\mathbf{r}\|,\ \langle\mathbf{R},\mathbf{P}\rangle = 0\\ \mathbf{R}\times\mathbf{P} = -eg\mathbf{r}/r\end{matrix}\right\}, \quad (2.15)$$

as a consequence of (2.13) with $\langle\mathbf{R}, \mathbf{P}\rangle = 0$. Then the projection $B_p^0 \to B_{\rm D}^0$ reads as $(\mathbf{r};\mathbf{R},\mathbf{P})\to\mathbf{r}$. Since $\mathbb{R}^3\setminus\{0\} = S^2\times\mathbb{R}^+$, the topology of the bundle $B_p^0\to B_{\rm D}^0$ is determined by its restriction to S^2. On the other hand, for $\|\mathbf{r}\| = 1$, $(\mathbf{r};\ \mathbf{R},\ \mathbf{P})\to(\mathbf{r};\ \hat{A},\ \hat{B})$ [where $\hat{A} = \mathbf{R}$, $\hat{B} = \mathbf{P}/(-eg)$, $\|\hat{A}\| = \|\hat{B}\| = 1$] maps this restriction isomorphically onto the bundle of oriented orthonormal frames of S^2, which is $\mathrm{SO}(3)\cong\mathrm{SU}(2)/Z_2$. To sum it up, we have proved the following theorem.

Theorem 3: The restriction of the bundle $B_p^0\to B_{\rm D}^0$ to S^2 is isomorphic to the Hopf bundle $\mathrm{SO}(3)\to S^2$.

It seems quite accidental that the "boundary fibering" leads to the same bundle that belongs to a Dirac monopole of charge $2eg = 2$. At this point it is easy to clarify the structure of the whole bundle $B_p\to B_{\rm D}$. To this let us introduce a map $B_p\to B_p^0$ by projecting the $\mathbf{P}$ part of any pair $(\mathbf{R}, \mathbf{P})\in B_p$ to the plane that is perpendicular to $\mathbf{R}$ and contains the origin. This is a bundle map, it commutes with $R_{g(\alpha)}$ acting both on B_p and B_p^0, for any $g(\alpha)\in\mathrm{U}(1)$. Hence it yields a map between the corresponding bases. As it is easy to see, this latter map is just the bundle projection from the line bundle $B_{\rm D}$ to its base manifold $B_{\rm D}^0\cong\mathbb{R}^3\setminus\{0\}$ sending $(\mathbf{r}, \mathbf{p})\in B_{\rm D}$ to $\mathbf{r}$. In conclusion, the bundle $B_p\to B_{\rm D}$ can be regarded as the pullback (see, e.g., Ref. 18, p. 60) of the bundle $B_p^0\to B_{\rm D}^0$ by the above natural map from $B_{\rm D}$ onto $B_{\rm D}^0$.

III. NONEXISTENCE OF THE TIME-INDEPENDENT, GLOBAL O(3,1) SYMMETRY ALGEBRA

Our purpose in this section is to answer the question of whether there exists a smooth, time-independent, conserved vector observable of the charge–monopole system that spans an O(3,1) Lie algebra together with the angular momentum or not. To find the answer, we apply, through the use of the related inverse square potential, a method[11] originally developed for the investigation of dynamical symmetries of spherically symmetric potential problems. First we shall restrict our considerations to $M_{\rm D}^0$ and derive the most general O(3,1) algebra of the above-mentioned type there, then examine its smoothness at the boundary $B_{\rm D}$.

An arbitrary phase point $(\mathbf{R}, \mathbf{P})\in M_p^0$ determines, as an initial value, a scatterng trajectory of the particle influenced by the potential $-e^2g^2/2mR^2$. For any $(\mathbf{R}, \mathbf{P})\in M_p^0$ let $\Gamma(\mathbf{R}, \mathbf{P})$ denote the unit vector pointing to the unique turning point of the actual three-space trajectory. After a bit of calculation one finds the explicit expression

$$\Gamma(\mathbf{R},\mathbf{P}) = (\cos\tau)\hat{R} + (\sin\tau)\hat{R}\times\hat{L} \quad (3.1)$$

with

$$\tau = \tau(\mathbf{R},\mathbf{P}) = \frac{L}{\sqrt{L^2 - e^2g^2}}\arctan\frac{\langle\mathbf{R},\mathbf{P}\rangle}{\sqrt{L^2 - e^2g^2}}. \quad (3.2)$$

(Here and in the following the "hatted" symbols stand for the corresponding unit vectors.) In this way we have obtained a conserved vector, without explicit time dependence, for the restricted potential problem, that is, $\{\Gamma_i,H_p\}_p = 0$, $\{L_i,\Gamma_k\}_p = \epsilon_{ikn}\Gamma_n$ hold on M_p^0. In addition, we have the relation $\{\Gamma_i,\Gamma_k\}_p = 0$ as it is easy to check (it also follows from the general results of Ref. 11). The point is that using Γ we can write[11] the most general smooth, conserved vector function on M_p^0 as

$$\mathbf{C} = \varphi_1\Gamma + \varphi_2\hat{L}\times\Gamma + \psi\hat{L}, \quad (3.3)$$

where φ_1, φ_2, and ψ are arbitrary but smooth, time-independent scalar constants of motion, that is, C^∞ functions of H_p and L. The properties of the map λ: $M_{\rm D}^0\to M_p^0$ allow us to transfer this result from the potential to the monopole problem. So the general form of a conserved C^∞ vector observable of the restricted charge–monopole problem that does not depend explicitly on time is

$$\mathbf{D} = \Phi_1\Lambda + \Phi_2\hat{J}\times\Lambda + \Psi\hat{J}. \quad (3.4)$$

Here $\Lambda = \Gamma\circ\lambda$, $\Phi_1(H,J)$, $\Phi_2(H,J)$, and $\Psi(H,J)$ ($H = H_{\rm D}$) are optional functions that are smooth on the "half-infinite, open rectangle" $H>0$, $J>|eg|$, the inverse image of which under the map (H,J): $M_{\rm D}\to\mathbb{R}^2$ is just $M_{\rm D}^0$.

Now we proceed to single out those conserved vectors that, besides the trivially valid equations

$$\{J_i,J_k\}_{\rm D} = \epsilon_{ikn}J_n, \quad \{J_i,D_k\}_{\rm D} = \epsilon_{ikn}D_n,$$

satisfy the O(3,1) commutation relations $\{D_i,D_k\}_{\rm D} = -\epsilon_{ikn}J_n$ as well, leastways on $M_{\rm D}^0$. The prescription of these Poisson brackets restrains the choice of the conserved scalars present in (3.4) according to

$$\frac{\partial\Psi}{\partial J} + \frac{\Psi}{J} = 0, \quad \frac{\Psi^2}{J} - \frac{1}{2}\frac{\partial Q^2}{\partial J} + J = 0, \quad (3.5)$$

where $Q^2 = (\Phi_1)^2 + (\Phi_2)^2$. The general solution of this system of equations is given by

$$\Psi = A(H)/J, \quad Q^2 = J^2 - \Psi^2 - B(H), \quad (3.6)$$

where $A(H)$ and $B(H)$ are arbitrary functions smooth on the half-line $H>0$. By writing Φ_1 and Φ_2 in the form

$$\Phi_1 = Q\cos S, \quad \Phi_2 = Q\sin S \quad (Q = +\sqrt{Q^2}), \quad (3.7)$$

we introduce a new scalar constant of motion $S(H,J)$.

