A generalization and refinement of Madivanane's theorem on general involutional transformations

J. S. Lomont
Department of Mathematics, University of Arizona, Tucson, Arizona 85721

(Received 21 February 1985; accepted for publication 9 May 1985)

A result of Madivanane on "quasianticommuting" complex square matrices, which are roots of the identity matrix, is generalized to linear operators on infinite-dimensional complex vector spaces. A constant which appears in the result is shown to be nonzero by giving a simple formula for its absolute value. Also, a simple formula for the constant itself is given, which is valid for some cases.

I. INTRODUCTION

It was shown by Madivanane[1] that an $n \times n$ complex matrix T, which is defined in terms of two $n \times n$ complex matrices A and B having certain properties, has the property $T^n = cI$. The two principal objectives of this paper are to show (1) that T, A, and B need not be $n \times n$ matrices and (2) that $c \neq 0$. In addition, $|c|$ will be evaluated.

The generalization can be formulated either in terms of associative algebras, or in terms of linear operators on arbitrary (possibly infinite-dimensional) complex vector spaces. The latter course will be followed here.

II. A SEQUENCE OF "GAUSS-TYPE" SUMS

It will be shown that certain "Gauss-type" sums are not equal to zero. In addition, the absolute values of the "Gauss-type" sums will be evaluated.

Definition 1: Let n be a positive integer, let ω be any primitive nth root of 1, and let $\tilde{\omega}$ be any primitive $2n$th root of 1. Then

$$S_n = \begin{cases} \displaystyle\sum_{j=0}^{n-1} \omega^{-j(j-1)/2}, & \text{if } n \text{ is odd,} \\ \displaystyle\sum_{j=0}^{n-1} \tilde{\omega}^{-j^2}, & \text{if } n \text{ is even.} \end{cases}$$

The dependence of S_n on ω or $\tilde{\omega}$ will be suppressed. Thus,

n	1	2	3	4	5
S_n	1	$1-\tilde{\omega}$	$1-\omega$	$-2\tilde{\omega}^3$	$-2\omega-\omega^2-2\omega^3$
n	6				
S_n	$1+2\tilde{\omega}-2\tilde{\omega}^2-\tilde{\omega}^3$.				

Lemma 1: Let n be a positive odd integer $\geqslant 3$, and let ω be any primitive nth root of 1. Then

(1) $S_n = \displaystyle\prod_{k=1}^{(n-1)/2} (1-\omega^{2k-1}) \neq 0$,

(2) $|S_n| = \sqrt{n}$.

Proof: (1) This is a simple consequence of p. 86 of Ref. 2. To prove (2), we have

$$S_n = (1-\omega)(1-\omega^3)\cdots(1-\omega^{n-2}).$$

Define

$$t_n = (1-\omega^2)(1-\omega^4)\cdots(1-\omega^{n-1}).$$

By p. 87 of Ref. 2,

$$S_n t_n = \prod_{j=1}^{n-1} (1-\omega^j) = n.$$

Since

$$\begin{aligned} \overline{S}_n &= (1-\omega^{-1})(1-\omega^{-3})\cdots(1-\omega^{-n+2}) \\ &= (1-\omega^{n-1})(1-\omega^{n-3})\cdots(1-\omega^2) \\ &= t_n, \end{aligned}$$

$$|S_n|^2 = S_n t_n = n. \qquad \blacksquare$$

Lemma 2: Let n be a positive even integer, let $\tilde{\omega}$ be any primitive $2n$th root of 1, and let $\hat{\omega}$ be any primitive $8n$th root of 1 such that $\hat{\omega}^4 = \tilde{\omega}$. Then

(1) $S_n = \pm \hat{\omega}^{-n}\sqrt{n} \neq 0$,

(2) $|S_n| = \sqrt{n}$.

Proof: Clearly it suffices to prove that $S_n = \pm \hat{\omega}^{-n}\sqrt{n}$.

(a) To prove that $S_n = \frac{1}{2}\sum_{j=0}^{2n-1} \tilde{\omega}^{j^2}$, let $m = n/2$. Then

$$\begin{aligned} \sum_{j=n}^{2n-1} \tilde{\omega}^{j^2} &= \sum_{j=0}^{n-1} \tilde{\omega}^{(n+j)^2} = \sum_{j=0}^{n-1} \tilde{\omega}^{n^2+2nj+j^2} \\ &= \sum_{j=0}^{n-1} (\tilde{\omega}^{2n})^{m+j} \tilde{\omega}^{j^2} = \sum_{j=0}^{n-1} \tilde{\omega}^{j^2} = \overline{S}_n. \end{aligned}$$

(b) To prove that $S_n = \pm \hat{\omega}^{-n}\sqrt{n}$ if $\tilde{\omega} = \exp(r\pi i/n)$, $r \in \{1,2,\ldots,2n-1\}$, r and n are relatively prime, and $\hat{\omega} = \exp(r\pi i/4n)$, let $\tilde{\omega}$ be any primitive $2n$th root of 1. There is a positive integer $r \in \{1,2,\ldots,n-1\}$ such that r and n are relatively prime, and $\tilde{\omega} = \exp(r\pi i/n)$. By part (a),

$$2\overline{S}_n = \sum_{j=0}^{2n-1} \exp\left(\frac{j^2 r\pi i}{n}\right).$$

In the notation of Ref. 3, the last sum is $\phi(r,2n)$, so $2\overline{S}_n = \phi(r,2n)$. By p. 178 of Ref. 3,

$$\phi(r,2n) = \phi(1,2rn)/\phi(2n,r),$$

if $\phi(2n,r) \neq 0$. By p. 187 of Ref. 3.

$$\phi(2n,r) = \pm i^{(r-1)^2/4}\sqrt{r} \quad (\neq 0),$$

and by p. 177 of Ref. 3,

$$\phi(1,2rn) = (1+i)\sqrt{2rn}.$$

Therefore

$$2\overline{S}_n = \phi(r,2n) = \pm(1+i)\, i^{-(r-1)^2/4}\sqrt{2n}.$$

Now let $\hat{\omega} = \exp(r\pi i/n)$ (so $\hat{\omega}^4 = \tilde{\omega}$), and $m = n/2$. Since $r^8 \equiv 1 \pmod{16}$,

$$\hat{\omega}^{r^7 n} = e^{r^8 \pi i/4} = e^{\pi i/4} = (1/\sqrt{2})(1+i)$$

$$\hat{\omega}^{(r^7 + r - 2)m} = e^{(r^8 + r^2 - 2r)\pi i/8} = e^{(r^2 - 2r + 1)\pi i/8}$$

$$= e^{(r-1)^2 \pi i/8} = e^{(\pi i/2)(r-1)^2/2} = i^{(r-1)^2/4}.$$

Therefore

$$\overline{S}_n = \pm \hat{\omega}^{r^7 n} \hat{\omega}^{-(r^7 + r - 2)m} \sqrt{n}$$

$$= \pm \hat{\omega}^{(r^7 - r + 2)m} \sqrt{n}.$$

Since $r^2 \equiv 1 \pmod 8$,

$$r^7 - r = r(r^2 - 1)(r^4 + r^2 + 1) \equiv 0 \pmod 8.$$

Define

$$t = \tfrac{1}{8}(r^7 - r), \quad \omega = \tilde{\omega}^2.$$

Then

$$\hat{\omega}^{r^7 - r} = \hat{\omega}^{8t} = \omega^t,$$

so

$$\hat{\omega}^{(r^7 - r)m} = (\omega^m)^t = (-1)^t.$$

Therefore

$$\overline{S}_n = \pm \hat{\omega}^{(r^7 - r + 2)m} \sqrt{n} = \pm \hat{\omega}^{2m} \sqrt{n} = \pm \hat{\omega}^n \sqrt{n}.$$

(c) To prove that $s_n = \pm \hat{\omega}^{-n} \sqrt{n}$ for any fourth root $\hat{\omega}$ of $\tilde{\omega}$, let $\hat{\omega}_0 = \exp(r\pi i/n)$, so by part (b), $S_n = \pm \hat{\omega}_0^{-n} \sqrt{n}$. Since $\hat{\omega} = \pm \hat{\omega}_0$, or $\hat{\omega} = \pm i\hat{\omega}_0$,

$$\hat{\omega}^n = \pm \hat{\omega}_0^{\,n},$$

so

$$S_n = \pm \hat{\omega}^{-n} \sqrt{n}. \qquad \blacksquare$$

III. TWO SEQUENCES OF COMPLEX NUMBERS

Definition 2: Let n be a positive integer, let ω be a primitive nth root of 1, and let $\tilde{\omega}$ be a primitive $2n$th root of 1. Then

$$\gamma_n = \begin{cases} \prod_{k=0}^{n-1} \omega^{k(k+1)/2}, & \text{if } n \text{ is odd,} \\ \prod_{k=0}^{n-1} \tilde{\omega}^{k^2}, & \text{if } n \text{ is even.} \end{cases}$$

The dependence of γ_n on ω or $\tilde{\omega}$ is suppressed. Thus,

n	1	2	3	4	5	6	8	9
γ_n	1	$\tilde{\omega}$	ω	$\tilde{\omega}^6 = -\tilde{\omega}^2$	1	$\tilde{\omega}^7 = -\tilde{\omega}$	$\tilde{\omega}^{12} = -\tilde{\omega}^4$	ω^3.

The following result is elementary.

Lemma 3: Let n be a positive integer, let ω be a primitive nth root of 1, and let $\tilde{\omega}$ be a primitive $2n$th root of 1. Then,

$$\gamma_n = \begin{cases} \omega^{n(n^2 - 1)/6}, & \text{if } n \text{ is odd,} \\ \tilde{\omega}^{n(n-1)(2n-1)/6}, & \text{if } n \text{ is even.} \end{cases}$$

Remark 1: Let n be a positive odd integer, and let ω be a primitive nth root of 1.

(i) If n and 3 are relatively prime, then $\gamma_n = 1$.

(ii) If $n \equiv 0 \pmod 3$, and $m = n/3$, then $\gamma_n = \omega^m$.

Proof: (i) Since n is odd, n and 3 are relatively prime, and $n(n^2 - 1)/6$ is an integer, it follows that $(n^2 - 1)/6$ is an integer. By Lemma 3,

$$\gamma_n = \omega^{n(n^2 - 1)/6} = (\omega^n)^{(n^2 - 1)/6} = 1.$$

(ii) Define $q = (3m^2 - 1)/2$, so

$$n(n^2 - 1)/6 = m + qn.$$

By Lemma 3,

$$\gamma_n = \omega^{m + qn} = \omega^m (\omega^n)^q = \omega^m. \qquad \blacksquare$$

Definition 3: Let n be a positive integer, let ω be a primitive nth root of 1, and let $\tilde{\omega}$ be a primitive $2n$th root of 1 such that $\tilde{\omega}^2 = \omega$. Then

$$c_n = \gamma_n (S_n)^n.$$

Thus,

n	1	2	3	4	5	6
c_n	1	2	$6 + 3\omega$	16ω	$25(1 + 2\omega^2 + 2\omega^3)$	$-216\omega^2$

Lemma 4:

(i) $c_n \neq 0, \quad n = 1,2,3,4,\ldots$.

(ii) $|c_n| = n^{n/2}, \quad n = 1,2,3,4,\ldots$.

(iii) $c_n = \omega^{n(n-2)(4n-1)/24} n^{n/2}$, if n is even.

Proof: (ii) By Definition 3, $|c_n| = |\gamma_n|\,|S_n|^n$. By Definition 2, $|\gamma_n| = 1$, so $|c_n| |S_n|^n$. By Lemmas 1 and 2, $|S_n| = \sqrt{n}$, so $|c_n| = (\sqrt{n})^n$.

(i) By part (ii), $c_n \neq 0$.

(iii) If n is even, then by Lemmas 3 and 2,

$$c_n = \tilde{\omega}^{n(n-1)(2n-1)/6} \hat{\omega}^{-n^2} n^{n/2}$$

$$= \tilde{\omega}^{n(n-1)(2n-1)/6} \tilde{\omega}^{-n^2/4} n^{n/2}$$

$$= \tilde{\omega}^{n(n-2)(4n-1)/12} n^{n/2}$$

$$= \omega^{n(n-2)(4n-1)/24} n^{n/2}$$

[since $n(n-2)(4n-1) \equiv 0 \pmod{24}$]. $\blacksquare$

IV. TWO SEQUENCES OF COMPLEX POLYNOMIALS

Definition 4: Let n be a positive integer, let ω be any primitive nth root of 1, and let $\tilde{\omega}$ be any primitive $2n$th root of 1. Then

$$P_n(z) = \begin{cases} \sum_{j=0}^{n-1} \omega^{-j(j-1)/2} z^j, & \text{if } n \text{ is odd,} \\ \sum_{j=0}^{n-1} \tilde{\omega}^{-j^2} z^j, & \text{if } n \text{ is even.} \end{cases}$$

The dependence of P_n or ω or $\tilde{\omega}$ is suppressed. Thus,

n	$P_n(z)$
1	1
2	$1 - \tilde{\omega}z$

3	$1+z+\omega^2z^2$
4	$1-\tilde\omega^3z-z^2-\tilde\omega^3z^3$
5	$1+z+\omega^4z^2+\omega^2z^3+\omega^4z^4$
6	$1-\tilde\omega^5z-\tilde\omega^2z^2+\tilde\omega^3z^3-\tilde\omega^2z^4-\tilde\omega^5z^5.$

Lemma 5:

(i) $P_n(1)=S_n\neq 0, \quad n=1,2,3,4,\ldots.$

(ii) $c_n=\gamma_n[P_n(1)]^n, \quad n=1,2,3,4,\ldots.$

Lemma 6: Let n be an odd positive integer $\geqslant 3$, let $k\in\{0,1,2,\ldots,n-1\}$, and let ω be the primitive nth root of 1, which is used in the definition of $P_n(z)$. Then

$$P_n(\omega^k)=\omega^{k(k+1)/2}P_n(1)=\omega^{k(k+1)/2}S_n.$$

Proof: The assertion is trivial if $k=0$. Assume $k\geqslant 1$. By Definition 4,

$$P_n(\omega^k)=\sum_{j=0}^{n-1}\omega^{-j(j-2k-1)/2},$$

so

$$\omega^{-k(k+1)/2}P_n(\omega^k)=\sum_{j=0}^{n-1}\omega^{-(1/2)(j-k)(j-k-1)}=\sum_{l=-k}^{k}\omega^{-l(l-1)/2}.$$

From

$$\omega^{(1/2)(n+l)(n+l-1)}=\omega^{(1/2)[n^2+(2l-1)n+l(l-1)]}=(\omega^n)^{(n+2l-1)/2}\omega^{l(l-1)/2}=\omega^{l(l-1)/2},$$

it follows that

$$\sum_{l=-k}^{-1}\omega^{-l(l-1)/2}=\sum_{l=-k}^{-1}\omega^{-(1/2)(n+l)(n+l-1)}=\sum_{j=n-k}^{n-1}\omega^{-j(j-1)/2}.$$

Therefore

$$\begin{aligned}\omega^{-k(k+1)/2}P_n(\omega^k)&=\sum_{l=0}^{n-k-1}\omega^{-l(l-1)/2}+\sum_{l=-k}^{-1}\omega^{-l(l-1)/2}\\&=\sum_{j=0}^{n-k-1}\omega^{-j(j-1)/2}+\sum_{j=n-k}^{n-1}\omega^{-j(j-1)/2}\\&=\sum_{j=0}^{n-1}\omega^{-j(j-1)/2}=P_n(1)\\&=S_n.\end{aligned}$$

■

Lemma 7: Let n be a positive even integer, let $k\in\{0,1,2,\ldots,n-1\}$, let $\tilde\omega$ be the primitive $2n$th root of 1 used in the definition of $P_n(z)$, and let $\omega=\tilde\omega^2$. Then

$$P_n(\omega^k)=\tilde\omega^{k^2}P_n(1)=\tilde\omega^{k^2}S_n.$$

Proof: The assertion is trivial if $k=0$. Assume $k\geqslant 1$. By Definition 4,

$$P_n(\omega^k)=\sum_{j=0}^{n-1}\tilde\omega^{-j^2+2kj},$$

so

$$\tilde\omega^{-k^2}P_n(\omega^k)=\sum_{j=0}^{n-1}\tilde\omega^{-(j^2-2kj+k^2)}=\sum_{j=0}^{n-1}\tilde\omega^{-(j-k)^2}=\sum_{l=-k}^{n-k-1}\tilde\omega^{-l^2}.$$

Define $m=n/2$. From

$$\tilde\omega^{(n+l)^2}=\tilde\omega^{n^2+2nl+l^2}=(\tilde\omega^{2n})^{m+l}\,\tilde\omega^{l^2}=\tilde\omega^{l^2},$$

it follows that

$$\sum_{l=-k}^{-1}\tilde\omega^{-l^2}=\sum_{l=-k}^{-1}\tilde\omega^{-(n+l)^2}=\sum_{j=n-k}^{n-1}\tilde\omega^{-j^2}.$$

Therefore

$$\begin{aligned}\tilde\omega^{-k^2}P_n(\omega^k)&=\sum_{l=0}^{n-k-1}\tilde\omega^{-l^2}+\sum_{l=-k}^{-1}\tilde\omega^{-l^2}\\&=\sum_{j=0}^{n-k-1}\tilde\omega^{-j^2}+\sum_{j=n-k}^{n-1}\tilde\omega^{-j^2}\\&=\sum_{j=0}^{n-1}\tilde\omega^{-j^2}=P_n(1)=S_n.\end{aligned}$$

■

Lemma 8: Let n be a positive integer: if n is odd, let ω be the primitive nth root of 1 used in the definition of $P_n(z)$; and if n is even, let $\tilde\omega$ be the primitive $2n$th root of 1 used in the definition of $P_n(z)$, and $\omega=\tilde\omega^2$. Then

$$c_n=\prod_{k=0}^{n-1}P_n(\omega^k).$$

Proof: (1) Case n odd. The assertion is trivial if $n=1$. Assume $n\geqslant 3$. By Lemma 5, Definition 2, and Lemma 6,

$$c_n=\gamma_n\,[P_n(1)]^n=\prod_{k=0}^{n-1}\omega^{k(k+1)/2}P_n(1)=\prod_{k=0}^{n-1}P_n(\omega^k).$$

(2) Case n even. By Lemma 5, Definition 2, and Lemma 7,

$$c_n=\gamma_n[P_n(1)]^n=\prod_{k=0}^{n-1}\tilde\omega^{k^2}P_n(1)=\prod_{k=0}^{n-1}P_n(\omega^k).$$

■

Definition 5: Let n be a positive integer; if n is odd, let ω be the primitive nth root of 1, which is used in the definition of $P_n(z)$; and if n is even, let $\tilde\omega$ be the primitive $2n$th root of 1 that is used in the definition of $P_n(z)$, and $\omega=\tilde\omega^2$. Then

$$Q_n(z)=\prod_{k=0}^{n-1}P_n(\omega^kz).$$

Thus,

n	$Q_n(z)$
1	1
2	$1+z^2$
3	$1+(4+3\omega)z^3+z^6$
4	$1+(1+8\omega)z^4-(1-8\omega)z^8+z^{12}.$

Lemma 9: Let n, ω, $\tilde\omega$, and $Q_n(z)$ be as in Definition 5. Then (1) $Q_n(1)=c_n$; (2) $Q_n(\omega z)=Q_n(z)$ for all $z\in\mathbb{C}$; (3) $Q_n(z)$ is a polynomial in z^n, i.e., there is a finite sequence $\{a_j\}_0^{n-1}\subset\mathbb{C}$ such that

$$Q_n(z) = \sum_{j=0}^{n-1} a_j z^{nj}.$$

Proof: (1) By Definition 5 and Lemma 8,

$$Q_n(1) = \prod_{k=0}^{n-1} P_n(\omega^k) = c_n.$$

(2) By Definition 5,

$$Q_n(\omega z) = \prod_{k=1}^{n} P_n(\omega^k z).$$

Since $\omega^n = 1$,

$$Q_n(\omega z) = \prod_{k=0}^{n-1} P_n(\omega^k z) = Q_n(z).$$

(3) Let $\{\alpha_l\}_0^{n(n-1)} \subset \mathbb{C}$ be such that

$$Q_n(z) = \sum_{l=0}^{n(n-1)} \alpha_l z^l.$$

From the preceding equation and

$$Q_n(z) = Q_n(\omega z) = \sum_{l=0}^{n(n-1)} \omega^l \alpha_l z^l,$$

it follows that

$$\alpha_l = \omega^l \alpha_l, \quad l = 0,1,2,\ldots,n(n-1).$$

Therefore $\alpha_l = 0$ unless $\omega^l = 1$. ∎

V. NOTATION

The following notation will be used in the next two sections (VI and VII).

(1) X is an arbitrary complex vector space of dimension $\geqslant 1$.

(2) $H(X)$ is the complex vector space of all linear operators on X.

(3) $I \in H(x)$ is the identity operator.

(4) $\mathrm{GL}(X) = \{S \in H(x): S \text{ is bijective}\}$.

(5) n is a positive integer.

(6) ω is a primitive nth root of 1.

(7) $A, B \in H(X)$ are such that (i) $A^n = B^n = I$ [so $A, B \in \mathrm{GL(X)}$] and (ii) $AB = \omega BA$.

(8) $T = \sum_{j=0}^{n-1} A^{n-j-1} B^j$.

(9) $\tilde{\omega}$ is a primitive $2n$th root of 1 such that $\tilde{\omega}^2 = \omega$.

(10) $\hat{B} \in \mathrm{GL}(X)$ is defined by

$$\hat{B} = \begin{cases} A^{-1}B, & \text{if } n \text{ is odd,} \\ \tilde{\omega} A^{-1} B, & \text{if } n \text{ is even.} \end{cases}$$

Thus,

n	1	2	3	4
T	I	$A+B$	A^2+AB+B^2	$A^3+A^2B+AB^2+B^3$

VI. ELEMENTARY IDENTITIES

Following are some elementary identities:

(1) $A^jB^k = \omega^{jk}B^kA^j$, $\quad j,k = 1,2,3,\ldots,$

(2) $BA^{-j} = \omega^j A^{-j}B$, $\quad j = 1,2,3,\ldots,$

(3) $A^{-j}BA^j = \omega^{-j}B$, $\quad j = 1,2,3,\ldots,$

(4) $A^{-j(n-1)}BA^{j(n-1)} = \omega^j B$, $\quad j = 1,2,3,\ldots,$

(5) $(A^{-1}B)^j = \omega^{j(j-1)/2}A^{-j}B^j$, $\quad j = 1,2,3,\ldots$.

It is easily shown that

(6) $\hat{B}^n = I$ [so $B \in \mathrm{GL}(X)$],

(7) $A\hat{B} = \omega \hat{B} A$.

Consequently the identities (1)–(5) remain true when B is replaced by $\hat{B}$.

Lemma 10: $T = A^{n-1}P_n(\hat{B})$, if the ω or $\tilde{\omega}$ used in defining $P_n(z)$ is the same as in this section.

Proof: (1) We prove for the case when n is odd. By identity 5,

$$\begin{aligned} T &= \sum_{j=0}^{n-1} A^{n-j-1}B^j = A^{n-1}\sum_{j=0}^{n-1} A^{-j}B^j \\ &= A^{n-1}\sum_{j=0}^{n-1} \omega^{-j(j-1)/2}(A^{-1}B)^j \\ &= A^{n-1}\sum_{j=0}^{n-1} \omega^{-j(j-1)/2}\hat{B}^j = A^{n-1}P_n(\hat{B}). \end{aligned}$$

(2) We prove for the case when n is even. As in part (1),

$$\begin{aligned} T &= A^{n-1}\sum_{j=0}^{n-1} \omega^{-j(j-1)/2}(A^{-1}B)^j \\ &= A^{n-1}\sum_{j=0}^{n-1} \omega^{-j(j-1)/2}\tilde{\omega}^{-j}(\tilde{\omega}A^{-1}B)^j \\ &= A^{n-1}\sum_{j=0}^{n-1} \tilde{\omega}^{-j^2}\hat{B}^j = A^{n-1}P_n(\hat{B}). \end{aligned}$$ ∎

VII. THE THEOREM

Theorem:

(1) $T^n = c_n I$,

where c_n is given by Definition 3.

(2) $c_n \neq 0$.

(3) $|c_n| = n^{n/2}$.

(4) $c_n = \omega^{n(n-2)(4n-1)/24} n^{n/2}$, if n is even.

Proof: (1) By Lemma 10 and Identity 4,

$$\begin{aligned} T^n &= [A^{n-1}P_n(\hat{B})]^n = A^{n-1}P_n(\hat{B})A^{n-1}P_n(\hat{B})\cdots A^{n-1}P_n(\hat{B})A^{n-1}P_n(\hat{B}) \\ &= A^{n(n-1)}[A^{-(n-1)^2}P_n(\hat{B})A^{(n-1)^2}]\cdots[A^{-2(n-1)}P_n(\hat{B})A^{2(n-1)}][A^{-(n-1)}P_n(\hat{B})A^{n-1}]P_n(\hat{B}) \\ &= P_n[A^{-(n-1)^2}\hat{B}A^{(n-1)^2}]\cdots P_n[A^{-2(n-1)}\hat{B}A^{2(n-1)}]P_n[A^{-(n-1)}\hat{B}A^{n-1}]P_n(\hat{B}) \\ &= P_n(\omega^{n-1}\hat{B})\cdots P_n(\omega^2\hat{B})P_n(\omega\hat{B})P_n(\hat{B}) = Q_n(\hat{B}). \end{aligned}$$

By Lemma 9, and Identity 6,

$$T^n = Q_n(\hat{B}) = Q_n(I) = Q_n(1)I = c_nI.$$

(2), (3), (4) These assertions come from Lemma 4. ■

VIII. A FINAL REMARK

Remark 2: Let X be an arbitrary complex vector space, let n be a positive integer, let A and B be linear operators on X such that $A^n = B^n = I$, and let $T = \sum_{j=0}^{n-1} A^{n-j-1} B^j$. Then $AT = TB$.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor J. D. Brillhart for several helpful discussions.

[1]S. Madivanane, J. Math. Phys. **25**, 2807 (1984).

[2]H. Rademacher, *Lectures on Elementary Number Theory* (Blaisdell, New York, 1964).

[3]T. Nagell, *Introduction to Number Theory* (Wiley, New York, 1959).

Nontrivial zeros of weight 1 3*j* and 6*j* coefficients: Relation to Diophantine equations of equal sums of like powers

Simcha Brudno
Chicago College of Osteopathic Medicine, Chicago, Illinois 60615
James D. Louck
Los Alamos National Laboratory, Theoretical Division, Los Alamos, New Mexico 87545

(Received 15 April 1985; accepted for publication 31 May 1985)

The nontrivial zeros of weight 1 3*j* and 6*j* coefficients given previously are shown to be the set of all such zeros. The relation of these zeros to the solutions of well-known Diophantine equations is also discussed.

I. INTRODUCTION

Nontrivial zeros of the 3*j* and 6*j* coefficients of weight 1 were given earlier by one of us.[1] These zeros originate from coefficients in which only two terms occur in the alternating sum expressions for the coefficients. They are called "linear zeros" in Ref. 1. The present terminology "weight 1" refers to this same class of coefficients and is based on the property that at least one 1 appears in the Regge[2] array notation and, respectively, in the Bargmann[3] array notation for these coefficients. This notation is used in Ref. 4 in the discussion of the symmetries of these coefficients and is given by

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \begin{bmatrix} j_1+m_1 & j_2+m_2 & j_3+m_3 \\ j_1-m_1 & j_2-m_2 & j_3-m_3 \\ j_2+j_3-j_1 & j_3+j_1-j_2 & j_1+j_2-j_3 \end{bmatrix}, \tag{1}$$

$$\begin{Bmatrix} a & b & e \\ d & c & f \end{Bmatrix} = \begin{pmatrix} d+f-b & c+f-a & c+d-e \\ a+f-c & b+f-d & a+b-e \\ d+e-c & b+e-a & b+d-f \\ a+e-b & c+e-d & a+c-f \end{pmatrix}. \tag{2}$$

The term "weight" is suggested by the properties of the Wigner and Racah operators associated with these coefficients.[4,5] A weight 1 coefficient is thus one in which a quantum number is chosen to be one number off a "stretched" coefficient; for example, $j_3 = j_1 + j_2 - 1$ (or $m_1 = j_1 - 1$) in the 3*j* symbol or $e = a + b - 1$ in the 6*j* symbol (or any symmetry equivalent of these). A stretched or boundary coefficient (for example, $e = a + b$) is a weight 0 coefficient and consists of "one term." These coefficients possess no nontrivial zeros. By definition, a nontrivial zero of the 3*j* coefficient (resp. of the 6*j* coefficient) is a set of quantum numbers $(j_1,j_2,j_3,m_1,m_2,m_3)$ [resp. (a,b,c,d,e,f)] such that all the domains of definition are satisfied and for which the coefficient has value zero. (The domains of definition of the quantum numbers are well known and consist of the integer and half-integer rules, the triangle rules for the angular momenta, and the projection rules for the m labels.)

The purpose of the present paper is twofold: (i) we relate the determination of the nontrivial zeros of weight 1 3*j* and 6*j* coefficients to the solution of classic Diophantine equations; and (ii) we prove that the enumeration of zeros given in Ref. 1 is complete, that is, that all such zeros of weight 1 3*j* and 6*j* coefficients are obtained by the method described there.

II. ZEROS OF WEIGHT 1 COEFFICIENTS IN TERMS OF STANDARD DIOPHANTINE EQUATIONS

The condition (necessary and sufficient) for a nontrivial zero of a weight 1 3*j* or 6*j* coefficient can be derived directly from the explicit alternating sum expression for the respective coefficient as described in detail in Ref. 1. Because of the symmetries of these coefficients, the conditions for a zero can be stated in various equivalent ways. One such statement is the following.

The nontrivial zeros of weight 1 3*j* coefficients are given by

$$\begin{pmatrix} (x+u)/2 & (y+v)/2 & (x+y+u+v-2)/2 \\ (x-u)/2 & (y-v)/2 & (-x-y+u+v)/2 \end{pmatrix} = \begin{bmatrix} 1 & x+u-1 & y+v-1 \\ u+v-1 & x & y \\ x+y-1 & u & v \end{bmatrix} = 0, \tag{3a}$$

for every set (x, y,u,v) of positive integers that satisfy the relation

$$(x+u)(y-v) = (x-u)(y+v). \tag{3b}$$

The nontrivial zeros of weight 1 6*j* coefficients are given by

$$\begin{Bmatrix} (x+u+v-1)/2 & (y+u+w-1)/2 & (x+y+v+w-2)/2 \\ (x+w)/2 & (y+v)/2 & (x+y+u-1)/2 \end{Bmatrix} = \begin{pmatrix} 1 & x & y \\ u & x+u-1 & y+u-1 \\ v & x+v-1 & y+v-1 \\ w & x+w-1 & y+w-1 \end{pmatrix} = 0, \tag{4a}$$

for every set (x, y,u,v,w) of positive integers that satisfy the relation

$$xy(x+y+u+v+w) = uvw. \tag{4b}$$

We have introduced the (x, y,u,v,w) variables into the 3*j* and 6*j* coefficients in Eqs. (3) and (4) for two reasons: (i) this simplifies the entries in the Regge and Bargmann arrays; and, more importantly, (ii) it reduces all domain requirements on the entries to the condition that the variables $(x,...,w)$ be nonnegative integers, in general, and positive integers for the problem at hand. The physical angular momentum quantum numbers can be recovered by comparing the entries in the $3j$ symbol in Eq. (3a) [resp. in the $6j$ symbol in Eq. (4a)] with those in the $3j$ symbol in Eq. (1) [resp. in the $6j$ symbol in Eq. (2)].

Next, we transform conditions (3b) [resp. (4b)] for a zero of a $3j$ coefficient [resp. $6j$ coefficient] to a standard-type Diophantine equation. For this, we use the identities

$$4AB=(A+B)^2-(A-B)^2, \tag{5a}$$

$$\begin{aligned}24ABC=&(A+B+C)^3+(A-B-C)^3\\&+(-A+B-C)^3+(-A-B+C)^3.\end{aligned} \tag{5b}$$

In Eq. (3b), we apply identity (5a) to each side of the relation, rearrange terms, and obtain

$$X^2+Y^2=U^2+V^2, \tag{6a}$$

where

$$\begin{aligned}&X=x+y+u-v, \quad Y=x-y-u-v,\\&U=x+y-u+v, \quad V=x-y+u+v,\end{aligned} \tag{6b}$$

with inverse

$$\begin{aligned}&x=(X+Y+U+V)/4, \quad y=(X-Y+U-V)/4,\\&u=(X-Y-U+V)/4, \quad v=(-X-Y+U+V)/4.\end{aligned} \tag{6c}$$

In Eq. (4b), we first define the variable z by

$$z=x+y+u+v+w, \tag{7}$$

apply identity (5b) to $xyz=uvw$, eliminate z from the resulting expression, and obtain

$$X^3+Y^3+Z^3=U^3+V^3+W^3, \tag{8a}$$

$$X+Y+Z=U+V+W, \tag{8b}$$

where

$$\begin{aligned}&X=2x+u+v+w, \quad U=u+v-w,\\&Y=2y+u+v+w, \quad V=u-v+w,\\&Z=u-v-w, \quad W=2x+2y+u+v+w.\end{aligned} \tag{9a}$$

The variables (x,y,z,u,v,w) may be obtained from (X,Y,Z,U,V,W) by

$$\begin{aligned}&x=(W-Y)/2, \quad u=(U+V)/2,\\&y=(W-X)/2, \quad v=(U-Z)/2,\\&z=(X+Y)/2, \quad w=(V-Z)/2.\end{aligned} \tag{9b}$$

Since these relations and $X+Y+Z=U+V+W$ imply $z=x+y+u+v+w$, the "inverse relations" (9b) give values of $(x,...,w)$ uniquely, although their expression in terms of $(X,...,W)$ is not unique.

We see from the transformation (6b) [resp. (9a)] that every integral solution (x,y,u,v) [resp. (x,y,u,v,w)] of $(x+u)(y-v)=(x-u)(y+v)$ [resp. of $xy(x+y+u+v+w)=uvw$] is also a solution of $X^2+Y^2=U^2+V^2$ [resp. $X^3+Y^3+Z^3=U^3+V^3+W^3$, $X+Y+Z=U+V+W$], but the converse need not be true for every solution. We turn next to the discussion of the solutions of these equations and the relations between them.

III. THE ZEROS OF WEIGHT 1 3j COEFFICIENTS

The complete result for the nontrivial zeros of weight 1 $3j$ coefficients is given by the following theorem.

Theorem 1: All nontrivial zeros of weight 1 $3j$ coefficients, that is, all zeros of

$$\begin{bmatrix}1 & x+u-1 & y+v-1\\ u+v-1 & x & y\\ x+y-1 & u & v\end{bmatrix}=0 \tag{10a}$$

are given by

$$x=\alpha\beta, \quad y=\beta\delta, \quad u=\alpha\gamma, \quad v=\gamma\delta, \tag{10b}$$

where α, β, γ, and δ assume all positive integral values.

Proof: The set of all solutions of the Diophantine equation $X^2+Y^2=U^2+V^2$ was given in 1906 by P. Pasternak in a form suitable for the present discussion [see Dickson[6] for this result (p. 252) and related references]. It is, however, easy to obtain the proof of Theorem 1 by using Eq. (3b) directly. Accordingly, we first give this proof and then show how the same result can be obtained from Pasternak's solution.

We verify directly that the (x,y,u,v) defined by Eq. (10b) satisfy $(x+u)(y-v)=(x-u)(y+v)$. Thus, we may complete the proof by showing that every positive solution of this Diophantine equation may be put in the form given by Eq. (10b). Let p denote the greatest common divisor of the positive integers $x+u$ and $y+v$, so that $x+u=\alpha p$ and $y+v=\delta p$, where α and δ are relatively prime positive integers. It follows from Eq. (3b) that $\alpha(y-v)=\delta(x-u)$, and, since α and δ are relatively prime, we find $x-u=\alpha q$ and $y-v=\delta q$ for some integer q with $|q|<p$. Defining β and γ by $\beta=(p+q)/2$, $\gamma=(p-q)/2$, we thus obtain $x=\alpha\beta$, $y=\beta\delta$, $u=\alpha\gamma$, and $v=\gamma\delta$. Since x and y are positive integers and α,δ are relatively prime, it follows that β is a positive integer; hence, p and q have the same parity. Thus, γ is also an integer, which is positive, since the property $x+u>x-u$ for all positive integers u implies that $p>q$. Thus, every positive integer solution (x,y,u,v) of Eq. (3b) can be written in the form (10b). This completes the direct proof of the theorem.

A proof based on Pasternak's result may be given as follows: Pasternak proved that all solutions of the Diophantine equation $X^2+Y^2=U^2+V^2$ are given by

$$\begin{aligned}&X=kr+ls, \quad Y=lr-ks,\\&U=kr-ls, \quad V=lr+ks,\end{aligned}$$

where k, l, r, s are integers. Substitution of these relations into Eqs. (6c) yields

$$\begin{aligned}&x=r(k+l)/2, \quad y=r(k-l)/2,\\&u=s(k+l)/2, \quad v=s(k-l)/2.\end{aligned}$$

Thus, the set of all positive integral solutions of $(x+u)(y-v)=(x-u)(y+v)$ is obtained from the subset of all solutions of $X^2+Y^2=U^2+V^2$ by choosing k, l, r, s to be positive integers with $k>l$ and (i) either $k+l$ odd and r,s both even, or (ii) $k+l$ even. This gives exactly the solution (10b). ■

Remarks: (a) The solution (10) for the nontrivial zeros of weight 1 $3j$ coefficients was given in Ref. 1. Here we have shown, in addition, that this solution gives all such zeros.

(b) It is interesting that the zeros of weight 1 $3j$ coefficients occur for $m_1/j_1 = m_2/j_2$. [This is Eq. (3b) written in terms of the angular momentum quantum numbers.] Classically, this condition occurs when the angular momenta $\mathbf{J}_1$ and $\mathbf{J}_2$ have the same projection on the "z axis" (belong to the same cone).

IV. THE ZEROS OF WEIGHT 1 6*j* COEFFICIENTS

We require several preliminary results before giving the zeros of the weight 1 $6j$ coefficients. For this we introduce the following notations: Consider the 3×3 array (within the box) of positive integers $k,l,...,t$ given by

$$\begin{array}{|ccc|c} \hline k & l & m & x \\ n & p & q & y \\ r & s & t & z\,. \\ \hline u & v & w & \end{array} \tag{11}$$

The integers x,y,z at the right end of a row are defined to be the *product* of the entries in the corresponding row ($x = klm$, etc.); similarly, u,v,w are defined to be the product of the entries in the corresponding column. We next defined the set $\mathbf{K}^6$ by

$$\mathbf{K}^6 = \begin{cases} \text{set of all 6-tuples } (x,y,z,u,v,w) \text{ obtained} \\ \text{by letting } k,l,...,t \text{ in the array (11)} \\ \text{assume all positive integral values.} \end{cases} \tag{12}$$

We also denote the set of all 6-tuples with positive integral entries by $\mathbf{N}^6_+$. Clearly, $\mathbf{K}^6 \subset \mathbf{N}^6_+$.

We can now prove the following theorem.

Theorem 2: The set of all positive solutions of the Diophantine equation $xyz = uvw$ is given by $\mathbf{K}^6$.

Proof: Each element (x,y,z,u,v,w) in $\mathbf{K}^6$ clearly satisfies $xyz = uvw$. Thus, the principal part of the proof is in showing that every 6-tuple (x,y,z,u,v,w) in $\mathbf{N}^6_+$ that satisfies $xyz = uvw$ belongs to $\mathbf{K}^6$; that is, in showing the existence of positive integers $k,l,...,t$ in the array (11) that yield each solution of the Diophantine equation.

Let (x,y,z,u,v,w) denote a positive solution of the Diophantine equation $xyz = uvw$ and define $N = xyz$. Suppose (induction hypothesis) that for each $n = 1,2,...,N-1$, we have proved that each positive solution of $n = x'y'z' = u'v'w'$ is obtained from an array of the form (11). We now extend this result to $n = N$.

Suppose at least one of the integers x, y, z has a common divisor with at least one of the integers u,v,w. Without loss of generality, we can assume these to be x and u, so that $x = \alpha x'$ and $u = \alpha u'$ for some positive integer $\alpha > 1$. Then, since $xyz = uvw$, we find that $x'yz = u'vw = n < N-1$. Hence, by the induction hypothesis, each positive solution of this latter Diophantine equation may be obtained from an array (11), say, with entries $k',l',...,t'$. The array with entries $\alpha k',l',...,t'$ then yields the solution $(x,y,...,w)$. If each of the integers x,y,z is relatively prime to each of the integers u,v,w, this argument does not apply. In this case, we suppose that at least one of the integers x,y,z has a common divisor with at least one of the integers uv,uw,vw. Without loss of generality, we can assume these to be x and uv, so that $x = \alpha\beta x'$, $u = \alpha u'$, $v = \beta v'$ for some positive integers $\alpha > 1$ and $\beta > 1$. The existence of the two divisors, α of u and β of v, is a consequence of the assumption that x and u, as well as x and v, are relatively prime. Thus, we find that $x'yz = u'v'w = n < N-1$. Hence, by the induction hypothesis, each positive solution of this latter Diophantine equation may be obtained from an array (11), say, with entries $k',l',...,t'$. The array with entries $\alpha k',\beta l',m',...,t'$ then yields the solution $(x,y,...,w)$. If each of the integers x,y,z is relatively prime to each of the integers uv,uw,vw, the argument up to this point does not apply. In this case, it follows from $yz = uvw/x$ and the assumption that x does not divide uv,uw,vw that there exist positive integers α,β,γ each greater than 1, such that $u = \alpha u'$, $v = \beta v'$, $w = \gamma w'$, and $x = \alpha\beta\gamma$. Thus, $yz = u'v'w' = n < N-1$. Hence, by the induction hypothesis, each positive solution to this latter Diophantine equation may be obtained from an array (11) with entries $1,1,1,n',p',...,t'$. The array with entries $\alpha,\beta,\gamma,n',...,t'$ then yields the solution $(x,y,...,w)$. This last step completes the induction loop; that is, the induction hypothesis and the property $xyz = uvw = N$ imply the validity of the hypothesis at level N. Since we easily verify that every positive solution of $xyz = uvw$ can be obtained from an array (11) for small N, say, $N = 1,2,...,7$, the theorem is proved. ∎

An immediate consequence of Theorem 2 and Eqs. (4), which define a nontrivial zero, is the following theorem.

Theorem 3: All nontrivial zeros of weight 1 $6j$ coefficients, that is, all nontrivial zeros of

$$\begin{pmatrix} 1 & x & y \\ u & x+u-1 & y+u-1 \\ v & x+v-1 & y+v-1 \\ w & x+w-1 & y+w-1 \end{pmatrix} = 0 \tag{13a}$$

are given by the points (x,y,z,u,v,w) in $\mathbf{K}^6$ that obey the condition

$$z = x + y + u + v + w. \tag{13b}$$

Remarks: (a) The nontrivial zeros of weight 1 $6j$ coefficients given in Theorem 3 were obtained in Ref. 1. Here we have shown that this solution gives all such zeros.

(b) The solution for the nontrivial zeros of weight 1 $6j$ coefficients given in Theorem 3 is not fully explicit, since the subset of points in $\mathbf{K}^6$ that satisfy the auxiliary condition (13b) has not been determined.

We next consider the relationship of the nontrivial zeros given in Theorem 3 to the solution of the pair of Diophantine equations

$$X^3 + Y^3 + Z^3 = U^3 + V^3 + W^3, \tag{14a}$$

$$X + Y + Z = U + V + W. \tag{14b}$$

From Eqs. (9b) we find that each solution of Eqs. (14a) and (14b) with

$$X,Y,...,W \text{ of the same parity} \tag{14c}$$

and such that

$$W > X > Y, \quad U > V > Z \tag{14d}$$

yields a positive solution $(x,y,...,w)$ of $xyz = uvw$,

$z = x + y + u + v + w$. (It is no restriction to assume that $X > Y > Z$ and $U > V > W$.) Moreover, the set of all solutions of Eqs. (14) yields the set of all zeros $(x, y, ..., w)$ of weight 1 $6j$ coefficients.

It appears that the general solution of the pair of Diophantine equations (14a) and (14b) is not known. A two-parameter set of solutions was given in 1915–16 by Gérardin (see Dickson,[6] pp. 565 and 713), but the survey paper on "Equal sums of like powers" by Lander *et al.*[7] does not include Eqs. (14a) and (14b) among their list of solved problems. The special parametric solution by Gérardin is, however, very useful for the present problem, as we next discuss.

Gérardin's solution (with an appropriate renaming of variables) is

$$\begin{aligned} X &= 2p^2 - 10pq + 12q^2, \quad U = 2pq, \\ Y &= p^2 - 5pq + 6q^2, \quad V = pq, \\ Z &= -2p^2 + 9pq - 6q^2, \quad W = p^2 - 9pq + 12q^2, \end{aligned} \tag{15a}$$

where p and q are arbitrary integers. Conditions (14c) and (14d) are satisfied for all

$$p \text{ even}, \quad q > p. \tag{15b}$$

An array (11) corresponding to the solution (15a) may be verified to be

$$\begin{array}{ccc|c} q & 3q-2p & 1 & x \\ p/2 & 1 & q-p & y \\ 3 & (2q-p)/2 & 3q-p & z \\ \hline u & v & w & \end{array} \tag{16}$$

Thus, for all p even and all $q > p$, the point $(x, y, ..., w)$ obtained from this array is a zero of a weight 1 $6j$ coefficient.

V. CONCLUSIONS

We have shown that the nontrivial zeros of weight 1 $3j$ and $6j$ coefficients given in Ref. 1 are all such zeros. We have also related the Diophantine equations that occur in the direct formulation of the conditions for nontrivial zeros [Eqs. (3) and (4)] to well-known Diophantine equations considered in the literature on number theory. For the zeros of weight 1 $3j$ coefficients, we have been able to give a complete solution; for weight 1 $6j$ coefficients, the solution is not fully explicit, although complete. A parametric family of zeros of weight 1 $6j$ coefficients has been given explicitly by using results from the literature on number theory.

ACKNOWLEDGMENTS

The work of one of us (J.D.L.) was performed under the auspices of the U. S. Department of Energy under Contract No. W-7405-ENG-36 and the Office of Basic Energy Sciences, Department of Applied Mathematics.

[1]S. Brudno, "Nontrivial zeros of the Wigner (3-j) and Racah (6-j) coefficients. I. Linear solutions," J. Math. Phys. **26**, 434 (1985).
[2]T. Regge, "Symmetry properties of Clebsch–Gordan coefficients," Nuovo Cimento **10**, 544 (1958).
[3]V. Bargmann, "On the representations of the rotation group," Rev. Mod. Phys. **34**, 829 (1962).
[4]L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics, Encyclopedia of Mathematics and Its Applications*, Vol. 8, edited by G-C. Rota (Addison–Wesley, Reading, MA, 1981).
[5]S. Brudno and J. D. Louck, "Nontrivial zeros of the Racah quadrupole invariant," J. Math. Phys. **26**, 1125 (1985).
[6]L. E. Dickson, *Theory of Numbers* (Chelsea, New York, 1952), Vol. II.
[7]L. J. Lander, T. R. Parkin, and J. L. Selfridge, "A survey of equal sums of like powers," Math. Comp. **21**, 446 (1967).

Locally operating realizations of transformation Lie groups

Jose F. Cariñena
Departamento de Física Teórica, Facultad de Ciencias, Universidad de Zaragoza, 50009 Zaragoza, Spain
Mariano A. Del Olmo[a),b)] and Mariano Santander
Departamento de Física Teórica, Facultad de Ciencias, Universidad de Valladolid, 47005 Valladolid, Spain

(Received 9 August 1984; accepted for publication 15 March 1985)

Using the Mackey theory of induced representations, a systematic study of the locally operating multiplier realizations of a connected Lie group G that acts transitively on a space-time manifold is presented. We obtain a mathematical characterization of the locally operating multiplier realizations and a reduction of the problem of multiplier locally operating realizations to linear ones via a splitting group $\bar{G}$ for G. In this way the locally operating multiplier realizations are obtained by induction from finite-dimensional linear representations of a well-determined subgroup of $\bar{G}$. Some examples, such as the two-dimensional Euclidean group, the Galilei group, and the one-dimensional Newton–Hooke group, are given.

I. INTRODUCTION

In a recent paper[1] the appropriate mathematical framework for the study of linear locally operating representations of transitive transformations Lie groups has been established, a concept whose relevance had been pointed out by Hoogland[2] some years ago, based on some previous comments of Bargmann and Wigner.[3] The character of rays rather than vectors of the mathematical objects describing the pure states of a quantum system[4,5] suggests the convenience of studying the projective representations corresponding to such locally operating representations of transitive transformation groups. This is the problem we have dealt with in a recent paper,[6] where we investigated the possibility of finding one group $\bar{G}$ such that any locally operating (multiplier) realization of G can be lifted to a linear representation of $\bar{G}$. It is to be remarked that we physicists are used to handling vectors instead of rays and consequently linear operators in place of projective ones.[4,5] This explains the interest of the so-called multiplier (or up to a factor) representations in physics. They are not homomorphisms because of the presence of a factor system ω of G. The point is that not every factor system of the group G can arise in a locally operating realization as was analyzed in Ref. 7, and the knowledge of such factor systems is a necessary step for the determination of a "minimal" splitting group for local realizations, also called a local splitting group.[6] We follow the method proposed by Schur[8] which has been used more recently in Refs. 9–12 and in Ref. 6 when only locally operating realizations are considered. The concept of equivalence deserves a deeper analysis and it will be studied in this paper according to the method pointed out.

The organization of this paper is as follows: In Sec. II we recall the main definitions about locally operating realizations and the way of dealing with the linear case[13] to which the general case can be reduced, as Sec. III proves. In Sec. III we also analyze the way of lifting locally operating realizations of G to linear ones of a splitting group $\bar{G}$; the study of the local or gauge equivalence is carried out in Sec. IV. Section V is devoted to giving an explicit way of building up a complete set of representatives of the gauge equivalence classes and in Sec. VI we present some particular examples for illustrating how the theory we have developed works.

[a)] In part from his Ph. D. thesis, Universidad de Valladolid, 1983.
[b)] Present address: CRM, Université de Montreal, Case Postale 6128 Succursale "A," Montreal, Province of Quebec, H3C 3J7, Canada.

II. LOCALLY OPERATING REALIZATIONS OF TRANSFORMATION GROUPS

Let G be a connected Lie group acting transitively on a differentiable manifold X. The isotopy group Γ of a fixed point x_0 is closed and the homogeneous space G/Γ can be endowed[14] with a differentiable structure in such a way that the projection $\pi:G\to G/\Gamma$ is differentiable and there are local differentiable sections $s:G/\Gamma\to G$ for π, i.e., $\pi\circ s=\mathrm{id}_{G/\Gamma}$. When endowed with this differentiable structure, G/Γ is diffeomorphic to X and will be identified to it.

The concepts of projective and multiplier representations of G that we are going to use are those of a previous paper[12] plus some additional topological requirements:linear and projective representations are continuous but multiplier representations are only Borel maps. This is so because Borel multiplier representations are related to continuous projective representations as indicated in Ref. 6.

Definition 1: A locally operating realization (or locally operating multiplier representation) of G is a Borel multiplier representation of G, U, in which the representation space is made up by vector-valued functions, $f:X\to\mathbb{C}^n$, and the representation of G is given by

$$[U(g)f](gx)=A(g,x)f(x),$$

where A is a matrix-valued Borel function, $A:G\times X\to \mathrm{GL}(n,\mathbb{C})$, satisfying

$$A(g_2,g_1x)A(g_1,x)=\omega(g_2,g_1)A(g_2g_1,x).$$

The function $\omega:G\times G\to T$ is the factor system of the representation $U,\omega\in Z^2(G,T)$, and A is called a gauge matrix.

The operators $U(g)$ of the representation are local operators because the support of $U(g)f$ is contained in the image by g of the support of f. As far as equivalence is concerned, such a local character must be preserved and this leads to a modification of the usual concept of equivalence, which will be called local equivalence or gauge equivalence. We recall that

 0022-2488/85/092096-11$02.50

for multiplier representations the relevant concept is that of pseudoequivalence.[12]

Definition 2: Two locally operating realizations of G, U and U', both operating on the space of vector-valued functions, are called gauge pseudoequivalent if there are a Borel function $\lambda: G \to T$ and a linear operator τ in the representation space which acts locally [i.e., $(\tau f)(x) = S(x) f(x)$, with S a nonsingular matrix] and such that $U'(g) = \lambda(g) \tau U(g) \tau^{-1}$, $\forall g \in G$. The corresponding gauge matrices will be related by $A'(g,x) = \lambda(g) S(gx) A(g,x) S^{-1}(x)$.

The particular case of linear representations, in which $\omega = 1$ and λ in Definition 2 does not appear, has recently been considered[1,13] and it was shown that any locally operating linear representation of G is gauge equivalent to the representation of G induced by the representation $\sigma(\gamma) = A(\gamma, x_0)$ of Γ. On the other hand, the problem of the determination of the projective and multiplier representations of a group is solved by looking for a "splitting" group[11,12] $\bar{G}$, i.e., a group $\bar{G}$ (which is not uniquely defined) such that any projective (or multiplier) representation of G can be lifted to a linear representation of $\bar{G}$. When only local realizations of G are considered, only some factor systems can arise and the corresponding splitting and representation groups can eventually be chosen to be of lower dimension.[6] The point is that the correspondence between equivalence classes of projective representations of G and linear representations of $\bar{G}$ is not one-to-one and in the case of locally operating representations the difference is more relevant and will be studied next.

III. LIFTING LOCALLY OPERATING REALIZATIONS OF G

Let $(\bar{G}, p)$ be a local splitting group for G, i.e., $p: \bar{G} \to G$ is an epimorphism and $\bar{G}$ is such that any projective representation of G defined by a locally operating (multiplier) realization of G can be lifted to a linear representation of $\bar{G}$ mapping $\operatorname{Ker} p$ in the circle group T. If the action of $\bar{G}$ on X is defined via the projection p, the isotopy group $\bar{\Gamma} = p^{-1}(\Gamma)$ is the middle group of an extension of Γ by $\operatorname{Ker} p$. We are going to prove that the lifting to $\bar{G}$ of any locally operating realization of G is a locally operating linear representation of $\bar{G}$ and therefore when $\bar{G}$ is a Lie group the theory developed in Ref. 1 can be carried out for finding the locally operating realizations of G.

Definition 3: Let $\mathscr{A}$ be the kernel of the morphism $p: \bar{G} \to G$. A linear represenation R of $\bar{G}$ is said to be $\mathscr{A}$-split if $R(\mathscr{A}) \subset T$. These representations define by quotient projective representations of G.

Proposition 1: For each normalized Borel section $\rho: G \to \bar{G}$, and for each $\mathscr{A}$-split locally operating linear representation R of $\bar{G}$ there is a locally operating (multiplier) realization U of G given by $U(g) = (R \circ \rho)(g)$, $\forall g \in G$. The corresponding factor system of U will be $R \circ W$, with W the factor system of the topological extension $1 \to \mathscr{A} \to \bar{G} \underset{\rho}{\overset{p}{\rightleftarrows}} G \to 1$, defined by the Borel section ρ.

Proof: According to the definition of U, we have

$$(U(g)f)(gx) = [(R \circ \rho)(g) f](gx)$$
$$= (R(\rho(g)) f)(\rho(g) x) = \bar{A}(\rho(g), x) f(x),$$

and if we denote by $A(g,x)$ the matrix $\bar{A}(\rho(g),x)$ we see that

$$A(g_1, g_2 x) A(g_2, x) A^{-1}(g_1 g_2, x) = A((W(g_1, g_2), 1), x)$$

and the identification $\omega(g_1, g_2) 1 = \bar{A}((W(g_1, g_2), 1), x)$ follows because of the relation

$$[R(W(g_1, g_2), 1) f]((W(g_1, g_2), 1) x)$$
$$= [\bar{R}((W(g_1, g_2), 1)) f](x)$$
$$= \bar{A}((W(g_1, g_2), 1), x) f(x)$$

and because R is an $\mathscr{A}$-split realization when $R((W(g_1, g_2), 1)) \in T$.

If we would take different normalized Borel sections of G into $\bar{G}$ the locally operating realizations of G obtained by means an $\mathscr{A}$-split locally operating linear representation of $\bar{G}$ would be similar and they would give the same projective representation of G.

The converse of Proposition 1 is the following one.

Proposition 2: For each locally operating (multiplier) realization U and G there is another similar one which can be lifted to an $\mathscr{A}$-split locally operating linear representation of $\bar{G}$.

Proof: Let R be a linear representation of $\bar{G}$ lifting U and ρ a Borel section, $\rho: G \to \bar{G}$. Then $U' = R \circ \rho$ is a multiplier representation of G similar to U.

Therefore there exists a function $\lambda: G \to T$ such that $U'(g) = \lambda(g) U(g)$ and then U' is also a local realization of G with a gauge matrix $A'(g,x)$. Now, if $\bar{g} = a\rho(g) \in \bar{G}$, with $a \in \mathscr{A}$, then R being an $\mathscr{A}$-split representation

$$[R(\bar{g}) f](\bar{g} x) = R((a,1)) A'(g,x) f(x)$$

and it is enough to define the gauge matrix function $\bar{A}(\bar{g},x) = R((a,1)) A'(g,x)$, which is a linear gauge matrix function of $\bar{G}$, i.e., its factor system is trivial.

The above propositions, whose results were announced in a previous paper,[13] give us the main lines for obtaining the locally operating realization of G by means of the $\mathscr{A}$-split locally operating linear representations of $\bar{G}$, which can be obtained inducing from some linear representations of $\bar{\Gamma}$ to be characterized in the next proposition.

Proposition 3: The linear representations of $\bar{\Gamma}$ inducing $\mathscr{A}$-split representations of $\bar{G}$ are those representations mapping $\mathscr{A} \subset \bar{\Gamma}$ in T.

Proof: Let $\bar{\sigma}$ be a representation of $\bar{\Gamma}$ inducing an $\mathscr{A}$-split locally operating linear representation R of $\bar{G}$. The gauge matrix will be $\bar{A}(\bar{g},x) = \bar{\sigma}(\bar{s}^{-1}(\bar{g}x) \bar{g} \bar{s}(x))$, with $\bar{s}$ being a normalized Borel section $\bar{s}: \bar{G}/\bar{\Gamma} \to \bar{G}$. Then $\bar{A}(a,x) = \bar{\sigma}(\bar{s}^{-1}(x)\, a \bar{s}(x)) = \bar{\sigma}(a)$, because $\bar{G}$ is a central extension and $\mathscr{A} \subset \bar{\Gamma}$. Consequently $[R(a) f](x) = \bar{\sigma}(a) f(x)$ and the necessary and sufficient condition for R to be $\mathscr{A}$-split is $\bar{\sigma}(\mathscr{A}) \subset T$.

Proposition 3 means that the interesting representations of $\bar{\Gamma}$ (i.e., those inducing $\mathscr{A}$-split representations of $\bar{G}$) are precisely the $\mathscr{A}$-split representations of $\bar{\Gamma}$, according to our Definition 3.

IV. GAUGE EQUIVALENCE

It is well known that inequivalent linear representations of $\bar{G}$ can give rise to equivalent projective representations of G and therefore the relevant concept is that of pseudoequivalence of representations of $\bar{G}$. If the local character is also to be taken into account we are led to the following definition.

Definition 4: Two locally operating linear representations of $\bar{G}$, R, and R' are said to be gauge pseudoequivalent if there is an operator τ on the support space of the functions of X into $\mathbb{C}^n$ that acts locally and there is a homomorphism $\bar{\lambda}:\bar{G}\to T$, such that

$$R'(\bar{g}) = \bar{\lambda}(\bar{g})\tau R(\bar{g})\tau^{-1}, \quad \forall \bar{g}\in\bar{G}.$$

The gauge matrices associated to R and R' are related by

$$\bar{A}'(\bar{g},x) = \bar{\lambda}(\bar{g})S(\bar{g}x)\bar{A}(\bar{g},x)S^{-1}(x),$$
$$\forall \bar{g}\in\bar{G}, \quad \forall x\in X,$$

where S is the nonsingular matrix valued function, $S:X\to \mathrm{GL}(n\mathbb{C})$ defining the local operator τ by $(\tau f)(x) = S(x)f(x)$, $\forall x\in X$.

With this definition and a straightforward generalization of the results of Ref. 12 we can state the following theorem.

Theorem 1: If $(\bar{G},p)$ is a local splitting group for G, for each normalized Borel section $\rho:G\to\bar{G}$, we can define a one-to-one correspondence between the gauge equivalence classes of the locally operating (multiplier) realizations of G and the gauge pseudoequivalence classes of the $\mathscr{A}$-split locally operating linear representations of $\bar{G}$.

As a consequence of this theorem, if we take different (normalized) Borel sections, $\bar{s}$ and $\bar{s}'$, of X into $\bar{G}$, with a fixed finite-dimensional linear representation $\bar{\sigma}$ of $\bar{\Gamma}$ and a (normalized) Borel section $\rho:G\to\bar{G}$, we will obtain gauge pseudoequivalent locally operating (multiplier) realizations of G.

On the other hand, if we take two different normalized Borel sections of G into $\bar{G}$, ρ and ρ', and with a fixed $\mathscr{A}$-split locally operating linear representation R of $\bar{G}$, the realizations $R\circ\rho$ and $R\circ\rho'$ of G are gauge pseudoequivalent.

The fundamental property concerning gauge pseudoequivalence of representations of $\bar{G}$ when inducing from representations of $\bar{\Gamma}$ is given by the following theorem.

Theorem 2: Two $\mathscr{A}$-split pseudoequivalent finite-dimensional linear representations of $\bar{\Gamma}$ induce gauge pseudoequivalent $\mathscr{A}$-split locally operating linear representations of $\bar{G}$ if and only if the homomorphism defining the pseudoequivalence of the linear representations of $\bar{\Gamma}$ can be extended to a homomorphism of $\bar{G}$ on T.

Proof: The condition is necessary. In fact, if R and R' are gauge pseudoequivalent, there exists $\bar{\lambda}\in \mathrm{Hom}(\bar{G},T)$ and an invertible local operator τ such that $R'(\bar{g}) = \lambda(\bar{g})\tau \times R(\bar{g})\tau^{-1}$. The gauge matrices of both representations will be related by $\bar{A}'(\bar{g},x) = \bar{\lambda}(\bar{g})S(\bar{g}x)\bar{A}(\bar{g},x)S^{-1}(x)$ with S being the matrix function defining the local operator τ. If we look at the restrictions of the gauge matrices to $\bar{\Gamma}\times\{x_0\}$ we find that these linear representations of $\bar{\Gamma}$ inducing R and R' are pseudoequivalent; the homomorphism of $\bar{\Gamma}$ on T realizing the equivalence is $\bar{\lambda}_{|\bar{\Gamma}}$, which can obviously be extended to the homomorphism $\bar{\lambda}$ of $\bar{G}$ on T.

Conversely, let $\bar{\sigma}$ and $\bar{\sigma}'$ be two pseudoequivalent finite-dimensional $\mathscr{A}$-split linear representations of $\bar{\Gamma}$. There exist a matrix V and a homomorphism Λ of $\bar{\Gamma}$ on T such that

$$\bar{\sigma}'(\bar{\gamma}) = \Lambda(\bar{\gamma})V\bar{\sigma}(\bar{\gamma})V^{-1}, \quad \forall\bar{\gamma}\in\bar{\Gamma}.$$

Let $\bar{\lambda}\in \mathrm{Hom}(\bar{G},T)$ to be such that $\bar{\lambda}_{|\bar{\Gamma}} = \Lambda$. If R and R' are the $\mathscr{A}$-split locally operating linear representations of $\bar{G}$ induced by $\bar{\sigma}$ and $\bar{\sigma}'$, respectively, we have that

$$\begin{aligned}(R'(\bar{g})f)(\bar{g}x) &= \bar{A}'(\bar{g},x)f(x) = \bar{\sigma}'(\bar{s}^{-1}(\bar{g}x)\overline{gs}(x))f(x)\\ &= \Lambda(\bar{s}^{-1}(\bar{g}x)\overline{gs}(x))V\bar{\sigma}(\bar{s}^{-1}(\bar{g}x)\overline{gs}(x))V^{-1}f(x),\end{aligned}$$

and, since $\bar{\lambda}_{|\bar{\Gamma}} = \Lambda$, we find

$$\begin{aligned}(R'(\bar{g})f)(\bar{g}x) &= \bar{\lambda}(\bar{s}^{-1}(\bar{g}x)\overline{gs}(x))V\bar{\sigma}(\bar{s}^{-1}(\bar{g}x)\overline{gs}(x))V^{-1}f(x)\\ &= \bar{\lambda}(\bar{s}^{-1}(\bar{g}x))\bar{\lambda}(\bar{g})\bar{\lambda}(\bar{s}(x))\\ &\quad\times V\bar{\sigma}(\bar{s}^{-1}(\bar{g}x)\overline{gs}(x))V^{-1}f(x).\end{aligned}$$

Now we can define the local operator τ by $(\tau f)(x) = \bar{\lambda}(\bar{s}^{-1}(x))Vf(x)$ and this operator and the homomorphism $\bar{\lambda}$ give the gauge pseudoequivalence of R and R'.

The result of this theorem suggests the following definition.

Definition 5: Two $\mathscr{A}$-split finite-dimensional linear representations of $\bar{\Gamma}$ are called superequivalent if they are pseudoequivalent and this equivalence can be realized by means of a homomorphism of $\bar{\Gamma}$ on T that can be extended to a homomorphism of $\bar{G}$ on T.

The superequivalence of the linear representations of $\bar{\Gamma}$ is an equivalence relation that splits each pseudoequivalence class of $\mathscr{A}$-split linear representations of $\bar{\Gamma}$ into subclasses of superequivalent representations.

A complete set of superequivalence classes of $\mathscr{A}$-split finite-dimensional linear representations of $\bar{\Gamma}$ with representatives $\bar{\sigma}' = \bar{\lambda}\bar{\sigma}$ is obtained when $\bar{\lambda}$ runs through all the equivalence classes of one-dimensional representations of $\bar{\Gamma}$ modulo the representations that can be extended to one-dimensional representations of $\bar{G}$, and $\bar{\sigma}$ runs all the pseudoequivalence classes of finite-dimensional $\mathscr{A}$-split linear representations of $\bar{\Gamma}$.

V. THE EXPLICIT CONSTRUCTION OF THE GAUGE EQUIVALENCE CLASSES

The method for the construction of a complete set of gauge pseudoequivalence classes of locally operating representations of $\bar{G}$ is based on the earlier properties and it is summarized as follows.

Theorem 3: We obtain a complete set of gauge pseudoequivalence classes of $\mathscr{A}$-split locally operating linear representations of $\bar{G}$ induced by finite-dimensional $\mathscr{A}$-split linear representations of $\bar{\Gamma}$ with representatives with gauge matrices given by

$$\bar{A}(\bar{g},x) = \bar{\sigma}(\bar{s}^{-1}(\bar{g}x)\overline{gs}(x)),$$

when $\bar{\sigma}$ is a representative running through the superequivalence classes of $\mathscr{A}$-split representations of $\bar{\Gamma}$, and $\bar{s}$ is a normalized Borel section of $\bar{G}/\bar{\Gamma}$ on $\bar{G}$.

As we also know the method for building up the representatives of all the superequivalence classes of the representations of $\bar{\Gamma}$, it is straightforward to prove that their gauge matrices have the form

$$\bar{A}(\bar{g},x) = \bar{\lambda}(\bar{\gamma}(\overline{gs}(x)))\bar{\sigma}(\bar{s}^{-1}(\bar{g}x)\overline{gs}(x)),$$

where $\bar{\lambda}$ runs through the set of equivalence classes of one-dimensional representations of $\bar{\Gamma}$ modulo those that can be extended to $\bar{G}$ and $\bar{\sigma}$ the pseudoequivalence classes of $\mathscr{A}$-split representations of $\bar{\Gamma}$.

In relation to the gauge pseudoequivalence classes of locally operating (multiplier) realizations of G, if we fix a normalized Borel section $\rho: G \to \bar{G}$, we obtain a representative U of each gauge pseudoequivalence class by $U = R \circ \rho$, when R runs the set of gauge pseudoequivalence classes of $\mathscr{A}$-split locally operating linear representations of $\bar{G}$.

The gauge matrices of the (multiplier) realizations of G will be

$$A(g,x) = \bar{A}(\rho(g),x)$$
$$= \bar{\lambda}(\bar{\gamma}(\rho(g)\bar{s}(x)))\bar{\sigma}(\bar{s}^{-1}(gx)\rho(g)\bar{s}(x)) \quad (1)$$

and the associated factor system is

$$\omega(g_1,g_2)1 = \bar{\lambda}((W_\rho(g_1,g_2),1))\bar{\sigma}((W_\rho(g_1,g_2),1)).$$

It is worthwhile to note that for any $\bar{W} \in H^2(G,\mathscr{A})$ and any arbitrary Borel section $s: X = G/\Gamma \to G$ we can find[6] a lift $W \in Z^2(G,\mathscr{A})$ such that $W|_{s(X)\times\Gamma} = 1$. Actually, if W' is a lift of $\bar{W}$ and $\mu: G \to T$ is defined by $\mu(g) = W'(s(gx_0),\gamma(g))$ with $g = s(gx_0)\gamma(g)$, one easily checks that $W = W'\delta\mu$ satisfies $W|_{s(X)\times\Gamma} = 1$ and $W|_{\Gamma\times\Gamma} = W'|_{\Gamma\times\Gamma}$.

Let us choose $\bar{s}(x) = (1,s(x))$, with s a normalized Borel section of G/Γ on G; if we make use of the earlier properties, then the gauge matrices appearing in formula (1) take the form

$$A(g,x) = \bar{\lambda}((W(g,s(x)),1))\bar{\sigma}((W(g,s(x)),1))$$
$$\times \bar{\lambda}((1,\gamma(gs(x))))\bar{\sigma}((1,\gamma(gs(x)))), \quad (2)$$

where

$$\omega(g,s(x)) = \bar{\lambda}((W(g,s(x)),1))\bar{\sigma}((W(g,s(x)),1))$$

is the factor system of G associated to the corresponding local (multiplier) realization.

The gauge matrices given by (2) are x_0 centered according to Hoogland's definition[2] [i.e., $A(g,x_0) = A(\gamma(g),x_0)$].

The following theorem summarizes all our results.

Theorem 4: Let $\bar{G}$ be a splitting group for G ($\rho: \bar{G} \to G$), W the factor system associated to the normalized Borel section $\rho: G \to \bar{G}$, s a normalized Borel section $s: G/\Gamma \to G$, and $\bar{\sigma}$ a representative of a class of superequivalence of finite-dimensional $\mathscr{A}$-split linear representations of $\bar{\Gamma}$. Then we obtain a complete set of the gauge pseudoequivalence classes of locally operating (multiplier) realizations of G, with representatives whose associated gauge matrices are x_0 centered, by means of

$$A(g,x) = \bar{\sigma}((1,s(gx))^{-1}\rho(g)(1,s(x)))$$
$$= \bar{\sigma}((W(g,s(x)),1))\bar{\sigma}((1,\gamma(gs(x)))),$$

when $\bar{\sigma}$ runs through all the superequivalence classes of finite-dimensional $\mathscr{A}$-split linear representations of $\bar{\Gamma}$. The factor system of the induced (locally operating) multiplier realizations of G is $\bar{\sigma} \circ W$.

An alternative form for Theorem 4 according to Definition 5 is the following.

Theorem 4a: Under the hypotheses of Theorem 4, a complete set of gauge pseudoequivalence classes of locally operating (multiplier) realizations of G is obtained with representatives whose associated gauge matrices have the following expression:

$$A(g,x) = \bar{\lambda}((W(g,s(x)),1))\bar{\sigma}((W(g,s(x)),1))$$
$$\times \bar{\lambda}((1,\gamma(gs(x))))\bar{\sigma}((1,\gamma(gs(x)))), \quad (3)$$

when $\bar{\lambda}$ and $\bar{\sigma}$ are representatives of every equivalence class of one-dimensional representations of $\bar{\Gamma}$ on T modulo those that can be extended to $\bar{G}$ and of pseudoequivalence classes of finite-dimensional $\mathscr{A}$-split linear representations of $\bar{\Gamma}$, respectively.

We also note that the locally operating representations and realizations are not irreducible, in general, even if the linear representations of the corresponding isotopy group inducing them were irreducible. We also remark that in certain cases (Mackey's theory[15] or Kirillov's theory[16]) we can obtain by induction irreducible canonical (not local) representations or realizations of Lie groups and this can be useful in the problem of finding irreducible subspaces under the local realizations. One way to solve this problem[17] is to compare the local realizations with the irreducible canonical realizations and in this way one can obtain the conditions that the functions of the locally operating realization must satisfy in order to belong to an irreducible subspace.

Finally, we want to remark that when the groups G and Γ are connected and simply connected Lie groups, the earlier results are simplified and our results coincide with those of Hoogland.[2] In this case the factor systems of Γ are to be equivalent to the trivial factor system and only the factor systems of G whose restrictions to Γ are equivalent to the trivial one could be associated to a (multiplier) locally operating realizations of G (see Ref. 7) We have that $H^2(G,T) = \mathbb{R}^m$, with $m \in \mathbb{N}$ and $H^2_{\rm loc}(G,T) = \mathbb{R}^n$, $n \in \mathbb{N}$, and $n < m$. A local splitting group $\bar{G}$ is a central extension of G by $\mathbb{R}^n$ [the dual of $H^2_{\rm loc}(G,T) = \mathbb{R}^n$ is $\mathbb{R}^n$ (see Ref. 6)] and $\bar{\Gamma} = \mathbb{R}^n \otimes \Gamma$. It follows from the structure of the direct product of $\bar{\Gamma}$ that its linear representations are tensorial products of one of $\mathbb{R}^n$ and another of Γ. Consequently, the classes of one-dimensional representations of $\bar{\Gamma}$ on T modulo extendible to $\bar{G}$ become the classes of one-dimensional representations of Γ on T modulo those extensible to G. Thus, the classes of pseudoequivalence of finite-dimensional linear representations of $\bar{\Gamma}$ agree with the equivalence classes of linear representations of Γ. Therefore, if we take the formula (3), we have

$$A(g,x) = \xi(W(g,s(x)))\Lambda(\gamma(g,s(x)))\sigma(\gamma(g,s(x))),$$

where ξ is a representative of every class of one-dimensional representations of $\mathbb{R}^n$, Λ is that of every class of one-dimensional representations of Γ modulo those extendable to G and σ that of every equivalence class of finite-dimensional linear representations of Γ.

We note that $\xi(W(g,s(x)))$ is a factor system of G whose restriction to $\Gamma \times \Gamma$ is trivial, and if we choose $W \in Z^2(G,\mathbb{R}^n)$ such that $W|_{s(X)\times\Gamma} = 1$ then the factor system is $(G \times \Gamma)$ trivial,[2] i.e., $\xi|_{G\times\Gamma} = 1$.

VI. EXAMPLES

A. The Euclidean group $E(2)$

The locally operating multiplier realizations of this group have been studied by Hoogland[18] in order to prove the relevance of the locally operating realizations and the gauge equivalence, but the method used by him is different from ours. The natural action of $E(2)$ on $\mathbf{R}^2$ given by

$$(\mathbf{a},\phi):\begin{pmatrix}x_1\\x_2\end{pmatrix}\to\begin{pmatrix}x_1'\\x_2'\end{pmatrix}=\begin{pmatrix}x_1\cos\phi-x_2\sin\phi+a_1\\x_1\sin\phi+x_2\cos\phi+a_2\end{pmatrix}$$

is transitive, with $\mathbf{a}=\begin{pmatrix}a_1\\a_2\end{pmatrix}\in\mathbf{R}^2$ and $\phi\in[0,2\pi)$. The composition law of this group is $(\mathbf{a}',\phi')\,(\mathbf{a},\phi)=(\mathbf{a}'+\mathbf{a}^{\phi'},\phi'+\phi)$, where

$$\mathbf{a}^{\phi}=\begin{pmatrix}a_1\cos\phi-a_2\sin\phi\\a_1\sin\phi+a_2\cos\phi\end{pmatrix}.$$

The isotopy group is SO(2) and its second cohomology group is trivial. On the other side, $H^2(E(2),T)\simeq\mathbf{R}$. The classes of extensions of $E(2)$ are denoted by $[\beta]$, with $\beta\in\mathbf{R}$, and their Lie algebras are given by the following nonvanishing commutation relations[4]:

$$[J,P_1]=P_2,\quad [J,P_2]=-P_1,\quad [P_1,P_2]=-\beta I,$$

a cocycle lifting $[\beta]$ being

$$\omega_\beta((\mathbf{a}',\phi'),(\mathbf{a},\phi))=\exp\{-i\tfrac{1}{2}\beta(\mathbf{a}'\wedge\mathbf{a}^{\phi'})_3\}.$$

The second cohomology group of SO(2) being trivial, then $H^2_{\rm loc}(E(2),T)=H^2(E(2),T)$ (see Ref. 6). If we consider the homomorphic section $s:H^2(E(2),T)\to Z^2(E(2),T)$ given by $s([\beta])\;((\mathbf{a}',\phi'),\;(\mathbf{a},\phi))=\exp\{-i\tfrac{1}{2}\beta(\mathbf{a}'\wedge\mathbf{a}\phi')_3\}$, the identity isomorphism on $H^2(E(2),T)$ and the usual topology on $\mathbf{R}$, the maps $W_{{\rm id},s}\;((\mathbf{a}',\phi'),(\mathbf{a},\phi)):H^2(E(2),T)\to T$, defined as

$$W_{{\rm id},s}((\mathbf{a}',\phi'),(\mathbf{a},\phi))[\beta]=\exp\{-i\tfrac{1}{2}\beta(\mathbf{a}'\wedge\mathbf{a}^{\phi'})_3\},$$

are continuous and $W_{{\rm id},s}\in Z^2(E(2),\widehat{H^2(E(2),T)})$ defines a central extension of $E(2)$ by $\widehat{H^2(E(2),T)}$, which is a (local) representation group for $E(2)$. The elements of $\widehat{H^2(E(2),T)}$, denoted by α, are defined by $\alpha([\beta])=e^{i\alpha\beta}$. This group $\widehat{H^2(E(2,T))}$ is a Lie group and the local representation group is a Lie group, denoted $\bar{E}(2)$, with the following composition law:

$$(\alpha',\mathbf{a}',\phi')(\alpha,\mathbf{a},\phi)$$
$$=(\alpha'+\alpha-\tfrac{1}{2}(\mathbf{a}'\wedge\mathbf{a}^{\phi'})_3,\mathbf{a}'+\mathbf{a}^{\phi'},\phi'+\phi).$$

(For more details, see Ref. 6.)

The canonical epimorphism $p:\bar{E}(2)\to E(2)$ is defined by $p(\alpha,\mathbf{a},\phi)=(0,\mathbf{a},\phi)$ and $p^{-1}({\rm SO}(2))=\overline{\rm SO}(2)$; this group is isomorphic to the direct product group $\mathbf{R}\otimes{\rm SO}(2)$. If we consider the action of $\bar{E}(2)$ on $\mathbf{R}^2$ via the epimorphism p, the isotopy subgroup of $\mathbf{R}^2$ is $\overline{\rm SO}(2)$.

1. The locally operating realization of $E(2)$

The one-dimensional unitary representations of $\overline{\rm SO}(2)$, to be denoted $\sigma_{\beta,n}$ with $\beta\in\mathbf{R}$ and $n\in\mathbf{Z}$, are given by

$$\sigma_{\beta,n}(\alpha,0,\phi)=e^{i\beta\alpha}\,e^{in\phi}.$$

The one-dimensional representations of $\bar{E}(2)$ are those of $\bar{E}(2)/\bar{E}'(2)\simeq{\rm SO}(2)$, where $\bar{E}'(2)$ is the derived subgroup of $\bar{E}(2)$. They are characterized by an integer number n and will be denoted by Λ_n:

$$\Lambda_n(\alpha,\mathbf{a},\phi)=e^{in\phi},\quad n\in\mathbf{Z}.$$

Therefore, the one-dimensional representations of $\overline{\rm SO}(2)$ that can be extended to $\bar{E}(2)$ are $\sigma_{0,n}$, i.e., $\sigma_{\beta,n}$ with $\beta=0$. Then, the classes of one-dimensional representations of $\overline{\rm SO}(2)$ modulo those admitting an extension to $\bar{E}(2)$, are only characterized by $\beta\in\mathbf{R}$. We can take as a representative of each class the representation $\sigma_{\beta,0}$.

As the irreducible representations of $\overline{\rm SO}(2)$ are all the $\sigma_{\beta,n}$ we see that all of them are pseudoequivalent and determine just one class of pseudoequivalence; we can consider as a representative the representation $\sigma_{0,0}=\mathbf{1}(\beta=n=0)$.

If we take the section $\bar{s}_0:\bar{E}(2)/\overline{\rm SO}(2)\to\bar{E}(2)$ associated to the section $s_0:E(2)/{\rm SO}(2)\to E(2)$, we get

$$\bar{s}_0(\mathbf{x})=(0,s_0(\mathbf{x}))=(0,\mathbf{x},0),$$

and when we compute the term $\bar{s}_0^{-1}(\bar{g}\mathbf{x})\bar{g}\bar{s}_0(\mathbf{x})$, with $\bar{g}=(\alpha,\mathbf{a},\phi)$, we obtain that it is equal to $(\alpha-\tfrac{1}{2}(\mathbf{a}\wedge\mathbf{x}^{\phi})_3,\mathbf{0},\phi)$. The gauge pseudoequivalence classes of locally operating linear representations of $\bar{E}(2)$ induced by $\sigma_{0,0}$ are characterized by a real number β, and a representative is given by

$$(R_\beta(\bar{g})f)(\bar{g}\mathbf{x})=\exp\{i\beta\,[\alpha-\tfrac{1}{2}(\mathbf{a}\wedge\mathbf{x}^{\phi})_3]\,\}f(\mathbf{x}),$$

with $\bar{g}=(\alpha,\mathbf{a},\phi)$.

The corresponding expression for a representative of each gauge pseudoequivalence class of (multiplier) realizations of $E(2)$ is

$$[U_\beta(g)f](g\mathbf{x})=\exp\{-i\tfrac{1}{2}\beta(\mathbf{a}\wedge\mathbf{x}^{\phi})_3\}f(\mathbf{x}),$$

where $g=p(\bar{g})=(\mathbf{a},\phi)$.

The infinitesimal generators of R_β and U_β will be given by $\hat{P}_1=-i\,\partial_x+\tfrac{1}{2}\beta y$; $\hat{P}_2=-i\,\partial_y-\tfrac{1}{2}\beta x$; $\hat{J}=-i(x\,\partial_y-y\,\partial_x)$; $\hat{I}=-\beta$.

2. The irreducible realizations of $E(2)$

As a first step we are going to obtain the irreducible linear representations of $\bar{E}(2)$. This group is a semidirect product $\bar{E}(2)=\bar{\mathscr{T}}_2\odot{\rm SO}(2)$, with $\bar{\mathscr{T}}_2=(\mathbf{R}\odot\mathscr{T}^1)\odot\mathscr{T}^2$, where $\mathscr{T}^1$ and $\mathscr{T}^2$ denote the translation one-parameter subgroups of translations along spatial directions. The representations of $\bar{\mathscr{T}}_2$ are obtained by making use of Mackey's theory[15] of the induced representations for semidirect product groups. Nevertheless, as the $\bar{\mathscr{T}}_2$ is a nilpotent group we will follow Kirillov's theory[16] for these groups. We recall that the group $\bar{\mathscr{T}}_2$ is isomorphic to the oscillator group, which has been studied by Streater.[19]

First of all, we analyze the orbits under $\bar{\mathscr{T}}_2$ in the coadjoint representation. These are of two types:

A: The orbit is a point $(0,p_1',p_2')$;

B: The orbit of the point $(\beta,0,0)$ with $\beta\neq0$ is the full set $\{(\beta,p_1',p_2')\}$.

The isotopy group of the orbits of type A is $\bar{\mathscr{T}}_2$ and that of type B is the subgroup generated by I. We will study the orbits of type B because we are interested in the realizations with nontrivial factor systems. If we choose the point $(\beta,0,0)$, the maximal subalgebra of $\bar{\mathscr{T}}_2$ whose derived subalgebra is annihilated by β will be two-dimensional and it can be generated either by $\{P_1,I\}$ or $\{P_2,I\}$. We choose the subgroup K generated by $\{P_2,I\}$, but the choice of $\{P_1,I\}$ would give equivalent representations. The character of the subgroup K is

$$\chi_\beta(e^{I\alpha}e^{P_2y})=e^{i\alpha\beta},\quad\text{with }\beta\in\mathbf{R}^*=\mathbf{R}-\{0\}.$$

The more interesting representations of $\bar{\mathcal{T}}_2$ are those induced by χ_β with support space the functions $f(x)$ in $\mathcal{L}^2(\mathbb{R})$. The left cosets are $\{(\alpha',a_1',a_2')\times(\alpha,0,a_2)=(\alpha'+\alpha,a_1',a_2'+a_2),\ \forall\alpha,a_2\}$ and we choose as a representative the element $(0,x,0)\equiv x$ and the section $\bar{s}:\bar{\mathcal{T}}_{2/K}\to\bar{\mathcal{T}}_2$ defined by $\bar{s}(x)=(0,x,0)$; then, every element $\bar{g}=(\alpha,a_1,a_2)$ of $\bar{\mathcal{T}}_2$ is decomposed, in a unique way, as $\bar{g}=(\alpha,a_1,a_2)=(0,a_1,0)(\alpha,0,a_2)=\bar{s}(a_1)\bar{\gamma}(\bar{g})$. A straightforward calculation leads to

$$\bar{s}^{-1}(\bar{g}x)\bar{g}\bar{s}(x)=(\alpha+a_2x,0,a_2).$$

Then, the irreducible representations of $\bar{\mathcal{T}}_2$ characterized by $\beta\in\mathbb{R}^*$ are given by

$$(D_\beta(\alpha,a_1,a_2)f)((\alpha,a_1,a_2)x)=\chi_\beta(\alpha+a_2x,0,a_2)f(x)=e^{i\beta(\alpha+a_2x)}f(x).$$

The Hermitian generators of the representation are then $\hat{P}_1=-id/dx,\hat{P}_2=-\beta x,\hat{I}=-\beta$.

The action of the subgroup SO(2) on the space of the representation classes of $\bar{\mathcal{T}}_2$ is given by

$$\phi:D_\beta(\alpha,a_1,a_2)\mapsto D_\beta^\phi(\alpha,a_1,a_2)=D_\beta(\phi^{-1}(\alpha,a_1,a_2)\phi),$$

where D_β is a representative of each class, and

$$\begin{aligned}\phi^{-1}(\alpha,a_1,a_2)\phi&=(0,0,0,-\phi)(\alpha,a_1,a_2,0)(0,0,0,\phi)\\&=(\alpha'-(\tfrac{1}{2}a_1^2-\tfrac{1}{2}a_2^2)\cos\phi\sin\phi\\&\quad-a_1a_2\sin^2\phi,a_1\cos\phi+a_2\sin\phi,\\&\quad-a_1\sin\phi+a_2\cos\phi).\end{aligned}$$

There is an operator $W(\phi)$ that realizes the equivalence between D_β and D_β^ϕ, i.e., $D_\beta^{-\phi}(\alpha,a_1,a_2)=W(\phi)\times D_\beta(\alpha,a_1,a_2)W^{-1}(\phi)$, and both D_β and D_β^ϕ are in the same orbit. The little group of each orbit is SO(2), the isotopy group G_β of the class of the representation D_β of $\bar{\mathcal{T}}_2$ is $\bar{E}(2)$, and the induction from G_β to $\bar{E}(2)$ is unnecessary. The irreducible linear representations of $\bar{E}(2)$ are

$$\mathcal{R}_{\beta,n}(\alpha,a_1,a_2,\phi)=e^{in\phi}D_\beta(\alpha,a_1,a_2)W(\phi).$$

The operator $W(\phi)$, unique up to a factor, is

$$W(\phi)=e^{(i/2\beta)(\hat{P}_1^2+\hat{P}_2^2+C)\phi},$$

with C an arbitrary constant. Nevertheless as the representation D_β is irreducible, $W(\phi')W(\phi)$ and $W(\phi'+\phi)$ must differ in a complex number, and $W(\phi)$ defines a multiplier realization of SO(2). All these realizations are equivalent to linear ones and we can make a choice $W'(\phi)$ of $W(\phi)$, such that W' is a linear representation of SO(2); this condition limits the possible values of the parameter C. The expression $\hat{P}_1^2+\hat{P}_2^2=-d^2/dx^2+\beta^2x^2$ is proportional to the Hamiltonian of a harmonic oscillator with angular frequency $\omega_0=\beta\hbar/m$, $\hat{P}_1^2+\hat{P}_2^2=(2m/\hbar^2)H_0$ and therefore $\exp\{(i/2\beta)\times(\hat{P}_1^2+\hat{P}_2^2)2\pi\}=-1$ and consequently $C/\beta=2m+1$, with $m\in\mathbb{Z}$. The irreducible linear representations of $\bar{E}(2)$ we obtain are

$$\mathcal{R}_{\beta,n,m}(\alpha,a_1,a_2,\phi)=e^{in\phi}D_\beta(\alpha,a_1,a_2)e^{(i/2\beta)(\hat{P}_1^2+\hat{P}_2^2)\phi}e^{(i/2)(2m+1)\phi}.$$

As this depends only on $n+m$ we can write $\mathcal{R}_{\beta,n}$ (with a slight change in notation) for the representation $\mathcal{R}_{\beta,n,m}$ as

$$\mathcal{R}_{\beta,n}(\alpha,a_1,a_2,\phi)=D_\beta(\alpha,a_1,a_2)e^{(i/2\beta)(\hat{P}_1^2+\hat{P}_2^2)\phi}e^{(i/2)(2n+1)\phi}.$$

The operator $\hat{J}$ that generates rotations is

$$\hat{J}=-(1/2\beta)(\hat{P}_1^2+\hat{P}_2^2)-\tfrac{1}{2}(2n+1).$$

It is to be remarked that $\mathcal{C}=P_1^2+P_2^2-2IJ$ is a Casimir operator, whose values in an irreducible representation $\mathcal{R}_{\beta,n}$ are $\mathcal{C}=(2n+1)\beta$. All the irreducible representations of $\bar{E}(2)$ are obtained by this method[20] and will be characterized by the value of the Casimir operators I and $\mathcal{C}$.

The linear representations $\mathcal{R}_{\beta,n}$ and $\mathcal{R}_{\beta,n'}$ are equivalent, and the pseudoequivalence classes are then parametrized by $\beta\in\mathbb{R}^*$. The multiplier canonical realization $\mathcal{U}_{\beta,n}$ of $\bar{E}(2)$ associated to $\mathcal{R}_{\beta,n}$ is given by

$$\begin{aligned}\mathcal{U}_{\beta,n}(a_1,a_2,\phi)&=\mathcal{R}_{\beta,n}(0,a_1,a_2\phi)\\&=D_\beta(0,a_1,a_2)e^{(i/2\beta)(\hat{P}_1^2+\hat{P}_2^2)\phi}\\&\quad\times e^{(i/2)(2n+1)\phi}\end{aligned}$$

and it is pseudoequivalent to $\mathcal{U}_{\beta,n'}$ for any $n'\in\mathbb{Z}$.

3. Relation between the locally operating and canonical realizations

We note that the parametrization of $E(2)$ in the case of locally operating realizations is different from that of canonical realizations. But this is not a problem because this discrepancy is due to a different choice for the section $\rho:E(2)\to\bar{E}(2)$ and a change of this section implies the pseudoequivalence of the corresponding realizations.

The locally operating realizations of $E(2)$ obtained are not irreducible and we must select irreducible support subspaces in the realizations U_β. We will see what irreducible canonical realizations $\mathcal{U}_{\beta',n}$ are contained in a locally operating realization U_β. As the irreducible canonical representations are labeled by the values of the Casimir operators $\mathcal{C}$ and I, in the locally operating realization U_β only can appear the canonical realizations $\mathcal{U}_{\beta,n}$ and every $\mathcal{U}_{\beta,n}$ will appear in the subspace of $L^2(\mathbb{R})$ defined by the solutions of the differential equation

$$(\hat{\mathcal{C}}f)(x,y)=\beta(2n+1)f(x,y),$$

where $\hat{\mathcal{C}}$ is obtained when P_1,P_2,I, and J are substituted in the expression of $\mathcal{C}$ by $\hat{P}_1,\hat{P}_2,\hat{I}$, and $\hat{J}$, respectively. After a straightforward calculation we find the equation

$$\{(-i\partial_x-\tfrac{1}{2}\beta y)^2+(-i\partial_y+\tfrac{1}{2}\beta x)^2\}f(x,y)=\beta(2n+1)f(x,y). \tag{4}$$

The functions $F(t)$ supporting irreducible canonical realizations $\mathcal{U}_{\beta,n}$ which can be written as a "linear combination" of functions $f(x,y)$ belonging to a subspace irreducible under the locally operating realization U_β, are defined by $F(t)=\int\int dx\,dy\,K(x,y;t)f(x,y)$.

Hoogland[18] has given the integration kernel for the "linear combination" in the opposite direction. He also proved that the irreducible locally operating subrepresentations $U_{\beta,n}$ for different n's were gauge inequivalent because there is not any local operator realizing the equivalence of the subspaces of the functions $f(x,y)$ solutions of Eq. (4) for different values of n.

Remarks: The physical interest of the two-dimensional Euclidean group comes because this group is a subgroup of the symmetry group of a charged particle in a uniform electromagnetic field both in the relativistic and the nonrelativistic case. If we do not consider the discrete symmetries (inversions), the symmetry group of the electromagnetic field **B** is a direct product of the two-dimensional Euclidean group in a plane orthogonal to **B** and the (1 + 1) Galilei (or Poincaré) group in a line parallel to **B** (see Ref. 21).

If we only consider the two-dimensional Euclidean group and redefining $\beta = q|\mathbf{B}|$, $\mathscr{E} = 2mE$, with q the charge, and m the mass of the particle, we obtain the relation

$$E = \mathbf{P}^2/2m + (q/m)|\mathbf{B}|J,$$

with $\mathbf{P}^2 = P_1^2 + P_2^2$. Since $\mathscr{E} = (2n+1)\beta$, we find the quantization condition

$$E = (n + \tfrac{1}{2})(q/m)|\mathbf{B}|.$$

The particle energy can only take some discrete values. This spectrum coincides with the Landau's levels (see, e.g., Ref. 22, p. 496). These Landau levels belong to different physical states; their corresponding representations are pseudoequivalent in the usual way, but not gauge pseudoequivalent. We remark that the change n by $-n$ is related with a charge conjugation, q going to $-q$. For more details see Hoogland's papers.[18,21] Hoogland gets the same results as we do. However, in our method the discrete values of $\mathscr{E}$ (or E) appear in a natural way when studying the canonical realizations of $E(2)$ by means of the linear representations of the local splitting group $\bar{E}(2)$, while Hoogland obtains the discrete spectrum as a product of the search for the transformation connecting the irreducible canonical realizations with the irreducible locally operating ones.

B. The Galilei group

As a further example we will study the Galilei group G. It is the group of transformations of the four-dimensional Newtonian space-time

$$(b,\mathbf{a},\mathbf{v},R):\begin{pmatrix}\mathbf{x}\\ t\end{pmatrix}\to\begin{pmatrix}R\mathbf{x}+t\mathbf{v}+\mathbf{a}\\ t+b\end{pmatrix},$$

with $g = (b,\mathbf{a},\mathbf{v},R)$ denoting a generic element of G. The composition law of this group is

$$g'g = (b',\mathbf{a}',\mathbf{v}',R')(b,\mathbf{a},\mathbf{v},R)$$
$$= (b'+b,\mathbf{a}'+R'\mathbf{a}+\mathbf{v}'b,\mathbf{v}'+R'\mathbf{v},R'R).$$

The action of G on the space-time manifold ($\mathbb{R}^4$) is transitive; if we choose the point $x = (0,0)$ of $\mathbf{R}^4$ the isotopy group of this point is $\Gamma = \mathscr{V} \odot \mathrm{SO}(3)$, where $\mathscr{V}$ is the subgroup generated by the boosts and SO(3) is the group of rotations of the three-dimensional space. The subgroup Γ is called the homogeneous Galilei group.[23] The homogeneous space G/Γ is identified with the space-time manifold. We choose the normalized Borel section $s_0: G/\Gamma \to G$, defined by $s_0(t,\mathbf{x}) = (t,\mathbf{x},0,1)$. Every element of G is factorized in a unique way as $g = (b,\mathbf{a},\mathbf{v},R) = (b,\mathbf{a},\mathbf{0},1)\,(0,0,\mathbf{v},R) = s_0(gx_0) \times \gamma(g)$.

The second cohomology group of G is $H^2(G,T) = \mathbb{R} \otimes \mathbb{Z}_2$. The cohomology classes are labeled by $[M,l]$, with $M \in \mathbb{R}$ and $l \in \mathbb{Z}_2$. A lifting of the class $[M,l]$ is given by[24–26]

$$\omega_{M,l}((b',\mathbf{a}',\mathbf{v}',R'),(b,\mathbf{a},\mathbf{v},R))$$
$$= \omega_l^{(\mathrm{SO}(3))}(R',R)\exp\{iM(\tfrac{1}{2}b\mathbf{v}'^2 + \mathbf{v}'\cdot R'\mathbf{a})\},$$

where $\omega_l^{(\mathrm{SO}(3))}$ is a factor system of SO(3) lifting the class $[l]$ of $H^2(\mathrm{SO}(3),T)$, defined by

$$\omega_1^{(\mathrm{SO}(3))}(R',R) = 1,$$
$$\omega_{-1}^{(\mathrm{SO}(3))}(R',R) = \sigma(R')\sigma(R)\sigma^{-1}(R'R),$$

with σ any normalized Borel section, $\sigma:\mathrm{SO}(3)\to\mathrm{SU}(2)$.

The second cohomology group of Γ is $H^2(\Gamma,T) = \mathbf{Z}_2$. Furthermore $H^2_{\mathrm{loc}}(G,T) = H^2(G,T)$ and the local representation group for G will coincide with the representation group.

Next, we are going to build up the representation group for G (see Ref. 6). The mapping $s: H^2(G,T) \to Z^2(G,T)$, defined by

$$s([M,l])((b',\mathbf{a}',\mathbf{v}',R'),(b,\mathbf{a},\mathbf{v},R))$$
$$= \omega_l^{(\mathrm{SO}(3))}(R',R)\exp\{iM(\tfrac{1}{2}b\mathbf{v}'^2 + \mathbf{v}'\cdot R'\mathbf{a})\},$$

is a homomorphism. When considering the identity as the automorphism of $H^2(G,T)$ the mapping $W_{\mathrm{id},s} \in Z^2(G,\widehat{H^2(G,T)})$, given by

$$W_{\mathrm{id},s}((b',\mathbf{a}',\mathbf{v}',R'),(b,\mathbf{a},\mathbf{v},R))[M,l]$$
$$= \omega_l^{(\mathrm{SO}(3))}(R',R)\exp\{iM(\tfrac{1}{2}b\mathbf{v}'^2 + \mathbf{v}'\cdot R'\mathbf{a})\},$$

defines a central extension of G by $\widehat{H^2(G,T)}$, such that its middle group $\bar{G}$ is a (local) representation group for G. The group $\bar{G}$ is also an 11-dimensional Lie group, whose composition law is

$$\bar{g}'\bar{g} = (\Theta',\alpha',b',\mathbf{a}',\mathbf{v}',R')(\Theta,\alpha,b,\mathbf{a},\mathbf{v},R)$$
$$= (\Theta'+\Theta+\tfrac{1}{2}b\mathbf{v}'^2+\mathbf{v}'\cdot R'\mathbf{a},\alpha'\alpha W^{\mathrm{SO}(3)}(R',R),$$
$$b'+b,\mathbf{a}'+R'\mathbf{a}+\mathbf{v}'b,\mathbf{v}'+R'\mathbf{v},R'R),$$

where $W^{\mathrm{SO}(3)}(R',R)$ is the nontrivial lifting of $H^2(\mathrm{SO}(3),H^2(\mathrm{SO}(3),T))$ given by

$$W^{\mathrm{SO}(3)}(R',R) = \begin{cases} 1, & \text{if } \omega_{-1}^{(\mathrm{SO}(3))}(R',R) = 1, \\ -1, & \text{if } \omega_{-1}^{(\mathrm{SO}(3))}(R',R) = -1. \end{cases}$$

The subgroup $\{(0,\alpha,0,\mathbf{0},\mathbf{0},\mathrm{R})\}$ is topologically isomorphic to SU(2), the universal covering of SO(3); denoting by R^* the elements of SU(2) and $R^*\mathbf{a}, R^*\mathbf{v},\ldots$ the transformed of $\mathbf{a},\mathbf{v},\ldots$ by R^* via the epimorphism SU(2)$\to$SO(3), the new composition law for $\bar{G}$ is

$$\bar{g}'\bar{g} \equiv (\Theta',b',\mathbf{a}',\mathbf{v}',R'^*)(\Theta,b,\mathbf{a},\mathbf{v},R^*)$$
$$= (\Theta'+\Theta+\tfrac{1}{2}b\mathbf{v}'^2+\mathbf{v}'\cdot R'^*\mathbf{a},b'+b,\mathbf{a}'+R'^*\mathbf{a}+\mathbf{v}'t,$$
$$\mathbf{v}'+R'^*\mathbf{v}',R'^*).$$

The canonical epimorphism $p:\bar{G}\to G$ is defined by $p((\Theta,b,\mathbf{a},\mathbf{v},R^*)) = (b,\mathbf{a},v,R)$ and therefore $p^{-1}(\Gamma) = \bar{\Gamma} = \mathbf{R} \otimes (\mathscr{V} \odot \mathrm{SU}(2))$. The section of $\bar{G}/\bar{\Gamma}$ on $\bar{G}$ associated to $s_0: G/\Gamma \to G$ is $\bar{s}_0(x) = (0,s_0^*(x))$, where s_0^* is a reminder that we are making reference to SU(2) instead of SO(3). By means of straightforward calculation we obtain

$$\bar{s}_0^{-1}(\bar{g}x)\bar{g}\bar{s}_0(x) = (\Theta + \tfrac{1}{2}\mathbf{v}^2t + \mathbf{v}R^*\mathbf{x},0,0,\mathbf{v},R^*).$$

Now, we will discuss locally operating realizations of G. The one-dimensional representations of $\bar{\Gamma}$ are $\lambda_\mu(\Theta,0,0,\mathbf{v},R^*) = e^{i\Theta\mu}$, with $\mu \in \mathbb{R}$, and those of $\bar{G}$ are $\Lambda_E(\Theta,b,\mathbf{a},\mathbf{v},R^*) = e^{ibE}$, with $E \in \mathbb{R}$. Consequently the representations of $\bar{\Gamma}$ that can be extended to $\bar{G}$ are those labeled by

$\mu = 0$, that is, $\lambda_0(\Theta,0,0,\mathbf{v},R^*) = 1$. The equivalence classes of one-dimensional representations of $\bar{\Gamma}$ modulo those that can be extended to $\bar{G}$ are labeled by $\mu \in \mathbb{R}$. A representative of every class will be λ_μ, with μ running through the set of the real numbers.

On the other hand, the classes of pseudoequivalence of (finite-dimensional) linear representations of $\bar{\Gamma}$ are only characterized by the labels of the classes of equivalence of (finite-dimensional) linear representations of Γ because the direct product structure $\bar{\Gamma} = R \otimes \Gamma$ of $\bar{\Gamma}$.

A particularly interesting representation from the physical viewpoint arises from a four-dimensional faithful representation of Γ, which we denote $\mathscr{D}_{1/2}$, given explicitly by

$$\mathscr{D}_{1/2}(\mathbf{v},R^*) = \begin{pmatrix} D_{1/2}(R^*) & 0 \\ \frac{1}{2}\boldsymbol{\sigma}\cdot\mathbf{v}D_{1/2}(R^*) & D_{1/2}(R^*) \end{pmatrix},$$

where $D_{1/2}(R^*)$ is the usual spin one-half representation of SU(2). The representations of $\bar{\Gamma}$ obtained by direct product of any one-dimensional representation of $\mathbb{R}$ and $\mathscr{D}_{1/2}$ of Γ are pseudoequivalent; and a representative in this class, to be denoted $\sigma_{0,1/2}$, is given by $\sigma_{0,1/2} = 1 \otimes \mathscr{D}_{1/2}$.

For one-dimensional representations of $\bar{\Gamma}$, as we have seen before, there is one pseudoequivalence class, and we can take as representative $\lambda_0(\Theta,0,0,\mathbf{v},R) = 1$.

The gauge pseudoequivalence classes of representations of $\bar{G}$ induced by the (classes of) representations λ_0 and $\sigma_{0,1/2}$ will be denoted by (μ) and $(\mu,\frac{1}{2})$, respectively, with $\mu \in \mathbb{R}$. Representatives of each of them are

$$(R_\mu(\bar{g})f)(\bar{g}x) = \exp[i\mu(\Theta + \tfrac{1}{2}\mathbf{v}^2t + \mathbf{v}\cdot R^*\mathbf{x})]f(x),$$

$$(R_{\mu,1/2}(\bar{g})f)(\bar{g}x) = \exp[i\mu(\Theta + \tfrac{1}{2}\mathbf{v}^2t + \mathbf{v}\cdot R^*\mathbf{x})]\mathscr{D}_{1/2}(\mathbf{v},R^*)f(x),$$

respectively, with $\bar{g} = (\Theta,b,\mathbf{a},\mathbf{v},R^*) \in \bar{G}$. Such (classes of) representations are $\mathscr{A}$-split. The gauge pseudoequivalence classes of the locally operating multiplier realizations of G obtained from the gauge pseudoequivalence classes of the locally operating linear representations of $\bar{G}$ will be denoted $[\mu]$ and $[\mu,\frac{1}{2}]$, respectively. Representatives of them take the following explicit expressions:

$$(U_\mu(g)f)(gx) = \exp\{i\mu(\tfrac{1}{2}\mathbf{v}^2t + \mathbf{v}\cdot R\mathbf{x})\}f(x), \tag{5}$$

$$(U_{\mu,1/2}(g)f)(gx) = \exp\{i\mu(\tfrac{1}{2}\mathbf{v}^2t + \mathbf{v}\cdot R\mathbf{x})\}\mathscr{D}_{1/2}(\mathbf{v},R)f(x). \tag{6}$$

The canonical realizations of the Galilei group have been studied earlier.[23,26,27] The irreducible canonical realizations of G with a physical interest are characterized by $[m,U,s]$, with m, $U \in \mathbb{R}$ and s an integer or a half-odd number.

The explicit form of these realizations is

$$(\mathscr{U}_{m,U,s}(g)F)(\mathbf{p}) = \exp\{i[(\mathbf{p}^2/2m + U)b - \mathbf{p}\cdot\mathbf{a}]\} \times D_s(R)F(R^{-1}(\mathbf{p} - m\mathbf{v})), \tag{7}$$

with $D_s(R)$ an irreducible realization of SO(3). The parameters $m \neq 0$ and s can be interpreted as the mass and the spin, respectively, of the elementary particle described by this realization, and U can be associated with the internal energy.[23] The case $m = 0$ has no meaningful physical interpretation.[28,29]

The locally operating realization can be related with the canonical ones by means of a Fourier transform. We will denote $\hat{f}(k)$, where $k = (E,\mathbf{p})$, the Fourier transform of $f(x)$,

$$\hat{f}(k) = f(E,\mathbf{p}) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}^4} d^3\mathbf{x}\, dt\, e^{i(Et - \mathbf{p}\cdot\mathbf{x})} f(t,\mathbf{x}).$$

These functions $\hat{f}(k)$ are transformed under the action of G as

$$(\mathscr{U}(g)\hat{f})(k) = e^{i(Eb - \mathbf{p}\cdot\mathbf{a})}\mathscr{D}(\mathbf{v},R)\hat{f}(k'), \tag{8}$$

where $k' = (E',\mathbf{p}')$, with $E' = E + \frac{1}{2}\mu\mathbf{v}^2 - \mathbf{p}\cdot\mathbf{v}$ and $\mathbf{p}' = R^{-1}(\mathbf{p} - \mu\mathbf{v})$. It is easy to see that if $\mu \neq 0$ then $2\mu E - \mathbf{p}^2 = 2\mu E' - \mathbf{p}'^2 = \rho$ is a constant. As $\mathscr{D}(\mathbf{v};R) = \mathscr{D}(\mathbf{v};1)\mathscr{D})(0;R)$, we can define the function $\bar{f}(k)$ by $\bar{f}(E,\mathbf{p}) = \mathscr{D}(-\mathbf{p}/\mu,1)\hat{f}(E,\mathbf{p})$ for every $\hat{f}(k)$ (see Ref. 26), and the action of G over these new functions becomes

$$(\mathscr{U}(g)\bar{f})(E,\mathbf{p}) = e^{i(Eb - \mathbf{p}\cdot\mathbf{a})}\mathscr{D}(0,R)\bar{f}(E',\mathbf{p}')$$

and $\mathscr{D}_{1/2}|_{SO(3)} \simeq D_{1/2} \oplus D_{1/2}$.

A comparison between (7) and (8) allows identification of μ with m, the mass of the system, and $\rho = 2mE - \mathbf{p}^2$ with the internal energy. Since realizations with different U's are pseudoequivalent we can take $U = 0$.

The relation $2mE - \mathbf{p}^2 = \rho$ gives us a necessary condition for irreducibility of the locally operating realization, namely the function $\hat{f}(E,\mathbf{p})$ vanishes when $2mE - \mathbf{p}^2$ is not a constant value U. When taking $U = 0$ we obtain

$$(2mE - \mathbf{p}^2)\hat{f}(E,\mathbf{p}) = 0$$

and for the locally operating realizations this condition becomes a Schrödinger equation

$$i\,\partial_t f(t,\mathbf{x}) = -(\nabla^2/2m)f(t,\mathbf{x}). \tag{9}$$

When we consider the realization given by (5) the standard Schrödinger equation appears and the realization is unitary with the usual inner product. Moreover, in the case corresponding to $\mathscr{D}_{1/2}(\mathbf{v},R)$, expression (6), it leads to the Lévy–Leblond equation for particles with mass m and spin one-half (see Refs. 26 and 30). These functions $f(t,\mathbf{x})$ verifying Eq. (9) support an irreducible locally operating realization, and the equation may be rewritten as

$$i\,\partial_t u(t,\mathbf{x}) + i(\boldsymbol{\sigma}\cdot\nabla)w(t,\mathbf{x}) = 0,$$

$$i(\boldsymbol{\sigma}\cdot\nabla)u(t,\mathbf{x}) + 2mw(t,\mathbf{x}) = 0,$$

with

$$f(t,\mathbf{x}) = \begin{pmatrix} u(t,\mathbf{x}) \\ w(t,\mathbf{x}) \end{pmatrix},$$

where u and w are spinors. This representation is unitary with respect to the inner product

$$\langle f',f\rangle = \int_{\mathbb{R}^3} d^3\mathbf{x}\, u'^+(t,\mathbf{x})u(t,\mathbf{x}).$$

More details about the Lévy–Leblond equation and the corresponding locally operating realization can be found in Refs. 26 and 30.

Notice that in this case of the Galilei group the fact that the gauge pseudoequivalence coincides with the standard pseudoequivalence is related to the semidirect product structure of the Galilei group as $G = \mathbb{R}^4 \odot (\mathscr{V} \odot \mathrm{SO}(3))$.

Another interesting point to be mentioned is that even if the linear representations of Γ are not unitary, the induced realizations of G are unitary.

C. The Newton–Hooke group

Besides the Galilei and Poincaré groups, there are other interesting kinematical groups as the Newton–Hooke groups. We consider, by simplicity, the one-dimensional $(1+1)$ case associated to an oscillating universe.

The oscillating Newton–Hooke $(1+1)$ group is a three-dimensional connected Lie group, whose Lie algebra is characterized by the nonvanishing commutators

$$[P,H] = -(1/\tau^2)K, \quad [K,H] = P.$$

The composition law is

$$\begin{aligned} h'h &= (b',a',v',)(b,a,v) \\ &= (b'+b, a'\cos(b/\tau) + v'\tau\sin(b/\tau) + a, \\ &\quad v'\cos(b/\tau) - (a/\tau)\sin(b/\tau) + v). \end{aligned}$$

Note that the set $\{(b,a,0)\}$ of the space-time translation does not close a subgroup.

The action of this group on the space-time manifold ($\simeq \mathbf{R}^2$) is transitive and is defined by

$$(b,a,v): \begin{pmatrix} t \\ x \end{pmatrix} \to \begin{pmatrix} t+b \\ x + v\tau\sin(t/\tau) + a\cos(t/\tau) \end{pmatrix}.$$

The isotopy group of the point $x_0 = (0,0)$ is $\mathscr{V} = \{(0,0,v)\}$, the subgroup of the boosts. The quotient space $H/\mathscr{V}$ is diffeomorphic to $X(\simeq \mathbf{R}^2)$. We choose the normalized Borel cross section $s_0: H/\mathscr{V} \to H$ defined by $s_0(t,x) = (t,x,0)$, and every element $h \in H$ is decomposed in a unique way as a product

$$h = (b,a,v) = (b,a,0)(0,0,v) = s_0(hx_0)\gamma(h).$$

1. The local representation group for H

The second cohomology group of H is $H^2(H,T) = \mathbf{R}$. The cohomology classes are labeled by $[m]$, with $m \in \mathbf{R}$ ($[P,K] = -mI$). A lifting of the class $[m]$ is given by[31]

$$\omega_m(h',h) = \exp\left\{ im\left[\frac{1}{2}\left(v'^2 - \frac{1}{\tau^2}a'^2\right)\tau\cos\frac{b}{\tau}\sin\frac{b}{\tau} + a\left(v'\cos\frac{b}{\tau} - \frac{a'}{\tau}\sin\frac{b}{\tau}\right) - v'a'\sin^2\frac{b}{\tau} \right]\right\}.$$

Since the isotopy subgroup is $\mathscr{V} \simeq \mathbf{R}$, then $H^2(\mathscr{V},T) = \{1\}$, and hence $H^2_{\rm loc}(H,T) = H^2(H,T) = \mathbf{R}$. Thus, the local representation group for H coincides with the representation group. By a straightforward calculation we obtain that the local representation group $\bar{H}$ for H is a central extension of H by $\widehat{H^2(H,T)}$ with composition law

$$\begin{aligned} \bar{h}'\bar{h} &= (\alpha',b',a',v')(\alpha,b,a,v) \\ &= \Big(\alpha' + \alpha + \frac{1}{2}\left(v'^2 - \frac{1}{2}a'^2\right)\tau\cos\frac{t}{\tau}\sin\frac{t}{\tau} \\ &\quad + a\left(v'\cos\frac{b}{\tau} - \frac{a'}{\tau}\sin\frac{b}{\tau}\right) - v'a'\sin^2\frac{b}{\tau}, \\ &\quad b'+b,\ a'\cos\frac{b}{\tau} + v'\tau\sin\frac{b}{\tau} + a, \\ &\quad v'\cos\frac{b}{\tau} - \frac{a'}{\tau}\sin\frac{b}{\tau} + v\Big). \end{aligned}$$

The group $\bar{H}$ can be decomposed as a semidirect product $\bar{H} = (\bar{\mathscr{S}} \odot \mathscr{V}) \odot \mathscr{E}$ with $\bar{\mathscr{S}} = \mathscr{S} \otimes \mathbf{R}$, or $\bar{H} = (\mathscr{S} \odot \bar{\mathscr{V}}) \odot \mathscr{E}$, with $\bar{\mathscr{V}} = \mathbf{R} \otimes \mathscr{V}$, where $\mathscr{S}$ is the one-parameter subgroup of the spatial translations and $\mathscr{E}$ is the one-parameter subgroup of the time translations.

The canonical projection is $p: \to \bar{H}, p(\alpha,b,a,v) = (b,a,v)$, and then $p^{-1}(\mathscr{V}) = \{(\alpha,0,0,v)\} = \bar{\mathscr{V}}$. Note that the action of $\bar{H}$ on X is transitive and the isotopy group is, evidently, $\bar{\mathscr{V}}$, then $\bar{H}/\bar{\mathscr{V}}$ is diffeomorphic to X. A normalized Borel cross section $\bar{s}_0: \bar{H}/\bar{\mathscr{V}} \to \bar{H}$ is $\bar{s}_0(t,x) = (0,t,x,0)$ and then $\bar{h} = (\alpha,b,a,v) = (0,b,a,0)(\alpha,0,0,v) = \bar{s}_0(\bar{h}\bar{x}_0)\bar{\gamma}(\bar{h})$, $\forall \bar{h} \in \bar{H}$, where $\bar{x}_0 = (0,0)$ and $\bar{\gamma}: \to \bar{H} \to \bar{\mathscr{V}}$ is a Borel map.

By a straightforward calculation we obtain

$$\bar{s}_0^{-1}(\bar{h}\bar{x})\bar{h}\bar{s}_0(\bar{x}) = \left(\alpha + W(h,s_0(\bar{x})), 0,0, v\cos\frac{t}{\tau} - \frac{a}{\tau}\sin\frac{t}{\tau}\right), \tag{10}$$

where $s_0(\bar{x}) = (t,x,0)$ and

$$\begin{aligned} W(h,s_0(\bar{x})) &= W((b,a,v),(t,x,0)) \\ &= \frac{1}{2}\left(v^2 - \frac{1}{\tau^2}a^2\right)\tau\cos\frac{t}{\tau}\sin\frac{t}{\tau} \\ &\quad + x\left(v\cos\frac{t}{\tau} - \frac{a}{\tau}\sin\frac{t}{\tau}\right) - va\sin^2\frac{t}{\tau}. \end{aligned} \tag{11}$$

2. Locally operating realization of H

The irreducible linear representations of $\bar{\mathscr{V}} = \mathbf{R} \otimes \mathscr{V} (\mathscr{V} \simeq \mathbf{R})$ are one dimensional, and they can be expressed by

$$\sigma_{\eta,\beta}: (\alpha,0,0,v) \to e^{i(\eta\alpha - \beta v)}, \quad \text{with } \eta,\beta \in \mathbf{R}.$$

The linear representations of $\bar{\mathscr{V}}$ labeled by different pairs of real numbers (η,β) are not equivalent but they are pseudoequivalent and there is a unique class of pseudoequivalence of irreducible linear representations of $\bar{\mathscr{V}}$. As a representative of this class we take $\sigma_{0,0}$, i.e., the trivial representation $\sigma_{0,0}(\alpha,0,0,v) = 1$. The one-dimensional representations of $\bar{H}$ are $\Lambda_\mu[(\alpha,a,b,v)] = e^{i\mu b}$, with $\mu \in \mathbf{R}$. Thus, the equivalence classes of the one-dimensional representations of $\bar{\mathscr{V}}$ modulo "to be extendable to $\bar{H}$" are labeled by the pair of real numbers $[\eta,\beta]$. As a representative of a generic class $[\eta,\beta]$ we can take

$$\lambda_{\eta,\beta}[(\alpha,0,0,v)] = \exp\{i(\eta\alpha - \beta v)\}.$$

For that, in every class $[\beta,\eta]$ of gauge pseudoequivalence of locally operating linear representations of $\bar{H}$, we can take the representative

$$\begin{aligned} &(R_{\eta,\beta}(\bar{h})f)(\bar{h}\bar{x}) \\ &\quad = \lambda_{\eta,\beta}\left(W(h,s_0(\bar{x})) + \alpha, 0,0, v\cos\frac{t}{\tau} - \frac{a}{\tau}\sin\frac{t}{\tau}\right)f(\bar{x}), \end{aligned}$$

with $\lambda_{\eta,\beta}$ a representative of every class of one-dimensional representations of $\bar{\mathscr{V}}$ modulo "to be extendable to $\bar{H}$."

Finally, the explicit form of a representative of the gauge pseudoequivalence class $[\eta,\beta]$ of locally operating (multiplier) realizations of H is

$$(U_{\eta,\beta}(h)f)(h\bar{x}) = \exp\left\{ i\left[\eta W(h,s_0(\bar{x})) - \beta\left(v\cos\frac{t}{\tau} - \frac{a}{\tau}\sin\frac{t}{\tau}\right)\right]\right\} f(\bar{x}). \tag{12}$$

The Hermitian operators $\hat{P},\hat{K},\hat{H},\hat{I}$ corresponding to the realization associated to $R_{\eta,\beta}$ and $U_{\eta,\beta}$ are given by

$$\hat{P} = -i\cos\frac{t}{\tau}\,\partial_x + \left(\frac{\eta}{\tau}\,x\sin\frac{t}{\tau} - \frac{\beta}{\tau}\sin\frac{t}{\tau}\right),$$

$$\hat{K} = i\tau\sin\frac{t}{\tau}\,\partial_x + \left(\eta x\cos\frac{t}{\tau} - \beta\cos\frac{t}{\tau}\right),$$

$$\hat{H} = i\partial_t, \quad \hat{I} = \eta.$$

3. The canonical realizations of H

We can consider the factorization of $\bar{H}$ as $\bar{H} = (\bar{\mathscr{S}}\odot\mathscr{V})\odot\mathscr{E}$, and we will make use of Mackey's theory[16] for the obtention of the linear representations of $\bar{H}$.

The irreducible linear representations of $\bar{\mathscr{S}}$ are one dimensional because $\bar{\mathscr{S}}$ is Abelian, and as $\bar{\mathscr{S}}$ is a direct product $\bar{\mathscr{S}} = \mathbb{R}\otimes\mathscr{S}$, these representations are $\lambda_{\mu,p}(\alpha,0,a,0) = \exp\{i(\mu\alpha - pa)\}$, with $\mu,p\in\mathbb{R}$. The pairs $(\mu,p)\in\bar{\mathscr{S}}$ characterize the representations.

The orbits of $\bar{\mathscr{S}}$ under $\mathscr{V}$, via the action of $\mathscr{V}$ into $\bar{\mathscr{S}}$ $v{:}(\mu,p)\mapsto(\mu,p+\mu v)$, are of two types: type I, (μ,p), with a fixed $\mu\neq 0$; this orbit is homeomorphic to $\mathbb{R}$; and type II, $(0,p)$, each orbit has only one point.

We will only consider the orbits of type I because the representations with physical meaning are related to them.[32,33] The representations obtained from these orbits are labeled by $\mu\in\mathbb{R}^*$, with support space the functions $F{:}\mathbb{R}\to\mathbb{C}$, such that $\int_{\mathbb{R}} F^*(p)F(p)dp < +\infty$, where dp is the invariant measure on these orbits.

If we select the point $(\mu,0)$ in the corresponding orbit $((\mu,p))$, the isotopy group of this point is $\{(0,0,0,0)\}$, then the little group of this orbit is trivial. If we choose a cross section of $\bar{\mathscr{S}}$ into $\mathscr{V}$, $(\mu,p)\mapsto(0,0,0,p/\mu)$, we have

$$(0,0,0,p/\mu)^{-1}(\alpha,0,a,v)(0,0,0,p/\mu) = (\alpha - pa/\mu,0,a,0).$$

The representations of $\bar{\mathscr{S}}\odot\mathscr{V}$, up to equivalence, are $(D_\mu(\alpha,a,v)F)(p) = e^{i(\mu\alpha - pa)}F(p-\mu v)$. The equivalence classes of these representations are labeled by $[\mu]$ with $\mu\in R^*$.

As the subgroup $\bar{\mathscr{S}}\odot\mathscr{V}$ is not Abelian, the set $\widehat{\bar{\mathscr{S}}\odot\mathscr{V}}$ is made up by all the equivalence classes of the representations of $\bar{\mathscr{S}}\odot\mathscr{V}$, that is, $\widehat{\bar{\mathscr{S}}\odot\mathscr{V}} = \{[\mu],\mu\in\mathbb{R}^*\}$. The action of the subgroup $\mathscr{E}$ on $\widehat{\bar{\mathscr{S}}\odot\mathscr{V}}$ is

$$b{:}D_\mu(\alpha,a,v)\mapsto D_\mu(b^{-1}(\alpha,a,v)b) = D_\mu^{\,b}(\alpha,a,v),$$

with D_μ a representative of the class $[\mu]\in\widehat{\bar{\mathscr{S}}\odot\mathscr{V}}$. Then

$$\begin{aligned}
&[D_\mu^{\,b}(\alpha,a,v)F](p)\\
&= \Big[D_\mu\Big(\Big(\alpha + \frac{1}{2}\Big(v^2 - \frac{1}{2}a^2\Big)\tau\sin\frac{b}{\tau}\cos\frac{b}{\tau}\\
&\quad - av\sin^2\frac{b}{\tau}, a\cos\frac{b}{\tau} + v\tau\sin\frac{b}{\tau},\\
&\quad v\cos\frac{b}{\tau} - \frac{a}{\tau}\sin\frac{b}{\tau}\Big)\Big)F\Big](p)\\
&= \exp\Big\{i\Big[\mu\Big(\alpha + \frac{1}{2}\Big(v^2 - \frac{1}{\tau^2}a^2\Big)\tau\sin\frac{b}{\tau}\cos\frac{b}{\tau}\\
&\quad - av\sin^2\frac{b}{\tau}\Big) - p\Big(a\cos\frac{b}{\tau} + v\tau\sin\frac{b}{\tau}\Big)\Big]\Big\}\\
&\quad\times F\Big(p - \mu\Big(v\cos\frac{b}{\tau} - \frac{a}{\tau}\sin\frac{b}{\tau}\Big)\Big).
\end{aligned}$$

It is easy to check that the representations D_μ and $D_\mu^{\,b}$ are equivalent, i.e.,

$$D_\mu^{\,b^{-1}}(\alpha,a,v) = W(b)D_\mu(\alpha,a,v)W^{-1}(b), \quad \forall b\in\mathscr{T},$$

with $W(b) = e^{i(b/2\mu)(P^2 + (1/\tau^2)K^2 + C)}$, where $C\in\mathbb{R}$ gives a phase. Then D_μ and $D_\mu^{\,b}$ are equivalent and are in the same class $[\mu]$. Thus, the orbit of the class $[\mu]$ under the action of the subgroup $\mathscr{T}$ is $[\mu]$ itself. The little group of each orbit $[\mu]$ is $\mathscr{T}$.

Finally, the irreducible linear representations of $\bar{H}$ are

$$\begin{aligned}
(\mathscr{R}_{\mu,C,\epsilon}(\alpha,b,a,v)F) &= \exp[i(\mu\alpha - pa)]\exp(i\epsilon b)\\
&\quad\times\exp[i(b/2\mu)(P^2 + (1/\tau^2)K^2 + C)]\\
&\quad\times F(p-\mu v).
\end{aligned}\tag{13}$$

Because $e^{i\epsilon b}$ acts as a phase factor we can incorporate it into the factor C. Thus, $\mathscr{R}_{\mu,C,\epsilon}$ and $\mathscr{R}_{\mu,C,0}$ are pseudoequivalent and the multiplier realizations of H, $\mathscr{U}_{\mu,C,\epsilon}$ and $\mathscr{U}_{\mu,C,0}$, are also pseudoequivalent. The explicit form of $\mathscr{U}_{\mu,c}$ is

$$\begin{aligned}
(\mathscr{U}_{\mu,\mathscr{C}}(b,a,v)F)(p) &= \exp(-ipa)\\
&\quad\times\exp[i(b/2\mu)(P^2 + (1/\tau^2)K^2 - \mathscr{C})]\\
&\quad\times F(p-\mu v),
\end{aligned}$$

with $\mathscr{C} = -C$.

A realization of the Hermitian infinitesimal operators $\hat{P},\hat{K},\hat{H},\hat{I}$ associated to the realizations $\mathscr{R}_{\mu,\mathscr{C}}$ or $\mathscr{U}_{\mu,\mathscr{C}}$ is

$$\hat{P} = p, \quad \hat{K} = i\mu\, d/dp, \quad \hat{H} = \frac{1}{2\mu}\left(p^2 - \frac{\mu^2}{\tau^2}\frac{d^2}{dp^2} - \mathscr{C}\right),$$
$$\hat{I} = \mu. \tag{14}$$

Note that I and $\mathscr{C} = P^2 + (1/\tau^2)K^2 - 2IH$ are Casimir operators. Making use of the expressions (14) we can write the representation $\mathscr{U}_{\mu,\mathscr{C}}$ (see Ref. 13) as follows:

$$\begin{aligned}
(\mathscr{U}_{\mu,\mathscr{C}}(b,a,v)F)(p) &= \exp(-ipa)\\
&\quad\times\exp[(i/2\mu)(p^2 - (\mu^2/\tau^2)(d^2/dp^2)\\
&\quad - \mathscr{C})b]F(p-\mu v).
\end{aligned}\tag{15}$$

As the irreducible canonical realizations $\mathscr{U}_{\mu,\mathscr{C}}$ are labeled by the eigenvalues of the Casimir operators I and $\mathscr{C}$, then in the locally operating realization $U_{\eta,\beta}$, the realizations $U_{\mu,\mathscr{C}}$ can only appear if $\mu = \eta$, and every one of these will be in the subspace of $\mathscr{L}^2(\mathbb{R})$ defined by the solutions of the differential equation $(\hat{\mathscr{C}}f)(t,x) = \mathscr{C}f(t,x)$, with $\mathscr{C}$ a real number and $\hat{\mathscr{C}}$ obtained when in the expression of the Casimir operator $\mathscr{C}$, P,K,H, and I are changed to $\hat{P},\hat{K},\hat{H}$, and $\hat{I}$, respectively.

Explicitly, this equation is

$$\begin{aligned}
i\,\partial_t f(t,x) &= \left\{\frac{1}{2\eta}(-1\,\partial_x)^2 + \frac{1}{2}\frac{\eta}{\tau^2}x^2\right.\\
&\quad\left. - \frac{\beta}{\tau^2}x + \frac{\beta^2}{2\eta\tau^2} - \frac{\mathscr{C}}{2\eta}\right\}f(t,x).
\end{aligned}\tag{16}$$

As all realizations with different $\mathscr{C}$'s are pseudoequivalent, we can give it the value zero. From a physical viewpoint the real number $\mathscr{C}$ is related with the internal energy of the corresponding particle, and with a change of the origin of energy we can make it zero. The parameters μ and η can be also identified with the particle mass and then

$\mu = \eta = m \in \mathbf{R}^*$. Finally, we can show that the functions supporting the irreducible canonical realization $\mathscr{U}_{\mu,\mathscr{C}}$ can be written as a "linear combination" of the functions $f(t,x)$ belonging to a subspace irreducible under the locally operating realization $U_{\mu,\beta}$, that is

$$F(p) = \iint_{-\infty}^{+\infty} dt\, dx\, K(t,x;p) f(t,x).$$

Hoogland[18] has obtained the integration kernel of the transformation $F(t) \to f(t,x)$ in the opposite direction.

If in Eq. (16) we take $\beta = 0$ and $\mathscr{C} = 0$, we will obtain

$$i\,\partial_t f(t,x) = \left\{ \frac{1}{2m}(-i\,\partial_x)^2 + \frac{1}{2}\frac{m}{\tau^2}x^2 \right\} f(t,x).$$

This equation is the one-dimensional Schrödinger equation for a free particle of mass m in a Newton–Hooke universe.[2,34]

If $\beta \neq 0$, the corresponding equation describes a particle with mass m in a Newton–Hooke universe acted on by an external force field (β).

When we consider the two-dimensional reducible matricial representation of $\mathscr{V}$

$$\sigma(0,0,v) = \begin{pmatrix} 1 & 0 \\ -\frac{1}{2}v & 1 \end{pmatrix},$$

we obtain the following locally operating realization of H:

$$[U_{m,\beta}(h)f][h(t,x)]$$

$$= \exp\left\{ i\left[mW(h,s_0(x)) - \beta\left(v\cos\frac{t}{\tau} - \frac{a}{\tau}\sin\frac{t}{\tau} \right) \right] \right\}$$

$$\times \begin{pmatrix} 1 & 0 \\ -\frac{1}{2}(v\cos(t/\tau) - (a/\tau)\sin(t/\tau)) & 1 \end{pmatrix} f(t,x).$$

When $\beta = 0$ the two-component functions $f(t,x)$ verify a Dirac-like equation,[2,34] and if $\beta \neq 0$ the terms related with the external forced field appear.

These equations have been obtained by Dubois[34] using other's arguments also based on the Newton–Hooke group.

ACKNOWLEDGMENTS

This work was partially supported by Comisión Asesora de Investigación Cientifica y Tecnica.

[1] M. Asorey, J. F. Cariñena, and M. A. del Olmo, J. Phys. A **16**, 1603 (1983).
[2] H. Hoogland, Nuovo Cimento B **32**, 427 (1976); Ph.D. thesis, University of Nijmegen, 1977.
[3] V. Bargmann and E. P. Wigner, Proc. Nat. Acad. Sci. **34**, 211 (1948).
[4] V. Bargmann, Ann. Math. **59**, 1 (1954).
[5] E. P. Wigner, Ann. Math. **40**, 149 (1939).
[6] J. F. Cariñena, M. A. del Olmo, and M. Santander, J. Phys. A **17**, 3091 (1984).
[7] J. F. Cariñena, M. A. del Olmo, and M. Santander, *Lecture Notes in Physics, Vol. 201*, edited by G. Denardo (Springer, Berlin, 1983), p. 30.
[8] J. Schur, Z. Reine Angew. Math **127**, 20 (1904).
[9] T. Janssen, J. Math. Phys. **13**, 342 (1972).
[10] J. F. Cariñena and M. Santander, J. Math Phys **16**, 1416 (1975).
[11] U. Cattaneo, J. Math. Phys. **19**, 452 (1978).
[12] J. F. Cariñena and M. Santander, J. Math. Phys. **20**, 2168 (1979).
[13] J. F. Cariñena, M. A. del Olmo, and M. Santander, Physica A **114**, 420 (1982).
[14] F. W. Warner, *Foundations of Differentiable Manifolds and Lie Groups* (Scott, Foresman, Glenview, IL, 1971).
[15] G. W. Mackey, Ann. Math. **55**, 101 (1952). [See also D. J. Simms, *Lie Groups and Quantum Mechanics Lecture Notes in Math, Vol. 52* (Springer, Berlin, 1969).]
[16] A. A. Kirillov, Usp. Math. Nauk **10**, 57 (1962).
[17] M. Asorey, L. J. Boya, and J. F. Cariñena, Rep. Math. Phys. **21**, 35 (1984).
[18] H. Hoogland J. Phys. A **11**, 1557 (1978).
[19] R. F. Streater, Commun. Math. Phys. **4**, 217 (1967).
[20] N. Giovannini, Hel. Phys. Acta **50**, 349 (1977).
[21] H. Hoogland, J. Phys. A **11**, 797 (1978).
[22] L. Landau and E. M. Lifshitz, *Mécanique Quantique* (Mir, Moscow, 1967).
[23] J. M. Levy-Leblond, *Galilei Group and Galilean Invariance in Group Theory and Its Applications*, edited by E. M. Loebl (Academic, New York, 1971), Vol. 2.
[24] H. Brennich, Ann. Inst. H. Poincaré A **13**, 137 (1970).
[25] J. F. Cariñena and M. Santander, J. Math. Phys. **22**, 1548 (1981).
[26] J. F. Cariñena and M. Santander, J. Phys. A **15**, 343 (1982).
[27] M. Santander, Ph.D. thesis, Universidad de Valladolid, 1977 (unpublished).
[28] E. Inonu and E. P. Wigner, Nuovo Cimento **9**, 705 (1952).
[29] M. Hamermesh, Ann. Phys. (NY) **32**, 157 (1960).
[30] J. M. Levy-Leblond, Commun. Math. Phys. **6**, 286 (1967).
[31] J. R. Derome and J. G. Dubois, Nuovo Cimento B **9**, 351 (1972).
[32] J. G. Dubois, Nuovo Cimento B **15**, 1 (1973).
[33] M. Valls, Tesina de Licenciatura, Universidad de Valladolid, 1977 (unpublished).
[34] J. G. Dubois, Can. J. Phys. **51**, 1757 (1973).

Boson realization of sp(4). I. The matrix formulation

O. Castaños
Centro de Estudios Nucleares, Universidad Nacional Autonoma de México, Apartado Postal 70-543, México, D. F. 01000, Mexico

E. Chacón and M. Moshinsky[a)]
Instituto de Física, Universidad Nacional Autonoma de México, Apartado Postal 20-364, México, D.F. 01000, Mexico

C. Quesne[b)]
Physique Théorique et Mathématique CP229, Université Libre de Bruxelles, Brussels, Belgium

(Received 5 November 1984; accepted for publication 19 April 1985)

Holstein and Primakoff derived long ago the boson realization of a su(2) Lie algebra for an arbitrary irreducible representation (irrep) of the SU(2) group. The corresponding result for su(1,1)≅sp(2) is also well known. This raises the question of whether it is possible to obtain in an explicit, analytic, and closed form, and for any integer d, the boson realization of a sp($2d$) Lie algebra for an arbitrary irrep of the Sp($2d$) group, which is a problem of considerable physical interest. The case $d = 2$ already illustrates the problem in its full generality and thus in this paper we concentrate on sp(4). The Dyson realization is well known, and the passage to bosons satisfying the appropriate Hermiticity conditions can be done by a similarity transformation through an operator K. What we want, though, is an explicit boson realization for sp($2d$) similar to the one that exists for sp(2). In Sec. VI we show how we can get it for sp(4) if the operator K is known. Unfortunately while the matrix form of K^2 can be explicitly derived from definite recursion relations, the same cannot be said of K as it involves, in general, the solution of algebraic equations of high degree. Thus the conclusion, corroborated also by a classical analysis where K does not appear, is that an explicit, analytic, and closed boson realization of sp(4), and thus also of sp($2d$), is only possible for particular irreps of the corresponding groups.

I. INTRODUCTION AND SUMMARY

Already in 1940 Holstein and Primakoff[1] had obtained a realization of the su(2) Lie algebra in terms of boson creation and annihilation operators for a given value of the Casimir operator, i.e., for a definite irreducible representation (irrep) of the SU(2) group. It is easy to extend this realization to the su(1,1)≅sp(2) Lie algebra for a definite irrep of the Sp(2) group.[2,3] This simple result immediately suggests the following question: Is it possible to obtain in an explicit analytic and closed form the boson realization of a sp($2d$) Lie algebra for a given irrep of the Sp($2d$) group when d is an arbitrary integer? We will be dealing in this article with the noncompact version of the symplectic Lie group, which is usually referred to in the literature as Sp($2d,R$) or Sp(d,R).

The answer to this question is of physical interest. For $d = 2$ Mlodinow and Papanicolau[4] have shown that it is relevant for a class of generalized helium Hamiltonians while recently Hecht has applied it to the proton–neutron quasispin group. For $d = 3$ Rosensteel and Rowe[5] have indicated its importance for the microscopic analysis of collective motions in nuclei and an extensive literature exists in this field,[6–11] either working directly with Sp(6) or its complementary group[12] O(n). Other applications are in progress, though the present authors do not have yet information on published material related to them.

Where do we stand at present in connection with the boson realization of a sp($2d$) Lie algebra for a given irrep of the Sp($2d$) group? To briefly review the extensive work in this field it is convenient first to express the generators of the sp($2d$) Lie algebra in terms of the creation η_{is} and annihilation ξ_{is} operators of a system of n particles, which are associated with the index $s = 1,2,...,n$ in a d-dimensional harmonic oscillator for which the component index is $i = 1,2,...,d$. The generators of sp($2d$) are then[10,13]

$$B^{\dagger}_{ij} = \sum_{s=1}^{n} \eta_{is}\eta_{js}, \tag{1.1a}$$

$$C_{ij} = \frac{1}{2}\sum_{s=1}^{n} (\eta_{is}\,\xi_{js} + \xi_{js}\eta_{is}) = \sum_{s=1}^{n} \eta_{is}\,\xi_{js} + \frac{n}{2}\delta_{ij}, \tag{1.1b}$$

$$B_{ij} = \sum_{s=1}^{n} \xi_{is}\,\xi_{js}, \tag{1.1c}$$

where in (1.1b) we used the commutation rule $[\xi_{jt},\eta_{is}] = \delta_{ij}\delta_{st}$. As is customary[14] one can divide the set (1.1) into raising, weight, and lowering generators, given by

$$B^{\dagger}_{ij};\quad C_{ij}\quad \text{with } i<j, \tag{1.2a}$$

$$C_{ii},\quad i = 1,2,...,d, \tag{1.2b}$$

$$B_{ij};\quad C_{ij}\quad \text{with } i>j. \tag{1.2c}$$

The lowest weight (l.w.) state denoted by the ket $|\text{l.w.}\rangle$ satisfies then the equations

$$B_{ij}|\text{l.w.}\rangle = 0, \tag{1.3a}$$

$$C_{ij}|\text{l.w.}\rangle = 0,\quad \text{for } i>j, \tag{1.3b}$$

$$C_{ii}|\text{l.w.}\rangle = (\omega_i + n/2)|\text{l.w.}\rangle, \tag{1.3c}$$

where we note that in (1.3c) the ω_i are the integer eigenvalues

[a)] Member of El Colegio Nacional and Instituto de Investigaciones Nucleares.
[b)] Maitre de Recherches F. N. R. S.

 0022-2488/85/092107-17$02.50

of the number operators $\Sigma_{s=1}^{n}\eta_{is}\xi_{is}$, $i=1,2,...,d$, which, furthermore, from (1.3b) satisfy the inequality[14] $0\leqslant\omega_1\leqslant\omega_2\cdots\leqslant\omega_d$. The eigenvalues of C_{ii} characterize then the irrep of Sp($2d$) and they can be put in the order

$$[\omega_1+n/2,\omega_2+n/2,...,\omega_d+n/2]. \tag{1.4}$$

The question raised in the previous paragraph refers now to our knowledge of the boson realizations of the sp($2d$) Lie algebras associated with the irrep (1.4) of the Sp($2d$) group. When the number of particles n goes to infinity the boson realization of the Lie algebra sp(6)$\supset$u(3) was obtained by Rosensteel and Rowe,[15] while a discussion for the chain sp(6)$\supset$sp(2)$\times$o(3) was given by Castaños and Frank.[16] It is easy to generalize these results from $d=3$ to arbitrary d.

For the much more difficult case when n is finite the problem was first attacked for the irrep (1.4) in which all the ω_i are equal, i.e., $\omega_i=\omega$, $i=1,2,...,d$. For $d=2$ a boson realization is already available in a paper of Mlodinow and Papanicolau,[4] though a more explicit procedure was carried out by Moshinsky and Seligman,[17] who also outlined an approach for a general d. Deenen and Quesne[13] gave a full and explicit discussion of the boson realization of the sp($2d$) Lie algebra when the irrep is given by (1.4) with $\omega_i=\omega$, $i=1,2,...,d$, which is sometimes referred to in the literature[11] as the case of "closed shells" to which it corresponds when $d=3$.

What happens, though, for the general irrep (1.4) when not all of the ω_i are equal? This could be called[11] the case of "open shells," and is a problem that has interested a number of researchers in the last few years.

Deenen and Quesne[18] found a realization in what is known as the Dyson formulation. As previously they had established the relation between the Dyson and Holstein–Primakof formulations in the case of "closed shells,"[13] they proceeded, using the language of coherent states, to outline the derivation of the boson realization of the sp($2d$) Lie algebra for the case of "open shells."[19] A simplified version of their approach was given by Moshinsky.[20]

Independently, Rowe, Rosensteel, and their collaborators[21,22] established the connection, in the case of "open shells," between the Dyson and Holstein–Primakoff boson realizations of sp(6), where the former seemed known to them since some time ago. Their analysis can be extended to arbitrary d.

It would seem then that the question raised in the first paragraph of this article has been completely answered. This turns out not to be the case because we ask about an *explicit*, *analytic*, and *closed* form for the boson realization of the sp($2d$) Lie algebra for the general irrep (1.4) such as exists, for example, for sp(2) (see Refs. 2 and 3).

Thus in the present paper we analyze how far we can go in the pursuit of an explicit, analytic, and closed form for the boson realization when $d=2$, i.e., the sp(4) Lie algebra. This problem is also of interest for the following reasons.

(1) The sp(4) algebra is of physical interest.[4]

(2) The sp(4) case, for which $d=2$, corresponds to the lowest value of d in which we can have "open shells," i.e., $\omega_1\neq\omega_2$.

(3) The sp(4) algebra has a u(2) subalgebra and thus the whole analysis can be made using the familiar results of the su(2) Racah algebra, rather than those of su(d) for the general case.

We proceed now to summarize the contents of the paper. In Sec. II we express the generator $B^\dagger$, C, B of sp(2) [given by (1.1) where $i=j=1$ and thus can be suppressed] in terms of boson operators $b^\dagger$ and b, which are Hermitian conjugate and satisfy $[b,b^\dagger]=1$. For this expression we use the general approach developed in recent publications,[19–21] and it will illustrate in an elementary fashion the procedure that one can follow for sp($2d$) and, in particular, for sp(4). We note, though, that while the method expresses the generators $B^\dagger$, C, B of sp(2) as explicit, analytic, and closed functions of $b^\dagger$ and b, this will not be true in general, as discussed in the following sections for the case sp(4), and thus also for sp($2d$).

In Sec. III we consider the ten generators of sp(4), nine of them in a vector notation, i.e., $B_i^\dagger$, J_i, B_i, $i=1,2,3$, plus the number operator $\mathscr{N}$, and construct a set of nonorthonormal basis states characterized by the irreps of the chain Sp(4)$\supset$U(2).

In Sec. IV we give the Dyson realization of the sp(4) Lie algebra in terms of creation β_i^+ and annihilation β_i operators, plus the s_i associated with an independent su(2) Lie algebra, where $i=1,2,3$. These operators have the commutation rules associated with a direct sum of a three-dimensional Weyl Lie algebra and a unitary unimodular Lie algebra, i.e., $w(3)\oplus$su(2). It is important to note that β_i^+ is *not* the Hermitian conjugate of β_i.

In Sec. V we consider the operators $b_i^\dagger$, b_i, S_i, $i=1,2,3$, in which the $b_i^\dagger$ is the Hermitian conjugate of b_i and S_i is Hermitian. We relate these operators to the β_i^+, β_i, s_i of the previous section through a similarity transformation involving an operator K which is Hermitian and invariant under u(2). Expressing the generators of sp(4) in terms of $b_i^\dagger$, b_i, S_i, and K we obtain an operator equation that allows us to determine K^2 in an appropriate boson basis.[19–21] In Appendix A we give an explicit algorithm for determining the matrix elements of K^2. Note that now $b_i^\dagger$, b_i, S_i, $i=1,2,3$, are bona fide generators of the Lie algebra $w(3)\oplus$su(2).

In Sec. VI we show how we can get the $B_i^\dagger$, J_i, B_i, and $\mathscr{N}$ as functions of the operators $b_i^\dagger$, b_i, S_i if we know K as an Hermitian operator function of the latter that is invariant under the U(2) group. The operator K can be determined if its matrix form is known explicitly with respect to an appropriate boson basis. Unfortunately to go from the matrix of K^2 to that of K we need to solve, in general, algebraic equations of high degree. Thus we can only get the operator K explicitly for low eigenvalues of the S^2 associated with the operator S_i which we call the spin. We discuss the cases of spin $0,\frac{1}{2},1$ (where we note that 0 corresponds to "closed shells") and give the explicit, analytic, and closed expressions for the $B_i^\dagger$, J_i, B_i and $\mathscr{N}$ for $s=0$ and $\frac{1}{2}$.

In Sec. VII we discuss the corresponding classical problem and show again that it does not seem feasible to obtain an explicit, analytic, and closed expression of the generators sp(4), in terms of those of $w(3)\oplus$su(2). In here all the generators are not operators but classical observables.

Finally in Sec. VIII we conclude that while recent developments in the field provide an algorithm for the calculation of the matrix elements of the generators of sp($2d$) in the sp(2d) $\supset$ u(d) basis of states,[19–21] as discussed in Appendix B for $d=2$, they cannot give us, in general, an explicit, analytic, and closed operator or functional relation between the generators of sp(2d) and those of the Holstein–Primakoff version of $w[(d/2)(d+1)]\oplus \mathrm{su}(d)$, when we have an arbitrary irrep (1.4) of Sp(2d).

As indicated above, in the present paper we give an explicit procedure for the determination of the *matrix* form of the operator K^2, through the use of appropriate recursion relations. We later realized that it is possible to get an *explicit and closed* form of the K^2 as a kernel with respect to appropriate coherent states. Thus we added to the title of this paper the words "I. The matrix formulation," and in a sequel we shall discuss the generating kernel formulation for the boson realization of sp(4), and give an independent procedure for deriving from it the matrix elements of K^2.

II. THE BOSON REALIZATION OF sp(2)

As we mentioned in the Introduction, the boson realization of the sp(2) Lie algebra for a given irrep of the Sp(2) group is both well known and simple.[2,3] The reason that we rederive it in this section is that we shall proceed to do it by exactly the same steps that one can use for sp(2d).

As indicated in (1.1) and (1.2) the raising, weight, and lowering generators of sp(2) can be denoted by $B^\dagger$, C, B, in which $B^\dagger$ is the Hermitian conjugate of B and C is Hermitian. Besides, from (1.1) and the commutation rules $[\xi_s,\eta_t]=\delta_{st}$ we conclude that $B^\dagger$, C, B satisfy the commutation relations[3]

$$[C,B^\dagger]=2B^\dagger, \tag{2.1a}$$

$$[C,B]=-2B, \tag{2.1b}$$

$$[B,B^\dagger]=4C. \tag{2.1c}$$

The Casimir operator[2,3] can be defined as

$$G\equiv \tfrac14[B^\dagger B-C(C-2)], \tag{2.2}$$

as from (2.1) it commutes with $B^\dagger$, C, B.

From (1.3) we conclude that the lowest weight state for a basis of a given irrep of the Sp(2) group, which we will now denote by the ket$|\omega\rangle$, is then characterized by

$$B\,|\omega\rangle=0, \tag{2.3a}$$

$$C\,|\omega\rangle=(\omega+n/2)|\omega\rangle. \tag{2.3b}$$

The full non-normalized basis for the irrep is then given by the states[3]

$$|\nu,\omega\rangle=(B^\dagger)^\nu|\omega\rangle,\quad \nu=0,1,2,\ldots,\quad |0,\omega\rangle\equiv|\omega\rangle, \tag{2.4}$$

and from (2.2) and (2.3) is characterized by the eigenvalue

$$-\tfrac14(\omega+n/2)(\omega+n/2-2) \tag{2.5}$$

of the Casimir operator G.

Our purpose now is to express $B^\dagger$, C, B, for a given value of the Casimir operator G of (2.2), i.e., a definite ω, in terms of Hermitian conjugate operators of creation $b^\dagger$ and annihilation b that satisfy the commutation rule

$$[b,b^\dagger]=1. \tag{2.6}$$

To implement our objective we start by noting that $B^\dagger$, C, B are given by (1.1), in which $i,j=1$ and thus can be suppressed. As then $B^\dagger$, C, B are given in terms of η_s, ξ_s, $s=1,2,\ldots,n$, we first consider the effect of the latter operators on the states $|\nu\omega\rangle$, which are the bases for an irrep of Sp(2). As $[\eta_s,(B^\dagger)^\nu]=0$ and the commutator $[\xi_s,(B^\dagger)^\nu]=\partial(B^\dagger)^\nu/\partial\eta_s$, we immediately obtain that

$$\eta_s|\nu\omega\rangle=(B^\dagger)^\nu\eta_s|\omega\rangle, \tag{2.7a}$$

$$\xi_s|\nu\omega\rangle=\frac{\partial(B^\dagger)^\nu}{\partial\eta_s}|\omega\rangle+(B^\dagger)^\nu\xi_s|\omega\rangle. \tag{2.7b}$$

Applying these results twice and summing with respect to s we then have

$$B^\dagger|\nu\omega\rangle=B^\dagger(B^\dagger)^\nu|\omega\rangle, \tag{2.8a}$$

$$\begin{aligned}C\,|\nu\omega\rangle&=2B^\dagger\frac{\partial(B^\dagger)^\nu}{\partial B^\dagger}|\omega\rangle+(B^\dagger)^\nu C\,|\omega\rangle\\&=2B^\dagger\frac{\partial(B^\dagger)^\nu}{\partial B^\dagger}|\omega\rangle+\left(\omega+\frac n2\right)(B^\dagger)^\nu|\omega\rangle,\end{aligned} \tag{2.8b}$$

$$\begin{aligned}B\,|\nu\omega\rangle&=4B^\dagger\frac{\partial^2(B^\dagger)^\nu}{\partial B^{\dagger 2}}|\omega\rangle+4\frac{\partial(B^\dagger)^\nu}{\partial B^\dagger}C\,|\omega\rangle+(B^\dagger)^\nu B\,|\omega\rangle\\&=4B^\dagger\frac{\partial^2(B^\dagger)^\nu}{\partial B^{\dagger 2}}|\omega\rangle+4\left(\omega+\frac n2\right)\frac{\partial(B^\dagger)^\nu}{\partial B^\dagger}|\omega\rangle,\end{aligned} \tag{2.8c}$$

where for the result on the right-hand side of (2.8b) and (2.8c) we used (2.3).

From Eqs. (2.8) we see that the effect of the generators of sp(2) on the states $|\nu\omega\rangle$ of (2.4) can be expressed in terms of an operator $B^\dagger$ that acts multiplicatively on $(B^\dagger)^\nu$ and of a differential operator $\partial/\partial B^\dagger$ acting also on $(B^\dagger)^\nu$. To express $B^\dagger$, C, B in terms of these new operators it is then convenient to introduce the definitions

$$\beta^+\equiv B^\dagger, \tag{2.9a}$$

$$\beta\equiv\frac{\partial}{\partial B^\dagger}, \tag{2.9b}$$

for which obviously

$$[\beta,\beta^+]=1. \tag{2.10}$$

The operators β^+,β have, from (2.10), the commutation rules associated with a one-dimensional Weyl algebra $w(1)$ though clearly β^+ is *not* the Hermitian conjugate of β.

From (2.8) we can now express the generators of sp(2) as

$$B^\dagger=\beta^+, \tag{2.11a}$$

$$C=2\beta^+\beta+(\omega+n/2), \tag{2.11b}$$

$$B=4\beta^+\beta^2+4\beta(\omega+n/2), \tag{2.11c}$$

and from (2.10) we immediately check that the commutation rules (2.1) are satisfied and that the Casimir operator (2.2) takes the value (2.5). The expressions (2.11) are known as the Dyson[3,20] realization associated with the Barut–Girardello representation[3] of the sp(2) Lie algebra when its irrep is characterized by $(\omega+n/2)$.

The Holstein–Primakoff realization would be in terms of operators $b^\dagger$ and b satisfying

$$[b,b^\dagger]=1, \tag{2.12}$$

but in which $b^\dagger$ is the Hermitian conjugate of b, and so they are actually the generators of $w(1)$. How can we get this realization from the knowledge of (2.11)? It seems natural to assume that $b^\dagger,b$ and β^+,β are related by a similarity transformation with some operator K, i.e.,

$$b^\dagger = K^{-1}\beta^+ K, \quad b = K^{-1}\beta K, \tag{2.13}$$

as in this way (2.12) follows immediately from (2.10). Furthermore we can impose the additional condition[19,20,21] that K should be Hermitian and an invariant of U(1) subgroup of Sp(2), whose generator in the boson representation is the number operator $N = b^\dagger b$, i.e.,

$$K^\dagger = K, \tag{2.14a}$$

$$[N,K] = 0. \tag{2.14b}$$

From (2.11), (2.14b) we have then

$$B^\dagger = Kb^\dagger K^{-1}, \tag{2.15a}$$

$$C = 2N + \omega + n/2, \tag{2.15b}$$

$$B = K[4b^\dagger b^2 + 4b(\omega + n/2)]K^{-1}. \tag{2.15c}$$

As (1.1) indicates that $B^\dagger$ is the Hermitian conjugate of B we have from (2.14a) and (2.15a) that

$$B = K^{-1}bK, \tag{2.16}$$

and from this expression and (2.15c) we obtain the relation

$$bK^2 = K^2[4b^\dagger b^2 + 4b(\omega + n/2)]. \tag{2.17}$$

We now wish to obtain explicitly from (2.17) the operator form of K. For this purpose we note that a complete normalized set of states associated with the Holstein–Primakoff boson operators $b^\dagger$, b is given by

$$|\nu) = [\nu!]^{-1/2}(b^\dagger)^\nu|0), \tag{2.18}$$

which we denote by the round bracket $|\nu)$ to distinguish them from the angular one $|\nu\omega\rangle$ associated with an irrep $(\omega + n/2)$ of Sp(2). We then take matrix elements of the right- and left-hand sides of (2.17) with respect to these states to obtain

$$\begin{aligned}(\nu - 1|b\,|\nu)(\nu|K^2|\nu) &\\ = (\nu - 1|K^2|\nu - 1)(\nu - 1|4b^\dagger b^2 + 4b(\omega + n/2)|\nu),&\end{aligned} \tag{2.19}$$

as from (2.14b), K and thus K^2 are diagonal in ν.

As the matrix elements of $b^\dagger$, b in the basis $|\nu)$ are trivial we immediately obtain the recursion relation

$$(\nu|K^2|\nu) = (\nu - 1|K^2|\nu - 1)4(\nu - 1 + \omega + n/2). \tag{2.20}$$

As we can assume without loss of generality that $(0|K^2|0) = 1$, we obtain from (2.20) that

$$(\nu|K^2|\nu) = 2^\nu[(2\nu - 2 + 2\omega + n)!!][2\omega + n - 2)!!]^{-1}, \tag{2.21}$$

which implies that

$$\begin{aligned}(\nu|K\,|\nu)&\\ = 2^{\nu/2}[(2\nu - 2 + 2\omega + n)!!]^{1/2}[(2\omega + n - 2)!!]^{-1/2}.&\end{aligned} \tag{2.22}$$

Turning now our attention to Eq. (2.15a) we see that its matrix form with respect to the states $|\nu)$ is

$$\begin{aligned}(\nu + 1|B^\dagger|\nu) &= (\nu + 1|Kb^\dagger K^{-1}|\nu)\\ &= (\nu + 1|K\,|\nu + 1)(\nu + 1|b^\dagger|\nu)(\nu|K^{-1}|\nu)\\ &= 2^{1/2}(2\nu + 2\omega + n)^{1/2}(\nu + 1|b^\dagger|\nu)\\ &= (\nu + 1|2b^\dagger(N + \omega + n/2)^{1/2}|\nu),\end{aligned} \tag{2.23}$$

where we have made use of (2.22).

As the $|\nu)$, $\nu = 0,1,2,\ldots$, are a complete set of states, we conclude from (2.23) that

$$B^\dagger = 2b^\dagger(N + \omega + n/2)^{1/2}, \tag{2.24a}$$

while its Hermitian conjugate is

$$B = 2(N + \omega + n/2)^{1/2}b, \tag{2.24b}$$

and from (2.15b)

$$C = 2N + \omega + n/2, \tag{2.24c}$$

where $N = b^\dagger b$. The expressions (2.24) give an *explicit, analytic,* and *closed* realization of the generators $B^\dagger$, C, and B of the Lie algebra sp(2) in terms of the bosons $b^\dagger, b$ that are generators of a $w(1)$ Lie algebra.

It can be immediately checked that $B^\dagger$, C, and B satisfy the commutation rules (2.1) and that the Casimir operator (2.2) takes the value (2.5). For this we only need to note that for an arbitrary function $f(N)$ we have

$$f(N)b^\dagger = b^\dagger f(N + 1), \tag{2.25a}$$

$$f(N)b = bf(N - 1), \tag{2.25b}$$

which is derived when we calculate the matrix elements of the left- and right-hand side of (2.25) with respect to the states $|\nu)$ of (2.18).

As a last point we note that with the help of (2.24c) we can invert (2.24a) and (2.24b) to get

$$b^\dagger = B^\dagger(2C + 2\omega + n)^{-1/2}, \tag{2.26a}$$

$$b = (2C + 2\omega + n)^{-1/2}B, \tag{2.26b}$$

thus obtaining the boson operators in terms of the generators $B^\dagger$, C, B in an explicit, analytic, and closed form. From (1.1) we see that these boson operators will be functions of the creation and annihilation operators η_s, ξ_s, $s = 1,2,\ldots,n$, of the n particles in a space of one dimension.

We now turn our attention to the sp(4) Lie algebra to see whether it is possible there for an arbitrary irrep (1.4) for $d = 2$, to arrive at expressions equivalent to (2.24).

III. GENERATORS OF sp(4) AND THE BASIS STATES

For sp(4) the generators are given by (1.1) when $d = 2$, but rather than write them as $B^\dagger_{ij}$, C_{ij}, B_{ij}, $i,j = 1,2$, it is convenient to express nine of the ten generators in vector form through the definitions

$$B^\dagger_1 = -\tfrac{1}{2}(B^\dagger_{11} - B^\dagger_{22}), \tag{3.1a}$$

$$B^\dagger_2 = (i/2)(B^\dagger_{11} + B^\dagger_{22}), \tag{3.1b}$$

$$B^\dagger_3 = B^\dagger_{12}, \tag{3.1c}$$

$$J_1 = \tfrac{1}{2}(C_{12} + C_{21}), \tag{3.1d}$$

$$J_2 = -(i/2)(C_{12} - C_{21}), \tag{3.1e}$$

$$J_3 = \tfrac{1}{2}(C_{11} - C_{22}), \tag{3.1f}$$

$$B_1 = -\tfrac{1}{2}(B_{11} - B_{22}), \tag{3.1g}$$

$$B_2 = -(i/2)(B_{11} + B_{22}), \tag{3.1h}$$

$$B_3 = B_{12}, \tag{3.1i}$$

while the last generator is the scalar number operator defined by

$$\mathcal{N} = \tfrac{1}{2}(C_{11} + C_{22}). \tag{3.1j}$$

From the definitions (1.1) for $d=2$ and the commutation relations $[\xi_{jt},\eta_{is}]=\delta_{ij}\delta_{st}$ it follows that the operators defined in (3.1) satisfy the following commutation relations:

$$[B_i^\dagger,B_j^\dagger]=0, \tag{3.2a}$$

$$[B_i,B_j]=0, \tag{3.2b}$$

$$[B_i,B_j^\dagger]=-2i\epsilon_{ijk}J_k+2\delta_{ij}\mathscr{N}, \tag{3.2c}$$

$$[J_i,B_j^\dagger]=i\epsilon_{ijk}B_k^\dagger, \tag{3.2d}$$

$$[J_i,B_j]=i\epsilon_{ijk}B_k, \tag{3.2e}$$

$$[J_i,J_j]=i\epsilon_{ijk}J_k, \tag{3.2f}$$

$$[\mathscr{N},B_i^\dagger]=B_i^\dagger, \tag{3.2g}$$

$$[\mathscr{N},B_i]=-B_i, \tag{3.2h}$$

$$[\mathscr{N},J_i]=0, \tag{3.2i}$$

where from now on i, j, k take the values 1,2,3, ϵ_{ijk} is the antisymmetric tensor, and repeated indices are summed from 1 to 3. Note that the J_i, $i=1,2,3$, are the generators of the su(2) subalgebra of sp(4) with the standard properties of angular momentum, while, with respect to the J_i, the $B_i^\dagger$, B_i, $i=1,2,3$, behave as ordinary three-dimensional vectors.

The set of ten generators of sp(4) can be divided into three subsets of raising, weight, and lowering type, which are separated below by semicolons:

$$B_i^\dagger,J_+;\quad \mathscr{N},J_0;\quad B_i,J_-,\quad i=1,2,3, \tag{3.3}$$

where

$$J_\pm=\mp(1/\sqrt{2})(J_1\pm iJ_2),\quad J_0=J_3. \tag{3.4}$$

The lowest weight state, which we designate as $|\omega s\rangle$, can now be characterized by

$$B_i|\omega s\rangle=0,\quad i=1,2,3, \tag{3.5a}$$

$$J_-|\omega s\rangle=0, \tag{3.5b}$$

$$\mathscr{N}|\omega s\rangle=(\omega+n/2)|\omega s\rangle, \tag{3.5c}$$

$$J_0|\omega s\rangle=-s|\omega s\rangle. \tag{3.5d}$$

From the expression (3.1f, j) for $\mathscr{N}$, J_0 we see that we can also take as weight generators C_{11} and C_{22} of (1.1), which from (1.3c) have the eigenvalues

$$C_{11}|\omega s\rangle=(\omega_1+n/2)|\omega s\rangle, \tag{3.6a}$$

$$C_{22}|\omega s\rangle=(\omega_2+n/2)|\omega s\rangle, \tag{3.6b}$$

so that from (3.1f), (3.1j), (3.5c), and (3.5d), we obtain

$$\omega_1=\omega-s, \tag{3.7a}$$

$$\omega_2=\omega+s. \tag{3.7b}$$

As ω_1, ω_2 are integers, we have that both ω,s are either integers or half-integers.

A complete set of basis states for the irrep

$$[\omega_1+n/2,\omega_2+n/2] \tag{3.8}$$

of Sp(4), which corresponds to (2.4) for Sp(2), is obtained by applying polynomials in the raising generators $B_1^\dagger$, $B_2^\dagger$, $B_3^\dagger$, J_+ to $|\omega s\rangle$. Using the commutation relations (3.2) we can put powers of the J_+ on the right of polynomials $P(B_i^\dagger)$ and use these powers to define lowest weight states of arbitrary projection of the angular momentum, i.e.,

$$|\omega s\sigma\rangle=\left[\frac{(s-\sigma)!2^{s+\sigma}}{(2s)!(s+\sigma)!}\right]^{1/2}(-J_+)^{s+\sigma}|\omega s\rangle, \tag{3.9a}$$

$$|\omega s-s\rangle\equiv|\omega s\rangle, \tag{3.9b}$$

where $\sigma=-s,-s+1,\dots,s$ and $|\omega s\sigma\rangle$ is an eigenket of $\mathscr{N}$, J^2,J_0 with eigenvalues $(\omega+n/2)$, $s(s+1)$, and σ.

The $B_i^\dagger$, $i=1,2,3$, are the components of a vector $\mathbf{B}^\dagger$ with respect to the J_q, $q=\pm,0$, and thus polynomials in this vector can be characterized by their degree ν in its components, by an irrep l of the SU(2) group whose generators are J_q, $q=\pm,0$ and by an irrep μ of the O(2) subgroup whose generator is J_0, i.e.,[23]

$$P_{\nu l\mu}(\mathbf{B}^\dagger)=(\mathbf{B}^\dagger\cdot\mathbf{B}^\dagger)^{(\nu-l)/2}\mathscr{Y}_{l\mu}(\mathbf{B}^\dagger), \tag{3.10}$$

where $\mathscr{Y}_{l\mu}$ is of the solid harmonic $\mathscr{Y}_{l\mu}(\mathbf{r})=r^lY_{l\mu}(\theta,\phi)$. Clearly[24]

$$[J_q,P_{\nu l\mu}(\mathbf{B}^\dagger)]$$
$$=[l(l+1)]^{1/2}\langle l\mu,1q|l,\mu+q\rangle P_{\nu l\mu+q}(\mathbf{B}^\dagger), \tag{3.11}$$

where $\langle\ |\ \rangle$ is a Clebsch–Gordan coefficient and thus $P_{\nu l\mu}(\mathbf{B}^\dagger)$ is a Racah tensor[24] of order l and projection μ.

The basis for the irrep (3.8) of Sp(4) can then be written as

$$P_{\nu l\mu}(\mathbf{B}^\dagger)|\omega s\sigma\rangle\equiv P|\omega s\rangle, \tag{3.12a}$$

where

$$\nu=0,1,2,\dots,\quad l=\nu,\nu-2,\dots,1\text{ or }0,$$
$$\mu=l,l-1,\dots,-l,\quad \sigma=s,s-1,\dots,-s, \tag{3.12b}$$

and the right-hand side is a shorthand notation we shall employ later.

The application of J_q to the states (3.11) implies

$$J_qP|\omega s\rangle=[J_q,P]|\omega s\rangle+PJ_q|\omega s\rangle, \tag{3.13}$$

and thus from (3.9) and (3.11) we immediately conclude that the kets

$$|\omega\nu[ls]jm\rangle\equiv\sum_{\mu,\sigma}\langle l\mu,s\sigma|jm\rangle P_{\nu l\mu}(\mathbf{B}^\dagger)|\omega s\sigma\rangle, \tag{3.14}$$

are an alternative basis to (3.12a) with the advantage that they are eigenkets of the Hermitian operators $\mathscr{N}$, J^2, J_0 with eigenvalues $(\nu+\omega+n/2)$, $j(j+1)$, $m=j,j-1,\dots,-j$.

Note that while the kets $|\omega\nu[l,s]jm\rangle$ are orthogonal if they differ in the eigenvalues ν, j, or m, this will not be the case for the l which satisfies the conditions

$$|j-s|\leqslant l\leqslant j+s,\quad l=\nu,\nu-2,\dots,1\text{ or }0. \tag{3.15}$$

Thus we have a complete but nonorthonormal basis in which the l is a kind of "multiplicity index."[14]

IV. THE DYSON BOSON REALIZATION

To derive the Dyson form of the boson realization of sp(4) we shall proceed in exactly the same form as was done for sp(2) between Eqs. (2.7) and (2.11). We first require the effect of the creation and annihilation operators appearing in (1.1) on the states (3.12). We designate these operators as $\eta_{\alpha t}$, $\xi_{\alpha t}$, $\alpha=1,2$, $t=1,2,\dots,n$, as now indices i,j,k take the values 1,2,3 and the s appears in lowest weight state $|\omega s\rangle$. We then have

$$\eta_{\alpha t}P|\omega s\rangle=P\eta_{\alpha t}|\omega s\rangle, \tag{4.1}$$
$$\xi_{\alpha t}P|\omega s\rangle=[\xi_{\alpha t},P]|\omega s\rangle+P\xi_{\alpha t}|\omega s\rangle$$

$$= \left(\frac{\partial P}{\partial B_i^\dagger}\frac{\partial B_i^\dagger}{\partial \eta_{\alpha t}}\right)|\omega s\rangle + P\xi_{\alpha t}|\omega s\rangle, \tag{4.2}$$

where repeated indices i, j, k will be summed from 1 to 3. Applying these results twice and summing over the index t from 1 to n, as well as using the definitions (3.1), we obtain

$$B_i^\dagger(P|\omega s\rangle) = (B_i^\dagger P)|\omega s\rangle, \tag{4.3a}$$

$$J_i(P|\omega s\rangle) = \left(-i\epsilon_{ijk}B_j^\dagger\frac{\partial P}{\partial B_k^\dagger}\right)|\omega s\rangle + PJ_i|\omega s\rangle, \tag{4.3b}$$

$$\begin{aligned} B_i(P|\omega s\rangle) = {} & \left(-B_i^\dagger\frac{\partial^2 P}{\partial B_j^\dagger \partial B_j^\dagger}\right)|\omega s\rangle \\ & + \left[\left(2B_j^\dagger\frac{\partial}{\partial B_j^\dagger} + 2\omega + n\right)\frac{\partial P}{\partial B_i^\dagger}\right]|\omega s\rangle \\ & - \left[2i\epsilon_{ijk}\frac{\partial P}{\partial B_j^\dagger}J_k\right]|\omega s\rangle, \end{aligned} \tag{4.3c}$$

$$\mathcal{N}(P|\omega s\rangle) = \left[\left(B_j^\dagger\frac{\partial P}{\partial B_j^\dagger}\right) + \omega + \frac{n}{2}\right]|\omega s\rangle, \tag{4.3d}$$

where again repeated indices i, j, k are summed from 1 to 3.

As in the discussion following Eq. (2.8) for sp(2), we see that the effect of the generators of sp(4) on the states (3.12) can be expressed in terms of operators $B_i^\dagger$, $i = 1,2,3$, acting multiplicatively on P, of differential operators $\partial/\partial B_i^\dagger$ acting also on P and, besides, of operators J_i acting only on the intrinsic states $|\omega s\rangle$. To express the generators of sp(4) in terms of these operators it is then convenient to introduce the definitions

$$\beta_i^+ \equiv B_i^\dagger, \tag{4.4a}$$

$$\beta_i \equiv \frac{\partial}{\partial B_i^\dagger}, \tag{4.4b}$$

as well as a new vector operator

$$s_i, \quad i = 1,2,3, \tag{4.4c}$$

acting exclusively on the intrinsic states. These new operators will then satisfy the commutation relations

$$[\beta_i,\beta_j] = [\beta_i^+,\beta_j^+] = [\beta_i,s_j] = [\beta_i^+,s_j] = 0, \tag{4.5a}$$

$$[\beta_i,\beta_j^+] = \delta_{ij}, \tag{4.5b}$$

$$[s_i,s_j] = i\epsilon_{ijk}s_k, \tag{4.5c}$$

which are the same as those of a Lie algebra that is the direct sum of a Weyl algebra in three dimensions and an independent two-dimensional unitary unimodular Lie algebra, i.e., $w(3) \oplus su(2)$.

From (4.3) we can then express the generators of sp(4) in vector notation as follows:

$$\mathbf{B}^\dagger = \boldsymbol{\beta}^+, \tag{4.6a}$$

$$\mathbf{J} = \mathbf{l} + \mathbf{s}, \tag{4.6b}$$

$$\mathbf{B} = -\boldsymbol{\beta}^+(\boldsymbol{\beta}\cdot\boldsymbol{\beta}) + (2\mathfrak{N} + 2\omega + n)\boldsymbol{\beta} - 2i(\boldsymbol{\beta}\times\mathbf{s}), \tag{4.6c}$$

$$\mathcal{N} = \mathfrak{N} + (\omega + n/2), \tag{4.6d}$$

where

$$\mathbf{l} = -i(\boldsymbol{\beta}^+\times\boldsymbol{\beta}), \tag{4.7a}$$

$$\mathfrak{N} = \boldsymbol{\beta}^+\cdot\boldsymbol{\beta}. \tag{4.7b}$$

It is easy to check that if β_i^+, β_i, s_i satisfy the commutation rules (4.5) then $B_i^\dagger, J_i, B_i, \mathcal{N}$ satisfy the commutation rules (3.2).

Equations (4.6) for the generators of sp(4) are equivalent to (2.11) for sp(2) and we note again that here β_i^+ *is not* the Hermitian conjugate of β_i so that we require the passage from the Dyson to the Holstein–Primakoff realization.

V. THE HOLSTEIN–PRIMAKOFF BOSON REALIZATION

The Holstein–Primakoff realization would be in terms of operators

$$b_i^\dagger, \quad b_i, \quad S_i, \quad i = 1,2,3, \tag{5.1}$$

satisfying the same commutation rules as in (4.5) but where now $b_i^\dagger$ is the Hermitian conjugate of b_i and S_i is Hermitian.

As in (2.13) for the case of sp(2) we assume that the operators (5.1) and (4.4) are related by a similarity transformation with an operator K, i.e.,

$$b_i^\dagger = K^{-1}\beta_i^+K, \tag{5.2a}$$

$$b_i = K^{-1}\beta_iK, \tag{5.2b}$$

$$S_i = K^{-1}s_iK, \tag{5.2c}$$

as in this way the appropriate commutation rules for $b_i^\dagger, b_i, S_i$ follow immediately from (4.5). As $(\beta_i^+)^\dagger \neq \beta_i$, while $(b_i^\dagger)^\dagger = b_i$, the K is not unitary but we can assume, without loss of generality, that it is Hermitian. We shall furthermore impose the additional condition that K should be an invariant[18,19,21] of the u(2) subalgebra of sp(4), i.e.,

$$[K,\mathcal{N}] = [K,J_i] = 0. \tag{5.3}$$

Through Eqs. (4.6) and (5.2) we then obtain in vector notation

$$\mathbf{B}^\dagger = K\mathbf{b}^\dagger K^{-1}, \tag{5.4a}$$

$$\mathbf{J} = \mathbf{L} + \mathbf{S}, \tag{5.4b}$$

$$\mathbf{B} = K[-\mathbf{b}^\dagger(\mathbf{b}\cdot\mathbf{b}) + (2N + 2\omega + n)\mathbf{b} - 2i(\mathbf{b}\times\mathbf{S})]K^{-1}, \tag{5.4c}$$

$$\mathcal{N} = N + (\omega + n/2), \tag{5.4d}$$

where

$$\mathbf{L} = -i(\mathbf{b}^\dagger\times\mathbf{b}), \tag{5.5a}$$

$$N = \mathbf{b}^\dagger\cdot\mathbf{b}. \tag{5.5b}$$

To obtain (5.4b) and (5.4d) we passed K, K^{-1} to the left-hand side and from (5.3) wrote $K^{-1}\mathbf{J}K = \mathbf{J}$, $K^{-1}\mathcal{N}K = \mathcal{N}$.

From (1.1) and (3.1) we see that $\mathbf{B}^\dagger$ is the Hermitian conjugate of $\mathbf{B}$ and as K is Hermitian we get

$$\mathbf{B} = K^{-1}\mathbf{b}K, \tag{5.6}$$

which combined with (5.4c) gives us the equation

$$\mathbf{b}K^2 = K^2[-\mathbf{b}^\dagger(\mathbf{b}\cdot\mathbf{b}) + (2N + 2\omega + n)\mathbf{b} - 2i(\mathbf{b}\times\mathbf{S})], \tag{5.7}$$

which is the equivalent for sp(4) of the relation (2.17) for sp(2).

We now wish to see whether from (5.7) we can get the operator form of K. This requires, as in the arguments following (2.17), a complete set of boson states. As now besides the $b_i^\dagger$, $i = 1,2,3$, we have the completely independent elements S_i of a su(2) Lie algebra, our boson states will be formed by coupling the standard harmonic oscillator states in three dimensions[23]

$$A_{\nu l}\, P_{\nu l\mu}(\mathbf{b}^\dagger)|0), \tag{5.8}$$

where $P_{\nu l\mu}(\mathbf{b}^\dagger)$ is given by (3.10) with $b^\dagger$ replacing $B^\dagger$ and[23]

$$A_{\nu l} = (-1)^{(\nu - l)/2}[4\pi/(\nu - l)!!(\nu + l + 1)!!]^{1/2}, \tag{5.9}$$

with the orthonormalized states $|s\sigma)$ for an independent su(2) Lie algebra, i.e.,

$$|\nu[l,s]jm) = \sum_{\mu,\sigma} \langle l\mu,s\sigma|jm\rangle \{[A_{\nu l}\, P_{\nu l\mu}(b^\dagger)|0)]\,|s\sigma)\}. \tag{5.10}$$

We denote these boson states by round kets to distinguish them from the angular kets (3.14) that are basis of the irrep (3.8) of sp(4). We furthermore notice that, from construction, the states (5.10) are orthonormal in all their quantum numbers as the kets (5.10) are eigenkets of the Hermitian operators N, L^2, S^2, J^2, J_0 with eigenvalues ν, $l(l+1)$, $s(s+1)$, $j(j+1)$, m. Note that here we interpret J_i as the *boson operator* $J_i = L_i + S_i$, $i = 1,2,3$, which, from (5.4b), equals the generator J_i of the su(2) subalgebra of sp(4).

Our next step would be to take the matrix elements of the left- and right-hand sides of (5.7) with respect to the states (5.10). Before proceeding we first rewrite (5.7) in a more convenient form using the relations

$$i(\mathbf{L}\times\mathbf{b}) = N\mathbf{b} - \mathbf{b}^\dagger(\mathbf{b}\cdot\mathbf{b}), \tag{5.11a}$$

$$[L^2,\mathbf{b}] = -2i(\mathbf{L}\times\mathbf{b}) - 2\mathbf{b}, \tag{5.11b}$$

$$\begin{aligned}[J^2,\mathbf{b}] &= -2i(\mathbf{J}\times\mathbf{b}) - 2\mathbf{b}\\ &= -2i(\mathbf{L}\times\mathbf{b}) - 2i(\mathbf{S}\times\mathbf{b}) - 2\mathbf{b},\end{aligned} \tag{5.11c}$$

to get

$$\mathbf{b}K^2 = K^2\{(N + 2\omega + n - 1)\mathbf{b} + \tfrac{1}{2}[L^2,\mathbf{b}] - [J^2,\mathbf{b}]\}. \tag{5.12}$$

We introduce the notation

$$(\nu[l's]jm|K|\nu[ls]jm) \equiv K_{l'l}(\nu,j,s), \tag{5.13}$$

which from (5.3) are the only matrix elements different from zero and in which l',l are restricted by (3.15). Furthermore we denote the matrix elements for the corresponding operator K^2 appearing in (5.12) as

$$(\nu[l's]jm|K^2|\nu[ls]jm) \equiv \mathscr{M}_{l'l}(\nu,j,s). \tag{5.14}$$

Taking now the reduced matrix elements of the left- and right-hand side of (5.12) with respect to the states (5.10), we get the following recursion relation for (5.14):

$$\begin{aligned}&\sum_{\bar l} (\nu - 1[l's]j'\|\mathbf{b}\|\nu[\bar l s]j)\mathscr{M}_{\bar l l}(\nu,j,s)\\ &\quad= \sum_{\bar l'} \mathscr{M}_{l'\bar l'}(\nu - 1,j',s)\{(\nu + 2\omega + n - 2)\\ &\qquad+ \tfrac{1}{2}[\bar l'(\bar l' + 1) - l(l+1)] - [j'(j'+1) - j(j+1)]\}\\ &\qquad\times(\nu - 1[\bar l's]j'\|\mathbf{b}\|\nu[ls]j).\end{aligned} \tag{5.15}$$

From the fact that $b^\dagger$ is the Hermitian conjugate of b we get that[24]

$$\begin{aligned}&(\nu - 1[l's]j'\|\mathbf{b}\|\nu[ls]j)\\ &\quad= (-1)^{j-j'}[(2j+1)/(2j'+1)]^{1/2}\\ &\qquad\times(\nu[ls]j\|\mathbf{b}^\dagger\|\nu - 1[l's]j'),\end{aligned} \tag{5.16}$$

while standard Racah algebra[24] gives us

$$\begin{aligned}&(\nu[ls]j\|\mathbf{b}^\dagger\|\nu - 1[l's]j')\\ &\quad= (-1)^{s+1-l'-j}[(2l+1)(2j'+1)]^{1/2}W(l'lj'j;1s)\\ &\qquad\times(\nu l\|\mathbf{b}^\dagger\|\nu - 1l')\end{aligned} \tag{5.17}$$

with W being a Racah coefficient, and the last reduced matrix element takes the value[23]

$$\begin{aligned}&(\nu l\|\mathbf{b}^\dagger\|\nu - 1l')\\ &\quad= [(\nu + l + 1)l/(2l+1)]^{1/2}\delta_{l',l-1}\\ &\qquad+ [(\nu - l)(l+1)/(2l+1)]^{1/2}\delta_{l',l+1}.\end{aligned} \tag{5.18}$$

In Appendix A we discuss the recursion relation (5.15) for $\mathscr{M}_{l'l}(\nu,j,s)$ and show that for a given spin s they can be solved in an analytic fashion though, in general, rather laboriously. The main problem is how to extract from $\mathscr{M}_{l'l}(\nu,j,s)$ the $K_{l'l}(\nu,j,s)$ that we need to establish in explicit, analytic, and closed form, through $\mathbf{B}^\dagger = K\mathbf{b}^\dagger K^{-1}$, the generators of sp(4) in terms of those of $w(3) \oplus$ su(2). In the next section we show that this is not possible in general as it requires the solution of an algebraic equation of high degree. We can though carry it out for low values of the spin as we also show in Sec. VI.

VI. DETERMINATION OF THE MATRIX FORM OF K FOR PARTICULAR SPINS AND THE CORRESPONDING BOSON REALIZATIONS

In Appendix A we show how to get, for a given value of the spin s, the explicit matrix elements $\mathscr{M}_{l'l}(\nu,j,s)$ with respect to the states (5.10), associated with the operator K^2. For fixed ν, j we have then the finite real symmetric matrices

$$\mathbf{M}(\nu,j,s) = \|\mathscr{M}_{l'l}(\nu,j,s)\|, \tag{6.1}$$

where l',l are restricted by (3.15).

The matrix

$$\mathbf{K}(\nu,j,s) = \|K_{l'l}(\nu,j,s)\| \tag{6.2}$$

can then be obtained by first diagonalizing $\mathbf{M}(\nu,j,s)$ through an orthogonal transformation

$$\mathbf{M} = \tilde{\mathscr{O}}\mathbf{D}\mathscr{O}, \tag{6.3}$$

where $\mathbf{D}$ is diagonal with all elements positive and $\tilde{\mathscr{O}}$ is the transpose of $\mathscr{O}$. We take the square root of this diagonal matrix to get

$$\mathbf{K} = \tilde{\mathscr{O}}\mathbf{D}^{1/2}\mathscr{O} \tag{6.4}$$

as then $\mathbf{K}^2 = \mathbf{M}$.

Immediately one problem becomes apparent. The explicit form of $\mathbf{M}$ for a given spin s can be obtained through the procedures of Appendix A. To diagonalize $\mathbf{M}$ we require the solution of the secular equation

$$\det(\mathbf{M} - \lambda\,\mathbf{I}) = 0, \tag{6.5}$$

which, in general, implies an algebraic equation in λ of large degree. Thus the present approach can give us, in general, the matrix $\mathbf{K}$ only numerically.

There are, though, values of s such as $s = 0,\frac{1}{2},1$, where the secular equation *can be solved analytically* as, from the discussion in Appendix A, it does not exceed the second degree. We expect then that for these cases it is possible to get a boson realization of sp(4) similar to that given in (2.24) for sp(2). We proceed to show that this is the case for $s = 0, \frac{1}{2}$, while for $s = 1$ we just discuss the matrix elements $K_{l'l}(\nu, j, 1)$, as the realization follows from them in a similar fashion as in the case $s = 0, \frac{1}{2}$.

Before dealing with the particular cases mentioned, we show that besides the relation $\mathbf{B}^\dagger = \mathbf{K}\mathbf{b}^\dagger\mathbf{K}^{-1}$, we have an alternative way of expressing $\mathbf{B}^\dagger$ in terms of elements of the enveloping algebra of $w(3) \oplus \mathrm{su}(2)$, which is relevant to the boson realization. We note that $\mathbf{b}^\dagger$ is a vector, which, in the ket $|\nu[ls]jm)$ of (5.10), changes $\nu \to \nu + 1$; $j \to j' = j \pm 1, j$, while K is a scalar that does not affect ν, j. As $\mathbf{B}^\dagger$ is also a vector we have the reduced matrix element relation

$$(\nu + 1[l's]j'\|B^\dagger\|\nu[ls]j)$$
$$= \sum_{\bar{l},\bar{l}'} \{K_{l'\bar{l}'}(\nu + 1, j', s) \times (\nu + 1[\bar{l}'s]j'\|b^\dagger\|\nu[\bar{l}s]j) K_{\bar{l}l}^{-1}(\nu, j, s)\}, \tag{6.6}$$

where $K_{l'l}(\nu, j, s)$ is given by (5.13) while the reduced matrix element of $b^\dagger$ appears in (5.17) and (5.18).

The three-dimensional vector $\mathbf{B}^\dagger$ can also be expanded in terms of three independent vectors with the properties of $\mathbf{b}^\dagger$, which we could take as

$$\mathbf{b}^\dagger, \quad [L^2, \mathbf{b}^\dagger], \quad [J^2, \mathbf{b}^\dagger]. \tag{6.7}$$

The coefficients in this expansion, which we denote by F, G, H, have to be invariants of the u(2) Lie subalgebra of sp(4) if we want to satisfy the commutation relations (3.2).

We thus have

$$\mathbf{B}^\dagger = \mathbf{b}^\dagger F + [L^2, \mathbf{b}^\dagger] G + [J^2, \mathbf{b}^\dagger] H. \tag{6.8}$$

An alternative form for $\mathbf{B}^\dagger$ can be obtained using the relation

$$[L^2, \mathbf{b}^\dagger] = -2i(\mathbf{L} \times \mathbf{b}^\dagger) - 2\mathbf{b}^\dagger, \tag{6.9a}$$

$$[J^2, \mathbf{b}^\dagger] = -2i(\mathbf{J} \times \mathbf{b}^\dagger) - 2\mathbf{b}^\dagger, \tag{6.9b}$$

as well as $\mathbf{J} = \mathbf{L} + \mathbf{S}$. We can then write

$$\mathbf{B}^\dagger = \mathbf{b}^\dagger f + i(\mathbf{L} \times \mathbf{b}^\dagger) g + i(\mathbf{S} \times \mathbf{b}^\dagger) h, \tag{6.10}$$

where

$$f = F - 2G - 2H, \quad g = -2(G + H), \quad h = -2H. \tag{6.11}$$

We now take the reduced matrix element of both sides of (6.8) with respect to the states $|\nu[ls]jm)$ of (5.10) and obtain

$$(\nu + 1[l's]j'\|\mathbf{B}^\dagger\|\nu[ls]j)$$
$$= \sum_{\bar{l}} (\nu + 1[l's]j'\|\mathbf{b}^\dagger\|\nu[\bar{l}s]j) \times \{F_{\bar{l}l}(\nu, j, s) + [l'(l'+1) - \bar{l}(\bar{l}+1)] G_{\bar{l}l}(\nu j s) + [j'(j'+1) - j(j+1)] H_{\bar{l}l}(\nu, j, s)\}, \tag{6.12}$$

where

$$F_{\bar{l}l}(\nu, j, s) = (\nu[\bar{l}s]jm|F|\nu[ls]jm), \tag{6.13}$$

and similarly for $G_{\bar{l}l}(\nu, j, s)$ and $H_{\bar{l}l}(\nu, j, s)$, as F, G, and H are invariant with respect to the u(2) subalgebra of sp(4).

If we equate (6.6) and (6.12) we get a system of linear equations in the unknowns $F_{\bar{l}l}(\nu, j, s)$, $G_{\bar{l}l}(\nu, j, s)$, $H_{\bar{l}l}(\nu, j, s)$, in which we consider ν, j, s fixed while $j' = j \pm 1, j$ and l' takes all values consistent with $\nu + 1, j'$ through the rules (3.15) while the same holds for $\bar{l}, l$ with respect to ν, j.

Once we have determined the unknowns we can identify them with matrix elements of definite operators that are functions of the generators of the sp(4) that are invariant under the u(2) Lie subalgebra of sp(4), thus getting in the form (6.10) the boson realization for given spin.

A. Boson realization for $s = 0$

When the spin $s = 0$ we have that $\mathbf{J} = \mathbf{L}$ and thus the states (5.10) can be written as

$$|\nu[l0]lm) \equiv |\nu lm\}. \tag{6.14}$$

The matrix elements $K_{l'l}(\nu, j, s)$ for $s = 0$ become

$$K_{ll}(\nu, l, 0) = \mathscr{M}_{ll}^{1/2}(\nu, l, 0) \equiv K(\nu, l), \tag{6.15}$$

and from (A10) for $s = 0$ they become

$$K(\nu l) = \left[\frac{(\nu + l + 2\omega + n - 2)!!(\nu - l + 2\omega + n - 3)!!}{(2\omega + n - 2)!!(2\omega + n - 3)!!}\right]^{1/2}. \tag{6.16}$$

From (6.6) and (6.14) we now have

$$\{\nu + 1, l'\|\mathbf{B}^\dagger\|\nu l\} = K(\nu + 1, l')\{\nu + 1, l'\|\mathbf{b}^\dagger\|\nu l\} K^{-1}(\nu, l), \tag{6.17}$$

and from (6.16) we obtain for the two possibilities $l' = l \pm 1$ the equations

$$\{\nu + 1, l + 1\|\mathbf{B}^\dagger\|\nu l\} = (\nu + l + n + 2\omega)^{1/2}\{\nu + 1, l + 1\|\mathbf{b}^\dagger\|\nu l\}, \tag{6.18a}$$

$$\{\nu + 1, l - 1\|\mathbf{B}^\dagger\|\nu l\} = (\nu - l + n + 2\omega - 1)^{1/2}\{\nu + 1, l - 1\|\mathbf{b}^\dagger\|\nu l\}. \tag{6.18b}$$

Turning now our attention to the expression (6.8) for $\mathbf{B}^\dagger$, we note that as $\mathbf{S} = 0$ and $\mathbf{J} = \mathbf{L}$ we can write it just as

$$\mathbf{B}^\dagger = \mathbf{b}^\dagger F + [L^2, \mathbf{b}^\dagger] G, \tag{6.19}$$

from which we get the matrix element

$$\{\nu + 1, l'\|\mathbf{B}^\dagger\|\nu l\} = \{\nu + 1, l'\|\mathbf{b}^\dagger\|\nu l\} \times \{F(\nu, l) + [l'(l'+1) - l(l+1)] G(\nu, l)\}, \tag{6.20a}$$

where

$$F(\nu, l) = \{\nu lm|F|\nu lm\}, \tag{6.20b}$$

as it is independent of m, and similarly for $G(\nu, l)$.

Substituting $l' = l \pm 1$ in (6.20) and comparing with (6.18) we get the equations

$$F(\nu, l) + 2(l + 1)G(\nu, l) = (\nu + l + 2\omega + n)^{1/2}, \tag{6.21a}$$

$$F(\nu, l) - 2lG(\nu, l) = (\nu - l + 2\omega + n - 1)^{1/2}, \tag{6.21b}$$

from which

$$F(\nu, l) = (2l + 1)^{-1}\{l(\nu + l + 2\omega + n)^{1/2} + (l + 1)(\nu - l + 2\omega + n - 1)^{1/2}\}, \tag{6.22a}$$

$$G(\nu, l) = [2(2l + 1)]^{-1}\{(\nu + l + 2\omega + n)^{1/2} - (\nu - l + 2\omega + n - 1)^{1/2}\}. \tag{6.22b}$$

From (6.9a) we see that Eq. (6.19) can also be written as

$$\mathbf{B}^\dagger = \mathbf{b}^\dagger f + i(\mathbf{L} \times \mathbf{b}^\dagger) g, \tag{6.23}$$

where

$$f = F - 2G, \quad g = -2G. \tag{6.24}$$

We can now replace in (6.22) ν, l by the operators

$$N, \quad L \equiv (\mathbf{L}^2 + \tfrac{1}{4})^{1/2} - \tfrac{1}{2}, \tag{6.25}$$

which give precisely these eigenvalues when acting on $|\nu lm\}$. Thus we obtain

$$f(N,L) = \{(L-1)(2L+1)^{-1}(N+L+2\omega+n)^{1/2} + (L+2)(2L+1)^{-1}(N-L+2\omega+n-1)^{1/2}\}, \tag{6.26a}$$

$$g(N,L) = (2L+1)^{-1}\{(N-L+2\omega+n-1)^{1/2} - (N+L+2\omega+n)^{1/2}\}, \tag{6.26b}$$

where the square roots and reciprocals cause us no trouble as they are operators diagonal in the basis $|\nu lm\}$.

The expressions (6.23) and (6.26) give us then the boson realization of sp(4) for the spin $s = 0$ when they are supplemented with

$$\mathbf{B} = (f - 2g)\mathbf{b} - ig(\mathbf{L}\times\mathbf{b}), \tag{6.23'}$$

as well as with

$$\mathbf{J} = \mathbf{L}, \tag{6.27a}$$

$$\mathscr{N} = N + (\omega + n/2), \tag{6.27b}$$

where (6.27a) and (6.27b) come from (5.4b) and (5.4d) when $\mathbf{S} = 0$, while (6.23′) follows from (6.23) by Hermitian conjugation, where we note that from (6.26) f, g are Hermitian.

The expressions (6.23) and (6.27) are then the boson realization of sp(4) for $\mathbf{S} = 0$ and they correspond to (2.24) for sp(2). We would like now to invert them to get $\mathbf{b}^\dagger$ and $\mathbf{b}$ in terms of $\mathbf{B}^\dagger$, $\mathbf{B}$, $\mathbf{J}$, and $\mathscr{N}$, as was done in (2.26) for sp(2). For this purpose we need only the equation

$$i(\mathbf{J}\times\mathbf{B}^\dagger) = \mathbf{b}^\dagger(\mathbf{L}^2 - 2)g + i(\mathbf{L}\times\mathbf{b}^\dagger)(f - 3g), \tag{6.28}$$

where to get the right-hand side we use (6.23) and (6.27a). From (6.23) and (6.28) we then obtain

$$\mathbf{b}^\dagger = [\mathbf{B}^\dagger(f-3g) - i(\mathbf{J}\times\mathbf{B}^\dagger)g][f^2 - 3fg - (J^2-2)g^2]^{-1}, \tag{6.29}$$

where f and g are given by (6.26) in which, from (6.27a) and (6.27b), we replace N by $\mathscr{N} - (\omega + n/2)$ and $\mathbf{L}$ by $\mathbf{J}$, where now

$$J \equiv (\mathbf{J}^2 + \tfrac{1}{4})^{1/2} - \tfrac{1}{2}. \tag{6.30}$$

Taking the Hermitian conjugate of (6.29) we finally obtain

$$\mathbf{b} = [f^2 - 3fg - (J^2-2)g^2]^{-1}[(f-g)\mathbf{B} + ig(\mathbf{J}\times\mathbf{B})], \tag{6.29'}$$

and thus we have in (6.29) and (6.29′) the boson operators for the "closed shell" case corresponding to $s = 0$.

B. Boson realization for $s = \frac{1}{2}$

When the spin $s = \frac{1}{2}$ the states (5.10) take the form

$$|\nu[l\tfrac{1}{2}]jm), \tag{6.31}$$

where $l = j \pm \frac{1}{2}$. Thus the matrices $\mathbf{M}$ and $\mathbf{K}$ of (6.1) and (6.2) for fixed ν, j are 2×2 and besides they are diagonal, as from (3.15) both $\nu - l'$ and $\nu - l$ must be even. We have then that

$$K_{ll}(\nu,j,\tfrac{1}{2}) = \mathscr{M}_{ll}^{1/2}(\nu,j,\tfrac{1}{2}), \tag{6.32}$$

and from (A22) and (A23) for $s = \frac{1}{2}$ we have that

$$K_{j+1/2,j+1/2}(\nu,j,\tfrac{1}{2}) \equiv K_+(\nu,j) = \left[\frac{(\nu+j+2\omega+n-\frac{5}{2})!!(\nu-j+2\omega+n-\frac{5}{2})!!}{(2\omega+n-1)!!(2\omega+n-4)!!}\right]^{1/2}, \tag{6.33a}$$

$$K_{j-1/2,j-1/2}(\nu,j,\tfrac{1}{2}) \equiv K_-(\nu,j) = \left[\frac{(\nu+j+2\omega+n-\frac{3}{2})!!(\nu-j+2\omega+n-\frac{7}{2})!!}{(2\omega+n-1)!!(2\omega+n-4)!!}\right]^{1/2}, \tag{6.33b}$$

where we have made use of in (A22) and (A23) that $\mathscr{M}_{00}(0,\frac{1}{2},\frac{1}{2}) = 1$ and $\mathscr{M}_{11}(1,\frac{1}{2},\frac{1}{2}) = (2\omega + \mathrm{n} - 2)$.

From the diagonal character of the matrix $\mathbf{K}(\nu,j,\frac{1}{2})$ we see that the summation in the reduced matrix elements of $\mathbf{B}^\dagger$ given by (6.6) disappears and thus we have

$$(\nu+1[l'\tfrac{1}{2}]j'\|\mathbf{B}^\dagger\|\nu[l\tfrac{1}{2}]j) = K_{l'l'}(\nu+1,j',\tfrac{1}{2})(\nu+1[l'\tfrac{1}{2}]j'\|\mathbf{b}^\dagger\|\nu[l\tfrac{1}{2}]j) \times K_{ll}^{-1}(\nu,j,\tfrac{1}{2}). \tag{6.34}$$

We note furthermore that from (6.13) $F_{l'l}(\nu,j,\frac{1}{2})$ is also diagonal in the indices l', l as again $l', l = j \pm \frac{1}{2}$ and $\nu - l, \nu - l'$ are even. As this applies also to G and H, we see that the sum in the reduced matrix element of $\mathbf{B}^\dagger$ given by (6.12) also disappears and we have

$$(\nu+1[l'\tfrac{1}{2}]j'\|B^\dagger\|\nu[l\tfrac{1}{2}]j) = (\nu+1[l'\tfrac{1}{2}]j'\|b^\dagger\|\nu[l\tfrac{1}{2}]j) \times\{F_{ll}(\nu,j,\tfrac{1}{2}) + [l'(l'+1) - l(l+1)]G_{ll}(\nu,j,\tfrac{1}{2}) + [j'(j'+1) - j(j+1)]H_{ll}(\nu,j,\tfrac{1}{2})\}. \tag{6.35}$$

Equating (6.34) and (6.35) we see that the reduced matrix element of $\mathbf{b}^\dagger$ will cancel and thus we get

$$F_{ll}(\nu,j,\tfrac{1}{2}) + (l'-l)(l+l'+1)G_{ll}(\nu,j,\tfrac{1}{2}) + (j'-j)(j+j'+1)H_{ll}(\nu,j,\tfrac{1}{2}) = K_{l'l'}(\nu+1,j',\tfrac{1}{2})K_{ll}^{-1}(\nu,j,\tfrac{1}{2}). \tag{6.36}$$

We note from (3.15) that for $l = j + \frac{1}{2}$ we have for (l', j') the admissible values

$$(l',j') = (j+\tfrac{3}{2}, j+1), \quad (j-\tfrac{1}{2}, j), \quad (j-\tfrac{1}{2}, j-1), \tag{6.37a}$$

while for $l = j - \frac{1}{2}$ they become

$$(l',j') = (j+\tfrac{1}{2}, j+1), \quad (j+\tfrac{1}{2}, j), \quad (j-\tfrac{3}{2}, j-1). \tag{6.37b}$$

Substituting these values in (6.36) and using (6.33) as well as the notation

$$F_{ll}(\nu,j,\tfrac{1}{2}) = F_{j\pm1/2,j\pm1/2}(\nu,j,\tfrac{1}{2}) \equiv F_\pm(\nu,j), \tag{6.38}$$

and similar ones for G, H, we obtain two sets of three linear equations, one for the unknowns F_+, G_+, H_+ and the other for F_-, G_-, H_-. Thus we easily determine the six unknowns as functions of ν, j in a similar fashion as we did in (6.22) for $s = 0$.

We want though to express $F_\pm, G_\pm, H_\pm$ not as functions of ν, j but as operators invariant under the u(2) subalgebra of sp(4). For this purpose we note that, for fixed ν, j, the 2×2 matrix

$$\mathbf{F}(\nu,j) = \|F_{l'l}(\nu,j,\tfrac{1}{2})\|, \quad l',l = j \pm \tfrac{1}{2}, \tag{6.39}$$

becomes

$$\mathbf{F}(\nu,j) = \begin{bmatrix} F_+(\nu,j) & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{if } \nu - j - \tfrac{1}{2} \text{ is even}, \tag{6.40a}$$

$$\mathbf{F}(\nu,j) = \begin{bmatrix} 0 & 0 \\ 0 & F_-(\nu,j) \end{bmatrix}, \quad \text{if } \nu - j - \tfrac{1}{2} \text{ is odd}. \tag{6.40b}$$

Thus if we want to write it in operator form we need projection operators that distinguish between the two cases. Using $\mathbf{L}^2$, whose eigenvalues are $l(l+1)$ with $l = j \pm \frac{1}{2}$, we immediately notice that the operators

$$[-(2J-1)/4 + \mathbf{L}^2/(2J+1)], \tag{6.41a}$$

$$[(2J+3)/4 - \mathbf{L}^2/(2J+1)], \tag{6.41b}$$

where J is given by (6.30), have the appropriate projection property as for fixed ν, j, (6.41a) has the matrix representation

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{if } \nu - j - \tfrac{1}{2} \text{ even},$$

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{if } \nu - j - \tfrac{1}{2} \text{ odd}, \tag{6.42a}$$

while for (6.41b) we obtain

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{if } \nu - j - \tfrac{1}{2} \text{ even},$$

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad \text{if } \nu - j - \tfrac{1}{2} \text{ odd}. \tag{6.42b}$$

The operator F can then be written as

$$F = F_+(N,J)\left[-\frac{2J-1}{4} + \frac{\mathbf{L}^2}{2J+1}\right] + F_-(N,J)\left[\frac{2J+3}{4} - \frac{\mathbf{L}^2}{2J+1}\right], \tag{6.43}$$

where the $F_\pm(N,J)$ are obtained when we replace in $F_\pm(\nu,j)$, ν by $N = \mathbf{b}^\dagger \cdot \mathbf{b}$ and j by J of (6.30). A similar result holds for the operators G and H also appearing in (6.8).

Writing now $\mathbf{B}^\dagger$ in the form (6.10), where f,g,h are related with F,G,H by (6.11) we finally obtain that the former have the operator form

$$\begin{aligned} f = {} & (4J)^{-1}[(2J+3)(N - J + 2\omega + n - \tfrac{1}{2})^{1/2} \\ & + (2J-3)(N + J + 2\omega + n - \tfrac{1}{2})^{1/2}] \\ & \times[-(2J-1)/4 + \mathbf{L}^2/(2J+1)] \\ & + [4(J+1)]^{-1}[(2J+5)(N - J + 2\omega + n - \tfrac{3}{2})^{1/2} \\ & + (2J-1)(N + J + 2\omega + n + \tfrac{1}{2})^{1/2}] \\ & \times[2J+3)/4 - \mathbf{L}^2/(2J+1)], \end{aligned} \tag{6.44a}$$

$$\begin{aligned} g = {} & (2J)^{-1}[(N - J + 2\omega + n - \tfrac{1}{2})^{1/2} \\ & - (N + J + 2\omega + n - \tfrac{1}{2})^{1/2}] \\ & \times[-(2J-1)/4 + \mathbf{L}^2/(2J+1)] \\ & + [2(J+1)]^{-1}[(N - J + 2\omega + n - \tfrac{3}{2})^{1/2} \\ & - (N + J + 2\omega + n + \tfrac{1}{2})^{1/2}] \\ & \times[(2J+3)/4 - \mathbf{L}^2/(2J+1)], \end{aligned} \tag{6.44b}$$

$$h = 2g. \tag{6.44c}$$

The boson realization of sp(4) for $s = \frac{1}{2}$ is given by the $\mathbf{B}^\dagger$ of (6.10), in which f,g,h have the form (6.44), by the $\mathbf{B}$ which is the Hermitian conjugate of $\mathbf{B}^\dagger$, by $\mathbf{J} = \mathbf{L} + \mathbf{S}$, and by $\mathscr{N} = N + \omega + n/2$. Thus we have an explicit, analytic, and closed realization for one case of "open shells."

C. The matrix form of K for $s = 1$

We saw from (6.6) and (6.12) that a boson realization of sp(4) is possible for any spin s for which we know the matrix elements $K_{l'l}(\nu,j,s)$. This was corroborated in the previous subsections for $s = 0,\frac{1}{2}$. For $s = 1$ the analysis is similar though longer and we shall only indicate here how we arrive at the matrix elements of $\mathbf{K}$.

When the spin $s = 1$, the states (5.10) have $l = j \pm 1, j$, and $\nu - l$ must be even. Thus the matrix $\mathbf{M}$ of (6.1), whose elements we designate, as in Appendix A, by

$$M_{\tau'\tau}(\nu,j,1) = \mathscr{M}_{j+\tau',j+\tau}(\nu,j,1), \quad \tau,\tau' = 1,0,-1, \tag{6.45}$$

vanishes except for $M_{00}(\nu,j,1)$ when $\nu - j$ is even, while for $\nu - j$ odd only the matrix elements with $\tau,\tau' = \pm 1$ remain. The set of equations (A7) lead us then to recursion relations. We use one of them to find the element denoted by 1 later on in Fig. 2, which is given explicitly by (A10) when we put in it $s = 1$ and thus for $\nu - j$ odd we have

$$\begin{aligned} M_{1,-1}(\nu,j,1) & = 3(2j+1)^{-1}[j(j+1)(\nu - j + 1)(\nu + j + 2)/20]^{1/2} \\ & \times \frac{(\nu + j + 2\omega + n - 3)!!}{(2\omega + n)!!} \frac{(\nu - j + 2\omega + n - 4)!!}{(2\omega + n - 3)!!} \\ & \times M_{1,-1}(2,1,1), \end{aligned} \tag{6.46a}$$

where, from another of the recursion relations (A7) we get

$$M_{1,-1}(2,1,1) = -4\sqrt{5}[3(2\omega + n - 2)]^{-1}. \tag{6.46b}$$

For $\nu - j$ even the recursion relations lead to the explicit expression (A13), which for $s = 1$ gives

$$M_{00}(\nu,j,1) = \frac{(\nu - j + 2\omega + n - 3)!!(\nu + j + 2\omega + n - 2)!!}{(2\omega + n - 3)!!(2\omega + n)!!}, \tag{6.47}$$

where we took $M_{00}(1,1,1) = 1$.

We use other recursion relations (A7) to find for $\nu - j$ odd the matrix elements

$$\begin{aligned} M_{11}(\nu,j,1) = {} & \Big\{(\nu - j + 2\omega + n - 3) \\ & + \frac{3}{\sqrt{20}} \frac{(\nu - j + 1)(j+1)}{(2j+1)} M_{1,-1}(2,1,1)\Big\} \\ & \times M_{00}(\nu - 1,j,1), \end{aligned} \tag{6.48}$$

$$\begin{aligned} M_{-1,-1}(\nu,j,1) = {} & \Big\{(\nu + j + 2\omega + n - 2) \\ & + \frac{3}{\sqrt{20}} \frac{(\nu + j + 2)j}{(2j+1)} M_{1,-1}(2,1,1)\Big\} \\ & \times M_{00}(\nu - 1,j,1), \end{aligned} \tag{6.49}$$

and as $M_{1,-1}(2,1,1)$,$M_{00}(\nu-1,j,1)$ are already given above we have determined all the matrix elements of $\mathbf{M}$ as $M_{-1,1}(\nu,j,s)=M_{1,-1}(\nu,j,s)$ because of the symmetric character of $\mathbf{M}$. One easily checks that the $M_{\tau'\tau}(\nu,j,1)$ obtained above satisfy *all* of the recursion relations (A7).

We turn now our attention to the matrix $\mathbf{K}$ of (6.2) whose elements we also write in the notation (6.45), i.e.,

$$K_{\tau'\tau}(\nu,j,1)\equiv K_{j+\tau',j+\tau}(\nu,j,1). \tag{6.50}$$

For $\nu-j$ even, as $\mathbf{M}$ is a diagonal matrix with only $M_{00}(\nu,j,1)$ nonvanishing, we get

$$K_{00}(\nu,j,1)=M_{00}^{1/2}(\nu,j,1), \tag{6.51}$$

with all other elements being zero. For $\nu-j$ odd we consider the 2×2 matrix made up of the nonvanishing elements of $M_{\tau'\tau}(\nu,j,1)$, i.e.,

$$\begin{bmatrix} M_{1,1}(\nu,j,1) & M_{1,-1}(\nu,j,1)\\ M_{-1,1}(\nu,j,1) & M_{-1,-1}(\nu,j,1)\end{bmatrix}=\Delta\begin{bmatrix}\alpha & \gamma\\ \gamma & \delta\end{bmatrix}, \tag{6.52}$$

where from (6.46), (6.48), and (6.49) we have

$$\Delta=[(2\omega+n-2)(2j+1)]^{-1}M_{00}(\nu-1,j,1), \tag{6.53a}$$

$$\alpha=[(2j+1)(2\omega+n-2)(\nu-j+2\omega+n-3) - 2(\nu-j+1)(j+1)], \tag{6.53b}$$

$$\gamma=-2[j(j+1)(\nu-j+1)(\nu+j+2)]^{1/2}, \tag{6.53c}$$

$$\delta=[(2j+1)(2\omega+n-2)(\nu+j+2\omega+n-2) - 2j(\nu+j+2)]. \tag{6.53d}$$

The eigenvalues of the matrix (6.52) will be denoted by

$$\lambda_{\pm}^2=\Delta/2\{(\alpha+\delta)\pm[(\alpha-\delta)^2+4\gamma^2]^{1/2}\}, \tag{6.54}$$

while the orthogonal 2×2 matrix that diagonalizes it depends only on an angle ϕ given by

$$\cos 2\phi=(\alpha-\delta)/[(\alpha-\delta)^2+4\gamma^2]^{1/2}$$

or (6.55)

$$\sin 2\phi=2\gamma/[(\alpha-\delta)^2+4\gamma^2]^{1/2}.$$

Thus for $\nu-j$ odd the 3×3 matrix K of (6.4) takes the form

$$\mathbf{K}=\begin{bmatrix}\frac{1}{2}[\lambda_++\lambda_-+(\lambda_+-\lambda_-)\cos 2\phi] & 0 & \frac{1}{2}(\lambda_+-\lambda_-)\sin 2\phi\\ 0 & 0 & 0\\ \frac{1}{2}(\lambda_+-\lambda_-)\sin 2\phi & 0 & \frac{1}{2}[\lambda_++\lambda_--(\lambda_+-\lambda_-)\cos 2\phi]\end{bmatrix}, \tag{6.56}$$

while for $\nu-j$ even, its only nonvanishing matrix element is given by $K_{00}(\nu,j,1)$ of (6.51).

With the help of these K's the matrix elements in (6.6) can be evaluated, and thus from (6.12) we can in turn get the matrix elements of the operators F, G, and H in the basis (5.10). These matrices can be transformed into operators but now they will not only be functions of $N,J,\mathbf{L}^2$, as in (6.44), but will depend also on other invariants of the u(2) Lie subalgebra of sp(4) that we can form from $b_i^\dagger$, b_i, and S_i, as will be suggested by the classical analysis presented in the next section.

VII. THE CLASSICAL LIMIT

In the present section we want to consider both the ten generators of sp(4) in (3.1) and the nine generators of $w(3)\oplus\mathrm{su}(2)$ in (5.1), as classical observables. We wish then to express the generators of sp(4) as functions of those of $w(3)\oplus\mathrm{su}(2)$ and of the parameter ω in such a way that they satisfy the classical counterpart of the commutation relations in (3.2), i.e., when we replace there the commutators by Poisson brackets

$$[F,G]\to i\{F,G\}. \tag{7.1}$$

As in the Introduction, we start by discussing the corresponding problem for sp(2), which will not only serve as a guideline for the sp(4) case, but will also show us how much more complicated is the latter case as compared with the former.

We deal in sp(2) with the observables $B^\dagger$, C, B, which from (2.1) satisfy the Poisson bracket relations

$$\{C,B^\dagger\}=-2iB^\dagger, \tag{7.2a}$$

$$\{C,B\}=2iB, \tag{7.2b}$$

$$\{B,B^\dagger\}=-4iC. \tag{7.2c}$$

Furthermore we would like to express $B^\dagger$, C, B as functions of the boson observables $b^\dagger$ and b (whose Poisson bracket is $\{b^\dagger,b\}=i$) as well as a parameter ω which characterizes the irrep of sp(2). From the relation between standard Poisson brackets and those in terms of creation and annihilation observables[25] we see that

$$\{F,G\}=i\left(\frac{\partial F}{\partial b^\dagger}\frac{\partial G}{\partial b}-\frac{\partial F}{\partial b}\frac{\partial G}{\partial b^\dagger}\right). \tag{7.3}$$

To achieve our purpose we first note that the classical counterpart of the relation between C and $N=b^\dagger b$, given in (2.24c), is

$$C=2N+\omega, \tag{7.4}$$

as the term $n/2$ appearing in (2.24c) comes from $\frac{1}{2}\Sigma_{s=1}^{n}(\xi_s\eta_s-\eta_s\xi_s)$, when $[\xi_s,\eta_t]=\delta_{st}$, while classically this term vanishes.

Turning now our attention to $B^\dagger$, B, considered as classical observables, we immediately see from (7.3) that

$$B^\dagger=f(N)b^\dagger, \tag{7.5a}$$

$$B=f(N)b \tag{7.5b}$$

will satisfy the Poisson bracket relations (7.2a) and (7.2b) for an arbitrary real function $f(N)$. To determine f we could use the relation (7.2c) but it is more direct to employ the classical Casimir operator of sp(2), which has the form[2]

$$G=\tfrac{1}{4}(B^\dagger B-C^2), \tag{7.6}$$

as, from (7.2), the Poisson bracket of G with $B^\dagger$, B, and C vanishes. The value of this Casimir operator[2] can be obtained from (2.5) when we disregard n and 2 as compared with ω, i.e.,

$$G=-\tfrac{1}{4}\omega^2. \tag{7.7}$$

From (7.5)–(7.7) we then get

$$B^\dagger B - C^2 = Nf^2 - (2N+\omega)^2 = -\omega^2, \tag{7.8}$$

from which we obtain

$$f(N) = 2(N+\omega)^{1/2}. \tag{7.9}$$

Thus the classical realization of sp(2) is given by

$$B^\dagger = 2(N+\omega)^{1/2} b^\dagger, \tag{7.10a}$$

$$C = 2N + \omega, \tag{7.10b}$$

$$B = 2(N+\omega)^{1/2} b, \tag{7.10c}$$

as would also follow from (2.24) if we disregard there $n/2$. From (7.10) we see that the Casimir operator (7.6) takes the value $-\omega^2/4$ and, from (7.3), the generators of sp(2) satisfy the Poisson bracket relations (7.2).

We wish now to see how far we can go in extending the previous analysis to the generators of sp(4), i.e., in expressing the observables $B_i^\dagger, J_i, B_i, \mathcal{N}$ [that satisfy the Poisson brackets obtained from (3.2) and (7.1)] in terms of $b_i^\dagger, b_i, S_i$. Note that now the Poisson bracket relation in terms of the latter variables becomes[25,26]

$$\{F,G\} = i\left(\frac{\partial F}{\partial b_i^\dagger}\frac{\partial G}{\partial b_i} - \frac{\partial F}{\partial b_i}\frac{\partial G}{\partial b_i^\dagger}\right) + \frac{\partial F}{\partial S_i}\frac{\partial G}{\partial S_j}\epsilon_{ijk}S_k, \tag{7.11}$$

where repeated indices i, j, k are summed from 1 to 3. The expression (7.11) corresponds for sp(4) to (7.3) for sp(2).

To achieve our purpose we first note that from (5.4b) and (5.4d) the classical expressions for the generators $J_i, \mathcal{N}$ for the u(2) Lie subalgebra of sp(4) are given by

$$\mathbf{J} = \mathbf{L} + \mathbf{S}, \tag{7.12a}$$

$$\mathcal{N} = N + \omega, \tag{7.12b}$$

where

$$\mathbf{L} = -i(\mathbf{b}^\dagger \times \mathbf{b}), \tag{7.12c}$$

$$N = \mathbf{b}^\dagger \cdot \mathbf{b}. \tag{7.12d}$$

We turn now our attention to $\mathbf{B}^\dagger$ considered as a classical observable. As it is a vector, we can expand it in terms of three independent vectors and from (6.10) we choose these to be

$$\mathbf{b}^\dagger, \quad i(\mathbf{L}\times\mathbf{b}^\dagger), \quad i(\mathbf{S}\times\mathbf{b}^\dagger). \tag{7.13}$$

We can then write

$$\mathbf{B}^\dagger = f\mathbf{b}^\dagger + gi(\mathbf{L}\times\mathbf{b}^\dagger) + ih(\mathbf{S}\times\mathbf{b}^\dagger), \tag{7.14}$$

where, from the discussion in Sec. VI, we see that we can take f, g, and h as real and functions of $\mathbf{b}^\dagger$, $\mathbf{b}$, and $\mathbf{S}$ that are invariant under the u(2) Lie subalgebra of sp(4). From (7.12a), (7.12b), and (7.14) we see then immediately that

$$\{J_i, B_j^\dagger\} = \epsilon_{ijk} B_k^\dagger, \tag{7.15a}$$

$$\{\mathcal{N}, B_j^\dagger\} = -iB_j^\dagger, \tag{7.15b}$$

as

$$\{J_i, f\} = \{L_i + S_i, f\} = 0, \tag{7.16a}$$

$$\{\mathcal{N}, f\} = \{N, f\} = 0, \tag{7.16b}$$

and similarly for g and h. Taking the conjugate of (7.14) we then obtain

$$\mathbf{B} = f\mathbf{b} - ig(\mathbf{L}\times\mathbf{b}) - ih(\mathbf{S}\times\mathbf{b}), \tag{7.17}$$

where (7.14) and (7.17) are for sp(4) what (7.5) was for sp(2).

Our next task is to try to determine f, g, and h, and, following the discussion for sp(2), we would like to use for this purpose the Casimir operators of sp(4), which, in the classical picture, have the form[27]

$$G_2 = \mathcal{N}^2 + J^2 - \mathbf{B}^\dagger \cdot \mathbf{B}, \tag{7.18}$$

$$\begin{aligned} G_4 = {} & \mathcal{N}^2 J^2 + i\mathcal{N}[\mathbf{J}\cdot(\mathbf{B}^\dagger\times\mathbf{B})] - \tfrac{1}{4}(\mathbf{B}^\dagger\times\mathbf{B})^2 \\ & - (\mathbf{J}\cdot\mathbf{B}^\dagger)(\mathbf{J}\cdot\mathbf{B}), \end{aligned} \tag{7.19}$$

where the indices 2 and 4 are used to indicate their degrees in the generators of sp(4).

We now need to determine the value of these Casimir operators in terms of the ω, s that characterize, through (3.7) and (3.8), the irrep of sp(4). For this purpose we go back to the quantum picture and apply these operators to the lowest weight state $|\omega s\rangle$ of (3.5). As $B_i|\omega s\rangle = 0$ we have only to consider the action of $\mathcal{N}^2, J^2$ on this state and, going back to the classical limit where $n/2$ or 1 can be disregarded as compared to s or ω, we get

$$G_2 = \omega^2 + s^2, \tag{7.20a}$$

$$G_4 = \omega^2 s^2, \tag{7.20b}$$

which are the equations that for sp(4) correspond to (7.7) for sp(2).

As the value of s^2 is also associated with the observable S^2 in the enveloping algebra of $w(3) \oplus$ su(2), we can then write two algebraic equations for the three unknowns f, g, h in the form

$$G_2 - \omega^2 - S^2 = 0, \tag{7.21a}$$

$$G_4 - \omega^2 S^2 = 0, \tag{7.21b}$$

where in G_2, G_4 we replace $\mathbf{J} = \mathbf{L} + \mathbf{S}$, $\mathcal{N} = N + \omega$ and $\mathbf{B}^\dagger$, $\mathbf{B}$ by (7.14) and (7.17). The coefficients of powers of f, g, h in these equations are necessarily invariant of the u(2) Lie subalgebra of sp(4) and in fact only the following six invariants appear:

$$\begin{aligned} & N, \quad L^2, \quad \mathbf{L}\cdot\mathbf{S}, \quad S^2, \\ & A \equiv (\mathbf{b}^\dagger\cdot\mathbf{S})(\mathbf{b}\cdot\mathbf{S}), \\ & D \equiv [(\mathbf{b}^\dagger\cdot\mathbf{b}^\dagger)(\mathbf{b}\cdot\mathbf{S})^2 + (\mathbf{b}\cdot\mathbf{b})(\mathbf{b}^\dagger\cdot\mathbf{S})^2]. \end{aligned} \tag{7.22}$$

The two algebraic equations given below are then those corresponding to (7.21a) and (7.21b), i.e.,

$$\begin{aligned} & f^2N - 2fgL^2 - 2fh(\mathbf{L}\cdot\mathbf{S}) + g^2NL^2 + 2ghN(\mathbf{L}\cdot\mathbf{S}) \\ & \quad + h^2NS^2 - h^2A - (N^2 + 2\omega N \\ & \quad + L^2 + 2\mathbf{L}\cdot\mathbf{S}) = 0, \end{aligned} \tag{7.23a}$$

$$\begin{aligned} & \tfrac{1}{4}[f^2N - 2fgL^2 - 2fh(\mathbf{L}\cdot\mathbf{S}) + g^2NL^2 + 2ghN(\mathbf{L}\cdot\mathbf{S}) \\ & \quad + h^2(NS^2 - A)]^2 \\ & \quad - \tfrac{1}{4}\{(N^2 - L^2)[f^2 - g^2L^2 - 2gh\,\mathbf{L}\cdot\mathbf{S} - h^2S^2]^2 \\ & \quad + h^4A^2 + h^2[f^2 - g^2L^2 - 2gh\,\mathbf{L}\cdot\mathbf{S} - h^2S^2]D\} \\ & \quad - (N+\omega)\{2fhA + gh(D - 2NA) + [h^2(\mathbf{L}\cdot\mathbf{S}) \\ & \quad - 2fhN + 2ghL^2](S^2 + \mathbf{L}\cdot\mathbf{S}) \\ & \quad + (f^2 - 2fgN + g^2L^2)[L^2 + \mathbf{L}\cdot\mathbf{S}]\} \\ & \quad - \{[f + (h-g)N]^2A \\ & \quad + (h-g)^2(N^2 - L^2)A - [f + (h-g)N](h-g)D\} \end{aligned}$$

$$+\{S^2(N^2+2N\omega)+(N^2+2N\omega+\omega^2)\times[L^2+2(\mathbf{L}\cdot\mathbf{S})]\}=0. \tag{7.23b}$$

These equations play for sp(4) the role that (7.8) played for sp(2) and we see how much more complicated is the present problem. Furthermore they already show us that it will not be possible to obtain explicit, analytic, and closed expressions for f, g, and h as functions of the six u(2) invariants in (7.22). This follows from the fact that the quadratic equation (7.23a) allows us, for example, to express h in terms of f and g with one square root. Substituting in the fourth-order equation (7.23b) and squaring to eliminate the square root we get an eighth-degree algebraic equation in f and g which cannot be solved analytically to give, for example, g as a function of f.

But even if h and g could be given as explicit closed functions of f, we would still have the problem of determining the latter. We no longer have at our disposal Casimir operators, so we would have to use Poisson brackets of the generators of sp(4). Rather than employ them directly it is more convenient to consider scalar functions of them such as

$$(\mathbf{J}\cdot\mathbf{B}^\dagger), \tag{7.24a}$$

$$(\mathbf{J}\cdot\mathbf{B}), \tag{7.24b}$$

whose Poisson bracket is, from (3.2) and (7.1), given by

$$\{(\mathbf{J}\cdot\mathbf{B}^\dagger),(\mathbf{J}\cdot\mathbf{B})\}=-[\mathbf{J}\cdot(\mathbf{B}^\dagger\times\mathbf{B})]+2iJ^2\mathscr{N}. \tag{7.25}$$

From (7.11), (7.14), and (7.17) we obtain

$$(\mathbf{J}\cdot\mathbf{B}^\dagger)=f(\mathbf{b}^\dagger\cdot\mathbf{S})+i(h-g)[(\mathbf{L}\times\mathbf{S})\cdot\mathbf{b}^\dagger], \tag{7.26a}$$

$$(\mathbf{J}\cdot\mathbf{B})=f(\mathbf{b}\cdot\mathbf{S})-i(h-g)[(\mathbf{L}\times\mathbf{S})\cdot\mathbf{b}], \tag{7.26b}$$

$$\begin{aligned}-i[\mathbf{J}\cdot(\mathbf{B}^\dagger\times\mathbf{B})] &= [f^2-2fgN+g^2L^2+2gh(\mathbf{L}\cdot\mathbf{S})]\\ &\quad\times L^2+[f^2-2fhN+h^2(\mathbf{L}\cdot\mathbf{S})-2fgN+g^2L^2]\\ &\quad\times(\mathbf{L}\cdot\mathbf{S})+gh(D-2NA)+[-2fhN+2ghL^2\\ &\quad+h^2(\mathbf{L}\cdot\mathbf{S})]S^2+2fhA,\end{aligned} \tag{7.26c}$$

and thus, from (7.12a), (7.12b), and (7.26), we see that (7.25) gives us a partial differential equation involving f, g, and h, which together with the equations (7.23a) and (7.23b) could determine these three unknown functions. It is clear, though, from the above discussion, that this will not give us f, g, and h as explicit, analytic, and closed functions of the six u(2) invariants of (7.22), thus corroborating the quantum mechanical discussion of Sec. VI.

VIII. CONCLUSION

The discussion carried out in this paper on the boson realization of sp(4) covers two main aspects. In the first one we illustrate in a simple fashion how to obtain the Dyson boson realization and then we pass it to Holstein–Primakoff form through a similarity transformation with an operator K, getting the equation (5.7) that K^2 satisfies. In Appendix A we show that for definite irreps of Sp(4) the matrix representation of K^2 can be found in an explicit, analytic, and closed form.

The second aspect, discussed in Sec. VI, shows through Eqs. (6.6) and (6.12) that an explicit, analytic, and closed form for the boson realization of sp(4) is possible if the matrix expression of K is known. The problem, though, is that to get the matrix representation of the operator K from that of K^2 implies, in general, the solution of algebraic equation of high degree which cannot be done analytically. Thus the boson realization of sp(4) equivalent to the simple and well-known expression for sp(2) rederived in (2.24), is only possible in some special cases when K can be obtained analytically as illustrated explicitly for $s=0,\frac{1}{2}$ in Sec. VI.

A discussion for the classical limit, in which the operator K does not appear, illustrates again through the high-degree algebraic equations associated with the Casimir operators, that a boson realization in an explicit, analytic, and closed form is, in general, not feasible.

The analysis carried out for sp(4) is clearly susceptible to generalization[19–21] to sp($2d$), so the conclusions indicated in the previous paragraphs are also applicable to the more general case. We note, though, that while the matrix representation K cannot be obtained from K^2 analytically, there is no problem for getting it numerically and thus the procedures developed in previous references[19–21] are applicable to the determination of the matrix elements of the generators of the Lie algebra sp($2d$) in a basis associated with an irrep of the chain of groups Sp($2d$)$\supset$U(d). Because of its simplicity we illustrate this well-known point in Appendix B for the cases $d=1$ and 2.

ACKNOWLEDGMENTS

Two of the authors (O. Castaños and M. Moshinsky) would like to thank Consejo Nacional de Ciencia y Tecnología (México) for a grant that allowed their travels to Brussels and Kingston where part of this work was done. They would like also to thank Professor M. Demeur, Professor B. Castel, and Professor J. D. Rowe for the hospitality extended to them in Brussels, Kingston, and Toronto.

APPENDIX A: MATRIX REPRESENTATION OF K^2

We start by considering in the recursion relation (5.15) the notation

$$l=j+\tau, \tag{A1}$$

in which τ goes from $-s$ to s in steps of 1 but subject to the restriction (3.15). We then can write

$$\mathscr{M}_{l'l}(\nu,j,s)\equiv M_{\tau'\tau}(\nu,j,s), \tag{A2}$$

and express (5.15) as

$$\sum_{\bar\tau}R_{\tau'\bar\tau}(\nu,j,j')M_{\bar\tau\tau}(\nu,j,s)=\sum_{\bar\tau'}M_{\tau'\bar\tau'}(\nu-1,j',s)T_{\bar\tau'\tau}(\nu,j,j'), \tag{A3}$$

where

$$R_{\tau'\bar\tau}(\nu,j,j')=\langle\nu-1[j'+\tau',s]j'\|b\|\nu[j+\bar\tau,s]j\rangle, \tag{A4a}$$

$$\begin{aligned}T_{\bar\tau'\tau}(\nu,j,j') &= \{\nu+2\omega+n-2+\tfrac{1}{2}[(j'+\bar\tau')(j'+\bar\tau'+1)\\ &\quad-(j+\tau)(j+\tau+1)]\end{aligned}$$

$$-j'(j'+1)+j(j+1)\}R_{\bar{\tau}'\tau}(\nu,j,j'), \tag{A4b}$$

where the reduced matrix elements ($\|b\|$) are given by (5.16)–(5.18).

The evaluation of the matrices **R** and **T** for the three cases $j'=j\pm 1$ and $j'=j$ is immediate so we have the following.

(a) For $j'=j\pm 1$,

$$R_{\tau'\tau}(\nu,j,j+\lambda)=a_{\tau'}^{\lambda}(\nu,j)\delta_{\tau'\tau}+b_{\tau'}^{\lambda}(\nu,j)\delta_{\tau',\tau-2\lambda}, \tag{A5a}$$

$$T_{\tau',\tau}(\nu,j,j+\lambda)=u_{\tau'}^{\lambda}(\nu,j)\delta_{\tau'\tau}+v_{\tau'}^{\lambda}(\nu,j)\delta_{\tau',\tau-2\lambda}, \tag{A5b}$$

where the coefficients $a_{\tau'}^{\lambda}$, $b_{\tau'}^{\lambda}$, $u_{\tau'}^{\lambda}$, and $v_{\tau'}^{\lambda}$, with $\lambda=\pm 1$, are given by

$$a_{\tau'}^{\lambda}(\nu,j) = -\left[\frac{(2j+\tau'+s+\lambda+2)(2j+\tau'+s+\lambda+1)(2j+\tau'-s+\lambda+1)(2j+\tau'-s+\lambda)(\nu-\lambda\tau'-\lambda j-(\lambda-1)/2)}{4(2j+2\tau'+2\lambda+1)(2j+2\tau'+1)(2j+2\lambda+1)(j+(\lambda+1)/2)}\right]^{1/2}, \tag{A5c}$$

$$b_{\tau'}^{\lambda}(\nu,j)=-\left[\frac{(s+\lambda\tau'+2)(s+\lambda\tau'+1)(s-\lambda\tau')(s-\lambda\tau'-1)(\nu+\lambda j+\lambda\tau'+(\lambda+5)/2)}{4(2j+2\tau'+4\lambda+1)(2j+2\tau'+2\lambda+1)(2j+2\lambda+1)(j+(\lambda+1)/2)}\right]^{1/2}, \tag{A5d}$$

$$u_{\tau'}^{\lambda}(\nu,j)=[\nu-\lambda j+\lambda\tau'+2\omega+n-(\lambda+5)/2]a_{\tau'}^{\lambda}(\nu,j), \tag{A5e}$$

and

$$v_{\tau'}^{\lambda}(\nu,j)=[\nu-3\lambda j-\lambda\tau'+2\omega+n-\tfrac{1}{2}(\lambda+3)]b_{\tau'}^{\lambda}(\nu,j). \tag{A5f}$$

(b) For $j'=j$,

$$R_{\tau'\tau}(\nu,j,j)=a_{\tau'}^{0}(\nu,j)\delta_{\tau',\tau-1}+b_{\tau'}^{0}(\nu,j)\delta_{\tau',\tau+1}, \tag{A6a}$$

$$T_{\tau'\tau}(\nu,j,j)=u_{\tau'}^{0}(\nu,j)\delta_{\tau',\tau-1}+v_{\tau'}^{0}(\nu,j)\delta_{\tau',\tau+1}, \tag{A6b}$$

where the coefficients $a_{\tau'}^{0}$, $b_{\tau'}^{0}$, $u_{\tau'}^{0}$, and $v_{\tau'}^{0}$ are given by

$$a_{\tau'}^{0}(\nu,j)=-\left[\frac{(2j+\tau'+s+2)(2j+\tau'-s+1)(s+\tau'+1)(s-\tau')(\nu+j+\tau'+2)}{4(2j+2\tau'+3)(2j+2\tau'+1)j(j+1)}\right]^{1/2}, \tag{A6c}$$

$$b_{\tau'}^{0}(\nu,j)=\left[\frac{(2j+\tau'+s+1)(2j+\tau'-s)(s+\tau')(s-\tau'+1)(\nu-j-\tau'+1)}{4(2j+2\tau'-1)(2j+2\tau'+1)(j+1)j}\right]^{1/2}, \tag{A6d}$$

$$u_{\tau'}^{0}(\nu,j)=(\nu-j-\tau'+2\omega+n-3)a_{\tau'}^{0}(\nu,j), \tag{A6e}$$

and

$$v_{\tau'}^{0}(\nu,j)=(\nu+j+\tau'+2\omega+n-2)b_{\tau'}^{0}(\nu,j). \tag{A6f}$$

Substituting the matrices **R** and **T** in Eq. (A3) we get three sets of recursion relations for **M**. These are

$$a_{\tau'}^{+}(\nu,j)M_{\tau'\tau}(\nu,j,s)+b_{\tau'}^{+}(\nu,j)M_{\tau'+2,\tau}(\nu,j,s)=M_{\tau'\tau}(\nu-1,j+1,s)u_{\tau}^{+}(\nu,j)+M_{\tau',\tau-2}(\nu-1,j+1,s)v_{\tau-2}^{+}(\nu,j), \tag{A7a}$$

$$a_{\tau'}^{0}(\nu,j)M_{\tau'+1,\tau}(\nu,j,s)+b_{\tau'}^{0}(\nu,j)M_{\tau'-1,\tau}(\nu,j,s)=M_{\tau',\tau-1}(\nu-1,j,s)u_{\tau-1}^{0}(\nu,j)+M_{\tau',\tau+1}(\nu-1,j,s)v_{\tau+1}^{0}(\nu,j), \tag{A7b}$$

and

$$a_{\tau'}^{-}(\nu,j)M_{\tau'\tau}(\nu,j,s)+b_{\tau'}^{-}(\nu,j)M_{\tau'-2,\tau}(\nu,j,s)=M_{\tau'\tau}(\nu-1,j-1,s)u_{\tau}^{-}(\nu,j)+M_{\tau',\tau+2}(\nu-1,j-1,s)v_{\tau+2}^{-}(\nu,j), \tag{A7c}$$

where we define $a_{\tau'}^{\pm}\equiv a_{\tau'}^{\pm 1}$, $b_{\tau'}^{\pm}\equiv b_{\tau'}^{\pm 1}$, $u_{\tau'}^{\pm}\equiv u_{\tau'}^{\pm 1}$, and $v_{\tau'}^{\pm}\equiv v_{\tau'}^{\pm 1}$.

First of all, we illustrate in Fig. 1 what elements of the matrix **M** are connected by each set of recursion relations. We use the symbols $+$, 0, and $-$ to represent the elements connected by the sets (A7a), (A7b), and (A7c), respectively.

To solve the recursion relations (A7) we have to distinguish two different cases: one of them when the spin s of the intrinsic state is an integer and the other when s is a half-integer. This is due to the fact that the matrix $\|M_{\tau'\tau}(\nu,j,s)\|$ has elements different from zero only when $\tau'-\tau$ is an even number.

1. The case of integer spin

We shall begin by discussing the procedure to solve the recursion relations for the case when s is an integer.

We first consider the matrix elements $M_{\tau'\tau}$ when $(\tau',\tau)=(s,-s)$ in Eq. (A7a) and $(\tau',\tau)=(-s,s)$ in Eq. (A7c), i.e.,

$$M_{s,-s}(\nu,j,s)=[u_{-s}^{+}(\nu,j)/a_{s}^{+}(\nu,j)]M_{s,-s}(\nu-1,j+1,s), \tag{A8}$$

$$M_{-s,s}(\nu,j,s)=[u_{s}^{-}(\nu,j)/a_{-s}^{-}(\nu,j)]M_{-s,s}(\nu-1,j-1,s), \tag{A9}$$

where $a_{\tau'}^{\lambda}$, u_{τ}^{λ} are given by Eqs. (A5c) and (A5e). It is easily proved that the matrix **M** is symmetric so that we can solve the homogeneous recursion relations (A8) and (A9) if we define $x=\nu+j$, $y=\nu-j$, and $W(x,y)\equiv M_{s,-s}(\nu,j,s)$. The solution is given by

$$M_{s,-s}(\nu,j,s) = \left\{ \frac{(2j-2s+1)!!(2j+2s)!!(\nu-j+s)!!(4s-1)!!(\nu+j+s+1)!!}{(2j+1)(2s)!(2j-2s)!!(2j+2s-1)!!(\nu-j-s)!!(4s+1)!(\nu+j-s+1)!!} \right\}^{1/2}$$
$$\times \frac{(2s+1)!!(\nu-j-s+2\omega+n-3)!!(\nu+j-s+2\omega+n-2)!!}{(2\omega+n-3)!!(2s+2\omega+n-2)!!} M_{s,-s}(2s,s,s). \tag{A10}$$

We continue by establishing the recursion relations (A7a) and (A7c) for the matrix element $M_{s-1,-s+1}$, i.e.,

$$M_{s-1,-s+1}(\nu,j,s) = [u^{+}_{-s+1}(\nu,j)/a^{+}_{s-1}(\nu,j)]M_{s-1,-s+1}(\nu-1,j+1,s), \tag{A11}$$

$$M_{s-1,-s+1}(\nu,j,s) = [u^{-}_{s-1}(\nu,j)/a^{-}_{-s+1}(\nu,j)]M_{s-1,-s+1}(\nu-1,j-1,s), \tag{A12}$$

where a^{λ}_{τ}, and u^{λ}_{τ} are given by Eqs. (A5c) and (A5e). These homogeneous recursion relations can be solved in a form similar to that applied to (A8) and (A9), so that we get

$$M_{s-1,-s+1}(\nu,j,s)$$
$$= \left[\frac{(2s+1)(2s+2)!(\nu-j+s-1)!!(2j-2s+3)!!(2j+2s)!!(2j-1)!!(2j-2)!!(\nu+j+s)!!}{3(4s-1)(4s)!!(2j-2s)!!(2j+2s-3)!!(2j+1)!!(2j+2)!!(\nu-j-s+1)!!(\nu+j-s+2)!!} \right]^{1/2}$$
$$\times \frac{(\nu-j-s+2\omega+n-2)!!(\nu+j-s+2\omega+n-1)!!}{(2s-2)!!(2\omega+n-3)!!(2s+2\omega+n-2)!!} M_{s-1,-s+1}(2s-1,s,s). \tag{A13}$$

Until now we have found the matrix elements $M_{s,-s}$ and $M_{s-1,-s+1}$, which are denoted in Fig. 2 by 1 and 2 to stress that they were first to be evaluated. Before we can solve the inhomogeneous recursion relations associated to $M_{s-2,-s+2}(\nu,j,s)$, we need to know $M_{s,-s+2}(\nu,j,s)$ and $M_{s-2,-s}(\nu,j,s)$, which we designate by 3 and 3′ to indicate that they are the next to be filled. These elements are obtained from Eq. (A7b) with the values $(\tau',\tau) = (s,-s+1)$ and $(\tau',\tau) = (s-1,-s)$, i.e.,

$$M_{s,-s+2}(\nu,j,s) = [b^{0}_{s}(\nu+1,j)/v^{0}_{-s+2}(\nu+1,j)]M_{s-1,-s+1}(\nu+1,j,s) - [u^{0}_{-s}(\nu+1,j)/v^{0}_{-s+2}(\nu+1,j)]M_{s,-s}(\nu,j,s) \tag{A14}$$

and

$$M_{s-2,-s}(\nu,j,s) = [v^{0}_{-s+1}(\nu,j)/b^{0}_{s-1}(\nu,j)]M_{s-1,-s+1}(\nu-1,j,s) - [a^{0}_{s-1}(\nu,j)/b^{0}_{s-1}(\nu,j)]M_{s,-s}(\nu,j,s), \tag{A15}$$

where the terms b^{0}_{τ}, v^{0}_{τ}, u^{0}_{-s}, and a^{0}_{s-1} are given in (A6c)–(A6f), $M_{s-1,-s+1}$ is given in Eq. (A13), and $M_{s,-s}$ in Eq. (A10).

Now, the recursion relations for $M_{s-2,-s+2}(\nu,j,s)$ are the following:

$$M_{s-2,-s+2}(\nu,j,s) - \frac{u^{+}_{-s+2}(\nu,j)}{a^{+}_{s-2}(\nu,j)} M_{s-2,-s+2}(\nu-1,j+1,s)$$
$$= \frac{v^{+}_{-s}(\nu,j)}{a^{+}_{s-2}(\nu,j)} M_{s-2,-s}(\nu-1,j+1,s) - \frac{b^{+}_{s-2}(\nu,j)}{a^{+}_{s-2}(\nu,j)} M_{s,-s+2}(\nu,j,s) \tag{A16a}$$

and

$$M_{s-2,-s+2}(\nu,j,s) - \frac{u^{-}_{s-2}(\nu,j)}{a^{-}_{-s+2}(\nu,j)} M_{s-2,-s+2}(\nu-1,j-1,s)$$
$$= \frac{v^{-}_{s}(\nu,j)}{a^{-}_{-s+2}(\nu,j)} M_{s,-s+2}(\nu-1,j-1,s) - \frac{b^{-}_{-s+2}(\nu,j)}{a^{-}_{-s+2}(\nu,j)} M_{s-2,-s}(\nu,j,s). \tag{A16b}$$

We see that they can be written symbolically in the form

$$W(x,y) = f_1(x,y)W(x,y-2) + g_1(x,y), \tag{A17a}$$

$$W(x,y) = f_2(x,y)W(x-2,y) + g_2(x,y), \tag{A17b}$$

where we define $x = \nu + j$, $y = \nu - j$, and $W(x,y) \equiv M_{s-2,-s+2}(\nu,j,s)$. These two relations are of the type

$$Z_{n+1} = a_n Z_n + b_n, \tag{A18}$$

whose solution is given by

$$Z_{n+1} = \prod_{k=0}^{n} a_{n-k} Z_0 + \sum_{t=0}^{n-1} \prod_{k=0}^{n-1-t} a_{n-k} b_t + b_n. \tag{A19}$$

Then we can obtain the element $M_{s-2,-s+2}(\nu,j,s)$, which we denote by 4 in Fig. 2. To continue we would like to get $M_{s-3,-s+3}(\nu,j,s)$, for this we need to know the matrix elements denoted by 5 and 5′ in Fig. 2. The latter can be obtained from Eq. (A7b) in terms of matrix elements already known. Once we have them, we can deduce directly the elements $M_{s,-s+4}$ and $M_{s-4,-s}$ indicated by 6 and 6′ as well as $M_{s-3,-s+3}$, denoted by 7, by solving recursion relations of the same type as (A17).

The cases we have discussed so far suggest a general algorithm to obtain the remaining elements of the matrix $\|M_{\tau'\tau}\|$, namely the elements outside the secondary diagonal are obtained successively from Eq. (A7b) in terms of three other elements previously known, and the elements in the secondary diagonal are obtained from inhomogeneous relations of the same type as

(A17), in which enter some of the elements outside the secondary diagonal that were determined at a previous stage.

2. The case of half-integer spin

When s is a half-integer we first obtain the matrix element $M_{s,-s+1}(\nu,j,s)$, which appears in Fig. 3 denoted by 1. The recursion relations for it have the form

$$M_{s,-s+1}(\nu,j,s)=[u^{+}_{-s+1}(\nu,j)/a^{+}_{s}(\nu,j)]M_{s,-s+1}(\nu-1,j+1,s), \tag{A20}$$

$$M_{s,-s+1}(\nu,j,s)=[u^{-}_{s}(\nu,j)/a^{-}_{-s+1}(\nu,j)]M_{s,-s+1}(\nu-1,j-1,s), \tag{A21}$$

where a^{λ}_{τ} and u^{λ}_{τ} are given by Eqs. (A5c) and (A5e). Their solution is given by

$$\begin{aligned}M_{s,-s+1}(\nu,j,s)&\\ &=\left[\frac{(2j-2s+3)!!!(2j+2s)!!!(\nu-j+s-1)!!!(\nu+j+s+1)!!!(2j)!!!(2j-1)!!!(2s+1)(2s+2)(2s+2)!!}{3(2j+2)!!!(2j-2s)!!!(2s-1)!!!(2j+2s-1)!!!(4s)!!!(2j+1)!!!(4s+1)(\nu+j-s+2)!!!(\nu-j-s)!!}\right]^{1/2}\\ &\times\frac{(\nu+j-s+2\omega+n-2)!!!(\nu-j-s+2\omega+n-2)!!}{(2s+2\omega+n-2)!!!(2\omega+n-2)!!}M_{s,-s+1}(2s,s,s).\end{aligned} \tag{A22}$$

The matrix element $M_{s-1,-s}(\nu,j,s)$, denoted by 2 in Fig. 3, is obtained solving recursion relations similar to those of Eqs. (A20) and (A21). Thus we have

$$\begin{aligned}M_{s-1,-s}(\nu,j,s)&\\ &=\left[\frac{(2j-2s+1)!!!(2j+2s)!!!(\nu-j+s)!!!(\nu+j+s)!!!(2s+1)(2s)(2s)!!}{(\nu-j-s+1)!!!(2s-1)!!!(4s-1)(4s)!!!(2j+1)(2j)(2j-2s)!!!(2j+2s-3)!!!(\nu+j-s+1)!!}\right]^{1/2}\\ &\times\frac{(\nu-j-s+2\omega+n-3)!!}{(2\omega+n-4)!!}\,\frac{(\nu+j-s+2\omega+n-1)!!}{(2s+2\omega+n-2)!!}M_{s-1,-s}(2s-1,s,s).\end{aligned} \tag{A23}$$

Next we can extablish, with Eqs. (A7a) and (A7c), the recursion relations for the matrix element $M_{s-1,-s+2}(\nu,j,s)$, which is denoted by 3 in Fig. 3, i.e.,

$$M_{s-1,-s+2}(\nu,j,s)-[u^{+}_{-s+2}(\nu,j)/a^{+}_{s-1}(\nu,j)]M_{s-1,-s+2}(\nu-1,j+1,s)=[v^{+}_{-s}(\nu,j)/a^{+}_{s-1}(\nu,j)]M_{s-1,-s}(\nu-1,j+1,s), \tag{A24}$$

$$M_{s-1,-s+2}(\nu,j,s)-[u^{-}_{s-1}(\nu,j)/a^{-}_{-s+2}(\nu,j)]M_{s-1,-s+2}(\nu-1,j-1,s)=-[b^{-}_{-s+2}(\nu,j)/a^{-}_{-s+2}(\nu,j)]M_{s-1,-s}(\nu,j,s). \tag{A25}$$

These are of the same form as those in (A16). Thus, their solution has the form (A19). Using Eq. (A7b), we can evaluate directly the matrix elements denoted in Fig. 3 by 4 ($M_{s-2,-s+1}$), 4′ ($M_{s,-s+3}$), and 5 ($M_{s-3,-s}$) in terms of matrix elements already known. Once we have 4 and 4′, by means of recursion relations of type (A17), we can deduce 6 ($M_{s-2,-s+3}$).

The cases we have discussed so far suggest a general algorithm to obtain the remaining elements of the matrix $\|M_{\tau'\tau}\|$ namely the elements $M_{\tau,-\tau+1}$ are obtained from inhomogeneous relations of the same type as (A17), in which some of the previous matrix elements enter; the elements that differ from $M_{\tau,-t+1}$ are obtained successively from Eq. (A6b) in terms of other three elements previously known.

APPENDIX B: MATRIX ELEMENTS OF THE GENERATORS OF SP(4)

Again, as in Sec. II and the beginning of Sec. VII, we start by discussing first the case of sp(2). The normalized[28] states that are the basis for an irrep $\omega+n/2$ are, from (2.4), given by

$$|\nu\omega]=\left[\frac{(2\omega+n-2)!!}{(2\nu)!!(2\nu+2\omega+n-2)!!}\right]^{1/2}|\nu,\omega\rangle$$

$$=\left[\frac{(2\omega+n-2)!!}{(2\nu)!!(2\nu+2\omega+n-2)!!}\right]^{1/2}(B^{\dagger})^{\nu}|\omega\rangle \tag{B1}$$

so that the matrix elements of $B^{\dagger}$ become

$$[\nu+1,\omega|\,B^{\dagger}\,|\nu,\omega]=[(2\nu+2)(2\nu+2\omega+n)]^{1/2}, \tag{B2}$$

while that of B can be obtained by Hermitian conjugation and the one of C is trivial.

On the other hand, if we take

$$B^{\dagger}=Kb^{\dagger}K^{-1} \tag{B3}$$

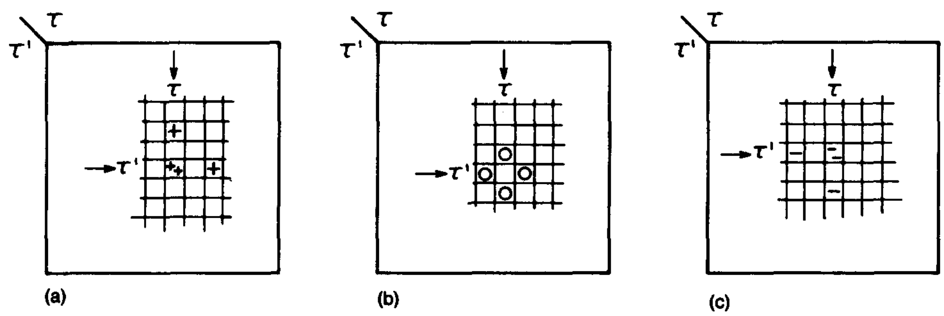


FIG. 1. We indicate symbolically the matrix elements $M_{\tau'\tau}(\nu,j,s)$ that are connected by the recursion relations in equations (a) (A7a), (b) (A7b), and (c) (A7c). We write double symbols when the matrix element appears two times in the recursion relation.

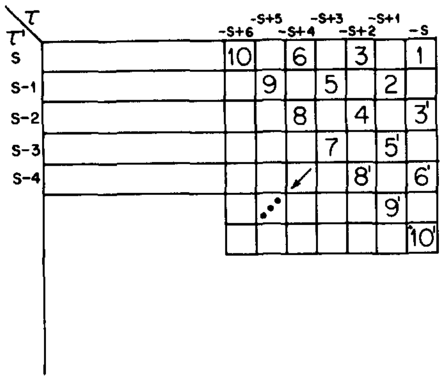


FIG. 2. We denote the matrix elements $M_{\tau'\tau}(\nu,j,s)$ for integer s by ordinal numbers 1,2,... to indicate the order in which they are evaluated.

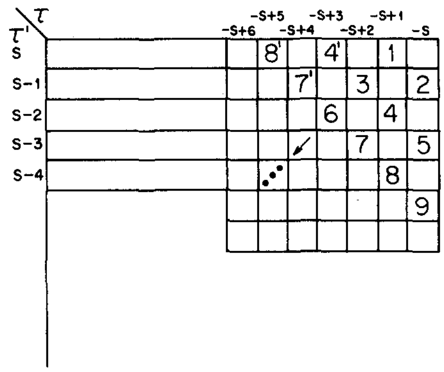


FIG. 3. We denote the matrix elements $M_{\tau'\tau}(\nu,j,s)$ for half-integer s by ordinal numbers 1,2,... to indicate the order in which they are evaluated.

and we use the basis $|\nu)$ of (2.18) as well as the expression of the matrix element of K given in (2.22), we immediately obtain

$$(\nu+1|\,B^\dagger\,|\nu) = [(2\nu+2)(2\nu+2\omega+n)]^{1/2}, \tag{B4}$$

which coincides with (B2) as we expect.

From the analysis for sp(2) we immediately conclude that in the case of sp(4) we get the matrix elements of $B^\dagger$ if we use

$$B_i^\dagger = K b_i^\dagger K^{-1} \tag{B5}$$

and the basis of states $|\nu[ls]jm)$ of (5.10), i.e.,

$$\begin{aligned}(\nu+1\,[l's]\,j'||\,B^\dagger\,||\nu[ls]\,j) &= \sum_{\bar l,\bar l'}\{K_{l'\bar l'}(\nu+1,j',s)(\nu+1\,[\bar l's]\\ &\quad\times j'||\,b^\dagger\,||\nu[\bar l s]\,j)\,K^{-1}_{\bar l l}(\nu,j,s)\},\end{aligned} \tag{B6}$$

where the reduced matrix element of $b^\dagger$ given in (5.17), (5.18), and $K_{l'l}(\nu,j,s)$ is discussed in Sec. VI. The matrix elements of B_i are obtained by Hermitian conjugation and those of J_i, $\mathcal{N}$ are trivial.

[1]T. Holstein and H. Primakoff, Phys. Rev. **58**, 1098 (1940).

[2]P. A. Mello and M. Moshinsky, J. Math. Phys. **16**, 2017 (1976).

[3]J. Deenen and C. Quesne, J. Math. Phys. **23**, 878 (1982).

[4]L. D. Mlodinow and N. Papanicolau, Ann. Phys. (N.Y.) **128**, 314 (1980); **131**, 1 (1981).

[5]G. Rosensteel and D. J. Rowe, Ann. Phys. (N.Y.) **96**, 1 (1976); **123**, 36 (1978); **126**, 198 (1980).

[6]O. L. Weaver, R. Y. Cusson, and L. C. Biedenharn, Ann. Phys. (N.Y.) **102**, 493 (1976); B. Buck, L. C. Biedenharn, and R. Y. Cusson, Nucl. Phys. A **317**, 205 (1979).

[7]G. F. Filippov and V. I. Ovcharenko, Yad. Fiz. **30**, 646 (1979); V. C. Vasilevsky, Ya. F. Smirnov, and G. F. Filippov, Yad. Fiz. **32**, 987 (1980); G. F. Filippov, V. I. Ovcharenko, and Ya. F. Smirnov, *Microscopic Theory of Collective Excitations in Nuclei* (Naukova Dumka, Kiev, 1981) (in Russian).

[8]V. Vanagas, "The Microscopic Nuclear Theory" in *Lecture Notes in Physics* (University of Toronto Press, Toronto, 1977); V. Vanagas, "The Microscopic Theory of Collective Motions in Nuclei," in *Group Theory and its Applications in Physics—1980*, edited by T. H. Seligman (AIP, New York, 1980).

[9]J. Carvalho, P. Park, D. J. Rowe, and G. Rosensteel, Phys. Lett. B **119**, 249 (1982), M. Vassanji and D. J. Rowe, Phys. Lett. B **115**, 77 (1982); P. Park, J. Carvalho, M. Vassanji, and D. J. Rowe, Nucl. Phys. A **414**, 93 (1984); G. Rosensteel, J. P. Draayer, and K. J. Weeks, Nucl. Phys. A **419**, 1(1984); J. P. Draayer, K. J. Weeks, and G. Rosensteel, Nucl. Phys. A **413**, 215 (1984).

[10]O. Castaños, E. Chacón, A. Frank, P. Hess, and M. Moshinsky, Phys. Rev. C **25**, 1611 (1982); J. Math. Phys. **23**, 2537 (1982).

[11]M. Moshinsky, J. Math. Phys. **25**, 1555 (1984).

[12]M. Moshinsky and C. Quesne, J. Math. Phys. **11**, 1631 (1970); **12**, 1772 (1971).

[13]J. Deenen and C. Quesne, J. Math. Phys. **23**, 2004 (1982), J. Phys. A: Math. Gen. **17**, L404 (1984).

[14]M. Moshinsky, *Group Theory and the Many Body Problem* (Gordon and Breach, New York, 1968).

[15]G. Rosensteel and D. J. Rowe, Phys. Rev. Lett. **47**, 223 (1981); D. J. Rowe and G. Rosensteel, Phys. Rev. C **25**, 36 (1982).

[16]O. Castaños and A. Frank, J. Math. Phys. **25**, 388 (1984).

[17]M. Moshinsky and T. H. Seligman, J. Math. Phys. **22**, 1526 (1981); M. Moshinsky, Physica A **114**, 322 (1982).

[18]J. Deenen and C. Quesne, J. Phys. A: Math. Gen. **16**, 2095 (1983); *Lecture Notes in Physics 180* (Springer, Berlin, 1983), p. 444.

[19]J. Deenen and C. Quesne, J. Math. Phys. **25**, 1638 (1984).

[20]M. Moshinsky, "Symmetry, constrained bosons and collectivity," in *Proceedings of the 1983 Suzhou Workshop on Nuclear Collective States*, Nucl. Phys. A **421**, 81 (1984); M. Moshinsky, E. Chacón, and O. Castaños, "Constrained bosons for collective states in open shell nuclei," in *Proceedings of the XIII Colloquium on Group Theoretical Methods in Physics*, University of Maryland, 1984 (World Scientific, Singapore, 1984).

[21]D. J. Rowe, G. Rosensteel, and R. Carr, J. Phys. A: Math. Gen. **17**, L399 (1984); D. J. Rowe, J. Math. Phys. **25**, 2662 (1984).

[22]G. Rosensteel, J. Math. Phys. **21**, 924 (1980); G. Rosensteel and D. J. Rowe, J. Math. Phys. **24**, 2461 (1983).

[23]M. Moshinsky, *The Harmonic Oscillator in Modern Physics: From Atoms to Quarks* (Gordon and Breach, New York, 1969).

[24]E. M. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957).

[25]J. Louck, K. B. Wolf, and M. Moshinsky, J. Math. Phys. **14**, 692 (1973).

[26]M. Moshinsky and C. Quesne, Ann. Phys. (N.Y.) **148**, 462 (1983).

[27]S. Goshen and H. J. Lipkin, "On the Applications of Groups Sp (4) or R (5) to Nuclear Structure," in *Spectroscopic and Group Theoretical Methods in Physics. Racah Memorial Volume*, edited by F. Bloch, S. Cohen, A. de Shalit, S. Sambursky and I. Talmi (North-Holland, Amsterdam, 1968), pp. 245–273.

[28]O. Castaños, E. Chacón, and M. Moshinsky, J. Math. Phys. **25**, 1211 (1984).

A set of commuting missing label operators for SO(5)⊃SO(3)

H. De Meyer,[a] G. Vanden Berghe, J. Van der Jeugt,[b] and P. De Wilde[c]
Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit Gent, Krijgslaan 281, gebouw S9, B-9000 Gent, Belgium

(Received 28 December 1984; accepted for publication 8 March 1985)

The two-missing label problem for basis vectors of an SO(5) irreducible representation reduced according to the principal SO(3) subalgebra is considered. A pair of commuting Hermitian operators which are scalars with respect to the SO(3) subalgebra are explicitly constructed within the SO(5) enveloping algebra. One label generating operator is of fourth order and the other of sixth order in the SO(5) basis elements.

I. INTRODUCTION

In the reduction of irreducible representations (IRT's) of a Lie algebra G into IR's of a subalgebra H, the latter usually does not provide enough labels to specify the basis states uniquely. By the introduction of extra labels to distinguish the states unambiguously, in general an orthogonalization procedure must be carried out. This difficulty is conceptually resolved by taking as bases the common eigenstates of a complete set of commuting Hermitian operators. Therefore, besides the Casimir invariants of the algebra and subalgebra one has to construct within the enveloping algebra of G an appropriate number of commuting missing label operators which are subalgebra scalars, i.e., which commute with all elements of H.

The simplest two-missing label problem is that of SU(4)⊃SO(4). A pair of independent commuting SO(4) scalars has been found by Moshinsky and Nagel.[1] Another independent set of commuting operators has also been constructed by Quesne[2] and Partensky and Maguin.[3] The commuting scalars for this problem are of third or fourth degree in the basis elements of SU(4). Recently, the two-missing label problem $G_2\supset SU(2)\times SU(2)$ has been solved by Hughes and Van der Jeugt.[4] From their work it becomes apparent that in order to construct missing label operators one should not necessarily try to express them in terms of a class of functionally independent subalgebra scalars. As we know from a theorem due to Peccia and Sharp[5] these are in number twice the number of missing labels.

In this paper we study the two-missing label problem which is associated to the chain SO(5)⊃SO(3), where SO(3) denotes the principal SO(3) subalgebra of SO(5). In a first step we have constructed all the linearly independent SO(3) scalars up to sixth order in the basis elements in the SO(5) enveloping algebra. From these we were able to find a pair of commuting operators which are both linear combinations of the SO(3) scalars, hence SO(3) scalars themselves. One of the operators turns out to be of fourth degree in the SO(3) tensor components, the other one of sixth degree.

It should be noted that the internal labeling of SO(5) states, which we obtain here, is clearly different from the one proposed by Bincer[6] on account of the reduction SO(5) ⊃SO(3)×SO(2), because the SO(3) subalgebra is not principal there. Also, one cannot make use of a procedure of Van der Jeugt[7] for obtaining a pair of commuting scalars for $G\supset [SU(2)]^n$.

Finally, if the SO(5) IR's are restricted to the symmetrical representations exclusively, the problem reduces into a one-missing label problem that has been exhaustively considered by the present authors elsewhere.[8]

II. THE SO(3) SCALARS IN SO(5)

A basis for the Lie algebra SO(5) may be chosen to consist of the generators $(l_0, l_{\pm 1})$ of the principal subalgebra SO(3) together with a seven-dimensional irreducible tensor representation q_μ $(\mu = -3, -2,..., +3)$ of SO(3). These satisfy the commutation relations

$$[l_0,l_{\pm 1}] = \pm l_{\pm 1}, \quad [l_{+1},l_{-1}] = -l_0, \tag{2.1}$$

$$\begin{aligned}&[l_{\pm 1},q_\mu] = \mp[(3\mp\mu)(4\pm\mu)/2]^{1/2}q_{\mu\pm 1},\\ &[l_0,q_\mu] = \mu q_\mu,\end{aligned} \tag{2.2}$$

together with the mutual commutation relations of the q_μ,

$$\begin{aligned}
&[q_0,q_{\pm 1}] = \pm 10^{-1/2}q_{\pm 1} \pm (6^{1/2}/10)l_{\pm 1},\\
&[q_0,q_{\pm 2}] = 10^{-1/2}q_{\pm 2},\\
&[q_0,q_{\pm 3}] = \mp 10^{-1/2}q_{\pm 3},\\
&[q_{\mp 1},q_{\pm 2}] = \mp(5^{-1/2}/2)l_{\pm 1},\\
&[q_{\mp 1},q_3] = \mp 5^{-1/2}q_{\pm 2},\\
&[q_{\mp 2},q_{\pm 3}] = \mp 5^{-1/2}q_{\pm 1} \pm (3^{1/2}/10)l_{\pm 1},\\
&[q_{\mp 1},q_{\mp 2}] = \mp 5^{-1/2}q_{\mp 3},\\
&[q_{-1},q_{+1}] = 10^{-1/2}q_0 + \tfrac{1}{10}l_0,\\
&[q_{-2},q_{+2}] = -10^{-1/2}q_0 - \tfrac{1}{5}l_0,\\
&[q_{-3},q_{+3}] = -10^{-1/2}q_0 + \tfrac{3}{10}l_0.
\end{aligned} \tag{2.3}$$

It should be noticed from (2.1) and (2.2) that the SO(3) subalgebra is realized by means of the components of a three-dimensional spherical tensor. Hence, we can immediately define SO(3) scalars by coupling the l and q tensors to rank zero tensors. We have developed a Pascal program for the construction of such scalars as homogeneous linear combinations of SO(5) generator strings. Thereby we have exploited the concept of recursivity which is implemented in the Pascal language. Let us describe an SO(3) scalar in the SO(5) enveloping algebra by (a,b), where a and b denote the degrees

[a] Senior Research Associate N.F.W.O., Belgium.
[b] Research Assistant N.F.W.O., Belgium.
[c] Research Assistant I.W.O.N.L., Belgium.

 0022-2488/85/092124-03$02.50

in the q and l generators, respectively. We define

$$
\begin{aligned}
&(0,2)=(ll)^0, \quad (2,0)=(qq)^0,\\
&(1,3)=(q((ll)^2l)^3)^0, \quad (2,2)=((qq)^2(ll)^2)^0,\\
&(3,1)=(((qq)^2q)^1l)^0,\\
&(2,4)=((qq)^4(((ll)^2l)^3l)^4)^0,\\
&(3,3)=(((qq)^2q)^3((ll)^2l)^3)^0,\\
&(4,0)=(((qq)^2q)^3q)^0,\\
&(4,2)=((((qq)^2q)^1q)^2(ll)^2)^0,\\
&(5,1)=(((((qq)^2q)^1q)^2q)^1l)^0,\\
&(6,0)=(((((qq)^2q)^1q)^2q)^3q)^0.
\end{aligned}
\tag{2.4}
$$

These are the elementary SO(3) scalars up to sixth degree in the SO(5) generators.[9]

In order to express our final result in rational form we have redefined the SO(5) Lie algebra (2.1)–(2.3) such that all the structure constants become integers, i.e., we introduce the generators

$$
\begin{aligned}
&l_{\pm} = \mp\sqrt{2}l_{\pm 1}, \quad q'_0 = \sqrt{10}q_0 \quad q'_{\pm 1} = \sqrt{30}q_{\pm 1},\\
&q'_{\pm 2} = \sqrt{300}q_{\pm 2}, \quad q'_{\pm 3} = \sqrt{50}q_{\pm 3}.
\end{aligned}
\tag{2.5}
$$

Reexpressing the SO(3) scalars (2.4) in terms of the new basis elements l_0, $l_{\pm}$, and q'_{μ} ($|\mu|\leqslant 3$) it suffices to separate an overall irrational factor such that the remaining linear combination of generator strings contains integer coefficients only. Hence, we define rescaled scalars by

$$
\begin{aligned}
&[0,2] = -2\sqrt{3}(0,2), \quad [2,0] = -300\sqrt{7}(2,0),\\
&[1,3] = -60\sqrt{7}(1,3), \quad [2,2] = -120\sqrt{70}(2,2),\\
&[3,1] = -12\,600\sqrt{2}(3,1),\\
&[2,4] = -50\,400\sqrt{55}(2,4),\\
&[3,3] = -100\,800\sqrt{5}(3,3),\\
&[4,0] = -50\,400\sqrt{5}(4,0),\\
&[4,2] = -126\,000\sqrt{42}(4,2),\\
&[5,1] = -2\,646\,000\sqrt{30}(5,1),\\
&[6,0] = -5\,292\,000\sqrt{3}(6,0).
\end{aligned}
\tag{2.6}
$$

These scalars denoted by $[a,b]$ are still homogeneous polynomials of degree a in the q'_{μ}'s and of degree b in the l_0, $l_{\pm}$ generators, respectively. All the coefficients are integers and the smallest coefficient equals in absolute value unity.

III. TWO COMMUTING SCALARS X_1 AND X_2

In order to construct besides the SO(3) and SO(5) Casimir invariants, two commuting SO(3) scalars, we need to consider the commutators of the scalars (2.6), of which we are assured that they possess only integer coefficients, too. Due to the Poincaré–Birkhoff–Witt theorem we can compare the the commutator results by bringing the generator strings into a predefined standard ordering of the generators. We have chosen here the left to right ordering

$$
q'_{+3}q'_{+2}q'_{+1}q'_0q'_{-1}q'_{-2}q'_{-3}l_+l_-l_0, \tag{3.1}
$$

which is rather convenient for calculational purposes.

A FORTRAN 77 program has been developed which replaces a nonstandard ordered polynomial form by the corresponding standard ordered polynomial together with all the extra lower-degree polynomial forms, themselves standard ordered, incurred by interchanging basis elements and making use of the commutation relations in terms of the structure constants. This program is a revised and optimalized version of a former FORTRAN IV program which we used in previous work and which has been used by Hughes and Van der Jeugt[4] in solving the $G_2 \supset SU(2)\times SU(2)$ state labeling problem.

In a first step we have established, in the l_0, $l_{\pm}$, q'_{μ} basis, the SO(3) Casimir L^2 and the SO(5) Casimirs I_2 and I_4 of second and fourth degree, respectively. We obtained

$$
\begin{aligned}
&L^2 = [0,2]/2,\\
&I_2 = ([2,0] + 15[0,2])/2,\\
&I_4 = ([4,0] - 4[3,1] - 6[2,2] + 2[0,2][2,0]\\
&\qquad + 24[1,3] - 24[0,2][0,2] - 384[0,2])/8.
\end{aligned}
\tag{3.2}
$$

Again we have in these forms factorized a rational number such that the lowest coefficient in absolute value equals 1. In the search for two additional SO(3) scalars we could readily demonstrate that at least one of them should be of degree higher than 4 in the basis elements. Hence, we conjectured that one scalar would be of fourth degree, the other of sixth degree, and therefore we considered first the commutators of fourth-order scalars with the sixth-order scalars constructed out of the set (2.6) and brought into standard order. We then searched for linear combinations of these commutators in which the highest or ninth-degree terms vanished, and then adjustments were made to eliminate the lower-degree terms. Proceeding this way, our initial hypothesis concerning the degree of two commutating scalars fortunately proved to be valid. Moreover, we found that the solution is certainly not unique. The simplest one contains as a fourth-order scalar X_1, a linear combination of two homogeneous scalars (2.6) only, namely

$$
X_1 = 3[2,2] + 4[1,3]. \tag{3.3}
$$

The second scalar X_2 of sixth degree that commutes with X_1 is then given by

$$
\begin{aligned}
X_2 = {}& [6,0] + 24[5,1] + 28[3,1]I_2 + 12[4,2] + \alpha[2,2]I_2\\
&- 864[3,3]/5 + 2172[3,1]L^2/5\\
&+ (4\alpha - 1296)[1,3]I_2/3\\
&+ 576[2,4]/35 + \beta[2,2]L^2\\
&+ (4\beta - 116208/7)[1,3]L^2/3,
\end{aligned}
\tag{3.4}
$$

where α and β are free parameters.

IV. CONCLUSIONS

Solutions of internal missing label problems generated by state reduction according to a canonical chain of algebra subalgebra inclusions usually lack physical relevance. The reason is that the canonical chains in general do not end up with an SO(3) or $SU(2)\times SU(2)$ subalgebra, which is maximal in the algebra considered.

In the present paper we have shown, following the ideas already expounded by Hughes and Van der Jeugt[4] in their solution of the $G_2 \supset SU(2) \times SU(2)$ missing label problem, that also the $SO(5) \supset SO(3)$ two-missing label problem can be completely solved.

ACKNOWLEDGMENTS

We would like to thank J. W. B. Hughes (Queen Mary College, London) for valuable discussions and encouragement and especially for suggesting ways to improve the FORTRAN programs.

We also acknowledge the support of the computing Centre (C.D.R.) of our University by allowing us to consume without any restriction the requested CPU time.

[1]M. Moshinsky and J. G. Nagel, Phys. Lett. **5**, 1973 (1963).
[2]C. Quesne, J. Math. Phys. **17**, 1452 (1976).
[3]A. Partensky and C. Maguin, J. Math. Phys. **19**, 511 (1978).
[4]J. W. B. Hughes and J. Van der Jeugt, J. Math. Phys. **26**, 894 (1985).
[5]A. Peccia and R. T. Sharp, J. Math. Phys. **17**, 1313 (1976).
[6]A. M. Bincer, J. Math. Phys. **24**, 1695 (1983).
[7]J. Van der Jeugt, J. Math. Phys. **25**, 1219 (1984).
[8]G. Vanden Berghe and H. De Meyer, J. Math. Phys. **21**, 1902 (1980); H. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 1906 (1980).
[9]R. Gaskell, A. Peccia, and R. T. Sharp, J. Math. Phys. **19**, 727 (1978).

Branching index sum rules for simple Lie algebras

Susumu Okubo
Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

(Received 11 December 1984; accepted for publication 24 May 1985)

Let L and L_0 be a simple Lie algebra and its sub-Lie algebra, respectively. Then, a given irreducible representation ω of L decomposes into a direct sum of irreducible components of L_0, which is called the branching rule. The general Dynkin indices introduced earlier satisfy many sum rules for the branching rule. These are found to be strong enough to uniquely determine the branching rule for many cases we have studied. The sum rules are especially useful for cases of exceptional Lie algebras.

I. INTRODUCTION AND SUMMARY OF MAIN RESULTS

Let G and L be a compact Lie group and its associated Lie algebra, respectively. Any irreducible representation ω of G (or L) restricted to its closed subgroup G_0 (or its sub-Lie algebra L_0) will be, in general, reducible and will be decomposed as a direct sum

$$\omega \supset \rho = \sum_j \oplus \rho_j \tag{1.1}$$

of irreducible representation components ρ_j's of G_0 (or L_0). In many problems of physics, it is very important to determine this decomposition, which will be called the branching rule (hereafter referred to as BR) for the reduction $G \downarrow G_0$ (or $L \downarrow L_0$). If G is one of the classical groups $U(k)$, $SU(k)$, $SO(k)$, and $Sp(2n)$, the use of the Schur function method[1,2] provides a general framework of obtaining the BR for many subgroups G_0. Indeed, the general BR's are already found in this manner for $U(k) \downarrow U(k-1)$ by Weyl,[3] for $SO(k) \downarrow SO(k-1)$ by Boerner,[4] for $Sp(2k) \downarrow Sp(2k-2)$ by Miller[5] and Hegerfeldt,[6] for $U(p+q) \downarrow U(p) \times U(q)$ and $U(pq) \downarrow U(p) \times U(q)$ by Whippman[7] and by Itzykson and Nauenberg,[8] for $SO(2k) \downarrow U(k)$, $SO(2k+1) \downarrow U(k)$, and $Sp(2k) \downarrow U(k)$ by King,[9] and for $SO(2p+2q) \downarrow SO(2p) \times SO(2q)$ and $SO(2p+2q+2) \downarrow SO(2p+1) \times SO(2q+1)$ by Black and Wybourne.[10] More exhaustive references as well as further elaborations on this method can be found in articles by King and El-Sharkaway[11] and by Black, King, and Wybourne.[12]

However, the method described above appears to become quite unmanageable and impractical, in general, when G is one of the exceptional Lie groups G_2, F_4, E_6, E_7, and E_8, or when G_0 contains an exceptional subgroup as its factor. For such a case, it is more convenient to deal with the reduction $L \downarrow L_0$ of Lie algebras rather than $G \downarrow G_0$ of Lie groups. Hereafter, we restrict ourselves for cases of L being a simple Lie algebra and of L_0 being reductive, i.e., a direct product of a semisimple Lie algebra and an Abelian Lie algebra. There are many methods[13–18] for finding the BR's for such a case. Unfortunately, hand calculations based upon these methods are still impractical for all but a few low-dimensional representations. For any simple Lie algebra L of rank less than or equal to 8, a large table for the BR's of L into its maximal sub-Lie algebras L_0 has been tabulated by McKay and Patera.[19] Also, the branching rule for $E_8 \downarrow D_8$ has been computed by Wybourne[20] and by Bélanger[21] for all irreducible representations of E_8 with dimensions less than 76 271 625.

The purpose of this paper is to demonstrate a fact that the notion of general indices introduced elsewhere[22] [hereafter referred to as (I)] is quite useful to test the validity of BR's. As we shall see shortly, there are many branching index sum rules to be satisfied, and these are essentially sufficient to uniquely determine the BR's for many cases in a mechanical way without investigating structures of L and L_0.

Let L be a simple Lie algebra, and let $\{t_\mu\}$ ($\mu = 1,2,\ldots$) be a basis of L with Lie multiplication table

$$[t_\mu, t_\nu] = C^\lambda_{\mu\nu} t_\lambda, \tag{1.2}$$

where $C^\lambda_{\mu\nu}$ is the structure constant of L and the standard summation convention on repeated indices is hereafter understood. It is known[23] that any simple Lie algebra L with rank r has precisely r independent Casimir invariants which we call fundamental. Let J_p be the pth-order fundamental Casimir invariant of L of form

$$J_p = g^{\mu_1\mu_2\cdots\mu_p} t_{\mu_1} t_{\mu_2} \cdots t_{\mu_p}, \tag{1.3}$$

where $g^{\mu_1\mu_2\cdots\mu_p}$ is completely symmetric in p indices $\mu_1,\mu_2,\ldots,\mu_p$, and satisfies orthogonality conditions[22] such as

$$g^{\mu\nu\alpha\beta} g_{\mu\nu} g_{\alpha\beta} = 0, \tag{1.4}$$

for the special case of $p = 4$. Here, $g_{\mu\nu}$ is the Killing form except for normalization, i.e.,

$$g_{\mu\nu} = C\,\mathrm{Tr}(\mathrm{ad}\, t_\mu\ \mathrm{ad}\, t_\nu), \tag{1.5}$$

for an unspecified nonzero constant C and we lower and raise Greek indices $\mu,\nu,\ldots$ by $g_{\mu\nu}$ and its inverse $g^{\mu\nu}$. The explicit form of J_4 can be found in Ref. 24. As we emphasized in (I), J_p is either identically zero, or unique apart from its overall normalization constant, except for the case of the Lie algebra D_p corresponding to the $SO(2p)$ group where we have two fundamental pth-order Casimir invariants J_p and $\hat{J}_p$. Let ω be a representation of L. The pth-order Dynkin index $D_p(\omega)$ is now defined by

$$D_p(\omega) = \mathrm{Tr}^{(\omega)}(J_p), \tag{1.6}$$

where the trace refers to the representation ω. Note that we have changed our notation from $D^{(p)}(\omega)$ in (I) to $D_p(\omega)$. If ω is irreducible, then we have

$$D_p(\omega) = d(\omega) J_p(\omega), \tag{1.7}$$

 0022-2488/85/092127-11$02.50

where $d(\omega)$ and $J_p(\omega)$ are the dimension and eigenvalue of J_p in the irreducible representation ω, respectively.

Now, consider the branching rule Eq. (1.1) for $L \downarrow L_0$. If L_0 is also simple, then it has been shown in (I) that we must have a branching sum rule

$$\xi_p D_p(\omega) = \sum_j D_p^{(0)}(\rho_j), \tag{1.8}$$

except for special case of $L = D_p$ (p being even and $p \geqslant 4$), corresponding to the SO($4l$) group ($p = 2l$ being even). Here, $D_p^{(0)}(\rho_j)$ is the pth-order index of the sub-Lie algebra L_0 in its irreducible component ρ_j, and ξ_p is a constant, which may depend upon p, L, and L_0 but *not* on ω. Therefore, once ξ_p is computed for a particular representation ω_0 of L from Eq. (1.8), then we can use Eq. (1.8) to test the BR for any irreducible representation ω. We note that Eq. (1.8) is valid also for $L = D_q$ if $p \neq q$. The Lie algebra D_p ($p \geqslant 4$) (p being even) possess two fundamental pth-order indices $D_p(\omega)$ and $\hat{D}_p(\omega)$, as we noted in (I) and in Ref. 24. For such a case, Eq. (1.8) must be replaced by

$$\xi_p D_p(\omega) + \hat{\xi}_p \hat{D}_p(\omega) = \sum_j D_p^{(0)}(\rho_j)$$
$$(L = D_p, \quad p = \text{even} \geqslant 4), \tag{1.9}$$

introducing two unknown constants ξ_p and $\hat{\xi}_p$. However, we will not consider such a case hereafter. We simply remark here that the sum rule Eq. (1.8) for the special case of $p = 2$ has been originally noted by Dynkin.[25]

Next, let $\overline{D}_4(\omega)$ be another fourth-order index defined by

$$\overline{D}_4(\omega) = \mathrm{Tr}^{(\omega)} I_4, \tag{1.10a}$$

$$I_4 = I_2[I_2 - \tfrac{1}{6} I_2(\omega_0)], \tag{1.10b}$$

where ω_0 refers hereafter to the adjoint representation of L. When ω is irreducible, then

$$\overline{D}_4(\omega) = D_2(\omega)\left[\frac{D_2(\omega)}{d(\omega)} - \frac{1}{6}\frac{D_2(\omega_0)}{d(\omega_0)}\right]. \tag{1.11}$$

Here $d(\omega_0)$ is the dimension of L. Now, we have also the following fourth-order sum rule

$$\eta_4 D_4(\omega) + \theta \overline{D}_4(\omega) = \sum_j \overline{D}_4^{(0)}(\rho_j), \tag{1.12}$$

for any simple Lie algebras other than $L = D_4$ corresponding to the SO(8) group. Here, η_4 is a new unknown constant but θ is related to ξ_2 by

$$\theta = \{d(\omega_0)[d_0(\rho_0) + 2]/d_0(\rho_0)[d(\omega_0) + 2]\}(\xi_2)^2 \tag{1.13}$$

in terms of ξ_2 appearing in Eq. (1.8) with $p = 2$, and ρ_0 is the adjoint representation of L_0 with dimension $d_0(\rho_0)$. Note that $\overline{D}_4^{(0)}(\rho)$ for the irreducible representation ρ of L_0 is defined similarly to Eq. (1.11) by

$$\overline{D}_4^{(0)}(\rho) = D_2^{(0)}(\rho)\left\{\frac{D_2^{(0)}(\rho)}{d_0(\rho)} - \frac{1}{6}\frac{D_2^{(0)}(\rho_0)}{d_0(\rho_0)}\right\}, \tag{1.14}$$

when we define $d_0(\rho)$ to be the dimension of ρ. We can also prove the validity of the following fifth-order mixed sum rule

$$\eta_5 D_5(\omega) + \xi_2\xi_3 \frac{d_0(\rho_0) + 6}{d_0(\rho_0)}\frac{d(\omega_0)}{d(\omega_0) + 6} \times D_3(\omega)\left[\frac{D_2(\omega)}{d(\omega)} - \frac{1}{4}\frac{D_2(\omega_0)}{d(\omega_0)}\right]$$
$$= \sum_j \left[\frac{D_2^{(0)}(\rho_j)}{d_0(\rho_j)} - \frac{1}{4}\frac{D_2^{(0)}(\rho_0)}{d_0(\rho_0)}\right] D_3^{(0)}(\rho_j). \tag{1.15}$$

Since the new constants η_4 and η_5 are independent of the generic irreducible representation ω, sum rules (1.12) and (1.15) can be used to check the BR, once the values of η_4 and η_5 are computed from a particular known BR. Although we can find a mixed six-order sum rule, we will discuss it later. These mixed index sum rules are analogs of similar sum rules for Kronecker products given in (I).

If L_0 is a direct product

$$L_0 = L_A \times L_B \tag{1.16}$$

or two simple Lie algebras L_A and L_B, we write the branching rule as

$$\omega \supset \rho = \sum_j \oplus \rho_j = \sum_j \oplus (\rho_a^{(A)} \otimes \rho_b^{(B)}), \tag{1.17}$$

where $\rho_a^{(A)}$ and $\rho_b^{(B)}$ are irreducible representations of L_A and L_B, respectively, with $\rho_j = \rho_a^{(A)} \otimes \rho_b^{(B)}$. First, considering $L \downarrow L_A$ and $L \downarrow L_B$, we evidently have

$$d(\omega) = \sum_j d^{(A)}(\rho_a) d^{(B)}(\rho_b), \tag{1.18a}$$

$$\xi_p^{(A)} D_p(\omega) = \sum_j d^{(B)}(\rho_b) D_p^{(A)}(\rho_a), \tag{1.18b}$$

$$\xi_p^{(B)} D_p(\omega) = \sum_j d^{(A)}(\rho_a) D_p^{(B)}(\rho_b). \tag{1.18c}$$

In addition, we find the following mixed sum rules:

$$\eta_4^{(AB)} D_4(\omega) + \frac{d(\omega_0)}{2 + d(\omega_0)} \xi_2^{(A)} \xi_2^{(B)} \overline{D}_4(\omega)$$
$$= \sum_j D_2^{(A)}(\rho_a) D_2^{(B)}(\rho_b), \tag{1.19}$$

$$\eta_5^{(AB)} D_5(\omega) + \frac{d(\omega_0)}{6 + d(\omega_0)} \xi_2^{(A)} \xi_3^{(B)} \times \left\{\frac{D_2(\omega)}{d(\omega)} - \frac{1}{4}\frac{D_2(\omega_0)}{d(\omega_0)}\right\} D_3(\omega)$$
$$= \sum_j D_2^{(A)}(\rho_a) D_3^{(B)}(\rho_b), \tag{1.20}$$

for new constants $\eta_4^{(AB)}$ and $\eta_5^{(AB)}$.

We remark that Eqs. (1.19) and (1.20) are also valid even when at least one of L_A and L_B is simple and the other is an Abelian Lie algebra. If L_A is an Abelian Lie algebra, then the corresponding $g_{\mu\nu}^{(A)}$ cannot be the Killing metric, but we can always find a suitable $g_{\mu\nu}^{(A)}$ such that

$$I_2^{(A)} = g^{(A)\mu\nu} t_\mu^{(A)} t_\nu^{(A)}$$

is a Casimir invariant with a property that the inverse matrix $g_{\mu\nu}^{(A)}$ of $g^{(A)\mu\nu}$ exists. This is because L_A is still a reductive Lie algebra.[26] However, we will not explicitly discuss such a case hereafter.

For special cases, we can say more. For example, E_6 has no seventh-order fundamental Casimir invariant. Then, this gives the trace identity of the seventh order,[27] which implies a validity of mixed branching index sum rules of a new type. Similarly, E_8 has no genuine sixth-order fundamental Casimir invariant, leading to complicated new sum rules. These will be discussed shortly.

Let $\square$ be a fixed irreducible representation of L. If $D_p(\square)\neq 0$, then it is convenient to set

$$D_p(\omega)=Q_p(\omega)D_p(\square). \tag{1.21}$$

We emphasize the fact that $Q_p(\omega)$ is well defined only when $J_p\neq 0$. For cases of L being classical Lie algebras $A_n(n\geqslant 1)$, $B_n(n\geqslant 2)$, $C_n(n\geqslant 2)$, and $D_n(n\geqslant 3)$, the explicit formula of $Q_p(\omega)$ for any irreducible representation ω has been calculated elsewhere[28] [hereafter referred to as (II)], where we choose $\square$ to be the defining representation of L. However, if L is one of the exceptional Lie algebras G_2, F_4, E_6, E_7, and E_8, then it is very difficult to give the general formula of $D_p(\omega)$ or $Q_p(\omega)$, although we can easily evaluate them for a few low-dimensional representations (see the Appendix). Regardless, this fact causes the branching index sum rules such as Eqs. (1.8) and (1.12) to be less useful in practice for exceptional Lie algebras. We can circumvent the problem by using the other indices $l_2(\omega)$ and $l_4(\omega)$ introduced by Patera *et al.*,[29] which may be defined as

$$l_{2p}(\omega)=\sum_M (M,M)^p \quad (p=0,1,2,\ldots). \tag{1.22}$$

Here, the summation extends over all weights M of ω and (M,M) is the standard scalar product in the root space of L. An extensive table of $d(\omega)$, $l_2(\omega)$, and $l_4(\omega)$ has been tabulated in Ref. 19 for any simple Lie algebra L with rank less than or equal to 8. First, it is easy to prove

$$D_2(\omega)=r^{-1}d(\omega_0)l_2(\omega) \tag{1.23}$$

for any simple Lie algebra L with rank r. Then Eq. (1.11) gives

$$\overline{D}_4(\omega)=\left[\frac{d(\omega_0)}{r}\right]^2 l_2(\omega)\left[\frac{l_2(\omega)}{d(\omega)}-\frac{1}{6}\frac{l_2(\omega_0)}{d(\omega_0)}\right]. \tag{1.24}$$

The relation between $l_4(\omega)$ and $D_4(\omega)$ is more complicated, but we find

$$l_4(\omega)=\eta D_4(\omega)+\frac{r(r+2)}{d(\omega_0)[2+d(\omega_0)]}\overline{D}_4(\omega), \tag{1.25}$$

except for the case of $L=D_4$ corresponding to the SO(8) group where the relation is slightly more involved. In Eq. (1.25), η is a constant whose explicit value has been calculated in Ref. 24 for a suitable normalization of $D_4(\omega)$. It is found that η is nonzero for all classical Lie algebras $A_n(n\geqslant 3)$, $B_n(n\geqslant 3)$, $C_n(n\geqslant 3)$, and $D_n(n\geqslant 5)$.

For any exceptional Lie algebras as well as A_1 and A_2, we know $D_4(\omega)=0$ identically.[24] Then Eqs. (1.24) and (1.25) imply the validity of

$$l_4(\omega)=\frac{(r+2)d(\omega_0)}{r[2+d(\omega_0)]}l_2(\omega)\left\{\frac{l_2(\omega)}{d(\omega)}-\frac{1}{6}\frac{l_2(\omega_0)}{d(\omega_0)}\right\} \tag{1.26}$$

for these algebras. This relation is also valid for $B_2(=C_2)$ (because of $\eta=0$) as well as special classes of irreducible representations ω for the Lie algebra D_4 by reasons explained in Ref. 24. We can verify Eq. (1.26) for many ω's from the table of Ref. 19 for these cases. Hereafter, we restrict ourselves to the consideration of L being one of the exceptional Lie algebras, unless it is stated otherwise. We now set

$$\bar{\xi}_2=(r_0/r)[d(\omega_0)/d_0(\rho_0)]\xi_2, \tag{1.27}$$

where r_0 and $d_0(\rho_0)$ are the rank and the dimension of the simple sub-Lie algebra L_0, respectively. Then, first, Eq. (1.8) for $p=2$ is rewritten as

$$\bar{\xi}_2 l_2(\omega)=\sum_j l_2^{(0)}(\rho_j) \tag{1.28}$$

for the branching rule equation (1.1). Also, since any exceptional Lie algebra possesses no third-order and fourth-order fundamental Casimir invariants, we have $D_3(\omega)=D_4(\omega)=0$, which leads to the sum rules

$$\sum_j D_3^{(0)}(\rho_j)=0, \tag{1.29}$$

$$\frac{r}{r+2}\frac{r_0+2}{r_0}(\bar{\xi}_2)^2 l_4(\omega)=\sum_j l_4^{(0)}(\rho_j), \tag{1.30}$$

$$\frac{d_0(\rho_0)+2}{d_0(\rho_0)}\frac{r}{r+2}(\bar{\xi}_2)^2 l_4(\omega)+\frac{1}{6}\frac{l_2^{(0)}(\rho_0)}{d_0(\rho_0)}\bar{\xi}_2 l_2(\omega)=\sum_j\frac{[l_2^{(0)}(\rho_j)]^2}{d_0(\rho_j)}. \tag{1.31}$$

We note that Eq. (1.31) is not independent of Eq. (1.30) if L_0 is one of $A_1, A_2, B_2, G_2, F_4, E_6, E_7, E_8$. Together with the dimensional sum rule

$$d(\omega)=\sum_j d_0(\rho_j), \tag{1.32}$$

these formulas are found to be in general sufficient to determine the branching rule uniquely for many low-dimensional representations ω. The important exception for this statement is cases of representations of E_6, which are not self-contragradient. This is because a representation ρ and its contragradient representation ρ^* have exactly the same $d_0(\rho)$, $l_2^{(0)}(\rho)$, and $l_4^{(0)}(\rho)$ so that these sum rules cannot distinguish ρ_j from ρ_j^*. Since odd index $D_3(\rho)$ changes its sign for $\rho\to\rho^*$, the sum rule Eq. (1.29) can reduce the ambiguity for the case when L_0 contains one of A_n as its factor. However, we cannot resolve the overall contragradiency of representations. We can resolve the whole matter by considering the $D_5(\omega)$ sum rule, which we rewrite as

$$\bar{\xi}_5 Q_5(\omega)=\sum_j Q_5^{(0)}(\rho_j), \tag{1.33}$$

in terms of $Q_5(\omega)$ defined by Eq. (1.21). As we shall see in Sec. III, the validity of Eq. (1.33) can settle the ambiguity of choosing ρ_j or ρ_j^*, when we note

$$Q_5^{(0)}(\rho^*)=-Q_5^{(0)}(\rho). \tag{1.34}$$

Next, let us consider the case $L_0=L_A\times L_B$ as in Eq. (1.16). First, Eqs. (1.18) and (1.19) with $D_4(\omega)=0$ lead to

$$\bar{\xi}_2^{(A)}l_2(\omega)=\sum_j d^{(B)}(\rho_b)l_2^{(A)}(\rho_a), \tag{1.35a}$$

$$\bar{\xi}_2^{(B)}l_2(\omega)=\sum_j d^{(A)}(\rho_a)l_2^{(B)}(\rho_b), \tag{1.35b}$$

$$\frac{r}{r+2}\frac{2+r_A}{r_A}(\bar{\xi}_2^{(A)})^2 l_4(\omega)=\sum_j d^{(B)}(\rho_b)l_4^{(A)}(\rho_a), \tag{1.36a}$$

$$\frac{r}{r+2}\frac{2+r_B}{r_B}(\bar{\xi}_2^{(B)})^2 l_4(\omega)=\sum_j d^{(A)}(\rho_a) l_4^{(B)}(\rho_b), \tag{1.36b}$$

$$\frac{r}{r+2}\bar{\xi}_2^{(A)}\bar{\xi}_2^{(B)} l_4(\omega)=\sum_j l_2^{(A)}(\rho_a) l_2^{(B)}(\rho_b), \tag{1.37}$$

while Eq. (1.20) with $D_3(\omega)=0$ may be rewritten as

$$\eta_5^{(AB)} D_5(\omega)=\sum_j D_2^{(A)}(\rho_a) D_3^{(B)}(\rho_b), \tag{1.38a}$$

$$\eta_5^{(BA)} D_5(\omega)=\sum_j D_2^{(B)}(\rho_b) D_3^{(A)}(\rho_a). \tag{1.38b}$$

Similarly, we find

$$\sum_j d^{(A)}(\rho_a) Q_5^{(B)}(\rho_b)=\bar{\xi}_5^{(B)} Q_5(\omega), \tag{1.39a}$$

$$\sum_j l_2^{(A)}(\rho_a) Q_5^{(B)}(\rho_b) =\bar{\xi}_2^{(A)}\bar{\xi}_5^{(B)}\frac{d(\omega_0)}{d(\omega_0)+10}\left\{\frac{l_2(\omega)}{d(\omega)}-\frac{5}{12}\frac{l_2(\omega_0)}{d(\omega_0)}\right\} Q_5(\omega), \tag{1.39b}$$

$$\sum_j D_3^{(A)}(\rho_a) D_4^{(B)}(\rho_b)=0, \tag{1.39c}$$

$$\sum_j D_3^{(A)}(\rho_a)\overline{D}_4^{(B)}(\rho_b) =2\xi_2^{(B)}\eta_5^{(BA)}\frac{d^{(B)}(\rho_0)+2}{d^{(B)}(\rho_0)}\frac{d(\omega_0)}{d(\omega_0)+10} D_5(\omega) \times\left\{\frac{D_2(\omega)}{d(\omega)}-\frac{5}{12}\frac{D_2(\omega_0)}{d(\omega_0)}\right\}, \tag{1.39d}$$

as the fifth- and seventh-order sum rules involving $D_5(\omega)$ [or $Q_5(\omega)$]. From Eqs. (1.39c) and (1.39d), we can also derive a calculationally more convenient sum rule:

$$\sum_j D_3^{(A)}(\rho_a) l_4^{(B)}(\rho_b) =2\bar{\xi}_2^{(B)}\eta_5^{(BA)}\frac{r_B+2}{d_B(\rho_0)}\frac{d(\omega_0)}{d(\omega_0)+10} D_5(\omega) \times\left\{\frac{l_2(\omega)}{d(\omega)}-\frac{5}{12}\frac{l_2(\omega_0)}{d(\omega_0)}\right\}. \tag{1.39e}$$

We remark that Eqs. (1.39) are also valid for $L=D_5$ corresponding to the SO(10) group. The validity of the sum rules Eqs. (1.35)–(1.39) turns out to be sufficiently restrictive so as to uniquely determine the BR's of many ω's of any exceptional Lie algebras as far as can be checked. Examples will be given in Sec. III.

When L_0 is a direct product of three simple Lie algebras,

$$L_0=L_A\times L_B\times L_C,$$

with

$$\omega\supset\rho=\sum_j\oplus\rho_j=\sum_j\oplus(\rho_a^{(A)}\otimes\rho_b^{(B)}\otimes\rho_c^{(C)}),$$

the knowledge of the six-order index $D_6(\omega)$ is convenient. For this case, we can prove the sum rule of the form

$$\sum_j l_2^{(A)}(\rho_a) l_2^{(B)}(\rho_b) l_2^{(C)}(\rho_c) =\eta_6 D_6(\omega)+\{[d(\omega_0)]^2/[d(\omega_0)+2][d(\omega_0)+4]\} \times\bar{\xi}_2^{(A)}\bar{\xi}_2^{(B)}\bar{\xi}_2^{(C)} R_6(\omega), \tag{1.40}$$

where we have set

$$R_6(\omega)=l_2(\omega)\left\{\left[\frac{l_2(\omega)}{d(\omega)}\right]^2-\frac{1}{2}\frac{l_2(\omega)}{d(\omega)}\frac{l_2(\omega_0)}{d(\omega_0)} +\frac{1}{12}\left[\frac{l_2(\omega_0)}{d(\omega_0)}\right]^2\right\}. \tag{1.41}$$

We may also note that we have a six-order sum rule for $L\downarrow L_0$ of the form

$$\frac{[d_0(\rho_0)]^2}{[d_0(\rho_0)+2][d_0(\rho_0)+4]}\sum_j\frac{[l_2^{(0)}(\rho_j)]^3}{(d_0(\rho_j))^2} =\eta_6' D_6(\omega)+\frac{[d(\omega_0)]^2}{[d(\omega_0)+2][d(\omega_0)+4]}(\bar{\xi}_2)^3\frac{[l_2(\omega)]^3}{[d(\omega)]^2} +\frac{r}{2(r+2)}(\bar{\xi}_2)^2\left\{\frac{l_2^{(0)}(\rho_0)}{d_0(\rho_0)+4}-\frac{l_2(\omega_0)}{d(\omega_0)+4}\bar{\xi}_2\right\} l_4(\omega), \tag{1.42}$$

if L_0 is simple.

Also for $L_0=L_A\times L_B$, we find

$$\sum_j l_2^{(A)}(\rho_a) l_4^{(B)}(\rho_b) =\eta_6'' D_6(\omega)+\{[d(\omega_0)]^2/[d(\omega_0)+2][d(\omega_0)+4]\} \times\bar{\xi}_2^{(A)}[\bar{\xi}_2^{(B)}]^2[(r_B+2)/r_B] R_6(\omega). \tag{1.43}$$

For $L=E_8$, we have $D_6(\omega)=0$ identically, so that Eqs. (1.40), (1.42), and (1.43) do not contain new unknown constants η_6, η_6', and η_6''.

Finally, we simply mention here a mixed seventh-order index sum rule for $L\downarrow L_0=L_A\times L_B\times L_C$ of the form

$$\sum_j l_2^{(A)}(\rho_a) D_3^{(B)}(\rho_b) l_2^{(C)}(\rho_c) =\frac{d(\omega_0)}{d(\omega_0)+10}\left\{\frac{r_B}{d^{(A)}(\rho_0)}\bar{\xi}_2^{(C)}\eta_5^{(AB)}+\frac{r_C}{d^{(C)}(\rho_0)} \times\bar{\xi}_2^{(A)}\eta_5^{(CB)}\right\} D_5(\omega)\left\{\frac{l_2(\omega)}{d(\omega)}-\frac{5}{12}\frac{l_2(\omega_0)}{d(\omega_0)}\right\}, \tag{1.44}$$

where $\eta_5^{(AB)}$ or $\eta_5^{(CB)}$ is the same coefficient $\eta_5^{(AB)}$ or $\eta_5^{(CB)}$ appearing in Eq. (1.38a) when we consider $L\downarrow L_A\times L_B$, or $L\downarrow L_C\times L_B$. Equation (1.44) is again useful to resolve possible contragradiency ambiguity which may occur for the case of $L=E_6$. It is also valid for the Lie algebra $L=D_5$.

II. DERIVATION OF MAIN FORMULA

First, let

$$X_\mu=\omega(t_\mu) \tag{2.1}$$

be the representation matrix of element $t_\mu\in L$ in the irreducible representation ω. Then as is well known,[30] $\mathrm{Tr}(X_\mu X_\nu)$ is proportional to $g_{\mu\nu}$, if L is simple:

$$\mathrm{Tr}(X_\mu X_\nu)=[1/d(\omega_0)] D_2(\omega) g_{\mu\nu}, \tag{2.2}$$

where ω_0 is the adjoint representation of L and

$$D_2(\omega)=\mathrm{Tr}(g^{\mu\nu}X_\mu X_\nu). \tag{2.3}$$

Restricting X_μ and X_ν to be elements of a Cartan subalgebra H_a and H_b in Eq. (2.2) and multiplying g^{ab}, we find Eq. (1.23), i.e.,

$$D_2(\omega) = r^{-1} d(\omega_0) l_2(\omega), \tag{2.4}$$

where r is the rank of L and $l_2(\omega)$ is defined by Eq. (1.22). Since (M,M) in Ref. (19) is normalized by the condition $(\alpha,\alpha)_{\max} = 2$ for roots α of L, Eq. (2.4) imposes a suitable choice of the normalization constant C in Eq. (1.5).

Next let us briefly review the derivation of the index branching sum rules Eq. (1.8) for $p = 2$ and $p = 4$. Consider

$$\frac{1}{4!} \sum_P \mathrm{Tr}(X_\mu X_\nu X_\alpha X_\beta), \tag{2.5}$$

where P stands for 4! permutations of indices μ, ν, α, and β. If L is not the Lie algebra D_4 corresponding to the SO(8) group, then by reasons explained in (I) and in Ref. 24 we can express it as a linear combination

$$\frac{1}{4!} \sum_P \mathrm{Tr}(X_\mu X_\nu X_\alpha X_\beta) = F_1(\omega) g_{\mu\nu\alpha\beta} + F_2(\omega) \{ g_{\mu\nu} g_{\alpha\beta} + g_{\mu\alpha} g_{\nu\beta} + g_{\mu\beta} g_{\nu\alpha} \}, \tag{2.6}$$

for some $F_1(\omega)$ and $F_2(\omega)$ to be determined. Multiplying $g^{\mu\nu\alpha\beta}$ and $g^{\mu\nu} g^{\alpha\beta}$ to both sides of Eq. (2.6) and noting the orthogonality condition Eq. (1.4), we obtain

$$\begin{aligned} F_1(\omega) g_{\mu\nu\alpha\beta} g^{\mu\nu\alpha\beta} &= D_4(\omega), \\ F_2(\omega) d(\omega_0)[d(\omega_0) + 2] &= \overline{D}_4(\omega). \end{aligned} \tag{2.7}$$

If $g_{\mu\nu\alpha\beta} g^{\mu\nu\alpha\beta} = 0$, then this implies $D_4(\omega) = 0$ for all irreducible representations ω of L and hence $J_4 = 0$ by Harish-Chandra's theorem,[23] which in turn implies $g^{\mu\nu\alpha\beta} = 0$. Therefore, we may rewrite Eq. (2.6) as

$$\begin{aligned} \frac{1}{4!} \sum_P \mathrm{Tr}(X_\mu X_\nu X_\alpha X_\beta) &= C_1 D_4(\omega) g_{\mu\nu\alpha\beta} + \{ 1/d(\omega_0)[d(\omega_0) + 2] \} \\ &\quad \times \overline{D}_4(\omega) \{ g_{\mu\nu} g_{\alpha\beta} + g_{\mu\alpha} g_{\nu\beta} + g_{\mu\beta} g_{\nu\alpha} \}. \end{aligned} \tag{2.8}$$

Here, the constant C_1 is given by

$$C_1 = 1/g_{\mu\nu\alpha\beta} g^{\mu\nu\alpha\beta}, \tag{2.9}$$

if J_4 is not identically zero, while C_1 is arbitrary if $J_4 = 0$ identically. In either case, the constant C_1 is regarded to be independent of the generic irreducible representation ω. When L is the Lie algebra D_4, we have two independent fourth-order Casimir invariants, so that we have to modify Eq. (2.8) by adding another term $\widehat{C}_1 \widehat{D}_4(\omega) \hat{g}_{\mu\nu\alpha\beta}$ to the right side of Eq. (2.8), as in Ref. 24. If we restrict ourselves to Cartan subalgebra elements of $X_\mu = H_a$, $X_\nu = H_b$, $X_\alpha = H_c$, and $X_\beta = H_d$ and multiply $g^{ab} g^{cd}$ to both sides of Eq. (2.8), we find

$$l_4(\omega) = \eta D_4(\omega) + \{ r(r+2)/d(\omega_0)[d(\omega_0) + 2] \} \overline{D}_4(\omega), \tag{2.10}$$

$$\eta = C_1 g^{ab} g^{cd} g_{abcd}. \tag{2.11}$$

Clearly, the constant η is independent of the generic irreducible representation ω, and this reproduces Eq. (1.25). The explicit value of η has been computed in Ref. 24 for a suitable normalization of $D_4(\omega)$.

Now, let L_0 be a simple sub-Lie algebra of L. We rearrange the basis of L so that the base of L_0 is spanned by its first m elements $t_1, t_2, \ldots, t_m$ of L. In order to distinguish the basis of L_0 from that of L, we use the notation $\{ t_j \}$ $(j = 1,2, \ldots, m)$ and $\{ t_\mu \}$ $(\mu = 1,2, \ldots, d)$ for bases of L_0 and L, respectively. The pth-order fundamental index $D_p^{(0)}(\rho)$ of L_0 is defined by

$$\begin{aligned} D_p^{(0)}(\rho) &= \mathrm{Tr}^{(\rho)} J_p^{(0)} \\ &= g^{(0) j_1 j_2 \cdots j_p} \mathrm{Tr}^{(\rho)}(X_{j_1} X_{j_2} \cdots X_{j_p}), \end{aligned} \tag{2.12}$$

where

$$J_p^{(0)} = g^{(0) j_1 j_2 \cdots j_p} t_{j_1} t_{j_2} \cdots t_{j_p} \tag{2.13}$$

is the pth-order fundamental Casimir invariant of L_0 and ρ is a representation of L_0, which may not be necessarily irreducible. In Eq. (2.2) we now restrict indices to those of the subalgebra L_0 with $\mu = i$ and $\nu = j$, and multiply $g^{(0)ij}$ to both sides. When we note

$$D_p^{(0)}(\rho) = \sum_j D_p^{(0)}(\rho_j)$$

for $\rho = \Sigma_j \oplus \rho_j$, this reproduces Eq. (1.8) for $p = 2$ as

$$\xi_2 D_2(\omega) = \sum_j D_2^{(0)}(\rho_j), \tag{2.14a}$$

$$\xi_2 = \frac{1}{d(\omega_0)} g^{(0)ij} g_{ij} = \frac{d_0(\rho_0)}{d(\omega_0)} \eta_2, \tag{2.14b}$$

where we have defined η_2 by

$$g_{ij} = \eta_2 g_{ij}^{(0)}, \tag{2.15}$$

and $d_0(\rho_0)$ is the dimension of L_0. Similarly, we restrict ourselves in Eq. (2.8) to L_0 with $\mu = i$, $\nu = j$, $\alpha = k$, and $\beta = l$, and multiply $g^{(0)ijkl}$ to both sides. Noting the orthogonality condition

$$g^{(0)ijkl} g_{ij}^{(0)} g_{kl}^{(0)} = 0,$$

we find similarly

$$\xi_4 D_4(\omega) = \sum_j D_4^{(0)}(\rho_j), \tag{2.16a}$$

$$\xi_4 = C_1 g^{(0)ijkl} g_{ijkl}. \tag{2.16b}$$

Clearly both ξ_2 and ξ_4 are independent of the generic irreducible representation ω of L. However, if L is the Lie algebra D_4, then Eq. (2.16a) must be modified by Eq. (1.9) with $p = 4$ since the Lie algebra D_4 has now two independent fourth-order fundamental indices $D_4(\omega)$ and $\widehat{D}_4(\omega)$ as in Ref. 24.

Next, we multiply $g^{(0)ij} g^{(0)kl}$ to both sides of Eq. (2.8) with $\mu = i$, $\nu = j$, $\alpha = k$, and $\beta = l$. Then, after some calculations, we obtain now the mixed fourth-order sum rule Eq. (1.12)

$$\sum_j \overline{D}_4^{(0)}(\rho_j) = \eta_4 D_4(\omega) + \theta \overline{D}_4(\omega), \tag{2.17a}$$

$$\eta_4 = C_1 g^{(0)ij} g^{(0)kl} g_{ijkl}, \tag{2.17b}$$

$$\begin{aligned} \theta &= \{ d_0(\rho_0)[d_0(\rho_0) + 2]/d(\omega_0)[d(\omega_0) + 2] \} (\eta_2)^2 \\ &= \{ d(\omega_0)[d_0(\rho_0) + 2]/d_0(\rho_0)[d(\omega_0) + 2] \} (\xi_2)^2. \end{aligned} \tag{2.17c}$$

We remark[24] that g_{ijkl} may not necessarily be proportional to $g^{(0)}_{ijkl}$, although it can be expressed as a linear combination of $g^{(0)}_{ijkl}$ and $g^{(0)}_{ij}g^{(0)}_{kl}+g^{(0)}_{ik}g^{(0)}_{jl}+g^{(0)}_{il}g^{(0)}_{kl}$. Hence, η_4 defined by Eq. (2.17b) is not necessarily zero but may be regarded as a new unknown constant, although it can be in principle computable.

If L is one of A_1, A_2, G_2, F_4, E_6, E_7, and E_8, we know[24] $D_4(\omega)=0$ identically for all ω. In that case, Eqs. (2.10) and (2.17a) are rewritten as Eqs. (1.26) and (1.31), when we note Eq. (2.14a). The mixed sum rule Eq. (1.30) can be derived also as follows. Assuming $L_0\neq D_4$, Eq. (1.25) for L_0 with ω being replaced by ρ_j gives

$$l_4^{(0)}(\rho_j)=\eta_0 D_4^{(0)}(\rho_j)+\frac{r_0(r_0+2)}{d_0(\rho_0)[d_0(\rho_0)+2]}\overline{D}_4^{(0)}(\rho_j). \tag{2.18}$$

Summing over j, this leads to

$$\sum_j l_4^{(0)}(\rho_j)=\eta_0\sum_j D_4^{(0)}(\rho_j)+\frac{r_0(r_0+2)}{d_0(\rho_0)[d_0(\rho_0)+2]}\times\sum_j \overline{D}_4^{(0)}(\rho_j).$$

Using Eqs. (2.16a), (2.17a), and (2.10), we can rewrite this as

$$\sum_j l_4^{(0)}(\rho_j)=\overline{C}_4 D_4(\omega)+\frac{r_0(r_0+2)}{r(r+2)}\frac{d(\omega_0)[d(\omega_0)+2]}{d_0(\rho_0)[d_0(\rho_0)+2]}\times\theta l_4(\omega), \tag{2.19a}$$

$$\overline{C}_4=\eta_0\xi_4+\{r_0(r_0+2)/d_0(\rho_0)[d_0(\rho_0)+2]\}\times\{\eta_4-(d(\omega_0)[d(\omega_0)+2]/r(r+2))\theta\eta\}. \tag{2.19b}$$

If L is one of $A_1, A_2, G_2, F_4, E_6, E_7$, and E_8, then $D_4(\omega)=0$ so that Eq. (2.19a) reproduces Eq. (1.30), when we further note Eqs. (1.27) and (2.17c). Although we have assumed $L_0\neq D_4$ for this derivation, this assumption is actually unnecessary for the validity of Eq. (2.19a). However, we will not go into its detail here, since it will be straightforward to prove it with a slight modification. We remark here that Eqs. (1.30) and (1.31) will not be independent of each other when L_0 is one of A_1, A_2, B_2, G_2, F_4, E_6, E_7, and E_8, since $l_4^{(0)}(\rho_j)$ will be expressed in terms of $l_2^{(0)}(\rho_j)$ as in Eq. (1.26) with replacement of $\omega\to\rho$ and $\omega_0\to\rho_0$.

Next, consider the derivation of Eq. (1.15). Because of the reason explained in (I), we can express $\Sigma_P \mathrm{Tr}(X_\mu X_\nu X_\lambda X_\alpha X_\beta)$ also as

$$\frac{1}{5!}\sum_P \mathrm{Tr}(X_\mu X_\nu X_\lambda X_\alpha X_\beta)=F_3(\omega)g_{\mu\nu\lambda\alpha\beta}+F_4(\omega)\frac{1}{5!}\sum_P g_{\mu\nu}g_{\lambda\alpha\beta}, \tag{2.20}$$

where P stands now for 5! permutations of indices μ, ν, λ, α, and β. Multiplying $g^{\mu\nu\lambda\alpha\beta}$ to both sides of Eqs. (2.20) and noting the orthogonality condition

$$g^{\mu\nu\lambda\alpha\beta}g_{\mu\nu}g_{\lambda\alpha\beta}=0,$$

it gives

$$D_5(\omega)=F_3(\omega)g^{\mu\nu\lambda\alpha\beta}g_{\mu\nu\lambda\alpha\beta}. \tag{2.21a}$$

Similarly, multiplying $g^{\mu\nu}g^{\lambda\alpha\beta}$ and noting $g^{\mu\nu}g_{\lambda\mu\nu}=0$, we find

$$D_3(\omega)\left[\frac{D_2(\omega)}{d(\omega)}-\frac{1}{4}\frac{D_2(\omega_0)}{d(\omega_0)}\right]=\tfrac{1}{10}[d(\omega_0)+6]g^{\lambda\alpha\beta}g_{\lambda\alpha\beta}F_4(\omega), \tag{2.21b}$$

which determines $F_3(\omega)$ and $F_4(\omega)$ in terms of indices. We now restrict Greek suffices μ, ν, λ, α, and β in Eq. (2.20) to the subindices i, j, k, l, and p of L_0, and multiply $g^{(0)ijklp}$ or $g^{(0)ij}g^{(0)klp}$ to both sides of Eq. (2.20). After some calculations, we find Eq. (1.8) for $p=5$ and Eq. (1.15), respectively, with $C_5=g^{(0)ij}g^{(0)klp}g_{ijklp}/g^{\mu\nu\lambda\alpha\beta}g_{\mu\nu\lambda\alpha\beta}$.

We will next discuss the case where L_0 is a direct product of two simple Lie algebras L_A and L_B, i.e., $L_0=L_A\times L_B$. We label bases of L_A and L_B, respectively, as $\{t_i\}$ and $\{t_a\}$. Considering $L\downarrow L_A$ and $L\downarrow L_B$, we have clearly

$$\xi_p^{(A)}D_p(\omega)=\sum_j d^{(B)}(\rho_b)D_p^{(A)}(\rho_a), \tag{2.22a}$$

$$\xi_p^{(B)}D_p(\omega)=\sum_j d^{(A)}(\rho_a)D_p^{(B)}(\rho_b), \tag{2.22b}$$

when we decompose ω as

$$\omega\supset\rho=\sum_j\oplus\rho_j=\sum_j\oplus[\rho_a^{(A)}\otimes\rho_b^{(B)}] \tag{2.23}$$

of irreducible components $\rho_j=\rho_a^{(A)}\otimes\rho_b^{(B)}$ of $L_A\times L_B$. Next, we study the cases of $p=4$ and 5 in more detail. In Eq. (2.8), we choose $X_\mu=X_i$, $X_\nu=X_j\in L_A$ and $X_\alpha=X_a$, $X_\beta=X_b\in L_B$, and multiply $g^{(A)ij}g^{(B)ab}$ to find

$$\sum_j D_2^{(A)}(\rho_a)D_2^{(B)}(\rho_b)=\eta_4^{(AB)}D_4(\omega)+\frac{d(\omega_0)}{d(\omega_0)+2}\xi_2^{(A)}\xi_2^{(B)}\overline{D}_4(\omega), \tag{2.24a}$$

$$\eta_4^{(AB)}=C_1 g^{(A)ij}g^{(B)ab}g_{ijab}. \tag{2.24b}$$

Similarly, we choose $X_\mu=X_i$, $X_\nu=X_j\in L_A$, $X_\lambda=X_a$, $X_\alpha=X_b$, $X_\beta=X_c\in L_B$ in Eq. (2.20) and multiply $g^{(A)ij}g^{(B)abc}$. In this way, we find

$$\sum_j D_2^{(A)}(\rho_a)D_3^{(B)}(\rho_b)=\eta_5^{(AB)}D_5(\omega)+\frac{d(\omega_0)}{d(\omega_0)+6}\xi_2^{(A)}\xi_3^{(B)}\times D_3(\omega)\left[\frac{D_2(\omega)}{d(\omega)}-\frac{1}{4}\frac{D_2(\omega_0)}{d(\omega_0)}\right], \tag{2.25a}$$

$$\eta_5^{(AB)}=g^{(A)ij}g^{(B)abc}g_{ijabc}/g^{\mu\nu\lambda\alpha\beta}g_{\mu\nu\lambda\alpha\beta}, \tag{2.25b}$$

which is Eq. (1.20). If $g^{\mu\nu\lambda\alpha\beta}g_{\mu\nu\lambda\alpha\beta}=0$, then $\eta_5^{(AB)}$ is arbitrary.

Hereafter, we assume L to be one of the exceptional Lie algebras G_2, F_4, E_6, E_7, and E_8 unless it is stated otherwise. We now have $D_3(\omega)=D_4(\omega)=D_7(\omega)=0$ identically for all ω, since exceptional Lie algebras do not possess any fundamental third, fourth, and seventh Casimir invariants. Also, we have $D_5(\omega)=0$ except for the case of $L=E_6$. Moreover, $D_6(\omega)=0$ for $L=E_8$. Equations (1.37) and (1.38a) are direct consequences of Eqs. (2.24a) and (2.25a), respectively, when we note

$$l_4(\omega)=\frac{r(r+2)}{d(\omega_0)[d(\omega_0)+2]}\bar{D}_4(\omega)$$
$$=\frac{(r+2)d(\omega_0)}{r[d(\omega_0)+2]}l_2(\omega)\left\{\frac{l_2(\omega)}{d(\omega)}-\frac{1}{6}\frac{l_2(\omega_0)}{d(\omega_0)}\right\},\qquad(2.26a)$$
$$\bar{\xi}_2=\frac{r_0}{r}\eta_2=\frac{r_0}{r}\frac{d(\omega_0)}{d_0(\rho_0)}\xi_2.\qquad(2.26b)$$

Consider next the sixth and seventh index sum rules. First, we express

$$\frac{1}{6!}\sum_P \mathrm{Tr}(X_\mu X_\nu X_\lambda X_\alpha X_\beta X_\gamma)$$
$$=C_6D_6(\omega)g_{\mu\nu\lambda\alpha\beta\gamma}+F_6(\omega)\frac{1}{6!}\sum_P g_{\mu\nu}g_{\lambda\alpha}g_{\beta\gamma},\qquad(2.27)$$

since we have $J_3=J_4=0$ for the present case. Multiplying $g^{\mu\nu}g^{\lambda\alpha}g^{\beta\gamma}$ to both sides of Eq. (2.27), we find

$$F_6(\omega)=\frac{15}{d(\omega_0)[d(\omega_0)+2][d(\omega_0)+4]}D_2(\omega)\left\{\left[\frac{D_2(\omega)}{d(\omega)}\right]^2\right.$$
$$\left.-\frac{1}{2}\frac{D_2(\omega_0)}{d(\omega_0)}\frac{D_2(\omega)}{d(\omega)}+\frac{1}{12}\left[\frac{D_2(\omega_0)}{d(\omega_0)}\right]^2\right\}.\qquad(2.28)$$

Choosing $\mu=i$, $\nu=j$, $\lambda=k$, $\alpha=l$, $\beta=p$, and $\gamma=q\in L_0$, and multiplying $g^{ij}g^{kl}g^{pq}$ to both sides of Eq. (2.27), we find

$$\sum_j R_6^{(0)}(\rho_j)=\eta_6''D_6(\omega)+\left[\frac{d(\omega_0)}{d_0(\rho_0)}\right]^2$$
$$\times\frac{[d_0(\rho_0)+2][d_0(\rho_0)+4]}{[d_0(\omega_0)+2][d(\omega_0)+4]}(\bar{\xi}_2)^3R_6(\omega),$$

which gives Eq. (1.42). Similarly, we can derive Eq. (1.40). Analogously, we obtain

$$\frac{1}{7!}\sum_P \mathrm{Tr}(X_\mu X_\nu X_\lambda X_\alpha X_\beta X_\gamma X_\tau)$$
$$=C_5F_7(\omega)\frac{1}{7!}\sum_P g_{\mu\nu}g_{\lambda\alpha\beta\gamma\tau},\qquad(2.29a)$$
$$F_7(\omega)=\frac{21}{[d(\omega_0)+10]}D_5(\omega)\left\{\frac{D_2(\omega)}{d(\omega)}-\frac{5}{12}\frac{D_2(\omega_0)}{d(\omega_0)}\right\},\qquad(2.29b)$$
$$C_5=1/g^{\lambda\alpha\beta\gamma\tau}g_{\lambda\alpha\beta\gamma\tau}.\qquad(2.29c)$$

Actually, C_5 is meaningless when $D_5(\omega)=0$ identically. However, we can assume then C_5 to be arbitrary but finite so that we have $C_5F_7(\omega)=0$ anyway. From Eqs. (2.29), we can derive the mixed seventh-order sum rules Eqs. (1.39b)–(1.39d) and (1.43). Combining Eqs. (1.39c) and (1.39d) and using Eq. (1.25) for L_0, we find also Eq. (1.39e)

$$\sum_j D_3^{(A)}(\rho_a)l_4^{(B)}(\rho_b)$$
$$=2\bar{\xi}_2^{(B)}\eta_5^{(BA)}\frac{r_B+2}{d_B(\rho_0)}\frac{d(\omega_0)}{d(\omega_0)+10}$$
$$\times D_5(\omega)\left\{\frac{l_2(\omega)}{d(\omega)}-\frac{5}{12}\frac{l_2(\omega_0)}{d(\omega_0)}\right\}.\qquad(2.30)$$

In passing, we note that the validity of Eq. (2.29) is related to the trace identity of the seventh order. Let $t\in L$ be an arbitrary element of L so that

$$t=\xi^\mu t_\mu,\qquad(2.31a)$$

for arbitrary real or complex numbers ξ^μ. Corresponding to this, we have

$$X=\omega(t)=\xi^\mu X_\mu.\qquad(2.31b)$$

Then, multiplying $\xi^\mu\xi^\nu\xi^\lambda\xi^\alpha\xi^\beta\xi^\gamma\xi^\tau$ to both sides of Eq. (2.29a), we find[27]

$$\mathrm{Tr}\,X^7=\frac{d(\omega_0)}{D_2(\omega)D_5(\omega)}F_7(\omega)\mathrm{Tr}\,X^2\,\mathrm{Tr}\,X^5,\qquad(2.32)$$

which is also valid for the Lie algebra D_5. Similarly for the Lie algebra E_8, we have $D_6(\omega)=0$ identically since E_8 has no fundamental sixth-order Casimir invariant. Then, multiplying $\xi^\mu\xi^\nu\xi^\lambda\xi^\alpha\xi^\beta\xi^\gamma$ to Eq. (2.29), it leads to

$$\mathrm{Tr}\,X^6=\left[\frac{d(\omega_0)}{D_2(\omega)}\right]^3F_6(\omega)(\mathrm{Tr}\,X^2)^3,\qquad(2.33)$$

as has been already noted in Ref. 31 and (I). It may be needless to mention that $D_3(\omega)=D_4(\omega)=0$ implies the validity of trace identities[27]

$$\mathrm{Tr}\,X^3=0,\qquad(2.34)$$
$$\mathrm{Tr}\,X^4=\frac{1}{2[2+d(\omega_0)]}\left\{6\frac{d(\omega_0)}{d(\omega)}-\frac{D_2(\omega_0)}{D_2(\omega)}\right\}(\mathrm{Tr}\,X^2)^2.\qquad(2.35)$$

III. SOME APPLICATIONS

We shall adopt here the lexicographical ordering convention of the simple root system of L as in Ref. 19. Let Λ be the highest weight of the irreducible representation ω of L. We will often write $d(\Lambda)$, $l_{2p}(\Lambda)$, $D_p(\Lambda)$, and $Q_p(\Lambda)$ instead of $d(\omega)$, $l_{2p}(\omega)$, $D_p(\omega)$, and $Q_p(\omega)$ hereafter. If $\Lambda_1,\Lambda_2,\ldots,\Lambda_r$ are the fundamental weight system of L, then we can express[32]

$$\Lambda=m_1\Lambda_1+m_2\Lambda_2+\cdots+m_r\Lambda_r\qquad(3.1)$$

in terms of r non-negative integers $m_1,m_2,\ldots,m_r$, so that we may write also

$$\omega=\Lambda=(m_1,m_2,\ldots,m_r).\qquad(3.2)$$

It is often more convenient for us to use the Young tableau notation[33] for classical Lie algebras. First consider the Lie algebra A_r corresponding to the SU$(r+1)$ group. Let $(f_1,f_2,\ldots,f_{r+1})$ be the Young tableau symbol where f_j, $(1\leqslant j\leqslant r+1)$ are non-negative integers satisfying

$$f_1\geqslant f_2\geqslant\cdots\geqslant f_{r+1}\geqslant 0.\qquad(3.3)$$

Then, m_j's given by Eq. (3.1) are related to f_j's by

$$m_1=f_1-f_2,\quad m_2=f_2-f_3,\quad\ldots,\quad m_r=f_r-f_{r+1}.\qquad(3.4)$$

Similarly, consider the Lie algebra D_r corresponding to the SO$(2r)$ group, whose irreducible representation ω may be labeled by r real numbers $f_1,f_2,\ldots,f_r$ satisfying

$$f_1\geqslant f_2\geqslant\cdots\geqslant f_{r-1}\geqslant|f_r|.\qquad(3.5)$$

Here, f_j $(1\leqslant j\leqslant r)$ are all simultaneously integers or half-integers, corresponding to tensor and spinor representations, respectively. They are related to m_j by

$$m_j=f_j-f_{j+1}\quad(1\leqslant j\leqslant r-1),$$
$$m_r=f_{r-1}+f_r\quad(j=r).\qquad(3.6)$$

We remark that there exists an extensive table [19] for $d(\omega)$, $l_2(\omega)$, and $l_4(\omega)$ for simple Lie algebras of rank less than or equal to 8. However, there is no corresponding table available for higher-order indices. Because of this, we first calculate values of $Q_p(\omega)$ for a few low-dimensional representations ω for a later purpose. Let $\square$ be the basic (or nontrivial lowest-dimensional) representation of L. In general, we have $D_p(\square)\neq 0$, if J_p is not identically zero. For such a case, we normalize $D_p(\omega)$ by Eq. (1.21), so that

$$Q_p(\square) = 1.$$

For example, $\square = \Lambda_1$ for all classical Lie algebras with possible exception of $L = D_p$, and $\square = \Lambda_1$, for E_6 and E_8 while $\square = \Lambda_2$ for G_2, $\square = \Lambda_4$ for F_4, and $\square = \Lambda_6$ for E_7. In this note, we consider mostly the Lie algebra D_p with odd p. If p is even, then $D_p(\omega)$ here really represents $\widehat{D}_p(\omega)$, which is one of two pth-order indices. For either case, we have $D_p(\Lambda_1) = 0$ so that we have to choose $\square$ to be the fundamental spinor representation Λ_p with normalization

$$Q_p(\Lambda_p) = 1, \quad \text{for} \quad L = D_p. \tag{3.7}$$

The explicit form of Q_p in this case can be readily computed as follows. Define $l_j (1 \leqslant j \leqslant p)$ here with $p = r$ by

$$l_j = f_j + j - 1 \quad (1 \leqslant j \leqslant p). \tag{3.8}$$

Then, we calculate

$$Q_p(\Lambda) = \frac{2 l_1 l_2 \cdots l_p}{(2p-1)!!} d(\Lambda) \quad (L = D_p,\ p = r). \tag{3.9}$$

For other cases of any classical Lie algebras, the explicit formulas for $Q_p(\Lambda)$ are found in II. For example, if L is the Lie algebra A_r with $p \leqslant r+1$, then

$$\begin{aligned} Q_p(\Lambda_2) &= (r+1)^2 - 2^{p-1}, \\ Q_p(\Lambda_3) &= \tfrac{1}{2}(r+1)^2 - \tfrac{1}{2}[1 + 2^p](r+1) + 3^{p-1}. \end{aligned} \tag{3.10}$$

We will next evaluate $Q_p(\Lambda)$ for exceptional Lie algebras for low-dimensional representations. This is not difficult. First, let $N = d(\square)$. Then in (II), we noted the validity of

$$Q_p\left(\square\square\right) = N + 2^{p-1}, \tag{3.11a}$$

$$Q_p\left(\begin{array}{c}\square\\\square\end{array}\right) = N - 2^{p-1}, \tag{3.11b}$$

$$Q_p\left(\square\square\square\right) = \tfrac{1}{2}N^2 + \tfrac{1}{2}[1 + 2^p]N + 3^{p-1}, \tag{3.11c}$$

$$Q_p\left(\begin{array}{c}\square\\\square\\\square\end{array}\right) = \tfrac{1}{2}N^2 - \tfrac{1}{2}[1 + 2^p]N + 3^{p-1}, \tag{3.11d}$$

$$Q_p\left(\begin{array}{ll}\square & \square\\\square & \end{array}\right) = N^2 - 3^{p-1}, \tag{3.11e}$$

etc. As a matter of fact, Eq. (3.10) for $L = A_r$ corresponds to $N = r + 1$ in Eqs. (3.11b) and (3.11d), since

$$\left(\begin{array}{c}\square\\\square\end{array}\right) = \Lambda_2 \quad \text{and} \quad \left(\begin{array}{c}\square\\\square\\\square\end{array}\right) = \Lambda_3$$

in this case. To be definite, let us consider the case of $L = E_6$ with $\square = \Lambda_1$ and $N = 27$. Then, we have,

$$\begin{aligned} &\left(\begin{array}{c}\square\\\square\end{array}\right) = \Lambda_2, \quad \left(\begin{array}{c}\square\\\square\\\square\end{array}\right) = \Lambda_3, \\ &\left(\begin{array}{c}\square\\\square\\\square\\\square\end{array}\right) = (\Lambda_4 + \Lambda_6), \quad \left(\square\square\right) = \Lambda_5 \oplus (2\Lambda_1), \\ &\left(\square\square\square\right) = (0) \oplus (\Lambda_1 + \Lambda_5) \oplus (3\Lambda_1), \\ &\left(\begin{array}{ll}\square & \square\\\square & \end{array}\right) = \Lambda_6 \oplus (\Lambda_1 + \Lambda_5) \oplus (\Lambda_1 + \Lambda_2), \end{aligned} \tag{3.12}$$

as we may verify by using the method explained in (II). Next, we note the fact that we have

$$Q_p(\Lambda^*) = (-1)^p Q_p(\Lambda), \tag{3.13}$$

where Λ^* is the contragradient representation of Λ, which may be obtained from Λ by interchanges of

$$m_1 \leftrightarrow m_5, \quad m_2 \leftrightarrow m_4, \quad m_3 \leftrightarrow m_3, \quad m_6 \leftrightarrow m_6,$$

for the case of $L = E_6$ and $m_4 \leftrightarrow m_5$ for the case of D_5. Especially, we have for E_6

$$Q_p(\Lambda_5) = (-1)^p Q_p(\Lambda_1) = (-1)^p, \tag{3.14a}$$

$$Q_p(\Lambda_4) = (-1)^p Q_p(\Lambda_2), \tag{3.14b}$$

$$Q_p(\Lambda_3) = (-1)^p Q_p(\Lambda_3). \tag{3.14c}$$

Then, Eqs. (3.11), (3.12), and (3.14) lead to

$$Q_p(\Lambda_2) = (-1)^p Q_p(\Lambda_4) = 27 - 2^{p-1}, \tag{3.15a}$$

$$Q_p(\Lambda_3) = 27(13 - 2^{p-1}) + 3^{p-1}, \tag{3.15b}$$

$$Q_p(2\Lambda_1) = (-1)^p Q_p(2\Lambda_5) = 2^{p-1} + 27 - (-1)^p, \tag{3.15c}$$

as well as a sum rule

$$Q_p(\Lambda_1 + \Lambda_5) + Q_p(3\Lambda_1) = 27(14 + 2^{p-1}) + 3^{p-1}. \tag{3.15d}$$

We emphasize the fact that these relations are meaningful only for those values of p where E_6 has nontrivial pth-order fundamental Casimir invariant J_p, i.e., only for $p = 2,5,6,8,9$, and 12 but not for other values of p. For example, Eq. (3.15b) for $p = 5$ and $p = 9$ gives $Q_5(\Lambda_3) = Q_9(\Lambda_3) = 0$ in accordance with Eq. (3.14c) but it leads to nonsensical answers $Q_3(\Lambda_3) \neq 0$ and $Q_7(\Lambda_3) \neq 0$ for $p = 3$ and 7 in contradiction with Eq. (3.14c). The other values of $Q_p(\Lambda)$ can be similarly calculated from these, once the Kronecker decomposition

$$\omega_A \times \omega_B = \sum_j \oplus\, \omega_j \tag{3.16a}$$

for two irreducible representations ω_A and ω_B of E_6 is known. In that case, we have the sum rule[22]

$$d(\omega_A)Q_p(\omega_B) + d(\omega_B)Q_p(\omega_A) = \sum_j Q_p(\omega_j). \qquad (3.16b)$$

Examples for E_6 are $\Lambda_1 \times \Lambda_6 = (\Lambda_1 + \Lambda_6) \oplus \Lambda_1 \oplus \Lambda_4$ and $\Lambda_1 \times \Lambda_5 = (\Lambda_1 + \Lambda_5) \oplus \Lambda_6 \oplus 0$. For a later purpose, we compute values of $Q_5(\Lambda\ )$ ($p = 5$) for a few low-dimensional representations of E_6

$$\begin{aligned}
&Q_5(\Lambda_1) = -Q_5(\Lambda_5) = 1, \quad Q_5(\Lambda_2) = -Q_5(\Lambda_4) = 11,\\
&Q_5(\Lambda_3) = Q_5(\Lambda_6) = Q_5(\Lambda_1 + \Lambda_5) = 0,\\
&Q_5(2\Lambda_1) = -Q_5(2\Lambda_5) = 44,\\
&Q_5(\Lambda_1 + \Lambda_6) = -Q_5(\Lambda_5 + \Lambda_6) = 88,\\
&Q_5(\Lambda_1 + \Lambda_2) = 648, \quad Q_5(3\Lambda_1) = 891.
\end{aligned} \qquad (3.17)$$

Similarly, some values of $Q_5(\Lambda)$ for the Lie algebra A_5 are given by

$$\begin{aligned}
&Q_5(\Lambda_1) = -Q_5(\Lambda_5) = 1, \quad Q_5(\Lambda_2) = -Q_5(\Lambda_4) = -10,\\
&Q_5(\Lambda_3) = 0, \quad Q_5(\Lambda_1 + \Lambda_3) = -Q_5(\Lambda_3 + \Lambda_5) = 10,\\
&Q_5(\Lambda_1 + \Lambda_4) = -Q_5(\Lambda_2 + \Lambda_5) = 76,\\
&Q_5(2\Lambda_1) = -Q_5(2\Lambda_5) = 22,\\
&Q_5(2\Lambda_2) = -Q_5(2\Lambda_4) = -320,
\end{aligned} \qquad (3.18)$$

while for $L = D_5$, Eq. (3.9) enables us to compute

$$\begin{aligned}
&Q_5(\Lambda_4) = -Q_5(\Lambda_5) = -1,\\
&Q_5(\Lambda_j) = Q_5(\Lambda_j + \Lambda_k) = 0\\
&\quad (1 \leqslant j \leqslant k \leqslant 3),\\
&Q_5(\Lambda_1 + \Lambda_5) = -Q_5(\Lambda_1 + \Lambda_4) = 11,\\
&Q_5(\Lambda_2 + \Lambda_5) = -Q_5(\Lambda_2 + \Lambda_4) = 55,\\
&Q_5(2\Lambda_5) = -Q_5(2\Lambda_4) = 32.
\end{aligned} \qquad (3.19)$$

Now, we would like to demonstrate the usefulness of our index branching sum rules. First, we have to know values of ξ_p or $\bar{\xi}_p$. This can be readily evaluated once a branching rule for $\square$ is known. If L_0 is any simple sub-Lie algebra of L with the same rank r as L, then it has been found that we have always $\bar{\xi}_2 = 1$ except for the case of $L = C_r$ and $L_0 = A_r$, where we have $\bar{\xi} = 2$. For cases of $L = G_2, F_4, E_7$, and E_8, the situation is simple, since any representation of these Lie algebras is self-contragradient.[34] Consider as an example the case of $F_4 \downarrow B_4$. Then as we have already remarked, we have $\bar{\xi} = 1$ by studying the simplest known BR of $\Lambda_1 \to \Lambda_2 \oplus \Lambda_4$. Sum rules Eqs. (1.28), (1.30), and (1.31) are now written as

$$l_2(\omega) = \sum_j l_2^{(0)}(\rho_j), \qquad (3.20a)$$

$$l_4(\omega) = \sum_j l_4^{(0)}(\rho_j), \qquad (3.20b)$$

$$\frac{19}{27} l_4(\omega) + \frac{7}{27} l_2(\omega) = \sum_j \frac{[l_2^{(0)}(\rho_j)]^2}{d_0(\rho_j)}, \qquad (3.21)$$

where we used $r = r_0 = 4$, $d_0(\rho_0) = 36$, and $l_2^{(0)}(\rho_0) = 56$ with $\rho_0 = \Lambda_2$ for Eq. (3.21c). We have also the dimensional sum rule

$$d(\omega) = \sum_j d_0(\rho_j). \qquad (3.22)$$

These are more than sufficient to establish branching rules[19] of

$$\begin{aligned}
&\Lambda_4 \to \Lambda_4 \oplus \Lambda_1 \oplus (0),\\
&\Lambda_3 \to (\Lambda_1 + \Lambda_4) \oplus \Lambda_1 \oplus \Lambda_2 \oplus \Lambda_3 \oplus \Lambda_4,\\
&(2\Lambda_4) \to (2\Lambda_4) \oplus (\Lambda_1 + \Lambda_4) \oplus (2\Lambda_1) \oplus \Lambda_1 \oplus \Lambda_4 \oplus (0),
\end{aligned}$$

etc., if we use values of $d_0(\rho_j)$, $l_2^{(0)}(\rho_j)$, and $l_4^{(0)}(\rho_j)$ of B_4 as well as of $d(\omega)$, $l_2(\omega)$, and $l_4(\omega)$ of F_4 tabulated in Ref. 19.

The case of $L = E_6$ is slightly more involved since its representation may not necessarily be self-contragradient. Consider first the case of $E_6 \downarrow D_5$ as an illustration. Then, the simplest BR for this case is $\omega = \Lambda_1$ with the BR $\Lambda_1 \to \Lambda_5 \oplus \Lambda_1 \oplus 0$ from the table of Ref. 19. Using the values of $l_2(\Lambda_1) = 6 \times 6$, $l_2^{(0)}(\Lambda_1) = 5 \times 2$, and $l_2^{(0)}(\Lambda_5) = 5 \times 4$ again from the table of Ref. 19, we calculate $\bar{\xi}_2 = \frac{5}{6}$, since we must have

$$\bar{\xi}_2 l_2(\Lambda_1) = l_2^{(0)}(\Lambda_5) + l_2^{(0)}(\Lambda_1).$$

The branching sum rules Eqs. (1.28), (1.30), and (1.31) are now read as

$$\frac{5}{6} l_2(\omega) = \sum_j l_2^{(0)}(\rho_j), \qquad (3.23a)$$

$$\frac{35}{48} l_4(\omega) = \sum_j l_4^{(0)}(\rho_j), \qquad (3.23b)$$

$$\frac{5}{27}\left\{\frac{47}{16} l_4(\omega) + \frac{4}{3} l_2(\omega)\right\} = \sum_j \frac{[l_2^{(0)}(\rho_j)]^2}{d_0(\rho_j)}, \qquad (3.23c)$$

when we note $r = 6$, $r_0 = 5$, $d_0(\rho_0) = 45$, and $l_2^{(0)}(\rho_0) = 80$ for the present case. We may readily verify the validity of Eqs. (3.23b) and (3.23c) for $\Lambda_1 \to \Lambda_5 \oplus \Lambda_1 \oplus (0)$ again. Together with the dimensional sum rule Eq. (3.22), these sum rules are sufficient for most cases to determine the BR uniquely except for the question of the contragradiency, i.e., how to distinguish ρ_j and ρ_j^*. As an example, consider the case of $\omega = \Lambda_2$ and $2\Lambda_1$, both of which have the same dimension $d(\omega) = 351$. The application of our sum rules Eqs. (3.23a)–(3.23c) is sufficient to establish the branching rules of

$$\Lambda_2 \to (\Lambda_1 + \Lambda_5) \oplus \Lambda_1 \oplus \Lambda_2 \oplus \Lambda_3 \oplus \Lambda_4 \oplus \Lambda_5, \qquad (3.24a)$$

$$2\Lambda_1 \to (2\Lambda_5) \oplus (\Lambda_1 + \Lambda_5) \oplus (2\Lambda_1) \oplus \Lambda_5 \oplus \Lambda_1 \oplus (0), \qquad (3.24b)$$

except for ambiguities of contragradiency, since Λ_j and its contragradient representation Λ_j^* for D_5 possess exactly the same values $d_0(\Lambda_j^*) = d_0(\Lambda_j)$, $l_2^{(0)}(\Lambda_j^*) = l_2^{(0)}(\Lambda_j)$, and $l_4^{(0)}(\Lambda_j^*) = l_4^{(0)}(\Lambda_j)$. Therefore, our use of the sum rules does not rule out the replacement $\Lambda_1 + \Lambda_5 \to \Lambda_1 + \Lambda_4$ in Eq. (3.24a) and $2\Lambda_5 \to 2\Lambda_4$, $\Lambda_1 + \Lambda_5 \to \Lambda_1 + \Lambda_4$ and/or $\Lambda_5 \to \Lambda_4$ in Eq. (3.24b), since they are contragradient to each other. In order to eliminate this ambiguity, we need consider the odd-order branching sum rules Eq. (1.33). Again from the known branching rule $\Lambda_1 \to \Lambda_5 \oplus \Lambda_1 \oplus 0$, we evaluate $\bar{\xi}_5$ from Eq. (1.33) to be $\bar{\xi}_5 = 1$, where we used the values given by Eqs. (3.17) and (3.19) for $Q_5(\Lambda)$ and $Q_5^{(0)}(\Lambda)$. The fifth-order sum rule is now read as

$$Q_5(\omega) = \sum_j Q_5^{(0)}(\rho_j). \qquad (3.25)$$

Using the table in Eqs. (3.17) and (3.19), we can readily verify now that correct decompositions of Λ_2 and $2\Lambda_1$ are indeed uniquely specified by Eqs. (3.24a) and (3.24b), respectively.

Next, let us consider $E_6 \downarrow A_1 \times A_5$. From the simplest BR of $\Lambda_1 \to (0 \otimes \Lambda_4) \oplus (\Lambda_1 \otimes \Lambda_1)$, we evaluate

$$\bar{\xi}_2^{(A)} = \tfrac{1}{6}, \quad \bar{\xi}_2^{(B)} = \tfrac{5}{6}, \quad \bar{\xi}_5^{(B)} = 12, \quad \bar{\xi}_5^{(A)} = 0$$

for this case. Then, for example, the branching rule

$$\Lambda_2 \to [0 \otimes (\Lambda_3 + \Lambda_5)] \oplus [\Lambda_1 \otimes (\Lambda_1 + \Lambda_4)]$$
$$\oplus [2\Lambda_1 \otimes \Lambda_2] \oplus [0 \otimes 2\Lambda_1] \oplus [\Lambda_1 \otimes \Lambda_5]$$

can be similarly established, solving the contragradiency problem.

The situation is slightly different, however, for the BR of $E_6 \downarrow A_2 \times G_2$, since both A_2 and G_2 have no fundamental fifth-order indices. However, we can resolve the ambiguity due to contragradiency for this case now by uses of mixed sum rules Eq. (1.38b) and/or (1.39e) as well as Eq. (1.29), although we will not go into detail. Note that A_2 (but not G_2) has nontrivial third-order index $D_3(\rho)$.

For other exceptional Lie algebras G_2, F_4, E_7, and E_8, the situation is more simple, since we need not be concerned about the contragradiency ambiguity for these cases because of self-contragadiency of representations of these Lie algebras. In the Appendix, we compute values of $Q_p(\Lambda)$ for some low-dimensional representations of these algebras.

In summary, our branching sum rules are quite useful in determining and checking BR's for exceptional Lie algebras. Many cases we have so far studied have been found to have unique solutions for their BR's without studying structures of L and L_0.

Finally, we would like to make the following remark. The notion of the general Dynkin indices is evidently applicable also for Lie superalgebras. It can be used to determine Kronecker products as well as the branching rule for Lie superalgebras. So far, only the branching rule of the Lie superalgebras U(N/M) appears to have been studied in the literature.[35] Also, the general Dynkin indices are intimately related to the anomaly[36] appearing in gauge theories. Some comments on this problem will be discussed elsewhere.[37]

ACKNOWLEDGMENT

This paper is in part supported by the U.S. Department of Energy under Contract No. DE-AC02-76ER13065.

APPENDIX: TABLE OF INDICES FOR EXCEPTIONAL LIE ALGEBRAS

We compute here $Q_p(\Lambda)$ for some low-dimensional irreducible representation with the highest weight Λ of exceptional Lie algebras. We recall

$$D_p(\Lambda) = Q_p(\Lambda) D_p(\square),$$

where $\square$ is the nontrivial lowest-dimensional representation of these Lie algebras. Using the method described in Sec. III, we find the following.

(1) G_2 $(p = 2,6)$, $\square = \Lambda_2$, and $\rho_0 = \Lambda_1$;

$$Q_p(\Lambda_2) = 1, \quad Q_p(\Lambda_1) = 6 - 2^{p-1},$$
$$Q_p(2\Lambda_2) = 7 + 2^{p-1},$$
$$Q_p(3\Lambda_2) = 27 + 7 \times 2^{p-1} + 3^{p-1},$$
$$Q_p(\Lambda_1 + \Lambda_2) = 8\{6 - 2^{p-1}\},$$
$$Q_p(2\Lambda_1) = 77 - 9 \times 2^{p-1} - 4^{p-1}.$$

(2) F_4 $(p = 2,6,8,12)$, $\square = \Lambda_4$, and $\rho_0 = \Lambda_1$;

$$Q_p(\Lambda_4) = 1, \quad Q_p(2\Lambda_4) = 25 + 2^{p-1},$$
$$Q_p(\Lambda_1) = (274 - 13 \times 2^p + 3^{p-1})/(77 - 2^{p-1}),$$
$$Q_p(\Lambda_2) = (51 - 2^{p-1}) Q_p(\Lambda_1),$$
$$Q_p(\Lambda_3) = (26 - 2^{p-1}) - Q_p(\Lambda_1),$$
$$Q_p(\Lambda_1 + \Lambda_4) = 25 + 2^{p-1} + 27 Q_p(\Lambda_1).$$

(3) E_6 $(p = 2,5,6,8,9,12)$, $\square = \Lambda_1$, and $\rho_0 = \Lambda_6$;

$$Q_p(\Lambda_1) = (-1)^p Q_p(\Lambda_5) = 1,$$
$$Q_p(\Lambda_2) = (-1)^p Q_p(\Lambda_4) = 27 - 2^{p-1},$$
$$Q_p(\Lambda_3) = 27(13 - 2^{p-1}) + 3^{p-1},$$
$$Q_p(2\Lambda_1) = (-1)^p Q_p(2\Lambda_5) = 2^{p-1} + 27 - (-1)^p,$$
$$Q_p(\Lambda_6) = \frac{Q_p(\Lambda_3)}{77 - 2^{p-1}} = \frac{27(13 - 2^{p-1}) + 3^{p-1}}{77 - 2^{p-1}},$$
$$Q_p(\Lambda_1 + \Lambda_2) = 27\{26 - (-1)^p\} - 3^{p-1},$$
$$Q_p(\Lambda_1 + \Lambda_5) = 27\{1 + (-1)^p\} - Q_p(\Lambda_6),$$
$$Q_p(2\Lambda_6) = \{79 + 2^{p-1}\} Q_p(\Lambda_6) - 27\{1 + (-1)^p\},$$
$$Q_p(3\Lambda_1) = 27\{13 - (-1)^p + 2^{p-1}\}$$
$$+ 3^{p-1} + Q_p(\Lambda_6),$$
$$Q_p(\Lambda_1 + \Lambda_6) = 77 - (-1)^p [27$$
$$- 2^{p-1}] + 27 Q_p(\Lambda_6),$$
$$Q_p(\Lambda_4 + \Lambda_6) = 2925 - 351 \times 2^{p-1}$$
$$+ 3^{p+2} - 4^{p-1}.$$

(4) E_7 $(p = 2,6,8,10,12,14,18)$, $\square = \Lambda_6$, and $\rho_0 = \Lambda_1$;

$$Q_p(\Lambda_6) = 1, \quad Q_p(\Lambda_2) = (132 - 2^{p-1}) Q_p(\Lambda_1),$$
$$Q_p(\Lambda_3) = 27\,664 - 1539 \times 2^{p-1}$$
$$+ 56 \times 3^{p-1} - 4^{p-1},$$
$$Q_p(\Lambda_4) = 1539 - 56 \times 2^{p-1} + 3^{p-1},$$
$$Q_p(\Lambda_5) = 56 - 2^{p-1},$$
$$Q_p(\Lambda_1) = Q_p(\Lambda_3)/(8513 - 133 \times 2^{p-1} + 3^{p-1}),$$
$$Q_p(\Lambda_7) = [856 + 2^{p-1} + 2^p Q_p(\Lambda_1)$$
$$+ Q_p(\Lambda_3)]/(856 - 2^{p-1}),$$
$$Q_p(2\Lambda_6) = 56 + 2^{p-1} - Q_p(\Lambda_1),$$
$$Q_p(\Lambda_1 + \Lambda_6) = 132 + 56 Q_p(\Lambda_1) - Q_p(\Lambda_7),$$
$$Q_p(3\Lambda_6) = 1595 + 56 \times 2^{p-1}$$
$$+ 3^{p-1} - Q_p(\Lambda_1 + \Lambda_6),$$
$$Q_p(\Lambda_5 + \Lambda_6) = 3003 - 3^{p-1} - 56 Q_p(\Lambda_1),$$

$$Q_p(2\Lambda_1) = (133 + 2^{p-1})Q_p(\Lambda_1) - Q_p(\Lambda_5).$$

(5) E_8 $(p = 2,8,12,14,18,20,24,30)$, $\square = \rho_0 = \Lambda_1$,

$$Q_p(\Lambda_1) = 1, \quad Q_p(\Lambda_2) = 247 - 2^{p-1},$$

$$Q_p(\Lambda_3) = 30\,133 - 248 \times 2^{p-1} + 3^{p-1}.$$

In deriving these results, we utilized Eq. (3.11) as well as

$$Q_p\left(\begin{array}{|c|}\hline \\ \hline \\ \hline \\ \hline \\ \hline \end{array}\right) = \frac{N(N-1)(N-2)}{6} - \frac{N(N-1)}{2} \times 2^{p-1} + N \times 3^{p-1} - 4^{p-1},$$

for $N = d(\square)$. In spite of fractional expressions for some $Q_p(\Lambda)$, these are really integers except for the cases of $p = 14$ and 18 for E_7. For example, we compute $Q_6(\Lambda_1) = -2$, $Q_8(\Lambda_1) = 10$, $Q_{10}(\Lambda_1) = -2$, $Q_{12}(\Lambda_1) = -30$ and $Q_6(\Lambda_7) = -10$, $Q_8(\Lambda_7) = -82$, $Q_{10}(\Lambda_7) = 230$, $Q_{12}(\Lambda_7) = -2082$, but $Q_{14}(\Lambda_1) = \frac{542}{29}$ and $Q_{18}(\Lambda_1) = -\frac{111\,658}{1229}$. The reason why $Q_p(\Lambda_1)$ and $Q_p(\Lambda_7)$ ($p = 14$ and 18) for E_7 are fractional is unclear.

As applications of our formulas, we discuss the branching rule for $E_7 \downarrow E_6$. Considering the simplest decomposition $\Lambda_6 \rightarrow \Lambda_1 \oplus \Lambda_5 \oplus 2(0)$, we find $\bar{\xi}_p = 2$. Then, the branching sum rule

$$2Q_p(\Lambda) = \sum_j Q_p^{(0)}(\Lambda_j) \quad (p = 2,6,8,12),$$

for $E_7 \downarrow E_6$ can be verified for many cases.

We have computed here only $Q_p(\Lambda_j)$ $(1 \leqslant j \leqslant 3)$ for E_8. The calculations of $Q_p(\Lambda_j)$ $(8 \geqslant j \geqslant 4)$ turn out to be quite complicated and appear to require a considerable effort. This has to wait for further study.

[1]R. C. King, J. Phys. A **8**, 429 (1975), and references quoted therein.

[2]R. C. King, L. Dehuai, and B. G. Wybourne, J. Phys. A **14**, 2509 (1981).

[3]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Methun, London, 1931), see p. 391.

[4]H. Boerner, *Representation of Groups* (North-Holland, Amsterdam, 1970), 2nd ed., pp. 267 and 269.

[5]W. Miller, Pac. J. Math. **16**, 341 (1966).

[6]G. C. Hegerfeldt, J. Math. Phys. **8**, 1195 (1967).

[7]M. L. Whippman, J. Math. Phys. **6**, 1534 (1965).

[8]C. Itzykson and M. Nauenberg, Rev. Mod. Phys. **38**, 95 (1966).

[9]R. C. King, J. Phys. A **8**, 429 (1975).

[10]G. R. E. Black and B. G. Wybourne, J. Phys. A **16**, 2405 (1983).

[11]R. C. King and N. G. I. El-Sharkaway, J. Phys. A **17**, 19 (1984).

[12]G. R. E. Black, R. C. King, and B. G. Wybourne, J. Phys. A **16**, 1555 (1983).

[13]A. Navon and J. Patera, J. Math. Phys. **8**, 489 (1967).

[14]P. H. Butler, Appendix in B. C. Wybourne, *Symmetry Principles and Atomic Spectroscopy* (Wiley–Interscience, New York, 1970).

[15]A. van Daele, J. Math. Phys. **11**, 3275 (1970).

[16]B. G. Wybourne and M. J. Bowick, Aust. J. Phys. **30**, 259 (1977).

[17]B. G. Wybourne, Aust. J. Phys. **32**, 417 (1979).

[18]G. Feldman, T. Fulton, and P. T. Matthews, J. Math. Phys. **25**, 1230 (1984).

[19]W. G. McKay and J. Patera, *Tables of Dimensions, Indices, and Branching Rules for Representations of Simple Lie Algebras* (Dekker, New York, 1981). We have, however, changed the notation $I^{(2p)}$ there to $l_{2p}(\rho)$ here.

[20]B. G. Wybourne, J. Phys. A **17**, 1397 (1984).

[21]G. Bélanger, J. Phys. A: Math. Gen. **16**, 3421 (1983).

[22]S. Okubo and J. Patera, J. Math. Phys. **25**, 219 (1984).

[23]J. Diximer, *Enveloping Algebras* (North-Holland, Amsterdam, 1977).

[24]S. Okubo, J. Math. Phys. **23**, 8 (1982).

[25]E. B. Dynkin, Mat. USSR Sb. **30**, 349 (1952) [Am. Math. Soc. Trans. Ser. 2 **6**, 111 (1957)].

[26]N. Bourbaki, *Lie Groups and Lie Algebras, Part I* (Addison–Wesley, Reading, MA, 1975).

[27]S. Okubo, J. Math. Phys. **20**, 586 (1979).

[28]S. Okubo and J. Patera, J. Math. Phys. **24**, 2722 (1983).

[29]J. Patera, R. T. Sharp, and P. Winternitz, J. Math. Phys. **17**, 1972 (1976); Erratum **18**, 1519 (1977).

[30]See, e.g., N. Jacobson, *Lie Algebras* (Interscience, New York, 1962).

[31]Y. Tosa, R. E. Marshak, and S. Okubo, Phys. Rev. D **27**, 444 (1983).

[32]See, e.g., J. E. Humphrey, *Introduction to Lie Algebras and Representation Theory* (Springer, Berlin, 1972).

[33]See, e.g., M. Hammermesh, *Group Theory and its Application to Physical Problems* (Addison–Wesley, Reading, MA, 1962).

[34]J. Tits, *Lecture Notes in Mathematics*, Vol. 40 (Springer, New York, 1967); M. L. Mehta, J. Math. Phys. 7, 1824 (1966); A. K. Bose and J. Patera, *ibid.* **11**, 2231 (1970).

[35]B. G. Wybourne, J. Phys. A **17**, 1573 (1984).

[36]J. Patera and R. T. Sharp, J. Math. Phys. **22**, 2352 (1981); P. H. Frampton and T. W. Kephart, Phys. Rev. Lett. **50**, 1343, 1347 (1983); L. Alvarez-Gaume and E. Witten, Nucl. Phys. B **234**, 269 (1983); M. B. Green and J. H. Schwarz, Phys. Lett. B **149**, 117 (1984).

[37]S. Okubo and J. Patera, Phys. Rev. D **31**, 2669 (1985).

Evaluation of the self-energy of a droplet interacting via a Yukawa force

Hans J. Krappe
Bereich Kern- und Strahlenphysik, Hahn-Meitner-Institut für Kernforschung Berlin, 1000 Berlin 39-Wannsee, West Germany

(Received 1 February 1984; accepted for publication 22 February 1985)

The binding energy of a homogeneously charged, classical droplet is considered, whose charge elements interact via a Yukawa force. The six-dimensional self-energy integral is reduced to a three-dimensional integral for axially symmetric, but otherwise arbitrarily shaped, droplets. This integral is brought into a form particularly convenient for numerical calculations. Two integrals involving products of Bessel functions are evaluated, which are either not listed in standard tables of integrals or given in an erroneous form.

I. INTRODUCTION

In the nuclear liquid-drop model one has to calculate the binding energy of a droplet interacting via a short-range force. Using in particular the Yukawa potential

$$Y(r) = (4\pi a^2 r)^{-1} e^{-r/a}$$

for the interaction, the six-dimensional self-energy integral

$$I(a) = \int_V Y(|\mathbf{r} - \mathbf{r}'|) d^3\mathbf{r}\, d^3\mathbf{r}' \tag{1}$$

has to be evaluated, where the integration extends over the volume V of the nuclear droplet in both variables.[1] Accounting for the diffuse nuclear surface, the zero-point motion of the nucleons, and the Pauli principle in the Thomas–Fermi approximation, the surface energy can be shown[2] to have the form

$$E_{\text{surf}} = -2\gamma \frac{dI(a)}{da},$$

where γ is the surface energy constant. The same formulas hold for any liquid interacting via a saturating, short-range force.

The Coulomb self-energy is also obtained from $I(a)$. In the limit $a \to \infty$ the Coulomb energy is given by

$$2\pi\rho_0^2 \lim_{a\to\infty} a^2 I(a) = \frac{\rho_0^2}{2} \int_V \frac{1}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r}\, d^3\mathbf{r}', \tag{2}$$

ρ_0 being the electric charge density.

In nuclear physics one is mostly interested in the binding energy of axially symmetric droplets. In the next section we show that Fourier transformation techniques can be used to transform $I(a)$ into a three-dimensional integral. This requires some integrals over products of Bessel functions, which are evaluated in the last two sections.

II. REDUCTION OF THE SIX-DIMENSIONAL SELF-ENERGY INTEGRAL TO A THREE-DIMENSIONAL INTEGRAL

Introducing the three-dimensional step function θ, which shall be 1 inside the nuclear surface and 0 outside, the integral $I(a)$ may be written as

$$\int_V Y(|\mathbf{r} - \mathbf{r}'|) d^3\mathbf{r}\, d^3\mathbf{r}' = (2\pi)^{-3} \int F[\theta] F[Y] F[\theta]^* d^3q, \tag{3}$$

where the Fourier transforms

$$F[\theta] = \int \theta(\mathbf{r}) e^{i\mathbf{qr}} d\mathbf{r}, \tag{4}$$

$$F[Y] = \int Y(r) e^{i\mathbf{qr}} d^3\mathbf{r} = (1 + a^2 q^2)^{-1} \tag{5}$$

have been introduced. In cylindrical coordinates the surface of an axially symmetric droplet may be generated by turning the shape function $\rho = P(z)$ around the z axis. The Fourier transforms (4) and (5) are then given in terms of the shape function $P(z)$ by

$$F[\theta] = 2\pi \int_{-\infty}^{\infty} e^{izq_z} P(z) J_1(q_\perp P(z)) q_\perp^{-1} dz, \tag{6}$$

$$F[Y] = [1 + a^2(q_z^2 + q_\perp^2)]^{-1}, \tag{7}$$

where q_z and $q_\perp$ denote the components of $\mathbf{q}$ in the direction of the symmetry axis and perpendicular to it. Inserting (6) and (7) into (3) the expression

$$\int_V Y(|\mathbf{r} - \mathbf{r}'|) d^3r\, d^3r' = \int_{-\infty}^{\infty} P(z) \int_{-\infty}^{\infty} P(z') \int_{-\infty}^{\infty} e^{iq_z(z-z')} \times \int_0^\infty \frac{J_1(q_\perp P(z)) J_1(q_\perp P(z'))}{q_\perp [1 + a^2(q_z^2 + q_\perp^2)]} dq_\perp\, dq_z\, dz'\, dz \tag{8}$$

is obtained after reordering the sequence of integrations. The innermost integral will be shown in the next section to have the value

$$\int_0^\infty \frac{J_1(qP) J_1(qP')}{q(b^2 + q^2)} dq = -b^{-2} I_1(bP_<) K_1(bP_>) + \tfrac{1}{2} b^{-2} P_< P_>^{-1}, \tag{9}$$

where I_1 and K_1 are modified Bessel and Hankel functions of order 1, respectively,[3] and $P_<$ is the smaller and $P_>$ the larger of the two numbers P and P'. The last term on the right-hand side (rhs) of (9) is missing in Gradshteyn–Ryzhik.[4] There, reference is made to Bateman's collection of integrals,[5] where the result is already given in the erroneous form. Note that the integral on the left-hand side (lhs) of (9) is not a special case of Hankel's integral, as given in Sec. 13.53 of Ref. 6,

 0022-2488/85/092138-03$02.50

$$\frac{1}{2\pi i}\int_0^\infty [\mathscr{C}_\mu(xP)H_\nu^{(1)}(xP') - e^{\rho\pi i}\mathscr{C}_\mu(xPe^{\pi i})$$
$$\times H_\nu^{(1)}(xP'e^{\pi i})]\,\frac{x^{\rho-1}}{(x^2-r^2)^{m+1}}\,dx$$
$$= \frac{1}{2^{m+1}m!}\left(\frac{d}{r\,dr}\right)^m [r^{\rho-2}\mathscr{C}_\mu(rP_<)H_\nu^{(1)}(rP_>)]\,,$$

for $\mathscr{C}_\mu(z) = J_\mu(z)$, $\mu = \nu = 1$, $m = 0$, $r = ib$, since the formula is only valid if

$$|\mathscr{R}\nu| - \mathscr{R}\mu < \mathscr{R}\rho\,.$$

The latter condition excludes the case $\rho = 0$ which is of interest here. Only for $\rho > 0$ would there be no second term in (9). That the condition is in fact necessary is easily seen from the argument presented in Sec. 13.53 of Ref. 6. Unfortunately, it is not properly taken into account in Refs. 4 and 5.

That the second term on the rhs of (9) is required for $\rho = 0$ is easily checked in the limit $b = 1$, $P = P' \to 0$. If for real positive z we define

$$A = \max_z \,[(J_1(z))^2/z]\,,$$

the integral on the lhs of (9) can be estimated by

$$0 \leqslant P\int_0^\infty \frac{(J_1(qP))^2}{qP(1+q^2)}\,dq \leqslant PA\int_0^\infty \frac{dq}{1+q^2} = \frac{\pi}{2}PA\,,$$

which tends to zero for $P \to 0$. On the other hand,

$$\lim_{P\to 0} I_1(P)K_1(P) = \tfrac{1}{2}\,,$$

so that (9) is satisfied in this limit only with the inclusion of the second term on the lhs.

Inserting (9) into (8) the q_z integration can be brought into a considerably more convenient form by using the identity

$$\int_{-\infty}^\infty e^{iq_z(z-z')}\,b^{-2}[2I_1(bP_<)K_1(bP_>) - P_<P_>^{-1}]\,dz$$
$$= -2aPP'\int_0^\pi \sin^2\phi R^{-2}\left(\exp\left(\frac{-|z-z'|}{a}\right) - \exp\left[-\frac{\sqrt{(z-z')^2+R^2}}{a}\right]\right)d\phi\,, \tag{10}$$

with

$$b^2 = a^{-2} + q_z^2\,,$$
$$R^2 = P^2(z) + P^2(z') - 2P(z)P(z')\cos\phi\,.$$

This result will be derived in the third section.

Inserting (10) into (8) leads to a reduction of the original six-dimensional integral (1) to a three-dimensional one:

$$\int_V Y(|\mathbf{r}-\mathbf{r}'|)d^3\mathbf{r}\,d^3\mathbf{r}'$$
$$= a^{-1}\int_{-\infty}^\infty P^2(z)\int_{-\infty}^\infty P^2(z')\int_0^\pi \sin^2\phi$$
$$\times R^{-2}(\exp(-|z-z'|/a)$$
$$-\exp[-\sqrt{(z-z')^2+R^2/a}])\,d\phi\,dz'\,dz\,.$$

For general shape functions $P(z)$ this has to be evaluated numerically. The singular behavior of the integrand for $z = z'$ suggests the introduction of new variables t and t' by

$$t = (z - z_{\min})/z_0,\quad tt' = (z' - z_{\min})/z_0,\quad 0 \leqslant t,t' \leqslant 1\,,$$

with

$$z_0 = z_{\max} - z_{\min}\,,$$

$z_{\max}$ and $z_{\min}$ denoting the largest and the smallest zero of $P(z)$ (the end points of the droplet), respectively. This transformation, introduced by Lawrence[7] for a similar purpose, maps the line of singularity $z = z'$ onto the boundary line $t' = 1$, which is more convenient for numerical purposes. The resulting expression for the self-energy integral is

$$I(a) = 2\frac{z_0^2}{a}\int_0^1 dt\,tP^2(z)\int_0^1 dt'\,P^2(z')\int_0^\pi d\phi$$
$$\times \sin^2\phi R^{-2}G(t,t',\phi)\,,$$

with

$$G(t,t',\phi) = \exp[-(z_0/a)t(1-t')]$$
$$-\exp\left[-a^{-1}\sqrt{z_0^2t^2(1-t')^2+R^2}\right]\,,$$
$$z = z_0t + z_{\min}\,,\quad z' = z_0tt' + z_{\min}\,.$$

Using the relation

$$\lim_{a\to\infty} aG = \sqrt{z_0^2t^2(1-t')^2+R^2} - z_0t(1-t')\,,$$

the Coulomb limit (2) is readily obtained from this result

$$E_{\text{Coul}} = 4\pi\rho_0^2 z_0^3\int_0^1 dt\,tP^2(z)\int_0^1 dt'\,P^2(z')\int_0^\pi d\phi$$
$$\times \sin^2\phi R^{-2}\{\sqrt{t^2(1-t')^2+R^2/z_0^2} - t(1-t')\}\,.$$

This expression for the Coulomb energy agrees with the one obtained by Lawrence[7] after use has been made of the identity

$$(z_0/R)^2[\sqrt{t^2(1-t')^2+(R/z_0)^2} - t(1-t')]$$
$$= [t(1-t') + \sqrt{t^2(1-t')^2+(R/z_0)^2}]^{-1}\,.$$

III. EVALUATION OF THE INTEGRAL $\int_0^\infty q^{-1}J_n(qP) \times J_n(qP')(b^2+q^2)^{-1}\,dq$

Writing

$$J_n(qP)\,J_n(qP')$$
$$= \tfrac{1}{2}J_n(qP_<)(H_n^{(1)}(qP_>) + H_n^{(2)}(qP_>))$$

and choosing the branch cut for the Hankel function along the negative real axis, we first consider the integral

$$\int_{A+B+C}\frac{J_n(qP_<)H_n^{(1)}(qP_>)}{q(b^2+q^2)}\,dq$$
$$= -i\pi b^{-2}J_n(ibP_<)H_n^{(1)}(ibP_>)$$
$$= -2b^{-2}I_n(bP_<)K_n(bP_>)\,, \tag{11}$$

for integer $n > 0$ over the closed contour A + B + C in the complex q plane (cf. Fig. 1 for the definition of integration paths) and

$$\int_{A+D+E+F}\frac{J_n(qP_<)H_n^{(2)}(qP_>)}{q(b^2+q^2)}\,dq$$
$$= i\pi b^{-2}J_n(-ibP_<)H_n^{(2)}(-ibP_>)$$
$$= -2b^{-2}I_n(bP_<)K_n(bP_>)\,, \tag{12}$$

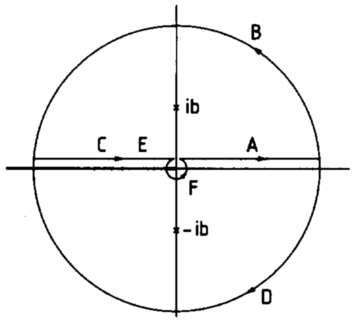


FIG. 1. Integration path used in (11) and (12). The branch cut of the integrand is along the negative real axis. There are two closed contours C + A + B and E + F + A + D of which the part E is in the second sheet.

over the closed contour A + D + E + F. Since

$$H_n^{(2)}(xe^{-\pi i}) = H_n^{(2)}(xe^{\pi i}) - 4J_n(xe^{\pi i}),$$

we have

$$\int_E \frac{J_n(qP_<)H_n^{(2)}(qP_>)}{q(b^2+q^2)}\,dq = \int_C \frac{J_n(qP_<)(H_n^{(2)}(qP_>) - 4J_n(qP_>))}{q(b^2+q^2)}\,dq. \quad (13)$$

For $P_< < P_>$ the asymptotic behavior of the Bessel functions for large arguments implies a vanishing contribution to the integrals (11) and (12) from the circular parts B and D of the contour when $|q| \to \infty$. At the origin

$$J_n(z) = (1/n!)(z/2)^n + \mathscr{O}(z^{n+2}),$$

$$H_n^{(2)}(z) = -\frac{2i}{\pi}J_n(z)\ln\frac{z}{2} + \frac{i}{\pi}(n-1)!\left(\frac{2}{z}\right)^n + \mathscr{O}(z^{2-n}).$$

Therefore

$$\int_F \frac{J_n(qP_<)H_n^{(2)}(qP_>)}{q(b^2+q^2)}\,dq = -\frac{2}{nb^2}\left(\frac{P_<}{P_>}\right)^n, \quad (14)$$

for $|q| \to 0$, $n > 0$. Adding (11) and (12) and using (13) and (14) yields finally

$$4\int_0^\infty \frac{J_n(qP_<)J_n(qP_>)}{q(b^2+q^2)}\,dq - \frac{2}{nb^2}\left(\frac{P_<}{P_>}\right)^n = -4b^{-2}I_n(bP_<)K_n(bP_>),$$

for integer $n > 0$.

IV. TRANSFORMATION OF THE INTEGRAL $\int_{-\infty}^{\infty} e^{i\kappa\zeta}I_1(bP_<)K_1(bP_>)b^{-2}\,d\kappa$

For the product of two modified Bessel functions an integral representation can be obtained from the summation formula[6]

$$\frac{K_n(bR)}{(bR)^n} = 2^n\Gamma(n)\sum_{m=0}^{\infty}(n+m)\frac{K_{n+m}(bP_>)}{(bP_>)^n} \times [I_{n+m}(bP_<)/(bP_<)^n]\,C_m^n(\cos\phi), \quad (15)$$

where

$$R^2 = P_<^2 + P_>^2 - 2P_< P_> \cos\phi,$$

and the $C_m^n(x)$ are Gegenbauer polynomials. Multiplying (15) with $C_0^1 \sin^2\phi$ and taking $n = 1$, one obtains after integrating over ϕ

$$\frac{1}{\pi}\int_0^\pi \frac{K_1(bR)}{bR}\sin^2\phi\,d\phi = \frac{K_1(bP_>)}{bP_>}\frac{I_1(bP_<)}{bP_<}.$$

Using the Fourier transform[8]

$$\frac{1}{\pi}\int_{-\infty}^{\infty} e^{i\kappa\zeta}\frac{K_1(bR)}{bR}\,d\kappa = \frac{a}{R^2}\exp\left(\frac{-\sqrt{\zeta^2+R^2}}{a}\right),$$

with $b^2 = \kappa^2 + a^{-2}$, therefore leads to

$$\int_{-\infty}^{\infty} e^{i\kappa\zeta}I_1(bP_<)K_1(bP_>)b^{-2}\,d\kappa = aP_< P_>\int_0^\pi \frac{\sin^2\phi}{R^2}\exp\left(\frac{-\sqrt{\zeta^2+R^2}}{a}\right)d\phi.$$

Since

$$\frac{1}{\pi}\int_{-\infty}^{\infty} e^{i\kappa\zeta}b^{-2}\,d\kappa = a\exp(-|\zeta|/a)$$

and[9]

$$\frac{P_<}{P_>} = P_< P_> \frac{2}{\pi}\int_0^\pi \frac{\sin^2\phi}{R^2}\,d\phi,$$

the identity (10) is seen to hold.

[1]H. J. Krappe and J. R. Nix, in *Proceedings of the Third IAEA Symposium on the Physics and Chemistry of Fission*, Rochester, New York, 1973 (IAEA, Vienna, 1974), Vol. 1, p. 159.

[2]H. J. Krappe, J. R. Nix, and A. J. Sierk, Phys. Rev. C **20**, 992 (1979); H. J. Krappe, in *Nuclear Physics, Proceedings of the Nuclear Physics Workshop of the ICTP, Trieste 1981*, edited by C. H. Dasso (North-Holland, Amsterdam, 1982), p. 473.

[3]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1965), 4th translated ed., p. 961.

[4]Reference 3, p. 679, Eq. 6.541 (2).

[5]A. Erdelyi, *Tables of Integral Transforms* (McGraw–Hill, New York, 1954), Vol. 2, p. 49.

[6]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U. P., Cambridge, England, 1966), Chap. 11, Sec. 41, Eq. 8.

[7]J. N. P. Lawrence, Phys. Rev. B **139**, 1227 (1965).

[8]F. Oberhettinger, *Tabellen zur Fourier Transformation* (Springer, Berlin, 1957), p. 93.

[9]Reference 3, p. 366, Eq. 3.613 (3).

On the phase retrieval problem in two dimensions

I. Sabba Stefanescu
Institut für Theoretische Kernphysik, Universität Karlsruhe, Karlsruhe, West Germany

(Received 9 August 1984; accepted for publication 8 March 1985)

The paper contains a discussion of the phase retrieval problem in two dimensions and proposes criteria to select those resolutions of the discrete ambiguity of the zero trajectories which are compatible with the analyticity in two variables of the scattered field.

I. INTRODUCTION

It has been realized over the last few years that the problem of phase retrieval from a given intensity distribution has features that are qualitatively different in higher dimensions ($d \geqslant 2$), as compared to the one-dimensional situation. Loosely speaking, it appears much easier to achieve uniqueness of the reconstructed phase (up to an overall phase factor) in higher dimensions.

The reasons for this occurrence may be most easily understood in the artificial problem of the reconstruction of a polynomial $P(u,v)$ of degree (M,N) with respect to u and v from knowledge of its modulus in the Re u-Re v plane. If $P(u,v)$ may be decomposed into Q prime factors $P_i(u,v)$, then the modulus squared of $P(u,v)$ in the Re u-Re v plane

$$M(u,v) = P(u,v)P^*(u^*,v^*) = \prod_{i=1}^{Q} P_i(u,v)P_i^*(u^*,v^*) \tag{1.1}$$

contains $2Q$ such factors, and the ambiguity in the reconstruction[1,2] of $P(u,v)$ knowing $M(u,v)$ is clearly 2^Q (if the Q factors are distinct). The remarkable feature is that the number Q bears in general no relation to the degree of $P(u,v)$; it depends on the specific polynomial under consideration. This is in contrast to the one-dimensional case[3,4] where any polynomial $P(u)$ of degree N may always be written as a product of N factors $(u - u_i)$ and the ambiguity is invariably 2^N.

The fact that in two dimensions there exists a much richer variety of prime factors than in one dimension leads to several problems: (i) given a polynomial $M(u,v)$, positive in the Re u-Re v plane and of degree $2M,2N$, one needs a method to decide whether it may be written as a product of pairs of prime factors, complex conjugate to each other. In two dimensions, there exist irreducible positive polynomials [e.g., $P(u,v) = u^2 + v^2 + 1$] and thus only a subset of the positive polynomials makes up admissible intensity distributions. (ii) The extension of this question is that of finding the ambiguity in the phase determination from an admissible intensity distribution $M(u,v)$, given numerically in the Re u-Re v plane, i.e., to determine the number of its prime factors. (iii) If $M(u,v)$ is admissible, one wishes to construct all polynomials $P(u,v)$ so that (1.1) is obeyed. (iv) The coefficients of a solution $P(u,v)$ of the phase problem are determined from data only within errors; every polynomial $P(u,v)$ of degree (M,N) may be regarded as an element of $\mathbb{R}^{(M+1)(N+1)}$ with the Euclidean metric and the question arises whether all polynomials $\widetilde{P}(u,v)$ in a sufficiently small ball around $P(u,v)$

$$\|\widetilde{P} - P\|_{\mathbb{R}^{(M+1)(N+1)}} < \epsilon \tag{1.2}$$

have the same number of prime factors. This is in fact wrong, since reducibility may be easily destroyed (as an example, $u^2 + v^2$ is reducible over $\mathbb{C}$, but $u^2 + v^2 + \epsilon$ is not), and the problem is to describe the situation in precise terms.

A solution to this problem is provided by Ref. 5, where it is shown that the set of reducible polynomials is of measure zero in $\mathbb{R}^{(M+1)(N+1)}$ and that, as a consequence of the method of proof, for each irreducible polynomial $P(u,v)$, there exists a neighborhood $\mathscr{U}$ of it, such that each element of $\mathscr{U}$ is irreducible. In fact, one can prove, following Ref. 5, that each polynomial of degree (M,N) with Q prime factors has a neighborhood $\mathscr{U}_1$ in $\mathbb{R}^{(M+1)(N+1)}$ such that each element of $\mathscr{U}_1$ has at most Q factors.

The purpose of this paper is to explore and partly formulate answers to the questions above in the framework of entire functions of exponential type in two variables. These functions occur naturally in the description of scattered fields in electromagnetic theory (see Sec. II); we shall call a function in this class "truly entire" if it is not equal to a polynomial, and it is with such functions that we shall be concerned. The solutions to problems (i)–(iii) for polynomials will appear in the text (Sec. IV), however, as special cases. It may be surprising that it will turn out that it is not yet possible, in general, for truly entire functions, to relate the extent of the ambiguity to the number of prime factors (as assumed by some authors[2]).

In Sec. II of this paper, we recall and render more precise known results concerning the solution of the phase retrieval problem in the framework of entire functions; we also give an analysis of the stability of the determination of the ambiguities, as discussed above [under (iv)].

In Sec. III, we give an answer (which might still seem impractical) to questions (i)–(iii) above for entire functions, using the zero trajectories of the modulus distribution. In two variables, there appears an "ambiguity of trajectories" instead of one of zeros, as in the one-variable case, but not every resolution of this ambiguity turns out to be compatible with two-variable analyticity (see also Ref. 2).

In Sec. IV, we give a quick approximate method of grouping trajectories into irreducible factors. This method has already been used in the analysis of scattering data in high-energy physics, in order to discard solutions that are incompatible with local analyticity in two variables.[6,7] Further, we apply the results of the previous sections to the case of polynomials, as used by Bruck and Sodin.[1]

Appendix A contains the proofs of some statements in the text, Appendix B proves explicitly that reducibility is unstable for zero sets of entire functions, and Appendix C contains the construction of a truly entire irreducible (i.e., not admissible) intensity distribution.

 0022-2488/85/092141-20$02.50

II. THE AMBIGUITY OF THE SOLUTIONS OF THE PHASE RETRIEVAL PROBLEM

We consider the field $F(u,v)$ in the plane at infinity, obtained from the scattering of a monochromatic parallel beam of light, normally incident on an aperture with finite extent $\mathscr{D}$. In the Kirchhoff approximation,[8] $F(u,v)$ is related to the wave function $\psi(x,y)$ in the aperture plane by

$$F(u,v) = \text{const} \int_{\mathscr{D}}\int \exp i(ux+vy)\psi(x,y)dx\,dy\,,$$
$$u = kp, \quad v = kq, \quad k = 2\pi/\lambda\,, \tag{2.1}$$

where λ is the wavelength and p,q are the direction cosines of the scattered wave with respect to orthogonal x, y axes in the aperture plane and a z axis normal to it. One verifies that $F(u,v)$ is an entire function of exponential type in u and v, with a square-integrable modulus over the Re u-Re v plane. The rate of increase of $F(u,v)$ in an imaginary direction (λ_1,λ_2) [i.e., $\sup_{\alpha_1,\alpha_2} \overline{\lim}(\log|F(\alpha_1 - i\lambda_1 r, \alpha_2 - i\lambda_2 r)|/r)$] is determined by the convex hull of the aperture $\mathscr{D}$ (Ref. 9).

A generalization of the Paley–Wiener theorem is true[9] and states that any entire function of exponential type, with square-integrable modulus in the Re u-Re v plane, is the Fourier transform of a square-integrable function with bounded support. The convex hull of the support can be read off the growth properties of $F(u,v)$ in imaginary directions.

As a consequence of this theorem, the phase retrieval problem is that of finding constraints on the phase of an entire function of exponential type $F(u,v)$ given its square-integrable modulus on the subset Im $u = 0 \otimes$ Im $v = 0$.

A central part is played in the study of this problem by the set $\mathscr{Z}$ of points $(u,v)\in\mathbb{C}^2$ where the entire function $F(u,v)$ vanishes. This has been pointed out also in Ref. 2.

The set $\mathscr{Z}$ may be written as a union over several irreducible analytic sets $\mathscr{M}_i$ (Ref. 10, p. 265). An irreducible analytic set can be defined as follows: consider a point (u,v) on $\mathscr{Z}$, so that the equation $F(u,v) = 0$ can be solved with respect to v, say, in a neighborhood $\mathscr{U}$ of $u{:}v = v_0(u)$; the set of all points $(u,v(u))\in\mathbb{C}^2$, where $v(u)$ is obtained by the analytic continuation of $v_0(u)$ in the u plane along all possible paths, is clearly included in $\mathscr{Z}$, and makes up an irreducible component of it.