At this stage we have a wealthy collection of time-independent O(3,1) symmetry algebras on $M_{\rm D}^0$, with two optional C^∞ functions of the Hamiltonian $A(H)$ and $B(H)$ and with a third conserved scalar $S(H,J)$, which must be a function of the type ensuring the smoothness of Φ_1 and Φ_2. But, ultimately, we are interested in the existence of a proper O(3,1) symmetry, we want the vector $\mathbf{D}$ to be well defined and as smooth as possible on the whole of $M_{\rm D}$. As a first step in the derivation of an O(3,1) algebra smooth on $M_{\rm D}$, let us suppose that the function $\mathbf{D}$ remains continuous at the boundary $B_{\rm D}$. Decompose it as $\mathbf{D} = \mathbf{D}_\parallel + \mathbf{D}_\perp$, with $\mathbf{D}_\parallel = \langle\mathbf{D},\hat{J}\rangle\hat{J}$. It will be useful to consider simultaneously $\mathbf{C} = \mathbf{D}\circ\sigma$: $\bar{M}_p^0\to\mathbb{R}^3$, too. Here $\mathbf{C}$ is continuous on $\bar{M}_p^0$ on account of the continuity of $\sigma: \bar{M}^0_p \to M_D$. Furthermore, $\mathbf{C}$ is a vector function on $\bar{M}^0_p$:

$$\mathbf{C}(g\mathbf{R},g\mathbf{P}) = g\mathbf{C}(\mathbf{R},\mathbf{P}) \tag{3.8}$$

is valid for any $(\mathbf{R}, \mathbf{P}) \in \bar{M}^0_p$ and $g \in SO(3)$ acting on three-vectors in the usual manner. The fact that $\sigma: M^0_p \to M^0_D$ is an SO(3) equivariant diffeomorphism implies (3.8) on M^0_p and then it holds on the boundary B_p as well because of the continuity of $\mathbf{C}$. The decomposition of $\mathbf{C}$ corresponding to that of $\mathbf{D}$ is $\mathbf{C} = \mathbf{C}_\parallel + \mathbf{C}_\perp$, where $\mathbf{C}_\parallel = \mathbf{D}_\parallel \circ \sigma = \langle \mathbf{C}, \hat{L} \rangle \hat{L}$. On the other hand, by its very definition, $\mathbf{C}$, when restricted to the boundary B_p, must be invariant with respect to the bundle action of U(1) described in Theorem 2 of Sec. II. Combining this condition with (3.8) we obtain that $\mathbf{C}_\perp$ vanishes on B_p since $R_{g(\alpha)}$ [see (2.14)], for any $g(\alpha) \in U(1)$, rotates any vector around the axis $\hat{L}$. Consequently, returning to M_D, the equation

$$\|\mathbf{D}_\perp\|^2{}_{|J^2 = e^2g^2} = 0 \tag{3.9}$$

must be satisfied for any continuous vector function $\mathbf{D}$: $M_D \to \mathbb{R}^3$.

In order to get a vector that is smooth on the whole of M_D from the formula (3.4) we assume that the functions A and B introduced by (3.6) are C^∞ on the closed half-line $H \geqslant 0$, not only in its interior $H > 0$. This is a natural assumption in the light of the relations

$$\langle \mathbf{J}, \mathbf{D} \rangle = A(H), \quad J^2 - D^2 = B(H), \tag{3.10}$$

which are derived from (3.4) and (3.6). As to the scalar function $S(H, J)$ in (3.7), notice that the ansatz $S = 0$ should yield a smooth $\mathbf{D}$ on M_D if one could obtain such a $\mathbf{D}$ at all, since in that case $\mathbf{D}_\perp$ would point in the physically distinguished direction of $\mathbf{\Lambda}$ on M^0_D. So we cannot lose much by making the convenient assumption that S is smooth on the half-infinite, closed rectangle $H \geqslant 0$, $J \geqslant |eg|$. Now, translate condition (3.9) into

$$Q^2 \ (H, J = |eg|) = 0. \tag{3.11}$$

This equation together with (3.6) provides us with an algebraic constraint between $A(H)$ and $B(H)$. Explicitly, (3.6) and (3.11) give rise to

$$B(H) = e^2g^2 - A^2(H)/e^2g^2, \quad Q = lF, \tag{3.12}$$

where we introduced the notation

$$F = F(H, J) = [1 + A^2(H)/e^2g^2J^2]^{1/2} \tag{3.13}$$

and $l = (J^2 - e^2g^2)^{1/2}$. Let us observe that the requirement (3.11) is not only necessary but also enough for $\mathbf{D}$ to be continuous on the whole of M_D provided that that scalar functions present in its formuala (3.4) are continuous for $H \geqslant 0$, $J \geqslant |eg|$. Thus at this point the conserved vector under inquiry is certainly smooth on M^0_D, spans there an O(3,1) Lie algebra together with $\mathbf{J}$ and is at least continuous on the whole of M_D.

The task ahead of us now is to decide whether the optional functions that left $A(H)$ and $S(H, J)$, which are assumed to be C^∞ for $H \geqslant 0$, $J \geqslant |eg|$, can be chosen in a way ensuring the C^∞ or at least the C^1 character of $\mathbf{D}$ on the whole of M_D. Unfortunately, we have no further "criterion" at hand to restrict the form of $\mathbf{D}$, so we are forced to investigate directly the behavior of the partial derivatives of $\mathbf{D}$ at the boundary B_D. First of all, by collecting our previous equations, we rewrite $\mathbf{D}$ as

$$\begin{aligned}\mathbf{D} = \mathbf{r}\left\{ -\frac{egA}{rJ^2} + Fp_r \sin(\nu - S) + \frac{Fl^2}{rJ}\cos(\nu - S)\right\} \\ + \mathbf{r}\times\mathbf{p}\left\{\frac{A}{J^2} + eg\frac{F}{J}\cos(\nu - S)\right\} \\ - \mathbf{p}\{Fr\sin(\nu - S)\}.\end{aligned} \tag{3.14}$$

Here $p_r = \langle \mathbf{r}, \mathbf{p} \rangle / r$, $l = \|\mathbf{r}\times\mathbf{p}\|$ and, in accordance with (3.2), $\nu = \tau \circ \lambda: M^0_d \to \mathbb{R}$ is given by

$$\nu = (J/l)\arctan(rp_r/l). \tag{3.15}$$

We note that (3.14) reproduces (apart from a sign error in Ref. 4) the conserved vector of Golo when taken with the ansatz $S = 0$, $A = 0$, $F = 1$.

Turning to our problem let $(\mathbf{r}, p_r \cdot \hat{r})$ be an arbitrarily fixed phase point in B_D. Furthermore, let us consider $\mathbf{p}(l) = p_r\hat{r} + (l/r)\hat{n}$, where $\hat{n} = (n_1, n_2, n_3)$ is some fixed unit vector which is orthogonal to $\hat{r}$. For $l \neq 0$ $(\mathbf{r}, \mathbf{p}(l))$ is in M^0_D since $\|\mathbf{r}\times\mathbf{p}(l)\|^2 = l^2$, and

$$\lim_{l\to 0}(\mathbf{r}, \mathbf{p}(l)) = (\mathbf{r}, p_r\hat{r}).$$

For simplicity, in the following we denote $(\partial D_i/\partial p_k)(\mathbf{r}, \mathbf{p}(l))$ by $Y_{ik}(l)$. We are going to show that, in general, $Y_{ik}(l)$ does not approach a limit as $l \to 0$. This will prove that D_i cannot be a C^1 function on M_D. It is convenient to put $Y_{ik}(l)$ in the form

$$Y_{ik}(l) = E_{ik}(l) + G_{ik}(l)\sin(\nu - S) + K_{ik}(l)\cos(\nu - S), \tag{3.16}$$

because of the form of $\mathbf{D}$ in (3.14). By an elementary but rather tedious calculation we obtain

$$\begin{aligned} E_{ik}(l) &= \mathscr{O}(l) - \epsilon_{ikn}r^n\frac{A}{e^2g^2} - \frac{r_ir_k}{r^2eg}\frac{dA}{dH}\frac{p_r}{m}, \\ G_{ik}(l) &= \mathscr{O}(l) + G^{(0)}_{ik} + l^{-1}G^{(-1)}_{ik}, \\ K_{ik}(l) &= \mathscr{O}(l) + K^{(0)}_{ik} + l^{-1}K^{(-1)}_{ik}, \end{aligned} \tag{3.17}$$

where

$$\begin{aligned} G^{(0)}_{ik} &= F\{(r_ir_k - r^2\delta_{ik})/r + (\pi/2)(\operatorname{sgn} p_r)r_in_k\}, \\ G^{(-1)}_{ik} &= F(\pi/2)eg(\operatorname{sgn} p_r)n_k\epsilon_{ijm}r_jn_m, \\ K^{(0)}_{ik} &= -F(\operatorname{sgn} eg)\epsilon_{ikj}r_j, \\ K^{(-1)}_{ik} &= F(\pi/2)|eg|(\operatorname{sgn} p_r)rn_in_k. \end{aligned} \tag{3.18}$$

The A and dA/dH in (3.17) and F in (3.18) are taken at the arbitrarily fixed phase point $(\mathbf{r}, p_r\hat{r}) \in B_D$. Assuming that $p_r \neq 0$, we see from (3.15) that $\lim_{l\to 0}|\nu| = \infty$. Therefore both $\sin(\nu - S)$ and $\cos(\nu - S)$ oscillate as $l \to 0$. From this, from the l independence and nonvanishing of $G^{(j)}_{ik}$, $K^{(j)}_{ik}$ $(j = 0, -1)$ it follows that, in general, $Y_{ik}(l)$ really does not have any limit as $l \to 0$. In conclusion, $\mathbf{D}$ cannot be C^1 on the whole of M_D no matter how one chooses the C^∞ scalar constants of motion $A(H)$ and $S(H, J)$. Thus there is no global, time-independent O(3,1) symmetry algebra of the form $(\mathbf{J}, \mathbf{D})$ for the charge–monopole problem.

Let us make some remarks about this nonexistence statement. First, the same reasoning would have lead to the same conclusion if we had taken the conserved scalar functions appeared in the demonstration as C^1 and not necessarily C^∞ functions. Second, the statement obviously remains valid without the smoothness assumption made at $H=0$. Third, let us notice that, in principle, one could allow $S(H, J)$ in (3.7) to be a function defined only modulo 2π for which $\sin S$, $\cos S$, $\partial \sin S/\partial r^i,\ldots,\partial \cos S/\partial p^k$ are continuous on $M^0_{\rm D}$. Perhaps some kind of singularity of S at $J=|eg|$ could compensate the singularity of **D**, which arises, after all, from the singularity of v given by (3.15)? This is not the case. It can be shown (analyzing the explicit form of $\partial D_i/\partial p^k$) that $\lim_{l\to 0} Y_{ik}(l)$ cannot exist in that case either.

It is well known that for the Kepler problem the Runge–Lenz vector gives rise to smooth O(3,1) and O(4) algebras only on the positive- and negative-energy submanifolds of the phase space, respectively, and not on the whole of it. Notwithstanding, these symmetry algebras are very useful in the quantum mechanics of the Kepler problem (e.g., Refs. 9 and 10). Our "almost global" symmetry algebras (**J**, **D**), however, are quite useless for the quantum mechanics of the charge–monopole scattering. The reason is that $B_{\rm D}$, where **D** has its inevitable singularity, has nonempty intersection with every constant energy submanifold of $(M_{\rm D}, H_{\rm D})$.

IV. DISCUSSION AND CONCLUDING REMARKS

Here we would like to mention some further results connected with the transformation analyzed in Sec. II and with the symmetries of the charge–monopole interaction and to point out some problems. Let us suppose that, beyond the magnetic field of the monopole, the test particle is also influenced by an additional spherically symmetric potential, that is, the Hamiltonian reads as

$$H'_{\rm D} = (1/2m)\|\mathbf{p}\|^2 + V(r) \tag{4.1}$$

instead of the simple kinetic Hamiltonian (2.3). Obviously, the transformation described in (2.6) and (2.8) relates this problem to a pure potential one governed by the Hamiltonian

$$H'_p = \frac{1}{2m}\|\mathbf{P}\|^2 + V(R) - \frac{e^2g^2}{2m}\frac{1}{R^2}. \tag{4.2}$$

For example, by putting in (4.1)

$$V(r) = -\frac{\alpha}{r} + \frac{e^2g^2}{2m}\frac{1}{r^2} \quad (\alpha > 0), \tag{4.3}$$

one can describe the charge moving in the field of a dyon and also influenced by an extra $1/r^2$ potential, or with another interpretation of the constants α, e, g: the interaction of two dually charged particles.[14,15] The quantum mechanical version of this problem[14,15] possesses a dynamical O(4) invariance, the extra repelling potential was introduced in Ref. 14 just to get a highly symmetric system. This symmetry is understandable from the fact[13] that in this case the "trick" of Boulware *et al.*[12] leads to a Kepler problem. We have carried through an analysis to answer the question of what the most general time-independent O(4) symmetry algebra of the form (**J**, **D**) is on the negative-energy submanifold of the phase space of the system governed by $H'_{\rm D}$ with potential (4.3). This is parallel to the analysis in Sec. III and that of in Ref. 11 for the Kepler problem. Therefore we give only the result

$$\mathbf{D} = X(-2mH'_{\rm D})^{-1/2} \times\left\{(\cos S)\left[-\mathbf{p}\times\mathbf{J} + m\alpha\hat{r} - \frac{m\alpha eg}{J^2}\mathbf{J}\right] + (\sin S)\hat{J}\times[-\mathbf{p}\times\mathbf{J} + m\alpha\hat{r}] + \frac{sm\alpha eg}{J^2}\mathbf{J}\right\}. \tag{4.4}$$

Here $S=S(H'_{\rm D}, J)$ is an arbitrary smooth scalar constant of motion on the $H'_{\rm D}<0$ part of the phase space and $s=\pm 1$. From (4.4) one gets the conserved vector used in Refs. 13 and 14 by taking $S=\pi$, $s=-1$. The general expression (4.4) or its simple special cases with $S=k\pi/2$ ($k=1,2,\ldots$) seem to have escaped notice so far. It would be interesting to investigate the integrability and the quantum mechanical role of the general O(4) algebra given here. Otherwise, the potential (4.3) appears in quite a natural way[19] in the mechanics of an isospin-carrying test particle feeling the large distance field of the Prasad–Sommerfield monopole.

As a second matter, notice that the transformations λ, σ described in (2.6), (2.8) can be trivially completed to transformations between the corresponding evolutional spaces[16]: by taking the identity map for the time coordinate. In this way one can relate time-dependent properties of monopole and pure potential problems. For example, the conformal O(2,1) symmetry of the inverse square potential[20] can be translated[21] into the O(2,1) symmetry of the monopole discovered by Jackiw.[1] It is worth noting that the rotational SO(3)$\times$conformal O(2,1) symmetry of these systems represents that part of the maximal kinematical invariance or Schrödinger group[22,23] of the free particle that remains undestroyed by coupling the particle to a monopole or/and to an inverse square potential (for details see Refs. 5, 8, and 24). On the other hand, the SO(3)$\times$O(2,1) symmetry algebra of the charge–monopole problem is partly contained in a spectrum and charge-quantization generating O(4,2) algebra[2,3,15,25] containing conserved as well as nonconserved observables. The origin of this, from the viewpoint of analytical mechanics, is the fact[26] that the symplectic manifold $(M_{\rm D}, W_{\rm D})$ is canonically isomorphic to a coadjoint orbit of mass-0, spin-eg of the restricted Poincaré group that carries a natural, invariant symplectic form [cf. our Eqs. (2.1), (2.2) and Eq. (17.145) in Ref. 16, p. 309]. On this basis it is clear that the charge–monopole quantum mechanics and that of the massless spinning particles are parallel[2,3,27]; they arise from quantizations[16] on isomorphic symplectic manifolds.