For each irreducible component $\mathscr{M}_i$ of $\mathscr{Z}$, one can construct an entire function $G_i(u,v)$, vanishing only on $\mathscr{M}_i$ and nowhere else[10] (Cousin's second theorem). This leads to a decomposition of $F(u,v)$ into "prime factors"

$$F(u,v) = \prod_{i=1}^{n} G_i(u,v)\,, \tag{2.2}$$

where n may be infinite. Each factor $G_i(u,v)$ is defined up to a nonvanishing entire function. In an obvious manner, one may introduce multiplicities p_i for each manifold $\mathscr{M}_i$ and corresponding powers for each $G_i(u,v)$ in (2.2). One might suspect that, if F is of exponential type, then each of the $G_i(u,v)$ can be chosen to be of exponential type. This is far from obvious and will be discussed below.

Similarly to the one-variable case, the relevance of the set of zeros of $F(u,v)$ is seen if one constructs from the available modulus distribution the entire function

$$M(u,v) = F(u,v)F^*(u^*,v^*)\,. \tag{2.3}$$

For (u,v) on Im $u = 0 \otimes$ Im $v = 0$, the (real) values of $M(u,v)$ are known. They can be used to construct $M(u,v)$ everywhere, e.g., by means of the Lagrange interpolation formula (see Sec. III). In particular, one can find in principle the set of zeros $\mathscr{Z}_M$ of $M(u,v)$. It contains, apart from $\mathscr{Z}$, the set $\mathscr{Z}^*$ [$(u,v)\in\mathscr{Z}^*$ if $(u^*,v^*)\in\mathscr{Z}$] and nothing else. If the irreducible components of $\mathscr{Z}$ are $\mathscr{M}_i$, then those of $\mathscr{Z}^*$ are $\mathscr{M}_i^*$ (which are also analytic). Now, knowing the $\mathscr{M}_i$'s, we can in principle use Cousin's method to construct the corresponding G_i's and thus F. We expect thus that the only ambiguities of the phase retrieval problem arise from the possible replacement of $G_i(u,v)$ by $G_i^*(u^*,v^*)$, which vanishes on $\mathscr{M}_i^*$.

We agree to call two solutions distinct if they do not differ by a constant phase factor $\exp(i\phi_0)$; we can then make the above statements precise through the following.

Proposition 1: If two solutions $F_1(u,v)$, $F_2(u,v)$ of the phase problem have the same set of zeros, then their Fourier transforms differ at most by a translation.

Proof: Let F_1,F_2 be the two distinct solutions; their quotient $R(u,v) = F_2(u,v)/F_1(u,v)$ has unit modulus for u,v real and is holomorphic and free of zeros in all of $\mathbb{C}^2$. It is therefore an entire function which may be written as $\exp(iP(u,v))$ with $P(u,v)$ a real analytic entire function.

Now, at each fixed u, we have the following (i) $\ln|F_2(u,v)|$ is bounded from above by const $|v|$; (ii) for any given $\epsilon > 0$ and $\sigma > 0$, there exists an r_0 such that, for $|v| > r_0$, $\ln|F_1(u,v)| > -|v|^{1+\epsilon}$, if v lies outside circles centered at the zeros v_n of $F_1(u,v)$ and of radius larger than $1/|v_n|^\sigma$ (see Ref. 11, Lemma 2.6.18).

Consequently, at each u, for $|v|$ sufficiently large, and outside those circles,

$$\ln|R(u,v)| < k\,|v|^{1+\epsilon}\,. \tag{2.4}$$

We can choose σ (e.g., $\sigma > 1$) such that, for any v with $|v|$ sufficiently large [$|v| > v_0(\sigma,\epsilon,u)$], there exists a circle of radius $k_v|v|$ with, e.g., $2 < k_v < 3$, so that (2.4) is valid on that circle. Applying Caratheodory's inequality (Ref. 11, p. 3) to the function $iP(u,v) \equiv \ln R(u,v)$ whose real part is bounded by (2.4) on the boundary of disks of radius $k_v|v|, |v| > v_0$, we obtain

$$|P(u,v)| < \text{const}|v|^{1+\epsilon}\,, \tag{2.5}$$

for all $|v|$ sufficiently large. Choosing $\epsilon < 1$, (2.5) implies that $P(u,v)$ is a polynomial of order at most unity, at each fixed u.

The same reasoning can be done at fixed v, and it follows that

$$R(u,v) = \exp[i(\gamma u + \delta v)]\,, \tag{2.6}$$

with γ,δ real. Proposition 1 follows then from a known property of Fourier transforms.

We can thus part the solutions $F(u,v)$ of the phase problem into equivalence classes modulo translations.

Proposition 2: The solutions in two different classes differ by the replacement of at least one irreducible zero set $\mathscr{M}_i$ by its complex conjugate $\mathscr{M}_i^*$.

This may be regarded as rather obvious, since it is equivalent to saying that irreducible sets cannot be "broken" if they are to be zero sets of some entire function. Nevertheless, we prove this in detail below.

Proof: According to Proposition 1, the sets of zeros of the two solutions $F_1(u,v)$, $F_2(u,v)$ cannot be identical. As a consequence of Weierstrass' preparation theorem (Ref. 10, p. 86), there must exist a point (u_0,v_0) and a neighborhood $\mathscr{U}_0\times\mathscr{V}_0$ of it, so that the sets of zeros of $F_1(u,v)$ and $F_2(u,v)$ in $\mathscr{U}_0\times\mathscr{V}_0$ are described by different pseudopolynomials (one of them may be a constant). We may assume $F_1(u_0,v_0)=0$ and write, for $(u,v)\in\mathscr{U}_0\times\mathscr{V}_0$,

$$\begin{aligned} F_1(u,v) &= W_1(u,v)\Omega_1(u,v),\\ F_2(u,v) &= W_2(u,v)\Omega_2(u,v), \end{aligned} \tag{2.7}$$

with

$$\begin{aligned} W_i(u,v) &= (v-v_0)^{n_i} + a_{1i}(u)(v-v_0)^{n_i-1} + \cdots\\ &\quad + a_{ni}(u), \quad i=1,2,\\ a_{ki}(u_0) &= 0, \quad n_1\geqslant 1, \quad n_2\geqslant 0, \end{aligned} \tag{2.8}$$

$a_{ki}(u)$ holomorphic in $\mathscr{U}_0$, and $\Omega_i(u,v)$ nonvanishing in $\mathscr{U}_0\times\mathscr{V}_0$.

We choose the indices 1 and 2 so that $W_2(u,v)$ may be a divisor of $W_1(u,v)$, but not conversely. There exists then an open set $\mathscr{O}\subset\mathscr{U}_0\times\mathscr{V}_0$, where $R(u,v)=F_1(u,v)/F_2(u,v)$ vanishes at all points where at least one of the irreducible factors (see Ref. 10, p. 104) $W_1'(u,v)$ of $W_1(u,v)$ vanishes [with the exception of at most a finite number of points where $W_1'(u,v)=0$].

However, $R(u,v)$ obeys on the plane $\operatorname{Im} u=0\otimes\operatorname{Im} v=0$ the identify

$$R(u,v)R^*(u^*,v^*)=1. \tag{2.9}$$

Since this plane is nonanalytic (Ref. 12, p. 93), Eq. (2.9) holds throughout $\mathbb{C}^2$. In particular, if $(u_1,v_1)\in\mathscr{O}$ and $R(u_1,v_1)=0$, it follows from (2.9) that $R(u_1^*,v_1^*)$ cannot be finite, i.e., $F_2(u_1^*,v_1^*)=0$. Thus $F_2(u,v)=0$ on the set of points (u,v) fulfilling $W_1'(u^*,v^*)=0$. Let $v=v(u)$ be the solution of $W_1'(u,v)=0$ in the neighborhood of a point in $\mathscr{O}$, where $\partial W_1'/\partial v\neq 0$ (such points exist). Then, for all analytic continuations of $v(u)$, $F_1(u,v(u))=0$. This defines the irreducible set $\mathscr{M}_1$ and it follows from the above that F_2 must vanish on $\mathscr{M}_1^*$.

If $F_2(u,v)$ does not vanish on $\mathscr{M}_1$, the proof is finished, since F_1 and F_2 differ by the "reflection" of the irreducible set $\mathscr{M}_1$. If F_2 vanishes on $\mathscr{M}_1$, then, by assumption, $F_1(u,v)$ vanishes with a higher order, so that $R(u,v)=0$, for $(u,v)\in\mathscr{M}_1$. Then, however, (2.9) implies that $F_2(u,v)$ vanishes on $\mathscr{M}_1^*$ with an order higher than F_1, which proves the assertion completely.

As a consequence, if $\mathscr{Z}=\cup_{i=1}^n\mathscr{M}_i$ is the set of zeros of a solution $F_0(u,v)$ of the phase problem, there are at most 2^n distinct solutions of the problem. One would like to show that, as in the case of polynomials, there exist precisely 2^n such solutions. Such a statement is difficult to prove because the Cousin construction leading to Eq. (1.2) cannot ensure that the solutions constructed with the reflected manifolds are indeed of exponential type. It is *a priori* not even clear that they are of order unity.

Using a generalization of the Weierstrass product, due to Lelong,[13] it is possible to answer this last point in the affirmative.

Proposition 3: To each set $\mathscr{Z}'$ obtained from $\mathscr{Z}$ by the reflection of an arbitrary number of manifolds, there exists an entire function of order unity $F(u,v)$ vanishing on $\mathscr{Z}'$ and nowhere else, and with the same modulus as $F_0(u,v)$ in the Re u-Re v plane.

The proof of this statement is given in Appendix A. Clearly, one needs further to show that, to each $\mathscr{Z}'$, there exists a function of exponential type vanishing on it. A necessary and sufficient condition for this, generalizing Lindelöf's theorem (Ref. 11, p. 27) of the one-variable case, is provided in the same paper[13] by Lelong. Namely, the set $\mathscr{Z}'$ admits of a function of exponential type vanishing on it and nowhere else if and only if the coefficients of u and v of the polynomials

$$P_R(u,v)=\int_{\|a\|<R}\nabla_a^2\log|F(a_1,a_2)|\,\frac{a_1u+a_2v}{\|a\|^4}\,dv_a \tag{2.10}$$

are bounded, as $R\to\infty$. In (2.10), $F(a_1,a_2)$ is any entire function vanishing on $\mathscr{Z}'$ [e.g., obtained from (2.2) by replacing factors $G_i(u,v)$ by their conjugates $G_i^*(u^*,v^*)$]; it is easy to verify that (2.10) is independent of the choice of $F(a_1,a_2)$ so that it refers indeed to $\mathscr{Z}'$ only; dv_a is the volume element in the real space $a_{1x}, a_{1y}, a_{2x}, a_{2y}$. The integral in (2.10) may be written as a sum over the polynomials $P_{R,i}$ corresponding to each manifold $\mathscr{M}_i$ (if they intersect the ball $\|a\|<R$). The replacement of $\mathscr{M}_i$ by $\mathscr{M}_i^*$ causes the complex conjugation of the coefficients of $P_{R,i}(u,v)$.

In the one-variable case, one can prove that, if u_i are the zeros of the Fourier transform $F(u)$, then $\Sigma_i\,|\operatorname{Im} u_i|/|u_i|^2$ converges, so that the reflection of zeros does not harm Lindelöf's criterion. However, the analogous statement in two dimensions, namely that the family of integrals $(0<R<\infty)$

$$I_i(R)=\int_{\substack{\|a\|<R\\ \operatorname{Im} a_i>0}}\nabla_a^2\log|F(a_1,a_2)|\,\frac{\operatorname{Im} a_i}{\|a\|^4}\,dv_a\,, \quad i=1,2 \tag{2.11}$$

is bounded, is wrong. A simple example where (2.11) diverge as $R\to\infty$ is given in Appendix A. Therefore, we can only state the following.

Proposition 4: Let $\mathscr{Z}$ consist of a finite number of irreducible manifolds $\mathscr{M}_i$ and suppose each of them has the property that the imaginary parts of the coefficients of u and v in the polynomials $P_{R,i}(u,v)$ defined by (2.10) on $\mathscr{M}_i$ are bounded as $R\to\infty$. Then, if $\mathscr{Z}$ is the zero set of a solution of the phase retrieval problem, to each set $\mathscr{Z}'$ obtained from $\mathscr{Z}$ by reflecting any number of manifolds $\mathscr{M}_i$, there corresponds a distinct solution of the phase retrieval problem, vanishing on $\mathscr{Z}'$ and nowhere else.

Clearly, a statement on ambiguities without the restriction above on $P_{R,i}$ is desirable. Indeed, one should notice that Proposition 4 does not contain the simple case of a purely radial dependence of $\psi(x,y)=\tilde\psi(r)$. In that case, the intensity distribution is given by the Bessel transform of $\tilde\psi(r)$, which is an entire function of order $\frac{1}{2}$ in u^2+v^2. Consequently, it contains infinitely many zeros and may be written as a convergent product over irreducible factors of the form $(1-(u^2+v^2)/\alpha_i)$, $\alpha_i\neq 0$, complex. Any number of these factors may be replaced by their conjugates $(1-(u^{*2}+v^{*2})/\alpha_i)^*$, still giving rise to a solution of the phase problem.

A remark on the existence of solutions to the phase problem is now appropriate. In the one-variable situation, it is true that, if an entire function $M(u)$ of exponential type τ is positive and integrable on the real axis, then a function $F(u)$ of type $\tau/2$ exists (in fact an infinity of them) so that $M(u) = F(u)F^*(u)$, $u \in \mathbb{R}$. This follows from Theorem 7.5.1 of Ref. 11, since the conditions on $M(u)$ ensure that $\Sigma_n |\mathrm{Im}\, 1/u_n| < \infty$ (u_n being its zeros). Thus, one can give a complete characterization of those functions $M(u)$ that are Fourier transforms of autocorrelation functions.

In two variables, it is not true that a function which is positive in the Re u-Re v plane and entire necessarily contains pairs of irreducible manifolds, conjugate to each other (by analogy to the conjugate zeros in one variable). As an example, consider $M(u,v) = \exp(u) + \exp(v) + 1$, whose zeros consist of one irreducible manifold. It is more complicated to find functions $M(u,v)$ that also belong to $L^1(\mathbb{R}^2)$ and for which irreducibility can be proven, but we construct such an example in Appendix C.

From the construction of Appendix C, one expects that, in general, there exist many functions $M(u,v)$, positive and with integrable modulus in the Re u-Re v plane whose Fourier transforms, although they have compact support, are not autocorrelation functions. From Proposition 4, we may read that $M(u,v)$ is admissible if it has a finite number of irreducible zero sets $\mathscr{M}_i$, which may be parted into two classes so that (i) to each $\mathscr{M}_i$ in one class, there corresponds $\mathscr{M}_i^*$ in the other class and (ii) each class separately can support an entire function of exponential type.

With Proposition 4, we have established that, if $\mathscr{Z}$ consists of N irreducible zero sets, the ambiguity of the phase problem is at most 2^N. The next problem is whether statements are possible if the modulus $M(u,v)$, Eq. (2.3), is affected by sufficiently small errors, still staying admissible.

It is sufficient to this end to study the variation in the number of prime factors of two Fourier transforms $F_1(u,v)$, $F_2(u,v)$, Eq. (2.1), that have values close to each other in the Re u-Re v plane, in the sense that

$$\|F_1 - F_2\|^2 = \int |F_1(u,v) - F_2(u,v)|^2\, du\, dv < \epsilon^2\,, \qquad (2.12)$$

and are further such that the original functions $\psi_1(x,y)$, $\psi_2(x,y)$ have support contained in a rectangle $[-a,a]\times[-b,b]$. Equation (2.12) is a norm in a Hilbert space $E_{a,b}$ of entire functions of exponential type less than a, at each fixed v, and less than b at each fixed u, as a consequence of the Paley–Wiener theorem. [One cannot use Eq. (2.12) instead of (1.2) for polynomials, since (2.12) would imply that the two polynomials differ by a small constant.]

By means of Schwartz's inequality and Parseval's theorem, one may verify that, if $F_1(u,v)$, $F_2(u,v) \in E_{a,b}$ and $(u,v) \in \mathbb{C}^2$ then

$$|F_1(u,v) - F_2(u,v)| < \exp[a|\mathrm{Im}\, u|]\exp[b\,|\mathrm{Im}\, v|]\, \|F_1 - F_2\| \times 2(ab)^{1/2}\,. \qquad (2.13)$$

Equation (2.13) shows that within any strip $|\mathrm{Im}\, u| < A$, $|\mathrm{Im}\, v| < B$, we can make the absolute departure of F_1 from F_2 as small as we wish, if $\|F_1 - F_2\|$ is chosen small enough, i.e., the analytic extrapolation within $E_{a,b}$ off an error corridor (2.12) is stable, inside any strip of finite width. This statement may seem to contradict those of Ref. 14, where an investigation of the instabilities in zero location is performed; our statement refers to the situation when $F(u,v)$ is available on the whole Re u-Re v plane with errors (2.12).

As we shall see in Secs. III and IV, the decision whether a manifold is irreducible or not rests with the ability of locating branch points in the u plane of the functions $v = v(u)$, the solutions of the equation $F(u,v) = 0$. In view of the finite errors (2.12), it is not artificial to replace the notion of irreducibility in the large by that of irreducibility with respect to a compact set $D(A_1,A_2,B_1,B_2)$: $|\mathrm{Re}\, u| < A_1$; $|\mathrm{Re}\, v| < A_2$; $|\mathrm{Im}\, u| < B_1$; $|\mathrm{Im}\, v| < B_2$; $D = D_1 \times D_2$. We call a subset $\mathscr{M}_{D,i}$ of the intersection $\mathscr{Z}_D = \mathscr{Z} \cap D$ of the zero set $\mathscr{Z}$ of $F(u,v)$ with D irreducible with respect to D if it consists of all points with coordinates $(u,v(u))$, where $v(u)$ is obtained by analytic continuation from a function element $v_0(u)$, $u \in D_1$, $v_0(u) \in D_2$ along all possible paths for which $u \in D_1$, $v \in D_2$. Clearly, the number N'_D of irreducible sets $\mathscr{M}_{D,i}$ with respect to D is larger, in general, than the number N_D of irreducible sets of $\mathscr{Z}$ which intersect D.

We can now state the following.

Proposition 5: Let $F_0(u,v)$ belong to $E_{a,b}$ and consider the compact rectangle $D(A_1,A_2,B_1,B_2)$. There exists then a neighborhood $\mathscr{U}$ of F_0 in $E_{a,b}$ so that any other F in $\mathscr{U}$ has at most the same number of irreducible zero sets with respect to D as $F_0(u,v)$.

In other words, the number of irreducible zero sets with respect to D can at most decrease under perturbation. The meaning of this statement concerning ambiguities is, however, rather complicated; that part of the ambiguity of the phase problem coming from the N_D irreducible zero sets that intersect D is at most 2^{N_D}. Proposition 5 makes a statement about an upper bound $2^{N'_D}$ of the latter estimate; it decreases under sufficiently small perturbations.

Proof: We consider the set $\mathscr{S}$ of points (u_{0i},v_{0i}) of $\mathscr{Z}_D$ with the property that several roots $v_{j1}(u),\ldots,v_{jr}(u)$, $r > 1$ of the equation $F_0(u,v) = 0$ assume the common value $v_{0,i}$ at $u = u_{0,i}$. The set consists either of (a finite number of) isolated points or contains an irreducible zero set of F_0. We assume first, for simplicity, that the former is the case and surround each of these points by a polydisk: d_i^0: $|u - u_{0,i}| < r_i$, $|v - v_{0,i}| < r_i$, with r_i chosen so that no two points of $\mathscr{S}$ belong to the same d_i^0.

To each point $(\tilde{u},\tilde{v})$ of D lying outside $\cup_i d_i^0$ we attach by the implicit function theorem a polydisk $d(\tilde{u},\tilde{v}) = d_u(\tilde{u},\tilde{v}) \otimes d_v(\tilde{u},\tilde{v})$, $d_u: |u - \tilde{u}| < r_1(\tilde{u},\tilde{v})$, $d_v: |v - \tilde{v}| < r_2(\tilde{u},\tilde{v})$, so that, for each u in d_u there exists one and only one root of $F_0(u,v) = 0$ lying strictly inside d_v. The covering of $\mathscr{Z}'_D = \mathscr{Z}_D \setminus \cup_i d_i^0$ that is thus obtained can be chosen so as to be contained in a rectangle

$$\begin{aligned} D_k: &|\mathrm{Re}\, u| < kA_1, \quad |\mathrm{Re}\, v| < kA_2,\\ &|\mathrm{Im}\, u| < kB_1, \quad |\mathrm{Im}\, v| < kB_2, \end{aligned} \qquad (2.14)$$

for any given $k > 1$. Now $\mathscr{Z}'_D$ is a compact set and we can extract from the set of $d(\tilde{u},\tilde{v})$ a finite cover of it; let its elements be $d_j = d_{u,j} \otimes d_{v,j}$, $j = 1,2,\ldots,N_1$. It is only at this stage that we use the finite extension of D in the Re u, Re v direction.

We shall now show that we can choose perturbations $\delta F(u,v)$ of $F_0(u,v)$ such that (i) the set of zeros of any function $(F_0+\delta F)(u,v)$ is entirely contained in $(\cup_i d_i^0)\cup(\cup_j d_j)$ and (ii) that the d_j's maintain the property that, for each u in $d_{u,j}$, there exists one and only one root of $(F_0+\delta F)(u,v)=0$ lying inside $d_{v,j}$. To this end, let

$$\mu_j = \inf_{\substack{u\in \bar{d}_{u,j}\\ |v-\tilde{v}_j| = r_{2,j}}} |F_0(u,v)| , \tag{2.15}$$

with $\tilde{v}_j$ the center of $d_{v,j}$. By our choice of $d_u(\tilde{u},\tilde{v})$, $\mu_j\neq 0$. Further, let

$$M_j = \sup_{\substack{u\in \bar{d}_{u,j}\\ |v-\tilde{v}_j| = r_{2,j}}} \left(\left|\frac{\partial F_0}{\partial v}(u,v)\right| + |F_0(u,v)|\right) \tag{2.16}$$

and let

$$\epsilon' < \inf_j \left[\mu_j^2/(\mu_j + M_j r_{2,j})\right] . \tag{2.17}$$

Now let D_k' be defined by (2.14) with $k'>k$ and choose $\delta F(u,v)\in E_{a,b}$ so that

$$\sup_{D_k'}|\delta F(u,v)| < \epsilon', \quad \sup_{D_k'}\left|\frac{\partial \delta F}{\partial v(u,v)}\right| < \epsilon' . \tag{2.18}$$

Then conditions (i) and (ii) above are fulfilled. This is verified by evaluating in each $d_{u,j}$ the differences of the values of the integer-valued functions

$$N(F) = \frac{1}{2\pi i}\oint_{|v-\tilde{v}_j| = r_{2,j}} \frac{\partial F/\partial v}{F}(u,v)dv , \tag{2.19}$$

evaluated for F_0 and $F_0+\delta F$. Clearly, $N(F_0)=1$ and (2.17) ensures that $|N(F_0)-N(F_0+\delta F)|<1$. Thus, condition (ii) is fulfilled; (i) is also verified by applying the reasoning above to a point outside $(\cup_i d_i^0)\cup(\cup_j d_j)$ and we conclude that it is impossible for $(F_0+\delta F)(u,v)$ to vanish there. It is also true that all points $(u,v(u))$ of the zero set $\mathscr{Z}_{\delta F,D}$ of $(F_0+\delta F)(u,v)$ lying inside $\cap_j d_j$ are regular points, i.e., $v(u)$ is holomorphic in $d_{u,j}$.

We can now show that $\mathscr{Z}_{\delta F,D}$ contains at most as many irreducible manifolds with respect to D as $\mathscr{Z}_D$ [corresponding to $F_0(u,v)=0$].

To this end, consider the set $\mathscr{K}$ of those polydisks $d_{j,k}$ which are part of the covering of an irreducible manifold $\mathscr{M}_{1,D}$ of $\mathscr{Z}_D$ (with respect to D). We shall argue that any two points of $\mathscr{Z}_{\delta F,D}$ lying in $\mathscr{K}$ can be connected by analytic continuation along a path such that $(u,v(u))$ stays inside D. Indeed, let the two points of $\mathscr{Z}_{\delta F,D}$ be $(u_i,\tilde{v}_i(u_i))$, $(u_f,\tilde{v}_F(u_F))$, lying in the polydisks d_i and d_f, in turn and let $\tilde{v}_i(u)$, $\tilde{v}_f(u)$ be the function elements defining $\mathscr{Z}_{\delta F,D}$ for u in d_{ui} and d_{uf}. By the construction above there exist two other function elements defining $\mathscr{Z}_D$, $v_i(u)$, $v_f(u)$, defined on d_{ui} and d_{uf} and such that $v_i(u)\in d_{vi}$ if $u\in d_{ui}$, $v_f(u)\in d_{vf}$ if $u\in d_{uf}$. Further, $v_f(u)$ may be obtained by the analytic continuation of $v_i(u)$ along a path $\mathscr{P}$ contained in D_u; let $d_1\equiv d_i$, $d_2,\dots,d_n\equiv d_f$ be the polydisks covering $\mathscr{P}$ and $(u,v_k(u))$, $(u,\tilde{v}_k(u))$ the coordinates of points of $\mathscr{Z}_D$ and $\mathscr{Z}_{\delta F,D}$ contained in the polydisk d_k for $u\in d_{u,k}$. Now, since to each u in $d_{u,k}$ there corresponds just one root of $(F_0+\delta F)(u,v)=0$, it follows that in the overlap $d_{u1}\cap d_{u2}$, $\tilde{v}_1(u)\equiv\tilde{v}_2(u)$. This means that $\tilde{v}_2(u)$ is the analytic continuation of $\tilde{v}_1(u)$ and our assertion is proved by repeating this reasoning from d_2 to d_n.

As a consequence $\mathscr{Z}_{\delta F,D}$ cannot contain more irreducible zero sets than $\mathscr{Z}_D$. However, it is possible that, by using paths of continuation that intersect $\cup_i d_i^0$, we may connect points of $\mathscr{Z}_{\delta F,D}$ whose correspondents on $\mathscr{Z}_D$ (outside $\cup_i d_i^0$) were unrelated. Therefore, in general, if $N_{F,D}$ denotes the number of irreducible zero sets of F with respect to D, then $N_{F_0+\delta F}\leqslant N_{F_0}$. Clearly, using Cauchy's inequalities, we can ensure both inequalities (2.18) by requiring $\sup|\delta F(u,v)|<\epsilon''$, for (u,v) on the boundary of $D_{k''}$, for sufficiently small ϵ'' and $k''>k'$. According to (2.13), this may in turn be obtained by choosing $\|\delta F\|$ sufficiently small.

If the set of points $(u_{0,i},v_{0,i})$ with $v_{0,i}$ multiple roots of the equation $F(u_{0,i},v)=0$ does not consist only of isolated points in D, it must contain an irreducible zero set $\tilde{\mathscr{M}}$, on which $F_0(u,v)$ has at each fixed u, a zero of multiplicity $p>1$. In this case, we cover $\tilde{\mathscr{M}}$ as above with polydisks, avoiding those points where the multiplicity of the root is higher. Under a sufficiently small perturbation, p roots of $F_0+\delta F$ stay inside the covering $\mathscr{K}_1$ of $\tilde{\mathscr{M}}$ (except for some special points) with polydisks. In general, we cannot ensure that the p roots are distinct at all points of the projection of $\mathscr{K}_1$ onto the u plane. Thus there will appear $p_1\leqslant p$ irreducible zero sets under perturbation, as asserted.

Next follow some comments.

(i) Assume $N_{F_0}=1$. Then, according to Proposition 5, there exists a neighborhood U of F_0 in the topology above, so that $N_{F_0+\delta F}=1$ for any $F_0+\delta F$ in U. This means that irreducible zero sets are stable under small perturbations and this is sufficient for the analysis of the ambiguity problem.

One would like, however, to show also that reducible situations are in some sense unstable. In fact, one expects that, if (u_0,v_0) is a point which is common to two irreducible zero sets of a function $F_0(u,v)$, then in any neighborhood of F_0 (in the topology above) there are functions $F_0+\delta F$ which have just one irreducible zero set in a sufficiently small domain containing (u_0,v_0). We show that this is the case in Appendix B. Several examples for the generation of branch points under the perturbation of reducible zero sets are given in Ref. 6.

(ii) If we let the rectangle D increase in the imaginary direction then (a) the maximum $\|\delta F\|$ allowed so that the inequalities (2.18) are fulfilled tends to zero, according to (2.13) and (b) the number of polydisks required to cover $\mathscr{Z}_D$ increases, so that ϵ' in (2.18) may tend to zero. Only the second problem occurs if we let D increase in the real directions and it is conceivable that one may show, by a more careful treatment, that ϵ' stays nevertheless finite (at least for Fourier transforms).

(iii) The fact that $F_0(u,v)$ is entire was used only through (2.13). In the next proposition, we shall use the fact that, according to the proof above, we may also state the following: for any $F_0(u,v)$, holomorphic in $D_{k''}$, there exists an ϵ'', so that all $F(u,v)$, holomorphic in $D_{k''}$ and satisfying $|F(u,v)-F_0(u,v)|<\epsilon''$ for (u,v) in $D_{k''}$, have the same number of irreducible zero sets with respect to D as $F_0(u,v)$ $(D\subset D_{k''})$.

(iv) A question of interest is the following: given $\epsilon>0$ and the function $F_0(u,v)$ in $E_{a,b}$, what is the number of irreducible zero sets with respect to D of a function $F_1(u,v)$ in $E_{a,b}$ departing from $F_0(u,v)$ by less than ϵ, in the sense of (2.12)? This is answered by the following.

Proposition 6: For any $\epsilon > 0$ and rectangle D [given by (2.14) with $k = 1$], there exists $\mathscr{N}_0(\epsilon,D)$ such that the number of irreducible zero sets with respect to D of any $F(u,v)$ in $E_{a,b}$ and obeying $\|F - F_0\| < \epsilon$, for some F_0 in $E_{a,b}$ is less than $\mathscr{N}_0$.

Proof: Let $\mathscr{F}$ be the set of functions $F(u,v)$ in $E_{a,b}$ obeying $\|F - F_0\| < \epsilon$ and let $k_1 > k''$ (of the previous proposition). According to (2.13), there exists a number Q, so that all F in $\mathscr{F}$ also obey $|F(u,v) - F_0(u,v)| < Q$ for all (u,v) in D_{k_1}. By Montel's principle, $\mathscr{F}$ is a compact set in the space $H(D_{k_1})$ of all functions holomorphic in D_{k_1}, with the topology of uniform convergence on compact subsets of D_{k_1}. Further, by remark (iii) above, we can attach to each F in $\mathscr{F}$ a set of functions $\mathscr{B}(F)$ holomorphic in D_{k_1} and having the same number $N(F)$ of irreducible zero sets with respect to D as $F(u,v)$. The functions $\widetilde{F}(u,v)$ of the set $\mathscr{B}(F)$ obey $|\widetilde{F}(u,v) - F(u,v)| < \epsilon''$ for all u,v in $D_{k''}$ and make up an open set of $H(D_{k_1})$. The set of all $\mathscr{B}(F)$ makes up a covering of the compact set $\mathscr{F}$ and we may therefore extract from it a finite subcover. Let the numbers $N(F)$ associated to this subcover be N_i. Then $\mathscr{N}_0(\epsilon,D) = \max_i N_i$. This ends the proof.

We recall that $2^{\mathscr{N}_0}$ is an upper bound on the ambiguity of the phase problem coming from the reflection of those irreducible zero sets that intersect D. The computation of $\mathscr{N}_0(\epsilon,D)$ may be difficult. Its existence is, however, significant since it may happen that it is much smaller than the number of roots $v(u)$ of $F(u,v) = 0$ at fixed u in $|\mathrm{Re}\, v| < A_2$, $|\mathrm{Im}\, v| < B_2$. The claims that two-variable analyticity leads to a reduction of ambiguities rest on this occurrence.

As is apparent from the above discussion, the present author disagrees with the remarks made in Ref. 2 concerning the "number" of irreducible functions, as compared to that of reducible ones. No statistical arguments are pertinent to this problem, except the relation to a topology defined on the space $E_{a,b}$ of entire functions, e.g., by means of Eq. (2.12); reducible situations are in this sense unstable. However, one cannot conclude that, consequently, the "chance" of meeting them is zero. The relevant question is the one under (iv) above and it may well be that quite common intensity distributions and error channels lead to a high $\mathscr{N}_0(\epsilon,D)$ (for a reasonably large D); simple examples are offered by the weak perturbation of the diffraction patterns on rectangular or circular apertures.

III. ZERO TRAJECTORIES

The discussion of the preceding section relies on methods of construction of $F(u,v)$ (Cousin's method and Lelong's generalization of the Weierstrass product) which uses its whole set of zeros in $\mathbb{C}^2$. The set $\mathscr{Z}$ is completely determined, in principle, from knowledge of $M(u,v)$ in the Re u-Re v plane, but its actual computation is prohibitive. Thus, these methods are suited for a discussion of the ambiguity problem, but not for the actual construction of $F(u,v)$ from $M(u,v)$.

Instead, it is natural to use, e.g., at each fixed real v, representations in one variable of $F(u,v)$ in the complex u plane.

(a) First, we use a Weierstrass product of order of growth unity (Ref. 11, p. 22)

$$F(u,v) = C_0(v) \exp(a(v)u) \prod_{n=1}^{\infty} \left(1 - \frac{u}{u_n(v)}\right) \times \exp(u/u_n(v)) \qquad (3.1)$$

(b) Second, we use a conditionally convergent Titchmarsh product,[15] valid for Fourier transforms of functions with compact support,

$$F(u,v) = C_0(v) \exp(i\alpha u) \prod_{n=1}^{\infty} \left(1 - \frac{u}{u_n(v)}\right). \qquad (3.2)$$

Let the smallest x strip containing the support of $\psi(x, y)$, Eq. (2.1), be $a_1 < x < a_2$. Then, by writing the integral in (2.1) as an iterated integral, we may verify that the smallest interval containing the support of the Fourier transform at fixed v of $F(u,v)$ is $[a_1,a_2]$ and is independent of v. It follows then from Ref. 15 that, in Eq. (3.2), $\alpha = (a_1 + a_2)/2$, independently of v. Since we are interested in solutions that are equivalent modulo translations, there is no loss of generality to assume $\alpha = 0$.

Equations (3.1) and (3.2) require for their evaluation the intersection of the set $\mathscr{Z}$ of zeros of $F(u,v)$ with Im $v = 0$. The complex functions $u = u_n(v)$, v real, are called v-zero trajectories. They are continuous functions of v, except for cusps, where several of them can meet (assume the same value). By a small continuation clockwise in the complex v plane at such cusps, it is possible to identify (connect with each other) unambiguously the zero trajectories on the whole real v axis.

At first sight, knowledge of $M(u,v)$ at each fixed real v in the complex u plane fixes the v-zero trajectories up to a twofold ambiguity: $u_n(v) \to u_n^*(v)$. Further, $|C_0(v)|$ is fixed by $M(u,v)$, but its phase requires the value of $F(0,v)$, which in turn depends on the resolution of the discrete ambiguity of the zeros in the complex v plane, at $u = 0$.

On the other hand, we have seen in Sec. II that use of the analyticity of $M(u,v)$ in two variables implies an upper bound on the number of solutions, which might be considerably smaller than the infinite one above. In particular, it may be that no solutions exist at all. This means that part of the resolutions of the discrete ambiguity of the zero trajectories and of the zeros in the v plane at $u = 0$ lead by means of (3.1) or (3.2) to functions $F(u,v)$ that, although continuous and holomorphic in one variable, will fail to be analytic in two variables or will violate the exponential bounds.

In this section, we set up necessary and sufficient conditions for the resolutions of the discrete ambiguity of the v-zero trajectories and of the zeros at $u = 0$ to lead, via Eq. (3.2), to solutions of the phase retrieval problem. In doing so, we shall give an answer to questions (i)–(iii) of the Introduction, in the framework of entire functions.

Before starting on this, we notice the following.

Lemma: Consider the smallest rectangle with sides parallel to the x and y axes, that contains the support of the Fourier transform of $M(u,v)$. Let $4a_0$, $4b_0$ be the lengths of its sides. Then, if solutions to the phase retrieval problem exist, each class of solutions contains an entire function of exponential type a_0 at each fixed v and of exponential type b_0 at each fixed u.

Clearly, it is enough to show that the smallest rectangle with sides parallel to the x and y axes that contains the support $\mathscr{D}$ of the object function $\psi(x,y)$ [Eq. (2.1)] has sides of lengths $2a_0, 2b_0$. To see this, notice that (i) the support $\mathscr{T}$ of the autocorrelation function is symmetrical; if $\mathbf{y} \in \mathscr{T}$, $-\mathbf{y} \in \mathscr{T}$, and (ii) a point $\mathbf{y} = \|\mathbf{y}\|\hat{y}$ on the boundary of $\mathscr{T}$ has the property $\|\mathbf{y}\| = \max\|\mathbf{r}_2 - \mathbf{r}_1\|$, over $\mathbf{r}_1, \mathbf{r}_2 \in \mathscr{D}$, $\hat{y} \times (\mathbf{r}_2 - \mathbf{r}_1) = 0$. Let now $2b_1$ be the length of the vertical side of the smallest rectangle containing $\mathscr{D}$ (with sides parallel to x and y) and let $\mathbf{y}_0$ be a point on the boundary of $\mathscr{T}$ at which the horizontal side of the smallest rectangle touches $\mathscr{T}$. If $\mathbf{e}$ is the unit vector in the direction of the y axis, we may choose $\mathbf{y}_0$ such that $(\mathbf{e}, \mathbf{y}_0) > 0$ and write

$$
\begin{aligned}
2b_0 &= (\mathbf{e}, \mathbf{y}_0) = |(\mathbf{e}, \mathbf{y}_0)| = \sup_{\mathbf{y} \in \mathscr{T}} (\mathbf{e}, \mathbf{y}) = (\mathbf{e}, \mathbf{r}_{20} - \mathbf{r}_{10}) \\
&\leqslant \sup_{\mathbf{r}_2 \in \mathscr{D}} (\mathbf{e}, \mathbf{r}_2) - \inf_{\mathbf{r}_1 \in \mathscr{D}} (\mathbf{e}, \mathbf{r}_1) = 2b_1 = (\mathbf{e}, \mathbf{r}_2' - \mathbf{r}_1') \\
&= (\mathbf{e}, \hat{y}_1)\|\mathbf{r}_2' - \mathbf{r}_1'\| \\
&\leqslant |(\mathbf{e}, \hat{y}_1)| \sup_{\substack{\mathbf{r}_1, \mathbf{r}_2 \in \mathscr{D} \\ (\mathbf{r}_1 - \mathbf{r}_2) \times \hat{y}_1 = 0}} \|\mathbf{r}_2 - \mathbf{r}_1\| \\
&= |(\mathbf{e}, \mathbf{y}_2)| \leqslant (\mathbf{e}, \mathbf{y}_0) = 2b_0 .
\end{aligned} \tag{3.3}
$$

In (3.3), $\mathbf{y}_1 = \|\mathbf{r}_2' - \mathbf{r}_1'\|\hat{y}_1$, $\mathbf{y}_2 = \hat{y}_1 \sup\|\mathbf{r}_2 - \mathbf{r}_1\|$. The same reasoning can be done for the horizontal side. This ends the proof.

In the following, we shall assume that the modulus function $M(u,v)$ belongs in the Re u-Re v plane to $L^1(R^2)$ and that its Fourier transform is contained in a rectangle of size $4a_0, 4b_0$. We shall need the following theorem of Plancherel and Polya.[9]

Theorem: If the entire function of exponential type $M(u,v)$ belongs to $L^1(\mathbf{R}^2)$ in the Re u-Re v plane, then the series

$$
\sum_{m,n} \left| M\left(\frac{m\pi}{a}, \frac{n\pi}{b}\right)\right| \tag{3.4}
$$

is convergent for any $a > 0$, $b > 0$ (see Ref. 9, Sec. 46).

To state the necessary and sufficient conditions for the correct resolution of the ambiguity of v-zero trajectories and zeros at $u = 0$, we consider the set of points $(u_m, v_n) = (m\pi/a, n\pi/b)$ with $a > a_0$, $b > b_0$, and a_0, b_0 of the Lemma. At each $v = n\pi/b$, we construct a function $F(u, n\pi/b)$ in the u plane by means of the Titchmarsh product (3.2) and some resolution of the discrete ambiguity for v-zero trajectories and for the zeros in the $u = 0$ plane. With this, we construct a function of the Lagrange–Valiron type

$$
L(u,v) = \sin bv \sum_{n=-\infty}^{\infty} \frac{(-1)^n F(u, n\pi/b)}{bv - n\pi} . \tag{3.5}
$$

Concerning (3.5) we prove the following.

Proposition 7: Let $a > a_0$, $b > b_0$; then the function $L(u,v)$ defined in (3.5) exists and is an entire function of u and v; at each fixed u, it is of exponential type less than b with respect to v, and it is of type less than a with respect to u at each fixed v.

It is worth pointing out that this statement is true regardless of whether the choices of the discrete ambiguity are "correct" or "wrong" in the sense discussed above, and in fact independent of whether a solution exists or not. According to the Lemma, we can read off the Fourier transform of $M(u,v)$ the allowed values of a and b.

Proof: We show first that the functions $F(u, n\pi/b)$ exist and are of exponential type less than a. It will be assumed that $M(0, n\pi/b) \neq 0$; the modifications needed to treat this situation are obvious. We proceed as follows.

(i) Carleman's theorem (see, e.g., Ref. 11, p. 2) applied to $M(u, n\pi/b)$ shows that (see also Ref. 15)

$$
\sum_k \left| \operatorname{Im} \frac{1}{u_k(n\pi/b)} \right| < \infty , \tag{3.6}
$$

where $u_k(n\pi/b)$ are the roots of $M(u, n\pi/b)$ lying, e.g., in the upper half plane. [Since $M(u, n\pi/b)$ is real, all its zeros fall in complex conjugate pairs or are real.]

(ii) $M(u, n\pi/b)$ obeys the Titchmarsh representation (3.2); in particular, the series

$$
S = \sum_{k=1}^{\infty} \frac{1}{u_k(n\pi/b)} \tag{3.7}
$$

converges. If we replace any number of u_k's by their conjugates, (3.6) shows that the series still converges. Thus, we may define $F(u, n\pi/b)$ by the Titchmarsh product constructed from an arbitrary resolution of the ambiguity of the v-zero trajectories; the product is certainly convergent. [We take $\alpha = 0$ in Eq. (3.2).]

(iii) The function $F(u, n\pi/b)$ is of exponential type. To see this, we use the fact that, if $n(r)$ represents the number of zeros of $M(u, n\pi/b)$ of modulus less than r, then $\lim_{r\to\infty} n(r)/r$ is a nonzero constant (see, e.g., Ref. 15, Theorem III). As a consequence, a constant C exists, so that for all k, $|u_k(n\pi/b)| \geqslant k/C$ [if $M(0, n\pi/b) \neq 0$]. We write then

$$
\begin{aligned}
\ln|F(u, n\pi/b)| &= \sum_{k=1}^{N} \ln\left|1 - \frac{u}{u_k(n\pi/b)}\right| \\
&\quad + \sum_{k \geqslant N+1} \ln\left|1 - \frac{u}{u_k(n\pi/b)}\right| = T_1 + T_2 ,
\end{aligned} \tag{3.8}
$$

with N chosen so that $|u/u_k(n\pi/b)| \leqslant C_1$, for some C_1 and all $k \geqslant N$. With the help of the inequality $\ln(1+x) \leqslant x$ $(x > -1)$, we can write for the second term in (3.8)

$$
T_2 \leqslant |u|(|\operatorname{Re} S_{N+1}| + |\operatorname{Im} S_{N+1}|) + \frac{1}{2}|u|^2 \sum_{k \geqslant N+1} \frac{1}{|u_k|^2} . \tag{3.9}
$$

The last term in (3.9) is $O(|u|)$ and S_{N+1} in the first term represents the rest of (3.7) after removing the first N terms. Comparing with an integral, we verify that T_1 is also $O(|u|)$ as $|u| \to \infty$, which proves our statement.

(iv) Since $\int_{-\infty}^{\infty} |F(u, n\pi/b)|^2\, du$ exists, it follows that $F(u)$ is the Fourier transform of a function with compact support. From Titchmarsh's representation, we see that the support has the form $(-\beta, \beta)$, for some $\beta > 0$. We now show that, in fact, $\beta = a_0$.

To this end, we use the fact (Theorem 7.5.1 of Ref. 10) that there exists an entire function $\omega(u)$ with modulus squared equal to $M(u, n\pi/b)$, of exponential type $a_0/2$ and with all its zeros in the upper half u plane. Such a function satisfies the Titchmarsh representation and $F(u, n\pi/b)$ differs from it by the flipping of a certain (maybe infinite) number of zeros, $u_{k_i}\,(n\pi/b)$. We can write

$$\ln\left|F\left(u,\frac{n\pi}{b}\right)\right| = \ln|\omega(u)| + \ln\prod_{k_i}\frac{|1-u/u_{k_i}^*|}{|1-u/u_{k_i}|}, \tag{3.10}$$

where the product extends over the set of zeros that have been flipped. The Ahlfors–Heins theorem (Ref. 11, p. 115) ensures then that, if $B(u)$ denotes the second term, then

$$\limsup_{\substack{|u|\to\infty \\ u=|u|\exp(i\theta)}} \ln(|B(u)|/|u|) = 0, \tag{3.11}$$

for almost all directions θ. It follows that

$$\limsup_{\substack{|u|\to\infty \\ u=|u|\exp(i\theta)}} (\ln(|F(u,n\pi/b)|)/|u|) = \beta|\sin\theta|$$

$$= \limsup_{\substack{|u|\to\infty \\ u=|u|\exp(i\theta)}} \frac{\ln|\omega(u)|}{|u|} = a_0|\sin\theta|, \tag{3.12}$$

i.e., that $\beta = a_0$.

As a consequence of this, $F(u,n\pi/b)$ satisfies a Lagrange–Valiron representation $(a > a_0)$

$$F\left(u,\frac{n\pi}{b}\right) = \sin au \sum_{m=-\infty}^{\infty}\frac{(-1)^m F(m\pi/a,\, n\pi/b)}{au - m\pi}. \tag{3.13}$$

It is important that, by construction

$$\left|F\left(\frac{m\pi}{a},\frac{n\pi}{b}\right)\right|^2 = M\left(\frac{m\pi}{a},\frac{n\pi}{b}\right), \tag{3.14}$$

and the theorem cited at the beginning of this proof ensures that

$$\sum_{m,n}\left|F\left(\frac{m\pi}{a},\frac{n\pi}{b}\right)\right|^2 < \infty, \tag{3.15}$$

i.e., $F(m\pi/a,\, n\pi/b)$ are the Fourier coefficients of a function $\psi_1(x,y)$ in $L^2(D)$, where $D = [-a,a]\times[-b,b]$

$$\psi_1(x,y) = \sum_{m,n} F\left(\frac{m\pi}{a},\frac{n\pi}{b}\right)\exp\left(i\left(\frac{m\pi}{a}x + \frac{n\pi}{b}y\right)\right). \tag{3.16}$$

Consider then

$$L_1(u,v) = \frac{1}{2\pi}\int_D \psi_1(x,y)\exp(i(ux+vy))\,dx\,dy, \tag{3.17}$$

which is, at each fixed u, of exponential type at most equal to b with respect to v, and at most equal to a with respect to u at fixed v. Expanding the exponential in a Fourier series in D and integrating term by term (which is permitted for square-integrable ψ_1), we obtain the Lagrange–Valiron expansion of $L_1(u,v)$

$$L_1(u,v) = \sin au \sin bv \times\sum_{m,n}\frac{(-1)^{m+n}F(m\pi/a,\, n\pi/b)}{(au-m\pi)(bv-n\pi)}. \tag{3.18}$$

It is easy to verify that, in view of (3.15), the series in (3.18) converges absolutely and uniformly on every compact subset of $\mathbb{C}^2$. Using a theorem on the iteration of double series (Ref. 16), the absolute convergence of the series in (3.18) entails the convergence of the iterated series to the same limit; one first keeps n fixed and obtains $F(u,n\pi/b)$ by (3.10) and then sums over m. Thus $L_1(u,v) = L(u,v)$, which ends the proof.

We can now formulate a criterion of the type announced in the Introduction.

Proposition 8: The function $L(u,v)$, Eq. (3.5), constructed from a certain resolution of the discrete ambiguity at each fixed $v = k\pi/b$ and at $u = 0$ is a solution of the phase retrieval problem if and only if

$$\left|L\left(\frac{m\pi}{2a},\frac{n\pi}{2b}\right)\right|^2 = M\left(\frac{m\pi}{2a},\frac{n\pi}{2b}\right), \tag{3.19}$$

for all m and all odd n.

Proof: The function $\mathscr{L}(u,v) = L(u,v)L^*(u^*,v^*)$ is at fixed u of exponential type at most equal to $2b$ in v, and at fixed v of exponential type at most equal to $2a$ with respect to u. So is $M(u,v)$, since $2a > 2a_0$, $2b > 2b_0$. The two functions are absolutely integrable over the Re u-Re v plane and, therefore (Ref. 9, Sec. 15), also square integrable. By the Paley–Wiener theorem, they are Fourier transforms of square-integrable functions with support contained in $[-2a,2a]\times[-2b,2b]$ and, therefore, with the reasoning following Eq. (3.17), satisfy representations like (3.18), with a,b replaced by $2a,2b$. It follows that the two functions coincide if they coincide at all points $(u,v) = (m\pi/2a,\, n\pi/2b)$, with (m,n) ranging over all integers (see also Ref. 9, Sec. 26). But $\mathscr{L}(u,v)$ and $M(u,v)$ already coincide by construction at those points with n even. Since $L(u,v)$ is entire, of exponential type, and has the correct modulus, it is a solution of the phase problem.

Conversely, if $L(u,v)$ is a solution of the phase problem, then Eq. (3.19) must be fulfilled. This ends the proof.

Next follow some comments.

(i) If condition (3.19) is fulfilled, a solution of the phase retrieval problem exists and is explicitly constructed. The resolution of the discrete ambiguity of the v-zero trajectories is done not only in a manner compatible with two-variable analyticity, but also with the constraints of Lindelöf's theorem.

(ii) The parameter b is arbitrary, as long as $b > b_0$. It follows that, by allowing b to increase, we can make $\mathscr{L}(u,v) = L(u,v)L^*(u^*,v^*)$ coincide with $M(u,v)$ at points $(u,n\pi/b)$ as closely spaced as we wish. Now, the type of $M(u,v)$ with respect to v is $2b_0$; one might wonder whether, by choosing $b > 2b_0$, we may not force the two functions to coincide for all v at fixed u, and thus conclude that $L(u,v)$ is always a solution of the phase problem, independently of whether it is constructed with a correct or wrong resolution of the discrete ambiguities. The reason why this is wrong is that the type of $\mathscr{L}(u,v)$ at fixed u is larger than b in the case of wrong resolutions and less than b (for $b > 2b_0$) only in the case of correct resolutions. In other words, if we consider the Fourier transform of $L(u,v)$ with respect to v [cf. (3.13)]

$$\psi_1(u,y) = \sum_{n=-\infty}^{\infty} F\left(u,\frac{n\pi}{b}\right)\exp\left(-i\frac{n\pi}{b}y\right), \tag{3.20}$$

then for a correct resolution of the discrete ambiguity, $\psi_1(u,y)$ has support contained in $-b_0 < y < b_0$; for an incorrect resolution, the support of $\psi_1(u,y)$ increases with b.

(iii) We inquire what happens if, instead of being satisfied exactly, Eqs. (3.19) are satisfied only on the average, i.e.,

$$\sum_m\sum_n\left|\mathscr{L}\left(\frac{m\pi}{2a},\frac{n\pi}{2b}\right) - M\left(\frac{m\pi}{2a},\frac{n\pi}{2b}\right)\right|^2 \leqslant \epsilon^2, \tag{3.21}$$

for some ϵ. Application of the Paley–Wiener theorem and of Parseval's equality shows that

$$\int_{R^2} |\mathscr{L}(u,v) - M(u,v)|^2 \, du \, dv \leqslant \epsilon^2 , \tag{3.22}$$

i.e., that $L(u,v)$ is a solution of the phase problem, within the errors ϵ^2. Thus, Eqs. (3.19) are stable against errors.

Equations (3.19) represent sufficient conditions for the compatibility between a resolution of the ambiguity of the v-zero trajectories, that of the zeros in the $u=0$ plane, analyticity in two variables, and the exponential type of the solution. They do not allow, however, a simple manner to fix the ambiguity of the zeros in the $u=0$ plane once a solution for that of the v-zero trajectories has been chosen, although they show that such a correlation exists. This correlation means that the positions of at least some of the zeros in the complex v plane at $u=0$ should be accessible from knowledge of the trajectories $u=u(v)$ by analytic continuation along some path in the v plane. In general, we cannot expect to obtain all zeros by such a continuation, since $F(u,v)$ may contain, e.g., zero sets of the form $v=v_0=$ const. These zero sets lead to ambiguities that are independent of those of the v-zero trajectories. In particular, the criterion (3.19) will be satisfied at a replacement $v_0 \to v_0^*$.

The problem is whether we can find in a more direct manner the positions of all other zeros in the $u=0$ plane, having chosen a resolution of the ambiguity of the v-zero trajectories. A method is clearly afforded by the following.

Remark: If $F_0(u,v)$ is a solution of the phase problem, then the function

$$Z(u,v) = \prod_{n=1}^{\infty} \left(1 - \frac{u}{u_n(v)}\right) \tag{3.23}$$

is a meromorphic function of v, for every fixed u; its poles are among the zeros of $F_0(0,v)$.

Indeed, from the fact that, at every fixed u, $F_0(u,v)$ is square integrable on the real axis and of exponential type equal to a constant b_0, independent of u, it follows from the Titchmarsh representation (3.2) that (we have seen we may assume $\alpha=0$)

$$Z(u,v) = F_0(u,v)/F_0(0,v) \tag{3.24}$$

which makes the remark obvious.

Thus, the discrete ambiguity of the zeros in the $u=0$ plane has to be solved in such a manner as to render $Z(u,v)F_0(0,v)$ holomorphic in v for all u.

Loosely speaking, we expect, e.g., for real u, the function $Z(u,v)$ to show structures as a function of real v; these structures have to disappear if we multiply them by the correct resolution of the zeros in $F_0(0,v)$. Assume now we start from an arbitrary resolution of the ambiguity of the v-zero trajectories. The obvious problem is to find out for how many u's we have to check that the combination $F(0,v)Z(u,v)$ is entire, in order to make sure that the resolution was correct. This is answered by the following.

Proposition 9: If for every integer m, $-\infty < m < \infty$, the functions $G_m(v) = F(0,v)Z(m\pi/a,v)$ are entire and of exponential type less than b, then there exists a solution of the phase problem, which vanishes at every real v at the same points as $Z(u,v)$.

Proof: Clearly, for real v, $|G_m(v)|^2 = M(m\pi/a,v)$, and since $M(m\pi/a,v)$ belongs to $L^1(\mathbf{R})$, $G_m(v)$ is the Fourier transform of a function with support contained in $(-b,b)$. Further, since $|G_m(n\pi/b)|^2 = M(m\pi/a, n\pi/b)$, the convergence of the series (3.4) implies that Eq. (3.18) with $F(m\pi/a, n\pi/b)$ replaced by $G_m(n\pi/b)$ defines an entire function of u and v, which we call $\widetilde{G}(u,v)$. We now sum the series first over n, leaving m fixed; the result of the summation is unchanged, since the double series converges absolutely. But $G_m(v)$ can be written as a Lagrange–Valiron expansion

$$G_m(v) = \sum_{n=-\infty}^{\infty} \frac{(-1)^n G_m(n\pi/b)}{bv - n\pi} \sin bv . \tag{3.25}$$

We obtain thus $\widetilde{G}(u,v)$ at each fixed v, as an entire function of type less than a in u, coinciding with $G_m(v)$ at each $u = m\pi/a$. To show that $\widetilde{G}(u,v)$ is a solution of the phase problem, we still have to show that it has the correct modulus in the Re u-Re v plane. But, at each real v, the function $G(u,v) \equiv F(0,v)Z(u,v)$ is, according to the first part of the proof of Proposition 7, entire of exponential type less than a with respect to u and coincides at $u = m\pi/a$ with $G_m(v)$. Thus, at every real v, $G(u,v) \equiv \widetilde{G}(u,v)$, for all u. However, $|G(u,v)|^2 \equiv M(u,v)$ for real u,v and this shows that $\widetilde{G}(u,v)$ is indeed a solution to the phase problem. This ends the proof.

This sufficient condition for the existence of a solution can be presumably much relaxed. It is indeed only under very special circumstances that one expects $G_m(v)$ to be entire for some values of m, but not for others. This is discussed in the next section in relation with the polynomial example.

To summarize this section, we notice we are now able to answer questions (i)–(iii) of the Introduction, for entire functions: (i) a modulus distribution $M(u,v)$ is admissible if Eqs. (2.19) are satisfied for some resolution of the discrete ambiguity of the zero trajectories; (ii) the extent of the ambiguity of the phase problem is given by the number of distinct resolutions of the ambiguity of trajectories, such that (3.19) is satisfied; (iii) the function $L(u,v)$, Eq. (3.5), represents the explicit solution of the phase problem, for each "correct" resolution of the ambiguity of trajectories.

In the above statements, an alternative condition to Eq. (3.19) for a correct resolution of the ambiguity of trajectories is available in the statement of Proposition 9.

IV. APPROXIMATE METHODS

One may object that the criteria proposed in the previous section for picking out solutions to the phase problem are not practical; they require checking an infinite number of equalities. Within the framework of entire functions, one can hardly expect simpler results.

In this section, we consider two practical simplifications: (i) an approximate method to find the correct resolution of the discrete ambiguity of the zero trajectories[6,7]; and (ii) the approximation of $F(u,v)$ by a polynomial in e^{iu}, e^{iv} (following Bruck and Sodin[1]).

(i) The v-zero trajectories are the sections of the various irreducible zero sets of $F(u,v)$ by the plane Im $v=0$. An irreducible zero set gives rise, in general, to several trajectories which, as a consequence, cannot be reflected independently of each other.

If we disregard the situation when zero trajectories $u = u_i(v)$ intersect on the plane $\mathrm{Im}\, v = 0$, they are holomorphic functions of v, in a certain neighborhood $d_r(v_0)$: $|v - v_0| < r(v_0)$ of any point v_0 on $\mathrm{Im}\, v = 0$ (the subscript i has been dropped). The radius $r(v_0)$ varies with the point v_0. In the limit of zero errors, the values of the zero trajectory $u(v)$ determine the radius $r(v_0)$ unambiguously. The disk $d_r(v_0)$ extends until the nearest algebraic branch point $\bar{v}$ of $u(v)$. If we continue around $\bar{v}$ and return to $\mathrm{Im}\, v = 0$, we obtain another zero trajectory $u_1(v)$, whose disk of analyticity around v_0 may also extend up to $\bar{v}$.

Assume now there exists a certain domain $D_u \times D_v$ of $\mathbb{C}^2$, so that D_v contains a segment of $\mathrm{Im}\, v = 0$, and has the property that for each v in D_v, there are only two solutions $u_1(v)$, $u_2(v)$ of $F(u,v) = 0$ in D_u. When two v-zero trajectories come close to each other in the u plane and are well separated from the others, one may expect that such domains exist. In this situation, the set of zeros of $F(u,v)$ in $D_u \times D_v$ is described by the equation

$$P(u,v) = u^2 + a_1(v)u + a_2(v) = 0\,, \tag{4.1}$$

with coefficients $a_i(v)$ that are holomorphic in D_v. The analyticity domain of the individual zero trajectories may be much smaller than D_v. However, the sum $u_1(v) + u_2(v)$ and the product $u_1(v)u_2(v)$ are holomorphic in D_v. On the other hand, if we replace, e.g., $u_2(v)$ by $u_2^*(v)$ the resulting sum and product are *not* holomorphic in D_v.

Therefore, in the limit of zero errors, one expects an increase of the radius of the disk of analyticity $d_r(v_0)$ as one moves from the individual trajectories to their sum and product, if they are "coupled" correctly. When the errors around the data $h(v)$ (individual trajectories or their sum and product) are of finite size ϵ, this test can be carried out only within some approximation; quantities $M_0(h,\epsilon)$ can be constructed, depending on the domain D_v, with the property that, as $\epsilon \to 0$, $M_0(h,\epsilon) \to \infty$ if the function h cannot be extended to a holomorphic function in D_v, whereas $M_0(h,\epsilon) \to$const if it can. The use of these tests has been described in many places.[17,18] In fact, most of the decisions can be taken by simply inspecting the local form of the zero trajectories: the nearby branch points can be seen clearly if the zero trajectories are close to the physical region, and the way they must be situated with respect to each other may be guessed immediately. In Sec. V, we give a rather extreme example of this.

With these tests, one is able to fix the relative position of some of the trajectories and exhaust this way the constraints coming from local two-variable analyticity. Once this is done, one can use Eqs. (3.19) to verify that the selected solutions are indeed correct.

The positions of the zeros in the $u = 0$ plane may be obtained by a simple extrapolation of the v-zero trajectories (see Refs. 6 and 7 for details at "intersections"). According to the remark of the previous section, one must in fact solve the discrete ambiguity of the zeros in the $u = 0$ plane, so that the singularities of $Z(u,v)$, Eq. (3.23), are removed.

Before closing this discussion of point (i), one should remember that the first paper in which the importance of zero trajectories for phase shift analysis in high energy physics was made clear is the one by Barrelet.[19]

(ii) We consider now the model used by Bruck and Sodin[1]

$$\Psi(x,y) = \sum_{m,n=1}^{M,N} a_{mn}\, \delta(x - m\Delta)\delta(y - n\Delta)\,, \tag{4.2}$$

$$F(u,v) = \sum_{m,n=1}^{M,N} a_{mn} \exp(imu\Delta)\exp(jnv\Delta)\,, \tag{4.3}$$

and introduce the variables $s = \exp(iu\Delta)$, $t = \exp(iv\Delta)$

$$F(u,v) \equiv P(s,t) = \sum_{m,n=1}^{M,N} a_{mn} s^m t^n\,, \tag{4.4}$$

so that the measured intensity distribution is a polynomial in s and t (divided by $s^M t^N$) of degree $2M$ and $2N$

$$M(s,t) = P(s,t)P^*(1/s^*, 1/t^*)\,. \tag{4.5}$$

An easy example of an irreducible (nonadmissible) intensity distribution is

$$M(s,t) = (s + it + 1)(1/s - i/t + 1) + 1\,. \tag{4.6}$$

In any case, at each fixed t, $|t| = 1$, the $2M$ zeros $s_i(t)$ of $M(s,t)$ may be grouped in pairs, complex conjugate with respect to $|s| = 1$, $s_{2k-1}(t) = 1/s_{2k}^*(t)$, $k = 1,\ldots,M$. When t leaves $|t| = 1$, $s_{2k-1}(t) = 1/s_{2k}^*(1/t^*)$. The question is now how to pick one trajectory out of each pair, if at all possible, so as to obtain a polynomial $P(s,t)$, obeying (4.5). If a choice has been made, a tentative solution reads

$$P(s,t) = \gamma(t) \prod_{i=1}^{M} (s + 1 - (s_i(t) + 1))\,, \tag{4.7}$$

where $\gamma(t)$ is determined by

$$P(-1,t) = \gamma(t)(-1)^M \prod_{i=1}^{M} (1 + s_i(t))\,. \tag{4.8}$$

In (4.8), we also assume that a choice has been made for the discrete ambiguity of the zeros in the t plane for $P(-1,t)$.

Now, expression (4.7) is a solution of the phase problem if $P(s,t)$ is a polynomial with respect to t. The analytic continuation of the functions $s_i(t)$ has branch points in the t plane and clearly $P(s,t)$ will be a solution if all these branch points cancel out in (4.7). By substituting (4.8) into (4.7) one verifies that this latter will be the case if the symmetric combinations

$$\begin{aligned}
\alpha_1(t) &= \sum_{i=1}^{M} \frac{1}{(1 + s_i(t))}\,, \\
\alpha_2(t) &= \sum_{i>j} \frac{1}{(1 + s_i(t))(1 + s_j(t))}\,, \\
&\vdots \\
\alpha_M(t) &= \prod_{k=1}^{M} \frac{1}{(1 + s_k(t))}
\end{aligned} \tag{4.9}$$

are meromorphic functions of t, with poles at those values for which $s_i(t) = -1$, i.e., at some of the zeros of $P(-1,t)$. Clearly, the resolution of the discrete ambiguity of the zeros of $P(-1,t)$ should be such that these poles are cancelled. Apart from the zeros needed for these cancellations, $P(-1,t)$ may contain, say, Q additional zeros which appear in $P(s,t)$ as s-independent factors of the form $(t - t_i)$. They generate a 2^Q ambiguity of the problem, which cannot be further reduced.

We conclude that a necessary condition for a correct resolution of the discrete ambiguity of the trajectories and of the zeros of $P(-1,t)$ is that the combinations

$$A_p(t) = P(-1,t)\alpha_p(t), \quad p = 1,2,\dots,M \tag{4.10}$$

are polynomials in t, of degree at most equal to N.

In analogy to Proposition 9, it is easy to see that this latter condition is also sufficient, i.e., if it is fulfilled, the function

$$\overline{P}(s,t) = \sum_{p=0}^{M} A_p(t)(s+1)^p(-1)^p \quad (A_0(t) = P(-1,t)) \tag{4.11}$$

is a solution of the phase problem. To see this, we only have to verify that $\overline{P}(s,t)$ has the correct modulus on $|s| = 1 \otimes |t| = 1$. But by construction, the function $\overline{M}(s,t) = \overline{P}(s,t)P^*(1/s^*,1/t^*)$ vanishes for $|t| = 1$ in the complex s plane at the same points as $M(s,t)$. Thus, the ratio $R(s,t) = \overline{M}(s,t)/M(s,t)$ is independent of s for each t on $|t| = 1$ and is thus independent of s for all t:$R(s,t) = g(t)$, and $g(t)$ can be fixed by letting $s = -1$. From (4.11) we see, however, that $g(t) \equiv 1$, and thus $P(s,t)$ is indeed a solution.

We can express the condition that $P(s,t)$ be a solution of the phase problem in a manner similar to Eq. (3.19). Choose to this end $N+1$ values t_k of t, $|t_k| = 1$ and construct by means of (4.11) the polynomials in s,$P(s,t_k)$. Let $P_1(s,t)$ be the polynomial in s and t that interpolates them. Then $P_1(s,t)$ is a solution of the phase problem if and only if, for $2M+1$ distinct values s_i of s on $|s| = 1$ and N further values t_k of t, $|t_k| = 1$

$$|P_1(s_i,t_k)|^2 = M(s_i,t_k), \quad i = 1,2,\dots,2M+1, \quad k = N+2,\dots,2N+1. \tag{4.12}$$

An obvious question that may be raised is the following: assume we know that a solution $P(s,t)$ exists. Is it then necessary to check that the $A_p(t)$, Eq. (4.10), are polynomials in t, for all p, or is it enough to consider only, say, $A_1(t)$? Unless special situations occur, we expect the latter to be the case. To support this, we consider the situation of a polynomial of the second degree in s and t and inquire, e.g., under what conditions, although the root $s_2(t)$ is the analytic continuation of $s_1(t)$ around one of its branch points at t_0, and thus the product $s_1(t)s_2(t)$ is holomorphic near t_0, nevertheless, $s_1(t) \times 1/s_2^*(1/t^*)$ is also holomorphic near t_0. This can happen only if $s_1(t)$ has a branch point also at $1/t_0^*$, with an appropriate discontinuity. This is clearly exceptional. It happens, however, in the example

$$P(s,t) = s^2(4t + i(1-t^2)) - 2s(1-t)^2 + (4t - i(1-t^2)). \tag{4.13}$$

It is only the sum $s_1(t) + s_2(t)$ which distinguishes the wrong from the correct solution.

We now apply the method of Ref. 20 to the problem of finding a polynomial $P_0(s,t)$ that vanishes for $|t| = 1$ on the zero trajectories $s = s_i(t)$. The method is more involved than the previous procedure, and actually suited for more complicated topologies.

One starts by drawing the curves $\gamma_i : s = s_i(t)$, $|t| = 1$, $i = 1,2,\dots,M$ in the complex s plane and parting it this way into (a finite number of) domains $\mathscr{D}_j$, $j = 1,2,\dots$. One can show (see Ref. 20) that the number of roots $t = t_j(s)$ of the equation $P_0(s,t) = 0$, lying inside (or outside) the unit disk $|t| \leqslant 1$ is constant as long as s lies inside one $\mathscr{D}_j$. Further, knowing the number of roots of $P_0(s,t) = 0$ for $s = -1$ in the t plane, i.e., in one $\mathscr{D}_j$, say $j = 1$, containing $s = -1$, we can deduce the number N_j of such roots in all $\mathscr{D}_j$.

Assume further $P_0(s,t)$ is known and consider, e.g., those roots lying outside $|t| \leqslant 1$; in each $\mathscr{D}_j$, the sums

$$S_{k,j}(s) = \sum_{l=1}^{N_j} \frac{1}{(t_l(s))^k}, \tag{4.14}$$

over the N_j zeros lying outside $|t| \leqslant 1$ for s in $\mathscr{D}_j$, are holomorphic functions of s for s in $\mathscr{D}_j$. If one of the $\mathscr{D}_j$'s extends to infinity, $S_{k,j}(s) \to \text{const}$, as $|s| \to \infty$. It is possible to show that one can write in each $\mathscr{D}_j$ a Cauchy integral for $S_k(s)/(s+1)$

$$S_{k,j}(s) = \delta_{j,-1} S_k(-1) + \frac{(1+s)}{2\pi i} \oint_{\partial \mathscr{D}_j} \frac{S_{k,j}(s')}{(1+s')(s'-s)} ds', \tag{4.15}$$

where $\delta_{j,-1}$ means that the term is to be counted only if $-1 \in \mathscr{D}_j$ and $\partial \mathscr{D}_j$ is the boundary of $\mathscr{D}_j$. Since the discontinuities over $\partial \mathscr{D}_j$ are known, one can add all equations (4.15) to obtain (see Ref. 20)

$$\Omega_k(s) = S_k(-1) + \sum_j \frac{1}{2\pi i} \int_{\gamma_j} \frac{(s+1)}{t_j^k(s')(s'-s)(s'+1)} ds', \tag{4.16}$$

where the integration is performed in the sense induced on γ_j by the clockwise motion on $|t| = 1$ and $t_i(s')$ is defined by $t_i(s_i(t')) = t'$, $|t'| = 1$. The function $\Omega_k(s)$ has the property that, in each $\mathscr{D}_j$,

$$\Omega_k(s) = S_{k,j}(s). \tag{4.17}$$

From the $S_{k,i}(s)$ one can obtain algebraically the symmetric combinations $\alpha_{k,i}(s)$ so that, for s in each $\mathscr{D}_j$, the roots $t = t_i(s)$ of $P_0(s,t) = 0$, lying outside $|t| \leqslant 1$, are obtained by solving the equation

$$P_{1,j}(s,t) = \sum_{p=0}^{N_j} \alpha_{p,j}(s) \left(\frac{1}{t}\right)^{N_j - p}, \quad \alpha_{0,j} = 1. \tag{4.18}$$

In a similar manner, we obtain a polynomial in t, $P_{2,j}(s,t)$, with coefficients $\beta_{p,j}(s,t)$ that vanish in each $\mathscr{D}_j$ at those roots of $P_0(s,t) = 0$ that lie inside $|t| \leqslant 1$. Then, one may verify that

$$P(s,t) \equiv P_{1,j}(s,t)P_{2,j}(s,t)t^{N_j}/\alpha_{N_j,j}(s) \tag{4.19}$$

is a polynomial of constant degree in t, independent of j, with the coefficient of the highest power of t equal to unity and with the other coefficients rational functions of s, independent of j. This polynomial vanishes for $|t| = 1$ on $s = s_i(t)$. Multiplying by the common denominator of the coefficients, we obtain a polynomial in s and t which differs from $P_0(s,t)$ at most by factors containing s only.

Now, the $\Omega_k(s)$, Eq. (4.16), may be constructed, both for $|t| \leqslant 1$ and for $|t| \geqslant 1$ starting from any resolution of the discrete ambiguity of the trajectories $s_i(t)$ and from knowledge of the $S_k(-1)$, as above. Also, given the $s_i(t)$, one can always assign numbers N_j of roots $t = t_j(s)$ in $|t| \leqslant 1$ (and $|t| \geqslant 1$, in turn) of the equation $P(s,t) = 0$ for s in each domain $\mathscr{D}_j$, starting, e.g., from the domain that contains $s = -1$. One would like to know where the construction leading to (4.19) fails if we start from a wrong resolution or if the constants $S_k(-1)$ in Eq. (4.16) are assigned wrong values.

The answer is that the manner in which $\Omega_k(s)$ was constructed contains no information about the number of roots $t_j(s)$ of $P(s,t)=0$ for s in $\mathscr{D}_j$, that appear in the sums (4.14) and (4.17). This number is determined by different methods and *a priori* an incompatibility might appear. Indeed, we recall the coefficients $\alpha_{p,j}(s)$ in Eq. (4.18) are obtained from the sums $S_{p,j}(s)$ by means of Newton's formulas (Ref. 21)

$$\alpha_{0,j}(s)=1,$$

$$\alpha_{p,j}(s)=\frac{(-1)^p}{p!}\times\begin{vmatrix} S_{1,j}(s) & 1 & \cdots & 0 \\ S_{2,j}(s) & S_{1,j}2 & \cdots & 0 \\ \vdots & & & p-1 \\ S_{p,j}(s) & S_{p-1,j} & \cdots & S_{1,j}\end{vmatrix}. \tag{4.20}$$

If the sum $S_{p,j}(s)$ contains N_j roots, then (4.20) leads to

$$\alpha_{p,j}(s)\equiv 0, \quad \text{for } p\geqslant N_j+1\,. \tag{4.21}$$

However, given a sequence $S_{k,j}(s)$, $k=1,2,\ldots$ of numbers, as is generated by (4.16) with an arbitrary choice of $S_k(-1)$ and of the resolution of the discrete ambiguity of the $s_i(t)$, condition (4.21) will fail to be satisfied.

In fact, we shall show that, if (4.21) is satisfied, then one can indeed construct a polynomial $P(s,t)$ by (4.19), which vanishes for $|t|=1$ precisely on $s=s_i(t)$ and is thus a solution to the problem. We assume for simplicity that the curves γ_j have unit multiplicity so that for two adjacent domains, $N_j-N_{j-1}=1$ (see Ref. 20) ($\mathscr{D}_j$ lies to the left of γ_j). By assumption, we can then write polynomials (4.18) of degree N_j and N_j-1 for $\mathscr{D}_j$ and $\mathscr{D}_{j-1}$ in turn, so that the difference of the corresponding sums of roots, Eqs. (4.14), for s on γ_j, is

$$S_{j,k}(s)-S_{j-1,k}(s)=1/(t_j(s))^k, \quad k=1,2,\ldots,N\,. \tag{4.22}$$

It follows then that the two polynomials (4.18) for $\mathscr{D}_j$ and $\mathscr{D}_{j-1}$ have, for s on γ_j, all roots equal, except for one, which is equal to $t_j(s)$, $|t_j(s)|=1$. Thus, as $s\to s_0\in\gamma_i$, one of the zeros of $P_{1,j}(s,t)$ tends to $|t|=1$, so that $s_i(t)=s_0$. The same is then true for $P_{2,j}(s,t)$; it follows that the combination $P(s,t)$ in Eq. (4.19) is independent of j, with coefficients that are holomorphic functions of s, except for isolated singularities, at those points with $\alpha_{N_p,j}(s)=0$. It can be argued that these singularities are poles, so that the coefficients are rational functions of s, as desired.

Notice, conditions (4.21) fix the possible values of the sums $S_k(-1)$ and thus of the extrapolation of the trajectories to $s=-1$. In the conclusions of Ref. 20, condition (4.21) was missed as one of the sufficient conditions for a resolution of the discrete ambiguity of the zero trajectories to be compatible with two-variable analyticity.

V. DISCUSSION OF AN EXAMPLE AND CONCLUSIONS

Loosely speaking, the violations of two-variable analyticity through "wrong" reflections of zero trajectories stem from the presence of "uncompensated" branch points in each trajectory. In Ref. 6, in another context, it was shown in examples how such branch points lead to easily recognizable distortions of the trajectories near "intersections," if these occur sufficiently close to the Re u-Re v plane.

Examples in optics in which intersections occur near to the Re u-Re v plane are easily manufactured by perturbing the diffraction pattern on a rectangular aperture. We take, for instance,

$$F(u,v)=\sin u\,\sin v/uv - \lambda\,[(u\cos u-\sin u)(v\cos v-\sin v)/u^2v^2]\,, \tag{5.1}$$

for small complex values of λ. There are intersections near the points $u=m\pi$, $m\neq 0$, $v=n\pi$, $n\neq 0$. The pattern of trajectories near intersections appears in its typical form if we move over to the variables

$$s=(u+v)/2, \quad t=(u-v)/2 \tag{5.2}$$

and look, e.g., for t-zero trajectories, that is, we find at each fixed t, those complex s values where $F(s,t)=0$.

Figure 1 shows the projections onto the Re s-Re t plane of the t-zero trajectories, for $\lambda=0.1\exp(i\pi/4)$. Figure 2

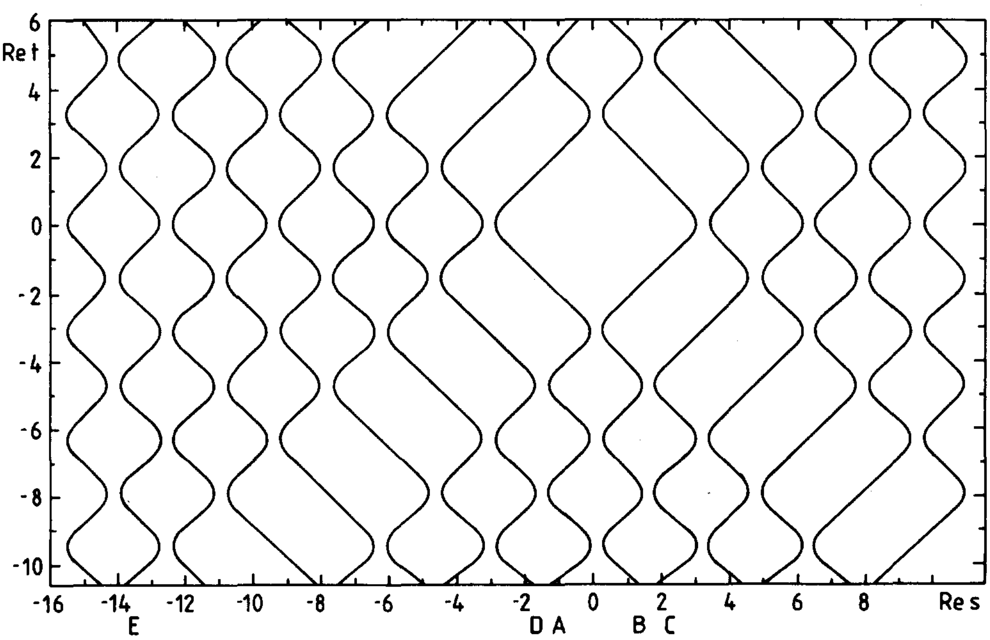


FIG. 1. Pattern of the projections of the t-zero trajectories for the model (5.1) onto the Re s-Re t plane.

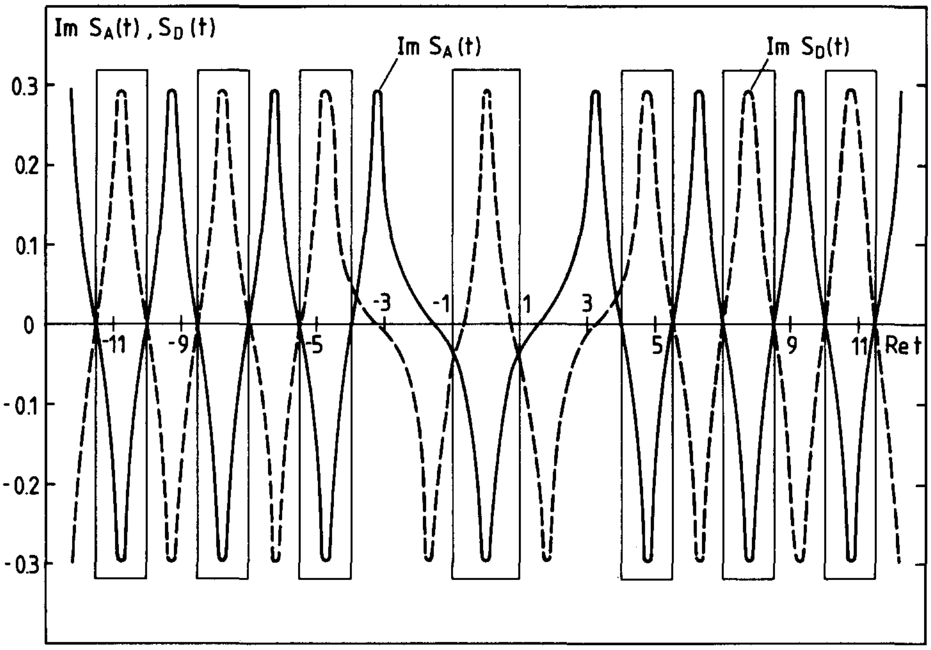


FIG. 2. The imaginary parts of the t-zero trajectories A and D as a function of t. The intersection regions are enclosed in a rectangle.

shows the imaginary part of the trajectories labeled A and B (dotted) for the same λ and a certain range of t. The intersection regions have been surrounded by a thick line. The other structures are due to branch points at intersections with other trajectories. In Figs. 3(a) and 3(b), we show the imaginary parts of the sum and product of the two trajectories [the coefficients $a_1(v)$, $a_2(v)$ in (4.1), with v replaced by t] and the dotted lines show the same quantities if trajectory B were reflected and, thus, its branch points left "uncompensated." Any test of analyticity will indicate that the thick line is compatible with quite smooth analytic functions inside a disk centered at $t_0 = -4.7$, whereas the dotted curves require "wild" analytic approximants.

These qualitative tests indicate that, in fact, the intensity distribution pertaining to (4.1) admits only of the trivial ambiguity $F(u,v) \mapsto F^*(u^*,v^*)$ in the determination of the scattered field.

This is supported by an application of the "exact" tests (3.19) or (3.22). In a numerical calculation, a truncation of the Lagrange–Valiron series, Eq. (3.5), has to be performed and the construction of $F(u,n\pi/b)$ must be achieved by means of a finite Titchmarsh product. The convergence of the latter is very slow and other methods (e.g., integral representations for $\log|F|$) have to be used in order to construct $F(u,n\pi/b)$. This was unnecessary, as a test of Eqs. (3.19) or (3.22), since the exact expression of the solution $F(u,n\pi/b)$ was available. The equalities (3.19) give too-detailed information. We use (3.22) with the integral restricted to the domain $-3\pi \leqslant u, v \leqslant 3\pi$ and assume a mean error $\epsilon = 10^{-3}$ for $M(u,v)$. The integral in (3.22) must then be divided by ϵ^2. Table I gives an example of the increase of the integral if one or several trajectories are reflected in the "wrong" manner or if the ambiguity of the zeros in the $u = 0$ plane is solved in a "wrong" way. The fact that even for the correct choice of the ambiguities we do not obtain strictly zero is due to the truncation of the Lagrange–Valiron series.

As pointed out in Sec. III, the Fourier transform with respect to v of $L(u,v)$, Eq. (3.4), at fixed u, has a support that is independent of the value of the parameter b used in the construction, if the resolution of the ambiguity of the trajectories was correctly performed. On the contrary, the support increases with increasing b if this is not the case. This effect is shown in Fig. 4, for the example (5.1), at $s = -1.1$ and the reflection of the trajectories A and C.

Figure 5 shows an example of a pole at complex v in $Z(u,v)$ and the way it is removed in the complete function, in relation with Proposition 9. The dotted line shows the effect of the multiplication by a "wrong" zero of $F(0,v)$.

To summarize the difference between the one- and two-dimensional versions of the phase problem, it is profitable to part the latter into four more precise questions (similar to those of the Introduction), that are typical of inverse problems[22]: (i) uniqueness; (ii) reconstruction; (iii) characterization; and (iv) construction (stability). Question (i) concerns the extent of the ambiguity allowed by an intensity distribution: $I(u,v) = |F_0(u,v)|^2$, with $F_0 \in E_{a,b}$; at each fixed v (or u), the one-dimensional ambiguity is infinite; in two dimensions, it is less than or equal to 2^{N_z} with N_z the number of irreducible zero sets of $F_0(u,v)$ (Sec. II). It is reasonable to conjecture that it is in fact equal to 2^{N_z}, but the proof, if it exists, is not a generalization of the one-variable case (see Appendix A). Question (ii) requires the construction of all solutions $F(u,v)$ in $E_{a,b}$ with the same modulus as $F_0(u,v)$. The answer is simple at fixed v (or u). In two dimensions, this may be done by the construction in Appendix A, taken from Ref. 11, but this requires knowledge of $F_0(u,v)$ in all of $\mathbb{C}^2$. An alternative construction is given in Sec. III, based on the zero trajectories of $F_0(u,v)$, supplemented by a criterion to select the correct resolution (Propositions 7–9). We have argued that this method may be improved, if one resorts to approximate methods of grouping trajectories into irreducible sets (Sec. IV). Question (iii) requires criteria for a modulus $M(u,v)$

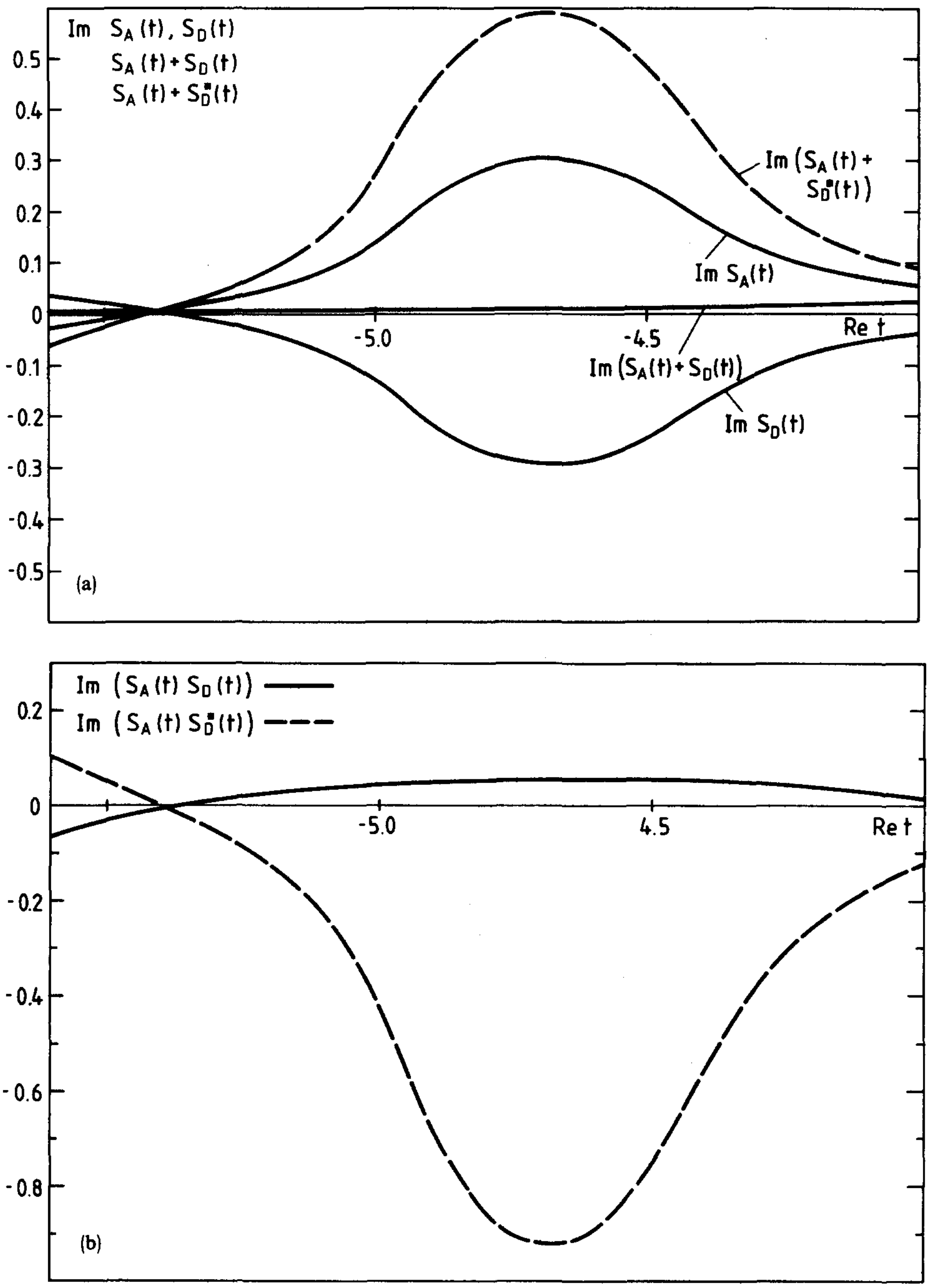


FIG. 3. (a) The imaginary parts of the A and D trajectories in a certain interval of t values; also shown are their sum for the correct and a "wrong" resolution (dotted line) of the discrete ambiguity. (b) The imaginary part of the product of the A and D trajectories for a correct and a "wrong" resolution (dotted line) of the discrete ambiguity of the zero trajectories.

whose Fourier transform has a compact support, to be an admissible intensity. No such criteria are needed in one dimension. In two dimensions, a necessary condition is that the set of zeros of $M(u,v)$ should consist of pairs of irreducible zero sets, conjugate to each other, which is a rare occurrence (see Appendix C). No conditions are as yet established that are both necessary and sufficient, since it is not clear that at

TABLE I. Values of the integral (3.22).

Situation	Integral
No reflection	1.91
Reflection of A	360.2
Reflection of A and C	654.2
Reflection of E	12.1

least one choice of elements of each of these pairs can be made, so that one obtains a zero set of a function of exponential type. This is a very interesting problem. A (rather complicated) way to verify whether $M(u,v)$ is admissible, using zero trajectories, is provided again by Propositions 7–9 in Sec. III.

Question (iv) requires the construction of all solutions inside a certain error channel around an experimental $I(u,v)$. This is, of course, very difficult to answer exactly. Part of this problem consists of establishing the degree of ambiguity admitted by the various moduli surrounding $I(u,v)$. This degree is unstable in two dimensions. The fact that it decreases under perturbations is shown in Ref. 3 for polynomials; Propositions 5 and 6 generalize this to some extent for entire functions (see also Appendix B).

Finally, it would be of considerable interest to understand the difference between the one- and two-dimensional

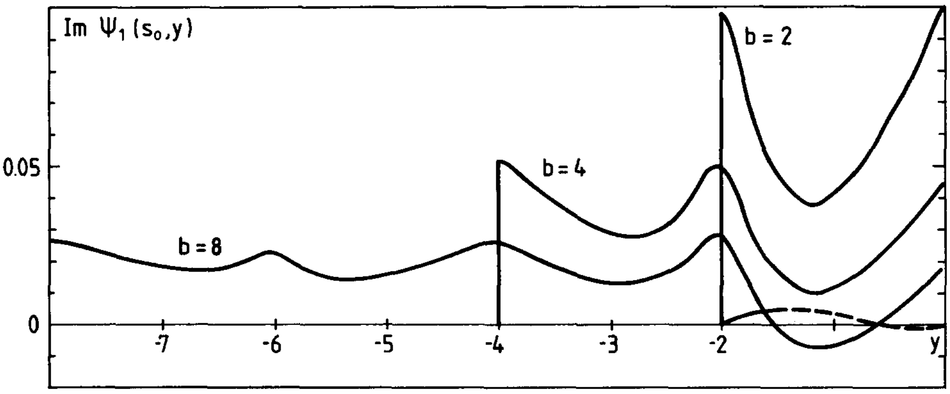


FIG. 4. The increase of the support of the imaginary part of the Fourier transform $\psi_1(s_0,y)$ of $L(s_0,t)$ Eq. (3.5), for increasing values of b ($s_0 = -1.1$) and the model (5.1).

ambiguities in terms of more standard, iterative approaches to the solution of the phase problem. It has been reported that such algorithms (Refs. 23–25) show a better convergence to the correct phase in two dimensions, than in one. However, since these procedures have no connection with the irreducibility of zero sets, they suffer from the drawback that they cannot distinguish whether a given intensity distribution is admissible or not. It is in this direction that, in the opinion of the author, an improvement is necessary and he hopes to return to these questions in the future.

APPENDIX A: ON PROPOSITIONS 3 AND 4

We first reformulate, using the notation of the present paper, the generalization to n dimensions of Weierstrass' representation for entire functions, due to Lelong (Theorem 5 of Ref. 13). We consider only $n = 2$; here are some definitions.

Let $\mathscr{Z}_F$ be the set of zeros of some entire function $F(u,v)$. Then $[z = (u,v),\ z \in \mathbb{C}^2]$

$$\nu(a) = \frac{1}{2\pi^2 a^2} \int_{\|z\|<a} \nabla^2 \log|F(u,v)|\,du\,dv \tag{A1}$$

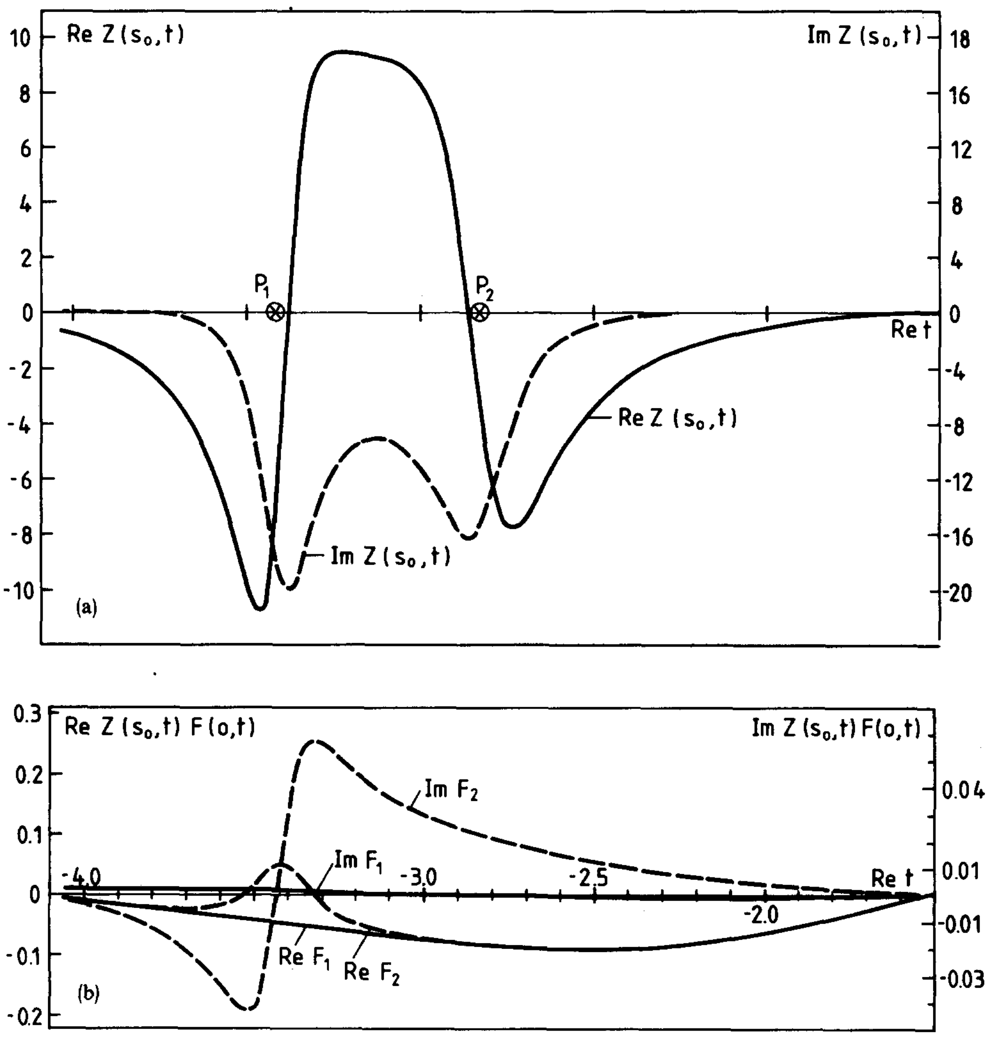


FIG. 5. (a) The poles present in $Z(s_0,t)$ for $s_0 = \pi/4$. The points P_1 and P_2 show the real part of the pole positions. (b) The result of the multiplication of $Z(s_0,t)$ by $F(0,t)$ with the correct resolution of the discrete ambiguity of the zeros (F_1) and with the zero whose real part is given by P_1 replaced by its complex conjugate (F_2).

is called the mean degree of the set of zeros of $F(u,v)$ over the ball of radius a and

$$\chi_{\mathscr{Z}} = \overline{\lim_{a\to\infty}} \, [\log \nu(a)/\log a] \tag{A2}$$

is called the order of increase of $\mathscr{Z}_F$ as $a\to\infty$. An intuitive definition of $\nu(a)$ is given in Ref. 26, where it is also shown that it is an increasing function of a. Define further, for any entire function $G(u,v)$, with

$$M(a) = \sup_{\|(u,v)\| < a} |G(u,v)|\,, \tag{A3}$$

$$\alpha_G = \overline{\lim}\,[\log M(a)/\log a] \tag{A4}$$

as the order of increase of G as $a\to\infty$.

Further, let $h_2(a,z)$ be the Green's function with a singularity at a and vanishing at infinity, for Laplace's equation in $\mathbf{R}^4$

$$h_2(a,z) = \|a-z\|^{-2}\,. \tag{A5}$$

Denote

$$\cos v = (a,z)/(\|a\|\,\|z\|), \quad u = \|z\|/\|a\|\,. \tag{A6}$$

Then for small u

$$h_2(a,z) = \|a\|^{-2}(1 + C_1^1(\cos v)u + C_2^1(\cos v)u^2 + \cdots)\,, \tag{A7}$$

where the $C_n^1(\cos v)$ are Gegenbauer polynomials (see Ref. 27). One defines the truncated kernel of order q, with singularity at a, as

$$e_2(a,z,q) = -h_2(a,z) + \|a\|^{-2}(1 + C_1^1(\cos v)u + \cdots + C_q^1(\cos v)u^q)\,. \tag{A8}$$

It is still a Green's function of Laplace's equation, but with a more rapid increase at infinity. However, if $\rho(a)$ is a "charge distribution" in $\mathbf{R}^4$, the potential

$$V(z) = \frac{1}{4\pi^2}\int \rho(a)e_2(a,z,q)d^4a \tag{A9}$$

converges even if $\rho(a)$ increases at infinity like $a^{q-1-\epsilon}$, for any $\epsilon > 0$. The analogy to one complex dimension is obtained by letting

$$h_1(a,z) = -\log|a-z|\,. \tag{A10}$$

One defines the genus of $\mathscr{Z}_F$ as the smallest integer q for which

$$\int t^{-q}\,d\nu(t) \tag{A11}$$

converges (it is assumed that $0 \notin \mathscr{Z}_F$). If $\chi_{\mathscr{Z}}$ is not an integer, then $q = [\chi - 1]$. Now we can state the following.

Theorem[13]: If $\mathscr{Z}_F$ has order of increase χ, then we have the following.

(i) There exists an entire function $F_0(u,v)$ with order of increase χ, and having $\mathscr{Z}_F$ as its zero set.

(ii) $F_0(z)$ is defined by $(z = (u,v))$, $z \notin \mathscr{Z}_F$

$$\log|F_0(z)| = \frac{1}{4\pi^2}\int \nabla^2 \log|F(a)|e_2(a,z,q)d^4a\,, \tag{A12}$$

where q is the genus of $\mathscr{Z}_F$

$$F_0(z) = 2\int_0^z \frac{\partial \log|F_0(a_1,a_2)|}{\partial a_1}\,da_1 + \frac{\partial \log|F_0(a_1,a_2)|}{\partial a_2}\,da_2\,, \tag{A13}$$

where the integration is done along a path joining 0 to (u,v) and avoiding $\mathscr{Z}_F$.

(iii) Any other entire function $G(u,v)$ vanishing at every point of $\mathscr{Z}_F$ has order of increase at least equal to χ and is divisible by $F_0(u,v)$.

To avoid confusion, it should be stated that one must show not only that (A12) verifies the Poisson equation in $\mathbf{R}^4$

$$\nabla^2 \log|F_0| \equiv \left(\frac{\partial^2}{\partial u\,\partial \bar u} + \frac{\partial^2}{\partial v\,\partial \bar v}\right)\log|F_0| = \nabla^2 \log|F|\,, \tag{A14}$$

(which it does by construction), but that it can be the real part of the logarithm of a solution of the second Cousin problem. This means $\log|F_0|$ must satisfy

$$\frac{\partial^2 \log|F_0|}{\partial z_i\,\partial \bar z_j} = \frac{\partial^2 \log|F|}{\partial z_i\,\partial \bar z_j}, \quad i,j = 1,2, \quad z_1 = u, z_2 = v\,. \tag{A15}$$

The proof of Proposition 3 is easy with the help of the Theorem above. Indeed, we assume that $\mathscr{Z}_F$ is the set of zeros of a solution of the phase retrieval problem, i.e., of an entire function of exponential type. It follows from part (iii) of the Theorem that the order of increase of $\mathscr{Z}_F$ is at most unity.

Now, $\nu(a)$, Eq. (A1), can be written as the sum of the contributions of each irreducible zero set that intersects $\|z\| < a$ $[z = (u,v)]$. There can only be a finite number of such sets, for each finite a. We verify next that $\nu(a)$ is unchanged by the replacement of any of these zero sets by their complex conjugates. To this end, we describe $\mathscr{Z}_F$ as the zero set of the product (2.2), where each factor corresponds to an irreducible subset of $\mathscr{Z}_F$. (The order of the product need not be unity.) The reflection of such a subset is described by the replacement of a factor $G_i(u,v)$ in (2.1) by $G_i^*(u^*,v^*)$. We can then write, with $u,v = r_i \exp(i\theta_i)$, $i = 1,2$, $0 < \theta_i < 2\pi$, and the substitution $\chi_i = -\theta_i + 2\pi$

$$\nu_{G_i^*}(a) = \frac{1}{2\pi^2 a^2}\int_{r_1^2+r_2^2<a^2} \nabla^2 \log|G_i^*(r_1e^{-i\theta_1},r_2e^{-i\theta_2})| \times r_1\,dr_1\,d\theta_1\,r_2\,dr_2\,d\theta_2$$

$$= \frac{1}{2\pi^2a^2}\int \nabla^2 \log|G_i(r_1e^{i\theta_1},r_2e^{i\theta_2})| \times r_1\,dr_1\,r_2\,dr_2\,d\theta_1\,d\theta_2 = \nu_{G_i}(a)\,. \tag{A16}$$

This verifies that $\nu(a)$ is unchanged if a number of zero sets are reflected. By part (i) of the Theorem, there exists an entire function of order unity, $F_0'(u,v)$, vanishing on the new zero set, obtained from $\mathscr{Z}_F$ by the replacement of $G_i(u,v)$ by $G_i^*(u^*,v^*)$, and nowhere else, as asserted by Proposition 3.

Such a function is given by (A12) and (A13), where F may be taken as the product (2.2) with any number of factors $G_i(u,v)$ replaced by $G_i^*(u^*,v^*)$. Writing (A12) as a sum over the irreducible zero sets of $F_0(u,v)$ $[q = 1, z = (u,v)]$

$\log|F_0(z)|$

$$= \lim_{R\to\infty} \sum_{\mathscr{M}_i \cap \|z\| < R} \frac{1}{4\pi^2} \int \nabla^2 \log|G_i(a)| e_2(a,z,1) d^4 a \,, \tag{A17}$$

one can verify that, under the replacement $G_i(a) \to G_i^*(a^*)$, the value of the integral stays unchanged for u,v real, so that $\log|F_0(u,v)| = \log|F_0'(u,v)|$ in the $\operatorname{Im} u = 0 \otimes \operatorname{Im} v = 0$ plane. This completes the proof of Proposition 3.

We next explain why we cannot exclude, by arguments similar to the one-variable case, that reflections of irreducible zero sets may lead to functions of infinite type. To this end, we evaluate the integrals $[z = (u,v)]$

$$I(R) = \int_{\substack{\|z\|<R \\ \operatorname{Im} u>0}} \nabla^2 \log|F(u,v)| \frac{\operatorname{Im} u}{\|z\|^4} du\, dv \tag{A18}$$

and show that they diverge for the special case when $F(u,v) = F_1(u)F_2(v)$ with F_1 and F_2 one-dimensional Fourier transforms. To be sure, this does not disprove the conjecture that to the reflection of an arbitrary number of irreducible zero sets there corresponds a solution of the phase problem.

To prove our statement, we separate (A17) into two integrals corresponding to the partition

$$\log|F(u,v)| = \log|F_1(u)| + \log|F_2(v)| \,, \tag{A19}$$

and use the volume element

$$du_x\, du_y\, dv_x\, dv_y = r_1\, dr_1\, r_2\, dr_2\, d\theta_1\, d\theta_2 \,. \tag{A20}$$

The integral containing $\log|F_1(u)|$ reduces to the sum of the imaginary parts of the zeros of $F_1(u)$, lying in $\operatorname{Im} u \geqslant 0$, as $R \to \infty$. This is convergent if $F_1(u)$ is a Fourier transform. The second integral leads, however, to the evaluation of

$$S = \sum_{r_{2n}<R} \left[\frac{1}{2r_{2n}} \arctan\left(\frac{(R^2 - r_{2n}^2)^{1/2}}{r_{2n}} \right) - \frac{(R^2 - r_{2n}^2)^{1/2}}{2R^2} \right], \tag{A21}$$

where r_{2n} are the radii of the positions of the zeros of $F_2(v)$. Using the fact that $r_{2n} \sim n$ for Fourier transforms, one verifies that the sum over the second terms in brackets converges, as $R \to \infty$, whereas that over the first one diverges logarithmically with R. Therefore, we may not yet exclude that irreducible zero sets of $F_0(u,v)$ exist, such that individually they support only entire functions of order 1 and of infinite type, and satisfy the generalization of Lindelöf's criterion, Eq. (2.11), only when they occur in pairs.

APPENDIX B: THE INSTABILITY OF REDUCIBLE ZERO SETS

It follows from Ref. 5 that in any neighborhood (in $\mathbf{R}^{(N+1)(M+1)}$) of a given reducible polynomial $P(u,v)$ [of maximal degree (M,N)], there exist irreducible polynomials, i.e., that reducibility is an unstable property. In this Appendix, we wish to prove a similar, but weaker result for entire functions (and in fact, for functions holomorphic in some domain) namely: let (u_0,v_0) be a point of the zero set $\mathscr{Z}$ of $F_0(u,v)$, so that $\mathscr{Z}$ is reducible with respect to any (sufficiently small) neighborhood of (u_0,v_0). Then, in any neighborhood of $F_0(u,v)$ in the sense of Eq. (2.12), there exists $F(u,v)$ so that the set $\mathscr{Z}$ of zeros of F is irreducible with respect to a (sufficiently small) neighborhood $\mathscr{U}$ of (u_0,v_0).

The method of proof may be used also to establish the result of Ref. 5 for polynomials. It should also be added that, for a reader of, e.g., Ref. 10, the statement above is not unexpected. However, the present author was unable to find it anywhere, in this or a related (accessible) form.

We start by applying Weierstrass' preparation theorem to $F_0(u,v) \in E_{a,b}$ at (u_0,v_0) and write it in some neighborhood of (u_0,v_0) as

$$F_0(u,v) = P_0(u,v)\Omega(u,v) \,, \tag{B1}$$

where, by assumption, $P_0(u,v)$, a polynomial in v of degree N, with peak at (u_0,v_0), is reducible to

$$P_0(u,v) = P_{0,1}(u,v)P_{0,2}(u,v) \tag{B2}$$

and $\Omega(u_0,v_0) \neq 0$ and is holomorphic. As a consequence of (B2) not only $\partial P_0/\partial v(u_0,v_0) = 0$ but also $\partial P_0/\partial u(u_0,v_0) = 0$. We shall assume for simplicity that $P_{0,1}(u,v), P_{0,2}(u,v)$ are irreducible and distinct, but it is very easy to allow for any number of factors.

Consider now a perturbation $\delta F = \lambda \Omega_1(u,v)$ of F_0, with $\Omega_1 \in E_{a,b}$. The zeros of $F_0 + \delta F$ are, near (u_0,v_0), the same as those of $P_0(u,v) + \lambda G(u,v)$, with $G(u,v) = \Omega_1(u,v)/\Omega(u,v)$, holomorphic near (u_0,v_0). One may choose now a number of ϵ and a neighborhood $U \times V$ of (u_0,v_0) such that, for every λ with $|\lambda| < \epsilon$ and $u \in U$, $P_0 + \lambda G$ vanishes precisely N times inside V. We can then write a polynomial in $(v - v_0)$ of degree N

$$P(u,v,\lambda) = (v - v_0)^N + a_1(u,\lambda)(v - v_0)^{N-1} + \cdots + a_N(u,\lambda) \,, \tag{B3}$$

with coefficients depending on Ω_1, holomorphic in u and λ, $u \in U$, $|\lambda| < \epsilon$ and such that $P(u,v,0) = P_0(u,v)$, $a_i(0,0) = 0$, $i = 1,\ldots,N$ and $P(u,v,\lambda) = 0$ there, where $P_0 + \lambda G = 0$ for (u,v) in $U \times V$.

Our purpose is to find the constraints on $\Omega_1(u,v)$, which ensure that the corresponding $P(u,v,\lambda)$, Eq. (B3), is irreducible for any λ small enough. To this end, we shall show that (i) if, for some sequence $\lambda_n \to 0$, $P(u,v,\lambda)$ is reducible, then, for high enough n, it may have only two factors with the same degree with respect to v as the two factors of $P(u,v)$; (ii) consequently, there must exist points $(\tilde{u}_n,\tilde{v}_n) \to (u_0,v_0)$, so that

$$\frac{\partial P}{\partial u}(\tilde{u}_n,\tilde{v}_n,\lambda_n) = \frac{\partial P}{\partial v}(\tilde{u}_n,\tilde{v}_n,\lambda_n) = P(\tilde{u}_n,\tilde{v}_n,\lambda_n) = 0 \,; \tag{B4}$$

and (iii) for all $\Omega_1(u,v) \in E_{a,b}$, with the exception of a set of codimension 1, there is a neighborhood $W = U_1 \times V_1 \times \{|\lambda| < \epsilon_1\}$ of $(u_0,v_0,0)$ so that (B4) is impossible for $(u_n,v_n,\lambda_n) \in W$. This will establish our statement.

To prove (i), we proceed as in the proof of Proposition 5; given two numbers $r_1,r_2 > 0$, so that the set C_{12}: $r_1 < |u - u_0| < r_2$ is contained in U, we can find a covering of the set of zeros of $P_0(u,v)$ contained in $C_{12} \times V$ by a finite number of polydisks $d_{\delta,j} = d_{\delta,ju} \times d_{\delta,jv}$, of radius δ, so that at each fixed $u \in d_{\delta,ju}$, there exists just one root of $P_0(u,v)$ contained in $d_{\delta,jv}$. Further, we can choose ϵ_2 as a function of δ so small that the set of zeros of $P(u,v,\lambda)$ in $C_{12} \times V$ stays contained inside $\cup d_\delta$, for $|\lambda| < \epsilon_2(\delta)$.

Now let $P_{n,k}(u,v)$, $k=1,2,\ldots$, be the (nontrivial) irreducible factors of $P(u,v,\lambda_n)$. We may index as $P_{n,1}(u,v)$ that factor for which one of the roots $v_{n,1}(u)$ lies in the same disk $d_{\delta,jv}$ for a fixed j, as a root $v_{0,1}(u)$ of $P_{0,1}(u,v)$, for u in $d_{\delta,ju}$. Then, however, as argued in Proposition 5, the analytic continuation of $v_{n,1}(u)$ to C_{12} stays close to the continuation of $v_{0,1}(u)$ throughout C_{12}; that is, if to $u\in C_{12}$ we associate $d_{\delta k}$ so that $u\in d_{\delta k,u}$, $v_{0,1}(u)\in d_{\delta k,v}$, then $v_{n,1}(u)\in d_{\delta k,v}$. It follows that, if $N_{n,1}$ is the degree of $P_{n,1}(u,v)$ with respect to v, and $N_{0,1}$ that of $P_{0,1}(u,v)$, then $N_{n,1}\geqslant N_{0,1}$. Similarly, there must exist $P_{n,2}(u,v)$ so that its roots $v_{n,2}(u)$ stay close (in the above sense) to those of $P_{0,2}(u,v)$ for u in C_{12}, and $N_{n,2}\geqslant N_{0,2}$. But $N_{n,1}+N_{n,2}=N=N_{0,1}+N_{0,2}$, so that the only possibility is $N_{n,1}=N_{0,1}$ and $N_{n,2}=N_{0,2}$, as claimed in (i).

To prove (ii), we show first that the coefficients $a_{n,k,1}(u)$ of the powers of v in $P_{n,1}(u,v)$ tend to the corresponding coefficients $a_{0,k,1}(u)$ in $P_{0,1}(u,v)$, uniformly in the disk $|u-u_0|\leqslant r_2$. To see this, we express $a_{n,k,1}(u)$, $a_{0,k,1}(u)$ as symmetric combinations of the roots of $P_{n,1}(u,v)$ and $P_{0,1}(u,v)$ and conclude from the proof of (i) that $a_{n,k,1}(u)\to a_{0,k,1}(u)$, uniformly in $r_1\leqslant|u-u_0|\leqslant r_2$. However, both $a_{n,k}(u)$ and $a_{0,k}(u)$ are holomorphic functions in $|u-u_0|\leqslant r_2$ and from Cauchy's theorem, we conclude that, in fact, $a_{n,k,1}(u)\to a_{0,k,1}(u)$, uniformly in $|u-u_0|\leqslant r_2$, as announced.

Now, we can show that, if n is sufficiently large, there exists (u_n,v_n), $|u_n-u_0|<r_2$, $v_n\in V$, so that $P_{n,1}(u_n,v_n)=0$, $P_{n,2}(u_n,v_n)=0$. To verify this, we construct the resultant $R_n(u)$ (eliminant[21]) of $P_{n,1}(u,v)$, $P_{n,2}(u,v)$; it consists of sums and products of coefficients of $P_{n,1}(u,v)$ and $P_{n,2}(u,v)$ and is thus a holomorphic function of u in $|u-u_0|\leqslant r_2$. Clearly, $R_n(u)$ tends to $R_0(u)$ [obtained from $P_{0,1}(u,v)$ and $P_{0,2}(u,v)$], uniformly in $|u-u_0|\leqslant r_2$, and consequently, for n large enough, it vanishes the same number of times as $R_0(u)$ in this disk. By assumption, however, $R_0(u)$ has at least a simple zero in $|u-u_0|<r_2$, and consequently, so has $R_n(u)$, for n large enough. Let (u_n,v_n) be one of these roots. At such a point, Eq. (B4) is clearly fulfilled. This ends the proof of (ii).

To show (iii), we notice that, if Eq. (B4) is satisfied for some values of u_n, v_n, λ_n, then these are also roots of the three equations

$$P_0(u,v)+\lambda G(u,v)=0,$$
$$\frac{\partial P_0}{\partial u}+\lambda\frac{\partial G}{\partial u}=0\,,\quad \frac{\partial P_0}{\partial v}+\lambda\frac{\partial G}{\partial v}=0\,. \tag{B5}$$

At $u=u_0$, $v=v_0$, $\lambda=0$, the Jacobian of these three functions with respect to u,v,λ is

$$J(u_0,v_0,0)=G(u_0,v_0)\left(\frac{\partial^2 P_0}{\partial u^2}\frac{\partial^2 P_0}{\partial v^2}-\left(\frac{\partial^2 P_0}{\partial u\,\partial v}\right)^2\right). \tag{B6}$$

If the bracket does not vanish, then any $G(u,v)$ [and thus $\Omega_1(u,v)$] with $G(u_0,v_0)\neq 0$ leads to $J(u_0,v_0,0)\neq 0$, and thus to the conclusion that the root $(u_0,v_0,0)$ of (B5) is isolated. This proves (iii) for this particular case. If the bracket vanishes, we have to show that, nevertheless, G may be chosen so that Eqs. (B5) admit of a unique solution [in a neighborhood of $(u_0,v_0,0)$]. If Eqs. (B5) admit of a sequence of points $(u_n,v_n,\lambda_n)\to(u_0,v_0,0)$ among their solutions, one may eliminate λ_n and conclude that (u_n,v_n) satisfies the equations

$$\left(P_0\frac{\partial G}{\partial u}-G\frac{\partial P_0}{\partial u}\right)(u_n,v_n)=0\,,$$
$$\left(P_0\frac{\partial G}{\partial v}-G\frac{\partial P_0}{\partial v}\right)(u_n,v_n)=0\,. \tag{B7}$$

It is easy to verify that (B7) for an infinite sequence of n implies that there exists an irreducible factor $P(u,v)$, which divides both functions in (B7) in a neighborhood of (u_0,v_0). We may try to determine the leading coefficient a_1 in the expansion $v=v_0+a_1(u-u_0)^\mu+\cdots$, with μ some rational number. This leads to an algebraic equation for a_1, with coefficients that are linear in (a certain number) of the few first Taylor coefficients in the expansion of $P_0(u,v)$,$G(u,v)$ around u_0,v_0. They admit of a solution only if their resultant vanishes. This leads to a quadratic equation for the first few Taylor coefficients of G, whose precise form depends on $P_0(u,v)$. This ends the proof of (iii), since we only have to choose G so that this equation is not fulfilled. Thus the assertion of the beginning of this Appendix is justified.

APPENDIX C: AN IRREDUCIBLE INTENSITY DISTRIBUTION

Since it is easy to find polynomials in two variables which are positive over the Re u-Re v plane and are not reducible (e.g., u^2+v^2+1), it is not surprising that there exist truly entire intensity distributions (i.e., entire functions of exponential type, positive and integrable over the Re u-Re v plane), which do not correspond to any amplitude. It is, however, somewhat difficult to prove that a given intensity distribution is irreducible, or even to show how such distributions may arise. In this Appendix, we show how to generate such distributions, in some simple special cases.

To this end, let, for α real

$$F(u,\alpha)=\sin(u-i\alpha)\sin(u+i\alpha)/(u^2+\alpha^2)\,, \tag{C1}$$

and consider

$$I(u,v,\alpha_i,\gamma)=F(u,\alpha_1)F(v,\alpha_3)+\gamma F(u,\alpha_2)F(v,\alpha_4)\,. \tag{C2}$$

We shall show how we may choose α_i, $i=1,\ldots,4$, γ so that $I(u,v,\alpha,\gamma)$, which is clearly positive in the Re u-Re v plane, for positive γ, integrable and of exponential type, is irreducible.

To show that $I(u,v)$ is irreducible, we have to prove that, if $v=v_1(u)$ is an analytic function, defined in the neighborhood of some point u_1 in the u plane, such that $F(u,v_1(u))\equiv 0$, and $v_2(u)$ a similar function element, defined near another point u_2 (it is not necessary that $u_2\neq u_1$) with $F(u,v_2(u))=0$, then there exists a path $\mathscr{P}$ in the u plane so that $v_2(u)$ is the analytic continuation of $v_1(u)$ along $\mathscr{P}$. Further, $I(u,v)$ should not vanish along any line $v=v_0$ or $u=u_0$.

Let us choose first $\alpha_1\neq\alpha_2$, and $\alpha_3\neq\alpha_4$. Then, clearly $I(u,v)$ does not vanish along any line $u=u_0$ or $v=v_0$. Further, we may obtain a parametric representation of the set of zeros of $I(u,v)$ as follows: let $u=\tilde{u}(\lambda)$ be a solution of the equation

$$F(u,\alpha_1)/F(u,\alpha_2)=\lambda \tag{C3}$$

and $v=\tilde{v}(\mu)$ be a solution of the equation

$$F(v,\alpha_3)/F(v,\alpha_4)=\mu\,. \tag{C4}$$

It is clear that the set of points $u = \tilde{u}(\lambda)$, $v = \tilde{v}(-\lambda/\gamma)$ describes part of the set of zeros of $I(u,v)$ as λ moves over the complex plane. To show that we obtain this way the whole set of zeros, we have to prove that Eqs. (C3), (C4) define irreducible functions of u and λ (μ). But this is obvious, since (i) $\lambda = \lambda(u)$ defined in (C3) is a meromorphic, one-valued function of u, and (ii) by assumption $F(u,\alpha_1)$, $F(u,\alpha_2)$ do not vanish at the same points. The same is true for $\mu = \mu(v)$ of (C4). Certainly, the functions $\tilde{u}(\lambda)$, $\tilde{v}(-\lambda/\gamma)$ have an infinity of branch points in the complex λ plane, since, e.g., at fixed λ, the function $F(u,\alpha_1) - \lambda F(u,\alpha_2)$ vanishes infinitely many times in the u plane.

Now, we claim that, if the set $\{\lambda_i\}_{i=1}^{\infty}$ of branch points of $u = \tilde{u}(\lambda)$ is completely disjoint from the set $\{\lambda'_i\}$ pertaining to $\tilde{v}(-\lambda/\gamma)$, then the manifold defined by $u = \tilde{u}(\lambda)$, $v = \tilde{v}(-\lambda/\gamma)$ is irreducible. To see this, first let $v = v_1(u)$, $v = v_2(u)$ be two roots of $I(u,v) = 0$ defined in a neighborhood of u_1 and $\lambda_1 = \lambda(u_1)$ be defined by (C3). Since all the zeros of $F(v,\alpha_3) + \lambda/\gamma F(v,\alpha_4) = 0$ are obtained by analytic continuation in the λ plane, let $\mathscr{P}_\lambda$ be the closed path of continuation needed to obtain $v_2(-\lambda/\gamma)$ from $v_1(-\lambda/\gamma)$, both defined near λ_1. The essential point is that, if the set of branch points $\{\lambda_i\}$ and $\{\lambda'_i\}$ are distinct, we can choose $\mathscr{P}_\lambda$ so that its image $\mathscr{P}$ under $\tilde{u}(\lambda)$ is also a closed path in the u plane; the path $\mathscr{P}_\lambda$ must avoid the set $\{\lambda_i\}$ and not turn around any of its points, although it may turn around the λ'_i, as required. If $\mathscr{P}_\lambda$ is chosen as shown, then $\mathscr{P}$ is the path of continuation in the u plane leading from $v_1(u)$ to $v_2(u)$.

The generalization to arbitrary u_1 and u_2 and roots $v_1(u)$, $v_2(u)$ defined near u_1 and u_2 is obvious; let $\lambda_1 \equiv \lambda(u_1)$, $\lambda_2 \equiv \lambda(u_2)$ and $\mathscr{P}'_\lambda$ be a path joining λ_1 to λ_2 and on which $u = \tilde{u}(\lambda)$ is holomorphic, one-valued, and with a one-valued inverse. The result of the continuation of $\tilde{v}_1(-\lambda/\gamma)$ along $\mathscr{P}'_\lambda$ up to λ_2 is not, in general, $v_2(\tilde{u}(\lambda_2))$, but another root $\tilde{v}'_2(-\lambda_2/\gamma)$. Supplement then $\mathscr{P}'_\lambda$ by as many cycles passing through λ_2 as are needed in order to obtain $v_2(\tilde{u}(\lambda_2))$. The cycles may be chosen so that their images under $u = \tilde{u}(\lambda)$ are again cycles $\mathscr{P}''$. The path of continuation in the u plane needed to obtain $v_2(u)$ from $v_1(u)$ is made up of $\mathscr{P}'$, the image of $\mathscr{P}'_\lambda$ and the cycles $\mathscr{P}''$.

With this, the next task is to show that we can choose α_i, $i = 1,\ldots,4$ and γ so that, indeed, the sets $\{\lambda_i\}$, $\{\lambda'_i\}$, defined by (C3) and (C4) are disjoint.

Clearly, the set of branch points $\{\lambda_i\}$ is among the solutions of the set of equations

$$F(u,\alpha_1) - \lambda F(u,\alpha_2) = 0\,,$$
$$\frac{\partial F}{\partial u}(u,\alpha_1) - \lambda \frac{\partial F}{\partial u}(u,\alpha_2) = 0\,. \tag{C5}$$

Let $u_n(\alpha_1,\alpha_2)$ be the roots of the equation

$$F(u,\alpha_1)\frac{\partial F}{\partial u}(u,\alpha_2) - F(u,\alpha_2)\frac{\partial F}{\partial u}(u,\alpha_1) = 0\,. \tag{C6}$$

Then, the set $\{\lambda_i\}$ is contained in the set given by

$$\lambda_n = \frac{F(u_n(\alpha_1,\alpha_2),\alpha_1)}{F(u_n(\alpha_1,\alpha_2),\alpha_2)}\,. \tag{C7}$$

This set must be compared with the one given by

$$\lambda'_n = -\gamma \frac{F(v_n(\alpha_3,\alpha_4),\alpha_3)}{F(v_n(\alpha_3,\alpha_4),\alpha_4)}\,, \tag{C8}$$

where $v_n(\alpha_3,\alpha_4)$ are the roots of an equation in v, analogous to (C6)

$$F(v,\alpha_3)\frac{\partial F}{\partial v}(v,\alpha_4) - F(v,\alpha_4)\frac{\partial F}{\partial v}(v,\alpha_3) = 0\,. \tag{C9}$$

Let us notice that, if $\alpha_i \neq 0$, for all i, the values $\lambda = 0$ and $\lambda = \infty$ are not contained in the sets (C7) or (C8). Indeed, $\lambda_n = 0$ for some n implies $F(u_n(\alpha_1,\alpha_2),\alpha_1) = 0$. But, if $\alpha_1 \neq 0$, all roots of $F(u,\alpha_1)$ are simple, so that (C6) implies $F(u,\alpha_2) = 0$ at $u = u_n(\alpha_1,\alpha_2)$. This is, however, forbidden by the fact that $\alpha_1 \neq \alpha_2$. The same argument shows that $\lambda = \infty$ is not in (C7) or (C8).

We shall now investigate the asymptotic distribution of the zeros of (C6) and (C9) and show that, apart possibly from a finite number of points, the two sets (C7) and (C8) are indeed distinct. A simple calculation shows that the zeros of (C6) are the roots of

$$\begin{aligned} &u(\alpha_1^2 - \alpha_2^2)\sin^2 2u + (u^2 + \alpha_1^2)(u^2 + \alpha_2^2)(\cosh 2\alpha_1 \\ &\quad - \cosh 2\alpha_2)\sin 2u + (\alpha_1^2 - \alpha_2^2)u(\cosh 2\alpha_1 \\ &\quad + \cosh 2\alpha_2)\cos 2u - (\alpha_1^2 - \alpha_2^2) \\ &\quad \times u(1 + \cosh 2\alpha_1 \cosh 2\alpha_2) = 0\,. \end{aligned} \tag{C10}$$

One can verify that the roots of (C10) fall asymptotically into two classes

(i) $\quad u_n = n\pi/2 + O(n^{-3})\,, \qquad$ (C11)

(ii) $\quad u_n = n\pi \mp \pi/4 \pm \frac{3}{2} i \ln k|n| + O(\ln|n|/n)\,. \qquad$ (C12)

In (C12) the upper/lower sign in front of $\pi/4$ holds if n is positive/negative and

$$k = \pi[2(\cosh 2\alpha_1 - \cosh 2\alpha_2)/(\alpha_1^2 - \alpha_2^2)]^{1/3}\,.$$

A way of deriving (C11) and (C12) is sketched at the end of the Appendix.

With this, it is easy to check that the subset of λ_n's, Eq. (C7), generated by (C11) is made up of real points, accumulating at $\lambda_\pm$ ($\lambda_\pm > 0$)

$$\lambda_\pm = (\cosh 2\alpha_1 \pm 1)/(\cosh 2\alpha_2 \pm 1)\,. \tag{C13}$$

The other subset of λ_n's generated by (C12) is made up of pairs of complex conjugate points, accumulating at $\lambda_0 = +1$. It is now obvious that, if $\gamma > 0$, the set (C8) accumulates at $-\gamma\lambda_\pm \neq \lambda_\pm$ and $-\gamma \neq +1$, so that it may have at most a finite number of points in common with (C7), as announced. Now, since $\lambda = 0$ is not contained in either (C7) or (C8), any choice of $\gamma > 0$, except possibly for a finite number of values, makes the sets $\{\lambda_n\}$ and $\{-\gamma\lambda'_n\}$ disjoint and leads thus to irreducible intensities.

To verify, e.g., (C11), one proceeds as follows: one substitutes $u_n = n\pi/2 + \beta$ in (C10), expresses $\sin 2\beta$ appearing in the second term (which is dominant for high n) as a function of β

$$\beta = \tfrac{1}{2}\arcsin\phi(\beta) \equiv \Gamma(\beta)\,, \tag{C14}$$

proves then explicitly that, for $|\beta| \leqslant \pi/4$ and n high enough, $|\Gamma'(\beta)| \leqslant 1$, so that (C14) can be solved by iteration, and finally evaluates the error committed by taking $\beta = 0$ on the right-hand side of (C14)

$$|\beta_{ex}| < |\Gamma(0)|/(1 - \sup_{|\beta| < \pi/4} |\Gamma'(\beta)|) . \tag{C15}$$

Equation (C12) is obtained in a similar manner. It is also true that Eqs. (C11) and (C12) represent asymptotically all the roots of Eq. (C10).

[1]Yu. M. Bruck and L. G. Sodin, Opt. Commun. **30**, 304 (1979).

[2]I. Manolitsakis, J. Math. Phys. **23**, 2291 (1982).

[3]B. J. Hoenders, J. Math. Phys. **16**, 1719 (1975).

[4]H. A. Ferwerda, in *Inverse Source Problems in Optics*, Vol. 9 of *Topics of Current Physics* (Springer, Berlin, 1978).

[5]M. H. Hayes and J. H. McClellan, Proc. IEEE, **70**, 197 (1982).

[6]G. Höhler and I. S. Stefanescu, Z. Phys. C **2** 253 (1979).

[7]C. B. Lang and I. S. Stefanescu, Phys. Lett. B **79**, 479 (1978).

[8]M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York 1964), p. 378.

[9]M. Plancherel and G. Polya, Commun. Math. Helv. **9**, 224 (1937); **10**, 110 (1938).

[10]W. F. Osgood, *Funktionentheorie* (Chelsea, New York, 1965), Vol. II.

[11]R. P. Boas, *Entire Functions* (Academic, New York, 1954).

[12]H. Behnke and P. Thullen, *Theorie der Funktionen Mehrerer Komplexer Veränderlicher* (Springer, Berlin, 1970).

[13]P. Lelong, J. Anal. **12**, 365 (1964).

[14]G. Atkin and G. Ross, "Object restoration with superresolution," preprint, Queen Elisabeth College, 1983.

[15]E. C. Titchmarsh, Proc. London Math. Soc. **25**, 283 (1926).

[16]G. M. Fichtenholtz, *Differential und Integral Rechnung* (VEB, Berlin, 1963), Chap. XI, Sec. 5.

[17]I. Caprini, S. Ciulli, A. Pomponiu, and I. S. Stefanescu, Phys. Rev. D **5**, 1658 (1972); I. Caprini, Nuovo Cimento Lett. **5**, 443 (1972); Nucl. Phys. B **68**, 561 (1974); I. S. Stefanescu, Nucl. Phys. B **56**, 287 (1973).

[18]S. Ciulli, C. Pomponiu, and I. S. Stefanescu, Phys. Rep. C **17**, 135; (1975) I. S. Stefanescu, J. Math. Phys. **21**, 175 (1980).

[19]E. Barrelet, Nuovo Cimento A **8**, 331 (1972).

[20]I. S. Stefanescu, J. Math. Phys. **25**, 2052 (1984).

[21]H. W. Turnbull, *Theory of Equations* (Oliver and Boyd, Edinburgh, 1957), p. 73; J. V. Uspensky, *Theory of Equations* (McGraw-Hill, New York, 1948), p. 257.

[22]M. Cheney, J. Math. Phys. **25**, 95 (1984).

[23]J. R. Fienup, Opt. Lett. **13**, 27 (1978).

[24]M. A. Fiddy, B. J. Brames, and J. C. Dainty, Opt. Lett. **8**, 96 (1983).

[25]H. A. Arsenault and K. Chalasinka Marcukow, Opt. Commun. **47**, 380 (1983).

[26]P. Lelong, Ann. Ec. Normale Super. **67**, 393 (1950).

[27]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970), Chap. 22, p. 794.

Some nodal theorems for noncentral forces

J. S. Levinger
Department of Physics, Rensselaer Polytechnic Institute, Troy, New York 12181

(Received 13 February 1985; accepted for publication 25 April 1985)

Two lemmas are proved for local noncentral forces in multidimensional space. First, the lowest partial wave for the ground state is nodeless. Second, the lowest partial wave for the first excited state has at least one node. Ballot–Fabre de la Ripelle perturbation theory is also used to show that higher partial waves for the ground state have nodes near the positions of nodes (if any) in the corresponding element of the matrix element of the noncentral potential.

I. INTRODUCTION

There is a well-known nodal theorem[1–3] for a one-dimensional Sturm–Liouville problem, such as the solution of the Schrödinger equation for a bound state for a one-dimensional local potential $V(x)$. The theorem states that the ordering of solutions by numbers of nodes is the same as the ordering by energy eigenvalue. In particular, the ground state has no nodes (for finite x) and is therefore nondegenerate. Of course a central potential in spherical coordinates gives equivalent one-dimensional problems, and the nodal theorem applies separately to each of the three functions $R(r)$, $\Theta(\theta)$, and $\Phi(\phi)$, which, when multiplied together, give the wave function $\psi(\mathbf{r})$.

The generalization of the nodal theorem to multidimensional space[1] was given many years ago. The proof is repeated later below in this note, since a number of physicists are unaware of the Courant–Hilbert theorem. The theorem states that for a local potential $V(\mathbf{r})$ the ground state has no nodal surfaces, and the first excited state has one nodal surface. The number of nodal surfaces is not determined for the second or higher excited states. Here $\mathbf{r}$ is a vector in multidimensional space: we are concerned with examples of three-dimensional, six-dimensional, and nine-dimensional spaces.

We prove below two lemmas concerning the lowest partial wave for the ground state and first excited state eigenfunctions, respectively. (The partial waves are ordered by a quantum number, which in three dimensions is the orbital angular momentum l. In multidimensional space, we order using the "grand orbital" L. The "lowest partial wave" has $l=0$ or $L=0$.) (i) The lowest partial wave for the ground state eigenfunction is nodeless, and (ii) the lowest partial wave for the first excited state has at least one node, provided that the nodal surface is closed and does not go through the origin.

We also use the perturbation theory formalism of Ballot and Fabre de la Ripelle[4] (BF) to examine nodes of higher partial waves for the ground state ($l>0$ or $L>0$.) We find that for a nearly diagonal potential matrix—the case for which the BF approximation is valid—a given partial wave has nodes if and only if the appropriate off-diagonal matrix element has a node.

The two lemmas and the perturbation theory approximation are used to check solutions of a number of examples: the anisotropic harmonic oscillator problem solved in spherical coordinates; the trinucleon solved in six-dimensional space using hyperspherical harmonics (H.H.); the hypertriton solved by the same techniques; the alpha particle solved in nine-dimensional space for both the ground and first excited states; and the helium tetramer $^4He_4$ solved in nine-dimensional space with a model atom–atom potential.

Finally we review the importance of the assumption that we are dealing with a *local* potential, and give examples where this assumption is not valid. We also consider two related problems for which we do not have solutions: Is the lowest partial wave nodeless for the ground state solution of two coupled differential equations (CDE's), and is the S wave $u(r)$ nodeless for the deuteron with a local tensor force?

II. NODAL THEOREMS, LEMMAS, AND APPROXIMATIONS

The Courant–Hilbert theorem[1] for two solutions of a Sturm–Liouville problem is proved in multidimensional space with the use of Green's formula. They choose $\psi_0(\mathbf{r})$ as the twice continuously differentiable normalizable solution of

$$-(\hbar^2/2m)\nabla^2\psi_0(\mathbf{r}) + V(\mathbf{r})\psi_0(\mathbf{r}) = E_0\psi_0(\mathbf{r}). \tag{1}$$

Here, $\psi_0(\mathbf{r})$ vanishes on the boundary Γ of some region B, and is positive in B. The eigenfunction for a higher energy E_1 obeys the equation

$$-(\hbar^2/2m)\nabla^2\psi_1(\mathbf{r}) + V(\mathbf{r})\psi_1(\mathbf{r}) = E_1\psi_1(\mathbf{r}). \tag{2}$$

Multiply Eq. (1) by $\psi_1(\mathbf{r})$, multiply Eq. (2) by $\psi_0(\mathbf{r})$, subtract the first equation from the second, and integrate over B. Then

$$\frac{\hbar^2}{2m}\int_B(\psi_1\nabla^2\psi_0 - \psi_0\nabla^2\psi_1)d^{\,n}r = (E_1-E_0)\int_B\psi_0(\mathbf{r})\psi_1(\mathbf{r})d^{\,n}r. \tag{3}$$

Here $d^{\,n}r$ is the volume element and ∇^2 is the Laplacian in n-dimensional space. The left-hand side of (3) is evaluated using Green's formula (valid in multidimensional space), giving us

$$\frac{\hbar^2}{m}\int_\Gamma\psi_1\frac{\partial\psi_0}{\partial n}d^{\,n-1}s = (E_1-E_0)\int_B\psi_0(r)\psi_1(r)d^{\,n}r, \tag{4}$$

where $d^{\,n-1}s$ is the differential element on a surface or hypersurface Γ, and $\partial/\partial n$ is the normal derivative, the outward normal being positive.

If we assume that $\psi_1(\mathbf{r})$ is positive throughout B, we find that Eq. (4) cannot hold. The left side is negative, since the normal derivative $\partial\psi_0/\partial n$ is negative on Γ. But the right side is positive, since we assumed $E_0<E_1<0$. Therefore, $\psi_1(r)$

 0022-2488/85/092161-04$02.50

must have at least one nodal surface.

Courant and Hilbert conclude that the eigenfunction with the lowest eigenvalue will have no nodal surfaces in B; and the first excited state will have one nodal surface. Note that the proof depends on the assumption of the locality of $V(\mathbf{r})$, since we used the cancellation of $\psi_0(\mathbf{r})\ V(\mathbf{r})\psi_1(\mathbf{r})$ and $-\ \psi_1(\mathbf{r})\ V(\mathbf{r})\psi_0(\mathbf{r})$ in writing Eq. (3). (A "nonlocal" potential is illustrated below: it uses Majorana exchange.)

Our first lemma is easily proved by expanding $\psi_0(\mathbf{r})$ in a complete orthonormal set of angular functions (spherical or hyperspherical harmonics) the expansion coefficients being the partial waves. In three dimensions,

$$r\psi_0(\mathbf{r}) = \sum_{l,m} u_l(r) Y_l^m(\theta,\phi), \tag{5}$$

using the standard spherical harmonics. In six dimensions[4] we write an analogous equation using hyperspherical harmonics (H.H.), which are functions of five angular variables, denoted by Ω. We use r as the hyperradius. The subscript $[L]$ stands for five quantum numbers. We have

$$r^{5/2}\psi_0(\mathbf{r}) = \sum_{[L]} u_L(r) Y_{[L]}(\Omega). \tag{6}$$

The partial waves $u_L(r)$ depend only on a single quantum number L, called the grand orbital. (Simonov[5] uses K as this quantum number and refers to "K-harmonics.")

We solve for the lowest partial wave $u_0(r)$ by inverting the series (5) or (6). For instance, in three-dimensional space,

$$u_0(r) = \int r\psi_0(\mathbf{r}) Y_0^0(\theta,\phi)\, d^2\Omega. \tag{7}$$

(We integrate over $d^2\Omega = \sin\theta\, d\theta\, d\phi$.) But the lowest spherical harmonic Y_0^0 is a constant, which, by convention, is positive. Therefore the integrand in (7) is positive for all r and Ω, and so the integral is also positive. We repeat this argument in six dimensions (or nine dimensions, etc.). The inversion of (6) uses the orthonormality of the H.H. The lowest H.H. $Y_{[0]}(\Omega)$ is again nodeless and, again by convention, is positive. Then

$$u_0(r) = r^{5/2} \int \psi_0(\mathbf{r}) Y_{[0]}(\Omega)\, d^5\Omega > 0. \tag{8}$$

Our second lemma concerns the partial wave $u_0^*(r)$ for the first excited state, with eigenfunction $\psi_1(\mathbf{r})$. Courant and Hilbert show that this eigenfunction has one nodal surface, or hypersurface. *If* this surface does not go through the origin, and does not go to infinity, then we show below that the lowest partial wave [which we denote by $u_0^*(r)$] must have at least one node. We must first investigate the behavior of the nodal surface.

First, we assume that the excited state in question has a lowest partial wave $u_0^*(r)$ that is not identically zero. (That is, we exclude problems such as an odd parity excited state of an anisotropic harmonic oscillator, which cannot have any S state, because of parity arguments.) Second, we assume that this lowest partial wave does not "happen" to "be extra small" near the origin. Let us illustrate in three dimensions, where $u_0^*(r)$ is expected to be small, of order r, near the origin; and "extra small" means still smaller. In six dimensions, $u_0^*(r)$ is expected to be of order $r^{5/2}$; in nine dimensions of order r^4, etc. Since higher partial waves are excluded from the origin by the centrifugal barrier, the behavior of $\Psi_1(\mathbf{r})$ near the origin is determined by the behavior of $u_0^*(r)$ in this region. The validity of this second assumption can be investigated numerically for any specified noncentral potential.

The nodal surface cannot go to infinity for a *bound* excited state, since the wave function must change sign at the nodal surface; but it must go to zero at infinity for a bound state.

Now that we know (subject to conditions given above) that $\psi_1(\mathbf{r})$ has a nodal surface that does not go through the origin, or go to infinity, we can prove our lemma on nodes in the lowest partial wave $u_0^*(r)$. We first write an equation analogous to (7), for a three-dimensional problem:

$$u_0^*(r) = \int r\psi_1(\mathbf{r}) Y_0^0(\theta,\phi)\, d^2\Omega. \tag{9}$$

Since $\psi_1(\mathbf{r})$ has a closed nodal surface that does not go through the origin it has a minimum radius r denoted by r_1, and a maximum denoted by r_2. Assume, for instance, that $\psi_1(\mathbf{r})$ is positive for $0 < r < r_1$, and $\psi_1(\mathbf{r})$ is negative for $r_2 < r < \infty$. If we evaluate (9) for $0 < r < r_1$ we clearly obtain a positive value for the lowest partial wave $u_0^*(r)$. On the other hand, (9) gives us a negative $u_0^*(r)$ for $r_2 < r < \infty$. Thus we have proved that $u_0^*(r)$ changes sign in the region $r_1 < r < r_2$, and therefore has one node (or possibly more than one). This proof clearly generalizes to six-dimensional or any multidimensional space.

Finally, we use the Ballot–Fabre de la Ripelle perturbation theory to examine the nodal character of the partial wave $u_L(r)$, for the ground state triton eigenfunction, for positive grand orbital L. Ballot and Fabre de la Ripelle treat off-diagonal terms in the potential matrix $[V]$ as a perturbation, and find that to lowest order in this perturbation the partial wave $u_L(r)$ is given by

$$u_L(r) \approx [u_0(r) V_{0,L}(r)/L(L+4)](r^2 m/\hbar^2). \tag{10}$$

The factor $\hbar^2 L(L+4)/(mr^2)$ is the difference of centrifugal barriers in six dimensions for a state with grand orbital L and a state with grand orbital zero. [In three dimensions one should use $l(l+1)$, and in nine dimensions one should use $L(L+7)$; but these numerical differences do not change our conclusions below.]

The small matrix element $V_{0L}(r)$ is

$$V_{0L}(r) = \int Y_{[0]}(\Omega) V(\mathbf{r}) Y_L(\Omega)\, d^5\Omega. \tag{11}$$

This expression holds in six-dimensional space, with $Y_{[0]}(\Omega)$ the lowest H.H., which we used in (8), and $Y_L(\Omega)$ a suitable combination of H.H. for grand orbital L. Here $V(\mathbf{r})$ is the total potential energy of the trinucleon, a function of the six-dimensional $\mathbf{r}$.

We see that all terms on the right side of Eq. (10) are positive, with the possible exception of the matrix element $V_{0L}(r)$, defined by (11). Hence if the approximation (10) is sufficiently accurate to use for the determination of nodes, the number of nodes in the partial wave $u_L(r)$ will be just the same as the number of nodes in the corresponding matrix element $V_{0L}(r)$, and they will occur at the same values of hyperradius r.

We use this approximation to show the existence of nodes in higher partial waves for the ground state eigenfunction in certain specified cases. Assume that all off-diagonal elements $V_{0L}(r)$ are sufficiently small that the BF approximation is valid. Then if $V_{0L}(r)$ has a node for a specified L, in this case we can use (10) with confidence, and conclude that $u_L(r)$ also has a node for this value of the grand orbital. We find below that quantitative use of (10) to find nodes in $u_L(r)$ works moderately well for finite but small values of $V_{0L}(r)$ such as those for the ground state of the triton or alpha particle using the Volkov potential.

For convenience I summarize our two lemmas and one approximation.

First lemma: The lowest partial wave for the ground state for a local noncentral potential has no modes.

Second lemma: The lowest partial wave for an excited state for a local noncentral potential has at least one node.

Approximation: For small off-diagonal terms in the potential matrix, the nodes in the partial wave $u_L(r)$ for the ground state wave function occur at the same hyperradius as the nodes (if any) in $V_{0L}(r)$.

III. ILLUSTRATIONS AND DISCUSSION

We present briefly five illustrations of the first lemma.

(i) The ground state wavefunction of an anisotropic axially symmetric oscillator is clearly nodeless in Cartesian coordinates, and so it is also nodeless in spherical coordinates. Elminyawi[6] and Elminyawi and Levinger[7] calculated the lowest partial wave solution, and found both analytically and in numerical work that the s wave $u_0(r)$ is also nodeless.

(ii) Ballot *et al.*[8] found partial waves numerically for the ground state of the triton, for several different choices of nucleon–nucleon potential. In each case they found that $u_0(r)$ was nodeless.

(iii) Clare[9] and Clare and Levinger[10] made an analogous calculation for the partial waves for the ground state of the hypertriton, and found that $u_0(r)$ was again nodeless.

(iv) Ballot *et al.*,[8] Elminyawi,[6] and Elminyawa and Levinger[7] found numerical solutions for the lowest partial wave for the alpha particle: again $u_0(r)$ was nodeless.

(v) Elminyawi[6] and Elminyawi and Levinger[7] found numerical values for the partial waves for the ground state of the helium tetramer He_4 (composed of atoms with ^4He nuclei) for a model potential, and found that for this case $u_0(r)$ did have a node, but that the ratio of minimum u_0 to maximum u_0 was only about -1%.

The reader may well wonder, "Why bother to give examples of a mathematical lemma?" First, they are illustrations of its applicability to a variety of problems. Second, the lemma is useful as a check on the accuracy of numerical (or for one case analytical) work in determination of the lowest partial wave. Third, the lemma applies to $u_0(r)$ determined from Eqs. (7) or (8): that is, to $u_0(r)$ as determined from the solution of an *infinite* set of CDE's. But all numerical work is based on solution of a *finite* set of M CDE's; e.g., for the tetramer $M = 8$.

The agreement between numerical solutions and our lemma for the first four illustrations shows we can answer "yes" to the following two questions: (i) is the numerical work of high enough accuracy and (ii) does the exact solution of M CDE's show the same nodeless behavior of $u_0(r)$ as the solution of an infinite set of CDE's? The fifth case—the tetramer—shows that one (or both) of these questions must be answered with "no, not exactly." Note that the potential matrix for the model potential used has very large off-diagonal matrix elements, thus posing a severe strain on the numerical procedures used.

We present one illustration of the second lemma: the lowest partial wave for the first excited state of the helium nucleus for a Volkov nucleon–nucleon potential. Ballot *et al.*[8] and also Elminyawi[6] found that the lowest partial wave $u_0^*(r)$ is not smaller than $r^{5/2}$near the origin and has a node at a hyperradius of 4.5 fm. (Unlike the tetramer case, the node is *not* due to numerical errors or possible failure of convergence of the H.H. expansion. The ratio [minimum $u_0^*(r)$]/[maximum $u_0^*(r)$] is -60%.) One can understand this node in a naive way using a wave function as a single H.H. The ground state is the nodeless solution $u_0(r)$ of an ordinary differential equation; and the excited state is the solution $u_0^*(r)$ of the same equation, at the higher energy. The node persists as the solution $u_0^*(r)$ for M CDE's.

We present two illustrations of the BF approximation method of finding nodes in higher partial waves of the ground state wave function. In each example, we use the Volkov potential. The first case is for the ground state of the triton. The positions of the nodes in $u_4(r)$, $u_6(r)$, and $u_8(r)$ are taken from Ballot *et al.*[8] and compared in Table I with the positions of the nodes of the Volkov potential. We find perfect agreement in the number of nodes; for this case one and only one node for each higher partial wave. We see quite good agreement in the nodal positions for each partial wave, showing that the Ballot–Fabre de la Ripelle approximation scheme is successful for this problem. The second case is for the ground state of the alpha particle again with the Volkov potential. Again we have perfect agreement in the number of nodes: namely one for each higher partial wave. The table shows that for this case the positions of nodes in V_{0L} (see Ref. 6) are in only fair agreement with the positions[8] of nodes in the partial waves $u_4(r)$, $u_6(r)$, and $u_8(r)$.

Of course the Courant–Hilbert theorem for the nodeless character of the ground state wave function need not apply for a nonlocal potential $V(\mathbf{r},\mathbf{r}')$. We give several illustrations of nonlocal potentials.

The easiest example to calculate is that of a strong Majorana space exchange force between two nucleons, chosen with the opposite sign to the one present in the real world.

TABLE I. Positions of nodes in higher partial waves and in the corresponding element of the potential matrix.

	Triton nodes		Alpha nodes	
Grand orbital	Wave function	Potential matrix	Wave function	Potential matrix
4	3.5 fm	3.5 fm	3.9 fm	2.8 fm
6	4.8	4.6	5.6	4.1
8	5.7	5.8	6.4	5.3

That is, assume the exchange force is *repulsive* in the S state and *attractive* in the P state. A sufficiently strong Majorana force of this sign could clearly overcome the effect of the P state centrifugal potential, and give us a ground state which is a P state, with a nodal surface! This violation of the nodal theorem occurs because a Majorana force is a (complicated) example[11,12] of a nonlocal potential. Note that it takes a *strong Majorana* force (of appropriate sign) to invert the normal order of S and P states.

Another example is Tabakin's[13] separable (nonlocal) nucleon–nucleon potential. He chose a potential so that the $2S$ state was bound, and the $1S$ state was in the continuum: another reversal of the order given by the nodal theorem. Like the first example, such a strong nonlocality seems unlikely to exist in the real world.

A third example is a shell model for a system of fermions. Here the Pauli principle forces us to put particles into higher shells, which have nodes in their wave functions, so the wave function of the system has nodal surfaces[2] even for the ground state. But (at least in the atomic case in the nonrelativistic approximation) the potential is local! If we neglected the Pauli principle we would find lower energy eigenvalues; and the lowest would indeed have a nodeless eigenfunction. But the Pauli principle forbids our use of any wave function that is not completely antisymmetric for space and spin exchange of two electronic coordinates.

Finally, the reader may ask, "What is so special about coordinate space that $\psi(\mathbf{r})$ is nodeless but in general the corresponding momentum space wave function $\phi(\mathbf{p})$ does have nodes?" At first glance this asymmetry between coordinate and momentum space is strange, since we could solve the Schrödinger equation for the ground state in either space. The asymmetry enters because we generally use a local potential $V(\mathbf{r})$ in coordinate space, which, when we transform to momentum space, becomes a nonlocal potential $V(\mathbf{p}, \mathbf{p}')$.

We have tried without success to extend the two lemmas. First, consider the lowest partial wave $u_0^{(M)}(r)$ found from the solution of M CDE's for the ground state of a local noncentral potential. Can we show that $u_0^{(M)}(r)$ is nodeless, for all M, $1 < M < \infty$? (The case $M = 1$ follows for solutions of an ordinary differential equation; the case of infinite M is just our first lemma.) The assertion that $u_0^{(M)}(r)$ is nodeless is certainly plausible, since it seems unlikely that a function that is nodeless for one "CDE" would develop a node for some finite number M of CDE's and then lose this node as M goes to infinity.

A very similar, or perhaps identical, problem is the following: Is the S-wave radial function $u(r)$ nodeless for the solution of the deuteron problem with a local tensor force in three dimensions? Here $u(r)$ and the D wave $w(r)$ are the solution of two CDE's, so this problem is closely related to the problem above for $M = 2$. We have never seen a node in $u(r)$ for any of the very large number of published numerical solutions of the deuteron problem with tensor forces; but of course this literature survey does *not* constitute a mathematical proof.

The second lemma does *not* extend to the second excited state $\Psi_2(\mathbf{r})$, and its lowest partial wave. The reason is that the ordering theorem for the number of nodal surfaces and the rank order of eigenvalues breaks down in general, when we go beyond the first excited state. An anisotropic harmonic oscillator in three dimensions provides a simple example of the failure of the ordering theorem. Suppose $\omega_z \gg \omega_x$, where the "spring constants" are $k_z = \frac{1}{2}m\omega_x^2$, and $k_x = \frac{1}{2}m\omega_x^2$. Then the state with one quantum excitation for z motion ($n_z = 1$) with one nodal surface (the plane $z = 0$) would lie in the vicinity of many quantum excitations for x motion (n_x some large integer) with many nodal surfaces.

ACKNOWLEDGMENT

The author wishes to thank many physicists for help in discussion of these problems, and in providing illustrations: R. B. Clare, F. Cuypers, I. Elminyawi, M. Fabre de la Ripelle, and R. M. Lichtenstein.

[1]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953) Vol. I, Chap. VI, Sec. 6.
[2]L. D. Landau and E. M. Lifshitz, *Quantum Mechanics, Non-relativistic Theory* (Addison–Wesley, Reading, MA, 1958), pp. 18 and 61.
[3]A. Messiah, *Quantum Mechanics* (Interscience, New York, 1961), Vol. I, Chap. III, Sec. 12.
[4]J. Ballot and M. Fabre de la Ripelle, Ann. Phys. (NY) **127**, 62 (1980).
[5]Yu. A. Simonov, Sov. J. Nucl. Phys. **3**, 461 (1966); **3**, 755 (1966).
[6]I. Elminyawi, Ph.D. thesis, Rensselaer Polytechnic Institute, August 1984 (unpublished).
[7]I. Elminyawi and J. S. Levinger, J. Chem. Phys. **82**, 905 (1985).
[8]J. Ballot, M. Beiner, and M. Fabre de la Ripelle, *Proceeding of the International Symposium on the Present Status and Novel Developments in the Nuclear Many Body Problems*, Rome 1972, edited by F. Calogero and C. Ciofi degli Atti (Editrice Compositori, Bologna, 1974).
[9]R. B. Clare, Ph.D. thesis, Rensselaer Polytechnic Institute, August 1983 (unpublished).
[10]R. B. Clare and J. S. Levinger, Phys. Rev. C **31**, 2303 (1985).
[11]J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (Wiley, New York, 1952), Chap. III, Sec. 3.
[12]J. A. Wheeler, Phys. Rev. **50**, 643 (1936).
[13]F. Tabakin, Phys. Rev. **174**, 1208 (1968).

Exact symmetries of unidimensional self-similar flow

B. Gaffet
Centre d'Etudes Nucléaires de Saclay, Service d'Astrophysique, 91191 Gif-sur-Yvette Cedex, France

(Received 11 February 1985; accepted for publication 10 May 1985)

In addition to the symmetries that are known to apply to arbitrary flow, the self-similar equations may present other symmetries of their own. We present here such a symmetry of the self-similar unidimensional flow of an adiabatic inviscid fluid, with arbitrary polytropic index and arbitrary power-law entropy distribution. The new symmetry can be extended to the non-self-similar case if the flow is assumed isentropic. A connection with the theory of Riemann invariants is also discussed.

I. INTRODUCTION

We consider here the one-dimensional flow of an adiabatic ideal gas; since we will essentially concentrate on the self-similar case, we assume a constant polytropic index γ and a power-law entropy distribution, which may be written in the form

$$\sigma(M) \equiv P/\rho^{\gamma} \propto M^{-\gamma(1+1/\gamma')}, \tag{1.1}$$

where M is the Lagrangian mass $[dM = \rho(dr - v\,dt)]$.

We start (Sec. II) with the canonical formulation of the self-similar problem, as derived by Gaffet[1] (hereafter referred to as Paper I). We then show (Sec. III) that the self-similar equations present a previously unnoticed symmetry, here denoted by (T_s), which leaves the two indices γ, γ' invariant.

In Sec. IV we show that, at least for isentropic flow, the transformation can be defined for the most general flow without assuming self-similarity. A connection with the theory of Riemann invariants is also discussed.

II. DIMENSIONLESS FORMALISM

We recall here some basic results derived in Paper I concerning the dimensionless form of the Euler equations and the derivation of the self-similar equations. The system of the Euler equations reads

$$\frac{\partial v}{\partial r} + \frac{d \ln \rho}{dt} = 0, \quad \rho\frac{dv}{dt} + \frac{\partial P}{\partial r} = 0,$$
$$P = \rho^{\gamma}\sigma(M), \quad \frac{d}{dt} \equiv \frac{\partial}{\partial t} + v\frac{\partial}{\partial r}, \tag{2.1}$$

where r, t, v, P, ρ are the Eulerian coordinates, velocity, pressure, and density, respectively. The sound velocity is $c = \sqrt{\gamma P/\rho}$. We introduce the following three dimensionless variables:

$$\xi_1 \equiv -Avt, \quad \xi_2 \equiv A(r + \gamma' M/\rho), \quad \xi_3 \equiv A(vt - r), \tag{2.2}$$

where $A \equiv \rho/(\gamma' M) \propto P^{1/\gamma} M^{1/\gamma'}$. We note the identity

$$\sum_{i=1}^{3} \xi_i = 1. \tag{2.3}$$

Two more dimensionless variables also come into play:

$$K \equiv \rho v(vt - r)/\gamma' Mc,$$
$$\tilde{K} \equiv \rho^2 ct(r + \gamma' M/\rho)/\gamma'^2 M^2. \tag{2.4}$$

They satisfy the identity

$$K\tilde{K} = -\xi_1\xi_2\xi_3. \tag{2.5}$$

It turns out to be also useful to introduce the following quantities, differing from ξ by a fixed translation:

$$u_i \equiv \xi_i - \alpha_i,$$
$$\alpha_1 = 2/(\gamma + \gamma'), \quad \alpha_2 = (\gamma' - 1)/(\gamma + \gamma'),$$
$$\alpha_3 = (\gamma - 1)/(\gamma + \gamma'). \tag{2.6}$$

Equation (2.3) implies the relation

$$\sum_i u_i = 0. \tag{2.7}$$

Starting from the Euler equations (2.1), it was shown in Paper I that, for arbitrary indices γ, γ', the following dimensionless partial differential equations hold:

$$2\,\partial_\alpha(K + \tilde{K}) + (\gamma' - 3)\tilde{K}\,\partial_\alpha \ln \xi_2 + (\gamma - 3)K\,\partial_\alpha \ln \xi_3$$
$$= +(\gamma + \gamma')(u_1\partial_\alpha u_2 - u_2\,\partial_\alpha u_1),$$
$$2\,\partial_\beta(K + \tilde{K}) + (\gamma' - 3)\tilde{K}\,\partial_\beta \ln \xi_2 + (\gamma - 3)K\,\partial_\beta \ln \xi_3$$
$$= -(\gamma + \gamma')(u_1\,\partial_\beta u_2 - u_2\,\partial_\beta u_1). \tag{2.8}$$

Here, α, β are characteristic coordinates. The latter are, as usual, defined by[2]

$$d\alpha \propto (v + c)dt - dr,$$
$$d\beta \propto (v - c)dt - dr, \tag{2.9}$$

whence

$$\partial_\alpha r = (v - c)\partial_\alpha t, \quad \partial_\beta r = (v + c)\partial_\beta t.$$

We assume self-similarity from now on. Since all quantities appearing in Eq. (2.8) are dimensionless, they depend on a single variable (the self-similar coordinate), and the partial derivatives ∂_α, ∂_β may be replaced by the total differential symbol d (Paper I, Sec. IV B). In this way, Eqs. (2.8) give rise to two ordinary differential equations

$$\text{(a)}\ \ 2\,d(K + \tilde{K}) + (\gamma' - 3)\tilde{K}\,d \ln \xi_2 + (\gamma - 3)K\,d \ln \xi_3 = 0,$$
$$\text{(b)}\ \ u_1 du_2 - u_2 du_1 = 0. \tag{2.10}$$

Together with Eqs. (2.5)–(2.7), Eqs. (2.10) constitute a complete system of ordinary differential equations for the determination of the self-similar solutions. Since Eq. (2.10) (b) is integrable in closed form, namely,

 0022-2488/85/092165-03$02.50

$$u_2 = m u_1 \quad (m \text{ constant}), \tag{2.11}$$

the system is *of the first-order* only, as is well known.[2]

III. THE INVARIANCE TRANSFORMATION (T_s)

A. A geometrical transformation in three dimensions

In a three-dimensional Euclidean space of coordinates ξ_i $(i = 1, 2, 3)$, let $\boldsymbol{\alpha}$ be a vector with components α_i such that

$$\sum_i \alpha_i = 1, \tag{3.1}$$

and let us denote $\boldsymbol{\beta}$ the vector product

$$\boldsymbol{\beta} = \boldsymbol{\alpha} \wedge \boldsymbol{\xi}. \tag{3.2}$$

We introduce the geometrical transformation defined by the homogeneous formulas

$$\begin{aligned} \xi'_1 &= -\alpha_1 \xi_1 u_2 u_3/\beta_2\beta_3, \\ \xi'_2 &= -\alpha_2 \xi_2 u_3 u_1/\beta_3\beta_1, \\ \xi'_3 &= -\alpha_3 \xi_3 u_1 u_2/\beta_1\beta_2, \end{aligned} \tag{3.3}$$

where the vector $\mathbf{u}$ of components u_1, u_2, u_3 is

$$\mathbf{u} = \boldsymbol{\xi} - \Big(\sum_i \xi_i\Big)\boldsymbol{\alpha} \quad \Big(\sum_i u_i = 0\Big), \tag{3.4}$$

and a prime denotes transformed quantities. The formulas are obviously invariant under circular permutation of the indices.

The straight line through the origin parallel to $\boldsymbol{\alpha}$ is, since $\boldsymbol{\beta} = \mathbf{0}$ there, a singular line for the transformation, which may be called the *axis*. An essential property of the transformation is that, on straight lines intersecting the axis, it reduces to *an affine transformation* of the coordinates ξ_i, since β_i/u_i $(i = 1, 2, 3)$ are then constants; in particular such straight lines transform into other straight lines.

By substitution we derive the identity

$$\sum_i \xi'_i = \sum_i \xi_i, \tag{3.5}$$

which means that the planes of the equation $\xi_1 + \xi_2 + \xi_3 = \text{const}$ are globally invariant.

The resulting tranformation formulas for the u_i's are obtained as follows:

$$u'_1 = -\alpha_1 \Phi/\beta_2\beta_3 \quad \text{(plus circular permutation)}, \tag{3.6}$$

where

$$\Phi = u_1 u_2 u_3 - \{\alpha_3(\alpha_1 + \alpha_2)u_2^2 + 2\alpha_2\alpha_3 u_2 u_3 + \alpha_2(\alpha_1 + \alpha_3)u_3^2\}.$$

Taking account of the property $\Sigma_i u_i = 0$, it is easily verified that Φ is invariant under circular permutation of the indices. Finally, the transformation formulas for $\boldsymbol{\beta}$ read as follows:

$$\beta'_1 = \frac{\alpha_2\alpha_3 u_1 \Phi}{\beta_1\beta_2\beta_3} \quad \text{(plus circular permutation)}. \tag{3.7}$$

The property $\beta_i/u_i = \text{const}$ $(i = 1, 2, 3)$ characterizing straight lines intersecting the axis is thus invariant; therefore, the transformed straight lines also intersect the axis.

Using Eqs. (3.3), (3.6), and (3.7) we immediately see that *the transformation squared is the identity*, i.e., it coincides with its inverse.

B. The invariance transformation (T_s)

Identifying $\boldsymbol{\xi}$, $\mathbf{u}$, $\boldsymbol{\alpha}$ with the dimensionless quantities defined in Sec. II, and remembering the relation (2.3) $(\Sigma_i \xi_i = 1)$, the transformation formulas (3.3) read as follows:

$$\xi'_1 = \alpha_1\xi_1(\xi_2 - \alpha_2)(\xi_3 - \alpha_3)/(\alpha_2\xi_1 - \alpha_1\xi_2)(\alpha_3\xi_1 - \alpha_1\xi_3) \tag{3.8}$$

(plus circular permutation).

This is, in effect, an affine transformation of the coordinates ξ_i in the case of self-similar flow, since such flows are represented (in these coordinates) by straight lines intersecting the axis, as shown in Sec. II [Eq. (2.11)]. The transformation is an *exact symmetry* of the complete system of the self-similar equations [Eqs. (2.5)–(2.7) and (2.10)], as may be verified by direct substitution into the equations. We will denote this symmetry by the symbol (T_s).

The transformation formulas for the remaining dimensionless variables K and $\tilde{K}$, defined by Eqs. (2.5) and (2.10) (a), reduce to a rescaling by a constant factor λ:

$$K' = \lambda K, \quad \tilde{K}' = \lambda \tilde{K}, \tag{3.9}$$

where λ is determined by the condition (2.5):

$$\lambda = \left(\frac{\xi'_1\xi'_2\xi'_3}{\xi_1\xi_2\xi_3}\right)^{1/2} = \left(\frac{u_1u_2u_3}{\beta_1\beta_2\beta_3}\right)(-\alpha_1\alpha_2\alpha_3)^{1/2}. \tag{3.10}$$

This is a constant in the case of self-similar flow, since, as already observed, the ratios u_i/β_i $(i = 1, 2, 3)$ are then constant. Equation (2.10) (a) is thus manifestly invariant.

C. The transformation formula for dimensional variables

When the dimensionless variables are known the dimensional variables can be deduced by quadrature [see, e.g., Paper I, Eq. (4.5)]; thus the time coordinate t is, assuming a self-similar flow, given by

$$t = \alpha\beta \exp \int \frac{K}{(K - \tilde{K})} d \ln\left(\frac{\xi_1}{\xi_3}\right), \tag{3.11}$$

where α, β constitute a particular choice of characteristic coordinates, normalized in such a way that

$$\frac{\beta}{\alpha} = \exp \int \frac{d\xi_2}{(K - \tilde{K})}. \tag{3.12}$$

Since (T_s) amounts to a rescaling of K, $\tilde{K}$ and of the ξ_i's, the transformed characteristic coordinates α', β' are, according to (3.12), such that

$$\beta'/\alpha' = (\beta/\alpha)^k \quad (k \text{ constant}). \tag{3.13}$$

In order to make the characteristic curves invariant, we choose the solution $\alpha' = \alpha^k$, $\beta' = \beta^k$, thus completing the definition of the symmetry (T_s) (see Ref. 3).

In the same way, it is evident that the integral in (3.11) does not change under (T_s); therefore the transformed time coordinate reads as follows:

$$\begin{aligned} t' &= (\alpha\beta)^k \exp \int \frac{K}{(K - \tilde{K})} d \ln\left(\frac{\xi_1}{\xi_3}\right) \\ &= (\alpha\beta)^{(k-1)} t. \end{aligned}$$

It is convenient to choose $\alpha^{(k-1)}$, $\beta^{(k-1)}$ as new characteristic coordinates, so that the transformation formula for t simply reads as follows:

$$t' = \alpha\beta t. \tag{3.14}$$

The above result is of rather uncommon and interesting form and suggests a relation with the theory of Riemann invariants, as shown in Sec. IV.

IV. THE GENERALIZATION OF (T_s) TO NON-SELF-SIMILAR (ISENTROPIC) FLOW

It would be of great interest to have a generalization of (T_s) that remains valid for non-self-similar flow. Equation (3.14) then would determine the product $\alpha\beta$ of the characteristic coordinates, i.e., the product of two Riemann invariants, in closed form. This would constitute a very important result, since, when two Riemann invariants are known, *the general solution can be explicitly constructed in closed form* [see, e.g., Gaffet,[4] Eq. (5.3)]. We presently show that (T_s) can indeed be so generalized, at least in the case of *isentropic* flow.

We treat the case of a polytrope $\gamma = 3$. By the classical transformation of isentropic flow (see Landau and Lifshitz[5]), corresponding results can be derived for all adiabatic indices of the general form

$$\gamma = (2n+3)/(2n+1) \quad (m = 1, 2, 3, \ldots, \infty).$$

The characteristic form of the Euler equations is, assuming isentropic flow and $\gamma = 3$ (see Refs. 2 and 4),

$$\text{(a)}\quad \partial_\alpha(v - c) = 0, \quad \partial_\beta(v + c) = 0,$$
$$\text{(b)}\quad \partial_\alpha M = -c^2\,\partial_\alpha t, \quad \partial_\beta M = c^2\,\partial_\beta t. \tag{4.1}$$

The first two equations may be solved without loss of generality in the form

$$v = \alpha + \beta, \quad c = \alpha - \beta,$$

so that the resulting system for M and t reads as follows:

$$\begin{aligned} \partial_\alpha M &= -(\alpha - \beta)^2\,\partial_\alpha t, \\ \partial_\beta M &= +(\alpha - \beta)^2\,\partial_\beta t. \end{aligned} \tag{4.2}$$

Eliminating M yields a second-order equation for t:

$$(\alpha - \beta)\partial^2_{\alpha\beta} t + \partial_\beta t - \partial_\alpha t = 0. \tag{4.3}$$

We now introduce the following transformation, denoted by (T):

$$\text{(a)}\quad v' = v/(v^2 - c^2),$$
$$\text{(b)}\quad c' = -c/(v^2 - c^2), \tag{4.4}$$
$$\text{(c)}\quad t' = (v^2 - c^2)t.$$

It is clear that $(T)^2$ is the identity; also, the pair of equations (4.1) (a) is obviously invariant under (T). Noting that $c' = (\beta - \alpha)/(4\alpha\beta)$, $t' = 4\alpha\beta t$, the second pair of equations (4.1) (b) determines the transformed quantity M' as follows:

$$4\,\partial_\alpha M' = -[(\beta - \alpha)^2/\alpha^2\beta](t + \alpha\partial_\alpha t),$$
$$4\,\partial_\beta M' = +[(\beta - \alpha)^2/\alpha\beta^2](t + \beta\partial_\beta t).$$

The Cauchy condition of integrability $[\partial_\beta(\partial_\alpha M') = \partial_\alpha(\partial_\beta M')]$ precisely coincides with Eq. (4.3) and is thus automatically satisfied. This completes the proof that (T) is an exact symmetry of isentropic flow, with $\gamma = 3$. Corresponding symmetries for the cases $\gamma = \frac{5}{3}, \frac{7}{5}$, etc. may be deduced by application of the above-mentioned transformation of Landau and Lifshitz.

It is straightforward to show that (T) reduces, for self-similar flow, to the transformation (T_s) introduced in the preceding sections. Comparing Eqs. (3.14) and (4.4) (c) we observe that, as predicted, the ratio $t'/t \equiv v^2 - c^2$ is the product of two Riemann invariants.

V. CONCLUSION

Self-similar solutions are widely used for solving problems in fluid mechanics and it is thus desirable to have, as far as possible, a complete knowledge of the symmetries presented by the self-similar ordinary differential equations. We have considered here the case of unidimensional flow of an adiabatic inviscid polytropic fluid.

The list of symmetries will in the first place include all those which are already known to exist independently of the self-similar assumption: there are essentially three such nontrivial symmetries, valid for arbitrary entropy distributions, whose properties have recently been reviewed by Gaffet[1,4,6]; they are denoted by the symbols (T'), $(\bar{T})$, (T^*).

We present here an additional symmetry, denoted (T_s), which applies to the self-similar flow of *arbitrary* polytropes with *arbitrary* power-law entropy distribution. We have been able to obtain a generalization also valid for non-self-similar flow, in the isentropic case (Sec. IV).

It is of interest to note that, should one generalize (T_s) to non-self-similar flow, one would thereby [see Eqs. (3.14) and (4.4) (c)] be able to derive the product of two Riemann invariants explicitly, and thus would obtain the general solution of the corresponding Euler equations in closed form.

[1] B. Gaffet, J. Math. Phys. **25**, 245 (1984) (Paper I).

[2] R. Courant and K. O. Friedrichs *Suspersonic Flow and Shock Waves* (Interscience, New York, 1948).

[3] The operation of (T_s) on dimensional variables was, until now, undetermined.

[4] B. Gaffet, J. Fluid Mech. **134**, 179 (1983).

[5] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Addison–Wesley, Reading, MA, 1959), Chap. X, p. 488, Eq. 98.7.

[6] B. Gaffet, Physica D **11**, 287 (1984).

Generalized Stäckel matrices

E. G. Kalnins
Mathematics Department, University of Waikato, Hamilton, New Zealand
W. Miller, Jr.
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

(Received 27 March 1985; accepted for publication 10 May 1985)

Stäckel and differential-Stäckel matrices are generalized so that the matrix elements may be functions of the derivatives of the dependent variable as well as the independent variable. The inverses of these matrices are characterized and it is shown that for significant classes of linear and nonlinear partial differential equations, variable separation is accomplished via this generalized Stäckel mechanism.

I. INTRODUCTION

In Ref. 1 the authors introduced a general definition of additive variable separation for a partial differential equation

$$H(x_I,u,u_{I,i}) = E, \tag{1.1}$$

where E is a parameter, $x_1,\ldots,x_N$ are the independent variables, u is the dependent variable, and $u_{I,i} = \partial^i_{x_I} u$, $i = 1,2,\ldots$. A separable solution of (1.1) is a solution of the form $u = \Sigma_{J=1}^N S^{(J)}(x_J,E)$. Our definition is a straightforward extension of Levi-Civita's definition for first-order equations.[2] We let n_I be the largest number l such that $\partial_{u_{I,l}} H \equiv H_{u_{I,l}} \neq 0$. To avoid discussion of degenerate cases we require $n_I > 0$ for each I (but n_I is finite).

Let the truncated differentiation operator $\widetilde{D}_I$ be defined by

$$\widetilde{D}_I = \partial_{x_I} + u_{I,1}\,\partial_u + u_{I,2}\,\partial_{u_{I,1}} + \cdots + u_{I,n_I}\partial_{u_{I,n_I-1}}. \tag{1.2}$$

In Ref. 1 we showed that every separable solution u of (1.1) satisfies the integrability conditions

$$\begin{aligned} &H_{u_{I,n_I}} H_{u_{J,n_J}} (\widetilde{D}_I\widetilde{D}_J H) + H_{u_{I,n_I}u_{J,n_J}} (\widetilde{D}_I H)(\widetilde{D}_J H) \\ &\quad - H_{u_{J,n_J}} (\widetilde{D}_I H)(\widetilde{D}_J H_{u_{I,n_I}}) - H_{u_{I,n_I}} (\widetilde{D}_J H)(\widetilde{D}_I H_{u_{J,n_J}}) = 0, \\ &\quad 1 \leqslant I < J \leqslant N. \end{aligned} \tag{1.3}$$

If (1.3) is an identity in the dependent variables $u, u_{K,k}$, we say that $\{x_I\}$ is a *regular separable coordinate system* for $H = E$. In this case the separable solutions involve $\Sigma_{J=1}^N n_J + 1$ independent parameters: u and the derivations $u_{I,i}$, $1 \leqslant I \leqslant N$, $1 \leqslant i \leqslant n_I$ can be prescribed arbitrarily at a given point $\mathbf{x}^0$. If conditions (1.3) do not hold identically then the separation is *nonregular* and separable solutions, if they exist, will involve strictly fewer parameters than the regular case. (The standard examples of variable separation for the differential equations of mathematical physics all correspond to regular separation.) Multiplicative separation is handled in this framework by passing to a new dependent variable $v = \ln u$. There is a modified definition of variable separation for (1.1) when $E \equiv 0$, which we will not discuss here.[1,3]

It would be of great interest to know the general solution of (1.3) so that the mechanism of variable separation could be determined in all cases. However, the general solution is not even known for the Levi-Civita case $n_I = \cdots = n_N = 1$. (The solution has recently been worked out for Hamilton–Jacobi equations on pseudo-Riemannian manifolds.[4])

Historically, the fundamental mechanism for variable separation has been the Stäckel matrix.[4] However, the Stäckel mechanism is not sufficient to encompass all types of separation given by solutions of (1.3). In order to describe the solutions of the integrability conditions for additive separation of linear equations $Lu = E$ and $Lu = 0$ the authors introduced differential-Stäckel (D-Stäckel) matrices, a nontrivial extension of Stäckel matrices.[3] Here, we further extend D-Stäckel matrices by permitting the matrix elements to be functions of the derivatives $u_{I,i}$ as well as the independent variables x_I. (For ordinary Stäckel matrices this is a straightforward extension. For D-Stäckel matrices it is more difficult.)

In Sec. II we define generalized D-Stäckel matrices and characterize their inverse matrices via a system of partial differential equations. This section is modeled on Ref. 3 (in which ordinary D-Stäckel matrices are treated) but Theorem 1 leads to some complications.

In Sec. III we present several classes of linear and nonlinear partial differential equations for which we can characterize the possible mechanisms of variable separation, and we show that they all correspond to generalized D-Stäckel form. For all cases treated we have $H_{u_{I,n_I}u_{J,n_J}} = 0$ in (1.3) for $I \neq J$. (The cases where the mixed partial derivatives do not vanish are much more complicated.) Even with this restriction we do not yet know if generalized D-Stäckel form is sufficient to describe all variable separation or if additional mechanisms exist.

All functions appearing in this paper are assumed to be locally real analytic. Furthermore, functions $f(x_K,u_{K,k})$ are assumed to be analytic as functions of the $u_{K,k}$ in the neighborhood of $u_{K,k} = 0$. If we require that a nonzero f is invertible we mean that $f(x_K,0) \neq 0$ for the $\{x_k\}$ in some neighborhood on R^N so that $f^{-1}(x_K,u_{K,k})$ is also analytic.

II. GENERALIZED D-STÄCKEL MATRICES

Consider a coordinate set $x_1,\ldots,x_N$ and let $n_1,\ldots,n_N$ be positive integers with $n = \Sigma_{I=1}^N n_I$. Let $S = (S_{(I,i),l}(x_I))$ be an $n \times n$ nonsingular matrix with the properties

(1) $S_{(I,i),l}(x_I) = \frac{d^{i-1}}{dx_I^{i-1}} S_{(I,1),l}(x_I), \quad i = 1,2,...,n_I;$ (2.1)

(2) $T^{1,(J,j)} \neq 0, \quad J = 1,...,N, \quad j = 1,...,n_J,$

where $T = S^{-1}$, i.e.,

$$\sum_{l=1}^{u} S_{(I,i),l}(x_I) T^{l,(J,j)} = \delta_{(I,i)}^{(J,j)}. \tag{2.2}$$

We call a matrix S with these properties a *differential-Stäckel matrix* (D-Stäckel matrix). [Here the rows of S are designated by the index (I,i), where $I = 1,...,N$, $i = 1,...,n_I$, whereas the columns of S are designated by the index $l = 1,2,...,n$. Row (I,i) depends only on x_I and is the $i-1$ derivative of row $(I,1)$. The index notation for T is defined similarly but with rows and columns interchanged.] If $n_I = 1$ for all I so $n = N$, then S is an ordinary Stäckel matrix.[5]

Set $H_{(J,j)} = T^{1,(J,j)}$. In Ref. 3 it is shown that for S a D-Stäckel matrix the system of equations

$$\partial_I \rho_{(J,j)} = (\rho_{(I,n_I)} - \rho_{(J,j)}) \partial_I \ln H_{(J,j)} + (\rho_{(I,n_I)} - \rho_{(J,j-1)}) \frac{H_{(J,j-1)}}{H_{(J,j)}} \delta_I^J, \tag{2.3}$$

$$I, J = 1,...,N, \quad h = 1,...,n_J,$$

admits a full linearly independent set of n vector valued solutions $\{\rho_{(J,j)}^l\}$, $l = 1,...,n$. Conversely if the n nonzero functions $\{H_{(J,j)}\}$ are such that (2.3) admits a linearly independent set of vector valued solutions then there is an $n \times n$ D-Stäckel matrix S such that $H_{(J,j)} = T^{1,(J,j)}$. See Refs. 6 and 7 for similar treatments of ordinary Stäckel matrices.

The integrability conditions for (2.3) are

$$\partial_{IJ} H_{(P,p)} - \partial_I H_{(P,p)} \partial_J \ln H_I - \partial_J H_{(P,p)} \partial_I H_J = 0,$$
$$P \neq I, J, \quad p = 1,...,n_P, \tag{2.4a}$$

$$\partial_{IJ} H_{(J,j)} - \partial_I H_{(J,j)} \partial_J \ln H_I - \partial_J H_{(J,j)} \partial_I \ln H_J = H_{(J,j-1)} \partial_I \ln H_J - \partial_I H_{(J,j-1)}, \quad j = 1,...,n_J, \tag{2.4b}$$

where $I \neq J$, $H_I \equiv H_{(I,n_I)}$, and $H_{(J,0)} \equiv 0$. By Theorem 1 of Ref. 3, conditions (2.4) are necessary and sufficient that the n nonzero functions $\{H_{(J,j)}\}$ can be expressed in the form $H_{(J,j)} = T^{1,(J,j)}$ for T the inverse of a D-Stäckel matrix S.

When Eqs. (2.4) hold the partial differential equation

$$\sum_{I=1}^{N} \sum_{i=1}^{n_I} (D_I^{i-1} Q_I) H_{(I,i)} = \mathrm{E} \tag{2.5}$$

admits regular additive separation in the coordinates $x_1,...,x_N$, where E is a parameter (which could be zero), $Q_I(x_I, u_{I,i})$ is a function of x_I, and a finite number of derivatives $u_{I,1}, u_{I,2},...,u_{I,q_I}$ with $m_I \geqslant 1$, $\partial_{u_{I,q_I}} Q_I \neq 0$, and

$$D_I = \partial_{x_i} + \sum_{i=1}^{\infty} u_{I,i+1} \partial_{u_{I,i}} \tag{2.6}$$

is the Ith total derivative. Indeed the separation equations are

$$D_I^{i-1} Q_I + \sum_{l=1}^{n} S_{(I,i),l}(x_I) \lambda_l = 0, \tag{2.7}$$

$$1 \leqslant I \leqslant N, \quad 1 \leqslant i \leqslant n_I, \quad \lambda_1 = -E,$$

where the λ_l are the separation parameters. The separable solutions $u = \sum_{I=1}^N u^{(I)}(x_I, \lambda_l)$ are obtained by integrating N ordinary differential equations (of order m_I)

$$Q_I(x_I, u_{I,i}) + \sum_{l=1}^{n} S_{(I,1),l}(x_I) \lambda_l = 0. \tag{2.8}$$

The remaining $n - N$ equations are redundant. The number of parameters in the solution u is $\Sigma_I q_I + n - N + 1$.

Stäckel and D-Stäckel matrices can clearly be generalized to include dependent variables, thus incorporating a wider class of separable partial differential equations than (2.5). For this we consider coordinates $x_1,...,x_N$, let $n = \sum_{I=1}^N n_I$, where the n_I are positive integers, and let $\Omega_1,...,\Omega_N$ be non-negative integers. Then a nonsingular $n \times n$ matrix $S = (S_{(I,i),l}(x_I, u_{I,j}))$, with the properties

(1) $S_{(I,i),l}(x_I, u_{I,j}) = D_I^{i-1} S_{(I,1),l}(x_I, u_{I,j})$,

$i = 1,2,...,n_I$,

(2) $T^{1,(J,j)} \neq 0, \quad J = 1,...,N, \quad j = 1,...,n_J,$

where $T = S^{-1}$,

(3) $S_{(I,1),l} = S_{(I,1),l}(x_I, u_{I,1},...,u_{I,\Omega_I})$,

with $\partial_{u_{I,\Omega_I}} S_{(I,1),l} \neq 0$ for some l if $\Omega_I > 0$,

is a *(generalized) D-Stäckel matrix.*

A generalized D-Stäckel matrix S can be used to construct partial differential equations that permit regular separation in the coordinates x_I. Set $H_{(J,j)} = T^{1,(J,j)}$. It is then easy to show that equations of the form (2.5) permit regular separation.

Characterization of generalized D-Stäckel form in terms of differential equations satisfied by the $H_{(J,j)}$ is not particularly difficult. In analogy with the derivation of Eq. (1.3) in Ref. 3, we can easily show that

$$D_I \rho_{(J,j)} = (\rho_{(I,n_I)} - \rho_{(J,j)}) D_I \ln H_{(J,j)} + (\rho_{(I,n_I)} - \rho_{(J,j-1)})(H_{(J,j-1)}/H_{(J,j)}) \delta_I^J, \tag{2.9}$$

$$I, J = 1,...,N, \quad j = 1,...,n_J,$$

where $\rho_{(J,j)}$ (as well as $H_{(J,j)}$) depends only on the variables $x_I, u_{I,1} u_{I,2},...,u_{I,s_I}$, $I = 1,...,N$, and $s_I = n_I + \Omega_I - 1$ if $\Omega_I > 0$, $s_I = 0$ if $\Omega_I = 0$, admit a full linearly independent set of n vector valued solutions $\{\rho_{(J,j)}^l\}$, $l = 1,...,n$ if and only if the nonzero functions $H_{(J,j)}$ are obtainable from a generalized D-Stäckel matrix S.

The total differential equation (2.9) is equivalent to a sequence of partial differential equations in which the left-hand side assumes the form $\partial_{u_{I,i}} \rho_{(J,j)}$, $i = 0,...,s_I$. (We make the convention that $x_I \equiv u_{I,0}$.) Indeed, equating coefficients of u_{I,s_I+1} on both sides of (2.9) we have

$$\partial_{u_{I,s_I}} \rho_{(J,j)} = (\rho_{(I,n_I)} - \rho_{(J,j)}) \partial_{u_{I,s_I}} \ln H_{(J,j)}. \tag{2.10}$$

We can obtain the derivatives $\partial_{u_{I,i}} \rho_{(J,j)}$, $i = 1,...,s_I - 1$ recursively from (2.9) and (2.10) through the relation $\partial_{u_{I,i-1}} = [\partial_{u_{I,i}}, D_I] = \partial_{u_{I,i}} D_I - D_I \partial_{u_{I,i}}$. Finally, $\partial_{x_I} = D_I - \sum_{i=1}^{s_I} u_{I,i+1} \partial_{u_{I,i}}$, when applied to $\rho_{(J,j)}$.

We can obtain integrability conditions for Eq. (2.9) by computing $D_J(D_I \rho_{(K,k)}) = D_I(D_J \rho_{(K,k)})$, $I \neq J$, and equating coefficients of $\rho_{(L,l)}$ on each side of the resulting expression:

$$D_I D_J H_{(P,p)} - D_I H_{(P,p)} D_J \ln H_I - D_J H_{(P,p)} D_I \ln H_J = 0,$$
$$P \neq I, J, \quad p = 1,...,n_P, \tag{2.11a}$$
$$D_I D_J H_{(J,j)} - D_I H_{(J,j)} D_J \ln H_I - D_J H_{(J,j)} D_I \ln H_J$$
$$- H_{(J,j-1)} D_I \ln H_J + D_J H_{(J,j-1)} = 0,$$
$$j = 1,...,n_J. \tag{2.11b}$$

Here $H_J \equiv H_{(J,n_J)}$, $H_{(J,0)} \equiv 0$. It is not entirely clear, however, that Eqs. (2.11) are the complete set of integrability conditions. For these we need to compute $\partial_{u_{J,j}}(\partial_{u_{I,i}} \rho_{(K,k)}) = \partial_{u_{I,i}}(\partial_{u_{J,j}} \rho_{(K,k)})$, $i = 0,...,s_I$, $j = 0,...,s_J$.

Theorem 1: Conditions (2.11a) and (2.11b) are necessary and sufficient for complete integrability of Eqs. (2.9), hence they are the necessary and sufficient conditions for the existence of a generalized D-Stäckel matrix S such that $H_{(J,j)} = T^{1,(J,j)} \neq 0$.

Proof: It is already evident that conditions (2.11) are necessary for the existence of a generalized D-Stäckel matrix. To prove they are sufficient we consider the integrability conditions ($I \neq J$)

$$D_J(D_I \rho_{(K,k)}) - D_I(D_J \rho_{(K,k)})$$
$$= -\rho_{(J,n_J)} \frac{\overleftarrow{D}_I \overrightarrow{D}_J H_{(K,k)}}{H_{(K,k)}} + \rho_{(I,n_I)} \frac{\overleftarrow{D}_I \overrightarrow{D}_J H_{(K,k)}}{H_{(K,k)}}, \tag{2.12}$$

where $\overleftarrow{D}_I \overrightarrow{D}_J H_{(K,k)}$ is the left-hand side of expression (2.11a) if $K = P \neq I, J$ or expression (2.11b) if $K = J$. The left-hand side of (2.12) is computed directly from (2.9). Clearly (2.12) vanishes for a complete set of solutions ρ if and only if $\overleftarrow{D}_I \overrightarrow{D}_J H_{(K,k)} = 0$. The integrability condition

$$\partial_{u_{J,s_J}}(\partial_{u_{I,s_I}} \rho_{(K,k)}) - \partial_{u_{I,s_I}}(\partial_{u_{J,s_J}} \rho_{(K,k)}) \tag{2.13}$$

can be obtained from (2.12) by equating coefficients of $u_{J,s_J+1} u_{I,s_I+1}$:

$$\overleftarrow{\partial}_{u_{I,s_I}} \overrightarrow{\partial}_{u_{J,s_J}} H_{(K,k)} = 0. \tag{2.14}$$

Furthermore, equating coefficients of u_{J,s_J+1} and u_{I,s_I+1}, respectively, we find the conditions corresponding to $\partial_{u_{J,s_J}}(D_I \rho)$,

$- D_I(\partial_{u_{J,s_J}} \rho)$ and $D_J(\partial_{u_{I,s_I}} \rho) - \partial_{u_{I,s_I}}(D_J \rho)$:

$$\overleftarrow{D}_I \overrightarrow{\partial}_{u_{J,s_J}} H_{(K,k)} = 0, \ \overleftarrow{\partial}_{u_{I,s_I}} \overrightarrow{D}_J H_{(K,k)} = 0. \tag{2.15}$$

Note that conditions (2.14) and (2.15) can be obtained directly from (2.11) by equating coefficients of $u_{J,s_J+1} u_{I,s_I+1}$, and u_{J,s_J+1} and u_{I,s_I+1}, respectively.

We can now derive the conditions corresponding to $\partial_{u_{J,j}}(D_I \rho) - D_I(\partial_{u_{J,j}} \rho)$ recursively from the above expression through repeated application of the identity

$$\partial_{u_{J,j}} = [\partial_{u_{J,j+1}}, D_J], \quad j = 1,2,...,s_J - 1. \tag{2.16}$$

From this result and (2.13) we can obtain the conditions corresponding to

$$\partial_{u_{J,j}}(\partial_{u_{I,i}} \rho) - \partial_{u_{I,i}}(\partial_{u_{J,j}} \rho) \tag{2.17}$$

recursively through application of the identity $\partial_{u_{I,i}} = [\partial_{u_{I,i+1}}, D_I]$, $i = 1,2,...,s_I - 1$. At each stage of this process the integrability conditions are linear combinations of (2.11), (2.14), (2.15), and their derivatives, hence they are implied by (2.11). Finally the conditions $\partial_{x_J}(\partial_{x_I} \rho) - \partial_{x_I}(\partial_{x_J} \rho)$ can be obtained from

$$[\partial_{x_I}, \partial_{x_J}] = \left[D_{x_I} - \sum_{i=1}^{s_I} u_{I,i+1} \partial_{u_{I,i}}, \quad D_J - \sum_{j=1}^{s_J} u_{J,j+1} \partial_{u_{J,j}} \right].$$

Again the integrability conditions are implied by (2.11a) and (2.11b). Q.E.D.