In general, having a symmetry algebra for a spherically symmetric potential problem that is smooth on M^0_p and possibly depends on time, too, one can pull it through to $M^0_{\rm D}$ and derive in this manner a symmetry algebra for the related monopole problem, which is smooth on $M^0_{\rm D}$, and vice versa. But in most cases local symmetries are not of particular interest, all the symplectic manifolds of equal dimension are locally isomorphic after all. Unfortunately, there is no simple relation between the differentiability of a function $f\in C^\infty(M^0_{\rm D})\cap C(M_{\rm D})$ at $B_{\rm D}$ and that of the corresponding function $f\circ\sigma\in C^\infty(M^0_p)\cap C(M^0_p\cup B_p)$ at the boundary B_p. The point is that while σ maps the manifolds M^0_p onto $M^0_{\rm D}$ and B_p onto $B_{\rm D}$ in a C^∞ manner, the whole map σ: $M^0_p\cup B_p\to M_{\rm D}$ in only continuous but not differentiable at the boundary. This is a map from a manifold with boundary onto a manifold in the proper sense and so its differentiability at a boundary point makes sense in terms of some local differentiable extension (e.g., Ref. 28, pp. 248–251), which does not exist for σ because of the appearance of the square roots in its formula (2.8). This pathology forces a case for case smoothness investigation when one sets up a correspondence between symmetries of monopole and potential problems.

It would be of great interest to clarify the structure of the space of motions[16] for the charge–monopole scattering but we "succeeded" only in realizing that naive approaches run into difficulties. These, as our nonsmoothness statement for the O(3,1) algebras, find their origin in the charge–monopole collisions. What is really needed is some kind of regularization, perhaps analogous to that of the Kepler problem.[29]

ACKNOWLEDGMENTS

The author would like to thank Z. Horváth, P. T. Nagy, and L. Palla for their interest in the present work and for reading the manuscript. Special thanks are due to P. A. Horváthy for sending me his unpublished manuscripts and a great many other papers among the references, for stimulating discussions, and for his helpful suggestions concerning the research and the preparation of the manuscript.

[1] R. Jackiw, Ann. Phys. (NY) **129**, 183 (1980).
[2] H. Bacry, J. Phys. A: Math. Gen. **14**, L73 (1981).
[3] A. O. Barut, J. Phys. A: Math. Gen. **14**, L267 (1981).
[4] V. L. Golo, JETP Lett. **35**, 535 (1982).
[5] C. Duval, These de doctorat d'Etat, Marseille, 1982.
[6] P. A. Horváthy, Lett. Math. Phys. **7**, 353 (1983); J. Geom. Phys. **1**(3), 39 (1984).
[7] E. D'Hoker and L. Vinet, Phys. Lett. B **137**, 72 (1984); Lett. Math. Phys. **8**, 439 (1984); Commun. Math. Phys. **97**, 341 (1985); Phys. Rev. Lett. **55**, 1043 (1985).
[8] V. Hussin and S. Sinzinkayo, J. Math. Phys. **26**, 1072 (1985).
[9] M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330 (1966).
[10] M. J. Englefield, *Group Theory and the Coulomb Problem* (Wiley, New York, 1972).
[11] H. Bacry, H. Ruegg, and J. M. Souriau, Commun. Math. Phys. **3**, 323 (1966).
[12] D. G. Boulware, L. S. Brown, R. N. Cahn, S. D. Ellis, and C. Lee, Phys. Rev. D **14**, 2708 (1976).
[13] J. F. Schonfeld, J. Math. Phys. **21**, 2528 (1980).
[14] D. Zwanziger, Phys. Rev. **176**, 1480 (1968).
[15] A. O. Barut and G. L. Bornzin, J. Math. Phys. **12**, 841 (1971).
[16] J. M. Souriau, *Structure des Systemes Dynamiques* (Dunod, Paris, 1970).
[17] R. Abraham and J. Marsden, *Foundations of Mechanics* (Benjamin, London, 1978).
[18] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. 1.
[19] L. Gy. Fehér, J. Phys. A: Math. Gen. **19**, 1259 (1986).
[20] V. de Alfaro, S. Fubini, and G. Furlan, Nuovo Cimento A **34**, 569 (1976).
[21] P. A. Horváthy (private communication).
[22] U. Niederer, Helv. Phys. Acta **45**, 802 (1972).
[23] C. R. Hagen, Phys. Rev. D **5**, 377 (1972).
[24] U. Niederer, Helv. Phys. Acta **47**, 167 (1974).
[25] A. O. Barut and R. Wilson, Ann. Phys. (NY) **164**, 223 (1985).
[26] C. Duval and P. A. Horváthy (unpublished).
[27] A. O. Barut and A. J. Bracken, Lett. Math. Phys. **7**, 407 (1983).
[28] W. M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry* (Academic, New York, 1975).
[29] J. M. Souriau, Sympos, Math. **14**, 343 (1974); "Geometrie globale du probleme a deux corps," Marseille Preprint, CPT-82 1434, 1982.

The Riemann–Hilbert problem for the supersymmetric constraint equations

J. Tafel[a),b)]
Physikalisches Institut der Universität Bonn, 5300 Bonn 1, Federal Republic of Germany

(Received 28 March 1986; accepted for publication 11 July 1986)

Different formulations of the Riemann–Hilbert problem related to the constraint equations in supersymmetric Yang–Mills theories are discussed.

I. INTRODUCTION

The supersymmetric constraint equations[1] yield a superfield formulation of N-extended supersymmetric gauge theories (see Ref. 2 for a review). For $N \geqslant 3$ these equations put the theory on shell.[3–6] Thus they can be used for describing classical solutions of the $N = 3,4$ supersymmetric Yang–Mills equations. This is especially important since the constraint equations possess some features of the completely integrable systems.[4,7,8] In this paper we discuss possible formulations of the related Riemann–Hilbert problem. In our opinion the formulation with two spectral parameters is the most suitable one.

We consider the Yang–Mills superpotential $\mathscr{A}$ with the components $\mathscr{A}_\mu$ (even), $\mathscr{A}^i_A$ (odd), $\mathscr{A}_{\dot{A}i}$ (odd) depending on x^μ, θ^A_i, $\bar{\theta}^{\dot{A}i}$, where x^μ are the Cartesian coordinates of Minkowski space and θ^A_i, $\bar{\theta}^{\dot{A}i}$ are anticommuting variables, $A, \dot{A} = 1,2$ being the spinor indices and $i = 1,\dots,N$. The components of $\mathscr{A}$ are superpotentials with values in the Lie algebra of the gauge group $\mathscr{G} \subset \mathrm{GL}(n,C)$. They define the following covariant derivatives:

$$\mathscr{D}_\mu = \partial_\mu + \mathscr{A}_\mu, \quad \mathscr{D}^i_A = D^i_A + \mathscr{A}^i_A,$$
$$\mathscr{D}_{\dot{A}i} = \bar{D}_{\dot{A}i} + \mathscr{A}_{\dot{A}i},$$

where

$$D^i_A = \frac{\partial}{\partial \theta^A_i} + i\bar{\theta}^{\dot{B}i}\partial_{A\dot{B}}, \quad \bar{D}_{\dot{A}i} = -\frac{\partial}{\partial \bar{\theta}^{\dot{A}i}} - i\theta^B_i \partial_{B\dot{A}},$$
$$\partial_\mu = \frac{\partial}{\partial x^\mu}, \quad \partial_{A\dot{B}} = \sigma^\mu{}_{A\dot{B}}\,\partial_\mu$$

and σ^μ are the Pauli matrices. The constraint equations of Grimm, Sohnius, and Wess[1] impose some restrictions on the curvature of $\mathscr{A}$, namely,

$$F^{ij}_{(AB)} = D^i_{(A}\mathscr{A}^j_{B)} + D^j_{(B}\mathscr{A}^i_{A)} + \{\mathscr{A}^i_{(A}, \mathscr{A}^j_{B)}\} = 0, \tag{1a}$$

$$F_{(\dot{A}i\dot{B})j} = \bar{D}_{(\dot{A}i}\mathscr{A}_{\dot{B})j} + \bar{D}_{(\dot{B}j}\mathscr{A}_{\dot{A})i} + \{\mathscr{A}_{(\dot{A}i}, \mathscr{A}_{\dot{B})j}\} = 0, \tag{1b}$$

$$F^i_{A\dot{B}j} = D^i_A\mathscr{A}_{\dot{B}j} + \bar{D}_{\dot{B}j}\mathscr{A}^i_A + \{\mathscr{A}^i_A, \mathscr{A}_{\dot{B}j}\} + 2i\,\delta^i_j\mathscr{A}_{A\dot{B}} = 0. \tag{1c}$$