Now that we have succeeded in characterizing generalized D-Stäckel matrices in terms of the first column of their inverse matrices, we can extend the notion of a Stäckel multiplier to this situation. Suppose n nonzero functions $H_{(J,j)}(x_I, u_{I,i})$ satisfy conditions (2.11), and hence determine a generalized D-Stäckel matrix S. A nonzero function $f(x_I, u_{I,i})$ such that the functions $\widetilde{H}_{(J,j)} = H_{(J,j)}/f$ also satisfy conditions (2.11) is a *(generalized) D-Stäckel multiplier* for the system $\{H_{(J,j)}\}$.

Theorem 2: The following are equivalent characterizations of D-Stäckel multipliers f: (1) f satisfies the equations

$$D_I D_J f - D_I \ln H_J D_J f - D_J \ln H_I D_I f = 0,$$
$$I \neq J, \quad H_I = H_{(I,n_I)}. \tag{2.18a}$$

(2) there exist N functions $\varphi^J(x_J, u_{J,j})$, $D_I \varphi^J = 0$ for $I \neq J$, such that

$$f(x_K, u_{K,k}) = \sum_{J=1}^{N} \sum_{j=1}^{n_J} (D_J^{j-1} \varphi^J) H_{(J,j)}. \tag{2.18b}$$

Proof: It is obvious from Theorem 1 that (2.18a) is equivalent to the definition of a D-Stäckel multiplier. Now suppose f is a D-Stäckel multiplier. Then there is an $n \times n$ D-Stäckel matrix $\widetilde{S}$ (for $\widetilde{H}_{(J,j)} = H_{(J,j)}/f$) such that

$$\widetilde{T}^{1,(J,j)} = H_{(J,j)}/f. \tag{2.19}$$

The elements in the first column of the D-Stäckel matrix S can be denoted

$$\widetilde{S}_{(J,j),1} = \frac{d^{j-1}}{dx_J^{j-1}} \varphi^J(x_J, u_{J,l})$$

for N functions φ^J. Multiplying both sides of (2.19) by $\widetilde{S}_{(J,j),1}$ and summing over the index (J,j) we obtain (2.18b).

Conversely, suppose f is defined by (2.18b) for some functions $\varphi^J(x_J, u_{J,l})$. From this expression and conditions (2.10) it is straightforward to verify that f satisfies Eq. (2.18a). Hence f is a D-Stäckel multiplier. Q.E.D.

Although Theorem 1 is valid only when $H_{(J,j)} \neq 0$ for all (J,j), Eqs. (2.11) make sense as long as $H_{(J,n_J)} \equiv H_J \neq 0$, even if some of the remaining $H_{(J,j)}$ vanish. We need to determine the significance of those solutions of (2.11) for which it is only required that $H_J \neq 0$. Furthermore it will be useful to determine the effect on the solutions of replacing each H_J by $g_J(x_J, u_{J,j}) H_J$, where g_J is invertible in a neighborhood of the point $(x_J^0, 0)$, so that g_J^{-1} will also be analytic in the $u_{(J,j)}$ in a neighborhood of the point.

To answer these questions it is useful to write Eqs. (2.11) in the form

$$A_{IJ} H_{(P,p)} = 0, \quad P \neq I, J,$$
$$A_{IJ} H_{(J,j)} = B_{IJ} H_{(J,j-1)}, \quad H_{(J,0)} = 0, \quad I \neq J, \tag{2.20}$$

where

$$A_{IJ} = D_I D_J - D_J \ln H_I D_I - D_I \ln H_J D_J,$$
$$B_{IJ} = -D_I + D_I \ln H_J, \quad H_J = H_{(J,n_J)} \neq 0, \tag{2.21}$$
$$1 \leqslant J \leqslant N, \quad 1 \leqslant j \leqslant n_J, \quad \sum_J n_J = n.$$

We require that the $H_{(J,j)}$ and the other functions appearing in the lemmas depend only on the variables x_I and $u_{I,i}$, $1 \leqslant I \leqslant N$, $1 \leqslant i \leqslant s_I$.

Suppose we are given N nonzero functions H_J satisfying $A_{IJ} H_P = 0$ for $P \neq I, J$, $I \neq J$, and N nonzero functions g_J satisfying $D_I g_J = 0$ for $I \neq J$. Our task will be to construct a finite set of functions $\mathscr{H}_{(J,j)}$ with $H_{(J,n_J)} = g_J H_J$ such that Eqs. (2.12) are satisfied. Initially the value of n_J is unknown.

The construction process is based on the second equation (2.20) which we rewrite as follows:

$$D_I\left(\frac{H_{(K,k-1)}}{H_K}\right) = \frac{-A_{IK} H_{(K,k)}}{H_K}, \quad I \neq K. \tag{2.22}$$

If $H_{(K,k)}$ is known we can construct $H_{(K,k-1)}$ from (2.22) by quadrature.

Lemma 1: Suppose the N nonzero functions H_P satisfy $A_{IJ} H_P = 0$ for $P \neq I,J$ and the function $H_{K,k}$ (fixed K,k) satisfies $A_{IJ} H_{K,k} = 0, K \neq I, J, I \neq J$. Then the $N-1$ equations (2.20) are compatible and have the general solution

$$H_{(K,k-1)} = \widetilde{H}_{(K,k-1)} + f^{(k-1)}(x_K, u_{K,i}) H_K, \tag{2.23}$$

where $\widetilde{H}_{(K,k-1)}$ is a particular solution and $f^{(k-1)}$ is an arbitrary function of $x_K, u_{K,i}$. The solution satisfies

$$A_{IJ} H_{(K,k-1)} = 0, \quad K \neq I, J, \quad I \neq J. \tag{2.24}$$

It follows that for each K we can always construct functions $H_{(K,k-1)}$ through a recursive procedure using (2.22) such that the first equation (2.20) is automatically satisfied. At each step the solution $H_{(K,k-1)}$ is arbitrary up to the additive term $f^{(k-1)}(x_K, u_{K,i}) H_K$ and we choose one of these solutions. Thus we generate an infinite sequence $\{H_{(K,k)} = H_K^{(l)}\}$, $l = 0,1,2,\ldots$, where $n_K - l = k$ (but n_K is unknown) and

$$A_{IK} H_K^{(l)} = B_{IK} \mathrm{H}_K^{(l+1)}, \quad I \neq K, \quad H_K = H_K^{(0)}. \tag{2.25}$$

Suppose there is a smallest finite positive integer n_K for which functions $f_{(i)}(x_K, u_{K,j})$ exist such that

$$H_K^{(m_K)} = \sum_{i=0}^{m_K - 1} f_{(i)}(x_K, u_{K,j}) H_K^{(i)}. \tag{2.26}$$

The following lemmas can be verified by straightforward induction using the properties

(1) $B_{IK} F(x_j, u_{J,j}) = 0$, for all $I \neq K$,
$$\Leftrightarrow F = f(x_K, u_{K,j}) H_K \tag{2.27}$$

(2) $A_{IK}(f(x_K, u_{K,j}) H_K^{(l)}) = B_{IK}(f H_K^{(l+1)} - D_K f H_K^{(l)}).$ (2.28)

Lemma 2: Each $H_K^{(m_K+s)}$, $s = 0,1,2,\ldots$, is a linear combination of the finite set $\{H_K^{(l)}: l = 0,\ldots,m_K - 1\}$ with coefficients that are functions of $x_K, u_{K,j}$.

Lemma 3: Let $\{\mathscr{H}_K^{(l)}\}$, $\{h_K^{(l)}\}$, $l = 0,1,2,\ldots$, be two sequences constructed by the procedure (2.22), (2.23) such that $\mathscr{H}_K^{(0)} \equiv \mathscr{H}_K = g_0 h_K$, where g_0 is invertible and $D_I g_0 = 0$ for $I \neq K$. Then there is a sequence $g_1, g_2, \ldots$ with $D_I g_i = 0$ for $I \neq K$ and expressions $L_{i,j}(g_0, g_1, \ldots, g_{i-j-1})$ with $L_{l,0} = 0$ for $l \geqslant 1$, $L_{i,i-1} = D_K g_0$, and $L_{i+1,j} = L_{i,j-1} + D_K g_{i-j} - D_K L_{i,j}$ such that

$$\mathscr{H}_K^{(i)} = g_0 h_K^{(i)} + \sum_{j=0}^{i-1} (g_{i-j} - L_{i,j}) h_K^{(j)}, \quad i = 0,1,2,\ldots. \tag{2.29}$$

Any such sequence $\{g_l\}$ together with $\{h_K^{(l)}\}$ determines a new sequence of solutions $\{\mathscr{H}_K^{(l)}\}$.

Let $\{H_K^{(l)}\}$ be the solution sequence with property (2.25). Then setting $h_K^{(i)} = H_K^{(i)}$ in (2.28) choosing $g_0 = 1, g_1, \ldots, g_{m_K - 1}$ recursively such that

$$-f_{(j)} = g_{m_K - j} - L_{m_{K,j}}, \quad j = 0,1,\ldots,m_K - 1,$$

we have $\mathscr{H}_K^{(m_K)} = 0$. Thus there is a solution sequence $\{\mathscr{H}_K^{(l)}\}$ with $\mathscr{H}_K^{(0)}, \ldots, \mathscr{H}_K^{(m_K - 1)}$ nonzero and all further terms zero. By Lemma 3, all other solution sequences are linear combinations of these m_K nonzero terms.

Lemma 4: The integer m_K, if it exists, is unique.

In particular, modifying H_K, to $g_K H_K$ with g_K invertible and $D_I g_K = 0$ for $I \neq K$ does not change m_K.

Based on the preceding results, given any solution $\{H_{(K,k)}\}$ of Eq. (2.11) we can determine the integers m_K such that $1 \leqslant m_K \leqslant n_K$. Then there is another solution $\{\mathscr{H}_{(K,k)}\}$ with $m = \Sigma_{K=1}^N m_K$ nonzero terms such that each $H_{(K,k)}$ is a linear combination of the $\mathscr{H}_{(K,k)}$. Thus the original solution is associated with an $m \times m$ generalized D-Stäckel matrix.

III. SEPARABILITY CONDITIONS

Suppose we are given a partial differential equation

$$H(x_I, u, u_{I,i}) = E, \tag{3.1}$$

which admits regular additive separability in the coordinates x_I. (Unless otherwise specified we will adhere to the notation and conventions for separation listed in the Introduction.) That is, suppose the integrability equations (1.3) are satisfied identically in u, $u_{I,i}$. What is the form of the separation and how can the separation equations be determined from (1.3)? In this section we will identify some classes of linear and nonlinear differential equations where the separation is achieved via generalized D-Stäckel matrices.

Our method of approach is exemplified by the following observation concerning (3.1).

Lemma 5: Suppose $\partial_u H = \partial_{u_{I,i}} \partial_{u_{J,j}} H = 0$ for all $I \neq J$ and $\partial_{u_{J,n_J}} H \equiv H_J(x_K, u_{K,k})$ is invertible for $1 \leqslant K \leqslant N$, $1 \leqslant k \leqslant n_K - 1$. Further suppose the $\{H_J\}$ generate a D-Stäckel matrix via the process (2.22). Then the differential equation $H = E$ is regular separable if and only if H is a generalized D-Stäckel multiplier.

Proof: The integrability conditions (1.3) for H are, in this case, equivalent to

$$(D_I D_J - D_I \ln H_J D_J - D_J \ln H_I D_I) H = 0, \quad I \neq J,$$

the condition that H be a D-Stäckel multipler.

Theorem 3: Suppose H takes the form

$$H = \sum_{J=1}^N H_J(x_K, u_{K,k}) \mathscr{P}_J(x_J, u_{J,n_J}, u_{J,j}) + V(x_K, u_{K,k}),$$
$$1 \leqslant K \leqslant N, \quad 1 \leqslant k \leqslant n_K - 1, \tag{3.2}$$

where H_J is invertible, $D_I \mathscr{P}_J = 0$ for $I \neq J$, and

$\partial^2_{u_{J,n_J}} \mathscr{P}_J \neq 0$. Then the equation $H = E$ is regular separable if and only if the $\{H_J\}$ are in generalized Stäckel form ($m_J = 1$) and H is a generalized Stäckel multiplier with respect to this form.

Proof: In this case the integrability conditions (1.3) are equivalent to

$$A_{IJ} H_P = 0, \quad I \neq J, \quad 1 \leqslant I, J, P \leqslant N, \quad A_{IJ} H = 0,$$

where A_{IJ} is defined by (2.21). Q.E.D.

Theorem 4: Suppose H takes the form

$$H = \sum_{J=1}^{N} H_J(x_K) \mathscr{P}_J(x_J, u_{J,j}) u_{J,n_J} + V(x_K, u_{K,k}),$$
$$1 \leqslant K \leqslant N, \quad 1 \leqslant k \leqslant n_K - 1, \quad 1 \leqslant j \leqslant n_J - 1, \tag{3.3}$$

where $H_J \mathscr{P}_J$ is invertible. Then the quasilinear equation $H = E$ is regular separable if and only if the functions H_J determine an (ordinary) D-Stäckel matrix via the process (2.22) and H is a D-Stäckel multiplier with respect to this form.

Proof: The integrability conditions (1.3) for $H = E$ are

$$\begin{aligned}
&(1) \quad \hat{A}_{IJ} H_P = 0, \quad P \neq I, J, \\
&(2) \quad \hat{A}_{IJ}(H_I \mathscr{P}_I) = \hat{B}_{IJ}(\partial_{u_{I,n_I-1}} V), \\
&(3) \quad \hat{A}_{IJ} V = 0, \\
&(4) \quad \partial_{u_{I,n_I-1}} \partial_{u_{J,n_J-1}} V = 0,
\end{aligned} \tag{3.4}$$

for all $I \neq J$, where

$$\begin{aligned}
\hat{A}_{IJ} &= \hat{D}_I \hat{D}_J - \hat{D}_J \ln H_I \hat{D}_I - \hat{D}_I \ln H_J \hat{D}_J, \\
\hat{B}_{IJ} &= -\hat{D}_I + \hat{D}_I \ln H_J, \\
\hat{D}_I &= \partial_{x_I} + \sum_{i=1}^{n_I-2} u_{I,i+1} \partial_{u_{I,i}}.
\end{aligned} \tag{3.5}$$

Note that

$$D_I \ln(H_J \mathscr{P}_J) = D_I \ln H_J = \partial_{x_I} \ln H_J, \tag{3.6}$$

for $I \neq J$. Using (4) and differentiating (2) with respect to u_{J,n_J-1}, $u_{J,n_J-2}, \ldots, u_{J,1}$, recursively, we obtain $\partial_{u_{I,n_I-1}} \partial_{u_{J,j}} V = 0$. Then differentiating (3) with respect to $u_{I,i}$ and $u_{J,j}$ recursively, we obtain $\partial_{u_{I,i}} \partial_{u_{J,j}} V = 0, I \neq J$. Thus we can write V uniquely in the form

$$V = V_0(x_K) + \sum_{J=1}^{N} V_J(x_K, u_{J,j}), \quad 1 \leqslant K \leqslant N, \quad 1 \leqslant j \leqslant n_J - 1, \tag{3.7}$$

where $V_J(x_K, 0) = 0$, and (2) becomes

$$\hat{A}_{IJ}(H_I \mathscr{P}_I) = \hat{B}_{IJ}(\partial_{u_{I,n_I-1}} V_I). \tag{3.8}$$

Differentiating (3) with respect to $u_{I,i}$ we find

$$\hat{A}_{IJ}(\partial_{u_{I,i}} V_I) = \hat{B}_{IJ}(\partial_{u_{I,i-1}}) V_I, \tag{3.9}$$

where the right-hand side of (3.9) vanishes for $i = 1$. Then, using (2) we can verify the formulas

$$A_{IJ}(H_P \mathscr{P}_P) = 0, \quad P \neq I, J,$$
$$A_{IJ}(H_I \mathscr{P}_I) = B_{IJ}(\partial_{u_{I,n_I-1}} V_I + H_I \partial_{u_{I,n_I-1}} \mathscr{P}_I)$$
$$A_{IJ}(\partial_{u_{I,i}} V_I + H_I \partial_{u_{I,i}} \mathscr{P}_I) = B_{IJ}(\partial_{u_{I,i-1}} V_I + H_I \partial_{u_{I,i-1}} \mathscr{P}_I),$$
$$1 \leqslant i \leqslant n_I - 1,$$

where $\partial_{u_{I,0}} V_I - H_I \partial_{u_{I,0}} \mathscr{P}_I \equiv 0$. Here the truncated derivatives $\hat{D}_J$ have been replaced by total derivatives D_J. These formulas agree with (2.11a) and (2.11b) for $\mathscr{H}_I = H_I \mathscr{P}_I$, $\mathscr{H}_{(I,i)} = \partial_{u_{I,i}} V_I + H_I \partial_{u_{I,i}} \mathscr{P}_I$, $1 \leqslant i \leqslant n_I - 1$. It follows from (2.6) and Lemma 3 that the H_I also generate solutions of (2.11) that are independent of the $u_{K,k}$, since $\mathscr{P}_I$ is invertible. Hence the integrability conditions (1.3) imply that $H_{(I,i)}$ determine a D-Stäckel matrix and that H is a D-Stäckel multiplier with respect to this form. Q.E.D.

Due to the property (3.6) it is not necessary to assume in Theorems 3 and 4 that the functions $\mathscr{P}_I(x_I, u_{I,i})$ are invertile in the strong sense of the Introduction; they may be permitted to vanish for $u_{I,i} = 0$.

Corollary 1: If in (3.3) we have $\mathscr{P}_I \neq 0$ for all I then the functions $H_I(x_K)$ determine an (ordinary) D-Stäckel matrix and H is an (ordinary) D-Stäckel multiplier with respect to this form.

This result follows from Lemma 3.

Corollary 2: Consider the differential equation $L\psi = E\psi$, where L is the linear nth-order partial differential operator ($n > 1$)

$$L = \sum_{J=1}^{N} H_J(x_K) \partial^n_{x_J} + \sum_{\substack{a_1 + \cdots + a_N < n \\ a_I \geqslant 0}} H_{a_1, \ldots, a_N}(x_K) \cdot \partial^{a_1}_{x_1}, \ldots, \partial^{a_N}_{x_N}, \tag{3.10}$$

with $H_J \neq 0$ for each J. This equation admits regular *multiplicative* separation in the coordinates x_K if and only if

$$L = \sum_{J=1}^{N} H_J(x_K) \left(\partial^n_{x_J} + \sum_{a=0}^{N-1} f^a_J(x_J) \partial^a_{x_J} \right), \tag{3.11}$$

where $\partial_{x_I} f^a_J = 0$ for $I \neq J$ and the $\{H_p\}$ are in (ordinary) Stäckel form.

Proof: The equation $L\psi = E\psi$ admits regular multiplicative separation (by definition) provided the equation $H = E$, obtained by setting $\psi = e^u$ in $L\psi/\psi = E$, admits regular additive separation:

$$H = \sum_{J=1}^{N} H_J(x_K) u_{J,n} + V(x_K, u_{K,k}), \quad 1 \leqslant k < n. \tag{3.12}$$

Here V is an nth-order polynomial in the derivatives $u_{K,k}$ whose nth-order terms take the form

$$\sum_{J=1}^{N} H_J(x_K) u^n_{J,1}.$$

Equating coefficients of $u^n_{P,1}$ on both sides of the integrability conditions $A_{IJ} H = 0$, we find $A_{IJ} H_P = 0$ for all P. Thus, $\{H_P\}$ is in (ordinary) Stäckel form. Furthermore, from (2.7) we see that there can be no cross terms in the potential V. This means that in (3.10) we can require $H_{a_1, \ldots, a_N} = 0$ if more than one a_I is nonzero. Since V must be a Stäckel multiplier with respect to the Stäckel form $\{H_P\}$ we obtain (3.11). Q.E.D.

Having brought up multiplicative separation of linear eigenvalue equations we might as well mention the additive separation case.

Proposition: The equation $Lu = Eu$, where

$$Lu = \sum_{J=1}^{N} \sum_{j=1}^{n_J} H_{(J,j)}(x_K) u_{J,j} + H_{(0)} u$$

and $H_J \equiv H_{(J, n_J)} \neq 0$ admits regular additive separation in the coordinates x_K if and only if $\partial_{x_I} H_{(J,j)} = 0$ for $I \neq J$ and $\partial_{x_I} H_{(0)} = 0$ for all I.

The proof is a straightforward application of the integrability conditions to $H = Lu/u$. Although additive separation is not very interesting for $Lu = Eu$, in the case of homogeneous equations $Lu = 0$ nontrivial D-Stäckel additive separation occurs even in coordinate systems for which there is no multiplicative separation.[3]

ACKNOWLEDGMENT

One of us (W.M.) was partially supported by National Science Foundation Grant No. MCS82-19847.

[1]E. G. Kalnins and W. Miller, "Intrinsic characterization of variable separation for the partial differential equations of mechanics," in *Proceedings of IUTAM-ISIMM Symposium on Modern Development in Analytical Mechanics*, Turin 1982, in Atti. Accad. Sci. Torino, Turin, Suppl. **117**, 511 (1983).

[2]J. Levi-Civita, "Sulla integrazione della equazione di Hamilton–Jacobi per separazione di variabili," Math. Ann. **59**, 383 (1904).

[3]E. G. Kalnins and W. Miller, "Differential Stäckel Matrices," J. Math. Phys. **26**, 1560 (1985).

[4]S. Benenti, "Separability structures on Riemannian manifolds," in *Proceedings of the Conference on Differential Geometrical Methods in Mathematical Physics, Salamanca, 1979, Lecture Notes in Mathematics,* Vol. 836 (Springer, Berlin, 1980).

[5]P. Stäckel, "Über die integration der Hamilton Jacobischen differentialgeleichung mittels separation der variabelen," Habilitationsschrift (thesis), Halle, 1891.

[6]L. P. Eisenhart, "Separable systems of Stäckel," Ann. Math. **35**, 284 (1934).

[7]J. W. Koornwinder, "A precise definition of separation of variables," in *Proceedings of the Scheveningen Conference of Differential Equations, August, 1979, Lecture Notes in Mathematics,* Vol. 810 (Springer, Berlin, 1980).

Modified equations, rational solutions, and the Painlevé property for the Kadomtsev–Petviashvili and Hirota–Satsuma equations

John Weiss

Center for Studies of Nonlinear Dynamics, La Jolla Institute, 8950 Villa La Jolla Drive, Suite 2150, La Jolla, California 92037 and Institute for Pure and Applied Physical Science,[a] University of California, San Diego, La Jolla, California 92093

(Received 31 October 1984; accepted for publication 22 February 1985)

We propose a method for finding the Lax pairs and rational solutions of integrable partial differential equations. That is, when an equation possesses the Painlevé property, a Bäcklund transformation is defined in terms of an expansion about the singular manifold. This Bäcklund transformation obtains (1) a type of modified equation that is formulated in terms of Schwarzian derivatives and (2) a Miura transformation from the modified to the original equation. By linearizing the (Ricati-type) Miura transformation the Lax pair is found. On the other hand, consideration of the (distinct) Bäcklund transformations of the modified equations provides a method for the iterative construction of rational solutions. This also obtains the Lax pairs for the modified equations. In this paper we apply this method to the Kadomtsev–Petviashvili equation and the Hirota–Satsuma equations.

I. INTRODUCTION

In Ref. 1 we have formulated a procedure for calculating the Lax pair and rational solutions of partial differential equations that possess the Painlevé property. That is, for an equation with the Painlevé property, a Bäcklund transformation is defined in terms of an expansion about the "singular manifold." This Bäcklund transformation obtains (1) a type of "modified equation" that can be expressed in terms of Schwarzian derivatives and (2) a Miura transformation from the modified to the original equation. By linearizing the Ricati-type Miura transformation (and the modified equations), the Lax pair is found. Then, further consideration of the Bäcklund transformations for the modified equations provides a method for the iterative construction of "rational" solutions, and finds the Lax pair for the modified equations as well.

We recall that the partial differential equation is said to possess the Painlevé property[2–7] when the solutions of the partial differential equation (pde) are "single valued" about the movable, singularity manifold and the singularity manifold is "noncharacteristic." To be precise, if the singularity manifold is determined by

$$\varphi(z_1,z_2,\ldots,z_n)=0, \tag{1.1}$$

and $u=u(z_1,\ldots,z_n)$ is a solution of the pde, then we require that

$$u=\varphi^{\alpha}\sum_{j=0}^{\infty}u_j\varphi^{j}, \tag{1.2}$$

where $u_0\neq 0$, $\varphi=\varphi(z_1,\ldots,z_n)$, and $u_j=u_j(z_1,\ldots,z_n)$ are analytic functions of (z_j) in a neighborhood of the manifold (1.1) and α (the leading-order exponent) is a (negative) rational number. The requirement that the manifold (1.1) be noncharacteristic (for the pde) insures that the expansion (1.2) will be well defined, in the sense of the Cauchy–Kovalevskaya theorem.[8] Substitution of (1.2) into the pde determines that value(s) of α, and defines the recursion relations for u_j, $j=0,1,2,\ldots$. When the expansion (1.2) is well defined and contains the maximum number of arbitrary functions allowed at the "resonances,"[2,9,10] the pde is said to possess the Painlevé property and is conjectured to be integrable. Informally, the resonances are the values of j for which the u_j are not "fixed" by the recursion relations (i.e., are arbitrary).

The Bäcklund transformation is defined by truncating the expansion (1.2) at the constant level term. That is, we set

$$u=u_0\varphi^{-n}+u_1\varphi^{-n+1}+\cdots+u_n, \tag{1.3}$$

and find, from the recursion relations for u_j and the condition that u_j vanish for $j>n$, a system of equations for (φ, u_j, $j=0,1,\ldots,n$), where u_n will satisfy the (original) pde. This system of equations will, in general (depending on the values of the resonances), be overdetermined. Upon solving this system, it is found, for those equations considered, the φ satisfies an equation formulated in terms of Schwarzian derivatives[3]:

$$\{\varphi;x\}=\frac{\partial}{\partial x}\left(\frac{\varphi_{xx}}{\varphi_x}\right)-\frac{1}{2}\left(\frac{\varphi_{xx}}{\varphi_x}\right)^2. \tag{1.4}$$

This equation, or system of equations, we regard as a type of modified equation. By the invariance of (1.4) under the Moebius group,

$$\varphi=(a\psi+b)/(c\psi+d),\quad \{\varphi;x\}=\{\psi;x\}, \tag{1.5}$$

the "modified" equations allow the Bäcklund transformation (1.5).

The above procedure may now be reapplied to the "modified" (or equivalent) equations to find different forms of Bäcklund transformations. These Bäcklund transformations may take the form of discrete symmetries,[1,5,6] reductions,[1] or, as we shall see, more complicated structures. The group of Bäcklund transformations for the modified equations may be conveniently employed to iteratively construct sequences of rational solutions. Also, by linearizing the Miura transformation from modified to original equation we propose to calculate the Lax pair.[1,6]

In this paper we consider the Kadomtsev–Petviashvili

[a] Permanent address.

 0022-2488/85/092174-07$02.50

(KP) equation and the Hirota–Satsuma equations. The modified equations are derived, their (modified) Bäcklund transformations are calculated, and the sequences of rational solutions are found.

II. THE KADOMTSEV–PETVIASHVILI EQUATION

The Kadomtsev–Petviashvili equation

$$U_{yy} + \frac{\partial}{\partial x}(U_t + UU_x + U_{xxx}) = 0 \tag{2.1}$$

possesses the Painlevé property.[2] The expansion about the singular manifold ($\varphi = 0$) is

$$U = \varphi^{-2} \sum_{j=0}^{\infty} U_j \varphi^j, \tag{2.2}$$

with resonances at

$$j = -1,4,5,6. \tag{2.3}$$

Therefore, subject to the "noncharacteristic" condition ($\varphi_x \neq 0$ when $\varphi = 0$), $\{\varphi, U_4, U_5, U_6\}$ are arbitrary functions of (x,t) in the expansion (2.2).

The Bäcklund transformation[2,3] for Eq. (2.1) is

$$U = U_0\varphi^{-2} + U_1\varphi^{-1} + U_2, \tag{2.4}$$

which obtains

$$\begin{aligned} &U_0 = -12\varphi_x^2, \quad U_1 = 12\varphi_{xx}, \\ &U_2 + \frac{\varphi_t}{\varphi_x} + 4\frac{\varphi_{xxx}}{\varphi_x} - 3\frac{\varphi_{xx}^2}{\varphi_x^2} + \frac{\varphi_y^2}{\varphi_x^2} = 0, \\ &\varphi_{xt} + \varphi_{xxxx} + \varphi_{yy} + \varphi_{xx}u_2 = 0. \end{aligned} \tag{2.5}$$

We note that the system (2.5) is not overdetermined since (U_4, U_5, U_6) may vanish without restriction. From (2.4) and (2.5) it is found that

$$U = 12\frac{\partial^2}{\partial x^2}\ln\varphi + U_2, \tag{2.6}$$

$$U_2 + \frac{\varphi_t}{\varphi_x} + 4\frac{\varphi_{xxx}}{\varphi_x} - 3\frac{\varphi_{xx}^2}{\varphi_x^2} + \frac{\varphi_y^2}{\varphi_x^2} = 0, \tag{2.7}$$

and

$$\frac{\partial}{\partial y}\left(\frac{\varphi_y}{\varphi_x}\right) + \frac{\partial}{\partial x}\left(\frac{\varphi_t}{\varphi_x} + \{\varphi;x\} + \frac{1}{2}\frac{\varphi_y^2}{\varphi_x^2}\right) = 0, \tag{2.8}$$

where

$$\{\varphi;x\} = \frac{\partial}{\partial x}\left(\frac{\varphi_{xx}}{\varphi_x}\right) - \frac{1}{2}\left(\frac{\varphi_{xx}}{\varphi_x}\right)^2. \tag{2.9}$$

In terms of our procedure, Eq. (2.8) is the "modified" equation formulated in terms of the Schwarzian derivative (2.9) and Eq. (2.7) is a "Miura" transformation from Eq. (2.8) to Eq. (2.1). Equation (2.8) is invariant under the Moebius group

$$\varphi = (a\psi + b)/(c\psi + d), \tag{2.10}$$

where $ad - bc \neq 0$.

To investigate the group of Bäcklund transformations for Eq. (2.8) it is convenient to study various forms of "modified" equations that are equivalent to (2.8). To begin we let

$$V = \varphi_{xx}/\varphi_x, \quad W = \varphi_y/\varphi_x, \quad Z = \varphi_t/\varphi_x, \tag{2.11}$$

and find, from (2.8) and (2.11), the system of modified equations

$$\begin{aligned} &W_y + \frac{\partial}{\partial x}\left(Z + \frac{W^2}{2} + V_x - \frac{1}{2}V^2\right) = 0, \\ &V_y = \frac{\partial}{\partial x}(W_x + VW), \quad V_t = \frac{\partial}{\partial x}(Z_x + VZ), \end{aligned} \tag{2.12}$$

where

$$W_t + WZ_x = Z_y + ZW_x. \tag{2.13}$$

Equations (2.12) and (2.13) are overdetermined. Equation (2.13) arises from the condition $V_{yt} = V_{ty}$. This system allows singularities of the form

$$V \sim V_0\epsilon^\alpha, \quad W \sim \omega_0\epsilon^\beta, \quad Z \sim Z_0\epsilon^\gamma, \tag{2.14}$$

where

$$\text{(i)} \quad \alpha = -1, \quad \beta = \gamma = 0, \quad V_0 = 0, -2\epsilon_x; \tag{2.15}$$

$$(ii) \quad \alpha = \beta = \gamma = -1, \quad V_0 = \epsilon_x, \quad \omega_0^2 = 3\epsilon_x^2; \tag{2.16}$$

$$\begin{aligned} (iii) \quad &\alpha = -1, \quad \beta = -2, \quad \gamma = 0, \\ &V_0 = 2\epsilon_x, \quad \omega_0 = \epsilon_y, \quad Z_0 = 4\epsilon_x. \end{aligned} \tag{2.17}$$

For (2.16) and (2.17) the resonances are

$$j = -1,0,2,2,2,3 \tag{2.18}$$

and

$$j = -2,-1,1,2,3,4, \tag{2.19}$$

respectively. The expansion about the singularity (2.16) contains the arbitrary functions $(\epsilon, Z_0, V_2, W_2, Z_2, \{V_3, W_3\})$. The Bäcklund transformation is

$$\begin{aligned} &V = V_0\epsilon^{-1} + V_1, \quad W = W_0\epsilon^{-1} + W_1, \\ &Z = Z_0\epsilon^{-1} + Z_1, \end{aligned} \tag{2.20}$$

where

$$V_0 = \epsilon_x, \quad W_0 = a\epsilon_x, \quad a^2 = 3, \quad Z_0 = H\epsilon_x, \tag{2.21}$$

$$\begin{aligned} &V_1 = \frac{1}{2}\left(-\frac{\epsilon_{xx}}{\epsilon_x} + a\frac{\epsilon_y}{\epsilon_x} + \frac{H}{2}\right), \\ &W_1 = \frac{1}{2}\left(-a\frac{\epsilon_{xx}}{\epsilon_x} - \frac{\epsilon_y}{\epsilon_x} - \frac{aH}{2}\right), \end{aligned} \tag{2.22}$$

$$\begin{aligned} &Z_1 = \frac{\epsilon_t}{\epsilon_x} - H_x - \frac{H^2}{4} - \frac{1}{2}\frac{\epsilon_{xx}}{\epsilon_x}H - \frac{a}{2}\frac{\epsilon_y}{\epsilon_x}H, \\ &H_y + \frac{\partial}{\partial x}\left(a\left(H_x + \frac{H^2}{4}\right) + 2\frac{\epsilon_y}{\epsilon_x}H - 2a\frac{\epsilon_t}{\epsilon_x}\right) = 0, \end{aligned} \tag{2.23}$$

and

$$\frac{\partial}{\partial y}\left(\frac{\epsilon_y}{\epsilon_x}\right) + \frac{\partial}{\partial x}\left(\frac{\epsilon_t}{\epsilon_x} + \{\epsilon;x\} + \frac{1}{2}\left(\frac{\epsilon_y}{\epsilon_x}\right)^2\right) = 0. \tag{2.24}$$

To simplify the above let

$$V = \epsilon_{xx}/\epsilon_x, \quad W = \epsilon_y/\epsilon_x, \quad Z = \epsilon_t/\epsilon_x, \tag{2.25}$$

and find

$$\begin{aligned} &\text{(i)} \quad a(V + V_1) = W - W_1, \\ &\text{(ii)} \quad V_1 - V = H + a(W + W_1), \\ &\text{(iii)} \quad Z - Z_1 = H_x + (V + V_1)H, \end{aligned} \tag{2.26}$$

where the auxiliary function H satisfies (2.23). Now (2.26) constitutes a somewhat awkward Bäcklund transformation (BT) for Eqs. (2.12) and (2.13), which can be simplified by identifying (2.12) with (2.8) through

$$\begin{aligned} &V=\psi_{xx}/\psi_x\,,\quad V_1=\varphi_{xx}/\varphi_x\,,\\ &W=\psi_y/\psi_x\,,\quad W_1=\varphi_y/\varphi_x\,,\\ &Z=\psi_t/\psi_x\,,\quad Z_1=\varphi_t/\varphi_x\,. \end{aligned} \tag{2.27}$$

Thus, after simplification, eliminating H in Eq. (2.26) obtains the BT

$$\psi_y=a\psi_{xx}+A\psi_x\,,\quad \psi_t=-4\psi_{xxx}-2aA\psi_{xx}+B\psi_x\,, \tag{2.28}$$

where (ψ,φ) satisfy (2.8) and

$$\begin{aligned} &a^2=3\,,\quad A=\varphi_y/\varphi_x+a(\varphi_{xx}/\varphi_x)\,,\\ &B=\frac{\varphi_t}{\varphi_x}-2\left(\frac{\varphi_{xxx}}{\varphi_x}+a\frac{\varphi_{xy}}{\varphi_x}\right). \end{aligned} \tag{2.29}$$

That (ψ,φ) satisfy (2.8) is found from the conditions $(\varphi_{yt}=\varphi_{ty},\,\psi_{yt}=\psi_{ty})$, respectively. Having found (2.28) we discontinue consideration of the system (2.12) and instead consider the system in (A,B) obtained from (2.28) by the condition, $\psi_{yt}=\psi_{ty}$,

$$\begin{aligned} &A_y+\frac{\partial}{\partial x}\left(aA_x+\frac{1}{2}A^2+B\right)=0\,,\\ &B_y+BA_x-4A_{xxx}-2aAA_{xx}=aB_{xx}+AB_x+A_t\,. \end{aligned} \tag{2.30}$$

Now, the expression

$$\Omega=B+A^2=\frac{\varphi_t}{\varphi_x}-2\{\varphi;x\}-2a\frac{\partial}{\partial x}\left(\frac{\varphi_y}{\varphi_x}\right)+\left(\frac{\varphi_y}{\varphi_x}\right)^2 \tag{2.31}$$

is invariant under (2.10). This suggests defining the system in (A,Ω):

$$\begin{aligned} &A_y+\frac{\partial}{\partial x}\left(aA_x-\frac{A^2}{2}+\Omega\right)=0\,,\\ &\Omega_y-A_t=\frac{\partial}{\partial x}\left(4A_{xx}+\frac{A^3}{3}-2aAA_x+a\Omega_x-A\Omega\right). \end{aligned} \tag{2.32}$$

Note that Eqs. (2.30) or (3.32) are "properly posed" in comparison to Eqs. (2.12) in that they are not overdetermined and have the same order as Eq. (2.1) or Eq. (2.8). From the Miura transformation (2.7) and the above,

$$\begin{aligned} -U_2&=\frac{\varphi_t}{\varphi_x}+4\frac{\varphi_{xxx}}{\varphi_x}-3\frac{\varphi_{xx}^2}{\varphi_x^2}+\frac{\varphi_y^2}{\varphi_x^2}\\ &=B+2aA_x+A^2=\Omega+2aA_x\,. \end{aligned} \tag{2.33}$$

By construction (2.28) constitute a Lax pair for Eqs. (2.30). The Lax pair for (2.32) is found by substituting for B in (2.30) using (2.31).

Equations (2.32) allow two Bäcklund transformations:

$$\text{(i)}\quad A=2a\frac{\psi_x}{\psi}+A_1\,,\quad \Omega=-12\frac{\partial^2}{\partial x^2}\ln\psi+\Omega_2\,, \tag{2.34}$$

where

$$A_1=\frac{\psi_y}{\psi_x}-a\frac{\psi_{xx}}{\psi_x}\,,\quad \Omega_2=\frac{\psi_t}{\psi_x}+\frac{\psi_y^2}{\psi_x^2}+4\frac{\psi_{xxx}}{\psi_x}-3\frac{\psi_{xx}^2}{\psi_x^2}\,, \tag{2.35}$$

and ψ satisfies (2.8).

$$\text{(ii)}\quad A=-2a(\varphi_x/\varphi)+A_1\,,\quad \Omega=\Omega_2\,, \tag{2.36}$$

where φ satisfies (2.8) and

$$\begin{aligned} &A_1=\frac{\varphi_y}{\varphi_x}+a\frac{\varphi_{xx}}{\varphi_x}\,,\\ &\Omega=\frac{\varphi_t}{\varphi_x}-2\{\varphi;x\}-2a\frac{\partial}{\partial x}\left(\frac{\varphi_y}{\varphi_x}\right)+\left(\frac{\varphi_y}{\varphi_x}\right)^2. \end{aligned} \tag{2.37}$$

Equations (2.35) reobtain the Lax pair for (2.32), where the identification (2.33),

$$-U_2=\Omega_2\,. \tag{2.38}$$

Equations (2.37) provide (2.29) and, with (2.35), obtain (2.28).

Now consider the BT/Lax pair (2.28). We have the following.

Lemma: For fixed (A,B) let (ψ_1,ψ_2) be two linearly independent solutions of Eqs. (2.28), then

$$\psi=\psi_{1x}/\psi_{2x} \tag{2.39}$$

will satisfy Eqs. (2.28) with

$$A\to A'\,,\quad B\to B'\,,$$

where $(A,B),(A',B')$ satisfy Eqs. (2.30) and

$$\begin{aligned} A'&=A+2a(\psi_{2xx}/\psi_{2x})\,,\\ B'&=B-2aA'_x-A'^2+A^2\\ &=B-2aA_x-\frac{\psi_{2xx}}{\psi_{2x}}A-12\frac{\psi_{2xxx}}{\psi_{2x}}\,. \end{aligned} \tag{2.40}$$

Proof: By direct calculation, which we omit.

To investigate the iterative application of the BT (2.28) we define a double sequence

$$\begin{aligned} &\varphi_{j+1,y}=a\varphi_{j+1,xx}+A_j\varphi_{j+1,x}\,,\\ &\psi_{j+1,y}=a\psi_{j+1,xx}+A_j\psi_{j+1,x}\,, \end{aligned} \tag{2.41}$$

$$\begin{aligned} &\varphi_{j+1,t}=-4\varphi_{j+1,xxx}-2aA_j\varphi_{j+1,xx}+B_j\varphi_{j+1,x}\,,\\ &\psi_{j+1,t}=-4\psi_{j+1,xx}-2aA_j\psi_{j+1,xx}+B_j\psi_{j+1,x}\,, \end{aligned} \tag{2.42}$$

where

$$A_j=\frac{\varphi_{j,y}}{\varphi_{j,x}}+a\frac{\varphi_{j,xx}}{\varphi_{j,x}}=A_{j-1}+2a\frac{\varphi_{j,xx}}{\varphi_{j,x}}\,, \tag{2.43}$$

$$B_j=\frac{\varphi_{j,t}}{\varphi_{j,x}}-2\left(\frac{\varphi_{j,xx}}{\varphi_{j,x}}+a\frac{\varphi_{j,xx}}{\varphi_{j,x}}\right), \tag{2.44}$$

$$B_j=B_{j-1}-2aA_{j,x}-A_j^2+A_{j-1}^2\,. \tag{2.45}$$

Then, by the lemma, it is consistent to set

$$\varphi_{j+1}=\psi_{j,x}/\varphi_{j,x}\,. \tag{2.46}$$

Now let

$$A_0=B_0=0 \tag{2.47}$$

or

$$\varphi_{1y} = a\varphi_{1xx}\,, \quad \varphi_{1t} = -4\varphi_{1xxx}\,. \tag{2.48}$$

By (2.43) and (2.45),

$$A_j = 2a\sum_{k=1}^{j}\frac{\varphi_{k,xx}}{\varphi_{k,x}}\,, \quad B_j = -2a\frac{\partial}{\partial x}\sum_{l=1}^{j}A_l - A_j^2\,, \tag{2.49}$$

and by (2.33), for $j>1$,

$$\begin{aligned} U_{2j} &= -B_j - 2aA_{j,x} - A_j^2 = 2a\frac{\partial}{\partial x}\sum_{l=1}^{j-1}A_l\,,\\ U_{2j} &= 12\frac{\partial^2}{\partial x^2}\ln\prod_{1}^{j-1}\varphi_{l,x}^{(j-l)}\,, \end{aligned} \tag{2.50}$$

$$U_j = 12\frac{\partial^2}{\partial x^2}\ln\varphi_j + U_{2_j} = 12\frac{\partial^2}{\partial x^2}\ln\left\{\psi_{j-1,x}\prod_{1}^{j-2}\varphi_{l,x}^{(j-l)}\right\}. \tag{2.51}$$

In effect, to iterate (2.28) two solutions are "interpolated" to produce one new solution at each step. From N linearly independent solutions at one level, fixing one solution as the denominator in (2.46) produces $N-1$ solutions of (2.41) and (2.42) at the next level. However, from the linearity of (2.48), it is possible to generate an infinity of linearly independent solutions. For instance, to find rational solutions, let

$$\varphi_y = a\varphi_{xx}\,, \quad \varphi_t = -4\varphi_{xxx}\,, \tag{2.52}$$

and

$$\varphi^{(n)} = \sum_{j=0}^{n} b_j x^j\,, \tag{2.53}$$

where

$$\begin{aligned} &b_{j-2,y} = aj(j-1)b_j\,, \quad b_{j-3,t} = -4j(j-1)(j-2)b_j\,,\\ &ab_{j-1,t} = -4jb_{j,y}\,. \end{aligned} \tag{2.54}$$

This obtains

$$\begin{aligned} &\varphi_1^0 = 1\,, \quad \varphi_1^1 = x\,, \quad \varphi_1^2 = x^2 + 2ay\,,\\ &\varphi_1^3 = x^3 + 6ayx - 24t\,,\\ &\varphi_1^4 = x^4 + 12ayx^2 - 96tx + 36y^2\,,\\ &\varphi_1^5 = x^5 + 20ayx^3 - 960tx^2 + 180y^2x - 288ayt\,. \end{aligned} \tag{2.55}$$

Using the identity

$$j\varphi_x^{(j)} = \varphi^{(j-1)}\,, \tag{2.56}$$

and letting

$$\varphi_1 = \varphi_1^5\,,$$

a first application of (2.41), (2.42), and (2.46) finds the appropriate set of solutions at the next level. That is,

$$\begin{aligned} &\varphi_2^1 = 1/\varphi_1^4\,, \quad \varphi_2^2 = \varphi_1^1/\varphi_1^4\,,\\ &\varphi_2^3 = \varphi_1^2/\varphi_1^4\,, \quad \varphi_2^4 = \varphi_1^3/\varphi_1^4\,. \end{aligned} \tag{2.57}$$

Then, with

$$\varphi_2 = \varphi_2^1 = 1/\varphi_1^4\,, \tag{2.58}$$

Eq. (2.46) obtains

$$\begin{aligned} &\varphi_3^1 = 4\varphi_1^1 - \varphi_1^4/\varphi_1^3\,, \quad \varphi_3^2 = 4\varphi_1^2 - 2\varphi_1^1\varphi_1^4/\varphi_1^3\,,\\ &\varphi_3^3 = 4\varphi_1^3 - 3\varphi_1^2\varphi_1^4/\varphi_1^3\,, \end{aligned} \tag{2.59}$$

etc.

An identical procedure can be applied to *any* linearly independent set of solutions of Eqs. (2.52).

Finally, the KP equation (2.1) is invariant under the Gallilean transformation

$$x' = x + \alpha y + \beta t\,, \quad y' = y - 2at\,, \tag{2.60}$$

$$t' = t\,, \quad u' = u - \alpha^2 - \beta\,. \tag{2.61}$$

Also, Eqs. (2.8), (2.30), and (2.32) are invariant under (2.60), where

$$A\to A+\alpha\,, \quad B\to B - 2\alpha A + \beta\,, \quad \Omega\to\Omega + \alpha^2 + \beta\,, \tag{2.62}$$

which is consistent with their definitions (2.29) and (2.31). Obviously, (2.60) preserves the form of the rational solutions and may be applied directly to, say, (2.55)–(2.59).

III. THE HIROTA–SATSUMA EQUATIONS

The Hirota–Satsuma equations

$$u_t = \tfrac{1}{2}u_{xxx} + 3uu_x - 6\omega\omega_x\,, \quad \omega_t = -\omega_{xxx} - 3u\omega_x \tag{3.1}$$

have the Painlevé property[6] about singularities of the form

$$\text{(i)}\quad u = \psi^{-2}\sum_{j=0}^{\infty}u_j\psi^j\,, \quad \omega = \psi^{-1}\sum_{j=0}^{\infty}\omega_j\psi^j\,, \tag{3.2}$$

with resonances

$$j = -1,0,1,4,5,6\,; \tag{3.3}$$

$$\text{(ii)}\quad u = \psi^{-2}\sum_{j=0}^{\infty}u_j\psi^j\,, \quad \omega = \psi^{-2}\sum_{j=0}^{\infty}\omega_j\psi^j\,, \tag{3.4}$$

with resonances

$$j = -2,-1,3,4,6,8\,. \tag{3.5}$$

The Bäcklund transformation about (3.2) is

$$u = 2\frac{\partial^2}{\partial x^2}\ln\psi + u_2\,, \quad \omega = \frac{\omega_0}{\psi} + \omega_1\,, \tag{3.6}$$

where, as found in Ref. 6,

$$\psi_t + \psi_{xxx} + 3\psi_x u_2 = 2\psi_x\vartheta\,, \tag{3.7}$$

$$\omega_0 = \psi_x H\,, \tag{3.8}$$

$$\psi_t/\psi_x - \tfrac{1}{2}\{\psi;x\} = \tfrac{3}{4}H^2 + \vartheta\,, \tag{3.9}$$

$$\vartheta_x^2 = (\lambda^2 + \vartheta^2)H^2\,, \tag{3.10}$$

$$\omega_1 = -\tfrac{1}{2}\omega_{0x}/\psi_x - \tfrac{1}{3}(\lambda^2 + \vartheta^2)^{1/2}\,, \tag{3.11}$$

and

$$H_t + \frac{\partial}{\partial x}\left(H_{xx} + \frac{H^3}{4} + \vartheta H + \frac{3}{2}\{\psi;x\}H\right) = 0\,. \tag{3.12}$$

In Ref. 6 we have found the Lax pair for (3.1) by "linearizing" the Miura transformation, (3.7) and (3.11), from the "modified equations" (3.9) and (3.12). To review, we let

$$W = \psi_{xx}/\psi_x\,, \tag{3.13}$$

and find the "modified" equations

$$W_t = \frac{1}{2}\frac{\partial}{\partial x}\left(W_{xx} - \frac{W^3}{2} + 3\left(H_x + \frac{WH}{2}\right)H + 2(\vartheta_x + W\vartheta)\right), \tag{3.14}$$

$$H_t + \frac{\partial}{\partial x}\left(H_{xx} + \frac{1}{4}H^3 + \vartheta H + \frac{3}{2}\left(W_x - \frac{1}{2}W^2\right)H\right) = 0, \tag{3.15}$$

where

$$\vartheta_x^2 = (\lambda^2 + \vartheta^2)H^2.$$

The Miura transformations are

$$\begin{aligned} -2u_2 &= W_x + \tfrac{1}{2}W^2 + \tfrac{1}{2}H^2 - \tfrac{2}{3}\vartheta, \\ -2\omega_1 &= H_x + WH + \tfrac{2}{3}(\lambda^2 + \vartheta^2)^{1/2}. \end{aligned} \tag{3.16}$$

Then letting

$$\begin{aligned} &W + H = 2(\epsilon_x/\epsilon), \quad W - H = 2(\beta_x/\beta), \\ &\vartheta = \lambda \sinh \alpha, \quad \alpha = \ln(\epsilon/\beta), \end{aligned} \tag{3.17}$$

obtains the Lax pair from (3.14)–(3.16) (see Ref. 6).

We now proceed to study the Bäcklund transformations of (3.14) and (3.15) when

$$\vartheta = \lambda = 0. \tag{3.18}$$

The relevant equations are

$$\psi_t/\psi_x - \tfrac{1}{2}\{\psi;x\} = \tfrac{3}{4}H^2 \tag{3.19}$$

and

$$\begin{pmatrix} W \\ H \end{pmatrix}_t = \begin{pmatrix} \frac{\partial}{\partial x} & 0 \\ 0 & \frac{\partial}{\partial x} \end{pmatrix} \times \begin{pmatrix} \frac{W_{xx}}{2} - \frac{W^3}{4} + \frac{3}{2}\left(H_x + \frac{WH}{2}\right)H \\ -H_{xx} - \frac{H^3}{4} - \frac{3}{2}\left(W_x - \frac{W^2}{2}\right)H \end{pmatrix}. \tag{3.20}$$

Equations (3.20) allow the following singularities:

$$W \sim W_0\varphi^{-1}, \quad H \sim H_0\varphi^{-1}, \tag{3.21}$$

$$\text{(i)} \quad W_0 = -2\varphi_x, \quad H_0 = 0; \tag{3.22}$$

$$\text{(ii)} \quad W_0 = 2\varphi_x, \quad H_0 = 0, \quad \pm 4\varphi_x; \tag{3.23}$$

$$\text{(iii)} \quad W_0 = \varphi_x, \quad -3\varphi_x, \quad H_0^2 = \varphi_x^2. \tag{3.24}$$

When

$$W_0 = \varphi_x, \quad H_0^2 = \varphi_x^2, \tag{3.25}$$

the resonances are

$$j = -1,1,2,3,3,4 \tag{3.26}$$

and the Bäcklund transformation is

$$W = \varphi_x/\varphi + W_1, \quad H = a(\varphi_x/\varphi) + H_1, \tag{3.27}$$

where

$$a^2 = 1, \tag{3.28}$$

$$aH_1 + W_1 = -\varphi_{xx}/\varphi_x, \tag{3.29}$$

$$\frac{\varphi_t}{\varphi_x} - \frac{1}{2}\{\varphi;x\} = \frac{3}{2}a\left(H_{1x} - \frac{\varphi_{xx}}{\varphi_x}H_1 - aH_1^2\right), \tag{3.30}$$

and

$$H_{1t} + \frac{\partial}{\partial x}\left(H_{1xx} + \frac{1}{4}H_1^3 + \frac{3}{2}\left(W_{1x} - \frac{1}{2}W_1^2\right)H_1\right) = 0. \tag{3.31}$$

We note that Eqs. (3.29)–(3.31) imply that (W_1,H_1) satisfy Eq. (3.20). Now, consistent with Eqs. (3.19) and (3.20), we define the variable ψ so that

$$W_1 = \psi_{xx}/\psi_x, \tag{3.32}$$

$$\psi_t/\psi_x - \tfrac{1}{2}\{\psi;x\} = \tfrac{3}{4}H_1^2, \tag{3.33}$$

$$H_{1t} + \frac{\partial}{\partial x}\left(H_{1xx} + \frac{H_1^3}{4} + \frac{3}{2}\{\psi;x\}H_1\right) = 0. \tag{3.34}$$

Note that Eqs. (3.33) and (3.34) define an equation for ψ formulated in terms of the Schwarzian derivative.

We find from (3.29) and (3.30), using (3.32) and (3.33), that

$$\begin{aligned} \frac{\varphi_t}{\varphi_x} &= -\frac{\varphi_{xxx}}{\varphi_x} + \frac{3}{4}\frac{\varphi_{xx}^2}{\varphi_x^2} - \frac{3}{2}\frac{\psi_{xxx}}{\psi_x} - \frac{3}{2}\frac{\varphi_{xx}}{\varphi_x}\frac{\psi_{xx}}{\psi_x}, \\ \frac{\psi_t}{\psi_x} &= \frac{1}{2}\frac{\psi_{xxx}}{\psi_x} + \frac{3}{2}\frac{\varphi_{xx}}{\varphi_x}\frac{\psi_{xx}}{\psi_x} + \frac{3}{4}\frac{\varphi_{xx}^2}{\varphi_x^2}. \end{aligned} \tag{3.35}$$

These equations may be written in the form

$$\psi_t = A\psi_{xx} + B\psi_x, \quad \psi_{xxx} = -A\psi_{xx} + (C - B)\psi_x, \tag{3.36}$$

where

$$A = \varphi_{xx}/\varphi_x, \quad C = -\varphi_t/\varphi_x - \{\varphi;x\}, \quad B = \tfrac{1}{2}A^2 - \tfrac{1}{3}C. \tag{3.37}$$

The compatibility condition $(\psi_{txxx} = \psi_{xxxt})$ of the linear equations, (3.36), for ψ obtains the equation for φ,

$$2\frac{\partial}{\partial t}\left(\frac{\varphi_t}{\varphi_x} + \{\varphi;x\}\right) = \frac{\partial}{\partial x}\begin{pmatrix} \frac{\partial}{\partial x^2}\left(\frac{\varphi_t}{\varphi_x} + \{\varphi;x\}\right) - \frac{9}{2}\{\varphi;x\}^2 \\ + 6\{\varphi;x\}\left(\frac{\varphi_t}{\varphi_x} + \{\varphi;x\}\right) - \left(\frac{\varphi_t}{\varphi_x} + \{\varphi;x\}\right)^2 \end{pmatrix}, \tag{3.38}$$

which is formulated in terms of the Schwarzian derivative. The equivalent condition $(\varphi_{txxx} = \varphi_{xxxt})$ obtains, from (3.35), that ψ satisfies Eqs. (3.33) and (3.34), which is distinct from (3.38). Therefore (3.35) defines a Bäcklund transformation between two equations formulated in terms of the Schwarzian derivative.

Letting

$$\Omega = \varphi_t/\varphi_x + \{\varphi;x\}, \quad W = \varphi_{xx}/\varphi_x, \tag{3.39}$$

the Miura transformation (3.16) is

$$\begin{aligned} -2u_2 &= W_{1x} + \tfrac{1}{2}(W_1^2 + H_1^2) = -\tfrac{2}{3}\Omega, \\ -3a\omega_1 &= \tfrac{3}{2}a(H_{1x} + W_1H_1) = \Omega - \tfrac{3}{2}(W_x - \tfrac{1}{2}W^2). \end{aligned} \tag{3.40}$$

Rather than consider Eqs. (3.38) and (3.34) separately we will define the new variables

$$V=\varphi_{xx}/\varphi_x\,,\quad W=\psi_{xx}/\psi_x\,, \tag{3.41}$$

and find from (3.35) the equations

$$V_t=\frac{\partial}{\partial x}\circ\left(\frac{\partial}{\partial x}+V\right)\circ\left\{-V_x+\frac{1}{2}V^2-\frac{3}{2}W_x-\frac{3}{2}W^2-\frac{3}{2}WV-\frac{3}{4}V^2\right\}, \tag{3.42}$$

$$W_t=\frac{\partial}{\partial x}\left(\frac{\partial}{\partial_x}+W\right)\circ\left\{\frac{1}{2}\left(W_x-\frac{1}{2}W^2\right)+\frac{3}{4}(W+V)^2\right\}.$$

These equations allow the following singularities:

$$V\sim V_0\epsilon^{-1}\,,\quad W\sim W_0\epsilon^{-1} \tag{3.43}$$

[where for simplicity $\epsilon=x+f(t)$, j is resonance]

$$\begin{aligned}
&\text{(i)} && V_0=0\,, && \omega_0=1\,, && j=-1,1,2,3,3,4\,;\\
&\text{(ii)} && V_0=-2\,, && W_0=-1\,, && j=-1,1,2,3,3,4\,;\\
&\text{(iii)} && V_0=4\,, && W_0=-3\,, && j=-2,-1,3,3,4,5\,;\\
&\text{(iv)} && V_0=2\,, && W_0=2\,, && j=-5,-1,3,3,4,8\,;\\
&\text{(v)} && V_0=2\,, && W_0=-2\,, && j=-1,1,2,3,3,4\,;\\
&\text{(vi)} && V_0=2\,, && W_0=-3\,, && j=-2,-1,3,3,4,5\,;\\
&\text{(vii)} && V_0=-2\,, && W_0=2\,, && j=-1,-1,3,3,4,4\,;\\
&\text{(viii)} && V_0=-6\,, && W_0=2\,, && j=-5,-1,3,3,4,8\,.
\end{aligned} \tag{3.44}$$

The Bäcklund transformations for (3.42) are of the form

$$V=V_0\epsilon^{-1}+V_1\,,\quad W=W_0\epsilon^{-1}+W_1\,. \tag{3.45}$$

For (i), (ii), and (v), with resonances at $j=-1,1,2,3,3,4$, (3.45) obtains a system of five equations for the five unknowns (ϵ, V_0, W_0, V_1, W_1). For (vii) there are six equations in five unknowns. We consider each in turn.

For BT1, we have

$$V=V_1\,,\quad W=\epsilon_x/\epsilon+W_1\,, \tag{3.46}$$

where ϵ satisfies (3.38),

$$\begin{aligned}
&\epsilon_x=1/\varphi_x\psi_x^2\,,\\
&\frac{\epsilon_t}{\epsilon_x}=\frac{\psi_t}{\psi_x}+\frac{\partial}{\partial x}\left(\frac{\varphi_{xx}}{\varphi x}\right)-\frac{\varphi_{xx}^2}{\varphi_x^2}-\frac{\psi_{xx}}{\psi_x}\frac{\varphi_{xx}}{\varphi_x}\,,
\end{aligned} \tag{3.47}$$

$$V_1=\varphi_{xx}/\varphi_x\,,\quad W_1=\psi_{xx}/\psi_x\,, \tag{3.48}$$

and both (ϵ,ψ) and (φ,ψ) satisfy (3.35). Note (3.47) defines a Bäcklund transformation for Eq. (3.35),

$$(\epsilon,\psi)\leftrightarrow(\varphi,\psi)\,, \tag{3.49}$$

while (3.46) defines the BT

$$(\varphi,\psi)\leftrightarrow(\varphi,\psi')\,, \tag{3.50}$$

where

$$\psi'_x=\epsilon\psi_x\,,\quad \frac{\psi'_t}{\psi'_x}=\frac{\psi_t}{\psi_x}+\frac{\epsilon_x}{\epsilon}\frac{\varphi_{xx}}{\varphi_x}\,. \tag{3.51}$$

For BT2, we have

$$V=-2(\epsilon_x/\epsilon)+V_1\,,\quad W=\epsilon_x/\epsilon+W_1\,, \tag{3.52}$$

where (ϵ,ψ) satisfy (3.35):

$$V_1=\epsilon_{xx}/\epsilon_x\,,\quad W_1=\psi_{xx}/\psi_x\,. \tag{3.53}$$

Letting

$$V=\varphi'_{xx}/\varphi'_x\,,\quad W=\psi'_{xx}/\psi'_x\,, \tag{3.54}$$

then

$$\varphi'=-1/\epsilon\,, \tag{3.55}$$

$$\psi'_x=\epsilon\psi_x\,,\quad \psi'_t=\epsilon\psi_t-\psi_x\epsilon_{xx}-2\psi_{xx}\epsilon_x\,, \tag{3.56}$$

where (φ',ψ') satisfy (3.35).

For BT5, we have

$$V=2(\epsilon_x/\epsilon)+V_1\,,\quad W=-2(\epsilon_x/\epsilon)+W_1\,, \tag{3.57}$$

where

$$V_1=\varphi_{xx}/\varphi_x\,,\quad W_1=\epsilon_{xx}/\epsilon_x \tag{3.58}$$

obtains (φ,ϵ) satisfying (3.35) and, with (3.54),

$$\varphi'_x=\epsilon^2\varphi_x\,,\quad \varphi'_t=\epsilon^2\varphi_t+4\varphi_x\epsilon\epsilon_{xx}+2\varphi_{xx}\epsilon\epsilon_x-2\varphi_x\epsilon_x^2\,, \tag{3.59}$$

$$\psi'=-1/\epsilon. \tag{3.60}$$

Finally, for BT7, we have

$$V=-2(\epsilon_x/\epsilon)+V_1\,,\quad W=2(\epsilon_x/\epsilon)+W_1\,, \tag{3.61}$$

where

$$V_1=\epsilon_{xx}/\epsilon_x\,,\quad W_1=\psi_{xx}/\psi_x=-\epsilon_{xx}/\epsilon_x\,, \tag{3.62}$$

$$\epsilon_t/\epsilon_x=\tfrac{1}{2}\{\epsilon;x\}\,. \tag{3.63}$$

Note the restriction

$$\psi_x=\epsilon_x^{-1}\,. \tag{3.64}$$

With (3.54),

$$\varphi'=-1/\epsilon\,,\quad \psi'_x=\epsilon^2\psi_x=\epsilon^2/\epsilon_x\,, \tag{3.65}$$

where (φ',ψ') satisfy (3.63). As expected BT7 is a reduction of (3.35) (to the KdV equation[4]) which is preserved by (3.61).

The Bäcklund transformations, BT1, BT2, BT5, and BT7, form a group under composition that properly restricts to (3.35); that is, maps solutions into solutions. The following identities are easily verified [where $\binom{\varphi'}{\psi'}=\mathrm{BT}\binom{\varphi}{\psi}$]:

$$\begin{aligned}
&\text{(i)} && \mathrm{BT1}\circ\mathrm{BT1}=\mathrm{BT2}\circ\mathrm{BT2}=\begin{pmatrix}1&0\\0&-1\end{pmatrix},\\
&\text{(ii)} && \mathrm{BT5}\circ\mathrm{BT5}=\begin{pmatrix}1&0\\0&1\end{pmatrix}=I\,,\\
&\text{(iii)} && \mathrm{BT1}\circ\mathrm{BT2}=\mathrm{BT2}\circ\mathrm{BT1}\,,\\
&\text{(iv)} && \mathrm{BT7}\circ\mathrm{BT7}=I\,,
\end{aligned} \tag{3.66}$$

and subject to the restriction on the domain of BT7,

$$\psi_x=\varphi_x^{-1}\,, \tag{3.67}$$

we also have

$$\text{(v)}\quad \mathrm{BT1}\circ\ \mathrm{BT2}\circ\mathrm{BT7}=I\,. \tag{3.68}$$

Furthermore, BT5 preserves the KdV restriction, (3.67), while BT1 and BT2 do not [preserve (3.67)]. The Bäcklund transformation

$$K=\mathrm{BT5}\circ\mathrm{BT7} \tag{3.69}$$

generates a sequence of KdV solutions[4]

$$\begin{pmatrix}\varphi_j\\ \psi_j\end{pmatrix} = K^j\begin{pmatrix}\varphi_0\\ \psi_0\end{pmatrix}, \qquad (3.70)$$

where (φ_0,ψ_0), (φ_j,ψ_j) satisfy (3.63) and $(\psi_{0x} = \varphi_{0x}^{-1})$, $(\psi_{j,x} = \varphi_{j,x}^{-1})$. For instance,

$$\varphi_0 = \psi_0 = x \qquad (3.71)$$

obtains a sequence of rational KdV solutions.[4]

In general, the (φ_j,ψ_j) are distinct in that $K^n \neq I$ for any $n > 0$. Note that

$$K = \mathrm{BT5}\circ\mathrm{BT7} \neq K' = \mathrm{BT7}\circ\mathrm{BT5}. \qquad (3.72)$$

Also,

$$\mathrm{BT1}\circ\mathrm{BT5}\neq\mathrm{BT5}\circ\mathrm{BT1}, \quad \mathrm{BT2}\circ\mathrm{BT5}\neq\mathrm{BT5}\circ\mathrm{BT2}. \qquad (3.73)$$

Application of the group of transformations generated by (BT1, BT2, BT5) to (3.70) produces a "lattice" of solutions of the Hirota–Satsuma equations (3.35). It can be shown that (BT1, BT2, BT5, BT7) map rational solutions into rational solutions [of (3.35)]. Therefore, from (3.70) and (3.71), rational solutions of the Hirota–Satsuma equations are found.

Direct calculation obtains a few solutions:

$$\begin{pmatrix}\varphi_0\\ \psi_0\end{pmatrix} = \begin{pmatrix}x\\ x\end{pmatrix} \xrightarrow{\mathrm{BT2}} \begin{pmatrix}-\dfrac{1}{x}\\ \dfrac{x^2}{2}\end{pmatrix} \xrightarrow{\mathrm{BT5}} \begin{pmatrix}\dfrac{x^3-24t}{12}\\ -\dfrac{2}{x^2}\end{pmatrix} \xrightarrow{\mathrm{BT2}} \begin{pmatrix}-\dfrac{12}{x^3-24t}\\ \dfrac{x^3+12t}{3x^2}\end{pmatrix} \xrightarrow{\mathrm{BT5}} \begin{pmatrix}-\dfrac{4}{x}\dfrac{x^3-6t}{x^3-24t}\\ -\dfrac{3x^2}{x^3+12t}\end{pmatrix}, \qquad (3.74)$$

$$\begin{pmatrix}\varphi_0\\ \psi_0\end{pmatrix} = \begin{pmatrix}x\\ x\end{pmatrix} \xrightarrow{\mathrm{BT5}} \begin{pmatrix}\dfrac{x^3-6t}{3}\\ -\dfrac{1}{x}\end{pmatrix} \xrightarrow{\mathrm{BT2}} \begin{pmatrix}-\dfrac{3}{x^3-6t}\\ \dfrac{x^3+12t}{6x}\end{pmatrix} \xrightarrow{\mathrm{BT5}} \begin{pmatrix}\dfrac{x}{4}\dfrac{x^3-24t}{x^3-6t}\\ -\dfrac{6x}{x^3+12t}\end{pmatrix} \xrightarrow{\mathrm{BT2}} \begin{pmatrix}-\dfrac{4}{x}\dfrac{x^3-6t}{x^3-24t}\\ -\dfrac{3x^2}{x^3+12t}\end{pmatrix}, \qquad (3.75)$$

$$\begin{pmatrix}\varphi_0\\ \psi_0\end{pmatrix} = \begin{pmatrix}x\\ x\end{pmatrix} \xrightarrow{\mathrm{BT1}} \begin{pmatrix}x\\ \dfrac{x^2}{2}\end{pmatrix} \xrightarrow{\mathrm{BT5}} \begin{pmatrix}\dfrac{x^5}{20}\\ -\dfrac{2}{x^2}\end{pmatrix} \xrightarrow{\mathrm{BT1}} \begin{pmatrix}\dfrac{x^5}{20}\\ \dfrac{x^3+12t}{12x^2}\end{pmatrix} \xrightarrow{\mathrm{BT5}} \begin{pmatrix}\dfrac{1}{576}\left(\dfrac{x^7}{7}+6tx^4+144t^2x\right)\\ -\dfrac{12x^2}{x^3+12t}\end{pmatrix}, \qquad (3.76)$$

$$\begin{pmatrix}\varphi_0\\ \psi_0\end{pmatrix} = \begin{pmatrix}x\\ x\end{pmatrix} \xrightarrow{\mathrm{BT5}} \begin{pmatrix}\dfrac{x^3-6t}{3}\\ -\dfrac{1}{x}\end{pmatrix} \xrightarrow{\mathrm{BT1}} \begin{pmatrix}\dfrac{x^3-6t}{3}\\ \dfrac{x^3+12t}{6x}\end{pmatrix} \xrightarrow{\mathrm{BT5}} \begin{pmatrix}\dfrac{1}{36}\left(\dfrac{x^7}{7}+6tx^4+144t^2x\right)\\ -\dfrac{6x}{x^3+12t}\end{pmatrix} \xrightarrow{\mathrm{BT1}} \begin{pmatrix}\dfrac{1}{36}\left(\dfrac{x^7}{7}+6tx^4+144t^2x\right)\\ -\dfrac{3x^2}{x^3+12t}\end{pmatrix}, \qquad (3.77)$$

$$\begin{pmatrix}\varphi_0\\ \psi_0\end{pmatrix} = \begin{pmatrix}x\\ x\end{pmatrix} \xrightarrow{\mathrm{BT1}} \begin{pmatrix}x\\ \dfrac{x^2}{2}\end{pmatrix} \xrightarrow{\mathrm{BT5}} \begin{pmatrix}\dfrac{x^5}{20}\\ -\dfrac{2}{x^2}\end{pmatrix} \xrightarrow{\mathrm{BT2}} \begin{pmatrix}-\dfrac{20}{x^5}\\ \dfrac{x^3+30t}{15}\end{pmatrix} \xrightarrow{\mathrm{BT5}} \begin{pmatrix}\dfrac{4}{9}\left(\dfrac{x^6-30tx^3-180t^2}{x^5}\right)\\ -\dfrac{15}{x^3+30t}\end{pmatrix}. \qquad (3.78)$$

From (3.74) and (3.75),

$$(\mathrm{BT5}\circ\mathrm{BT2})^2\circ\begin{pmatrix}\varphi_0\\ \psi_0\end{pmatrix} = (\mathrm{BT2}\circ\mathrm{BT5})^2\circ\begin{pmatrix}\varphi_0\\ \psi_0\end{pmatrix}, \qquad (3.79)$$

and, from (3.76) and (3.77),

$$D\circ(\mathrm{BT5}\circ\mathrm{BT1})^2\circ\begin{pmatrix}\varphi_0\\ \psi_0\end{pmatrix} = (\mathrm{BT1}\circ\mathrm{BT5})^2\circ\begin{pmatrix}\varphi_0\\ \psi_0\end{pmatrix}, \qquad (3.80)$$

where

$$D = \begin{pmatrix}16 & 0\\ 0 & \frac{1}{4}\end{pmatrix}, \qquad (3.81)$$

and

$$\begin{pmatrix}\varphi_0\\ \psi_0\end{pmatrix} = \begin{pmatrix}x\\ x\end{pmatrix}. \qquad (3.82)$$

The periodicities (3.79) and (3.80) are not verified in general $\left[\text{for arbitrary } \binom{\psi_0}{\varphi_0}\right]$. Determination of relationships of the form (3.79) and (3.80) when $\binom{\psi_0}{\varphi_0}$ belongs to the KdV sequence, (3.70), may provide a method for "classifying" the Hirota–Satsuma solutions.

ACKNOWLEDGMENTS

This work is supported by the Department of Energy under Contract No. DOE-DE-AC03-81ER10923 and the Air Force Office of Scientific Research under Grant No. AFOSR84-0195.

[1]J. Weiss, J. Math. Phys. **26**, 258 (1985).

[2]J. Weiss, M. Tabor, and G. Carnevale, J. Math. Phys. **24**, 522 (1983).

[3]J. Weiss, J. Math. Phys. **24**, 1405 (1983).

[4]D. V. Chudnovsky, G. V. Chudnovsky, and M. Tabor, Phys. Lett. A **97**, 268 (1983).

[5]J. Weiss, J. Math. Phys. **25**, 13 (1984).

[6]J. Weiss, J. Math. Phys. **25**, 2226 (1984).

[7]P. Radmore, D. Wood, J. D. Gibbon, and M. Tabor, Stud. Appl. Math. **72**, 39 (1985).

[8]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol. II.

[9]M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715 (1980).

[10]H. Yoshida, Celestial Mech. **81**, 363, 381 (1983).

Geometrical interpretation of the solutions of the sine–Gordon equation

J. J. Klein
Department of Physics, The University of Regina, Regina, Saskatchewan, S4S OA2, Canada[a)]

(Received 31 January 1985; accepted for publication 9 May 1985)

The sine–Gordon equation is known to possess solutions that correspond to solitons, that is, localized entities that maintain their shape after collisions, and have certain properties characteristic of elementary particles. Although the algebraic structure of these solutions is well known, their geometric interpretation as surfaces of constant negative curvature has not been previously illuminated. We discuss these surfaces herein. Curves drawn on these surfaces along the asymptotic directions at each point simulate solutions of the nonlinear wave equation $\phi_{xx} - \phi_{tt} = \sin\phi$.

I. INTRODUCTION

The sine–Gordon equation

$$\phi_{xx} - \phi_{tt} = \sin\phi \tag{1}$$

has many applications in physics, such as pulse propagation in one-dimensional media in optics,[1] the theory of superconductivity,[2] and dislocation movement in metal crystals.[3] It was exhaustively studied about ten years ago, when it was discovered that it belongs to a special class of nonlinear wave equations possessing "soliton" solutions. As this name implies, the wave equation admits as solutions localized entities of unique form, which do not spread out as time passes, and moreover, pass through each other without being altered in shape (except for a phase shift) in a manner simulating that of one-dimensional billiard-ball collisions in classical mechanics. In 1976, it was pointed out by the author[4] that the "breather" solitons, which the equation possesses, behave also like elementary particles and have an inherent time dependence which gives the soliton a dual nature; that is, the localized pulse is always accompanied by a plane wave when viewed in a moving frame. These solutions have the form

$$\phi = 4\tan^{-1}\left(\frac{b}{a}\,\frac{\cos at}{\cosh bx}\right), \quad a^2 + b^2 = 1, \tag{2}$$

in the rest frame of the soliton.

A Lorentz transformation of the above gives

$$\phi = 4\tan^{-1}\left(\frac{b}{a}\,\frac{\cos a(\gamma t - \beta\gamma x)}{\cosh b(\gamma x - \beta\gamma t)}\right), \quad \gamma = (1-\beta^2)^{-1/2}. \tag{3}$$

When two or more such entities collide, a phase shift appears in the plane wave factor, although the solitons are otherwise unaltered. Thus the particles possess an inherent dualism reminiscent of that of elementary particles in quantum theory. Moreover, when a soliton is trapped in a "well," i.e., boundary conditions are imposed at finite points in the one-dimensional medium, the particlelike aspect disappears as the soliton can form only a standing wave, periodic in space and time. (See Ref. 4, especially the part "Standing Wave Solutions," p. 1385.) Because of these properties of the solutions of the sine–Gordon equation, it has value as a model for teaching students about the dualism of wave and particle that is evident in quantum theory.

The sine–Gordon equation in the form

$$\frac{\partial^2\omega}{\partial u^2} - \frac{\partial^2\omega}{\partial v^2} = \sin\omega\cos\omega \tag{4}$$

has been used in the study of surfaces of constant negative Gaussian curvature,[5] K, which may be taken as -1. The above equation arises when one writes the metric in the "asymptotic form" natural for constant negative curvature surfaces:

$$ds^2 = \cos^2\omega\, du^2 + \sin^2\omega\, dv^2, \tag{5}$$

where (u,v) are coordinates such that $u = \text{const}$, $v = \text{const}$ represent the lines of principal curvature of the surface, and ω is the angle between the u axis and the asymptotic direction on the surface. The asymptotic direction is that direction in which the normal curvature is zero. The stipulation that the Gaussian curvature be -1 leads immediately to Eq. (4). Mathematicians have usually taken (4) as the canonical form, while physicists have used (1). It is evident that if $\phi = 2\omega$ these equations are identical.

The algebraic structure of the many-soliton solutions of (1) has been presented in various forms, by various authors.[6] The particular representation used by Hirota is in some ways the most convenient. From his equations one can write down explicit formulas for the one-, two-, three-, etc. soliton solutions of (1). (I am referring to the breather solitons, which in Hirota's terms, correspond to his $N = 2, 4, 6$, etc. cases.) It will be shown below that the two-breather solution behaves, at $t \to \infty$, as a single soliton, that is, as if we are in the rest frame of one of the solitons and we are observing the second soliton approach, pass by, and recede as t varies from $-\infty$ to $+\infty$. The solitons have equal "mass," so no recoil is evident, and only a phase shift in the particle's time-oscillatory factor $\cos(bt)$ indicates passage of the other soliton.

The geometrical nature of the surfaces, imbedded in ordinary three-dimensional space, which corresponds to these solutions of (1) or, if you wish, (4), seems not to have been elucidated. [Although it is possible that a visualization of the surface corresponding to a single breather may have been presented in the older mathematical literature during the past century, it is in any case not widely known to physicists. The many-soliton surface probably was not even considered by mathematicians, since the existence of these solutions of (4) was not known until 1973.]

The purpose of this paper is to give explicit formulas for

[a)] Address starting September 1985: Department of Physics, St. Cloud State University, St. Cloud, Minnesota 56301.

imbedding the single-breather soliton in three-space so as to visualize why such entities exist as solutions of the sine–Gordon equation and how a many-soliton solution can be visualized as generated by bending the single-soliton surface. We will show why the sine–Gordon equation possesses the many-soliton nature, and why it is probably unique in differential geometry in this regard.

II. IMBEDDING THE SINGLE-SOLITON BREATHER IN EUCLIDEAN THREE-SPACE

We will use the older notation of differential geometry with (u,v) as the curvature coordinates on a surface of constant negative curvature, with the understanding that in modern soliton theory the crucial equation

$$\frac{\partial^2\omega}{\partial u^2}-\frac{\partial^2\omega}{\partial v^2}=\sin\omega\cos\omega \tag{4}$$

has the interpretation of a nonlinear wave equation, with (u,v) being space and time variables (x,t), and the asymptotic angle ω is to be interpreted as the dependent variable with physical meaning. Certain parameters of the solutions are designated by (a,h,k) in Darboux' treatise and in modern soliton theory are denoted by (κ,ω), with the stipulation $\kappa^2+\omega^2=1$ imposed.

The method described by Darboux and attributed to Enneper is too lengthy to be described in detail here.[7] The outcome of their reasoning is the conclusion that spherical surfaces cutting the $K=-1$ surface at 90° angles must have their centers along a straight line, which is designated as the X axis. This preferred axis would for the ordinary pseudosphere (a tractrix of revolution) be simply an axis of rotational symmetry. Darboux does not concern himself with this (degenerate) case; nor shall we, since it represents a soliton $\omega=2\tan^{-1}e^{-u}$, that is, time independent in its rest frame. Other time-independent (i.e., v-independent) solutions involving elliptic integrals are also known and not dealt with here (for a pictorial representation of these axially symmetric surfaces, the reader is referred to Eisenhart[8]).

Darboux then sets up a cylindrical coordinate system (X,λ,ψ) in ordinary three-space in which the non-axially symmetric $K=-1$ surfaces are to be imbedded. We cite his findings as follows: the imbedding coordinates are

$$X=\left\{z\cos\omega-\int z^2\,du\right\}\frac{1}{a^2}, \tag{6}$$

$$\lambda=[(1+z_1^2)^{1/2}/a^2]\sin\omega, \tag{7}$$

and

$$\psi=a^2\int\frac{dv}{1+z_1^2},\quad a^2+b^2=1, \tag{8}$$

where ω, of course, is a solution of the equation

$$\frac{\partial^2\omega}{\partial u^2}-\frac{\partial^2\omega}{\partial v^2}=\sin\omega\cos\omega, \tag{4}$$

and $z(u)$ and $z_1(v)$ are solutions of the auxiliary equations

$$z'^2=z^4+az^2+2h \tag{9}$$

and

$$z_1'^2=(z_1^2+1)^2+a(z_1^2+1)+2h. \tag{10}$$

To obtain the usual breather soliton in my present notation one must rewrite, in (9) and (10), the parameter a as $(-2a^2)$, and $2h$ as a^4. Then these equations become simply

$$z'^2=z^4-2a^2z^2+a^4,\quad\text{whereby}\quad z=a\tanh au, \tag{11}$$

and

$$z_1'^2=(z_1^2+1)^2-2a^2(z_1^2+1)+a^4, \tag{12}$$

so $z_1=b\tan bv$.

Darboux' results indicate only that a solution of the equation $\omega_{uu}-\omega_{vv}=\sin\omega\cos\omega$ may be represented by

$$\cos\omega=(z'-z_1')/(z^2-z_1^2-1), \tag{13}$$

where z and z_1 satisfy (9) and (10). In the special case dealt with here, it is easy to show that (11) and (12) lead to the expression for ω:

$$\omega=2\tan^{-1}((a/b)\cos bv/\cosh au) \tag{14}$$

with $a^2+b^2=1$.

This is recognizable as the usual breather soliton in modern work on the sine–Gordon equation.[9] Inserting (14) into (6)–(8) yields

$$X=\frac{2z}{a^2}\left\{\frac{b^2+z_1^2}{1+z_1^2-z^2}\right\}-u, \tag{15}$$

$$\lambda=\frac{(1+z_1^2)^{1/2}}{a^2}\frac{2(b^2+z_1^2)^{1/2}(a^2-z^2)^{1/2}}{1+z_1^2-z^2}, \tag{16}$$

$$\psi=v-\tan^{-1}(b\tan bv), \tag{17}$$

where $z=a\tanh au$ and $z_1=b\tan bv$.

A profile of the surface can be obtained by setting ψ equal to some constant value, say $\psi=0$. Then $v=0$ and $z_1=0$ also.

The $\psi=0$ profile is then given as

$$\lambda=[2b/(1-z^2)][(a^2-z^2)^{1/2}/a^2], \tag{18a}$$

$$X=\frac{2zb^2}{a^2(1-z^2)}-\frac{1}{a}\tanh^{-1}\left(\frac{z}{a}\right). \tag{18b}$$

Consider the case $a=b=1/\sqrt{2}=0.707$. Then the $\psi=0$ profile is

$$\lambda=2(1-2z^2)^{1/2}/(1-z^2), \tag{18c}$$

$$X=[2z/(1-z^2)]-\sqrt{2}\tanh^{-1}(\sqrt{2}z), \tag{18d}$$

$0<z<1/\sqrt{2}$. The asymptotic angle ω on the $\psi=0$ section is

$$\omega=\pm2\tan^{-1}(1-2z^2)^{1/2}.$$

The profile is a self-intersecting curve, symmetric about $X=0$ and tending toward the X axis, as $X\to\pm\infty$.

The quantity ω, which in differential geometry is an asymptotic direction on a surface, plays the role of the wave function in the nonlinear sine–Gordon wave equation. Thus, Fig. 1 can be construed, in a certain sense, as a "snapshot" of the wave taken at time $t=0$. The wave function goes from 0 to $\pi/2$ and back to 0 as

$$z=a\tanh au=(1/\sqrt{2})\tanh(u/\sqrt{2})$$

varies (u plays the role of distance x). We wish to emphasize that a snapshot of a solution of a wave equation normally displays the wave function ϕ versus distance x. If one follows the behavior of ω along the looped curve in Fig. 1, the depen-

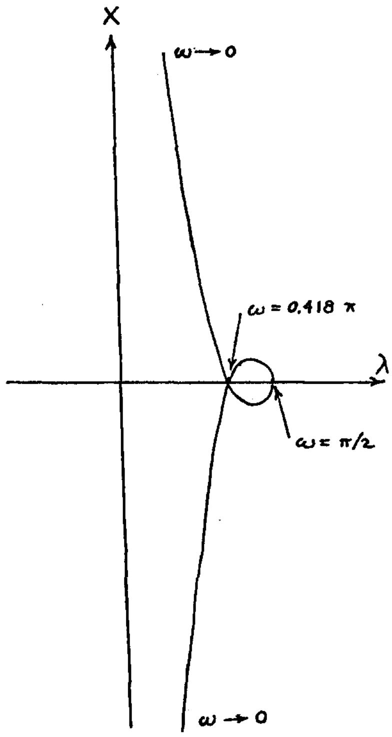


FIG. 1. Profile $\psi = 0$ of $a = b = 1/\sqrt{2}$ surface.

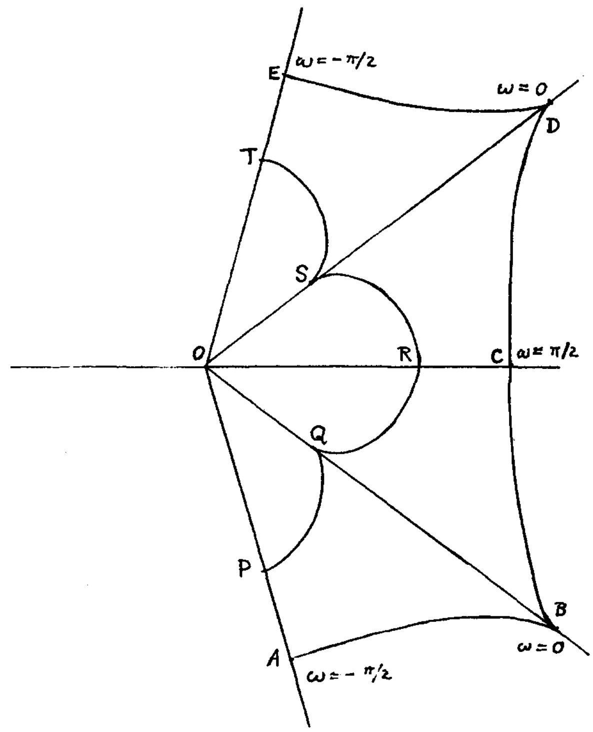


FIG. 2. Cross section $X = 0$ of $a = b = 1/\sqrt{2}$ surface.

dence of ω on u is displayed implicitly.

It is evident that a surface having two parts is represented by Eq. (15)–(17). Let us pursue this further by looking at the cross section $X = 0$. The polar equations expressing λ and ψ parametrically in terms of z_1 (rather than v) are, for $a = b = 1/\sqrt{2}$, *with z taken as zero*,

$$\lambda = 2(1 + 2z_1^2)^{1/2}/(1 + z_1^2)^{1/2} \tag{19a}$$

and

$$\begin{aligned}\psi &= v - \tan^{-1}\left(\frac{1}{\sqrt{2}}\tan\frac{v}{\sqrt{2}}\right)\\ &= \sqrt{2}\tan^{-1}(\sqrt{2}\,z_1) - \tan^{-1} z_1,\end{aligned} \tag{19b}$$

where z_1 may take on *any* value from 0 to ∞.

However, the equation

$$X = \frac{2}{a^2} z\left\{\frac{b^2 + z_1^2}{1 + z_1^2 - z^2}\right\} - \frac{1}{a}\tanh^{-1}\left(\frac{z}{a}\right) = 0,$$

with $a^2 = b^2 = \frac{1}{2}$ has a *second* solution besides the obvious $z = 0$ solution. Again taking z_1 as a parameter we have

$$4z\{(\tfrac{1}{2} + z_1^2)/(1 + z_1^2 - z^2)\} - \sqrt{2}\tanh^{-1}(\sqrt{2}\,z) = 0, \tag{20a}$$

$$\lambda = 4(1 + z_1^2)^{1/2}(\tfrac{1}{2} + z_1^2)^{1/2}(\tfrac{1}{2} - z^2)^{1/2}/(1 + z_1^2 - z^2), \tag{20b}$$

$$\psi = \sqrt{2}\tan^{-1}(\sqrt{2}\,z_1) - \tan^{-1} z_1. \tag{20c}$$

Again, z_1 may take on any value from 0 to ∞.

A sketch of the $X = 0$ cross section with $a^2 = b^2 = \frac{1}{2}$ is given in Fig. 2.

The figure suggests a five-pointed star or a five-lobed rose depending on whether one considers ABCDE$\cdots$ or PQRST$\cdots$ (it is evident that the figure shows only a portion of the non-axially symmetric surface; it actually continues around the X axis, which is perpendicular to the paper, an infinite number of times).

For certain choices of b, the star/rose closes on itself, and the asymptotic angle ω has the same values when ψ attains the same values (mod 2π). For instance, if $b = \frac{3}{5}, \frac{4}{6}, \frac{5}{7}, \frac{6}{8}$, ..., the figure has threefold, fourfold, fivefold, and sixfold symmetry (reminiscent of ring-shaped organic compounds). However, only the "even" values $b = \frac{4}{6}, \frac{6}{8}$,... yield strict periodicity in ω, as a function of angle ψ; that is, if one follows an asymptotic curve around the X axis as the curve wends its way over the surface, the angle ω will be reversed in sign if $b = \frac{3}{5}$, or $\frac{5}{7}$, when $\psi = 2\pi$.

Zeros of Hirota's f and g functions give rise to $\tan(\omega/2) = 0, \infty$ so that $\omega = 0, \pi$. This condition implies that the asymptotic curves lie parallel to $v = \text{const}$ curves; this in turn implies a cuspidal edge in the surface. These can be inferred also from Fig. 2. As $u \to \pm\infty$, the asymptotic curves on the inner part of the surface encompassing the X axis spiral around this axis in such a way that $\omega \to 0$. The values $u \to \pm\infty$ yield singular points analogous to those of the pseudosphere. The X value tends to infinity also at these points, as is evident from (18d).

As mentioned above, the surface whose asymptotic directions satisfy (14) has two parts, one part bounded by ABCDE$\cdots$ and PQRST$\cdots$, which bears a resemblance to a chain of "apple cores" joined end-to-end along the cusps QB, SD, and so on around the X axis; and the other part, bounded on the outside by the rose curve PQRST$\cdots$, and which extends all along the X axis out to infinity, tapered to a peak at $X = \pm\infty$ so that the Gaussian curvature is everywhere -1.

For arbitrary a,b the $\psi = 0$, $v = 0$, $z_1 = 0$ profile is

$$\omega = 2\tan^{-1}((a/b)1/\cosh au) \tag{21}$$

or

$$\omega = 2\tan^{-1}((a/b)(1-(z^2/a^2))^{1/2}). \quad (22)$$

This is a looped curve similar to that shown in Fig. 1. To reiterate, the $\psi = \text{const}$ profile reflects the spatial dependence of the soliton. The $X = 0$ cross section brings out the time dependence, that is, the v dependence of ω, as it varies from 0 at the cuspidal edges, to $\omega_{\max}$ or $-\omega_{\max}$ at the innermost points of the star curve. It should be emphasized that although the v dependence is periodic in (14), the surface representing this soliton intersects itself as it whorls around the X axis [except for certain special choices of (a,b) which cause it to close]. Thus, it is convenient to visualize the surface as extending from $\psi = -\infty$ to $\psi = +\infty$, around the X axis. (Mathematically this is certainly permissible, although an engineer accustomed to dealing with actual surfaces might object to a self-intersecting surface of many sheets.)

It should be pointed out that this imbedding of the surface of constant negative Gaussian curvature -1 in ordinary Euclidean three-space can be explicitly verified. One starts with the Euclidean metric in cylindrical coordinates

$$ds^2 = dX^2 + d\lambda^2 + \lambda^2 d\psi^2, \quad (23)$$

substitutes the expressions (6), (7), and (8) for X, λ, ψ, and finds that a metric involving u and v is obtained of the form

$$ds^2 = \cos^2\omega\, du^2 + \sin^2\omega\, dv^2,$$

where $\omega(u,v)$ is a solution (14) of the sine–Gordon equation (!). The algebra involved is lengthy, but straightforward, and gives one confidence that the imbedding is really correct.

III. MANY-SOLITON SOLUTIONS

The formulas giving ω as a function of (u,v) allow one to describe the surface in a natural manner. Every solution of the equation

$$\frac{\partial^2\omega}{\partial u^2} - \frac{\partial^2\omega}{\partial v^2} = \sin\omega\cos\omega \quad (4)$$

describes a surface of constant negative curvature that can be imbedded in ordinary three-space.

Geometrically, it seems plausible that bending the single-soliton surface (14) will yield another type of solution. Gauss pointed out that bending a surface does not change its intrinsic curvature K. In the present context, this means that if the Gaussian curvature was originally -1, it is still -1 after bending. (For the reader unfamiliar with the concepts of differential geometry, this means that although the principal curvatures κ_1 and κ_2 change, their product $K = \kappa_1\kappa_2 = -1$ does not.)

In order to shed light on this question, we examine the solution of (4) that represents a two-soliton solution. We mean here two-breather solutions, each breather soliton possessing the dualism inherent in solutions of the type (14). Following the prescription given by Hirota,[10] one writes

$$\omega = 2\tan^{-1}(g/f). \quad (24)$$

The auxiliary functions g and f used by Hirota satisfy certain (nonlinear) partial differential equations whose solutions can be cast into a standard form consisting of certain sums over products of purely exponential functions involving certain parameters, designated by the symbols (κ,ω). These constant parameters eventually turn out to be identical with parameters designated as (a,b) in our discussion above involving the differential-geometric interpretation of the soliton solutions.

To obtain breather solitons, the exponentials must be combined pairwise; the procedure is straightforward, but complicated algebraically. The two-breather solution (Hirota's $N = 4$ case) is (using a,b instead of κ,ω)

$$f = -b'a\cosh(bt)\cosh(a'x) + iba'\cos(b't)\cosh(ax), \quad (25)$$

$$g = a'a\{\lambda_1\cosh(bt)\cos(b't) + i\lambda_2\sinh(bt)\sin(b't)\} - ib'b\{\mu_1\cosh(ax)\cosh(a'x) + \mu_2\sinh(a'x)\sinh(ax)\}, \quad (26)$$

with the constraints $a^2 + b^2 = 1$, $a'^2 - b'^2 = 1$.

The constants λ_1, λ_2 and μ_1, μ_2 depend on the a's and b's; specifically,

$$\lambda_1 = (b'^2 - b^2)/(b'^2 + b^2), \quad (27a)$$

$$\lambda_2 = 2ib'b/(b'^2 + b^2). \quad (27b)$$

Analogous expressions can be written for μ_1 and μ_2. One finds, as $t \to +\infty$, that the ratio (g/f) simplifies to

$$\tan\frac{\omega}{2} = \frac{g}{f} \to \frac{-a'}{b'}\,\frac{\sin(b't-\delta)}{\cosh(a'x)}; \quad (28a)$$

and as $t \to -\infty$,

$$\tan\frac{\omega}{2} = \frac{g}{f} \to \frac{-a'}{b'}\,\frac{\sin(b't+\delta)}{\cosh(a'x)}. \quad (28b)$$

The phase shift is found to be

$$\delta = \tan^{-1}((b'^2 - b^2)/2b'b). \quad (29)$$

It seems very difficult to write down explicit formulas giving the X, λ, and ψ coordinates of the surface described by

$$\tan(\omega/2) = g/f, \quad (30)$$

with the g, f given by complicated many-soliton expressions such as (25) and (26). Even if the methods used by Darboux or Enneper were feasible, it seems likely that physical interpretation would be awkward.

Instead, we note from the asymptotic behavior at $t \to \pm\infty$ (in geometrical terms, $v \to \pm\infty$) that the surface is identical with the star/rose surface discussed in Sec. II. To visualize the surface, we imagine it to be an infinitely many-sheeted surface whose asymptotic angle ω has the forms (28a) and (28b) when the v coordinate approaches $\pm\infty$. For values of v which are not large, the periodic behavior is lost, and a perturbation of the surface produces nonperiodicity in the v coordinate. Interpreting v as a timelike coordinate t, this perturbation represents in soliton theory the disturbance produced as a second soliton approaches, passes through, and recedes. The soliton seen at $t \to \pm\infty$ is the soliton at rest in the frame of the observer. It suffers a phase shift when the moving soliton "collides" with it. (There is no recoil.)

The single-soliton surface described by (14) is capable of being deformed, that is, *bent* into the double-soliton surface implied by (30) using Hirota's f and g formulas. By the theorem of Gauss—the *theorema egregium*—the curvature $K = \kappa_1\kappa_2 = -1$ remains invariant under bending. The single-soliton surface, under bending, deforms in a manner analogous to that of an accordion; the perturbation introduced into the many-sheeted surface (14) generates the two-soliton surface described by (25), (26), and (30), which, of course, is again a surface of Gaussian curvature -1.

Thus, it becomes clearly understandable why a simple solution of

$$\frac{\partial^2 \omega}{\partial u^2} - \frac{\partial^2 \omega}{\partial v^2} = \sin \omega \cos \omega \tag{4}$$

can generate more complex solutions under deformation, that is, bending. When the above equation is interpreted as a nonlinear wave equation

$$\frac{\partial^2 \phi}{\partial x^2} - \frac{\partial^2 \phi}{\partial t^2} = \sin \phi, \tag{1}$$

$\phi = 2\omega$, this geometrical property (invariance of Gaussian curvature) finds algebraic expression in the many-soliton formulas of Hirota, Ablowitz, Caudrey, and Takhadzhyan.

IV. DISCUSSION

One might ask whether there are other types of surfaces dealt with in differential geometry that have properties similar to the surfaces of constant negative curvature, in that their defining equations might have solutions that display soliton behavior. Minimal surfaces have average curvature $\frac{1}{2}(\kappa_1 + \kappa_2)$ equal to zero, but their Gaussian curvature varies from point to point on the surface. The asymptotic angle ω always equals $\pi/4$ ($\tan \omega = 1$) everywhere on a minimal surface. It is natural to use a metric of the form

$$ds^2 = e^q(du^2 + dv^2), \tag{31}$$

where q is a function of both u and v, in general. It can be shown that most, if not all, minimal surfaces satisfy the equation

$$\frac{\partial^2 q}{\partial u^2} + \frac{\partial^2 q^2}{\partial v^2} = e^{-2q}. \tag{32}$$

The minus sign in the exponential on the right is crucial. This equation is the natural analog, for minimal surfaces, of Eq. (4), which represents the defining equation of $\kappa_1 \kappa_2 = -1$ surfaces. One obvious difference between (4) and (32) is that the latter is *not* a wave equation. Furthermore, inspection of various minimal surfaces such as catenoids or Enneper's minimal surface reveals that, although they have negative curvature everywhere (as they must, since an asymptotic direction does not exist wherever the Gaussian curvature is positive), they do not have the unique attributes under a bending deformation required for generating entities that could be construed as solitons, in any sense of the word.

Therefore, although minimal surfaces are in certain ways complementary to constant (negative) curvature surfaces, they are, in another sense fundamentally and intrinsically different. Other negative curvature surfaces, such as ruled surfaces, an example of which is the hyperboloid of a one-sheet, do not appear to have the necessary attributes. They can be bent, but nothing simple and mathematically beautiful happens. We conclude that the sine–Gordon equation is unique.

V. CONCLUSION

The rationale for this study of the sine–Gordon equation is that it provides, in one dimension, a model for understanding the nature and behavior of elementary particles. The wave function, which in differential geometry is an angle representing the asymptotic direction at a point on a surface, is related to the arc tangent of a product of two factors. In the rest frame, one factor, $\mathrm{sech}(ax)$, gives the wave its localization in space; the other factor, $\cos(bt)$, gives the wave an inherent time dependence. (This is the origin of the phrase "breather soliton.") As mentioned in the Introduction, a Lorentz transformation can always be carried out since the sine–Gordon equation is invariant under such a transformation of the variables (x,t). The effect of this is to give a Lorentz contraction to the spatial factor, and to convert the time factor to the form of a "plane wave." The dualism appears, almost magically, when the soliton is seen in a moving frame. Second, these disturbances behave exactly like particles moving in a one-dimensional space in so far as they always retain their original shape after colliding. This is the canonical property of any sort of soliton. Third, previous work has shown that sine–Gordon solitons form purely standing-wave type entities when trapped in a one-dimensional enclosure. Their sharp, spatial localization disappears when boundary conditions are imposed.

The behavior of the asymptotic angle ω in terms of its functional dependence on curvature coordinates (u,v) provides a simulation of the solutions of the sine–Gordon wave equation. The differential geometry of surfaces of constant negative curvature[11] is worth studying for this reason.

[1] *Propagation of Ultra-Short Optical Pulses, In Honor of Philip Morse*, edited by H. Feshbach and K. V. Ingard (M.I.T, Cambridge, MA, 1969), Chap. 7, pp. 88–105; G. L. Lamb, Rev. Mod. Phys. **43**, 99 (1971). Multisoliton theory is developed, using Bäcklund transforms, in order to picture a variety of picosecond or subpicosecond pulses of light.

[2] P. Lebwohl and M. J. Stephen, Phys. Rev. **163**, 376 (1967).

[3] A. Kochendorfer and A. Seeger, Z. Phys. **127**, 553 (1950).

[4] J. J. Klein, Can. J. Phys. **54**, 1383 (1976).

[5] G. Darboux, *Theorie Generale des Surfaces* (reprinted by Chelsea, New York, 1972), Vol. VII, Chaps. X–XIII.

[6] A. C. Scott, F. Y. F. Chu, and D. W. McLaughlin, Proc. IEEE **61**, 1443 (1973); the work of Caudrey *et al.*, Ablowitz *et al.*, and Hirota is cited in Appendices A and C. See also, L. A. Takhadzhyan Sov. Phys. JETP **39**, 228 (1974).

[7] Reference 5, Vol. VII, Chap. XIII, Sec. 814–820.

[8] L. P. Eisenhart, *An Introduction to Differential Geometry* (Princeton U. P., Princeton, NJ, 1940), pp. 283 and 284.

[9] M. J. Ablowitz, D. J. Kaup, G. F. Newell, and H. Segur, Phys. Rev. Lett. **30**, 1262 (1973).

[10] R. Hirota, J. Phys. Soc. Jpn. **33**, 1459 (1972).

[11] R. Steuerwald, Abh. Bayerische Acad. Wiss. München **40**, 1 (1936); "An extensive treatment of surfaces of constant negative curvature," available at the Yale University Library, Call No. A32m 0180 II 40.

Harmonic analysis of the Euclidean group in three-space. II

Jung Sik Rno
R. Walters College, University of Cincinnati, Cincinnati, Ohio 45236

(Received 21 February 1985; accepted for publication 26 April 1985)

We develop the harmonic analysis for spinor functions which are defined by the matrix elements of the unitary irreducible representations of E(3) with the representation space on the translation subgroup.

I. INTRODUCTION

This paper is a sequel to our previous paper[1] (called hereafter Paper I). Paper I treated the harmonic analysis on the (simply connected) twofold universal covering group of the Euclidean group in three-space, proving the explicit Plancherel formula. In this paper this covering group is, for the sake of simplicity, named the Euclidean group in three-space (E(3)). Miller[2] introduced spinor functions through the group representations of E(3). However, no harmonic analysis on spinor functions, to the best of our knowledge, has been attempted. This paper fills this gap.

Spinor functions are the vector-valued, generalized spherical Bessel functions and they are solutions[3] of the Helmholtz, or reduced wave, equation. Aside from the importance of spinor functions, there are at least two obvious motivations of this paper in contemporary physics. The harmonic expansion is an effective tool to study high- (larger than four-) dimensional theories, i.e., Kaluza–Klein and supergravity theories.[4] The harmonic analysis also plays a major role in the formalisms of the stochastic quantum mechanics and stochastic quantum field theory.[5–7]

Section II contains the résumé of the group E(3). Section III outlines the construction of unitary irreducible representations (UIR's) of E(3). Section IV presents the necessary summary of spinor functions by the matrix elements of UIR's of E(3). Section V deals with the harmonic analysis involving spinor functions, deriving, among others, the generalized Parseval's formula.

II. THE GROUP E(3)

The group E(3) is the semidirect product space $R^3 \times_\eta \mathrm{SU}(2)$ relative to the homomorphism η of SU(2) into the group of automorphisms of R^3. For simplification we skip the bold-faced notation of vectors. The matrices $\pm A \in \mathrm{SU}(2)$ determine the same rotation $\eta(A)$ given by

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

where σ stands for the Pauli matrices

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

The multiplication law for E(3) is

$$\{r_1,A_1\}\{r_2,A_2\} = \{r_1 + \eta(A_1)r_2, A_1A_2\}. \tag{3}$$

If

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

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

$$\eta(A) = \begin{pmatrix} (a^2 - b^2 + \bar a^2 - \bar b^2)/2 & i(\bar a^2 + \bar b^2 - a^2 - b^2)/2 & \bar a\bar b + ab \\ i(a^2 - b^2 - \bar a^2 + \bar b^2)/2 & (\bar a^2 + \bar b^2 + a^2 + b^2)/2 & i(-\bar a\bar b + ab) \\ -(\bar a b + a\bar b) & i(-\bar a b + a\bar b) & a\bar a - b\bar b \end{pmatrix}. \tag{4}$$

III. THE UIR OF E(3)

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

$$P^3 = \bigcup_{\dot p \in K} \Omega(\dot p) \equiv \bigcup_{\rho > 0} \Omega_\rho, \tag{5}$$

where $K = \{\dot p = (0,0,\rho):\ \rho \geqslant 0\}$.

Hence there are only two stability groups (little groups)

$$G_{\dot p} = \mathrm{SU}(2),\quad \text{for } \dot p \in \Omega_0,$$
$$G_{\dot p} = \mathrm{SO}(2),\quad \text{for } \dot p \in \Omega_\rho \quad (\rho > 0), \tag{6}$$

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

$$\Gamma^s\left(\begin{bmatrix} e^{i\psi/2} & 0 \\ 0 & e^{-i\psi/2} \end{bmatrix}\right) = e^{is\psi}, \tag{7}$$

where $2s = 0, \pm 1, \pm 2, \ldots$.

The UIR's associated with the trivial orbit $\Omega_0 = \{0\}$ are of no interest in the present work. The UIR's (ρ,s) of E(3) associated with an orbit $\Omega_\rho\ (\rho > 0)$ are given by

$$[U^{\rho,s}(a,A)f](p) = e^{ip\cdot a}(\Gamma^s \uparrow \mathrm{SU}(2))(p,A)f(A^{-1}p), \tag{8}$$

where $\uparrow$ denotes "induced."

The carrier space of (ρ,s) is $H(\rho,s)$, the Hilbert space of Lebesgue square-integrable functions on the manifolds Ω_ρ with the inner product

 0022-2488/85/092186-03$02.50

$$\langle f, g\rangle = \int_{\Omega_\rho} f(p)\,\overline{g(p)}dw(p), \quad f,g\in H(\rho,s), \tag{9}$$

where $dw(p) = \sin\theta\, d\theta\, d\varphi$ for

$$p = (\rho\sin\theta\cos\varphi, \rho\sin\theta\sin\varphi, \rho\cos\theta)\in\Omega_\rho.$$

Further details are given in Ref. 8.

IV. SPINOR FUNCTIONS

We first determine the matrix elements of the operator $U^{\rho,s}(r,A)$ with respect to the orthonormal basis $\{h^u_m(\theta,\varphi)\}$ of $H(\rho,s)$. For fixed u the functions $\{h^u_m(\theta,\varphi)\}, m = -u, -u+1,\ldots, +u$, form a canonical basis for the irreducible representations $D(u)$ of SU(2). In the following we write $U(r,A)$ instead of $U^{\rho,s}(r,A)$. Since $U(r,A) = U(r,I)U(0,A)$, it suffices to calculate the matrix elements of $U(r,I)$ and $U(0,A)$, respectively. The latter are nothing but those of SU(2)

$$\langle h^v_n, U(0,A)h^u_m\rangle = T^u_{nm}(A)\cdot\delta_{v,u}. \tag{10}$$

By definition, the matrix elements of $U(r,I)$ are[8]

$$\begin{aligned}\langle h^v_n, &U(r,I)h^u_m\rangle \\ &\equiv [v,n|\rho,s|u,m](r) \\ &= \sqrt{4\pi}\sum_{l=|u-v|}^{u+v} i^{-l}\left[\frac{(2u+1)(2l+1)}{(2v+1)}\right]^{1/2} C(l,0;u,s|v,s) \\ &\quad\times C(l,n-m;u,m|v,n)\,j_l(\rho r)\,\overline{Y^{n-m}_l(\theta r,\varphi r)},\end{aligned} \tag{11}$$

where $C(\cdot;\cdot|\cdot)$ are the Clebsch–Gordan coefficients of SU(2) and the $j_l(\rho r)$ are spherical Bessel functions. The

$$j^{v,u}_{s,n}(\rho r) = i^{u-v}[v,n|\rho,s|u,m]((0,0,r))$$

are called generalized spherical Bessel functions.[9] In particular, $j^{l,0}_{0,0}(\rho r) = j_l(\rho r)$.

The $[v,n|\rho,s|u,m](r)$ for fixed v are called v-spinor functions, which have $2v+1$ components

$$\begin{aligned}\chi^{(\rho,s)}_{v;u,m}(r) &= ([v,n|\rho,s|u,m](r)), \\ n &= -v, -v+1,\ldots, +v,\end{aligned} \tag{12}$$

for some u and m. They are solutions[3] of the Helmholtz, or reduced wave, equation. In the following we often call them spinor functions instead of v-spinor functions.

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

$$\begin{aligned}\{v,n|&\rho,s|u,m\}(r,A) \\ &\equiv\langle h^v_n, U(r,A)h^u_m\rangle = \sum_{m'=-u}^{u}[v,n|\rho,s|u,m'](r)T^u_{m',m}(A) \\ &= \sum_{n'=-v}^{v} T^v_{n,n'}[v,n'|\rho s|u,m](A^{-1}r).\end{aligned} \tag{13}$$

Equation (13) shows $\chi^{(\rho,s)}_{v;u,m}(r)$ transforms like a spinor field of weight v; in fact, under the action of SU(2) it transforms like the eigenvector h^u_m of the irreducible representation $D(u)$. Furthermore, the components of $\chi^{(\rho,s)}_{v;u,m}(r)$ satisfy the Helmholtz, or reduced wave, equation.

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

$$\begin{aligned}\int_{R^3} d^3r\int_{\mathrm{SU(2)}} dA\,&\{v,n|\rho,s|u,m\}(r,A)\,\overline{\{v',n'|\rho',s'|u',m'\}(r,A)} \\ &= (4\pi^2/\rho^2)\cdot\delta(\rho-\rho')\cdot\delta_{s,s'}\cdot\delta_{u,u'}\cdot\delta_{v,v'}\cdot\delta_{m,m'}\cdot\delta_{n,n'}.\end{aligned} \tag{14}$$

It is well known that the matrix elements $T^u_{m,n}(A)$ defined for $2u = 0,1,2,\ldots;\ m,n = -u, -u+1,\ldots, +u$ satisfy the orthogonality relations

$$\begin{aligned}\int_{\mathrm{SU(2)}} &T^u_{n,m}(A)\cdot T^{u\prime}_{n',m'}(A)dA \\ &= (1/2u+1)\delta_{u,u'}\cdot\delta_{m,m'}\cdot\delta_{n,n'},\end{aligned} \tag{15}$$

where dA is the normalized Haar measure on SU(2). Making use of (13)–(15), we can easily obtain

$$\begin{aligned}\int_{R^3} d^3r &\sum_{n=-v}^{v}[v,n|\rho,s|u,m](r) \\ &\times\overline{[v,n|\rho',s'|u',m'](r)} \\ &= (4\pi^2/\rho^2)\cdot\delta(\rho-\rho')\cdot\delta_{s,s'}\cdot\delta_{u,u'}\cdot\delta_{m,m'}.\end{aligned} \tag{16}$$

V. HARMONIC ANALYSIS OF SPINOR FUNCTIONS

We first construct the Hilbert space $H_v \equiv L^2_v(R^3)$, the elements of which are vector-valued functions

$$\psi(r) = \begin{pmatrix}\Psi_v(r)\\ \cdot\\ \cdot\\ \cdot\\ \Psi_{-v}(r)\end{pmatrix} = \sum_{n=-v}^{v}\Psi_n(r)\cdot e_n, \tag{17}$$

where e_n is the column vector with a one in row n and zeros everywhere else. The vector $\psi(r)\in H_v$ if

$$\int_{R^3}\Psi^t(r)\overline{\Psi}(r)d^3r = \int_{R^3}\sum_{n=-v}^{v}|\Psi_n(r)|^2\,d^3r < \infty, \tag{18}$$

where the superscript t denotes "transposed" and the bar on $\Psi(r)$ signifies its complex conjugate. The inner product is

$$\begin{aligned}\langle\Phi,\Psi\rangle &\equiv \int_{R^3}\Phi^t(r)\overline{\Psi}(r)d^3r \\ &= \int_{R^3}\sum_{n=-v}^{v}\Phi_n(r)\overline{\Psi}_n(r)d^3r.\end{aligned} \tag{19}$$

We remark that H_v is the ordinary state space of a single nonrelativistic particle with spin v.

We shall establish a generalized Fourier transform of elements in H_v. We define the unitary (left) regular representation V of E(3) in H_v by

$$[U(a,A)\Psi](r) = T^v(A)\Psi(A^{-1}(r-a)), \tag{20}$$

or in components

$$[U(a,A)\Psi_m](r) = \sum_{n=-v}^{v} T^v_{mn}(A)\Psi_n(A^{-1}(r-a)). \tag{21}$$

We apply the Fourier transform $\mathscr{F}$ to $\Psi_n(r)$

$$(\mathscr{F}\cdot\Psi_n)(p) \equiv \widehat{\Psi}_n(p) = \frac{1}{(2\pi)^{3/2}}\int_{R^3} e^{ip\cdot r}\Psi_n(r)d^3r. \tag{22}$$

The inverse Fourier transform $\mathscr{F}^{-1}$ is given by

$$(\mathscr{F}^{-1}\cdot\hat{\Psi}_n)(r)\equiv\Psi_n(r)\,\frac{1}{(2\pi)^{3/2}}\int_{P^3}e^{-ir\cdot p}\hat{\Psi}_n(p)d^3p. \quad (23)$$

Then obviously we have

$$\int_{R^3}|\Psi_n(r)|^2\,d^3r=\int_{P^3}|\hat{\Psi}_n(p)|^2\,d^3p. \quad (24)$$

The regular representation U in (20) induces in $\hat{H}_v(R^3)\equiv H_v(P^3)$

$$[\hat{U}(a,A)\hat{\Psi}](p)=e^{ia\cdot p}T^v(A)\hat{\Psi}(A^{-1}p), \quad (25)$$

or in components

$$[\hat{U}(a,A)\hat{\Psi}_m](p)=e^{ia\cdot p}\sum_{n=-v}^{v}T^v_{mn}(A)\Psi_n(A^{-1}p). \quad (26)$$

Clearly U and $\hat{U}$ are unitarily equivalent via $U=\mathscr{F}^{-1}\hat{U}\mathscr{F}$.

Similarly, as in Paper I, we set

$$\hat{\Psi}^\rho(p)=\hat{\Psi}(p),\quad \text{for}\quad p\in\Omega_\rho. \quad (27)$$

Then we have

$$[\hat{U}(a,A)\hat{\Psi}^\rho_m](p)=e^{ia\cdot p}\sum_{n=-v}^{v}T^v_{mn}(A)\hat{\Psi}^\rho_n(A^{-1}p). \quad (28)$$

We also have

$$\int_{P^3}|\hat{\Psi}_n(p)|^2\,d^3p=\int_0^\infty\rho^2\,d\rho\int_{\Omega_\rho}|\hat{\Psi}^\rho_n(p)|^2\,dw(p), \quad (29)$$

where $dw(p)$ is the invariant Haar measure of Ω_ρ. Making use of (24) and (29), we can derive

$$\int_{R^3}\Psi^t(r)\overline{\Psi}(r)d^3r=\int_0^\infty\rho^2\,d\rho\sum_{n=-v}^{v}\int_{\Omega_\rho}|\hat{\Psi}^\rho_n(p)|^2\,dw(p). \quad (30)$$

We now turn to v-spinor functions. Equations (16) and (19) lead to

$$\langle\chi^{(\rho,s)}_{v;u,m},\chi^{(\rho,s)}_{v;u',m'}\rangle=\delta_{u,u'}\cdot\delta_{m,m'}. \quad (31)$$

Exploiting (31) we can define a generalized Fourier transform for functions in $\hat{H}_v(P^3)$

$$\hat{\Psi}_v(p)=\int_0^\infty\rho^2\,d\rho\times\sum_{s=-l}^{l}\sum_{2u=0}^{\infty}\sum_{m=-u}^{u}\hat{\Psi}^{(\rho,s)}_{v;u,m}\cdot\chi^{(\rho,s)}_{v;u,m}(\rho), \quad (32)$$

or in components

$$\hat{\Psi}_{v,n}(p)=\int_0^\infty\rho^2\,d\rho\times\sum_{s=-l}^{l}\sum_{2u=0}^{\infty}\sum_{m=-u}^{u}\hat{\Psi}^{(\rho,s)}_{v,n;u,m}\cdot\chi^{(\rho,s)}_{v,n;u,m}(p), \quad (33)$$

where l is the smaller of u and v, and

$$\hat{\Psi}^{(\rho,s)}_{v,n;u,m}=\langle\hat{\Psi}_{v,n},\chi^{(\rho,s)}_{v,n;u,m}\rangle.$$

From (33) we can easily derive

$$\int_{P^3}|\hat{\Psi}_{v,n}(p)|^2\,d^3p=\int_0^\infty\rho^2\,d\rho\sum_{s=-l}^{l}\sum_{2u=0}^{\infty}\sum_{m=-u}^{u}|\Psi^{(\rho,s)}_{v,n;u,m}|^2. \quad (34)$$

Combining (30) and (34) we can write the generalized Parseval's formula

$$\int_{R^3}\Psi^t_v(r)\hat{\Psi}_v(r)d^3r=\int_0^\infty\rho^2\,d\rho\times\sum_{s=-\min(u,v)}^{\min(u,v)}\sum_{n=-v}^{v}\sum_{2u=0}^{\infty}\sum_{m=-u}^{u}\hat{\Psi}^{(\rho,s)}_{v,n;u,m}\,\overline{\hat{\Psi}^{(\rho,s)}_{v,n;u,m}}. \quad (35)$$

ACKNOWLEDGMENTS

The author has benefited from discussions with F. Mansouri, W. Miller, Jr., F. E. Schroeck, Jr., and L. Witten. He is grateful to S. Sprague for her superb secretarial help.

This research has been supported in part by R. Walters College, University of Cincinnati.

[1] J. S. Rno, J. Math. Phys. **26**, 675 (1985).
[2] W. Miller, Jr., Commun. Pure Appl. Math. **17**, 527 (1964).
[3] W. Miller, *Lie Theory and Special Functions* (Academic, New York, 1968).
[4] Reference 1 and references listed therein.
[5] S. P. Gudder, *Stochastic Methods in Quantum Mechanics* (North-Holland, New York, 1979).
[6] E. Prugovecki, *Stochastic Quantum Mechanics and Quantum Spacetime* (Reidel, Dordrecht, Holland, 1984).
[7] F. E. Schroeck, Jr., "Quantum fields for reproducing kernel Hilbert spaces," Florida Atlantic University preprint, 1984.
[8] J. S. Rno, J. Math. Phys. **15**, 2042 (1974).
[9] H. Ogura, J. Phys. Soc. Jpn. **25**, 586 (August 1968).

Dynamic systems driven by Markov processes

Andrzej Korzeniowski
Department of Mathematics, University of Texas at Arlington, Arlington, Texas 76019

(Received 6 February 1985; accepted for publication 10 May 1985)

Consider a differential equation $\dot{Y} = V(X(t))Y(t)$, where $X(t)$ is a random function. Sufficient conditions for asymptotic stability of the solution in terms of a generator of the stochastic process $X(t)$ are given. The results are illustrated by several examples.

I. INTRODUCTION

The main objective in this paper is to study the behavior of the solution $Y(t)$ to $\dot{Y} = V(X(t))Y(t)$, where $X(t)$ is a stochastic process and $V(x)$ is a real-valued function. A motivation comes from the dynamic reliability discussed by Ladde and Siljak.[1] We are mainly interested in finding conditions for convergence of $Y(t)$ in some sense defined below, as $t \to \infty$. In particular asymptotic stability of the stochastic solution is discussed. For probabilistic facts we refer the reader to Doob.[2]

Before making any statements it is instructive to consider several examples that will illuminate the nature of the problem.

II. EXAMPLES AND RESULTS

Example 1: Let $X(t)$ be a two-state Markov process with values a, $-a$ for some $a > 0$, and infinitesimal generator

$$A = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}, \quad \text{for some } \lambda,\mu > 0,$$

then

$$Y(t) = Y(0)\exp\left(\int_0^t V(X(s))ds\right).$$

One can analyze $Y(t)$ in two ways: pathwise or in the mean by conditioning on $X(0) = x$. The first case is easy to handle because $X(t)$ has a stationary measure $m = [\mu/(\lambda + \mu), \lambda/(\lambda + \mu)]$, which means $A^T m = 0$. Take, for simplicity, $V(x) = x$, then by the ergodic theorem (see Ref. 3, p. 121)

$$\frac{1}{t}\int_0^t X(s)ds \to \int_{\{-a,a\}} x\, dm(x) = \frac{a(\mu - \lambda)}{\mu + \lambda}.$$

If one assumes that $\lambda - \mu > 0$, which means that the intensity of passing from $\{a\}$ to $\{-a\}$ is greater than that of passing from $\{-a\}$ to $\{a\}$ or equivalently the tendency of staying at $\{-a\}$ dominates staying at $\{a\}$, then one gets pathwise asymptotic stability because $Y(t) \to 0$ almost surely as $t \to \infty$. This is not surprising when taking into account the fact that in the long run $Y(t)$ should be more likely to approach $Y(0)e^{-at}$ than $Y(0)e^{at}$.

On the other hand, for $X(t)$ with $n + 1$ possible states $x_0,\ldots,x_n$ and a generator $A = (a_{ij})$, $i, j = 0,\ldots,n$, it is known (see Ref. 3, p. 299) that

$$M(x_i,t) = E_{x_i}Y(t) = Y(0)E_{x_i}\exp\left(\int_0^t V(X(s))ds\right),$$

$$X(0) = x_i, \quad i = 0,\ldots,n,$$

satisfy

$$\frac{\partial M}{\partial t} = (A + IV)M,$$

where IV is an $(n+1)\times(n+1)$ matrix with $V(x_i)$ on diagonal and 0 elsewhere. For a diffusion process $X(t)$ with the generator L, the above reads

$$\frac{\partial M}{\partial t} = (L + V)M, \quad M(x,0) = f(x),$$

and the solution

$$M(x,t) = E_x f(X(t))\exp\left(\int_0^t V(X(s))ds\right)$$
$$= \exp[t(L + V)]f(x)$$

is given by the generalized Feynman–Kac (FK) formula (see Ref. 4, p. 314), which for $X(t)$ = Brownian motion with the generator $L = \frac{1}{2}\Delta$ and V replaced by $-V$ reduces to the standard FK formula.[5]

Taking in our example $n = 1$, $V(x) = x$, $x_0 = -a$, $x_1 = a$, and $A = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}$ one has

$$M(\pm a,t) = Y(0)e^{\lambda^{\pm} t},$$

$$\lambda^{\pm} = \tfrac{1}{2}[-(\mu + \lambda) \pm \sqrt{4\lambda\mu + (\lambda - \mu - 2a)^2}],$$

whence to get asymptotic stability we must assume $\lambda^{+} < 0$ or equivalently $0 < a < \lambda - \mu$. Notice that pathwise asymptotic stability requires only $0 < \lambda - \mu$. In general, the stability should be expressed in terms of $V(x)$ and the generator A of the Markov process $X(t)$.

Example 2: Let $X(t) = B(t)$ be a Brownian motion $V(x) = x$, $Y(0) = 1$, $X(0) = x$, then the mean value of $Y(t)$ solving $\dot{Y}(t) = B(t)Y(t)$ is given by

$$M(x,t) = E_x \exp\left(\int_0^t B(s)ds\right) = E\exp\left(\int_0^t (x + B(s))ds\right)$$
$$= \exp(xt + t^3/6) \to \infty,$$

as $t \to \infty$. The last equality comes from the fact that by the FK formula

$$\frac{\partial M}{\partial t} = \left(\frac{1}{2}\frac{\partial^2}{\partial x^2} + x\right)M,$$

which easily checks.

To get the asymptotic stability one needs a drift term for Brownian motion of order t^2. Namely, consider the diffusion

 0022-2488/85/092189-03$02.50

process $X(t)$ with the generator $L = \frac{1}{2}(\partial^2/\partial x^2) + 2\alpha t(\partial/\partial x)$, or equivalently $X(t) = B(t) + \alpha t^2$, then, by the above,

$$M(x,t) = \exp[xt + (\alpha/3)t^3 + t^3/6] \to 0 \quad \text{as } t \to \infty,$$

as long as $2\alpha < -1$. In general, if $X(t) = \sigma(t)B(t) + \mu(t)$, and $\sigma(t) > 0$,

$$\begin{aligned} M(x,t)E_x \exp\left(\int_0^t X(s)ds\right) &= E\exp\left(\int_0^t (x + \mu(s) + \sigma(s)B(s))ds\right) \\ &= \exp\left(xt + \int_0^t \mu(s)ds + \int_0^t \sigma(v)\left[\int_0^v u\sigma(u)du\right]dv\right), \end{aligned}$$

so knowing μ and σ it is possible in some cases to check whether $M(x,t) \to 0$ as $t \to \infty$.

Example 3: Let $X(t) = B(t)$,

$$V(x) = \operatorname{sgn} x = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0, \end{cases}$$

then

$$Y(t) = Y(0)\exp\left(\int_0^t V(B(s))ds\right), \quad B(0) = 0.$$

Notice that $T = \int_0^t V(B(s))ds$ = time spent by the Brownian particle in the positive half-line minus time spent by the Brownian particle in the negative half-line. Due to symmetry of Brownian motion one could expect that $Y(t) \to Y(0)$ as $t \to \infty$ because T should go to 0 almost surely. Quite the opposite is true. Since X is the time up to time t spent by a particle in the positive half-line it enjoys arcsin law $(2/\pi)\times\arcsin\sqrt{x/t}$ with the density $(1/\pi)[1/\sqrt{x(t-x)}]$, $0<x<t$, therefore conditioning on X one gets

$$\begin{aligned} Ee^T &= \int_0^t E(e^T|X = x)\frac{1}{\pi\sqrt{x(t-x)}}dx \\ &= \int_0^t e^{2x-t}\frac{1}{\pi\sqrt{x(t-x)}}dx \geqslant \frac{1}{3}e^{t/2}, \end{aligned}$$

and thus there is no stability.

Remark 1: If $V(x) \leqslant 0$ then $Y(t)$ is decreasing $[Y(0) > 0]$ and always has a limit as $t \to \infty$, which is 0 whenever $\int_\infty^0 V(X(s))ds = -\infty$. When this happens almost surely then also $E_x Y(t) \to 0$ for each x as $t \to \infty$. In terms of the diffusion process $X(t)$, pathwise asymptotic stability means that the probability of terminating the process at time t, while traveling along a given path $\{X(s)|0 \leqslant s \leqslant t\}$, which is equal to

$$\exp\left(\int_0^t V(X(s))ds\right),$$

goes to 0. In other words the longer the particle survives the smaller the chance of ever being killed.

Now let $X(t)$ be a birth–death process with $x_0,\ldots,x_n$ possible states, whose generator is the matrix

$$A = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & \\ & & \mu_{n-1} & -(\lambda_{n-1}+\mu_{n-1}) & \lambda_{n-1} \\ & & 0 & \mu_n & -\mu_n \end{bmatrix}$$

and stationary measure $m = (m_0,\ldots,m_n)$ with

$$m_i = 1 + \sum_{k=1}^n \frac{\lambda_0\cdots\lambda_{k-1}}{\mu_1\cdots\mu_k},$$

and

$$m_i = (\lambda_0\lambda_1\cdots\lambda_{i-1})\left[\mu_1\mu_2\cdots\mu_i\left(1 + \sum_{k=1}^n \frac{\lambda_0\cdots\lambda_{k-1}}{\mu_1\cdots\mu_k}\right)\right]^{-1}$$

$$i = 1,\ldots,n$$

(see Ref. 6, p. 154).

Theorem 1: The solution $Y(t)$ of $\dot{Y}(t) = V(X(t))Y(t)$ has the following properties.

(i) $Y(t) \to 0$ a.s. (almost surely) as $t \to \infty$ whenever

$$\sum_{i=0}^n V(x_i)m_i = \int_{\{x_0,\ldots,x_n\}} V(x)dm(x) < 0.$$

(ii) $E_x Y(t) \to 0$ as $t \to \infty$ whenever real parts of the eigenvalues of $A + IV$ are negative.

(iii) $V(x_i) < \lambda_i - \lambda_{i-1} + \mu_i - \mu_{i+1}$,

$$i = 0,\ldots,n, \quad \lambda_{-1} = 0, \quad \mu_{n+1} = 0.$$

Proof: Property (i) generalizes Example 1. Property (ii) is the classical condition for stability of a linear system. To show (iii) we apply the logarithmic norm estimate[7] to the linear system $(\partial/\partial t)M = (A + IV)M$ and get

$$\|M\| \leqslant \|M(0)\|e^{\nu(A+IV)t},$$

where

$$\|(b_0,\ldots,b_n)\| = \sum_{i=0}^n |b_i|,$$

$\nu(B)$ = the logarithmic norm of $(n+1)\times(n+1)$ matrix (b_{ij})

$$= \sup_{0 \leqslant 1 \leqslant n}\left[b_{kk} + \sum_{\substack{i=0 \\ i\neq k}}^n b_{i_k}\right],$$

and require $\nu(A + IV) < 0$, which is equivalent to (iii).

Remark 2: Property (i) extends to $n = \infty$ provided

$$\sum_{n=1}^\infty \frac{\lambda_0\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n} < \infty,$$

which ensures the existence of stationary measure m. Property (iii) can also be considered for $n = \infty$ provided that $(\partial M/\partial t) = (A + IV)M$ has a solution satisfying $\sum_{i=0}^\infty |M(x_i,t)| < \infty$ while (ii) extends with no additional assumptions. The next theorem discusses the case of a time-homogeneous diffusion process $X(t)$ with the generator $L = a(x)(\partial^2/\partial x^2) + b(x)(\partial^2/\partial x^2)$.

Theorem 2: The solution $Y(t)$ of $\dot{Y}(t) = V(X(t))Y(t)$ satisfying $Y(0) = y, X(0) = x$ is pathwise asymptotically stable, i.e., $Y(t) \rightarrow 0$ almost surely as $t \rightarrow \infty$ whenever the following conditions are met.

(i) There exists a positive eigenfunction $\psi(x) = e^{-h(x)}$ and a negative eigenvalue λ solving

$$(L + V)\psi = \lambda\psi.$$

(ii) $\lim_{t\rightarrow\infty} -h(X(t)+x)/t > \lambda$ almost surely.

Proof: Define $u(x,t) = u(x,0) = \psi(x)$, then

$$\frac{\partial u}{\partial t} = (L + V - \lambda)u,$$

whence by the FK formula

$$\psi(x) = E_x \psi(X(t))\exp\left(\int_0^t (V(X(s)) - \lambda)ds\right)$$

$$= E\psi(x + X(t))\exp\left(-\lambda t + \int_0^t V(x + X(s))ds\right).$$

On the other hand,

$$\exp[-h(x + X(t)) - \lambda t]\exp\left(\int_0^t V(x + X(s))ds\right)$$

is a positive martingale (see Ref. 8, p. 168) whose expected value is $\psi(x) < \infty$. Therefore it is convergent almost surely as $t \rightarrow \infty$ (see Ref. 2, p. 354). Since the exponential of the first factor tends by (ii) to ∞ a.s., therefore the second factor tends to 0 a.s., which concludes the proof.

Example 4: Let $X(t) = B(t)$ by a Brownian motion and assume that e^{-h} is the eigenfunction corresponding to the ground state of the Schrödinger equation, i.e.,

$$\left(-\frac{1}{2}\frac{d^2}{dx^2} - V(x)\right)e^{-h} = -Ee^{-h}$$

and

$$-E = \inf \operatorname{spectrum}_{L_2}\left(-\frac{1}{2}\frac{d^2}{dx^2} - V(x)\right) > 0.$$

If $-V(x) = ((h')^2 - h'') \sim |x|^\alpha$ for large x, for some $0 < \alpha < 2$, then $h(x) \sim |x|^{1+\alpha/2}$ (see Ref. 9, p. 487). Consequently one gets

$$(\sqrt{2t\log\log t})^{1+\alpha/2}\left|\frac{x + B(t)}{\sqrt{2t\log\log t}}\right|^{1+\alpha/2} t^{-1} \rightarrow 0 \quad \text{a.s.}$$

by the law of the iterated logarithm and thus $Y(t) \rightarrow 0$ a.s. by our theorem by taking $\lambda = E$ and noting that (ii) holds.

Remark 3: The harmonic oscillator, i.e., $V(x) = -kx^2$, cannot be handled by our theorem because then $h(x) = x^2$ and (ii) does not hold. However, in this case the exact solution is known:

$$u(x,t) = E_x \exp\left(-k\int_0^t B^2(s)ds\right)$$

$$= \frac{\exp\{-\sqrt{k/2}x^2 \tanh t\sqrt{2k}\}}{\sqrt{\cosh t\sqrt{2k}}}.$$

This satisfies

$$\frac{\partial u}{\partial t} = \left(\frac{1}{2}\frac{\partial^2}{\partial x^2} - kx^2\right)u,$$

and for $B(0) = 0$, $k = \frac{1}{2}$ reduces to

$$\frac{1}{\sqrt{\cosh t}} = E\exp\left(-\frac{1}{2}\int_0^t B^2(s)ds\right),$$

i.e., the one-dimensional Cameron–Martin formula (see Ref. 5, p. 261). Obviously we have here asymptotic stability in the mean.

ACKNOWLEDGMENT

The author would like to thank Professor G. S. Ladde for introducing him to the subject as well as discussions on stability.

[1]G. S. Ladde and D. D. Siljak, "Multiplex control systems: stochastic stability and dynamic reliability," Int. J. Control **38**, 515 (1983).
[2]J. Doob, *Stochastic Processes* (Wiley, New York, 1964), 5th ed.
[3]R. Z. Khasminski, *Stochastic Stability of Differential Equations* (Sijthoff and Nordhoff, Groningen, The Netherlands, 1980) (English translation).
[4]S. Karlin and H. M. Taylor, *A Second Course in Stochastic Processes* (Academic, New York, 1981).
[5]R. Durret, *Brownian Motion and Martingales in Analysis* (Wadsworth Mathematics Series, Belmont, CA, 1984).
[6]S. M. Ross, *Stochastic Processes* (Wiley, New York, 1983).
[7]G. S. Ladde, "Logarithmic norm and stability of linear systems with random parameters," Int. J. Systems Sci. **8**, 1057 (1977).
[8]P. Feinsilver, *An Introduction to Operator Calculus and Representation Theory* (to appear).
[9]B. Simon, "Schrödinger semigroup," Bull. Am. Math. Soc. **7**, 447 (1982).

Noncoordinated basis and Schild's solution

P. Palou, J. Stela, and L. Mas
Departament de Fisica Teorica, Facultat de Ciencies, Palma de Mallorca, Spain

(Received 8 March 1984; accepted for publication 19 April 1985)

In this paper we show the paradoxical consequences which appear in Schild's solution, as seen from a noninertial frame (NIF). We propose a treatment for the noninertial frame which resolves the ambiguities.

I. INTRODUCTION

Schild's solution for the electromagnetic two-body problem can be described from an inertial frame (IF) as two charges q_p and q_e which describe concentrical circular trajectories of radius a and b, at the same circular speed ω, so that the simultaneous positions seen from this frame are diametrically opposite. When the coordinates of one charge are $(x^0,a,0,0)$, the coordinates of the other one are $(x^0,-b,0,0)$. We take the center of the circumferences as the origin of space coordinates (Sec. II).

If this situation is observed from a rotating frame turning around the same axis and with the same angular speed as that of the two charges, then the description of the solution will be as follows: there are two electric charges at a distance $a+b$ from each other; they interact electrically and remain motionless without any constraint. This situation demands a consideration of the observed facts.

The incompatibility between Maxwell's equations which provide an electric field of Coulombian type and those of Lorentz, which would demand the absence of a field due to the static equilibrium, cannot be solved with Schild's hypothesis of the time-symmetric, half-advanced–half-retarded field (Sec. III).

To resolve this incompatibility the implicit anholonomicity at the basis linked to the noninertial frame (NIF) has to be taken into account in the particular way of calculating the four-potential and the electromagnetic tensor (Sec. IV).

It is convenient to emphasize that the location of advanced and retarded positions is very different in each frame. Even being stationary in the IF, the charge system does not have the property of maintaining the same point distribution of charges at any instant (this would happen if there were two charged rings instead of two point charges), but the recognized situation in the rotating frame is static.

II. SCHILD'S SOLUTIONS

Schild's solution[1] requires the accomplishment of the condition

$$\begin{aligned} -m_e v_e/\omega\sqrt{1-v_e^{2}} &= [q_e q_p/(\theta+v_e v_p \sin\theta)^3]\,[(v_e+v_p\cos\theta) \\ &\quad\times(1-v_e^{2})(1-v_p^{2})+(v_e\theta+v_p\sin\theta) \\ &\quad\times(\theta+v_e v_p\sin\theta)], \end{aligned} \tag{1}$$

and a formally identical expression which is obtained by changing the subindex e for p and vice versa, where $v_e=\omega\times b$, $v_p=\omega\times a$, m_e and m_p are the masses of the charges q_e and q_p, respectively, and θ is the advanced or retarded angle.

To arrive at this condition the following expression has been used for the four-potential[2]:

$$A_\mu{}^{\pm}(x)=\pm q_p\dot z^{\mu}/\rho|_{\psi_\pm=0}, \tag{2}$$

where z^μ represents the four-vector position of q_p and where $\psi_\pm=0$ is the advanced or retarded condition

$$\psi_\pm=|x^\mu-z^\mu|=a^2+b^2+2ab\cos\theta-\theta^2/\omega^2 \tag{3}$$

and ρ is the scalar

$$\rho=(x^\alpha-z^\alpha)\dot z_\alpha. \tag{4}$$

III. THE OBSERVER IN THE NONINERTIAL FRAME

In the NIF, the rotating basis associated with the particle system, obtained from Frenet's formulas, is related to the Cartesian one of the IF, where we describe the solution by

$$\mathbf{e}_a=h_a{}^{\mu}\mathbf{e}_\mu, \tag{5}$$

where Latin subindices indicate the basis in the NIF, Greek subindices indicate the basis on the IF, and where

$$h_a{}^{\mu}=\begin{pmatrix}\gamma & -\gamma\omega r\sin\varphi & \gamma\omega r\cos\theta & 0\\ 0 & \cos\varphi & \sin\varphi & 0\\ \gamma r^2\omega & -\gamma r\sin\varphi & \gamma r\cos\theta & 0\\ 0&0&0&1\end{pmatrix}. \tag{6}$$

The commutation coefficients for this basis are the same as used by Corum[3]

$$\Omega^{0}_{01}=-\tfrac{1}{2}r\omega^2\gamma^2, \tag{7a}$$

$$\Omega^{0}_{12}=\gamma^2 r\omega, \tag{7b}$$

$$\Omega^{2}_{12}=-\tfrac{1}{2}r\omega^2\gamma^2, \tag{7c}$$

according to the definition $[\mathbf{e}_a,\mathbf{e}_b]\equiv\nabla_{\mathbf{e}_a}\mathbf{e}_b-\nabla_{\mathbf{e}_b}\mathbf{e}_a=\Omega^{c}_{ab}\mathbf{e}_c$.

From the point of view of the NIF, and disregarding any kind of information about the noninertiality of the frame, an observer would deal with the following system of particles (which corresponds to Schild's solution for an observer that turns at the same angular speed):

$$x^a=(x^0,b,0,0),\quad z^a=(x^0,-a,0,0),$$

$$\dot x^a=(1,0,0,0),\quad \dot z^a=(1,0,0,0),$$

$$\ddot x^a=(0,0,0,0),\quad \ddot z^a=(0,0,0,0).$$

The advanced and retarded positions of the particle m_p are

$$z^a_\pm=(x^0-\epsilon\tau,-a,0,0),$$

where $\pm$ indicates the advanced or retarded position, and ϵ is a value factor of $+1$ for the advanced position and -1 for the retarded position, so that in the expression

$$F_{ab} = \tfrac{1}{2}(F^{+}_{ab} + F^{-}_{ab}),$$

terms with odd powers of ϵ will be eliminated. Here, τ is the time (in the NIF) which light takes for traveling the spatial distance $a+b$, and in units $c=1$, is just $a+b$.

So,

$$z^a_{\pm} = (x^0 - \epsilon(a+b), -a, 0, 0),$$
$$\dot{z}^a_{\pm} = (1,0,0,0).$$

The scalar ρ, which appears in the denominator of (2), in the NIF will not coincide with the same one of the IF, because it does not have the same meaning. Let us remember that the scalar ρ represents the spatial distance which separates the advanced and retarded positions of a body from another at a determinate instant. On the other hand, we can not apply the matrix $h_\mu{}^a$ to the vector $(x^\mu - z^\mu)$, because there is a double choice for the point where the value of the matrix terms have to be calculated.

The calculation of ρ in $x^0 = 0$ gives the result

$$\rho = \epsilon(a+b).$$

In this way we find that

$$A^{\pm}_a = \epsilon(q_p \dot{z}_{a\pm}/\rho) = \epsilon\delta^0_a[q_p/\epsilon(a+b)]. \tag{8}$$

Therefore,

$$A^{\pm}_0 = q_p/(a+b), \tag{9a}$$
$$A^{\pm}_i = 0 \quad (i=1,2,3). \tag{9b}$$

The four-potential obtained is that of Coulomb. The values of F_{ab} can, also, be calculated from

$$F^{\pm}_{ab} = \pm(q_p/\rho^3)[r_a\dot{z}_b - r_b\dot{z}_a)(1 - r^c\ddot{z}_c) + \rho(r_a\ddot{z}_b - r_b\ddot{z}_a)]_{\psi_\pm = 0}. \tag{10}$$

If we call $r^a = (x^a - z^a)$, then we have $(r_m\ddot{z}_n - r_n\ddot{z}_m) = 0$ for all n,m. The only F_{mn} not null will be those where $(r_m\dot{z}_n - r_n\dot{z}_m)$ is not null. Then we will only have terms with $m=0$ and $n=0$, as only component 0 is not null in $\dot{x}_m$ and $\dot{z}_m$. So the only F_{mn} not null are F_{01} and its opposite, because F_{mn} is an antisymmetric tensor

$$F^{\pm}_{01} = F_{01} = -q_p/(a+b)^2, \tag{11}$$

which is the electric field produced by a static charge q_p at a distance $a+b$.

The Lorentz equation for inertial frames

$$m_e\ddot{x}^a = q_e F^a_b \dot{x}^b \tag{12}$$

is erroneous with this expression for F_{ab}.

IV. THE CORRECT PLANNING IN THE NIF

The anholonomicity makes it necessary to add the terms of anholonomicity $2\Omega^a_{bc}$ to the F_{ab} values, calculated in (10). Thus,

$$\widehat{F}_{ab} = F_{ab} + 2\Omega^d_{ba}A_d, \tag{13}$$

but it has also to be kept in mind that the four-potential cannot be calculated as in an IF; it is necessary to transform it by matrices $h_\mu{}^a$ and to consider that scalars depend on the retarded time; somehow this fact has to be taken into account. Corum[4] presents the problem of the change of the IF to a rotating frame with cylindrical symmetry, which does not make it necessary to worry about what the situation is in the retarded instant because this situation is not only dynamically stationary but its aspect is the same in any instant.

The process we follow from an NIF makes it clear from the very start that the observer has deduced from the inconsistency of the former results that they are found in an NIF. That means that Maxwell's and Lorentz's equations are incompatible, in the sense that the expression (12) is not adequate for a noninertial frame.

At this stage we may say that this observer tries a possible description in his frame and that, as a test of validity, has to prove that the equations he obtains correpond to known solutions (in this case that of Schild[1]).

First we shall construct the matrix $h_\mu{}^a$ at any given point. We shall apply it to A^μ at the point z^μ, for any x^μ (remember that z^μ are the components of the space-time four-vector for m_p and x^μ those of m_e) and we shall have A^a as a function of the polar components of x^μ. Finally, we shall calculate F_{ab} deduced according to an anholonomic basis. We shall see that, if the terms of anholonomicity are added to the result, the expression (1) calculated from an IF is reproduced.

The matrix $h_a{}^\mu$ shall be (6) and $h_\mu{}^a$ its inverse; φ is equal to $\omega \cdot x^0$.

Now, r and φ are variable values in relation to which the derivatives will be calculated. We also assume that the denominator that we have in (2) is a scalar, and therefore it does not depend on the frame. A different value had been calculated beforehand because it had either forgotten or disregarded that we were in an NIF, and we had given to it a significance that it did not have. But it is not a constant scalar because it is a function of r or τ. From now on, τ is the time of advance or retard.

From (3) in $\psi_\pm = 0$, one obtains

$$\frac{\partial\tau}{\partial r} = \frac{r + a\cos\omega\tau}{\tau + rv_p\sin\omega\tau}. \tag{14}$$

We also have

$$\frac{\partial\gamma}{\partial r} = r\omega^2\gamma^3. \tag{15}$$

From (2) one obtains

$$A^\mu = \omega\dot{z}^\mu q_p/\gamma_p(\omega\tau + \omega^2 ar\sin\omega\tau)|_{\psi_\pm = 0}, \tag{16}$$

which, transformed by $h_\mu{}^a$, gives

$$A^a = h_\mu{}^a A^\mu, \tag{17}$$

where the Latin index in A^a indicates components of the four-potential in the directions of the anholonomic tetrad basis, that is to say, the projections of the four-potential which, in this way, are referred to basis (5).

These projections are

$$A^0 = q_p\gamma(1 + v_p\omega r\cos\omega\tau)/(\tau + \omega ar\sin\omega\tau), \tag{18}$$
$$A^1 = -\epsilon q_p v_p \sin\omega\tau/(\tau + \omega ar\sin\omega\tau), \tag{19}$$
$$A^2 = -q_p\gamma(\omega r + v_p\cos\omega\tau)/r(\tau + \omega ar\sin\omega\tau), \tag{20}$$
$$A^3 = 0. \tag{21}$$

Let us remember that the variation in x^3 has not been taken into account because the solution is coplanar by hypothesis, but also and especially because, since $\dot{x}^3 = 0$ and $\ddot{x}^3 = 0$, no $F_{\mu 3}$ will appear in the equations of motion.

Now, if we want to write $A_0, A_1, A_2 (A_3 = 0)$ we will take

$$A_a = g_{ab} A^b , \tag{22}$$

with the metric of this NIF, which is[3]

$$g_{ab} = \mathrm{diag}(1, -1, -r^2, -1), \tag{23}$$

and therefore

$$A_0 = A^0, \quad A_1 = -A^1, \quad A_2 = -r^2 A^2. \tag{24}$$

To facilitate the calculation of derivatives, we first calculate the derivative of the present expression for A^a with respect to r

$$\frac{\partial}{\partial r}(\tau + \omega ar \sin \omega\tau)^{-1} = -(\tau + \omega ar \sin \omega\tau)^{-3} \times [(r + a \cos \omega\tau)(1 + \omega^2 ar \cos \omega\tau) + \omega a \sin \omega\tau(\tau + \omega ar \sin \omega\tau)]. \tag{25}$$

Now let us calculate derivatives of A; if we now situate ourselves on the point $r = b$, $\gamma = \gamma_e$, and $\omega\tau = \theta$, then

$$A_{0,4} = \omega^2 q_p \gamma_e \rho^{-3} [v_e \rho^2 \gamma_e{}^2 (1 + v_e v_p \cos \theta) + \rho^2 v_p \cos \theta - \rho v_e v_p \times \sin \theta (v_e + v_p \cos \theta) - (1 + v_e v_p \cos \theta) \times \{(v_e + v_p \cos \theta)(1 + v_e v_p \cos \theta) + \rho v_p \sin \theta\}]. \tag{26}$$

On the other hand,

$$A_{1,0} = 0. \tag{27}$$

Therefore,

$$\widehat{F}_{01} = -A_{0,1} + 2\Omega^{a}_{10} A_a . \tag{28}$$

The only Ω^{a}_{10} non-null is the opposite to (7a) and so

$$\widehat{F}_{01} = q_p \gamma_e \omega^2 \rho^{-3} [-v_e \rho^2 \gamma_e{}^2 (1 + v_e v_p \cos \theta) - \rho^2 v_p \cos \theta + \rho v_e v_p \sin \theta (v_e + v_p \cos \theta) + (1 + v_e v_p \cos \theta)^2 (v_e + v_p \cos \theta) + v_p \rho \sin \theta (1 + v_e v_p \cos \theta) + \gamma_e^2 v_e \rho^2 (1 + v_e v_p \cos \theta)]. \tag{29}$$

The term suppressed by the presence of Ω^{0}_{10} is that one which came from deriving γ with respect to r.

Since $A_{2,0} = A_{0,2} = A_{2,3} = A_{3,2} = 0$, we have $\widehat{F}_{02} = \widehat{F}_{03} = 0$ and also $\widehat{F}_{23} = 0$. Finally, $\widehat{F}_{12}$ is

$$\widehat{F}_{12} = A_{2,1} - A_{1,2} + 2\Omega^{0}_{21} A_0 + 2\Omega^{2}_{21} A_2, \tag{30}$$

but it is not necessary to develop it further as it does not appear in the equations of motion.

Equations (12) are the following:

for $a = 0$,

$m_e \ddot{x}^0 = q_e \widehat{F}^0{}_1 \dot{x}^1$, identically null,

for $a = 1$,

$$-m_e \ddot{x}^1 = q_e \widehat{F}^1{}_0 \dot{x}^0 + q_e \widehat{F}^1{}_2 \dot{x}^2, \tag{31}$$

for $a = 2$,

$m_e \ddot{x}^2 = q_e \widehat{F}^2{}_1 \dot{x}^1$, identically null,

but (31) being $\dot{x}^2 = 0$, they are reduced to

$$-m_e \gamma_e^{22} v_e \omega = q_e \widehat{F}^1{}_0 \dot{x}^0, \tag{32}$$

or

$$-m_e v_e / \omega\sqrt{1 - v_e{}^2} = [q_p q_e / (\theta + v_e v_p \sin \theta)^3][\theta v_p (\sin \theta - \theta \cos \theta) + v_p \cos \theta (1 + v_e{}^2 + v_e v_p \cos \theta) + v_e{}^2 v_p \cos \theta + v_e + v_e v_p{}^2 (1 + v_e{}^2 + v_e v_p \cos \theta) + v_e{}^2 v_p \theta \sin \theta], \tag{33}$$

which is exactly (1).

This shows that

$$\widehat{F}^1{}_0 = (F^1{}_0 + F^1{}_2 v_e)\gamma_e, \tag{34}$$

so we see that, in the NIF, the same relation between the parameters of the problem contemplated on the IF can be obtained.

Adding the terms of anholonomicity amounts to including two components of $F^\mu{}_\nu$ in one component of $F^a{}_b$. This is justified if we keep in mind that in the frame where the two charges remain static, there is not only an electric field in the direction of the segment which joins the charges ($\widehat{F}^1{}_0$), but actual components of the electromagnetic field as well: the components $F^1{}_0$ (electric field) and $F^1{}_2$ (magnetic field, on the term $F^1{}_2 v_e$) of the tensor $F^\mu{}_\nu$ calculated in the IF where the charges are in motion.

Therefore, we have followed the following steps.

We have seen that the inertial form of Lorentz's and Maxwell's equations are incompatible.

We have tried a possible description in the reference frame where we are and verified if the equations correspond to the well-known solution in the IF.

We have constructed the matrix $h_a{}^\mu$ for any given point.

We have transformed the four-potential.

The four-potential and so the electromagnetic field tensor in the NIF are obtained; therefore, the derivatives are taken accordingly with the anholonomic basis which we had chosen.

We have fixed point values.

We have added to F_{ab} values the terms of anholonomicity which correspond to the basis used.

We have verified that the equations of motion are compatible with well-known solutions.

V. ENERGY OF THE SYSTEM

We can observe that in the NIF the energy corresponding to the two masses m_e and m_p of charges q_e and q_p, at rest and separated by a distance $a + b$, is

$$P^a = [m_e \dot{x}^a + q_e A^a]_A + [m_p \dot{z}^a + q_p \overline{A}^a]_{\overline{A}} + 2q_e q_p \times \left[\int_A^\infty \int_{-\infty}^{\overline{A}} - \int_{-\infty}^{A} \int_{\overline{A}}^{\infty}\right] \times \delta'[(x - z)^2](x^a - z^a)\dot{x}^b \dot{z}_b \, ds \, d\overline{s}, \tag{35}$$

according to[1] and in which $\overline{A}^a$ represent the four-potential on q, and A and $\bar{A}$ are, respectively, points of the line of the universe of q_e and q_p. Thus,

$$\dot{x}^b\dot{z}_b = 1,$$

and on the other hand,

$$\int \delta'(x)f(x)dx = f'(0),$$

they cause the integral term to be

$$\left.\frac{d(x^a - z^a)}{dx^0}\right|_{\psi_\pm = 0} = \dot{x}^a - \dot{z}^a = 0.$$

One can obtain

$$P^0 = m_e + m_p + 2q_e q_p/(a+b), \quad P^1 = P^2 = P^3 = 0, \tag{36}$$

from (8).

But if we take into account the effects of noninertiality, even though the integral term is not modified by this consideration, we will have the following result: $\bar{A}^a = h_\mu{}^a \bar{A}^\mu$,

$$\bar{A}^0 = \omega q_e \gamma_p (1 + v_e v_p \cos\theta)/(\theta + v_e v_p \sin\theta),$$

$$\bar{A}^1 = f(\epsilon^1), \quad A^1 = f(\epsilon'),$$

and therefore they do not take part in the time-symmetric case

$$\bar{A}^2 = -\omega^2 q_e \gamma_p (v_p + v_e \cos\theta)/v_p(\theta + v_e v_p \sin\theta).$$

Thus,

$$P^0 = m_e + m_p + \frac{\omega q_e q_p (1 + v_e v_p \cos\theta)}{\theta + v_e v_p \sin\theta}(\gamma_e + \gamma_p)$$

$$= m_e + m_p + L^{12}\omega(\gamma_e + \gamma_p), \tag{37}$$

$$P^2 = \frac{-\omega^2 q_e q_p}{v_e v_p}\cdot\frac{v_e v_p(\gamma_e + \gamma_p) + \cos\theta\,(v_p{}^2\gamma_e + v_e{}^2\gamma_p)}{\theta + v_e v_p \sin\theta},$$

$$P^3 = 0,$$

where L^{12} is the non-null component of the angular momentum in the IF according to Schild.[1]

VI. CONCLUSION

We have seen a possible treatment in an NIF for a problem in special relativity without making use of an infinite series of inertial frames. Formally the expressions are the same; noninertiality is deeply related to the concept of an anholonomic basis and inertial forms of Lorentz's and Maxwell's equations are not modified. Nevertheless, in the Maxwell equations the anholonomicity terms are explicit, a fact unusual in special relativity because inertial frames are in general more convenient and in these frames the use of a coordinate basis makes null the anholonomicity terms.

The calculation of the energy in the NIF also explains the noninertiality effects due to P^2 being non-null and to the term proportional to $L^{12}\omega$ which appears in P^0.

ACKNOWLEDGMENT

The authors have been supported partly through funds provided by the Comision Asesora de Investigacion Cientifica y Tecnica.

[1]A. Schild, "Electromagnetic two body problem," Phys. Rev. **131**, 2765 (1963).

[2]J. L. Anderson, *Principles on Relativity Physics* (Academic, New York, 1967), p. 219.

[3]J. F. Corum, "Relativistic rotation and the anholonomic object," J. Math. Phys. **18**, 770 (1977).

[4]J. F. Corum, "Relativistic covariance and rotational electrodynamics," J. Math. Phys. **21**, 2360 (1980).

Splitting methods for time-independent wave propagation in random media

J. W. Evans
Applied Mathematical Sciences, Ames Laboratory, Iowa State University, Ames, Iowa 50011

(Received 7 February 1984; accepted for publication 2 May 1985)

Time-independent wave propagation is treated in media where the index of refraction contains a random component, but its mean is invariant with respect to translation in some direction distinguishing the wave propagation. Abstract splitting operators are used to decompose the wave field into forward and backward traveling components satisfying a coupled pair of equations. Mode-coupled equations follow directly from these after implementing a specific representation for the abstract splitting operators. Here we indicate a formal solution to these equations, concentrating on the diffusion regime, where we estimate the forward- and backscattering contributions to the mode specific diffusion coefficients. We consider, in detail, random media with uniform (random atmosphere) and square law (stochastic lense) mean refractive indices.

I. INTRODUCTION

Time-independent scalar wave propagation in random media can be described by the 3-D Helmholtz equation with a random-valued refractive index $n(\mathbf{x})$. [Throughout we suppress the dependence of random-valued functions, e.g., $n(\mathbf{x})\equiv n(\mathbf{x},\omega)$, on the probability space variable ω.] We concentrate on propagation in a distinguished direction, chosen along the x axis in a Cartesian coordinate system, (x_1,x_2,x_3), with $(x_1,x_2)\equiv\mathbf{x}_\perp$, $x_3\equiv x$, for the case where the mean of the refractive index, $\langle n(\mathbf{x})\rangle\equiv\bar{n}(\mathbf{x}_\perp)$, is independent of x. (Here $\langle\ \rangle$ denotes the average over the statistical ensemble.) In particular, we are interested in the random atmosphere (where $\bar{n}$ is constant), and stochastic lenses [where $\bar{n}(\mathbf{x}_\perp)$ increases from large $|\mathbf{x}_\perp|$ asymptotic value(s) to a maximum near $\mathbf{x}_\perp=\mathbf{0}$]. The Helmholtz equation is naturally written here as

$$\frac{d^2}{dx^2}\Psi+S\Psi=0, \tag{1.1}$$

where $S=S(\mathbf{x})\equiv\Delta_\perp+k^2(\mathbf{x})$. Here $\Delta_\perp=\partial^2/\partial\mathbf{x}_\perp^2$ is the transverse Laplacian, and $k(\mathbf{x})\equiv kn(\mathbf{x})$ is the position-dependent wave number, where $k>0$ is arbitrary. We have in mind cases where the random fluctuations, $\beta(\mathbf{x})\equiv k(\mathbf{x})^2-\bar{k}(\mathbf{x}_\perp)^2$ with $\bar{k}(\mathbf{x}_\perp)=k\bar{n}(\mathbf{x}_\perp)$, are "small," and $\beta(\mathbf{x})$ is a stationary process so one has $\langle\beta(\mathbf{x})\beta(\mathbf{x}+\delta\mathbf{x})\rangle=R(\delta\mathbf{x})$, where $R(\mathbf{y})=R(-\mathbf{y})$.

The approach implemented here is to somewhat arbitrarily split the wave field Ψ into right (x-increasing), Ψ^+, and left (x-decreasing), Ψ^-, propagating components that satisfy a coupled set of equations. Such procedures are useful in relating various unidirectional propagation approximations, associated with the zeroth-order decoupled solutions, to the exact solutions expressed as Bremmer-type series.[1] Here we implement "reference" splitting in terms of naturally chosen deterministic, unbounded, abstract, self-adjoint operators (cf. Refs. 2–4) such that Ψ^+ and Ψ^- are decoupled when the random fluctuations are set to zero. In Sec. II, we detail this procedure and rearrange the coupled $\Psi^\pm$ equations so that a *first-order smoothing approximation*[5] can be conveniently applied to obtain a closed equation for the mean, $\langle\Psi^+\rangle$, of Ψ^+. Using the representation provided by the splitting operator eigenfunctions, one obtains explicit mode-coupled equations. Since we are primarily interested here in the slow decay of $\langle\Psi^+\rangle$ induced by small-amplitude stochastic fluctuations in $n(x)$ (the *diffusion regime*), a *long-range Markovian approximation* is applied to these to obtain estimates of the forward- and backscattering contributions to the mode-specific decay rates (termed, here, diffusion coefficients).[6] We continue to detail specific applications of this general procedure to the random atmosphere in Sec. III, and to the stochastic square law medium (lense) in Sec. IV. Concluding remarks are made in Sec. V.

II. REFERENCE SPLITTING APPLIED TO RANDOM MEDIA

Reference splitting[3,4] of the wave field Ψ into right, Ψ^+, and left, Ψ^-, traveling components with respect to an x-independent operator, S_0, on $L^2(\mathbf{x}_\perp)$ (described in detail below), is given by

$$\Psi^\pm(x)=\frac{1}{2}\left[\Psi(x)\mp\hat{i}S_0^{-1/2}\frac{d}{dx}\Psi(x)\right] \tag{2.1}$$

(where here, and in the following, we suppress all $\mathbf{x}_\perp$ dependence). Note that $\Psi\equiv\Psi^++\Psi^-$, and that the $\Psi^\pm$ satisfy the coupled set of equations

$$\frac{d}{dx}\Psi^\pm\mp iS_0^{1/2}\Psi^\pm=\pm(\hat{i}/2)S_0^{-1/2}(S(x)-S_0)(\Psi^++\Psi^-). \tag{2.2}$$

Naturally, here, S_0 is chosen to be deterministic (i.e., nonrandom). A scalar choice $S_0=\bar{k}(\mathbf{0})^2$, after ignoring $\pm$ coupling in (2.2), produces a parabolic-type approximation.[1] Equations (2.2) could be iterated about this approximation but instead we prefer to start with a "more complete," but abstract, choice of splitting associated with

$$S_0=\langle S(x)\rangle=\Delta_\perp+\bar{k}(\mathbf{x}_\perp)^2, \tag{2.3}$$

for which $S(x)-S_0\equiv\beta(x)$, i.e., the random fluctuations. Notice that S_0 is a deterministic, unbounded, self-adjoint operator on the space $L^2(\mathbf{x}_\perp)$. Its spectral theory for uniform or focusing media is naturally represented in terms of an (assumed) complete set of guided-mode eigenfunctions, and radiation-mode "weak" eigenfunctions.[4] For a random at-

 0022-2488/85/092196-05$02.50

mosphere, where $\bar{k}$ is constant, one simply takes the Fourier transform with respect to $\mathbf{x}_\perp$ in (2.2) and (2.3), corresponding to expanding with respect to a complete set of (trivial) transverse plane wave radiation-mode eigenfunctions. It is convenient to introduce a generic mode label κ for the S_0 eigenfunctions $\Psi_\kappa(\mathbf{x}_\perp)$, with S_0 eigenvalues λ_κ. Also $\sum\!\!\!\!\int d\kappa$ will denote a sum/integral over eigenmodes.

Here we treat only the explicit choice of boundary conditions: $\Psi^+(x=0)$ specified and deterministic, and $\Psi^-(x=\infty)=0$. Equation (2.2) is then readily integrated, and expressed in abstract operator form, as

$$\Psi^+ = \phi^+ + (\hat{i}/2)S_0^{-1/2}G_0^+\,\beta(\Psi^+ + \Psi^-)\,, \tag{2.4a}$$

$$\Psi^- = (\hat{i}/2)S_0^{-1/2}G_0^-\,\beta(\Psi^+ + \Psi^-)\,, \tag{2.4b}$$

where

$$\phi^+(x) = \exp[\hat{i}S_0^{1/2}x]\,\Psi^+(x=0)\,,$$

$$(G_0^+ f)(x) = \int_0^x dx' \exp[\hat{i}S_0^{1/2}(x-x')]\,f(x')\,,$$

and

$$(G_0^- f)(x) = \int_x^\infty dx' \exp[\hat{i}S_0^{1/2}(x'-x)]\,f(x')\,.$$

Equation (2.4b) can be solved formally for Ψ^- to give

$$\Psi^- = (\hat{i}/2)S_0^{-1/2}G_0^-\,\beta\,(1-(\hat{i}/2)S_0^{-1/2}G_0^-\,\beta)^{-1}\Psi^+\,. \tag{2.5}$$

If (2.5) is substituted into (2.4a), this results in a closed equation for Ψ^+, which, after some simplification, reduces to

$$\Psi^+ = \phi^+ + (\hat{i}/2)S_0^{-1/2}G_0^+\,\beta \times (1-(\hat{i}/2)S_0^{-1/2}G_0^-\beta)^{-1}\Psi^+\,. \tag{2.6}$$

Equation (2.6) could be iterated to obtain Ψ^+ in terms of the boundary condition ϕ^+, but since β is not spatially confined, the expansion contains secular terms (with respect to x).[5] Since we are not primarily interested in short distances $x=O(1)$, but rather $\beta^2x=O(1)$, for $\beta \ll 1$ (the diffusion regime[6]), a resummation of this series is required.[5] However, the lowest order of the resummed expression for $\langle\Psi^+\rangle$, the mean of Ψ^+, can be obtained, in a certain statistical approximation, more succinctly by a different approach indicated below.

We start with the differential form of (2.6), which is obtained by substituting (2.5) and (2.6) into (2.2) for Ψ^+. After taking the mean of this equation, one obtains

$$\begin{aligned}\frac{d}{dx}\big(e^{-\hat{i}S_0^{1/2}x}\langle\Psi^+(x)\rangle\big) &= -\tfrac{1}{4}S_0^{-1/2}e^{-\hat{i}S_0^{1/2}x}\langle\beta(x)S_0^{-1/2}[(G_0^+ + G_0^-)\beta(1-(\hat{i}/2)S_0^{-1/2}G_0^-\,\beta)^{-1}\Psi^+](x)\rangle\\ &= -\tfrac{1}{4}S_0^{-1/2}e^{-\hat{i}S_0^{1/2}x}\langle\beta(x)S_0^{-1/2}\,[(G_0^+ + G_0^-)\beta\Psi^+](x)\rangle + O(\beta^3)\,.\end{aligned} \tag{2.7}$$

This, of course, is not a closed equation for $\langle\Psi^+(x)\rangle$, but such can be provided by making a *first-order smoothing approximation*,[5] which replaces Ψ^+ on the right-hand side (rhs) by $\langle\Psi^+\rangle$. One then obtains, for the "slowly varying," nonrandom function $u^+(x)=e^{-\hat{i}S_0^{1/2}x}\langle\Psi^+(x)\rangle$, the equation

$$\begin{aligned}\frac{d}{dx}u^+(x) = &-\frac{1}{4}S_0^{-1/2}\int_0^x dx'\langle[e^{-\hat{i}S_0^{1/2}x}\beta(x)e^{+\hat{i}S_0^{1/2}x}]S_0^{-1/2}[e^{-\hat{i}S_0^{1/2}x'}\beta(x')e^{+\hat{i}S_0^{1/2}x'}]\rangle u^+(x')\\ &-\frac{1}{4}S_0^{-1/2}\int_x^\infty dx'\langle[e^{-\hat{i}S_0^{1/2}x}\beta(x)e^{-\hat{i}S_0^{1/2}x}]S_0^{-1/2}[e^{+\hat{i}S_0^{1/2}x'}\beta(x')e^{+\hat{i}S_0^{1/2}x'}]\rangle u^+(x') + O(\beta^3)\,.\end{aligned} \tag{2.8}$$

The second term in (2.8) can be thought of as a backscattering correction to the first, forward-scattering term. It is a straightforward matter, using (2.7), to write out explicitly higher-order terms, but here we concentrate on the $O(\beta^2)$ ones.

To proceed further, it is convenient to express (2.8) in the natural representation provided by the S_0 eigenfunctions, $\Psi_\kappa(\mathbf{x}_\perp)$. One then obtains a coupled set of equations for the corresponding components, $u_\kappa^+(x)$, of u^+, where $u^+(x)=\sum\!\!\!\!\int d\kappa\, u_\kappa^+(x)\,\Psi_\kappa(\mathbf{x}_\perp)$. Specifically, one has

$$\begin{aligned}\frac{d}{dx}u_\kappa^+(x) = &-\frac{1}{4}\lambda_\kappa^{-1/2}\sum\!\!\!\!\!\!\int d\kappa'\sum\!\!\!\!\!\!\int d\kappa''\,\lambda_{\kappa'}^{-1/2}\\ &\times\Big\{\int_0^x dx'\exp[\hat{i}(\lambda_{\kappa'}^{1/2}-\lambda_\kappa^{1/2})(x-x')]\,\bullet\\ &+\int_x^\infty dx'\exp[\hat{i}(\lambda_{\kappa'}^{1/2}+\lambda_\kappa^{1/2})(x-x')]\,\bullet\Big\}\\ &\times R_{\kappa,\kappa',\kappa''}(|x-x'|)\exp[\hat{i}(\lambda_{\kappa''}^{1/2}-\lambda_\kappa^{1/2})x']\\ &\times u_{\kappa''}^+(x') + O(\beta^3)\,,\end{aligned} \tag{2.9}$$

where the correlation functions $R_{\kappa,\kappa',\kappa''}(|\delta x|)\equiv\langle\beta_{\kappa\kappa'}(x)\times\beta_{\kappa'\kappa''}(x+\delta x)\rangle$, are $O(\beta^2)$, $\beta_{\kappa,\kappa'}=\int d\mathbf{x}_\perp\,\Psi_\kappa^*\beta\Psi_{\kappa'}$, and we have exploited the stationarity of β. In cases of interest here, there will be coupling of propagating ($\lambda_\kappa>0$) to evanescent ($\lambda_\kappa<0$) modes. For the latter, we must use the convention, $\lambda_\kappa^{1/2}=\hat{i}|\lambda_\kappa|^{1/2}$, to ensure that the evanescent components of Ψ^+ are exponentially decreasing to the right. Henceforth, we implicitly concentrate on propagating modes assumed to be far removed from the cutoff mode κ_c, where $\lambda_{\kappa_c}=0$.

The details of the manipulations, from here on, depend on the specific $\tilde{n}$ for the medium under consideration. For the random atmosphere, we show that $R_{\kappa,\kappa',\kappa''}\propto\delta_{\kappa,\kappa''}$. For the stochastic lense, we pick boundary conditions so that $u^+(0)$ corresponds to a single low-order propagating guided mode κ^*. Then, in the $u_{\kappa^*}^+$ equation (2.9), we neglect $O(\beta^3)$ terms, which include coupling terms $\kappa''\neq\kappa^*$, since an iterative solution of (2.9) shows that such $u_{\kappa''}^+$ are $O(\beta^2)$. Thus, in both cases, we shall be dealing with a decoupled form of (2.9), where $\kappa''=\kappa$ [and we shall neglect $O(\beta^3)$ terms].

Such decoupling affects the important simplification of reducing the kernel of (2.9) to the *convolution type*. If we also neglect the backscattering term $\int_x^\infty dx'$, then Laplace transform techniques can be used to simply solve the resulting equation. If the backscattering term is retained, then Weiner–Hopf techniques[7] may be useful, assuming sufficiently fast decay of the correlations. Here, however, we concentrate on the diffusion regime $\beta \ll 1$, $\beta^2 x = O(1)$, where, clearly, u_κ^+ will vary little over the characteristic $O(1)$ correlation length(s) of the $R_{\kappa,\kappa',\kappa}$. This motivates the *long-range Markovian approximation* to the decoupled form of (2.9), which replaces $u_\kappa^+(x')$, on the rhs, with $u_\kappa^+(x)$. Neglecting $O(\beta^3)$ terms, and letting $x \to \infty$ in evaluating the coefficient on the rhs [a good approximation, since $x = O(\beta^{-2})$], yields a "corrected" diffusion approximation

$$u_\kappa^+(x) \approx \exp(-\gamma_\kappa x) u_\kappa^+(0), \tag{2.10}$$

and the mode κ diffusion coefficient γ_κ can be decomposed as the sum, $\gamma_\kappa = \gamma_\kappa^f + \gamma_\kappa^b$, of a forward-scattering part

$$\gamma_\kappa^f = \tfrac{1}{4} \lambda_\kappa^{-1/2} \sum\!\!\!\!\!\!\int d\kappa' \, \lambda_{\kappa'}^{-1/2} \int_0^\infty dx \, e^{\hat{i}(\lambda_{\kappa'}^{1/2} - \lambda_\kappa^{1/2})x} R_{\kappa,\kappa',\kappa}(x), \tag{2.11a}$$

and a backscattering correction

$$\gamma_\kappa^b = \tfrac{1}{4} \lambda_\kappa^{-1/2} \sum\!\!\!\!\!\!\int d\kappa' \, \lambda_{\kappa'}^{-1/2} \int_0^\infty dx \, e^{\hat{i}(\lambda_{\kappa'}^{1/2} + \lambda_\kappa^{1/2})x} R_{\kappa,\kappa',\kappa}(x). \tag{2.11b}$$

Often in forward propagation approximations, one is interested in the short-wavelength (high "k") regime. Here we elucidate the corresponding behavior of the low-mode (high S_0 eigenvalue) diffusion coefficients. Let us assume that $\lambda_\kappa = \bar{n}_m^2 k^2 + \alpha_\kappa k + \beta_\kappa + O(1/k)$, as $k \to \infty$, where α_κ, $\beta_\kappa = O(1)$, and $\bar{n}_m \equiv \max \bar{n}(\mathbf{x}_\perp)$ (noting that the α_κ are identically zero for the random atmosphere case). Then one has

$$\lambda_{\kappa'}^{1/2} - \lambda_\kappa^{1/2} = \frac{\alpha_{\kappa'} - \alpha_\kappa}{2\bar{n}_m} + \frac{1}{2\bar{n}_m k} \times \left[\beta_{\kappa'} - \beta_\kappa + \frac{\alpha_{\kappa'}^2 - \alpha_\kappa^2}{4\bar{n}_m^2} \right] + O\left(\frac{1}{k^2}\right), \tag{2.12a}$$

and

$$\lambda_{\kappa'}^{1/2} + \lambda_\kappa^{1/2} = 2\bar{n}_m k + \frac{\alpha_{\kappa'} - \alpha_\kappa}{\bar{n}_m} + O\left(\frac{1}{k}\right), \tag{2.12b}$$

so

$$\int_0^\infty dx \, e^{\hat{i}(\lambda_{\kappa'}^{1/2} + \lambda_\kappa^{1/2})x} R_{\kappa,\kappa',\kappa}(x) \sim \frac{R_{\kappa,\kappa',\kappa}(0)}{2\hat{i}\bar{n}_m k} + O\left(\frac{1}{k^2}\right), \tag{2.13}$$

and, thus, one concludes that

$$\gamma_\kappa^b / \gamma_\kappa^f = O(1/k), \quad \text{as } k \to \infty. \tag{2.14}$$

III. THE RANDOM ATMOSPHERE

For the random atmosphere $\bar{k}$ is constant, and since $S_0 = \Delta_\perp + \bar{k}^2$, it is natural to apply to (2.7) and (2.8), the transverse Fourier transform

$$\hat{f}(x, \mathbf{p}) = \frac{1}{2\pi} \int d\mathbf{x}_\perp e^{\hat{i}\mathbf{p}\cdot\mathbf{x}_\perp} f(\mathbf{x}), \tag{3.1}$$

i.e., to use the representation based on the S_0-plane wave eigenfunctions, $\Psi_\mathbf{p}(\mathbf{x}_\perp) = (2\pi)^{-1} e^{-\hat{i}\mathbf{p}\cdot\mathbf{x}_\perp}$, with eigenvalues $\lambda_\mathbf{p} = \bar{k}^2 - p^2$, where $p^2 = \mathbf{p}\cdot\mathbf{p}$. Note that $\widehat{(F(S_0)f)} = F(\bar{k}^2 - p^2)\hat{f}$, and that the Fourier transform produces convolution integrals associated with the products $\cdots\beta\cdots\beta\cdots u^+$ [represented by the $\sum\!\!\!\!\!\!\int$ sum/integrals in (2.9)]. Stationarity of $\beta(\;)$, here, implies that

$$R_{\mathbf{p},\mathbf{p}',\mathbf{p}''}(|x - x'|) = \langle \hat{\beta}(x, \mathbf{p} - \mathbf{p}') \hat{\beta}(x', \mathbf{p}' - \mathbf{p}'') \rangle = 2\pi\delta(\mathbf{p} - \mathbf{p}'') \hat{R}(x - x', \mathbf{p} - \mathbf{p}'), \tag{3.2}$$

where $\delta(\;)$ is the Dirac delta function, and this results in the following simplified form for (2.9):

$$\begin{aligned}
&-8\pi(\bar{k}^2 - p^2)^{1/2} \frac{d}{dx} u^+(x, \mathbf{p}) \\
&= \int_0^x dx' \left[\int d\mathbf{p}' (\bar{k}^2 - p'^2)^{-1/2} \exp\{\hat{i}[(\bar{k}^2 - p'^2)^{1/2} - (\bar{k}^2 - p^2)^{1/2}](x - x')\} \hat{R}(x - x', \mathbf{p} - \mathbf{p}') \right] u^+(x', \mathbf{p}) \\
&+ \int_x^\infty dx' \int d\mathbf{p}' \left[(\bar{k}^2 - p'^2)^{-1/2} \exp\{\hat{i}[(\bar{k}^2 - p'^2)^{1/2} + (\bar{k}^2 - p^2)^{1/2}](x' - x)\} \hat{R}(x - x', \mathbf{p} - \mathbf{p}') \right] \\
&\times u^+(x', \mathbf{p}) + O(\beta^3).
\end{aligned} \tag{3.3}$$

We remark that the exact decoupling, with respect to $\mathbf{p}$, manifested in (3.3) also occurs for all higher-order terms obtained from a formal expansion of (2.7). It suffices to observe that the integrand of the multiple convolution integral terms in this expansion involve factors

$$\begin{aligned}
\left\langle \prod_{i=1}^n \hat{\beta}(x_i, \mathbf{p}_i - \mathbf{p}_{i+1}) \right\rangle
&= 2\pi\delta(\mathbf{p}_1 - \mathbf{p}_{n+1}) \hat{R}(x_2 - x_1, \mathbf{p}_2 - \mathbf{p}_3; \\
&\qquad x_3 - x_1, \mathbf{p}_3 - \mathbf{p}_4; \\
&\qquad \ldots; x_n - x_1, \mathbf{p}_n - \mathbf{p}_{n+1}),
\end{aligned} \tag{3.4}$$

where

$$\left\langle \prod_{i=1}^n \beta(\mathbf{x}_i) \right\rangle \equiv R(\mathbf{x}_2 - \mathbf{x}_1; \mathbf{x}_3 - \mathbf{x}_1; \ldots; \mathbf{x}_n - \mathbf{x}_1)$$

using stationarity of $\beta(\;)$, and $\hat{R}$ denotes the $(n-1)$-fold multiple Fourier transform of R with respect to the transverse variables. Of course, the delta-function factor $\delta(\mathbf{p}_1 - \mathbf{p}_{n+1})$ is responsible for the above-mentioned decoupling.

The long-range Markovian approximation to (3.3) has the solution [cf. (2.10)]

$$u^+(x, \mathbf{p}) \approx \exp[-(\gamma_\mathbf{p}^f + \gamma_\mathbf{p}^b)x] u^+(0, \mathbf{p}), \tag{3.5}$$

where the forward- (back-) scattering contributions, $\gamma_\mathbf{p}^f(\gamma_\mathbf{p}^b)$, to the diffusion coefficients, for wave number $\mathbf{p}$, are obtained simply from (3.3) by replacing $u^+(x', \mathbf{p})$ with $u^+(x, \mathbf{p})$, and letting $x \to \infty$ in the coefficient $\int_0^x dx' \cdots (\int_x^\infty dx' \cdots)$. Explicitly, one finds that

$$4\bar{k}(\bar{k}^2-p^2)^{1/2}\gamma_{\mathbf{p}}^f = \int_0^\infty dy\left(\frac{1}{2\pi}\int d\mathbf{q}\right)\exp\left(\frac{\hat{i}y\mathbf{p}\cdot\mathbf{q}}{\bar{k}}\right)$$
$$\times\left[\left(1-\frac{|\mathbf{p}-\mathbf{q}|^2}{\bar{k}^2}\right)^{-1/2}\exp\left\{-\hat{i}\left(\frac{q^2}{2\bar{k}}+O\left(\frac{1}{\bar{k}^3}\right)\right)y\right\}\hat{R}(y,\mathbf{q})\right]$$
$$\equiv\int_0^\infty dy\, R^f_{\mathbf{p},\bar{k}}\left(y,\frac{-y\mathbf{p}}{\bar{k}}\right), \tag{3.6a}$$
$$4\bar{k}(\bar{k}^2-p^2)^{1/2}\gamma_{\mathbf{p}}^b = \int_0^\infty dy\, e^{2\hat{i}\bar{k}y}\left(\frac{1}{2\pi}\int d\mathbf{q}\right)\exp\left(\frac{\hat{i}y\mathbf{p}\cdot\mathbf{q}}{\bar{k}}\right)$$
$$\times\left[\left(1-\frac{|\mathbf{p}-\mathbf{q}|^2}{\bar{k}^2}\right)^{-1/2}\exp\left\{-\hat{i}\left(\frac{2p^2+q^2}{2\bar{k}}+O\left(\frac{1}{\bar{k}^3}\right)\right)y\right\}\hat{R}(y,\mathbf{q})\right]$$
$$\equiv\int_0^\infty dy\, e^{2\hat{i}\bar{k}y} R^b_{\mathbf{p},\bar{k}}\left(y,\frac{-y\mathbf{p}}{\bar{k}}\right)=\frac{\hat{i}}{2\bar{k}}R^b_{\mathbf{p},\bar{k}}(0,0)+\frac{\hat{i}}{2\bar{k}}\int_0^\infty dy\, e^{2\hat{i}\bar{k}y}\frac{\partial}{\partial y}R^b_{\mathbf{p},\bar{k}}\left(y,\frac{-y\mathbf{p}}{\bar{k}}\right), \tag{3.6b}$$

where $R^{f,b}_{\mathbf{p},\bar{k}}$ are inverse Fourier transforms, with respect to the $\mathbf{q}$ variable, of the corresponding expressions in the square parentheses [both of which are products of $\hat{R}(y,\mathbf{q})$ and a slowly varying function ~ 1]. Some care must be taken here in determining large-$\bar{k}$ behavior since straightforward expansion, with respect to $1/\bar{k}$, of functions appearing in the square parentheses, can lead to divergent $\mathbf{q}$ integrals. The difficulty here is that an even function $R(y,\mathbf{x}_\perp)$ may have a slope discontinuity at $\mathbf{x}_\perp = O$, which leads to slow decay of $\hat{R}(y,\mathbf{q}) = O(q^{-2})$, as $q\to\infty$. In any case, it is clear that

$$\gamma_{\mathbf{p}}^f = \frac{1}{4\bar{k}^2}\int_0^\infty dy\, R(y,0)+O\left(\frac{1}{\bar{k}^3}\right),\quad \gamma_{\mathbf{p}}^b = \frac{\hat{i}R(0,0)}{8\bar{k}^3}+O\left(\frac{1}{\bar{k}^4}\right), \tag{3.7}$$

which should be compared with similar results in Refs. 8 and 9.

IV. THE STOCHASTIC SQUARE LAW MEDIUM (LENSE)

A square law medium is described by $\bar{n}(\mathbf{x}_\perp)^2 = 1 - B^2x_\perp^2$, where B is a constant, and here we assume that the random fluctuations have the form[10]

$$\eta(\mathbf{x}) = n(\mathbf{x})^2 - \bar{n}(\mathbf{x}_\perp)^2 = -\sum_{i=1,2}\eta_i(x)x_i - \sum_{i,j=1,2}\eta_{ij}(x)x_i\,x_j\,. \tag{4.1}$$

The centered random variables η_i and η_{ij} are associated with random misalignment and random focusing "width," respectively. This choice of $n(\mathbf{x})$, though unphysical for large $\mathbf{x}_\perp$, provides a reasonable model for the propagation of low-order modes in certain optical fibers.[10]

The orthonormal eigenfunctions and eigenvalues of the self-adjoint reference splitting operator, $S_0 = \Delta_\perp + k^2(1-B^2x_\perp^2)$, are naturally enumerated as $\Psi_{m_1m_2}(\mathbf{x}_\perp) = \Psi_{m_1}(x_1)\Psi_{m_2}(x_2)$, where

$$\Psi_m(y) = \left(\frac{2^{-m}}{m!}\right)^{1/2}\left(\frac{kB}{\pi}\right)^{1/4} \times H_m(k^{1/2}B^{1/2}y)\exp\left(\frac{-kBy^2}{2}\right),$$

and H_m is the mth Hermite polynomial[4], and $\lambda_{m_1m_2} = k^2 - 2kB(m_1+m_2+1)$. The only other model specific input required for the coupled equations (2.2), or the equations following, is the matrix elements of $\beta(\mathbf{x}) = k^2\eta(\mathbf{x})$ with respect to S_0 eigenfunctions. Because of the special form of (4.1), these are obtained simply and explicitly from the identities

$$\int_{-\infty}^{\infty} dy\,\Psi_m(y)\,y\Psi_n(y) = (2kB)^{-1/2}\{(n+1)^{1/2}\times\delta_{m,n+1} + n^{1/2}\delta_{m,n-1}\}, \tag{4.2a}$$

$$\int_{-\infty}^{+\infty} dy\Psi_m(y)\,y^2\Psi_n(y) = (2kB)^{-1}\{(n+2)^{1/2}(n+1)^{1/2}\delta_{m,n+2} + (2n+1)\delta_{m,n} + n^{1/2}(n-1)^{1/2}\delta_{m,n-2}\}\,. \tag{4.2b}$$

From (4.2), it is clear that random misalignment provides the dominant coupling in the high-k regime. In the absence of misalignment, and if the η_{ij} are independent of x, then clearly one can make a (generally) new choice of transverse x_i axes such that there will only be coupling between modes (m,n) and (m',n') for even $m-m'$ and $n-n'$.

Here we consider only the behavior of the right propagating mean field, u^+, for the special choice of initial conditions corresponding to a single right propagating mode $(m,n) = (0,0)$, i.e., a *Gaussian beam*, at $x=0$. Thus one has $u^+_{mn}(x=0)\propto\delta_{m,0}\delta_{n,0}$. Furthermore, we restrict our attention to the long-range Markovian approximation of the first-order smoothed equation (2.9), after neglecting terms $O(\beta^3)$, which implies that [cf. (2.10)]

$$u^+_{00}(x)\approx\exp[-(\gamma^f_{00}+\gamma^b_{00})x]\,u^+_{00}(0)\,. \tag{4.3}$$

Here the forward- and backscattering contributions γ^f_{00}, γ^b_{00} to the (0,0)-mode diffusion coefficient can be determined from the correlation matrix elements

$$R_{00,m_1,m_2,00}(\delta x)$$
$$= \langle \beta_{00,m_1,m_2}(x)\beta_{m_1m_2,00}(x+\delta x)\rangle$$
$$= (2kB)^{-1}\sum_{i=1,2} R_i(\delta x)\delta_{m_i,1}\delta_{m_j,0}$$
$$+ (2kB)^{-2}[R_{12,12}(\delta x) + 2R_{12,21}(\delta x)$$
$$+ R_{21,21}(\delta x)]\delta_{m_1,1}\delta_{m_2,1}$$
$$+ 2(2kB)^{-2}\sum_{i=1,2} R_{ii,ii}(\delta x)\delta_{m_i,2}\delta_{m_j,0}$$
$$+ (2kB)^{-2}[R_{11,11}(\delta x) + 2R_{11,22}(\delta x)$$
$$+ R_{22,22}(\delta x)]\delta_{m_1,0}\delta_{m_2,0}\ , \tag{4.4}$$

where $\{i,j\} = \{1,2\}$, $R_i(\delta x) = k^4\langle\eta_i(x)\eta_i(x+\delta x)\rangle$, and $R_{ij,kl}(\delta x) = k^4\langle\eta_{ij}(x)\eta_{kl}(x+\delta x)\rangle$ (so $R_{11,22} = R_{22,11}$ and $R_{12,21} = R_{21,12}$). It is clear that one can make the decompositions, $\gamma^f_{00} = \gamma^f_{00}(\text{mis}) + \gamma^f_{00}(\text{width})$, $\gamma^b_{00} = \gamma^b_{00}(\text{mis}) + \gamma^b_{00}(\text{width})$, into contributions associated with random misalignment and random width, respectively. Then one obtains from (2.11),

$$\gamma^f_{00}(\text{mis}) = \frac{1}{8k^3B}\left(1-\frac{2B}{k}\right)^{-1/2}\left(1-\frac{4B}{k}\right)^{-1/2}$$
$$\times\int_0^\infty dx\,\exp\{\hat{i}[(k^2-4kB)^{1/2}$$
$$-(k^2-2kB)^{1/2}]x\}[R_1(x)+R_2(x)]$$
$$=\frac{1}{8k^3B}\int_0^\infty dx\,e^{-\hat{i}Bx}[R_1(x)+R_2(x)]$$
$$\times\left[1+\frac{3B}{k}-\frac{3\hat{i}B^2x}{2k}+O\left(\frac{1}{k^2}\right)\right],$$
$$\gamma^b_{00}(\text{mis}) = \frac{1}{8k^3B}\left(1-\frac{2B}{k}\right)^{-1/2}\left(1-\frac{4B}{k}\right)^{-1/2}$$
$$\times\int_0^\infty dx\,\exp\left\{\hat{i}k\left[\left(1-\frac{4B}{k}\right)^{1/2}\right.\right.$$
$$\left.\left.+\left(1-\frac{2B}{k}\right)^{1/2}\right]x\right\}[R_1(x)+R_2(x)]$$
$$=\frac{\hat{i}}{16k^4B}\left[R_1(0)+R_2(0)+\frac{1}{2k}\{\hat{i}R'_1(0)+\hat{i}R'_2(0)\right.$$
$$\left.+9BR_1(0)+9BR_2(0)\}+O\left(\frac{1}{k^2}\right)\right],$$
$$\text{as } k\to\infty\ , \tag{4.5}$$

where $R'_i(0)$ denotes the (right-sided) derivative of $R_i(x)$, at $x=0$, and similarly,

$$\gamma^f_{00}(\text{width}) = \frac{1}{16k^4B^2}\int_0^\infty dx\,e^{-2\hat{i}Bx}[R_{12,12}(x)$$
$$+2R_{12,21}(x)+R_{21,21}(x)+2R_{11,11}(x)$$
$$+2R_{22,22}(x)]\left[1+\frac{4B}{R}-\frac{4\hat{i}B^2}{k}x+O\left(\frac{1}{k^2}\right)\right]$$
$$+\frac{1}{16k^4B^2}\left[1+\frac{2B}{k}+O\left(\frac{1}{k^2}\right)\right]\int_0^\infty dx$$
$$\times[R_{11,11}(x)+2R_{11,22}(x)+R_{22,22}(x)]\ ,$$

$$\gamma^b_{00}(\text{width})$$
$$=(\hat{i}/32k^5B^2)[R_{12,12}(0)+2R_{12,21}(0)+R_{21,21}(0)$$
$$+3R_{11,11}(0)+3R_{22,22}(0)+2R_{11,22}(0)$$
$$+0(1/k)],\quad \text{as } k\to\infty\ . \tag{4.6}$$

Note that the dominant terms in Re $\gamma^f_{00}(\ )$ are clearly positive (assuming monotonically decreasing correlations). Similar expressions are readily calculated for diffusion coefficients for other modes. One should compare the structure of γ^f_{00} with results of Besieris[11] for a related stochastic lense problem from a parabolic-type approximation.