It was shown by Volovich[7] that Eqs. (1) are satisfied if and only if the system of linear equations for a matrix superfield ψ,

$$z^A\mathscr{D}^i_A\psi = 0, \tag{2a}$$

$$w^{\dot{A}}\mathscr{D}_{\dot{A}i}\psi = 0, \tag{2b}$$

$$z^A w^{\dot{B}}\mathscr{D}_{A\dot{B}}\psi = 0, \tag{2c}$$

possesses a nonsingular (i.e., ψ^{-1} exists) solution for arbitrary complex parameters z^A and $w^{\dot{A}}$. Since the equations (2) are invariant under scaling of z^A and $w^{\dot{A}}$ it is sufficient to assume that

$$z^A = (1,\lambda), \quad w^{\dot{A}} = (1,\xi),$$

where

$$\lambda,\xi \in C \cup \infty.$$

II. THE RIEMANN–HILBERT PROBLEM WITH ONE PARAMETER

A common property of many completely integrable equations is the existence of a linear system [analogous to (2)], which can be reduced to the Riemann–Hilbert problem in one complex variable.[9] Solutions of the original equations can then be obtained using the Atiyah–Ward ansatz[10,11] and its generalizations[12,13] or soliton generating transformations of Zakharov and Shabat[14] (see also Ref. 9). The question arises whether we can assume $\xi = \xi(\lambda)$ and relate (2) to the Riemann–Hilbert problem with respect to λ. In most papers on the integrability of the constraint equations it is assumed that $\xi = \lambda^2$ (and often a part of the equation is solved, but it does not change the following discussion). Then the equations (1) remain exactly the integrability conditions for (2). If $\mathscr{A}$ satisfies (1) and ψ_0, ψ_1 are nonsingular solutions of (2), analytic in λ for $|\lambda| < 1 + \epsilon$ (where ϵ is a positive constant) and $1 - \epsilon < |\lambda| \leqslant \infty$, respectively, then

$$G = \psi_0^{-1}\psi_1 \tag{3}$$

is analytic in a neighborhood of the circle $|\lambda| = 1$ and satisfies

$$z^A D^i_A G = 0, \quad w^{\dot{A}}\bar{D}_{\dot{A}i}G = 0, \quad z^A w^{\dot{B}}\partial_{A\dot{B}}G = 0. \tag{4}$$

On the other hand, for a given G that satisfies (4) and is decomposable according to (3), we obtain, in particular, that $w^{\dot{A}}(\bar{D}_{\dot{A}i}\psi_1)\psi_1^{-1}$ has, for $|\lambda| < 1$, the analytic continuation $w^{\dot{A}}(\bar{D}_{\dot{A}i}\psi_0)\psi_0^{-1}$. Therefore it can have at most a (second-order) pole at $\lambda = \infty$, i.e.,

$$w^{\dot{A}}(\bar{D}_{\dot{A}i}\psi_1)\psi_1^{-1} = -\mathscr{A}_{\dot{1}i} + \lambda\mathscr{B}_i - \lambda^2\mathscr{A}_{\dot{2}i}$$
$$= -w^{\dot{A}}\mathscr{A}_{\dot{A}i} + \lambda\mathscr{B}_i,$$

where $\mathscr{A}_{\dot{A}i}$, $\mathscr{B}_i$ are arbitrary coefficients. Since in general the $\mathscr{B}_i$ do not vanish the Riemann–Hilbert splitting (3) do not imply (2b) and, consequently, do not yield a solution of the constraint equations (1) (in contrast to the results of Chau *et al.*[15]). The equations for $\mathscr{A}$ and $\mathscr{B}_i$, which follow from (4), were considered by Aref 'eva and Volovich.[16]

[a)] Alexander von Humboldt Fellow.
[b)] On leave from Institute of Theoretical Physics, University of Warsaw, Warsaw, Poland.

 0022-2488/87/010240-04$02.50

They are nonrelativistic and they do not imply the supersymmetric Yang–Mills equations.

There is no problem with additional fields resulting from (3) if $\xi = \lambda$ or $\xi = 1/\lambda$, but then the integrability conditions for (2) are weaker then the constraint equations. We can improve the situation by imposing specific boundary conditions on ψ_0 and ψ_1, e.g., for the gauge group GL(n,C)

$$D^i_1\psi_0(0) = \overline{D}_{2i}\psi_1(\infty) = 0, \quad \text{if } \xi = \lambda, \tag{5a}$$

$$D^i_1\psi_0(0) = \overline{D}_{1i}\psi_1(\infty) = 0, \quad \text{if } \xi = 1/\lambda. \tag{5b}$$

Unfortunately, because of these extra conditions we are unable to apply known solution generating techniques to (3).

The assumption $\xi = \lambda^2$ was also exploited in searching for the Bäcklund transformation between solutions of (1). Unfortunately the transformation given in Ref. 15 leads to fields, which do not satisfy (1), and the one obtained by Devchand[8] is appropriate only for the linearized equations.

III. THE RIEMANN-HILBERT PROBLEM WITH TWO PARAMETERS

In this and the next section we assume that the parameters λ and ξ are independent and take arbitrary values in $C\cup\infty$. The space of pairs (λ,ξ) is isomorphic to $CP^1\times CP^1$ and can be covered by four overlapping regions U_a $(a=0,1,2,3)$ such that $|\lambda|, |\xi| < 1+\epsilon$, for $(\lambda,\xi)\in U_0$, $|\lambda| > 1-\epsilon$ and $|\xi| < 1+\epsilon$, for $(\lambda,\xi)\in U_1$, $|\lambda| < 1+\epsilon$ and $|\xi| > 1-\epsilon$, for $(\lambda,\xi)\in U_2$ and $|\lambda|, |\xi| > 1-\epsilon$ for $(\lambda,\xi)\in U_3$, where ϵ is a positive constant. Since Eqs. (2) are linear and for each U_a can be written in an analytic form with respect to $(\lambda,\xi)\in U_a$ it seems reasonable to assume that given a local solution $\mathscr{A}$ of (1) there are nonsingular (as matrices) solutions ψ_a of (2), which are defined locally in Minkowski space and analytic in the corresponding regions U_a. If we define in $U_a\cap U_b$ the patching function

$$G_{ab} = \psi_a^{-1}\psi_b, \tag{6}$$

then it follows from (2) that

$$z^A D^i_A G_{ab} = 0, \quad w^{\dot A}\overline{D}_{\dot A i} G_{ab} = 0, \quad z^A w^{\dot B}\partial_{A\dot B} G_{ab} = 0. \tag{7}$$

Hence each G_{ab} depends on x^μ, θ^A_i, $\bar\theta^{\dot A i}$ only through the following combinations:

$$w_{\dot B}x_-^{A\dot B}, \quad z_A x_+^{A\dot B}, \quad z_A\theta^A_i, \quad w_{\dot B}\bar\theta^{\dot B i}, \tag{8}$$

where

$$x_\pm^{A\dot B} = x^{A\dot B} \pm 2i\theta^A_k\bar\theta^{\dot B k}, \quad z_A = (-\lambda,1), \quad w_{\dot A} = (-\xi,1).$$

The functions (8) are not completely independent since

$$z_A(w_{\dot B}x_-{}^{A\dot B}) - w_{\dot B}(z_A x_+{}^{A\dot B}) = 4i(z_A\theta^A_k)(w_{\dot B}\bar\theta^{\dot B k}).$$

In the rest of this section we shall assume that GL(n,C) is the gauge group. Then it is easy to show that given a set of G_{ab} satisfying (7) and admitting the splitting (6) there are potentials $\mathscr{A}^i_A$, $\mathscr{A}_{\dot A i}$, $\mathscr{A}_{A\dot B}$ such that the equations (2) are satisfied (with $\psi = \psi_a$) and hence the potentials satisfy (1). Thus (6) and (7) describe the regular Riemann–Hilbert problem, which implies (and probably is equivalent to) the constraint equations.