V. DISCUSSION

By exploiting a powerful albeit abstract splitting technique, we have simply and succinctly obtained equations which provide a practical basis for the perturbative analysis of wave propagation in random media. The emphasis here is not on a detailed or rigorous analysis of the statistical assumptions, the diffusion limit or convergence of perturbative expansions. Rather, we simply elucidate the effects of randomness on the wave field and, in particular, the backscattering contribution to the diffusion coefficients.

ACKNOWLEDGMENTS

The author would like to thank Dr. J. P. Corones for motivating this analysis, and Dr. D. W. McLaughlin and Dr. J. P. Corones for providing copies of their unpublished work.

Ames Laboratory is operated for the U. S. Department of Energy by Iowa State University under Contract No. W-7405-Eng-82. This work was supported by the Office of Basic Energy Sciences.

[1]See, for example, H. Bremmer, Commun. Pure Appl. Math. **4**, 105 (1951); F. V. Atkinson, J. Math. Anal. Appl. **1**, 255, (1960); J. A. Arnaud, Bell Syst. Tech. J. **49**, 2311 (1970); J. Corones, J. Math. Anal. Appl. **50**, 361 (1975); D. Dudley and A. P. Wang, J. Math. Phys. **24**, 1470 (1983).

[2]L. Fishman and J. J. McCoy, in *Applications of Mathematics to Modern Optics*, edited by W. H. Carter (SPIE, Bellingham, WA, 1982), Vol. 358, p. 168; J. Math. Phys. **25**, 285 (1984).

[3]J. P. Corones, M. E. Davison, and R. J. Krueger, " A general approach to splitting and invariant imbedding techniques for linear wave equations," in *Inverse Optics, Proceedings of the SPIE 413*, edited by A. J. Devaney (SPIE, Bellingham, WA, 1983), p. 102; J. P. Corones and R. J. Krueger, J. Math. Phys. **24**, 2301 (1983); M. E. Davison, "A general approach to splitting and invariant imbedding techniques for linear wave equations," Ames Laboratory Preprint, 1982.

[4]J. W. Evans and J. P. Corones, "Splitting methods for time-independent wave propagation in focussing media and wave guides," Ames Laboratory Preprint, 1985.

[5]U. Frisch, in *Probabilistic Methods in Applied Mathematics*, edited by A. T. Bharucha-Reid (Academic, New York, 1968), Vol. 1.

[6]R. Z. Khasminskii, Theory Probab. Appl. **11**, 390 (1966); G. C. Papanicolaou and J. B. Keller, Siam J. Appl. Math. **20**, 287 (1971); W. Kohler and G. C. Papanicolaou, in "Wave propagation and underwater acoustics," in *Lecture Notes in Physics, Vol. 70*, edited by J. B. Keller and J. S. Papadakis (Springer, New York, 1977); G. C. Papanicolaou and W. E. Kohler, Commun. Pure Appl. Math. **27**, 641 (1974); W. E. Kohler, Wave Motion **4**, 243 (1982).

[7]J. Mathews and R. L. Walker, *Mathematical Methods of Physics* (Benjamin, Reading, MA, 1964), 2nd ed.

[8]V. I. Klyatskin and V. I. Tatarskii, Sov. Phys. JETP **31**, 335 (1970).

[9]J. Corones and D. W. McLaughlin, "The parabolic approximation in focusing media" (unpublished manuscript); D. W. McLaughlin, "Accuracy of the parabolic and diffusion approximation" (unpublished manuscript).

[10]G. C. Papanicoloau, D. McLaughlin, and R. Burridge, J. Math. Phys. **14**, 84 (1973); D. Marcuse, Bell Syst. Tech. J. **53**, 195 (1974).

[11]I. M. Besieris, J. Math. Phys. **19**, 2533 (1978).

Singularities of the continuation of fields and validity of Rayleigh's hypothesis

D. Maystre and M. Cadilhac

Laboratoire d'Optique Electromagnétique, Equipe de Recherche Associée au Centre National de la Recherche Scientifique no. 597, Faculté des Sciences et Techniques, Centre de Saint-Jérôme, 13397 Marseille Cedex 13, France

(Received 17 July 1984; accepted for publication 18 April 1985)

To formulate general results concerning the validity of the Rayleigh hypothesis, we first introduce a definition of the foci and antifoci of an analytic curve. Then, we state two lemmas on the properties of an analytic or harmonic function satisfying given conditions on an analytic curve. This allows us to predict the behavior of the analytic continuation of the field in electrostatics. The use of a conformal mapping permits the generalization of this method in electromagnetics and acoustics. As a consequence, we are able to predict the limit of validity of the Rayleigh hypothesis.

I. INTRODUCTION

At the beginning of the century, the Rayleigh method had been the first attempt at solving the problem of diffraction by gratings.[1] This method has been used for many other problems of electromagnetism and acoustics. Rayleigh made an assumption, the so-called Rayleigh hypothesis, which remained unquestioned for almost 50 years, but provoked considerable controversy thereafter. At present, there is no doubt that the Rayleigh hypothesis is neither always valid, nor always invalid. The interested reader may consult recent reviews in this field.[2,3]

However, the controversial aspect of the Rayleigh hypothesis has not died down, due to a second question: in what conditions may the Rayleigh theory be used to determine the field diffracted by a scattering object, even though the Rayleigh hypothesis fails? In this paper, we are not concerned with this second question. Our aim is to establish a mathematical property which allows us to state a very simple and general result concerning the validity of the Rayleigh hypothesis in electromagnetism and acoustics, when the profile of a diffracting object is given by an analytic curve. To this end, we first deal with the Neumann and Dirichlet problem in electrostatics, since it has been shown that the validity of the Rayleigh hypothesis in electromagnetism or acoustics is linked with the properties of the analytical continuation of the field in the corresponding problems of electrostatics.[4]

II. DEFINITION OF THE FOCI AND ANTIFOCI OF AN ANALYTIC CURVE

The notion of foci is well known for conics. Here, we propose a generalization of this notion to analytic curves. Moreover, we introduce the notion of antifoci.

First, let us recall the definition of an analytic curve Γ: let D_t be a domain (open connected set) of the complex t plane and $I \subset D_t$ a real interval. An analytic curve Γ is the image of I through a transformation

$$z = \zeta(t), \tag{1}$$

ζ being a nonconstant analytic function defined in D_t.

Now, if there exists a point $t_0 \in D_t$, such that $\bar{t}_0 \in D_t$, satisfying

$$\zeta'(t_0) = 0, \tag{2}$$

$$\zeta'(\bar{t}_0) \neq 0, \quad \zeta' \text{ being the derivative of } \zeta, \tag{3}$$

the images $z_0 = \zeta(t_0)$ and $\tilde{z}_0 = \zeta(\bar{t}_0)$ of t_0 and $\bar{t}_0$ will be called the associated focus and antifocus of Γ, respectively.

For instance, let us consider the case of a parabola given by the function

$$z = \zeta(t) = 2t + it^2. \tag{4}$$

Its focus z_0 will be obtained by setting

$$\zeta'(t_0) = 2 + 2it_0 = 0, \tag{5}$$

which means that $t_0 = i$, thus

$$z_0 = i, \tag{6}$$

$$\tilde{z}_0 = -3i. \tag{7}$$

Finally, the antifocus is symmetrical to the focus with respect to the directrix of the parabola.

More generally, it can be verified that the notion of focus given here identifies with the classical one in the case of conics (except for a circle!). When the analytic curve Γ is given by the equation

$$G(x,y) = 0, \tag{8}$$

where G is an analytic function of the variables x and y ($z = x + iy$), it can be shown that a focus z_0 is obtained by

$$z_0 = x_1 + iy_1, \tag{9}$$

where x_1 and y_1 are complex numbers satisfying the system

$$G(x,y) = 0, \tag{10}$$

$$\frac{\partial G}{\partial x} + i\frac{\partial G}{\partial y} = 0, \quad \text{with } \frac{\partial G}{\partial x} \neq 0. \tag{11}$$

In addition, the associated antifocus is given by

$$\tilde{z}_0 = \bar{x}_1 + i\bar{y}_1. \tag{12}$$

We shall set

$$z_0 = x_0 + iy_0, \quad \tilde{z}_0 = \tilde{x}_0 + i\tilde{y}_0,$$

where x_0, y_0, $\tilde{x}_0$, $\tilde{y}_0$, the Cartesian coordinates of the focus and the antifocus, are real.

We define a focal line as the image $\zeta(L)$ of a curve L (a) joining t_0 to $\bar{t}_0$ in D_t, (b) symmetrical with respect to the real axis, and (c) intersecting I. For example, in the case of a parabola, the segment $[z_0, \tilde{z}_0]$ is a focal line.

A domain D will be called a focal domain if (a)$D \subset D_z = \zeta(D_t)$ and (b) whenever D contains an antifocus $\tilde{z}_0$, it includes an associated focal line.

It is interesting to notice that with the new variables

$$z = x + iy, \tag{13}$$

$$\tilde{z} = x - iy, \tag{14}$$

the focus is given by

$$H(z, \tilde{z}) = 0, \tag{15}$$

$$\frac{\partial H}{\partial \tilde{z}} = 0 \quad \text{and} \quad \frac{\partial H}{\partial z} \neq 0, \tag{16}$$

with

$$H(z, \tilde{z}) = G(x,y).$$

It is worth noting that a system of parametric equations similar to (1) may be deduced from (10) by integrating the system (Hamilton's canonical equations!)

$$\frac{dx}{dt} = -\frac{\partial G}{\partial y}, \tag{17}$$

$$\frac{dy}{dt} = \frac{\partial G}{\partial x}, \tag{18}$$

with arbitrary initial conditions.

With the new variables defined in (13) and (14), these equations become

$$\frac{dz}{dt} = 2i \frac{\partial H}{\partial z}, \tag{19}$$

$$\frac{d\tilde{z}}{dt} = -2i \frac{\partial H}{\partial \tilde{z}}. \tag{20}$$

III. LEMMAS

Lemma 1: An analytic curve Γ being given, let $F(z)$ be an analytic function in a focal domain D and $\tilde{z}_0$ an antifocus in D. If, for $z \in \Gamma \cap D$, $F(z)$ is real, then $F'(\tilde{z}_0) = 0$.

Proof: The function

$$\theta(t) = F(\zeta(t)) \tag{21}$$

is analytic in the connected component of $\zeta^{-1}(D)$ which contains $t_0, \bar{t}_0$. If $t \in I$, $\theta(t)$ is real, thus $\theta'(t)$ is real, too, and therefore, from a well-known symmetry property,

$$\theta'(\bar{t}_0) = \overline{\theta'(t_0)}. \tag{22}$$

But,

$$\theta'(t_0) = \zeta'(t_0)F'(z_0), \tag{23}$$

$$\theta'(\bar{t}_0) = \zeta'(\bar{t}_0)F'(\tilde{z}_0), \tag{24}$$

and from (2) and (3)

$$F'(\tilde{z}_0) = 0. \tag{25}$$

Lemma 2: An analytic curve Γ being given, let $u(x,y)$ be a harmonic function in a focal domain D and $\tilde{z}_0$ an antifocus in D. If, for $z \in \Gamma$, $u(x,y)$ (or its normal derivative) vanishes, then $\tilde{z}_0$ is a saddle point of $u(x,y)$

$$\frac{\partial u}{\partial x}(\tilde{x}_0, \tilde{y}_0) = \frac{\partial u}{\partial y}(\tilde{x}_0, \tilde{y}_0) = 0. \tag{26}$$

Proof: D can be supposed to be simply connected without loss of generality. There exists an analytic function $F(z)$ such that

$$u(x,y) = \mathrm{Im}(F(z)) \quad [\text{or } u(x,y) = \mathrm{Re}(F(z))],$$

where $F(z)$ fulfills the conditions of Lemma 1. Hence $F'(\tilde{z}_0) = 0$, which is equivalent to (26).

IV. EXAMPLES OF APPLICATION

(1) Let D be a domain intersecting an analytic curve Γ and containing an antifocus $\tilde{z}_0$. Let F be analytic in D and real on Γ. Then, an analytic continuation of F cannot be made along a focal line up to the associated focus z_0, unless $F'(\tilde{z}_0) = 0$.

Such an analytic continuation can be deduced from the symmetry property of $F(\zeta(t))$.

(2)Let us consider a Jordan domain Ω with analytic boundary Γ and a conformal mapping $Z = \phi(z)$ of the exterior of Γ on the exterior of the unit circle C (Fig. 1). We have locally $\phi(z) = \exp(iF(z))$, where F is real on Γ. Moreover, F' is analytic and different from 0 outside $\Omega + \Gamma$. This entails that the foci of Γ located in Ω are singularities of the analytic continuation of ϕ along the focal lines.

(3) A third example consists of the homogeneous Dirichlet and Neumann problems for the Laplace equations.

Now, we shall restrict ourselves to the case where Γ separates the space in two complementary regions Ω_1 and Ω_2. These regions are unbounded if Γ goes to infinity, but one of them, Ω_2, is bounded (the interior region) if Γ is a Jordan curve.

We consider a harmonic function $u(x,y)$ defined in Ω_1 and which satisfies a homogeneous Dirichlet or Neumann condition on Γ. If Ω_1 contains an antifocus $\tilde{z}_0$, then the continuation of u across Γ along a focal line will not be possible at the associated focus z_0 if $(\partial u/\partial x)(\tilde{x}_0, \tilde{y}_0) \neq 0$ or $(\partial u/\partial y)(\tilde{x}_0, \tilde{y}_0) \neq 0$.

Indeed, if this continuation were possible, $u(x,y)$ would be harmonic in a focal domain containing z_0 and $\tilde{z}_0$, a fact which entails that the partial derivatives of u with respect to x and y vanish at the point $(\tilde{x}_0, \tilde{y}_0)$.

V. VALIDITY OF SOME EXPANSIONS OF THE FIELD USED IN ELECTROMAGNETICS AND ACOUSTICS

We consider the Helmholtz equation

$$\nabla^2 u(x,y) + k^2 u(x,y) = 0, \quad \text{in } \Omega_1, \tag{27}$$

with the homogeneous Dirichlet or Neumann conditions on Γ [notations of Sec. IV, example (3)].

It has been shown[5] that the use of a conformal mapping $Z = \Phi(z)$ which maps Ω_1 on the upper Z half-plane or on the

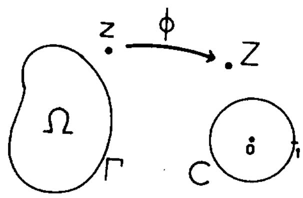


FIG. 1. A property of the conformal mapping.

exterior of the unit disk allows one to define an equivalent problem in the Z complex plane, where $v(X,Y) = u(x,y)$ satisfies the Dirichlet or Neumann boundary conditions on the real axis or the unit circle and a new Helmholtz equation

$$\nabla^2 v(X,Y) + k^2 \left| \frac{dz}{dZ} \right|^2 v(X,Y) = 0. \tag{28}$$

We have already seen that the continuation of Φ in Ω_2 is singular at a focus z_0 of Γ. This entails that even though $v(X,Y)$ is regular in the upper half-plane, we can expect a singularity of the continuation of $u(x,y)$ in Ω_2 at the focus since dZ/dz is singular at this point. Of course, this rule is not general since we have no information about the value of $v(X,Y)$ at the image of the focus.

It is clear that our criterion gives a means to locate *some* of the singularities of the conformal mapping. Other singularities may well exist in the complementary domain. On the other hand, we emphasize that the criterion does not *guarantee* a singularity at the focus in Ω_2.

The location of the singularity of the analytical continuation of the field in Ω_2 allows one to predict the validity of some expansions of the field used in electromagnetics and acoustics. The most famous of these expansions has been used by Lord Rayleigh to represent the field diffracted by a grating.[1] The reader interested in the study of the validity of Rayleigh's hypothesis may refer to recent reviews in this field (for instance, see Ref. 3 and included references).

Here, we first deal with the more general case where Γ is a modulated two-dimensional surface extending to infinity (Fig. 2), obtained by deforming a mirror placed on the $0x$ axis. An incident wave u^i propagating in Ω_1 is impinging on Γ. The equivalent of Rayleigh's hypothesis is to assume that in Ω_1, the diffracted field $u^d = u - u^i$ (where u denotes the total field) can be expressed in the form of a sum of plane waves

$$u^d = \int_{-\infty}^{\infty} a(\alpha) \exp(i\alpha x + i\beta y) d\alpha, \tag{29}$$

with $\beta = \sqrt{k^2 - \alpha^2}$ or $i\sqrt{\alpha^2 - k^2}$, the time dependence of the field being in $\exp(-i\omega t)$.

Let us show briefly the great interest of this kind of representation of the field. Indeed, the right-hand member of Eq. (29) obviously satisfies the Helmholtz equation and the outgoing wave condition at infinity. So, if this representation is valid everywhere above Γ, it can be used to express the

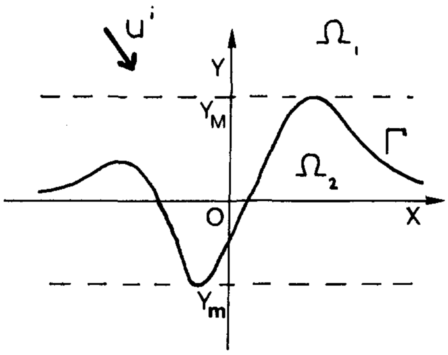


FIG. 2. Validity of the plane wave expansion in the problem of modulated surface.

third condition of the boundary value problem, viz., the boundary condition on Γ. This gives a very simple tool to solve the diffraction problem. It is not so for other rigorous methods which can lead to the solving of integral equations or differential systems of infinite order.

It can be demonstrated that the integral in the right-hand side of (29) actually represents u^d above the top y_M of Γ (the demonstration of this property and those used in the following can be found in Ref. 3 for the particular case of diffraction gratings). Below y_M, the integral is equal to the diffracted field or its analytic continuation in Ω_2, *provided it converges*. Obviously, this integral cannot converge below a focus (except if this focus is not a singularity of the continuation of u). Indeed, since $\exp(i\beta y)$ behaves like $\exp - |\alpha| y$ when $|\alpha| \to \infty$, this integral cannot converge at a point of ordinate y' if it diverges at a point of ordinate $y > y'$.

So, it can be expected that the expansion of u^d given by (29) cannot represent the diffracted field in Ω_1 if a focus is located above the bottom y_M of Γ. This means that a method using this integral to express the boundary condition on Γ fails, at least from a theoretical point of view. Finally we can state the following rule: The plane wave expansion given by the right-hand side of (29) in general cannot represent the diffracted field in Ω_1 when a focus of Γ in Ω_2 is located above the bottom of Γ.

It must be remarked that, in the particular case where Γ is a periodic curve, a profile of a diffraction grating, similar criterion have been given by some authors using conformal mapping[6,7] or the steepest descent method.[8,9]

For instance, let us consider the curve Γ given by

$$y = 2a/\cosh x, \quad \text{with } a > 0, \tag{30}$$

located above the $0x$ axis.

From Eqs. (10) and (11), we deduce that the foci are given by the equation

$$\sin^2 v + 2a \sin v - 1 = 0, \quad \text{where } v = ix. \tag{31}$$

There exists an infinity of foci. From the point of view of the validity of the Rayleigh expansion, the most important is

$$z_0 = iy_0, \tag{32}$$

with

$$y_0 = (2a\sqrt{1 + a^2} + 2a^2)^{1/2} - \arcsin(\sqrt{1 + a^2} - a). \tag{33}$$

This focus is located on the imaginary axis ($y_0 \to -\pi/2$ for $a \to 0$) and crosses the real axis for $a = 0.280\,548...$ (the corresponding antifocus being located in Ω_1).

So, we can expect a failure of the plane-wave expansion method for larger values of a. The study of the other foci does not modify this conclusion.

Now, let us consider a second kind of curve: the Jordan curve (Fig. 3). In that case, it can be shown that, if an incident wave propagates in Ω_1, the field outside a circle of radius ρ_M centered on 0 can be represented by a series

$$u^d(P) = \sum_{-\infty}^{\infty} a_n H_n^{(1)}(kr) \exp(in\theta), \tag{34}$$

a_n being complex coefficients, $H_n^{(1)}$ Hankel functions, and (r,θ) the polar coordinates of a point P.

Considerations similar to those described for modulated surfaces demonstrate the following rule: The expansion given by the right-hand side of (34) in general cannot represent the diffracted field everywhere in Ω_1 when a focus of Γ in Ω_2 is located between the two dotted circles of Fig. 3, of radius ρ_M and ρ_m.

Let us apply this rule to the curve Γ given by

$$x^4 + y^4 = 1. \tag{35}$$

To find the foci of Γ, we use Eqs. (15) and (16), and remarking that (35) becomes

$$H(z,\tilde{z}) = \tfrac{1}{32}(z^4 + 6z^2\tilde{z}^2 + \tilde{z}^4 - 8) = 0, \tag{36}$$

it turns out that $12z^2\tilde{z} + 4\tilde{z}^3 = 0$, i.e.,

$$\tilde{z} = 0, \tag{37}$$

or

$$\tilde{z}^2 + 3z^2 = 0. \tag{38}$$

Putting (37) into (36) shows that $z^4 = 8$, and the associated foci are given by

$$z_0 = 2^{3/4}\exp(in(\pi/2)), \quad n = 0,1,2,3. \tag{39}$$

These foci are located in Ω_1 and have no interest for our problem. Now, Eqs. (36) and (38) lead to the equation $z^4 = -1$, which means that the second set of foci is given by

$$z_0 = \exp[i(\pi/4 + n\pi/2)], \quad n = 0,1,2,3. \tag{40}$$

We are led to an amazing conclusion: four foci are just located on the circle of radius $\rho_m = 1$, which means that the expansion (34) actually can represent the field in Ω_1, but diverges just below the points of Γ located on the two axes of coordinate and placed on the circle $r = \rho_m$.

It is worth noting that Eqs. (19) and (20) allow one to find parametric equations associated with Eq. (35), using elliptic functions.

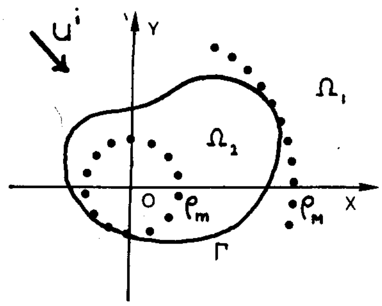


FIG. 3. Validity of a simple representation of the field for a Jordan curve.

VI. CONCLUSION

Introducing the notion of focus and antifocus has allowed us to state in a very simple and general form a property of the singularities of the continuation of the field. As a consequence, we can predict the theoretical limits of some simple expansions used to solve a large class of boundary problems in electromagnetics and acoustics.

[1]Lord Rayleigh (J. W. Strutt), Proc. R. Soc. London Ser. A **79**, 399 (1907).
[2]M. Cadilhac, "Some mathematical aspects of the grating theory," in *Electromagnetic Theory of Gratings*, edited by R. Petit (Springer, Berlin, 1980), pp. 53–62.
[3]D. Maystre, "Rigorous vector theories of diffraction gratings," in *Progress in Optics*, Vol. 21, edited by E. Wolf (North-Holland, Amsterdam, 1984).
[4]R. F. Millar, Radio Sci. **8**, 785 (1973).
[5]M. Neviere and M. Cadilhac, Opt. Commun. **2**, 235 (1970).
[6]P. M. Van der Berg and J. T. Fokkema, J. Opt. Soc. Am. **69**, 27 (1979).
[7]P. M. Van der Berg and J. T. Fokkema, IEEE Trans. Antennas Propag. **AP-27**, 577 (1979).
[8]N. R. Hill and V. Celli, Phys. Rev. B **17**, 2478 (1978).
[9]J. A. De Santo, Radio Sci. **16**, 1315 (1981).

Comments on the perturbed sine–Gordon equation

N. J. Papastamatiou
Physics Department, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin 53201
H. Matsumoto and H. Umezawa
Theoretical Physics Institute, University of Alberta, Edmonton, Alberta, Canada T6G 2J1

(Received 7 September 1984; accepted for publication 17 May 1985)

We examine the sine–Gordon equation with a perturbation $\lambda\Delta V$. We derive necessary conditions on ΔV such that the perturbed equation has solutions with finite energy, analytic in λ, and which reduce to the static soliton when the perturbation is removed ($\lambda\to 0$). Several examples illustrating these conditions are presented.

I. INTRODUCTION

The study of integrable nonlinear equations with additional interactions is an important topic in the theory of extended object dynamics. Such additional interaction terms occur naturally when the theory is quantized and quantum corrections are included, or when finite-temperature effects are considered. An important question in such cases refers to the "stability" of the integrability property, i.e., the ease with which a perturbation destroys the integrability of the original equation.

Soliton systems with nontrivial additional interactions have been studied by several authors. These studies show that a static soliton often can be forced to become time dependent by an additional interaction and that a meaningful perturbation should be built not on the original static soliton, but on a modified, time-dependent unperturbed solution.[1]

The purpose of this paper is to present some comments about the conditions for the existence of perturbative solutions. We examine the perturbed sine–Gordon equation

$$\phi_{tt}-\phi_{xx}+m^2\sin\phi=-\lambda\frac{\partial\Delta V}{\partial\phi}, \tag{1}$$

with $\lambda\Delta V$ the perturbing potential. For $\lambda=0$, Eq. (1) possesses the well-known static solution (soliton)

$$\phi_0(x)=4\tan^{-1}e^{mx}. \tag{2}$$

The question we address is the following: What conditions must be imposed on ΔV so that Eq. (1) possesses solutions analytic in λ, which reduce to $\phi_0(x)$ in the limit $\lambda\to 0$? In the following, we present two theorems dealing with this question. Each theorem gives a necessary condition on ΔV for the existence of solutions of the type discussed above. We also present examples which illustrate these theorems.

II. THEORY

We are interested in solutions $\phi(x,t)$ of (1) which satisfy the following criteria: (i) they are analytic in λ and reduce to $\phi_0(x)$ in the limit $\lambda\to 0$; and (ii) they lead to finite energy.

Criterion (i) implies that $\phi(x,t)$ can be expanded as

$$\phi(x,t)=\phi_0(x)+\lambda\phi_1(x,t)+\cdots, \tag{3}$$

with $\phi_0(x)$ and $\phi_1(x,t)$ obeying, respectively,

$$-\phi_{0xx}+m^2\sin\phi_0=0, \tag{4}$$

$$(\partial_t^2-\partial_x^2+m^2\cos\phi_0)\phi_1=-\frac{\partial\Delta V}{\partial\phi}\bigg|_{\phi=\phi_0}. \tag{5}$$

Next we turn to criterion (ii): the energy

$$E=\int_{-\infty}^{+\infty}dx\left[\frac{1}{2}\phi_t^2+\frac{1}{2}\phi_x^2+m^2(1-\cos\phi)+\lambda\Delta V\right] \tag{6}$$

can be expanded in powers of λ as

$$E=E_0+\lambda E_1+\cdots, \tag{7}$$

where E_0 is the soliton energy

$$E_0=8m$$

and E_1 is given by the expression

$$E_1=\int_{-\infty}^{+\infty}dx\,[\phi_{0x}\phi_{1x}+m^2\sin\phi_0\cdot\phi_1+\Delta V(\phi_0)]. \tag{8}$$

For E_1 to be finite, the integrand must tend to zero as $x\to\pm\infty$; therefore

$$\lim_{x\to\pm\infty}4me^{-m|x|}(\phi_{1x}\mp m\phi_1)+\Delta V_\pm=0. \tag{9}$$

In deriving this result, we used Eq. (2) and defined

$$\Delta V_\pm=\lim_{x\to\pm\infty}\Delta V(\phi_0). \tag{10}$$

Since ϕ_0 tends to constants for $x\to\pm\infty$, $\Delta V_\pm$ are constants. Equation (9) has the solution

$$\phi_1\sim(-(\Delta V_\pm/4m)x+c_\pm)e^{m|x|}+f_\pm\quad(x\to\pm\infty), \tag{11}$$

where the $c_\pm$ are functions of t only and the $f_\pm$ are functions of x and t satisfying

$$\lim_{x\to\pm\infty}e^{-m|x|}f_\pm(x,t)=0. \tag{12}$$

The asymptotic form of (5) for $x\to\pm\infty$ is

$$(\partial_t^2-\partial_x^2+m^2)\phi_1=-\frac{\partial\Delta V_\pm}{\partial\phi}, \tag{13}$$

where we defined

$$\frac{\partial\Delta V_\pm}{\partial\phi}=\lim_{x\to\pm\infty}\frac{\partial\Delta V}{\partial\phi}\bigg|_{\phi=\phi_0}. \tag{14}$$

If we combine Eqs. (11) and (13), we find that

$$\frac{d^2}{dt^2}c_\pm=\mp\frac{1}{2}\Delta V_\pm, \tag{15a}$$

 0022-2488/85/092205-03$02.50

$$\frac{\partial^2}{\partial t^2}f_\pm - \frac{\partial^2}{\partial x^2}f_\pm + m^2 f_\pm = -\left(\frac{\partial \Delta V}{\partial \phi}\right)_\pm . \qquad (15b)$$

We are not insisting that ϕ_1 be square integrable. However, from Eq. (11), the function

$$\tilde{\phi}_1 = \phi_1 - \left[\left(-\frac{\Delta V(\phi_0)}{2m}x + 2c\right)\cosh mx + f\right] \equiv \phi_1 - F, \qquad (16)$$

with

$$\lim_{x\to\pm\infty} c = c_\pm , \quad \lim_{x\to\pm\infty} f = f_\pm , \qquad (17)$$

is square integrable. It satisfies the equation

$$(\partial_t^2 - \partial_x^2 + m^2 \cos\phi_0)\tilde{\phi}_1 = -\frac{\partial \Delta V}{\partial \phi}\bigg|_{\phi=\phi_0} - (\partial_t^2 - \partial_x^2 + m^2\cos\phi_0)F. \qquad (18)$$

Let ψ_k, ω_n be the normalized eigenfunctions and eigenvalues of the stability equation

$$(-\partial_x^2 + m^2\cos\phi_0)\psi_n = \omega_n^2\psi_n . \qquad (19)$$

Then we can expand

$$\tilde{\phi}_1 = \sum_n b_n(t)\psi_n(x), \qquad (20)$$

and the coefficients $b_n(t)$ obey

$$\frac{d^2}{dt^2}b_n + \omega_n^2 b_n = -g_n, \qquad (21)$$

$$g_n = \int_{-\infty}^{+\infty} dx\left[-\frac{\partial\Delta V}{\partial\phi}\bigg|_{\phi=\phi_0} + (\partial_t^2 - \partial_x^2 + m^2\cos\phi_0)F\right]\psi_n(x). \qquad (22)$$

As is well known, the stability equation possesses a zero eigenvalue ($\omega_0 = 0$) whose eigenfunction is proportional to ϕ_{0x},

$$\psi_0 = A\phi_{0x} .$$

The corresponding g_0 is obtained from (22) as

$$g_0 = A(\Delta V_+ - \Delta V_-) + \tilde{g}_0,$$

$$\tilde{g}_0 = A\int_{-\infty}^{+\infty} dx\,[(\partial_t^2 - \partial_x^2 + m^2\cos\phi_0)F\,]\phi_{0x} . \qquad (23)$$

Let us examine the integrand in Eq. (23) for $x\to\pm\infty$; using Eqs. (12)–(17), we find that

$$(\text{integrand}) \to \pm m\Delta V_\pm \quad (x\to\pm\infty).$$

Therefore $\tilde{g}_0$ is finite, and perturbation theory works to first order, only if

$$\Delta V_+ = \Delta V_- = 0.$$

This is the content of the first theorem.

Theorem I: The modified sine–Gordon equation (1) has a finite energy solution analytic in λ which reduces to the static soliton ϕ_0 only if

$$\lim_{x\to+\infty}\Delta V(\phi_0) = \lim_{x\to-\infty}\Delta V(\phi_0) = 0. \qquad (24)$$

The above condition is necessary, but not sufficient. Of course, the important part of Eq. (24) is the equality of the two limits; their value always can be adjusted to zero by adding a constant to ΔV.

The condition (24) arises also as the necessary condition for the existence of a *time-independent* correction ϕ_1; indeed, for ϕ_1 independent of t, Eq. (5) becomes

$$(-\partial_x^2 + m^2\cos\phi_0)\phi_1 = -\frac{\partial\Delta V}{\partial\phi}\bigg|_{\phi=\phi_0} .$$

The equation has solutions if the right-hand side is orthogonal to the zero-eigenvalue eigenfunction of the operator on the left-hand side. This eigenfunction is proportional to ϕ_{0x}; therefore one needs the condition

$$0 = \int_{-\infty}^{+\infty} dx\,\phi_{0x}\frac{\partial\Delta V}{\partial\phi}\bigg|_{\phi=\phi_0} = \Delta V_+ - \Delta V_-,$$

which is the same as (24).

The second result we wish to present depends on the boson transformation method of constructing solutions to nonlinear equations. For the present case, it is sufficient to examine the Yang–Feldman equation corresponding to Eq. (1):

$$\phi(x,t) = f(x,t) + \int d^2y\, G(x-y, t-t_y) \times\left[-m^2\sin\phi - \lambda\frac{\partial\Delta V}{\partial\phi} - \mu^2\phi\right]. \qquad (25)$$

In this equation, f and G obey

$$(\partial^2 + \mu^2)f = 0,$$

$$(\partial^2 + \mu^2)G(x,t) = \delta(x)\delta(t).$$

In order for a perturbative solution of Eq. (25) to exist, μ must be chosen so that the quantity in the square bracket contains no term linear in ϕ. If ΔV contains a quadratic (mass) term, this implies that μ^2 is a function of λ. On the other hand, the solution of (25) will depend on x through the combination μx, and therefore an expansion in λ will involve powers of x and will become nonperturbative for large $|x|$. This is the content of our second theorem.

Theorem II: If the perturbing potential ΔV contains a quadratic term, the modified sine–Gordon equation (1) has no perturbative solution.

The analogy between this purely classical result and Haag's theorem in quantum field theory is amusing.

The remainder of this paper is devoted to examples illustrating the above theorems.

An example satisfying the conditions of both theorems is provided by the potential

$$\Delta V = \tfrac{1}{2}m^2(t - \cos\phi)^2.$$

Equation (1) has the static solution

$$\phi = -2\cot^{-1}(\sqrt{1+\lambda}\,\sinh mx),$$

whose expansion in powers of λ is

$$\phi = \tan^{-1}e^{mx} + \lambda\frac{\tanh mx}{\cosh mx} - \lambda^2(1 + 2\tanh^2 mx)\frac{\tanh mx}{4\cosh mx} + \cdots .$$

The coefficients of the expansion are well behaved, bounded functions, of x. The energy associated with this solution is

$$E = 4m[1 + [(1+\lambda)/\sqrt{\lambda}]\sin^{-1}\sqrt{\lambda/(1+\lambda)}],$$

and has also a good perturbative expansion

$$E = 8m[1 + \tfrac{1}{3}\lambda + \tfrac{1}{15}\lambda^2 + \cdots].$$

An example which violates the assumptions of Theorem I is given by the double-sine–Gordon equation,[2] for which

$$\Delta V = \tfrac{1}{2}m^2(1 - \cos(\phi/2))^2.$$

The static solution of Eq. (1) is

$$\tan(\phi/4) = -2/(\sqrt{\lambda}\sinh mx).$$

Clearly, the solution has no perturbative expansion in λ. Furthermore, the classical energy

$$E = 8m\left[2 + \frac{\lambda}{4\sqrt{1-\lambda/4}}\ln\left(\frac{8-\lambda+8\sqrt{1-\lambda/4}}{\lambda}\right)\right]$$

has no perturbative expansion either. However, the limit

$$\lim_{x\to 0} E = 16m$$

exists and is twice the soliton mass. Finally, we present an example which violates the conditions of Theorem II:

$$\Delta V = -\tfrac{1}{2}m^2\sin^2\phi.$$

Notice that the "mass" (coefficient of the bilinear term) is now given by

$$\mu = m\sqrt{1-\lambda}.$$

The solution is

$$\phi = 4\tan^{-1}\{(\sqrt{1-\lambda+\sinh^2\mu x} + \sinh\mu x)/\sqrt{1-\lambda}\}.$$

The expansion in powers of λ looks like

$$\phi = 4\tan^{-1}e^{mx} - \frac{2}{1+e^{2mx}}[mx - \tanh mx]\lambda + O(\lambda^2).$$

For x very large and negative, the second term dominates the first for any value of λ and perturbation theory breaks down. On the other hand, the energy

$$E = 4m\{\sqrt{1-\lambda} + (1/\sqrt{\lambda})\sin^{-1}\sqrt{\lambda}\}$$

has a good expansion in powers of λ:

$$E = 8m\{1 - \tfrac{1}{6}\lambda - \tfrac{1}{40}\lambda^2 + \cdots\}.$$

This is consistent with the fact that energy considerations did not enter in the proof of Theorem II.

ACKNOWLEDGMENTS

One of the authors (NJP) wishes to acknowledge the hospitality received from the Institute of Theoretical Physics at The University of Alberta where this work was done.

This work was supported by the Natural Sciences and Engineering Research Council, Canada and the Dean of Science, Faculty of Science, The University of Alberta, Edmonton, Alberta, Canada.

[1]D. W. McLaughlin and A. C. Scott, Phys. Rev. A **18**, 1652 (1978); V. I. Karpman and V. V. Solov'ev, "The influence of external perturbations on solitons in Josephon junctions," IZMIRAN Preprint No. 28, Moscow, 1980, p. 294; M. Salerno and A. C. Scott, Phys. Rev. B **26**, 2474 (1982); M. Salerno, M. P. Soerensen, O. Skovgaard, and P. L. Christiansen, Wave Motion **5**, 49 (1983); V. I. Karpman and V. V. Solov'ev, Physica D **3**, 487 (1981); Y. Leblanc, H. Matsumoto, H. Umezawa, and F. Mancini, Phys. Rev. B **30**, 5958 (1984).

[2]T. Uchiyama, Phys. Rev. D **14**, 3520 (1976).

Wavelength-dependent electromagnetic parameters for coherent propagation in correlated distributions of small-spaced scatterers

Victor Twersky
Mathematics Department, University of Illinois, Chicago, Illinois 60680

(Received 19 October 1984; accepted for publication 26 April 1985)

Earlier results for coherent propagation of electromagnetic waves in pair-correlated random distributions of scatterers (of radius a and physical parameters ϵ', μ') with minimum separation of centers $b \geqslant 2a$ small compared to wavelength $(2\pi/k)$ are generalized to obtain polarization, refraction, and absorption terms to order k^2. The development includes multiple scattering and multipole coupling by electric and magnetic dipoles, as well as quadrupoles to appropriate order. The correlation aspects are determined by simple integrals of the statistical mechanics radial distribution function f for impenetrable particles (spheres, cylinders, and slabs) of diameter b. For slab scatterers, in terms of the exact Zernike–Prins f, the correlation integrals are expressed as algebraic functions of the volume fraction w; the resultant bulk values reduce to those of one particle at full packing, $w = 1$. Similar results are obtained for spheres in terms of the Wertheim–Thiel solution of the Percus–Yevick approximation of f at the unrealizable bound $w = 1$.

I. INTRODUCTION

We use earlier forms for the coherent electromagnetic field (the ensemble average) in correlated random distributions[1,2] of scatterers (with radius a, volume v, and physical parameters ϵ', μ') to obtain additional explicit terms of the associated bulk index of refraction ($\eta^2 = \epsilon\mu$) and bulk parameters (ϵ, μ) for minimum separation $(b \geqslant 2a)$ of centers small compared to wavelength $(\lambda = 2\pi/k)$. The results are limited to applications specified by the average number ρ of centers per unit volume and by the statistical mechanics radial distribution function $f(r)$ for impenetrable particles of diameter b.

The earlier explicit approximations for spheres, cylinders, and slabs ($m = 3,2,1$, respectively) consist of essentially two terms.[1] The leading term, corresponding to refraction (or polarization) and absorption, is independent of k (at least explicitly) and exhibits distributional aspects only in the volume fraction $w = \rho v$ occuped by scatterers. The second term, corresponding to incoherent scattering losses, depends as well on $(ka)^m$ and on the low frequency limit of the structure factor $\mathscr{W}(W)$ with $W = w(b/2a)^m$ as the volume fraction of cocentered (electromagnetically transparent) statistical particles. The present sequel provides $(ka)^2$ corrections which include additional correlation integrals $\mathscr{N}(W)$ and multipole coupling between electric and magnetic dipoles as well as quadrupoles. The development for $m = 3$ (spheres) is relatively detailed because no comparable results exist, but for $m = 2$ and 1 (normal incidence on parallel cylinders and parallel slabs), we need only reinterpret recent[3] acoustic results. If μ' or ϵ' equals unity, then η^2 equals ϵ or μ, as analyzed for the simpler one-parameter optical cases.[4]

Using existing statistical mechanics results[5–14] for f, the packing functions $\mathscr{W}_m$ and $\mathscr{N}_m$ can be evaluated numerically,[9,10] but simple closed forms are available for all but $\mathscr{N}_2$. The closed forms for $m = 1$ (polynomials in W) are exact, and those for $m = 3$ and for $\mathscr{W}_2$ (ratios of polynomials) are approximate. For $W = w$, the explicit approximations of the bulk values for slabs reduce to the single particle values at $w = 1$ (full packing) as required physically. For spheres, although only $w \lesssim 0.63$ is realistic,[14] the behavior of the approximations at the unrealizable $w = 1$ is the same as for slabs (which appears consistent with the implicit statistical mechanics and multiple scattering approximations). We also find that the correction and scattering loss terms of η for the extreme case of perfectly conducting spheres vanish at the unrealizable bound. For cylinders, only $w \lesssim 0.84$ is realistic,[15] but we use the unrealizable bound to infer the first two terms of a consistent form of $\mathscr{N}_2$ for large w.

The present approximations for larger kb than before,[1,2] plus recent asymptotic results[16] for large kb, provide relatively simple forms that display the explicit dependence on all parameters for many practical applications. In these ranges of kb, the results help delineate the fundamental physical processes, and obviate elaborate machine computations.

Sections II and III introduce notation, list available results for spheres,[1,4,17–20] and outline the scope of the paper. The development is based on representation theorems[1,2,21] for ϵ and μ, and an approximation for η obtained from the ensemble average of the functional equation relating the multiple and isolated scattering amplitudes of particles in an arbitrary configuration.[20] The theorems follow from the first of the system of hierarchy integrals of the ensemble, and the approximation from essentially the second. Replacing the average scattering amplitude with two particles fixed by that with one fixed, analogous to Lax's procedure[22] for the effective exciting field, truncated the system and led directly to a determinate functional equation[2,1] for η. Lax's original interpretation,[22] comparison of the iterated expansion of such truncated systems with the average of the successive scattering series,[23] and Keller's procedures[24] not based on hierarchy integrals, provide insight for the closure approximation. Analytical aspects, essentially as for the analogous statistical mechanics problems,[5] are unresolved.

See Refs. 1 and 2 for detailed derivations of the forms we analyze, as well as for simplifications introduced to delineate relations to earlier results of Rayleigh,[25] Reiche,[26] Foldy,[27] and Lax.[22] Such simplifications were not used in the subse-

 0022-2488/85/092208-10$02.50

quent analytical developments,[1-4] nor are they used in the present sequel which derives the corresponding $(ka)^2$ terms for the two-parameter electromagnetic cases.

In the following, we use (1:45) for Eq. (45) of Ref. 1, etc., and modify the earlier notation slightly to suppress numerical factors or emphasize key parameters.

II. PRELIMINARY CONSIDERATIONS

For a slab-region distribution and a normally incident wave $\phi e^{-i\omega t}$ (representing either the electric or magnetic component) we write

$$\phi = \hat{\mathbf{x}} e^{ikz}, \quad k = 2\pi/\lambda = 2\pi\eta_0/\lambda_0, \quad \eta_0^2 = \epsilon_0\mu_0, \tag{1}$$

with η_0 as the index of refraction of the embedding medium. The corresponding bulk values

$$K = k\eta_b/\eta_0 = k\eta, \quad \eta_b^2 = \epsilon_b\mu_b, \quad \eta^2 = \epsilon\mu \tag{2}$$

specify propagation of the ensemble averaged[1] field $\langle\Psi\rangle$. Using (1:113) for η and (1:45) and (1:46) for ϵ and μ, and the known scattering amplitude for an isolated particle, we express the bulk relative values in terms of $w, f, x \equiv ka$, and the particles' relative parameters

$$\epsilon' = \epsilon_p/\epsilon_0, \quad \mu' = \mu_p/\mu_0, \quad \eta'^2 = \epsilon'\mu', \qquad \epsilon' - 1 \equiv \delta_e, \quad \mu' - 1 \equiv \delta_m, \tag{3}$$

which may depend implicitly on ω. These are complex in general, but to facilitate discussion we call the δ's decrimants and use the term depolarization (electric or magnetic) as if they were positive.

Essentially as before,[1] we work with

$$\langle\Psi\rangle = \begin{Bmatrix}\mathbf{E}\\ \mathbf{H}\end{Bmatrix}, \quad C = \begin{Bmatrix}\epsilon\\ \mu\end{Bmatrix}, \quad B = \begin{Bmatrix}\mu^{-1}\\ \epsilon^{-1}\end{Bmatrix}, \quad \eta^2 = \frac{C}{B}, \tag{4}$$

as well as with $\overline{C} = B^{-1}$ for a complementary parameter; the same holds for the corresponding particle parameters C', B'. We refer to the E case ($C = \epsilon$, $\overline{C} = B^{-1} = \mu$) for nomenclature, but use collective forms or either case when convenient. Since existing[3] explicit results (η^2, C, B) for slabs and cylinders may be applied to electromagnetics (and we exhibit the forms subsequently for comparison), we emphasize the case of spheres.

For a distribution of either electric or magnetic dipoles, to lowest orders in x for the real and imaginary parts inclusive of incoherent scattering losses, from (1:89) we have the one-parameter form $\eta^2 = C$ equalling ϵ or μ with

$$C = C_1 + iC_s(x^3) + O(x^2), \quad x = ka. \tag{5}$$

We write the k-independent C_1, given originally by Maxwell[17] and derived analytically for a cubic lattice by Rayleigh[18] (i.e., the Clausius–Massotti result), in terms of $w = \rho 4\pi a^3/3$ as

$$C_1 = 1 + w\delta/D, \quad \delta = C' - 1, \qquad D = 1 + (1-w)\delta/3 = D(\delta), \tag{6}$$

with D as the depolarization denominator (for $\delta > 0$). The corresponding incoherent scattering loss term

$$C_s = x^3 w \mathscr{W} 2\delta^2/9D^2 \tag{7}$$

depends on the packing factor determined by the second moment of the total correlation function[5] $F = f - 1$,

$$\mathscr{W} = 1 + 4\pi\rho\int_0^\infty F(r) r^2\,dr, \quad F(r) = f(r) - 1. \tag{8}$$

From the scaled particle equation of state[6] or from the Wertheim–Thiel solution of the Percus–Yevick equation,[7]

$$\mathscr{W} \approx (1-W)^4/(1+2W)^2, \quad W = w(b/2a)^3, \qquad w = \rho 4\pi a^3/3, \tag{9}$$

as discussed earlier in detail.[14-15]

More generally, for the one-parameter case (either $\mu' = 1$ or $\epsilon' = 1$) correct to $O(x^3)$, from (4:41) based on (1:113) for scatterers consisting of dominant electric dipoles plus weak magnetic dipoles and electric quadrupoles, we have $\eta^2 = C$ with

$$C = C_1 + C_c + iC_s + O(x^4), \tag{10}$$

$$C_c = -x^2 w\delta\left\{\frac{1}{D^2}\left[\frac{1-\delta}{5} + \frac{\delta N}{9}\left(2 + \frac{C_1}{5}\right)\right] - \frac{C_1}{10} - \frac{C_1(2C_1+3)^2}{50(2C'+3)}\right\}. \tag{11}$$

The correction term $C_c = O(x^2)$ depends on the first moment of F,

$$N = -4\pi\rho a\int_0^\infty F(r) r\,dr, \tag{12}$$

with closed form[7,8]

$$N \approx \frac{2a}{b}\frac{6W}{1+2W}\left[1 - \frac{W}{5} + \frac{W^2}{10}\right]. \tag{13}$$

For the unrealizable value $W = w = 1$, the results based on (9) and (13) reduce to $C = C'$, which appears consistent with the implicit statistical mechanics and multiple scattering closure approximations. Note that N of (4:41) should be multiplied by δ.

If each scatterer is both an electric plus magnetic dipole, then from (1:96) to lowest orders in x for the real and imaginary parts,

$$\eta^2 = (\epsilon_1 + i\epsilon_s)(\mu_1 + i\mu_s) + O(x^2) = \eta_1^2 + i\eta_s^2 + O(x^2),$$

with

$$\eta_1^2 = \epsilon_1\mu_1 = (1 + w\delta_e/D_e)(1 + w\delta_m/D_m), \qquad D_e = D(\delta_e), \quad D_m = D(\delta_m), \tag{14}$$

$$\eta_s^2 = \epsilon_1\mu_s + \mu_1\epsilon_s = \eta_1^2(\epsilon_s/\epsilon_1 + \mu_s/\mu_1), \qquad \frac{\eta_s^2}{\eta_1^2} = \frac{x^3 w\mathscr{W}2}{9}\left(\frac{\delta_e^2}{\epsilon_1 D_e^2} + \frac{\delta_m^2}{\mu_1 D_m^2}\right), \tag{15}$$

in terms of the forms (6) and (7). The attenuation via incoherent scattering losses is given by $\mathrm{Re}(\eta_s^2/2\eta_1)$, approximately.

In the following, we apply (1:113) to scatterers consisting of strong electric and magnetic dipoles and weak quadrupoles to obtain

$$\eta^2 = \eta_1^2 + \eta_c^2 + i\eta_s^2 + O(x^4), \tag{16}$$

with the correction η_c^2 of $O(x^2)$. Using (16) and $\eta^2 = \epsilon\mu$ with each bulk parameter in the form (10), we require

$$\eta_c^2 = \epsilon_1\mu_c + \mu_1\epsilon_c, \quad \frac{\eta_c^2}{\eta_1^2} = \frac{\epsilon_c}{\epsilon_1} + \frac{\mu_c}{\mu_1}, \tag{17}$$

where each correction term depends on both particle parameters ϵ' and μ'. However, a simple decomposition of the explicit result for η_c^2 does not determine ϵ_c and μ_c. Although η^2 must be symmetrical in ϵ' and μ', and the interchange of ϵ' and μ' in ϵ must produce μ, direct factorization of η^2 as a product form does not yield results for the parameters satisfying the mean value theorems (1:40) and (1:41). The theorems require that $C-1$ be proportional to $C'-1$ (so that if one of the particle's relative parameters is unity, then so is the corresponding bulk parameter), but terms of ϵ and μ that insure this behavior cancel in the product $\epsilon\mu = \eta^2$. To obtain the explicit x^2 corrections for the bulk parameters, we use the theorems to construct electromagnetic analogs of the determinate forms (2:99) and (2:100) introduced for the scalar problems (and applied recently[3] in detail).

The next section summarizes required aspects of the development of the determinate electromagnetic equation for η for arbitrary sized spheres, and derives corresponding results for C and B. Subsequent sections consider applications and comparisons with slabs and cylinders.

III. GENERAL RESULTS FOR SPHERES

From (1:82), we write the forward ($\hat{r} = \hat{k} = \hat{\mathbf{z}}$) value of the dyadic scattering amplitude of an isolated sphere as

$$\tilde{g}(\mathbf{k},\mathbf{k}) = (\tilde{\mathbf{I}} - \hat{\mathbf{z}}\hat{\mathbf{z}})g(k,k), \quad g(k,k) = \sum_{n=1}^{\infty} (b_n + c_n), \tag{18}$$

where $\tilde{\mathbf{I}}$ is the unit dyadic, and the two sets of scattering coefficients b_n and c_n (electric and magnetic multipoles, respectively, for the E case) are well known.[19] Collectively, for $a_n = b_n, c_n$,

$$a_n = \frac{n(n+1)}{2}\frac{a_n'}{1 - a_n'/d_n}, \quad d_n = \frac{2n+1}{n(n+1)}, \tag{19}$$

with $a_n' = a_n'(B')$ in the form

$$a_n' = id_n \frac{j_n(y)[xj_n(x)]' - j_n(x)[yj_n(y)]'B'}{j_n(y)[xn_n(x)]' - n_n(x)[yj_n(y)]'B'}, \tag{20}$$

where $y = \eta' x$, and the prime on a bracket indicates differentiation with respect to argument. For the E case, $a_n'(1/\epsilon') = b_n'$, $a_n'(1/\mu') = c_n'$; we obtain corresponding coefficients for the perfect conductor by the formal procedure of letting $\epsilon' \to \infty$, $\mu' \to 0$, and regarding η' as bounded.

The development for the pair-correlated distribution involved the average of the general functional equation (20:127) relating the multiple and isolated scattering amplitudes of particles in an arbitrary configuration. We replaced[1] the ensemble averaged multiple scattering amplitude with two particles fixed by that with one fixed to obtain the determinate functional equation (1:65) for η, and then expanded the amplitudes in spherical harmonics to construct a corresponding algebraic system for the scattering coefficients. From (1:84), we write the multiple scattering analog of (18) as

$$g(k\,|K) = \sum (B_n + C_n), \tag{21}$$

such that the homogeneous algebraic system (1:86) for the coefficients B_n, C_n in terms of the known b_n, c_n and the correlation integrals $\mathscr{H}_n$ of (1:80) determines η and all but one of the coefficients.

To facilitate manipulation and interpretation, (1:111)–(1:113) recast the original system as an inhomogeneous system in terms of normalized coefficients

$$P_n = b_n\eta^{2n}\Big[1 + \sum (P_m h_{nm} + M_m \bar{h}_{nm})\Big],$$
$$M_n = c_n\eta^{2n}\Big[1 + \sum (P_m \bar{h}_{nm} + M_m h_{nm})\Big], \tag{22}$$

$$-\frac{(\eta^2 - 1)}{c} = \sum (P_n + M_n), \quad c = \frac{ip4\pi}{k^3} = \frac{i3w}{x^3}, \tag{23}$$

$$\frac{P_n}{B_n} = \frac{M_n}{C_n} = -\frac{(\eta^2-1)\eta^n}{c\Sigma(B_m + C_m)\eta^m}. \tag{24}$$

The coupling factors h_{nm} for like poles (PP,MM) and $\bar{h}_{nm}$ for unlike (PM,MP) are given in (1:114) in terms of η, c, and $\mathscr{H}_n$. For detailed considerations, we suppress self-coupling in the sense $a_n a_n$ by the form

$$\mathscr{A}_n\eta^{2n} = a_n\eta^{2n}/(1 - a_n\eta^{2n}h_{nn})$$
$$= \{\mathscr{B}_n\eta^{2n}, \mathscr{C}_n\eta^{2n}\} = \{p_n, m_n\} \tag{25}$$

and work with the coefficients $\mathscr{B}_n(b_n)$, $\mathscr{C}_n(c_n)$ as well as p_n, m_n. See (1:88)–(1:110) and (4:36)–(4:41) for earlier applications.

The bulk index η is fully determined by (22) and (23), but additional relations are required to determine the bulk parameters as well as the remaining coefficient of (21). To construct the relations, we represent g of (18) as an integral over the sphere's volume (v) in terms of the internal field $\psi(\eta'k) = \psi(K') = \tilde{\psi}\cdot\hat{\mathbf{x}}$ and the dyadic plane wave $\tilde{\phi}(k) = (\tilde{\mathbf{I}} - \hat{\mathbf{z}}\hat{\mathbf{z}})e^{-ikz}$ by the form

$$g(k,k) = \frac{ik}{4\pi}\hat{\mathbf{x}}\cdot\int [(C'-1)k^2\tilde{\phi}(k)\cdot\psi(K')$$
$$+ (B'-1)(\nabla\times\tilde{\phi})\cdot(\nabla\times\psi)]dv$$
$$\equiv [\tilde{\phi}(k), \psi(K')]. \tag{26}$$

The integral is a special case of (1:23), and the present symbolic form is the x component of the original. If we replace k by K in $\tilde{\phi}$, then

$$g(K,k) = [\tilde{\phi}(K), \psi(K')] = \sum (\bar{b}_n + \bar{c}_n),$$
$$\bar{b}_n \equiv \beta_n b_n, \quad \bar{c}_n \equiv \gamma_n c_n. \tag{27}$$

To obtain the ratios β_n, γ_n of isolated scattering coefficients we decompose the general dyadic version of the volume integral into two terms, and convert them to surface integrals by

$$(K^2 - K'^2)\int \tilde{\phi}(\mathbf{K})\cdot\tilde{\psi}(K')dv$$
$$= \int [(\tilde{\phi}\times\hat{n})\cdot(\nabla\times\tilde{\psi}) - (\nabla\times\tilde{\phi})\cdot(\hat{n}\times\tilde{\psi})]ds$$
$$\equiv S_1 - S_2, \tag{28}$$

$$(K^2 - K'^2)\int (\nabla\times\tilde{\phi})\cdot(\nabla\times\tilde{\psi})dv = -K^2S_1 + K'^2S_2, \tag{29}$$

where $\tilde{\Phi}(\mathbf{K}) = (\tilde{\mathbf{I}} - \hat{K}\hat{K})e^{-i\mathbf{K}\cdot\mathbf{r}}$ and $ds = a^2\, d\Omega\,(\hat{r})$. Using (20:38) and (20:97) for $\tilde{\Phi}$ and $\tilde{\psi}$, and the orthogonality relations (20:33), we construct

$$\begin{aligned}
&\gamma_n = V_n(\eta,\eta')/V_n(1,\eta'),\\
&V_n(\eta,\eta') = -(C'-1)I_n(\eta,\eta') + (B'-1)L_n(\eta,\eta'),\\
&\beta_n = W_n(\eta,\eta')/W_n(1,\eta'),\\
&W_n(\eta,\eta') = -[(C'-1)/\eta\eta']L_n(\eta,\eta')\\
&\qquad\qquad + (B'-1)\eta\eta' I_n(\eta,\eta'),\\
&(\eta^2-\eta'^2)\begin{Bmatrix} I_n \\ L_n \end{Bmatrix} = \begin{Bmatrix} 1 \\ \eta^2 \end{Bmatrix} J_n(\eta,\eta') - \begin{Bmatrix} 1 \\ \eta'^2 \end{Bmatrix} J_n(\eta',\eta),\\
&J_n(\eta,\eta') = j_n(\eta x)[\eta' x j_n(\eta' x)]'.
\end{aligned} \tag{30}$$

These expressions plus the representation theorems (1:40)–(1:47) help make all aspects of the multiple scattering problem determinate.

For the special case of spheres we need consider only simplified versions of (1:45) and (1:46)

$$\begin{aligned}
&C-1 = (C'-1)\rho\hat{\mathbf{x}}\cdot\int \tilde{\Phi}(K)\cdot\Psi(K')dv,\\
&\tilde{\Phi}(K) = (\tilde{\mathbf{I}} - \hat{\mathbf{z}}\hat{\mathbf{z}})e^{-iKz},
\end{aligned} \tag{31}$$

$$\eta^2(B-1) = -(B'-1)\frac{\rho}{k^2}\hat{\mathbf{x}}\cdot\int(\nabla\times\tilde{\Phi})\cdot(\nabla\times\Psi)dv, \tag{32}$$

where (except for translational factors) $\Psi = \tilde{\Psi}\cdot\hat{\mathbf{x}}$ represent the ensemble averaged internal field within one fixed sphere.

Subtracting (32) from (31), we obtain the corresponding version of (1:47)

$$\begin{aligned}
\eta^2 - 1 = \frac{\rho}{k^2}\hat{\mathbf{x}}\cdot\int&[(C'-1)k^2\tilde{\Phi}\cdot\Psi\\
&+(B'-1)(\nabla\times\tilde{\Phi})\cdot(\nabla\times\Psi)]dv,
\end{aligned} \tag{33}$$

with the integral proportional to $[\tilde{\Phi}(K),\Psi(K')]$ in the form defined in (26). Thus (33) relates η^2 to a multiple scattered analog of (27)

$$\begin{aligned}
&-\frac{(\eta^2-1)}{c} = g(K\,|K) = \sum(\overline{B}_n + \overline{C}_n),\\
&\overline{B}_n = \beta_n B_n, \quad \overline{C}_n = \gamma_n C_n,
\end{aligned} \tag{34}$$

in terms of β_n, γ_n of (30); this additional relation for η^2 helps determine all the multiple scattering coefficients.

From (31) and (32) in terms of (28)–(30), the bulk parameters are determined by

$$\begin{aligned}
&\frac{C-1}{c} = \sum(B_n\beta_n^C + C_n\gamma_n^C),\\
&\beta_n^C = (C'-1)\frac{L_n(\eta,\eta')}{\eta\eta' W_n(1,\eta')},\\
&\gamma_n^C = (C'-1)I_n(\eta,\eta')/V_n(1,\eta'),
\end{aligned} \tag{35}$$

$$\begin{aligned}
&\eta^2[(B-1)/c] = \sum(B_n\beta_n^B + C_n\gamma_n^B),\\
&\beta_n^B = (B'-1)\eta\eta'[I_n(\eta,\eta')/W_n(1,\eta')],\\
&\gamma_n^B = (B'-1)[L_n(\eta,\eta')/V_n(1,\eta')],
\end{aligned} \tag{36}$$

such that $\beta_n = -\beta_n^C + \beta_n^B$, $\gamma_n = -\gamma_n^C + \gamma_n^B$.

Using (34) and (23) in (24), we construct

$$\frac{P_1}{B_1\eta} = \frac{\Sigma(P_n\beta_n + M_n\gamma_n)/\eta^n}{\Sigma(P_m + M_m)} \equiv \frac{1}{R}, \tag{37}$$

which with (24) determines all coefficients B_n, C_n in terms of P_n, M_n, and enables us to rewrite (35) and (36) as

$$C - 1 = cR\sum(P_n\beta_n^C + M_n\gamma_n^C)(\eta^{-n}), \tag{38}$$

$$\eta^2(B-1) = cR\sum(P_n\beta_n^B + M_n\gamma_n^B)(\eta^{-n}). \tag{39}$$

The analogous development for the scalar case (2:92)–(2:100) is simpler in that only one set of multipoles and one set of ratios (corresponding essentially to the γ set) are involved.

These results, and the relation $\eta^2 = C/B$ plus the two sets in (4), provide for alternative derivations (as well as for checks) of mutually consistent results for the bulk values η^2, ϵ, and μ in an unbounded distribution. See Refs. 1, 2, and 21 for additional discussion and for comparison with interface (equivalent slab) approximations.

IV. DIPOLES PLUS QUADRUPOLES

To evaluate P_n, M_n of (22) for spheres specified by the four isolated scattering coefficients b_n, c_n represented by $n = 1,2$ in (19), we require only the coupling factors

$$\begin{aligned}
&h_{11} = (2c + 2\mathscr{H}_0 + \mathscr{H}_2)/3\eta^2,\\
&\overline{h}_{11} = c/\eta(\eta+1) + \mathscr{H}_1/\eta^2,\\
&h_{12} = (3c\eta + 3\mathscr{H}_1 + 2\mathscr{H}_3)/5\eta^3,\\
&\overline{h}_{12} = c/\eta(\eta+1) + \mathscr{H}_2/\eta^3,\\
&h_{22} = [c(14+19\eta^2) + 14\mathscr{H}_0 + 5\mathscr{H}_2 + 16\mathscr{H}_4]/35\eta^4,
\end{aligned} \tag{40}$$

in terms of the correlation integrals of (2:148)

$$\begin{aligned}
&\mathscr{H}_n = 4\pi\rho\int_0^\infty F(r)j_n(Kr)h_n^{(1)}(kr)r^2\,dr,\\
&F(r) = f(r) - 1.
\end{aligned} \tag{41}$$

The isolated scattering coefficients

$$a_1 = a_1'/(1 - a_1' 2/3), \quad a_2 = 3a_2'/(1 - a_2' 6/5) \tag{42}$$

are given to $O(x^6)$ by

$$\begin{aligned}
&a_1' = ix^3(\delta - x^2 t)/3(1 + \delta/3 - x^2 d) + O(x^7),\\
&\quad 10t = (\eta'^2 + 1)\delta - C'\overline{\delta},\\
&\quad 30d = (10 + \eta'^2)\delta + 5C'\overline{\delta},
\end{aligned} \tag{43}$$

$$a_2' = ix^5\delta/18(2C'+3) + O(x^7). \tag{44}$$

Here $\delta = C' - 1$, and $\overline{\delta} = 1/B' - 1 = \overline{C}' - 1$ is the complement; for the E case, we have $\delta = \epsilon' - 1$ and $\overline{\delta} = \mu' - 1$ for b_n', and conversely for c_n'. The remaining a_n' are $O(x^{2n+1})$.

For small-spaced scatterers (small $kb = xb/a$), from (2:149)

$$\begin{aligned}
&\mathscr{H}_0 = \mathscr{W} - 1 + iN/x + O(x),\\
&\mathscr{H}_n = i\eta^n N/x(2n+1) + O(x),
\end{aligned} \tag{45}$$

with $\mathscr{W}, N$ as in (8) and (12). To $O(x^3)$ for the bulk values, we need retain only

$$h_{11} \approx \frac{1}{3\eta^2}\left[2c + 2(\mathscr{W} - 1) + \frac{iN}{x}\left(2 + \frac{\eta^2}{5}\right)\right],$$

$$\bar{h}_{11} \approx c/\eta(\eta + 1) + iN/3x\eta, \tag{46}$$

$$h_{12} \approx 3c/5\eta^2 \equiv h, \quad \bar{h}_{12} \approx c/\eta(\eta + 1) \equiv \bar{h}.$$

From (42) and (46) we write $\mathscr{A}_1$ of (25) initially as

$$\mathscr{A}_1 = a_1/(1 - a_1\eta^2 h_{11})$$
$$= a_1' \left\{1 - \frac{a_1'}{3}\left[\frac{i6w}{x^3} + \frac{iN}{x}\left(2 + \frac{\eta^2}{5}\right) + 2\mathscr{W}\right]\right\}^{-1}$$
$$+ O(x^7). \tag{47}$$

Substituting (43) we obtain to $O(x^3)$ in terms of C_1 and C_s of (6) and (7)

$$c\mathscr{A}_1 = -\frac{w\delta}{DC_1} + \frac{wx^2}{(DC_1)^2}\left[\mathscr{I} + \frac{\delta^2 N}{9}\left(2 + \frac{\eta^2}{5}\right)\right] - i\frac{C_s}{C_1^2}, \tag{48}$$

$$\mathscr{I} = tDC_1 - \delta(d + 2wt/3) = \delta(1-\delta)/5 - C'^2\bar{\delta}/10.$$

Equivalently, with η_1^2 as in (14)

$$1 + c\mathscr{A}_1 = (C_1 - x^2 T + iC_s)^{-1}$$
$$= \frac{1}{C_1} + x^2\frac{T}{C_1^2} - i\frac{C_s}{C_1^2} + O(x^4),$$

$$T = \frac{w}{D^2}\left\{\delta\left[\frac{1-\delta}{5} + \frac{\delta N}{9}\left(2 + \frac{\eta_1^2}{5}\right)\right] - \bar{\delta}\frac{C'^2}{10}\right\} \tag{49}$$
$$\equiv \delta V + \bar{\delta} R,$$

where T consists of essentially two terms, one proportional to δ and one to the complement $\bar{\delta}$. The corresponding quadrupoles are given by

$$\mathscr{A}_2 = 3a_2' + O(x^{10})$$
$$= ix^5\delta/6(2C' + 3) + O(x^7), \tag{50}$$
$$c\mathscr{A}_2 = -wx^2\delta/2(2C' + 3) + O(x^4).$$

Thus from (22) to $O(x^6)$ in terms of $\{\mathscr{B}_n\eta^{2n}, \mathscr{C}_n\eta^{2n}\} = \{p_n, m_n\}$ of (25)

$$P_1 = p_1(1 + m_1\bar{h}_{11} + \mathscr{P})/\mathscr{D},$$
$$M_1 = m_1(1 + p_1\bar{h}_{11} + \mathscr{M})/\mathscr{D}, \quad \mathscr{D} = 1 - p_1 m_1 \bar{h}_{11}^2,$$
$$\mathscr{P} = P_2(h_{12} + m_1\bar{h}_{11}\bar{h}_{12}) + M_2(\bar{h}_{12} + m_1\bar{h}_{11}h_{12})$$
$$\approx P_2(h + m_1\bar{h}^2) + M_2\bar{h}(1 + m_1 h), \tag{51}$$
$$P_2 = p_2[\mathscr{D} + p_1(1 + m_1\bar{h}_{11})h_{12}$$
$$+ m_1(1 + p_1\bar{h}_{11})\bar{h}_{12}]/\mathscr{D}$$
$$\approx p_2(1 + p_1 h)(1 + m_1\bar{h})/(1 - p_1 m_1\bar{h}^2).$$

Corresponding results for $\mathscr{M}, M_2$ follow by interchanging p and m in $\mathscr{P}, P_2$. To the required accuracy in P_2 and M_2 forms we retain only the leading terms of p_1 and m_1 factors and use $\bar{h}$ for $\bar{h}_{11}$ as well as for $\bar{h}_{12}$.

V. BULK PROPAGATION INDEX

Substituting (51) into (23) we write initially

$$-(\eta^2 - 1)/c = P_1 + M_1 + P_2 + M_2$$
$$= (p_1 + m_1 + 2p_1 m_1 \bar{h}_{11} + U)/\mathscr{D}, \tag{52}$$
$$U = \frac{p_2(1 + p_1 h)^2(1 + m_1\bar{h})^2 + m_2(1 + m_1 h)^2(1 + p_1\bar{h})^2}{1 - p_1 m_1 \bar{h}^2},$$

which simplifies to

$$-\frac{\eta^2 - 1}{c} = p_1 + m_1 + p_1 m_1 \frac{c}{\eta^2}\left(1 + \frac{2\mathscr{H}_1}{c\eta}\right) + U, \tag{53}$$
$$2\mathscr{H}_1/c\eta \approx 2x^2 N/9w.$$

Thus to $O(x^3)$ in terms of forms (49) and (50)

$$1/\eta^2 = (1 + c\mathscr{B}_1)(1 + c\mathscr{C}_1)$$
$$+ \{c^2\mathscr{B}_1\mathscr{C}_1(2Nx^2/9w) + uc\eta^2\}, \tag{54}$$
$$u = [\mathscr{B}_2(5 + 3c\mathscr{B}_1)^2(\eta + 1 + \eta c\mathscr{C}_1)^2$$
$$+ \mathscr{C}_2(5 + 3c\mathscr{C}_1)^2(\eta + 1 + \eta c\mathscr{B}_1)^2]$$
$$\times \{25[(\eta + 1)^2 - \eta^2 c^2 \mathscr{B}_1\mathscr{C}_1]\}^{-1},$$

where in $\{\ \}$ we replace $c\mathscr{A}_1$ by $1/C_1 - 1$ and η by η_1. Using the first two of the elementary relations

$$2\eta + \mu + \epsilon = \frac{(\mu + \eta)^2}{\mu} = \frac{(\epsilon + \eta)^2}{\epsilon} = \frac{(\epsilon + \eta)(\mu + \eta)}{\eta} \tag{55}$$

and the abbreviations

$$T_e = \delta_e V_e + \delta_m R_e,$$
$$V_e = \frac{w}{D_e^2}\left[\frac{1 - \delta_e}{5} + \frac{\delta_e N}{9}\left(2 + \frac{\eta_1^2}{5}\right)\right],$$
$$R_e = -\epsilon'^2 w/10D_e^2, \tag{56}$$
$$S = \frac{(\epsilon_1 - 1)(\mu_1 - 1)N}{9w} = \frac{w\delta_e\delta_m N}{9D_e D_m},$$
$$Q_e = w(2\epsilon_1 + 3)^2/50(2\epsilon' + 3),$$

and Q_m, T_m obtained by interchanging ϵ' and μ', we rewrite (54) as

$$\eta_1^2/\eta^2 = 1 + x^2(T_e/\epsilon_1 + T_m/\mu_1) - i(\epsilon_s/\epsilon_1 + \mu_s/\mu_1)$$
$$+ x^2 2S - x^2\eta_1^2(\delta_e Q_e/\epsilon_1 + \delta_m Q_m/\mu_1). \tag{57}$$

Thus we obtain the form (16) in terms of

$$-\frac{\eta_c^2}{x^2\eta_1^2} = \frac{T_e}{\epsilon_1} + \frac{T_m}{\mu_1} + 2S - \eta_1^2\left(\frac{\delta_e Q_e}{\epsilon_1} + \frac{\delta_m Q_m}{\mu_1}\right) \tag{58}$$

and η_1^2, η_s^2 as in (14) and (15).

For $b = 2a$ and the unrealizable value $w = 1$, the closed forms (9) and (13) give $\mathscr{W} = 0$ and $N = \frac{9}{5}$; then $\eta_c^2 = \eta_s^2 = 0$ and η^2 reduces to the single particle value η'^2 which seems consistent with the implicit truncation approximations. If one of the particle's parameters equals unity, we obtain $\eta^2 = C$ as in (10) in terms of (6), (7), and (11). To emphasize the structure of the generalization (58), we indicate the essential features involved in the reduction to the one-parameter case (11) for $C=\epsilon=\eta^2$. The overall electric dipole contribution $T_e=\delta_e V_e+\delta_m R_e$ is not purely electric and (like b_1) depends on μ' in addition to ϵ'; the same applies for the complement $T_m=\delta_m V_m+\delta_e R_m$. The S term, corresponding to coupling of electric and magnetic dipoles, is symmetric in ϵ' and μ'. The overall quadrupole terms Q (with dipole coupling implicit) depend only on either ϵ' or μ'. Thus, if $\mu'\to 1$, then $T_e\to\delta_e V_e$, $T_m\to -\delta_e w/10$, $\eta_1^2\to\epsilon_1$, and S and Q_m vanish; $\eta^2\to\epsilon$ and (58) reduces to

$$\eta_c^2=\epsilon_c=-x^2\delta_e[V_e-\epsilon_1(w/10+Q_e)], \tag{59}$$

corresponding to (11) for $C'=\epsilon'$.

For moderate parametric contrasts, retaining only to second order terms in δ in (14), (15), and (58), we have

$$\eta_1^2=1+w(\delta_e+\delta_m)+w^2\delta_e\delta_m-w(1-w)(\delta_e^2+\delta_m^2)/3,$$
$$\eta_s^2=x^3 2w\mathscr{W}(\delta_e^2+\delta_m^2)/9, \tag{60}$$
$$\eta_c^2=x^2w(6+3w-5N)[11(\delta_e^2+\delta_m^2)+10\delta_e\delta_m]/225.$$

As discussed for (3:83), for $b=2a$, the function

$$\mathscr{V}=\frac{6+3w-5N}{6}=\frac{(2-w)(1-w)^2}{2(1+2w)}=\frac{2-w}{2}\mathscr{W}^{1/2}, \tag{61}$$

based on (13) and (9), is similar to $\mathscr{W}(w)$ in that both decrease monotonically from 1 to 0 as w increases from 0 to 1. The product $w\mathscr{W}$ has a maximum at $w\approx 0.129$, and $w\mathscr{V}$ has a maximum at $w\approx 0.221$. We have $\eta\approx\eta_1+(\eta_c^2+i\eta_s^2)/2\eta_1$, and to $O(\delta^2)$ for real δ's these maxima specify the maxima of the attenuation via scattering losses Im η and of the correction to Re η.

We also apply (14), (15), and (58) to the extreme case of perfectly conducting particles by using the formal procedure indicated for the single sphere after (20), i.e., we let $\epsilon'\to\infty$, $\mu'\to 0$ to obtain

$$\eta_1^2=2(1+2w)/(2+w), \tag{62}$$
$$\eta_s^2=x^3 2w(5+4w)\mathscr{W}/(1-w)(2+w)^2. \tag{63}$$

For $b=2a$ and $\mathscr{W}(w)$ as in (9)

$$\eta_s^2=x^3 2w(5+4w)(1-w)^3/(2+w)^2(1+2w)^2 \tag{64}$$

has a maximum at $w\approx 0.147$; the corresponding attenuation $\eta_s^2/2\eta_1$ has a maximum at $w\approx 0.135$.

We write the correction to η_1^2 initially as

$$\frac{\eta_c^2(1-w)}{wx^2}=\frac{\eta_1^2}{2+w}\left\{\frac{9}{5}+N-\frac{\eta_1^2}{10}\left[N+\frac{(10-w)^2}{30}\right]\right\}+\frac{\eta_1^2}{1+2w}\left\{\frac{9}{10}-2N-\frac{\eta_1^2}{5}\left[N-\frac{(5+w)^2}{20}\right]\right\}, \tag{65}$$

where the first/second term is dominantly magnetic/electric; the η_1^2 factors indicate dipole effects, and the η_1^4 factors involve both dipole and quadrupole effects. Substituting for η_1^2 and combining terms

$$\eta_c^2(1-w)(2+w)^3 75/wx^2$$
$$=135(4+5w)(2+w)+(1+2w)(50-45w+75w^2+w^3)-30(35+29w+8w^2)N,$$

and using N of (13) for $b=2a$, we obtain the simple result

$$\eta_c^2=\frac{x^2w(1-w)}{15(2+w)^3}[226+33w-42w^2-28w^3], \tag{66}$$

where we also have

$$[\ ]=189+135(1-w)-126(1-w)^2+28(1-w)^3.$$

The correction term has a maximum at $w\approx 0.351$, and the corresponding correction $\eta_c^2/2\eta_1$ to η_1 in the form Re $\eta\approx\eta_1+\eta_c^2/2\eta_1$ has a maximum at $w\approx 0.316$.

VI. BULK PARAMETERS

From (37) in terms of (30), to the required accuracy,

$$\frac{1}{R}=\left[1+\frac{c(P_1+P_2)}{\eta(1+\eta)}\right]+\frac{x^2}{10}c\left\{M_1+\frac{P_1}{\eta}(1+r\eta'^2)\right\}, \tag{67}$$

$$r=(B'-1)/(C'-1),$$

where we may also use $[\ ]=1/\eta-c(M_1+M_2)/\eta(\eta+1)$. Similarly from (38) and (39) in terms of the factors defined in (35) and (36)

$$-(C-1)\eta/cR=P_1+P_2-(x^2/10)\times[P_1(\eta^2-1-r\eta'^2)+M_1\eta/r], \tag{68}$$

$$\eta^2(B-1)/cR=M_1+M_2-(x^2/10)\times[M_1(\eta^2-1-1/r)+P_1r\eta\eta'^2], \tag{69}$$

which satisfy

$$[-C+1+\eta^2(B-1)]/c=P_1+M_1+P_2+M_2=-(\eta^2-1)/c$$

to $O(x^6)$. Thus from (51) for the coefficients we can evaluate the bulk parameters explictly to $O(x^3)$. The B form (69) is the simpler to use to evaluate both ϵ_c and μ_c because to the required accuracy the derivation involves only combinations of leading terms $(\eta_1,\epsilon_1,\mu_1)$ and the particle's parameters. The C form (68) also involves η_c^2 of (58) and requires more manipulations. From either form, by elementary operations based on (51) and (55), we obtain initially for comparison with (54) and (57) for E-case interpretation

$$\frac{1}{\epsilon}=1+c\mathscr{B}_1+\frac{x^2}{\epsilon_1}\left[L(\epsilon',\mu')+S-\frac{\eta_1^2\delta_e Q_e}{\epsilon_1}\right],\qquad 1+c\mathscr{B}_1=\frac{1}{\epsilon_1}\left(1+\frac{x^2T_e}{\epsilon_1}-\frac{i\epsilon_s}{\epsilon_1}\right), \tag{70}$$

$$\frac{1}{\mu}=1+c\mathscr{C}_1+\frac{x^2}{\mu_1}\left[L(\mu',\epsilon')+S-\frac{\eta_1^2\delta_m Q_m}{\mu_1}\right],\qquad 1+c\mathscr{C}_1=\frac{1}{\mu_1}\left(1+\frac{x^2T_m}{\mu_1}-\frac{i\mu_s}{\mu_1}\right), \tag{71}$$

with

$$L=L(\epsilon',\mu')=\frac{1}{10}\left[\frac{(\epsilon_1-1)\epsilon'(\mu'-1)}{\epsilon'-1}-\frac{(\mu_1-1)\mu'(\epsilon'-1)}{\mu'-1}\right]=(w/10)[\epsilon'\delta_m/D_e-\mu'\delta_e/D_m]=-L(\mu',\epsilon'), \tag{72}$$

and T, S, and Q as in (56). The L's cancel in the product (70)×(71) to reproduce (57) to the required accuracy, so that a simple factorization of the bulk index form (57) does not lead to correct values of the bulk parameters.

From (70)–(72), we obtain ϵ_1 and μ_1 in the form (6), ϵ_s and μ_s in the form (7), and write the corrections initially as

$$\epsilon_c = -x^2(T_e + \epsilon_1 L + \epsilon_1 S - \eta_1^2 \delta_e Q_e), \tag{73}$$

$$\mu_c = -x^2(T_m - \mu_1 L + \mu_1 S - \eta_1^2 \delta_m Q_m), \tag{74}$$

to indicate that relation (17) is satisfied. The L terms of $\epsilon_c/\epsilon_1 + \mu_c/\mu_1$ cancel to reproduce η_c^2/η_1^2 of (58). To show that the parameters satisfy theorems (31) and (32), we combine the T and L terms and factor δ. Thus

$$\frac{\epsilon_c}{-x^2 w\delta_e} = \frac{1}{D_e^2}\left[\frac{1-\delta_e}{5} - \frac{\delta_m \epsilon'(1-w)}{15} + \frac{\delta_e N}{9}\left(2 + \frac{\eta_1^2}{5}\right)\right] - \frac{\epsilon_1 \mu'}{10 D_m} + \frac{\epsilon_1 \delta_m N}{D_e D_m 9} - \eta_1^2 \frac{(2\epsilon_1 + 3)^2}{50(2\epsilon' + 3)}, \tag{75}$$

$$\frac{\mu_c}{-x^2 w\delta_m} = \frac{1}{D_m^2}\left[\frac{1-\delta_m}{5} - \frac{\delta_e \mu'(1-w)}{15} + \frac{\delta_m N}{9}\left(2 + \frac{\eta_1^2}{5}\right)\right] - \frac{\mu_1 \epsilon'}{10 D_e} + \frac{\mu_1 \delta_e N}{D_e D_m 9} - \eta_1^2 \frac{(2\mu_1 + 3)^2}{50(2\mu' + 3)}, \tag{76}$$

where $\delta_e = \epsilon' - 1$, $\delta_m = \mu' - 1$, $D_i = 1 + \delta_i(1-w)/3$ for $i = e$ or m, $\epsilon_1 = 1 + w\delta_e/D_e$, $\mu_1 = 1 + w\delta_m/D_m$, and $\eta_1^2 = \epsilon_1\mu_1$. The correlation integral N is defined in (12) in terms of the first moment of the total correlation function, and its closed form based on the Percus–Yevick approximation is given in (13). We include both (75) and (76), and repeat the definitions of all elements to facilitate comparisons and applications. However, (76) is merely (75) with μ' and ϵ' interchanged and it suffices to consider the collective form

$$\frac{-C_c}{x^2 w\delta} = \frac{1}{D^2}\left[\frac{1-\delta}{5} - \frac{\bar{\delta} C'(1-w)}{15} + \frac{\delta N}{9}\left(2 + \frac{\eta_1^2}{5}\right)\right] - \frac{C_1 \bar{C}'}{10\bar{D}} + \frac{C_1 \bar{\delta} N}{D\bar{D} 9} - \frac{\eta_1^2}{50}\frac{(2C_1 + 3)^2}{(2C' + 3)}, \tag{77}$$

where $\eta_1^2 = C_1 \bar{C}_1$ and the bar indicates that the complement is involved. For the one-parameter case we take $\bar{C}' = 1$; then $\bar{\delta} = 0$, $\bar{D} = 1$, $\eta_1^2 = C_1$, and (77) reduces to (11).

Although C_c depends on both C' and $\bar{C}'$, it vanishes if $\delta = C' - 1 = 0$, as does C_s of (7). Thus the form C of (10) in terms of (6), (7), and (77) reduces to $C = 1$ if $C' = 1$, as required by theorem (31). For $b = 2a$, in terms of the closed forms for the correlation integrals, for the unrealizable bound $w = 1$, all scattering effects vanish and C reduces to C'.

For moderate parametric contrast, we retain only terms to $O(\delta^2)$ in (6), (7), and (77) to obtain

$$C_1 = 1 + w\delta - w(1-w)\delta^2/3,$$
$$C_s = x^3 2w\mathscr{W}\delta^2/9, \tag{78}$$
$$C_c = x^2 w\delta(11\delta + 5\bar{\delta})(6 + 3w - 5N)/225.$$

Comparison of the corresponding ϵ_i and μ_i (for $i = 1, s$, and c) with the analogous results for η_i^2 in (60) shows $\eta_1^2 = \epsilon_1\mu_1 + O(\delta^3)$, $\eta_s^2 = \epsilon_s + \mu_s$, and $\eta_c^2 = \epsilon_c + \mu_c$ as required to $O(\delta^2)$.

VII. SLABS AND CYLINDERS

As discussed before,[1] electromagnetic results for normal incidence on parallel slabs or cylinders are obtained by applying results for the corresponding acoustic problems.[2] The analogs of the leading terms in (5)–(17) follow from Ref. 2, and the analogs of the corrections (10)–(13) for the one-parameter cases are given inRef. 4. Now we emphasize the corrections for the two-parameter (ϵ', μ') cases based on the explicit two-parameter results (C',B') in Ref. 3. We use the general forms (3:8), (3:12), and (3:13) to display the slab set (3:32)–(3:34) and the cylinder set (3:50)–(3:52) in forms analogous to the present (57), (75), and (76).

A. Slab scatterers

The leading terms for the one-dimensional problem of slab scatters corresponding to (5)–(7) and (14) and (15) are

$$C_1 = 1 + w\delta, \quad C_s = xw\mathscr{W}\delta^2; \tag{79}$$

$$\eta_1^2 = (1 + w\delta_e)(1 + w\delta_m),$$
$$\eta_s^2/\eta_1^2 = xw\mathscr{W}(\delta_e^2/\epsilon_1 + \delta_m^2/\mu_1), \tag{80}$$
$$\eta_s^2 = xw\mathscr{W}[\delta_e^2 + \delta_m^2 + w\delta_e\delta_m(\delta_e + \delta_m)].$$

From the Laplace transform[12] of the exact[11] F, or from the equation of state,[13,6]

$$\mathscr{W} = 1 + 2\rho\int_0^\infty F\,dr = (1-w)^2,$$
$$W = (b/2a)w, \quad w = 2\rho a, \tag{81}$$

with F as the rigorous Zernike–Prins[11] distribution function; the closed form is exact. For this case, the scatterers are characterized solely by electric and magnetic dipoles, and there are no depolarization effects (i.e., $D = 1$ because there are no field components along the finite dimension of the slabs).

The correction terms involve the first moment

$$N = -\frac{2\rho}{a}\int_0^\infty Fr\,dr = \frac{b}{a}W\left(1 - \frac{4}{3}W + \frac{W^2}{2}\right). \tag{82}$$

(The present N, defined as in Ref. 3, is the negative of that used in Ref. 4.) The closed form is exact, and follows from the Laplace transform of F. Using the abbreviations

$$T = w[\delta(1 + 2\delta) - C'^2\bar{\delta} - 3\delta^2 N]/3,$$
$$S = (C_1 - 1)(\bar{C}_1 - 1)N/w = \delta_e\delta_m N w, \tag{83}$$

with $\delta = \delta_e, \delta_m$ and $\bar{\delta} = \delta_m, \delta_e$ for $T = T_e, T_m$, we rewrite (3:32) as

$$-\eta_c^2/\eta_1^2 x^2 = T_e/\epsilon_1 + T_m/\mu_1 + 2S, \tag{84}$$

which is an analog of (58) without quadrupole effects.

If one of the particle's parameters equals unity, then $T = w\delta(1 + 2\delta - 3\delta N)/3$, $\bar{T} = -w\delta/3$, and (84) reduces to

$$\eta_c^2 = C_c = -x^2 w\delta^2(2 - w - 3N)/3, \tag{85}$$

as in (4:21) for $\delta=\delta_e$ and $\eta_c^2=\epsilon_c$ in terms of the present sign for N. For $b=2a$,

$$\begin{aligned}2-w-3N&=2-7w+8w^2-3w^3\\&=(2-3w)(1-w)^2\\&=2\mathscr{V},\end{aligned}\tag{86}$$

where $\mathscr{V}$, $\mathscr{W}$ and $w\mathscr{V}$, $w\mathscr{W}$ are discussed after (3:40).

For the general case (84) we may regroup to obtain

$$\begin{aligned}3\eta_c^2/x^2w=&-(\delta_e-\delta_m)^2(2-w-3N)\\&+\delta_e\delta_m(\delta_e+\delta_m+2w\delta_e\delta_m)(1-3wN),\end{aligned}\tag{87}$$

where for $b=2a$,

$$1-3wN=1-w^2(6-8w+3w^2)=(1+3w)(1-w)^3.\tag{88}$$

If $\mu'=\epsilon'$ then the leading term is $O(\delta^3)$ instead of $O(\delta^2)$. From (3:33) and (3:34), the analogs of (73) and (74) are

$$\begin{aligned}\epsilon_c&=-x^2[T_e+\epsilon_1L+\epsilon_1S],\\\mu_c&=-x^2[T_m-\mu_1L+\mu_1S],\end{aligned}\tag{89}$$

$$\begin{aligned}L=L(\epsilon',\mu')&=\frac{1}{3}\left[\frac{\epsilon_1-1}{\delta_e}\epsilon'\delta_m-\frac{\mu_1-1}{\delta_m}\mu'\delta_e\right]\\&=(w/3)[\epsilon'\delta_m-\mu'\delta_e]=-L(\mu',\epsilon'),\end{aligned}\tag{90}$$

where L has the same symmetry as for the sphere. Using (84) and (89), we see that (17) is satisfied because the L terms cancel. Combining the T and L terms and factoring δ we obtain the analog of (77)

$$\begin{aligned}-3C_c/x^2w\delta=&\,1+2\delta-C'\bar{\delta}(1-w)-3N\delta\\&-C_1\bar{C}'+3C_1\bar{\delta}N,\end{aligned}\tag{91}$$

which satisfies the corresponding version of (31). Using $\bar{C}'-3\bar{\delta}N=1+\bar{\delta}(1-3N)$, etc., the form simplifies to

$$-3C_c/x^2w\delta=(\delta-\bar{\delta})(2-w-3N)+\delta\bar{\delta}(3wN-1),\tag{92}$$

in terms of the statistical functions in (86) and (88).

If $b=2a$, then for $w=1$ (full packing) we have $\mathscr{W}=0$, $N=\frac{1}{3}$, and the bulk values reduce to the single particle values as required by elementary physical considerations.

For the corresponding acoustic problem, we reinterpret ϵ' as a scatterer's relative compressibility and μ' as its relative mass density for the simplest case (but both parameters may be complex), and similarly for the bulk values.

B. Cylindrical scatterers

For the two cases of E perpendicular ($C=\epsilon,\bar{C}=\mu$) or parallel ($C=\mu,\bar{C}=\epsilon$) to the axes, we represent the leading terms of the real and imaginary parts collectively by

$$\begin{aligned}&C_1=1+w\delta/D,\quad D=1+\delta(1-w)/2,\\&C_s=\pi x^2w\mathscr{W}\delta^2/8D^2,\end{aligned}\tag{93}$$

$$\bar{C}_1=1+w\bar{\delta},\quad \bar{C}_s=\pi x^2w\mathscr{W}\bar{\delta}^2/4,\tag{94}$$

$$\eta_1^2=C_1\bar{C}_1,\quad \frac{\eta_s^2}{\eta_1^2}=\frac{\pi x^2w\mathscr{W}}{4}\left[\frac{\delta^2}{C_12D^2}+\frac{\bar{\delta}^2}{\bar{C}_1}\right].\tag{95}$$

As discussed earlier in detail[15]

$$\mathscr{W}=1+2\pi\rho\int_0^\infty Fr\,dr,\tag{96}$$

$$\mathscr{W}\approx(1-W)^3/(1+W),\quad W=(b/2a)^2w,\quad w=\pi a^2\rho.\tag{97}$$

The form of the parameter C (which involves either electric or magnetic depolarization effects for a field component along the cylinder's finite dimension) is an analog of the sphere form, and that of the complementary $\bar{C}$ is an analog of the slab form. Equation 97 is based on the scaled particle equation of state.[6]

The correction terms also involve[3,4]

$$M=-\ln\frac{b}{a}-\mathscr{W}\ln\frac{2}{c'kb}+8W\int_0^\infty(F\ln u)u\,du,\tag{98}$$

where $c'=1.781...$, and $u=r/b$. (The present M, defined as in Ref. 3, is the negative of that used in Ref. 4.) No closed form of M is available, but as before,[3] in order that the bulk values for cylinders mimic those for spheres at the unrealizable value $w=1$, we require

$$M=\tfrac{3}{4}-2(1-w)/(3+w)+O(1-w)^\nu,\quad \nu>1.\tag{99}$$

The roles of the terms are displayed in the following.

In distinction to the symmetrical odd-dimensional cases in which $\mathscr{W}$ appeared solely in the scattering loss terms and N in the correction terms, for the present case of cylinders the correction terms depend on both M and $1-\mathscr{W}$. As can be seen from (2:71) and (2:73), the M term is associated with a J_0N_0 logarithmic contribution of a correlation integral, and $1-\mathscr{W}$ with a J_1N_1 regular contribution.

We use the abbreviations

$$\begin{aligned}T&=(w/16D^2)\{\delta(4+\delta)-4C'^2\bar{\delta}\\&\qquad+\delta^2[4M+\eta_1^2(1-\mathscr{W})]\},\\\bar{T}&=(w/8)[\bar{\delta}(2-\bar{\delta})-\bar{C}'^2\delta+4\bar{\delta}^2M],\\S&=(C_1-1)(\bar{C}_1-1)(1-\mathscr{W})/4w\\&=w\delta\bar{\delta}(1-\mathscr{W})/4D,\\Q&=(w/16)[(C_1+1)^2/(C'+1)],\end{aligned}\tag{100}$$

where T and Q are analogs of the spherical dipole and quadrupole functions in (56), and $\bar{T}$ of the slab function in (83). From (3:50), we have

$$-\frac{\eta_c^2}{\eta_1^2x^2}=\frac{T}{C_1}+\frac{\bar{T}}{\bar{C}_1}+2S-\eta_1^2\,\delta\frac{Q}{C_1},\tag{101}$$

which is intermediate to forms (58) and (84). For the unrealizable $w=1$, the closed form of $\mathscr{W}$ vanishes and $\eta_s^2=0$; in order that $\eta_c^2=0$ and η^2 reduce to the single particle value η'^2, we require $M=\frac{3}{4}$ as indicated in (99).

The present result differs essentially from (58) and (84) in that (101) represents two distinct physical situations determined by the direction of the field. This also applies for the corresponding one-parameter cases. Thus if $C'=1$, then $T=-w4\bar{\delta}/16$, $\bar{T}=w\bar{\delta}(2-\bar{\delta}+4\bar{\delta}M)/8$, and (101) reduces to

$$\eta_c^2=\bar{C}_c=x^2w\bar{\delta}^2(1+2w-4M)/8,\tag{102}$$

which for $\overline{C}' = \epsilon'$ (i.e., **E** parallel to the axes) is the same as in (4:26) in terms of the present sign for M. If $\overline{C}' = 1$, then $T = w\delta\{4+\delta+\delta[4M + C_1(1-\mathscr{W})]\}/16D^2, \overline{T} = -w\delta/8$, and (101) reduces to

$$\eta_c^2 = C_c = -\frac{x^2 w\delta}{8}\left\{\frac{4+\delta+\delta[4M+C_1(1-\mathscr{W})]}{2D^2} - C_1 - \frac{C_1(C_1+1)^2}{2(C'+1)}\right\}, \tag{103}$$

which for $C' = \epsilon'$ (i.e., **E** perpendicular to the axes) is as in (4:33) in terms of the present sign for M.

For moderate parametric contrast in (101), to second order in δ, the two analogs of (60) are represented by

$$\begin{aligned}\eta_1^2 &= 1 + w(\delta+\overline{\delta}) + \delta\overline{\delta}w^2 - \delta^2 w(1-w)/2,\\ \eta_s^2 &= (\pi x^2 w\mathscr{W}/8)(2\overline{\delta}^2+\delta^2),\\ \eta_c^2 &= (x^2 w/16)[(2\overline{\delta}^2+\delta^2)(1+2w-4M)\\ &\quad + \delta(8\overline{\delta}+\delta)\mathscr{W}],\end{aligned} \tag{104}$$

where the next terms are $O(\delta^3)$.

We also apply (95) and (101) for **E** perpendicular to the axes of perfect conductors by proceeding as for (62) ff. We let $C' = \epsilon' \to \infty$ and $\overline{C}' = \mu' \to 0$ to obtain

$$\eta_1^2 = 1 + w, \tag{105}$$

$$\eta_s^2 = \pi x^2 w\mathscr{W}(3+w)/4(1-w). \tag{106}$$

For $b = 2a$ in terms of the closed form $\mathscr{W}$ of (97),

$$\eta_s^2 = \pi x^2 w(3+w)(1-w)^2/4(1+w) \tag{107}$$

has a maximum at $w \approx 0.301$ and the attenuation $\eta_s^2/2\eta_1$ has a maximum at $w \approx 0.275$, i.e., at volume fractions approximately twice as large as those for spheres given after (64). The correction reduces to

$$\frac{\eta_c^2 8}{wx^2} = \frac{1 + 11w - 4M(3+w) - 6\mathscr{W}(1+w)}{1-w}. \tag{108}$$

For $b = 2a$, we use the closed form of $\mathscr{W}$ to rewrite the result as

$$\frac{\eta_c^2 8}{wx^2} = \left[\frac{(3-4M)(3+w) - 8(1-w)}{1-w}\right] - 6(1-w)^2. \tag{109}$$

In order that (109) behave similarly at large w as (66) for spheres, the function in brackets must vanish for the unrealizable value $w = 1$; this leads to (99), which may facilitate development of a closed form approximation for M.

The results (105)–(109) and the inference (99) are essentially the same as obtained from (3:95)–(3:99) for the acoustic case of rigid cylinders. Corresponding results for **E** parallel to the axes of perfectly conducting cylinders are the same as acoustic results for pressure release (free surface) cylinders; as discussed before (Ref. 15, 1978), alternative procedures than those now under consideration are required to obtain more than the leading terms in w.

From (3:51) and (3:52), we write the bulk parameters initially in terms of

$$\begin{aligned}C_c &= -x^2(T + C_1 L + C_1 S - \eta_1^2\delta Q),\\ \overline{C}_c &= -x^2(\overline{T} - \overline{C}_1 L + \overline{C}_1 S),\end{aligned} \tag{110}$$

$$L = L(C', \overline{C}') = (w/8)[2C'\overline{\delta}/D - \overline{C}'\delta], \tag{111}$$

which satisfy (17) expressed as $\eta_c^2/\eta_1^2 = C_c/C_1 + \overline{C}_c/\overline{C}_1$. The present cases lack the symmetry of the sphere and slab problems, so that $L(\overline{C}', C')$ is not the negative of $L(C', \overline{C}')$.

Combining the T and L terms, we obtain

$$\begin{aligned}-C_c 16/x^2 w\delta &= \{4 + \delta - 2\overline{\delta}C'(1-w)\\ &\quad + \delta[4M + \eta_1^2(1-\mathscr{W})]\}/D^2\\ &\quad - 2C_1\overline{C}' + 4C_1\overline{\delta}(1-\mathscr{W})/D\\ &\quad - \eta_1^2[(1+C_1)^2/(1+C')],\end{aligned} \tag{112}$$

$$\begin{aligned}-\overline{C}_c 8/x^2 w\overline{\delta} &= 2 - \overline{\delta} - \overline{C}'\delta(1-w) + 4\overline{\delta}M\\ &\quad - (\overline{C}_1 2/D)[C' - \delta(1-\mathscr{W})],\end{aligned} \tag{113}$$

where the factored decrimant shows that the appropriate versions of (31) and (32) are satisfied. If $\overline{C}' = 1$, then $\overline{C}_c$ vanishes and (112) reduces to (103); if $C' = 1$, then C_c vanishes, and (113) reduces to (102). The present forms facilitate comparison and checks, but we may also use $C' - \delta(1-\mathscr{W}) = 1 + \delta\mathscr{W}$, etc. For the corresponding acoustic problem for the simplest case we reinterpret $\overline{C}$ as relative compressibility and C as relative mass density.[3] For $b = 2a$ and the unrealizable $w = 1$, we supplement $\mathscr{W} = 0$ with $M = \frac{3}{4}$ to reduce the bulk parameters $\overline{C}, C$ to the corresponding single particle values $\overline{C}', C'$.

For moderate parametric contrast, to second order in δ,

$$\begin{aligned}C_1 &= 1 + \delta w - \delta^2 w(1-w)/2,\\ C_s &= \pi x^2 w\mathscr{W}\delta^2/8,\\ C_c &= (x^2 w\delta/16)[\delta(1+2w-4M+\mathscr{W}) + 4\overline{\delta}\mathscr{W}],\end{aligned} \tag{114}$$

$$\begin{aligned}\overline{C}_1 &= 1 + \overline{\delta}w, \quad \overline{C}_s = \pi x^2 w\mathscr{W}\overline{\delta}^2/4,\\ \overline{C}_c &= (x^2 w\overline{\delta}/8)[\overline{\delta}(1+2w-4M) + 2\delta\mathscr{W}].\end{aligned} \tag{115}$$

We compare the bulk values of a given parameter (ϵ or μ) for polarization parallel and perpendicular to the axes by forming the differences

$$\begin{aligned}\overline{C}(\delta) - C(\delta) &= C_{\parallel} - C_{\perp}\\ &= \delta^2 w(1-w)/2 + (x^2 w\delta^2/16)\\ &\quad \times(1 + 2w - 4M - \mathscr{W})\\ &\quad + i\pi x^2 w\mathscr{W}\delta^2/8,\end{aligned} \tag{116}$$

which for $C = \epsilon$ equals (4:35) with the present sign of M, i.e., the cross terms cancel to second order. The corresponding difference of the indices of refraction for **E** parallel and perpendicular to the axes is given by

$$\begin{aligned}\eta_{\parallel} - \eta_{\perp} &= [(\delta_e^2 - \delta_m^2)w/4][1 - w + (x^2/8)\\ &\quad \times(1 + 2w - 4M - \mathscr{W}) + i\pi x^2\mathscr{W}/4].\end{aligned} \tag{117}$$

See (4:62) ff for discussion of birefringence and dichroism. Note that M of (4:62) should be replaced by $4M$, and 16 by 8.

ACKNOWLEDGMENT

This work was supported in part by grants from the National Science Foundation and the U. S. Office of Naval Research.

[1]V. Twersky, "Coherent electromagnetic waves in pair-correlated random distributions of aligned scatterers," J. Math. Phys. **19**, 215 (1978).
[2]V. Twersky, "Coherent scalar field in pair-correlated random distributions of aligned scatterers," J. Math. Phys. **18**, 2468 (1977).
[3]V. Twersky, "Wavelength-dependent bulk parameters for coherent sound in correlated distributions of small-spaced scatterers," J. Acoust. Soc. Am. **77**, 29 (1985).
[4]V. Twersky, "Wavelength-dependent refractive and absorptive terms for propagation in small-spaced correlated distributions," J. Opt. Soc. Am. **73**, 1562 (1983).
[5]H. L. Frisch and J. L. Lebowitz, *The Equilibrium Theory of Fluids* (Benjamin, New York, 1964); R. L. Baxter, "Distribution functions," in *Physical Chemistry*, edited by H. Eyring, D. Henderson and W. Jost (Academic, New York, 1971), Vol. VIII A, Chap. 4, pp. 267–334.
[6]H. Reiss, H. L. Frisch, and J. L. Lebowitz, "Statistical mechanics of rigid spheres," J. Chem. Phys. **31**, 369 (1959); E. Helfand, H. L. Frisch, and J. L. Lebowitz, "The theory of the two- and one-dimensional rigid sphere fluids," J. Chem. Phys. **34**, 1037 (1961).
[7]M. S. Wertheim, "Exact solution of the Percus–Yevick integral equation for hard spheres," Phys. Rev. Lett. **10**, 321 (1963); E. Thiele, "Equation of state for hard spheres," J. Chem. Phys. **39**, 474 (1963).
[8]H. D. Jones, "Method for finding the equation of state of liquid metals," J. Chem. Phys. **55**, 2640 (1971).
[9]G. L. Throop and R. J. Bearman, "Numerical solution of the Percus–Yevick equation for the hard-sphere potential," J. Chem. Phys. **42**, 2408 (1965); F. Mandel, R. J. Bearman, and M. Y. Bearman, "Numerical solution of the Percus–Yevick equation for the Lennard–Jones (6-12) and hard-sphere potentials," J. Chem. Phys. **52**, 3315 (1970); D. Levesque, J. J. Weis, and J. P. Hansen, "Simulation of classical fluids," in *Monte Carlo Methods in Statistical Physics* (Springer, New York, 1979), pp. 121–144.
[10]Y. Uehara, T. Ree, and F. H. Ree, "Radial distribution function for hard disks from the BGY2 theory," J. Chem. Phys. **70**, 1876 (1979); J. Woodhead-Galloway and P. A. Machin, "X-ray scattering from a gas of uniform hard-disks using the Percus–Yevick approximation," Mol. Phys. **32**, 41 (1976); F. Lado, "Equation of state for the hard-disk fluid from approximate integral equations," J. Chem. Phys. **49**, 3092 (1968).
[11]F. Zernike and J. A. Prins, "Die beuging von Rontgenstrahlen in flussigkeiten als effekt der molekulanordnung," Z. Phys. **41**, 184 (1927).
[12]V. Twersky, "Scattering by quasi-periodic and quasi-random distributions," IRE Trans. **AP-7**, S307 (1959); "Multiple scattering of waves by planar random distributions of cylinders and bosses," Institute of Mathematical Science, NYU report EM58, 1953, "Multiple scattering of sound by correlated monolayers," J. Acoust. Soc. Am. **73**, 68 (1983).
[13]L. Tonks, "The complete equation of state of one, two and three-dimensional gases of hard elastic spheres," Phys. Rev. **50**, 955 (1936).
[14]S. W. Hawley, T. H. Kays, and V. Twersky, "Comparison of distribution functions from scattering data on different sets of spheres," IEEE Trans. Antennas Propag. **AP-15**, 118 (1967).
[15]V. Twersky, "Transparency of pair-correlated, random distributions of small scatterers with applications to the cornea," J. Opt. Soc. Am. **65**, 524 (1975); "Acoustic bulk parameters in distribution of pair-correlated scatterers," J. Acoust. Soc. Am. **64**, 1710 (1978); "Propagation in pair-correlated distributions of small-spaced lossy scatterers," J. Opt. Soc. Am. **69**, 1567 (1979).
[16]V. Twersky, "Propagation in correlated distributions of large-spaced scatterers," J. Opt. Soc. Am. **73**, 313 (1983).
[17]J. C. Maxwell, *A Treatise on Electricity and Magnetism* (Cambridge, 1873; Dover, New York, 1954), Sec. 314.
[18]Lord Rayleigh (J. W. Strutt), "On the influence of obstacles arranged in rectangular order upon the properties of a medium," Philos. Mag. **34**, 481 (1892).
[19]J. A. Stratton, *Electromagnetic Theory* (McGraw–Hill, New York, 1941), p. 564 ff.
[20]V. Twersky, "Multiple scattering of electromagnetic waves by arbitrary configurations," J. Math. Phys. **8**, 589 (1967).
[21]V. Twersky, "Propagation parameters in random distributions of scatterers," J. Analyse Math. **30**, 498 (1976).
[22]M. Lax, "Multiple scattering of waves," Rev. Mod. Phys. **23**, 287 (1951); "The effective field in dense systems," Phys. Rev. **88**, 621 (1952).
[23]V. Twersky, "On propagation in random media of discrete scatterers," Proc. Symp. Appl. Math. **16**, 84 (1964).
[24]J. B. Keller, "Wave propagation in random media," Proc. Symp. Appl. Math. **13**, 227 (1962); "Stochastic equations and wave propagation in random media," *ibid.* **16**, 145 (1964); D. J. Vezzetti and J. B. Keller, "Refractive index, dielectric constant, and permeability for waves in a polarizable medium," J. Math. Phys. **8**, 1861 (1967).
[25]Lord Rayleigh (J. W. Strutt), "On the transmission of light through an atmosphere containing small particles in suspension and on the origin of the blue of the sky," Philos. Mag. **47**, 375 (1899).
[26]F. Reiche, "Zur theorie der dispersion in gasen und dampfen," Ann. Phys. **50**, 1, 121 (1916).
[27]L. L. Foldy, "The multiple scattering of waves," Phys. Rev. **67**, 107 (1945).

Core size effects, bound states, and scattering states of the Aharonov–Bohm problem

N. Gauthier and P. Rochon
Department of Physics, Royal Military College, Kingston, Ontario K7L 2W3, Canada

(Received 22 June 1984; accepted for publication 24 May 1985)

The mathematical solutions of Schrödinger's equation for an electron which is moving outside a hard cylindrical core which contains a hidden flux is studied and it is shown that there are two possible types of eigenstates or scattering states. The first type of state is cyclic about the hidden flux. Such states give rise to field-induced energy shifts, to probability density shifts, and to a divergent scattering cross section. The second type of state is noncyclic about the above axis but all the physical observables are independent of the hidden magnetic flux. The relationship to gauge invariance is discussed.

I. INTRODUCTION

In 1959, Aharonov and Bohm[1] analyzed the quantum-mechanical scattering of electrons by an infinitely long inaccessible whisker of magnetic flux of negligible radius R. Through a partial-wave analysis of this quantum-mechanical problem, Aharonov and Bohm independently rediscovered Ehrenberg and Siday's earlier proposal[2] that a magnetic field can have a quantum effect which is "not due to the magnetic field itself." This surprising effect has stirred a literature controversy which is still active today, because of its far-reaching consequences for physical theories.[3] Our purpose here is not to take part in the controversy but to present a mathematical study of core size effects on the bound states and scattering states of the problem. By assumption, the core is very hard and long, and its axis coincides with the z axis of a cylindrical coordinate system (ρ,ϕ,z).

In the first part of our paper, the bound state Aharonov–Bohm effect is studied. We show that there are two possible types of eigenstates: (1) eigenstates whose eigenvalues are independent of the hidden field but which are not invariant under rotations by multiples of 2π around the z axis, and (2) eigenstates which are cyclic in the angular coordinate ϕ but whose eigenvalues depend explicitly on the hidden flux. The local values of electron density and current density are shown to be flux independent in the first case and to depend on the hidden flux in the second one. The requirements of gauge invariance are discussed.

In the second part of our paper, we consider the important case of electron scattering by a hard core of finite radius R which contains a magnetic flux in its interior. Again, there are two possible forms of the total wave function, associated with the nonperiodic and the periodic partial-wave spectrum discussed in the first part of the paper. When the nonperiodic form is used for the partial-wave analysis, the total scattering cross section is finite and independent of the hidden flux. On the other hand, when the periodic form is used, there results a local change in the charge density and the total scattering cross section diverges for all values of the core size R; this poses a major problem in this case. Aharonov and his co-workers[4] recently published some considerations pertaining to scattering of a uniform electron beam by a core of finite radius. Their solution is an extension of Aharonov and Bohm's original solution and, as such, does not deal with the nonperiodic solution considered here. Aharonov and co-workers actually consider the finite core solution to see how it behaves when the core radius is small, but nonzero. However, as can be seen from our results for the periodic scattering state, the physical parameter which characterizes the core size is the dimensionless combination kR, where $k \equiv (2mE/\hbar^2)^{1/2}$ is the wave vector of the incident electrons, E is the energy, m is the mass, and $\hbar$ is Planck's constant divided by 2π. When $kR \ll 1$ the diffracting core is "small" and when $kR \gg 1$ the diffracting core is "large." Now, in experimental cases reported in the literature, workers have used energies[5] in the vicinity of 100 keV so that $k = 10^{12}$ m^{-1} or so. A typical core size[6] is of order $R = 10^{-6}$ m and so $kR \gg 1$, i.e., the diffracting core is large. To reduce kR to unity would require energies 10^{12} times smaller than those currently used if one still assumes that $R = 10^{-6}$ m. From a practical point of view, therefore, the theoretical region of small core size investigated originally by Aharonov and Bohm and more recently by Aharonov and co-workers is not very useful and a more general discussion of this problem is needed.

Corinaldesi and Rafeli[7] and Henneberger[8,9] have also done some work on the zero-size limit of the scattering problem under investigation here. The work of these researchers, particularly Hennebergers's, prompted us to construct a model where the wave function cannot reach the z axis and where a study of the periodicity and nonperiodicity of the wave function was made. It does not appear that the square-integrability argument used by Henneberger[9] is valid in our case because the wave function cannot reach the z axis, by construction.

II. BOUND-STATE AHARONOV–BOHM EFFECT: PERIODIC AND NONPERIODIC EIGENSTATES

In this section, we consider the eigenstates and eigenvalues of an electron (q) which is confined to the two-dimensional region $R < \rho < R'$, where ρ is the distance to the solenoid axis. The magnetic flux provided by the hidden field is $-\alpha h/q$, by assumption, and it is positioned at $\rho = 0$; α is assumed positive and it is a dimensionless quantity. The eigenstates of the system satisfy the differential equation

$$\rho^2\Big(\frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho} + k^2\Big)\psi = -\Big(\frac{\partial}{\partial\phi} + i\alpha\Big)^2\psi, \tag{1}$$

where k^2 is $2mE/\hbar^2$, with E the eigenvalues. In general, it is permissible to write

$$\psi = A^{\nu}_{\nu+\alpha} e^{i\nu\phi} f_{\nu+\alpha}(\rho), \tag{2}$$

where ν is a number to be determined, and $A^{\nu}_{\nu+\alpha}$ is a normalization constant. The function $f_{\nu+\alpha}(\rho)$ satisfies the differential equation

$$\Big(\frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho} + k^2 - \frac{(\nu+\alpha)^2}{\rho^2}\Big) f_{\nu+\alpha}(\rho) = 0, \tag{3}$$

whose general solution is a linear combination of the Bessel functions $J_{\nu+\alpha}(k\rho)$ and $Y_{\nu+\alpha}(k\rho)$. By construction, the eigenstates must vanish at $\rho = R$ and $\rho = R'$ and so it is permissible to write

$$f_{\nu+\alpha}(\rho) = J_{\nu+\alpha}(k\rho)Y_{\nu+\alpha}(kR) - J_{\nu+\alpha}(kR)Y_{\nu+\alpha}(k\rho). \tag{4}$$

The boundary condition $f_{\nu+\alpha}(R') = 0$ then gives an equation for the energy levels $E = \hbar^2k^2/2m$:

$$J_{\nu+\alpha}(kR')Y_{\nu+\alpha}(kR) = J_{\nu+\alpha}(kR)Y_{\nu+\alpha}(kR'). \tag{5}$$

For ν, α, R, and R' given, we then have eigenvalues $E^{l}_{\nu+\alpha}$, where l is a label attached to the distinct solutions of Eq. (5).

In order to determine the number ν, extra physical conditions must be imposed. If, following tradition, we impose that the eigenstates be cyclic in ϕ, then ν is an integer n. As a result, the energy levels of the particle depend explicitly on the hidden flux, through α. Furthermore, the probability density at ρ also depends on α;

$$|\psi|^2 = |A^{n}_{n+\alpha}|^2 |f_{n+\alpha}(\rho)|^2. \tag{6}$$

A similar dependence on α is found for the current density **j** at a point. On the other hand, on the grounds of gauge invariance,[10] we may require that the physical observables $|\psi|^2$, **j**, and E be unchanged by the hidden field. As a result, we must set $\nu + \alpha = n + \delta$ in the eigenstates and eigenvalues, with n an integer and δ a fixed number which is α independent. In particular, if $\alpha = 0$, we must obtain the zero flux solution and so $\delta = 0$ and the physical observables are the same as they would be in the absence of the hidden field. For example,

$$|\psi|^2 = |A^{n-\alpha}_{n}|^2 |f_n(\rho)|^2, \tag{7}$$

where $|A^{n-\alpha}_{n}|^2$ is found to be α independent upon normalization; similarly, the eigenvalues are given by the α-independent equation

$$J_n(kR')Y_n(kR) = J_n(kR)Y_n(kR'). \tag{8}$$

However, the eigenstates now contain the angular term $e^{in\phi}e^{-i\alpha\phi}$ and so they are nonperiodic in ϕ if α is noninteger. We will return to this point in our concluding remarks.

III. PERIODIC AND NONPERIODIC SCATTERING STATES

We now extend the previous considerations to the scattering context and use the partial wave bases of the previous section to formulate the problem. The radius of the inaccessible core is R, as previously, and the outer boundary $\rho = R'$ is removed. Furthermore, we assume that an incident plane wave

$$\Psi_0 = e^{ik\rho\cos\phi} = \sum_m i^{|m|} J_{|m|}(k\rho) e^{im\phi} \tag{9}$$

is traveling in the positive x direction and scatters off the hard core plus hidden field. As is customary in scattering theory, we require that the scattered wave Ψ_s behave like an outgoing cylindrical wave $e^{ik\rho}/\rho^{1/2}$, as $\rho \to \infty$. Because of the hard core, the total wave function

$$\Psi = \Psi_0 + \Psi_s \tag{10}$$

vanishes at $\rho = R$.

The Hankel function $H \equiv J + iY$ satisfies the requirements imposed on the scattered wave and so we write the cyclic version of the total wave function thus:

$$\Psi = \sum_m e^{im\phi}\big(A^{\alpha}_{m} H_{|m+\alpha|}(k\rho) + B^{\alpha}_{m} J_{|m+\alpha|}(k\rho)\big). \tag{11}$$

In this expression,

$$B^{\alpha}_{m} = (-1)^m(-i)^{|m+\alpha|} \tag{12}$$

and

$$A^{\alpha}_{m} = -B^{\alpha}_{m} J_{|m+\alpha|}(kR)/H_{|m+\alpha|}(kR), \tag{13}$$

as found through the boundary conditions on Ψ at $\rho = R$ and on Ψ_s as $\rho \to \infty$. The total cross section per unit length along the z axis σ is given by integrating

$$\frac{d\sigma}{d\phi} = \rho|\Psi - \Psi_0|^2 = \rho|\Psi_s|^2 \tag{14}$$

over ϕ, from 0 to 2π, assuming that $\rho \to \infty$; $\Psi_s = \Psi_s(\rho,\phi)$ is the asymptotic form of the scattered wave. We write Ψ_s as follows:

$$\Psi_s = \Psi_1 + \Psi_2 + \Psi_3, \tag{15}$$

where

$$\Psi_i = \sum_m e^{im\phi}\Psi_{im}, \quad i = 1,2,3, \tag{16}$$

by definition. Here,

$$\Psi_{1m} = A^{\alpha}_{m} H_{|m+\alpha|}(k\rho) \tag{17}$$

and

$$\Psi_{2m} = B^{\alpha}_{m} J_{|m+\alpha|}(k\rho) \tag{18}$$

are terms that determine the total wave function Ψ, and

$$\Psi_{3m} = -i^{|m|} J_{|m|}(k\rho) \tag{19}$$

determines the negative of the incident wave. In terms of these functions, the total scattering cross section is

$$\sigma = \sum_{i,j=1}^{3} \sigma_{ij}, \tag{20}$$

where

$$\sigma_{ij} = \pi\rho \sum_m (\Psi^*_{im}\Psi_{jm} + \Psi_{im}\Psi^*_{jm}). \tag{21}$$

Note that $\sigma_{ij} = \sigma_{ji}$ so that we only require a calculation of $\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{12} = \sigma_{21}, \sigma_{13} = \sigma_{31}$, and $\sigma_{23} = \sigma_{32}$. In order to make the discussion of the limiting case $R \to 0$ easier later on, we give here the terms which are core-size independent (σ_{22}, σ_{33}, and $\sigma_{23} = \sigma_{32}$):

$$\sigma_{22} = 2\pi\rho \sum_m J^2_{|m+\alpha|}(k\rho), \tag{22}$$

$$\sigma_{33} = 2\pi\rho \sum_m J^2_{|m|}(k\rho) = 2\pi\rho, \tag{23}$$

and

$$\sigma_{23} = -2\pi\rho \sum_m J_{|m+\alpha|}(k\rho) J_{|m|}(k\rho) \times \cos\left[(|m| - |m+\alpha|)\frac{\pi}{2}\right]. \tag{24}$$

One may be tempted to replace the Bessel functions of Eqs. (22)–(24) by the asymptotic form[11] $(2/\pi k\rho)^{1/2}\cos(k\rho - \pi/4 - \nu\pi/2)$, in order to make σ_{22}, σ_{33}, and σ_{23} independent of ρ. Such a procedure is often used to obtain the cross section whether it be in quantum-mechanical or electromagnetic scattering. Its application here is, however, not mathematically justified because the functions $J_\nu(k\rho)$ involved in the above series are not multiplied by a factor whose magnitude decreases as $|m|$ increases. The present coefficients, e.g., the cosine modulation in σ_{23}, are of order unity for all m and thus fail to provide the cutoff which is already built into the exact $J_\nu(k\rho)$. Indeed, as is known,[11] the following expression holds for $\nu \gg k\rho$:

$$J_\nu(k\rho) \sim (2\pi\nu)^{1/2}(ek\rho/2\nu)^\nu; \tag{25}$$

this property ensures the convergence of the above series to a finite, ρ-independent value in the asymptotic region. As a result, the sum of the core-size independent terms, $\sigma_a \equiv \sigma_{22} + \sigma_{33} + 2\sigma_{23}$, diverges linearly with ρ. One can verify this latter property exactly for the case $\alpha =$ odd integer since then $\sigma_{23} \equiv 0$ and $\sigma_{22} = \sigma_{33} = 2\pi\rho$. The explicit ρ dependence of σ_a in the asymptotic region leads to an explicit ρ dependence of the total cross section of Eq. (20) and is not consistent with our original requirement (see earlier in this section) that Ψ_s behave like an outgoing cylindrical wave $e^{ik\rho}/\rho^{1/2}$ in the asymptotic region. This lack of self-consistency of the theory is apparently due to the long-range nature of the vector potential. Indeed, the "strength" of the potential may be characterized by the circulation $\oint \mathbf{A}\cdot d\mathbf{r}$ over a closed circle of radius ρ centered on the solenoid axis; this strength does not vanish asymptotically and the imposed requirement that the full Ψ become the incident Ψ_0 as $\rho \to \infty$ is not met. The long-range behavior of the present solution does not decay into that of the incident beam and the theory is not self-consistent.

An *ad hoc* solution to the present dilemma might be to replace the original beam by a beam of finite extent. One could assume an incident beam of the form $\Sigma_m i^{|m|} e^{-\epsilon|m|} J_{|m|}(k\rho) e^{im\phi}$, where $\epsilon > 0$ is a positive number. Such a procedure does ensure that the anomalous part of the cross section is ρ independent asymptotically. Then, for example,

$$\sigma_{33} = 2\pi\rho \sum_m e^{-2\epsilon|m|} J^2_{|m|}(k\rho) \sim \frac{4}{k}\sum_m e^{-2\epsilon|m|}\cos^2\left(k\rho - \frac{\pi}{4} - \frac{|m|\pi}{2}\right), \tag{26}$$

and the sum is absolutely convergent. However, for $\epsilon \to 0^+$, the cross section is strongly dependent on the beam shape, through ϵ, and the total cross section diverges in that limit. Indeed, the anomalous cross section is simply obtained in closed form for $\epsilon \to 0$,

$$\sigma_a \propto \operatorname{csch}(2\epsilon), \tag{27}$$

and the result diverges in that limit.

The finite core contribution to the cross section is $\sigma_n = \sigma_{11} + 2\sigma_{12} + 2\sigma_{13}$ and it behaves properly for all values of ϵ. The limit $\epsilon \to 0^+$ presents no difficulties and so we take it at the outset. As a result, we find

$$\sigma_{11} = 2\pi\rho \sum_m M^2_{|m+\alpha|}(k\rho)\gamma^2_{\alpha m}(R) \sim \frac{4}{k}\sum_m \gamma^2_{\alpha m}(R), \tag{28}$$

$$\sigma_{12} = -2\pi\rho \sum_m M^2_{|m+\alpha|}(k\rho)\gamma_{\alpha m}(\rho)\gamma_{\alpha m}(R)\cos\lambda_{\alpha m}(\rho,R) \sim -\frac{4}{k}\sum_m \gamma_{\alpha m}(R)\gamma^a_{\alpha m}(\rho)\cos\lambda^a_{\alpha m}(\rho,R), \tag{29}$$

and

$$\sigma_{13} = 2\pi\rho \sum_m M^2_{|m+\alpha|}(k\rho)\gamma_{\alpha m}(R)\gamma_{\alpha m}(\rho)\cos\delta_{\alpha m}(\rho,R) \sim \frac{4}{k}\sum_m \gamma_{\alpha m}(R)\gamma^a_{\alpha m}(\rho)\cos\delta^a_{\alpha m}(\rho,R), \tag{30}$$

where

$$\theta_\nu(s) \equiv \tan^{-1}(Y_\nu(s)/J_\nu(s)) \tag{31}$$

and

$$M_\nu(s) \equiv (J^2_\nu(s) + Y^2_\nu(s))^{1/2} \tag{32}$$

are the phase and modulus of the Bessel functions J and Y.[11] Also,

$$\lambda_{\alpha m}(\rho,R) \equiv \theta_{|m+\alpha|}(kR) - \theta_{|m+\alpha|}(k\rho), \tag{33}$$

$$\gamma_{\alpha m}(s) \equiv \cos\theta_{|m+\alpha|}(ks), \tag{34}$$

and

$$\delta_{\alpha m}(\rho,R) \equiv \lambda_{\alpha m}(\rho,R) + (|m+\alpha| - |m|)\pi/2; \tag{35}$$

the quantities $\lambda^a_{\alpha m}$, $\gamma^a_{\alpha m}$, and $\delta^a_{\alpha m}$ refer to the above definitions, with $\theta_\nu(k\rho)$ replaced by its asymptotic expression[11]

$$\theta^a_\nu(k\rho) = k\rho - \pi/4 - \nu\pi/2. \tag{36}$$

The second line of Eqs. (28)–(30) gives the asymptotic form of the corresponding expression; in each case, the series involved contains roughly kR finite terms and the normal part of the cross section is of order R, as expected on a semiclassical basis when $kR \gg 1$. As is easily shown, the normal part of the cross section vanishes in the limit $kR \to 0$. As a result, the total cross section diverges due to the anomalous part σ_a [see Eq. (27)]. Henneberger[8] appears to have been the first to mention this anomaly of the AB cross section in his study of the zero-size core problem ($kR \to 0$). Henneberger had surmised that the origin of the divergence was due to the presence of the wave function on the solenoid axis. As shown here, however, the difficulty is deeper than this: the cross section still diverges for a finite core size.

The nonperiodic form of the scattering wave function is easily found to be

$$\Psi = e^{-i\alpha\phi}\Psi_{\alpha=0}, \tag{37}$$

where

$$\Psi_{\alpha=0} = \sum_m i^{|m|}\left[J_{|m|}(k\rho) - \frac{J_{|m|}(kR)}{H_{|m|}(kR)}H_{|m|}(k\rho)\right]e^{im\phi} \tag{38}$$

is the wave function with $\alpha = 0$. This represents scattering of a plane wave by a scalar-potential hard core at $\rho = R$ only, in the absence of hidden flux. The total scattering cross section for the state represented by Eqs. (37) and (38) is equal to $\sigma_{11} + 2\sigma_{12} + 2\sigma_{13}$, with σ_{ij} given by Eqs. (28)–(30) and $\alpha = 0$. The anomalous contribution vanishes and so the physical properties are well behaved. Henneberger also studied[9] some of the aspects associated with a noncyclic scattering wave function, for a core of zero size. Equations (37) and (38) represent the first generalization of his results for the case $R \neq 0$.

The scattering state of Eq. (37) is generally not single valued, as implied by its noncyclic character. Furthermore, the usual requirements that the full wave function Ψ decay into the incident beam "outside" the range of the potential is not met. Indeed, as $\rho \rightarrow \infty$, $\Psi \sim e^{i\alpha\phi}e^{ik\rho\cos\phi}$. This problem can again be traced back to the infinite range of the vector potential used in discussions of *AB* scattering.

IV. CONCLUDING REMARKS

It appears extremely difficult to ascribe a physical meaning to the observables computed by assuming that the eigenstates or scattering states of an *AB* system are periodic in the coordinate ϕ, i.e., single valued. In such a case, the gauge invariance of the theory appears to be broken and the scattering cross section is divergent. The divergence of the cross section is not unique to *AB* scattering, however, since this property also occurs in Coulomb scattering, due to the long-range nature of the interaction. What is really special here is the absence of a long-range *force* to cause the divergence. However, there is a long-range vector potential in *AB* scattering and if we accept Aharonov and Bohm's original suggestion that it, by itself ($R \rightarrow 0$), can scatter particles, then the finding that σ (anomalous) diverges for an incident beam of infinite extent may not be all that surprising. The total cross section of Eq. (20), σ (normal) + σ (anomalous), represents the effects of scattering by the scalar and the vector potentials in our problem. The normal part of σ is essentially due to scattering by the hard core, proper account being taken for the continuing interaction of those particles with the vector potential as they move out to infinity. Conservation of the number of such particles is reflected by the ρ independence and finite value of σ (normal). The anomalous contribution, on the other hand, takes into account the scattering of all those particles of the incident beam which would normally not "collide" with the core because their impact parameter ($|m|$ − values $\gg kR$) is too large. Because of the infinite range of the vector potential used in *AB* scattering, scattering events "cumulate" up to infinity and the number of scattered particles eventually diverges for a beam of infinite extent. As we have seen, the difficulty may be removed by limiting the spatial extent of the incident beam, but the strong beam-size dependence of the resulting cross section is hardly acceptable as a definitive solution. Failure to incorporate the tailing-off of the vector potential as $\rho \rightarrow \infty$ (as would be the case for a realistic situation) seems to imply that the usual assumption of scattering theory that the total Ψ eventually decays into the incident Ψ_0 is not met. This latter point is clearly shown by the nonperiodic solution of Eq. (37) and so both the present cyclic and noncyclic solutions are faced with serious difficulties. We believe that proper account of the tailing-off of the vector potential will lead to a correct physical solution of this question, even when an incident beam of infinite extent is used. Once this is done, the limit of the solenoid length going to infinity can be taken since then the boundary conditions on Ψ at $\rho \rightarrow \infty$ will have been properly satisfied. Such a solution is not presently available, however.

ACKNOWLEDGMENTS

The authors wish to thank T. J. Racey for his general interest and helpful suggestions.

[1]Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).

[2]W. Ehrenberg and R. E. Siday, Proc. Phys. Soc. London B **62**, 8 (1949).

[3]S. M. Roy, Phys. Rev. Lett. **44**, 111 (1980). This paper contains reference to many other papers on this subject.

[4]Y. Aharonov, C. K. Au, E. C. Lerner, and J. Q. Liang, Phys. Rev. D **29**, 2396 (1984). See also S. N. M. Ruijsenaars, Ann. Phys. (NY) **146**, 1 (1983).

[5]A. Tonomura, T. Matsuda, R. Suzuki, A. Fukuhara, N. Osakabe, H. Umezaki, J. Endo, K. Shinagawa, Y. Sugita, and H. Fujiwara, Phys. Rev. Lett. **48**, 1443 (1982).

[6]W. Bayh, Z. Phys. **169**, 492 (1962). See J. Woodilla and H. Schwarz's brief review of experiments on the Aharonov–Bohm effect, Am. J. Phys. **39**, 111 (1971).

[7]E. Corinaldesi and F. Rafeli, Am. J. Phys. **46**, 1185 (1978).

[8]W. C. Henneberger, Phys. Rev. A **22**, 1383 (1980).

[9]W. C. Henneberger, J. Math. Phys. **22**, 117 (1981).

[10]C. Cohen-Tannoudji, B. Diu, and F. Laloë, *Mécanique quantique* (Hermann, Paris, 1973), Vol. I.

[11]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970), 9th printing.

Quantum stochastic calculus, operation valued stochastic processes, and continual measurements in quantum mechanics

A. Barchielli and G. Lupieri
Dipartimento di Fisica dell'Università di Milano, Istituto Nazionale di Fisica Nucleare, Sezione di Milano, Via Celoria, 16-20133 Milano, Italy

(Received 5 November 1984; accepted for publication 17 May 1985)

The physical idea of a continual observation on a quantum system has been recently formalized by means of the concept of operation valued stochastic process (OVSP). In this article, it is shown how the formalism of quantum stochastic calculus of Hudson and Parthasarathy allows, in a simple way, for constructing a large class of OVSP's that in particular contains the quantum counting processes of Davies and Srinivas and continual "Gaussian" measurements. This result is obtained by means of a stochastic dilation of the OVSP's: at the level of the enlarged system probabilities turn out to be expressed in terms of projection valued measures associated with certain time-dependent, commuting, self-adjoint operators.

I. INTRODUCTION

Continual measurements can be consistently introduced in quantum mechanics by using the general framework of measurement theory. In this setup, *effect* (or positive operator) *valued measures* generalize the concept of observables (usually associated with projection valued measures) and *operations* generalize the Von Neumann reduction postulate.[1–4]

Up to now, two classes of continual measurements have been independently introduced and studied: counting processes (analogous to classical Poisson processes),[2,5–10] and the continual measurements of some observables with "Gaussian" instruments.[11–20] In both cases the dynamics of the continually measured system is given by a quantum dynamical semigroup as for quantum open systems. In Ref. 12, *operation valued stochastic processes* (OVSP's) have been introduced as the mathematical objects that formalize the concept of continual measurements in quantum mechanics.

Meanwhile, the analysis of quantum open systems has led to the introduction of the concept of quantum stochastic process[21–26] and the development of a quantum stochastic calculus,[27–29] which enable one to construct unitary dilations of quantum dynamical semigroups[27–29] and a wide class of quantum stochastic processes.[26]

In this paper we want to show how quantum stochastic calculus can be used for constructing a class of OVSP's which generalizes the results of Ref. 12 and includes and mixes the "Gaussian" and the "Poisson" cases. Moreover, in this way, we obtain a dilation of the irreversible dynamics to a reversible one and of the effect valued measure, describing the continual observation, to a projection valued one; at the level of the enlarged system the continually measured observables are associated with a set of self-adjoint operators commuting also at different times (in the Heisenberg picture).

The plan of the paper is as follows. In Sec. II we give the definition of OVSP and introduce the notion of *characteristic operator* of an OVSP. In Sec. III we sketch the formalism of quantum stochastic calculus, with emphasis on the structural properties which are needed in our work. In Sec. IV we construct a class of OVSP's by giving a dilation for them, in some sense similar to stochastic dilations of quantum dynamical semigroups. In Sec. V we study the meaning of this dilation. For the global system the standard formulation of quantum mechanics appears to hold: commuting (at all times) self-adjoint operators are associated with the continually measured quantities, after the measurement the state of the system is given by the Von Neumann reduction postulate, etc. In Sec. VI we show how the class of continual measurements studied by Davies and Srinivas[2,5–10] is included in the class of OVSP's found in Sec.IV.

II. OPERATION VALUED STOCHASTIC PROCESSES

The most general setup for speaking of continually measured quantities is that of generalized stochastic processes (GSP).[30,31] Let $\mathscr{D}$ be the nuclear space of the n-component, real C^∞ functions $\mathbf{h}(t)\equiv(h_1(t),...,h_n(t))$ on $\mathbb{R}$ with compact supports. The random variables are taken to be the elements of $\mathscr{D}'$, the topological dual space of $\mathscr{D}$; thus, $\mathscr{D}'$ is the "trajectory space" of the continually measured quantities. For $x\in\mathscr{D}'$ and $\mathbf{h}\in\mathscr{D}$, we denote by x_h the distribution x applied to the test function $\mathbf{h}$. The subsets of $\mathscr{D}'$ of the form $\{x\in\mathscr{D}':(x_{h^{(1)}},...,x_{h^{(s)}})\in B\}$, where B is a Borel subset of $\mathbb{R}^s$, are called cylinder sets. We equip $\mathscr{D}'$ with the σ algebra Σ generated by the cylinder sets and denote by $\Sigma_{(t_1,t_2)}$, $t_1<t_2$, the sub-σ-algebra generated by the cylinder sets defined by test functions with supports contained in the time interval (t_1,t_2).

Now, let $\mathscr{h}$ be the Hilbert space of the measured system and denote by $T(\mathscr{h})$ the Banach space of the trace-class operators on $\mathscr{h}$. An *operation valued stochastic process* (OVSP) is defined to be a set of linear maps $\mathscr{F}(t_2,t_1;N)$, $N\in\Sigma_{(t_1,t_2)}$, from $T(\mathscr{h})$ into itself with the following properties: (i) $\mathscr{F}(t_2,t_1;N)$, $\forall N\in\Sigma_{(t_1,t_2)}$ is completely positive; (ii) $\mathscr{F}(t_2,t_1;\cdot)$ is σ additive on $\Sigma_{(t_1,t_2)}$ (convergence in the strong sense); (iii) $\mathscr{F}(t_2,t_1;\mathscr{D}')$ is trace preserving (normalization); and (iv) the following Markov property holds:

$$\mathscr{F}(t_3,t_2;M)\mathscr{F}(t_2,t_1;N)=\mathscr{F}(t_3,t_1;M\cap N),$$
$$\forall N\in\Sigma_{(t_1,t_2)},\quad \forall M\in\Sigma_{(t_2,t_3)},\quad t_1<t_2<t_3. \qquad (2.1)$$

 0022-2488/85/092222-09$02.50

By properties (i)–(iii), $\mathscr{F}(t_2,t_1;\cdot)$ is a normalized *operation valued measure* (OVM)[2] on $\Sigma_{(t_1,t_2)}$ for any time interval (t_1,t_2) of measurement. The physical interpretation of these properties is the following: if ρ is the statistical operator for the system at time t_1, then

$$P(N\,|\,\rho,t_1) := \mathrm{Tr}\{\mathscr{F}(t_2,t_1;N)(\rho)\} \tag{2.2}$$

is the probability of finding the result $x\in N$ ($N\in\Sigma_{(t_1,t_2)}$) and

$$\rho' := \mathscr{F}(t_2,t_1;N)(\rho)/P(N\,|\rho,t_1) \tag{2.3}$$

is the state at time t_2, conditioned upon the result $x\in N$. Property (i) is mathematically stronger than the simple positivity of quantities (2.2) and (2.3); for a physical motivation of this condition, see for instance Ref. 3.

Property (iv) ensures the consistency of the OVM's referring to different time intervals. First, by Eq. (2.1) and property (iii), we have

$$\begin{aligned} P(N\,|\,\rho,t_1) &= \mathrm{Tr}\{\mathscr{F}(t_2,t_1;N)(\rho)\} \\ &= \mathrm{Tr}\{\mathscr{F}(t_3,t_1;N)(\rho)\}, \quad N\in\Sigma_{(t_1,t_2)}, \quad t_1<t_2<t_3; \end{aligned} \tag{2.4}$$

therefore probabilities do not depend on the future but only on the past. Then, if we introduce the conditional probability

$$\begin{aligned} P(M\,|N;\rho,t_1) &:= P(M\cap N\,|\,\rho,t_1)/P(N\,|\,\rho,t_1) \\ &= \frac{\mathrm{Tr}\{\mathscr{F}(t_3,t_1;M\cap N)(\rho)\}}{\mathrm{Tr}\{\mathscr{F}(t_3,t_1;N)(\rho)\}}, \\ N&\in\Sigma_{(t_1,t_2)},\ M\in\Sigma_{(t_2,t_3)},\ t_1<t_2<t_3, \end{aligned} \tag{2.5}$$

we have

$$P(M\,|N;\rho,t_1) = P(M\,|\,\rho',t_1), \tag{2.6}$$

where ρ' is given by Eq. (2.3); therefore, at any time, the whole information on the past can be represented by a statistical operator.

Note that the triple

$$\{E',\Sigma_{(t_1,t_2)},P(\cdot|\,\rho,t_1)\}, \quad \text{for given } \rho,t_1,t_2$$

is a GSP.[30,31] Just as a GSP is uniquely determined by its *characteristic functional*, an OVSP is uniquely determined by its *characteristic operator* $\mathscr{G}(t_2,t_1;[\varphi])$ (Refs. 12 and 15), defined as the mean value of $\exp(ix_\varphi)$, $\varphi\in\mathscr{D}_{(t_1,t_2)}$ [$\mathscr{D}_{(t_1,t_2)}$ is the subspace of $\mathscr{D}$ of the functions with supports contained in (t_1,t_2)] with respect to the OVM $\mathscr{F}(t_2,t_1;\cdot)$.

The characteristic operator $\mathscr{G}(t_2,t_1;[\varphi])$ of an OVSP has the following properties.

(a) It is a bounded linear operator from $T(\mathscr{h})$ into itself.

(b) It is normalized, i.e.,

$$\mathrm{Tr}\,(\mathscr{G}(t_2,t_1;[0])(X)) = \mathrm{Tr}(X), \quad \forall X\in T(\mathscr{h}). \tag{2.7}$$

(c) It is of completely positive type, i.e., the quantities

$$\sum_{i,j=1}^{n} \alpha_i^*\mathscr{G}(t_2,t_1;[\varphi_i-\varphi_j])\alpha_j$$

are completely positive for any choice of the integer n, of the complex numbers α_i, and of the test functions $\varphi_i(t)$.

(d) It is strongly continuous in φ ($\varphi\in\mathscr{D}_{(t_1,t_2)}$).

(e) It satisfies the following composition law:

$$\begin{aligned} &\mathscr{G}(t_3,t_2;[\varphi_2])\mathscr{G}(t_2,t_1;[\varphi_1]) = \mathscr{G}(t_3,t_1;[\varphi_1+\varphi_2]), \\ &\qquad \forall\varphi_1\in\mathscr{D}_{(t_1,t_2)}, \quad \forall\varphi_2\in\mathscr{D}_{(t_2,t_3)}, \quad t_3>t_2>t_1. \end{aligned} \tag{2.8}$$

Vice versa, given a set of operators $\mathscr{G}(t_2,t_1;[\varphi])$ with properties (a)–(e), there exists a unique OVSP whose characteristic operator is $\mathscr{G}(t_2,t_1;[\varphi])$. Note that, when $\mathscr{h}=\mathbb{C}$, these properties define the characteristic functional of a classical GSP with independent values at every time.[30]

In this paper we shall consider only *time-translation invariant* OVSP's. This notion is formalized in Ref. 12; when the Schrödinger picture is chosen, for the characteristic operator time-translation invariance becomes the following.

(f) If we put $\varphi^{(\bar t)}(t)=\varphi(t-\bar t)$, then, $\forall\varphi\in\mathscr{D}_{(t_1,t_2)}$, we have

$$\mathscr{G}(t_2+\bar t,t_1+\bar t;[\varphi^{(\bar t)}]) = \mathscr{G}(t_2,t_1;[\varphi]). \tag{2.9}$$

By these properties, the family of operators

$$\mathscr{G}(t_2-t_1) := \mathscr{G}(t_2,t_1;[0])\equiv\mathscr{F}(t_2,t_1;\mathscr{D}') \tag{2.10}$$

is a quantum dynamical semigroup[2] and gives the dynamics of the measured system [see Eq. (2.3) and Ref. 12].

Finally, let us recall that for any operator $\mathscr{A}$ in $T(\mathscr{h})$ it is possible to define its adjoint $\mathscr{A}'$ in $B(\mathscr{h})$ (Banach space of bounded operators on $\mathscr{h}$) using the duality relation between $T(\mathscr{h})$ and $B(\mathscr{h})$. In particular, in the following sections it will be useful to work with $\mathscr{G}'(t_2,t_1;[\varphi])$. Moreover, using this notation, we can write the probabilities (2.2) as

$$P(N\,|\,\rho,t_0) = \mathrm{Tr}\,(F(t_0;N)\rho), \quad N\in\Sigma_{(t_0,t_f)}, \tag{2.11}$$

where, denoting by $\mathbb{1}$ the identity operator on $\mathscr{h}$,

$$F(t_0;N) := \mathscr{F}'(t_f,t_0;N)(\mathbb{1}); \tag{2.12}$$

$F(t_0;\cdot)$ turns out to be an *effect* (or positive operator) *valued measure*.[1,2] If we consider cylinder sets and put

$$\widetilde F_{(h^{(1)},\dots,h^{(s)})}(t_0;B) := F(t_0;(x_{h^{(1)}},\dots,x_{h^{(s)}})\in B), \tag{2.13}$$

we have that $\widetilde F_{(\dots)}(t_0;\cdot)$ is an effect valued measure on the Borel σ algebra of $\mathbb{R}^s$ and it can be formally obtained from the characteristic operator by taking the Fourier transform

$$\begin{aligned} &\widetilde F_{(h^{(1)},\dots,h^{(s)})}(t_0;B) \\ &\quad = \int_B d_s\mathbf{x}\,\frac{1}{(2\pi)^s}\int d_s\mathbf{k}\,\exp(-i\mathbf{k}\cdot\mathbf{x}) \\ &\qquad \times\mathscr{G}'\Bigl(t_f,t_0;\Bigl[\sum_{r=1}^{s}k_r\mathbf{h}^{(r)}(t)\Bigr]\Bigr)(\mathbb{1}). \end{aligned} \tag{2.14}$$

By similar formulas, the operations $\mathscr{F}(\cdots)$ can be reconstructed from the characteristic operator.

III. QUANTUM STOCHASTIC CALCULUS

In this section we want to recall some results about quantum stochastic calculus and related topics,[26–29] mainly for fixing notations.

Let $\mathscr{L}_k(-\infty,+\infty)\equiv\Gamma(L^2_k(-\infty,+\infty))$ be the symmetric Fock space over the space of the square-integrable functions taking values in a complex separable Hilbert space k. We denote by $\psi(f)$, $f\in L^2_k(-\infty,+\infty)$, the exponential vectors in the Fock space

$$\psi(f) = (1,f,\dots,(n!)^{-1/2}f\otimes\cdots\otimes f,\dots); \tag{3.1}$$

we recall that $\{\psi(f);f\in L^2_k(-\infty,+\infty)\}$ is a total family in $\mathscr{L}_k(-\infty,+\infty)$ (Ref. 32), $\{\psi(f_i);\,i=1,\dots,n\}$ is a set of linearly independent vectors, if $f_i\neq f_j$ for $i\neq j$, and the following equality holds:

$$\langle \psi(g),\psi(f)\rangle_{\mathscr{L}_k} = \exp\left[\langle g,f\rangle_{L^2_k}\right]. \tag{3.2}$$

In these notations the Fock vacuum is $\psi(0)$.

Similarly, one can introduce the space $\mathscr{L}_k(s,t) \equiv \Gamma(L^2_k(s,t))$, for $s<t, s,t \in \mathbb{R}$. Then, there is the natural identification

$$\mathscr{L}_k(-\infty,+\infty) = \mathscr{L}_k(-\infty,s) \otimes \mathscr{L}_k(s,t) \otimes \mathscr{L}_k(t,+\infty) \tag{3.3}$$

in which, for $f_1 \in L^2_k(-\infty,s), f_2 \in L^2_k(s,t), f_3 \in L^2_k(t,+\infty)$,

$$\psi(f_1 \oplus f_2 \oplus f_3) = \psi(f_1) \otimes \psi(f_2) \otimes \psi(f_3). \tag{3.4}$$

Using the structure (3.3), we introduce the W^* algebras

$$\mathscr{C}(s,t) := \mathbf{1} \otimes B(\mathscr{L}_k(s,t)) \otimes \mathbf{1}. \tag{3.5}$$

In the following we shall use the property

$$\left\langle \psi(0), \prod_{j=1}^{n} C_j \psi(0)\right\rangle = \prod_{j=1}^{n} \langle \psi(0), C_j \psi(0)\rangle,$$
$$\text{for } C_j \in \mathscr{C}(t_j,t_{j+1}), \quad t_1<t_2<\cdots. \tag{3.6}$$

Now, we define the time-shift operators S_t on $\mathscr{L}_k(-\infty,+\infty)$ by[26]

$$S_\tau \psi(f) = \psi(f_\tau), \quad f_\tau(t) = f(t+\tau); \tag{3.7}$$

$\{S_t; t \in \mathbb{R}\}$ is a strongly continuous one-parameter group of unitary operators. Moreover, we have

$$S_t \psi(0) = \psi(0), \tag{3.8}$$

$$S_\tau^+ \mathscr{C}(s,t) S_\tau = \mathscr{C}(s+\tau,t+\tau). \tag{3.9}$$

In a series of papers (see, for instance, Refs. 27–29) Hudson and Parthasarathy have developed a noncommutative stochastic calculus with respect to the basic operator processes A_t, A_t^+ (quantum Brownian motion) and Λ_t (gauge process). For simplicity, consider $\mathscr{L}(0,+\infty) \equiv \Gamma(L^2_{\mathbb{C}}(0,+\infty))$; then the basic processes are defined as follows:

$$A_t := a(\chi_{[0,t]}), \quad A_t^+ := a^+(\chi_{[0,t]}), \quad t \geqslant 0, \tag{3.10}$$

$$\Lambda_t \psi(f) := -i \frac{d}{d\epsilon} \psi(\exp[i\epsilon\chi_{[0,t]}] f)|_{\epsilon=0}, \quad t \geqslant 0, \tag{3.11}$$

where $a(f)$ and $a^+(f)$ are the annihilation and creation operators in the Fock space defined by

$$a(f)\psi(g) = \langle f,g\rangle \psi(g), \tag{3.12a}$$

$$a^+(f)\psi(g) = \frac{d}{d\epsilon} \psi(g+\epsilon f)|_{\epsilon=0}. \tag{3.12b}$$

Note that, using the heuristic notation usually adopted in theoretical physics, we can write

$$A_t = \int_0^t d\tau\, a(\tau), \quad A_t^+ = \int_0^t d\tau\, a^+(\tau), \tag{3.13a}$$

$$\Lambda_t = \int_0^t d\tau\, a^+(\tau) a(\tau), \tag{3.13b}$$

where we have introduced the boson field $a(t)$, $a^+(t)$ satisfying the CCR's

$$[a(t),a^+(t')] = \delta(t-t'), \quad [a(t),a(t')] = 0. \tag{3.14}$$

Consider now the space $\mathscr{h} \otimes \mathscr{L}(0,+\infty)$ ($\mathscr{h}$ is the "initial" Hilbert space); we identify A_t and $\mathbf{1} \otimes A_t$, etc. In this context, stochastic differentials[28] are defined by

$$dM(t) = E(t) d\Lambda_t + F(t) dA_t + G(t) dA_t^+ + H(t) dt, \tag{3.15}$$

where $M(t), E(t), F(t), G(t), H(t)$ are adapted, which means $M(t) \in B(\mathscr{h}) \otimes \mathscr{C}(0,t)$, etc. (or, if unbounded, they are affiliated to this algebra). The differential of a product can be evaluated by means of the quantum Ito's formula

$$d(MN) = (dM)N + M(dN) + (dM)(dN) \tag{3.16}$$

and of the multiplication table

	$d\Lambda_t$	dA_t	dA_t^+	dt
$d\Lambda_t$	$d\Lambda_t$	0	dA_t^+	0
dA_t	dA_t	0	dt	0.
dA_t^+	0	0	0	0
dt	0	0	0	0

(3.17)

The following theorem allows for the construction of *stochastic evolutions.*

Theorem 3.1 (Hudson and Parthasarathy[27,28]): Let $R, W, H \in B(\mathscr{h})$, W unitary and H self-adjoint. Then, there exists a unique strongly continuous adapted family of unitary operators on $\mathscr{h} \otimes \mathscr{L}(0,+\infty)$ such that

$$dU(t) = \{(W-1)d\Lambda_t - R^+ W dA_t + R\, dA_t^+ - (\tfrac{1}{2}R^+R + iH)dt\} U(t), \tag{3.18a}$$

$$U(0) = \mathbf{1}. \tag{3.18b}$$

The family $\{U(t), t \geqslant 0\}$ gives a unitary dilation of a quantum dynamical semigroup on $B(\mathscr{h})$. Indeed, let us define the operators $\mathscr{T}_t$, $t \geqslant 0$ on $B(\mathscr{h})$ by

$$\mathscr{T}_t(X) = E_0(U(t)^+ X U(t)), \quad X \in B(\mathscr{h}), \tag{3.19}$$

where the vacuum conditional expectation map $E_0: B(\mathscr{h} \otimes \mathscr{L}(0,+\infty)) \to B(\mathscr{h})$ is defined by

$$\langle u, E_0(J) v\rangle = \langle u \otimes \psi(0), J\, v \otimes \psi(0)\rangle,$$
$$u,v \in \mathscr{h}, \quad J \in B(\mathscr{h} \otimes \mathscr{L}(0,+\infty)). \tag{3.20}$$

Then, the following theorem holds.

Theorem 3.2 (Hudson and Parthasarathy[28]): $\{\mathscr{T}_t, t \geqslant 0\}$ is a uniformly continuous semigroup of completely positive maps on $B(\mathscr{h})$, whose infinitesimal generator $\mathscr{L}$ is given by

$$\mathscr{L}(X) = i[H,X] - \tfrac{1}{2}(R^+RX + XR^+R) + R^+XR,$$
$$X \in B(\mathscr{h}). \tag{3.21}$$

In the following, the operators S_t on $\mathscr{L}(-\infty,+\infty)$ defined by Eq. (3.7) and the operators $U(t)$ on $\mathscr{h} \otimes \mathscr{L}(0,+\infty)$ defined by Eq. (3.18) will be identified with the operators $\mathbf{1}_{\mathscr{h}} \otimes S_t$ and $\mathbf{1}_{\mathscr{L}(-\infty,0)} \otimes U(t)$, acting on $\mathscr{h} \otimes \mathscr{L}(-\infty,+\infty)$, respectively. For $s \leqslant t \in \mathbb{R}$, we define a two-parameter family of unitary operators on $\mathscr{h} \otimes \mathscr{L}(-\infty,+\infty)$ by

$$U(t,s) = S_s^+ U(t-s) S_s. \tag{3.22}$$

In Ref. 26 it is shown that the following properties hold:

$$S_\tau^+ U(t,s) S_\tau = U(t+\tau,s+\tau), \tag{3.23a}$$

$U(t,s)$ is strongly continuous in s and t, (3.23b)

$$U(t,s) \in B(\mathscr{h}) \otimes \mathscr{C}(s,t), \tag{3.23c}$$

$$U(t,s)U(s,r) = U(t,r), \quad r \leqslant s \leqslant t. \tag{3.23d}$$

More generally, a two-parameter family of unitary operators $U(t,s)$ on $\mathscr{h} \otimes \mathscr{L}_k(-\infty, +\infty)$ with the properties (3.23) is called a *covariant adapted unitary evolution*[26] (see also Refs. 23 and 25). So, the results of Ref. 20 allow for constructing a large class of such evolutions when $k = \mathbb{C}$ or, with a trivial extension, when $k = \mathbb{C}^N$. The case of an infinite-dimensional k could be treated along the lines of Ref. 29. We note also that in the general case the equation [cf. Eq. (3.19)]

$$\mathscr{T}_t(X) = E_0(U(t,0)^+ X U(t,0)), \quad X \in B(\mathscr{h}) \tag{3.24}$$

defines a quantum dynamical semigroup.[26]

IV. CONSTRUCTION OF OPERATION VALUED STOCHASTIC PROCESSES

Using the formalism of quantum stochastic calculus, OVSP's can be constructed. First, we establish the main result of this paper by exploiting only general structural properties. Then, we consider a particular case and obtain an explicit construction of a large class of OVSP's.

Let $\{V[\varphi]; \varphi \in \mathscr{D}\}$ be a family of unitary operators on $\mathscr{L}_k(-\infty, +\infty)$ with the following properties:

$$V^+[\varphi] = V[-\varphi], \tag{4.1a}$$

$$V[\varphi_1 + \varphi_2] = V[\varphi_1] V[\varphi_2], \tag{4.1b}$$

$$V[\varphi] \in \mathscr{C}(t_1,t_2), \quad \text{if } \operatorname{supp}(\varphi) \subset (t_1,t_2), \tag{4.1c}$$

$$V[\varphi] \text{ is strongly continuous in } \varphi, \tag{4.1d}$$

$$S_{\bar{t}} V[\varphi^{(\bar{t})}] S_{\bar{t}}^+ = V[\varphi], \tag{4.1e}$$

where $\operatorname{supp}(\varphi) = \cup_{s=1}^{n} \operatorname{supp}(\varphi_s)$; Equation (4.1a) and (4.1b) and unitarity imply

$$V[0] = 1. \tag{4.2}$$

Note that, by these properties and Eq. (3.6),

$$L[\varphi] := \langle \psi(0), V[\varphi] \psi(0) \rangle \tag{4.3}$$

is a normalized, continuous, positive definite functional on $\mathscr{D}$ with the further property

$$L[\varphi_1 + \varphi_2] = L[\varphi_1] L[\varphi_2],$$
$$\text{for } \operatorname{supp}(\varphi_1) \cap \operatorname{supp}(\varphi_2) = \phi; \tag{4.4}$$

therefore, $L[\varphi]$ is the characteristic functional of a GSP with independent values at every time.[30]

Consider now the space $\mathscr{h} \otimes \mathscr{L}_k(-\infty, +\infty)$ and identify $V[\varphi]$ with $1_{\mathscr{h}} \otimes V[\varphi]$; we have the following result.

Theorem 4.1: Let $V[\varphi]$ be the family of unitary operators introduced above and $U(t,s)$ a covariant adapted evolution on $\mathscr{h} \otimes \mathscr{L}_k(-\infty, +\infty)$. We define the operator $\mathscr{G}'(t,t_0;[\varphi])$, for $\operatorname{supp}(\varphi) \subset (t_0, t)$, from $B(\mathscr{h})$ into itself, by

$$\mathscr{G}'(t,t_0;[\varphi])(X) = E_0(U(t,t_0)^+ X V[\varphi] U(t,t_0)). \tag{4.5}$$

Then, $\mathscr{G}'(t,t_0;[\varphi])$ is the adjoint of an operator $\mathscr{G}(t,t_0;[\varphi])$ on $T(\mathscr{h})$ with the properties (a)–(f) of Sec. II; therefore it defines a time-translation invariant OVSP.

Proof: Property (a): As we have

$$\operatorname{Tr}_{\mathscr{h}}\{\rho \mathscr{G}'(t,t_0;[\varphi])(X)\}$$
$$= \operatorname{Tr}_{\mathscr{h} \otimes \mathscr{L}}\{(\rho \otimes |\psi(0)\rangle\langle\psi(0)|) U(t,t_0)^+ \times X V[\varphi] U(t,t_0)\},$$

then $\mathscr{G}'(t,t_0;[\varphi])$ is a linear operator on $B(\mathscr{h})$, continuous in the ultraweak topology.[2] This guarantees (see for instance Ref. 33) that a bounded linear operator $\mathscr{G}(t,t_0;[\varphi])$ exists from $T(\mathscr{h})$ into itself such that $\forall X \in B(\mathscr{h})$, $\forall \rho \in T(\mathscr{h})$

$$\operatorname{Tr}_{\mathscr{h}}\{\rho \mathscr{G}'(t,t_0;[\varphi])(X)\} = \operatorname{Tr}_{\mathscr{h}}\{X \mathscr{G}(t,t_0;[\varphi])(\rho)\}.$$

Property (b): From the definition (4.5) and Eq. (4.2), we have

$$\mathscr{G}'(t,t_0;[0])(1) = E_0(U(t,t_0)^+ V[0] U(t,t_0)) = 1,$$

and, therefore, $\mathscr{G}(t,t_0;[0])$ is trace preserving.

Property (c): From Eqs. (4.1a) and (4.1b) and the fact that $V[\varphi]$ commutes with all $X \in B(\mathscr{h})$, we can write

$$\sum_{i,j=1}^{n} \alpha_i^* \alpha_j \, \mathscr{G}'(t_2,t_1;[\varphi_j - \varphi_i])(X)$$
$$= E_0\left\{U(t,t_0)^+ \left(\sum_{i=1}^{n} \alpha_i^* V[\varphi_i]^+\right) \times X \left(\sum_{j=1}^{n} \alpha_j V[\varphi_j]\right) U(t,t_0)\right\}.$$

By the fact that E_0 is a completely positive map, property (c) follows.

Property (d): Let $\{\varphi_\alpha\}_{\alpha \in I}$ be a net in $\mathscr{D}_{(t_1,t_2)}$ converging to φ; then $V[\varphi_\alpha] \xrightarrow{s} V[\varphi]$ by Eq. (4.1d). This implies that $\mathscr{G}(t_2,t_1;[\varphi_\alpha]) \xrightarrow{s} \mathscr{G}(t_2,t_1;[\varphi])$; indeed, we have

$$\|\mathscr{G}(t_2,t_1;[\varphi_\alpha])(\rho) - \mathscr{G}(t_2,t_1;[\varphi])(\rho)\|_1$$
$$= \sup_{\substack{X \in B(\mathscr{h}) \\ \|X\|=1}} |\operatorname{Tr}_{\mathscr{h}}\{X(\mathscr{G}(t_2,t_1;[\varphi_\alpha]) - \mathscr{G}(t_2,t_1;[\varphi]))(\rho)\}|$$
$$= \sup_{\substack{X \in B(\mathscr{h}) \\ \|X\|=1}} |\operatorname{Tr}_{\mathscr{h} \otimes \mathscr{L}}\{X(V[\varphi_\alpha] - V[\varphi]) U(t_2,t_1) \times (\rho \otimes |\psi(0)\rangle\langle\psi(0)|) U(t_2,t_1)^+\}|;$$

the last quantity goes to zero because $V[\varphi_\alpha] Y \to V[\varphi] Y$ in the trace norm, $\forall Y \in T(\mathscr{h} \otimes \mathscr{L})$ (see for instance Ref. 2, p. 6).

Property (e): By using Eqs. (3.23d), (3.23c), (4.1c), and (3.6), we have

$$\forall \varphi_1 \in \mathscr{D}_{(t_1,t_2)}, \quad \forall \varphi_2 \in \mathscr{D}_{(t_2,t_3)}, \quad t_1 < t_2 < t_3,$$
$$\mathscr{G}'(t_3,t_1;[\varphi_1 + \varphi_2])(X)$$
$$= E_0(U(t_2,t_1)^+ U(t_3,t_2)^+ X V[\varphi_2] \times U(t_3,t_2) V[\varphi_1] U(t_2,t_1))$$
$$= E_0(U(t_2,t_1)^+ E_0(U(t_3,t_2)^+ X V[\varphi_2] \times U(t_3,t_2)) V[\varphi_1] U(t_2,t_1))$$
$$= \mathscr{G}'(t_2,t_1;[\varphi_1]) \mathscr{G}'(t_3,t_2;[\varphi_2])(X).$$

Therefore, $\mathscr{G}(\cdots)$ satisfies Eq. (2.8).

Property (f): Using Eqs. (3.23a) and (4.1e) and the invariance of the vacuum under time shift, we have

$$\mathscr{G}'(t_2 + \bar{t}, t_1 + \bar{t}; [\varphi^{(\bar{t})}])(X)$$
$$= E_0(U(t_2 + \bar{t}, t_1 + \bar{t})^+ X V[\varphi^{(\bar{t})}] U(t_2 + \bar{t}, t_1 + \bar{t}))$$
$$= E_0(S_{\bar{t}}^+ U(t_2,t_1)^+ X S_{\bar{t}} V[\varphi^{(\bar{t})}] S_{\bar{t}}^+ U(t_2,t_1) S_{\bar{t}})$$
$$= E_0(U(t_2,t_1)^+ X V[\varphi] U(t_2,t_1))$$
$$= \mathscr{G}'(t_2,t_1;[\varphi])(X),$$

and, so, Eq. (2.9) holds. Q.E.D.

Now, we consider an explicit case that allows for the construction of a large and very interesting class of OVSP's. Take $k=\mathbb{C}^N$, so that the Hilbert space is $k\otimes\mathscr{L}_N(-\infty,+\infty)$, $\mathscr{L}_N(-\infty,+\infty)\equiv\Gamma(L^2_{\mathbb{C}^N}(-\infty,+\infty))$, $L^2_{\mathbb{C}^N}(-\infty,+\infty)\equiv\mathbb{C}^N\otimes L^2(\mathbf{R})$. In this space the basic differentials $dA_t^{(j)}, dA_t^{(j)}, dA_t^{(j)+}, dt$ $(j=1,\dots,N)$ are introduced; they can be defined for $t\geqslant 0$ as in Eqs. (3.10) and (3.11) and then shifted to all times by

$$S_\tau^+\, dA_t^{(j)} S_\tau = dA_{\tau+t}^{(j)}, \quad \text{etc.} \tag{4.6}$$

Products of differentials with different indices vanish, while products of differentials with the same index are given by the multiplication table (3.17).

First, we introduce in $k\otimes\mathscr{L}_N(-\infty,+\infty)$ a stochastic evolution by

$$\begin{aligned} dU(t,t_0) = \Big\{ \sum_{j=1}^{N} \big[(W_j-1)dA_t^{(j)} - R_j^+\, W_j\, dA_t^{(j)} \\ + R_j\, dA_t^{(j)+} - \tfrac{1}{2} R_j^+\, R_j\, dt \big] - iH\,dt \Big\} \\ \times U(t,t_0), \end{aligned} \tag{4.7a}$$

$$U(t_0,t_0)=\mathbf{1}. \tag{4.7b}$$

By a trivial extension of Theorem 3.1, the operators $U(t,t_0)$ are unitary and enjoy properties (3.23); thus, $\{U(t,s), s<t\in\mathbf{R}\}$ is a covariant adapted unitary evolution.

Then, we introduce the operators $V[\varphi]$ by giving their action on the exponential vectors and extending them by linearity and closure to the whole Hilbert space

$$\begin{aligned} V[\varphi]\psi(f) = \exp\int_{-\infty}^{+\infty} dt \Big\{ \sum_{j=1}^{N} \Big[\frac{\beta_j}{\lambda_j}\Big(\frac{\beta_j}{\lambda_j}+f_j(t)\Big) \\ \times\big(e^{i\varphi(t)\cdot\alpha^{(j)}\lambda_j}-1\big) - i\varphi(t)\cdot\alpha^{(j)}\frac{\beta_j^2}{\lambda_j}\Big] \\ + i\varphi(t)\cdot\mathbf{c} \Big\}\, \psi\big(g^{(f)}\big), \end{aligned} \tag{4.8}$$

where φ is a n-component real integrable function and

$$\alpha^{(j)},\mathbf{c}\in\mathbf{R}^n, \quad \lambda_j,\beta_j\in\mathbf{R}, \tag{4.9a}$$

$$g_j^{(f)}(t) = f_j(t) + \Big(f_j(t)+\frac{\beta_j}{\lambda_j}\Big)\big(e^{i\varphi(t)\cdot\alpha^{(j)}\lambda_j}-1\big). \tag{4.9b}$$

Note that one can have $\lambda_j=0$ for some j; in this case the right-hand sides of Eqs. (4.8) and (4.9b) are defined as the limit for $\lambda_j\to 0$; the same convention is adopted in all the following formulas.

The operators defined by Eqs. (4.8) and (4.9) enjoy all the properties (4.1). This can be easily proved by exploiting the definition (4.8) and (4.9); moreover, operators of this kind are extensively studied in Ref. 28. Therefore, by Eq. (4.5), we can define the characteristic operator of an OVSP. Now, we want to see what the explicit structure of such a characteristic operator is.

Let us put

$$V(t,t_0;[\varphi]) := V[\chi_{[t_0,t]}\varphi], \quad t\geqslant t_0, \quad \varphi\in\mathscr{D}. \tag{4.10}$$

By computing $\langle\psi(g), V(t,t_0;[\varphi])\,\psi(f)\rangle$ from Eqs. (4.8) and (4.9) and by differentiating this quantity with respect to t, we obtain a differential equation for $\langle\psi(g), V(t,t_0;[\varphi])\,\psi(f)\rangle$. By comparing this result with the definition of quantum stochastic differential for an adapted process,[28] we have that $V(t,t_0;[\varphi])$ satisfies the following stochastic differential equation:

$$\begin{aligned} dV(t,t_0;[\varphi]) \\ = \Big\{ \sum_{j=1}^{N} \Big[\big(e^{i\varphi(t)\cdot\alpha^{(j)}\lambda_j}-1\big)\Big(dA_t^{(j)} \\ + \frac{\beta_j}{\lambda_j}\big(dA_t^{(j)}+dA_t^{(j)+}\big) + \frac{\beta_j^2}{\lambda_j^2}dt \Big) \\ - i\varphi(t)\cdot\alpha^{(j)}\frac{\beta_j^2}{\lambda_j}dt \Big] \\ + i\varphi(t)\cdot\mathbf{c}\,dt \Big\}\, V(t,t_0;[\varphi]). \end{aligned} \tag{4.11}$$

Vice versa, $V(t,t_0;[\varphi])$ is the unique solution of Eq. (4.11) (Ref. 28), with the initial condition

$$V(t_0,t_0;[\varphi]) = \mathbf{1}. \tag{4.12}$$

Now, we define

$$\begin{aligned} \mathscr{G}'(t,t_0;[\varphi])(X) \\ = E_0\big(U(t,t_0)^+ X V(t,t_0;[\varphi]) U(t,t_0)\big), \quad \forall X\in B(k). \end{aligned} \tag{4.13}$$

By theorem 4.1, $\{\mathscr{G}'(t_2,t_1;[\varphi]), \varphi\in\mathscr{D}_{(t_1,t_2)}, t_1<t_2\in\mathbf{1}\}$ is the adjoint of the characteristic functional of an OVSP.

Using the quantum Itô's formula (3.16), the multiplication table (3.17), and the differential equations (4.7a) and (4.11), we obtain

$$\begin{aligned} d_{(t)}\big(U(t,t_0)^+ X V(t,t_0;[\varphi]) U(t,t_0)\big) \\ = U(t,t_0)^+ \Big\{ \sum_{j=1}^{N} \Big[\big(W_j^+ X W_j e^{i\varphi(t)\cdot\alpha^{(j)}\lambda_j} - X\big) d\Lambda_t^{(j)} \\ + \Big(\Big(R_j^+ + \frac{\beta_j}{\lambda_j}\Big) X e^{i\varphi(t)\cdot\alpha^{(j)}\lambda_j} - X\Big(R_j^+ + \frac{\beta_j}{\lambda_j}\Big)\Big) \\ \times W_j\, dA_t^{(j)} - W_j^+\Big(\Big(R_j+\frac{\beta_j}{\lambda_j}\Big)X - X\Big(R_j+\frac{\beta_j}{\lambda_j}\Big) \\ \times e^{i\varphi(t)\cdot\alpha^{(j)}\lambda_j}\Big) dA_t^{(j)+} + \Big(R_j^+ X R_j - \frac{1}{2}\{R_j^+ R_j, X\} \\ + \Big(R_j^+ + \frac{\beta_j}{\lambda_j}\Big) X \Big(R_j + \frac{\beta_j}{\lambda_j}\Big)\big(e^{i\varphi(t)\cdot\alpha^{(j)}\lambda_j}-1\big) \\ - i\varphi(t)\cdot\alpha^{(j)}\frac{\beta_j^2}{\lambda_j} X\Big) dt \Big] + i([H,X] \\ + \varphi(t)\cdot\mathbf{c}X)dt \Big\}\, V(t,t_0;[\varphi]) U(t,t_0), \end{aligned} \tag{4.14}$$

where $\{A,B\}$ is the anticommutator between A and B. By taking the vacuum conditional expectation of Eq. (4.14) all terms containing $dA_t^{(j)}, dA_t^{(j)}, dA_t^{(j)+}$ vanish and a differential equation for $\mathscr{G}'(\cdots)$ is found. For $\mathscr{G}(\cdots)$, this equation becomes

$$\frac{\partial}{\partial t}\mathscr{G}(t,t_0;[\varphi]) = \mathscr{K}(\varphi(t))\mathscr{G}(t,t_0;[\varphi]), \tag{4.15a}$$

$$\mathscr{G}(t_0,t_0;[\boldsymbol{\varphi}]) = \mathbf{1}, \tag{4.15b}$$

where, $\forall \rho \in T(\mathscr{h})$,

$$\mathscr{K}(\boldsymbol{\varphi})(\rho) = \mathscr{L}(\rho) + \sum_{j=1}^{N}\left[\left(R_j + \frac{\beta_j}{\lambda_j}\right)\rho\left(R_j^{+} + \frac{\beta_j}{\lambda_j}\right) \times (e^{i\boldsymbol{\varphi}\cdot\boldsymbol{\alpha}^{(j)}\lambda_j} - 1) - i\boldsymbol{\varphi}\cdot\boldsymbol{\alpha}^{(j)}\frac{\beta_j^2}{\lambda_j}\rho\right] + i\boldsymbol{\varphi}\cdot\mathbf{c}\rho, \tag{4.16}$$

$$\mathscr{L}(\rho) = \sum_{j=1}^{N}(R_j\rho R_j^{+} - \tfrac{1}{2}\{R_j^{+}R_j,\rho\}) - i[H,\rho]. \tag{4.17}$$

Equation (4.17) gives the infinitesimal generator of the quantum dynamical semigroup (2.10) associated with the OVSP we have constructed.

If we take $\lambda_j \neq 0$, for $j = 1,...,M$ and $\lambda_j = 0$, for $j = M+1,...,N$, by making the replacements

$$\mathbf{c} \to \mathbf{c} + \sum_{j=1}^{M}\boldsymbol{\alpha}^{(j)}\frac{\beta_j^2}{\lambda_j}, \tag{4.18a}$$

$$\boldsymbol{\alpha}^{(j)} \to \boldsymbol{\alpha}^{(j)}/\lambda_j, \quad j = 1,...,M,$$
$$\boldsymbol{\alpha}^{(j)} \to \boldsymbol{\alpha}^{(j)}/\beta_j, \quad j = M+1,...,N, \tag{4.18b}$$

$$R_j \to R_j - \beta_j/\lambda_j, \quad j = 1,...,M, \tag{4.18c}$$

$$H \to H + \frac{i}{2}\sum_{j=1}^{M}\frac{\beta_j}{\lambda_j}(R_j - R_j^{+}), \tag{4.18d}$$

we find that $\mathscr{K}(\boldsymbol{\varphi})$ can be written as

$$\mathscr{K}(\boldsymbol{\varphi})(\rho) = \mathscr{L}(\rho) + \sum_{j=1}^{M}R_j\rho R_j^{+}(e^{i\boldsymbol{\varphi}\cdot\boldsymbol{\alpha}^{(j)}} - 1) + \sum_{j=M+1}^{N}\left[(R_j\rho + \rho R_j^{+})i\boldsymbol{\varphi}\cdot\boldsymbol{\alpha}^{(j)} - \frac{1}{2}(\boldsymbol{\varphi}\cdot\boldsymbol{\alpha}^{(j)})^2\rho\right] + i\boldsymbol{\varphi}\cdot\mathbf{c}\rho, \tag{4.19}$$

whilst $\mathscr{L}$ does not change.

As the formal solution of Eqs. (4.15) is

$$\mathscr{G}(t_2,t_1;[\boldsymbol{\varphi}]) = T\exp\int_{t_1}^{t_2}dt\,\mathscr{K}(\boldsymbol{\varphi}(t)), \tag{4.20}$$

where T means the time-ordered product, we can identify in Eq. (4.19) a Poisson and a Gaussian contribution; therefore, Eq. (4.19) generalizes the results of Ref. 12, where only the Gaussian part was found. It can be also shown (see Sec. VI) that the pure Poisson case corresponds to the class of continual measurements studied by Davies and Srinivas.[2,5–10]

In Ref. 12 it is shown how the moments of the continually measured quantities $\mathbf{x}(t)$ can be obtained by functional differentiation of the characteristic operator. In particular, by Eq. (4.19), for the mean values we have

$$\langle x_i(t)\rangle_{(\rho,t_0)} = \mathrm{Tr}\left\{\left[\sum_{j=1}^{M}\alpha_i^{(j)}R_j^{+}R_j + \sum_{j=M+1}^{N}\alpha_i^{(j)}(R_j^{+} + R_j) + c_i\right]\rho(t)\right\}, \tag{4.21}$$

where

$$\rho(t) = \exp[(t-t_0)\mathscr{L}](\rho). \tag{4.22}$$

So, we can say that the continually measured quantity $x_i(t)$ is represented by the self-adjoint operator

$$C_i = \sum_{j=1}^{M}\alpha_i^{(j)}R_j^{+}R_j + \sum_{j=M+1}^{N}\alpha_i^{(j)}(R_j^{+} + R_j) + c_i. \tag{4.23}$$

V. THE DILATION

In the previous section a certain class of OVSP's on $T(\mathscr{h})$ has been constructed starting from quantities defined in $\mathscr{h}\otimes\Gamma(L^2_{\mathbb{C}^N}(\mathbf{R}))$; therefore we have also obtained a *dilation* of this class of OVSP's. In this section we want to show the main features of this dilation. For simplicity, we work only at a formal level, though rigorous statements could be given. By using the time-shift operators (3.7) we can define the operators

$$T(t) := \begin{cases} S_t U(t,0), & \text{for } t \geqslant 0, \\ U(|t|,0)^{+}S_{|t|}^{+}, & \text{for } t < 0; \end{cases} \tag{5.1}$$

$\{T(t), t\in\mathbf{R}\}$ is a strongly continuous one-parameter group of unitary operators on $\mathscr{h}\otimes\Gamma(L^2_{\mathbb{C}^N}(\mathbf{R}))$ (Ref. 26), which gives the dynamics of the global system. We have also

$$S_t U(t,t_0)S_{t_0}^{+} = T(t-t_0). \tag{5.2}$$

Then, consider the formal solution of Eq. (4.11), with the initial condition (4.12), which is given by

$$V(t_f,t_0;[\boldsymbol{\varphi}]) = \exp\int_{t\in(t_0,t_f)} i\sum_{i=1}^{n}\varphi_i(t)\left\{\sum_{j=1}^{N}\alpha_i^{(j)}[\lambda_j\,dA_t^{(j)} + \beta_j(dA_t^{(j)} + dA_t^{(j)+})] + c_i\,dt\right\}; \tag{5.3}$$

no time ordering is needed because the operators $\lambda_j dA_t^{(j)} + \beta_j(dA_t^{(j)} + dA_t^{(j)+})$ commute, also at different times.

Using Eqs. (5.3), (3.8), and (4.13), we have that the effect valued measure (2.14) can be written as

$$\widetilde{F}_{(h^{(1)},...,h^{(s)})}(t_0;B) = E_0(P_{(h^{(1)},...,h^{(s)})}(t_0;B)), \tag{5.4}$$

where

$$P_{(h^{(1)},...,h^{(s)})}(t_0;B) := \int_B d_s\mathbf{x}\frac{1}{(2\pi)^s}\int d_s\mathbf{k} \times\exp\left[i\sum_{r=1}^{s}k_r(O(\mathbf{h}^{(r)};t_0) - x_r)\right] \tag{5.5}$$

is the projection valued measure associated with the commuting self-adjoint operators $O(\mathbf{h}^{(r)};t_0)$ given by

$$O(\mathbf{h}^{(r)};t_0) := S_{t_0}U(t_f,t_0)^{+}\int_{t\in(t_0,t_f)}\sum_{i=1}^{n}h_i^{(r)}(t) \times\left\{\sum_{j=1}^{N}[\lambda_j\,dA_t^{(j)} + \beta_j(dA_t^{(j)} + dA_t^{(j)+})] + c_i\,dt\right\}U(t_f,t_0)S_{t_0}^{+}. \tag{5.6}$$

A more revealing expression can be given to the observables $O(\mathbf{h}^{(r)},t_0)$. For $t_0 \leqslant t < t_f$, we have

$$S_{t_0} U(t_f,t_0)^+ dA_t^{(j)} U(t_f,t_0) S_{t_0}^+$$
$$= S_{t_0} d_{(t)} (U(t_f,t_0)^+ A_{t,t_0}^{(j)} U(t_f,t_0)) S_{t_0}^+$$
$$= S_{t_0} d_{(t)} (U(t,t_0)^+ A_{t,t_0}^{(j)} U(t,t_0)) S_{t_0}^+$$
$$= S_{t_0} U(t,t_0)^+ (dA_t^{(j)} + R_j^+ W_j dA_t^{(j)} + W_j^+ R_j dA_t^{(j)+} + R_j^+ R_j dt) U(t,t_0) S_{t_0}^+$$
$$= T(t-t_0)^+ S_{-t}^+ (dA_t^{(j)} + R_j^+ W_j dA_t^{(j)} + W_j^+ R_j dA_t^{(j)+} + R_j^+ R_j dt) S_{-t} T(t-t_0), \quad (5.7)$$

where

$$A_{t,t_0}^{(j)} = \int_{t' \in (t_0,t)} dA_{t'}^{(j)}. \quad (5.8)$$

In the third step of Eq. (5.7) we have used Eq. (3.23d) and the fact that $U(t_f,t) \in B(\mathscr{h}) \otimes \mathscr{C}(t,t_f)$ and $A_{t,t_0}^{(j)} \in \mathbf{1} \otimes \mathscr{C}(t_0,t)$; in the fourth step we have used Eq. (4.7a) and the rules (3.16) and (3.17) of quantum stochastic calculus; in the fifth step Eq. (5.2) has been used. In an analogous way, we find

$$S_{t_0} U(t_f,t_0)^+ (dA_t^{(j)} + dA_t^{(j)+}) U(t_f,t_0) S_{t_0}^+$$
$$= T(t-t_0)^+ S_{-t}^+ [W_j dA_t^{(j)} + W_j^+ dA_t^{(j)+} + (R_j + R_j^+) dt] S_{-t} T(t-t_0). \quad (5.9)$$

Moreover, by using Eq. (4.6), we write

$$S_{-t}^+ dA_t^{(j)} S_{-t} = dA_0^{(j)}, \quad \text{etc.}, \quad (5.10)$$

where $dA_0^{(j)}$ means the differential of the gauge process in the time interval $(0,dt)$. Therefore, for the observables $O(\mathbf{h}^{(r)};t_0)$ we have

$$O(\mathbf{h}^{(r)};t_0)$$
$$= \sum_{i=1}^{n} \int_{t \in (t_0,t_f)} h_i^{(r)}(t) T(t-t_0)^+ \Big\{ \sum_{j=1}^{N} \alpha_i^{(j)} \Big[\lambda_j dA_0^{(j)} + (\lambda_j R_j^+ + \beta_j) W_j dA_0^{(j)} + W_j^+ (\lambda_j R_j + \beta_j) dA_0^{(j)+} + (\lambda_j R_j^+ R_j + \beta_j R_j + \beta_j R_j^+) dt \Big] + c_i\, dt \Big\} T(t-t_0). \quad (5.11)$$

By Eqs. (5.4), (5.5), and (5.11) and the arbitrariness of the test functions $\mathbf{h}^{(r)}(t)$, we can say that we have obtained a dilation of the effect valued measure associated with the considered OVSP to a projection valued measure, which is constructed along the lines of the standard formulation of quantum mechanics, starting from a set of self-adjoint operators; these operators, in the Heisenberg picture, commute also at different times.

The operation valued measure $\mathscr{F}(t_f,t_0;N)$ (see Sec. II), defining the OVSP, can be obtained from the characteristic operator by the Fourier transform

$$\mathscr{F}(t_f,t_0;(x_{h^{(1)}},\ldots,x_{h^{(s)}}) \in B)$$
$$= \int_B d_s\mathbf{x} \frac{1}{(2\pi)^s} \int d_s\mathbf{k}\, e^{-i\mathbf{k}\cdot\mathbf{x}} \times \mathscr{G}\Big(t_f,t_0; \Big[\sum_{r=1}^{s} k_r \mathbf{h}^{(r)}(t)\Big]\Big). \quad (5.12)$$

By using Eqs. (3.8), (4.1), (4.13), (5.3), and (5.6) we obtain for the adjoint of $\mathscr{F}(\cdots)$

$$\mathscr{F}'(t_f,t_0;(x_{h^{(1)}},\ldots,x_{h^{(s)}}) \in B)(X)$$
$$= \int_B d_s\mathbf{x} \frac{1}{(2\pi)^s} \int d_s\mathbf{k} \exp(-i\mathbf{k}\cdot\mathbf{x}) E_0\Big\{ S_{t_0} U(t_f,t_0)^+ V\Big(t_f,t_0; \Big[\frac{1}{2}\sum_{r=1}^{s} k_r \mathbf{h}^{(r)}\Big]\Big) \times X V\Big(t_f,t_0; \Big[\frac{1}{2}\sum_{r=1}^{s} k_r \mathbf{h}^{(r)}\Big]\Big) U(t_f,t_0) S_{t_0}^+ \Big\}$$
$$= \lim_{\epsilon\to 0^+} \int_B d_s\mathbf{x} \frac{1}{(2\pi)^s} \int d_s\mathbf{k} \Big(\frac{\epsilon}{4\pi}\Big)^{s/2} \int d_s\boldsymbol{\eta} \exp\Big[-\frac{\epsilon}{4}(|\mathbf{k}|^2 + |\boldsymbol{\eta}|^2) - i\mathbf{k}\cdot\mathbf{x}\Big] E_0\Big\{ S_{t_0} U(t_f,t_0)^+ \times V\Big(t_f,t_0; \Big[\frac{1}{2}\sum_{r=1}^{s} (k_r - \eta_r)\mathbf{h}^{(r)}\Big]\Big) X V\Big(t_f,t_0; \Big[\frac{1}{2}\sum_{r=1}^{s} (k_r + \eta_r)\mathbf{h}^{(r)}\Big]\Big) \times U(t_f,t_0) S_{t_0}^+ \Big\}$$
$$= \lim_{\epsilon\to 0^+} \int_B d_s\mathbf{x} \frac{1}{(2\pi)^s} \int d_s\mathbf{k} \Big(\frac{\epsilon}{4\pi}\Big)^{s/2} \int d_s\boldsymbol{\eta} \exp\Big[-\frac{\epsilon}{4}(|\mathbf{k}|^2 + |\boldsymbol{\eta}|^2) - i\mathbf{k}\cdot\mathbf{x}\Big] \mathrm{E}_0 \Big\{ \exp\Big[\frac{i}{2}\sum_{r=1}^{s} (k_r - \eta_r) \times O(\mathbf{h}^{(r)};t_0)\Big] S_{t_0} U(t_f,t_0)^+ S_{t_f}^+ X S_{t_f} U(t_f,t_0) S_{t_0}^+ \exp\Big[\frac{i}{2}\sum_{r=1}^{s} (k_r + \eta_r) O(\mathbf{h}^{(r)};t_0)\Big]\Big\}$$
$$= \lim_{\epsilon\to 0^+} \int_B d_s\mathbf{x} \frac{1}{(2\pi)^s} \int d_s\,\mathbf{k} \exp\Big[-\frac{\epsilon}{4}|\mathbf{k}|^2 - i\mathbf{k}\cdot\mathbf{x}\Big] E_0\Big\{ \exp\sum_{r=1}^{s} \Big(\frac{i}{2} k_r \{O(\mathbf{h}^{(r)};t_0),\cdot\} - \frac{1}{4\epsilon}[O(\mathbf{h}^{(r)};t_0),[O(\mathbf{h}^{(r)};t_0),\cdot]]\Big) (T(t_f-t_0)^+ X T(t_f-t_0))\Big\}$$
$$= \lim_{\epsilon\to 0^+} \Big(\frac{1}{\pi\epsilon}\Big)^{s/2} \int_B d_s\mathbf{x}\, E_0 \Big\{ \exp\Big[-\frac{1}{2\epsilon}\sum_{r=1}^{s} (x_r - O(\mathbf{h}^{(r)};t_0))^2\Big] T(t_f-t_0)^+ X T(t_f-t_0) \times \exp\Big[-\frac{1}{2\epsilon}\sum_{r=1}^{s} (x_r - O(\mathbf{h}^{(r)};t_0))^2\Big]\Big\}. \quad (5.13)$$

Consider now the case in which all the operators $O(\mathbf{h}^{(r)};t_0)$ have purely discrete spectra. Let P_i be the eigenprojectors and $\lambda_r^{(i)}$ the eigenvalues, i.e.,

$$O(\mathbf{h}^{(r)};t_0)P_i=\lambda_r^{(i)}P_i,\quad r=1,\dots,s. \tag{5.14}$$

In this situation Eq. (5.13) gives

$$\begin{aligned}&\mathscr{F}(t_f,t_0;(x_{h^{(1)}},\dots,x_{h^{(s)}})\in B)(\rho)\\&\quad=\sum_{\substack{i\\ \lambda^{(i)}\in B}}\mathrm{Tr}_\Gamma\{T(t_f-t_0)\\&\qquad\times P_i(\rho\otimes|\psi(0)\rangle\langle\psi(0)|)P_iT(t_f-t_0)^+\},\end{aligned} \tag{5.15}$$

where Tr_Γ is the partial trace over the Fock space. Equation (5.15) can be interpreted by saying that for the global system the state after the continual measurement is given by the Von Neumann reduction postulate. In the case of not purely discrete spectra, Eq. (5.13) gives, in some sense, a formal generalization of the reduction postulate.

In conclusion, as far as the global system is concerned, we can say that continual measurements are obtained by applying the standard formulation of quantum mechanics: observables are associated with commuting self-adjoint operators, projection valued measures describe measurements, reduction postulate holds, etc. Obviously, these statements are very formal: the operators involved are "distribution valued operators" (only time smoothed operators have meaning), the spectra are not always purely discrete, etc.

VI. THE PURE POISSON CASE

In this paper continual measurements are treated using the language of Refs. 11–18, where only the pure Gaussian case was studied; it is interesting to show explicitly how the formalism developed includes also the "quantum stochastic processes" (counting processes) of Davies and Srinivas.[2,5–10]

Let us consider the simplest Poisson case; take in Eq. (4.19) $M=N=1$, $n=1$, and $c=0$, so that the generator $\mathscr{K}(\varphi)$ can be written as

$$\mathscr{K}(\varphi)(\rho)=R\rho R^+e^{i\alpha\varphi}-\tfrac{1}{2}\{R^+R,\rho\}-i[H,\rho]. \tag{6.1}$$

Choose

$$\varphi(t')=(k/\alpha)\chi_{[t,t+\tau]}(t') \tag{6.2}$$

and compute $\mathscr{G}(t+\tau,t;[\varphi])$ from Eqs. (4.15); we obtain

$$\begin{aligned}&\mathscr{G}(t+\tau,t;[(k/\alpha)\chi_{[t,t+\tau]}])\\&\quad=\sum_{m=0}^{\infty}\int_0^\tau d\tau_m\int_0^{\tau_m}d\tau_{m-1}\cdots\int_0^{\tau_2}d\tau_1\,e^{imk}\\&\qquad\times\mathscr{U}_{\tau-\tau_m}J\mathscr{U}_{\tau_m-\tau_{m-1}}J\cdots\mathscr{U}_{\tau_2-\tau_1}J\mathscr{U}_{\tau_1},\end{aligned} \tag{6.3}$$

where

$$J(\rho)=R\rho R^+, \tag{6.4a}$$

$$\mathscr{U}_t(\rho)=\exp\Big(-\frac{t}{2}R^+R-itH\Big)\rho\times\exp\Big(-\frac{t}{2}R^+R+itH\Big). \tag{6.4b}$$

Now, let us introduce the cylinder set

$$C_B=\{x\in\mathscr{D}':x_{\chi_{[t,t+\tau)}/\alpha}\in B\}; \tag{6.5}$$

for the operation valued measure introduced in Sec. II we obtain, using Eq. (6.3),

$$\begin{aligned}&\mathscr{F}(t+\tau,t;C_B)\\&\quad=\int_B dx\frac{1}{2\pi}\int_{-\infty}^{+\infty}dk\,e^{-ikx}\mathscr{G}\Big(t+\tau,t;\Big[\frac{k}{\alpha}\chi_{[t,t+\tau)}\Big]\Big)\\&\quad=\sum_{m=0}^{\infty}\int_0^\tau d\tau_m\int_0^{\tau_m}d\tau_{m-1}\cdots\int_0^{\tau_2}d\tau_1\mathscr{U}_{\tau-\tau_m}\\&\qquad\times J\mathscr{U}_{\tau_m-\tau_{m-1}}J\cdots\mathscr{U}_{\tau_2-\tau_1}J\mathscr{U}_{\tau_1}\chi_B(m)\\&\quad=\sum_{m\in B}N_\tau(m),\end{aligned} \tag{6.6}$$

where the operations $N_\tau(m)$ are defined by

$$\begin{aligned}N_\tau(m):=&\int_0^\tau d\tau_m\int_0^{\tau_m}\tau_{m-1}\cdots\int_0^{\tau_2}d\tau_1\mathscr{U}_{\tau-\tau_m}J\mathscr{U}_{\tau_m-\tau_{m-1}}\\&\times J\cdots\mathscr{U}_{\tau_2-\tau_1}J\mathscr{U}_{\tau_1},\quad\text{for } m\geqslant 1,\end{aligned} \tag{6.7a}$$

$$N_\tau(0):=\mathscr{U}_\tau. \tag{6.7b}$$

The quantities $N_\tau(m)$ are the basic objects introduced by Davies and Srinivas [cf. Ref. 10, Eqs. (3.4)–(3.9), (3.14), (3.18), and (3.27)]; the meaning of these objects is given by the assumption that the probability of having m counts in the time interval $[t,t+\tau)$, when the system is in the state $\rho(t)$ at time t, is given by

$$P(m;\tau|\rho,t)=\mathrm{Tr}\{N_\tau(m)(\rho(t))\}. \tag{6.8}$$

Moreover, from the operations $N_\tau(m)$, more complicated joint probabilities can be obtained, coincidence experiments studied, etc.[10]

Therefore, from an OVSP of pure Poisson type we have obtained a counting process $\{N_\tau(m)\}$. Vice versa, given a counting process one can reconstruct an OVSP. Indeed, let $N_\tau(m)$ be defined by Eqs. (6.7); we can introduce the characteristic operator of this process by

$$\begin{aligned}&\mathscr{G}(t_0+\tau,t_0;[\varphi])\\&\quad=\lim_{k\to+\infty}\sum_{\{m_j\}}\exp\Big[i\sum_{j=1}^{k}m_j\varphi\Big(t_0+j\frac{\tau}{k}\Big)\Big]\\&\qquad\times N_{\tau/k}(m_k)\cdots N_{\tau/k}(m_2)N_{\tau/k}(m_1).\end{aligned} \tag{6.9}$$

The generator of this characteristic operator is given by

$$\begin{aligned}\mathscr{K}(\varphi(t))&=\lim_{k\to+\infty}\frac{k}{\tau}\Big\{\sum_{m=0}^{\infty}e^{im\varphi(t)}N_{\tau/k}(m)-1\Big\}\\&=\lim_{k\to+\infty}\frac{k}{\tau}\Big\{\mathscr{U}_{\tau/k}+e^{i\varphi(t)}\\&\qquad\times\int_0^{\tau/k}d\tau_1\,\mathscr{U}_{\tau/k-\tau_1}J\mathscr{U}_{\tau_1}-1\Big\}\\&=-\tfrac{1}{2}\{R^+R,\cdot\}-i[H,\cdot]+e^{i\varphi(t)}J.\end{aligned} \tag{6.10}$$

Apart from the quantity α, that can be reintroduced by rescaling the observables, the generators (6.10) and (6.1) coincide; therefore, we have reobtained an OVSP. More generally, one can see that the OVSP's of pure Poisson type [Eq. (4.19) with $M=N$] coincide with the general counting processes introduced in Refs. 2 and 5–9.

ACKNOWLEDGMENTS

The authors are particularly indebted to A. Frigerio for his constant interest in this work and for many stimulating discussions.

[1]G. Ludwig, *Foundations of Quantum Mechanics* (Springer, Berlin, 1982).

[2]E. B. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976).

[3]K. Kraus, "States, effects and operations," in *Lecture Notes in Physics*, Vol. 190 (Springer, Berlin, 1983).

[4]A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).

[5]E. B. Davies, "Quantum stochastic processes," Commun. Math. Phys. **15**, 227 (1969).

[6]E. B. Davies, "Quantum stochastic processes II," Commun. Math. Phys. **19**, 83 (1970).

[7]E. B. Davies, "Quantum stochastic processes III." Commun. Math. Phys. **22**, 51 (1971).

[8]E. B. Davies, "Quantum communication systems," IEEE Trans. Inf. Theory **23**, 530 (1977).

[9]M. D. Srinivas, "Quantum counting processes," J. Math. Phys. **18**, 2138 (1977).

[10]M. D. Srinivas and E. B. Davies, "Photon counting probabilities in quantum optics," Opt. Acta **28**, 981 (1981).

[11]A. Barchielli, L. Lanz, and G. M. Prosperi, "A model for macroscopic description and continual observations in quantum mechanics," Nuovo Cimento B **72**, 79 (1982).

[12]A. Barchielli, L. Lanz, and G. M. Prosperi, "Statistics of continuous trajectories in quantum mechanics: operation valued stochastic processes," Found. Phys. **13**, 779 (1983).

[13]A. Barchielli, "Continual measurements for quantum open systems," Nuovo Cimento B **74**, 113 (1983).

[14]G. Lupieri, "Generalized stochastic processes and continual observations in quantum mechanics," J. Math. Phys. **24**, 2329 (1983).

[15]G. M. Prosperi, "The quantum measurement process and the observation of continuous trajectories," in *Quantum Probability and Applications to the Quantum Theory of Irreversible Processes, Lecture Notes in Mathematics*, Vol. 1055, edited by L. Accardi, A. Frigerio, and V. Gorini (Springer, Berlin, 1984), pp. 301–326.

[16]A. Barchielli, L. Lanz, and G. M. Prosperi, "Statistics of continuous trajectories and the interpretation of quantum mechanics," in *Proceedings of the International Symposium on Foundations of Quantum Mechanics*, edited by S. Kamefuchi, H. Ezawa, Y. Murayama, M. Namiki, S. Nomura, Y. Ohnuki, and T. Yajima (Phys. Soc. Japan, Tokyo, 1984), pp. 165–179.

[17]A. Barchielli, "Continual observations in quantum mechanics and the problem of macroscopic description," to appear in *Proceedings of the Workshop on Fundamental Logical Concepts of Measurement*, Torino, September 5–6, 1983.

[18]A. Barchielli, "Continuous observations in quantum mechanics: an application to gravitational-wave detectors," Phys. Rev. D (to appear).

[19]A. S. Holevo, "On measurement of parameters of quantum processes," Teor, Mat. Fiz. **57**, 424 (1983) [Theor. Math. Phys. **7**, 1238 (1984)].

[20]A. Barchielli, "Markovian limit for a reduced operation valued stochastic process," Università di Milano preprint IFUM 301/FT, May, 1984.

[21]G. Lindblad, "Non-Markovian quantum stochastic processes and their entropy," Commun. Math. Phys. **65**, 281 (1979).

[22]L. Accardi, A. Frigerio, and J. T. Lewis, "Quantum stochastic processes," Publ. RIMS Kyoto Univ. **18**, 97 (1982).

[23]L. Accardi, "On the quantum Feynman–Kac formula," Rend. Semin. Mat. Fis. Milano **48**, 135 (1980).

[24]J. T. Lewis, "Quantum stochastic processes I," Phys. Rep. **77**, 339 (1981).

[25]R. L. Hudson, P. D. F. Ion, and K. R. Parthasarathy, "Time orthogonal unitary dilations and noncommutative Feynman–Kac formulae," Commun. Math. Phys. **83**, 261 (1982).

[26]A. Frigerio, "Covariant Markov dilations of quantum dynamical semigroups," Publ. RIMS Kyoto Univ. (to appear).

[27]R. L. Hudson and K. R. Parthasarathy, "Construction of quantum diffusions," in *Quantum Probability and Applications to the Quantum Theory of Irreversible Processes, Lecture Notes in Mathematics*, Vol. 1055, edited by L. Accardi, A. Frigerio, and V. Gorini (Springer, Berlin, 1984), pp. 171–198.

[28]R. L. Hudson and K. R. Parthasarathy, "Quantum Itô's formula and stochastic evolutions," Commun. Math. Phys. **93**, 301 (1984).

[29]R. L. Hudson and R. K. Parthasarathy, "Stochastic dilations of uniformly continuous completely positive semigroups," Acta Appl. Math. **2**, 353 (1984).

[30]I. M. Gel'fand and N. Ya. Vilenkin, *Generalized Functions*, Vol. 4, *Applications of Harmonic Analysis* (Academic, New York, 1964).

[31]M. C. Reed, "Functional analysis and probability theory," in *Lecture Notes in Physics*, Vol. 25 (Springer, Berlin, 1973), pp. 2–43.

[32]A. Guichardet, "Symmetric Hilbert spaces and related topics," in *Lecture Notes in Mathematics*, Vol. 261 (Springer, Berlin, 1972).

[33]R. Schatten, *Norm Ideals of Completely Continuous Operators* (Springer, Berlin, 1960).

Path integral quantization of the dyonium

H. Dürr[a)] and A. Inomata
Department of Physics, State University of New York at Albany, Albany, NY 12222

(Received 4 February 1985; accepted for publication 22 February 1985)

The dyonium is solved exactly by path integration. The Green's function for the dyonium is separated into the monopole harmonics and the radial path integral, and the radial Green's function is found in closed form. The exact energy spectrum is also obtained. Dirac's charge quantization condition is seen to be essential for performing path integration.

I. INTRODUCTION

For quantization of a charge-monopole system which involves the path-dependent integral of a vector potential, Feynman's path integral approach[1] is generally considered ideal, but no explicit path integral calculation has yet been made available. Apparently the computational complexity has hindered its application to the monopole problem. In this paper, we report that under Dirac's charge quantization condition[2] path integration can explicitly be carried out for the dyonium[3] which includes the charge-monopole system as a limit. The calculation is of course not at all straightforward. The various tricks recently developed[4–9] have to be effectively exploited. By presenting the path integral quantization of the dyonium, we shall achieve the following: (i) we provide a way to calculate a path integral involving the monopole potential; (ii) we separate for the first time the monopole harmonics and the radial path integral; (iii) we find the radial Green's function for the dyonium in a closed form; (iv) we derive the discrete energy spectrum of the dyonium; (v) we observe that Dirac's condition is essential for performing path integration; and (vi) we establish a unified path integral treatment of the dyonium, the hydrogen atom, and the charge-monopole system. Throughout this paper, we use units for which $\hbar = c = 1$.

II. PATH INTEGRAL FOR DYONIUM

We consider a dyonium whose Lagrangian is given by

$$L = \tfrac{1}{2}m\dot{\mathbf{r}}^2 + q\mathbf{D}\cdot\dot{\mathbf{r}} + \alpha/r, \tag{1}$$

where $\alpha = -(e_1e_2 + g_1g_2)$, $q = e_1g_2 - e_2g_1$, and

$$\mathbf{D}(\mathbf{r}) = (x\mathbf{j} - y\mathbf{i})[\pm 1 - (z/r)]/(r^2 - z^2), \tag{2}$$

with $r^2 = x^2 + y^2 + z^2$. Here, a light dyon mass m having a dual charge (e_1,g_1) is viewed as moving about a heavy dyon of (e_2,g_2) fixed at the origin $\mathbf{r} = 0$. Evidently the hydrogen atom $(e_1 = -e_2 = -e,\ g_1 = g_2 = 0)$ and the charge-monopole system $(e_1 = e,\ g_2 = g,\ e_2 = g_1 = 0)$ are special cases. The vector potential (2) has Dirac's singularity which we choose to be in either the negative or the positive z direction. The $\pm$ signs in (2) may as well correspond to the regions (a) and (b), respectively, of the Wu–Yang potential.[10]

For the dyonium (1), we intend to evaluate by path integration the Green's function

$$G(\mathbf{r}'',\mathbf{r}';E) = -i\int Q(\mathbf{r}'',\mathbf{r}';\tau)d\tau, \tag{3}$$

which is the Fourier transform of Feynman's propagator,[1] $K(\mathbf{r}'',\mathbf{r}';\tau) = \int \exp[i\int L dt]D\mathbf{r}(t)$. Naturally the integrand of (3) can be given as a path integral,

$$Q(\mathbf{r}'',\mathbf{r}';\tau) = \lim_{N\to\infty}\int \prod_{j=1}^{N}\exp[iW(\Delta t_j)] \times \prod_{j=1}^{N}\left(\frac{m}{2\pi i\Delta t_j}\right)^{3/2}\prod_{j=1}^{N-1}d\mathbf{r}_j \tag{4}$$

having an effective short time action,

$$W(\Delta t_j) = (m/2\Delta t_j)(\Delta\mathbf{r}_j)^2 + q\mathbf{D}_j\cdot\Delta\mathbf{r}_j + (E + \alpha/r_j)\Delta t_j, \tag{5}$$

where $\tau = t'' - t'$, $t'' = t_N$, $t' = t_0$, $\Delta t_j = t_j - t_{j-1}$, $\mathbf{r}_j = \mathbf{r}(t_j)$, and $\Delta\mathbf{r}_j = \mathbf{r}_j - \mathbf{r}_{j-1}$. Since the action (5) involves the monopole potential as well as the Coulomb potential, the path integral (4) cannot be calculated by standard techniques.[1] Fortunately we know the local time rescaling trick that has enabled us to treat the Coulomb problem, both nonrelativistic[4,5] and relativistic.[6] We also have a trick to handle the Aharonov–Bohm potential which depends on a single angular variable.[7]

III. SEPARATIONS OF MONOPOLE HARMONICS

In order to utilize these tricks in our problem, we first transform Cartesian variables into parabolic variables (ξ,η,ϕ): $x = \xi\eta\cos\phi$, $y = \xi\eta\sin\phi$, and $z = \tfrac{1}{2}(\xi^2 - \eta^2)$. This set of variables, containing a single angular variable, is suitable for describing the vector potential term of (5) in a form similar to that of the Aharonov–Bohm case.[7] Keeping in mind that $(\Delta q)^2$ of any generalized coordinate q is of the order of Δt in effect[5,8] and that terms of $O[(\Delta t)^2]$ can be ignored in a short time action,[1] we express (5) as

$$\begin{aligned} W_j = {} & \tfrac{1}{2}(m/\Delta t)(\bar{\xi}^2 + \bar{\eta}^2)[(\Delta\xi)^2 + (\Delta\eta)^2] \\ & + (m/\Delta t)(\hat{\xi}\hat{\eta})^2[1 - \cos(\Delta\phi)] \\ & + q\Delta\phi\,[\pm 1 - (\hat{\xi}^2 - \hat{\eta}^2)/(\bar{\xi}^2 + \bar{\eta}^2)] \\ & + 2\alpha\Delta t/(\bar{\xi}^2 + \bar{\eta}^2) + E\Delta t, \end{aligned} \tag{6}$$

where $\bar{\xi} = \tfrac{1}{2}(\xi_j + \xi_{j-1})$, $\hat{\xi} = (\xi_j\xi_{j-1})^{1/2}$, $\Delta\xi = \xi_j - \xi_{j-1}$, etc. For convenience, we have suppressed in (6) the subscript j which identifies the jth interval. Hereafter, we shall also employ the same convention as far as we can. By doing this, however, we do not mean that Δt_j are the same for all j.

Next, we change the short time interval Δt_j into a new interval Δs_j by $\Delta t = 2(\bar{\xi}^2 + \bar{\eta}^2)\Delta s$ as before,[5,6] and put (6) into two parts, $W_j = -i(A_j + B_j)$, where

[a)] Present address: Physikalisches Institut der Universität Würzburg, Würzburg, West Germany.

 0022-2488/85/092231-03$02.50

$$A_j = (im/4\Delta s)[(\Delta\xi)^2 + (\Delta\eta)^2 + (\hat{\xi}\hat{\eta}/\bar{\rho})^2] + 4i\Delta s(\alpha + E\bar{\rho}^2), \tag{7}$$

$$B_j = (m/4i\Delta s)(\hat{\xi}\hat{\eta}/\bar{\rho})^2[\cos(\Delta\phi) + \delta_\pm \Delta\phi], \tag{8}$$

with $\bar{\rho}^2 = \frac{1}{2}(\bar{\xi} + \bar{\eta}^2)$ and $\delta_\pm = (2q\Delta s/m)[\hat{\xi}^2 - \hat{\eta}^2 \mp (\bar{\xi}^2 + \bar{\eta}^2)]/(\hat{\xi}\hat{\eta})^2$. As Δs is kept small, so are $\delta_\pm$. Hence we can use the relation, $\cos(\Delta\phi) + \delta\Delta\phi = \cos(\Delta\phi - \delta) + \frac{1}{2}\delta^2$, to change (8) into

$$B_j = \frac{m}{4i\Delta s}\left(\frac{\hat{\xi}\hat{\eta}}{\bar{\rho}}\right)^2 \cos(\Delta\phi - \delta_\pm) - \frac{(2iq^2\Delta s/m)(\hat{\eta}/\hat{\xi})^{\pm 2}}{\bar{\rho}^2}. \tag{9}$$

Then, we use the asymptotic formula for large $|z|$, $\exp(x\cos\theta) = \Sigma \exp(i\nu\theta)\widetilde{I}_\nu(z)$, where $\widetilde{I}_\nu(z)$ is the asymptotic form of the modified Bessel function $I_\nu(z)$, that is,[9]

$$\widetilde{I}_\nu(z) = (2\pi z)^{-1/2}\exp[z - (\nu^2 - \tfrac{1}{4})/(2z)]. \tag{10}$$

After rearranging terms, we get

$$e^{iW_j} = (2\pi m\bar{\rho}^2/i\Delta s)^{1/2}\exp[(im/2\Delta s)(\overline{\xi^2} + \overline{\eta^2}) + 4i\Delta s(\alpha + E\bar{\rho}^2)]\exp(2iq^2\Delta s/m\bar{\rho}^2)f_\pm(\xi,\eta,\phi), \tag{11}$$

where $\overline{\xi^2} = \frac{1}{2}(\xi_j^2 + \xi_{j-1}^2)$, etc., and

$$f_\pm(\xi,\eta,\phi) = \sum_{\nu=-\infty}^{\infty} e^{i\nu\Delta\phi}\widetilde{I}_{\nu-q\mp q}\left(\frac{m\hat{\xi}^2}{2i\Delta s}\right)\widetilde{I}_{\nu+q\mp q}\left(\frac{m\hat{\eta}^2}{2i\Delta s}\right). \tag{12}$$

Use of Graf's addition formula[11] applied to the modified Bessel functions in the asymptotic form reduces (12) into the form

$$f_\pm(\xi,\eta,\phi) = e^{\pm iq\Delta\phi}(\zeta_+/\zeta_-)^{2q}\widetilde{I}_{\mp 2q}(m\zeta_+\zeta_-/2i\Delta s), \tag{13}$$

where $\zeta_\pm^2 = \hat{\xi}^2\exp(\mp i\Delta\phi) + \hat{\eta}^2\exp(\pm\frac{1}{2}i\Delta\phi)$. At this point, we must remark that Graf's formula is applicable to (12) provided that $\xi^2 \gtrless \eta^2$ for $f_\pm$, respectively.[11] Such constraints jeopardize the chance of completing path integration. However, (13) is constraint-free in the exceptional case when $2q$ = integer,[11] i.e., when Dirac's quantization condition is satisfied. Thus, in order to proceed with our calculation, we have to adopt Dirac's condition.

Now we turn parabolic variables to polar variables (r,θ,ϕ) with $r = \rho^2$ by $\xi = \sqrt{2}\rho\cos(\frac{1}{2}\theta)$, $\eta = \sqrt{2}\rho\sin(\frac{1}{2}\theta)$, and $\phi = \phi$. As a result, (11) becomes

$$e^{iW_j} = (2\pi m\bar{\rho}^2/i\Delta s)^{1/2}\exp(4i\alpha\Delta s)\exp[(im/\Delta s)\bar{\rho}^2 + 4iq^2\Delta s/(2m\bar{\rho}^2) + 4iE\bar{\rho}^2\Delta s]f_\pm(\rho,\theta,\phi), \tag{14}$$

with

$$f_\pm(\rho,\theta,\phi) = e^{\pm iq\Delta\phi}(\zeta_+/\zeta_-)^{2q} \times \widetilde{I}_{2q}[(m/i\Delta s)\hat{\rho}^2\cos(\tfrac{1}{2}\Theta)], \tag{15}$$

where $\cos\Theta = \cos\theta_j\cos\theta_{j-1} + \sin\theta_j\sin\theta_{j-1}\cos(\Delta\phi)$. Note that $I_{2q}(z) = I_{-2q}(z)$ for $2q$ = integer. The radial and angular variables in (15) can be separated with the aid of the expansion formula[12] for large $|z|$,

$$\widetilde{I}_{2q}\left[iz\cos\left(\frac{1}{2}\Theta\right)\right] = \frac{2}{iz}\sum_{l=|q|}^{\infty}(2l+1)d_q{}^l{}_q(\Theta)\widetilde{I}_{2l+1}(iz), \tag{16}$$

where $d_q{}^l{}_{q'}(\Theta)$ is the Wigner function having the property[13]

$$\left(\frac{\zeta_+}{\zeta_-}\right)^{2q} d_q{}^l{}_q(\Theta) = \sum_{\mu=-l}^{l} e^{i\mu\Delta\phi}d_\mu{}^l{}_{-q}(\theta_j)d_\mu{}^l{}_{-q}(\theta_{j-1}). \tag{17}$$

Namely, we have

$$f_\pm(\rho_j,\theta_j,\phi_j) = \frac{2i\Delta s_j}{m\hat{\rho}_j^2}\sum_{l=|q|}^{\infty}\sum_{\mu=-l}^{l}(2l+1)\widetilde{I}_{2l+1}\left(\frac{m\hat{\rho}_j^2}{i\Delta s}\right) \times d_\mu{}^l{}_{-q}(\theta_j)d_\mu{}^l{}_{-q}(\theta_{j-1})e^{i(\mu\pm q)\Delta\phi_j}. \tag{18}$$

Substitution of (18) into (14) yields

$$e^{iW_j} = e^{4i\alpha\Delta s_j}\frac{2i\Delta s_j}{m\hat{\rho}_j^2}\sum_l\sum_\mu(2l+1)e^{iS_j} \times d_\mu{}^l{}_{-q}(\theta_j)d_\mu{}^l{}_{-q}(\theta_{j-1})e^{i(\mu\pm q)\Delta\phi_j}, \tag{19}$$

where

$$S_j = (m/2\Delta s)(\Delta\rho)^2 - \lambda(\lambda+1)\Delta s/(2m\hat{\rho}^2) - \tfrac{1}{2}m\omega^2\hat{\rho}^2\Delta s, \tag{20}$$

with $\lambda = [(2l+1)^2 - 4q^2]^{1/2} - \frac{1}{2}$ and $\omega^2 = -8E/m$.

Let us now perform the integration of (4) for (19) on the polar coordinate basis. With $r = \rho^2$, we have $d\mathbf{r} = 2\rho^5\,d\rho\,\sin\theta\,d\theta\,d\phi$. The angular integration can easily be done by using the orthogonality relations[13]

$$\int\!\!\int\prod_{j=1}^{N}\frac{l_j+\frac{1}{2}}{2\pi}d_\mu{}^l{}_{-q}(\theta_j)d_\mu{}^l{}_{-q}(\theta_{j-1}) \times e^{i(\mu_j\pm q)\Delta\phi_j}\prod_{j=1}^{N-1}\sin\theta_j\,d\theta_j\,d\phi_j$$
$$= [(l''+\tfrac{1}{2})/(2\pi)]d_{\mu''}{}^{l''}{}_{-q}(\theta'')d_{\mu'}{}^{l'}{}_{-q}(\theta') \times e^{i(\mu'\pm q)(\phi''-\phi')}\prod_{j=1}^{N-1}\delta(\mu_{j+1},\mu_j)\delta(l_{j+1},l_j). \tag{21}$$

Consequently, we achieve the separation of three variables,

$$Q(\mathbf{r}'',\mathbf{r}';\tau) = \sum_{l=|q|}^{\infty}Q_l(r'',r';\tau)\sum_{\mu=-l}^{l}\frac{2l+1}{4\pi} \times d_\mu{}^l{}_{-q}(\theta')d_\mu{}^l{}_{-q}(\theta'')e^{i(\mu\pm q)(\phi''-\phi')}. \tag{22}$$

Hence, from (3), we obtain

$$G^{(\pm)}(\mathbf{r}'',\mathbf{r}';E) = \sum_{l=|q|}^{\infty}G_l(r'',r';E)\sum_{\mu=-l}^{l}Y_{ql\mu}^{(\pm)*}(\theta',\phi') \times Y_{ql\mu}^{(\pm)}(\theta'',\phi''), \tag{23}$$

where $G_l(r'',r';E) = -i\int Q_l(r'',r';\tau)d\tau$ is the radial Green's function and $Y^{(\pm)}(\theta,\phi) = [(2l+l)/4\pi]^{1/2}d(\theta)\exp[i(\mu \pm q)\phi]$ are the monopole harmonics.[14] The $\pm$ signs in (23) correspond to the regions (a) and (b) of the Wu–Yang potential.[10] Apparently, $G^{(+)} = G^{(-)}\exp[2iq(\phi''-\phi')]$.

IV. RADIAL PATH INTEGRATION

In (23), we have still to carry out the radial path integration in (22),

$$Q_l(r'',r';\tau) = \tfrac{1}{2}(\rho'\rho'')^{-3/2}e^{4i\alpha\sigma}K_\lambda(\rho'',\rho';\sigma), \tag{24}$$

where $r' = \rho'^2$, $r'' = \rho''^2$, $\sigma = \tau/(4\rho'\rho'')$, and

$$K_\lambda(\rho'',\rho';\sigma) = (\rho'\rho'')^{-1}\lim_{N\to\infty}\int_0^\infty\prod_{j=1}^{N}\exp[iS_j] \times\prod_{j=1}^{N}\left(\frac{m}{2\pi i\Delta s_j}\right)^{1/2}\prod_{j=1}^{N-1}d\rho_j. \tag{25}$$

This path integral is identical in form with that for the radial propagator of the three-dimensional harmonic oscillator, which has been evaluated exactly.[9] The radial Green's function in (23) has also been integrated for the hydrogen atom having l in the place of $\gamma = \frac{1}{2}(\lambda - \frac{1}{2})$. Thus, exploiting the result for the hydrogen atom[5,6] and replacing l by γ, we obtain the radial Green's function of the dyonium in closed form, expressed in terms of the Whittaker functions,

$$G_l(r'',r';E) = (2ikr'r'')^{-1}[\Gamma(p+\gamma+1)/\Gamma(2\gamma+2)] \times M_{-p,\gamma+1/2}(-2ikr')W_{-p,\gamma+1/2}(-2ikr''), \tag{26}$$

where $r'' > r'$, $k = (2mE)^{1/2}$, $p = -i(m\alpha^2/2E)^{1/2}$, and $\gamma = [(l+\frac{1}{2})^{1/2} - q^2]^{1/2} - \frac{1}{2}$. The poles of $\Gamma(p+\gamma+1)$ occur when $p+\gamma+1 = -n_r$ ($n_r = 0, 1, 2, \ldots$), giving rise to the discrete energy spectrum, $E_n = -m\alpha^2/(2n^2)$, where $n = n_r + \frac{1}{2} + [(l+\frac{1}{2})^2 - q^2]^{1/2}$ and $2q = 0, \pm 1, \pm 2, \ldots, \pm 2l$. This bound state energy formula is in agreement with that of Barut and Bornzin obtained by the algebraic method.[15] Obviously, in the limit $\alpha = e^2$ and $q = 0$, the radial Green's function and the energy spectrum coincide with those of the hydrogen atom.[5,6] In the algebraic approach, the pure dyonium case[15] ($\alpha \neq 0$) and the charge-monopole case[16] ($\alpha = 0$) have been treated separately, whereas the present path integral scheme provides a framework in which the dyonium, the charge-monopole system, and the monopolium as well as the hydrogenlike atom are treated in a unified manner. The techniques used for the separation of the monopole harmonics can also be applied to such problems as the Pöschl–Teller and the Rosen–Morse oscillators.[17]

ACKNOWLEDGMENTS

One of the authors (H.D.) wishes to thank Professor W. Gibson for his hospitality at SUNY—Albany and also gratefully acknowledges the support of the Deutscher Akademischer Austauschdienst.

[1] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw–Hill, New York, 1965).
[2] P. A. M. Dirac, Proc. R. Soc. London Ser. A **113**, 60 (1931).
[3] J. Schwinger, Science **165**, 757 (1969).
[4] I. H. Duru and H. Kleinert, Phys. Lett B **84**, 184 (1979); R. Ho and A. Inomata, Phys. Rev. Lett. **48**, 231 (1982).
[5] A. Inomata, Phys. Lett. A **101**, 253 (1984).
[6] M. A. Kayed and A. Inomata, Phys. Rev. Lett. **53**, 107 (1984).
[7] A. Inomata and V. A. Singh, J. Math. Phys. **19**, 2318 (1978); C. Bernido and A. Inomata, *ibid.* **22**, 715 (1981).
[8] D. W. McLaughlin and L. S. Schulman, J. Math. Phys. **12**, 2520 (1971).
[9] D. Peak and A. Inomata, J. Math. Phys. **10**, 1422 (1969); W. Langguth and A. Inomata, *ibid.* **20**, 499 (1979).
[10] T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975).
[11] G. N. Watson, *Theory of Bessel Functions* (Cambridge U.P., London, 1966), 2nd ed., p. 359.
[12] Reference 11, p. 140.
[13] J. D. Talman, *Special Functions* (Benjamin, New York, 1968), p. 160.
[14] T. T. Wu and C. N. Yang, Nucl. Phys. B **107**, 365 (1976); L. C. Biedenharn and J. D. Louck, *Racah-Wigner Algebra in Quantum Theory* (Addison–Wesley, Reading, MA, 1981), p. 218.
[15] A. O. Barut and G. L. Bornzin, J. Math. Phys. **12**, 841 (1971).
[16] R. Jackiw, Ann. Phys. (NY) **129**, 183 (1980).
[17] A. Inomata and M. A. Kayed, Phys. Lett. A **108**, 9 (1985); A. Inomata and M. A. Kayed, J. Phys. A: Math. Gen. **18**, L235 (1985)

Clifford- and Grassmann-like algebras—Old and new

A. K. Kwasniewski
Institute of Theoretical Physics, University of Wrocław, Cybulskiego 36, 50-205 Wrocław, Poland

(Received 15 March 1984; accepted for publication 15 February 1985)

Algebra extensions of $\Gamma = Z_k \oplus \cdots \oplus Z_k$ (n summands) are considered as old Clifford-like algebras. Grassmann-like algebras closely related to them are introduced. New Clifford-like and another Grassmann-like algebras are defined and discussed, the generalization consisting in considering k-linear structures instead of only bilinear ones. Several applications are listed.

I. INTRODUCTION

Generalizations of Clifford algebras were introduced quite independently by the authors of Refs. 1–4. In the present paper we derive, in a canonical way, an ultimate generalization of Clifford algebras so that the precedent $\mathscr{C}_n^{(k)}$ generalized Clifford algebras serve as an epimorphic image of the ones introduced by the author. These new $k - \mathscr{C}_n$ Clifford-like algebras are naturally of primary importance for those algebraic problems of physics in which universality of the arising algebra is crucial.

In parallel, we introduce the corresponding Grassmann-like algebras, which are expected to play a similar role with respect to Clifford-like algebras as the Grassmann ones with respect to the usual Clifford algebras.

First applications of generalized Clifford algebras date from the late 1960's,[5,7] then followed by other ones in another branch of physics.[8,9] The $\mathscr{C}_n^{(k)}$ algebras were also shown to play a decisive role in solving problems of general involutional transformations.[10]

Quite recently a remarkable application of these algebras was found while constructing and classifying the so-called ϵ-Lie Γ-graded algebras,[11,12] hence equivalently an application to modular quantization (see Ref. 13 and references therein).

It is already well known that algebra extensions (including $\mathscr{C}_n^{(k)}$ algebras) of finite groups as well as generalized Dirac groups form an excellent tool for deriving and classifying projective representations of finite groups; a subject of primary importance for the quantum theory of crystals. For that application see the review (Ref. 14) and Refs. 15 and 16.

Finally, let us mention some other applications. Namely, the generalized Pauli algebra and the corresponding Dirac groups were used in Ref. 17 to solve an inverse problem for harmonic vibrations of cyclic molecules not restricted to the closest neighbor approximation. It is also to be noted that generalized Clifford and Grassmann algebras are of potential importance for generalizations of the Ising model as the Onsager formula for the partition function can be transparently derived from the Clifford algebra algebraic properties only in the case of the Ising model.[18]

We close this incomplete list of existing or possible applications by noting that for reasons already implied by Ref. 1, the ultrageneralized complex analysis of Ref. 19 should make use of k-$\mathscr{C}_k$ algebras as these are the most general objects linearizing the equation that defines ultraanalyticity.

Out paper is organized as follows. In Sec. II we review generalized Clifford algebras[2] in the language of group algebras and algebra extensions,[4] and then (in Sec. III) in the language of commutative δ-Lie Γ-graded algebras.[11,12] The latter enables us to introduce a class of new Grassmann-like algebras G_ρ in Sec. III. An intermediate section (Sec. IV) serves to introduce some sort of algebras important for subsequent use in Sec. V. Finally, in Sec. V we introduce a new generalization of the usual $\mathscr{C}_n^{(2)}$ Clifford algebras and corresponding Grassmann-like ones. A class of their matrix representation is also given there.

The generalization of Sec. V consists in considering k-linear structures instead of only bilinear ones, and represents the main goal of the paper.

II. PRELIMINARIES

Let $D_{n,k}$ be the group generated by its $\omega, \gamma_1,\ldots,\gamma_n$ elements satisfying

$$\omega\gamma_i = \gamma_i\omega, \quad \gamma_i^k = \omega^k = 1, \quad \gamma_i\gamma_j = \omega\gamma_j\gamma_i,$$
$$i<j, \quad i,j = 1,\ldots,n.$$

$D_{n,k}$ is called the generalized Dirac group in the following. This is the meta-Abelian group of order k^{n+1} and $D_{n,k}$ is a special central group extension of Z_k by $\Gamma \equiv Z_k \oplus \cdots \oplus Z_k$ (n summands) because of an exact sequence

$$1 \to Z_k \to D_{n,k} \to \Gamma \to 1,$$

Z_k being a subgroup of the center of $D_{n;k}$ (we make no distinction between isomorphic structures). The matrix form of $\gamma_1,\ldots,\gamma_n \in D_{n;k}$ may be found in Ref. 2, while ω is the primitive kth root of unity. Here, $D_{n;2}$ is the well-known group of Dirac γ matrices for Euclidean spaces. Generalization of the "pseudo-Euclidean" Dirac γ matrices is also known.[2] The group $D_{n;k}$ belongs to the family of $k^{n(n+1)/2}$ inequivalent central group extensions of Z_k by Γ as they are in one-to-one correspondence[4(a)] with the elements of second cohomology groups $H^{(2)}(\Gamma, Z_k) \simeq Z_k \oplus \cdots \oplus Z_k$ [$n(n+1)/2$ summands].

The group algebra $\mathbb{C}[D_{n;2}]$ of $D_{n;2}$ over $\mathbb{C}$ contains the familiar Clifford algebra $\mathscr{C}_n^{(2)}$. Of course $\mathbb{C}[D_{n;k}]$ is semisimple and $\mathscr{C}_n^{(k)}$, the generalized Clifford algebra generated over $\mathbb{C}$ by $\gamma_1,\ldots,\gamma_n$ (See Ref. 2), is a two-sided ideal of $\mathbb{C}[D_{n;k}]$. Hence $\mathscr{C}_n^{(k)}$ is an irreducible (reducible) representation for $n = 2v$ ($n = 2v + 1$) of the group algebra $\mathbb{C}[D_{n,k}]$. While the generalized Dirac group $D_{n;k}$ is the special central group extension $1 \to Z_k \to G \to \Gamma \to 1$, the generalized Clifford algebra is a special case of algebra extension of Γ over $\mathbb{C}$ where[4(b)] the following definition holds.

Definition: An algebra $\mathscr{C}$ is an algebra extension of Γ

over $\mathbb{C}$ iff

(1) $\mathscr{C} = \oplus_{\alpha\in\Gamma} \mathscr{C}_\alpha$ (is Γ graded),

(2) $\dim \mathscr{C}_\alpha = 1,\quad \mathscr{C}_\alpha\mathscr{C}_\beta = \mathscr{C}_{\alpha+\beta},\quad \alpha,\beta\in\Gamma.$

It is known[4(a)] that there exists a bijective correspondence between isomorphic classes of algebra extensions of Γ over $\mathbb{C}$ and cohomology classes of $H^{(2)}(\Gamma,\mathbb{C})\simeq P_{as}(\Gamma,\mathbb{C}^*)$, where $P_{as}(\Gamma,\mathbb{C}^*)$ denotes the group of all antisymmetric pairings, i.e., mappings $\delta:\Gamma\times\Gamma\to\mathbb{C}^*$ which are (1) bimorphisms and (2) $\delta(\alpha,\alpha)=1, \alpha\in\Gamma$. As $H^2(\Gamma,\mathbb{C}^*)\simeq Z_k\oplus\cdots\oplus Z_k$ [$n(n-1)/2$ summands] we have $k^{n(n-1)/2}$ different algebra extensions $\mathscr{C}_\delta$ [$\delta\in P_{as}(\Gamma,\mathbb{C}^*)$] of Γ over $\mathbb{C}$. We can think of $\mathscr{C}_\delta$ as the algebra generated by generators $\gamma_1,\dots,\gamma_n$ satisfying

$$\gamma_i\gamma_j=\omega_{ij}\gamma_j\gamma_i,\quad \gamma_i^k=1,\quad i,j=1,\dots,n,$$

where $\omega_{ij}=\delta(s_i,s_j)$, while $\{s_i\}_{i=1}^n$ are generators of Γ and $\delta\in P_{as}(\Gamma,\mathbb{C})$. Because δ is an antisymmetric pairing, $\omega_{ij}=\omega^{\alpha_{ij}}$, where $\alpha_{ij}\in Z_k$, ω is the primitive kth root of unity and this $(\alpha_{ij})=(n\times n)$ matrix is antisymmetric in the sense of a Z_k ring. Note that the additive group of these (α_{ij}) matrices is isomorphic to $H^2(\Gamma,\mathbb{C}^*)$.

For any $\delta\in P_{as}(\Gamma,\mathbb{C}^*)$, we then have

$$\delta(\alpha,\beta)=\omega^{\langle\alpha|A\beta\rangle},\quad \alpha,\beta\in\Gamma,$$

where $A=(\alpha_{ij})$ matrix and $\langle\alpha|\beta\rangle=\Sigma_{i=1}^n\alpha_i\beta_i$, $\alpha_i,\beta_i\in Z_k$. The special choice of A, namely $\alpha_{ij}=1$ for $i<j$ gives $\mathscr{C}_n^{(k)}$ generalized Clifford algebras.

Though these Clifford-like algebras $\mathscr{C}_\delta$ are different as algebra extensions some of them are isomorphic to each other as algebras. Which? The notion of annihilator $N_\delta\subset\Gamma$ is crucial to answer that question. The subgroup N_δ, is defined as the set $N_\delta=\{\alpha\in\Gamma;A\alpha=0\}$, i.e., the "kernel" of the A matrix. It is trivial to note that if $N_\delta=\Gamma$ then $\mathscr{C}_\delta=\mathbb{C}\oplus\cdots\oplus\mathbb{C}$ (k^n summands). On the other hand, for $N_\delta=\{0\}$ (this holds iff $\det A$ is comprime with k) the algebra $\mathscr{C}_\delta$ is simple; therefore $\mathscr{C}_\delta\simeq M_{k^v}(\mathbb{C})$; $2v=n$ and $M_d(\mathbb{C})$ denotes the matrix algebra of all $(d\times d)$ complex matrices. In general, any algebra extension $\mathscr{C}_\delta$ of Γ over $\mathbb{C}$ is of the form[4(b),(c)]

$$\mathscr{C}_\delta\simeq M_d\oplus\cdots\oplus M_d\quad(\mu\text{ summands}),$$

where $k^n=\mu d^2$ and μ = order of N_δ. This condition follows from the fact that $\mathscr{C}_\delta$ exists iff Γ/N_δ is of symmetric type. Of course for any $\delta\in P_{as}(\Gamma,\mathbb{C}^*)$: $\dim\mathscr{C}_\delta=k^n$ and the center $Z(C_\delta)$ of the Clifford-like algebra C_δ is determined by N_δ. Namely,

$$Z(C_\delta)=\oplus_{x\in N_\delta}\mathbb{C}_x,\quad \mathbb{C}_x\simeq\mathbb{C}.$$

Finally, we give an example.

Example 1: Consider for illustration the case $\Gamma=Z_3\oplus Z_3\oplus Z_3$. We then have 27 different algebra extensions. However, we obtain that way only two nonisomorphic algebras; $\mathscr{C}_1\simeq\mathbb{C}\oplus\cdots\oplus\mathbb{C}$ (27 summands) and $\mathscr{C}_\delta\simeq\mathscr{C}_3^{(3)}\simeq M_3\oplus M_3\oplus M_3$ for $\delta\neq1$. It is easy to see that for all 28 δ's different from 1, $N_\delta\simeq Z_3$. ∎

These $\mathscr{C}_\delta$ Clifford-like algebras were applied to construct and classify[11,12] ϵ-Lie Γ-graded algebras of potential importance for physics, as defined below.

III. GRASSMANN-LIKE ALGEBRAS G_ρ

$\mathscr{C}_\delta$ Clifford-like algebras provide an example of δ-Lie Γ-graded commutative algebras[11,12] and this point of view enables one to introduce Grassmann-like algebras along the same lines. For completeness we start with the necessary definitions.

Definition: $\epsilon:\Gamma\times\Gamma\to\mathbb{C}^*$ is said to be a commutation factor iff (1) ϵ is a biomorphism, and (2) $\epsilon(\alpha,\beta)\epsilon(\beta,\alpha)=1$, $\alpha,\beta\in\Gamma$. ∎

In the case of Γ admitting a Γ_0 subgroup of index 2 one comes up with the following definition.

Definition: Let $\Gamma=\Gamma_0\cup\Gamma_1$, ϵ_0 is said to be the Grassmann commutation factor iff

$$\epsilon_0:\Gamma\times\Gamma\to\mathbb{C}^*,\quad \epsilon_0(\alpha,\beta)=\begin{cases}-1, & \alpha,\beta\in\Gamma_1,\\ 1, & \text{otherwise.}\end{cases}$$

∎

The group of commutation factors is given by either $P_{as}(\Gamma,\mathbb{C}^*)\cup\epsilon_0P_{as}(\Gamma,\mathbb{C}^*)$ or $P_{as}(\Gamma,\mathbb{C}^*)$ depending on whether Γ admits the Γ_0 subgroup of index 2 or not.[11,12]

The ϵ-Lie Γ-graded algebra is then defined as follows.

Definition: Let L be a Γ-graded vector space equipped with bilinear mapping $\langle\,,\,\rangle:L\times L\to L$. Let $x_\alpha, y_\beta, z_0\in L$, $\alpha,\beta,\gamma\in\Gamma$, denote homogeneous elements. We then say L is a ϵ-Lie Γ-graded algebra iff it is a Γ-graded algebra under $\langle\,,\,\rangle$ multiplication and

(1) $\alpha,\beta\in\Gamma,\quad \langle x_\alpha,y_\beta\rangle=-\epsilon(\alpha,\beta)\langle y_\beta,x_\alpha\rangle$ (ϵ skew symmetric),

(2) $\alpha,\beta,\gamma\in\Gamma,\quad \langle x_\alpha,\langle y_\beta,z_\gamma\rangle\rangle=\langle\langle x_\alpha,y_\beta\rangle,z_\gamma\rangle+\epsilon(\alpha,\beta)\langle y_\beta,\langle x_\alpha,z_\gamma\rangle\rangle$ (ϵ Jacobi identity),

where ϵ is a commutation factor. ∎

Still one more definition is necessary.

Definition: Let U be an associative Γ graded algebra. Then ass U is said to be an ϵ-Lie Γ-graded algebra *associated to* U iff (1) ass $U=U$ as Γ-graded vector spaces and (2) $\langle\,,\,\rangle$: ass $U\times$ ass $U\to$ ass U is defined via

$$\langle x_\alpha,y_\beta\rangle=x_\alpha y_\beta-\epsilon(\alpha,\beta)y_\beta x_\alpha.$$

∎

One sees now that ass $\mathscr{C}_\delta$ is a commutative δ-Lie Γ-graded algebra, hence $\mathscr{C}_\delta$ is an epimorphic image of the universal enveloping algebra of a commutative δ-Lie Γ graded algebra (see Ref. 20, p. 26). This is seen in the following way.

Let V be a maximally Γ-graded vector space, i.e., $V=\oplus_{\gamma\in\Gamma}V_\gamma$, div $V_\gamma=1$. Let $S_\delta=T/I_\delta$ be the δ symmetric algebra of V, where T is the tensor algebra of V while I_δ is an ideal of T generated by the elements

$$x_\alpha\otimes y_\beta-y_\beta\otimes x_\alpha\delta(\alpha,\beta),\quad \alpha,\beta\in\Gamma.$$

Of course the vector space V can be considered as a commutative δ-Lie Γ-graded algebra, where, by definition, $\langle x_\alpha,y_\beta\rangle=0$, $\alpha,\beta\in\Gamma$. Then S_δ is the universal enveloping algebra of this δ-commutative δ-Lie Γ-graded algebra V, and S_δ may be identified with the algebra of all polynomials in *the basis elements* $\{x_\alpha\}_{\alpha\in\Gamma}$ of V, which satisfy relations

$$x_\alpha x_\beta=\delta(\alpha,\beta)x_\beta x_\alpha.$$

Of course, $(x_\alpha)^k \in Z[S_\delta]$ as $\delta(0,\beta) = \delta(\beta,0) = 1$, $\beta \in \Gamma$. Also, S_δ is naturally Γ-graded with the grading by that of V, i.e.,

$$x_{\alpha_1} \cdots x_{\alpha_r} \in (S_\delta)_\gamma$$

$$\text{iff} \sum_{i=1}^{r} \alpha_i = \gamma, \quad \alpha_i, \gamma \in \Gamma, \quad S_\delta = \bigoplus_{\gamma \in \Gamma} (S_\delta)_\gamma.$$

The center $Z[S_\delta]$ is not trivial as it contains the subalgebra $W[\{Z_\alpha\}]$ of all polynomials in variables $\{Z_\alpha\}_{\alpha\in\Gamma}$ $Z_\alpha \equiv (x_\alpha)^k$, independently of the δ chosen.

Because of the commutative diagram

$$\begin{tikzcd} \text{ass } S_\delta \arrow[rr, "\pi_0"] \arrow[dr, "\pi"'] & & S_\delta \arrow[dl, "\tau"] \\ & \mathscr{C}_\delta & \end{tikzcd},$$

where π_0 is the trivial isomorphism ass $S_\delta \to$ ass S_δ, π is the epimorphism ass $S_\delta \to$ ass$\mathscr{C}_\delta$, and τ is the epimorphism of the Γ-graded associative algebras, $\mathscr{C}_\delta$ is a τ epimorphic image of S_δ, epimorphism τ being that sending $(x_\alpha)^k$ $\alpha \in \Gamma$ into $1 \in \mathscr{C}_\delta$.

With the Γ-graded vector space V given, this commutative diagram provides a definition of $\mathscr{C}_\delta$ with τ being that above, i.e.,

$\mathscr{C}_\delta \simeq S_\delta / \ker \tau$. ■

This seemingly affected presentation of $\mathscr{C}_\delta$ Clifford-like algebras opens the way to introduce Grassmann-like algebras $\mathscr{G}_\rho$ along the same lines. Namely, consider now $S_\rho = T/I_\rho$ with $\rho = \epsilon_0 \delta$; $\delta \in P_{\text{as}}(\Gamma, \mathbb{C}^*)$, where Γ admits a T_0 subgroup of index 2. Again we have the commutative diagram

$$\begin{tikzcd} \text{ass } S_\rho \arrow[rr, "\pi_0"] \arrow[dr, "\pi"'] & & S_\rho \arrow[dl, "\rho"] \\ & \mathscr{G}_\rho & \end{tikzcd},$$

where this time the epimorphism ρ of Γ graded algebras (S_δ onto $\mathscr{G}_\rho$) sends all $(x_\alpha)^k$, $\alpha \in \Gamma$ into zero. Hence

$\mathscr{G}_\rho \simeq S_\rho / \ker \rho$. ■

We shall consider now some examples of these new Grassmann-like algebras.

Example 2: Consider $\Gamma = Z_2 \oplus Z_2 \oplus Z_2$, $\Gamma = \Gamma_0 \cup \Gamma_1$, where $\Gamma_0 = \{(1,1,0), (1,0,1), (0,1,1), (0,0,0)\}$ and $\Gamma_1 = \{(1,0,0), (0,1,0), (0,0,1), (1,1,1)\}$. Define $\Theta_i = x_{s_i}$, $i = 1,2,3$ with s_i generators of Γ, $\Theta_i \in \mathscr{G}_\rho$.

(1) Take $\delta \equiv 1$, then $\{\Theta_i, \Theta_j\} = 0$ $i,j = 1,2,3$ and $\mathscr{G}_{\epsilon_0}$ is the usual Grassmann algebra generated by 1 and $\{\Theta_i\}_{i=1}^3$.

(2) Take $\delta \neq 1$, $\delta \leftrightarrow A = (a_{ij})$, $a_{ij} \in Z_2$, $A \neq 0$. Then $(\rho = \delta\epsilon_0) \mathscr{G}_\rho$ is the algebra generated by 1 and $\{\Theta_i\}_{i=1}^3$, which satisfy this time

$$\Theta_i \Theta_j + (-1)^{\alpha_{ij}} \Theta_j \Theta_i = 0, \quad i \neq j, \quad \Theta_i^2 = 0, \quad i,j = 1,2,3.$$

It is now clear how to construct $\mathscr{G}_\rho$ Grassmann-like algebras for any Γ admitting a Γ_0 subgroup of index 2. For a given Γ some Grassmann-like $\mathscr{G}_\rho$ algebras are isomorphic as algebras, the problem being similar to the one considered for $\mathscr{C}_\delta$ algebras.

Note: For any Γ of even order there exists a subgroup Γ_0 of index. 2.

Note: Any usual Grassmann algebra can be obtained as in Example 2 for $\Gamma = Z_2 \oplus \cdots \oplus Z_2$ (in summands) with Γ_0 and Γ_1 chosen according to the same rule as

$$\sum_{i=0}^{[n/2]} \binom{n}{2i+1} = 2^{n-1} = \text{order of } \Gamma_0 = \tfrac{1}{2} \text{ order of } \Gamma.$$

■

To end these considerations we find it interesting to give one more simple example.

Example 3: Let $\Gamma = Z_{2l}$; $\Gamma_0 = \{2i; i = 0,\ldots,l-1\}$, $\Gamma_l = \{2i+1; i = 0,\ldots,l-1\}$. Since $H^2(Z_k, \mathbb{C}^*) \simeq \{0\}$, we have only one δ and $\delta \equiv 1$. Therefore, we end up with only one Grassmann-like algebra $\mathscr{G}_{\epsilon_0}(l)$. Note that the ϵ_0 commutation factor now can be written as

$$\epsilon_0(\alpha,\beta) = (-1)^{\alpha\beta}, \quad \alpha,\beta \in Z_{2l}.$$

$\mathscr{G}_{\epsilon_0}(l)$ is the algebra generated by 1 and Θ, $\Theta^l = 0$. However, note that $\mathscr{G}_{\epsilon_0}(1) = \mathscr{G}_{\epsilon_0}(2)$ for $l > 2$ because for $l > 2$ $\Theta^4 = 0$. ■

IV. THE ALGEBRAS $\mathscr{C}^{(k)}_{\{\beta_1,\ldots,\beta_n\}}$

We consider now, for completeness, the dimodule algebra construction of $\mathscr{C}^{(k)}_n$ algebras.[21] The $\mathscr{C}^{(k)}_{\{\beta_1,\ldots,\beta_n\}}$ algebra, defined below, is important for the forthcoming application. Consider the generalized Pauli algebra $\mathscr{C}^{(k)}_2$. This is a special case of (central simple) generalized quaternion algebra $A_\omega(a,b)$, $a,b \in \mathbb{C}^*$ (discussed in Ref. 22, Sec. 15), i.e., $\mathscr{C}^{(k)}_2 = A_\omega(1,1)$.

In Ref. 21, it was observed that $A_\omega(a,b) = \mathscr{C}^{(k)}_a \# \mathscr{C}^{(k)}_{(b)}$, where $\mathscr{C}^{(k)}_{(a)}$ is the Z_k dimodule algebra generated by λ subjected to the relation $\lambda^k = a\mathbf{1}$, $a \in \mathbb{C}^*$, and $\#$ denotes the smash product of dimodule algebras.[21] The Z_k group action on $\mathscr{C}^{(k)}_{(a)}$ is defined via $\lambda \to \omega\lambda$. A similar generalization of $\mathscr{C}^{(k)}_n$ is an algebra $\#_{i=1}^n \mathscr{C}^{(k)}(a_i)$, $a_i \in \mathbb{C}^*$, the generators $\gamma_1,\ldots,\gamma_n$ of which satisfy relations $\gamma_i^k = \mathbf{1}a_i$, $\gamma_i\gamma_j = \omega\gamma_j\gamma_i$, $i<j$, $i,j = 1,\ldots,n$.

With the help of the smash product of Z_k-graded dimodule algebras one can obtain, besides $\mathscr{C}^{(k)}_n$-algebra extension of Γ, some other algebra extensions. Here they are as follows: let $\mathscr{C}^{(k)}_\beta$ be the k-dimensional algebra over $\mathbb{C}$ with the basis $1,\lambda,\ldots,\lambda^{k-1}$, where $\lambda^k = 1$. Take the grade of λ to be $\beta \in Z_k$ and let β be comprime with k. Define the Z_k action on $\mathscr{C}^{(k)}_\beta$ via $\kappa \in Z_k$, ${}^\kappa\lambda = \omega^\beta\lambda$. Then $\mathscr{C}^{(k)}_\beta$ becomes a dimodule algebra. Consider now the smash product

$$\mathscr{C}^{(k)}_{\beta_1} \# \mathscr{C}^{(k)}_{\beta_2} \# \cdots \# \mathscr{C}^{(k)}_{\beta_n} \equiv \mathscr{C}^{(k)}_{\{\beta_1,\ldots,\beta_n\}}.$$

Its generators,

$$\gamma_1 = \lambda_1 \# 1 \# \cdots \# 1,$$

$$\gamma_2 = 1 \# \lambda_2 \# 1 \cdots \# 1,$$

$$\vdots$$

$$\gamma_n = 1 \# \cdots \# 1 \# \lambda_n$$

then satisfy

$$\gamma_i\gamma_j = \omega^{\beta_j}\gamma_j\gamma_i, \quad i>j,$$
$$\gamma_i^k = \mathbf{1}, \quad i,j = 1,\dots,n.$$

For any choice of $\beta_1,\dots,\beta_n \in Z_k$ (β's comprime with k) we obtain "several" algebra extensions $\mathscr{C}^{(k)}_{\{\sigma(\beta_1),\dots,\sigma(\beta_n)\}}$, $\sigma\in S_n$. (The S_n group of permutations$\equiv$symmetric group.) All these extensions are isomorphic as algebras.

V. THE UNIVERSAL k-$\mathscr{C}_n$ CLIFFORD ALGEBRAS

Up to now we have presented Clifford-like albegras which generalized $\mathscr{C}_n^2$ algebras due to the observation that $\mathscr{C}_n^{(2)}$ is an algebra extension of $\Gamma = Z_2\oplus\cdots\oplus Z_2$ (n summands). In this section we introduce new Clifford-like algebras, denoted by k-$\mathscr{C}_n$, which are universal in the sense of the following commutative diagram:

$$\begin{tikzcd} \mathscr{V} \arrow[rr, "\alpha_0"] \arrow[dr, "\alpha"'] & & k\text{-}\mathscr{C}_n \arrow[dl, "\sigma"] \\ & A & \end{tikzcd}, \tag{5.1}$$

where $\mathscr{V}$ is an n-dimensional vector space, k-$\mathscr{C}_n$ and A are associative algebras, and α_0, α are corresponding monomorphisms with the property

$$[\alpha_0(x)]^k = Q_k(x)\mathbf{1}, \quad [\alpha(x)]^k = Q_k(x)\mathbf{1},$$

while $\sigma\in \mathrm{Hom}(k\text{-}\mathscr{C}_n, A)$. ■

For $k=2$, Q_2 is a quadratic form and 2-$\mathscr{C}_n \equiv \mathscr{C}_n^{(2)}$. For $k=3$, Q_3: $\mathscr{V}\longrightarrow\mathbb{C}$ becomes a cubic form, etc.

We also introduce in this section new Grassmann-like algebras.

The k-$\mathscr{C}_n$ algebras are defined by the commutative diagram (5.1) up to isomorphism and it is clear that

$$k\text{-}\mathscr{C}_n \simeq T(\mathscr{V})/I(Q_k),$$

where $T(\mathscr{V})$ is the tensor albegra of $\mathscr{V}$ and $I(Q_k)$ is the ideal of T generated by the elements

$$\{x \underset{k}{\otimes\cdots\otimes} x - Q_k(x)\mathbf{1}\}_{x\in\mathscr{V}},$$

The mapping Q_k ("k-ubic" form) is defined as follows.

Definition: For Q_k: $\mathscr{V}\longrightarrow\mathbb{C}$, we have the following:

(1) $Q_k(\lambda\mathbf{x}) = \lambda^k Q_k(\mathbf{x})$, $\mathbf{x}\in\mathscr{V}$, $\lambda\in\mathbb{C}$;

and (2) the mapping B_k: $\mathscr{V}\underset{k}{\times\cdots\times}\mathscr{V}\longrightarrow\mathbb{C}$,

$$k!B_k(\mathbf{x}_1,\dots,\mathbf{x}_k) = \sum_{l=1}^{k}(-1)^{k-l}\sum_{(i_1,\dots,i_l)\in S\binom{k}{l}} Q_k(\mathbf{x}_{i_1}+\cdots+\mathbf{x}_{i_l})$$

is k-linear. Here $S\binom{k}{l}$ denotes the family of subsets of $\{1,\dots,k\}$ that count l elements. ■

This generalized motion of quadratic and cubic (see Ref. 23, p. 114) forms is achieved by the process of polarization typical for multilinear structures. Clearly we have the identity

$$k!\,B_k(x,\dots,x) \equiv Q_k(x).$$

The Clifford-like algebra k-$\mathscr{C}_n$ has n generators. Namely, let $\{\hat{\gamma}_i\}_{i=1}^n$ be the "k-orthonormal" basis of $\mathscr{V}$, i.e.,

$$B(\hat{\gamma}_{i_1},\dots,\hat{\gamma}_{i_k}) = \delta(i_1,\dots,i_k), \quad i_1,\dots,i_k = 1,\dots,n,$$

where

$$\delta(i_1,\dots,i_k) = \begin{cases} 1, & i_1=\cdots=i_k, \\ 0, & \text{otherwise} \end{cases}.$$

Let ρ be a canonical epimorphism ρ: $T\to T/I(Q_k)$; then $[\rho(x)]^k = Q_k(x)\mathbf{1}$, $x\in\mathscr{V}$. [In the following we shall not distinguish $x\in\mathscr{V}$ from its monomorphic image $\rho(x)\in k$-$\mathscr{C}_n$.] It is easy to see now that

$$(1/k!)\,\{\hat{\gamma}_{i_1},\dots,\hat{\gamma}_{i_k}\} = \delta(i_1,\dots,i_k), \quad \forall\,\hat{\gamma}\in k\text{-}\mathscr{C}_n, \tag{5.2}$$

where

$$\{x_1,\dots,x_k\} \equiv \sum_{\sigma\in S_k} x_{\sigma(1)}x_{\sigma(2)}\cdots x_{\sigma(k)}.$$

Here, the k-$\mathscr{C}_n$ Clifford-like algebra is generated by $\{\hat{\gamma}_i\}_{i=1}^n$ satisfying (5.2). Clearly all this is due to the identity

$$k!\,B(x_1,\dots,x_k) \equiv \{x_1,\dots,x_k\}, \quad x_1,\dots,x_k\in\mathscr{V}\simeq\rho(\mathscr{V}).$$

We present now a class of matrix representations of k-$\mathscr{C}_n$ Clifford-like algebras. It is obvious and characteristic that (i) $\mathscr{C}_n^{(2)}$ is the faithful representation of 2-$\mathscr{C}_n$, and (ii) $\mathscr{C}_n^{(2)}$ is the only algebra extension of $Z_2\oplus\cdots\oplus Z_2$ (n summands) out of $2^{n(n-1)/2}$ possible, that has this property.

For $k>3$, $\mathscr{C}_n^{(k)}$ is also a representation of k-$\mathscr{C}_n$ but not faithful and neither is it the only algebra extension of $\Gamma\equiv Z_k\oplus\cdots\oplus Z_k$ ($n>1$ summands) representing k-$\mathscr{C}_n$. (Though it can be shown, for example, that all six algebra extensions of $Z_3\oplus Z_3\oplus Z_3$, $\mathscr{C}_3^{(3)}$ among them, representing 3-$\mathscr{C}_3$, are isomorphic as algebras.)

Let us prove what is stated above. Consider the diagram (5.1), $k>3$. The $\hat{\gamma},\dots,\hat{\gamma}_n$ satisfying (5.2) are $\alpha_0\equiv\rho|_{\mathscr{V}}$ images of a k-orthonormal basis in $\mathscr{V}$. Denote by $\gamma_1,\dots,\gamma_n$ the α_0 image of this basis in an algebra A. If A is chosen to be an algebra extension of Γ for which $\gamma_1,\dots,\gamma_n$ satisfy (5.2) then

$$0\neq\hat{\gamma}_1\hat{\gamma}_2^2 - \omega^{2\alpha_{12}}\hat{\gamma}_2^2\hat{\gamma}_1 \xrightarrow{\sigma} 0,$$

which shows that this very algebra A is not a faithful representation of k-$\mathscr{C}_n$. Now we have Lemma V.1.

Lemma V.1: Generators $\gamma_1,\dots,\gamma_n$ of $\mathscr{C}_n^{(k)}$ satisfy (5.2).

Proof: The case $i_1=\cdots=i_k$ in (5.2) is trivial, hence we assume otherwise. Let us consider first the case when $i_1=1$, $i_l\neq1$, $l=2,\dots,k$. Then

$$\{\gamma_1,\gamma_{i_2},\dots,\gamma_{i_k}\} \sim \sum_{l=0}^{k-1}\omega^l\sum_{\sigma\in S_{k-1}}\gamma_{\sigma(i_2)}\cdots\gamma_{\sigma(i_k)}\gamma_1 = 0.$$

Similarly for $i_l\neq1$, $l=3,\dots k$, we have

$$\{\gamma_1,\gamma_1,\gamma_{i_3},\dots,\gamma_{i_k}\} \sim \sum_{0<i<j<k-1}\omega^{i+j}\{\gamma_{i_3},\dots,\gamma_{i_k}\}\,\gamma_1\gamma_1 = 0,$$

and so on. The choice of $i_1=1$ (and so on) is replaced by the choice of other smallest number out of $\{i_1,\dots,i_k\}$ in the case $1\notin\{i_1,\dots,i_k\}$. The lemma is thus proved due to the famous zero

$$\sum_{0<i_1<\cdots<i_s<k-1}\omega^{i_1}\cdots\omega^{i_s} = 0, \quad s<k-1.$$

■

In the same manner one proves a more general lemma.

Lemma V.2: The $\gamma_1,...,\gamma_n$ generators of the $\mathscr{C}^{(k)}_{\{\beta_1,...,\beta_n\}}$-algebra extension of Γ (see Sec. IV) satisfy (5.2).

Analogously to the 2-$\mathscr{C}_n$ case one can define Grassmann-like (k-$\mathscr{G}_n$) algebras via k-$\mathscr{C}_n$ Clifford-like algebras.

Definition: k-$\mathscr{G}_v$ is the algebra generated by $1,\Theta_1,...,\Theta_v$, where

$$\{\Theta_{i_1},...,\Theta_{i_k}\} = 0, \quad i_1,...,i_k = 1,2,...,v. \tag{5.3}$$

■

One immediately gets the representation (denoted by $\mathscr{G}^{(k)}_v$) of k-$\mathscr{G}_v$ via $\mathscr{C}^{(k)}_{2v}$. Namely, let $\mathbb{C} \ni \kappa$, $\kappa^k = -1$, then define

$$\Theta_i \equiv \gamma_i + \kappa\gamma_{v+i}, \quad i = 1,...,v,$$

where the γ's are generators of $\mathscr{C}^{(k)}_{2v}$. Clearly these Θ's do satisfy (5.3).

It is also obvious that for the $\Theta^{(l)}$'s, $l = 0,...,k-1$,

$$\Theta^{(l)}_i = \gamma_i + \omega^l\kappa\gamma_{v+i}, \quad i = 1,...,v,$$

where the γ's are generators of $\mathscr{C}^{(k)}_{2v}$, represent generators k-$\mathscr{G}_v$.

It is then easy to show that

$$\left\{\Theta^{(l_1)}_{i_1},...,\Theta^{(l_k)}_{i_k}\right\} = (1 - \omega^{l_1 + \cdots + l_k})\delta(i_1,...,i_k),$$

and also we have a kind of "$\mathbf{Z}_k$-Witt decomposition":

$$\gamma_i = \frac{1}{k}\sum_{l=0}^{k-1} \Theta^{(l)}_i, \quad \gamma_{v+i} = \frac{1}{k}\sum_{l=0}^{k-1} \omega^{k-l}\,\Theta^{(l)}_i.$$

VI. FINAL REMARKS

A natural question arises how far one can pursue the investigation of 2-$\mathscr{C}_n = \mathscr{C}^{(2)}_n$ algebras,[24] but now for k-$\mathscr{C}_n$ Clifford-like algebras. Related question is to find out whether k-$\mathscr{G}_n$ is relevant to the same kind of projective geometry as $\mathscr{G}^{(2)}_n$ is via Plücker coordinates. These questions and other similar expectations (analogous to the case $k = 2$ geometrical facts) are, however, naive as the group of linear transformations leaving invariant the "k-ubic" ($k > 2$) form $Q_k(\mathbf{x}) = \Sigma^n_{i=1} x^k_i$ is finite—of order $k^n n!$. This linear group is given[1] by transformations

$$i = 1,...,n, \quad x_i \to \omega^{\alpha_i} x_{\sigma(i)}, \quad \alpha_i \in Z_k, \quad \sigma \in S_n.$$

The generators of k-$\mathscr{C}_n$ algebra, satisfying (5.2) linearize the "k-ubic" form Q_i; $\mathscr{V} \ni \mathbf{x} \to Q(\mathbf{x}) = \Sigma^n_{i=1} x^k_i$, i.e.,

$$1\sum_{i=1}^n x^k_i = \left(\sum_{i=1}^n x_i\hat{\gamma}_i\right)^k. \tag{6.1}$$

This property of $\{\hat{\gamma}_i\}^n_{i=1}$ is due to the obvious identity valid in any associative algebra A,

$$k!\left(\sum_{i=1}^n a_i\right)^k \equiv \sum_{i_1,...,i_k=1}^n \{a_{i_1},...,a_{i_k}\}, \quad a_i \in A.$$

As for the linearization of (6.1), any representation of (5.2) relations will do, of course—$\mathscr{C}^{(k)}_n$ algebra generators included. ■

Due to the property (6.1), algebras k-$\mathscr{C}_n$ can be applied to ultrageneralized complex analysis as in Ref. 19.

ACKNOWLEDGMENTS

The author would like to express his thanks and appreciation to Professor Luis Michel and Professor Jan Mozrzymas for an interesting discussion and some indications on how this work can be continued and applied.

[1]K. Morinaga and T. Nono, J. Sci. Hiroshima Univ. Ser. A. Math. Phys. Chem. **16**, 13 (1952); T. Nono, see Ref. 4(c).

[2]A. O. Morris, Q. J. Math. Oxford Ser. (2) **18**, 7 (1967); **19**, 289 (1968); E. Thomas, Glasgow Math. J. **15**, 74 (1974).

[3]I. Popovici and C. Ghéorghe, C. R. Acad. Sci. Paris **262**, 682 (1966); Rev. Roum. Math. Pures Appl. **XI**, 989 (1966).

[4]K. Yamazaki, (a) J. Fac. Sci. Univ. Tokyo Sect. 1 **10**, 147 (1964); (b) Sci. Pap. Coll. Gen. Edu. Univ. Tokyo **14**, 37 (1964); (c) in *Proceedings of the Conference on Clifford Algebra, Its Generalization and Applications*, edited by A. Ramakrishnan (Matscience, Madras, 1971), p. 20.

[5]A. Ramakrishnan, T. S. Santhanam, and P. S. Chandrasekaran, J. Math. Phys. Sci. Madras **3**, 307 (1969).

[6]T. S. Santhanam, P. S. Chandrasekaran, and N. B. Menon, J. Math. Phys. **12**, 377 (1971).

[7]A. Ramakrishnan see Ref. 4(c).

[8]T. S. Santhanam, Found. Phys. **7**, 121 (1977); T. S. Santhanam and K. B. Sinha, Aust. J. Phys. **31**, 233 (1978); T. S. Santhanam, Physica A **114**, 445 (1982).

[9]R. Jannathan and T. S. Santhanam, Theor. Phys. **21**, 363 (1982).

[10]T. S. Santhanam, P. S. Chandrasekaran, and N. B. Menon, J. Math. Phys. **12**, 377 (1971).

[11]M. Scheunert, J. Math. Phys. **20**, 712 (1979).

[12]A. K. Kwasniewski, Wrocław Univ. preprint No. 585, 1983.

[13]R. Kleeman, J. Math. Phys. **24**, 166 (1983).

[14]A. O. Morris, see Ref. 4(c).

[15]L. Michel, J. Mozrzymas, and B. Stawski, Lect. Notes Phys. **94**, 86 (1979).

[16]L. Michel and J. Mozrzymas, IHES/P/83/33; IHES/P/82/54; IHES/P/83/20.

[17]J. Mozrzymas, IHES/P/82/40.

[18]B. N. Valuev, Dubna Lectures 1977, Dubna P17-11020.

[19]P. J. McCarthy, Lett. Math. Phys. **4**, 509 (1980).

[20]N. Bourbaki, *Groupes et Algèbres de Lie* (Hermann, Paris, 1960), Chap. I.

[21]F. W. Long. J. London Math. Soc. (2) **13**, 438 (1976).

[22]J. Milnor, *Introduction to Algebraic K-Theory* (Annals of Mathematics Studies, No. 72, Princeton, 1971).

[23]C. C. Chevalley, *The Algebraic Theory of Spinors* (Columbia U.P., New York, 1955).

[24]M. F. Atiyah, R. Bott, and A. Shapiro, Topology **3**, 3 (1964).

Quantum-mechanical path integrals with Wiener measure for all polynomial Hamiltonians. II

Ingrid Daubechies[a)]
Theoretische Natuurkunde, Vrije Universiteit Brussel, B-1050 Brussels, Belgium
John R. Klauder
AT&T Bell Laboratories, Murray Hill, New Jersey 07974

(Received 8 January 1985; accepted for publication 3 May 1985)

The coherent-state representation of quantum-mechanical propagators as well-defined phase-space path integrals involving Wiener measure on continuous phase-space paths in the limit that the diffusion constant diverges is formulated and proved. This construction covers a wide class of self-adjoint Hamiltonians, including all those which are polynomials in the Heisenberg operators; in fact, this method also applies to maximal symmetric Hamiltonians that do not possess a self-adjoint extension. This construction also leads to a natural covariance of the path integral under canonical transformations. An entirely parallel discussion for spin variables leads to the representation of the propagator for an arbitrary spin-operator Hamiltonian as well-defined path integrals involving Wiener measure on the unit sphere, again in the limit that the diffusion constant diverges.

I. INTRODUCTION

For quantum systems the problem of providing a well-defined meaning for the heuristic and formal path-integral expressions for the propagator has attracted the attention of a number of workers.[1] The most commonly used prescription involves the continuum limit of a time-slicing formulation which, although perfectly correct,[2] is sometimes criticized as being far removed from the idealized desired goal of an integration over a space of paths defined for a continuous-time parameter. Unfortunately, in such quantum formulations, and unlike the Feynman–Kac formula, the orders of integration and the continuum limit cannot be interchanged to yield a formulation on continuous-time path spaces. Not only does this procedure fail for configuration-space path integrals, but seemingly even more so for the far more widely applicable phase-space path integrals.[3]

In this paper we propose an alternative to the time-slicing and continuum-limit procedure to define path integrals that leads to the quantum-mechanical propagator being given by well-defined path integrals involving Wiener measure on continuous phase-space paths in the limit that the diffusion constant diverges.[4] We are able to prove the existence of this formulation for a wide class of quantum Hamiltonians (described below) which includes all those that are polynomials in (Cartesian) P's and Q's. Indeed, our construction leads to a natural definition for the propagator even in cases where the Hamiltonian operator is maximal symmetric and admits *no* self-adjoint extension. Moreover, a formulation in terms of continuous phase-space paths permits one to make a transformation of integration variables, such as that involved in canonical transformations, with much greater care than usual (see the end of this section). We feel this possibility is just one of several advantages offered by our approach.

A. Motivation, summary of principal results, and outline of the paper

We begin by giving a heuristic overview of our formulation of quantum-mechanical phase-space path integrals. In terms of the canonical coherent states, defined in Dirac notation for all $(p,q) \in \mathbb{R}^2$ as

$$|p,q\rangle \equiv e^{i(pQ-qP)}|0\rangle,$$

where $|0\rangle$ is the normalized ground state of $(P^2+Q^2)/2$, one can write the following formal expression for the coherent state matrix elements of $\exp(-iTH)$ (see Ref. 5):

$$\langle p'',q''|e^{-iTH}|p',q'\rangle = \mathcal{N}^{-1}\int \exp\left[i\frac{1}{2}\int(p\dot q - q\dot p)dt - i\int H(p,q)dt\right]\prod_t dp(t)dq(t). \tag{1.1}$$

This is only a formal expression because there is no well-defined measure underlying this "integral"; $\mathcal{N}$ stands for a formal (actually infinite) "normalization constant." The function $H(p,q)$ was defined in Ref. 5 as the diagonal coherent state matrix element of H

$$H(p,q) = \langle p,q|H|p,q\rangle,$$

which, in the terminology of pseudodifferential operators, is equivalent with the "ordered symbol" corresponding to the operator H.

It is possible to give meaning to the formal expression (1.1) by inserting an extra factor

$$\exp\left[-\frac{1}{2\nu}\int(\dot p^2+\dot q^2)dt\right] \tag{1.2}$$

into the integrand, and redefining $\mathcal{N}$ in such a way that

$$\mathcal{N}^{-1}\exp\left[-\frac{1}{2\nu}\int(\dot p^2+\dot q^2)dt\right]\prod_t dp(t)dq(t)$$

can be interpreted as a Wiener measure with diffusion constant ν. The measure is pinned at p',q' at the initial time and at p'',q'' at the final time, a conditioning made possible by the use of the overcomplete coherent states. Since $\int(p\,dq - q\,dp)$ is a well-defined stochastic integral for this Wiener measure (and in fact the Itô and Stratonovich rules give the same result), then the function

[a)] "Bevoegdverklaard navorser" at the Nationaal Fonds voor Wetenschappelijk Onderzoek, Belgium.

 0022-2488/85/092239-18$02.50

$$\exp\left[i\frac{1}{2}\int(p\,dq-q\,dp)-i\int H(p,q)dt\right]$$

is integrable with respect to the Wiener measure, and the resulting expression is a well-defined path integral.

In the limit $\nu\to\infty$ the extra regularizing factor (1.2) formally tends to unity and the ν-dependent path integrals revert to the original formal expression. This entirely formal argument suggests that the coherent state matrix element $\langle p'',q''|\exp(-iTH)|p',q'\rangle$ might be considered as the limit, as the diffusion constant ν tends to ∞, of well-defined phase-space path integrals with Wiener measure.

Our main result is that this heuristic argument indeed contains some truth. More precisely, we will show that

$$\begin{aligned}\langle p'',q''|&\exp(-iTH)|p',q'\rangle\\ &=\lim_{\nu\to\infty}2\pi e^{\nu T/2}\int\exp\left[i\frac{1}{2}\int(p\,dq-q\,dp)\right.\\ &\quad\left.-i\int h(p,q)dt\right]d\mu_W^\nu(p,q),\end{aligned}\tag{1.3}$$

where μ_W^ν is the product of two independent Wiener measures (one in p, one in q) with diffusion constant ν, pinned at p',q' for $t=0$, and at p'',q'' for $t=T$. The normalization of the measure is given by

$$\begin{aligned}\int &d\mu_W^\nu(p,q)\\ &=[2\pi\nu T]^{-1}\exp\left\{-\frac{(p''-p')^2+(q''-q')^2}{2\nu T}\right\}.\end{aligned}\tag{1.4}$$

Its connected covariance is (x is either p or q) ($t_1\leqslant t_2$)

$$\begin{aligned}\langle x(t_1)x(t_2)\rangle^c&=\langle x(t_1)x(t_2)\rangle-\langle x(t_1)\rangle\langle x(t_2)\rangle\\ &=\nu t_1(1-t_2/T),\end{aligned}\tag{1.5}$$

where $\langle(\cdot)\rangle\equiv\int(\cdot)d\mu_W^\nu/\int d\mu_W^\nu$. The formula is valid for all self-adjoint Hamiltonians for which the finite linear span D_c of the harmonic oscillator eigenstates is a core, and which can be written as

$$H=\int\frac{dp\,dq}{2\pi}h(p,q)|p,q\rangle\langle p,q|.\tag{1.6}$$

The function $h(p,q)$ must satisfy, for all $\alpha>0$, the bound

$$\int dp\,dq|h(p,q)|^2\exp[-\alpha(p^2+q^2)]<\infty.$$

The class of Hamiltonians satisyfing these conditions contains all Hamiltonians polynomial in P and Q.

The function $h(p,q)$ used in the integrand in (1.3) is defined by (1.6). The relation between $h(p,q)$ and the diagonal matrix element $H(p,q)$ is given by

$$h(p,q)=\exp[-\tfrac{1}{2}(\partial_p^2+\partial_q^2)]H(p,q).\tag{1.7}$$

From (1.7) one sees that generally $h(p,q)\neq H(p,q)$; equality only holds when $H(p,q)$ is linear in p and q. In pseudodifferential operator terminology, $h(p,q)$ is equivalent to the "antiordered" symbol. From the difference between h and H one sees that (1.3) is more than just a "regularization" of (1.1) by (1.2). We shall return later (at the end of Sec. II) to the role played by $h(p,q)$.

As a matter of fact, our approach can also handle symmetric operators which are *not* self-adjoint. Formula (1.3) still holds if the closure of $H|_{D_c}$ is maximal symmetric, where we then have to write either $\exp(-iHT)$ or $\exp(-iH^*T)$ in the matrix element on the left-hand side, according to which deficiency index of $H|_{D_c}$ is zero (see Theorem 2.4 in Sec. II C). Here H is again defined by (1.6), and the growth restriction on h ensures that H is well defined on D_c.

Note also that the regularization procedure which consists of inserting terms of type (1.2) into (1.1) in order to obtain (1.3) cannot work for the ordinary configuration-space path integral (whereas we assert here that it *does* work for the coherent-state, phase-space path integral). The reason for this is that the configuration-space path integral contains (formally) factors of the type $\exp(i\frac{1}{2}\int\dot q^2\,dt)$ in the integrand. This cannot be regularized by inserting an extra factor $\exp(-\frac{1}{2}\nu^{-1}\int\dot q^2\,dt)$; an old argument[6] shows that it is impossible to define the Brownian measure with a nonreal diffusion constant [or, alternatively, $\exp(i\frac{1}{2}\int\dot q^2\,dt)$ is not a measurable function with respect to a Wiener measure]. One could imagine inserting $\exp(-\frac{1}{2}\nu^{-1}\int\ddot q^2\,dt)$; however, the additional data needed at the initial and final times are outside the scope of the configuration-space approach (it is more nearly like the coherent-state approach; compare, however, Itô, Ref. 1).

For the proof of (1.3) we shall first show that the path integral in the right-hand side of (1.3) can be considered, for finite ν, as the integral kernel of a contraction operator on $L^2(\mathbb{R}^2)$, the set of square-integrable functions on phase space. This will be done in Sec. II B, after we have defined all the necessary machinery in Sec. II A. In Sec. II C we take the limit $\nu\to\infty$, and prove (1.3) (Theorem 2.4). For reasons of simplicity we will restrict ourselves to the case of one degree of freedom, i.e., to a two-dimensional phase space. Everything we do can be trivially extended to any finite number of degrees of freedom.

In Sec. III we discuss path integrals for Hamiltonians containing spin operators. Again we consider path-integral expressions for coherent-state matrix elements of the evolution operators corresponding to these Hamiltonians. The coherent states used here are associated with SU(2) rather than with the Heisenberg group, and are labeled by elements of S^2 rather than of $\mathbb{R}^2$. In our construction we shall be able to treat an *arbitrary* Hamiltonian written, analogously to (1.6), as a superposition of diagonal dyadic operators in the spin coherent states (this representation has been studied and used before; its first use in the construction of path integrals for spin systems was by Lieb[7]). Once the appropriate definitions are formulated (Sec. III A), the analysis of Sec. II carries over to the spin case without any problem, and we therefore shall only state the result, without detailed proofs (Secs. III B and III C).

We have already announced our principal results in Ref. 4, in a slightly weaker version. The proofs outlined in Ref. 4 are, however, different from the ones we give here, though there is some connection. In the Appendix we compare the two versions, and show how our previous approach fits into the present framework.

B. Canonical transformations

As an illustration of our path-integral formalism we conclude this Introduction with a few remarks about how time-independent canonical transformations appear in our approach. For this purpose it is useful to interpret all stochastic integrals and stochastic differential equations in the sense of Stratonovich,[8] and this we shall do in this subsection. We introduce new canonical coordinates $\bar{p} = \bar{p}(p,q)$ and $\bar{q} = \bar{q}(p,q)$, which are classically connected, for example, by the relation

$$p\,dq - q\,dp = \bar{p}\,d\bar{q} - \bar{q}\,d\bar{p} + 2\,dF(\bar{p},\bar{q};p,q).$$

The stochastic variables $\bar{p}$ and $\bar{q}$ satisfy the stochastic differential equations given by

$$d\bar{p} = \frac{\partial \bar{p}}{\partial p}\,dp + \frac{\partial \bar{p}}{\partial q}\,dq, \quad d\bar{q} = \frac{\partial \bar{q}}{\partial p}\,dp + \frac{\partial \bar{q}}{\partial q}\,dq,$$

and their solution determines a new, generally non-Gaussian measure $\bar{\mu}^\nu(\bar{p},\bar{q})$ according to $d\bar{\mu}^\nu(\bar{p},\bar{q}) = d\mu_W^\nu(p,q)$. In the new canonical coordinates (1.3) becomes

$$\langle \bar{p}'',\bar{q}''|e^{-iTH}|\bar{p}',\bar{q}'\rangle = \lim_{\nu\to\infty} 2\pi e^{\nu T/2} \int \exp\Big[i\frac{1}{2}\int (\bar{p}\,d\bar{q} - \bar{q}\,d\bar{p}) - i\int \bar{h}(\bar{p},\bar{q})dt \Big] d\bar{\mu}^\nu(\bar{p},\bar{q}), \tag{1.8}$$

where $\bar{h}(\bar{p},\bar{q}) \equiv h(p(\bar{p},\bar{q}),q(\bar{p},\bar{q}))$, and where we have incorporated the effects of F by defining the states

$$|\bar{p},\bar{q}\rangle \equiv \exp[iF(\bar{p},\bar{q};p(\bar{p},\bar{q}),q(\bar{p},\bar{q}))]\,|p(\bar{p},\bar{q}),q(\bar{p},\bar{q})\rangle.$$

With this phase convection, (1.8) is canonically equivalent to (1.3); the phase is still given by the classical action for stochastic phase-space paths; what is different is the weighting of those paths by the integration measure. Note that the measures $\bar{\mu}^\nu$ and μ_W^ν are typically mutually singular, as is already the case if $\bar{p} = ap$, $\bar{q} = q/a$, for $a > 0$, $a \neq 1$.

It is straightforward to extend the foregoing discussion to time-dependent canonical transformations.

II. THE CANONICAL CASE

A. Definitions and basic properties

We start by a review of the definition and some of the properties of the canonical coherent states. Let $\mathscr{H}$ be a separable complex Hilbert space carrying an irreducible, strongly continuous unitary representation $W(p,q)$ of the Weyl commutation relations

$$W(p',q')W(p'',q'') = \exp[i\tfrac{1}{2}(p'q'' - p''q')]W(p'+p'',q'+q'').$$

The position operator Q and the momentum operator P are the infinitesimal generators of the strongly continuous unitary groups $W(p,0)$, $W(0,-q)$, respectively; one has

$$W(p,q) = \exp[i(pQ - qP)] = \exp(-i\tfrac{1}{2}pq)\exp(ipQ)\exp(-iqP). \tag{2.1}$$

We define $\omega \in \mathscr{H}$ to be the normalized ground state of the harmonic oscillator Hamiltonian

$$\tfrac{1}{2}(P^2 + Q^2 - 1)\omega = 0.$$

The canonical coherent states (cs) are defined as

$$\omega^{p,q} = W(p,q)\omega.$$

They form an overcomplete set of vectors in $\mathscr{H}$ with "overlap function"

$$\langle \omega^{p'',q''},\omega^{p',q'}\rangle = \exp[i\tfrac{1}{2}(p'q'' - p''q') - \tfrac{1}{4}(p''-p')^2 - \tfrac{1}{4}(q''-q')^2]. \tag{2.2}$$

They also give rise to the following "resolution of unity":

$$\int \frac{dp\,dq}{2\pi} \langle \psi,\omega^{p,q}\rangle\langle\omega^{p,q},\phi\rangle = \langle\psi,\phi\rangle. \tag{2.3}$$

This can be viewed as a special case of

$$\int \frac{dp\,dq}{2\pi} \langle \psi_1,W(p,q)\psi_2\rangle\langle W(p,q)\phi_1,\phi_2\rangle = \langle\psi_1,\phi_2\rangle\langle\phi_1,\psi_2\rangle. \tag{2.4}$$

Note: In the usual Schrödinger representation, one has $\mathscr{H} = L^2(\mathbb{R})$. The $W(p,q)$ act then as follows:

$$[W(p,q)f](x) = \exp(-i\tfrac{1}{2}pq + ipx)f(x-q).$$

The vector $\omega^{p,q}$ is given by the familiar functions

$$\omega^{p,q}(x) = \pi^{-1/4}\exp[-i\tfrac{1}{2}pq + ipx - \tfrac{1}{2}(x-q)^2].$$

Setting $p = q = 0$ gives $\omega(x)$.

We shall also use the harmonic oscillator excited states ω_k, defined by

$$\tfrac{1}{2}(P^2 + Q^2 - 1)\omega_k = k\omega_k. \tag{2.5}$$

In analogy with the definition of the cs we define

$$\omega_k^{p,q} = W(p,q)\omega_k.$$

In order to alleviate many of the expressions in what follows, we shall often make use of Dirac's bra-ket notation in scalar products, matrix elements, and dyadic operators involving the coherent states. We shall write, e.g.,

$$\begin{aligned}
&\langle p,q|\phi\rangle \equiv \langle\omega^{p,q},\phi\rangle \quad (\phi\in\mathscr{H}),\\
&\langle k|\phi\rangle \equiv \langle\omega_k,\phi\rangle,\\
&\langle p,q;k|\phi\rangle \equiv \langle\omega_k^{p,q},\phi\rangle,\\
&\langle p'',q''|p',q'\rangle \equiv \langle\omega^{p'',q''},\omega^{p',q'}\rangle,\\
&|p,q\rangle\langle p,q| \equiv \omega^{p,q}\langle\omega^{p,q},\cdot\,\rangle,\\
&|p,q;k\rangle\langle p,q;l| \equiv \omega_k^{p,q}\langle\omega_l^{p,q},\cdot\,\rangle.
\end{aligned}$$

In these notations (2.3), e.g., can be written as

$$\int \frac{dp\,dq}{2\pi} |p,q\rangle\langle p,q| = \mathbf{1}_{\mathscr{H}}, \tag{2.6}$$

where the integral converges weakly, according to (2.3). As a matter of fact, (2.6) also converges strongly; see, e.g., the remark following Lemma 2.3 in Sec. II C. Equation (2.4) implies

$$\int \frac{dp\,dq}{2\pi} |p,q;k\rangle\langle p,q;l| = \delta_{kl}\mathbf{1}_{\mathscr{H}}. \tag{2.7}$$

For matrix elements $\langle\omega^{p'',q''},A\,\omega^{p',q'}\rangle$ we shall use the notation $\langle p'',q''|A\,p',q'\rangle$ rather than the more common bra-ket notation $\langle p'',q''|A|p',q'\rangle$ (i.e., we use a space instead of the second vertical bar) in order to avoid confusion in case A is not symmetric.

In the next section we shall interpret the right-hand side of (1.3) as the integral kernel of an operator on $L^2(\mathbb{R}^2)$; this operator can be constructed explicitly, and its limit for $\nu \to \infty$ can then be taken later. In order to do all this, we shall need the following definitions and constructions.

We shall use the notation $L^2(V)$ for the Hilbert space $L^2(\mathbb{R}^2)$ with the normalization

$$\|f\|^2 = \int \frac{dp\,dq}{2\pi} |f(p,q)|^2.$$

For $\psi \in \mathscr{H}$, we shall denote by f_ψ the function

$$f_\psi(p,q) = \langle p,q|\psi\rangle.$$

It follows from (2.3) that the map $\psi \to f_\psi$ is isometric from $\mathscr{H}$ into $L^2(V)$; the image of $\mathscr{H}$ under this map is a closed subspace $\mathscr{H}_o$ of $L^2(V)$. The properties of $\mathscr{H}_o$ are well known[9]; its elements are products of analytic functions in $p + iq$ with the Gaussian $\exp[-\frac{1}{4}(p^2+q^2)]$. We shall denote the isomorphism between $\mathscr{H}$ and $\mathscr{H}_o$ by U

$$U: \mathscr{H} \to \mathscr{H}_o, \quad (U\psi)(p,q) = \langle p,q|\psi\rangle. \tag{2.8}$$

We shall also make use of the operator $\widehat{U}: \mathscr{H} \to L^2(V)$, which is defined as $\widehat{U} = I \circ U$, where I is the natural embedding of $\mathscr{H}_o$ into $L^2(V)$. The orthogonal projection operator in $L^2(V)$, onto $\mathscr{H}_o$, will be denoted by P_o.

Define also

$$h_{kl}(p,q) \equiv \langle W(p,q)\omega_k, \omega_l\rangle = \langle p,q;k\,|l\,\rangle. \tag{2.9}$$

These functions can be explicitly calculated; they are related to the generalized Laguerre functions, and can all be written as the product of a polynomial in p,q with $\exp[-(p^2+q^2)/4]$. One easily sees from (2.4) that the h_{kl} are orthonormal in $L^2(V)$; as a matter of fact, they form a complete orthonormal basis for $L^2(V)$ (see Ref. 10). From (2.8) one then sees that the h_{ol} are a complete orthonormal basis for $\mathscr{H}_o$. We shall use the notation D for the set of finite linear combinations of the h_{kl}. Note that for any $\psi \in \mathscr{H}$

$$\begin{aligned}\langle h_{kl}, \widehat{U}\psi\rangle &= \int \frac{dp\,dq}{2\pi} \langle l\,|p,q;k\,\rangle\langle p,q;0|\psi\rangle \\ &= \delta_{ko}\langle l\,|\psi\rangle.\end{aligned}$$

Suppose that R is a (bounded) operator on $\mathscr{H}$. The unitary map U transports this operator to URU^{-1} on $\mathscr{H}_o$. A simple way to extend URU^{-1} to all of $L^2(V)$ is to "fill in zeros," i.e., we define $\widehat{R}$ on $L^2(V)$ such that

$$\begin{aligned}&\widehat{R}f = 0, \quad \text{if } f \perp \mathscr{H}_o, \\ &\widehat{R}f = \widehat{U}RU^{-1}f, \quad \text{if } f \in \mathscr{H}_o.\end{aligned} \tag{2.10}$$

It turns out that $\widehat{R}$ is an integral operator on $L^2(V)$

$$\begin{aligned}&\langle h_{kl}, \widehat{R}h_{rs}\rangle \\ &= \delta_{ok}\delta_{or}\langle h_{ol}, \widehat{R}h_{os}\rangle \\ &= \delta_{ok}\delta_{or}\langle \omega_l, R\omega_s\rangle \\ &= \delta_{ok}\delta_{or}\int \frac{dp''\,dq''}{2\pi}\int \frac{dp'\,dq'}{2\pi} \\ &\quad \times \langle l\,|p'',q''\rangle\langle p'',q''|R\,p',q'\rangle\langle p',q'|s\rangle \\ &= \int \frac{dp''\,dq''}{2\pi}\int \frac{dp'\,dq'}{2\pi} \\ &\quad \times \langle l\,|p'',q'';k\,\rangle\langle p'',q'';0|R\,p',q',0\rangle\langle p',q';r|s\rangle \\ &= \int \frac{dp''\,dq''}{2\pi}\int \frac{dp'\,dq'}{2\pi} \\ &\quad \times \overline{h_{kl}(p'',q'')}\langle p'',q''|R\,p',q'\rangle h_{rs}(p',q').\end{aligned} \tag{2.11}$$

Hence $\widehat{R}$ has integral kernel $\langle p'',q''|R\,p',q'\rangle$; note that this integral kernel (since it is a cs matrix element in $\mathscr{H}$) is a smooth function of p'',q'',p',q'.

For H a (possibly unbounded) self-adjoint operator on $\mathscr{H}$, we have

$$\widehat{(e^{itH})} = P_o e^{it\widehat{H}} P_o \tag{2.12}$$

(here we have extended our construction of $\widehat{H}$ to unbounded operators; for the Hamiltonians we shall consider, however, this is not a problem). It is necessary to introduce P_o in (2.12) because $\widehat{H}(1-P_o) = 0$, hence $[\exp(-it\widehat{H})](1-P_o) = 1 - P_o$, whereas $\widehat{[\exp(-itH)]}(1-P_o) = 0$ [see (2.10)].

For the special class of operators R and H, which can be written as

$$R = \int \frac{dp\,dq}{2\pi} r(p,q)|p,q\rangle\langle p,q|,$$

another natural extension, different from $\widehat{R}$, is possible. Define the multiplication operator

$$R_V: L^2(V) \to L^2(V), \quad (R_V f)(p,q) = r(p,q)f(p,q). \tag{2.13}$$

Then

$$\begin{aligned}\langle h_{ol}, R_V h_{os}\rangle &= \int \frac{dp\,dq}{2\pi} \langle l\,|p,q\rangle r(p,q)\langle p,q|s\rangle \\ &= \langle l\,|R\,s\rangle,\end{aligned}$$

which shows that $P_o R_V P_o|_{\mathscr{H}_o} = URU^{-1}$, hence

$$P_o R_V P_o = \widehat{R}. \tag{2.14}$$

We are now ready to tackle our path integral. In the next subsection we shall see that, for finite ν, the path integral in the right-hand side of (1.3) can be interpreted as the integral kernel of an operator on $L^2(V)$, which we can construct explicitly.

B. Interpretation of the path integral (for finite ν) as an integral kernel on $L^2(V)$

Let us introduce the symbol $\mathscr{P}_\nu(h)$ for the expression in the right-hand side of (1.3)

$$\begin{aligned}&\mathscr{P}_\nu(h;p'',q'',t'';p',q',t') \\ &= 2\pi e^{\nu(t''-t')/2}\int \exp\left[i\frac{1}{2}\int(p\,dq - q\,dp)\right. \\ &\quad \left. - i\int h(p,q)dt\right] d\mu_W^\nu(p,q),\end{aligned} \tag{2.15}$$

where again the measure μ_W^ν is the product of two independent Wiener measures with diffusion constant ν, and pinned at p',q' for $t = t'$ and at p'',q'' for $t = t''$, respectively $(t'' > t')$.

If we put $h = 0$, $\mathscr{P}_\nu$ can be calculated explicitly; the result is (for the case $\nu = 1$, this calculation was carried out in Ref. 11)

$$\mathscr{P}_\nu(h=0;p'',q'',t'',p',q',t')$$
$$=\frac{e^{\nu(t''-t')/2}}{2\sinh[\nu(t''-t')/2]}\exp\left\{\frac{i}{2}(p'q''-p''q') - \frac{1}{4}\coth\frac{\nu(t''-t')}{2}\times[(p''-p')^2+(q''-q')^2]\right\}. \quad (2.16)$$

By their definition, these $\mathscr{P}_\nu(h=0)$ have a semigroup property, as can also be checked by explicit calculation

$$\int\frac{dp'\,dq'}{2\pi}\mathscr{P}_\nu(h=0;p'',q'',t'';p',q',t') \times\mathscr{P}_\nu(h=0;p',q',t';p,q,t) =\mathscr{P}_\nu(h=0;p'',q'',t'',p,q,t). \quad (2.17)$$

From (2.16) one sees that

$$|\mathscr{P}_\nu(h=0;p'',q'',t'';p',q',t')| \leqslant e^{\nu(t''-t')/2}[\nu(t''-t')]^{-1} \times\exp\{-[(p''-p')^2+(q''-q')^2]/[2\nu(t''-t')]\}. \quad (2.18)$$

Since for all $\alpha>0$ the function $\rho_\alpha(p,q)=\exp[-\alpha(p^2+q^2)]$ is in $L^1(\mathbf{R}^2)$, the upper bound (2.18) implies that for $t''\neq t'$, $\mathscr{P}_\nu(h=0)$ is the integral kernel of a bounded operator $E^o(\nu;t'',t')$ on $L^2(V)$

$$\mathscr{P}_\nu(h=0;p'',q'',t'';p',q',t') =[E^o(\nu;t'',t')](p'',q'';p',q'). \quad (2.19)$$

The bound (2.18) implies $\|E^o(\nu;t'',t')\|\leqslant e^{\nu(t''-t')}$. From the semigroup relation (2.17) we see that

$$E^o(\nu;t'',t)E^o(\nu;t,t')=E^o(\nu;t'',t'). \quad (2.20)$$

It is also easy to check from (2.16) that

$$[E^o(\nu;t'',t')]^*=E^o(\nu;t'',t'). \quad (2.21)$$

As t'' tends to $t'(t''\to t')$, it is clear from (2.16) that $\mathscr{P}_\nu(h=0;p'',q'',t'';p',q',t')$ tends to $2\pi\delta(p''-p')\delta(q''-q')$ in the sense of distributions. Using (2.20), (2.21), and $\|E^o(\nu;t'',t')\|\leqslant e^{\nu(t''-t')}$, this implies s-$\lim_{t''\to t'}E^o(\nu;t'',t')=\mathbf{1}$ on $L^2(V)$. Moreover, one sees from the explicit expression (2.16) that the $\mathscr{P}_\nu(h=0)$ depend on the initial and final times t',t'' only through the difference $(\nu t''-\nu t')$. Putting all this together, we conclude that, for fixed ν, the $E^o(\nu,t)\equiv E^o(\nu;t,0)$ form a strongly continuous semigroup (in t) of bounded operators. Hence

$$E^o(\nu;t'',t')=E^o(\nu,t''-t')=\exp[-\nu A(t''-t')], \quad (2.22)$$

where A is a self-adjoint operator on $L^2(V)$. The bound $\|E^o(\nu,t)\|\leqslant e^{\nu t}$ implies $A\geqslant -1$. The operator A can be calculated explicitly from (2.16). One finds

$$A=\tfrac{1}{2}[-(\partial_p^2+\partial_q^2)+i(p\,\partial_q-q\,\partial_p)+\tfrac{1}{4}(p^2+q^2)-1] =\tfrac{1}{2}[(-i\,\partial_p+q/2)^2+(-i\,\partial_q-p/2)^2-1]. \quad (2.23)$$

It is particularly interesting to note, if A is interpreted as a Hamiltonian on $L^2(\mathbf{R}^2)$, that it describes a two-dimensional particle in the presence of a constant magnetic field orthogonal to the plane of motion. Indeed, exactly such a Hamiltonian arises in the two-dimensional quantized Hall effect, and good use has been made of coherent state techniques in the study of this problem.[12] Because of this magnetic field analogy, we know immediately that A has a purely discrete spectrum (the Landau levels for the corresponding magnetic field). As a matter of fact, we already have an expression for a set of eigenvalues and eigenvectors for A. Recalling the definition (2.9) of the h_{kl}, and using the definitions (2.5) and (2.1) of the ω_k and $W(p,q)$, respectively, one finds

$$Ah_{kl}=kh_{kl}. \quad (2.24)$$

This will be useful for our analysis of the $\nu\to\infty$ limit below. It follows immediately from (2.24) that D, the set of finite linear combinations of the h_{kl}, is a core for A. Note that (2.24) also implies $A\geqslant 0$. We have therefore $\|E^o(\nu,t)\| = \|\exp(-\nu At)\|\leqslant 1$, which means that the $E^o(\nu,t)$ are a strongly continuous *contraction* semigroup.

Let us now look at the case where h is not identically zero. We shall consider functions h satisfying, for all $\alpha>0$, the condition

$$\int dp\,dq|h(p,q)|^2\exp[-\alpha(p^2+q^2)]<\infty. \quad (2.25)$$

This is automatically fulfilled if, e.g.,

$$\int dp\,dq|h(p,q)|^2\exp[-\beta(p^2+q^2)^\gamma]<\infty,$$

for some $\beta>0$ and $0<\gamma<1$.

Condition (2.25) ensures that the path integral (2.15) is well defined. To see this, we only need to check that $|\int_0^T h(p(t),q(t))dt|$ is finite for almost all paths $(p(t),q(t))$ in the support of μ_W^ν. This is certainly true if

$$\int\left\{\int_0^T|h(p(t),q(t))|dt\right\}d\mu_W^\nu(p,q)<\infty.$$

Using the definition of μ_W^ν, we can rewrite this condition as

$$\int_0^T dt\int dp\,dq|h(p,q)|(2\pi\nu t)^{-1} \times\exp\left\{-\frac{(p-p')^2+(q-q')^2}{2\nu t}\right\} \times[2\pi\nu(T-t)]^{-1}\exp\{-[(p-p'')^2 +(q-q'')^2]/[2\nu(T-t)]\}<\infty. \quad (2.26)$$

For h satisfying condition (2.25), we can use the Cauchy–Schwarz inequality to bound the left-hand side of (2.26) by

$$C\int_0^T t^{-1}(T-t)^{-1}\int dp\,dq\exp\left\{-\frac{(p-p')^2+(q-q')^2}{\nu t} -\frac{(p-p'')^2+(q-q'')^2}{\nu(T-t)}+\frac{(p^2+q^2)}{\nu T}\right\}$$

$$\leqslant C\int_0^T dt\,\nu T[T^2-t(T-t)]^{-1} \times\exp\left\{\frac{2T(|p'||p''|+|q'||q''|)}{\nu[T^2-t(T-t)]}\right\}. \quad (2.27)$$

Since, for $t \in [0,T]$, $T^2 - t(T-t) \geqslant 3T^2/4$, one immediately sees that expression (2.27) is finite, and hence that $\mathscr{P}_\nu(h)$ is well defined for all p',q',p'',q''. From (2.15) we see then that, for all h satisfying (2.25),

$$|\mathscr{P}_\nu(h;p'',q'',t'';p',q',t')| < 2\pi e^{\nu(t''-t')/2} \int d\mu_W^\nu(p,q) < \frac{e^{\nu(t''-t')/2}}{\nu(t''-t')} \exp\left\{-\frac{(p''-p')^2+(q''-q')^2}{2\nu(t''-t')}\right\}. \quad (2.28)$$

Since h is time independent, $\mathscr{P}_\nu(h)$ will depend on t'',t' only through the difference $t''-t'$. Together with (2.28) this implies that $\mathscr{P}_\nu(h)$ is the integral kernel of a bounded operator $E(\nu,h;t''-t')$ on $L^2(V)$

$$\mathscr{P}_\nu(h;p'',q'',t'';p',q',t') = [E(\nu,h;t''-t')](p'',q'';p',q'), \quad (2.29)$$

with

$$\|E(\nu,h;t)\| < e^{\nu t/2}. \quad (2.30)$$

From the path integral definition (2.15) one immediately sees that the semigroup property (2.17) for the $\mathscr{P}_\nu(h)$ also holds for $h \neq 0$. This implies $[t_1,t_2 > 0; E(\nu,h;0) \equiv \mathbb{1}]$

$$E(\nu,h;t_1)E(\nu,h;t_2) = E(\nu,h;t_1+t_2), \quad (2.31)$$

i.e., the $E(\nu,h;t)$ form a semigroup.

We want to show that the $E(\nu,h;t)$ actually form a *strongly continuous* semigroup of *contractions* on $L^2(V)$ [which means we have to do better than (2.30)!]. In order to do this, we shall proceed in several steps. We shall first consider h in C_0^∞, the C^∞ functions vanishing at ∞. Then we shall extend our results to bounded h, and in a third step we generalize to all h satisfying condition (2.25).

For any h satisfying (2.25), we define H_V to be [as in (2.13)] the multiplication operator by $h(p,q)$

$$(H_V f)(p,q) = h(p,q) f(p,q). \quad (2.32)$$

We shall always assume that h is a real function, which implies that H_V is self-adjoint. If h is not bounded, the domain $D(H_V)$ of H_V consists of all $f \in L^2(V)$ for which hf is still square integrable. Because of the special form of the h_{kl} we have

$$|h_{kl}(p,q)| \leqslant C(1+p^2+q^2)^n \exp[-(p^2+q^2)/4]$$

(C and n depend on k and l); together with (2.25) this implies that $h_{kl} \in D(H_V)$ for all k,l.

Let us now consider $h \in C_0^\infty$. Then H_V is a bounded operator. On the domain $D(A)$ of A we can define $\nu A + iH_V$. This is a closed operator, which is the generator of a strongly continuous contraction semigroup (both νA and iH_V are generators, and H_V is A bounded with relative bound zero; see, e.g., Kato's book,[13] p. 499). The integral kernel of $\exp[-(\nu A + iH_V)T]$ is given by the Trotter product formula

$$\begin{aligned}
&\{\exp[-(\nu A + iH_V)T]\}(p'',q'';p',q') \\
&= \lim_{N\to\infty} \int \cdots \int \prod_{j=0}^{N} \{[\exp(-\nu A\epsilon)](p_{j+1},q_{j+1};p_j,q_j) \\
&\quad \times \exp[-ih(p_j,q_j)\epsilon]\} \prod_{j=1}^{N} \left(\frac{dp_j\, dq_j}{2\pi}\right) \\
&= \lim_{N\to\infty} 2\pi \int \cdots \int \prod_{j=0}^{N} \left(\frac{\exp(\nu\epsilon/2)}{4\pi \sinh(\nu\epsilon/2)}\right. \\
&\quad \times \exp\{(i/2)(p_j q_{j+1} - p_{j+1} q_j) \\
&\quad - \tfrac{1}{4}\coth(\nu\epsilon/2)[(p_{j+1}-p_j)^2 + (q_{j+1}-q_j)^2]\} \\
&\quad \left.\times \exp[-ih(p_j,q_j)\epsilon]\right) \prod_{j=1}^{N} (dp_j\, dq_j)
\end{aligned}$$

[use (2.16) and (2.17)].

Here we have used the notations $p_o = p'$, $q_o = q'$, $p_{N+1} = p''$, $q_{N+1} = q''$, and $\epsilon \equiv T/(N+1)$. In the limit for $N\to\infty$, we can replace $[\sinh(\nu\epsilon/2)]^{-1}$ and $\coth(\nu\epsilon/2)$ by their first-order approximation $2/\epsilon\nu$ (higher-order terms do not contribute in the limit). This leads to

$$\begin{aligned}
&\{\exp[-(\nu A + iH_V)T]\}(p'',q'';p',q') \\
&= 2\pi e^{\nu T/2} \lim_{N\to\infty} \int \cdots \int \prod_{j=0}^{N} \left(\exp\left\{\frac{1}{2} i[p_j(q_{j+1}-q_j) - q_j(p_{j+1}-p_j)]\right\} \exp[-ih(p_j,q_j)\epsilon]\right) \\
&\quad \times \prod_{j=0}^{N} \left([2\pi\nu\epsilon]^{-1} \exp\left\{-\frac{(q_{j+1}-q_j)^2 + (p_{j+1}-p_j)^2}{2\nu\epsilon}\right\}\right) \prod_{j=1}^{N} (dp_j\, dq_j) \\
&= 2\pi e^{\nu T/2} \int \exp\left[i\frac{1}{2}\int (p\,dq - q\,dp) - i\int h(p,q)dt\right] d\mu_W^\nu(p,q) \\
&= \mathscr{P}_\nu(h;p'',q'',T;p',q',0).
\end{aligned} \quad (2.33)$$

Here we have used the continuity of h in the limit so that as $N\to\infty$,

$$\exp\left[-i\sum_{j=0}^{N} h(p_j,q_j)\epsilon\right] \to \exp\left[-i\int h(p,q)dt\right].$$

Comparing (2.33) with (2.29) we see that for $h \in C_0^\infty$,

$$E(\nu,h;t) = \exp[-(\nu A + iH_V)t]. \quad (2.34)$$

This implies that for h in C_0^∞ we have achieved our goal: the $E(\nu,h;t)$ form a strongly continuous contraction semigroup.

For h in L^∞, the operator H_V defined by (2.32) is still bounded. The operator $\nu A + iH_V$, defined on $D(A)$, is therefore still a generator of a strongly continuous contraction semigroup. We can find functions h_n in C_0^∞ such that $|h_n(p,q)| \leqslant \|h\|_\infty$ for all p,q and $h_n(p,q) \to h(p,q)$ almost everywhere (a.e.). By the dominated convergence theorem, one sees immediately from (2.15) that this implies $\mathscr{P}_\nu(h_n)$

$\to_{n\to\infty} \mathscr{P}_\nu(h)$ pointwise. On the other hand, $(\nu A + iH_{V,n})f \to (\nu A + iH_V)f$ for all $f\in D(A)$, and therefore $\nu A + iH_{V,n}$ converges to $\nu A + iH_V$ in the strong resolvent sense (see Ref. 13, Theorem VIII 1.5), where we have used the notation $H_{V,n}$ for the multiplication operator by h_n on $L^2(V)$. Hence

$$\operatorname*{s-lim}_{n\to\infty} \exp[-(\nu A + iH_{V,n})t] = \exp[-(\nu A + iH_V)t].$$

Using (2.28) [note that this upper bound on $\mathscr{P}_\nu(h)$ is independent of h !] one can therefore apply the dominated convergence theorem to see that $[f, g \in L^2(V)]$

$$\begin{aligned}
&\langle f, \exp[-(\nu A + iH_V)t]\, g\rangle \\
&= \lim_{n\to\infty} \langle f, \exp[-(\nu A + iH_{V,n})t]\, g\rangle \\
&= \lim_{n\to\infty} \int \frac{dp''\,dq''}{2\pi} \int \frac{dp'\,dq'}{2\pi}\, \overline{f(p'',q'')} \\
&\quad \times \mathscr{P}_\nu(h_n; p'',q'',t; p',q',0)\, g(p',q') \\
&= \int \frac{dp''\,dq''}{2\pi} \int \frac{dp'\,dq'}{2\pi}\, \overline{f(p'',q'')} \\
&\quad \times \mathscr{P}_\nu(h; p'',q'',t; p',q',0)\, g(p',q') \\
&= \langle f, E(\nu,h;t) g\rangle.
\end{aligned}$$

Hence $E(\nu,h;t) = \exp[-(\nu A + iH_V)t]$ for $h\in L^\infty$, which implies again that the $E(\nu,h;t)$ form a strongly continuous contraction semigroup, now for all $h \in L^\infty$.

Finally, let us take a general function h satisfying (2.25). Let $h_n(p,q)$ be defined as

$$\begin{aligned}
&h_n(p,q) = h(p,q), \quad \text{if } |h(p,q)| \leqslant n, \\
&h_n(p,q) = 0, \quad \text{otherwise.}
\end{aligned} \tag{2.35}$$

Clearly $\lim_{n\to\infty} h_n(p,q) = h(p,q)$ (h_n converges pointwise to h, a.e.), while $|h_n(p,q)| \leqslant |h(p,q)|$ for all p,q. By the dominated convergence theorem we have therefore, for every path $[p(t),q(t)]$ for which $\int_0^T |h(p,q)|dt$ is finite, that

$$\lim_{n\to\infty} \int_0^T h_n(p,q)dt = \int_0^T h(p,q)dt. \tag{2.36}$$

Since $\int_0^T |h(p,q)|dt$ is finite a.e. with respect to μ_W^ν (see above), (2.36) implies, again by the dominated convergence theorem, that for all p',q',p'',q''

$$\begin{aligned}
&\lim_{n\to\infty} \mathscr{P}_\nu(h_n; p'',q'',T; p',q',0) \\
&\quad = \mathscr{P}_\nu(h; p'',q'',T; p',q',0).
\end{aligned} \tag{2.37}$$

Take now any $f\in L^2(V)$. We have then

$$\begin{aligned}
&\|[E(\nu,h;t) - E(\nu,h_n;t)]f\|^2 \\
&= \int \frac{dp\,dq}{2\pi} \int \frac{dp_1\,dq_1}{2\pi} \int \frac{dp_2\,dq_2}{2\pi}\, \overline{f(p_1,q_1)} \\
&\quad \times \overline{[\mathscr{P}_\nu(h; p,q,t; p_1,q_1,0) - \mathscr{P}_\nu(h_n; p,q,t; p_1,q_1,0)]} \\
&\quad \times [\mathscr{P}_\nu(h; p,q,t; p_2,q_2,0) - \mathscr{P}_\nu(h_n; p,q,t; p_2,q_2,0)] \\
&\quad \times f(p_2,q_2).
\end{aligned}$$

Using the pointwise convergence (2.37) and the upper bound (2.28) one sees that this integral converges to zero for $n\to\infty$, by the dominated convergence theorem. Hence

$$\operatorname*{s-lim}_{n\to\infty} E(\nu,h_n;t) = E(\nu,h;t). \tag{2.38}$$

Since $h_n \in L^\infty$ for all n, the $E(\nu,h_n;t)$ are contraction operators. Hence the strong convergence (2.38) implies

$$\|E(\nu,h;t)\| \leqslant 1. \tag{2.39}$$

Taking (2.31) into account, we therefore only need to prove still that

$$\operatorname*{s-lim}_{t\to 0} E(\nu,h;t) = 1 \tag{2.40}$$

in order to conclude that the $E(\nu,h;t)$ form a strongly continuous contraction semigroup. Since the $E(\nu,h;t)$ are uniformly bounded, and since D, the set of finite linear combinations of the h_{kl}, is dense, it is sufficient to prove, for all k,l,

$$\lim_{t\to 0} \|E(\nu,h;t)h_{kl} - h_{kl}\| = 0. \tag{2.41}$$

To prove (2.41) we shall again use the L^∞ functions h_n defined by (2.35). Since $h_n \in L^\infty$, we know that $E(\nu,h_n;t) = \exp[-(\nu A + iH_{V,n})t]$. Fix k,l. Since $h_{kl} \in D(A) = D(\nu A + iH_{V,n})$, $G_n(t) \equiv E(\nu,h_n;t)h_{kl}$ is differentiable, and

$$\begin{aligned}
\left\| \frac{d}{dt} G_n(t) \right\| &= \|E(\nu,h_n;t)[\nu A + iH_{V,n}]h_{kl}\| \\
&\leqslant \nu k + \| |h_n| h_{kl}\| \leqslant \nu k + \| |h| h_{kl}\|,
\end{aligned} \tag{2.42}$$

where we have used $\|E(\nu,h_n;t)\| \leqslant 1$, $Ah_{kl} = kh_{kl}$, and $|h_n(p,q)| \leqslant |h(p,q)|$. The fact that the upper bound (2.42) on $\|(d/dt)G_n(t)\|$ is independent of n implies that the G_n form an equicontinuous family of vector-valued functions of t. Since $G_n(t)$ converges to $E(\nu,h;t)h_{kl}$ for every t, the equicontinuity of the G_n implies that $E(\nu,h;t)h_{kl}$ is continuous in t. We have thus proved (2.41), and hence (2.40).

We have now achieved our goal, i.e., we have shown that for all functions h satisfying (2.25), $\mathscr{P}_\nu(h; p'',q'',t; p',q',0)$ is the integral kernel for a strongly continuous contraction semigroup $E(\nu,h;t)$. This contraction semigroup can be considered as a "perturbation" of the semigroup $E^0(\nu,T) = \exp(-\nu At)$. If h is a bounded function, the multiplication operator H_ν is bounded, and one sees from (2.34) that $E(\nu,h;t)$ satisfies the integral equation

$$E(\nu,h;T) = E^0(\nu,T) - i\int_0^T ds\, E(\nu,h;T-s)H_V E^0(\nu,s). \tag{2.43}$$

For a general unbounded h, (2.43) cannot be written as an operator equation, because of domain problems. However, we shall see that (2.43) is still true in a "weak" sense for functions h satisfying (2.25).

Take any (real) function h satisfying (2.25). Let $[p(t),q(t)]$ be a path in the support of μ_W^ν for which $\int_0^T |h(p,q)|dt$ is finite. Then $F(s) = \exp[-i\int_s^T h(p,q)dt]$ is a function of bounded variation on the interval $[0,T]$. Hence, F is differentiable a.e. and by the fundamental theorem of calculus

$$F'(s) = ih(p(s),q(s))F(s), \quad \text{a.e.}$$

Hence

$$F(T) = F(0) + i\int_0^T ds\, h(p(s),q(s))F(s),$$

or

$$\exp\left[-i\int_0^T h(p,q)dt\right] = 1 - i\int_0^T ds\, h(p(s),q(s))\exp\left[-i\int_s^T h(p,q)dt\right].$$

Since for h satisfying (2.25), $\int_0^T|h(p,q)|dt$ is finite a.e. with respect to μ_W^ν, we can insert the above expression into the definition (2.15) of $\mathscr{P}_\nu(h)$, which gives

$$\begin{aligned}\mathscr{P}_\nu(h;p'',q'',T;p',q',0) &= \mathscr{P}_\nu(h=0;p'',q'',T;p',q',0) - i2\pi e^{\nu T/2}\\ &\times\int\left\{\int_0^T ds\, h(p(s),q(s))\exp\left[i\frac{1}{2}\int(p\,dq - q\,dp)\right]\right.\\ &\left.\times\exp\left[-i\int_s^T h(p,q)dt\right]\right\} d\mu_{W;p',q',0}^{\nu;p'',q'',T}(p,q). \qquad (2.44)\end{aligned}$$

In order to avoid confusion we have, for this computation only, explicitly labeled the Brownian bridge measure μ_W^ν by its initial and final times, together with the pinned values of p,q at these times. For every s in $(0,T)$ we can write

$$d\mu_{W;p',q',0}^{\nu;p'',q'',T} = \int dp\,dq\, d\mu_{W;p,q,s}^{\nu;p'',q'',T} \otimes d\mu_{W;p',q',0}^{\nu;p,q,s}.$$

We insert this into (2.44). The multiple integral we thus obtain is absolutely convergent if h satisfies (2.25) [see (2.26)]. We are therefore allowed to change the order of the integrations, which yields

$$\begin{aligned}\mathscr{P}_\nu(h;p'',q'',T;p',q',0) &= \mathscr{P}_\nu(h=0;p'',q'',T;p',q',0)\\ &- i\int_0^T ds\int\frac{dp\,dq}{2\pi}\,\mathscr{P}_\nu(h;p'',q'',T;p,q,s)\\ &\times h(p,q)\mathscr{P}_\nu(h=0;p,q,s;p',q',0). \qquad (2.45)\end{aligned}$$

For $f_1,f_2\in D$, we multiply this expression by $\overline{f_1(p'',q'')}f_2(p',q')$, and integrate over p'',q'',p',q'. We can use the fact that for some C,n, $|f_j(p,q)| < C(1+p^2+q^2)^n\exp[-(p^2+q^2)/4]$ $(j=1,2)$ to show that the multiple integral converges absolutely; we may therefore again change the order of the integrations, which leads, for all $f_1,f_2\in D$, to

$$\langle f_1,E(\nu,h;T)f_2\rangle = \langle f_1,E^o(\nu,T)f_2\rangle - i\int_0^T ds\langle f_1,E(\nu,h;T-s)H_V E^o(\nu,s)f_2\rangle. \qquad (2.46)$$

Formula (2.46) holds for *all* functions h satisfying (2.25). Note that $E^o(\nu,t)$ leaves D invariant; since $D\subset D(H_V)$, all the terms in (2.46) are well defined.

Putting together all the preceding results, we see that we have proved the following proposition.

Proposition 2.1: Let h be a real function satisfying, for all $\alpha>0$,

$$\int dp\,dq|h(p,q)|^2\exp[-\alpha(p^2+q^2)] < \infty.$$

Then

$$\begin{aligned}\mathscr{P}_\nu(h;p'',q'',T;p',q',0) &= 2\pi e^{\nu T/2}\int\exp\left[i\frac{1}{2}\int(p\,dq - q\,dp)\right.\\ &\left.- i\int h(p,q)dt\right]d\mu_W^\nu\end{aligned}$$

is well defined. Here μ_W^ν is a Gaussian measure completely determined by its normalization (1.4) and its connected covariance (1.5). Moreover, there exists a strongly continuous contraction semigroup $E(\nu,h;t)$ on $L^2(V)$ such that we have the following.

(1) $E(\nu,h;t)$ is an integral operator, with kernel $[E(\nu,h;t)](p'',q'';p',q') = \mathscr{P}_\nu(h;p'',q'',t;p',q',0)$.

(2) For all $f,g\in D$,

$$\langle f,E(\nu,h;T)g\rangle = \langle f,E^o(\nu,T)g\rangle - i\int_0^T dt\,\langle f,E(\nu,h;T-t)H_V E^o(\nu,t)g\rangle.$$

This proposition will enable us, in the next section, to study the limit for $\nu\to\infty$.

From the path integral definition (2.15) for $\mathscr{P}_\nu(h)$ one can easily check that

$$\mathscr{P}_\nu(-h;p'',q'',T;p',q',0) = \overline{\mathscr{P}_\nu(h;p',q',T;p'',q'',0)}. \qquad (2.47a)$$

This implies

$$E(\nu,-h;t) = E(\nu,h;t)^*. \qquad (2.47b)$$

C. Taking the limit $\nu\to\infty$

Let us again first consider the case $h=0$. Since the h_{ol} span the subspace $\mathscr{H}_o$, we see from (2.24) that $e^{-\nu TA}\to P_o$ as $\nu\to\infty$. Hence, as $\nu\to\infty$,

$$\begin{aligned}\mathscr{P}_\nu(h=0;p'',q'',T;p',q',0) &= [\exp(-\nu AT)](p'',q'';p',q')\to P_o(p'',q'';p',q').\end{aligned} \qquad (2.48)$$

Since $P_o = (\mathbf{1}_{\mathscr{H}})\widehat{\ }$ [using the definition (2.10) of $\widehat{R}$ for $R\in\mathscr{B}(\mathscr{H})$], its integral kernel is given by [use (2.11)]

$$P_o(p'',q'';p',q') = \langle p'',q''|p',q'\rangle. \qquad (2.49)$$

Putting (2.48) and (2.49) together yields, as $\nu\to\infty$,

$$\mathscr{P}_\nu(h=0;p'',q'',T;p',q',0)\to\langle p'',q''|p',q'\rangle.$$

This is exactly statement (1.3), specialized to the case $h\equiv 0$.

For $h \neq 0$, the same will happen. If h is bounded, $\mathscr{P}_\nu(h)$ is the integral kernel of $\exp[-(\nu A + iH_V)T]$. Since $A(1-P_o) \geqslant 1 - P_o$, the effect of the $-\nu TA$ term in the exponent, in the limit $\nu \to \infty$, is that everything happening outside $\mathscr{H}_o = P_o L^2(V)$ gets damped out. An analogous phenomenon takes place for unbounded h. This is the content of the following proposition.

Proposition 2.2: Let h be a real function on $\mathbb{R}^2$ satisfying (2.25). Let H_V be the operator on $L^2(V)$ defined by

$$(H_V f)(p,q) = h(p,q) f(p,q).$$

Let $E(\nu,h;t)$ be the contraction semigroup given by Proposition 2.1. Define the operator $P_o H_V P_o$ on the domain $\{f; P_o f \in D(H_V)\}$. Obviously $D \subset D(P_o H_V P_o)$. Assume that $P_o H_V P_o$ is essentially self-adjoint on D. Then, for all $T > 0$,

$$\operatorname*{s-lim}_{\nu \to \infty} E(\nu,h;T) = P_o \exp(-iP_o H_V P_o T) P_o, \tag{2.50}$$

where, with a slight abuse of notation, we write $\exp(-iP_o H_V P_o T)$ for $\exp(-i\,\overline{P_o H_V P_o}\,T)$.

Remark: Note that the condition that $\overline{P_o H_V P_o|_D}$ be self-adjoint is an extra condition on h, which is not fulfilled by all h satisfying (2.25), not even if H_V is essentially self-adjoint on D, and $P_o D \subset D$ notwithstanding. It may happen that $P_o H_V P_o|_D$ has more than one self-adjoint extension [e.g., $h(p,q) = p^2 + (1-3\lambda)q^2 - \lambda q^4, \lambda > 0$] or none at all [e.g., $h(p,q) = pq^3 + \frac{3}{2}pq$]. The condition that $\overline{P_o H_V P_o|_D}$ be self-adjoint ensures that $\exp[-i\,\overline{P_o H_V P_o|_D}\,T]$ is well defined and unitary. This is needed in point (7) of the proof (see below). The condition on $P_o H_V P_o|_D$ may be weakened, however (see the remark following the proof of Proposition 2.1); it is sufficient to require that $\overline{P_o H_V P_o|_D}$ be maximal symmetric. In that case a slightly weaker conclusion than (2.50) holds: in some cases one has to substitute "weak limit" for "strong limit." Note that if $\overline{P_o H_V P_o|_D}$ is maximal symmetric (this includes the self-adjoint case), we automatically have $\overline{P_o H_V P_o|_D} = \overline{P_o H_V P_o}$.

Before starting on the proof of Proposition 2.2, we state the following lemma which we shall need. Since it is easy to prove, we omit the proof here (see also Ref. 13, Lemma V 1.2).

Lemma 2.3: Let $\mathscr{H}$ be any separable complex Hilbert space. Let B_n be a sequence of bounded operators on $\mathscr{H}$, with $\text{w-lim}_{n\to\infty} B_n = B$. suppose that

$\|B_n\| \leqslant 1$, all n,

$\|B\psi\| = \|\psi\|$, all $\psi \in \mathscr{H}$.

Then the B_n converge strongly to B.

Remark: A corollary to this lemma is that the integral in (2.6) actually converges in the strong sense; that is, for any increasing sequence of compact sets K_n ("increasing" means $K_n \subset K_{n+1}$ for all n) such that $\cup_n K_n = \mathbb{R}^2$, one has

$$\operatorname*{s-lim}_{n\to\infty} \int_{K_n} \frac{dp\,dq}{2\pi} |p,q\rangle\langle p,q| = 1$$

[take $B_n = \int_{K_n}(dp\,dq/2\pi)|p,q\rangle\langle p,q|, B = 1$. Since $\|B_n\| \leqslant 1$, all the conditions for the lemma are fulfilled, and the above statement follows].

We now proceed to prove proposition 2.2.

Proof of Proposition 2.2: (1) Since we shall work with one fixed h, we shall drop this label in our notation for $E(\nu,h;t)$:

$E(\nu,t) = E(\nu,h;t)$, $\quad E^o(\nu,t) = E(\nu,0;t) = \exp(-\nu t A)$.

Since the $E(\nu,t)$ are contractions, $\|E(\nu,t)\| \leqslant 1$, it suffices to prove the strong convergence on D. For $f, g \in D$, we have [see (2.46)], for all $t \geqslant 0$,

$$\langle f, E(\nu,t) g\rangle = \langle f, E^o(\nu,t) g\rangle - i\int_0^t ds \langle f, E(\nu,t-s) H_V E^o(\nu,s) g\rangle. \tag{2.51}$$

Hence, for $k > 0$,

$$\begin{aligned} |\langle f, E(\nu,t) h_{kl}\rangle| &\leqslant |\langle f, E^o(\nu,t) h_{kl}\rangle| \\ &\quad + \int_0^t ds |\langle f, E(\nu,t-s) H_V E^o(\nu,s) h_{kl}\rangle| \\ &\leqslant e^{-\nu k t}\|f\| + (\nu^{-1}k^{-1})\|f\|\,\|H_V h_{kl}\|. \end{aligned}$$

Since D is dense, this implies

$$\|E(\nu,t) h_{kl}\| \leqslant e^{-\nu k t} + (\nu^{-1}k^{-1})\|H_V h_{kl}\|. \tag{2.52}$$

Hence, for $k > 0$: $\|E(\nu,T) h_{kl}\| \to 0$ as $\nu \to \infty$ (since $T > 0$), which proves

$$\operatorname*{s-lim}_{\nu\to\infty} [E(\nu,T)(1-P_o)] = 0. \tag{2.53}$$

(2) We can use (2.52) to prove an estimate that will be useful below. From (2.52) we see that for any $f \in D$ there exists a constant C_f (depending on f) such that, for all $t \geqslant 0$,

$$\|E(\nu,t)(1-P_o)f\| \leqslant C_f(e^{-\nu t} + \nu^{-1}).$$

Take $g_1 \in L^2(V)$. For arbitrary $\epsilon > 0$, we can find $f \in D$ such that $\|f - g_1\| \leqslant \epsilon$. Hence

$$\|E(\nu,t)(1-P_o) g_1\| \leqslant \epsilon + C_f(e^{-\nu t} + \nu^{-1}).$$

This implies, for $g_2 \in L^2(V)$, that

$$\begin{aligned} &\int_0^T dt\, |\langle g_2, E(\nu,T-t)(1-P_o) g_1\rangle| \\ &\quad \leqslant \epsilon T \|g_2\| + C_f \|g_2\| [\nu^{-1}T + \nu^{-1}(1-e^{-\nu T})]. \end{aligned}$$

It is always possible to choose ν_o such that for $\nu \geqslant \nu_o$ the second term in the right-hand side of this inequality becomes smaller than ϵ. Since ϵ was arbitrary to start with, we have therefore proved for all $g_1, g_2 \in L^2(V)$ that

$$\lim_{\nu\to\infty} \int_0^T dt\, \langle g_2, E(\nu,T-t)(1-P_o) g_1\rangle = 0. \tag{2.54}$$

(3) We now concentrate on $P_o E(\nu,T) P_o$. Take any strictly increasing sequence $(\nu_n)_n, \nu_n \geqslant 0$, with $\lim_{n\to\infty} \nu_n = \infty$. Fix $t \geqslant 0$. Since $\|P_o E(\nu_n,t) P_o\| \leqslant 1$, there exists a subsequence $(\nu_{n(k)})_k$ of $(\nu_n)_n$ such that $P_o E(\nu_{n(k)},t) P_o$ converges weakly. By the standard diagonalization trick, one finds there exists a subsequence $(\hat{\nu}_k)_k$ of $(\nu_n)_n$ such that the $P_o E(\hat{\nu}_k,t) P_o$ converge weakly for all rational values of t in $\mathbb{R}_+$.

(4) For $f, g \in D_o = P_o D$ (= finite linear span of the h_{ol}), and any $t \in \mathbb{R}_+$, we see from (2.51) that [use $E^o(\nu,t)g = g$]

$$\left| \frac{d}{dt} \langle f, E(\hat{\nu}_k,t)g\rangle \right| = | -i\langle f, E(\hat{\nu}_k,t)H_V g\rangle | \leqslant \|f\| \, \|H_V g\|.$$

Since this bound is independent of k (and of t) this implies that for fixed $f, g \in D_o$, the functions $F_{k,f,g}(t) = \langle f, E(\hat{\nu}_k,t)g\rangle$ form a (uniformly) equicontinuous family of functions $\mathbb{R}_+ \to \mathbb{C}$. Since the F_k converge for $t \in \mathbb{Q}_+$, a dense set in $\mathbb{R}_+$, this implies that they converge for all $t \in \mathbb{R}_+$; namely, for all $t > 0$, $\langle f, E(\hat{\nu}_n,t)g\rangle \to F_{\infty,f,g}(t)$ as $\nu \to \infty$.

(5) for any t, $F_{\infty,f,g}$ is obviously sesquilinear in f, g. Moreover, since the $E(\nu,t)$ are all contractions, we have $|F_{\infty,f,g}(t)| \leqslant \|f\| \, \|g\|$. By Riesz' lemma this implies that there exist operators $L_t \in \mathscr{B}(\mathscr{H}_o)$ such that, for all $f, g \in \mathscr{H}_o$ and all $t \in \mathbb{R}_+$,

$$\langle f, L_t g\rangle = F_{\infty,f,g}(t) = \lim_{k\to\infty} \langle f, E(\hat{\nu}_k,t)g\rangle. \tag{2.55}$$

(6) Putting together (2.51) with some estimates, we can find an explicit form for the operators L_t. Let f, g be arbitrary elements of D_o. Then

$$\begin{aligned}
\langle f, L_t g\rangle
&= \lim_{k\to\infty} \langle f, E(\hat{\nu}_k,t)g\rangle \quad [\text{see (2.55)}] \\
&= \lim_{k\to\infty} \left[\langle f, g\rangle - i\int_0^T dt\, \langle f, E(\hat{\nu}_k, T-t)H_V g\rangle \right] \\
&\qquad [\text{use (2.51), together with } E^o(\nu,t)g = g] \\
&= \langle f, g\rangle - i \lim_{k\to\infty} \int_0^T dt\, \langle f, E(\hat{\nu}_k, T-t)(1-P_o)H_V g\rangle \\
&\quad - i \lim_{k\to\infty} \int_0^T dt\, \langle f, E(\hat{\nu}_k, T-t)P_o H_V g\rangle.
\end{aligned}$$

The second term is zero by (2.54); in the third term we can interchange the limit and the integration because of the dominated convergence theorem, which gives

$$\langle f, L_T g\rangle = \langle f, g\rangle - i\int_0^T dt\, \langle f, L_{T-t} P_o H_V g\rangle. \tag{2.56}$$

Equation (2.56) holds for $f, g \in D_o$. Introducing the operator $\hat{L}_t$ on $L^2(V)$, defined as the trivial extension of L_t

$$\hat{L}_t f = 0, \quad \text{if } f \perp \mathscr{H}_o,$$

$$\hat{L}_t f = L_t f, \quad \text{if } f \in \mathscr{H}_o,$$

we can rewrite (2.56), for all $f, g \in D$ and all $T > 0$, as

$$\langle f, \hat{L}_T g\rangle = \langle f, P_o g\rangle - i\int_0^T dt\, \langle f, \hat{L}_{T-t}\, P_o H_V P_o g\rangle. \tag{2.57}$$

Since $P_o H_V P_o$ is essentially self-adjoint on D, (2.57) implies for all $T > 0$ that

$$\hat{L}_T = P_o \exp[-iP_o H_V P_o T]P_o. \tag{2.58}$$

From (2.55) and the definition of $\hat{L}_T$, one sees that this implies

$$\operatorname*{w-lim}_{k\to\infty} P_o E(\hat{\nu}_k,T)P_o = P_o \exp[-iP_o H_V P_o T]P_o. \tag{2.59}$$

Since $(\hat{\nu}_k)_k$ was a well-chosen subsequence of an *arbitrary* increasing sequence $(\nu_n)_n$, with $\lim_{n\to\infty} \nu_n = \infty$, (2.59) implies

$$\operatorname*{w-lim}_{\nu\to\infty} P_o E(\nu,T)P_o = P_o \exp[-iP_o H_V P_o T]P_o. \tag{2.60}$$

(7) Equation (2.60) is still not quite what we want, since it only gives weak convergence, while we are interested in strong convergence. However, we shall see that we can, by applying Lemma 2.3, convert this weak into strong convergence. Let us restrict ourselves to $\mathscr{H}_o$. If we define $\widetilde{E}(\nu,t)$, $\widetilde{H}_V$ to be the obvious restrictions to $\mathscr{H}_o$ of $P_o E(\nu,t)P_o$, $P_o H_V P_o$, respectively, (2.60) can be rewritten as

$$\operatorname*{w-lim}_{\nu\to\infty} \widetilde{E}(\nu,T) = \exp(-i\widetilde{H}_V T).$$

Since $\exp(-i\widetilde{H}_V T)$ is a unitary operator on $\mathscr{H}_o$, and $\|\widetilde{E}(\nu,T)\| \leqslant 1$, we can apply Lemma 2.3, and conclude

$$\operatorname*{s-lim}_{\nu\to\infty} \widetilde{E}(\nu,T) = \exp(-i\widetilde{H}_V T).$$

This then implies, on $L^2(V)$,

$$\operatorname*{s-lim}_{\nu\to\infty} P_o E(\nu,T)P_o = P_o \exp[-iP_o H_V P_o T]P_o. \tag{2.61}$$

(8) Comparing (2.61) and (2.53) with (2.50) one sees that we only need to prove still that $(1-P_o)E(\nu,t)P_o$ converges to zero. This is an easy consequence of the fact that the $E(\nu,t)$ are contractions, while $\exp(-iP_o H_V P_o T)$ is unitary. Take $f \in L^2(V)$. Let ϵ be arbitrary, with $\epsilon > 0$. Because of (2.61) we know that there exists a ν_o such that

$$\|P_o[E(\nu,t) - \exp(-iP_o H_V P_o t)]P_o f\| \leqslant \epsilon, \quad \text{for all } \nu \geqslant \nu_o.$$

Since $P_o \exp(-iP_o H_V P_o t)P_o = \exp(-iP_o H_V P_o t)P_o$, this implies

$$\|e^{-iP_o H_V P_o t}P_o f\|^2 - \|P_o E(\nu,t)P_o f\|^2 \leqslant 2\epsilon\|P_o f\|.$$

Hence, for $\nu \geqslant \nu_o$,

$$\begin{aligned}
&\|(1-P_o)E(\nu,t)P_o f\|^2 \\
&= \|E(\nu,t)P_o f\|^2 - \|P_o E(\nu,t)P_o f\|^2 \\
&\leqslant \|P_o f\|^2 - \|e^{-iP_o H_V P_o t}P_o f\|^2 + 2\epsilon\|P_o f\| \\
&= 2\epsilon\|P_o f\|.
\end{aligned}$$

Since ϵ and f were arbitrary, this proves

$$\operatorname*{s-lim}_{\nu\to\infty} (1-P_o)E(\nu,t)P_o = 0, \tag{2.62}$$

and (2.50) now follows from (2.53), (2.61), and (2.62).

Remark: We already noted above that the condition on $P_o H_V P_o$ may be weakened. We only required that $P_o H_V P_o|_D$ be essentially self-adjoint in order to be allowed to make the transition from the integral equation (2.57) to the "integrated form" (2.58) for $\hat{L}_T$. There are, however, more general conditions under which this transition is still permitted.

We first make some general remarks. Let T be a closed symmetric operator on a complex Hilbert space $\mathscr{H}$. We define its deficiency indices $n_\pm$ as $n_\pm = \dim \operatorname{Ker}(T^* \pm i1)$. Let ξ be any strictly positive real number. One checks easily that $iT + \xi$ is closed, and for all $\phi \in D(T)$,

$$\|(iT+\xi)\phi\| \geqslant \xi \|\phi\|.$$

This implies that $\operatorname{Ran}(iT+\xi)$ is closed. If $\operatorname{Ran}(iT+\xi) = \mathscr{H}$ then $(iT+\xi)^{-1}$ exists, and $\|(iT+\xi)^{-1}\| \leqslant \xi^{-1}$. This is a necessary and sufficient condition for iT to be the generator of a strongly continuous contraction semigroup (this is the Hille–Yosida theorem; see, e.g., Kato[13]); we denote this semigroup by $\exp(-iTt)$. But $[\operatorname{Ran}(iT+\xi)]^\perp = \operatorname{Ker}(-iT^*+\xi)$, hence $\operatorname{Ran}(iT+\xi) = \mathscr{H}$ if and only if $n_+ = 0$. Therefore, iT generates a strongly continuous contraction semigroup if and only if $n_+ = 0$. It turns out, due to the fact that T is symmetric, that this semigroup consists entirely of isometries [for $\phi \in D(T)$, one checks that $(d/dt)\|\exp(-iTt)\phi\|^2 = 0$, hence $\|\exp(-iTt)\phi\| = \|\phi\|$. Since $D(T)$ is dense and $\exp(-iTt)$ a contraction, this extends to all of $\mathscr{H}$]. If $n_- = 0$, the same analysis as above holds for $-T$ instead of T. We have then that all strictly positive real numbers lie in the resolvent set of iT, and $\|(-iT+\xi)^{-1}\| \leqslant \xi^{-1}$ for $\xi > 0$. This implies $\|(iT^*+\xi)^{-1}\| \leqslant \xi^{-1}$, i.e., iT^* is a generator of a strongly continuous contraction semigroup $\exp(-iT^*t)$. In fact, $\exp(-iT^*t) = [\exp(iTt)]^*$.

Let us now specialize this to the case at hand. It is clear that we have to assume at least that $\overline{P_o H_V P_o|_D}$ is maximal symmetric. Hence either $n_+ = 0$ or $n_- = 0$. If $n_+ = 0$, $\overline{P_o H_V P_o|_D}$ is the generator of a contraction semigroup, and we are allowed to conclude (2.58) from (2.57). This then leads to the weak limit statement (2.59). Since $\exp(-iP_o H_V P_o T)$ is still an isometry, Lemma 2.3 can still be applied, and the arguments in points (7) and (8) of the proof of Proposition 2.2 still carry through.

If $n_- = 0$ we can apply the above to $-h$. We have thus

$$\operatorname*{s-lim}_{\nu\to\infty} E(\nu,-h;T) = P_o \exp[iP_o H_V P_o T] P_o.$$

Taking adjoints, we see that this implies [use (2.47)]

$$\operatorname*{w-lim}_{\nu\to\infty} E(\nu,h;T) = P_o \exp[-i(P_o H_V P_o)^* T] P_o,$$

where we have used $(P_o H_V P_o)^* = (P_o H_V P_o|_D)^*$ since $\overline{P_o H_V P_o|_D} = \overline{P_o H_V P_o}$. Note that we only can conclude weak convergence in this case (due to the taking of adjoints).

Summarizing, we see that if $\overline{P_o H_V P_o|_D}$ is maximal symmetric, then

$$\begin{aligned}
&\operatorname*{s-lim}_{\nu\to\infty} P_o E(\nu,h;T) P_o = P_o \exp[-iP_o H_V P_o T] P_o,\\
&\qquad \text{if } n_+(P_o H_V P_o) = 0,\\
&\operatorname*{w-lim}_{\nu\to\infty} P_o E(\nu,h;T) P_o = P_o \exp[-i(P_o H_V P_o)^* T] P_o,\\
&\qquad \text{if } n_-(P_o H_V P_o) = 0.
\end{aligned} \tag{2.63}$$

Ultimately we are interested in convergence of the $\mathscr{P}_\nu(h)$, the integral kernels of the operators $E(\nu,h;t)$, rather than in convergence of the operators themselves. As a consequence of the constructions we made in Sec. II A, we can easily conclude that the $\mathscr{P}_\nu(h)$ converge in the sense of the Schwartz distributions.

To see this, let us first define the operator H on $\mathscr{H}$ by

$$H = \int \frac{dp\,dq}{2\pi} |p,q\rangle h(p,q) \langle p,q|.$$

This definition is consistent with our earlier notations; the operator H_V on $L^2(V)$ associated to H by means of (2.13) coincides with the multiplication operator H_V defined by (2.32). We have, therefore, according to (2.14),

$$P_o H_V P_o = \hat{H}, \tag{2.64}$$

where $\hat{H}$ is defined by (2.10). This implies [use (2.12)]

$$P_o \exp(-iP_o H_V P_o T) P_o = [\exp(-iTH)]\hat{}.$$

According to (2.11), $[\exp-(iTH)]\hat{}$ is an integral operator, and its integral kernel is given by the cs matrix elements of $\exp(-iTH)$. Hence

$$\begin{aligned}
&[P_o \exp(-iP_o H_V P_o T) P_o](p'',q'';p',q')\\
&\quad = \langle p'',q''|\exp(-iHT)|p',q'\rangle.
\end{aligned}$$

We find therefore that both the $E(\nu,h;T)$ *and* their limiting operator $P_o \exp(-iP_o H_V P_o T) P_o$ are integral operators on $L^2(V)$. It is then easy to show, using the fact that the Schwartz test functions on $\mathbb{R}^2$ are elements of $L^2(V)$, that the convergence proved in Proposition 2.2 implies convergence, in the sense of the distributions, of the integral kernels $\mathscr{P}_\nu(h)$, for all $T > 0$, to the integral kernel of the limiting operator. We have thus (d-lim = limit in the sense of the Schwartz distributions)

$$\begin{aligned}
&\operatorname*{d-lim}_{\nu\to\infty} \mathscr{P}_\nu(h;p'',q'',T;p',q',0)\\
&\quad = \langle p'',q''|\exp(-iHT)|p',q'\rangle.
\end{aligned} \tag{2.65}$$

In the final theorem of this section we shall see that we can do better than this, i.e., that we can prove in addition *pointwise* convergence of the $\mathscr{P}_\nu(h)$, provided the function h satisfies a condition slightly stronger than (2.25). To prove this theorem we shall make use of formula (2.45) and of Proposition 2.2. Actually we shall only need weak convergence of the operators $E(\nu,h;t)$; this enables us to consider also operators H which are not self-adjoint, but only maximal symmetric [see (2.63)].

Theorem 2.4: Let H be a maximal symmetric operator on $\mathscr{H}$, which can be written as

$$H = \int \frac{dp\,dq}{2\pi} |p,q\rangle h(p,q) \langle p,q|.$$

Assume that D_c, the finite linear span of the harmonic oscillator eigenstates ω_k, is a core for H. Suppose that the function h satisfies the following.

(C1) For every $\alpha > 0$

$$\int dp\,dq |h(p,q)|^2 \exp[-\alpha(p^2+q^2)] < \infty.$$

(C2) For some $0 \leqslant \beta < 1$

$$\int dp\,dq |h(p,q)|^4 \exp\left[-\frac{\beta(p^2+q^2)}{2}\right] = C_\beta < \infty.$$

Then, for all p'',q'',p',q' in $\mathbb{R}$, and all $t'' > t'$,

$$\lim_{\nu\to\infty} 2\pi e^{\nu(t''-t')/2} \int \exp\left[\frac{i}{2}\int (p\,dq - q\,dp) - i\int h(p,q)dt\right] d\mu_W^\nu(p,q)$$

$$= \begin{cases} \langle p'',q''|\exp[-i(t''-t')H]\,p',q'\rangle, \\ \quad \text{if } n_+(H)=0, \\ \langle p'',q''|\exp[-i(t''-t')H^*]\,p',q'\rangle, \\ \quad \text{if } n_-(H)=0. \end{cases} \tag{2.66}$$

Here μ_W^ν is the product of two independent Wiener measures (one in p, one in q), pinned at p',q' for $t=t'$, and at p'',q'' for $t=t''$. The normalization of μ_W^ν is given by

$$\int d\mu_W^\nu(p,q) = [2\pi\nu(t''-t')]^{-1} \times\exp\left\{-\frac{(p''-p')^2+(q''-q')^2}{2\nu(t''-t')}\right\}$$

and the connected covariance is (x either p or q; $t_1<t_2$)

$$\langle x(t_1)x(t_2)\rangle^c \equiv \langle x(t_1)x(t_2)\rangle - \langle x(t_1)\rangle\langle x(t_2)\rangle = \nu t_1[1-t_2/(t''-t')],$$

where $\langle f\rangle = (\int d\mu_W^\nu f)/(\int d\mu_W^\nu)$. If the limit is taken in the sense of the Schwartz distributions, then (2.66) already holds if only (C1) is satisfied.

Proof: (1) We take, without loss of generality, $t'=0$ and $t''=T>0$.

(2) We shall use (2.45), relating $\mathscr{P}_\nu(h)$ with $\mathscr{P}_\nu(h=0)$. If we write (2.45) also for $\mathscr{P}_\nu(-h)$, take the complex conjugate, and apply (2.47a), we find another such integral equation for $\mathscr{P}_\nu(h)$. Combining this with (2.45) leads to

$$\begin{aligned}\mathscr{P}_\nu(h;p'',q'',T;p',q',0) &= \mathscr{P}_\nu(0;p'',q'',T;p',q',0) - i\int_0^T dt \int \frac{dp\,dq}{2\pi} \\ &\quad\times \mathscr{P}_\nu(0;p'',q'',T;p,q,t)h(p,q)\mathscr{P}_\nu(0;p,q,t;p',q',0) \\ &\quad - \int_0^T dt_1 \int_0^{t_1} dt_2 \int \frac{dp_1\,dq_1}{2\pi}\int\frac{dp_2\,dq_2}{2\pi} \\ &\quad\times \mathscr{P}_\nu(0;p'',q'',T;p_1,q_1,t_1)h(p_1,q_1) \\ &\quad\times \mathscr{P}_\nu(h;p_1,q_1,t_1;p_2,q_2,t_2)h(p_2,q_2) \\ &\quad\times \mathscr{P}_\nu(0;p_2,q_2,t_2;p',q',0).\end{aligned} \tag{2.67}$$

A calculation analogous to what was done above [see (2.26)] shows that all the integrals in (2.67) converge absolutely (for fixed ν).

(3) Let us introduce a new notation. For $p,q\in\mathbb{R},\nu,t>0$, we define $\phi_{p,q,\nu,t}\in L^2(V)$ by

$$\begin{aligned}\phi_{p,q,\nu,t}(p_1,q_1) &= \mathscr{P}_\nu(0;p_1,q_1,t;p,q,0) \\ &= (e^{\nu t/2}/(2\sinh[\nu t/2]))\exp\{(i/2)(pq_1-p_1q) \\ &\quad - \tfrac14\coth(\nu t/2)[(p-p_1)^2+(q-q_1)^2]\}.\end{aligned}$$

One easily calculates

$$\|\phi_{p,q,\nu,t}\| = (1-e^{-2\nu t})^{-1/2}.$$

Using (C1), one can check that $\phi_{p,q,\nu,t}\in D(H_V)$. As ν tends to ∞ (the other parameters remaining fixed), $\phi_{p,q,\nu,t}(p_1,q_1)$ converges pointwise to a familiar expression

$$\begin{aligned}\phi_{p,q,\nu,t}(p_1,q_1) &\to \exp\{(i/2)(pq_1-p_1q) - \tfrac14[(p-p_1)^2+(q-q_1)^2]\} \\ &= \langle p_1,q_1|p,q\rangle = (\hat U\omega^{p,q})(p_1,q_1),\end{aligned}$$

where we have used the notation of Sec. II A. An easy calculation shows that this convergence also holds in $L^2(V)$:

$$\|\phi_{p,q,\nu,t} - \hat U\omega^{p,q}\| = (e^{2\nu t}-1)^{-1/2}. \tag{2.68}$$

(4) With this new notation we can rewrite (2.67) as

$$\begin{aligned}\mathscr{P}_\nu(h;p'',q'',T;p',q',0) &= \mathscr{P}_\nu(0;p'',q'',T;p',q',0) \\ &\quad - i\int_0^T dt\,\langle\phi_{p'',q'',\nu,T-t}, H_V\phi_{p',q',\nu,t}\rangle \\ &\quad - \int_0^T dt_1\int_0^{t_1} dt_2 \langle H_V\phi_{p'',q'',\nu,T-t_1}, E(\nu,h;t_1-t_2) \\ &\qquad\times H_V\phi_{p',q',\nu,t_2}\rangle\end{aligned} \tag{2.69}$$

(5) One can derive a similar integral equation for $\langle p'',q''|\exp(-iTH)\,p',q'\rangle$ [we assume $n_+(H)=0$; a similar derivation can be made if $n_-(H)=0$]. Since (C1) ensures that the $\omega^{p,q}$ lie in the domain of H, one can write

$$\begin{aligned}\langle p'',q''|\exp(-iTH)p',q'\rangle &= \langle p'',q''|p',q'\rangle - i\int_0^T dt\,\langle p'',q''|\exp(-itH)H\,p',q'\rangle \\ &= \langle p'',q''|p',q'\rangle - i\int_0^T dt\,\langle p'',q''|H\,p',q'\rangle \\ &\quad - \int_0^T dt_1\int_0^{t_1} dt_2\langle p'',q''|H\exp[-i(t_1-t_2)H]\,H\,p',q'\rangle.\end{aligned} \tag{2.70}$$

Transposing the inner products in (2.70) to $L^2(V)$ by means of $\hat U$, and subtracting the resulting equation from (2.69), we obtain

$$\begin{aligned}&\mathscr{P}_\nu(h;p'',q'',T;p',q',0) - \langle p'',q''|\exp(-iTH)\,p',q'\rangle \\ &\quad= [\mathscr{P}_\nu(h=0;p'',q'',T;p',q',0) - \langle p'',q''|p',q'\rangle] \\ &\qquad - i\int_0^T dt\,\langle\phi_{p'',q'',\nu,T-t}, H_V(\phi_{p',q',\nu,t} - \hat U\omega^{p',q'})\rangle \\ &\qquad - i\int_0^T dt\,\langle\phi_{p'',q'',\nu,T-t} - \hat U\omega^{p'',q''}, H_V\hat U\omega^{p',q'}\rangle \\ &\qquad - i\int_0^T dt_1\int_0^{t_1} dt_2\langle H_V\phi_{p'',q'',\nu,T-t_1}, \\ &\qquad\qquad E(\nu,h;t_1-t_2)H_V(\phi_{p',q',\nu,t_2} - \hat U\omega^{p',q'})\rangle \\ &\qquad - i\int_0^T dt_1\int_0^{t_1} dt_2\langle H_V(\phi_{p'',q'',\nu,T-t_1} - \hat U\omega^{p'',q''}),\end{aligned}$$

$$E(\nu,h;t_1-t_2)H_V\hat{U}\omega^{p',q'}\rangle$$

$$-\int_0^T dt_1\int_0^{t_1} dt_2\langle H_V\hat{U}\omega^{p'',q''},\{E(\nu,h;t_1-t_2)$$

$$-P_o\exp[-iP_oH_VP_o(t_1-t_2)]P_o\}H_V\hat{U}\omega^{p',q'}\rangle.$$

Let us denote these six terms by $\Delta_1,\ldots,\Delta_6$ (in the above order). We shall see that each $\Delta_j\to 0$ as $\nu\to\infty$, which proves the theorem.

(6) Using the explicit expression (2.16) for $\mathscr{P}_\nu(h=0)$, one easily finds

$$|\Delta_1|<(e^{\nu t}-1)^{-1}[1+\tfrac{1}{2}(p''-p')^2+\tfrac{1}{2}(q''-q')^2]$$
$$\times\exp[-(p''-p')^2/4-(q''-q')^2/4],$$

hence $\Delta_1\to 0$.

(7) For Δ_3 we can use (2.68) and Cauchy–Schwarz, which leads to

$$|\Delta_3|<\int_0^T dt\,(e^{2\nu t}-1)^{-1/2}\|H_V\hat{U}\omega^{p,q}\|$$

$$<\nu^{-1}\|H_V\hat{U}\omega^{p,q}\|\int_0^\infty ds(e^s-1)^{-1/2}.$$

This implies $\Delta_3\to 0$.

(8) Let us now consider Δ_6. This term is the integral, on a bounded domain, of a function uniformly bounded by $2\|H_V\hat{U}\omega^{p'',q''}\|\cdot\|H_V\hat{U}\omega^{p',q'}\|$, and converging pointwise a.e. to zero for ν tending to ∞ [by (2.63)]. Hence, by the dominated convergence theorem, $\Delta_6\to 0$.

(9) For the remaining three terms Δ_2, Δ_4, and Δ_5 we need estimates of $\|H_V\phi_{p,q,\nu,t}\|$ and $\|H_V(\phi_{p,q,\nu,t}-\hat{U}\omega^{p,q})\|$, which we shall compute using (C2). We have

$$\|H_V\phi_{p',q',\nu,t}\|^2$$
$$=\int\frac{dp\,dq}{2\pi}|h(p,q)|^2\frac{e^{\nu t}}{4\sinh^2(\nu t/2)}$$
$$\times\exp\{-\tfrac{1}{2}\coth(\nu t/2)[(p-p')+(q-q')^2]\}$$
$$<\left\{\int\frac{dp\,dq}{2\pi}|h(p,q)|^4\exp\left[-\frac{\beta(p^2+q^2)}{2}\right]\right\}^{1/2}$$
$$\times e^{\nu t}\left[4\sinh^2\frac{\nu t}{2}\right]^{-1}\exp\left[\frac{\beta(p'^2+q'^2)}{2}\right]$$
$$\times\left\{\int\frac{dp\,dq}{2\pi}\exp\left[-\left(\coth\frac{\nu t}{2}-\beta\right)(p^2+q^2)\right]\right\}^{1/2}$$
$$=\frac{1}{2}\left(\frac{C_\beta}{\pi}\right)^{1/2}e^{\nu t}\left[4\sinh^2\frac{\nu t}{2}\right]^{-1}\left(\coth\frac{\nu t}{2}-\beta\right)^{-1/2}$$
$$\times\exp[\beta(p'^2+q'^2)/2]$$
$$<K\exp[\beta(p'^2+q'^2)/2](1-e^{-\nu t})^{-3/2}.$$

Hence

$$\|H_V\phi_{p',q',\nu,t}\|<K_1\exp[\beta(p'^2+q'^2)/4](1-e^{-\nu t})^{-3/4}.$$

An analogous calculation can be made for $\|H_V\phi_{p',q',\nu,t}-\hat{U}\omega^{p',q'})\|$. Putting $y=e^{\nu t}-1$, $a=2(1-\beta)<2$, one finds

$$\|H_V(\phi_{p',q',\nu,t}-\hat{U}\omega^{p',q'})\|^2<\tfrac{1}{2}(C_\beta/\pi)^{1/2}\exp[\beta(p'^2+q'^2)/2]\{y^{-3}a^{-1}(1+ay)^{-1}$$
$$\times(2+ay)^{-1}(3+ay)^{-1}(4+ay)^{-1}[(a^4-4a^3+12a^2-24a+24)y^3$$
$$+6a(a^2-2a+2)y^2+6a(11a-8)y+6a]\}^{1/2}$$
$$\leqslant K'\exp[\beta(p'^2+q'^2)/2]\,y^{-3/2}(1+y)^{-1/2}.$$

Hence

$$\|H_V(\phi_{p',q',\nu,t}-\hat{U}\omega^{p',q'})\|<K_2\exp[\beta(p'^2+q'^2)/4](e^{\nu t}-1)^{-3/4}.$$

(10) With the help of these two estimates we can now discuss Δ_2, Δ_4, and Δ_5. We give here the explicit estimate for Δ_4

$$|\Delta_4|<K_1K_2\exp\left[\frac{\beta(p'^2+p''^2+q'^2+q''^2)}{4}\right]\int_0^T dt_1\int_0^{t_1}dt_2[1-e^{-\nu(T-t_1)}]^{-3/4}[e^{\nu t_2}-1]^{-3/4}$$

$$<C\int_0^T ds(1-e^{-\nu s})^{-3/4}\nu^{-1}\int_0^\infty ds(e^x-1)^{-3/4}\leqslant C'\nu^{-1}\left\{\nu^{-1}\int_0^2 ds\left[s\left(1-\frac{s}{2}\right)\right]^{-3/4}+(1-e^{-2})^{-3/4}\left(T-\frac{2}{\nu}\right)\right\}.$$

Since C' does not depend on ν, we clearly have $\Delta_4\to 0$. Estimates for Δ_2,Δ_5 can be computed analogously; one also finds $\Delta_2\to 0$, $\Delta_5\to 0$.

(11) Since $\Delta_j\to 0$, $j=1,\ldots,6$, we have shown that

$$|\mathscr{P}_\nu(h,p'',q'',T;p',q',0)-\langle p'',q''|\exp(-iTH)|p',q'\rangle|\to 0.$$

This proves the main statement of the theorem.

(12) The fact that convergence in the sense of distributions follows already if only (C1) is satisfied was proved by the argument preceding the theorem. Note that there, too, only weak convergence was needed for the operators $E(\nu,h;T)$; the argument therefore extends trivially to maximal symmetric H. ■

Remarks: (1) In our formulation of Theorem 2.4, we have used initial and final times t' and t'', respectively, while in all our preceding analyses we took $t'=0$ (and $t''=T$). Since $h(p,q)$ and therefore also H are time independent, this simply amounts to a relabeling of t. It is certainly plausible that all the above also holds for time-dependent Hamiltonians, where the evolution operators are then taken to be time-ordered products. For quadratic Hamiltonians, where everything can be calculated explicitly, this is indeed the case.

(2) Strictly speaking, the pointwise limit proved in Theorem 2.4 is *not* stronger than the limit in the sense of the Schwartz distributions proved before. A close inspection of the proof shows indeed that our estimates of the difference functions Δ_j contain factors of the form $\exp[\frac{1}{4}\beta(p''^2+p'^2+q''^2+q'^2)]$, which are not tempered distributions. If h is polynomially bounded, better estimates can be made for the Δ_j, containing only polynomials in the p',q',p'',q''. These estimates then automatically imply convergence in the sense of the Schwartz distributions.

(3) Note that we have always restricted ourselves to strictly positive time intervals: $T>0$ in Proposition 2.2, and $t''>t'$ in Theorem 2.4. For $T=0$ or $t'=t''$ there is no hope of proving convergence, since

$$E(\nu,h;0)=\mathbf{1},$$

$$\mathscr{P}_\nu(h;p'',q'',0;p',q',0)=\delta(p''-p')\delta(q''-q')$$
$$\neq\langle p'',q''|p',q'\rangle.$$

(4) The construction above shows how the "antiordered symbol" $h(p,q)$ comes into play, rather than the more expected (and much more well-behaved) "ordered symbol" $H(p,q)$ (as defined in the Introduction). Note that for quadratic Hamiltonians a result similar to (1.3), but where the function $H(p,q)$ is used in the path integral instead of $h(p,q)$, also holds.[14] The price to pay for this change is that the measure has then to be replaced by a Wiener measure with drift terms (depending on H; see Ref. 14). This suggests that (1.3) might be one element of a family of related results, each with slightly different Hamiltonian functions in the action, and accordingly different measures.

III. THE SPIN CASE

The spin case can be treated completely analogously to the canonical case, modulo a change in the basic setting of course. In Sec. III A we define our notation, and in Sec. III B we reinterpret the spin path integral for finite ν as the integral kernel of an operator on $L^2(S^2)$. We state our final result (limit for $\nu\to\infty$) in Sec. III C, without proof since the proofs are the same as in Sec. II.

A. Notations and definitions

At the end of Sec. III C we shall see, analogously to (2.66) in the canonical case, that the matrix element between spin coherent states of the unitary evolution operator associated to a spin Hamiltonian for spin s can be written as the limit, for diverging diffusion constant, of path integrals on S^2 involving Wiener measure on the sphere. In all this the spin value s is fixed; s occurs also as a parameter in the path integral. In order to prove this relation we shall, however, also need matrix elements relating to other spin values than s (this is similar to the use of the $|p,q;k\rangle$ in the arguments in Sec. II, even though the final result involved only the $|p,q\rangle$). In order to make this distinction clear, we shall use the symbol j for any arbitrary integer or half-integer value, while s will be used only for the particular spin value (which can also be any integer or half-integer) for which we wish to construct a path integral.

Let $\mathscr{H}_j$ be a $(2j+1)$-dimensional complex Hilbert space $(j=\frac{1}{2},1,\frac{3}{2},...)$ carrying an irreducible representation of the Lie group SU(2). We denote the generators of the corresponding Lie algebra by S_k, $k=1,2,3$; one has

$$[S_1,S_2]=iS_3\quad\text{(plus cyclic permutations).}\tag{3.1}$$

It follows that

$$\sum_{k=1}^{3}S_k^2=j(j+1)\mathbf{1}_j,\tag{3.2}$$

where $\mathbf{1}_j$ is the unit operator in $\mathscr{H}_j$. Let $|j,m\rangle$ be a normalized vector in $\mathscr{H}_j$ such that

$$S_3|j,m\rangle=m|j,m\rangle,\quad m=-j,-j+1,...,j.$$

(We use here directly Dirac's bra-ket notation.) Let θ,ϕ denote the usual coordinates on the unit sphere, where $0\leqslant\theta\leqslant\pi$, $0\leqslant\phi<2\pi$. We define unitary operators on $\mathscr{H}_j$ by

$$U(\Omega)\equiv U(\theta,\phi)=\exp(-i\phi S_3)\exp(-i\theta S_2).$$

The spin coherent states (for state $|j,m\rangle$) are then given by[15]

$$|\Omega;j,m\rangle\equiv|\theta,\phi;j,m\rangle=U(\theta,\phi)|j,m\rangle.$$

We shall be more specifically interested in the $|\Omega;j,j\rangle$ $(m=j)$. We therefore also introduce the notation

$$|\Omega;j\rangle\equiv|\theta,\phi,j\rangle=U(\theta,\phi)|j,j\rangle.$$

For a given value of s, the states $|\Omega;s\rangle$ will be the analog of the states $|p,q\rangle$ in the canonical case, while the $|\Omega;j,m\rangle$ will play a role analogous to the $|p,q;n\rangle$. As in the canonical case, the $|\Omega;s\rangle$ form an overcomplete set in $\mathscr{H}_s$; their "overlap function" is given by

$$\begin{aligned}&\langle\theta'',\phi'',s|\theta',\phi';s\rangle\\&\quad=\{\cos[(\theta''-\theta')/2]\cos[(\phi''-\phi')/2]\\&\qquad+i\cos[(\theta''+\theta')/2]\sin[(\phi''-\phi')/2]\}^{2s}.\end{aligned}$$

As in the canonical case, the spin coherent states $|\Omega;j,m\rangle$ give rise to a resolution of the identity in $\mathscr{H}_j$ (for any m value)

$$N_j\int d\Omega\,|\Omega;j,m\rangle\langle\Omega;j,m|=\mathbf{1}_j,\tag{3.3}$$

where $d\Omega\equiv\sin\theta\,d\theta\,d\phi$, $N_j\equiv(2j+1)/4\pi$. One can also prove [from the orthogonality relations for the representations of SU(2)] that for $|j-j'|>0$ and integer,

$$\psi\in\mathscr{H}_j,\ \chi\in\mathscr{H}_{j'}\Rightarrow\int d\Omega\,\langle\psi|\Omega;j,m\rangle\langle\Omega;j',m|\chi\rangle=0.\tag{3.4}$$

Note well the same value of m!

One of the intermediate steps in Sec. II was the interpretation of the path integral as defining the integral kernel of an operator on the Hilbert space of square-integrable functions on the label space $\mathbb{R}^2$ for the $|p,q\rangle$. We shall need this here, too. We use the notation $L^2(S^2)$ for the square-integrable functions on S^2, with normalization

$$\|f\|^2=\int\frac{d\Omega}{4\pi}|f(\Omega)|^2,\quad d\Omega=\sin\theta\,d\theta\,d\phi.$$

For given spin value s (integer or half-integer, but fixed), we define functions h^s_{lm} by

$$h^s_{lm}(\theta,\phi) = \sqrt{2l+2s+1}\langle\Omega;l+s,s|l+s,m\rangle. \tag{3.5}$$

Here l takes all non-negative integer values ($l = 0,1,2,...$); for each value of l, m takes any of the $2(l+s)+1$ values $-(l+s)$, $-(l+s)+1,...,(l+s)$. One checks from (3.3) and (3.4) that the h^s_{lm} form an orthonormal set in $L^2(S^2)$, i.e.,

$$\int \frac{d\Omega}{4\pi} \overline{h^s_{l'm'}(\Omega)} h^s_{lm}(\Omega) = \delta_{ll'}\delta_{mm'}. \tag{3.6}$$

Note that for $s=0$, the $h^{s=0}_{lm}(\Omega) = (2l+1)^{1/2}\langle\Omega;l,0|l,m\rangle$ are exactly the familiar spherical harmonics

$$h^0_{lm} = Y_{lm}. \tag{3.7}$$

Hence the h^0_{lm} are not only orthonormal, they also form a complete set for $L^2(S^2)$. We shall see that this is true for *any* value for s (integer or half-integer).

Explicit calculation (see, e.g., Ref. 16) shows that the h^s_{lm} can be written as

$$h^s_{lm}(\theta,\phi) = K_{lm}e^{im\phi}(1+\cos\theta)^{|s+m|/2}(1-\cos\theta)^{|s-m|/2}$$
$$\times P^{(|s+m|,|s-m|)}_{l-\max(0,|m|-s)}(-\cos\theta), \tag{3.8}$$

where the K_{lm} are constant factors (normalization plus phase), and where the $P^{(\alpha,\beta)}_k$ are the Jacobi polynomials. Using the well-known fact that the $P^{(\alpha,\beta)}_k(x)$, $k=0,1,...$, are a complete orthogonal set on $L^2([-1,1])$ with respect to the weight functions $(1-x)^\alpha(1+x)^\beta$, one sees again from (3.8) that the h^s_{lm} are orthogonal. Since for every value of m the associated allowed l values range from $\max(0,|m|-s)$ to ∞ [this follows from $-(l+s) \leqslant m \leqslant l+s$], the lower index of the Jacobi polynomial in the right-hand side (rhs) of (3.8) ranges from 0 to ∞. This ensures that for fixed m and ϕ, the $h^s_{lm}(\theta,\phi)$ form a complete set in the θ variable. Consequently, the h^s_{lm} are a complete set in $L^2(S^2)$. Taking into account (3.6) also, we conclude that the h^s_{lm}, $l=0,1,2,...,m=-(l+s),-(l+s)+1,...,l+s$, form a complete orthonormal basis in $L^2(S^2)$.

For $\psi \in \mathscr{H}_s$, we define by f_ψ the function

$$f_\psi(\Omega) = \sqrt{2s+1}\langle\Omega;s|\psi\rangle.$$

It follows from (3.3) that the map $\psi \to f_\psi$ is isometric from $\mathscr{H}_s$ into $L^2(S^2)$. The image of $\mathscr{H}_s$ under this map is a closed subspace of $L^2(S^2)$, which we shall denote by $\mathscr{H}^s_o$ [this is the analog of $\mathscr{H}_o$ in Sec. II; we have introduced an extra superscript s because different s values lead, of course, to different subspaces $\mathscr{H}^s_o$ of $L^2(S^2)$]. We shall denote the isomorphism between $\mathscr{H}_s$ and $\mathscr{H}^s_o$ by U_s:

$$U_s:\mathscr{H}_s \to \mathscr{H}^s_o, \quad (U\psi)(\Omega) = \sqrt{2s+1}\langle\Omega;s|\psi\rangle,$$

where the notation $\widehat{U}_s$ stands for the operator $\mathscr{H}_s \to L^2(S^2)$ defined as $\widehat{U}_s = I_s \circ U_s$, where I_s is the natural embedding $\mathscr{H}^s_o \to L^2(S^2)$. The orthogonal projection operator in $L^2(S^2)$, mapping $L^2(S^2)$ onto $\mathscr{H}^s_o$, will be denoted by P^s_o.

Again, we define possible ways of transporting and extending an operator on $\mathscr{H}_s$ to an operator on $L^2(S^2)$. These two constructions are completely analogous to what we did in Sec. II [cf. (2.10)–(2.14)].

(1) Given $R \in \mathscr{B}(\mathscr{H}_s)$, we define $\widehat{R} \in \mathscr{B}[L^2(S^2)]$ by

$$\widehat{R}f = 0, \quad \text{if } f \perp \mathscr{H}^s_o,$$
$$\widehat{R}f = \widehat{U}_s R U_s^{-1} f, \quad \text{if } f \in \mathscr{H}^s_o.$$

Then $\widehat{R}$ is an integral operator on $L^2(S^2)$ with integral kernel $(2s+1)\langle\Omega'';s|R\,\Omega';s\rangle$

$$\langle f,\widehat{R}g\rangle = (2s+1)\int\frac{d\Omega''}{4\pi}\int\frac{d\Omega'}{4\pi}\overline{f(\Omega'')}$$
$$\times\langle\Omega'';s|R\,\Omega';s\rangle g(\Omega'). \tag{3.9}$$

(2) Given $R \in \mathscr{B}(\mathscr{H}_s)$, with

$$R = (2s+1)\int\frac{d\Omega}{4\pi}|\Omega;s\rangle\langle\Omega;s|r(\Omega) \tag{3.10}$$

[any operator in $\mathscr{B}(\mathscr{H}_s)$ can be written in this form, with r a smooth function], we define R_S on $L^2(S^2)$ by

$$(R_S f)(\Omega) = r(\Omega)f(\Omega). \tag{3.11}$$

(The index S stands for sphere.)

One checks again that

$$P^s_o R_S P^s_o = \widehat{R} \tag{3.12}$$

and

$$[\exp(-itH)]\widehat{\ } = P^s_o \exp(-it\widehat{H})P^s_o. \tag{3.13}$$

So much for our notation and definitions in the spin case. In the next subsection we introduce spin path integrals and show how they can be interpreted as integral kernels for operators in $L^2(S^2)$.

Note that since $\mathscr{H}_s$ is finite dimensional, we only have to deal with bounded operators this time (unlike the canonical case). Since the function $r(\Omega)$ in (3.10) can always be chosen as a smooth function, and since S^2 is compact, we also only have to consider bounded functions $r(\Omega)$. This simplifies the discussion considerably.

B. Definition of the path integral (for finite ν)

We define our ν-dependent spin path integral as

$$\mathscr{P}^s_{\nu,h}(\Omega'',t'';\Omega',t')$$
$$= 4\pi e^{\nu s(t''-t')/2}\int\exp\left[is\int\cos\theta\,d\phi\right.$$
$$\left.-i\int h(\theta,\phi)dt\right]d\mu^\nu_W(\theta,\phi),$$

where μ^ν_W is the Wiener measure on the sphere S^2, pinned at Ω'' for $t=t''$ and at Ω' for $t=t'$, and defined such that

$$\int d\mu^\nu_W(\theta,\phi) = (4\pi)^{-1}\left\{\exp\left[\frac{\nu(t''-t')\Delta}{2}\right]\right\}(\Omega'',\Omega')$$
$$= \sum_{l=0}^{\infty}\exp\left[-\frac{\nu(t''-t')l(l+1)}{2}\right]$$
$$\times\sum_{m=-l}^{l}Y_{lm}(\Omega'')Y^*_{lm}(\Omega'),$$

where the Y_{lm} are the standard spherical harmonic functions. For a restricted class of Hamiltonians, analogous path integrals for spin systems were already discussed in Ref. 15; the principal difference consists in the presence, in Ref. 15, of extra drift terms in the measure, which are absent here.

In the canonical case, in Sec. II, we could calculate $\mathscr{P}_\nu(h=0)$ explicitly, and thus identify the generator A of the semigroup with integral kernel $\mathscr{P}_\nu(h=0)$. In the present case, we have no explicit expression for $\mathscr{P}_\nu(h=0)$; we can, however, by standard techniques determine the partial differential equation associated with the above path integral. One finds thus that $\mathscr{P}_\nu(h;\Omega,t;\Omega',t')$ is a solution to the partial differential equation

$$\partial_t \mathscr{P}_\nu(h;\Omega,t) = -[\nu A_s + ih(\Omega)]\mathscr{P}_\nu(h;\Omega,t), \quad (3.14)$$

where

$$A_s = \frac{1}{2}\left[-\frac{1}{\sin\theta}\partial_\theta \sin\theta\,\partial_\theta - \frac{1}{\sin^2\theta}(\partial_\phi - is\cos\theta)^2 - s\right]. \quad (3.15)$$

The kernel $\mathscr{P}_\nu(h;\Omega,t;\Omega',t')$ is completely determined by (3.14) and the initial condition

$$\mathscr{P}_\nu(h;\Omega,t';\Omega',t') = 4\pi\delta(\Omega-\Omega') \equiv 4\pi\delta(\phi-\phi')\delta(\cos\theta-\cos\theta'). \quad (3.16)$$

In order to define A_s, given by (3.15), as an operator on $L^2(S^2)$, we need also to specify the domain $D(A_s)$ of A_s. We define

$$D(A_s) = \{f; e^{is\phi/2}f \in D(-\Delta)\},$$

where $D(-\Delta)$ is the usual domain of the Laplacian on the sphere. Note that, as in the canonical case, A_s is "almost" equal to $(-\frac{1}{2})$ times the Laplacian. In the present case we even recover the Laplacian if we put $s=0$. For half-integer values of s, the functions in $D(A_s)$ are not continuous at $\phi=0$. This poses no problem, however; we define ∂_ϕ on these functions, at the points $\phi=0$ or $\phi=2\pi$, as the suitable right or left derivative.

It is easy to check that A_s, defined by (3.15), with domain $D(A_s)$, is a self-adjoint operator on $L^2(S^2)$. This operator will be the analog, for the spin case, of the operator A in the canonical case. The property of A which turned out to be crucial in the proof of Proposition 2.2 was (2.24); this showed that any vector in $\mathscr{H}_o$ was an eigenvector of A with eigenvalue 0, while on $\mathscr{H}_o^\perp$ the spectrum of $A|_{\mathscr{H}_o^\perp}$ was bounded below by a strictly positive number. In the limit $\nu\to\infty$, this made everything collapse onto $\mathscr{H}_o$. The same is true here. It is not difficult to check, using (3.5), (3.1), (3.2), and

$$sh^s_{lm}(\theta,\phi) = \langle U(\theta,\phi)S_3 \quad l+s,s|l+s,m\rangle,$$

that

$$A_s h^s_{lm} = [l(l+2s+1)/2]h^s_{lm}. \quad (3.17)$$

Since obviously $h^s_{lm}\in D(A_s)$ for all l,m, and since the h^s_{lm} are a complete orthonormal set of vectors in $L^2(S^2)$, (3.17) tells us that A_s has a purely discrete spectrum; its eigenvalues and eigenvectors are given by (3.17). We have, as in the canonical case, that $A_s|_{\mathscr{H}^s_o}=0$; moreover, on $\mathscr{H}^{s\perp}_o$, A_s is bounded below by $\frac{1}{2}>0$. For h a real smooth function on S^2, we define the operator H_S on $L^2(S^2)$ by [as in (3.11)]

$$(H_S f)(\Omega) = h(\Omega)f(\Omega).$$

Since h is a bounded function, H_S is a bounded operator, and $\nu A_s + iH_S$, defined on $D(A_s)$, is a closed operator generating a contraction semigroup on $L^2(S^2)$. From (3.14) and (3.16) one then sees that the integral kernel for this semigroup is given by $\mathscr{P}_\nu(h)$

$$\{\exp[-(\nu A_s + iH_S)(t''-t')]\}(\Omega'',\Omega') = \mathscr{P}_\nu(h;\Omega'',t'';\Omega',t'). \quad (3.18)$$

C. The limit for $\nu\to\infty$

Exactly the same arguments as in Sec. II show that

$$\operatorname*{s-lim}_{\nu\to\infty} \exp[-(\nu A_s + iH_S)T] = P^s_o \exp(-iP^s_o H_S P^s_o T)P^s_o. \quad (3.19)$$

Note that since H_S is always bounded, $\exp(-iP^s_o H_S P^s_o T)$ is always well defined and unitary (unlike the canonical case).

The integral kernel of the operator in the rhs is given by

$$\begin{aligned}[P^s_o \exp(-iP^s_o H_S P^s_o T)P^s_o](\Omega'',\Omega') &= [P^s_o \exp(-i\hat{H}T)P^s_o](\Omega'',\Omega') \quad [\text{use } (3.12)],\\ &= [\exp(-iHT)]\hat{}\,(\Omega'',\Omega') \quad [\text{use } (3.13)],\\ &= (2s+1)\langle\Omega'';s|\exp(-iHT)\,\Omega';s\rangle \quad [\text{use } (3.9)],\end{aligned}$$

where

$$H = (2s+1)\int \frac{d\Omega}{4\pi}|\Omega;s\rangle\langle\Omega;s|h(\Omega).$$

Together with (3.18) this implies

$$\begin{aligned}&\operatorname*{d-lim}_{\nu\to\infty} 4\pi\exp[\nu s(t''-t')/2]\\ &\times\int \exp\left[is\int\cos\theta\,d\phi - i\int h(\theta,\phi)dt\right]d\mu^\nu_W(\theta,\phi)\\ &= (2s+1)\langle\Omega'';s|\exp[-iH(t''-t')]\,\Omega';s\rangle.\end{aligned}$$

Moreover, one can show, in a way completely analogous to the proof of Theorem 2.4, that the limit also holds pointwise. Note that no extra conditions on the function h are needed here, since h can always be chosen to be a continuous, bounded function.

Putting everything together, we can formulate our final result for spin path integrals.

Theorem 3.1: Let s be any integer or half-integer number (non-negative). Let H be any Hermitian operator on $\mathscr{H}_s$. Let H be generated by the smooth real function $h(\theta,\phi)$ by

$$H = (2s+1)\int \frac{d\Omega}{4\pi}|\Omega;s\rangle\langle\Omega;s|h(\Omega).$$

Then, for all Ω'',Ω' in S^2, and for all $t''>t'$,

$$\begin{aligned}&\lim_{\nu\to\infty}\frac{4\pi}{2s+1}\exp\left[\frac{\nu s(t''-t')}{2}\right]\int\exp\left[is\int\cos\theta\,d\phi\right.\\ &\left.-i\int h(\theta,\phi)dt\right]d\mu^\nu_W(\theta,\phi)\\ &= \langle\Omega'';s|\exp[-iH(t''-t')]\,\Omega';s\rangle.\end{aligned}$$

Here μ^ν_W is a Wiener measure on the sphere S^2, with diffusion constant ν, and pinned at $\Omega''=(\theta'',\phi'')$ for $t=t''$ and at $\Omega'=(\theta',\phi')$ for $t=t'$. The normalization of μ^ν_W is given by

$$\int d\mu_W^{\nu}(\theta,\phi) = \sum_{l=0}^{\infty} \exp\left[-\frac{\nu(t''-t')l(l+1)}{2}\right] \times \sum_{m=-l}^{l} Y_{lm}(\Omega'')Y^*_{lm}(\Omega').$$

ACKNOWLEDGMENTS

We would like to thank Ms. Terry Yulick for her patience and expertise in preparing the compuscript.

The authors also wish to thank Zentrum für interdisziplinäre Forschung (ZiF) (Bielefeld), where this work was begun, for its hospitality and support. One of us (I.D.) thanks AT&T Bell Laboratories, Murray Hill, where this work was finished, for hospitality and financial support.

APPENDIX: CONNECTION WITH PREVIOUS PROOF

Theorem 2.4 had been announced by us earlier,[4] in a weaker version, namely (see Ref. 4)

$$\lim_{\nu\to\infty} \int \frac{dp''\,dq''}{2\pi} \int \frac{dp'\,dq'}{2\pi} \Bigg(\langle \psi | p'',q'' \rangle \Bigg\{ 2\pi e^{\nu T/2} \times \int \exp\left[\frac{i}{2}\int (p\,dq - q\,dp) - i\int h(p,q)dt\right] d\mu_W^{\nu} \Bigg\} \langle p',q' | \phi \rangle \Bigg) = \langle \psi, \exp(-iTH)\phi \rangle. \tag{A1}$$

To see that (A1) is weaker than Theorem 2.4 let us go back to the properties of the cs. It follows from (2.5) that for any (bounded) operator B on $\mathscr{H}$ and any $\psi,\phi \in \mathscr{H}$

$$\langle \psi, B\phi \rangle = \int \frac{dp''\,dq''}{2\pi} \int \frac{dp'\,dq'}{2\pi} \times \langle \psi | p'',q'' \rangle \langle p'',q'' | B p',q' \rangle \langle p',q' | \phi \rangle. \tag{A2}$$

It is therefore clear that (2.37) implies (A1). The reverse is not true, however; due to the overcompleteness of the cs, the matrix element $\langle p'',q'' | B p',q' \rangle$ in formula (A2) can be replaced by any element of a large equivalence class of functions.

The proof for (A1), sketched in Ref. 4, was different from the one given here. The main difference lies in the interpretation of the path integral expression for finite ν (corresponding to Sec. II B in the present paper). Let us restrict ourselves, in this discussion of the difference between Ref. 4 and the present paper, to the case where h is a bounded function, hence H_V a bounded operator. In Ref. 4, we introduced an abstract Hilbert space

$$\hat{\mathscr{H}} = \bigoplus_{n=0}^{\infty} \mathscr{H}_n,$$

where each $\mathscr{H}_n$ was a copy of $\mathscr{H}_o$. It turns out that the Hilbert space $L^2(V)$ we have used here is nothing else than a concrete realization for this $\hat{\mathscr{H}}$; the subspace $\mathscr{H}_o \subset L^2(V)$ corresponds to the zeroth space $\hat{\mathscr{H}}_o$ in the construction of $\hat{\mathscr{H}}$. The nth subspace $\hat{\mathscr{H}}_n \subset \hat{\mathscr{H}}$ corresponds then to the closed linear span in $L^2(V)$ of the $(h_{nl}), l = 0,1,2,\ldots$. We also introduced vectors $|p,q;\beta\rangle\rangle$ in $\hat{\mathscr{H}}$ defined as

$$|p,q;\beta\rangle\rangle = \bigoplus_{n=0}^{\infty} \beta^{n/2} |p,q;n\rangle, \quad 0<\beta<1.$$

With the identification $\hat{\mathscr{H}} \leftrightarrow L^2(V)$, these can be written, in our present framework, as

$$|p,q;\beta\rangle\rangle = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \beta^{n/2}\, \overline{h_{nm}(p,q)}\, h_{nm}.$$

The operators, defined in Ref. 4,

$$\mathbf{A} = \bigoplus_{h=0}^{\infty} n\mathbb{1}_n, \quad \mathbf{I}(\beta) = \bigoplus_{n=0}^{\infty} \beta^n \mathbb{1}_n,$$

$$\mathbf{H} = \lim_{\beta\to 1} \int \frac{dp\,dq}{2\pi} |p,q;\beta\rangle\rangle\langle\langle p,q;\beta | \, h(p,q)$$

correspond to, respectively, A [as defined by (2.23)], $I(\beta) = \exp(A \ln\beta)$, and H_V. We also introduced in Ref. 4 the operators

$$\mathbf{E}_N = \int \frac{dp\,dq}{2\pi} e^{-i\epsilon_N h(p,q)} |p,q;\beta_N\rangle\rangle\langle\langle p,q;\beta_N|,$$

where

$$\epsilon_N = T/(N+1),$$

$$\beta_N = [1 - \nu T/2(N+1)]/[1 + \nu T/2(N+1)].$$

Finally, we rewrote $\mathscr{P}_\nu(h)$ as [using our present notations, and identifying $\hat{\mathscr{H}}$ and $L^2(V)$]

$$\int \frac{dp''\,dq''}{2\pi} \int \frac{dp'\,dq'}{2\pi} \langle \psi | p'',q'' \rangle \times \mathscr{P}_\nu(h; p'',q'',T; p',q',0) \langle p',q' | \phi \rangle = \lim_{N\to\infty} \langle \hat{U}\psi, I(\beta_N)(\mathbf{E}_N)^N I(\beta_N)\hat{U}\phi \rangle, \tag{A3}$$

with β_N as above, and $\hat{U}$ as defined in Sec. II A. In the limit $N\to\infty$, obviously, $\text{s-lim}_{N\to\infty} I(\beta_N) = \mathbb{1}$. The limit of $(\mathbf{E}_N)^N$ was more tricky, because of the complicated N dependence of $\mathbf{E}_N$. Using a theorem of Chernoff,[17] one can show, however, that

$$\operatorname*{s-lim}_{N\to\infty} (\mathbf{E}_N)^N = \exp[-T(\nu A + iH_V)]. \tag{A4}$$

This can intuitively be guessed already from the matrix elements of $\mathbf{E}_N$ between the h_{kl}

$$\langle h_{kl}, \mathbf{E}_N h_{rs} \rangle = \int \frac{dp\,dq}{2\pi} \overline{h_{kl}(p,q)} \exp[-i\epsilon_N h(p,q) - [(k+r)/2]\nu\epsilon_N + O(\epsilon_N^3)] h_{rs}(p,q).$$

Substituting (A4) into (A3) leads to the interpretation of $\mathscr{P}_\nu(h)$ as a generating function [in the sense of (A2)] for the operator $\hat{U}^* \exp[-T(\nu A + iH_V)]\hat{U}$. This conclusion is of course weaker than (2.33), and therefore led to the weaker result (A1) in Ref. 4.

[1]A representative sample of references is the following: I. M. Gel'fand and A. M. Yaglom, J. Math. Phys. **1**, 48 (1960); D. G. Babbitt, J. Math. Phys. **4**, 36 (1963); E. Nelson, J. Math. Phys. **5**, 332 (1964); 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); K. Gawedzki, Rep. Math. Phys. **6**, 327 (1974); A. Truman, J. Math. Phys. **17**, 1852 (1976); S. A. Albeverio and R. J. Hoegh-Krohn, *Mathematical Theory of Feynman Path Integrals* (Springer, Berlin, 1976); V. P. Maslov and A. M. Chebotarev, Theor. Math. Phys. **28**, 793 (1976); C. DeWitt-Morette, A. Maheshwari, and B. Nelson, Phys. Rep. **50**, 255 (1979); D. Fujiwara, Duke Math. J. **47**, 559 (1980); P. Combe, R. Hoegh-Krohn, R. Rodriguez, and M. Sirugue, Com-

mun. Math. Phys. 77, 269 (1980); F. A. Berezin, Sov. Phys. Usp. 23, 763 (1980); T. Ichinose, Proc. Jpn. Acad. Ser. A 58, 290 (1982); I. Daubechies and J. R. Klauder, J. Math. Phys. 23, 1806 (1982).
[2]E. Nelson, J. Math. Phys. 5, 332 (1964).
[3]See. e.g., L. Shulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).
[4]An announcement of our results appears in J. R. Klauder and I. Daubechies, Phys. Rev. 52, 1161 (1984).
[5]J. R. Klauder, in *Path Integrals*, edited by George J. Papadopoulos and J. T. Devreese (Plenum, New York, 1978).
[6]R. H. Cameron, J. Anal. Math. 10, 287 (1962/63).
[7]E. Lieb, Commun. Math. Phys. 31, 327 (1973).
[8]See. e.g, E. J. McShane, *Stochastic Calculus and Stochastic Models* (Academic, New York, 1974).
[9]V. Bargmann, Commun. Pure Appl. Math. 14, 187 (1961).
[10]J. E. Moyal, Proc. Camb. Philos. Soc. 45, 99 (1949); M. S. Bartlett and J. E. Moyal, Proc. Camb. Philos. Soc. 45, 545 (1949).
[11]I. Daubechies and J. R. Klauder, J. Math. Phys. 23, 1806 (1982).
[12]S. Girvin and T. Jach, Phys. Rev. B 29, 5617 (1984).
[13]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966).
[14]I. Daubechies and J. R. Klauder, Lett. Math. Phys. 7, 229 (1983).
[15]J. R. Klauder, J. Math. Phys. 23, 1797 (1982).
[16]See, e.g., I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications* (Pergamon, New York, 1983); or N. Ja. Vilenkin, *Special Functions and the Theory of Group Representations* (A. M. S., Providence, RI, 1968), Vol. 22.
[17]P. Chernoff, J. Funct. Anal. 2, 238 (1968).

Remarks concerning functional integration on compact spaces

Adrian Patrascioiu[a)]
Centre de Physique Théorique, Centre National de la Recherche Scientifique, Luminy Case 907, 13288 Marseille Cedex 9, France and Université de Provence, Marseille, France

Jean-Louis Richard
Centre de Physique Théorique, Centre National de la Recherche Scientifique, Luminy Case 907, 13288 Marseille Cedex 9, France

(Received 26 April 1984; accepted for publication 8 March 1985)

Certain technical subleties regarding the Feynman–Kac representation of the propagator for motion on compact spaces are discussed. A method of evaluation of the corresponding path integral is given. It produces the correct spectrum for the Hamiltonian, while the ordinary semiclassical method fails.

I. INTRODUCTION

The scope of this paper is to analyze functional integration on compact manifolds. This problem is of interest for understanding the mass spectrum of the nonlinear σ model. In a recent paper,[1] the case of a nonrelativistic particle moving on a sphere S^2 was analyzed. It was pointed out that both the derivation of a Feynman–Kac formula for the propagator and its evaluation were highly nontrivial, and that in the semiclassical approximation for this problem, one obtained the wrong spectrum for the Hamiltonian.

In the present paper, we would like to clarify certain points raised in Ref. 1 and outline a procedure of functional integration by which one can obtain the (qualitatively) correct spectrum.