The question arises whether (6) admits solution generating techniques. We are rather pessimistic concerning a generalization of the Zakharov–Shabat transformations.[14] Probably it is better to look for an ansatz for G_{ab} analogous to that of Atiyah and Ward. Following this direction let us note that due to (6) the G_{ab} satisfy the composition law $G_{ab}G_{bc} = G_{ac}$, hence $G_{ab}^{-1} = G_{ba}$ and G_{03} and G_{12} can be defined in terms of the remaining G_{ab},

$$G_{03} = G_{01}G_{13}, \quad G_{12} = G_{10}G_{02}. \tag{9}$$

Thus, G_{01}, G_{02}, G_{13}, and G_{23} can be chosen as the basic patching functions. Unfortunately they are not independent since

$$G_{01}G_{13} = G_{02}G_{23} \tag{10}$$

in $U_0\cap U_1\cap U_2\cap U_3$. The properties (7), (9), and (10) do not yet guarantee the splitting (6). To complete them to sufficient conditions one can assume

$$G_{01} = \chi_0^{-1}\chi_1, \quad G_{23} = \chi_2^{-1}\chi_3, \tag{11a}$$

$$G_{02}(\lambda_0) = \rho_0^{-1}\rho_2, \quad \partial_\lambda\rho_0 = \partial_\lambda\rho_2 = 0, \tag{11b}$$

where χ_a and ρ_a are nonsingular and analytic in U_a and G_{02} is taken at a fixed value λ_0 of λ. Then substituting (11) into (10) yields

$$\chi_0 G_{02}\chi_2^{-1} = \chi_1 G_{13}\chi_3^{-1}. \tag{12}$$

Hence $\chi_0 G_{02}\chi_2^{-1}$, as a function of λ, has an analytic continuation in the whole complex plane (including infinity) and therefore it cannot depend on λ,

$$\chi_0 G_{02}\chi_2^{-1} = Q, \quad \partial_\lambda Q = 0. \tag{13}$$

Now it follows from (12) and (13) taken at $\lambda = \lambda_0$ that

$$Q = (\chi_0(\lambda_0)\rho_0^{-1})(\rho_2\chi_2(\lambda_0)^{-1}), \tag{14}$$

where the expressions in brackets are analytic for $|\xi| < 1+\epsilon$ and $|\xi| > 1-\epsilon$, respectively. A direct consequence of (11a), (12), (13), and (14) is that the Riemann–Hilbert problem (6) is satisfied by

$$\begin{aligned}\psi_0 &= \rho_0\chi_0(\lambda_0)^{-1}\chi_0, \quad \psi_1 = \rho_0\chi_0(\lambda_0)^{-1}\chi_1,\\ \psi_2 &= \rho_2\chi_2(\lambda_0)^{-1}\chi_2, \quad \psi_3 = \rho_2\chi_2(\lambda_0)\chi_3.\end{aligned} \tag{15}$$

The corresponding gauge fields can be identified using (2) with $\psi = \psi_0$. Summarizing we can say that the Riemann–Hilbert splitting (6) exists and yields a solution of the constraint equations if and only if there exist λ, ξ-dependent matrix superfields G_{01}, G_{02}, G_{13}, G_{23} that (i) are nonsingular (in the matrix sense) and analytic with respect to λ,ξ in the corresponding regions $U_a\cap U_b$, (ii) depend on x^μ, θ^A_i, $\bar\theta^{\dot A i}$ only through the functions (8), (iii) admit the splittings (11), and (iv) satisfy the consistency condition (10).

The relevance of Eqs. (11) is based on the fact that they describe separate Riemann–Hilbert problems with respect to only one complex variable. The parameter ξ appears in (11a) as an external parameter and it is rather natural that the functions χ_a should depend on it in an analytic way since G_{01} and G_{23} do.

At the moment we are not able to give examples of G_{ab}, which depend on both parameters λ and ξ and lead to non-Abelian solutions of (1). The major difficulty is presented by Eq. (10). All other conditions can be satisfied by taking the Atiyah–Ward ansatz for G_{ab}.

If the G_{ab} do not depend on one of the parameters, say ξ, then without loss of generality one can assume

$$G_{02}=G_{13}=1,\quad G_{01}=G_{23}=G(z_Ax_+^{A\dot B},z_A\theta^A_i,\lambda)$$

and the problem (6) reduces to the splitting

$$G=\psi_0^{-1}\psi_1,\quad \partial_{\xi}\psi_0=\partial_{\xi}\psi_1=0. \tag{16}$$

The Atiyah–Ward ansatz and the Zakharov–Shabat method are both applicable in this case, e.g., for the gauge group SL(2,C) we can follow Corrigan *et al.*[11] in order to find solutions of (16). The resulting gauge fields can be called self-dual[17] since they satisfy equations that imply the self-duality of $F_{\mu\nu}$, namely,

$$F^{ij}_{(AB)}=0,\quad F_{Ai\dot Bj}=0,\quad F^{i}_{A\dot Bj}=0,\quad \text{for } N\geqslant 2, \tag{17a}$$

$$F_{AB}=0,\quad F_{\dot A\dot B}=0,\quad F_{A\dot B}=0,\quad F_{\mu\dot A}=0,\quad \text{for } N=1. \tag{17b}$$

The above equations are the integrability conditions for the linear system[18]

$$z^A\mathscr{D}^i_A\psi=0,\quad \mathscr{D}_{\dot Ai}\psi=0,\quad z^A\mathscr{D}_{A\dot B}\psi=0,$$

which follows from (2) under the assumption $\partial_{\xi}\psi=0$. A similar situation occurs when the G_{ab} do not depend on λ. Then it can be assumed that

$$G_{01}=G_{23}=1,\quad G_{02}=G_{13}=G(w_{\dot B}x_-^{A\dot B},w_{\dot A}\bar\theta^{\dot Ai},\xi).$$

The resulting gauge fields satisfy equations, obtained from (17) by the change of indices $({}^i_A)\leftrightarrow(\dot Ai)$. In the case of an Abelian gauge group the two types of solutions can be superimposed.

IV. REALITY CONDITIONS

In this section we consider the reduction of the gauge group GL(n,C) to SU(n). In terms of the gauge potentials it is described by the conditions

$$\mathscr{A}_\mu^+=-\mathscr{A}_\mu,\quad \operatorname{Tr}\mathscr{A}_\mu=0, \tag{18a}$$

$$\mathscr{A}_{\dot Ai}=-(C_{ij}\mathscr{A}^j_A)^+,\quad \operatorname{Tr}\mathscr{A}^i_A=\operatorname{Tr}\mathscr{A}_{\dot Ai}=0, \tag{18b}$$

where C_{ij} is a constant nonsingular Hermitian matrix (by a change of variables it can be reduced to a diagonal matrix with entries ± 1) and $+$ denotes transposition of matrices composed with the generalized complex conjugation, which maps θ^A_i into $C_{ij}\bar\theta^{\dot Aj}$ and reverses the order of anticommuting factors. With the above restrictions on $\mathscr{A}$ it is easy to show that if $\psi(\lambda,\xi)$ is a solution of (2), then $\det\psi$ satisfies (7) and $(\psi^+(\xi,\lambda))^{-1}$, where $\psi^+(\xi,\lambda)$ denotes $(\psi(\bar\xi,\bar\lambda))^+$, is also a solution of (2). Thus, in accordance with (2), we can assume

$$\psi(\lambda,\xi)\psi^+(\xi,\lambda)=1, \tag{19}$$

$$\det\psi=1. \tag{20}$$

Equations (2), (19), and (20) imply the constraint equations (1) and the reality conditions (18). Taking into account that in fact there are four functions ψ_a and four regions U_a we can rewrite (19), (20) in the following form:

$$\psi_0(\lambda,\xi)\psi_0^+(\xi,\lambda)=1, \tag{21a}$$

$$\psi_3(\lambda,\xi)\psi_3^+(\xi,\lambda)=1, \tag{21b}$$

$$\psi_2(\lambda,\xi)=(\psi_1^+(\xi,\lambda))^{-1}, \tag{21c}$$

$$\det\psi_a=1. \tag{22}$$

In virtue of (6) Eqs. (21) and (22) imply the following properties of the G_{ab}:

$$G_{02}(\lambda,\xi)=G^+_{10}(\xi,\lambda), \tag{23a}$$

$$G_{13}(\lambda,\xi)=G^+_{32}(\xi,\lambda), \tag{23b}$$

$$\det G_{01}=1,\quad \det G_{23}=1. \tag{24}$$

Thus G_{01} and G_{23} can be taken as the basic patching functions and the remaining G_{ab} are defined by (23) and (9). In terms of G_{01} and G_{23} the consistency condition (10) reads

$$G^+_{01}(\xi,\lambda)G_{01}(\lambda,\xi)=G_{23}(\lambda,\xi)G^+_{23}(\xi,\lambda). \tag{25}$$