We begin our discussion with a brief outline of the derivation of the Feynman–Kac formula for the propagator. For the Hamiltonian

$$H = \mathbf{L}^2/2MR^2, \tag{1}$$

$$\mathbf{L}^2|l,m\rangle = \hbar^2 l(l+1)|l,m\rangle, \quad l=0,1,..., \quad -l<m<l, \tag{2}$$

$$\langle l,m|\theta,\phi\rangle^* = Y_l^m(\theta,\phi), \tag{3}$$

one can easily derive

$$\langle\theta_f,\phi_f,T\,|\theta_0,\phi_0,0\rangle = \left(\prod_{i=1}^{N}\int_0^{\pi}\sin\theta_i\,d\theta_i\int_0^{2\pi}d\phi_i\right)\times\prod_{i=0}^{N}\left(\sum_{l_i=0}^{\infty}\frac{2l_i+1}{4\pi}P_{l_i}(\cos\Xi_{i+1,i})\times e^{-l_i(l_i+1)/2\beta N}\right), \tag{4}$$

with

$$\beta \equiv MR^2/\hbar T, \tag{5}$$

$$\cos\Xi_{i+1,i} = \cos\theta_{i+1}\cos\theta_i + \sin\theta_{i+1}\sin\theta_i \times\cos(\phi_{i+1}-\phi_i). \tag{6}$$

For $N\to\infty$ (β fixed), one has

$$\lim_{N\to\infty}\sum_{l=0}^{\infty}\frac{2l+1}{4\pi}P_l(\cos\theta)e^{-l(l+1)/2\beta N} = \lim_{N\to\infty}(\beta N/2\pi)e^{\beta N(\cos\theta-1)} + O(N^{-2}). \tag{7}$$

We will take the following representation for the propagator:

$$\langle\theta_f,\phi_f,T\,|\theta_0,\phi_0,0\rangle = \lim_{N\to\infty}\left(\frac{\beta N}{2\pi}\right)^{N+1}\left(\prod_{i=1}^{N}\int_0^{\pi}\sin\theta_i\,d\theta_i\int_0^{2\pi}d\phi_i\right)\times\exp\left(-\beta N\sum_{i=o}^{N}(1-\cos\Xi_{i+1,i})\right). \tag{8}$$

This formula displays the manifest rotational invariance present in the problem. Indeed, it says

$$\langle\theta_f,\phi_f,T\,|\theta_0,\phi_0,0\rangle = \lim_{N\to\infty}\left(\frac{\beta N}{2\pi}\right)^{N}\left(\prod_{i=1}^{N}\int d\mathbf{n}_i\right)\times\exp\left(-\sum_{i=0}^{N}\beta N\frac{(\mathbf{n}_{i+1}-\mathbf{n}_i)^2}{2}\right), \tag{9}$$

or formally

$$\langle\theta_f,\phi_f,T\,|\theta_0,\phi_0,0\rangle = \int D\mathbf{n}(t)e^{-1/\hbar\int_0^T dt\,\frac{MR^2\dot{\mathbf{n}}^2}{2}}. \tag{10}$$

The semiclassical approximation of this functional integral consists of finding all the classical solutions and integrating the Gaussian fluctuations around them. This calculation was carried out in Ref. 1, where it was found that it produced a qualitatively incorrect spectrum. We will discuss this failure later (see Sec. II), after first outlining the correct computation.

To calculate the functional integral in Eq. (8), we use the expansion

$$\exp[-\beta N\sin\theta_{i+1}\sin\theta_i\cos(\phi_{i+1}-\phi_i)] = \sum_{m_i=-\infty}^{\infty}e^{im_i(\phi_{i+1}-\phi_i)}I_{m_i}(\beta N\sin\theta_{i+1}\sin\theta_i). \tag{11}$$

The integrations over the ϕ_i's can be performed, and one obtains

$$\langle\theta_f,\phi_f,T\,|\theta_0,\phi_0,0\rangle = \lim_{N\to\infty}\left(\frac{\beta N}{2\pi}\right)\sum_{m=-\infty}^{N\infty}e^{im(\phi_f-\phi_0)}\times\prod_{i=1}^{N}\int_0^{\pi}\sin\theta_i\,d\theta_i\,e^{-\beta N(1-\cos\theta_{i+1}\cos\theta_i)}\times I_m(\beta N\sin\theta_{i+1}\sin\theta_i). \tag{12}$$

It is convenient to define the following function:

$$e^{-V_m(z)} \equiv (z/2\pi)^{1/2}e^{-z}I_m(z). \tag{13}$$

[a)] Permanent address: Physics Department, University of Arizona, Tucson, Arizona 85721.

We note for further reference that

$$\lim_{z\to\infty} V_m(z) = (m^2 - \tfrac{1}{4})/2z, \tag{14}$$

$$\lim_{z\to 0} V_m(z) = -(m + \tfrac{1}{2})\ln z. \tag{15}$$

In terms of V_m, Eq. (12) can be written as

$$\langle \theta_f,\phi_f,T\,|\theta_0,\phi_0,0\rangle$$

$$= (\sin\theta_f \sin\theta_0)^{1/2} \lim_{N\to\infty} (\beta N/2\pi)^{N/2}$$

$$\times \sum_{m=-\infty}^{\infty} e^{im(\phi_f-\phi_0)} \prod_{i=1}^{N} \int_0^{\pi} d\theta_i$$

$$\times \exp(-\beta N[1-\cos(\theta_{i+1}-\theta_i)]$$

$$- V_m(\beta N \sin\theta_{i+1} \sin\theta_i)). \tag{16}$$

Since we are letting $N\to\infty$, in Eq. (16)

$$V_m(\beta N \sin\theta_{i+1} \sin\theta_i) \to \frac{m^2-\frac{1}{4}}{2\beta N \sin\theta_{i+1}\sin\theta_i}, \tag{17}$$

for θ_i, θ_{i+1} not too close to 0 or π. We notice then that for $m\neq 0$, we have an attractive potential centered around $\theta = \pi/2$. Since our formalism has been manifestly rotational invariant, we can choose without loss of generality

$$\theta_0 = \theta_f = \pi/2,$$

and apply the semiclassical approximation to evaluate all the functional integrals with $m\neq 0$ in Eq. (16). The classical solution has $\theta_i = \pi/2$. We replace them in Eq. (16)

$$\theta_i \to \pi/2 + \delta_i,$$

using for V_m $m\neq 0$ the expression in Eq. (17). Expanding V_m to second order in δ_i and $\cos(\theta_{i+1}-\theta_i)$ to fourth order in $(\delta_{i+1}-\delta_i)$ (see Ref. 2), we obtain for $m\neq 0$

$$\langle \pi/2,\phi_f,T\,|(\pi/2),\phi_i,0\rangle_m$$

$$= \exp[im(\phi_f-\phi_0)]\exp\Big[-(1/2\beta)$$

$$\times\big(m^2 - \tfrac{1}{2} + \sqrt{m^2-\tfrac{1}{4}}\big)\Big]. \tag{18}$$

This result shows that the spectrum is discrete for $m\neq 0$. The actual values of E_m are reasonably accurate, the error decreasing to zero as $m\to\infty$.

The case $m=0$ cannot be treated semiclassically: as $N\to\infty$, $V_0(\delta)$ produces an infinitely attractive potential with minima at 0 and π. Since as $\theta\to 0$, $V(\theta)\sim -1/4\theta^2$, the potential is as singular as the kinetic energy; hence, it cannot be treated as a perturbation. We proceed with the calculation directly by observing that

$$\lim_{N\to\infty} \beta N e^{-\beta N(1-\cos\theta_{i+1}\cos\theta_i)} I_0(\beta N \sin\theta_{i+1}\sin\theta_i)$$

$$= \lim_{N\to\infty} \sum_{l_i=0}^{\infty} \frac{2l_i+1}{2} P_{l_i}(\cos\theta_{i+1})$$

$$\times P_{l_i}(\cos\theta_i) e^{-l_i(l_i+1)/2\beta N} + O(N^{-2}). \tag{19}$$

Using the orthogonality of the Legendre polynominals, we obtain immediately

$$\langle \theta_f,\phi_f,T\,|\theta_0,\phi_0,0\rangle_{m=0}$$

$$= \sum_{l=0}^{\infty} \frac{2l+1}{2} P_l(\cos\theta_f) P_l(\cos\theta_0) e^{-l_i(l_i+1)/2\beta} \tag{20}$$

This, of course, is the exact answer: in effect, we have traced back the steps leading from (4) to (8). This procedure is rather unsatisfactory. Yet the singular nature of $V_0(Z)$ has prevented all our efforts for finding a genuine approximation for $m=0$.

II. DISCUSSION

We have shown an approximate method for obtaining a correct result. With hindsight, we would like to go back and discuss why the naive semiclassical approximation failed. The error can be traced back to two basic difficulties.

(i) In Eq. (9), we may hope for a Gaussian approximation for $\beta\to\infty$. Since for finding the spectrum we want $\beta\to 0$ ($T\to\infty$), there is no *a priori* reason for the approximation to work, and it does not. On the contrary, it is well known[3] that it is correct for $\beta\to\infty$.

It may be instructive to emphasize that this difficulty is strictly connected with the compact nature of the sphere S^2. Indeed, taking the classical solutions along the equator, it is easy to verify that in the θ direction, the fluctuations occur in a potential (see Ref. 1)

$$(MR^2/2)[(2\pi m + \phi_f - \phi_0)/T]^2(\pi/2-\theta)^2.$$

Obviously, as $T\to\infty$, ignoring that the particle is on a sphere, hence, $0<\theta<\pi$, is incorrect.

(ii) The functional integral to be computed is not the naive one. Indeed, one may be tempted to interpret Eq. (10) as

$$\langle \theta_f,\phi_f,T\,|\theta_0,\phi_0,0\rangle$$

$$= \lim_{N\to\infty}\Big(\frac{\beta N}{2\pi}\Big)^N \Big(\prod_{i=1}^{N}\int d\Omega_i\Big)\exp\Big(\sum_{i=0}^{N}\frac{\beta N}{2}[(\theta_{i+1}-\theta_i)^2$$

$$+ \sin\theta_{i+1}\sin\theta_i(\phi_{i+1}-\theta_i)^2]\Big). \tag{21}$$

It has been pointed out by many authors, starting with DeWitt,[3] that to obtain the functional integral corresponding to Eq. (9), one must retain all the terms of order $\exp(1/N)$. This then replaces the naive action with an effective action [see Eq. (16)]. It may be tempting to think in the present case the effective action differs from the naive one by [see Eq. (17)]

$$-\sum_{i=1}^{N}\frac{1}{8\beta N}\Big(1+\frac{1}{\sin\theta_{i+1}\sin\theta_i}\Big). \tag{22}$$

This would be incorrect, since for $\beta\sin\theta_{i+1}\sin\theta_i \ll 1$, one must use Eq. (15) rather than (14). It can be checked that it is the fact that $V_m(z)\to\infty$ for $\delta\to 0$, which produces the correct spectrum.

To summarize, our conclusion is that the functional representation of the propagator is specified by the combined effect of the Hamiltonian and the parametrization of the manifold used for the integration. These specifications produce an effective action, generally different from the classical action. Any attempt to approximate the functional integration should use the former action. Therefore, if interested in a semiclassical approximation, one must find the effective action and its extrema, then verify that the Gaussian approximation does not *a priori* fail because the extremum is too flat. For the sphere S^2, this approximation works for $m \neq 0$ (even for $\beta \to 0$), but it is inapplicable for $m = 0$.

A final statement: We have discussed the case of S^2 in one dimension. Our conclusion would apply equally well to S^n. In particular, we doubt that the naive semiclassical approximation is exact for motion on Lie groups.[4] Similarly, they apply in two dimensions, a case which we are presently studying.

ACKNOWLEDGMENT

A. Patrascioiu would like to thank the Centre de Physique Théorique (Luminy–Marseille) for its hospitality.

[1] A. Patrascioiu and M. Viviniscu "The path integral for motions on compact spaces," preprint FT-235-89, Bucharest.
[2] S. F. Edwards and Y. Y. Guliaev, Proc. R. Soc. London Ser. A **279**, 229 (1964).
[3] B. B. DeWitt, Rev. Mod. Phys. **29**, 377 (1957).
[4] M. S. Marinov, Phys. Rep. **60**, 1 (1980).

A note on the classification of the zeros of angular momentum coefficients

K. Srinivasa Rao[a)]
Institut für Theoretische Kernphysik, Universitat Bonn, Bonn, West Germany

(Received 19 October 1984; accepted for publication 22 February 1985)

Closed form expressions for polynomial or "nontrivial" zeros of degree one of the Clebsch–Gordan (3-j) and the Racah (6-j) coefficients are presented and a conjecture regarding closed form formulas for the same is made.

I. INTRODUCTION

It is well known that the angular momentum coupling (3-j) and recoupling (6-j) coefficients have a series or a polynomial part, and that the polynomial parts of these coefficients can be rearranged into generalized hypergeometric functions of unit argument. For certain allowed values of the arguments of the 3-j or the 6-j coefficient, if the polynomial part becomes zero, then we have a "nontrivial" or polynomial zero of that coefficient. Further, if $n + 1$ indicates the number of terms in the polynomial part, which add up to a zero value, then we have a polynomial zero of degree n of that angular momentum coefficient.

Inspired by the suggestion of Koozekanani and Biedenharn[1] that realizations of exceptional Lie algebras might provide bases for explaining the polynomial or "nontrivial" zeros of the Racah (6-j) coefficient, Vanden Berghe *et al.*[2–4] have accounted for some of these zeros, in a continuing series of papers.

II. MULTIPLICATIVE FACTORS

On the other hand, closed form multiplicative factors can be obtained[5] for the polynomial zeros of degree 1, of the 3-j and the 6-j coefficients, which are independent of the numerical values of the arguments of the coefficients, whose zeros are being sought. In fact, the polynomial zeros of degree 1 of the 3-j coefficient can all be accounted for by the factor

$$(1-\delta_{x,y}\delta_{n,1}), \tag{1}$$

where n, which governs the degree of the polynomial zeros, is given by

$$n=\min(R_{2p},R_{3q},R_{1r}), \tag{2}$$

with R_{ik} being the elements of the square symbol

$$||R_{ik}|| = \left|\left|\begin{matrix} -j_1+j_2+j_3 & j_1-j_2+j_3 & j_1+j_2-j_3 \\ j_1-m_1 & j_2-m_2 & j_3-m_3 \\ j_1+m_1 & j_2+m_2 & j_3+m_3 \end{matrix}\right|\right|. \tag{3}$$

The values of x and y are then given by

$$x=R_{mr}R_{kp}, \quad y=R_{mp}R_{kr}, \tag{4}$$

with (lmk) and (pqr) corresponding to specific permutations of (123) and $n=R_{lq}$ (say).

Similarly, the multiplicative factor

$$(1-\delta_{X,Y}\delta_{n,1}), \tag{5}$$

in the definition[6] for the 6-j coefficient,

$$W(abcd;ef)=(-1)^{a+b+c+d}\begin{Bmatrix} a & b & e \\ d & c & f \end{Bmatrix}, \tag{6}$$

will account for all the polynomial zeros of degree 1 of this coefficient. In (5), the degree of the polynomial zeros is given by

$$n=\beta_0-\alpha_0, \tag{7}$$

where

$$\alpha_0=\max(\alpha_1,\alpha_2,\alpha_3,\alpha_4), \quad \beta_0=\min(\beta_1,\beta_2,\beta_3),$$

with

$$\begin{aligned} &\alpha_1=a+b+e, \quad \beta_1=a+b+c+d, \\ &\alpha_2=c+d+e, \quad \beta_2=a+d+e+f, \\ &\alpha_3=a+c+f, \quad \beta_3=b+c+e+f, \\ &\alpha_4=b+d+f, \end{aligned} \tag{8}$$

and

$$\begin{aligned} X&=(\beta_s-\alpha_0)(\beta_t-\alpha_0)(\beta_0+1), \\ Y&=(\beta_0-\alpha_k)(\beta_0-\alpha_l)(\beta_0-\alpha_m), \end{aligned} \tag{9}$$

where β_s and β_t correspond to two of the three β's other than β_0 and $\alpha_k,\alpha_l,\alpha_m$ correspond to three of the four α's other than α_0.

III. POLYNOMIAL ZEROS

We arrived at these closed form multiplicative factors, to be introduced into the definitions of the 3-j and the 6-j coefficients, by rearranging the series parts of these coefficients into binomial expansions, using generalized powers, in Ref. 5. Using these, we were able to classify the polynomial zeros according to their degree, and the explicit tables of these zeros are given in Ref. 7. It is to be noted that 21 out of 36 polynomial zeros of the 3-j coefficient, for $J(=j_1+j_2+j_3)\leqslant 27$ and 1174 out of 1420 polynomial zeros of the 6-j coefficient, for $(a,b,c,d,e,f)\leqslant 18.5$, are polynomial zeros of degree 1, accounted for by the factors (1) and (5), respectively. Hence, the polynomial zeros of degree one are to be considered as trivial zeros, and not as "nontrivial" zeros, as referred to hitherto.[1]

In Table I, we list the polynomial zeros of the 6-j coefficient which have been explained either due to the violation of the triangle rule for quasispin,[1] or due to the vanishings of fractional parentage coefficients (f.p.c.)[8] in the atomic g shell, or due to the realizations of the exceptional Lie algebras G_2, F_4, and E_6. In the last column of Table I, we give n, which corresponds to the degree of the polynomial zeros, given by (7). We note that of the 12 generic entries, 11 are

[a)]On leave of absence from The Institute of Mathematical Sciences, Madras—600 113, India.

TABLE I. Generic polynomial zeros of the 6-j coefficient $\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}$ for which explanations have been given.

j_1	j_2	j_3	l_1	l_2	l_3	Explanation	Ref.	n
2	2	2	1.5	1.5	1.5	quasispin[a]	1	1
5	5	3	3	3	3	$R_7 \supset G_2 \supset SO_3$	1	1
5	5	2	2	2	4	g shell	8	1
9	6	4	2	5	5	f.p.c.[b]	8	1
11	11	3	4	4	8	$SO_{26} \supset F_4 \supset SO_3$	2	1
11	11	9	8	4	8	$SO_{26} \supset F_4 \supset SO_3$	2	1
3	2	2	1	2	2	$F_4 \supset SO_3 \otimes SO_3$	3	1
7	4.5	4.5	2.5	4	4	$F_4 \supset SO_3 \otimes SO_3$	3	1
11	8	6	4	4	8	$E_6 \supset SO_3$	4	1
7	6	5	4	6	4	$E_6 \supset SO_3$	4	2
6	6	6	5	4	3	$E_6 \supset SO_3$	4	1
9	6	4	2	5	5	$E_6 \supset SO_3$	4	1

[a] Regge symmetries do not give rise to any other 6-j coefficient.

[b] $\begin{Bmatrix} 5 & 5 & 2 \\ 2 & 2 & 4 \end{Bmatrix} = \begin{Bmatrix} 5 & 4.5 & 1.5 \\ 2 & 2.5 & 4.5 \end{Bmatrix} = \begin{Bmatrix} 5 & 4.5 & 2.5 \\ 2 & 1.5 & 4.5 \end{Bmatrix}$ and $\begin{Bmatrix} 9 & 6 & 4 \\ 2 & 5 & 5 \end{Bmatrix} = \begin{Bmatrix} 8 & 6 & 5 \\ 1 & 5 & 6 \end{Bmatrix}$, follow from the Regge symmetries.

Note: The 24 6-j coefficient zeros which follow from the nine other generic zeros in this table can be found in Table I of Ref. 4.

polynomial zeros of degree 1, and only one is a polynomial zero of degree 2.

As has been pointed out by Vanden Berghe *et al.*,[4] one can continue the program of giving a group theoretical explanation for the polynomial zeros of the 6-j coefficient. While the basis for realizations of the exceptional Lie algebras is by itself fascinating, it being used to explain the zeros of the 6-j coefficient is likely to lead to only alternate explanations for more of the polynomial zeros of degree 1, which are trivial structure zeros represented by (5).

We conjecture that, in principle, one can find closed form formulas for the polynomial zeros of the 3-j and 6-j coefficients, provided we look upon these coefficients as generalized hypergeometric functions of unit argument,[9,10] which are analytic,[11] and extend the method of Siewert and Burniston[12] for the determination of zeros of analytic functions to the case of analytic ${}_pF_q(1)$s.

ACKNOWLEDGMENTS

The author thanks the Alexander von Humboldt Foundation for the award of a fellowship and Professor Dr. K. Bleuler and Professor Dr. M. G. Huber for their kind hospitality at the Institut für Theoretische Kernphysik of the University of Bonn.

[1] S. H. Koozekanani and L. C. Biedenharn, Rev. Mex. Fis. **23**, 327 (1974). Also, see L. C. Biedenharn and J. D. Louck, "The Racah–Wigner Algebra in Quantum Theory," in *Encyclopedia of Mathematics and its Applications* (Addison–Wesley, London, 1981), Vol. 9.

[2] J. Van der Jeugt, G. Vanden Berghe, and H. De Meyer, J. Phys. A: Math. Gen. **16**, 1377 (1983).

[3] H. De Meyer, G. Vanden Berghe, and J. Van der Jeugt, J. Math. Phys. **25**, 751 (1984).

[4] G. Vanden Berghe, H. De Meyer, and J. Van der Jeugt, J. Math. Phys. **25**, 2585 (1984).

[5] K. Srinivasa Rao and V. Rajeswari, J. Phys. A: Math. Gen. **17**, L243 (1984).

[6] See, for instance, L. C. Biedenharn and J. D. Louck, "Angular Momentum in Quantum Physics," in *Encyclopedia of Mathematics and its Applications* (Addison–Wesley, London, 1981), Vol. 8.

[7] K. Srinivasa Rao and V. Rajeswari, Rev. Mex. Fis. **31**, No. 4 (1985).

[8] B. R. Judd, in *Topics in Atomic and Nuclear Theory* (Caxton Press, Christchurch, New Zealand, 1970), pp. 42 and 43.

[9] K. Srinivasa Rao, T. S. Santhanam, and K. Venkatesh, J. Math. Phys. **16**, 1528 (1975).

[10] K. Srinivasa Rao, J. Phys. A: Math. Gen. **11**, L69 (1978).

[11] J. Raynal, J. Math. Phys. **19**, 467 (1978); **20**, 2398 (1979).

[12] See E. G. Anastasselou and N. I. Ioakimidis, J. Math. Phys. **25**, 2422 (1984) for references to the papers of C. E. Siewert and E. E. Burniston.

Bound states of N particles: A variational approach

J. Fernando Perez, C. P. Malta, and F. A. B. Coutinho
Departamento de Física Matemática, Instituto de Física da Universidade de São Paulo, CP 20516, 01000 São Paulo, SP, Brazil

(Received 9 October 1984; accepted for publication 8 March 1985)

Using a variational technique we provide sufficient conditions for the existence of a bound state in a system of N particles in one, two, and three dimensions. Our assumptions imply two bound clusters' thresholds.

I. INTRODUCTION

In this paper we explore the variational method for obtaining sufficient conditions for the existence of a bound state in a system of N particles in ν dimensions interacting via two-body potentials $V_{ij}(\mathbf{r}_i - \mathbf{r}_j)$.

Let H_N denote the Hamiltonian of the N-particle system [with center of mass (CM) kinetic energy removed] and ϵ_N the energy of its continuum threshold. If we can find a wave function Φ_N such that

$$(\Phi_N, H_N\,\Phi_N) < \epsilon_N(\Phi_N, \Phi_N), \tag{1}$$

the variational principle guarantees, then, the existence of at least one bound state of N particles (below the continuum).

The difficulty associated with the method lies in the determination of ϵ_N. For locally square-integrable two-body potentials vanishing at infinity, ϵ_N is given by Hunziker's theorem,[1,2] which involves the knowledge of the bound states of all subsystems of the whole system. Denoting by C a cluster, $C \equiv \{i_1,..., i_{n(C)}\} \subset \{1,..., N\}$, and by H^C the Hamiltonian of the subsystem formed by C, after the removal of CM kinetic energy, and by E_0^C the energy of the infimum of the spectrum of H^C, then Hunziker's theorem[1,2] gives

$$\epsilon_N = \inf_{\substack{\{C_1,\dots,C_l\} \\ \bigcup_{i=1}^{l} C_i = \{1,\dots,N\} \\ C_i \cap C_j = \phi,\ i \neq j \\ 1 < l < N}} \sum_{i=1}^{l} E_0^{C_i}, \tag{2}$$

where the infimum is taken over all possible decompositions of $\{1,..., N\}$ into disjoint clusters.

For the two-body problem, $\epsilon_2 = 0$ and therefore it is not difficult to find sufficient conditions on the potential for the existence of bound states. If $N \geqslant 3$, however, $\epsilon_N < 0$, in general, thus making the problem not so simple. To bypass this difficulty we use a recursive procedure on the particle number N to show that there exist simple sufficient conditions on the two-body potentials that ensure the existence of bound states for arbitrary N in ν dimensions ($\nu = 1$, 2, or 3).

The physical idea behind our method is dimension dependent. In three dimensions, it is basically a "two-body mechanism": for a certain class (to be made precise below) of purely attractive two-body interactions, given two *bound* clusters $C_1(n(C_1) = N_1)$ and $C_2(n(C_2) = N_2)$, it is possible to bind them together provided there is at least one pair of particles, each one in a different cluster, which can form a bound state. On the other hand, in one and two dimensions it is a "two-cluster mechanism": for "globally attractive" two-body potentials, i.e., $\int V_{ij}\, d^\nu x < 0$, given two *bound* clusters C_1 and C_2, it is possible to bind them together since the "effective" intercluster potential also satisfies $\int V^{\text{eff}}_{C_1 C_2}\, d^\nu x < 0$. The difference between $\nu = 1$, 2 and $\nu = 3$ cases has its origin in the fact that in one and two dimensions a two-body interaction with $\int V d^\nu x < 0$ always binds[3,4] two particles but in three dimensions this is not the case even if V is purely attractive. However, if the particles are identical (bosons or fermions) we can also show that a N-body system will exhibit bound states even if the two-body system has no bound states, provided N is big enough (along a subsequence). In fact, large particle number favors the existence of bound states of identical particles (bosons or fermions): classically catastrophic potentials (for instance, globally attractive potentials or potentials with an attractive core)[5] remain so in the quantum case, i.e., $\lim_{N\to\infty} (-E_N/N) = \infty$. Based on that, we show that for these potentials there exists an infinite sequence $2 \leqslant N_0 \leqslant N_1 \leqslant \cdots$ such that H_{N_i} has at least one bound state.

This paper is organized as follows: In Sec. II we derive useful sufficient conditions for a two-particle system to have a bound state with energy below $-\alpha^2$ in $\nu = 1$, 2, or 3 dimensions. For $\nu = 1$ or 2 and $\alpha = 0$ we recover the above-mentioned result that if the potential satisfies $\int V d^\nu x < 0$ the system always has a bound state. For $\nu = 3$, and $\alpha = 0$, we obtain a sufficient condition which is simpler than that obtained in Ref. 6; also, if the potential has spherical symmetry we recover Calogero's "best" sufficient condition.[7] We also derive sufficient conditions for the existence of a bound state of a given angular momentum. Moreover, we show that some of the sufficient conditions provided by Calogero[7] are improved by the variational approach.

In Sec. III we derive sufficient conditions for the existence of a bound state in the N-body problem in one, two, and three dimensions. Part of these results have been announced in Refs. 3 and 8.

In Sec. IV we prove our results for a large number of identical particles.

In the Appendix we collect some kinematical facts for N-body systems which are used in this work.

II. BOUND STATES IN THE TWO-BODY PROBLEM

A collection of sufficient conditions on V for the existence of bound states of H_2 with energy below $-\alpha^2$ is obtained by varying the trial function in the inequality (1) (with $\epsilon_2 = -\alpha^2$)

$$(\Phi_R^\alpha, H_2\,\Phi_R^\alpha) < -\alpha^2(\Phi_R^\alpha, \Phi_R^\alpha). \tag{3}$$

Particularly useful sufficient conditions are those expressed in terms of simple integrals of the two-body potential V (see Ref. 6). For instance, very simple conditions are obtained by

 0022-2488/85/092262-06$02.50

taking as a trial a function Φ_R^α that at short range behaves as $e^{-\alpha r}$ and at long range behaves as the solution of the free equation with energy $-\alpha^2$:

$$\Phi_R^\alpha(r) = e^{-\alpha r}\phi(r), \quad \text{for } r<R,$$
$$\Phi_R^\alpha(r) = e^{-\alpha R}\phi(r)\,[H(\alpha r)/H(\alpha R)], \quad \text{for } r>R \tag{4}$$

($r = |\mathbf{r}|$), where $\phi(r)$ is arbitrary and $H(\alpha r)$ is the solution of the modified Helmholtz equation,

$$(-\Delta + \alpha^2)H(\alpha r) = 0. \tag{5}$$

(We are taking energy in units of $\hbar^2/2\mu$, μ being the reduced mass of the two particles.)

So, in one dimension we take

$$\Phi_R^\alpha(x) = e^{-\alpha|x-x_0|}\phi_R(x-x_0), \tag{6}$$

where x_0 is arbitrary and $\phi_R(x-x_0) \in L^2(\mathbb{R}^1)$ is such that for $|x-x_0|<R$, $\phi_R(x-x_0)=1$ and for $|x-x_0|>R$, $\phi_R(x-x_0)$ starts at 1 and $\phi_R(x-x_0)\to 0$ as $|x-x_0|\to\infty$. By making the scaling $\phi_R(x-x_0)\to\beta^{1/2}\phi_R(\beta(x-x_0))$ and letting $\beta\to 0$ we obtain the condition

$$2\alpha + \int_{-\infty}^{\infty} e^{-2\alpha|x-x_0|}V(x)dx < 0.$$

Setting $\alpha = 0$ we recover the known result[3,4] that in one dimension a globally attractive potential, i.e., $\int V dx < 0$, always possesses at least one bound state.

In two dimensions we take

$$\Phi_R^\alpha(r) = e^{-\alpha r}, \quad \text{for } r<R,$$
$$\Phi_R^\alpha(r) = [e^{-\alpha R}/K_0(\alpha R)]\,K_0(\alpha r), \quad \text{for } r>R, \tag{7}$$

where $K_0(\alpha R)$ is the modified Bessel function, and obtain the following sufficient condition for the existence of a bound state of energy less than $-\alpha^2$:

$$-\frac{1}{2\pi}\int_0^{2\pi}d\theta\int_0^R e^{-2\alpha r}V(r,\theta)r\,dr$$
$$-\frac{1}{2\pi}\int_0^{2\pi}d\theta\int_R^\infty e^{-2\alpha R}\frac{K_0(\alpha r)}{K_0(\alpha R)}V(r,\theta)r\,dr$$
$$\geqslant\left(\frac{1-e^{-2\alpha R}}{2}\right) - \alpha R e^{-2\alpha R}\left\{1+\frac{K_0'(\alpha R)}{K_0(\alpha R)}\right\}, \tag{8}$$

where

$$K_0'(\alpha R) = \left.\frac{dK_0(\alpha r)}{d(\alpha r)}\right|_{\alpha r=\alpha R}.$$

Again, setting $\alpha = 0$ we recover the result[3,4] that in two dimensions a globally attractive potential ($\int V d^2x < 0$) always possesses at least one bound state.

Finally, in three dimensions we take

$$\Phi_R^\alpha(r) = (1/R^{1/2})e^{-\alpha r}, \quad \text{for } r<R,$$
$$\Phi_R^\alpha(r) = R^{1/2}(e^{-\alpha r}/r), \quad \text{for } r>R, \tag{9}$$

obtaining the following sufficient conditions for the existence of a bound-state of energy less than $-\alpha^2$:

$$-\frac{1}{4\pi}\int d\Omega\int_0^R\frac{e^{-2\alpha R}}{R}V(r,\Omega)r^2\,dr$$
$$-\frac{1}{4\pi}\int d\Omega\int_R^\infty R\,\frac{e^{-2\alpha r}}{r^2}V(r,\Omega)r^2\,dr \geqslant \frac{1-e^{-2\alpha R}}{2\alpha R}. \tag{10}$$

Setting $\alpha = 0$ obtains

$$-\frac{1}{4\pi}\int d\Omega\int_0^R\frac{1}{R}V(r,\Omega)r^2\,dr$$
$$-\frac{1}{4\pi}\int d\Omega\int_R^\infty RV(r,\Omega)dr \geqslant 1, \tag{11}$$

which is simpler than the condition obtained by Chadan and Martin.[6] In the particular case of spherically symmetric potential, condition (11) reduces to Calogero's[7] "best" sufficient condition.

We shall now present the variational version of the other sufficient conditions ($\alpha = 0$) derived by Calogero in Ref. 7. Taking as a trial function $\Phi_R(r) = R^{1/2}/(r+R)$, a sufficient condition for a spherically symmetric potential to hold a bound state is

$$\int_0^\infty\frac{R}{(r+R)^2}V(r)r^2\,dr < -\frac{1}{3}. \tag{12}$$

Taking $\Phi_R(r) = (R^{1/2}/r)(1-e^{-r/R})$ as the trial function we get the condition

$$\int_0^\infty\frac{R}{r^2}(1-e^{-r/R})^2\,V(r)r^2\,dr < -\frac{1}{2}. \tag{13}$$

Conditions (12) and (13) should be compared with Calogero's[7] condition (3.15) and (3.17), respectively. In both cases the variational method produced improvement.

Remark: The fact that, for $\alpha = 0$, the trial function is not square integrable is of no importance: if $\lim_{\alpha\to 0}(\Phi_R^\alpha, H_2\Phi_R^\alpha) < 0$ for a sequence Φ_R^α, then, for α sufficiently small, $(\Phi_R^\alpha, H_2\Phi_R^\alpha) < 0$ and the variational principle guarantees the existence of a bound state.

For spherically symmetric potentials, our method can be readily adapted to provide sufficient conditions for the existence of a bound state having energy less than $-\alpha^2$ and of a given angular momentum l. Our recipe in $\nu = 3$ is then to use as the trial function the regular and irregular solutions of the modified Bessel equation,

$$-\frac{d^2\chi_l}{dr^2} - \frac{2}{r}\frac{d\chi_l}{dr} + \frac{l(l+1)}{r^2}\chi_l + \alpha^2\chi_l = 0, \tag{14}$$

matched at an arbitrary point $r = R$. Thus, the trial function is

$$\Phi_R^l(r) = \frac{(\alpha R)^{1/2}}{K_{l+1/2}(\alpha R)}\frac{K_{l+1/2}(\alpha r)}{(\alpha r)^{1/2}}, \quad \text{for } r<R,$$
$$\Phi_R^l(r) = \frac{(\alpha R)^{1/2}}{I_{l+1/2}(\alpha R)}\frac{I_{l+1/2}(\alpha r)}{(\alpha r)^{1/2}}, \quad \text{for } r>R, \tag{15}$$

providing the following sufficient condition:

$$-\frac{1}{4\pi}\int d\Omega\int_0^\infty|\Phi_R^l|^2\,V(r,\Omega)r^2\,dr$$
$$\geqslant \alpha R^2\left\{\frac{(l+1)}{\alpha R} + \frac{K_{l-1/2}(\alpha R)}{K_{l+1/2}(\alpha R)}\right\}$$
$$+\alpha R^2\left\{\frac{l}{\alpha R}\frac{I_{l+3/2}(\alpha R)}{I_{l+1/2}(\alpha R)}\right\}. \tag{16}$$

In the limit $\alpha \to 0$, the above condition reduces to Calogero's "best"[6] sufficient condition

$$-\frac{1}{R}\int_0^R \left(\frac{r}{R}\right)^{2l} V(r) r^2\,dr - \frac{1}{R}\int^{\infty} \left(\frac{R}{r}\right)^{2l+2} V(r) r^2\,dr \geqslant 2l+1. \quad (17)$$

Remark: The technical assumption on V needed for the validity of our arguments is $V \in L^2_{\text{loc}}(\mathbb{R}^\nu)$ and $\lim_{r\to\infty} V(\mathbf{r}) = 0$. This ensures that (a) Hunziker's theorem,[1,2] i.e., $\epsilon_2 = 0$, holds and (b) since $L^2_{\text{loc}}(\mathbb{R}^\nu) \subset L^1_{\text{loc}}(\mathbb{R}^\nu)$ all integrals are well defined or equivalently, all trial wave functions Φ^α_R are in the form domain. Notice, also, that the condition of being "globally attractive," $\int V d^\nu x < 0$, can be generalized as follows: there exist $R > 0$ and $I > 0$ such that

$$\int_{|\mathbf{x}|<R'} V(\mathbf{x}) d^\nu x \leqslant -I, \quad \text{for all } R' \geqslant R, \quad (18)$$

thus avoiding the requirement of integrability.

III. BOUND STATES IN THE N-BODY PROBLEM[9]

Sufficient conditions for the existence of bound states for all $N \geqslant 2$ are derived inductively on N, that is, assuming that a certain sufficient condition for the existence of bound states for $N = 2$ is verified, and assuming the existence of a bound state of N particles, we prove the existence of a bound state of $N+1$ particles.

A. One and two dimensions

Let all two-body interactions V_{ij} be globally attractive, i.e., $\int V_{ij}\, d^\nu x < 0$. Consider now the $(N+1)$-body system. Denoting by $\mathbf{r}_i$ and m_i, $i = 1,\ldots, N+1$, the particles' coordinates and masses, respectively, and introducing Jacobi coordinates[10]

$$\xi_i = \mathbf{r}_{i+1} - \left(\sum_{j\leqslant i} m_j\right)^{-1} \left(\sum_{j \leqslant i} m_j\, \mathbf{r}_j\right), \quad i = 1,\ldots, N,$$

the Hamiltonian H_{N+1} reads

$$H_{N+1} = -\sum_{i=1}^{N} \frac{1}{2\mu_i} \Delta_{\xi_i} + \sum_{i<j} V_{ij}\,(\mathbf{r}_i - \mathbf{r}_j),$$

where

$$\mu_i^{-1} = m_{i+1}^{-1} + \left(\sum_{j\leqslant i} m_j\right)^{-1},$$

that is,

$$H_{N+1} = H_N - (2\mu_N)^{-1} \Delta_{\xi_N} + \sum_{i=1}^{N} V_{i,N+1}(\mathbf{r}_{N+1} - \mathbf{r}_i),$$

where the Hamiltonian H_N involves only the coordinates $\xi_1,\ldots, \xi_{N-1}$.

Let E_N be the energy of the bound state of N particles (with $E_N < \epsilon_N$) and $\Phi_N(\xi_1,\ldots, \xi_{N-1})$ its normalized wave function. Consider then

$$\Phi(\xi_1,\ldots, \xi_{N-1}, \xi_N) = \Phi_N(\xi_1,\ldots, \xi_{N-1})\, \phi(\xi_N),$$

where ϕ is going to be conveniently chosen.

It is clear that

$$(\Phi, H_{N+1}\, \Phi) = E_N + (\phi, (H_0 + U_N)\phi),$$

where $H_0 = -\Delta_{\xi_N}/2\mu_N$ and

$$U_N(\xi_N) = \sum_{i=1}^{N} \int |\Phi_N(\xi_1,\ldots, \xi_{N-1})|^2 \times V_{i,N+1}(\mathbf{r}_{N+1} - \mathbf{r}_i) d^\nu \xi_1 \cdots d^\nu \xi_{N-1}$$

is the "effective" potential seen by the $(N+1)$th particle in the presence of the bound state of the other N particles. The important property of the effective potential U_N is that it is also globally attractive. This follows from

$$\int U_N(\xi_N) d^\nu \xi_N = \sum_{i=1}^{N} \int d^\nu\xi_1 \cdots d^\nu \xi_{N-1}\, |\Phi_N(\xi_1,\ldots, \xi_{N-1})|^2 \times \int d^\nu\xi_N\, V_{i,N+1}(\xi_N + \mathbf{u}_i),$$

where $\mathbf{u}_i$ is a linear combination of $\xi_1,\ldots, \xi_{N-1}$. The integral in ξ_N can be performed and is independent of $\mathbf{u}_i$, and since $\int |\Phi_N|^2 d^\nu \xi_1 \cdots d^\nu \xi_{N-1} = 1$, we obtain

$$\int U_N(\xi_N) d^\nu \xi_N = \sum_{i=1}^{N} \left[\int V_{i,N+1}(\mathbf{x}) d^\nu \mathbf{x}\right],$$

which proves the statement. By a simple limiting argument, this result follows even if we use the more general definition of globally attractive potentials introduced in (18).

So, from the discussion in Sec. II it then follows that ϕ can be chosen such that $(\phi, (H_0 + U_N)\phi) < 0$. For this choice of ϕ we have

$$(\Phi, H_{N+1}\, \Phi) < E_N < \epsilon_N.$$

Notice that the arguments can be repeated for any decomposition of the $(N+1)$-body system into two clusters of N and 1 particle(s), respectively. So, numbering the particles in such a way that E_N is the smallest energy of all N-body bound states it then follows that (if $\epsilon_{N+1} = E_N$) $(\Phi, H_{N+1}\, \Phi) < \epsilon_{N+1}$. In the general case, the continuum is not necessarily given by E_N (not even by a two bound cluster decomposition). However, the discussion below will inductively imply that the threshold is given by a two bound cluster decomposition. Assume that there exists a decomposition of the system into two disjoint clusters

$$C_1 = \{i_1,\ldots, i_{N_1}\}, \quad C_2 = \{j_1,\ldots, j_{N_2}\}, \quad N_1 + N_2 = N,$$

both admitting bound states with energies E^{C_1} and E^{C_2} (below the respective continuum thresholds) such that the "effective" intercluster potential

$$V^{\text{eff}}_{C_1 C_2}(\xi) = \sum_{\substack{i\in C_1\\ j\in C_2}} V_{ij}(\xi) \quad (19)$$

is globally attractive and such that the continuum spectrum of H_N starts at $\epsilon_N = E^{C_1} + E^{C_2}$.

In fact, the Hamiltonian H_N can be written as

$$H_N = H^{C_1} + H^{C_2} + \left[H_0 + \sum_{\substack{i\in C_1\\ j\in C_2}} V_{ij}(\mathbf{r}_i - \mathbf{r}_j)\right],$$

where

$$H_0 = -\frac{\Delta_\xi}{2\mu}, \quad \mu^{-1} = \left(\sum_{i\in C_1} m_i\right)^{-1} + \left(\sum_{j\in C_2} m_j\right)^{-1},$$

and $\boldsymbol{\xi}$ denotes the position of the CM of C_2 with respect to the CM of C_1.

Consider now

$$\Phi(\xi_1^{C_1},\ldots,\xi_{N_1-1}^{C_1},\xi_1^{C_2},\ldots,\xi_{N_2-1}^{C_2},\xi) = \Phi^{C_1}(\xi_1^{C_1},\ldots,\xi_{N_1-1}^{C_1})\,\Phi^{C_2}(\xi_1^{C_2},\ldots,\xi_{N_2-1}^{C_2})\,\phi(\xi),$$

where $\xi_i^{C_l}$, $i=1,\ldots,N_{l-1}$, are the Jacobi coordinates for cluster C_l, $l=1,2$, Φ^{C_l} are the normalized eigenfunctions of H^{C_l} (with energies E^{C_l}) and, as before, ϕ will be conveniently chosen. For this trial function we have

$$(\Phi, H_N\,\Phi) = E^{C_1} + E^{C_2} + (\phi,(H_0 + V^{\rm eff}_{C_1C_2})\,\phi),$$

where

$$V^{\rm eff}_{C_1C_2}(\xi) = \sum_{\substack{i\in C_1\\ j\in C_2}} \int |\Phi^{C_1}(\xi_1^{C_1},\ldots,\xi_{N_1-1}^{C_1})|^2 \times |\Phi^{C_2}(\xi_1^{C_2},\ldots,\xi_{N_2-1}^{C_2})|^2\, V_{ij}(\mathbf{r}_i-\mathbf{r}_j) \times \prod_{k=1}^{N_1-1} d^\nu \xi_k^{C_1} \prod_{l=1}^{N_2-1} d^\nu \xi_l^{C_2}$$

is the "effective" intercluster potential (19). This follows from

$$\int V^{\rm eff}_{C_1C_2}(\xi)\,d^\nu\xi = \sum_{\substack{i\in C_1\\ j\in C_2}} \int \prod_{k=1}^{N_1-1} d^\nu\xi_k^{C_1}\, |\Phi^{C_1}(\xi_1^{C_1},\ldots,\xi_{N_1-1}^{C_1})|^2 \times \int \prod_{l=1}^{N_2-1} d^\nu\xi_l^{C_2}\, |\Phi^{C_2}(\xi_1^{C_2},\ldots,\xi_{N_2-1}^{C_2})|^2 \times \int d^\nu\xi\, V_{ij}(\xi + \mathbf{u}_i^{C_1} + \mathbf{u}_j^{C_2}),$$

where $\mathbf{u}_i^{C_l}$, $l=1,2$, are linear combinations of $\xi_1^{C_l},\ldots,\xi_{N_l-1}^{C_l}$. The integral in ξ can be performed and is independent of $\mathbf{u}_i^{C_l}$, and as the functions Φ_l^C are normalized, we obtain

$$\int V^{\rm eff}_{C_1C_2}(\xi)\,d^\nu\xi = \sum_{\substack{i\in C_1\\ j\in C_2}} \int V_{ij}(\xi)\,d^\nu\xi.$$

So, as the "effective" intercluster potential $V^{\rm eff}_{C_1C_2}$ is assumed to be globally attractive, from the discussion in Sec. II it follows that ϕ can be chosen such that $(\phi,(H_0+V^{\rm eff}_{C_1C_2})\phi) < 0$. For this choice of ϕ we have

$$(\phi, H_N\,\phi) < E^{C_1} + E^{C_2} = \epsilon_N,$$

thus concluding the proof.

Again, by a simple limiting argument this result follows even if we use the more general definition of globally attractive potentials introduced in (18).

B. Three dimensions

Let $C_1 = \{i_1,\ldots,i_{N_1}\}$ and $C_2 = \{j_1,\ldots,j_{N_2}\}$ be two disjoint clusters, $N_1+N_2=N$. The following set of relative coordinates will prove suitable for displaying the binding mechanism that we exploit in $\nu=3$:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{r}_{i_k} - \mathbf{r}_{i_{N_1}}, \quad i=1,\ldots,N_1-1,\\ \mathbf{y}_l &= \mathbf{r}_{j_l} - \mathbf{r}_{j_{N_2}}, \quad j=1,\ldots,N_2-1,\\ \mathbf{z} &= \mathbf{r}_{i_{N_1}} - \mathbf{r}_{j_{N_2}}, \end{aligned} \tag{20}$$

i.e., inside each cluster C_l, $l=1,2$, we single out a given particle (for simplicity we always make it the last one in the cluster) and take coordinates relative to these particles.

In the Appendix we show that the Hamiltonian H_N written in terms of (20) is given by

$$H_N = H^{C_1} + H^{C_2} + H^D + \widetilde{V}_{C_1C_2} + T_{H-E}, \tag{21}$$

where D is the two-body cluster

$$D = \{i_{N_1}, j_{N_2}\} \equiv \{N_1, N\},$$

$$\widetilde{V}_{C_1C_2} = \sum_{\substack{k=1,\ldots,N_1-1\\ l=1,\ldots,N_2-1}} V_{i_kj_l}(\mathbf{x}_k - \mathbf{y}_l + \mathbf{z}) \tag{22}$$

is the intercluster interaction excluding V_{N_1N} already included in H^D, and

$$T_{H-E} = \sum_{i=1}^{N_1-1} \frac{\mathbf{p}\cdot\mathbf{k}_i}{2m_{N_1}} - \sum_{j=1}^{N_2-1} \frac{\mathbf{p}\cdot\mathbf{q}_j}{2m_N} \tag{23}$$

is the Hughes–Eckart kinetic energy.[1] Here $\mathbf{k}_i$, $\mathbf{q}_j$, and $\mathbf{p}$ denote the canonically conjugate momenta to $\mathbf{x}_i$, $\mathbf{y}_j$, and $\mathbf{z}$, respectively.

Let us now assume that there are eigenstates Φ^{C_i} of H^{C_i}, with energies $E^{C_i} < \epsilon^{C_i}$, $i=1,2$, where ϵ^C denotes the continuum threshold of H^C.

Considering then a state Φ_N of the N-body system of the form

$$\Phi_N = \Phi^{C_1}(\mathbf{x}_1,\ldots,\mathbf{x}_{N_1-1})\,\Phi^{C_2}(\mathbf{y}_1,\ldots,\mathbf{y}_{N_2-1})\,\phi(\mathbf{z}), \tag{24}$$

we get

$$(\Phi_N, H_N\,\Phi_N) = E^{C_1} + E^{C_2} + (\phi, H^D\phi) + (\Phi_N, \widetilde{V}_{C_1C_2}\,\Phi_N) + (\Phi_N, T_{H-E}\,\Phi_N). \tag{25}$$

Now, for purely attractive potentials V_{ij},

$$(\Phi_N, \widetilde{V}_{C_1C_2}\,\Phi_N) \leqslant 0, \tag{26}$$

and, if it is possible to choose ϕ such that

$$(\phi, H^D\phi) \leqslant 0 \tag{27}$$

and

$$(\phi, \mathbf{p}\,\phi) = 0, \tag{28}$$

then

$$(\Phi_N, H_N\,\Phi_N) < E^{C_1} + E^{C_2}. \tag{29}$$

A few comments are in order.

(a) Condition (27) is a sufficient condition for the existence of a bound state for the cluster $D \equiv \{N_1, N\}$.

(b) Condition (28) follows from symmetry requirements on ϕ.

(c) Equation (27) is automatically satisfied if the V_{ij}'s are central potentials and ϕ is taken to be a bound state of H^D with well-defined angular momentum.

(d) For noncentral potentials a sufficient condition for the possibility of choosing ϕ satisfying (27) is, for instance, (11).

Under the above assumptions on V_{ij} we can draw the following conclusions.[8]

(i) The threshold for the continuum spectrum is given by a two-cluster breakup

$$\epsilon_N = \inf_{\substack{C_1 \cup C_2 = \{1,\ldots,N\} \\ C_1 \cap C_2 = \phi}} (E_0^{C_1} + E_0^{C_2}), \tag{30}$$

where E_0^C denotes the ground state energy (the infimum of the spectrum) of H^C.

(ii) For all $N \geqslant 2$ there exists an eigenstate of H_N with energy $E_N < \epsilon_N$.

Proof: (i) Suppose the minimum in (2) were attained at $l \geqslant 3$, i.e.,

$$\epsilon_N = E_0^{C_1} + E_0^{C_2} + \cdots + E_0^{C_l}.$$

Now, from (29) it follows that $E_0^{C_1 \cup C_2} < E_0^{C_1} + E_0^{C_2}$, therefore the only possibility left is $l = 2$.

(ii) Follows trivially from (1) and (29).

IV. IDENTICAL PARTICLES—LARGE N RESULTS

The results of the previous section have one thing in common: they all assume the existence of bound states of two-particles. However, even if the two-body interaction is not strong enough for binding two-particles, there is a class of interactions for which we can prove the existence of bound states of N particles, provided N is large enough. To prove this let us consider a system of N identical bosons of mass m interacting via a two-body potential V. Using relative coordinates with respect to the Nth particle, $\mathbf{x}_i = \mathbf{r}_i - \mathbf{r}_N$, $i = 1,\ldots, N-1$, the Hamiltonian H_N reads

$$H_N = \sum_{i=1}^{N-1} \frac{\mathbf{k}_i^2}{m} + \sum_{i=1}^{N-1} V(\mathbf{x}_i) + \frac{1}{2} \sum_{i \neq j = 1}^{N-1} V(\mathbf{x}_i - \mathbf{x}_j) + \frac{1}{2} \sum_{i \neq j = 1}^{N-1} \frac{\mathbf{k}_i \cdot \mathbf{k}_j}{2m}. \tag{31}$$

For the trial wave function

$$\psi_N(\mathbf{x}_1,\ldots, \mathbf{x}_{N-1}) = \prod_{i=1}^{N-1} \phi(\mathbf{x}_i), \tag{32}$$

we have

$$(\psi_N, H_N \psi_N) = (N-1)\{(\phi, H_2 \phi) + [(N-2)/2]u(\phi)\}, \tag{33}$$

where

$$u(\phi) = \int |\phi(\mathbf{x}_1)\phi(\mathbf{x}_2)|^2 V(\mathbf{x}_1 - \mathbf{x}_2)\, d^3x_1\, d^3x_2. \tag{34}$$

Notice that the Hughes–Eckart terms disappear by symmetry.

It is clear from (33) that if there is a $\phi \in L^2(\mathbb{R}^\nu)$ such that $u(\phi) < 0$, then for $N > N_0(\phi) = 2\{1 + (\phi, H_2 \phi)/|u(\phi)|\}$, it follows that $(\psi_N, H_N \psi_N) < 0$. So, for some $N_0 \leqslant \inf_{\|\phi\|} N_0(\phi)$ (by Hunziker's theorem) H_{N_0} will have a bound state. Sufficient conditions on V for this to happen are as follows.

(a) V is purely attractive ($V(\mathbf{x}) \leqslant 0$); in this case $u(\phi) \leqslant 0$ for all ϕ.

(b) V has an attractive core, i.e., $V(\mathbf{x}) \leqslant 0$ for $|\mathbf{x}| \leqslant R$. Choosing $\phi(\mathbf{x}) = 0$ for $|\mathbf{x}| \geqslant R/2$ we have $u(\phi) \leqslant 0$.

(c) V is globally attractive. Let $\nu = 3$ (the case $\nu = 1$ or 2 has been discussed in Sec. III A). For $\phi_\beta(\mathbf{x}) = \beta^{\nu/2}\phi(\beta\mathbf{x})$ we have

$$\begin{aligned} u(\phi_\beta) &= \beta^{2\nu} \int V(\mathbf{x} - \mathbf{y}) |\phi(\beta\mathbf{x})|^2 |\phi(\beta\mathbf{y})|^2 d^\nu x\, d^\nu y \\ &= \beta^\nu \int V\left(\mathbf{x} - \frac{\mathbf{y}}{\beta}\right) |\phi(\beta\mathbf{x})|^2 |\phi(\mathbf{y})|^2 d^\nu x\, d^\nu y \\ &= \beta^\nu \int d^\nu y |\phi(\mathbf{y})|^2 \left[\int V(\omega) |\phi(\beta\omega + \mathbf{y})|^2 d^\nu \omega\right], \end{aligned}$$

where $\omega = \mathbf{x} - \mathbf{y}/\beta$. Since

$$\lim_{\beta \to 0} \left[\int V(\omega) |\phi(\beta\omega + \mathbf{y})|^2 d^\nu \omega\right] = |\phi(\mathbf{y})|^2 \int V(\omega) d^\nu \omega,$$

if $\int V(\omega) d^\nu \omega < 0$ then $u(\phi_\beta) < 0$ for β small enough.

We now show the existence of an infinite sequence $2 \leqslant N_0 \leqslant N_1 \cdots \leqslant N_n$, such that H_{N_i} has at least one bound state. Suppose this sequence is finite, i.e., $\exists\, L$ such that for $N > N_L$, H_N has no bound state. In fact, by Hunziker's theorem,[1,2] the continuum threshold is given by

$$\epsilon_N = \sum_{i=1}^{L} k_i E_{N_i}, \quad k_i \text{ integer},$$

where

$$H_{N_i} \psi_{N_i} = E_{N_i} \psi_{N_i}.$$

Now, since

$$\sum_{i=1}^{L} k_i N_i \leqslant N,$$

it follows that

$$k_i \leqslant N, \quad i = 1,\ldots, L,$$

so that

$$\epsilon_N > N \sum_{i=1}^{L} E_{N_i}. \tag{35}$$

As we have seen, (33), in the case of bosons, $(\psi_N, H_N \psi_N)$ has $-CN^2$ $(C > 0)$ for an upper bound. Due to the linear dependence of ϵ_N on N it follows that there exists a $N > N_L$ such that H_N has at least one bound state, thus proving that the sequence of N_i such that H_{N_i} has at least one bound state is infinite.

In fact all these interactions are catastrophic or collapsing since the binding energy per particle diverges as $N \to \infty$. Any two-body interaction is catastrophic if it is not *stable*[5] [a two-body interaction V is said to be stable if there exists a constant $B \geqslant 0$ such that $U(\mathbf{r}_1,\ldots,\mathbf{r}_N) = \frac{1}{2}\sum_{i \neq j = 1}^{N} V(\mathbf{r}_i - \mathbf{r}_j) \geqslant -BN$ for all $N \geqslant 2$ and for all $\mathbf{r}_1,\ldots, \mathbf{r}_N \in \mathbb{R}^\nu$]. As proved in p. 35 of Ruelle's book,[5] if the interaction is not stable there exists a sequence of integers $N_1 < N_2 < \cdots < N_k < N_{k+1} \cdots$, a sequence of points $\xi_1,\ldots, \xi_k \in \mathbb{R}^\nu$, and constants $C > 0$, $\delta > 0$ such that

$$U(\mathbf{r}_1 = 0,\ldots,\mathbf{r}_{N_k}) \leqslant -CN_k^2, \quad \text{if } |\mathbf{r}_i - \xi_i| < \delta \quad \text{for all } i.$$

Then, it is easy to construct a sequence of normalized wave functions such that

$$(\psi_{N_k}, V_{N_k} \psi_{N_k}) < -CN_k^2$$

and

$$(\psi_{N_k}, H_0 \psi_{N_k}) \leqslant \begin{cases} C' N_k & \text{(bosons)}, \\ C'' N_k^{(1 + 2/\nu)} & \text{(fermions)}. \end{cases}$$

Therefore there exists a $k_0 \gg 1$ such that $(\psi, H_{N_{k_0}} \psi) < 0$ and so there exists $2 < N_0 < N_{k_0}$ such that for fermions H_{N_0} has at least a bound state if $\nu \geqslant 3$ (and for bosons if $\nu \geqslant 1$, as already seen).

ACKNOWLEDGMENTS

Two of us (J. F. P. and F. A. B. C.) were partially supported by CNPq.

APPENDIX: KINEMATICS FOR $\nu = 3$

Here we derive expression (21) of H_N in terms of the relative coordinates $\mathbf{x}_i$, $\mathbf{y}_j$, and $\mathbf{z}$ defined in (20). The only thing we have to do to obtain expression (21) is to express the kinetic energy in terms of the momenta $\mathbf{k}_i$, $\mathbf{q}_j$, and $\mathbf{p}$ canonically conjugate to $\mathbf{x}_i$, $\mathbf{y}_j$, and $\mathbf{z}$, respectively. This is easily done since the transformation from the momenta $\mathbf{k}_i$, $\mathbf{q}_j$, and $\mathbf{p}$ to the momenta $\mathbf{p}_{r_i}$ canonically conjugate to $\mathbf{r}_i$ is given by

$$\begin{bmatrix} \mathbf{p}_{r_1} \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{p}_{r_{N_1}} \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{p}_{r_N} \end{bmatrix} = [\tilde{B}] \begin{bmatrix} \mathbf{P}_{\mathrm{CM}} = 0 \\ \mathbf{k}_{i_1} \\ \vdots \\ \mathbf{k}_{i_{N_1-1}} \\ \vdots \\ \mathbf{q}_{j_i} \\ \vdots \\ \mathbf{q}_{j_{N_2-1}} \\ \mathbf{p} \end{bmatrix}, \tag{A1}$$

where $\mathbf{P}_{\mathrm{CM}} = 0$ is the CM momentum canonically conjugate to

$$\mathbf{R}_{\mathrm{CM}} = 0 = \sum_{i=1}^{N} \frac{m_i \mathbf{r}_i}{M} \quad \left(M = \sum_{i=1}^{N} m_i \right)$$

and the matrix $\tilde{B}$ is the transpose of the matrix B defining transformation (20):

$$\begin{bmatrix} \mathbf{R}_{\mathrm{CM}} = 0 \\ \mathbf{x}_{i_1} \\ \vdots \\ \mathbf{x}_{i_{N_1-1}} \\ \mathbf{y}_{j_1} \\ \vdots \\ \mathbf{y}_{j_{N_2-1}} \\ \mathbf{z} \end{bmatrix} = \begin{bmatrix} m_1/M & m_2/M & \cdots & \cdots & & m_{N_1}/M & & \cdots & \cdots & & m_N/M \\ 1 & 0 & \cdots & \cdots & 0 & -1 & 0 & \cdots & \cdots & & 0 \\ \vdots & \vdots & & & \vdots & \vdots & \vdots & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & 1 & -1 & 0 & \cdots & \cdots & & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & -1 \\ \vdots & & & & \vdots & \vdots & \vdots & & & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 0 & \cdots & \cdots & 1 & -1 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & 0 & \cdots & \cdots & 0 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{r}_{N_1} \\ \vdots \\ \vdots \\ \mathbf{r}_N \end{bmatrix}.$$

$$H_N = H^{C_1} + H^{C_2} + H^D + \tilde{V}_{C_1C_2} + T_{H-E},$$

where

$$H^{C_1} = \sum_{l=1}^{N_1-1} \frac{\mathbf{k}_l^2}{2\mu_{N_1 i_l}} + \sum_{k>l=1}^{N_1-1} \frac{\mathbf{k}_k \cdot \mathbf{k}_l}{m_{N_1}} + \sum_{l=1}^{N_1-1} V_{i_l N_1}(\mathbf{x}_l) + \sum_{k>l=1}^{N_1-1} V_{i_k i_l}(\mathbf{x}_k - \mathbf{x}_l),$$

$$H^{C_2} = \sum_{l=1}^{N_2-1} \frac{\mathbf{q}_l^2}{2\mu_{N j_l}} + \sum_{k>l=1}^{N_2-1} \frac{\mathbf{q}_k \cdot \mathbf{q}_l}{m_N} + \sum_{l=1}^{N_2-1} V_{j_l N}(\mathbf{y}_l) + \sum_{k>l=1}^{N_2-1} V_{j_k j_l}(\mathbf{y}_k - \mathbf{y}_l),$$

$$H^D = \frac{\mathbf{p}^2}{2\mu_{N_1 N}} + V(\mathbf{z}),$$

$$T_{H-E} = \sum_{l=1}^{N_1-1} \frac{\mathbf{p} \cdot \mathbf{k}_l}{2m_{N_1}} - \sum_{l=1}^{N_2-1} \frac{\mathbf{p} \cdot \mathbf{q}_l}{2m_N},$$

$$\tilde{V}_{C_1C_2} = \sum_{\substack{k=1,\ldots,N_1-1 \\ l=1,\ldots,N_2-1}} V_{i_k j_l}(\mathbf{x}_k - \mathbf{y}_l + \mathbf{z}),$$

μ_{ij} being the reduced masses.

[1] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1978), Vol. IV.

[2] W. Hunziker, Helv. Phys. Acta **39**, 451 (1966).

[3] F. A. B. Coutinho, C. P. Malta, and J. Fernando Perez, Phys. Lett. A **97**, 242 (1983).

[4] B. Simon, Ann. Phys. **97**, 279 (1976); M. Klaus, Ann. Phys. **108**, 288 (1977).

[5] D. Ruelle, *Statistical Mechanics*, Mathematical Physics Monograph Series (Benjamin, New York, 1969).

[6] K. Chadan and A. Martin, J. Phys. Lett. **41**, L-205 (1980).

[7] F. Calogero, J. Math. Phys. **6**, 161 (1965).

[8] F. A. B. Coutinho, C. P. Malta, and J. Fernando Perez, Phys. Lett. A **100**, 460 (1984).

[9] From now on we shall use $\hbar = 1$.

[10] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1979), Vol. III.

Static relativistic perfect fluids with spherical, plane, or hyperbolic symmetry

C. B. Collins
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

(Received 17 January 1985; accepted for publication 8 March 1985)

This article examines the Einstein field of equations of general relativity, when the source of the gravitational field is a perfect fluid, and the geometry is static and possesses spherical, plane, or hyperbolic symmetry. This examination unifies, extends, and amends some earlier works. It is shown that a previous qualitative treatment of static spherically symmetric perfect fluids that obey a γ-law equation of state can be extended to include the cases of plane and hyperbolic symmetry. In the case of plane symmetry, the exact solution is provided for general values of γ. This indicates defects in an earlier prescription that was given for a general equation of state.

I. INTRODUCTION

The first solution of Einstein's field equations of general relativity was provided by Schwarzschild[1] almost seventy years ago, when he published details of the static, spherically symmetric vacuum metric that now bears his name. Since that time, there have been numerous attempts to study the consequences of relaxing at least one of the conditions that originally characterized the Schwarzschild space-time (see, e.g., the articles cited in Ref. 2). In the particular case where a perfect fluid source for a static spherically symmetric gravitational field is introduced, there have been many exact solutions given, although, to the best of my knowledge, the most general solution has not been obtained. Frequently, for mathematical convenience, computations are performed prior to imposing conditions on the reasonability of the equation of state, with the unfortunate consequence that the resulting fluid might very well be physical in only a local region of space-time, or indeed it might be unphysical everywhere. An alternative line of attack is first to impose an equation of state, and then to attempt to solve, or somehow to analyze, the resulting field equations. This generally leads to little advance from an analytic standpoint, and so investigators usually resort directly to numerical techniques. However, if the fluid obeys a γ-law equation of state, i.e., the energy density μ and the pressure p of the fluid are related by an equation of state of the form $p = (\gamma - 1)\mu$, where γ is a constant (which, for physical reasons,[3] satisfies the inequality $1 \leqslant \gamma \leqslant 2$), some analytical progress can be made. Under the given circumstances, the field equations may in general be reduced to a "plane autonomous system," whose solutions can be examined qualitatively by means of fairly standard techniques (see, e.g., Refs. 4–6). This was the approach used earlier by the present author,[7] and it has several advantages. First, it highlights a special solution, in which the relevant variables are constant; this solution is often attributed to Misner and Zapolsky,[7–9] although it is a special case of the class VI solutions of Tolman,[10] and has been discussed elsewhere by Wyman[11] and others.[2] Another advantage is that the qualitative method yields a vivid pictorial description of *all* solutions, thereby providing an understanding of the "evolution" of the physical variables (in terms of spatial distance from the center of symmetry), and, in particular, giving fairly direct information about the "asymptotic" behavior (i.e., far from and near to the center of symmetry). It will be shown in the present article that this qualitative description can be readily extended to the cases of plane and hyperbolic symmetry. In fact, the case of hyperbolic symmetry is related by means of a complex transformation to that of spherical symmetry, and so the system of field equations reduces to that previously considered, although its examination must be made on a different domain. In the case of plane symmetry, the field equations simplify to a greater extent, thereby permitting an exact general solution.

The particular choice of variables that reduces the Einstein field equations to a plane autonomous system, and which thus renders them amenable to the qualitative technique, can be motivated by an examination of "quasihomologous" transformations, as was done in Ref. 7 in the spherically symmetric case. It is not of course necessary to do this, since the variables could always be introduced in an *ad hoc* manner. Nevertheless, these transformations will be considered, since the attendant discussion yields a more systematic derivation, and in particular provides a somewhat quantitative measure of the simplifications that result in the field equations in the plane symmetric case.

In the following, it will be assumed that the fluid obeys an equation of state $p = p(\mu)$. This represents only a minor restriction, since the symmetries that are involved necessitate such an equation, except in the case when μ is constant throughout space-time. In fact, it will be convenient to focus attention on those space-times in which neither μ nor p is identically constant; this thereby requires the existence of an equation of state $p = p(\mu)$, while at the same time it forbids the situation in which p is also constant, as occurs, for instance, with dust ($p = 0$). Furthermore, I shall suppose that the cosmological constant is zero. Thus the following discussion is not concerned with space-times such as the Schwarzschild interior solution, the generalized Einstein static solution, and their possible counterparts with plane or hyperbolic symmetry.[2] For physical reasons, I shall suppose that the condition $\mu + p \not\equiv 0$ is satisfied. Later, it will be assumed that the fluid obeys a γ-law equation of state, i.e., that $p = (\gamma - 1)\mu$, where $\mu > 0$ and $1 < \gamma \leqslant 2$, the case $\gamma = 1$ being rejected by the assumption of nonconstant p.

In summary, the space-times to be considered are those which are static and possess either spherical, plane, or hyperbolic symmetry, and in which the source of the gravitational

 0022-2488/85/092268-07$02.50

field is a perfect fluid whose energy density μ and pressure p are not identically constant. It is initially assumed that the condition $\mu + p \not\equiv 0$ holds; the additional assumption that $p = (\gamma - 1)\mu$ with $\mu > 0$ and $1 < \gamma \leqslant 2$ will later be made.

The plan of this article is as follows. In Sec. II, the general form of the space-time metric is given, and the field equations are provided. Some implications are then deduced, specifically with regard to the invariance properties of these equations, and to the homologous transformations that lead to a preferred choice of variables. The field equations are expressed in terms of these quantities as an autonomous system of ordinary differential equations. In Sec. III, plane autonomous systems are examined qualitatively in the cases of spherical, plane, and hyperbolic symmetry, when the fluid obeys a γ-law equation of state. This leads to a prescription for finding the general exact solution in the plane symmetric case, and thereby points to defects in an earlier prescription. Conformal diagrams are given for the various types of solutions that can arise in all three cases.

Throughout, geometric units are chosen, in which $8\pi G = c = 1$, where G is the Newtonian gravitational constant and c is the speed of light in a vacuum. The space-time metric g_{ij} has signature $(-\ +\ +\ +)$; the Riemann tensor $R^i{}_{jkl}$ is defined by the convention $u^i{}_{;l;k} - u^i{}_{;k;l} = R^i{}_{jkl} u^j$, where a semicolon denotes covariant differentiation; the Ricci tensor R_{ij} is defined by $R_{ij} = R^k{}_{ikj}$; and $R = R^i{}_i$ is the Ricci scalar.

II. THE FIELD EQUATIONS

In this section, a study will be made of Einstein's field equations of general relativity, viz.,

$$R_{ij} - \tfrac{1}{2} R g_{ij} = T_{ij}, \tag{2.1}$$

in the case when the energy-momentum tensor T_{ij} is that of a perfect fluid, and when the space-time is static and possesses either spherical, plane, or hyperbolic symmetry. The requirement of a perfect fluid means that

$$T_{ij} = (\mu + p) u_i u_j + p g_{ij}, \tag{2.2}$$

where μ is the fluid's energy density, p is the (isotropic) pressure, and $\mathbf{u}$ is the (timelike, future-pointing) unit vector that is tangent to the fluid flow-lines. With the assumed symmetries, the space-time metric is of the form

$$ds^2 = -g^2(x) dt^2 + dx^2 + f^2(x)(dy^2 + h^2(y) dz^2), \tag{2.3}$$

where the function h satisfies the differential equation

$$\frac{d^2 h}{dy^2} + K h = 0,$$

K is a constant with $K = +1$, 0, or -1 according as the geometry possesses spherical, plane, or hyperbolic symmetry, and where, without loss of generality, the functions f, g, and h are positive (see, e.g., Ref. 2). With no further loss of generality, the function $h = h(y)$ will be chosen as $\sin y$ ($K = +1$), y ($K = 0$) or $\sinh y$ ($K = -1$). The field equations (2.1) for the perfect fluid energy-momentum tensor (2.2) and metric (2.3) are

$$f''/f + g''/g + (f'/f)(g'/g) = p, \tag{2.4a}$$

$$2 f''/f + f'^2/f^2 - K/f^2 = -\mu, \tag{2.4b}$$

and

$$2(f'/f)(g'/g) + f'^2/f^2 - K/f^2 = p, \tag{2.4c}$$

where a prime (′) denotes differentiation with respect to x. Equations (2.4) are compatible whenever the contracted Bianchi identities are satisfied, i.e., when

$$(\mu + p)(g'/g) + p' = 0. \tag{2.5}$$

These equations provide for a uniform discussion of the class of solutions that is being studied. They have been given previously by various authors, including Hojman and Santamarina,[12] whose article will be discussed later. The equations are simple generalizations of those given for the spherically symmetric ($K = +1$) case in Ref. 13. From the difference of (2.4b) and (2.4c), it follows that $f' \not\equiv 0$, since otherwise $\mu + p \equiv 0$. It is immediately clear by inspection that Eqs. (2.4) are invariant under the two-parameter group of transformations

$$g \to \nu g, \tag{2.6a}$$

$$x \to \kappa^{-1} x, \quad f \to \kappa^{-1} f, \quad \mu \to \kappa^2 \mu, \quad p \to \kappa^2 p, \tag{2.6b}$$

where ν and κ are nonzero numbers; this generalizes the observation made by Ellis, Maartens, and Nel[14] in the spherically symmetric ($K = +1$) case (cf. Ref. 13). It is also clear that Eqs. (2.4) are invariant under the complex transformation $f \to if$, $K \to -K$, with g, x, μ, and p fixed. Hence the case of hyperbolic symmetry, where $K = -1$ (and f is real), may alternatively be treated formally as the spherically symmetric case, where $K = +1$, with f being purely imaginary. Following various authors (cf. Refs. 2, 13, and 15), Eq. (2.4b) may be written in the form

$$f'^2 = K - 2m(f)/f, \tag{2.7}$$

where

$$\frac{dm}{df} = \frac{1}{2} \mu f^2, \tag{2.8}$$

and the fact that $f' \not\equiv 0$ has been used. Equation (2.7) implies that

$$m(f)/f \leqslant \tfrac{1}{2} K. \tag{2.9}$$

In the spherically symmetric ($K = +1$) case, the quantity m may be physically interpreted in terms of the total mass within a given radius, assuming that there is a regular center of symmetry. Equations (2.4c) and (2.5) imply that

$$\frac{dp}{df} = -\frac{(\mu + p)[2m + p f^3]}{2 f (K f - 2m)} \tag{2.10}$$

(cf. the case when $K = +1$, in Ref. 13). The invariance of the field equations under the transformation (2.6b) suggests the introduction of the new variables $M := m/f$, $D := \tfrac{1}{2} \mu f^2$, and $P := \tfrac{1}{2} p f^2$, since these are each also invariant under (2.6b). Then

$$\frac{dM}{df} = \frac{1}{f}(D - M),$$

where use has been made of Eq. (2.8). Moreover, from (2.10),

$$\frac{dD}{df} = \frac{1}{f} \frac{1}{K - 2M} \left[2D(K - 2M) - \frac{(D + P)(P + M)}{dp/d\mu} \right].$$

If the fluid satisfies a γ-law equation of state, $p = (\gamma - 1)\mu$, with $1 < \gamma \leqslant 2$ and $\mu > 0$, the substitution $\tau = \ln f$ reduces these equations to the *autonomous* system

$$\frac{dD}{d\tau}=\frac{D}{K-2M}\left[2K-\frac{(5\gamma-4)}{\gamma-1}M-\gamma D\right] \tag{2.11a}$$

and

$$\frac{dM}{d\tau}=D-M, \tag{2.11b}$$

in agreement with Eqs. (3.4) of Ref. 12. When $K=+1$, this system is equivalent to (2.2) of Ref. 13. Note that the denominator $K-2M$ in Eq. (2.11a) cannot vanish identically, for, if it did, Eq. (2.7) would imply that $f'\equiv 0$.

While the system (2.11) is the one that will be employed in most of the remainder of this article, it might be thought that its derivation has been rather *ad hoc*, owing to a reliance on the observation of invariance of the field equations under the transformation (2.6b). An alternative and somewhat more systematic approach is to consider the invariance of the field equations (2.4) under "quasihomologous" transformations, along the lines of the study in Ref. 7 for the spherically symmetric case. Such an invariance arises from the fact that for Newtonian stars in equilibrium, a simple homologous family of solutions exists, the individual members being related to each other by rescalings of the radius, the density, and the total mass.[7,16] Because the radial coordinate is less well defined in general relativity, it is not necessarily justified to assert that homologous families may be specified by a similar rescaling. A generalization, which takes the ambiguity in the radial coordinate fully into account, has been considered previously.[7] In this, it is postulated that the individual members of the family should be related by a transformation that maps each physically relevant variable into a function of itself. It is then *proved* that the only allowable transformation involves a simple rescaling. This is done by making use of the result[17] that a system of ordinary differential equations

$$\frac{du^i}{dx}=f^i(x,\mathbf{u}),$$

where $1\leqslant i\leqslant n$ and $\mathbf{u}=(u^1,u^2,...u^n)$, is invariant under the action of the infinitesimal generator $\mathbf{X}:=\xi(x,\mathbf{u})\times(\partial/\partial x)+\eta^i(x,\mathbf{u})(\partial/\partial u^i)$ if and only if $[\mathbf{L},\mathbf{X}]=r\mathbf{L}$, where $[\,,\,]$ denotes the Lie bracket, $\mathbf{L}:=\partial/\partial x+f^i(\partial/\partial u^i)$, and $r=\mathbf{L}(\xi)$, which is obtained by action of the commutator on the independent variable x. In the particular case where $\mathbf{X}$ generates quasihomologous transformations, in which each variable is transformed into a function of itself, $\xi=\xi(x)$ and $\eta^i=\eta^i(u^i)$, with no sum over i, and the invariance requirement becomes

$$f^i\left[\frac{d\eta^i}{du^i}(u^i)-\frac{d\xi}{dx}(x)\right]=\mathbf{X}(f^i), \tag{2.12}$$

where $1\leqslant i\leqslant n$, and there is no sum over i. For any given system, Eqs. (2.12) can now be examined for compatibility in their functional dependence, and it was just such an examination[7] that led to the rescaling property for homologous families in the spherically symmetric $(K=+1)$ case.

In the present situation, it is possible to proceed in a similar fashion, starting with Eqs. (2.4). However, it is slightly more convenient to use Eqs. (2.8) and (2.10). The result of applying (2.12) leads to the following conclusions: *either* $K=\pm 1$ and $\mathbf{X}=0$ *or* $K=0$ and $\mathbf{X}=f(\partial/\partial f)+3m(\partial/\partial m)$; in any case, there is the extra symmetry $\mathbf{X}=f(\partial/\partial f)-2\mu\times(\partial/\partial\mu)+m(\partial/\partial m)$, if and only if $p=(\gamma-1)\mu$, for some constant γ.

Given an allowed symmetry vector $\mathbf{X}$, it is a straightforward matter to calculate a set of independent invariants, i.e., of functions of the given variables that are invariant under the finite transformation which $\mathbf{X}$ generates. This is done by solving the first-order differential equation $\mathbf{X}(F)=0$ for invariants F. Thus in the present case, the variables are f, μ, and m, and, if $K=\pm 1$, in general these form a set of independent invariants, whereas if $K=0$, there are in general only two independent invariants, e.g., m/f^3 and μ. If the equation of state is of the form $p=(\gamma-1)\mu$, for some constant γ, the number of independent invariants is reduced by one, and these invariants may be written as m/f and $\frac{1}{2}\mu f^2$ if $K=\pm 1$, and as $\frac{1}{2}(\mu/m)f^3$ if $K=0$. Division by f is allowable, since $f>0$ in the metric (2.3). The factor $\frac{1}{2}$ has been inserted for ease of comparison with other works.[2,13,15] The above provides a generalization of the result given previously[7] that pertains to the particular case when $K=\pm 1$. It also emphasizes the fact that while (2.6b) represents a transformation under which the general system of equations is always invariant, it is necessary to bear in mind that the rather innocuous assumption of an equation of state forces p to be an extraneously posited function of μ, which in general will be incompatible with the transformation (2.6b). The above results confirm that (2.6b) is indeed a valid transformation, but that it is only meaningful if there is a γ-law equation of state.

It is interesting to observe that in the *general* plane-symmetric $(K=0)$ case, i.e., without there necessarily being a γ-law equation of state, use of the invariants m/f^3 and μ yields a two-dimensional system, viz.,

$$\frac{d\mu}{d\tau}=\frac{(\mu+p)(2m/f^3+p)}{(4m/f^3)(dp/d\mu)}$$

and

$$\frac{d}{d\tau}\left(\frac{m}{f^3}\right)=\frac{1}{2}\mu-\frac{3m}{f^3},$$

but without an explicit equation of state little progress can be made in providing a qualitative description. However, for this situation, a prescription for an exact solution has already been given,[12] and some further comments on this will be made in Sec. III, which in the main deals with the particular cases that now ensue, viz., $K=\pm 1$ and $K=0$, there being a γ-law equation of state in each case. The invariants $M:=m/f$ and $D:=\frac{1}{2}\mu f^2$ will be used as basic variables. When $K=0$, use of the invariant $\tilde{D}:=\frac{1}{2}(\mu/m)f^3$ would reduce (2.11a) and (2.11b) to the single equation

$$\frac{d\tilde{D}}{d\tau}=\frac{\tilde{D}}{2}\left[\frac{7\gamma-6}{\gamma-1}-(2-\gamma)\tilde{D}\right]. \tag{2.13}$$

III. QUALITATIVE ANALYSIS

In this section, the system of equations (2.11) will be analyzed qualitatively for the cases $K=\pm 1$ and $K=0$, assuming an equation of state $p=(\gamma-1)\mu$, where γ is a constant satisfying $1<\gamma\leqslant 2$. The complex transformation that was observed in Sec. II could now be employed to give a simultaneous treatment of the two cases where $K\neq 0$. The complex transformation $f\to if$, $K\to -K$, with g, x, μ, and

p fixed, transforms $M:=m/f$ into $-M$, by (2.7), and $D:=\frac{1}{2}\mu f^2$ into $-D$. Since $\tau:=\ln f\rightarrow\tau+i(\pi/2)$, this means that Eqs. (2.11) are invariant under the transformation $K\rightarrow -K$, $D\rightarrow -D$, $M\rightarrow -M$, as may be readily checked, and that the case when $K=-1$ (with $D>0$) could be alternatively investigated by considering the case when $K=+1$ (with $D<0$). By (2.9), it follows that $M\leqslant -\frac{1}{2}$ if $K=-1$, so that in the transformed treatment, $M\geq +\frac{1}{2}$. Thus attention would be confined to the qualitative examination of the system (2.11) with $K=+1$, and with either $D>0$ and $M\leqslant\frac{1}{2}$ (corresponding to the case of spherical symmetry, where K really is equal to $+1$) or with $D<0$ and $M\geqslant\frac{1}{2}$ (corresponding to the case of hyperbolic symmetry, where K is really equal to -1). The former case was treated in Ref. 7, but with slightly different coordinates and notation, and so incorporation of the $K=-1$ case involves an extension to the region in which $D<0$ and $M\geqslant\frac{1}{2}$. For clarity and ease of discussion, the phase-plane diagrams for the cases $K=\pm 1$ are drawn separately, as Figs. 1(a) and 2(a), the symmetry under discussion manifesting itself in the fact that each figure is equivalent to the other under a reflection in the origin. It may also be noticed that the same symmetry applied to the $K=0$ case leads to the result that the corresponding diagram for the plane-symmetric case, Fig. 3(a), is symmetric under a reflection in the origin. Conformal Penrose diagrams[15,18] for the various types of solutions that can occur in each of the three cases are also provided, in Figs. 1(b), 2(b) and 3(b). In each case, these depict the character of the two-spaces $\{y=\text{const}, z=\text{const}\}$. The several possibilities will now be considered. In the following, it is interesting to note that there is one particular limit, in which D tends to zero and M becomes indefinitely large and negative, that is a common feature of all cases. From Eqs. (2.11), it follows that, in such a situation, the broad asymptotic behavior is independent of the sign of K, and that $|M|$ and D, respectively, grow and decay exponentially in τ. It follows from (2.7) that this occurs within a finite value of x, i.e., within a finite proper distance, and from (2.5) and the definition of D that the fluid's energy density μ tends to zero, while the fluid's acceleration $|g'/g|$ becomes infinite, as would be expected from a consideration of the infinite growth of $\ln\mu$ within a finite distance x. Under these circumstances, there is a scalar-curvature singularity[19] that is *not* a matter singularity.

A. $K=+1$

In the spherically symmetric case, there are three qualitative distinct types of solution. There is one degenerate case, represented by the center of the spiral in Fig. 1(a). In this case, $D=M=2(\gamma-1)/[(\gamma+2)^2-8]$, and this represents the Misner–Zapolsky solution[7–9] that was mentioned earlier. By converting back to the former variables, f, g, x, and μ, it is discovered that the central energy density (i.e., at $x=0$) in this solution is infinite. The second type of solution is depicted in Fig. 1(a) by the curve that starts at the origin (where f is zero) and which spirals into the central point. Such a solution has the property[7,13] that conditions are regular at the physical center at $x=0$, and the space-time extends out to infinite proper distances, becoming asymptotically like the Misner–Zapolsky solution. The third type of spherically

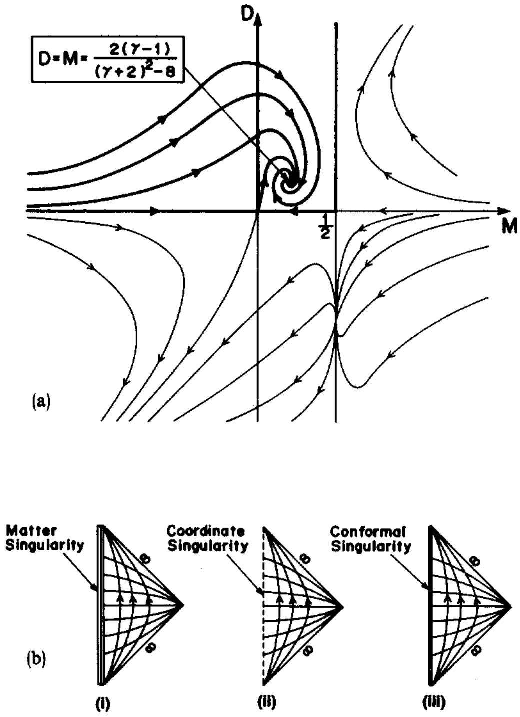


FIG. 1. Spherical symmetry ($K=+1$). Figure 1(a) is the phase-plane diagram for a static perfect fluid obeying a γ-law equation of state. Only the top left of the figure ($M\leqslant\frac{1}{2}$, $D>0$) represents physically realistic situations. Arrows are drawn in the direction of increasing values of f in the metric (2.3). The three distinct types of behavior [(i) the Misner–Zapolsky focal solution, (ii) the solution starting at the origin, and (iii) the remainder] are demarcated. For a fixed value of γ, each curve depicts the "evolution" (in terms of a radial variable) of a single space-time. Conformal Penrose diagrams of the maximal analytic extensions for the three possibilities (i), (ii), and (iii) are drawn in Figs. 1(b). Typically, each point in these diagrams represents a two-sphere. The family of curves with arrows indicates the fluid flow-lines, while the other family of curves represents the hypersurfaces $\{t=\text{const}\}$ that are orthogonal to the flow. Note the symmetry between Figs. 1(a) and 2(a).

symmetric solution is the most typical; the solution curves have the property that they cross the D axis at a finite value of τ, with the interpretation that $M=0$ at some finite nonzero radius. As observed in Ref. 7, this is impossible if there is to be a regular center of symmetry. Without this restriction, the central energy density (at $x=0$) is zero and the fluid's acceleration $|g'/g|$ becomes infinite, resulting in a scalar curvature singularity that is not a matter singularity. At large radial distances, the space-time behaves like the Misner–Zapolsky solution. The three types of behavior that result in the spherically symmetric case are depicted in conformal diagrams in Fig. 1(b). These show the nature of the two-surfaces $\{y=\text{const}, z=\text{const}\}$. It is interesting to observe that in the $\gamma=2$ Misner–Zapolsky solution, these two-surfaces are *flat*. In fact, regardless of the value of K, this is the only case when this is so.

B. $K=-1$

In the case of hyperbolic symmetry, there is essentially only one kind of behavior, since in Fig. 2(a) the solution curves of interest all begin at a large negative M with D close to zero, and end at the point where $M=-\frac{1}{2}$ and $D=-1/[2(\gamma-1)]$. At first sight, the situation is therefore

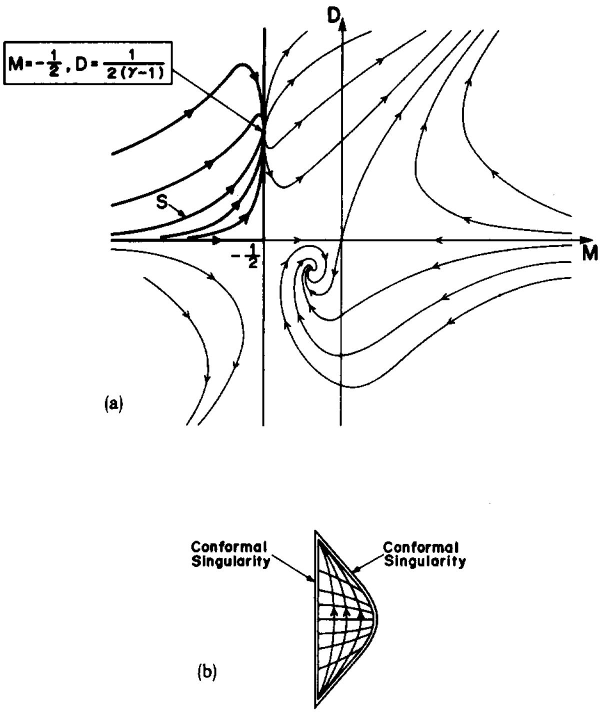


FIG. 2. Hyperbolic symmetry ($K = -1$). Figure 2(a) is the phase-plane diagram for a static perfect fluid obeying a γ-law equation of state. Only the top left of the figure ($M < -\frac{1}{2}$, $D > 0$) represents physically realistic situations. Arrows are drawn in the direction of increasing values of f in the metric (2.3). This may or may not be in the direction of increasing x, giving rise to the possibility of analytic extension through the point where $M = -\frac{1}{2}$ and $D = -1/[2(\gamma - 1)]$, against the direction of the arrows (for further details, see text). Apart from this, there is essentially only one type of behavior. For a fixed value of γ, each curve depicts the "evolution" (in terms of a spatial direction) of a single space-time. A conformal Penrose diagram of the maximal analytic extension for such a space-time is drawn in Fig. 2(b). Typically, each point in this diagram represents a two-surface that is diffeomorphic to $\mathbf{R}^2$. The family of curves with arrows indicates the fluid flow-lines, while the other family of curves represents the hypersurfaces $\{t = \text{const}\}$ that are orthogonal to the flow. Note the symmetry between Fig. 1(a) and 2(a).

considerably simpler. However, a detailed examination reveals that this latter limit does not necessarily signify an end to the space-time, since it corresponds to a finite limiting value of x (i.e., a finite proper distance) at which the metric coefficients f and g, and their first and second derivatives with respect to x, all approach finite limits [and hence, by the field equations (2.4), so also do the fluid variables μ and p]. The difficulty arises in the present formalism because the denominator in Eq. (2.11a) approaches zero (and so also does the numerator). It can be overcome by temporarily reverting to the formulation of the field equations that is provided by (2.4). In the limit in question, the derivative f' in (2.7) tends to zero, and so, by (2.4b), f'' tends to a negative value. Thus the limit may be equally well regarded either as one prior to which x and f are increasing, or as one prior to which x is decreasing and f is increasing, and in each case an analytic continuation through the limit would require f' to change sign. Thus it is possible to use the field equations (2.4) in order to obtain an analytic continuation through the limit, but since the construction of Fig. 2(a) was predicated upon the assumption of increasing τ, such an analytic continuation must be accompanied by a change in the direction of the arrows on passage through the limit point in the figure, where $M = -\frac{1}{2}$ and $D = -1/[2(\gamma - 1)]$. A detailed examination of the behavior of the solution curves near this point reveals that typically their slope becomes infinite, but that the derivative dD/dx tends to a nonzero finite limit, so that, in the analytic extension, D continues either to increase or to decrease. An exceptional situation arises for the one curve S that approaches the limiting point in Fig. 2(a) with a finite slope, and in this case dD/dx tends to zero. As may be expected, a further study shows that in the analytic extension, D now changes from an increasing to a decreasing function of x, or *vice versa*. Thus the special curve S in Fig. 2(a) that exhibits this property divides the remaining curves in such a way that analytic continuation of the space-time associated with a typical curve is represented by a smooth passage along the curve from above S to below S, or *vice versa*, while analytic extension of the space-time associated with the curve S itself is represented by a reflection at the point where $M = -\frac{1}{2}$ and $D = -1/[2(\gamma - 1)]$. The asymptotic behavior of the analytically extended solution is at both extremes depicted in Fig. 2(a) by M becoming indefinitely large and negative, and D tending to zero; the energy density approaches zero, and the fluid's acceleration $|g'/g|$ becomes infinite, within a proper distance x. Thus the analytically extended solution is confined by two scalar curvature singularities (that are not matter singularities), which are encountered in each hypersurface $\{t = \text{const}\}$ orthogonal to the fluid flow. This situation is depicted in a conformal diagram in Fig. 2(b), which exhibits the structure of any two-surface $\{y = \text{const}, z = \text{const}\}$.

C. $K = 0$

In the case of plane symmetry, the field equations are somewhat simpler. As observed in the previous section, it is possible to reduce them further to the single first-order equation (2.13). However, in order to afford a comparison of the solutions in this case with those when $K = \pm 1$, it will be more convenient to retain the system (2.11). The results of the qualitative analysis are depicted in Fig. 3(a). As in the case when $K = -1$, there is only one possible pattern of behavior. The curves start at large negative M and with D small, and run into the origin, in a manner that requires D to dominate M. A study of the details of this latter limit, using (2.7), reveals that for each curve it takes place as x becomes infinite, i.e., that it represents points at infinity in the associated space-time. The asymptotic behavior in the opposite direction along the curves is similar to that obtained in the cases $K = \pm 1$, and corresponds to a scalar curvature singularity (within a finite proper distance x) at which the fluid's energy density is zero, but the fluid's acceleration $|g'/g|$ is infinite. This situation is depicted in a conformal diagram in Fig. 3(b), which shows the structure of the two-surfaces $\{y = \text{const}, z = \text{const}\}$.

Some final remarks will now be made with regard to the case of plane symmetry ($K = 0$), and with particular reference to the prescription given by Hojman and Santamarina[12] for what is claimed to be the general solution with arbitrary cosmological constant Λ. It is claimed that, without loss of generality, Λ is zero. In a sense, this is always true for solu-

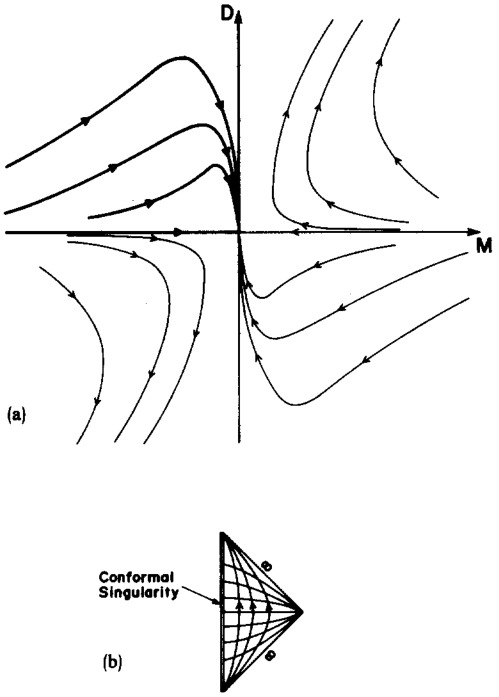


FIG. 3. Plane symmetry ($K = 0$). Figure 3(a) is the phase-plane diagram for a static perfect fluid obeying a γ-law equation of state. Only the top left of the figure ($M<0$, $D>0$) represents physically realistic situations. Arrows are drawn in the direction of increasing values of f in the metric (2.3). There is only one type of behavior. For a fixed value of γ, each curve depicts the "evolution" (in terms of a spatial direction) of a one-parameter family of space-times. A conformal Penrose diagram of the maximal analytic extension for such a space-time is drawn in Fig. 3(b). Typically, each point in this diagram represents a two-plane. The family of curves with arrows indicates the fluid flow-lines, while the other family of curves represents the hypersurfaces $\{t = \text{const}\}$ that are orthogonal to the flow. Note the symmetry of Fig. 3(a) under a reflection in the origin.

tions of Einstein's equations, since the physical effects of Λ can be incorporated in terms of a perfect fluid with $\mu + p = 0$ (see, e.g., Ref. 20). However, it must be remembered that not only does this involve an adjustment of the expressions for the energy density and pressure of a perfect fluid (as is done in Ref. 12), but also that this is such that it will not in general preserve any preassigned equation of state (a fact which is ignored in Ref. 12). Thus the system (3.4) of Ref. 12, which pertains to the case of arbitrary K and Λ, and a γ-law equation of state, is *not* in general equivalent to the subsystem obtained by putting $\Lambda = 0$. There is never a situation when, under the adjustment of μ and p, a γ-law equation of state is mapped into a γ-law, and, in fact, the only situation when an equation of state is preserved is when it is of the form $p = -\mu + \text{const}$. Henceforth, I shall therefore assume that $\Lambda = 0$ in referral to Ref. 12. The second point is that while, in the case of plane symmetry ($K = 0$), the prescription for the general solution is correct for $p \neq 0$, so that the function $G := m/p$ is defined, it is one which is analogous to some of those prescriptions that are employed in the spherically symmetric case and that were mentioned in Sec. I, i.e., as alluded to in Ref. 12, it is more adapted to a process that yields the equation of state at a final step, rather than being amenable to the imposition of an equation of state from the start. This means that is may be safely concluded that if the function G is assumed to be proportional to f^3, then the equation of state is a γ-law one, as observed in Ref. 12. However, it is also important to check that such an equation is of physical interest, i.e., that $\mu > 0$ and that $1 \leq \gamma \leq 2$, and it is easily shown that this is not the case. The most general function G that produces a γ-law equation of state $(1 < \gamma \leqslant 2)$ may be easily found from Eqs. (4.5) and (4.7) of Ref. 12, which in the present notation are

$$\frac{1}{p}\frac{dp}{df} = \frac{(f^2 + 2(dG/df))(f^3 + 2G)}{2G(f^3 - 2G)} \tag{3.1}$$

and

$$m = \frac{2}{f^2}\left(p\frac{dG}{df} + G\frac{dp}{df}\right), \tag{3.2}$$

respectively. Additional factors of 2 have been inserted, since the function m as defined in Ref. 12 differs by such a factor from the function m of the present article; it is to be noted that the denominator in (3.1) cannot vanish identically. Upon substitution of the relation $p = (\gamma - 1)\mu$, with $1 < \gamma \leqslant 2$, Eqs. (3.1) and (3.2) constrain G to satisfy the restriction

$$\frac{dG}{df} + \frac{\gamma}{2(\gamma - 1)f}G = \frac{(2-\gamma)f^2}{4(\gamma - 1)},$$

whose general solution is

$$G = \frac{(2-\gamma)}{7\gamma - 6}f^3 + Af^{-\gamma/2(\gamma - 1)}, \tag{3.3}$$

where A is an arbitrary constant. In this context, Hojman and Santamarina[12] considered the special case $A = 0$, when $G = \alpha f^3$ for some constant $\alpha = (2-\gamma)/(7\gamma - 6)$, where α is distinct from 0 and 1, because $G \not\equiv 0$ and $\gamma \neq 1$. However, Eq. (2.7) and the definition of G show that $pG/f \leqslant 0$, and with $G = \alpha f^3$ this means that $\alpha p \leqslant 0$. Now if $1 < \gamma \leqslant 2$, then $0 \leqslant \alpha < 1$, and since the case $\alpha = 0$ is inadmissible, the requirement that αp be nonpositive now forces p to be strictly negative, and hence so also is μ.

Finally, it may be noted that if there is a γ-law of state, the general form (3.3) of G allows one to determine, by quadratures, the functional forms for μ and p, using Eqs. (3.1) and (3.2). From these relationships, the metric functions f and g are determined by quadratures from Eqs. (2.5) and (2.7), the expression for F being in general implicit. An alternative procedure is readily obtained from Eq. (2.13), which was derived by exploiting the extra symmetry of the system (2.11) that exists when $K = 0$. It should first be recalled that since $\widetilde{D} = \frac{1}{2}(\mu/m)f^3$, Eq. (2.7) implies that $\widetilde{D} < 0$, and so the obvious solutions of (2.13) in which $\widetilde{D}$ is constant are of no further interest. In order to obtain the remaining solutions, the variables $\widetilde{D}$ and τ can be separated in (2.13), and the resulting expression integrated to give

$$\widetilde{D} = Bae^{a\tau}/(1 + Bbe^{a\tau}), \tag{3.4}$$

where $a := (7\gamma - 6)/[2(\gamma - 1)]$, $b := \frac{1}{2}(2-\gamma)$, and B is a negative constant. Equation (2.11b) can now be written in the form

$$\frac{dM}{d\tau} = M(\widetilde{D} - 1),$$

and integrated with the aid of (3.4), with the result that

$$M=\begin{cases} M_0 e^{-\tau}(1+Bbe^{a\tau})^{1/b} & (b\neq 0)\,, \\ M_0 \exp(-\tau+Be^{4\tau}) & (b=0)\,, \end{cases} \tag{3.5}$$

where M_0 is a negative constant. Since $f=e^{\tau}$, M is determined explicitly as a function of f, and, by (3.4), so also is $D=\widetilde{D}M$. Now $D=\frac{1}{2}\mu f^2$, and hence μ is given explicitly in terms of f. Inserting $p=(\gamma-1)\mu$ in Eq. (2.5) results in the relationship

$$g=g_0\mu^{-(\gamma-1)/\gamma}\,, \tag{3.6}$$

where g_0 is a nonzero constant. Having determined μ as a function of f, Eq. (3.6) yields g as a function of f. Thus all physical and geometrical variables are obtained explicitly as simple functions of f, which, by means of (2.7), may be used in place of x in the metric (2.3). In terms of coordinates (t,f,y,z) the general exact solution is then provided by the metric

$$ds^2=-g^2\,dt^2+df^2/(-2M)+f^2(dy^2+dz^2)\,.$$

Here, M is given by Eq. (3.5) in terms of $f=e^{\tau}$, where it should be recalled that $a=(7\gamma-6)/[2(\gamma-1)]$, $b=\frac{1}{2}(2-\gamma)$, and g is determined in the manner explained above, i.e.,

$$g=g_0[2BMaf^{a-2}/(1+Bbf^a)]^{-(\gamma-1)/\gamma}\,.$$

By an appropriate rescaling of the coordinates, it may be arranged that $M_0=-1$ and $g_0=1$, so that the solution depends on two essential parameters, γ and B. In the particular case when $\gamma=2$, the solution so obtained is readily seen to be equivalent to that of Tabensky and Taub,[21] by means of the transition $B\to -a^{8/3}/32, t\to aT/2^{1/2}, f\to 2Z^{1/2}/a^{1/6}$, $y\to(a^{2/3}/2)x$, and $z\to(a^{2/3}/2)y$, where the new coordinates (T, Z, x, y) and the new parameter a are those of Ref. 21. Similarly, in the particular case when $\gamma=\frac{4}{3}$, the above solution is readily seen to be equivalent to that obtained by Teixeira, Wolk, and Som,[22] either by means of the transition $B\to -\frac{486}{5}(\frac{4}{5})^{2/3}p_0{}^{5/3}$, $t\to t$, $f\to\frac{1}{3}(5/[4p_0])^{1/3}z$, $y\to 3(4p_0/5)^{1/3}x$, and $z\to 3(4p_0/5)^{1/3}y$, where the new coordinates (t, z, x, y) and the new parameter, p_0, are those of Ref. 2, or by means of the transition $B\to -\frac{486}{3125}\cdot 4^{2/3}p_0{}^{5/3}$, $t\to(3p_0)^{1/4}x^0$, $f\to\frac{5}{3}\cdot(4p_0)^{-1/3}f^{-1/2}=\frac{5}{3}\cdot 6^{1/5}/(4p_0)^{1/3}\times[1+5e^{6p_0^{1/2}x}]^{-1/5}$, $y\to\frac{3}{5}(4p_0)^{1/3}y$, and $z\to\frac{3}{5}(4p_0)^{1/3}z$, where the new function f, the new coordinates (x^0, x, y, z), and the new parameter p_0 are those of Ref. 22. In view of this equivalence, the additional conditions that were imposed in Ref. 22, and which relate to the values of the metric coefficients and their first derivatives on the plane $x=0$, are seen to be superfluous. It should also be noted that the general treatment in Ref. 2 that is attributed to Taub[23] amounts to no more than a formulation of the problem as the solution of two coupled nonlinear first-order ordinary differential equations, and so it cannot be regarded as a proper prescription for the general solution; this observation persists upon specialization to the case of a γ-law equation of state.

It is also possible to obtain a quadrature expression for f in terms of the variable x in the metric (2.3), by employing (2.7):

$$x=\pm\int\frac{1}{(-2M(f))^{1/2}}\,df+C, \tag{3.7}$$

where C is an arbitrary constant, which, without loss of generality, is zero. Thus all physical and geometrical variables of interest can also be expressed in the coordinates (t, x, y, z) of metric (2.3), at least up to a single quadrature given by (3.7), and subsequent inversion of the relationship so obtained.

ACKNOWLEDGMENTS

I wish to thank Helen Warren for her expert typing of the manuscript.

This work was partially supported by an operating grant from the Natural Sciences and Engineering Research Council of Canada under Grant No. A3978.

[1]K. Schwarzschild, Sitzungsber. Dtsch. Akad. Wiss. Berlin Kl. Math. Phys. Techn. 189 (1916).

[2]D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U. P., Cambridge, England, 1980).

[3]G. F. R. Ellis, "Relativistic cosmology," in *General Relativity and Cosmology*, Course XLVII, Varenna, Italy, 1969, edited by R. K. Sachs (Academic, New York, 1971).

[4]V. V. Nemytskii and V. V. Stepanov, *Qualitative Theory of Differential Equations* (Princeton U. P., Princeton, NJ, 1960).

[5]G. Sansone and R. Conti, *Non-Linear Differential Equations* (MacMillan, New York, 1964).

[6]S. Vojtášek and K. Janáč, *Solution of Non-Linear Systems* (Iliffe, London, 1969).

[7]C. B. Collins, J. Math. Phys. **18**, 1374 (1977).

[8]C. W. Misner and H. S. Zapolsky, Phys. Rev. Lett. **12**, 635 (1964).

[9]S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, New York, 1972).

[10]R. C. Tolman, Phys. Rev. **55**, 364 (1939).

[11]M. Wyman, Phys. Rev. **75**, 1930 (1949).

[12]R. Hojman and J. Santamarina, J. Math. Phys. **25**, 1973 (1984).

[13]C. B. Collins, J. Math. Phys. **24**, 215 (1983).

[14]G. F. R. Ellis, R. Maartens, and S. D. Nel, Mon. Not. R. Astron. Soc. **184**, 439 (1978).

[15]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1973).

[16]S. Chandrasekhar, *An Introduction to the Study of Stellar Structure* (Dover, New York, 1939).

[17]L. V. Ovsiannikov, *Group Analysis of Differential Equations* (Academic, New York, 1982).

[18]R. Penrose, "Conformal treatment of infinity," in *Relativity Groups and Topology*, Les Houches Summer School of Theoretical Physics, University of Grenoble, 1963, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1964).

[19]G. F. R. Ellis and B. G. Schmidt, Gen. Relat. Gravit. **8**, 915 (1977).

[20]C. B. Collins and J. Wainwright, Phys. Rev. D **27**, 1209 (1983).

[21]R. Tabensky and A. H. Taub, Commun. Math. Phys. **29**, 61 (1973).

[22]A. F. da F. Teixeira, I. Wolk, and M. M. Som, J. Phys. A **10**, 1679 (1977).

[23]A. H. Taub, "Plane-symmetric similarity solutions for self-gravitating fluids," in *General Relativity (Papers in Honour of J. L. Synge)*, edited by L. O'Raifeartaigh (Clarendon, Oxford, 1972).

Relativistic, stationary, axisymmetric perfect fluids. I. Reduction to a system of two equations

S. Bonanos and D. Sklavenites
N. R. C. "Demokritos," Aghia Paraskevi, Attiki, Greece

(Received 23 October 1984; accepted for publication 1 March 1985)

The introduction of a new system of coordinates permits a partial solution of Einstein's field equations for the interior of a uniformly rotating, stationary, axisymmetric perfect fluid, and reduces the problem to two second-order partial differential equations for two unknown functions.

I. INTRODUCTION

The study of the "interior" stationary axisymmetric problem in general relativity has been hampered by the fact that, contrary to what is true for the corresponding "exterior" problem, no significant reduction of Einstein's field equations to a simpler—but equivalent—system is known. Thus, while in the "exterior" problem there is a well-known choice of coordinates,[1] which reduces the problem to a system of two second-order partial differential equations for two unknown functions,[2] no such reduction for the "interior" problem has been achieved, except for the important cases of Einstein–Maxwell fields[3] and dust,[4] where the same coordinate choice can be made.

In this paper we present a different choice of coordinates which allows a similar reduction of the problem to two second-order partial differential equations for the case of a uniformly rotating perfect fluid (with nonconstant pressure).

II. THE CHOICE OF COORDINATES

The derivation of the form of the metric for stationary axially symmetric space times has been given repeatedly in the literature. We will therefore write down immediately the line element appropriate to our problem as follows[5]:

$$\begin{aligned} ds^2 &= g_{ab}\, dx^a\, dx^b \\ &= e^{2U}(dt + A\, d\varphi)^2 - e^{-2U} W^2\, d\varphi^2 \\ &\quad - e^{-2U}\gamma_{MN}\, dx^M\, dx^N, \end{aligned} \tag{1}$$

where $t\,(=x^0)$ and $\varphi(=x^1)$ are two coordinates adapted to the timelike and rotational Killing vectors implied by the assumed symmetries ($\xi = \partial_t, \eta = \partial_\varphi$), and the metric functions U, A, W, γ_{MN} depend only on the coordinates $x^M = (x^2, x^3)$ which label the points on the two-surfaces orthogonal to the Killing orbits. The existence of these two-surfaces is guaranteed by the assumption that the source is a perfect fluid rotating about the symmetry axis, i.e., its four-velocity $\mathbf{u}$ satisfies the condition[6] $u_{[a}\xi_b \eta_{c]} = 0$.

For uniformly rotating perfect fluids, i.e., when $u^\varphi = \Omega u^t$ with $\Omega = \text{const}$, the transformation $\varphi = \varphi' + \Omega t$, which preserves the form of (1), defines comoving coordinates where $u^{\varphi'} = 0$. Dropping the prime, we assume in the following that (1) is written in comoving coordinates. The equations of motion $T^{ab}{}_{;b} = 0$, where

$$T^{ab} = (\mu + p)u^a u^b - p g^{ab} \tag{2}$$

is the energy-momentum tensor of a perfect fluid, with μ the energy density and p the pressure, can be written in the form

$$dp + (\mu + p)dU = 0, \tag{3}$$

so that the surfaces of constant p(or μ) and U coincide.

The form of the metric (1) is unchanged under general coordinate transformations $x^M = x^M(x^{N'})$ in the two-space orthogonal to the Killing orbits. The usual choice of coordinates is to demand that $\gamma_{22} = \gamma_{33}$ and $\gamma_{23} = 0$ and allows the field equations to be written in a compact form. We shall make a different choice of coordinates in an attempt to reduce the mathematical complexity of the system of equations to be solved. That such a reduction might be possible is suggested by the following observation: the second derivatives of the metric functions enter four of the six nontrivial Einstein equations only through the second derivatives of W and a single combination of second derivatives of γ_{MN}—the scalar curvature of the two-space $\gamma_{MN}\, dx^M\, dx^N$ (see Ref. 7). This implies that choosing the function W as a coordinate makes three of these four equations first order in the unknowns. Similarly, choosing either U or A as a second coordinate[8] makes one of the remaining two field equations first order also. We shall choose U as the second coordinate so that our coordinate system remains valid even in the static case ($A = 0$).

The choice of W and U as coordinates can be made (locally) provided that

$$dW \wedge dU \neq 0. \tag{4}$$

Since W vanishes on the axis,[9] we expect this condition to hold at least near the axis for realistic configurations (stars) which contain part of the axis and for which the pressure [and by (3), U] varies between the center and the surface along the axis. If the configurations have an additional symmetry (equatorial) plane—as would be expected for isolated rotating objects—condition (4) would, by symmetry, be violated in that plane (the vectors ∇W and ∇U would have to be parallel there). We thus, *a priori*, expect our coordinate system to be good throughout the interior of realistic rotating stars except for points on a symmetry plane.

The space-times excluded by condition (4)—in addition to the case that W and U are functionally related which leads to unbounded configurations—are those of constant pressure (in particular, dust), where $U = \text{const}$. The case $W = \text{const}$ does not arise since W vanishes on the axis but cannot vanish throughout the space-time.

The choice of W and U as coordinates, in addition to making four of the six field equations first order in the un-

 0022-2488/85/092275-04$02.50

knowns, allows a simple description of the intrinsic geometry of the equipressure surfaces in the three-space orthogonal to the motion of the fluid. The metric on these two-surfaces is given by

$$d\sigma^2 = (g_{ab} - u_a u_b)dx^a\,dx^b|_{U=\text{const}} = -e^{-2U}\{\gamma_{WW}|_{U=\text{const}}\,dW^2 + W^2\,d\varphi^2\}\,. \tag{5}$$

Recalling that the metric of a surface of revolution in Euclidean three-space, described (in cylindrical coordinates ρ,z,φ) by the equation $z = z(\rho)$, is

$$d\sigma_E^2 = \left\{1 + \left(\frac{dz}{d\rho}\right)^2\right\}d\rho^2 + \rho^2\,d\varphi^2, \tag{6}$$

we can easily construct, for any given function $\gamma_{WW}(U,W)$ the surface $z(\rho)$ in Euclidean three-space, the intrinsic geometry of which is the same as that of fluid's equipressure surfaces.

III. THE FIELD EQUATIONS

Our choice of W and U as coordinates leaves A, γ_{WW}, γ_{UU}, and γ_{UW} as the four unknown functions to be determined by the field equations. It turns out that the field equations becomes simpler if we express γ_{MN} in terms of its inverse γ^{MN}. With the notation

$$\gamma^{WW}\equiv P,\quad \gamma^{UU}\equiv Q,\quad \gamma^{UW}\equiv R, \tag{7}$$

$$\det\gamma_{MN}\equiv\Delta^2 = 1/(PQ - R^2), \tag{8}$$

the line element (1) becomes

$$ds^2 = e^{2U}(dt + A\,d\varphi)^2 - e^{-2U}W^2\,d\varphi^2 - e^{-2U}\Delta^2\{P\,dU^2 + Q\,dW^2 - 2R\,dU\,dW\}\,. \tag{9}$$

We shall further rename our coordinates using lowercase letters ρ,h, as follows:

$$W = \rho,\quad U = -\ln h, \tag{10}$$

so that the line element finally becomes

$$ds^2 = (1/h^2)(dt + A\,d\varphi)^2 - h^2\rho^2\,d\varphi^2 - h^2\Delta^2\left\{P\left(\frac{dh}{h}\right)^2 + Q\,d\rho^2 + 2R\,\frac{dh}{h}\,d\rho\right\}. \tag{11}$$

Equation (3) now reads $dp = (\mu + p)(dh/h)$, and can be integrated once an equation of state has been given. To integrate Eq. (3) for an arbitrary equation of state we express both p and μ through an arbitrary function of h, $f(h)$, so that (3) is an identity. The following parametrization is convenient ($f' = df/dh$):

$$8\pi p = f/h^2,\quad 8\pi(\mu + p) = f'/h - 2f/h^2\,. \tag{12}$$

The field equations $E^a{}_b = G^a{}_b - 8\pi T^a{}_b = 0$, where $G^a{}_b = R^a{}_b - \frac{1}{2}R\delta^a_b$ is Einstein's tensor and $T^a{}_b$ the energy-momentum tensor (2), will now be written in the orthonormal frame defined by the one-forms[10]

$$\omega^0 = (1/h)(dt + A\,d\varphi),\quad \omega^2 = (1/\sqrt{Q})dh,$$
$$\omega^1 = h\rho\,d\varphi,\quad \omega^3 = \Delta\sqrt{Q}\,[h\,d\rho + (R/Q)dh\,]\,, \tag{13}$$

so that $ds^2 = (\omega^0)^2 - (\omega^1)^2 - (\omega^2)^2 - (\omega^3)^2$.

A straightforward calculation gives, for the nontrivial components of $E^a{}_b$,

$$-h^2(E^2_2 + E^3_3) = (1/\rho\Delta)\{(P\Delta)_\rho - h(R\Delta)_h\} - 2f = 0, \tag{14}$$

$$-h^2(E^2_2 - E^3_3) = \frac{1}{\rho\Delta Q}(1 - \Delta^2R^2)\left(\frac{1}{\Delta}\right)_\rho + \frac{\Delta Q}{\rho}h\left(\frac{R}{\Delta Q}\right)_h - 2Q + \frac{1}{2\rho^2h^4}\left\{Q(hA_h)^2 - 2RhA_hA_\rho + \left(R^2 - \frac{1}{\Delta^2}\right)\frac{A_\rho^2}{Q}\right\} = 0, \tag{15}$$

$$-h^2E^2_3 = \frac{R}{\rho Q}\left(\frac{1}{\Delta}\right)_\rho + \frac{1}{\rho\Delta}\frac{h(\Delta\sqrt{Q})_h}{\Delta\sqrt{Q}} + \frac{1}{2\rho^2h^4\Delta}A_\rho\left(hA_h - \frac{R}{Q}A_\rho\right) = 0, \tag{16}$$

$$\tfrac{1}{2}h^2(E^0_0 - E^1_1 - E^2_2 - E^3_3) = (1/\rho\Delta)\{(\rho R\Delta)_\rho - h(\rho Q\Delta)_h\} + 2\omega^2 - \tfrac{1}{2}hf' = 0, \tag{17}$$

$$2\Delta E^0_1 = \left\{\frac{\Delta}{\rho h^4}(PA_\rho - RhA_h)\right\}_\rho + h\left\{\frac{\Delta}{\rho h^4}(QhA_h - RA_\rho)\right\}_h = 0, \tag{18}$$

$$-h^2E^1_1 = -K + Q + \omega^2 - f = 0, \tag{19}$$

where for brevity we have written

$$K = -\frac{1}{\Delta}\left\{\left[P\Delta_\rho - Rh\Delta_h + \frac{1}{2}\Delta(P_\rho - hR_h)\right]_\rho - h\left[R\Delta_\rho - Qh\Delta_h + \frac{1}{2}\Delta(R_\rho - hQ_h)\right]_h + (\Delta R)_\rho\left(\frac{hP_h}{P} - \frac{hQ_h}{Q}\right) - h(\Delta R)_h\left(\frac{P_\rho}{P} - \frac{Q_\rho}{Q}\right)\right\} \tag{20}$$

for the scalar curvature of the two-space $\gamma_{MN}\,dx^M\,dx^N$, and

$$\omega^2 = (1/4\rho^2h^4)\{PA^2_\rho + Q(hA_h)^2 - 2RhA_hA_\rho\} = -h^2g_{ab}\omega^a\omega^b \tag{21}$$

for the magnitude (up to a factor h^2) of the vorticity vector of the fluid

$$\omega^a = (1/2\sqrt{-g})\epsilon^{abcd}u_bu_{c;d}\,. \tag{22}$$

As was anticipated in Sec. II, the first four field equations (14)–(17) involve only first derivatives of the metric functions A, P, Q, R. The second derivatives of A appear in Eq. (18) only, while the second derivatives of P, Q, R appear in Eq. (19) only. Using the Bianchi identities one can show that Eqs. (18) and (19) are satisfied identically when A, P, Q, R satisfy Eqs. (14)–(17).

IV. REDUCTION TO A SYSTEM OF TWO EQUATIONS

Eliminating $(1/\Delta)_\rho$ between Eqs. (14), (15), and (16) in two different ways and using the definition of Δ^2(Eq. 8), we obtain

$$(1/2\Delta Q)\{\Delta R\ [\text{Eq. (15)}] - (1-\Delta^2R^2)\ [\text{Eq. (16)}]\}$$
$$= \frac{hP_h}{4\rho} - P\frac{A_\rho A_h}{4\rho^2 h^3} - R\left\{1 - \left(\frac{A_h}{2\rho h}\right)^2\right\} = 0 \tag{23}$$

and

$$\tfrac{1}{2}[\text{Eq. (14)}] + \tfrac{1}{2}[\text{Eq. (15)}] + \Delta R\ [\text{Eq. (16)}]$$
$$= \frac{P_\rho}{2\rho} - f - P\left(\frac{A_\rho}{2\rho h^2}\right)^2 - Q\left\{1 - \left(\frac{A_h}{2\rho h}\right)^2\right\} = 0. \tag{24}$$

We observe that Eqs. (23) and (24) can be trivially satisfied by using them to express the unknowns R and Q in terms of P and A (and their 1st derivatives). If we now substitute for R and Q in Eqs. (14) and (17) the expressions[11]

$$R = [hP_h/4\rho - P(A_\rho A_h/4\rho^2 h^3)]/[1-(A_h/2\rho h)^2] \tag{23'}$$

and

$$Q = [P_\rho/2\rho - f - P(A_\rho/2\rho h^2)^2]/[1-(A_h/2\rho h)^2], \tag{24'}$$

we shall obtain two second-order partial differential equations for the remaining unknown functions P and A. Equivalently, we can substitute for R and Q in Eqs. (18) and (19) after using Eqs. (14) and (17) to eliminate the derivatives of Δ. A long calculation gives the equations

$$\begin{aligned} P\left(1-\frac{A_z^2}{\xi}\right)P_{\xi\xi} &+ \left(P_\xi - f - P\frac{A_\xi^2}{z^2}\right)\frac{z^2P_{zz}}{\xi} \\ &- \left(\frac{zP_z}{\xi} - 2P\frac{A_\xi A_z}{z\xi}\right)\left(zP_{z\xi} - zf_z - \frac{zP_z}{4\xi}\right) \\ &- \left(1-\frac{A_z^2}{\xi}\right)\Big\{(P_\xi - f)(P_\xi - 2f) \\ &- \frac{zP_z}{\xi}\left(P_\xi - f - \frac{1}{2}zf_z\right)\Big\} \\ &- \frac{1}{\xi}\{2A_z(P_\xi - f) - P_zA_\xi\} \\ &\times\left\{P_zA_\xi + \frac{1}{2}P\frac{A_z}{\xi} - \frac{zP_z}{\xi}A_z\right\} = 0, \end{aligned} \tag{25}$$

and

$$\begin{aligned} P\left(1-\frac{A_z^2}{\xi}\right)A_{\xi\xi} &+ \left(P_\xi - f - P\frac{A_\xi^2}{z^2}\right)\frac{z^2A_{zz}}{\xi} \\ &- \left(\frac{zP_z}{\xi} - 2P\frac{A_\xi A_z}{z\xi}\right)\left(zA_{z\xi} - \frac{zA_z}{2\xi}\right) \\ &- \left(1-\frac{A_z^2}{\xi}\right)\Big\{\frac{zA_z}{\xi}(P_\xi - f) - \frac{zP_z}{\xi}A_\xi \\ &+ \frac{z^2A_zf_z}{2\xi} - fA_\xi\Big\} = 0, \end{aligned} \tag{26}$$

where we have introduced the new independent variables

$$\xi \equiv \rho^2 \quad \text{and} \quad z \equiv h^2. \tag{27}$$

In these equations $f(z)$ is an arbitrary function specifying the equation of state. If we let $f=0$ we obtain the equations appropriate for a stationary axisymmetric vacuum space-time in our coordinates. In the static case ($A=0$), Eq. (26) is satisfied identically.

As might be expected, Eqs. (25) and (26) are derivable from a variational principle

$$\delta\int \mathscr{L}\,\frac{dz}{z}\,d\xi = 0.$$

It can be shown that an appropriate Lagrangian density is

$$\begin{aligned} \mathscr{L} &= (1 - A_z^2/\xi)/P\Delta \\ &= \left[\frac{P_\xi - f}{P}\left(1-\frac{A_z^2}{\xi}\right) - \frac{1}{\xi}\left(\frac{zP_z}{2P}\right)^2 \right. \\ &\left. - \left(\frac{A_\xi}{z}\right)^2 + \frac{P_zA_zA_\xi}{P\xi}\right]^{1/2}. \end{aligned} \tag{28}$$

V. BEHAVIOR NEAR THE AXIS

In the previous section it was shown that Einstein's equations for the interior of a uniformly rotating perfect fluid reduce in our formalism to Eqs. (25) and (26); and *any* functions $P(z,\xi)$, $A(z,\xi)$, and $f(z)$, satisfying these equations and the conditions $A \neq z\sqrt{\xi}$ and $PQ - R^2 \neq 0$, determine a space-time with $G^a_b = 8\pi T^a_b$. For the resulting metric to be nonsingular on the axis, however, we must impose on the functions P and A the boundary conditions

$$P \to 1 + p_1\xi \quad \text{and} \quad A \to a_1\xi \quad \text{as} \quad \xi \to 0, \tag{29}$$

where p_1 and a_1 are arbitrary functions of z. These conditions follow from the requirements of "elementary flatness"[12] and finite angular velocity of the locally nonrotating observers[13] on the axis. They also ensure that the magnitude of the fluid's acceleration and vorticity remain finite on the axis. Finally, the arbitrary functions p_1 and a_1 must satisfy $p_1 - f - (a_1/z)^2 > 0$, so that the metric has the required signature ($PQ\text{-}R^2 > 0$) on the axis, while $f(z)$ must lead to a realistic equation of state.

If we assume that P and A can be expanded in a power series in ξ near the axis, i.e.,

$$P = 1 + \sum_{n=1} p_n\xi^n, \quad A = \sum_{n=1} a_n\xi^n, \tag{30}$$

Then Eqs. (25) and (26) determine successively the coefficients p_n, a_n for $n \geqslant 2$ in terms of $p_{n-1},\ldots,p_1$ and $a_{n-1},\ldots,a_1$, and thus ultimately in terms of p_1 and a_1. The coefficients p_2 and a_2, for example, are given by (a prime denoting differentiation w.r.t. z)

$$\begin{aligned} 2p_2 = &-p_1''\{z^2(p_1 - f) - a_1^2\} + z^2p_1'(\tfrac{3}{4}p_1' - \tfrac{1}{2}f') \\ &- 2a_1a_1'(p_1' - f') + a_1'^2(p_1 - f) - zp_1'(p_1 - f) \\ &+ (p_1 - f)(p_1 - 2f), \end{aligned} \tag{31}$$

and

$$\begin{aligned} 2a_2 = &-a_1''\{z^2(p_1 - f) - a_1^2\} \\ &+ \tfrac{1}{2}za_1'(zp_1' - 2a_1a_1'/z) \\ &+ za_1'(p_1 - f) - zp_1'a_1 + (z^2/2)a_1'f' - fa_1. \end{aligned} \tag{32}$$

It can be verified that $P = 1 + p_1\xi$ and $A = 0$, p_1 satisfying Eq. (31) with $p_2 = a_1 = 0$, is an exact solution of Eq. (25), describing a static and spherically symmetric space-time for any equation of state $f(z)$.

Futher applications of the formalism presented in this paper are given in the following paper.

[1]A. Papapetrou, Ann. der Phys. **12**, 309 (1953).
[2]F. Ernst, Phys. Rev. **167**, 1175 (1968).
[3]F. Ernst, Phys. Rev. **168**, 1415 (1968).
[4]J. Winicour, J. Math. Phys. **16**, 1806 (1975).
[5]D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U. P., London, 1980), Eq. 17.15.
[6]B. Carter, "Black Hole Equilibrium States," in *Black Holes*, edited by C. DeWitt and B. S. DeWitt (Gordon and Breach, New York, 1973), p. 162.
[7]Ref. 5, Eqs. 17.28 and 17.29.
[8]The choice of U or A as one coordinate has been made before, although the condition $\gamma_{23} = 0$ has been maintained. See P. Carlson, Gen. Relativ. Gravit. **11**, 291 (1979); or E. Glass, J. Math. Phys. **18**, 708 (1977). We are not aware of a previous use of W as a coordinate when $T^2_2 + T^3_3 \neq 0$.
[9]Reference 6, p. 139.
[10]In this frame $u^a = \delta^a_0$ and $\dot{u}^a = u^b u^a_{;b} = (\sqrt{Q}/h)\delta^a_2$, so ω^0 and ω^2 are parallel to the fluid's velocity and acceleration, respectively. Then ω^1 is orthogonal to the velocity in the space spanned by the Killing vectors, while ω^3 is orthogonal to the acceleration in the orthogonal two-space.
[11]As shown in the next section, the denominator cannot vanish for a well-behaved manifold near the axis.
[12]The requirement that the ratio of circumference to radius of a small circle centered on the axis approaches 2π as the radius goes to zero.
[13]J. Bardeen, Astrophys. J. **162**, 71 (1970). These observers have angular velocity $\omega_{NR} = -g_{\varphi t}/g_{\varphi\varphi} = A/(\rho^2 h^4 - A^2)$. Equivalently one can demand finiteness of the vorticity ω^2 [(Eq. (21)] on the axis.

Relativistic, stationary, axisymmetric perfect fluids. II. Solutions with vanishing magnetic Weyl tensor

D. Sklavenites
N. R. C. "Demokritos," Aghia Paraskevi, Attiki, Greece

(Received 23 October 1984; accepted for publication 1 March 1985)

The formalism of the previous article is used to obtain solutions of Einstein's field equations for the interior of a rigidly rotating perfect fluid, with zero magnetic Weyl tensor. It is found that these solutions cannot represent isolated rotating masses.

I. INTRODUCTION

In the previous article (Ref. 1, hereinafter referred to as I) a formalism has been developed for the "interior" problem in the case of a rigidly rotating perfect fluid, with an arbitrary equation of state. By an appropriate choice of the coordinate system and a suitable definition of the four functions characterizing the metric, a simplification of the field equations is achieved: The four independent Einstein equations are reduced to a system of two differential equations of second order for two unknown functions, while the other two functions are obtained algebraically.

As an application of the formalism in I, we present here the solutions with vanishing magnetic Weyl tensor, which can be expressed in our coordinate system. Although the existence of such solutions has been proved recently,[2] their explicit expressions have not been given before. In Sec. II we show that the condition that a single component of the magnetic Weyl tensor vanishes restricts the equation of state to $\mu = p + \text{const}$ and the two field equations of Ref. I reduce to one fourth-order ordinary differential equation. In Sec. III the solutions of that equation are obtained and their properties are discussed. The solutions split into two families, the first of which is calculated analytically, and it is found that the equipressure surfaces are spherical. The second family of solutions is obtained up to quadratures and does not reduce to the spherically symmetric solution when there is no rotation. In both families it is found that the remaining components of the magnetic Weyl tensor also vanish.

II. SOLUTIONS OF THE FIELD EQUATIONS WITH VANISHING MAGNETIC WEYL TENSOR

In a recent paper[2] Collins has proved the existence of shear-free rotating perfect fluids with zero magnetic Weyl tensor. The "magnetic" part of the Weyl tensor with respect to the fluid flow is defined by

$$H_{ab} = \tfrac{1}{2}\epsilon_{ac}{}^{gh}C_{ghbd}u^c u^d, \tag{1}$$

where C_{abcd} is the Weyl tensor and u^a is the four-velocity of the fluid.

In the orthonormal frame defined in I [Eq. (13)],

$$\begin{aligned} &\omega^0 = (1/h)(dt + A\,d\varphi), \quad \omega^1 = h\rho\,d\varphi, \\ &\omega^2 = (1/\sqrt{Q})dh, \ \omega^3 = \Delta h\sqrt{Q}\,[d\rho + (R/hQ)dh\,], \end{aligned} \tag{2}$$

the nonzero components of H_{ab} are

$$H_{11} = R^2{}_{310} = -(1/2\rho^2 h^4\Delta)(hA_h - 2\rho A_\rho), \tag{3}$$

$$\begin{aligned} H_{22} = -H_{33} = R^1{}_{302} &= -\frac{1}{2\rho h^4\Delta}\left\{-hA_{\rho h} + 4A_\rho + \frac{R}{Q}\left(A_{\rho\rho} - \frac{A_\rho}{\rho}\right)\right. \\ &\left.\quad - \frac{hA_h Q_\rho}{2Q} - \frac{A^2{}_\rho}{2\rho h^4 Q}(hQA_h - RA_\rho)\right\}, \end{aligned} \tag{4}$$

$$\begin{aligned} H_{23} = H_{32} = \tfrac{1}{2}(R^1{}_{220} + R^1{}_{303}) &= \frac{1}{2\rho h^4 Q}\left\{QhN_h - RN_\rho\right. \\ &\quad + N\left(\frac{R}{\rho} + R_\rho - \frac{hQ_h}{2} - 6Q - \frac{A_\rho N}{2\rho h^4}\right) \\ &\left.\quad + \frac{1}{\Delta^2}\left(A_{\rho\rho} - \frac{A_\rho}{\rho} - A_\rho\frac{\Delta_\rho}{\Delta}\right)\right\}, \end{aligned} \tag{5}$$

where $N = RA_\rho - hQA_h$.

It is obvious that in a static field ($A = 0$) the conditions

$$H_{ab} = 0 \tag{6}$$

are satisfied identically. It has been proven[2] that even if the field is stationary ($A \neq 0$) there exist solutions of the field equations compatible with (6).

We shall use only the condition

$$H_{11} = 0, \tag{7}$$

and the field equations I [Eq. (25)] and I [Eq. (26)] to derive these solutions. Later it will be shown that all the conditions (6) are satisfied.

Following I, we put

$$\rho^2 = \xi,$$

so (7) becomes

$$(1/2\xi h^4\Delta)(hA_h - 4\xi A_\xi) = 0 \tag{8}$$

and can immediately be integrated to give

$$A = A(x), \quad x = \xi h^4, \tag{9}$$

where $A(x)$ is an arbitrary function of x. The dependence of A on the angular velocity Ω must be

$$A(x) = \Omega\cdot\tilde{A}(x), \tag{10}$$

where the function $\tilde{A}(x)$ may depend on Ω but is finite for $\Omega = 0$.

Now by changing the independent variable ξ to

$$\xi \to x = \xi h^4, \tag{11}$$

and noting that for any function $N(h,\xi)$

$$N_h(h,\xi) = N_h(h,x) + (4x/h)N_x(h,x),$$
$$N_\xi(h,\xi) = h^4\cdot N_x(h,x),$$

the field equation I [Eq. (26)] can be written[3] as

$$h^4(A'' + 2A'^3)\cdot(-hP_h + P) + A'(1-4xA'^2)(f - hf'/2) - 4xf(A'' + 2A'^3) = 0, \tag{12}$$

where

$$A' = \frac{dA}{dx}, \quad A'' = \frac{d^2A}{dx^2}, \quad f' = \frac{df}{dh}.$$

Two obvious solutions of (12) for any $P(h,x)$ can be easily excluded.

(i) $A'' + 2A'^3 = 0$ and $f - \frac{1}{2}hf' = 0$ because it gives as an equation of state $\mu + p = 0$, which is physically unacceptable.

(ii) $1 - 4xA'^2 = 0$, which implies $A'' + 2A'^3 = 0$, because it gives $A = \sqrt{x} + \text{const}$, which causes the denominators in I [Eq. (23′)] and I [Eq. (24′)] to vanish.

Setting

$$g(x) = A'(1-4xA'^2)/(A'' + 2A'^3) + 4x, \tag{13}$$

Eq. (12) can be written as

$$P_h - \frac{P}{h} = g(x)\frac{2f - hf'}{2h^5} - \frac{4x}{h^5}\left(2f - \frac{hf'}{2}\right). \tag{14}$$

This last equation is a linear first-order differential equation for P that can be integrated to give

$$P(h,x) = hT(x) + g(x) \times h\int \frac{2f - hf'}{2h^6}\,dh - 4xh\int \frac{4f - hf'}{2h^6}\,dh, \tag{15}$$

where $T(x)$ is an arbitrary function of x.

The function P must take the value 1 on the axis (I [Eq. (29)]), i.e., $P(h,x=0) = 1$. This regularity condition restricts the function $f(h)$, or equivalently the equation of state. We find that $f(h)$ must have the form

$$f = -\tfrac{3}{4}\kappa h^2 + [1/g(0)]h^4, \tag{16}$$

where $g(0) = g(x=0)$ and $\kappa = \text{const}$.

Using I [Eq. (12)], the equation of state is found to be

$$\mu = p + \tfrac{3}{2}\kappa/8\pi. \tag{17}$$

Now (15) can be rewritten in the form

$$P(h,x) = [1/g(0)]g(x) + h\Sigma(x) - \kappa x/h^2, \tag{18}$$

where $\Sigma(x)$ is an arbitrary function of x subject to the condition $\Sigma(0) = 0$.

Summarizing, the field equation I [Eq. (26)] has been satisfied with an equation of state given by (17) and functions A and P given by (9) and (18), respectively. We will examine now if there exists a choice of the functions $A(x)$ and $\Sigma(x)$ that will satisfy I [Eq. (25)]. Substituting the expressions found above for $f(h)$, $A(x)$, and $P(h,x)$ in I [Eq. (25)] we obtain a polynomial of fifth degree with respect to h, with coefficients which are functions of x. It is found that three of these coefficients vanish identically while the other three will vanish if $A(x)$ and $\Sigma(x)$ satisfy the equations

$$A'^2(2gg' - 4g'x + 8x - 4g) + g''(g - 4x) + g' - 2 = 0, \tag{19}$$

$$A'^2\Sigma(3g/4x - 3 - 4g') - (3/16x^2)\Sigma(g - 4x) = 0, \tag{20}$$

$$4x\Sigma' - \Sigma = 0, \tag{21}$$

where $g(x)$ is given by (13).

The choice $\Sigma(x) = 0$ satisfies Eqs. (20) and (21) (see Ref. 4) and we are left with Eq. (19) for the determination of $A(x)$. Putting

$$u(x) = 1/A'^2, \tag{22}$$

$$G(x) = g(x) - 2x, \tag{23}$$

and using (13) we have

$$g(x) = (2u + 8x - 4xu')/(4 - u') \tag{24}$$

and

$$G(x) = (2u - 2xu')/(4 - u'), \tag{25}$$

so that (19) becomes

$$(1/u)2GG' + G''(G - 2x) + G' = 0. \tag{26}$$

Equation (26) [or equivalently Eq. (19)] is an ordinary differential equation of fourth order for the function $A(x)$, as can be seen from (25) and (22). The function $P(h,x)$ is given by (18) with $\Sigma = 0$, i.e.,

$$P(h,x) = [1/g(0)]g(x) - \kappa x/h^2. \tag{27}$$

III. SOLUTIONS OF EQ. (26)

A. Case (a): $G = \text{const}$

When $G = \text{const}$, Eq. (26) is an identity, and Eq. (25) implies that $u(x)$ must be a linear function of x, while from (22) and (10) we may write for $A(x)$

$$A(x) = \Omega a(\sqrt{1 - b^2x} - 1), \quad a,b = \text{const}, \tag{28}$$

where the constants a and b are finite for $\Omega = 0$.

From (24) and (27) we find the function $P(h,x)$:

$$P(h,x) = 1 + b^2x(1 + b^2a^2\Omega^2) - x\kappa/h^2. \tag{29}$$

Finally from I [Eq. (24′)] and I [Eq. (23′)] we obtain $Q(h,x)$ and $R(h,x)$:

$$Q(h,x) = (h^2/4)\{-\kappa + b^2h^2(2 + b^2a^2\Omega^2)\}, \tag{30}$$

$$2\rho R = (x/h^2)\{-\kappa + b^2h^2(2 + b^2a^2\Omega^2)\} = (4x/h^4)Q. \tag{31}$$

We observe that the function $P(h,x)$ has the form that corresponds to the spherically symmetric solution

$$P = 1 + x\cdot\pi(h). \tag{32}$$

Although the above stationary solution was derived by using the vanishing of only one component of the magnetic Weyl tensor [Eq. (7)] and the field equations, it is easily checked that all the conditions (6) are satisfied. So for this solution, the magnetic Weyl tensor vanishes. Also the acceleration and vorticity four-vectors are parallel. These four-vectors, in our coordinate system, are given by

$$a^a = u^a{}_{;b}u^b = (1/h)(0,0,Q,-R/h),$$

$$\omega^a = (1/2\sqrt{-g})\epsilon^{abcd}u_b u_{c;d} = (1/2\rho h^3\Delta)(0,0,A_\rho,-A_h),$$

and it is seen from (8) and (31) that $hQA_h = RA_\rho$ so the acceleration and vorticity four-vectors are parallel. Because of the above properties this solution is Collins' case 2. The particular case in which there is no acceleration ($\dot u_2 = 0 = \dot u_1$ in Collins' notation) cannot be examined in our coordinate system because it implies $h = \text{const}$.

Another property of this solution is that, even though stationary, the equipressure surfaces are spherical. To prove that, let us look for an axially symmetric surface in Euclidean three-space described (in cylindrical coordinates R,z,φ) by $z = z(R)$ such that the geometry on that surface (I [Eq. (6)])

$$d\sigma_E^2 = \left\{1 + \left(\frac{dz}{dR}\right)^2\right\} dR^2 + R^2\, d\varphi^2 \tag{33}$$

is the same as that of the fluid's equipressure surfaces, $h = \text{const}$ (I [Eq. (5)]),

$$d\sigma^2 = \rho^2 h^2\, d\varphi^2 + \Delta^2 Q h^2\, d\rho^2. \tag{34}$$

From (33) and (34) we find

$$h\rho = R, \tag{35}$$

$$\Delta^2 Q - 1 = \left(\frac{dz}{dR}\right)^2. \tag{36}$$

Using (29)–(31) and the definition of $\Delta^2 = 1/(PQ - R^2)$, (36) is written

$$\left(\frac{dz}{dR}\right)^2 = h^2 b^2 \frac{R^2}{1 - h^2 b^2 R^2}. \tag{37}$$

The last equation can be integrated to give

$$Z = \pm\sqrt{1/h^2 b^2 - R^2} + \text{const}, \tag{38}$$

which is the equation of a sphere in the Euclidean three-space. Without loss of generality we may put const $= 0$.

B. Case (b): $G \neq$ const

When $G \neq$ const, we can partially integrate Eq. (26). In order to be able to find the static limit of this case it is convenient to write out explicitly the dependence of the various functions on the angular velocity Ω. In the following we shall denote by a tilde ($\tilde{}$) every function finite for $\Omega = 0$. By Eqs. (13) and (23) the functions $g(x)$ and $G(x)$ are finite for $\Omega = 0$. From (22) we can write for $u(x)$

$$u = (1/\Omega^2)\tilde{u}. \tag{39}$$

Defining the function $\tilde{\psi}(x)$ by

$$\tilde{u} = 4x\Omega^2 + (2x - G)\tilde{\psi}, \tag{40}$$

Eq. (25) gives

$$2x\tilde{\psi}' = (\tilde{\psi}G)'. \tag{41}$$

The latter is solved formally if we put

$$\tilde{\psi} = \tilde{L}', \tag{42}$$

and

$$G = 2(x - \tilde{L}/\tilde{L}'). \tag{43}$$

The function $\tilde{L}(x)$ will be determined from (26). Substituting for u and G from (40) and (43) we find the following equation for $\tilde{L}(x)$:

$$\frac{d}{dx}\left\{\frac{x\tilde{L}''^2}{\tilde{L}'^4}\cdot\Omega^2 + \frac{1}{2}\frac{\tilde{L}\tilde{L}''^2}{\tilde{L}'^4}\right\} = 0. \tag{44}$$

An integration of this equation gives

$$(\tilde{L}''^2/\tilde{L}'^4)(\Omega^2 x + \tfrac{1}{2}\tilde{L}) = \tilde{C}_1^2, \quad \tilde{C}_1 = \text{const}. \tag{45}$$

Now from (40), (42), and (43) we have

$$\tfrac{1}{2}\tilde{L} = \tilde{u}/4 - \Omega^2 x, \tag{46}$$

so (45) is written

$$(\tilde{u}''/16)\cdot\sqrt{\tilde{u}} = \tilde{C}_1(\tilde{u}'/4 - \Omega^2)^2. \tag{47}$$

The last equation can be integrated to give

$$\ln\left|\Omega^2 - \frac{\tilde{u}'}{4}\right| + \frac{\Omega^2}{\Omega^2 - \tilde{u}'4} = 2\tilde{C}_1\sqrt{\tilde{u}} + \tilde{C}_2,$$
$$\tilde{C}_2 = \text{const}. \tag{48}$$

Equation (48) determines, in principle, the function $u(x)$ and consequently $A(x)$. When $\tilde{C}_1 = 0$, this equation gives $\tilde{u}' = \text{const}$, which leads to case (a). The function $P(h,x)$ is given by (27) and Q and R by

$$Q = \frac{h^4}{g(0)}\cdot\left\{1 + 2\tilde{C}_1\sqrt{\tilde{u}} - \frac{1}{2}\frac{\Omega^2}{\Omega^2 - \tilde{u}'/4}\right\} - \frac{\kappa h^2}{4}, \tag{49}$$

$$2\rho R = \frac{2x}{g(0)}\left\{2 + 2\tilde{C}_1\sqrt{\tilde{u}} - \frac{\Omega^2}{\Omega^2 - \tilde{u}'/4}\right\} - \frac{\kappa x}{h^2}. \tag{50}$$

It is easily proved, using Eqs. (27) and (48)–(50) that the conditions (6) are satisfied, so for this solution the magnetic Weyl tensor vanishes. This is Collins' case 1. The particular case in which the acceleration is orthogonal to the vorticity vector ($\dot{u}_1 = 0$ in Collins' notation) means that $(1/\Delta)A_\rho = 0$. Either $1/\Delta = 0$, which cannot be examined in our coordinate system, or $A_\rho = 0$. The second possibility, together with (8), gives $A_h = 0$ so the field is static. The main property of the above solution is that it does not reduce to the spherically symmetric solution when there is no rotation. In order that $P(h,x)$ have the form $P = 1 + x\cdot\pi(h)$ of the static spherically symmetric field, $\tilde{u}(x)$ has to be a linear function of x. But this is not the case, when $\tilde{C}_1 \neq 0$, as can be seen from (48).

An isolated rotating perfect fluid mass is expected to have an oblate spheroidal shape which becomes spherical in the limit of no rotation. Neither of the solutions found above satisfies this criterion and hence neither can represent an isolated rotating mass.

[1]S. Bonanos and D. Sklavenites, J. Math. Phys. **26**, 2275 (1985).

[2]C. B. Collins, J. Math. Phys. **25**, 995 (1984).

[3]We use the field equations I [Eq. (25)] and I [Eq. (26)] putting $z = h^2$.

[4]The solution $\Sigma(x) = c\cdot x^{1/4}$, $c = \text{const}$, of Eq. (21) is not compatible with Eqs. (19) and (20).

On the viscous fluid interpretation of some exact solutions

J. Carot and J. Ibanez
Departament de Fisica Teorica, Facultat de Ciencies, Universitat de Palma de Mallorca, Palma de Mallorca, Spain

(Received 22 May 1984; accepted for publication 26 April 1985)

An example of the equivalence between a perfect fluid and a viscous fluid is presented, showing that the Schwarzschild interior solution obtained from a perfect fluid can also be derived from a viscous fluid with heat conduction. The equivalence between a scalar field and a viscous fluid is investigated, showing that under certain circumstances, both can generate, from Einstein's equations, the same space-time geometry. Some examples are presented and, in particular, it is shown that every plane-symmetric solution deduced from a scalar field can also be derived from a viscous fluid.

I. INTRODUCTION

Recently attention has been devoted to the problem of obtaining the same space-time geometry from different stress-energy tensors. As is well known, a metric tensor does not lead to a unique stress-energy tensor. It has been shown in several examples that, under certain conditions, an exact solution of Einstein's field equations can be interpreted as due to different material distributions satisfying the energy conditions. Tupper[1,2] and Raychaudhuri and Saha[3,4] found conditions for the equivalence between a perfect fluid and a viscous magnetohydrodynamic fluid with heat conduction, both leading to the same metric tensor and therefore the same space-time geometry.

Interesting cases arise when these ideas are applied to a number of important exact solutions to give them an alternative interpretation. For example, Coley and Tupper[5] showed that the zero-curvature Friedmann–Robertson–Walker (FRW) cosmological model can be deduced from a viscous fluid. On the other hand, there is an interest in viscous fluids as suitable models to describe, for example, the interior of a star in certain density ranges.[6,7] Also, in cosmology the effect of the viscosity has been taken into account by Belinskii *et al.*[8]

However, exact solutions for viscous fluids are difficult to find in the usual way. Therefore, the dual interpretation of known metrics can be seen as a useful tool either to better understand the effect of the viscosity or to obtain exact solutions for viscous fluid sources.

In this paper we first apply the results obtained by Tupper[1,2] and Coley and Tupper[5] to the Schwarzschild interior solution, so that it may be interpreted as due to a viscous fluid with heat conduction.

In the second part of the paper we investigate the analogy between a scalar field and a viscous fluid. Tabensky and Taub[9] considered the analogy between a scalar field and an irrotational stiff matter perfect fluid. However, such an analogy works only when the gradient of the scalar field (which is proportional to the velocity of the equivalent fluid) is timelike. Such a restriction upon the gradient of the scalar field is avoided when one considers a viscous fluid with heat conduction.

We illustrate these results with two examples; the first one concerns the plane-symmetric solutions which can be derived from a scalar field, as in the Tabensky and Taub paper, and we show that in this case the analogy is always possible, i.e., every plane-symmetric solution (derived from a scalar field) is equivalent to a viscous fluid. In the second example we study the scalar field solutions given by Tupper[10]; these solutions are static and spherically symmetric ones and they reproduce the same predictions as the Schwarzschild exterior solution with respect to the tests of general relativity (the three classical tests and the radar reflection experiment).

II. SCHWARZSCHILD INTERIOR SOLUTION

In this section we summarize the results obtained by Tupper,[1,2] particularized to a perfect fluid and a viscous fluid with heat conduction, and we apply them to the Schwarzschild interior solution.

The stress-energy tensor of a perfect fluid is given by

$$T_{ab} = (p+\rho)v_a v_b + p g_{ab}, \tag{1}$$

where $v^a v_a = -1$. The stress-energy tensor for a viscous fluid with heat conduction is[11,12]

$$\hat{T}_{ab} = (\hat{\rho}+\hat{p})u_a u_b + \hat{p} g_{ab} - 2\eta\sigma_{ab} + 2q_{(a}u_{b)}\,, \tag{2}$$

where $u^a u_a = -1$, q^a is the heat conduction vector, which is orthogonal to the velocity field ($q^a u_a = 0$), η is the shear viscosity ($\eta > 0$), and σ_{ab} is the shear tensor[11] satisfying

$$g^{ab}\sigma_{ab} = \sigma_{ab}u^b = 0\,. \tag{3}$$

The equivalence between both fluids, in the sense that both generate the same geometry, implies the equality between their respective stress-energy tensors; so, by equating (1) and (2), the following results are obtained (a detailed discussion is given in Tupper, Refs. 1 and 2):

$$\begin{aligned}
&\hat{\rho} = \alpha^2(p+\rho) - p\,, \quad \hat{p} = \tfrac{1}{3}(\hat{\rho}-\rho) + p\,,\\
&q_a = \alpha(p+\rho)(v_a - \alpha u_a)\,,\\
&-2\eta\sigma_{ab} = (p+\rho)v_a v_b - (\hat{\rho}+\hat{p})u_a u_b\\
&\qquad\qquad + (p-\hat{p})g_{ab} - 2q_{(a}u_{b)}\,,\\
&\alpha \equiv -v^a u_a\,.
\end{aligned} \tag{4}$$

The vector u^a is obtained from the differential equation resulting from the substitution of the definition of the shear tensor[11] into (4):

$$u_{(a;b)} + \mathring{u}_{(a}u_{b)} - (\theta/3)(g_{ab} + u_a u_b)$$

 0022-2488/85/092282-04$02.50

$$- [(p+\rho)/6\eta]\,[(\alpha^2-1)g_{ab} - (2\alpha^2+1)u_a u_b]$$
$$= [(p+\rho)/2\eta]\,[2\alpha u_{(a}v_{b)} - v_a v_b]\,. \tag{5}$$

Let us now apply these results to the particular case of the Schwarzschild interior solution, which is a static and spherically symmetric perfect fluid solution[13]

$$ds^2 = -\left(a - b\sqrt{1-\frac{r^2}{R^2}}\right)^2 dt^2 + \frac{dr^2}{1-r^2/R^2} + r^2\,d\Omega^2\,, \tag{6}$$

a and b being constants whose value is taken to match with the Schwarzschild exterior solution onto a surface $r=r_0$ ($2a = 3\sqrt{1-r_0^2/R^2}$, $b=\frac{1}{2}$). The perfect fluid corresponding to this solution is characterized by

$$\rho = 3/R^2 = \text{const}\,,\quad p = \frac{3b\sqrt{1-r^2/R^2} - a}{R^2(a - b\sqrt{1-r^2/R^2})}\,,$$
$$v^a = ([a - b\sqrt{1-r^2/R^2}]^{-1}, 0,0,0)\,. \tag{7}$$

From the definition of α and the requirement of spherical symmetry, we have

$$u_a = (-\alpha g_{00}^{1/2},\ \sqrt{\alpha^2-1}\,g_{11},\ 0,\ 0)\,. \tag{8}$$

Equations (4) read in this case

$$\hat{\rho} = R^{-2}(a - b\sqrt{1-r^2/R^2})^{-1}[a(1+2\alpha)^2 - 3b\sqrt{1-r^2/R^2}]\,,$$
$$\hat{p} = \tfrac{1}{3}R^{-2}(a - b\sqrt{1-r^2/R^2})^{-1}[a(2\alpha^2-5) + 9b\sqrt{1-r^2/R^2}]\,, \tag{9}$$
$$q_a = -\alpha(p+\rho)((\alpha^2-1)g_{00}^{1/2}, \alpha\sqrt{\alpha^2-1}\,g_{11}^{1/2}, 0,0)\,,$$

and (5) becomes a differential equation for α

$$\alpha\alpha' = -(1/2\eta)e^{\lambda}(p+\rho)(\alpha^2-1)^{3/2} + (\alpha^2-1)(1/r - \nu')\,. \tag{10}$$

To integrate (10) we suppose that η remains constant when $r<r_0$ and $\eta=0$ for $r \geqslant r_0$. This leads to

$$\alpha^2 = 1 + b^2\eta^2R^2\frac{r^2}{[Cb\eta(aR - b\sqrt{R^2-r^2}) - aR\,]^2}\,, \tag{11}$$

$C = \text{const}\,.$

So we have completely determined a viscous fluid (with constant shear viscosity) which generates the same metric tensor as the perfect fluid (7). Taking into account the above value for the constants a and b we obtain

$$\hat{p}(r_0) = q^a(r_0) = \sigma_{ab}(r_0) = 0\,,\quad u_a(r_0) = v_a\,. \tag{12}$$

Therefore, the solution (6) regarded as due to the viscous fluid can also match with the Schwarzschild exterior solution. Note that at the origin ($r=0$) both fluids coincide and are regular (there are no singularities).

III. EQUIVALENCE BETWEEN A SCALAR FIELD AND A VISCOUS FLUID

In this section we investigate the equivalence between a scalar field and a viscous fluid. The stress-energy tensor corresponding to a scalar field ϕ is given by

$$T_{ab} = \phi_a\phi_b - \tfrac{1}{2}\phi^c\phi_c g_{ab}\,, \tag{13}$$

where ϕ_a stands for the gradient of ϕ, and satisfies

$$\phi^a{}_{;a} = 0\,, \tag{14}$$

as a consequence of the conservation law $T^{ab}{}_{;b} = 0$.

Tabensky and Taub[9] showed that (13) is equivalent to a stiff matter perfect fluid, the velocity of the perfect fluid being proportional to ϕ_a. This equivalence is only valid if $\phi^a\phi_a < 0$.

As in the previous section, equivalence between (13) and the viscous fluid (2) implies

$$\phi_a\phi_b - \tfrac{1}{2}\phi^c\phi_c g_{ab} = (p+\rho)u_a u_b + p g_{ab} - 2\eta\sigma_{ab} + 2q_{(a}u_{b)}\,. \tag{15}$$

Contracting (15) with g^{ab}, $u^a u^b$, and $u^a q^b$, we get

$$3p - \rho = -\phi^a\phi_a\,,\quad \rho = \mathring{\phi}^2 + \tfrac{1}{2}\phi^a\phi_a\,,$$
$$q^a\phi_a = -Q^2/\mathring{\phi}\,, \tag{16}$$

where $Q^2 = q^a q_a$; $\mathring{\phi} = \phi^a u_a$ and we have assumed $\mathring{\phi} \neq 0$. Next, $\mathring{\phi}$ can be obtained by contracting (15) with $\phi^a u^b$

$$\mathring{\phi}^2 = \tfrac{1}{2}\{-\phi^a\phi_a + \sqrt{(\phi^a\phi_a)^2 + 4Q^2}\}\,, \tag{17}$$

where the positive sign for the square root has been taken to keep ρ positive.

From expressions (16) and (17) and after a straightforward calculation we obtain

$$\rho = \sqrt{(\phi^a\phi_a)^2 + 4Q^2}\,, \tag{18}$$
$$p = \tfrac{1}{3}(\rho - \phi^a\phi_a)\,, \tag{19}$$
$$q_a = -\mathring{\phi}(\phi_a + \mathring{\phi}u_a)\,. \tag{20}$$

After substituting (18)–(20) into (13), we have

$$2\eta\sigma_{ab} = \tfrac{1}{3}[\mathring{\phi}^2 + \phi^a\phi_a](g_{ab} + u_a u_b) - \mathring{\phi}^2 u_a u_b - 2\mathring{\phi}\phi_{(a}u_{b)} - \phi_a\phi_b\,. \tag{21}$$

This last expression constitutes a differential equation for the components of the velocity field u_a. Therefore, once it has been integrated, the viscous fluid equivalent to the scalar field is completely determined.

Depending on the character of ϕ_a (timelike, spacelike, or null) the pressure will vary continuously from values $p > \frac{1}{3}\rho$ in the case $\phi^a\phi_a < 0$ to values $p < \frac{1}{3}\rho$ when $\phi^a\phi_a > 0$, with $p = \frac{1}{3}\rho$ in the case $\phi^a\phi_a = 0$, as it can be easily seen from (19).

Note that now, conversely to the Tabensky and Taub results, the character of ϕ_a does not impose any other restriction, because from (20) it follows that $u^a u_a = -1$ is identically satisfied whatever the character of ϕ_a.

In the case $\phi^a\phi_a = 0$ we can see moreover that the vector $u^a + \hat{q}^a$ (where $\hat{q}^a$ stands for the space-like unit vector in the q^a direction) is lightlike and orthogonal to ϕ^a.

From (21) it is easy to see that q^a is an eigenvector of σ_{ab} with eigenvalue $-Q^2/3\eta\mathring{\phi}^2$.

In the special case in which $q^a = 0$ we can distinguish two different situations: in the case $\phi^a u_a \neq 0$, $2\eta\sigma_{ab} = 0$ and we trivially reproduce the Tabensky and Taub results.[9] The other case $\phi^a u_a = 0$ leads to $p = -\frac{1}{3}\rho$, namely, negative pressure, which seems to be an unphysical situation.

If we set $\phi^a u_a = 0$ in the general case $q^a \neq 0$, we obtain immediately $q^a = 0$ and the consequent results given above.

We have shown how we can determine a viscous fluid generating the same solution as a scalar field. Note that field equations can be written simply as $R_{ab} = \phi_a \phi_b$ in the case of a scalar field; these equations can be easily integrated in some cases (e.g., for plane-symmetric solutions,[9] Tabensky and Taub formulated the Cauchy problem for the field equations by using a Riemann–Volterra representation). Then, taking into account all that was stated above, exact solutions for viscous fluid sources can be obtained in a more simple way; in this sense the scalar field could be seen as an intermediate to obtain solutions for more complex stress-energy tensors.

To illustrate the above results we apply them to two solutions corresponding to a scalar field.

(a) The first is the general plane-symmetric solution for a scalar field[9]

$$ds^2 = (e^{\Omega}/\sqrt{t})(-dt^2 + dz^2) + t(dx^2 + dy^2), \tag{22}$$

where the scalar field satisfies

$$\phi_{zz} = \phi_{tt} + t^{-1}\phi_t, \tag{23}$$

and the derivatives of Ω are related to those of ϕ by

$$\Omega_t = t(\phi_t^2 + \phi_z^2), \quad \Omega_z = 2t\phi_t\phi_z. \tag{24}$$

Our purpose is to show that every plane-symmetric solution of type (22) can always be derived from a viscous fluid source. To do this, we start from (21) by calculating the components of the shear tensor according to this expression and equating them to those calculated from the definition[11] of the shear tensor.

The requirement of plane symmetry applied to the fluid imposes $u_x = u_y = 0$. So Eq. (21) leads only to one differential equation

$$\begin{aligned} u_z^2(u_{t,t} - \tfrac{1}{2}t^{-1}u_t) - u_{t,z}u_t u_z - (1/2\eta)(e^{\Omega}\sqrt{t}) \\ \times [\phi_t + \sqrt{t}\, e^{-\Omega} u_t(\phi_z u_z - \phi_t u_t)]^2 \\ = -(e^{\Omega}\sqrt{t})t\phi_t\phi_z u_z. \end{aligned} \tag{25}$$

The normalization condition for the velocity field reads

$$u_z^2 - u_t^2 = -e^{\Omega}/\sqrt{t}. \tag{26}$$

We take

$$u_t = (e^{\Omega}/\sqrt{t})^{1/2}\cosh\theta, \quad u_z = (e^{\Omega}/\sqrt{t})^{1/2}\sinh\theta. \tag{27}$$

Substituting (27) in (25) we obtain

$$\begin{aligned} (e^{\Omega}/\sqrt{t})^{1/2}\{(\Omega_t - \tfrac{1}{2}t^{-1} - 2\theta_z)\cosh\theta + (2\theta_t - \Omega_z)\sinh\theta\} \\ = \eta^{-1}(\phi_z\cosh\theta - \phi_t\sinh\theta)^2. \end{aligned} \tag{28}$$

Equation (28) constitutes a quasilinear partial differential equation.[14] Therefore every plane-symmetric solution obtained from a scalar field ϕ can also be derived from a viscous fluid.

As an example illustrating this general case, let us consider the scalar field $\phi = az$, which generates the line element[9]

$$ds^2 = (e^{a^2t^2/2}/\sqrt{t})(-dt^2 + dz^2) + t(dx^2 + dy^2). \tag{29}$$

Equation (25) can be easily integrated, assuming η constant, to give

$$\begin{aligned} u_t &= 2\eta\sqrt{t}\, e^{a^2t^2/2}, \\ u_z &= e^{a^2t^2/4}(4\eta^2 t e^{a^2t^2/2} - 1/\sqrt{t})^{1/2}. \end{aligned} \tag{30}$$

The other magnitudes characterizing the viscous fluid are easily calculated from Eqs. (18)–(20)

$$\begin{aligned} q_t &= -2\eta a^2 t[-1 + 2\eta t^{3/2}e^{a^2t^2/2}], \\ q_z &= -4\eta^2 a^2 t^2[4\eta^2 e^{a^2t^2} - (1/\sqrt{t})e^{a^2t^2/2}]^{1/2}, \\ \rho &= a^2\{16\eta^2 t^3(4\eta^2 t - e^{-a^2t^2/2}/\sqrt{t}) + te^{-a^2t^2}\}^{1/2}, \\ p &= \tfrac{1}{3}(\rho - a^2\sqrt{t}\, e^{-a^2t^2/2}). \end{aligned} \tag{31}$$

(b) Let us now study other solutions corresponding to a scalar field which have the interesting property of reproducing the tests of general relativity.[10] These solutions correspond to a static spherically symmetric asymptotically flat space-time with line element

$$ds^2 = -e^{2\nu(r)}dt^2 + e^{2\mu(r)}(dr^2 + r^2 d\Omega^2). \tag{32}$$

In order to satisfy the tests of general relativity, two solutions have been given.

$$\begin{aligned} \text{(i)}\quad \exp 2\mu &= (1 + M/r)^{2+m/M}(1 - M/r)^{2-m/M}, \\ \exp 2\nu &= (1 - M/r)^{m/M}(1 + M/r)^{-m/M}, \end{aligned} \tag{33}$$

which corresponds to the scalar field

$$\phi = (hm/2M)\ln[(1 - M/r)/(1 + M/r)], \tag{34}$$

where h, m, and M are constants. In this case the differential equation for the velocity field (21) is written

$$\begin{aligned} u_{r,r} + \frac{h^2}{2\eta}\frac{m^2r^4}{(r^2 - M^2)^4}\left(\frac{r - M}{r + M}\right)^{m/M}u_r^2 + \frac{h^2}{2\eta}\frac{m^2}{(r^2 - M^2)^2} \\ - [(r^2 + M^2 - 3mr^2)/r(r^2 - M^2)]\,u_r = 0. \end{aligned} \tag{35}$$

This equation always has a solution for every r different from $r = 0$ and $r = M$ (provided $\eta \neq 0$).

$$\begin{aligned} \text{(ii)}\quad \exp 2\mu &= \exp(2m/r), \\ \exp 2\nu &= \exp - (2m/r), \end{aligned} \tag{36}$$

which corresponds to the scalar field

$$\psi = -hm/r. \tag{37}$$

In this case the differential equation for the velocity field is

$$u_{r,r} + \frac{h^2m^2}{2\eta r^4}(1 + e^{-2m/r}u_r^2) - \left(\frac{1}{r} - \frac{3m}{r^2}\right)u_r = 0. \tag{38}$$

As in Eq. (35), this equation always has a solution, except in $r = 0$.

ACKNOWLEDGMENTS

We would like to thank Professor Ll. Mas for his attentive reading of the original and valuable comments.

We would like also to thank Dr. C. Bona for helpful discussions and suggestions.

This work was supported in part by C.A.Y.C.I.T.

[1] B. O. J. Tupper, Gen. Relativ. Gravit. 15, 849 (1983).
[2] B. O. J. Tupper, J. Math. Phys. 22, 2666 (1981).
[3] A. K. Raychaudhuri and S. K. Saha, J. Math. Phys. 22, 2237 (1981).
[4] A. K. Raychaudhuri and S. K. Saha, J. Math. Phys. 23, 2554 (1982).
[5] A. A. Coley and B. O. J. Tupper, Astrophys. J. 271 1 (1983).
[6] V. Canuto, "Neutron stars: General review," presented at the Solvay Con-

ference on Astrophysics and Gravitation, Brussels, 1973.
[7]M. Cosenza, L. Herrera, M. Esculpi, and L. Witten, J. Math. Phys. **22**, 118 (1981).
[8]V. A. Belinskii, E. S. Nikomarov, and S. M. Khalatnikov, Sov. Phys. JETP **50** (1979).
[9]R. R. Tabensky and A. H. Taub, Commun. Math. Phys. **29**, 61 (1973).
[10]B. O. J. Tupper, Nuovo Cimento B **19**, 135 (1974).
[11]G. F. R. Ellis, "Relativistic cosmology," in *General Relativity and Cosmology, XL VII Enrico Fermi Summer School Proceedings*, edited by R. K. Sachs (Academic, New York, 1971).
[12]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[13]D. Kramer, H. Stephani, M. Mc. Callum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (VEB, Berlin, 1980).
[14]F. John, *Partial Differential Equations* (Springer, New York, 1982).

Solutions to wave equations on black hole geometries. II

W. E. Couch
Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada T2N 1N4

(Received 17 January 1985; accepted for publication 9 May 1985)

Methods from the author's previous work and the classes of solutions which they produce are extended to the wave equations which govern the massive scalar field and the massless spin-$\frac{1}{2}$ field on the Kerr–Newman geometry and the massless fields of spin 1 and 2 on the Kerr geometry. The solutions found are exact and expressed in simple closed forms in terms of elementary functions, but they only exist when appropriate constraints hold on some of the black hole parameters and on the frequency of the field in some cases. The behavior on the horizon, at null infinity, and with respect to the angular variables is analyzed for some example solutions. For the examples studied, it is found that the ones having radial behavior of a normal mode are not free of angular singularities. An exact relation is established between the scalar wave equation on the extreme Kerr–Newman geometry and the Whittaker–Hill equation.

I. INTRODUCTION

In a previous paper[1] (hereafter denoted by I) the radial differential equations which govern fields on the Schwarzschild, Reissner–Nordström, and extreme Kerr black hole background geometries were examined with a view to finding exact solutions which may be written explicitly in terms of elementary functions in closed form. Several classes of simple closed-form solutions for the massless spin-zero field were displayed in detail. With a time dependence of $\exp(-i\omega t)$ for the field, these special solutions were found to exist for countably infinite sets of particular values among *some* of the quantities ω, mass M of the black hole, charge e, angular momentum a, and separation constant λ (for the angular differential equation resulting from separation of variables). For the extreme Reissner–Nordström and extreme Kerr geometries the special solutions discussed in I have singular angular dependence of the field, while some solutions on nonextreme Reissner–Nordström have nonsingular angular dependence.

In this paper the methods of I will be used to find closed-form solutions in situations more general than those of I. One generalization will be to a massive spin-0 field on the Kerr–Newman[2] background geometry, and a second generalization will be to fields with nonzero spin: spin $\frac{1}{2}$ on the Kerr–Newman geometry, and spin 1 and 2 on the Kerr geometry. We will also include a test electromagnetic field as part of the background in the spin-0 and spin-$\frac{1}{2}$ cases.

Our analysis will incidentally provide some of the proofs omitted in I.

The Kerr–Newman metric in Boyer–Lindquist coordinates is given by

$$\begin{aligned} ds^2 = {} & [1-(2Mr-e^2)/\Sigma]\,dt^2-(\Sigma^2/\Delta)\,dr^2-\Sigma\,d\theta^2 \\ & -[r^2+a^2+a^2\sin^2\theta(2Mr-e^2)/\Sigma]\sin^2\theta\,d\phi^2 \\ & +[2(2Mr-e^2)a\sin^2\theta/\Sigma]\,d\phi\,dt, \end{aligned} \qquad (1)$$

where $\Delta \equiv r^2-2Mr+a^2+e^2$ and $\Sigma \equiv r^2+a^2\cos^2\theta$. We restrict ourselves to physical values of the black hole parameters, namely real values satisfying $M^2 \geqslant a^2+e^2$, $M>0$.

Inclusion of a test electromagnetic field means that the covariant derivative ∇_ν with respect to (1) is replaced by $\nabla_\nu + iQ\mathscr{A}$ in the construction of covariant wave equations; here Q is the charge of the test field and the electromagnetic one-form is given by

$$\mathscr{A} = (er/\Sigma)(dt - a\sin^2\theta\,d\phi)\,. \qquad (2)$$

In Sec. II we list the radial differential equations which are considered in this paper, namely the spin-0 and spin-$\frac{1}{2}$ fields on the Kerr–Newman geometry and the spin-1 and spin-2 fields on the Kerr geometry. Both extreme and nonextreme geometries and all subclasses $ae=0$ are included. All fields considered are separable, so field components may be expressed as $f(r)S(\theta)\exp(-i\omega t+im\phi)$. The following constant parameters occur in the field equations of Sec. II: M, e, a, Q, ω, m, the mass μ of the field, and λ. We require $\mu^2>0$, m an integer for integer spin, and m a half-odd integer for spin $\frac{1}{2}$.

In order to avoid proliferating special cases we shall require $\omega\neq 0$ and thus not consider static fields. Static fields could be easily found by the methods of this paper, if desired, but in most, or maybe all, cases they are already known.

All of the differential equations on black hole geometries which we consider have the property that they take the form

$$\psi_{xx} + A\psi = 0\,, \qquad (3)$$

where A is a fourth-degree polynomial in r, and x is a coordinate defined by $dx = \delta\,dr/\Delta$, $\delta=$ const. [In some cases a factor transformation on the radial field is used in addition to achieve (3).] In terms of x, A is a fourth-degree polynomial in $\coth x$ in the exterior of nonextreme geometries ($\coth x$ becomes $\tanh x$ in the interior), and it is a fourth-degree polynomial in $1/x$ on extreme black hole geometries. The constant coefficients in A are different functions of $(M,e,a,Q,\omega,m,\mu,\lambda)$ in the various field equations of Sec. II.

The methods used here and in I for obtaining closed-form solutions make use of the property (3), and the solutions exist for infinite classes of special values of the coefficients in A, and so for special values of some (but not all) of the parameters $(M,e,a,Q,\omega,m,\mu,\lambda)$.

The extreme geometries ($M^2=a^2+e^2$ and subclasses $M^2=a^2$ and $M^2=e^2$) are dealt with in Sec. III. There it is shown that the field equations considered in this paper when taken on extreme geometries can, with restrictions on parameters in some cases, be further transformed from (3) to a standard equation of mathematical physics, namely the Whittaker–Hill equation. In particular it is shown that on the extreme Kerr–Newman geometry the massless scalar field (with no restrictions on parameters) is governed by the Whittaker–Hill equation, thus the known properties of Whittaker functions can be used to analyze the massless scalar field. This fact was shown in I for the cases $M^2 = a^2$ and $M^2 = e^2$. In order to obtain closed-form solutions we make use of special solutions to the Whittaker–Hill equation which exist for classes of special values of the constants in the equation and which can be obtained from finite trigonometric series known in the theory of the Whittaker–Hill equation. The special classes of Whittaker–Hill constants, their relation to $(M,e,a,Q,\omega,m,\mu,\lambda)$, and the corresponding solutions are given in Sec. III. This method for extreme geometries yields no solutions in the case of spins 1 and 2.

For nonextreme geometries the quantity A in (3) has the form

$$A = A_0 + A_1 y + A_2 y^2 + A_3 y^3 + A_4 y^4 , \tag{4}$$

where the coefficients A_i, $i = 0,1,2,3,4$, are constant, and y is $\coth x$ or $\tanh x$. In Sec. IV a method for obtaining closed-form solutions to (3) with A given by (4) is described and conditions sufficient for the applicability of the method are derived. By the method, a solution ψ is explicitly determined as a simple function of y from (42) when all A_i are specified, plus, for any non-negative integer n, $n+4$ additional parameters [α, β, γ, and $n+1$ coefficients a_i of a polynomial of degree n in $(y+1)$] are also known. The method gives a simple way in which these additional parameters may always be determined from the A_i; however, the determination puts restrictions on the specification of the A_i and thus on the parameters $(M,e,a,Q,\omega,m,\mu,\lambda)$. For nonextreme exterior geometries, r is linear in $\coth x$, and so the method determines, for each radial field, simple solutions $\psi(r)$ which are given explicitly by (62) as a function of r and of whichever of the parameters $(M,e,a,Q,\omega,m,\mu,\lambda)$ occur in the field equation.

In Sec. V the method of Sec. IV is applied to the field equations listed in Sec. II, the allowed classes of black hole parameter values are thus found, and thus the corresponding closed-form solutions. For all spins, it turns out that at each fixed n there are infinite classes of allowed parameters. The domain of these solutions, and those of Sec. III, may be taken to be the region exterior to the horizon of the black hole, but it may alternatively be taken to be the interior region.

In Sec. VI some properties of particular example solutions are discussed. Among the examples are the following. There is a massless scalar field (and also a spin-$\frac{1}{2}$ field) on the Kerr–Newman geometry which behaves radially as a normal mode but is angularly singular. Another such scalar field [given by (98)] fails to be a normal mode for the opposite reason—it is angularly nonsingular but its radial dependence is not that of a normal mode. This field has finite null data on past null infinity and on the past horizon but blows up on the future horizon. The existence of an exact solution with this behavior shows that in order to obtain a finite field from past data there must be conditions on the data in addition to finiteness. That the data be bounded as the retarded time on the past horizon increases to infinity would be sufficient to exclude this solution. On the extreme Kerr–Newman geometry there is a nonstatic massive field [given by (96)] which is angularly nonsingular and has the same behavior at null infinity and on the horizons as does a static massless field.

All examples in Sec. VI fail to be exact normal modes either in their radial behavior or their angular behavior. However, for $n > 0$ and $|a\omega|$ not small some of the solutions (27) or (62) may be normal modes, or may be exact illustrations of other interesting kinds of solutions which have been shown or suggested to exist for fields on black hole background geometries. We do not pursue an investigation of those possibilities here.

II. THE DIFFERENTIAL EQUATIONS

A. Massive spin-0 field on a Kerr–Newman background geometry

A spin-0 field Ψ of mass μ on the background (1) satisfies

$$(\nabla^\nu + iQ\mathscr{A}^\nu)(\nabla_\nu + iQ\mathscr{A}_\nu)\Psi + \mu^2\Psi = 0 . \tag{5}$$

This equation is separable, $\Psi = \psi(r)S(\theta)\exp(-i\omega t + im\phi)$, and the resulting radial equation is (subscripts r, x, and x' will denote differentiation)

$$\Delta(\Delta\psi_r)_r + \epsilon^2 M^2 A\psi = 0 , \tag{6}$$

where

$$\epsilon \equiv \sqrt{M^2 - a^2 - e^2}/M , \quad 0 \leqslant \epsilon \leqslant 1 , \tag{7}$$

$$\begin{aligned} \epsilon^2 M^2 A = -\{ & r^4(\mu^2 - \omega^2) + r^3(-2M\mu^2 + 2eQ\omega) \\ & + r^2[\mu^2(a^2+e^2) + \lambda - a^2\omega^2 - e^2Q^2] \\ & + r[-2M\lambda - 2Ma^2\omega^2 + 4Mam\omega \\ & + 2a^2\omega eQ - 2ameQ] \\ & + \lambda(a^2+e^2) + a\omega e^2(a\omega - 2m) - a^2m^2\} . \end{aligned} \tag{8}$$

The separation constant λ appears also in the angular differential equation for $S(\theta)$.

B. Spin-$\frac{1}{2}$ field on a Kerr–Newman background geometry

A field of spin $\frac{1}{2}$ satisfies the covariant Dirac equation and this equation has been shown to be separable[3–5] on the background (1). Both components of the radial field can be obtained from a single function $R(r)$ which satisfies a second-order differential equation. We use the equation given by Page,[4] which may be written

$$\begin{aligned} \sqrt{\Delta}(\sqrt{\Delta}R_r)_r - \frac{i\mu\Delta}{b + i\mu r}R_r + \Big\{ & \frac{K^2 + i(r-M)K}{\Delta} - 2i\omega r \\ & + ieQ - \frac{\mu K}{b + i\mu r} - \mu^2 r^2 - b^2 \Big\} R = 0 , \end{aligned} \tag{9}$$

where $K \equiv (r^2 + a^2)\omega - eQr - am$ and $b^2 = \lambda + a^2\omega^2 - 2am\omega$.

C. Fields of spin 1 and 2 on Kerr and Reissner–Nordström geometries

The physically interesting equations for fields of spin 1 and 2 in connection with a Kerr–Newman geometry are those governing electromagnetic and gravitational perturbations[3,6,7] of a charged, rotating black hole. Both perturbations are necessarily present and the equations governing them which have been derived in Refs. 3, 6, and 7 are coupled with respect to the two fields and do not by separation of variables split into radial and angular ordinary differential equations. So when $ae \neq 0$ we have no radial equation to which our methods might be applied.

When $e = 0$ (Kerr geometry) the perturbation equations are separable[8] and we use the Teukolski equation[8]

$$\Delta^{1-s}(\Delta^{1+s}R_r)_r + BR = 0, \tag{10}$$

where

$$B = (r^2 + a^2)^2\omega^2 - 4aMr\omega m + a^2m^2 + 2iams(r - M) - 2isM\omega(r^2 - a^2) + 2is\omega r\Delta - a^2\omega^2\Delta - \lambda\Delta, \tag{11}$$

and s is the spin weight ± 1, ± 2 for fields of spin 1 and 2, respectively.

When $a = 0$ (Reissner–Nordström) the equations for the spin-1 and -2 perturbations have been decoupled and separated by Chandrasekhar,[9] Moncrief,[10] and others. However, the method of obtaining (3) which works for (6), (9), and (10) does not work when applied to the radial equations of Chandrasekhar or Moncrief. The Moncrief equations are the simplest and most closely resemble (3), but for them, instead of being a fourth-degree polynomial in r, A is degree six plus $1/r$ and $1/r^2$ terms. There are Debye potentials for the spin-1 perturbations,[11] but their governing equation also does not take the form (3) but rather has A of degree four plus $1/r$ and $1/r^2$ terms. Thus for spin-1 and -2 fields we have no radial equations of the form to which our method is applicable when $e \neq 0$.

III. CLOSED-FORM SOLUTIONS ON EXTREME GEOMETRIES

For the extreme ($M^2 = a^2 + e^2$) black hole geometries the coordinate change from r to x' given by

$$x' = -M/(r - M), \tag{12}$$

coupled with a factor transformation on the radial field in some cases, will transform each of the field equations of Sec. II into

$$\psi_{x'x'} + A\psi = 0, \tag{13}$$

where A is a polynomial of fourth degree in $1/x'$, which we write as

$$A = d_0 + d_1/x' + d_2/x'^2 + d_3/x'^3 + d_4/x'^4. \tag{14}$$

[Equation (13) with (14) is the form of (3) on extreme geometries, the prime is on x merely to agree with the notation of I.] The method in I for obtaining closed-form solutions on extreme geometries is to transform (13) to the Whittaker–Hill equation

$$\chi_{xx} + (\theta_0 + 2\theta_1 \cos 2x + 2\theta_2 \cos 4x)\chi = 0, \tag{15}$$

because it has solutions expressible in terms of *finite* trigonometric sereis[1,12] when certain relations hold among the constants θ_0, θ_1, and θ_2.

The sequence of transformations $x' = c\exp(2z)$, $\psi = \chi\exp(z)$, $z = ix$, with c constant, will convert (13) into (15) if we impose conditions on the coefficients d_i, namely,

$$d_4 = c^4 d_0 \tag{16a}$$

and

$$d_3 = c^2 d_1, \tag{16b}$$

with the θ_i identified as

$$\theta_0 = 1 - 4d_2, \tag{17a}$$

$$\theta_1 = -4cd_1, \tag{17b}$$

$$\theta_2 = -4c^2 d_0, \tag{17c}$$

where c is known up to sign from (16b).

The solutions to (15) are most conveniently described in terms of parameters ξ, p, and λ_h defined by

$$\xi^2 = 16\theta_2, \tag{18a}$$

$$(p+1)\xi = -2\theta_1, \tag{18b}$$

$$\lambda_h = \theta_0 + 2\theta_2, \tag{18c}$$

and in terms of a quantity, V, defined by

$$\chi = V\exp(-\xi\cos 2x/4), \tag{19}$$

which satisfies Ince's equation

$$V_{xx} + \xi(\sin 2x)V_x + (\lambda_h - p\xi\cos 2x)V = 0. \tag{20}$$

For θ_0, θ_1, θ_2 real and $\theta_2 > 0$, (20) has four types of finite series solutions, which are[1,12]

$$V = \left(\sum_{k=0}^{n} a_k \cos 2kx,\ \sum_{k=0}^{n} a_k \cos(2k+1)x,\ \sum_{k=1}^{n} a_k \sin 2kx,\ \sum_{k=0}^{n} a_k \sin(2k+1)x\right), \tag{21}$$

and for each type in (21), p must satisfy the corresponding condition

$$p = (2n, 2n+1, 2n, 2n+1), \tag{22}$$

where n is any non-negative integer [$n \geqslant 1$ for type (iii)]. For each type, the column vector $\mathbf{a} = (a_0,\ldots,a_n)^T$ of coefficients and the quantity λ_h are determined by an eigenvalue problem

$$\mathscr{L}\mathbf{a} = \lambda_h \mathbf{a}, \tag{23}$$

where $\mathscr{L}$ is a tridiagonal matrix given correspondingly to the types in (21) by

$$\begin{pmatrix} - & (n+1)\xi, & (n+2)\xi, & (n+3)\xi, & \ldots, & 2n\xi \\ 0, & 4\cdot 1^2, & 4\cdot 2^2, & 4\cdot 3^2, & \ldots, & 4n^2 \\ 2n\xi, & (n-1)\xi, & (n-2)\xi, & (n-3)\xi, & \ldots\xi & - \end{pmatrix}, \tag{24a}$$

$$\begin{pmatrix} - & (n+2)\xi, & (n+3)\xi, & \ldots, & (2n+1)\xi \\ 1+(n+1)\xi, & 3^2, & 5^2, & \ldots, & (2n+1)^2 \\ n\xi, & (n-1)\xi, & (n-2)\xi, & \ldots\xi & - \end{pmatrix}, \tag{24b}$$

$$\begin{pmatrix} - & (n+2)\xi, & (n+3)\xi, & \dots, & 2n\xi \\ 4\cdot 1^2, & 4\cdot 2^2, & 4\cdot 3^2, & \dots, & 4n^2 \\ (n-1)\xi, & (n-2)\xi, & (n-3)\xi, & \dots\xi & - \end{pmatrix}, \tag{24c}$$

$$\begin{pmatrix} - & (n+2)\xi, & (n+3)\xi, & \dots, & (2n+1)\xi \\ 1+(n+1)\xi, & 3^2, & 5^2, & \dots, & (2n+1)^2 \\ n\xi, & (n-1)\xi, & (n-2)\xi, & \dots\xi & - \end{pmatrix}, \tag{24d}$$

The main diagonal of $\mathscr{L}$ is the middle row in each case of (24), the adjacent diagonals are the other two rows. All elements of each $\mathscr{L}$ not on one of these diagonals are zero. The matrix $\mathscr{L}$ is n by n for type (iii) and $(n+1)$ by $(n+1)$ for the other types.

Each matrix $\mathscr{L}$ is similar to a real, symmetric, Jacobi matrix and has $(n+1)$ real, distinct eigenvalues λ_h [n eigenvalues for type (iii)]. Hence, in each type, there are $(n+1)$ conditions (18c) [n conditions for type (iii)] on $\theta_0 + 2\theta_2$, and each condition gives a finite series (21) for V. Each condition can always be satisfied since we may consider θ_2 to be given, θ_1 to be defined by (18a) and (18b), and θ_0 to be defined by (18c). After each eigenvalue λ_h is found, all the corresponding coefficients a_k may be very easily generated recursively from (23); for example, starting at the last equation of (23) with $a_n = 1$, each successive equation determines a single a_k in terms of known quantities.

To insure that the transformation to (15) can actually be carried out and the types (24) obtained, we require (16)–(18) and

$$c \neq 0, \quad \theta_2 > 0, \quad \theta_0,\theta_1,\theta_2 \text{ real.} \tag{25}$$

Note that these imply $d_0 d_1 d_3 d_4 \neq 0$. As a separate case we sometimes allow $\theta_1\theta_2 = 0$ if it leads to a nontrivial solution, but then (16) and (18) imply $\theta_1 = \theta_2 = d_0 = d_1 = d_3 = d_4 = 0$, and (15) is not needed since ψ is easily obtained from the single powers of $|x'|$ which solve (13). In many cases allowing $\theta_0\theta_1 = 0$ leads only to static fields and we shall not mention each such occurrence in what follows.

In the following subsections the method described above will be applied to the field equations of Sec. II on the corresponding extreme geometry. This will be accomplished by obtaining the transformation to (15) and finding the d_i in terms of the parameters ($M,e,a,Q,\omega,m,\mu,\lambda$). Our conditions (16), (18a), and (25) are then conditions on these parameters, since θ_0, θ_1, θ_2, and ξ are determined in terms of the parameters via (17) and (18b). Each matrix $\mathscr{L}$ is then determined by (24), and for each n in each type the remaining conditions on the parameters are known from (18c) and

$$\det(\mathscr{L} - \lambda_h I) = 0 . \tag{26}$$

For each n in each type, χ is given by (19) and a solution for the field is given by

$$\psi = |x'|^{1/2}\chi = |x'|^{1/2} V \exp\left[\frac{-1}{(p+1)}\left(d_1 x' + \frac{d_3}{x'}\right)\right]. \tag{27}$$

Specializations of the parameters ($M,e,a,Q,\omega,m,\mu,\lambda$) other than (16)–(18) and (25) can lead to solutions in terms of standard functions; for example, when $d_3 = d_4 = 0$, (13) is solved by Coulomb functions, or when $d_0 = d_1 = d_3 = 0$, the solutions to (13) are simply expressed in terms of Bessel functions.

A. Massive spin-0 field on extreme Kerr–Newman geometry

In this case, Eq. (6) of Sec. II A is the relevant field equation which transforms into (13) with A of (14) given by

$$d_0 = T^2/M^2 , \tag{28a}$$

$$d_1 = -2T[2M\omega - eQ]/M , \tag{28b}$$

$$d_2 = 6M^2\omega^2 - M^2\mu^2 - 6M\omega eQ + a^2\omega^2 + e^2Q^2 - \lambda , \tag{28c}$$

$$d_3 = 2M^2\mu^2 - 4M^2\omega^2 + 2M\omega eQ , \tag{28d}$$

$$d_4 = M^2\omega^2 - M^2\mu^2 , \tag{28e}$$

where $T \equiv (M^2 + a^2)\omega - am - eQM$. With these coefficients, (16) can be satisfied in two (and only two) instances:

$$\mu = 0 , \tag{29a}$$

and

$$M^2\mu^2 = eQ(2M\omega - eQ) . \tag{29b}$$

No conditions beyond (29) are needed to obtain (15) and thus we have shown, as mentioned in the Introduction, that the scalar field is governed by the Whittaker–Hill equation if the field is massless [or also if (29b) holds].

We now give the results of applying our full conditions (16)–(18) and (25) to (28).

It turns out that the case (29b) is empty unless we allow $\theta_1\theta_2 = 0$, and then $\mu^2 = \omega^2$, $eQ = M\omega$, $a^2\omega = am$, $d_2 = a^2\omega^2 - \lambda$, and since all the d_i except d_2 vanish we have from (13)

$$\psi = |x'|^{(1 \pm \sqrt{1 + 4\lambda - 4a^2\omega^2})/2} . \tag{30}$$

If in this case $a \neq 0$, then $a\omega = m$ and we have the restriction $m^2 < M^2\omega^2 < M^2Q^2$.

In the case of (29a) our conditions (which include $\theta_1\theta_2 \neq 0$) can be shown to imply

$$2M\omega = eQ + i\tau , \tag{31}$$

where τ is real and nonzero so that ω is necessarily complex. Furthermore, two subclasses completely cover (29a): $a = 0$ and $aeQ = mM$.

In the first instance ($a = 0$) we have from (28) and (18)

$$\tau = \tfrac{1}{4}(p+1)^2 , \tag{32a}$$

$$\xi = 4j(\tau/|\tau|)\sqrt{\tfrac{1}{4}(p+1)^2 + e^2Q^2} , \tag{32b}$$

$$\lambda_h = 1 + 4\lambda + 2(p+1)^2 + 4e^2Q^2 , \tag{32c}$$

where $j = \pm 1$. For each value of eQ, each of the four types of $\mathscr{L}$ will now generate, via (26) and (32), a value for τ^2, a value for λ, and a solution to (13) given by

$$\psi = |x'|^{1/2} V \exp\left[-\frac{(p+1)}{4}\left(x' + \frac{1}{x'}\right) - ieQ\frac{\tau}{2|\tau|}\left(x' - \frac{1}{x'}\right)\right], \tag{33}$$

where V, given by one of (21), is a polynomial in x' and $1/x'$ [times $|x'|^{1/2}$ for types (ii) and (iv)]. If $Q=0$ is allowed in (32) and (33) they reduce to the results found in I for the extreme Reissner–Nordström geometry.

In the subcase $aeQ = mM$ we have

$$\tau^2 = \tfrac{1}{4}(p+1)^2, \tag{34a}$$

$$\xi = 4j(\tau/|\tau|)\sqrt{1+a^2/M^2}\sqrt{\tfrac{1}{4}(p+1)^2+e^2Q^2}, \tag{34b}$$

$$\lambda_h = 1 + 4\lambda + \tfrac{1}{4}(8+3a^2/M^2)(p+1)^2 + (4+a^2/M^2)e^2Q^2 - 2ia^2eQ\tau/M^2, \tag{34c}$$

so that $\operatorname{Im}\lambda = a^2eQ\tau/2M^2$. For given values of a/M and m, each type of $\mathscr{L}$ generates a value for τ^2, a value for λ, and a solution to (13) given by

$$\psi = |x'|^{1/2}V\exp\{-(\tau/2|\tau|)[x'(1+a^2/M^2)\times(\tau+ieQ)+(\tau-ieQ)/x']\}, \tag{35}$$

where again V is one of (21). Equations (34) and (35) reduce to the results of I where $e=m=0$.

In the case of (32) and (33) the angular function $S(\theta)$ is well behaved only when $\lambda = l(l+1)$ and l is a non-negative integer, but this can only occur if $\lambda_h \geqslant 1+2(p+1)^2+4e^2Q^2$. For types (ii)–(iv), it can be shown that for each n the eigenvalues λ_h are all bounded below $1+2(p+1)^2+4e^2Q^2$, so l is not real (though λ is) and $S(\theta)$ is singular. For type (i) it can be shown that at most one eigenvalue is greater than or equal to $1+2(p+1)^2+e^2Q^2$. These bounds on λ_h and consequent behavior of $S(\theta)$ are the same as found in I when $Q=0$.