Let us assume that G_{01} and G_{23} satisfy (25) and admit the splitting (11a) with $\det\chi_a=1$. If we define G_{02} and G_{13} according to (23) Eq. (11b) is satisfied by

$$\rho_0=\chi_0^+(\xi_0,\lambda_0)(\chi_0^+(\xi,\lambda_0))^{-1}, \tag{26a}$$

$$\rho_2=\chi_0^+(\xi_0,\lambda_0)(\chi_1^+(\xi,\lambda_0))^{-1}, \tag{26b}$$

where $(\lambda_0,\xi_0)\in U_0$ is a fixed point and $\chi_0^+(\xi_0,\lambda_0)$ is introduced for a more convenient normalization. Now it follows from the consideration after Eqs. (11) that $G_{ab}=\psi_a^{-1}\psi_b$, where the ψ_a are given by (15) and (26), in particular

$$\psi_0=\chi_0^+(\xi_0,\lambda_0)(\chi_0^+(\xi,\lambda_0))^{-1}\chi_0(\lambda_0,\xi)^{-1}\chi_0(\lambda,\xi). \tag{27}$$

Condition (22) is automatically satisfied and substituting $G_{ab}=\psi_a^{-1}\psi_b$ into (23) yields (in virtue of the analytic properties of ψ_a)

$$\psi_0(\lambda,\xi)\psi_0^+(\xi,\lambda)=\psi_2(\lambda,\xi)\psi_1^+(\xi,\lambda)=c, \tag{28a}$$

$$\psi_3(\lambda,\xi)\psi_3^+(\xi,\lambda)=\psi_1(\lambda,\xi)\psi_2^+(\xi,\lambda)=c^+, \tag{28b}$$

where $\partial_\lambda c=\partial_\xi c=0$. It follows from (27) that $\psi_0(\lambda_0,\xi_0)=1$, so $c=1$ and Eqs. (28) coincide with the reality conditions (21). Summarizing we can say that the Riemann–Hilbert splitting (6) exists and defines a SU(n) solution of (1) if there exist λ, ξ-dependent matrix superfields G_{01}, G_{23} that (i) are analytic in $U_0\cap U_1$ and $U_2\cap U_3$, respectively, (ii) depend on x^μ, θ^A_i, $\bar\theta^{\dot Ai}$ only through the functions (8), (iii) admit the splittings (11a) with $\det\chi_a=1$ (hence $\det G_{01}=\det G_{23}=1$), and (iv) satisfy the condition (25). We suspect that each SU(n) solution of (1) can be obtained from some G_{01}, G_{23}.

Unfortunately we are not able to give nontrivial, non-Abelian examples that satisfy the above conditions. For the gauge group U(1) the condition $\det\chi_a=1$ should be replaced by $\det\chi_a\neq 0$ and we can take

$$G_{01}=G_{23}=\exp f,$$

where

$$f=f(z_Ax_+^{A\dot B},z_A\theta^A_i,\lambda)$$

is analytic with respect to λ in a neighborhood of $|\lambda|=1$. Equations (11a) are now satisfied by

$$\chi_0=\chi_2=\exp(-f_0),$$

$$\chi_1=\chi_3=\exp(f_1),$$

where f_0 (f_1) is the part of the Laurent series of f containing only positive (resp. nonpositive) powers of λ. Setting $\lambda_0=\xi_0=0$ in (27) we obtain

$$\psi_0=\exp(f_0^+(\xi)-f_0(\lambda)).$$

Hence, in virtue of (2),

$$\mathscr{A}^i_{1} = 0, \quad \mathscr{A}^i_{2} = -D^i_{2}h, \tag{29a}$$

$$\mathscr{A}_{\dot{1}i} = 0, \quad \mathscr{A}_{\dot{2}i} = \overline{D}_{\dot{2}i}h^{+}, \tag{29b}$$

$$\begin{aligned} &\mathscr{A}_{1\dot{1}} = 0, \quad \mathscr{A}_{1\dot{2}} = D_{1\dot{2}}h^{+}, \quad \mathscr{A}_{2\dot{1}} = -D_{2\dot{1}}h, \\ &\mathscr{A}_{2\dot{2}} = D_{2\dot{2}}(h^{+} - h), \end{aligned} \tag{30}$$

where

$$h = f_1(\infty) = \oint_{|\lambda|=1} \frac{d\lambda}{2\pi i\lambda} f.$$

The above solutions are linear superpositions of two self-dual solutions (one of which could be called anti-self-dual). For $\theta^A_i = \overline{\theta}^{\dot{A}i} = 0$ the formulas (30) yield all analytic solutions of the Maxwell equations.[19] We do not know whether (29) and (30) represent all U(1) solutions of the supersymmetric constraint equations.

V. CONCLUDING REMARKS

We have shown that there is no simple way of transforming the constraint equations (1) into the Riemann–Hilbert problem in one complex variable. It can be done provided extraordinary boundary conditions, e.g., (5), are imposed on the wave functions, but then the soliton methods do not work. Such conditions can be avoided if we consider the Riemann–Hilbert problem (6) in two complex variables λ, ξ. Unfortunately, soliton techniques are not yet worked out for this problem. We have slightly relaxed the conditions on the patching functions G_{ab}, which assure the existence of solutions of (1) for the gauge group GL(n,C) (Sec. III) and SU(n) (Sec. IV). They can be easily satisfied if the G_{ab} do not depend on one of the parameters λ, ξ. Then the resulting gauge fields are self-dual or anti-self-dual (Sec. III). In the Abelian case they can be added in order to get real solutions of (1). Further examples are needed to prove the usefulness of the Riemann–Hilbert problem for solving the constraint equations. This work is in progress.

ACKNOWLEDGMENTS

I am grateful to W. Nahm for encouragement and to E. F. Corrigan and R. S. Ward for useful practical suggestions. I also thank R. Flume for helpful discussions.

A large part of this work was carried out during my stay at the Department of Mathematical Sciences and at Trevelyan College of the University of Durham. I wish to thank Trevelyan College for a Visiting Fellowship and its generous hospitality and the U. K. Science and Engineering Research Council for financial support.

[1] R. Grimm, M. Sohnius, and J. Wess, Nucl. Phys. B **133**, 275 (1978).
[2] S. J. Gates, M. T. Grisaru, M. Rocek, and W. Siegel, *Supersymmetry or One Thousand and One Lessons in Supersymmetry* (Benjamin, New York, 1983).
[3] M. Sohnius, Nucl. Phys. B **136**, 461 (1978).
[4] E. Witten, Phys. Lett. B 77, 394 (1978).
[5] I. Ya. Aref'eva and I. V. Volovich, "Higher conservation laws for supersymmetric gauge theories. II.," report presented at the Third Seminar "Quantum Gravity," Moscow, 23–25 October 1984.
[6] J. Harnad, J. Hurtubise, M. Légaré, and S. Shnider, "Constraint equations and field equations in supersymmetric $N = 3$ Yang–Mills theory," Montreal CRMA preprint, 1984.
[7] I. V. Volovich, Phys. Lett. B **129**, 429 (1983).
[8] C. Devchand, Nucl. Phys. B **238**, 333 (1984).
[9] S. Novikov, S. V. Manakov, L. P. Pitaevskii, and V. E. Zakharov, *Theory of Solitons: The Inverse Scattering Method* (Consultants Bureau, New York, 1984).
[10] M. F. Atiyah and R. S. Ward, Commun. Math. Phys. **55**, 117 (1977).
[11] E. F. Corrigan, D. B. Fairlie, P. Goddard, and R. G. Yates, Commun. Math. Phys. **58**, 223 (1978).
[12] R. S. Ward, Commun. Math. Phys. **80**, 563 (1981).
[13] J. Burzlaff, Lett. Math. Phys. **6**, 141 (1982).
[14] V. E. Zakharov and A. B. Shabat, Funkcional. Anal. Prilozen. **13**, 13 (1979).
[15] L.-L. Chau, M.-L. Ge, and Z. Popowicz, Phys. Rev. Lett. **52**, 1940 (1984).
[16] I. Ya. Aref'eva and I. V. Volovich, Lett. Math. Phys. **9**, 231 (1985).
[17] I. V. Volovich, Phys. Lett. B **123**, 329 (1983).
[18] I. V. Volovich, Lett. Math. Phys. **7**, 517 (1983).
[19] R. Penrose, J. Math. Phys. **10**, 38 (1969).

Relationship between specific surface area and spatial correlation functions for anisotropic porous media

James G. Berryman
Lawrence Livermore National Laboratory, P. O. Box 808 L-156, Livermore, California 94550

(Received 13 May 1986; accepted for publication 27 August 1986)

A result of Debye, Anderson, and Brumberger [P. Debye, H. R. Anderson, Jr., and H. Brumberger, J. Appl. Phys. **28**, 679 (1957)] for isotropic porous media states that the derivative of the two-point spatial correlation at the origin is equal to minus one-quarter of the specific surface area. This result is generalized for nonisotropic media by noting that the angular average of the anisotropic two-point spatial correlation function has the same relationship to the specific surface area.

I. INTRODUCTION

Debye, Anderson, and Brumberger[1] have shown that, for isotropic porous media, the derivative of the two-point spatial correlation at the origin is equal to minus one-quarter of the specific surface area. The two-point correlation function can be obtained from pictures of cross sections of a material using image processing techniques.[2] Then, the specific surface area can be measured using the result of Debye *et al.*[1] The practical importance of this result has been demonstrated recently by combining the measured values of specific surface area with a Kozeny–Carman relation to obtain estimates of fluid permeability.[3]

Since we do not always know that the porous material to be analyzed is isotropic, an important question arises concerning the applicability of the result of Debye *et al.* for possibly nonisotropic media. We generalize their result to anisotropic media by noting that the angular average of any two-point spatial correlation function has the same relationship to the specific surface area. Thus, if the two-point correlation function is computed from images by taking angular averages (as has typically been proposed[2]), the slope at the origin will provide a valid estimate of the specific surface area regardless of the degree of anisotropy of the sample.

II. ANISOTROPIC POROUS MEDIA

For a porous material, we define a characteristic function $f(\mathbf{x}) = 0$ or 1. Then, we say that void regions have $f = 1$, while material regions have $f = 0$. The first two void–void correlation functions are then given by

$$\hat{S}_2 = \langle f(\mathbf{x}) \rangle = \phi , \tag{1}$$

and

$$\hat{S}_2(\mathbf{r}_1,\mathbf{r}_2) = \langle f(\mathbf{x}+\mathbf{r}_1) f(\mathbf{x}+\mathbf{r}_2) \rangle . \tag{2}$$

The brackets $\langle \cdot \rangle$ indicate a volume average over the spatial coordinate $\mathbf{x}$. The void volume fraction (or porosity) is given by ϕ. We refer to these two correlation functions as the one- and two-point correlation functions, respectively. For isotropic materials, the one- and two-point correlations can in principle be measured by processing representative images of material cross sections. However, for anisotropic materials, multiple images in orthogonal planes are required to obtain all the necessary information. In general, we still assume that the porous medium of interest is statistically homogeneous so that on average only the differences in the coordinate values are significant (translational invariance). With these assumptions, the two-point correlation function simplifies to

$$\hat{S}_2(\mathbf{r}_1,\mathbf{r}_2) = S_2(\mathbf{r}_2 - \mathbf{r}_1) . \tag{3}$$

From Eq. (2), it follows that

$$S_2(\mathbf{r}) = \frac{1}{V} \int_V d^3x\, f(\mathbf{x}) f(\mathbf{x}+\mathbf{r}) , \tag{4}$$

where V is total volume of integration. Two important facts about the two-point correlations for applications to random media are

$$S_2(0) = \phi \tag{5}$$

and

$$\lim_{|\mathbf{r}| \to \infty} S_2(\mathbf{r}) = \phi^2 . \tag{6}$$

The theorem that we wish to prove states that

$$A_2'(0) = -s/4 , \tag{7}$$

where s is the specific surface area (internal surface area per unit volume) and the angular average of $S_2(\mathbf{r})$ is defined by

$$\begin{aligned} A_2(r) &= \frac{1}{4\pi} \int d\varphi\, d\theta \sin\theta\, S_2(r\hat{r}) \\ &= \frac{1}{4\pi V} \int d\varphi\, d\theta \sin\theta \int_V d^3x\, f(\mathbf{x}) f(\mathbf{x}+r\hat{r}) , \end{aligned} \tag{8}$$

where $\hat{r} = \hat{r}(\theta,\varphi)$ is the radial unit vector.

The derivation of Eq. (7) proceeds as follows: Taking the derivative of Eq. (8) gives

$$\frac{dA_2(r)}{dr} = \frac{1}{4\pi V} \int d\varphi\, d\theta \sin\theta \int_V d^3x\, f(\mathbf{x}) \frac{\partial f(\mathbf{x}+r\hat{r})}{\partial r} . \tag{9}$$

Defining the pore volume as V_p, we have

$$\frac{dA_2(r)}{dr} = \frac{1}{4\pi V} \int d\varphi\, d\theta \sin\theta\, \hat{r} \cdot \int_{V_p} d^3x\, \nabla f(\mathbf{x}+r\hat{r}) . \tag{10}$$

 0022-2488/87/010244-02$02.50

Then, if da_s is an infinitesimal element of the material surface area a_s, we have

$$\frac{dA_2(r)}{dr} = \frac{1}{4\pi V}\int_{a_s} da_s \int d\varphi\, d\theta \sin\theta \hat{r}\cdot \hat{n}_s f(\mathbf{x}_s + r\hat{r}) , \tag{11}$$

where $\hat{n}$ is the unit outward normal vector at the surface position given by $\mathbf{x}_s$. Now, if we let $r \to 0^+$ and center the coordinate system at $\mathbf{x}_s$ with $\hat{n} = \hat{z}$, we find that

$$\begin{aligned}\int d\varphi\, d\theta \sin\theta\hat{r}\cdot\hat{n}_s f(\mathbf{x}_s + 0^+\hat{r})\\ &= 2\pi \int d\theta \sin\theta\cos\theta f(\mathbf{x}_s + 0^+\hat{r})\\ &= 2\pi\int_{-1}^{0} d\mu\mu = -\pi ,\end{aligned} \tag{12}$$

and we obtain a definite result for Eq. (11) given by

$$\lim_{r\to 0^+}\frac{dA_2(r)}{dr} = -\frac{a_s}{4V} . \tag{13}$$

Since the specific surface area is defined as $s = a_s/V$, Eq. (13) is equivalent to Eq. (7). Debye *et al.*[1] used a more intuitive approach to obtain their result for isotropic media.

ACKNOWLEDGMENTS

I thank Professor R. A. Knight for asking a stimulating question.

This work was performed under the auspices of the U. S. Department of Energy by the Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48 and supported specifically by the Electronics Engineering Department Thrust Area in Remote Sensing, Image, and Signal Engineering.

[1]P. Debye, H. R. Anderson, Jr., and H. Brumberger, J. Appl. Phys. **28**, 679 (1957).
[2]J. G. Berryman, J. Appl. Phys. **57**, 2374 (1985).
[3]J. G. Berryman and S. C. Blair, J. Appl. Phys. **60**, 1930 (1986).