When $a\neq 0$ the angular function $S(\theta)$ is well behaved when λ is an eigenvalue of the spheroidal wave equation. Such eigenvalues are functions of $a^2\omega^2$ and may be determined numerically when $a^2\omega^2$ is given. To confine our discussion to real eigenvalues (this is possible even though $a\omega$ is pure imaginary) we take $Q=m=0$ in the case of (34) and (35). There the specification of a/M determines the value of $a^2\omega^2$ via (34a) and (31). Thus we will have a well-behaved $S(\theta)$ if there are any choices of types of $\mathscr{L}$, values of a/M, and values of n such that λ determined by (34c) coincides with an eigenvalue. Like in the case of $a=0$, bounds on λ_h and hence on λ may be established from the form of $\mathscr{L}$, but since the eigenvalues are not known explicitly when $a\neq 0$ these bounds do not immediately preclude the real value of λ determined by (34c) from being an eigenvalue.

B. Spin-$\frac{1}{2}$ field on extreme Kerr–Newman geometry

The field equation in this case is (9) and it turns out that to obtain (13) we must set $\mu=0$ and $R=\Delta^{1/4}\psi$. Then we have

$$d_0 = T^2/M^2, \tag{36a}$$

$$d_1 = -T[4M\omega - 2eQ + i]/M, \tag{36b}$$

$$d_2 = [(6M^2+a^2)\omega^2 - eQ(6M\omega - eQ) + \tfrac{1}{4} - \lambda], \tag{36c}$$

$$d_3 = -M\omega(4M\omega - 2eQ - i), \tag{36d}$$

$$d_4 = M^2\omega^2. \tag{36e}$$

This case of the spin-$\frac{1}{2}$ field on the Kerr–Newman geometry is degenerate in that the conditions (16)–(18) and (25) when applied to (36) do not yield solutions at each integer n. In fact, they imply that we must have $n=0$ and V constant with the parameters satisfying $\lambda_h=0$, $eQ=2M\omega$ (so ω is real), and $\lambda = 2am\omega - a^2\omega^2$. The solution for ψ is then

$$\psi = \sqrt{|x'|}\exp(iTx'/M - iM\omega/x'). \tag{37}$$

C. Spin-1 and -2 fields on extreme Kerr geometry

The field equation in this case is (10) and it is transformed into (13) when x' is defined by (12) and when $R=\Delta^{-s/2}\psi$. Then we have

$$d_0 = (2M^2\omega - am)^2/M^2, \tag{38a}$$

$$d_1 = -2(2M^2\omega - am)(2M\omega - is)/M, \tag{38b}$$

$$d_2 = 7M^2\omega^2 - s^2 - s - \lambda, \tag{38c}$$

$$d_3 = -2M\omega(2M\omega + is), \tag{38d}$$

$$d_4 = M^2\omega^2. \tag{38e}$$

These quantities are similar in structure to (36), but because of the absence of the eQ term the requirement (16) implies $M\omega s=0$ so that the method yields no solutions in this case.

IV. METHOD FOR OBTAINING CLOSED-FORM SOLUTIONS ON NONEXTREME GEOMETRIES

In this section a method of solving

$$\psi_{xx} + A\psi = 0, \tag{39}$$

$$A = A_0 + A_1y + A_2y^2 + A_3y^3 + A_4y^4, \tag{40}$$

with the A_i constant and y being $\coth x$ or $\tanh x$, will be described and sufficient conditions for the applicability of the method will be proved. In Sec. V the method will be used to construct closed-form solutions to the field equations of Sec. II on nonextreme geometries.

For definiteness we choose

$$y = \coth x, \tag{41}$$

and simple modifications of what follows would handle the choice $y=\tanh x$.

Introduce constants α, β, and γ, to be determined, and a function V by letting

$$\psi = (\sinh x)^\beta V\exp(\alpha x + \gamma y) = \exp(\gamma y)V/(y+1)^{(\beta-\alpha)/2}(y-1)^{(\alpha+\beta)/2}. \tag{42}$$

It follows that V must satisfy

$$(y-1)^2(y+1)^2\frac{d^2V}{dy^2} - 2(y+1)(y-1)\times[\alpha+\gamma+(\beta-1)y-\gamma y^2]\frac{dV}{dy} + \widetilde{A}V = 0, \tag{43}$$

where

$$\widetilde{A} = A + (\alpha+\gamma)^2 + \beta + 2(\alpha\beta+\gamma\beta-\gamma)y + (\beta^2-\beta-2\alpha\gamma-2\gamma^2)y^2 + 2\gamma(1-\beta)y^3 + \gamma^2y^4. \tag{44}$$

Define coefficients $\widetilde{A}_i$ in $\widetilde{A}$ by

$$\widetilde{A} = \widetilde{A}_0 - \widetilde{A}_1(y+1) + \widetilde{A}_2(y+1)^2 - \widetilde{A}_3(y+1)^3 + \widetilde{A}_4(y+1)^4. \tag{45}$$

Then

$$\widetilde{A}_0=(\alpha-\beta)^2+A_0-A_1+A_2-A_3+A_4\,, \tag{46a}$$

$$\widetilde{A}_1=2\beta^2-2(\alpha+1)\beta-4\gamma(\alpha-\beta+1)$$
$$-A_1+2A_2-3A_3+4A_4\,, \tag{46b}$$

$$\widetilde{A}_2=4\gamma^2+2\gamma(3\beta-\alpha-3)+\beta^2-\beta+A_2-3A_3+6A_4\,, \tag{46c}$$

$$\widetilde{A}_3=4\gamma^2+2(\beta-1)\gamma-A_3+4A_4\,, \tag{46d}$$

$$\widetilde{A}_4=\gamma^2+A_4\,. \tag{46e}$$

It is also useful to define coefficients $\widetilde{A}_i^-$ by rearranging (45) into

$$\widetilde{A}^-=\widetilde{A}_0^- -\widetilde{A}_1^-(y-1)+\widetilde{A}_2^-(y-1)^2$$
$$-\widetilde{A}_3^-(y-1)^3+\widetilde{A}_4^-(y-1)^4\,; \tag{47}$$

then

$$\widetilde{A}_0^-=\widetilde{A}_0-2\widetilde{A}_1+4\widetilde{A}_2-8\widetilde{A}_3+16\widetilde{A}_4\,, \tag{48}$$

and similarly for the other $\widetilde{A}_i^-$.

We want to require that (43) be satisfied when V is a polynomial in y of degree n. The simplest recursion relation for coefficients in V results if the coefficients are those of powers of $(y\pm 1)$ rather than of y, hence we write

$$V=\sum_{k=0}^{n}(-1)^k a_k(y+1)^k\,,\quad a_n\neq 0\,, \tag{49}$$

where the a_k are constants, and substitute into (43). Because of the form of (42) we may assume without loss of generality that $a_0\neq 0$; the substitution then yields

$$\widetilde{A}_4=0\,, \tag{50a}$$

$$\widetilde{A}_0=0\,, \tag{50b}$$

$$\widetilde{A}_3=2n\gamma\,, \tag{50c}$$

$$4(k+1)(k+1+\alpha-\beta)a_{k+1}$$
$$+[\widetilde{A}_1+2k(k+\alpha-3\beta-4\gamma+1)]a_k$$
$$+[\widetilde{A}_2+(k-1)(k-2\beta-8\gamma)]a_{k-1}$$
$$+2\gamma(n+2-k)a_{k-2}=0\,,$$
$$k=0,1,2,\ldots,n\,, \tag{51}$$

$$[\widetilde{A}_2+n(n+1-2\beta-8\gamma)]a_n+2\gamma a_{n-1}=0\,, \tag{52}$$

where $a_k\equiv 0$ if $k<0$ or $k\geqslant n+1$.

Note that (49) may be rearranged into a polynomial in $(y-1)$ whose leading term a_0^- is given by $a_0^-=\Sigma_{k=0}^n(-1)^k2^{k+1}a_k$.

The following result allows for a nice formulation of the determination of the a_k and hence of V and ψ: if (50) holds and if $\widetilde{A}_0^-a_0^-=0$ then (52) is linearly dependent on (51).

Proof: Let the left-hand sides of (51) and (52) be denoted by E_k for $k=0,1,2,\ldots,(n+1)$. Then direct calculation gives

$$\sum_{k=0}^{n+1}(-1)^k2^kE_k$$
$$=\left(\frac{1}{2}\widetilde{A}_1-\widetilde{A}_2+4n\gamma\right)\sum_{k=0}^{n}(-1)^k2^{k+1}a_k\,. \tag{53}$$

Since by (48) and (50),

$$\tfrac12\widetilde{A}_1-\widetilde{A}_2+4n\gamma=\tfrac12\widetilde{A}_1-\widetilde{A}_2+2\widetilde{A}_3=-\tfrac14\widetilde{A}_0^-\,,$$

we see that (53) is

$$\sum_{k=0}^{n+1}(-1)^k2^kE_k=-\frac{1}{4}\widetilde{A}_0^-a_0^-\,. \tag{54}$$

Thus we see that (52) is linearly dependent on (51) if

$$\widetilde{A}_0^-a_0^-=0\,. \tag{55}$$

We therefore have the following method of insuring that (49) solves (43) and of determining the a_k. We require (50), and we make (52) hold by requiring (51), $a_0^-\neq 0$, and

$$\widetilde{A}_0^-=0\,. \tag{56}$$

We also require

$$\det\mathscr{M}=0\,, \tag{57}$$

where $\mathscr{M}$ is the $(n+1)$ by $(n+1)$ matrix of coefficients of the a_k in (51) with elements

$$\mathscr{M}_{ki}=4(k+1)(k+1+\alpha-\beta)\delta^i_{k+1}$$
$$+[\widetilde{A}_1+2k(k+\alpha-3\beta-4\gamma+1)]\delta^i_k$$
$$+[\widetilde{A}_2+(k-1)(k-2\beta-8\gamma)]\delta^i_{k-1}$$
$$+2\gamma(n+2-k)\delta^i_{k-2}\,, \tag{58}$$

where $0\leqslant i,k\leqslant n$, and δ^i_k is the Kronecker delta. Then (51) may be considered to be $n+1$ linear homogeneous equations $\mathscr{M}\vec{a}=0$ which are guaranteed by (57) to have nonzero solutions for the column vector $\vec{a}=(a_0,a_1,\ldots,a_n)^T$.

Note that because of (46b) and (46c) the condition (57) is explicit in terms of the parameters α, β, γ, and the constants A_i as are our conditions (50) and (56) which may be written

$$\widetilde{A}_4=\gamma^2+A_4=0\,, \tag{59a}$$

$$\widetilde{A}_0=(\alpha-\beta)^2+A_0-A_1+A_2-A_3+A_4=0\,, \tag{59b}$$

$$\widetilde{A}_0^-=(\alpha+\beta)^2+A_0+A_1+A_2+A_3+A_4=0\,, \tag{59c}$$

$$\widetilde{A}_3-2n\gamma=2(\beta-n-1)\gamma-A_3=0\,. \tag{59d}$$

The conditions which are being imposed on the constants A_i are now clear. We may consider (59a)–(59c) to determine α, β, and γ and then (59d) is a constraint on the A_i, and we have as well the constraint (57).

After all conditions (57) and (59) are satisfied and α, β, and γ determined, and the a_k are found from (51) (easily done, for example, by iteration starting with $a_0=1$), then a solution to (39) is given by the elementary function (42).

The apparent alternative for obtaining linear dependence of (52) on (51) by allowing $a_0^-=0$ in (55) can actually be disregarded, because the form of (42) implies that any solution corresponding to some α, β, and n values with $a_0^-=0$ is the same as one with $a_0^-\neq 0$ and some appropriate other values of α, β, and n.

Note that since our method requires $\widetilde{A}_0=\widetilde{A}_0^-=0$, it makes all three terms in (43) have a factor of $(y+1)(y-1)$ so that the order of the regular singular points at $y=1$ and $y=-1$ is reduced.

V. CLOSED-FORM SOLUTIONS ON NONEXTREME GEOMETRIES

When $M^2>a^2+e^2$, each one of the differential equations in Sec. II may be transformed to (39) by a coordinate change from r to x given by

$$x = \tfrac{1}{2}\ln|(r-r_+)/(r-r_-)|, \tag{60}$$

where $r_\pm \equiv M \pm M\epsilon$, ϵ is given by (7) with $0<\epsilon\leqslant 1$, and we have

$$r = M(1-\epsilon y), \tag{61}$$

with $y = \coth x$. [In some cases a factor transformation on the radial field is also involved and in the case of (9) we also require $\mu = 0$.] Thus the method of Sec. IV may be applied to obtain closed-form solutions to the equations of Sec. II.

The application will be carried out in this section, and will consist of the following. The $\widetilde{A}_i$ will be displayed in terms of the parameters $M, \epsilon, a, Q, \omega, m, \mu$, and λ occurring in the radial field equations, and (59) will be solved algebraically for α, β, γ, and ω in terms of $M, \epsilon, a, Q, m, \mu$, and n. It turns out that λ enters only in (57) and not in (59), so α, β, γ, and ω can be expressed independently of λ. The fact that we are solving for ω in addition to α, β, and γ reflects the constraint which (59) puts on the parameters. After solving (59), (57) will be imposed; then it happens that at each n, (57) is equivalent to an eigenvalue problem for λ with respect to an $(n+1)$ by $(n+1)$ matrix whose elements may be expressed in terms of $M, \epsilon, a, Q, \omega, m$, and μ. Further explicit algebraic analysis of allowed parameter values cannot proceed for arbitrary n because that would entail solving (57) for λ with n arbitrary. The analysis may be done easily, however, for small n. For example, for $n = 0$ the condition is that which results from setting $\widetilde{A}_2 = 0$ in (44c). With a computer one may easily go up to $n = 30$ or more.

A solution for the field in every case is obtained once the algebraic steps just described are completed and the a_k's are found from (51). The solution is given by

$$\psi = \frac{\exp(-\gamma r/\epsilon M)}{|r-r_+|^{(\beta-\alpha)/2}|r-r_-|^{(\alpha+\beta)/2}} \sum_{k=0}^{n} a_k \left(\frac{r-r_+}{M\epsilon}\right)^k. \tag{62}$$

A. Massive spin-0 field on nonextreme Kerr–Newman geometry

From (8), (44), (59), and (61) we have

$$\widetilde{A}_4 = \gamma^2 + M^2\epsilon^2\omega^2 - M^2\epsilon^2\mu^2 = 0, \tag{63a}$$

$$\widetilde{A}_0 = (\alpha-\beta)^2 + (1/M^2\epsilon^2)[M^2(1+\epsilon)^2\omega + a^2\omega - am - eQM(1+\epsilon)]^2 = 0, \tag{63b}$$

$$\widetilde{A}_0^- = (\alpha+\beta)^2 + (1/M^2\epsilon^2)[M^2(1-\epsilon)^2\omega + a^2\omega - am - eQM(1-\epsilon)]^2 = 0, \tag{63c}$$

$$\widetilde{A}_3 - 2n\gamma = 4\gamma^2 + 2(\beta-1-n)\gamma + 4\epsilon(\epsilon+1)M^2\omega^2 - 2eQM\epsilon\omega - 2(2\epsilon+1)M^2\epsilon\mu^2 = 0, \tag{63d}$$

and the quantities $\widetilde{A}_1$ and $\widetilde{A}_2$ which specify $\mathscr{M}$ are given by

$$\begin{aligned}\widetilde{A}_1 = (2/\epsilon)\{&2(\epsilon+1)^3M^2\omega^2 + (\epsilon+2)a^2\omega^2 + e^2Q^2(\epsilon+1)\\ &-\epsilon(\epsilon+1)^2M^2\mu^2 - eQ[3(\epsilon+1)^2M\omega\\ &-am/M + a^2\omega/M]\\ &-2am\omega - \epsilon\lambda + 2(\beta-\alpha-1)\epsilon\gamma + \beta(\beta-\alpha-1)\epsilon\},\end{aligned} \tag{64}$$

$$\begin{aligned}\widetilde{A}_2 = 6(\epsilon+1)^2M^2\omega^2 &+ a^2\omega^2 - (5\epsilon+1)(\epsilon+1)M^2\mu^2\\ &+ e^2Q^2 - 6eQ(\epsilon+1)M\omega - \lambda\\ &+ 2\gamma(2\gamma+3\beta-\alpha-3) + \beta(\beta-1).\end{aligned} \tag{65}$$

Because of (64) and (65), $\mathscr{M}$ as given by (58) has the form $\mathscr{M} = \mathscr{M}' - \lambda\mathscr{N}$, where $\mathscr{M}'$ is independent of λ and $\mathscr{N}_{ij} = \delta_i^j + \delta_{i+1}^j$. Thus (51) becomes the eigenvalue problem $\mathscr{N}^{-1}\mathscr{M}'\vec{a} = \lambda\vec{a}$ and (57) becomes

$$\det(\mathscr{N}^{-1}\mathscr{M}' - \lambda I) = 0. \tag{66}$$

The conditions (59) are now expressed as (63) and these we solve for $(\alpha,\beta,\gamma,\omega)$. Equations (63b) and (63c) have four solutions for (α,β) given by

$$(\alpha,\beta) = [(\alpha_0,\beta_0),(\beta_0,\alpha_0),-(\alpha_0,\beta_0),-(\beta_0,\alpha_0)], \tag{67}$$

where α_0 and β_0 are defined by

$$\alpha_0 = i(2M\omega - eQ), \tag{68}$$

$$\beta_0 = -i[M^2(1+\epsilon^2)\omega + a^2\omega - am - eQM]/\epsilon M. \tag{69}$$

In the case $(\alpha,\beta) = (\alpha_0,\beta_0)$ the desired formulas for $(\alpha,\beta,\gamma,\omega)$ in terms of M, ϵ, a, Q, m, and μ are obtained as follows. Define $\widetilde{\omega}$ to be a root of the quartic

$$(\widetilde{\omega}^2 + \widetilde{\mu}^2)[(\epsilon^2+1-\widetilde{a}^2)\widetilde{\omega} - (\widetilde{a}\widetilde{m} + \epsilon + E)]^2 - \epsilon^2[\widetilde{\omega}(2\widetilde{\omega} - E) + \widetilde{\mu}^2]^2 = 0, \tag{70}$$

where $\widetilde{\mu} = M\mu/(n+1)$, $\widetilde{a} = a/iM$, $\widetilde{m} = m/(n+1)$, and $E = eQ/i(n+1)$. Then ω is given by

$$\omega = i(n+1)\widetilde{\omega}/M, \tag{71}$$

substitution into (68) and (69), respectively, produces α and β, and γ is given by

$$\gamma = \epsilon\{[\mu^2M^2 - M\omega(2M\omega - eQ)]/[\beta-(n+1)]\}, \tag{72}$$

where (63) requires $\beta \neq n+1$.

In the case $(\alpha,\beta) = (\beta_0,\alpha_0)$, ω is obtained from (71) and

$$4\widetilde{\omega}^3 + (1-2E)\widetilde{\omega}^2 + 2\widetilde{\mu}^2(2-E)\widetilde{\omega} - \widetilde{\mu}^2[\widetilde{\mu}^2 - (1-E)^2] = 0, \tag{73}$$

instead of (70), and γ is given by (72), where α and β are given by (69) and (68), respectively.

The results for the last two choices of (α,β) in (67) may be obtained by the formal replacements $(\omega,a,eQ) \to -(\omega,a,eQ)$ in the results just discussed for the first two choices. Thus the dependence of $(\alpha,\beta,\gamma,\omega)$ on the other parameters has been established explicitly in all cases for arbitrary n since the solutions to (70) and (73) may be written in terms of radicals.

We now give some illustrative results in special cases. When $\mu = 0$ possible values of $(\alpha,\beta,\gamma,\omega)$ are

$$\omega = -[i(n+1) + 2eQ]/4M = i\gamma/\epsilon M, \tag{74a}$$

$$\beta = (n+1)/2, \tag{74b}$$

$$\alpha = -\frac{[(\epsilon^2+1)M^2 + a^2](n+1) - 2i[2aMm + e^3Q]}{4\epsilon M^2}, \tag{74c}$$

and

$$\omega = \frac{i\epsilon M(n+1) + am + eQM(1\pm\epsilon)}{(\epsilon\pm 1)^2M^2 + a^2} = \frac{\mp i\gamma}{\epsilon M}, \tag{75a}$$

$$\beta = \frac{[(\epsilon^2+1)M^2+a^2](n+1) \pm i[2aMm+e^3Q]}{(\epsilon \pm 1)^2M^2+a^2}, \tag{75b}$$

$$\alpha = -\frac{2\epsilon M^2(n+1) - i[2aMm+e^3Q]}{(\epsilon \pm 1)^2M^2+a^2}, \tag{75c}$$

plus those values obtained upon replacing i by $-i$ everywhere in (74) and (75). Equations (74) and (75) provide generalizations of the Reissner–Nordström case with $Q=0$ given by Eqs. (46) and (47) in I to the present case of Kerr–Newman geometry with an electromagnetic test field possibly present.

Next consider $n=0$, $Q=0$. In that case there is a solution with real λ for which all quantities may be expressed in terms of a real quantity σ, constrained only by $\sigma \geqslant 1$, and we have

$$\gamma = -\epsilon\sigma^2/4, \quad \alpha = \sigma(\sigma-2)[M^2(1+\epsilon^2)+a^2]/4\epsilon M^2,$$
$$\beta = -\tfrac{1}{2}\sigma(\sigma-2), \quad \omega = i\sigma(\sigma-2)/4M,$$
$$\mu^2 = \sigma^2(\sigma-1)/4M^2, \quad \lambda = \sigma[a^2\sigma^3+4e^2\sigma-16M^2]/16M^2,$$
$$m=0.$$

Note that ω is pure imaginary and all other quantities are real in this case.

A set of parameter values with both ω and λ real can be found by requiring $n=0$, $\operatorname{Im}\omega=0$, $eQ \neq 0$. In that case we must have $0<\sigma<(\epsilon+1)^2/2\epsilon$, and then

$$e^2Q^2 = 8\epsilon\sigma[(\epsilon+1)^2-2\sigma\epsilon]/(\epsilon+1)^4, \quad \omega = eQ/2M,$$
$$\gamma = -2\epsilon^2\sigma/(\epsilon+1)^2, \quad \mu^2 = 2\epsilon\sigma/M^2(\epsilon+1)^2,$$
$$\alpha=0, \quad \beta=0,$$
$$\lambda = 2\epsilon\sigma\{-2\epsilon\sigma[2M^2(\epsilon^2-1)+a^2]+M^2(\epsilon+1)^2(\epsilon^2-3)$$
$$+a^2(\epsilon+1)^2\}/M^2(\epsilon+1)^4,$$
$$2Mma = -e^3Q.$$

The last equation is a constraint on M, e, a, and σ since we require m to be an integer.

A final simple case occurs when $\mu^2=\omega^2$. Then for all n we must have $\gamma=0$, $eQ=M\omega$, $\beta_0 = i[am-M^2\epsilon^2\omega-a^2\omega]/\epsilon M$, $\alpha_0 = iM\omega$. Checking λ when $n=0$ we find that λ is not real unless we choose $M^2\omega+a^2\omega-am=0$ and the first or third case of (67); then $\lambda = M^2\epsilon^2\omega^2+a^2\omega^2$ and $m\neq 0$.

B. Spin-$\frac{1}{2}$ field on nonextreme Kerr–Newman background geometry

The field equation in this case is (9) and we want to put it into the form of (39), but eliminating all first-derivative terms in (9) by a pure coordinate transformation to a complex x does not accomplish this. Eliminating all first-derivative terms by a coordinate transformation to x as given by (60) coupled with the appropriate factor transformation on R results in (39) only if $\mu=0$. Hence we restrict ourselves to $\mu=0$, and then we have (39) holding for $\psi = R/\Delta^{1/4}$, x given by (60), and A given by

$$M^2\epsilon^2A = K^2 + i(r-M)K - 2i\omega r\Delta + ieQ\Delta + \tfrac{1}{2}\Delta$$
$$-\tfrac{1}{4}(r-M)^2 - \lambda\Delta - a^2\omega^2\Delta + 2am\omega\Delta. \tag{76}$$

From (61) and (76) we obtain the form which (59), (46b), and (46c), take in this case

$$\widetilde{A}_4 = M^2\epsilon^2\omega^2+\gamma^2=0, \tag{77a}$$

$$\widetilde{A}_0 = (\alpha-\beta)^2 + (1/\epsilon^2M^2)\{[M^2(\epsilon+1)^2+a^2]\omega - am$$
$$-eQ(1+\epsilon)M + (i/2)M\epsilon\}^2 = 0, \tag{77b}$$

$$\widetilde{A}_0^- = (\alpha+\beta)^2 + (1/\epsilon^2M^2)\{[M^2(1-\epsilon)^2+a^2]\omega - am$$
$$-eQ(1-\epsilon)M - (i/2)M\epsilon\}^2 = 0, \tag{77c}$$

$$\widetilde{A}_3 = 4\epsilon(\epsilon+1)M^2\omega^2 - 2eQ\epsilon M\omega + 4\gamma^2$$
$$+2(\beta-1)\gamma - i\epsilon M\omega = 2n\gamma, \tag{77d}$$

$$\widetilde{A}_1 = (2/\epsilon)\{2(\epsilon+1)^3M^2\omega^2 + (\epsilon+2)a^2\omega^2 + (\epsilon+1)e^2Q^2$$
$$-eQ[3(\epsilon+1)^2M\omega - am/M + a^2\omega/M] - 2am\omega$$
$$-\epsilon\lambda + 2(\beta-\alpha-1)\epsilon\gamma + \beta(\beta-\alpha-1)\epsilon\}$$
$$-(i/\epsilon M)[(\epsilon^2-1)M^2\omega - a^2\omega + am + M\epsilon] + \tfrac{1}{2}, \tag{78}$$

$$\widetilde{A}_2 = 6(\epsilon+1)^2M^2\omega^2 + a^2\omega^2 + e^2Q^2 - 6eQ(\epsilon+1)M\omega - \lambda$$
$$+2\gamma(2\gamma+3\beta-\alpha-3) + \beta(\beta-1) - 3i\epsilon M\omega + \tfrac{1}{4}. \tag{79}$$

The condition which determines λ is (66) with $\widetilde{A}_1$ and $\widetilde{A}_2$ given by (78) and (79).

From (77) the quantities $(\alpha,\beta,\gamma,\omega)$ may be found in terms of n and the other parameters in the same manner as they were obtained in Sec. V A for the spin-zero case.

Defining $(\alpha_{1/2}, \beta_{1/2})$ by

$$\alpha_{1/2} = i_1[2M\omega - eQ + i/2], \tag{80}$$

$$\beta_{1/2} = -i_1[M^2(1+\epsilon^2)\omega + a^2\omega - am - eQM]/\epsilon M, \tag{81}$$

the solutions for $(\alpha,\beta,\gamma,\omega)$ are then the following three sets given by (82)–(84):

$$M\omega = [2i_2(n+1) - (2i_1i_2/\epsilon)(am/M+eQ) + 2eQ + i]$$
$$\times[4 - 2i_1i_2((\epsilon^2+1)/\epsilon + a^2/\epsilon M^2)]^{-1}, \tag{82a}$$

$$\gamma = i_2\epsilon M\omega, \tag{82b}$$

$$\beta = \beta_{1/2}, \tag{82c}$$

$$\alpha = \alpha_{1/2}, \tag{82d}$$

where there are four solutions in (82) corresponding to $i_1 = \pm i$ and, independently, $i_2 = \pm i$:

$$M\omega = -i_1(n+1)/4 + eQ/2, \tag{83a}$$

$$\gamma = -(n+1)\epsilon/4 - i_1\epsilon eQ/2, \tag{83b}$$

$$\beta = (n+1)/2 + ii_1/2, \tag{83c}$$

$$\alpha = -\frac{(n+1)}{4\epsilon}\left(\epsilon^2+1+\frac{a^2}{M^2}\right) + \frac{i_1e^3Q}{2\epsilon M^2} + \frac{i_1am}{\epsilon M}; \tag{83d}$$

$$\gamma = -i\epsilon M\omega, \tag{84a}$$

$$\beta = \alpha_{1/2}, \quad i_1 = -i, \tag{84b}$$

$$\alpha = \beta_{1/2}, \quad i_1 = -i, \tag{84c}$$

$$n = 0, \tag{84d}$$

where ω is arbitrary in (84).

C. Spin-1 and -2 fields on nonextreme Kerr geometry

The field equation in this case is (10) and it is transformed into (39) by introducing x again by (60) and letting $R = \Delta^{-s/2}\psi$. The result for A is then

$$A = (1/M^2\epsilon^2)[B - s\Delta - s^2(r-M)^2], \tag{85}$$

where B is given by (11).

The form which (59), (46b), and (46c) take in this case is the following:

$$\tilde{A}_0 = (\alpha-\beta)^2 + \frac{1}{M^2\epsilon^2}\{[(\epsilon+1)^2M^2 + a^2]\omega - am - is\epsilon M\}^2 = 0, \tag{86a}$$

$$\tilde{A}_0^- = (\alpha+\beta)^2 + \frac{1}{M^2\epsilon^2}\{[(1-\epsilon)^2M^2 + a^2]\omega - am + is\epsilon M\}^2 = 0, \tag{86b}$$

$$\tilde{A}_4 = \epsilon^2M^2\omega^2 + \gamma^2 = 0, \tag{86c}$$

$$\tilde{A}_3 = 4\epsilon(\epsilon+1)M^2\omega^2 + 2is\epsilon M\omega + 4\gamma^2 + 2(\beta-1)\gamma = 2n\gamma, \tag{86d}$$

$$\tilde{A}_2 = 6(\epsilon+1)^2M^2\omega^2 + a^2\omega^2 + 6is\epsilon M\omega - s^2 - s - \lambda + 2\gamma(2\gamma + 3\beta - 3 - \alpha) + \beta(\beta-1), \tag{87}$$

$$\tilde{A}_1 = (1/\epsilon)\{4(\epsilon+1)^3M^2\omega^2 + 2(\epsilon+2)a^2\omega^2 + is[4(\epsilon^2-1)M\omega + 2am/M] - 4am\omega - 2\epsilon s^2 - 2\epsilon s - 2\epsilon\lambda\} + 4(\beta-\alpha-1)\gamma + 2\beta(\beta-\alpha-1). \tag{88}$$

The condition for λ is again (66) with this $\tilde{A}_1$ and $\tilde{A}_2$.

Algebraic expressions for $(\alpha,\beta,\gamma,\omega)$ may be found also in this case of fields of spin 1 and 2. The solutions to (86) for $(\alpha,\beta,\gamma,\omega)$ are given by

$$M\omega = \frac{2is\epsilon + is\epsilon i_3(i_1+i_2) - 2i_3\epsilon(n+1) + i_3(i_1-i_2)am/M}{4\epsilon - 2\epsilon i_3(i_1+i_2) - 2i_3(i_1-i_2)}, \tag{89}$$

$$\gamma = i_3\epsilon M\omega, \tag{90a}$$

$$2M\epsilon\alpha = (i_1+i_2)(2M^2\omega - am) + (i_1-i_2)(2M^2\epsilon\omega - is\epsilon M), \tag{90b}$$

$$2M\epsilon\beta = (i_1+i_2)(-2\epsilon M^2\omega + is\epsilon M) + (i_1-i_2)(-2M^2\omega + am), \tag{90c}$$

where $s = \pm 1, \pm 2$, and i_1, i_2, and i_3 may independently take the values $\pm i$ with the following exceptions. For $(i_1,i_2,i_3) = (-,-,+)i$, (89) is to be disregarded and the solution is given by (90) with ω arbitrary but with the constraint $2s = n+1$ so that the only possible values of (s,n) in this special case are (1,1) and (2,3). For $a=0$ and $(i_1,i_2,i_3) = \pm(+,-,-)i$ the solutions are given by (90) with $\epsilon = 1$, ω arbitrary, and $(s,n) = [(\mp 1,0),(\mp 2,1)]$.

VI. BEHAVIOR OF CLOSED-FORM SOLUTIONS

Because the number of solutions found in this paper is so large, we do not give here an exhaustive description of their behavior. We do, however, summarize some of the main results, and for particularly interesting example solutions we point out some which have well-behaved angular dependence and some which do not, analyze field behavior as the conformal boundaries $\mathscr{H}^\pm$ and $\mathscr{I}^\pm$ of the space-time are approached, and compare our radial fields to normal modes.

Two interesting sorts of solutions which we might wish to find in closed form (but seemingly do not) would be (a) an exact physically reasonable radiation field or (b) a normal mode radial solution (one representing only outgoing radiation as $r\to\infty$ and satisfying the proper boundary conditions[13] as $r\to r_+$; the mode would be stable if Im $w<0$ and unstable if Im $w>0$). Since spherically symmetric geometries ($a=0$) have been shown to be stable to perturbation by massless fields no unstable normal modes can be found in those cases. For the Kerr geometry, stability with respect to massless fields probably holds although it is not yet rigorously proven,[14] so we might not expect to find any unstable normal modes; however, some authors do think that they exist.[15] Other authors[16] believe unstable normal modes exist for a *massive* perturbing scalar field of the Kerr geometry. None of the examples which we analyze prove to be of sorts (a) or (b), though some could be normal modes for large n or values of a at which it is difficult to discover whether λ is an angular eigenvalue. For those parameter values for which the analysis is easily completed we find some exact fields which have the proper normal mode radial behavior but fail to have λ be an eigenvalue at least for small $|a\omega|$, and we find the reverse—angularly well-behaved fields which fail to have normal mode behavior either as $r\to\infty$ or as $r\to r_+$. The care with which our solutions avoid being normal modes leads one to suspect that the conditions which have been imposed to obtain closed forms may put one outside the class of normal mode solutions.

Other types of fields which we find are ones which, though nonstatic, behave just as some static massless fields are known to do in the respect that they blow up on $\mathscr{H}^+$ or $\mathscr{H}^-$ if they are well behaved on $\mathscr{I}^+$ and $\mathscr{I}^-$ and blow up on $\mathscr{I}^+$ or $\mathscr{I}^-$ if finite on $\mathscr{H}^+$ and $\mathscr{H}^-$. These types of fields occur frequently when w has a nonzero imaginary part.

Another closed-form field which we have found has the properties that it goes to zero like $1/r$, is nonradiative at $\mathscr{I}^-$, is finite on $\mathscr{H}^-$, has outgoing radiation at $\mathscr{I}^+$ (with a profile exponentially increasing towards the future on $\mathscr{I}^+$), and is infinite on $\mathscr{H}^+$ but is infinite nowhere inside the boundaries $\mathscr{H}^\pm$ and $\mathscr{I}^\pm$. Such a field is not well behaved but has finite null data on $\mathscr{I}^-$ and finite but unbounded null data on $\mathscr{H}^-$.

To discuss behavior at $\mathscr{H}^\pm$ and $\mathscr{I}^\pm$ we define null coordinates u and v by $2u = t - r^*$, $2v = t + r^*$, where

$$r^* = r + \frac{r_+^2 + a^2}{2\epsilon M}\ln|r - r_+| - \frac{r_-^2 + a^2}{2\epsilon M}\ln|r - r_-|, \tag{91}$$

on nonextreme geometries, and

$$r^* = r - (M^2+a^2)/(r-M) + 2M\ln|r-M|, \tag{92}$$

on extreme geometries ($r_+ = r_- = M$). Then at fixed u we approach $\mathscr{I}^+$ via $v\to\infty$ (or $r\to\infty$) and $\mathscr{H}^-$ via $v\to-\infty$ (or $r\to r_+$), and at fixed v we approach $\mathscr{I}^-$ via $u\to-\infty$ (or $r\to\infty$) and $\mathscr{H}^+$ via $u\to+\infty$ (or $r\to r_+$).

On extreme geometries the behavior of

$$\psi \exp(-i\omega t) = \psi \exp(-2i\omega u - i\omega r^*)$$
$$= \psi \exp(-2i\omega v + i\omega r^*), \qquad (93)$$

at the conformal boundaries is easily obtained from (27) using (12) and (92). Similarly, on nonextreme geometries we use (60)–(62) and (91), from which we obtain

$$\psi \to \exp(-\gamma r^*/\epsilon M)/r^{\beta - n - 2\gamma/\epsilon}, \quad \text{as } r \to \infty, \qquad (94)$$

and

$$\psi \to \text{const} \exp(\epsilon M(\alpha - \beta) r^*/(r_+^2 + a^2)), \quad \text{as } r \to r_+. \qquad (95)$$

Our notation for behaviors of $\psi \exp(-i\omega t)$ at the conformal boundaries will be to list them in square brackets in the order $\mathscr{H}^-, \mathscr{H}^+, \mathscr{I}^-, \mathscr{I}^+$.

A. Solutions on extreme geometries

The variety of closed-form solutions generated on extreme geometries by the method of Sec. III is considerably less than that obtained on nonextreme geometries in Sec. V. For spins 1 and 2 on the extreme Kerr geometry the method employed gives no solutions, and for the spin-$\frac{1}{2}$ field on the extreme Kerr–Newman geometry it yields a solution, given by (37), only if the field is massless ($\mu = 0$), $n = 0$, and a nonvanishing electromagnetic test field is present. The spin-$\frac{1}{2}$ field has real ω, but its angular part $S(\theta)$ cannot be nonsingular for all $a\omega$ and l because we have $\lambda = 2am\omega - a^2\omega^2$ and this expression cannot be an eigenvalue since it goes to zero instead of $l(l+\frac{1}{2})^2$ as $a\omega$ goes to zero.

The only solution on an extreme geometry for a massive field ($\mu \neq 0$) is the spin-0 field, specialized by $\mu^2 = \omega^2$ and other conditions, on the extreme Kerr–Newman geometry. The radial field in this case is given by (30), ω is real, and the angular factor $S(\theta)$ is nonsingular. Though $\omega \neq 0$, this field behaves in the same fashion as the static field with $m = \mu = Q = 0$; namely, it goes as $[\infty, \infty, 0, 0]$ or $[0, 0, \infty, \infty]$ at the conformal boundaries. In fact, for $a = 0$, (30) gives exactly the static formulas for ψ, so we have

$$\psi \exp(-i\omega t)$$
$$= (1/(r-M)^{l+1}, (r-M)^l) \exp(-i\omega(u+v)). \qquad (96)$$

In particular, the first case of (96) has, for $l = 0$, a nontrivial radiation field at $\mathscr{I}^-$ and $\mathscr{I}^+$, and

$$\psi \exp(-i\omega t) \to \left[\infty, \infty, \exp(-i\omega v)\frac{\exp(-i\omega u)}{r}, \right.$$
$$\left. \exp(-i\omega u)\frac{\exp(-i\omega v)}{r}\right]. \qquad (97)$$

When the spin-0 field is massless, ω has a nonvanishing imaginary part and if $am \neq 0$ then $Q = 0$ so that the presence of an electromagnetic test field is required unless we restrict to extreme Reissner–Nordström or to only $m = 0$ modes. When $a = 0$ the radial massless scalar field is (33) and $S(\theta)$ is singular except possibly for only one eigenvalue at each n out of the $4n + 3$ eigenvalues of the four types of $\mathscr{L}$ in (24). When $a \neq 0$, the radial field is (35), we must have $aeQ = mM$, and a numerical investigation is needed to show whether $S(\theta)$ may be nonsingular.

B. Solutions on nonextreme geometries

On the appropriate nonextreme geometries there are closed-form solutions given in Sec. V for fields of any spins 0, $\frac{1}{2}$, 1, 2. Restrictions on the parameters are not as severe as with the extreme geometries; in particular, the presence of an electromagnetic test field is not required except in highly special cases. We are still restricted, however, to massless fields when the spin is nonzero. There are several classes of solutions with real ω, several with complex ω, and several with nonsingular $S(\theta)$. The explicit formula for the radial field is (62).

A massless scalar field on the Kerr–Newman geometry which results from (75) (with the lower sign chosen) provides an example of an angularly nonsingular field which behaves well at $\mathscr{H}^-$ and $\mathscr{I}^-$ but badly at $\mathscr{H}^+$ and has an exponential radiation profile on $\mathscr{I}^+$. This is easily seen explicitly by setting $n = Q = 0$ in (75); then from (62) the solution to (5) is

$$\psi \exp(-i\omega t) = \frac{|r - r_+|^{-iam/2M\epsilon}}{|r - r_-|^{1 - iam/2M\epsilon}} \exp(-2i\omega u), \qquad (98)$$

where $\omega = (am + i\epsilon M)/(r_-^2 + a^2)$, and the asymptotic behavior when $m = 0$ is

$$\psi \exp(-i\omega t)$$
$$\to [\text{const} \exp(-2i\omega u), \infty, 0/r, \exp(-2i\omega u)/r]. \qquad (99)$$

It can be shown [by examining the expression for λ which results from setting $\widetilde{A}_2 = 0$ in (65)] that for $m = 0$ and sufficiently small $|a\omega|$ there exists a choice of $\epsilon, 0 < \epsilon < 1$, such that λ is an eigenvalue in the angular differential equation and $S(\theta)$ is therefore nonsingular. This radial field has pure imaginary ω and proper behavior as $r \to \infty$, but it is not a normal mode as is clear from (98). If $am \neq 0$ in (75) then ω has a nonvanishing real part, as expected for a normal mode, but $S(\theta)$ may be singular.

When the upper sign is chosen in (75) we can show for $n = Q = m = 0$ that $\lambda < 0$ and so λ is not an eigenvalue as $a\omega \to 0$.

The scalar field characterized in Sec. V A by $n = Q = m = 0$, $\omega = i\sigma(\sigma - 2)/4M$, $\mu^2 = \sigma^2(\sigma - 1)/4M^2$, $\sigma \geqslant 1$ behaves as $\psi \to r^{-\sigma} \exp(i\sqrt{\omega^2 - \mu^2}\, r^*)$ for $r \to \infty$ and as $\psi \to \exp(-i\omega r^*)$ for $r \to r_+$. For $\sigma = 1$ the field is massless and behaves radially as a stable normal mode, but $\lambda < 0$ so the field is angularly singular at least as $a\omega \to 0$. The field is massive if $\sigma > 1$ and for $a\omega \to 0$ (or $a = 0$) it can be shown that for every e, $0 \leqslant e^2 < M^2$, there exists a value of σ, $\sigma > 2(1 + \sqrt{1 + l(l+1)}) + o(a) > 4$, such that λ is an eigenvalue with the exception that $a = e = 0$ is excluded. The field then behaves radially like an unstable normal mode except that the r falloff is at the faster rate $1/r^\sigma$, $\sigma > 4$, instead of $1/r$ as $r \to \infty$.

The scalar field discussed last in Sec. V A having $\mu^2 = \omega^2$, $n = 0$, and $eQm \neq 0$ is angularly nonsingular when we restrict ω by $\omega^2 = l(l+1)/(M^2 - e^2)$.

As an example of nonzero spin field consider the spin-$\frac{1}{2}$ field determined by (83) with $i_1 = i$ and $Q = 0$. Then from (62) and (91) we find

$$\psi = \exp(i\omega r^*)(|r - r_+|^{g_+}/|r - r_-|^{g_-})\, V, \qquad (100)$$

where $\omega = -i(n+1)/4M$ and $g_{\pm} \equiv -i\omega(r_{\pm}^2 + a^2)/\epsilon M + iam/2\epsilon M \pm \frac{1}{4}$. Recalling that $\psi = R\Delta^{1/4}$ we find $R \rightarrow \exp(i\omega r^*)$ as $r \rightarrow \infty$ and $R \rightarrow \mathrm{const}\,\Delta^{1/2} \exp(-i\kappa r^*)$ as $r \rightarrow r_+$, where $\kappa \equiv \omega - am/(2Mr_+ - e^2)$. For all n, this is the proper r dependence of a stable normal mode; but for $n = 0$ we find $\lambda = -(1+\epsilon^2)/4 + 3a^2\omega^2 + iam/2M$, which cannot be an eigenvalue at least as $a\omega \rightarrow 0$.

ACKNOWLEDGMENTS

The author wishes to thank C. A. Heisler for checking some of the equations and R. J. Torrence for useful discussions.

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada.

[1]W. E. Couch, J. Math. Phys. **22**, 1457 (1981).
[2]E. T. Newman, E. Couch, K. Chinnapared, A. Exton, A. Prakash, and R. Torrence, J. Math. Phys. **6**, 918 (1965).
[3]S. K. Bose, J. Math. Phys. **16**, 772 (1975).
[4]D. N. Page, Phys. Rev. D **14**, 1509 (1976).
[5]S. A. Teukolsky, Astrophys. J. **185**, 635 (1973).
[6]C. H. Lee, J. Math. Phys. **17**, 1226 (1976).
[7]D. M. Chitre, Phys. Rev. D **13**, 2713 (1976).
[8]S. A. Teukolsky, Phys. Rev. Lett. **29**, 1114 (1972).
[9]S. Chandrasekhar, Proc. R. Soc. London Ser. A **365**, 453 (1979).
[10]V. Moncrief, Phys. Rev. D **9**, 2707 (1974); **10**, 1057 (1974); **12**, 1526 (1975).
[11]G. Stephenson, J. Phys. A **10**, L7 (1977).
[12]I. M. Arscott, *Periodic Differential Equations* (MacMillan, New York, 1964).
[13]W. H. Press and S. A. Teukolsky, Astrophys. J. **185**, 649 (1973).
[14]J. M. Stewart, Proc. R. Soc. London Ser. A **344**, 65 (1975).
[15]S. Detweiler and R. Ore, Phys. Rev. Lett. **51**, 67 (1983).
[16]T. J. M. Zouros and D. M. Eardley, Ann. Phys. **118**, 139 (1979).

On hypersurface-homogeneous space-times

J. Hajj-Boutros
Physics Department, Lebanese University, Mansourieh El-Metn, P.O. Box 72, Lebanon

(Received 26 June 1984; accepted for publication 8 February 1985)

We propose a new method to build exact solutions of Einstein field equations in case of "hypersurface-homogeneous space-times." The energy-momentum tensor is of perfect fluid type. Starting from SE solutions we are able to build new classes of solutions which add to the rare solutions not satisfying the equation of state $p = (\gamma - 1)\mu$. We study the geometrical and physical properties of some of the solutions obtained.

I. INTRODUCTION

In a previous article[1] we proved the utility of a new technique to generate exact solutions of Einstein's field equations with spherical symmetry and an energy-momentum tensor of perfect fluid type.

Using the same idea we present a method to generate exact solutions (always for an energy-momentum tensor of perfect fluid type) for what is known as a homogeneous-hypersurface space time. These are spaces having a group of motions G_4 on V_3, the isotropy group being spatial rotations. The metric used is

$$ds^2 = -dt^2 + A^2(t)dx^2 + B^2(t)\left(dy^2 + \sum^2 (y,K)dz^2\right). \tag{1.1}$$

Starting from a metric of form (1.1), Stewart and Ellis[2] (SE) already obtained general solutions for the Einstein field equations satisfying the equation of state

$$p = (\gamma - 1)\mu, \tag{1.2}$$

where p is the pressure, μ is the energy density, and γ is a parameter such that

$$1 < \gamma < 2.$$

Starting from the SE (Stewart–Ellis) metric, and using our technique, we obtain new classes of solutions.

It is interesting to notice that the classes of metrics we obtain do not satisfy (1.2).

As far as we know, these classes are new and can be added to the rare perfect fluid solutions[3–6] not satisfying the relation (1.2).

II. FIELD EQUATIONS

The field equations in general relativity are

$$R_{ab} - \tfrac{1}{2}Rg_{ab} + \Lambda g_{ab} = \kappa_0 T_{ab}. \tag{2.1}$$

The energy momentum tensor T_{ab} for a perfect fluid is

$$T_{ab} = (\mu + p)u_a u_b + pg_{ab}, \quad u_a u^a = -1. \tag{2.2}$$

We consider metrics admitting a group of motions G_4 on V_3 which are locally rotationally symmetric (LRS).

Among the metrics mentioned above, we choose one which has been used often as a cosmological model[7] and whose importance is clear.

This metric is

$$ds^2 = -dt^2 + A^2(t)dx^2 + B^2(t)\left(dy^2 + \sum^2 (u,K)dz^2\right), \tag{2.3}$$

where $\Sigma(y,K) = \sin y,\ y,\ \sinh y$, respectively, when $K = 1,0,-1$. For a Ricci tensor of type [(111),1], in Segre notation, the field equations are[7]

$$2\frac{B''}{B} + \frac{B'^2}{B^2} + \frac{K}{B^2} = \Lambda - \kappa_0 p, \tag{2.4}$$

$$\frac{B''}{B} + \frac{A''}{A} + \frac{A'}{A} \times \frac{B'}{B} = \Lambda - \kappa_0 p, \tag{2.5}$$

$$2\frac{A'}{A} \times \frac{B'}{B} + \frac{B'^2}{B^2} + \frac{K}{B^2} = \Lambda + \kappa_0 \mu. \tag{2.6}$$

We have to consider two cases.

A. $K = 0$

Equations (2.4) and (2.5) give

$$\frac{B''}{B} + \frac{B'^2}{B^2} = \frac{A''}{A} + \frac{A'}{A} \times \frac{B'}{B}. \tag{2.7}$$

We associate to the couple (A,B) another one (Y,Z) given by

$$A'/A = Y, \quad B'/B = Z. \tag{2.8}$$

Using (2.8), Eq. (2.7) becomes

$$Y' + Y^2 + YZ = Z' + 2Z^2. \tag{2.9}$$

1. (2.9) is a Riccati equation in Y

By the change of function

$$Y = Y_0 + 1/U, \tag{2.10}$$

we obtain

$$U' - (2Y_0 + Z)U = 1, \tag{2.11}$$

where Y_0 is a particular solution of (2.9) with Y being the more general one. By quadrature (2.11) gives

$$U(t) = \left\{\exp \int (2Y_0 + Z)dt\right\} \times \left\{\int \left(\exp \int -(2Y_0 + Z)dt\right)dt + C\right\}, \tag{2.12}$$

C being a constant. From (2.8) and (2.12) we obtain

$$U(t) = A_0^2 B\left(\int \frac{dt}{A_0^2 B} + C_1\right). \tag{2.13}$$

Equations (2.10) and (2.13) yield

$$A(t) = A_0(t)\exp\left\{\int \frac{dt}{A_0^2 B\,[\int dt/A_0^2 B + C_1]} + C_2\right\}, \tag{2.14}$$

C_2 being a constant. Then from the couple $[A_0(t), B(t)]$ our method allows us to obtain $[A(t), B(t)]$, where $A(t)$ is given by (2.14) and $B(t)$ stays invariable.

2. (2.9) is a Riccati equation in Z

By the change of function

$$Z = Z_0 + 1/V, \tag{2.15}$$

we obtain

$$V' + V(Y - 4Z_0) = 2, \tag{2.16}$$

where Z_0 is a particular solution of (2.9) with Z being the more general one. By quadrature (2.16) gives

$$V(t) = \left\{\exp\int (4Z_0 - Y)dt\right\} \times \left\{2\int\left[\exp\int (Y - 4Z_0)dt\right]dt + C_3\right\}, \tag{2.17}$$

C_3 being a constant. From (2.8) and (2.17) we obtain

$$V(t) = \left(\frac{B_0^4}{A}\right)\left(\int 2\frac{A}{B_0^4}dt + C_4\right). \tag{2.18}$$

Equations (2.15) and (2.18) yield

$$B(t) = B_0(t)\exp\left\{\int \frac{dt}{(B_0^4/A)[\int(2A/B_0^4)dt + C_4]} + C_5\right\}, \tag{2.19}$$

where C_5 is a constant. Then from the couple $[B_0(t), A(t)]$ we obtain $[B(t), A(t)]$, where $B(t)$ is given by (2.19) and $A(t)$ stays invariable.

B. $K \neq 0$ $(K = \pm 1)$

From (2.4), (2.5), and (2.8) we obtain

$$Y' + Y^2 + YZ = Z' + 2Z^2 + K/B^2. \tag{2.20}$$

By a similar procedure to the one used in Sec. II A, Eq. (2.20), which is of Riccati type in Y, admits the solution

$$A(t) = A_0(t)\exp\left\{\int \frac{dt}{A_0^2 B\,[\int dt/A_0^2 B + C_6]} + C_7\right\}, \tag{2.21}$$

where C_6 and C_7 are two constants.

III. SOLUTIONS

We confine ourselves to particular solutions of the SE type, i.e.,

$$ds^2 = -dt^2 + A_0^2|t|dx^2 + B^2(t)[dy^2 + y^2 dz^2]. \tag{3.1}$$

Here, A, B, and t are given in parametric form such that

$$dt = 2C\,d\kappa, \tag{3.2}$$

$$C = \alpha_1 2^{\gamma/(2-\gamma)}(\cosh\kappa\,\sinh\kappa)^{\gamma/(2-\gamma)}, \tag{3.3}$$

$$A_0 = d_2 2^{2/3(2-\gamma)}(\cosh\kappa)^{2/(2-\gamma)}(\sinh\kappa)^{-2/3(2-\gamma)}, \tag{3.4}$$

$$B = \alpha_3 2^{2/3(2-\gamma)}(\sinh\kappa)^{4/3(2-\gamma)}, \tag{3.5}$$

where $\alpha_1,\alpha_2,\alpha_3$ are given in Ref. 7. For this case we must apply the results of Sec. II A.

(1) Starting from the results of Sec. II A 1, the formula (2.13) yields

$$U(t) = A_0^2 B \times I_1, \tag{3.6}$$

where I_1 is given by

$$I_1 = \int \frac{dt}{A_0^2 B} + C, \tag{3.7}$$

and can be written as

$$I_1 = \frac{\alpha_1}{\alpha_2^2\alpha_3}\left[\int \frac{1}{(\cosh\kappa)^2}(\tanh\kappa)^{\gamma/(2-\gamma)}d\kappa + C_1\right], \tag{3.8}$$

where

$$C_1 = C(\alpha_2^2\alpha_3/\alpha_1).$$

By quadrature we obtain

$$I_1 = \frac{\alpha_1}{\alpha_2^2\alpha_3}\left[\frac{(2-\gamma)}{2}(\tanh\kappa)^{2/(2-\gamma)} + C_1\right]. \tag{3.9}$$

Thus, (3.6) becomes

$$A_0^2 B \times I_1 = \alpha_1\{2^{2/(2-\gamma)}(\cosh\kappa)^{4/(2-\gamma)}\} \times\left\{\frac{2-\gamma}{2}(\tanh\kappa)^{2/(2-\gamma)} + C_1\right\}. \tag{3.10}$$

Evaluating

$$I_2 = \int \frac{dt}{A_0^2 B\,[\int dt/A_0^2 B + C_1]},$$

which is in the formula (3.14), we obtain

$$I_2 = \int \frac{(\sinh\kappa)^{\gamma/(2-\gamma)}d\kappa}{(\cosh\kappa)^{(4-\gamma)/(2-\gamma)}[[(2-\gamma)/2](\tanh\kappa)^{2/(2-\gamma)} + C_1]} \tag{3.11}$$

$$= \int \frac{(\tanh\kappa)^{\gamma/(2-\gamma)}d\kappa}{\cosh^2\kappa[[(2-\gamma)/2](\tanh\kappa)^{2/(2-\gamma)} + C_1]}. \tag{3.12}$$

Setting

$$w = (\tanh(\kappa))^{2/(2-\gamma)} + 2C_1/(2-\gamma),$$

we obtain

$$I_2 = \ln\left[\frac{(\tanh\kappa)^{2/(2-\gamma)} + 2C_1/(2-\gamma)}{C_2}\right], \tag{3.13}$$

where C_2 is a constant. From (3.14) and (3.13) we obtain

$$A(t) = A_0(t)\left\{\frac{(\tanh\kappa)^{2/(2-\gamma)} + 2C_1/(2-\gamma)}{C_2}\right\}, \tag{3.14}$$

where $A_0(t)$ is given by (3.4). In the case $C_1 = 0$, $C_2 = 1$ we obtain

$$A(t) = \alpha_2 2^{2/3(2-\gamma)}[\sinh(\kappa)]^{4/3(2-\gamma)}. \tag{3.15}$$

The new class of metrics reads

$$ds^2 = -4C^2 d\kappa^2 + A^2(f)dx^2 + B^2(t)[dy^2 + y^2 dz^2], \tag{3.16}$$

where C, A, B are given, respectively, by (3.3), (3.14), and (3.5). We call this class $F1$ $(1 < \gamma < 2)$.

In the case (3.15) we have

$$ds^2 = -4C^2\,d\kappa^2 + \alpha_2^2 2^{4/3(2-\gamma)}(\sinh\kappa)^{8/3(2-\gamma)}\,dx^2 + B^2(t)[dy^2 + y^2 dz^2]. \quad (3.17)$$

We call this class F2.

(2) Starting from the results of Sec. II A 2 and using the class F2 as a particular solution we can obtain from (2.18)

$$V(t) = (B_0^4/A)I_3, \quad B_0 = \alpha_3 2^{2/3(2-\gamma)}(\sinh\kappa)^{4/3(2-\gamma)},$$

where I_3 is given by

$$I_3 = \frac{\alpha_1\alpha_2}{\alpha_3^4}\left[\int \frac{(\cosh\kappa)^{\gamma/(2-\gamma)}}{(\sinh\kappa)^{(4-\gamma)/(2-\gamma)}}\,d\kappa + C_4\right]. \quad (3.18)$$

By suitable arrangement (3.18) becomes

$$I_3 = \frac{\alpha_1\alpha_2}{\alpha_3^4}\int \frac{(\coth\kappa)^{\gamma/(2-\gamma)}}{\sinh^2\kappa}\,d\kappa = \frac{\alpha_1\alpha_2}{\alpha_3^4}\left[-\frac{(2-\gamma)}{2}(\coth\kappa)^{2/(2-\gamma)} + C_4\right]. \quad (3.19)$$

Evaluating the integral

$$I_4 = \int \frac{dt}{(B_0^4/A)[\int 2A\,dl/B_0^4 + C_4]},$$

which arises in the formula (2.19), we have

$$I_4 = \int \frac{(\cosh\kappa)^{\gamma/(2-\gamma)}\,d\kappa}{2(\sinh\kappa)^{(4-\gamma)/(2-\gamma)}[-[(2-\gamma)/2](\coth\kappa)^{2/(2-\gamma)} + C_4]}. \quad (3.20)$$

By suitable arrangement (3.20) becomes

$$I_4 = \int \frac{(\coth\kappa)^{\gamma/(2-\gamma)}\,d\kappa}{2(\sinh\kappa)^2[-[(2-\gamma)/2](\coth\kappa)^{2/(2-\gamma)} + C_4]}. \quad (3.21)$$

By a change of variable,

$$\xi = (\coth\kappa)^{2/(2-\gamma)} - 2C_4/(2-\gamma), \quad (3.22)$$

(3.21) becomes

$$I_4 = \int \frac{d\xi}{2\xi}. \quad (3.23)$$

And by quadrature we obtain

$$I_4 = \ln\left[\frac{(\coth\kappa)^{2/(2-\gamma)} - 2C_4/(2-\gamma)}{C_5}\right]^{1/2}, \quad (3.24)$$

where C_5 is a constant. From (2.19) and (3.24) we obtain

$$B(t) = B_0(t)\left(\frac{(\coth\kappa)^{2/(2-\gamma)} - 2C_4/(2-\gamma)}{C_5}\right)^{1/2}, \quad (3.25)$$

and the new class of metrics obtained becomes

$$ds^2 = -4C^2\,d\kappa^2 + A^2(t)dx^2 + B^2(t)(dy^2 + y^2\,dz^2), \quad (3.26)$$

where C,A,B are given, respectively, by (3.3), (3.15), and (3.25). We call this class F3 ($1<\gamma<2$).

For $C_4 = 0$ and $C_5 = 1$ we obtain

$$ds^2 = -4C^2\,d\kappa^2 + \alpha_2^2 2^{4/3(2-\gamma)}(\sinh\kappa)^{8/3(2-\gamma)}\,d\kappa^2 + \alpha_3^2 2^{4/3(2-\gamma)}(\sinh\kappa)^{2/3(2-\gamma)} \times(\cosh\kappa)^{2/(2-\gamma)}[dy^2 + y^2\,dz^2]. \quad (3.27)$$

We call this class F4 ($1<\gamma<2$).

IV. PRESSURE p AND DENSITY μ

The pressure p and the energy density μ are given by the formulas (2.4) and (2.6). For solutions (SE) we obtain

$$\frac{A'}{A} = \left[\frac{\tanh\kappa}{2-\gamma} - \frac{\coth\kappa}{3(2-\gamma)}\right]\frac{2^{-\gamma/(2-\gamma)}}{\alpha_1(\cosh\kappa\sinh\kappa)^{\gamma/(2-\gamma)}}, \quad (4.1)$$

$$\frac{B'}{B} = \frac{2\coth\kappa}{3(2-\gamma)}\frac{2^{-\gamma(2-\gamma)}}{\alpha_1(\cosh\kappa\sinh\kappa)^{\gamma/(2-\gamma)}}, \quad (4.2)$$

$$\kappa_0\mu + \Lambda = (1-\gamma)[-\Lambda + 4/2^{2\gamma/(2-\gamma)}\alpha_1^2(\sinh\kappa\cosh\chi)^{\gamma/(1-\gamma)}], \quad (4.3)$$

$$-\kappa_0 p + \Lambda = (1-\gamma)[-\Lambda + 4/3(2-\gamma)^2\alpha_1^2 2^{2\gamma/(2-\gamma)} \times(\sinh\kappa\cosh\kappa)^{2\gamma/(2-\gamma)}] + \Lambda, \quad (4.4)$$

and therefore

$$p = (\gamma-1)\mu. \quad (4.5)$$

Starting from the F2 class, we evaluate the ratio A'/A and we obtain

$$\frac{A'}{A} = \frac{2\coth(\kappa)}{3(2-\gamma)\alpha_1 2^{\gamma/(2-\gamma)}(\cosh\kappa\sin\kappa)^{\gamma/(2-\gamma)}}, \quad (4.6)$$

and B'/B is given by (4.2). Using again the formulas (2.4) and (2.6) we obtain

$$\kappa_0\mu + \Lambda = \frac{4\coth^2(\kappa)}{3(2-\gamma^2)\alpha_1^2 2^{2\gamma/(2-\gamma)}(\cosh\kappa\sinh\kappa)^{2\gamma/(2-\gamma)}}, \quad (4.7)$$

$$-\kappa_0 p + \Lambda = (1-\gamma)[-\Lambda + 4/3\alpha_1^2(2-\gamma)^2 2^{2\gamma/(2-\gamma)} \times(\sinh\kappa\cosh\kappa)^{2\gamma/(2-\gamma)}] + \Lambda. \quad (4.8)$$

The equation of state is

$$\kappa_0\mu = [-\kappa_0 p/(1-\gamma) + \Lambda]\left[1 + 2/\sqrt{1 + \{\tfrac{3}{4}[\Lambda(1-\gamma) - \kappa_0 p/(1-\gamma)]\alpha_1^2(2-\gamma)^2\}^{\gamma/(2-\gamma)}} - 1\right]. \quad (4.9)$$

By a similar procedure we evaluate the pressure and energy density for the class of metric F4 and we obtain

$$\kappa_0\mu + \Lambda = [\alpha_1^2(2-\gamma)(\sinh\kappa\cosh\kappa)^{\gamma/(2-\gamma)}]^{-2} \times[\tfrac{1}{4}\coth^2\kappa + \tfrac{1}{4}\tanh^2\kappa + \tfrac{5}{6}], \quad (4.10)$$

$$-\kappa_0 p + \Lambda = [\alpha_1(2-\gamma)(\sinh\kappa\cosh\kappa)^{\gamma/(2-\gamma)}]^{-2} \times\{-\tfrac{1}{4}\coth^2\kappa - \tfrac{1}{4}\tanh^2\kappa + \tfrac{11}{6} - \tfrac{4\gamma}{3}\}.$$

Thus p and μ do not verify the equation of state $p = (\gamma-1)\mu$.

V. GEOMETRICAL AND PHYSICAL PROPERTIES OF THE SOLUTIONS UNDER STUDY

We discuss the properties of the shear tensor σ_{ij}. It has been pointed out by Collins and Wainwright[8] that the shear tensor σ_{ij} plays an important role in general relativistic cosmological and stellar models.

The shear tensor arises in the decomposition of four-vector velocity of the fluid; i.e.,

$$u_{a;b} = -\dot{u}_a u_b + \omega_{ab} + \sigma_{ab} + \Theta h_{ab}/3, \tag{5.1}$$

$$\dot{u}_a = u_{a;b}u^b, \quad \dot{u}_a u^a = 0, \tag{5.2}$$

$$\omega_{ab} = u_{[a;b]} + \dot{u}_{[a}u_{b]}, \quad \omega_{ab}u^b = 0, \tag{5.3}$$

$$h_{ab} = g_{ab} + u_a u_b, \quad h_{ab}u^b = 0, \tag{5.4}$$

$$\sigma_{ab} = u_{(a;b)} + \dot{u}_{(a}u_{b)} - \Theta h_{ab}/3, \ \sigma_{ab}u^b = 0, \tag{5.5}$$

and $\Theta = u^a_{;a}$, where $\dot{u}_a, \omega_{ab}, \Theta, \sigma_{ab}$ are called acceleration, rotation, expansion, and shear, respectively, and a ; means a covariant derivative.

If we use Cartesian coordinates in the orbits the metric (1.1) ($K = 0$) reads

$$ds^2 = A^2(t)dx^2 + B^2(dy^2 + dz^2) - dt^2. \tag{5.6}$$

A natural choice of a basis $\{\omega^\alpha\}$ of a one-form is

$$\omega^1 = A(t)dx, \quad \omega^2 = B(t)dy, \quad \omega^3 = B(t)dz, \quad \omega^4 = dt. \tag{5.7}$$

The tetrad metric g_{ab} reads

$$g_{ab} = (1,1,1,-1). \tag{5.8}$$

By straightforward calculations we get

$$\sigma_{11} = \frac{2A^2}{3}\left[\frac{\dot{A}}{A} - \frac{\dot{B}}{B}\right], \tag{5.9}$$

$$\sigma_{22} = \frac{B^2}{3}\left[\frac{\dot{B}}{B} - \frac{\dot{A}}{A}\right], \tag{5.10}$$

$$\sigma_{33} = \frac{B^2}{3}\left[\frac{\dot{B}}{B} - \frac{\dot{A}}{A}\right], \tag{5.11}$$

$\left(\cdot = \frac{d}{dt}\right)$, $\sigma_{ij} = 0$ for all other i and j.

Applying now the above formula for the class F2 we get

$$\sigma_{11} = \sigma_{22} = \sigma_{33} = 0.$$

Thus the class F2 is isotropic. Furthermore we are going to prove that F2 can be reduced by appropriate scale transformations to Friedmann–Roberston–Walker[9–13] (FRW) form.

It is useful to summarize here the physical properties of the FWR model. This universe is the same at all points in space (spatial homogeneity); all directions at a point are equivalent.

For geometrical properties and singularity we refer to Hawking and Ellis.[14]

For the class F4 we get

$$\sigma_{11} = \frac{\alpha_2^2 2^{4/3(2-\gamma)}(\sinh\kappa)^{8/3(2-\gamma)}}{3(2-\gamma)C(\kappa)}[\coth\kappa - \tanh\kappa], \tag{5.12}$$

$$\sigma_{22} = \sigma_{33} = \frac{\alpha_3^2 2^{4/3(2-\sigma)}(\sinh\kappa)^{2/(3-\gamma)}\cosh\kappa^{1/(2-\gamma)}}{6(2-\gamma)C(\kappa)} \times [\tanh\kappa - \coth\kappa], \tag{5.13}$$

where

$$C(\kappa) = \alpha_1 2^{\gamma/(2-\gamma)}(\cosh\kappa\sinh\kappa)^{\gamma/(2-\gamma)}. \tag{5.14}$$

Formulas (3.2) and (3.3) show us directly that t is a monotonic function of κ verifying the following properties:

$t \to 0$ when $\kappa \to 0$,

$t \to \infty$ when $\kappa \to \infty$.

From the above results we deduce the following facts.

(a) The shear tensor σ_{ij} is nonzero for all values of t $0 < t < \infty$; thus the model is anisotropic.

(b) At infinite time ($t \to \infty$) the shear tensor σ_{ij} drops to zero; the universe is then shear-free, there is no anisotropy.

We are now going to prove that F4 goes to a FRW model. For $t \to \infty$ ($\kappa \to \infty$), we have

$\cosh\kappa \to \sinh\kappa$.

Thus the class F4 reads

$$ds^2 = -4C^2\,d\kappa^2 + \alpha_2^2 2^{4/3(2-\gamma)}(\sinh\kappa)^{8/3(2-\gamma)}\,dx^2 + \alpha_3^2 2^{4/3(2-\gamma)}(\sinh\kappa)^{8/3(2-\gamma)}(dy^2 + dz^2). \tag{5.15}$$

By the scale transformations

$$dX = \alpha_2\,dx, \tag{5.16}$$

$$dY = \alpha_3\,dy, \tag{5.17}$$

$$dZ = \alpha_3\,dz, \tag{5.18}$$

$$dT = dt, \tag{5.19}$$

we get

$$ds^2 = 2^{4/3(2-\gamma)}(\sinh(\kappa))^{8/3(2-\gamma)}[dX^2 + dY^2 + dZ^2] - dT^2, \tag{5.20}$$

where κ is a function of time [(3.2), (3.3)].

A similar scale transformation is used for reduced F2 to a FWR model, so (5.20) is manifestly a FRW metric; the factor $2^{2/3(2-\gamma)}[\sinh(\kappa)]^{4/3(2-\gamma)}$ is a universal expansion factor or scale factor. At zero time ($t \to 0$) formulas (4.10) and (4.11) imply infinite pressure and density.

From formulas (3.2) and (3.3) we get (at first order)

$$t = \alpha_1(2-\gamma)2^{\gamma/(2-\gamma)}\kappa^{2/(2-\gamma)}. \tag{5.21}$$

The class F4 takes the limiting form ($t \to 0$)

$$ds^2 = \alpha_2^2[\alpha_1(2-\gamma)]^{-4/3}2^{4(1-\gamma)/3(2-\gamma)}t^{4/3}\,dx^2 + [\alpha_1(2-\gamma)]^{-1/3}\alpha_3^2 2^{(4-\gamma)/(2-\gamma)}t^{1/3}\,dy^2 + (\alpha_1(2-\gamma))^{-1/3}\alpha_3^2 2^{(4-\gamma)/3(2-\gamma)}t^{1/3}\,dz^2 - dt^2. \tag{5.22}$$

We recognize the well-known Kasner model[15] thus the singularity[16] is of Kasner's type.

Before studying if the pressure and density obey the usual energy conditions, i.e.,

$$-\mu < p_\alpha < \mu, \quad \mu > 0, \tag{5.23}$$

where $\alpha = 1,2,3$. We notice that the physical properties of p and μ cannot be largely affected if we set $\Lambda = 0$. This fact has been used to evaluate the correct value for perihelion precession by means of the Schwarzschild[17] exterior solution.[18]

From formula (4.7) we get ($\Lambda = 0$)

$$\kappa_0 \mu = \frac{4 \coth^2 \kappa}{3(2-\gamma)^2 \alpha_1^2 2^{2\gamma/(2-\gamma)} (\cosh \kappa \sinh \kappa)^{2\gamma/(2-\gamma)}}. \tag{5.24}$$

The relation $1 < \gamma < 2$ implies $2\gamma/(2-\gamma) > 2$, hence the energy density for the class F2 is always positive. For t approaching infinity $\kappa_0 \mu$ drops to zero. For t approaching zero $\kappa_0 \mu$ becomes infinite. Before analyzing the nature of the singularity present here we look for the pressure. The formula (4.8) yields

$$\kappa_0 p = 4/3(2-\gamma)^2 \alpha_1^2 2^{2\gamma/(2-\gamma)} (\sinh \kappa \cosh \kappa)^{2\gamma/(2-\gamma)}. \tag{5.25}$$

So p is always positive and goes to zero when $t \to \infty$. For $t \to 0$ we also have a singularity.

We conclude by noting the fact that $\mu + 3p$ is positive, and μ goes to an infinite value; this is the most striking feature of the FRW metric.[14] From formulas (5.24) and (5.25) we get

$$p = [(\gamma - 1)/\coth^2 \kappa]\, \mu, \tag{5.26}$$

so p verifies the inequality $p < \mu$ for any value of $\kappa(t)$. Thus the usual energy conditions are everywhere verified for the class F2. For the class F4 the density μ reads

$$\mu = \{\alpha_3 (2-\gamma)(\sinh \kappa \cosh \kappa)^{\gamma/(2-\gamma)}\}^{-2} \times [\tfrac{1}{4}(\coth^2 \kappa + \tanh^2 \kappa) + \tfrac{5}{6}]; \tag{5.27}$$

hence μ is always positive ($0 < t < \infty$) and goes to zero when $\kappa \to \infty$ ($t \to \infty$). For the pressure we have the formula

$$\kappa_0 p = [\alpha_1 (2-\gamma)(\sinh \kappa \cosh \kappa)^{\gamma/(2-\gamma)}]^{-2} \times [\tfrac{1}{4}(\coth^2 \kappa + \tanh^2 \kappa) + \tfrac{11}{6} + \tfrac{4}{3}\gamma]. \tag{5.28}$$

Thus p is always positive; furthermore, $p < \mu$ for $\gamma < 2$, so the energy conditions are always verified for the class F4.

We conclude by noting the relation between density μ and expansion factor $a(t)$ for the universe F2. From (4.7) and from the expression for the expansion factor $a(t)$ we have ($t \to 0$)

$$\kappa_0 \mu + \Lambda = \frac{4}{3(2-\gamma)^2 \alpha_1^2 2^{2(\gamma-1)/(2-\gamma)}} \frac{1}{a^3(t)}. \tag{5.29}$$

From the general formula[19]

$$\mu(t) = \mu_{m0} [a_0^3 / a^3(t)] + \mu_{r0} [a_0^4 / a^4(t)], \tag{5.30}$$

where μ_{m0} is the density of matter today, μ_{r0} is the density of radiation today, a_0 is the expansion factor for the universe today. We can deduce that the universe F2 is at time zero "matter dominated"; this differs from the so-called standard model of the universe, which becomes matter dominated for $\sim 10^5$ years. We can then introduce a temperature of matter[19] by the relation $T \propto 1/a^2(t)$, $a(t)$ being the expansion factor, hence at zero time we can relate μ and temperature T by means of the relation

$$\mu \propto T^{3/2},$$

which is to be compared with the γ law (for a perfect fluid).

[1] J. Hajj-Boutros, "A method for generating exact solutions of Einstein field equations," in *Lecture Notes in Physics, Gravitation, Geometry and Relativistic Physics* (Springer, Berlin, 1984), Vol. 212, p. 51.
[2] J. M. Stewart and G. F. R. Ellis, J. Math. Phys. **9**, 1072 (1968).
[3] K. P. Singh and D. N. Singh, Mon. Not. R. Astron. Soc. **140**, 453 (1968).
[4] K. P. Singh and Abdussattar, Curr. Sci. **43**, 372 (1974).
[5] K. P. Singh and Abdussattar, J. Phys. A **6**, 1090 (1973).
[6] L. K. Patel and P. C. Vaidya, Prog. Math. **3**, 158 (1969).
[7] D. Kramer, H. Stephani, M. A. H. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (VEB, Deutscher Verlag der Wissenschaften, Berlin, 1980), pp. 131–151.
[8] C. B. Collins and J. Wainwright, Phys. Rev. D **27** 1209 (1983).
[9] A. Friedmann, Z. Phys. **10**, 377 (1922).
[10] A. Friedmann, Z. Phys. **21**, 326 (1924).
[11] H. P. Robertson, Astrophys. J. **82**, 284 (1935).
[12] H. P. Robertson, Astrophys. J. **83**, 137 (1936).
[13] A. G. Walker, Proc. London. Math. Soc. **42**, 90 (1936).
[14] S. W. Hawking and G. F.R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U.P., London, 1973).
[15] E. Kasner, Am. J. Math. **43**, 217 (1921).
[16] For all about singularities see R. Geroch and G. Horowitz, in *General Relativity, an Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U.P., London, 1979).
[17] K. Schwarzschild, Sitz. Preuss. Akad. Wiss. **1916**, 424.
[18] E. A. Lord, *Tensors, Relativity and Cosmology* (McGraw–Hill, New Delhi, 1976), p. 112.
[19] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), pp. 726–765.

Isotropic and anisotropic charged spheres admitting a one-parameter group of conformal motions[a)]

L. Herrera[b)]
Departamento de Física, Facultad de Ciencias, Universidad Central de Venezuela, Caracas 1051, Venezuela

J. Ponce de León
Departamento de Física, Facultad de Ciencias, Universidad Central de Venezuela, Caracas 1051, Venezuela and Centro de Física, Instituto Venezolano de Investigaciones Cientificas, Apartado Postal 1827, Caracas 1010-A, Venezuela

(Received 28 November 1984; accepted for publication 1 March 1985)

Some exact analytical solutions of the static Einstein–Maxwell equations for perfect and anisotropic fluids were found under the assumption of spherical symmetry and the existence of a one-parameter group of conformal motions. All solutions are matched to the Reissner–Nordström metric and possess positive energy density larger than the stresses, everywhere within the sphere.

I. INTRODUCTION

Several papers have been focused on finding interior solutions to the Einstein–Maxwell equations corresponding to static charged spheres.[1–7] More recently, new interior solutions to these equations in spherical symmetry were found by introducing some assumptions on the equation of state or *ad hoc* functional relations between the metric coefficients.[8,9] It is our purpose in this paper to integrate the Einstein–Maxwell equations for spherically symmetric and static distributions of matter under the assumption that the space-time admits a one-parameter group of conformal motions, i.e.,

$$\underset{\xi}{L}\, g_{\mu\nu} = \psi g_{\mu\nu}, \tag{1}$$

where the left-hand side is the Lie derivative of the metric tensor with respect to the vector field ξ^μ, and ψ is an arbitrary function of the coordinates.

Furthermore we generalize the discussion on charged interior solutions, considering anisotropic matter (principal stresses unequal). The introduction of anisotropic matter is suggested by some theoretical works on more realistic equations of state and stellar models,[10,11] which indicate that compact objects could have anisotropic pressures.

The paper is organized as follows. The field equations as well as the conventions used are given in Sec. II. In Sec. III we integrate the field equations considering some models of isotropic charged matter and in Sec. IV, we display solutions for anisotropic charged fluid. The discussion of the results and conclusions are in Sec. V.

II. THE FIELD EQUATIONS AND CONVENTIONS

Let us consider a static distribution of matter represented by charged spherically symmetric fluid which may be anisotropic.

In Schwarzschild coordinates the line element takes the following form:

$$ds^2 = e^{\nu(r)}\, dt^2 - e^{\lambda(r)}\, dr^2 - r^2\, d\Omega^2, \tag{2}$$

with

$$d\Omega^2 = d\theta^2 + \sin^2\theta\, d\phi^2, \quad x^{0,1,2,3} \equiv t,r,\theta,\phi.$$

The total energy-momentum tensor T^ν_μ is assumed to be the sum of two parts, M^ν_μ and E^ν_μ, for matter and electromagnetic contributions, respectively, i.e.,

$$T^\mu_\nu = M^\mu_\nu + E^\mu_\nu. \tag{3}$$

The energy-momentum tensor for anisotropic matter has the usual expression

$$M^\mu_\nu = (\rho_m + p_\perp) U^\mu U_\nu - p_\perp \delta^\mu_\nu + (p_r - p_\perp)\chi^\mu \chi_\nu, \tag{4}$$

where U^μ is the four-velocity $U^\mu = \delta^\mu{}_0 e^{-\nu/2}$, χ^μ is the unit spacelike vector in the radial direction $\chi^\mu = \delta^\mu{}_1 e^{-\lambda/2}$, ρ_m is the energy density, p_r is the pressure in the direction of χ_μ, and $p_\perp$ is the pressure on the two-space orthogonal to χ_μ.

The electromagnetic contribution can be written as[12]

$$E^\mu_\nu = -(1/4\pi)(F_{\nu\beta}F^{\mu\beta} - \tfrac{1}{4}\delta^\mu_\nu F_{\alpha\beta}F^{\alpha\beta}), \tag{5}$$

where $F_{\mu\nu}$ is the electromagnetic field tensor defined in terms of the four-potential A_μ as

$$F_{\mu\nu} = A_{\nu;\mu} - A_{\mu;\nu}. \tag{6}$$

Because we are in the rest frame we adopt the gauge

$$A_\mu(\phi(r),0,0,0). \tag{7}$$

The combined Einstein–Maxwell equations can be expressed as[12]

$$R^\mu_\nu - \tfrac{1}{2}\delta^\mu_\nu R = 8\pi T^\mu_\nu, \tag{8}$$

$$F_{\mu\nu;\alpha} + F_{\nu\alpha;\mu} + F_{\alpha\mu;\nu} = 0, \tag{9}$$

$$F^{\mu\nu}_{\ \ ;\nu} = -4\pi J^\mu, \tag{10}$$

where J^μ is the four-current density that becomes $J^\mu = \bar\rho_e U^\mu$ ($\bar\rho_e$ is the proper charge density) for nonconducting fluids.

Using the line element (2), the field equations (8)–(10) read

$$8\pi\rho_m + E^2 = -e^{-\lambda}(1/r^2 - \lambda'/r) + 1/r^2, \tag{11}$$

[a)] This work was completed when one of the authors (L.H.) was on leave at the Department of Mathematics, University of California, Berkeley, California 94720.

[b)] Postal address: Apartado Postal 80793, Caracas 1080-A, Venezuela.

$$-8\pi p_r + E^2 = -e^{-\lambda}(1/r^2 + \nu'/r) + 1/r^2, \tag{12}$$

$$8\pi p_\perp + E^2 = \frac{e^{-\lambda}}{2}\left(\nu'' + \frac{\nu'^2}{2} + \frac{\nu' - \lambda'}{r} - \frac{\nu'\lambda'}{2}\right), \tag{13}$$

$$[r^2E(r)]' = 4\pi\rho_e r^2. \tag{14}$$

Primes denote differentiation with respect to r, and E is the usual electric field intensity defined as

$$F_{01}F^{01} = -E^2,$$
$$E(r) = -e^{-(\nu+\lambda)/2}\phi'(r), \tag{15}$$
$$\phi'(r) = F_{10} = -F_{01}.$$

The charge density ρ_e defined in Eq. (14) is related to the proper charge density $\bar{\rho}_e$ by

$$\rho_e = \bar{\rho}_e e^{\lambda/2}. \tag{16}$$

Now we shall assume that space-time admits a one-parameter group of conformal motions, i.e.,

$$\underset{\xi}{L}\, g_{\mu\nu} = \xi_{\mu;\nu} + \xi_{\nu;\mu} = \psi g_{\mu\nu}, \tag{17}$$

where ψ is an arbitrary function of r. Thus, using (2), Eq. (17) explicitly reads

$$\xi^1\nu' = \psi \tag{18}$$
$$\xi^0 = \bar{C} = \text{const}, \tag{19}$$
$$\xi^1 = \psi r/2, \tag{20}$$
$$\lambda'\xi^1 + 2\xi^1{}_{,1} = \psi, \tag{21}$$

where a comma denotes partial derivatives. It can be seen at once from (18)–(21) that

$$e^\nu = A^2r^2, \tag{22}$$
$$\psi = Be^{-\lambda/2}, \tag{23}$$
$$\xi^\mu = \bar{C}\delta^\mu_0 + (\psi r/2)\delta^\mu_1, \tag{24}$$

where A and B are constants of integration. Expressions (22)–(24) contain all the implications derived from the existence of the conformal motion.

Now substituting (22) and (23) into (11)–(13), we have

$$8\pi\rho_m + E^2 = (1/r^2)(1 - \psi^2/B^2) - 2\psi\psi'/B^2r, \tag{25}$$
$$-8\pi p_r + E^2 = (1/r^2)(1 - 3\psi^2/B^2), \tag{26}$$
$$8\pi p_\perp + E^2 = \psi^2/B^2r^2 + 2\psi\psi'/B^2r. \tag{27}$$

In what follows, it will be useful to define as a measure of anisotropy the function

$$\Delta = 4\pi(p_\perp - p_r), \tag{28}$$

in terms of which we can formally solve (25)–(27) in order to have

$$8\pi\rho_m = -3X'/2r + 1/2r^2 + \Delta, \tag{29}$$
$$E^2 = (1 - 2X)/2r^2 + X'/2r - \Delta, \tag{30}$$
$$8\pi p_r = (4X - 1)/2r^2 + X'/2r - \Delta, \tag{31}$$
$$X = \psi^2/B^2, \tag{32}$$

and the line element (2), using (22)–(23) and (32), becomes

$$ds^2 = A^2r^2\,dt^2 - dr^2/X - r^2\,d\Omega^2. \tag{33}$$

Thus, if the function ψ (or X) and an equation of state for the stresses are specified *a priori*, then the problem will be fully determined.

Let us now consider that the charged sphere extends to radius r_0. Then the solution of Einstein–Maxwell equations for $r > r_0$ is given by the Reissner–Nordström metric as

$$ds^2 = \left(1 - \frac{2M}{r} + \frac{q^2}{r^2}\right)dt^2 - \left(1 - \frac{2M}{r} + \frac{q^2}{r^2}\right)^{-1}dr^2 - r^2\,d\Omega^2 \tag{34}$$

and the radial electric field is

$$E = q/r^2, \tag{35}$$

where M and q are the total mass and charge, respectively. To match the line element (33) with the Reissner–Nordström metric across the boundary $r = r_0$, (i) we require continuity of gravitational potential $g_{\mu\nu}$ at $r = r_0$,

$$A^2r_0^2 = X(r_0) = (1 - 2M/r + q^2/r_0^2); \tag{36}$$

(ii) we require vanishing of the radial pressure at the boundary

$$P_r(r_0) = 0; \tag{37}$$

and (iii) in the absence of surface concentration of charge at $r = r_0$, we require the continuity of the field tensor $F_{\mu\nu}$. Thus, it follows from (16) and (36) the continuity of the radial electric field

$$E(r_0) = q/r_0^2. \tag{38}$$

Evaluating Eq. (26) at $r = r_0$ and using (38), the vanishing pressure condition gives us

$$q^2/r_0^2 = 1 - 3X(r_0). \tag{39}$$

Feeding this expression back into (36) we obtain

$$M/r_0 = 1 - 2X(r_0), \tag{40}$$

or eliminating $X(r_0)$,

$$M = r_0/3 + \tfrac{2}{3}q^2/r_0. \tag{41}$$

This equation gives us the increase of the total mass caused by the charge. For noncharged spheres the total mass is $r_0/3$ in agreement with previous results.[13]

Another relation between q and M is easily obtained integrating Eq. (31):

$$(4X - 1)Z^2 = 16\pi\int_0^Z (p_r + p_\perp)Z^2\,dZ + D, \tag{42}$$

where $Z = r^2$, and D is a constant. Assuming that $p_r, p_\perp$, and D are non-negative (in the examples we shall show, D is nonzero only for charged nonhomothetic dust) we have

$$X \geqslant \tfrac{1}{4}. \tag{43}$$

Finally, using this inequality together with the relation

$$q^2/M^2 = [1 - 3X(r_0)]/[1 - 2X(r_0)]^2, \quad X(r_0) < \tfrac{1}{3},$$

we obtain

$$q^2 \leqslant M^2. \tag{44}$$

We finish this section by noticing that all the above-mentioned configurations are outside the horizon.

In fact, using (39) and (40) the equation $g_{00}^{\text{ext}} = 0$ becomes

$$g_{00}^{\text{ext}} = (1 - 2M/r + q^2/r^2)$$
$$= \{1 - 2(r_0/r)[1 - 2X(r_0)] + (r_0/r)^2[1 - 3X(r_0)]\} = 0 \tag{45}$$

from which it is obvious [using the inequality $\frac{1}{4} < X(r_0) < \frac{1}{3}$] that $r_g < r_0$, where r_g is the radius of the horizon.

III. PERFECT FLUID SOLUTIONS

In order to determine the unknown functions ρ_m, E^2, P_r, $P_\perp$, and ψ, it is necessary to introduce additional assumptions. In this section, we consider that the fluid is locally isotropic, i.e., $P_r = P_\perp$. According to (29)–(32), different solutions may be obtained by specifying the choice of ψ.

(i) Homothetic charged spheres: Considering $X = C_1 = Cte$, from Eqs. (29)–(31) we get

$$8\pi\rho_m = 1/2r^2, \tag{46}$$

$$8\pi p = (4C_1 - 1)/2r^2, \tag{47}$$

$$E^2 = (1 - 2C_1)/2r^2. \tag{48}$$

Since p and E^2 must be non-negative throughout we have the following condition on C_1:

$$\tfrac{1}{4} \leqslant C_1 \leqslant \tfrac{1}{2}. \tag{49}$$

The state equation is

$$p = (4C_1 - 1)\rho_m. \tag{50}$$

Consequently the inequality (49) also ensures the fulfillment of $dp/d\rho_m \leqslant 1$.

Equation (47) implies that it is not possible to find a finite radius $r = r_0$ such that boundary condition (37) can be satisfied unless $C_1 = \frac{1}{4}$ when we have charged dust. For other values of C_1, Eqs. (46)–(50) represent a distribution of infinite extent. For $C_1 = \frac{1}{2}$, we recover a previously known solution[13] for neutral matter.

We would like to point out that some known charged perfect fluid solutions in the literature admit a one-parameter group of conformal motions. For example, the solutions of Pant and Sah,[5] with $n = 1$, and of Tikekar,[8] for $\alpha = \beta = 0$ and $(a + b) = 1$, are homothetic. Furthermore, the spherically symmetric solutions of Humi and Mansour given by Eqs. 2.21 and 2.22 and 4.1 and 4.2 of their paper,[9] with $K_1 = K_2 = 1$ and $K = 1$, respectively, have conformal symmetry.

In what follows, instead of arbitrarily assuming function X, we shall find some solutions from physical considerations.

(ii) Charged dust spheres: Let us consider the case of spheres of charged dust whose radius is $r = r_0$. If $p = 0$ in Eq. (31) we obtain an equation for X,

$$8\pi p = (4X - 1)/2r^2 + X'/2r = 0, \tag{51}$$

whose solution is

$$X = C/r^4 + \tfrac{1}{4}, \tag{52}$$

where C is a constant of integration.

Feeding (52) back into (29)–(30) we get

$$8\pi\rho_m = 6C/r^6 + 1/2r^2, \tag{53}$$

$$E^2 = (r^4 - 12C)/4r^6, \tag{54}$$

the requirement $\rho_m > 0$ throughout the distribution implies that C must be non-negative. However, (54) gives $E^2 < 0$ in the central region. Thus we have obtained an analytical solution of Einstein–Maxwell equations in the region $r^4 > 12C$ (for further discussion see Sec. V). If $C = 0$, we recover the homothetic charged dust solution.

Setting $C = \alpha r_0^4$ and using (39) and (40) we obtain the total mass and the total charge:

$$q^2/r_0^2 = (1 - 12\alpha)/4, \quad 0 \leqslant \alpha < \tfrac{1}{12}, \tag{55}$$

$$2M/r_0 = (1 - 4\alpha). \tag{56}$$

We note that $q^2 \leqslant M^2$ and notice that the equality will only be valid in case of homothetic ($\alpha = 0$) charged dust spheres.

Finally, using (33), (36), (52), and (55), it is easy to write the interior metric in terms of a parameter α:

$$ds^2 = \left(\alpha + \frac{1}{4}\right)\left(\frac{r}{r_0}\right)^2 dt^2 - \left[\alpha\left(\frac{r_0}{r}\right)^4 + \frac{1}{4}\right]^{-1} dr^2 - r^2\, d\Omega^2. \tag{57}$$

To obtain other solutions, we can specify, in some way, the electric field or charge distribution throughout the interior of the sphere. Then, we integrate Eq. (30) and get

$$(1 - 2X)/2r^2 + X'/2r = E^2(r) \tag{58}$$

to obtain

$$X = Cr^2 + \tfrac{1}{2} + r^2F(r), \tag{59}$$

where C is a constant of integration and

$$F(r) = 2\int \frac{E^2(r)}{r}\, dr. \tag{60}$$

(iii) Conformally symmetric charged spheres: First we shall give the electric field. For the sake of simplicity we suppose that the electric field is the same as in the homothetic solution. Putting (48) into (60) and integrating, we get

$$X = C_1 + Cr^2. \tag{61}$$

Feeding (61) back into (29)–(31), we get

$$8\pi\rho_m = 1/2r^2 + (4C_1 - 1)/r_0^2, \tag{62}$$

$$8\pi p = [(4C_1 - 1)/2r_0^2]\{(r_0/r)^2 - 1\}, \tag{63}$$

$$E^2 = (1 - 2C_1)/2r^2, \tag{64}$$

where r_0 is the vanishing pressure surface

$$r_0^2 = (1 - 4C_1)/6C. \tag{65}$$

Taking $\frac{1}{4} < C_1 < \frac{1}{2}$ we ensure the positiveness of p and E^2, with $C_1 = \frac{1}{2}$, we recover the previously known solution[13] for neutral matter. When $C_1 = \frac{1}{4}$ we have homothetic charged dust.

From the equation of state

$$p = (4C_1 - 1)\rho_m - (8C_1 - 1)(4C_1 - 1)/16\pi r_0^2, \tag{66}$$

we see that $dp/d\rho_m \leqslant 1$.

From the fulfillment of the boundary conditions, we obtain the mass and charge. Using (39) and (40),

$$q^2/r_0^2 = (1 - 2C_1)/2, \tag{67}$$

$$M/r_0 = \tfrac{2}{3}(1 - C_1). \tag{68}$$

Finally the interior metric is obtained from (33), (36), (61), (65), (67), and (68) as

$$ds^2 = ((1 + 2C_1)/6)(r/r_0)^2\, dt^2 - dr^2/[C_1 + [(1 - 4C_1)/6](r/r_0)^2] - r^2\, d\Omega^2. \tag{69}$$

It is obvious from the condition $\frac{1}{4} < C_1 < \frac{1}{2}$, that all the roots of the equation

$$C_1 + [(1 - 4C_1)/6](r/r_0)^2 = 0$$

are larger than r_0.

(iv) Uniform charge density: The simplest form of the charge distribution throughout the interior is the uniform charge density.

We integrate Eq. (14) with $\rho_e = \text{const}$, taking the arbitrary constant of integration equal to zero. We obtain

$$E = \omega r, \tag{70}$$

with $\omega = \frac{4}{3}\pi\rho_e$. Putting (70) into (60), we get from (59)

$$X = \tfrac{1}{2} + Cr^2 + \omega^2 r^4. \tag{71}$$

Feeding (71) back into (29) and (31), we get

$$8\pi\rho_m = \frac{1}{2r^2}\left\{1 + \left(\frac{r}{r_0}\right)^2\right\} + \frac{\alpha^2}{2r_0^2}\left\{1 - \frac{3}{2}\left(\frac{r}{r_0}\right)^2\right\}, \tag{72}$$

$$8\pi p = \frac{1}{2r^2}\left\{1 - \left(\frac{r}{r_0}\right)^2\right\}\left\{1 - \alpha^2\left(\frac{r}{r_0}\right)^2\right\}, \tag{73}$$

where r_0 is the radius of the configuration

$$4\omega^2 r_0^2 + 1/2r_0^2 = -3C, \tag{74}$$

and $\alpha^2 = 8r_0^4\omega^2$. Because ρ_m and p are decreasing functions of r, taking $\alpha^2 \leqslant 1$, we ensure their positiveness; in addition the derivative $dp/d\rho_m$ is not larger than unity, i.e.,

$$\frac{dp}{d\rho_m} = \left(\frac{dp}{dr}\right)\left(\frac{dr}{d\rho_m}\right) = \left[\frac{1 - \alpha^2(r/r_0)^4}{1 + \frac{3}{2}\alpha^2(r/r_0)^4}\right] \leqslant 1. \tag{75}$$

Using (71), (74), and (39) we obtain the total charge

$$q^2 = \omega^2 r_0^6 = (\alpha^2/8) r_0^2. \tag{76}$$

Of course the same result may be obtained from the expression

$$q = 4\pi\int_0^{r_0} \rho_e r^2\, dr. \tag{77}$$

Equation (41) gives the total mass as

$$M/r_0 = \tfrac{1}{3}(1 + \tfrac{1}{4}\alpha^2). \tag{78}$$

Finally the interior metric is obtained from (33), (36), (71), (74), (76), and (78) as

$$ds^2 = \frac{1}{3}\left(1 - \frac{\alpha^2}{8}\right)\left(\frac{r}{r_0}\right)^2 dt^2 - \left\{\frac{1}{2}\left[1 - \frac{1}{3}\left(\frac{r}{r_0}\right)^2\right] + \frac{\alpha^2}{8}\left(\frac{r}{r_0}\right)^2 \times\left[\left(\frac{r}{r_0}\right)^2 - \frac{4}{3}\right]\right\}^{-1} dr^2 - r^2\, d\Omega^2. \tag{79}$$

Until now we have considered that the charge is distributed continuously throughout the sphere. With the aim of better understanding the role played by the charge distribution we shall consider the following case.

(v) Shell of charge: Let us consider a spherical shell of charge which surrounds a neutral core of a self-similar perfect fluid. The radius of the core and the outer surface of the shell are r_1 and r_2, respectively.

Since the core (region I) contains a neutral perfect fluid, the metric functions are given by (33) and (59), with $F(r) = 0$, as

$$e^{\nu}_{\mathrm{I}} = A^2 r^2, \tag{80}$$

$$X = e_{\mathrm{I}}^{-\lambda} = \tfrac{1}{2} + Cr^2. \tag{81}$$

From (29) and (31) we get

$$8\pi\rho_m = 1/2r^2 - 3C, \tag{82}$$

$$8\pi p = 1/2r^2 + 3C. \tag{83}$$

The field equations for the space of the shell (region II) are given by (11)–(14) with $\rho_m = P_r = P_\perp = 0$. To integrate (14) we consider for simplicity $\rho_e = \text{const}$ and obtain

$$E_{\mathrm{II}}(r) = \omega r + D_1/r^2, \quad \omega = \tfrac{4}{3}\pi\rho_e, \tag{84}$$

where D_1 is a constant of integration.

Equation (11) (with $\rho_m = 0$) is easily integrated to give

$$e_{\mathrm{II}}^{-\lambda} = 1 - \frac{1}{r}\int E^2 r^2\, dr + \frac{D_2}{r}. \tag{85}$$

Now, putting (84) into (85), we take

$$e_{\mathrm{II}}^{-\lambda} = 1 - \frac{\omega^2 r^4}{5} + \frac{D_1^2}{r^2} - D_1\omega r + \frac{D_2}{r}. \tag{86}$$

Subtracting Eq. (12) from (11) we get

$$e^{\nu}_{\mathrm{II}} = D_3 e_{\mathrm{II}}^{-\lambda}, \tag{87}$$

where D_3 is a new constant of integration.

Finally, for $r > r_2$ (region III) we have the Reissner–Nordström solution

$$e^{\nu}_{\mathrm{III}} = e_{\mathrm{III}}^{-\lambda} = (1 - 2M/r + q^2/r^2), \tag{88}$$

$$E_{\mathrm{III}} = q/r^2. \tag{89}$$

Let us now apply the boundary conditions: continuity of $g_{\mu\nu}$ at $r = r_2$ gives

$$D_3 = 1, \tag{90}$$

$$-\frac{\omega^2 r_2^4}{5} + \frac{D_1^2}{r_2^2} - D_1\omega r_2 + \frac{D_2}{r_2} = -\frac{2M}{r_2} + \frac{q^2}{r_2^2}; \tag{91}$$

at $r = r_1$,

$$Cr_1^2 + \tfrac{1}{2} = A^2 r_1^2, \tag{92}$$

$$1 - \frac{\omega^2 r_1^4}{5} + \frac{D_1^2}{r_1^2} - D_1\omega r_1 + \frac{D_2}{r_1} = A^2 r_1^2. \tag{93}$$

From the condition $p(r_1) = 0$,

$$1/2r_1^2 + 3C = 0; \tag{94}$$

from the continuity of $F_{\mu\nu}$, at $r = r_1$,

$$\omega r_1 + D_1/r_1^2 = 0; \tag{95}$$

at $r = r_2$,

$$\omega r_2 + D_1/r_2^2 = q/r_2^2. \tag{96}$$

We have six equations (91)–(96) for the six unknown D_1, D_2, C, A, q, M.

Solving the system we obtain

$$3C = -1/2r_1^2, \tag{97}$$

$$A^2 = 1/3r_1^2, \tag{98}$$

$$D_1 = -\omega r_1^3, \tag{99}$$

$$D_2 = -\tfrac{2}{3}r_1 - \tfrac{9}{5}\omega^2 r_1^5, \tag{100}$$

$$q = \omega(r_2^3 - r_1^3) = 4\pi\int_{r_1}^{r_2} \rho_e r^2\, dr, \tag{101}$$

$$M = \frac{1}{3}r_1 + \frac{q^2}{2r_2} + \frac{\omega^2}{2}\left\{\frac{r_2^5}{5} + \frac{9}{5}r_1^5 - r_1^3 r_2^2 - \frac{r_1^6}{r_2}\right\}. \tag{102}$$

(vi) Surface charge: Let us now consider self-similar spheres for which the charge resides entirely on the surface. This case may be easily obtained[4] from the preceding example taking the limit $(r_2 - r_1) \to 0$, and defining the surface charge density σ as

$$\lim_{(r_2 - r_1) \to 0} \rho_e(r_2 - r_1) = \sigma. \tag{103}$$

Thus, from (101) we obtain the total charge

$$q = 4\pi r_0^2 \sigma \tag{104}$$

(as before r_0 denotes the radius of the sphere, so then $r_0 \equiv r_1$). Equation (102) gives the total mass as

$$M = \tfrac{1}{3}r_0 + q^2/2r_0. \tag{105}$$

The comparison between Eq. (41) and Eq. (105) shows that (with respect to surface charge) any continuous distribution of charge throughout the sphere leads to more massive configurations.

Now, let us consider two self-similar perfect fluid spheres with the same total charge and the same radius, one of these spheres has mass M_1 and a charge somehow distributed throughout while the other has mass M_2 with a charge distributed only on the surface. Then, the difference

$$\Delta M = M_1 - M_2 = \tfrac{1}{6}q^2/r_0 \tag{106}$$

is identifiable as the contribution to the total mass due to the electric field inside the sphere.

Finally we write the interior metric. Using (80), (81), (97), and (98),

$$ds^2 = \frac{1}{3}\left(\frac{r}{r_0}\right)^2 dt^2 - \frac{2\, dr^2}{[1 - \frac{1}{3}(r/r_0)^2]} - r^2\, d\Omega^2, \tag{107}$$

and the electric field on the surface will be

$$E = 4\pi\sigma. \tag{108}$$

IV. ANISOTROPIC SOLUTIONS

In this section we obtain some solutions for anisotropic matter, i.e., $p_r \neq p_\perp$.

(i) Homothetic spheres with constant anisotropy: For the sake of simplicity we begin by considering $X = C_1 = \text{const}$ and making a rough model of constant anisotropy throughout the interior:

$$\Delta = 4\pi(p_\perp - p_r) = k^2 = \text{const}. \tag{109}$$

Then, we obtain from (29)–(31),

$$8\pi\rho_m = \frac{1}{2r^2} + \frac{4C_1 - 1}{2r_0^2}, \tag{110}$$

$$8\pi p_r = \frac{4C_1 - 1}{2r_0^2}\left\{\left(\frac{r_0}{r}\right)^2 - 1\right\}, \tag{111}$$

$$8\pi p_\perp = \frac{4C_1 - 1}{2r_0^2}\left\{\left(\frac{r_0}{r}\right)^2 + 1\right\}, \tag{112}$$

$$E^2 = \frac{4C_1 - 1}{2r_0^2}\left\{\left(\frac{r_0}{r}\right)^2\left(\frac{1 - 2C_1}{4C_1 - 1}\right) - 1\right\}, \tag{113}$$

where r_0 is the radius of the configuration defined by

$$r_0^2 = (4C_1 - 1)/2k^2, \tag{114}$$

taking $\frac{1}{4} < C_1 \leqslant \frac{1}{3}$, we ensure the positiveness of ρ, p_r, $p_\perp$, and E^2. On the other hand we have

$$\frac{dp_r}{d\rho_m} \leqslant \frac{1}{3}, \quad \frac{dp_\perp}{d\rho_m} \leqslant \frac{1}{3}. \tag{115}$$

The total mass and charge are given by Eqs. (39) and (40) as

$$q^2/r_0^2 = 1 - 3C_1, \tag{116}$$

$$M/r_0 = 1 - 2C_1. \tag{117}$$

It is worthwhile to note that, unlike perfect fluid homothetic solutions (which have a boundary only in case of charged dust), the introduction of anisotropy allows us to fit all solutions with the Reissner–Nordström metric. Finally, we remark that for $C_1 = \frac{1}{3}$, the electric field vanishes at the surface $r = r_0$, and according to (116) the total charge is zero, i.e., Eqs. (110)–(117) will represent a sphere that as a whole is neutral. It is also interesting to note that the equation of state, in this case, will be $\rho_m = p_r + 2p_\perp$.

(ii) Homothetic ice spheres: As our second model we consider homothetic spheres sustained only by tangential pressures. These kinds of solutions have been considered by Lemaitre.[14]

Setting $X = C_1$ and $p_r = 0$ in (29)–(31), we obtain

$$8\pi\rho_m = 2C_1/r^2, \tag{118}$$

$$8\pi p_\perp = (4C_1 - 1)/r^2, \tag{119}$$

$$E^2 = (1 - 3C_1)/r^2, \tag{120}$$

$$\Delta = (4C_1 - 1)/2r^2, \tag{121}$$

taking $\frac{1}{4} < C_1 < \frac{1}{3}$ we ensure the positiveness of ρ_m, $p_\perp$, E^2 and we also get

$$\frac{dp_\perp}{d\rho_m} \leqslant \frac{2}{3}. \tag{122}$$

Now, from the fulfillment of the boundary conditions we obtain the total mass and charge of the sphere. They are given by the above equations (116) and (117). Finally we write the line element, which is the same for all homothetic bounded charged spheres, as

$$ds^2 = C_1(r/r_0)^2 dt^2 - dr^2/C_1 - r^2\, d\Omega^2. \tag{123}$$

(iii) Uniform charge density anisotropic spheres: As a final example we consider the model with uniform charge distribution. According to Eq. (71), in the case of an isotropic

fluid this consideration was sufficient for defining uniquely the function X (or ψ). Then it is obvious that any function of the form

$$X = \tfrac{1}{2} + Cr^2 + \omega^2 r^4 + H(r),$$

with $H(r)$ arbitrary, will correspond to anisotropic matter. For the sake of simplicity we shall take $H = \text{const}$. Thus, we have

$$X = C_1 + C_2 r^2 + \omega^2 r^4. \tag{124}$$

Feeding back this function into (29)–(31) and using that [from Eq. (70)] $E^2 = \omega^2 r^2$, we obtain

$$8\pi\rho_m = \frac{3C_1 - 1}{r_0^2}\left\{\left[\left(\frac{r_0}{r}\right)^2\left(\frac{1 - C_1}{3C_1 - 1}\right) + 1\right] + \alpha^2\left[1 - \frac{3}{2}\left(\frac{r}{r_0}\right)^2\right]\right\}, \tag{125}$$

$$8\pi p_r = \frac{3C_1 - 1}{r^2}\left[1 - \left(\frac{r}{r_0}\right)^2\right]\left[1 - \alpha^2\left(\frac{r}{r_0}\right)^2\right], \tag{126}$$

$$\Delta = 4\pi(p_\perp - p_r) = (1 - 2C_1)/2r^2, \tag{127}$$

where the radius r_0 is defined by the equation

$$4\omega^2 r_0^2 + (3C_1 - 1)/r_0^2 + 3C_2 = 0$$

and

$$\alpha^2 = 4\omega^2 r_0^4/(3C_1 - 1). \tag{128}$$

For the positiveness of ρ_m, p_r, and $p_\perp$ we choose $\frac{1}{3} < C_1 < \frac{1}{2}$ and $\alpha^2 \leqslant 1$. It is also easy to prove that

$$\frac{dp_r}{d\rho_m} < 1, \quad \frac{dp_\perp}{d\rho_m} < 1. \tag{129}$$

The total mass and charge are

$$q^2 = \alpha^2 r_0^2(3C_1 - 1)/4, \tag{130}$$

$$M = (r_0/3)\left[1 + \tfrac{1}{2}(3C_1 - 1)\alpha^2\right]. \tag{131}$$

V. DISCUSSION AND CONCLUSIONS

In Secs. III and IV we have explicitly displayed some interior solutions to the Einstein–Maxwell equations for isotropic and anisotropic matter, respectively. They have the following properties.

(i) The stresses p_r and $p_\perp$, the mass energy density ρ_m, and E^2 are non-negative throughout the matter.

(ii) The following inequalities hold:

$$\frac{dp_r}{d\rho_m} \leqslant 1, \quad \frac{dp_\perp}{d\rho_m} \leqslant 1.$$

(iii) e^ν and e^λ are positive, continuous, and nonsingular for $r < r_0$.

(iv) All solutions are matched with the Reissner–Nordström metric at $r = r_0$.

Finally, since our solutions are not valid at the center $r = 0$, we can consider the sphere as composed, in the central region, by a core, inside of which all physical quantities are finite, and outside, above the core, by a self-similar fluid described by any of the solutions of Secs. III and IV.

To illustrate this idea, let us match the interior Schwarzschild solution ($\rho_m = \text{const}$) with a charged dust solution given by Eqs. (51)–(57), across the surface $r = r_1$.

The interior Schwarzschild solution is given by[15]

$$ds^2 = [A - B(1 - a^2r^2)^{1/2}]^2\, dt^2 - (1 - a^2r^2)^{-1}\, dr^2 - r^2\, d\Omega^2, \tag{132}$$

$$\rho_m = (3/8\pi)a^2, \tag{133}$$

$$8\pi p = 3a^2\{2A/3/[A - B(1 - a^2r^2)^{1/2}] - 1\}, \tag{134}$$

where a, A, and B are constants to be determined from the boundary conditions.

Thus, from the continuity of $g_{\mu\nu}$ at $r = r_1$ and using (57), we have

$$[A - B(1 - a^2r_1^2)^{1/2}]^2 = (\alpha + \tfrac{1}{4})(r_1/r_0)^2, \tag{135}$$

$$(1 - a^2r_1^2) = [\alpha(r_0/r_1)^4 + \tfrac{1}{4}]; \tag{136}$$

using (134) and the continuity pressure condition at $r = r_1$, we get

$$2A/3/[A - B(1 - a^2r_1^2)^{1/2}] - 1 = 0. \tag{137}$$

Finally, from (54), the continuity of $F_{\mu\nu}$ gives

$$r_1^4 - 12C = 0 \tag{138}$$

(we remember that $C = \alpha r_0^4$).

Thus, we have four equations which can be solved to obtain A, B, α, and a^2 in terms of r_0 and r_1:

$$12C = r_1^4, \quad \alpha = \tfrac{1}{12}(r_1/r_0)^4, \tag{139}$$

$$a^2 = 2/3r_1^2, \quad \rho_m = 1/4\pi r_1^2, \tag{140}$$

$$A = \sqrt{3}B, \tag{141}$$

$$B = \frac{\sqrt{3}}{2}\left(\frac{r_1}{r_0}\right)\left[\frac{1}{12}\left(\frac{r_1}{r_0}\right)^4 + \frac{1}{4}\right]^{1/2}. \tag{142}$$

The total charge will be

$$q^2/r_0^2 = \tfrac{1}{4}[1 - (r_1/r_0)^4]. \tag{143}$$

The mass of the whole configuration will be

$$2M/r_0 = [1 - \tfrac{1}{3}(r_1/r_0)^4]. \tag{144}$$

ACKNOWLEDGMENT

One of us (L.H.) was partially supported by C.O.N. I.C.I.T.

[1] W. B. Bonnor, Z. Phys. **160**, 59 (1960).
[2] W. B. Bonnor, Mon. Not. R. Astron. Soc. **129**, 443 (1965).
[3] F. C. Cooperstock and V. De la Cruz, Gen. Relativ. Gravit. **9**, 835 (1978).
[4] A. L. Mehra and M. L. Bohra, Gen. Relativ. Gravit. **11**, 333, (1978).
[5] D. N. Pant and A. Sah, J. Math. Phys. **20**, 2537 (1979).
[6] P. G. Whitman and R. C. Burch, Phys. Rev. D **24**, 2049 (1981).
[7] Z. Shi-Chang, Gen. Relativ. Gravit. **15**, 293 (1982).
[8] R. Tikekar, J. Math. Phys. **25**, 1481 (1984).
[9] M. Humi and J. Mansour, Phys. Rev. D **29**, 1076 (1984).
[10] V. Canuto, "Neutron Stars: General Review," Solvay Conference on Astrophysics and Gravitation, Brussels, Belgium, 1973.
[11] M. Ruderman, Ann. Rev. Astron. Astrophys. **10**, 427 (1972).
[12] L. D. Landau and E. M. Lifschitz, *The Classical Theory of Fields* (Addison–Wesley, Reading, MA, 1951), paragraphs 90, 94, and 95.
[13] L. Herrera, J. Jiménez, L. Leal, J. Ponce de León, M. Esculpi, and V. Galina, J. Math. Phys. **25**, 3274 (1984).
[14] G. Lemaitre, Ann. Soc. Sci. Bruxelles A **53**, 97 (1933).
[15] R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (Mc Graw–Hill, New York, 1970), 2nd ed., pp. 470–472.

Exact solutions of an algebraically extended Kaluza–Klein theory

R. B. Mann[a)]
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138

(Received 4 October 1984; accepted for publication 2 May 1985)

The four-dimensional field equations in an algebraically extended Kaluza–Klein theory are solved in the static spherically symmetric case. Eight distinct classes of solutions are found, some of which are free of singularities in both the metric and the electromagnetic field.

I. INTRODUCTION

Recently a new technique for extending general relativity (GR) has been introduced.[1] This technique, called the method of algebraic extension (AE), assumes that tensor fields take their values not in the algebra of real numbers $\mathbb{R}$, but in some arbitrary algebra $\mathscr{A}$. Under some very general assumptions, it has been shown using AE that GR may be consistently extended for only five algebras, yielding five theories of gravity.[1–3] The algebras are $\mathbb{R}$, $\mathbb{C}$, $\mathbb{E}$, $\mathscr{Q}$, and $\mathbb{H}$ (denoting the real, complex, hypercomplex, quaternion, and hyperquaternion numbers, respectively); $\mathscr{A} = \mathbb{R}$ yields GR. Lagrangians for these theories have been constructed that are linear in the curvatures; these take their values in the real algebra $\mathbb{R}$ no matter what $\mathscr{A}$ is.[1]

The theories based on $\mathbb{C}$ and $\mathbb{E}$ have been investigated by Moffat and co-workers in some detail.[4] Although the theory based on $\mathbb{C}$ has ghosts, the theory based on $\mathbb{E}$ does not[5] and is, in fact, testable in the solar system.[6] A Kaluza–Klein[7] extension of the $\mathbb{E}$ theory has been constructed by Kalinowski, for both the Abelian[8] and non-Abelian cases.[9] The $\mathbb{E}$ theory is referred to in the literature as N.G.T.[4]

The Abelian case, which is a generalization of Einstein–Maxwell (EM) theory, is quite interesting. The unifying of the $\mathbb{E}$ algebra with the Kaluza–Klein extension yields new interference effects between the gravitational and electromagnetic fields which are not present in EM theory.[8] It has been shown that in the linear approximation the graviton–photon coupling is the same as in EM theory[10] and that the new effects do not contradict any present-day solar-system data.[8]

By finding an exact solution to the field equations of the AE Kaluza–Klein theory, Kalinowski and Kunstatter[11] were able to demonstrate the consequences of the new effects in the static spherically symmetric case. These are as follows.

(1) The electric field is nonsingular at $r = 0$ and has Coulomb-like behavior for large r. There is, therefore, a maximal value for the electric field. Asymptotically the full solution behaves like the Reissner–Nordstrom solution.[12]

(2) The energy distribution is nonsingular.

(3) The total energy of the solution is the same as the Newtonian mass (the mass seen at $r \to \infty$).

(4) At $r = 0$ (or anywhere else) there are no Coulomb-like or Newton-like first- and second-order poles with mass and charge as residues in the solution; hence, in this sense, the solution describes "mass without mass" and "charge without charge."

The aim of the present work is to investigate the field equations in the Kaluza–Klein $\mathbb{E}$ theory in the four-dimensional static spherically symmetric case. Several new exact solutions are found, some of which are nonsingular in both the metric and the electromagnetic field. All solutions reduce to GR in well-defined limits. The solution of Kalinowski and Kunstatter is found as a special case of a more general class of solutions.

The organization of the paper is as follows. In Sec. II the Kaluza–Klein $\mathbb{E}$ theory is outlined. The metric, energy-momentum tensor, and field equations are presented in Sec. III and are solved in a variety of special cases in Sec. IV. Section V contains a discussion of the results.

II. KALUZA–KLEIN $\mathbb{E}$ THEORY

In the hypercomplex ($\mathscr{A} = \mathbb{E}$) case the algebra has basis elements $(1, e_0)$, where $e_0^2 = +1$. Using the method of AE,[1,3] it was shown that an $\mathbb{E}$-valued metric may be given for a real four-dimensional manifold M, with $\mathbb{E}$-valued connection W such that

$$\nabla_\mu g_{\alpha\beta} = g_{\alpha\beta,\mu} - W^{\gamma}_{\mu\alpha} g_{\gamma\beta} + g_{\alpha\gamma} (W^{\gamma}_{\mu\beta})^* = 0, \tag{2.1}$$

where the operator "*" is defined by

$$(a + e_0 b)^* = a - e_0 b. \tag{2.2}$$

The ordering of tensor indices is important, but ordering of factors is not, since $\mathbb{E}$ is Abelian. The curvature and Lagrangian of the $\mathbb{E}$ theory are

$$R^{\ \alpha}_{\mu\nu\beta} = W^{\alpha}_{\mu\beta,\nu} - W^{\beta}_{\nu\beta,\mu} + W^{\sigma}_{\mu\beta} W^{\alpha}_{\nu\sigma} - W^{\sigma}_{\nu\beta} W^{\alpha}_{\mu\sigma}, \tag{2.3}$$

$$\mathscr{L} = \mathbf{g}^{\mu\nu} (A_1 R^{\ \ \sigma}_{\nu\sigma\mu} + A_2 R^{\ \ \sigma}_{\nu\mu\sigma}), \tag{2.4}$$

where A_1 and A_2 are real constants and $\mathbf{g}^{\mu\nu} = \sqrt{-g}\, g^{\mu\nu}$.

The Kaluza–Klein extension of the $\mathbb{E}$ theory is given in Ref. 8. An outline of its derivation is as follows. Let $\underline{P}$ be a principal fiber bundle with structure group U(1) over a space-time M with projection π and connection one-form α. A curvature form for α is

[a)] Now at the Department of Physics, University of Toronto, Toronto, Ontario, Canada M5S 1A7.

$$\Omega = d\alpha = \tfrac{1}{2}\pi^*(F_{\mu\nu}\bar{\Theta}^{\mu}\wedge\bar{\Theta}^{\nu}), \tag{2.5}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{2.6}$$

$$e^*\alpha = A\bar{\Theta}. \tag{2.7}$$

Here A is the electromagnetic four-potential, e is a local cross section of $\underline{P}$, $\bar{\Theta}^{\mu}$ is a frame on M, and $F_{\mu\nu}$ is the electromagnetic field. Using the metric of (2.1), with

$$g_{\alpha\beta}\, g^{\gamma\beta} = g_{\beta\alpha}\, g^{\beta\gamma} = \delta^{\gamma}_{\alpha}, \tag{2.8}$$

connections ω^{α}_{β} and w^{α}_{β} may be defined such that

$$\omega^{\alpha}_{\beta} = \Gamma^{\alpha}_{\beta\gamma}\bar{\Theta}^{\gamma}, \tag{2.9a}$$

$$w^{\alpha}_{\beta} = W^{\alpha}_{\beta\gamma}\bar{\Theta}^{\gamma}, \tag{2.9b}$$

$$w^{\alpha}_{\beta} = \omega^{\alpha}_{\beta} - \tfrac{2}{3}\delta^{\alpha}_{\beta}w, \tag{2.10}$$

where $W^{\alpha}_{\beta\gamma}$ is the connection in (2.1) and

$$w = W_{\gamma}\Theta^{\gamma} \equiv \tfrac{1}{2}(W^{\sigma}_{\gamma\sigma} - W^{\sigma}_{\sigma\gamma})\bar{\Theta}^{\gamma}. \tag{2.11}$$

By introducing a frame

$$\Theta^{A} = (\pi^*(\Theta^{\alpha}), \kappa\alpha = \Theta^{5}), \tag{2.12}$$

the metrization of $\underline{P}$ may now be carried out. From Ref. 8, quantities $\bar{\gamma}$ and $\underline{\gamma}$ may be defined so that

$$\bar{\gamma} = \pi^*(g_{\alpha\beta}\Theta^{\alpha}\otimes\Theta^{\beta}) - \Theta^{5}\otimes\Theta^{5}, \tag{2.13a}$$

$$\underline{\gamma} = \pi^*(g_{\alpha\beta}\Theta^{\alpha}\wedge\Theta^{\beta}). \tag{2.13b}$$

The constant $\kappa = 2\sqrt{G}/c^2$; in this paper units $G = c = 1$ are used. The metric γ_{AB} of $\underline{P}$ is given by[8]

$$\gamma_{AB} = \begin{pmatrix} g_{\alpha\beta} & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.14}$$

By requiring $\nabla\gamma = 0$, a metric connection W^{A}_{B} on $\underline{P}$ may be defined having components

$$w^{A}_{B} = \begin{pmatrix} \pi^*(w^{\alpha}_{\beta}) + g^{\gamma\alpha}H_{\gamma\beta}\Theta^{5} & H_{\beta\gamma}\Theta^{\gamma} \\ g^{\alpha\beta}(H_{\gamma\beta} + 2F_{\beta\gamma})\Theta^{\gamma} & 0 \end{pmatrix}, \tag{2.15}$$

where $H_{\alpha\beta}$ is defined in terms of $F_{\alpha\beta}$ by the equation

$$g_{\sigma\beta}g^{\gamma\sigma}H_{\gamma\alpha} + g_{\alpha\sigma}g^{\sigma\gamma}H_{\beta\gamma} = 2g_{\alpha\sigma}g^{\sigma\gamma}F_{\beta\gamma}, \tag{2.16}$$

with $H_{\alpha\beta} = -H_{\beta\alpha}$. A complete derivation of Eqs. (2.13)–(2.16) is given in Ref. 8. The w^{A}_{B} connection may be rewritten in terms of a connection ω^{A}_{B} as in Eq. (2.10). The four-dimensional Lagrangian for the theory is conveniently written in terms of the connection $\Gamma^{\alpha}_{\beta\gamma}$ and is[8]

$$\mathscr{L} = \mathbf{g}^{\mu\nu}R_{\mu\nu}(\Gamma) + \tfrac{2}{3}\mathbf{g}^{\beta\mu}W_{[\beta,\mu]} + (2(g^{\mu\nu}F_{\mu\nu})^2 - H^{\mu\nu}F_{\mu\nu})\sqrt{-g}, \tag{2.17}$$

where the constants A_1 and A_2 in Eq. (2.4) are chosen so that[13]

$$R_{\mu\nu}(\Gamma) \equiv \Gamma^{\beta}_{\mu\nu,\beta} - \tfrac{1}{2}(\Gamma^{\beta}_{(\mu\beta),\nu} + \Gamma^{\beta}_{(\nu\beta),\mu}) - \Gamma^{\beta}_{\alpha\nu}\Gamma^{\beta}_{\mu\beta} + \Gamma^{\beta}_{\alpha\beta}\Gamma^{\alpha}_{\mu\nu}. \tag{2.18}$$

The field equations which follow from (2.17) are

$$R_{\alpha\beta}(W) - \tfrac{1}{2}g_{\alpha\beta}R(W) = 8\pi T_{\alpha\beta}, \tag{2.19a}$$

$$g_{\mu\nu,\sigma} - g_{\gamma\nu}\Gamma^{\gamma}_{\mu\sigma} - g_{\mu\gamma}\Gamma^{\gamma}_{\sigma\nu} = 0, \tag{2.19b}$$

$$\mathbf{g}^{[\mu\nu]}{}_{,\nu} = 0, \tag{2.19c}$$

$$(\mathbf{H}^{\alpha\mu} - 2\mathbf{g}^{[\alpha\mu]}g^{\nu\beta}F_{\nu\beta})_{,\mu} = 0, \tag{2.19d}$$

using the notation $R_{\alpha\beta}(W) = R_{\alpha\beta}(\Gamma) - \tfrac{2}{3}W_{[\alpha,\beta]}$. The electromagnetic energy-momentum tensor is given by the expression

$$T_{\alpha\beta} = -(1/4\pi)(g_{\sigma\beta}H^{\mu\sigma}F_{\mu\alpha} - 2g^{\mu\nu}F_{\mu\nu}F_{\alpha\beta} - \tfrac{1}{4}g_{\alpha\beta}(H^{\mu\nu}H_{\mu\nu} - 2(g^{\mu\nu}F_{\mu\nu})^2)), \tag{2.20}$$

with

$$H^{\mu\alpha} = g^{\beta\mu}g^{\gamma\alpha}H_{\beta\gamma}. \tag{2.21}$$

It is easy to show that

$$g^{\alpha\beta}T_{\alpha\beta} = 0. \tag{2.22}$$

Note that a coupling of the form $g^{\mu\nu}F_{\mu\nu}$ appears. It is the presence of this term in Eqs. (2.19) and (2.20) that yields the interesting behavior for the gravitational and electromagnetic fields mentioned previously. Solutions of Eqs. (2.19) in the absence of this term may be found in Ref. 14.

The derivation of Eqs. (2.19d) and (2.20) is perhaps a bit subtle. From (2.16) it is clear that $H_{\mu\nu}$ implicitly depends on the metric and the electromagnetic potential; from this equation it can be shown that $H^{\mu\nu}F_{\mu\nu} = H^{\mu\nu}H_{\mu\nu}$. Variations of $H_{\mu\nu}$ therefore always occur as $g^{\tau\alpha}g^{\pi\gamma}(\delta H_{\gamma\alpha}H_{\pi\tau} + H_{\gamma\alpha}\delta H_{\pi\tau})$. This variational combination is straightforwardly solved using (2.16) for variations with respect to both $g_{\mu\nu}$ and A_{μ}. Equations (2.19d) and (2.20) correct the results of Refs. 8 and 11, in which variations of $H_{\mu\nu}$ were not appropriately accounted for[15]; this error is corrected in Ref. 16.

III. THE STATIC SPHERICALLY SYMMETRIC CASE

In the case of spherical symmetry the metric tensor has the form[17]

$$g_{\mu\nu} = \begin{pmatrix} -\alpha & 0 & 0 & \omega \\ 0 & -\beta & u\sin\theta & 0 \\ 0 & -u\sin\theta & -\beta\sin^2\theta & 0 \\ -\omega & 0 & 0 & \gamma \end{pmatrix}, \tag{3.1}$$

where α, β, γ, and ω are functions of r and t. The tensor $g^{\mu\nu}$ has components

$$g^{\mu\nu} = \begin{pmatrix} \dfrac{\gamma}{\omega^2 - \alpha\gamma} & 0 & 0 & \dfrac{\omega}{\omega^2 - \alpha\gamma} \\ 0 & \dfrac{-\beta}{\beta^2 + u^2} & \dfrac{u\csc\theta}{\beta^2 + u^2} & 0 \\ 0 & \dfrac{-u\csc\theta}{\beta^2 + u^2} & \dfrac{-\beta\csc^2\theta}{\beta^2 + u^2} & 0 \\ \dfrac{-\omega}{\omega^2 - \alpha\gamma} & 0 & 0 & \dfrac{-\alpha}{\omega^2 - \alpha\gamma} \end{pmatrix}. \tag{3.2}$$

The electromagnetic field tensor $F_{\mu\nu}$ has components

$$F_{14} = \mathscr{E}(r,t), \tag{3.3a}$$

$$F_{23} = \mathscr{B}(r,t)\sin\theta, \tag{3.3b}$$

all other components being zero. The tensor $H_{\alpha\beta}$ may be shown to be identical to $F_{\alpha\beta}$ using Eqs. (3.1)–(3.3) and (2.16). Using the relation

$$F_{\mu\nu,\sigma} + F_{\nu\sigma,\mu} + F_{\sigma\mu,\nu} = 0, \tag{3.4}$$

which follows from (2.6), the value of $\mathscr{B}(r,t)$ is found to be constant and shall be denoted by B_0. It corresponds to the magnetic charge of the particle.[18]

The components of $H^{\alpha\beta}$ are all zero except for

$$H^{14} = -\mathscr{E}/(\alpha\gamma - \omega^2), \tag{3.5a}$$

$$H^{23} = B_0 \csc\theta/(\beta^2 + u^2). \tag{3.5b}$$

The determinant of the metric is given by

$$\sqrt{-g} = \sin\theta\,(\alpha\gamma - \omega^2)^{1/2}(\beta^2 + u^2)^{1/2}. \tag{3.6}$$

Before writing down the components of $T_{\alpha\beta}$, it is useful to first solve Eqs. (2.19c) and (2.19d). Equation (2.19c) has the solution

$$\omega^2 = l^4\alpha\gamma/(\beta^2 + u^2 + l^4), \tag{3.7}$$

where l is an arbitrary constant of integration. Equation (2.19d) has the solution

$$\mathscr{E}(r,t) \equiv E = (\omega/l^2)[(-Q\rho^2 + 4uB_0)/(\rho^2 + 4l^4)], \tag{3.8}$$

where Q is an aribitrary constant corresponding to the electric charge of the particle[18] and

$$\rho^2 = \beta^2 + u^2. \tag{3.9}$$

The expression for E in (3.8) depends crucially on the presence of the $(g^{\mu\nu}F_{\mu\nu})$ term in (2.19) and (2.20). If this term were absent, then (3.8) would become

$$E = -Q\omega/l^2 = \sqrt{(\alpha\gamma - \omega^2)/(\beta^2 + u^2)}. \tag{3.10}$$

If $l = 0$, then (3.10) yields E with $\omega = 0$. Note that in (3.8) both the electric and magnetic charges of a particle contribute to the electric field E.

The components of $T_{\alpha\beta}$ follow straightforwardly from (2.20). They are

$$4\pi T_{11} = -\frac{1}{2}\left(\frac{El^2}{\omega}\right)^2\frac{\alpha}{\rho^2} - \frac{1}{2}\frac{\alpha B_0^2}{\rho^2} + 2\alpha\left(\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right)^2, \tag{3.11a}$$

$$4\pi T_{22} = \frac{1}{2}\frac{\beta B_0^2}{\rho^2} + \frac{1}{2}\left(\frac{El^2}{\omega}\right)^2\frac{\beta}{\rho^2} + 2\beta\left(\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right)^2, \tag{3.11b}$$

$$4\pi T_{33} = 4\pi T_{22}\csc^2\theta, \tag{3.11c}$$

$$4\pi T_{44} = -(\gamma/\alpha)4\pi T_{11}, \tag{3.11d}$$

$$4\pi T_{23}\csc\theta = \frac{7}{2}\frac{uB_0^2}{\rho^2} - 4\left(\frac{El^2}{\omega}\right)\frac{l^2B_0}{\rho^2} - \frac{1}{2}\left(\frac{El^2}{\omega}\right)^2\frac{u}{\rho^2} - 2u\left(\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right)^2, \tag{3.11e}$$

$$4\pi T_{14} = \frac{-7}{2}\left(\frac{El^2}{\omega}\right)^2\frac{\omega}{\rho^2} + \frac{4uB_0E}{\rho^2} + \frac{1}{2}\frac{\omega B_0^2}{\rho^2} - 2\omega\left(\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right)^2. \tag{3.11f}$$

The field equations (2.19a) may now be written down, using Eqs. (2.19b)–(2.19d) and (3.11). In the static case these are

$$1 + \left(\frac{uB' - \beta A'}{2\alpha}\right)' + B'\left(\frac{\beta B' + uA'}{2\alpha}\right) + \frac{1}{2}\left(\frac{uB' - \beta A'}{2\alpha}\right)\ln(\alpha\gamma U)' = \frac{\beta}{\rho^2}\left[B_0^2 + \left(\frac{El^2}{\omega}\right)^2\right] + 4\beta\left[\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right]^2, \tag{3.12a}$$

$$c + \left(\frac{\beta B' + uA'}{2\alpha}\right)' - B'\left(\frac{uB' - \beta A'}{2\alpha}\right) + \frac{1}{2}\left(\frac{\beta B' + uA'}{2\alpha}\right)\ln(\alpha\gamma U)' = 8B_0\left[\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right] - \frac{u}{\rho^2}\left[B_0^2 + \left(\frac{El^2}{\omega}\right)^2\right] - 4u\left[\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right]^2, \tag{3.12b}$$

$$-A'' + \tfrac{1}{2}(\ln\alpha)'A' - \tfrac{1}{2}[(A')^2 + (B')^2] - \tfrac{1}{2}\ln(\gamma U)'' + \tfrac{1}{4}\ln(\gamma U)'\ln(\alpha/\gamma U)' = \frac{-\alpha}{\rho^2}\left[B_0^2 + \left(\frac{El^2}{\omega}\right)^2\right] + 4\alpha\left(\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right)^2, \tag{3.12c}$$

$$\frac{\gamma}{2\alpha}\left[(1 - U)[(A')^2 + (B')^2] + \frac{1}{2}(\ln\gamma)'\ln\left(\frac{\gamma\rho^2}{\alpha}\right)' + \frac{1}{2}(\ln U)'\ln\left(\frac{\gamma\rho^4}{\alpha^2}\right)' + \frac{1}{2}\frac{(2U - 1)}{1 - U}[(\ln U)']^2 + \ln(\gamma U^2)''\right] = \frac{\gamma}{\rho^2}\left[B_0^2 + \left(\frac{El^2}{\omega}\right)^2\right] - 4\gamma\left(\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right)^2. \tag{3.12d}$$

The quantities A, B, and U are defined by

$$A \equiv \ln\rho, \quad B \equiv \tan^{-1}(\beta/u),$$
$$U \equiv \rho^2/(l^4 + \rho^2) = 1 - \omega^2/\alpha\gamma, \tag{3.13}$$

and

$$A' = \frac{\partial A}{\partial r}, \quad \text{etc.} \tag{3.14}$$

Furthermore, by using the definitions

$$x \equiv \rho^2/\alpha, \quad y \equiv \gamma U, \quad e^q \equiv e^{A + iB} = u + i\beta, \tag{3.15}$$

the field equations (3.12) may be written as

$$2A'' - (A')^2 + (B')^2 + A'\ln(x/y)' = 0, \tag{3.16}$$

$$(\ln y)'' + \tfrac{1}{2}(\ln y)'\ln(xy)' = (2/x)F, \tag{3.17}$$

$$q'' + \tfrac{1}{2}q'\ln(xy)' + 2(i + c)(e^q/x) = (e^q/x)G - (2/x)F, \tag{3.18}$$

where

$$G \equiv 16\left[\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right]\left[B_0 - e^{-q}\frac{uB_0 + Ql^2}{(\rho^2 + 4l^4)}\right], \tag{3.19}$$

$$F \equiv B_0^2 + (El^2/\omega)^2 - 4\rho^2[(uB_0 + Ql^2)/(\rho^2 + 4l^4)]^2. \tag{3.20}$$

The details of the calculations that are used in obtaining (3.12)–(3.20) are given in Appendix A.

Equations (3.16)–(3.18) are the field equations for the

AE Kaluza–Klein theory in the electromagnetic case. In the next section these equations will be solved in a variety of special cases.

IV. SOLUTIONS OF THE FIELD EQUATIONS

The field equations (3.16)–(3.18) are most easily solved by defining a quantity λ (see Ref. 19)

$$(y')^2 = \lambda(r)(y/x). \tag{4.1}$$

Equation (3.17) becomes

$$\lambda' = 4y'F. \tag{4.2}$$

By writing

$$q + z = p, \quad z = \ln y, \tag{4.3}$$

Eq. (3.18) becomes

$$2\frac{d^2p}{dz^2}\lambda + \frac{d\lambda}{dz}\frac{dp}{dz} + 4(i+c)e^p = 2Ge^p, \tag{4.4}$$

where (4.2) has been used. Equation (4.4) integrates to

$$\left(\frac{dp}{dz}\right)^2\lambda + 4(i+c)e^p = \int 2G\frac{de^p}{dz}dz + c_1, \tag{4.5}$$

where c_1 is a complex constant. This equation may be reinserted into (4.4); the resulting equation is then consistent with Eq. (3.16) provided

$$\frac{d\lambda}{dz} = 4e^zF = 2\,\mathrm{Re}\left[\int e^p\frac{dG}{dz}\right] - (\mathrm{Re}\,c_1) + \lambda, \tag{4.6}$$

which follows from comparison with the real part of (3.16).

The condition (4.6) on F and G makes the system of equations (3.16)–(3.18) difficult to solve in general. Solutions may, however, be easily obtained in two special cases: $B' = 0$ and $uB_0 = -Ql^2$. In each case exact solutions to the field equations will be obtained by considering a variety of boundary conditions. These equations are now Eqs. (3.17), (4.5), and (4.6). The fact that all variables now depend on r via y is merely a reflection of the coordinate degree of freedom left in the system. Any one of β, γ, or α may be chosen freely, corresponding to a choice of coordinates. Alternatively, coordinate invariance implies that the functions y may be chosen freely.

Case (A) $B'=0$: In this case (3.13) implies $u = u_0\beta$, and a comparison of (3.12a) and (3.12b) shows that $c = -u_0$. It is convenient in this case to choose coordinates so that $\beta = r^2$. In this case Eq. (3.16) integrates to

$$y/x = \alpha\gamma U = A_0e^A(A')^2, \tag{4.7}$$

which is

$$\alpha\gamma = 4A_0(1 + L^4/r^4)\sqrt{1+u_0^2}, \tag{4.8}$$

where

$$L^4 = l^4/(1+u_0^2). \tag{4.9}$$

By examining (4.6) [or alternatively (3.12a) and (3.12b)] one finds $B_0 = 0$ (otherwise $\beta = \text{const}$ and $\alpha\gamma = 0$). Solving (3.17) yields

$$(r/\alpha)' = 1 - [r^2Q^2/(1+u_0^2)](r^4 + 4L^4)^{-2}, \tag{4.10}$$

which integrates to

$$\alpha = (1 + \alpha_0/r + [Q^2/(1+u_0^2)]r^{-1}K(r,L))^{-1}, \tag{4.11}$$

where

$$K(r,L) = -\int r^2(r^4 + 4L^4)^{-2}\,dr. \tag{4.12}$$

This is a generalization of the solution obtained by Kalinowski and Kunstatter,[11,16] which in turn extended the solution of Papapetrou.[17] Here, $\alpha(r)$ is nonsingular if $\alpha_0 = 0$; letting $r \to \infty$ gives the Newtonian mass to be

$$M_0 = Q^2\pi/16\sqrt{2}l^4. \tag{4.13}$$

This solution is therefore

$$g_{\mu\nu} = \begin{pmatrix} \alpha & 0 & 0 & l^2/r^2\sqrt{1+u_0^2} \\ 0 & -r^2 & u_0r^2\sin\theta & 0 \\ 0 & -u_0r^2\sin\theta & -r^2\sin^2\theta & 0 \\ -l^2/r^2\sqrt{1+u_0^2} & 0 & 0 & (1 + l^4/(1+u_0^2)r^4)\alpha^{-1} \end{pmatrix}, \tag{4.14}$$

where α is given in (4.11). The Kalinowski–Kunstatter solution has $u_0 = 0$. All of the properties of their solution that were mentioned in the Introduction hold for this solution except that as $r \to \infty$, $u \to u_0r^2$. However, the line element still behaves like the Reissner–Nordstrom line element for large r. If the boundary conditions are such that $u \to 0$ as $r \to \infty$, then $u \equiv 0$ (i.e., $u_0 = 0$).

The case $uB_0 = -Ql^2$ shall now be examined. In this case (3.19) and (3.20) imply

$$G = 0, \quad F = B_0^2 + Q^2 = \text{const}. \tag{4.15}$$

Equation (4.6) then gives

$$\lambda = 4e^zF + \lambda_0, \quad (\mathrm{Re}\,c_1) = \lambda_0, \tag{4.16}$$

where λ_0 is a real constant. Finally, (4.4) yields

$$\frac{2}{\sqrt{c_1}}\sinh^{-1}\left[\sqrt{\frac{(i-c)c_1}{4(c^2+1)}}\,e^{-p/2}\right] = \int\frac{dy}{y(4yF+\lambda_0)^{1/2}}. \tag{4.17}$$

There are two distinct solutions to (4.17).

(a) $\lambda_0 \neq 0$: Here $c_1 = \lambda_0(1 + ic_0)$, and

$$(u + i\beta)y\sinh^2[(1+ic_0)^{1/2}(\sinh^{-1}(\sqrt{\lambda_0/4yF}) + a)] = \lambda_0(i-c)(1+ic_0)/4(1+c^2). \tag{4.18a}$$

(b) $\lambda_0 = 0$: Here $c_1 = ic_0$, and

$$(u + i\beta)y\sinh^2\left[\left(\frac{ic_0}{4F}\right)^{1/2}\left(\frac{1}{\sqrt{y}} + a\right)\right] = \frac{(ic_0)(i-c)}{4(1+c^2)}. \tag{4.18b}$$

In both cases a is a complex constant to be chosen to satisfy certain boundary conditions, and

$$\gamma=\left(\frac{l^4+u^2+\beta^2}{u^2+\beta^2}\right)y,\quad \alpha=\frac{(y')^2(u^2+\beta^2)}{y(4yF+\lambda_0)},$$
$$\omega=l^2(y')(4yF+\lambda_0)^{-1/2}, \tag{4.19}$$

where y is an arbitrary function of r.

The solutions (4.18) must now be chosen to satisfy physically realistic boundary conditions. These shall be taken to be, for $r^2\to\infty$,

$$\alpha\to 1,\quad \gamma\to 1,\quad \beta\to r^2. \tag{4.20}$$

The condition $\beta\to r^2$ implies $y\to 1$; for (4.18a) this means

$$a=-\sinh^{-1}\sqrt{\lambda_0/4F} \tag{4.21a}$$

and for (4.18b)

$$a=-1, \tag{4.21b}$$

since the right-hand side (rhs) of both sides must remain finite as $r^2\to\infty$. The two solutions may be power series expanded for large r [$y=1+\mathscr{O}(1/r)$]; an equation of the form

$$\sum_{m=1}^{\infty} a_n r^{-n}=\text{const} \tag{4.22}$$

is obtained, which can only be true if r is a constant implying $\alpha=0$. In fact, by series expanding u and β in powers of r it can be shown that u is constant only for $c_0=0$, which then yields $u\equiv 0$. Hence u is not a nonzero constant; since $uB_0=-Ql^2$, both the left- and right-hand sides separately vanish. There are four possibilities, leading to seven distinct solutions, denoted by cases (B)–(H).

Case (B) $u=l^2=0$: This case is general relativity. The solution is the Reissner–Nordstrom-type[12] solution with electric charge Q, magnetic charge B_0, and mass $M^2=\frac{1}{4}\lambda_0+F$. The function y is, from (4.18a)

$$y=1-2M/r+(Q^2+B_0^2)/r^2. \tag{4.23}$$

The solution (4.18b) yields (4.23) but with $M^2=F$. The functions α, γ, and ω are ($\beta=r^2$)

$$\gamma=y=1-2M/r+(Q^2+B_0^2)/r^2,\quad \alpha=\gamma^{-1},\quad \omega=0. \tag{4.24}$$

Case (C) $u=Q=0$: The solution takes the form

$$\gamma=(1+l^4/r^4)(1-2M/r+B_0^2/r^2),$$
$$\alpha=(1-2M/r+B_0^2/r^2)^{-1},\quad \omega=l^2/r^2 \tag{4.25}$$

in coordinates where $\beta=r^2$, where $M=\frac{1}{4}\lambda_0+B_0^2$. If $\lambda_0=0$ then (4.25) holds with $M=B_0$.

Cases (D)–(F) $B_0=Q=0$: This is a general case of a result obtained by Vanstone.[19] The functions u and β are most easily expressed in terms of the function y

$$\text{(D)}\quad u+i\beta=[\lambda_0(i-c)/4(1+c^2)y](1+ic_0)\times\operatorname{csch}^2[(\sqrt{1+ic_0/2})\ln y], \tag{4.26a}$$

if $\lambda_0\neq 0$ and by

$$\text{(E)}\quad u+i\beta=\frac{ic_0(i-c)}{4(1+c^2)}\operatorname{csch}^2\left(\frac{\sqrt{ic_0}}{2}(z-z_0)\right), \tag{4.26b}$$

for $c_0\neq 0$ and by

$$\text{(F)}\quad u+i\beta=[(i-c)/(1+c^2)](z-z_0)^{-2}, \tag{4.26c}$$

for $c_0=0$. For (4.26a), α, γ, and ω are given by (4.19), with y an arbitrary function of r, and for (4.28b) and (4.28c), α, γ, and ω are

$$\gamma=(1+l^4/(u^2+\beta^2))y_0, \tag{4.27a}$$
$$\alpha=(u^2+\beta^2)(z')^2, \tag{4.27b}$$
$$\omega=l^2\sqrt{y_0}(z'). \tag{4.27c}$$

The constants c, c_0, z_0, and y_0 are determined by the boundary conditions on u [and on α,γ,β from (4.20)].

Consider the case $u\to u_0$ as $r\to\infty$. Since $y\to 1+\mathscr{O}(1/r)$, Eq. (4.26a) yields

$$c=0,\quad c_0=12u_0, \tag{4.28}$$

and so u and β take the form

$$u=\frac{\lambda_0}{2y}\times\left[\frac{\sinh\phi_+\ \sin\phi_- +12u_0(1-\cosh\phi_+\ \cos\phi_-)}{(\cosh\phi_+ -\cos\phi_-)^2}\right], \tag{4.29a}$$

$$\beta=\frac{\lambda_0}{2y}\times\left[\frac{12u_0\sinh\phi_+\ \sin\phi_- -(1-\cosh\phi_+\ \cosh\phi_-)}{(\cosh\phi_+ -\cos\phi_-)^2}\right],$$

where

$$\phi_\pm=(1/\sqrt{2})(\sqrt{1+(12u_0)^2}\pm 1)^{1/2}(\ln y). \tag{4.30}$$

For Eq. (4.26b), Eq. (4.28) holds, but now

$$u=-6u_0\frac{\cosh((z-z_0)/\sqrt{2})\cos((z-z_0)/\sqrt{2})-1}{(\cosh((z-z_0)/\sqrt{2})-\cos((z-z_0)/\sqrt{2}))^2}, \tag{4.29b}$$

$$\beta=6u_0\frac{\sinh((1/\sqrt{2})(z-z_0))\sin((1/\sqrt{2})(z-z_0))}{(\cosh((z-z_0)/\sqrt{2})-\cos((z-z_0)/\sqrt{2}))^2}.$$

Finally, Eq. (4.26c) can only satisfy the given boundary conditions for $c=u_0=0$.

Note that in all three cases, the functions y (or z) may be chosen arbitrarily. It is possible to choose coordinates so that as $r\to 0$, $y\to y_0$ and $y'\to y_0'$. In this case α, β, u, and ω are all finite as $r\to 0$, and so the solutions obtained will at most contain coordinate singularities. For example, if z in (4.26c) is chosen to be

$$z=\sqrt{(r^2+r_1^2)/(r^2+r_2^2)}+z_0, \tag{4.31}$$

then α,β,γ, and ω are finite for all values of r in this coordinate system.

Finally, if boundary conditions are chosen so that $u\to u_0r^2$ as $r^2\to\infty$ then $c\neq 0\neq u_0$ in (4.26c) and so $u\neq 0$.

Cases (G) and (H) $B_0=l^2\equiv 0$: This is the case of an electrically charged, magnetically neutral particle. The functions u and β are, from (4.18),

$$u = +[\lambda_0/2y(1+c^2)]\left[\frac{(1-cc_0)\sinh\psi_+ \sin\psi_- + (c+c_0)(1-\cosh\psi_+ \cos\psi_-)}{(\cosh\psi_+ - \cos\psi_-)^2}\right],$$

$$\beta = [\lambda_0/2y(1+c^2)]\left[\frac{(c+c_0)\psi \sinh\psi_+ \sin\psi_- - (1-cc_0)(1-\cosh\psi_+ \cos\psi_-)}{(\cosh\psi_+ - \cos\psi_-)^2}\right], \tag{4.32a}$$

and

$$u = \frac{c_0}{2y(1+c^2)}\left[\frac{1-\cosh\psi\cos\psi - c\sinh\psi\sin\psi}{(\cosh\psi - \cos\psi)^2}\right],$$

$$\beta = \frac{c_0}{2y(1+c^2)}\left[\frac{c(1-\cosh\psi\cos\psi) + \sinh\psi\sin\psi}{(\cosh\psi - \cos\psi)^2}\right]. \tag{4.32b}$$

Here

$$\psi_\pm \equiv \sqrt{2}(\sqrt{1+c_0^2} \pm 1)^{1/2}[\sinh^{-1}(\sqrt{\lambda_0/4yF}) - \sinh^{-1}(\sqrt{\lambda_0/4F})], \tag{4.33a}$$

$$\psi = \sqrt{c_0/2F}(1/\sqrt{y} - 1), \tag{4.33b}$$

where now $F = Q^2$.

It is clear from (4.32a) and (4.32b) that u cannot be constant unless $u = 0$, since these equations would give y and β to be constants. The boundary conditions (4.20) along with $u \to u_0$ as $r \to \infty$ give the result (4.28) for c and c_0 in both of Eqs. (4.32). By series expanding $y = 1 - 2M/r + \cdots$, the mass parameter M is given by

$$M^2 = \tfrac{1}{4}\lambda_0 + Q^2, \tag{4.34}$$

for both solutions [$\lambda_0 = 0$ in (4.32b)]. Hence (4.32) becomes

$$u = \frac{\lambda_0}{2y}\left[\frac{\sinh\psi_+ \sin\psi_- + 12u_0(1-\cosh\psi_+ \cos\psi_-)}{(\cosh\psi_+ - \cos\psi_-)^2}\right],$$

$$\beta = \frac{\lambda_0}{2y} \times \left[\frac{12u_0 \sinh\psi_+ \sin\psi_- - (1-\cosh\psi_+ \cos\psi_-)}{(\cosh\psi_+ - \cos\psi_-)^2}\right], \tag{4.35a}$$

for $\lambda_0 \neq 0$, and

$$u = \frac{6u_0}{y}\left[\frac{1-\cosh\psi\cos\psi}{(\cosh\psi - \cos\psi)^2}\right],$$

$$\beta = \frac{6u_0}{y}\left[\frac{\sinh\psi\sin\psi}{(\cosh\psi - \cos\psi)^2}\right], \tag{4.35b}$$

where α and γ are given by (4.19) with $l = 0$ and $\omega = 0$.

Suppose now that $y \to 0$ as $r \to 0$. In this case both $\psi_\pm \to \infty$ and $\psi \to \infty$, and if $y \to r^{2\xi}$ for $\xi > 0$, then in (4.33a)

$$u \to \lambda_0 \chi^{-1}(1 - 12u_0) r^{\xi(K_+ - 2)} \to 0,$$

$$\beta \to \lambda_0 \chi^{-1}(12u_0 + 1) r^{\xi(K_+ - 2)} \to 0, \tag{4.36a}$$

where

$$\chi \equiv (2/(1 + \sqrt{1 + 4F/\lambda_0}))K_+ > 0,$$

$$K_+ = [2(1+c_0^2)^{1/2} + 2]^{1/2} > 2, \tag{4.37}$$

and in (4.33b)

$$u \to 12u_0 r^{-2\xi} \exp[-\sqrt{6u_0/F}\, r^{-\xi}] \to 0,$$

$$\beta \to 12u_0 r^{-2\xi} \exp[-\sqrt{6u_0/F}\, r^{-\xi}] \to 0. \tag{4.36b}$$

For both solutions $\gamma = y \to 0$ and

$$\alpha \to 2\lambda_0 \chi^{-2}(12u_0 - 1)^2 r^{2[\xi(K_+ - 1) - 1]} \to 0, \tag{4.38a}$$

$$\alpha \to [(12u_0)^2/2F]\, r^{-2-4\xi} \exp[-2\sqrt{6u_0/F}\, r^{-\xi}] \to 0. \tag{4.38b}$$

Note that in (4.38a), $\xi > \frac{1}{2}$ for $\alpha \to 0$. Hence the metric is nonsingular as $r \to 0$. The function y may be chosen so that only coordinate singularities are present. By choosing $y \to y_0$ as $r \to 0$ it is possible to have α, β, γ, and u all finite and nonvanishing in this limit.

Of course it is possible to work in a coordinate system in which $\beta = r^2$ and then (4.33) implicitly defines y and u in terms of β. For the $\lambda_0 = 0$ case it is relatively easy to express y as a function of r using power series; this is shown in Appendix B.

Finally, for (B) and (H) the electric field is given by Eq. (3.10)

$$E = -Q\omega/l^2 \tag{4.39}$$

and is nonsingular for the cases (D)–(F).

V. SUMMARY AND CONCLUSIONS

Several exact solutions to the $d = 4$ field equations of the AE Kaluza–Klein theory have been found in a variety of special cases. The results for the metric $g_{\mu\nu}$ given in (3.1) are summarized in Table I, and in Table II the values of various parameters in Table I are listed.

In Table I, solution (A) has noncoordinate singularities at the origin only for γ and ω. Solutions (B) and (C) have noncoordinate singularities in α, β, and ω, while solutions (D)–(H) have only coordinate singularities. This is most easily understood by examining Eqs. (4.18). If $u = 0$, then y must be singular when $\beta = 0$; in coordinates where $\beta = r^2$ this forces y (and hence α, γ, ω) to be singular at the origin. But if $u \neq 0$, as in solutions (D), (E), (G), and (H), this singularity can be avoided; hence these solutions have only coordinate singularities. In solution (A) singularities in α can be avoided by choosing α_0 as in Table II. In solution (F) the absence of a mass parameter allows a singularity-free solution, even though $u = 0$. It has been demonstrated that the $\mathbb{E}$ theory (N.G.T) has nonvanishing dynamical torsion[8,20]; hence the singularity theorems[21] may be avoided. In order to understand this more fully in the context of the solutions in Table I it would be of interest to calculate whether or not the curvatures associated with these solutions contain any singularities. This would involve the computation of the Kretsch-

TABLE I. Exact solutions.

Solution	α	β	u	γ	ω
(A) $B'=0$	$(1+Q^2K(r,L)/(1+u_0^2)r)^{-1}$	r^2	u_0r^2	$(1+L^4/r^4)\alpha^{-1}$	L^2/r^2
(B) $u=l=0$	$(1-2M/r+(Q^2+B_0^2)/r^2)^{-1}$	r^2	0	α^{-1}	0
(C) $u=Q=0$	$(1-2M/r+B_0^2/r^2)^{-1}$	r^2	0	$(1+l^4/r^4)\alpha^{-1}$	l^2/r^2
(D) $\lambda_0\neq0$, $B_0=Q=0$	$\dfrac{(y')^2(u^2+\beta^2)}{y(4yF+\lambda_0)}$	$\dfrac{\lambda_0}{2y}(\cosh\phi_+-\cos\phi_-)^{-2}\{12u_0\sinh\phi_+\sin\phi_- -(1-\cosh\phi_+\cos\phi_-)\}$	$\dfrac{\lambda_0}{2y}(\cosh\phi_+-\cos\phi_-)^{-2}\{12u_0(1-\cosh\phi_+\cos\phi_-)+\sinh\phi_+\sin\phi_-\}$	$\left(1+\dfrac{l^2}{u^2+\beta^2}\right)y$	$\dfrac{l^2y'}{\sqrt{4yF+\lambda_0}}$
(E) $\lambda_0=0$, $B_0=Q=0$, $c_0\neq0$	$2(u^2+\beta^2)(\varphi')^2$	$\dfrac{6u_0(\sinh\phi\sin\phi)}{(\cosh\phi-\cos\phi)^2}$	$\dfrac{6u_0(1-\cosh\phi\cos\phi)}{(\cosh\phi-\cos\phi)^2}$	$\left(1+\dfrac{l^4}{u^2+\beta^2}\right)y_0$	$l^2\sqrt{2y_0}\varphi'$
(F) $c_0=\lambda_0=0$, $B_0=Q=0$	$\beta^2(z')^2$	$(z-z_0)^{-2}$	0	$(1+l^4/\beta^2)y_0$	$l^2\sqrt{y_0}z'$
(G) $\lambda_0\neq0$, $B_0=l^2=0$	$\dfrac{(y')^2(u^2+\beta^2)}{y(4yF+\lambda_0)}$	$\dfrac{\lambda_0}{2y}(\cosh\psi_+-\cos\psi_-)^{-2}\{12u_0\sinh\psi_+\sin\psi_- -(1-\cosh\psi_+\cos\psi_-)\}$	$(\lambda_0/2y)(\cosh\psi_+-\cos\psi_-)^{-2}\{12u_0(1-\cosh\psi_+\cos\psi_-)+\sinh\psi_+\sin\psi_-\}$	y	0
(H) $\lambda_0=0$, $B_0=l^2=0$	$\dfrac{(y')^2(u^2+\beta^2)}{4y^2F}$	$\dfrac{6u\sinh\psi\sin\psi}{y(\cosh\psi-\cos\psi)^2}$	$\dfrac{6u_0(1-\cosh\psi\cos\psi)}{y(\cosh\psi-\cos\psi)^2}$	y	0

TABLE II. Notation.

Parameter	Value
M_0	$\pi Q^2/16\sqrt{2}l^4$
$K(r,L)$	$-\displaystyle\int\frac{r^2}{r^4+4L^4}dr=\frac{1}{\sqrt{2}L}J(\sqrt{2}L)$
$J(x)$	$\dfrac{1}{4\sqrt{2}}\left\{\ln\left(\dfrac{x^2+\sqrt{2}x+1}{x^2-\sqrt{2}x+1}\right)+2[\tan^{-1}(\sqrt{2}x+1)+\tan^{-1}(\sqrt{2}x-1)]\right\}$
L^4	$l^4(1+u_0^2)^{-1}$
M^2	$\frac{1}{4}\lambda_0+Q^2+B_0^2$ (Mass)
$\phi_\pm$	$(\frac{1}{2}\sqrt{1+(12u_0)^2}\pm\frac{1}{2})^{1/2}\ln y$
ϕ	$(1/\sqrt{2})(z-z_0)$
$\psi_\pm$	$(2\sqrt{1+(12u_0)^2}\pm2)^{1/2}(\sinh^{-1}(\sqrt{\lambda_0/4yF})-\sinh^{-1}\sqrt{\lambda_0/4F})$
ψ	$(6u_0/F)(1/\sqrt{y}-1)$
y,z	Arbitrary functions of r that equal unity when $r\to\infty$
u_0,λ_0,l^4	Constants of integration
Q^2	Electric charge
B_0^2	Magnetic charge

mann scalar $R_{\mu\nu\alpha\beta}R^{\mu\nu\alpha\beta}$ for each of these solutions, to see what singularity structure this object has. Work on this area is in progress.

Although the magnetic field (3.3b) is singular at $r=0$, the electric field (3.10) or (4.39) is not. This is because of the coupling between the electromagnetic field and the metric in (2.17) as discussed in Sec. III. Asymptotically, all solutions except for (F) have Schwarzchild/Reissner–Nordstrom-like behavior.[8,12] At the origin ($r=0$) in solutions (A), (D), (E), (G), and (H) there is no mass or electric charge; only the parameters l in (A) or (u_0) in (D)–(F) are present.

Although the Kaluza–Klein $\mathbb{E}$ theory described in Sec. II contains the ordinary Kaluza–Klein theory[7] (reducing to it when the skew part of the extended metric vanishes), the solutions obtained in this paper do not contain the static spherically symmetric solutions of the ordinary Kaluza–Klein theory.[22] In Ref. 22 the five-dimensional field equations were solved in the static spherically symmetric case. The solutions of Table I are solutions to the dimensionally reduced four-dimensional field equations of the extended $\mathbb{E}$ theory; as such they contain only the solutions of the dimensionally reduced four-dimensional field equations of the ordinary theory, i.e., Einstein–Maxwell theory.

It is shown in Ref. 8 that solution (A) could describe the field of an electron when $u_0=0$. The parameter l has been interpreted as a constant related to the fermion number of a body[20]; this interpretation is in agreement with all present-day solar-system data.[4–6] However, the parameter u_0 has not yet been given a physical interpretation. It seems to play a role in $g_{\mu\nu}$ somewhat analogous to that of B_0 in $F_{\mu\nu}$. It may be that u_0 is connected with certain topological invariants of the AE theory; this can only be decided by further investigation of the topological structure of AE theories.

ACKNOWLEDGMENTS

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, and by the National Science Foundation under Grant No. PHY82-15249.

APPENDIX A: DERIVATION OF THE FIELD EQUATIONS

In this appendix a derivation of Eqs. (3.16)–(3.19) is given.

The field equations (2.19) are most easily solved as follows. Equations (2.19c) and (2.19d) are straightforwardly solved [using (3.1)–(3.6)] to yield Eqs. (3.7) and (3.8). The set of equations (2.19b) determines the 64 $\Gamma^\lambda_{\mu\nu}$ (Ref. 23). The nonvanishing components have been found by Pant, in the static case, to be,[25] using (3.13),

$$\Gamma^1_{11} = \alpha'/2\alpha, \quad \Gamma^1_{22} = \csc^2\theta\,\Gamma^1_{33} = (1/2\alpha)(uB' - \beta A'),$$

$$\Gamma^1_{44} = \frac{\omega^2}{\alpha^2}\ln\left[\frac{U}{U-1}\right] + \frac{\gamma'}{2\alpha}, \quad \Gamma^2_{12} = \Gamma^3_{13} = \frac{1}{2}A',$$

$$\Gamma^2_{34} = -\Gamma^3_{24}\sin^2\theta = (\omega/2\alpha)B'\sin\theta,$$

$$\Gamma^4_{14} = (\omega^2/2\alpha\gamma)\ln[U/(U-1)] + \gamma'/2\gamma,$$

$$\Gamma^1_{14} = -\Gamma^1_{41} = (\omega/2\alpha)\ln[U/(U-1)],$$

$$\Gamma^1_{23} = -\Gamma^1_{32} = (\sin\theta/2\alpha)(uA' + \beta B'),$$

$$\Gamma^2_{24} = -\Gamma^2_{42} = -\Gamma^3_{43} = -(\omega/2\alpha)A',$$

$$\Gamma^2_{13} = -\Gamma^2_{31} = -\Gamma^3_{12}\sin^2\theta = \Gamma^3_{21}\sin^2\theta = \tfrac{1}{2}B'\sin\theta, \tag{A1}$$

where $X' = \partial x/\partial r$.

Equations (A1) are then used in (2.19a) to obtain (3.12). In the static case the nonvanishing components of (2.19a) yield

$$R_{22} = 8\pi T_{22}, \tag{A2a}$$

$$R_{23} + c\sin\theta = 8\pi T_{23}, \tag{A2b}$$

$$R_{11} = 8\pi T_{11}, \tag{A2c}$$

$$R_{44} = 8\pi T_{44}. \tag{A2d}$$

The equations involving R_{33} are identical to those involving R_{22}, and in the static case the equations involving R_{14} and T_{14} trivially yield $0 = 0$. Equations (A2) are identical to (3.12) when written out using (A1) and (3.11). In (A2b), c is an arbitrary constant.

Equations (A1) [or (3.12)] may be rewritten in a more compact form using the technique of Vanstone[19] and Wyman.[26] Multiplying (3.12a) by u and (3.12b) by β, and then (3.12a) by $(-\beta)$ and (3.12b) by u, respectively, yields after addition

$$u + c\beta + \frac{\rho^2 B'}{4\alpha}\ln(\alpha\gamma U)' + \left(\frac{\rho^2 B'}{2\alpha}\right)' - \frac{\rho^2}{2\alpha}(A'B') = 8\beta B_0[(uB_0 + Ql^2)/(\rho^2 + 4l^4)], \tag{A3}$$

$$uc - \beta + \frac{\rho^2 A'}{4\alpha}\ln(\alpha\gamma U)' + \left(\frac{\rho^2 A'}{2\alpha}\right)' - \frac{\rho^2(A')^2}{2\alpha} = 8uB_0\left[\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right] - 4\rho^2\left[\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right]^2 - \left[B_0^2 + \left(\frac{El^2}{\omega}\right)^2\right]. \tag{A4}$$

Using the definitions of q, x, y in (3.15) these two equations may be rewritten as

$$[q'' + \tfrac{1}{2}\ln(xy)'q' + 2(c+i)(e^q/x)](x/2) = \left[8e^q B_0 - 4\rho^2\left(\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right)\right]\left(\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right) - [B_0^2 + (El^2/\omega)^2], \tag{A5}$$

where (A5) is obtained by adding (A4) to i times (A3).

From (3.3) it follows that

$$(\ln U)' = 2(1-U)A', \tag{A6}$$

$$(\ln U)'' = 2(1-U)[A'' - 2U(A')^2]. \tag{A7}$$

Using these two equations and (3.15), from Eq. (3.12c) one obtains

$$2A'' - (A')^2 + (B')^2 + A'\ln(x/y)' + (\ln y)'' + \tfrac{1}{2}(\ln y)'\ln(xy)' = \frac{2\alpha}{\rho^2}\left[B_0^2 + \left(\frac{El^2}{\omega}\right)^2\right] - 8\alpha\left[\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right]^2, \tag{A8}$$

while (3.12d) becomes

$$(1-U)[2A'' - (A')^2 + (B')^2 + A'\ln(x/y)'] + (\ln y)'' + \tfrac{1}{2}(\ln y)'\ln(xy)' = \frac{2\alpha}{\rho^2}\left[B_0^2 + \left(\frac{El^2}{\omega}\right)^2\right] - 8\alpha\left[\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right]^2. \tag{A9}$$

These two equations are equivalent to the system

$$2A'' - (A')^2 + (B')^2 + A'\ln(x/y)' = 0, \tag{A10}$$

$$(\ln y)'' + \tfrac{1}{2}(\ln y)'\ln(xy)' = \frac{2}{x}\left[B_0^2 + \left(\frac{El^2}{\omega}\right)^2 - \rho^2\left(\frac{uB_0 + Ql^2}{\rho^2 + 4l^4}\right)^2\right]. \tag{A11}$$

By using the definitions of F and G in (3.19) and (3.20), it is easy to show that (A10), (A11), and (A5) correspond to Eqs. (3.16)–(3.19), respectively.

APPENDIX B: CHOICE OF COORDINATES

Suppose that in (4.35b) coordinates are chosen so that $\beta = r^2$. One then must solve for y in terms of r, and then for u in terms of r. From (4.35b),

$$(r^2/6u_0)(\cosh\psi - \cos\psi)^2 = y^{-1}\sinh\psi\sin\psi, \tag{B1}$$

which is, upon expansion in powers of ψ,

$$(r^2/6u_0)\psi^4(1 + (4/6!)\psi^4 + \cdots) = \psi^2(1 + 2x\psi + x^2\psi^2)(1 + (2/5! - 1/(3!)^2)\psi^4) + \cdots, \tag{B2}$$

where

$$x = \sqrt{F/6u_0} = \sqrt{Q^2/6u_0}. \tag{B3}$$

Inversion of the series (B2) yields

$$\left(\frac{1}{\sqrt{y}} - 1\right) = -\frac{1}{2}\left[1 + \frac{Q^2 - r^2}{4Q^2} + \frac{1}{8Q^4}(Q^2 - r^2)^2 + \frac{1}{64}\left[-\frac{1}{90}\left(\frac{6u_0}{Q^2}\right) + 5\left(\frac{Q^2 - r^2}{Q^2}\right)^3\right]\right] \tag{B4}$$

or

$$y = 4\left[1 + \frac{r^2 - Q^2}{4Q^2} - \frac{(r^2 - Q^2)^2}{8Q^4} + \frac{1}{64}\left[5\left(\frac{r^2 - Q^2}{Q^2}\right)^3 + \frac{1}{90}\left(\frac{6u_0}{Q^2}\right)^3\right] + \cdots\right]^{-2}. \tag{B5}$$

Note that as $r \to \infty$, $y \to 0$ instead of 1. This is because in obtaining (B5) it was necessary to cancel a power of ψ^2 on both sides of (B2); this is not valid when $\psi \to 0$, i.e., not for $r \to \infty$. This is a general feature of solutions with $u_0 \neq 0$: one cannot obtain y as a series in r valid for $0 \leqslant r < \infty$. Instead, (4.35b) implicitly defines β and u in terms of r via the function y which can be chosen arbitrarily through a choice of coordinates.

One could require $\beta = 0$ only at $r = 0$; in this case (4.35b) implies

$$0 \leqslant \psi < \pi \tag{B6}$$

or

$$(1 + \pi\sqrt{F/6u_0})^{-2} < y \leqslant 1. \tag{B7}$$

By choosing coordinates so that

$$y = ((r + a_1)/(r + a_0))^2, \tag{B8}$$

where

$$a_0 = a_1 + Q, \tag{B9a}$$

$$a_1 = 6u_0/(Q\pi^2 + 2\pi\sqrt{(6u_0/Q)}), \tag{B9b}$$

conditions (4.20) and (B7) are satisfied for all r. This choice of coordinates yields a metric and electric field that is free of singularities.

[1]R. B. Mann, Class. Quantum Gravity **1**, 561 (1984).
[2]R. B. Mann, Nucl. Phys. B **231**, 481 (1984).
[3]G. Kunstatter, J. W. Moffat, and J. Malzan, J. Math. Phys. **24**, 886 (1983); G. Kunstatter and R. Yates, J. Phys. A **14**, 847 (1981).
[4]For a review, see J. W. Moffat, Lectures from the Seventh International School of Gravitation and Cosmology, Erice, 1981.
[5]R. B. Mann and J. W. Moffat, Phys. Rev. D **26**, 1858 (1982).
[6]J. W. Moffat, Phys. Rev. Lett. **50**, 709 (1983); R. B. Mann and J. W. Moffat, Can. J. Phys. **59**, 1592 (1981); J. W. Moffat, Can. J. Phys. **59**, 283 (1981); Erratum **59**, 1289 (1981).
[7]T. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl. 966 (1921); O. Klein, Z. Phys. **37**, 895 (1926).
[8]M. W. Kalinowski, J. Math. Phys. **24**, 1835 (1983).
[9]M. W. Kalinowski, J. Phys. A **16**, 1669 (1983).
[10]M. W. Kalinowski and R. B. Mann. Class. Quantum Gravity **1**, 157 (1984).
[11]M. W. Kalinowski and G. Kunstatter, J. Math. Phys. **25**, 117 (1984).
[12]H. Reissner, Ann. Phys. (Leipzig) **50**, 106 (1916); G. Nordstrom, K. Ned. Akad. Wet. Versl. Gewone Vergad. Afd. Natuurkd. **20**, 1238 (1918).
[13]R. B. Mann, Ph.D. thesis, University of Toronto, May 1982.
[14]J. W. Moffat, Phys. Rev. D **19**, 3562 (1979); L. Mihich, Nuovo Cimento B **78**, 115 (1983).
[15]I am grateful to G. Kunstatter for pointing out the errors in Refs. 8 and 11 to me.
[16]G. Kunstatter, J. Math. Phys. **25**, 2691 (1984).
[17]A. Papapetrou. Proc. R. Ir. Acad. Sec. A **52**, 69 (1948).
[18]S. Coleman, Harvard University preprint HUTP-82/A032, 1982.
[19]J. R. Vanstone, Can. J. Math. **14**, 568 (1962).
[20]J. W. Moffat, J. Math. Phys. **21**, 1798 (1980); J. C. McDow and J. W. Moffat, J. Math. Phys. **23**, 634 (1980).
[21]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U. P., Cambridge, MA, 1973).
[22]A. Chodos and S. Detweiler, Gen. Relativ. Gravit. **14**, 879 (1982).
[23]The relationship between the two connections $W^\alpha_{\beta\lambda}$, $\Gamma^\alpha_{\beta\lambda}$ and the vector W_μ is discussed in Ref. 24. The 64 $\Gamma^\alpha_{\beta\lambda}$ are *not* independent, but instead are fully determined; however, it has been shown[5] that it is an auxiliary field having no propagating modes, and hence no physical degrees of freedom.
[24]G. Kunstatter, Gen. Relativ. Gravit. **12**, 373 (1980).
[25]D. N. Pant, Nuovo Cimento B **25**, 175 (1975).
[26]M. Wyman, Can. J. Math. **2**, 427 (1950).

Quantum angular momentum fluctuations

Bruno Nachtergaele[a)]
Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium

(Received 6 February 1985; accepted for publication 17 May 1985)

An exact derivation of the angular momentum fluctuations for quantum mechanical systems with two-body interactions is given under suitable cluster conditions. The deviation from the classical result in the presence of a phase transition is discussed.

I. INTRODUCTION

In classical statistical mechanics it is a standard exercise to calculate the total angular momentum of a system of N indistinguishable particles of mass m. Let $L_z = \Sigma_{i=1}^N L_z^i$, where $L_z^i = x_i p_{y,i} - y_i p_{x,i}$, then one checks that

$$\langle L_z^2 \rangle = \sum_{i=1}^{N} \langle x_i^2 \rangle \langle p_{y,i}^2 \rangle + \langle y_i^2 \rangle \langle p_{x,i}^2 \rangle. \tag{1}$$

This computation shows that there is a factorization of the momentum and position observables. Furthermore, using the equipartition law $\langle p_{x,i}^2 \rangle = mkT$ one gets

$$\frac{\langle L_z^2 \rangle}{\Sigma_{i=1}^N \langle x_i^2 + y_i^2 \rangle} = mkT, \tag{2}$$

where $\langle \cdot \rangle$ stands for the thermal expectation. Written in this form the result yields a direct and intrinsic method for the determination of the absolute temperature by means of a measurement of the angular momentum and of the quantity $\Sigma_{i=1}^N \langle x_i^2 + y_i^2 \rangle$, which is essentially related to the compressibility of the system.

Here we are interested in the quantum mechanical version of this law. Another motivation for the study of the angular momentum in quantum mechanics is situated in theories on superfluidity.[1] The deviation of the quantum mechanical law from the classical one is suggested to be the origin of this phenomenon. These authors give an elaborate calculation and discussion of the angular momentum of the free Bose gas, which is therefore assumed to be a genuine theory of superfluidity. The rigorous derivation of this result is found in Ref. 2.

The point of this paper is to give a rigorous derivation in quantum mechanics of the fact that the mean value of the angular momentum shows the classical behavior under fairly general conditions of the state. In particular we suppose that the state has integrable correlation functions and is translation and rotation invariant. As such our result is known for a long time.[1] Their argument was an intuitive one based on the assumption of a finite correlation length hiding the problems related to the boundary conditions. Here we take rigorously into account a variety of boundary conditions.

Basically we proceed as follows. We assume first that we have given a state of the infinite system (i.e., in the thermodynamic limit), which satisfies the cluster and invariance conditions. As our main result we show that this is sufficient to give the factorization of the momentum and position observables. If we suppose furthermore that the state is an equilibrium state, we get the quantum analog of formula (2). Our result is independent of the statistics.

In the last section then, we try to clarify the importance of the conditions in our main theorem (III.1). Especially we discuss some features of the boson gas with condensate, in which case one finds a deviation from the classical behavior. The fluctuations of the angular momentum do not factorize, and even diverge in the thermodynamic limit.

II. GENERAL SETTING

The problem at hand requires the calculation of expectations of observables that are essentially functions of the momentum p and the position q. So, in a certain sense, the algebra of observables relevant to our problem contains merely the quantum analogs of the classical observables. This makes it possible to construct such an algebra for infinite quantum systems, which is useful both for fermion and boson systems.

Take the one-particle configuration space to be $\mathscr{H} = L_2(\mathbb{R}^\nu)$, where ν denotes the space dimension (for simplicity we disregard internal degrees of freedom). Then we define $\mathscr{A}_1$, the local one-particle observables, as the $\star$ algebra generated by the operators of the form $R(p)f(q)$ [on an appropriate domain in $L_2(\mathbb{R}^\nu)$], where R is a polynomial with ν variables and f is an infinitely differentiable function on $\mathbb{R}^\nu$ with compact support. The p and q stand for the canonical momentum and position observables (i.e., $p_k = (1/i)(\partial/\partial x_k)$, q_l is the multiplication operator by x_l; $q = (q_1,...,q_\nu)$, $p = (p_1,..., p_\nu)$; $[q_l, p_k] = i\delta_{lk}$).

An n-particle observable is an element of $\mathscr{A}_n = S_n(\otimes \mathscr{A}_1)^n$, the symmetrized tensor product of n copies of $\mathscr{A}_1$. A general observable for the infinite quantum system is a sequence $A = \{A_n\}_{n=0}^\infty$ with only a finite number of A_n different from zero and for each n, A_n is an n-particle observable ($A_0 \in \mathscr{A}_0$ is a scalar). Summation and multiplication by scalars are defined on $\mathscr{A} = \{A \,|\, A = \{A_i\}_{i=0}^\infty$ as above$\}$ in the usual way. The vector space $\mathscr{A}$ is then generated by elements of the form

$$B = (0,0,...,S_n B_n, 0,...),$$

where B_n is a tensor product $b_1 \otimes \cdots \otimes b_n$, $b_i \in \mathscr{A}_1$ for $i = 1,...,n$ and S_n is the projection on the permutation symmetric operators on $(\otimes \mathscr{H})^n$. To make an algebra we have to define the appropriate product for the generators B only. Having in mind the meaning of this kind of operators, their

[a)] Onderzoeker Interuniversitair Instituut voor Kernwetenschappen (I.I.K.W.), Belgium.

 0022-2488/85/092317-07$02.50

product rule has to be compatible with the following representation in the usual n-body operator language of Bose and Fermi creation and annihilation operators. The n-particle component of A_n of A is then represented by

$$\frac{1}{n!}\int dx_1\cdots dx_n\, dy_1\cdots dy_n\,(x_1,\dots,x_n|A_n|y_1,\dots,y_n)$$
$$\times a^+(x_1)\cdots a^+(x_n)a(y_n)\cdots a(y_1).$$

If we bring back a product of two such operators, say an n- and an m-body operator, to this canonically ordered form (Wick ordering) by application of the commutation or anticommutation relations for $a^+(x)$ and $a(x)$, we obtain a list of k-body terms for $k=\max\{m,n\},\dots,m+n$. According to this compatibility we define a product on $\mathscr{A}$ as follows:

$$A=(0,\dots,0,S_n(a_1\otimes\cdots\otimes a_n),0,\dots),$$
$$a_i\in\mathscr{A}_1,\quad \text{for } i=1,\dots,n,$$
$$B=(0,\dots,0,S_n(b_1\otimes\cdots\otimes b_n),0,\dots),$$
$$b_j\in\mathscr{A}_1,\quad \text{for } j=1,\dots,n,$$
$$AB=\{(AB)_j\}_{j=0}^{\infty},$$
$$(AB)_j=0,\quad \text{if } j>m+n \text{ or } j<\max\{m,n\},$$
$$\begin{aligned}(AB)_j={}&\frac{j!}{n!m!(j-n)!(j-m)!(m+n-j)!}S_j\\&\times\sum_\pi\sum_\sigma a_{\pi(1)}\otimes\cdots\otimes a_{\pi(j-m)}\\&\otimes a_{\pi(j-m+1)}b_{\sigma(1)}\otimes\cdots\otimes a_{\pi(n)}b_{\sigma(m+n-j)}\\&\otimes b_{\sigma(m+n-j+1)}\otimes\cdots\otimes b_{\sigma(m)},\end{aligned}\tag{3}$$

if $\max\{m,n\}\leqslant j\leqslant m+n$, where π and σ run over the permutation groups of n and m elements.

A $\star$ operation is defined on $\mathscr{A}$ by taking the adjoint of operators.

A state for the infinite quantum system is defined as a normalized linear functional ω of $\mathscr{A}$, which is positive, that means that $\forall A\in\mathscr{A}\ \omega(A^\star A)\geq 0$. We assume that ω is given in terms of a set of correlation functions $\langle\rho^{(n)}p^\alpha\rangle(x_1,\dots,x_n)$, $n\in\mathrm{N}$, where p^α denotes a monomial of $n\nu$ variables, i.e., $\forall\alpha\in\mathrm{N}^{n\nu}\ p^\alpha=p_1^{\alpha_{11}}\cdots p_1^{\alpha_{1\nu}}p_2^{\alpha_{21}}\cdots p_n^{\alpha_{n\nu}}$. Then an expectation value of an observable of the form

$$\begin{aligned}A=(0,0,\dots,0,S_n(&R_1(p)f_1(q)\otimes R_2(p)f_2(q)\\&\otimes\cdots\otimes R_n(p)f_n(q)),0,\dots)\end{aligned}\tag{4}$$

is calculated by the formula

$$\begin{aligned}\omega(A)={}&\frac{1}{n!}\int dx_1\cdots dx_n\,f_1(x_1)\cdots f_n(x_n)\\&\times\langle\rho^{(n)}R^{(n)}\rangle(x_1,\dots,x_n),\end{aligned}\tag{5}$$

where $R^{(n)}$ is the monomial given by

$$R^{(n)}(p_1,\dots,p_n)=R_1(p_1)R_2(p_2)\cdots R_n(p_n),$$

and each $p_i=(p_{i,1},\dots,p_{i,\nu})$. Here ω is extended to $\mathscr{A}$ by linearity.

Such a system of correlation functions occurs when a state on $\mathscr{A}$ is described in terms of a family of reduced density matrices $\{\rho^{(n)}\}_{n=0}^{\infty}$, where the $\rho^{(n)}$ are positive linear operators on $(\otimes\mathscr{H})^n$, such that they have an infinitely differentiable kernel $(y_1,\dots,y_n|\rho^{(n)}|x_1,\dots,x_n)$. Then define

$$\langle\rho^{(n)}R(p_{1,1},\dots,p_{1,\nu};p_{2,1},\dots;p_{n,1},\dots,p_{n,\nu})\rangle(x_1,\dots,x_n)$$

as

$$\begin{aligned}&R\left(\frac{\partial}{\partial y_{1,1}},\dots,\frac{\partial}{\partial y_{1,\nu}},\dots,\frac{\partial}{\partial y_{n,1}},\dots,\frac{\partial}{\partial y_{n,\nu}}\right)\\&\quad\times(y_1,\dots,y_n|\rho^{(n)}|x_1,\dots,x_n)|_{y_1=x_1,\dots,y_n=x_n}.\end{aligned}\tag{6}$$

Thus, with the definitions (5) and (6),

$$\omega(A)=\sum_n\mathrm{Tr}(A_n\rho^{(n)}).$$

For infinite systems one has to define local densities of momentum, position fluctuations, and so on, because the observables themselves are not well defined. This is done by constructing a sequence $\{A_\Lambda\}_\Lambda$ of local approximations for the observable A under consideration, where Λ is a volume in $\mathbb{R}^\nu$, which is tending to infinity in a specified sense. In particular we will use the following local approximations for position and momentum.

Define a family $(f_R)_{R\in\mathbb{R}^+}$ of real functions on $\mathbb{R}^\nu$ such that

$$\forall R\in\mathbb{R}^+,\quad f_R\in C_0^\infty(\mathbb{R}^\nu),\quad 0\leqslant f_R\leqslant 1,$$
$$f_R(x)=\begin{cases}1, & \text{if } \|x\|\leqslant R,\\ 0, & \text{if } \|x\|\geqslant R+1,\end{cases}\tag{7}$$

and $\forall\alpha=1,\dots,\nu$, $\|\nabla^\alpha f_R\|_\infty<\infty$ uniformly in R. We do not specify these cutoff functions further, which would correspond to a particular choice of boundary conditions.

Now we define the following observables in $\mathscr{A}$, $\forall f\in C_0^\infty(\mathbb{R}^\nu)$, $i=1,\dots,\nu$,

$$P_i(f)=(0,\tfrac{1}{2}(p_if(q)+f(q)p_i),0,\dots),$$
$$Q_i(f)=(0,q_if(q),0,\dots).$$

It is clear that $P_i(f_R)$ and $Q_i(f_R)$ are observables corresponding to the bulk momentum and bulk position, located in the support of f_R, roughly a ball in $\mathbb{R}^\nu$ with radius R. When ν is 2 or 3 we can define the third component of the local angular momentum as

$$L_3(f_R)=Q_1(f_R)P_2(f_R)-Q_2(f_R)P_1(f_R),$$

which has, according to the product rule (3), nonvanishing one- and two-particle components. We denote them by $[L_3(f_R)]_1$ and $[L_3(f_R)]_2$. At any time it will be clear whether they have to be considered as an infinite sequence (an element of $\mathscr{A}$) or as the only nonvanishing component of such a sequence:

$$\begin{aligned}A_1\equiv[L_3(f_R)]_1={}&\tfrac{1}{2}(q_1f_R(q)p_2f_R(q)+q_1f_R(q)^2p_2\\&-q_2f_R(q)p_1f_R(q)-q_2f_R(q)^2p_1),\\A_2\equiv[L_3(f_R)]_2={}&2!S_2\tfrac{1}{2}(q_1f_R(q)\otimes p_2f_R(q)\\&+q_1f_R(q)\otimes f_R(q)p_2\\&-q_2f_R(q)\otimes p_1f_R(q)\\&-q_2f_R(q)\otimes f_R(q)p_1).\end{aligned}$$

Now, one computes $L_3^2(f_R)$, expanded in its different terms:

$$L_3^2(f_R) = (0, [A_1^2]_1, [A_1^2]_2 + [A_1A_2]_2 + [A_2A_1]_2 + [A_2^2]_2, [A_1A_2]_3 + [A_2A_1]_3 + [A_2^2]_3, [A_2^2]_4, 0,\ldots). \tag{8}$$

In order to apply (5), one has to write $L_3^2(f_R)$ as a sum of terms of the form (4).

The L_1-clustering conditions for a translation-invariant state can be expressed and used in an elegant way if we introduce the truncated correlation functions $\langle \rho^{(n)} p^\alpha \rangle_T$, which are recursively defined as follows:

$$\langle \rho^{(n)} p_1^{\beta_1} \cdots p_n^{\beta_n} \rangle(x_1,\ldots,x_n) = \sum_{p\in\mathscr{P}_n} \prod_{\pi\in p} \langle \rho^{(l)} p_1^{\beta_{\pi(1)}} \cdots p_l^{\beta_{\pi(l)}} \rangle_T (x_{\pi(1)},\ldots,x_{\pi(l)}),$$
$$\beta_i \in \mathbb{N}^\nu, \quad i = 1,\ldots,n, \tag{9}$$

where $\mathscr{P}_n$ is the set of all partitions p of $\{1,\ldots,n\}$. Such a partition has elements $\pi = \{\pi(1),\ldots,\pi(l)\}$ and the $\pi(i)$ are labeled such that $\pi(1) < \pi(2) < \cdots < \pi(l)$. It is easily seen that a translation-invariant state has translation-invariant truncated functionals $\langle \rho^n p^\alpha \rangle_T$ such that

$$\langle \rho^{(n)} p_{p_1}^{\beta_1} \cdots p_n^{\beta_n} \rangle(x_1,\ldots,x_n) = \sum_{p\in\mathscr{P}_n} \prod_{\pi\in p} \langle \rho^{(l)} p_1^{\beta_{\pi(1)}} \cdots p_l^{\beta_{\pi(l)}} \rangle_T \times (x_{\pi(1)} - x_{\pi(l)},\ldots,x_{\pi(l-1)} - x_{\pi(l)}).$$

An L_1-clustering condition is expressed as follows: $\langle \rho^{(n)} p^\alpha \rangle_T(\cdot,\cdot,\cdot,\cdot,0)$ (as a function of $n-1$ ν-dimensional variables) is $L_1(\mathbb{R}^{(n-1)\nu})$ integrable, for certain values of n and α. We now derive a lemma which is an immediate consequence of such a clustering condition. It shows how L_1 clustering is at the origin of the factorization of $\langle L_3^2 \rangle$.

Lemma II.1: Suppose that the truncated correlation functions $\langle \rho^{(n)} p^\alpha \rangle_T$ are translation invariant and are $L_1(\mathbb{R}^{(n-1)\nu})$. Let $\{A(R)\}_{R\in\mathbb{R}^+}$ be a sequence of observables

$$A(R) = (0,\ldots,0, S_n(p^{\alpha_1} f_1(q) \otimes \cdots \otimes p^{\alpha_n} f_n(q)),0,\ldots),$$

with

$$f_1(x_1)\cdots f_n(x_n) = x_{i,j}\, x_{k,l}\, f_R(x_1)\cdots f_R(x_n),$$
$$i,k \in \{1,\ldots,n\} \quad j,l \in \{1,\ldots,\nu\}.$$

The f_R is defined as above; let Σ_ν be the volume of the unit ball in ν dimensions. Then for $\epsilon > 0$,

$$\lim_{R\to\infty} \frac{\omega_T(A(R))}{\Sigma_\nu R^{\nu+2+\epsilon}} = \lim_{R\to\infty} \frac{1}{\Sigma_\nu R^{\nu+2+\epsilon}} \int d^\nu x_1 \cdots d^\nu x_n\, x_{i,j} x_{k,l} \times f_R(x_1)\cdots f_R(x_n) \langle \rho^{(n)} p_1^{\alpha_1} \cdots p_n^{\alpha_n} \rangle_T (x_1,\ldots,x_n) = 0.$$

Proof: By a simple substitution $x_j = x_j' - x_k$ $(j \neq k)$, $x_k' = x_k$, and application of the translation invariance of $\langle \rho^{(n)} p^\alpha \rangle_T$, we have

$$\int d^\nu x_1 \cdots d^\nu x_n\, x_{i,j} x_{k,l} f_R(x_1)\cdots f_R(x_n) \times \langle \rho^{(n)} p^\alpha \rangle_T (x_1,\ldots,x_n)$$
$$= \int d^\nu x_1 \cdots d^\nu x_{k-1}\, d^\nu x_{k+1} \cdots d^\nu x_n \times \langle \rho^{(n)} p^\alpha \rangle_T (x_1,\ldots,x_{k-1},0, x_{k+1},\ldots, x_n)$$
$$\times \int d^\nu x_k (x_i + x_k)_j\, x_{k,l}\, f_R(x_1 + x_k) \cdots f_R(x_{k-1} + x_k) f_R(x_k) f_R(x_{k+1} + x_k) \cdots f_R(x_n + x_k).$$

The second integral is majorized by $2\Sigma_\nu (R+1)^{\nu+2}$ and hence the integral

$$\frac{1}{R^{\nu+2+\epsilon}} \int d^\nu x_k (x_i + x_k)_j\, x_{k,l}\, f_R(x_1 + x_k) \cdots f_R(x_k) \times f_R(x_{k+1} + x_k) \cdots f_R(x_n + x_k),$$

as a function of $x_1,\ldots, x_{k-1}, x_{k+1},\ldots, x_n$, is bounded uniformly for $R \geqslant 1$ and for each $(x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_n) \in \mathbb{R}^{(n-1)\nu}$ it converges to zero. As a consequence of the dominated convergence theorem we obtain the result. ∎

Remark II.2: Obviously the result of the lemma remains true whenever the functions f_R are replaced by some functions g_R such that $\forall x\in\mathbb{R}^\nu, |g_R(x)| \leqslant C f_{(AR+B)}(x)$, $A,B,C > 0$, or also if we drop a factor $x_{i,j}$ in the integrand.

In the calculation of $\langle L_3^2 \rangle$ we will use frequently the following properties: $\langle \rho^{(1)} p^\beta \rangle$ is a constant $\forall \beta \in \mathbb{N}^\nu$ by translation invariance; by rotation invariance $\langle \rho^{(1)} p_j \rangle = 0$, $\forall j \in \{1,\ldots,\nu\}$, and $\langle \rho^{(1)} p_i p_j \rangle = \delta_{ij} \langle \rho^{(1)} (p_i)^2 \rangle$; and $\langle \rho^{(1)} \rangle$ is the usual particle density ρ.

It is clear that Lemma II.1 (if applicable) implies that a two-particle term in $L_3^2(f_R)$ factorizes. So it vanishes when one of its factors vanishes. For the same reason a three-particle contribution in $L_3^2(f_R)$ with all three one-particle factors vanishing, has zero expectation itself.

III. THE ANGULAR MOMENTUM

Now we are ready to formulate and prove our main result, namely the proof that the quantum mechanical angular momentum density behaves like the classical one [see formula (1)].

We remark that in the following theorem we prove that there is a factorization of the momentum and of the position observables and that this result is solely a consequence of the clustering properties of the state.

Theorem III.1: Let ω be a translation-, rotation-, and time reversal-invariant state of $\mathscr{A}$ with correlation functions satisfying the L_1-clustering conditions

$$\langle \rho^{(n)} p^\alpha \rangle_T \in L_1(\mathbb{R}^{(n-1)\nu}),$$

for $n = 1,2,3,4$, $\alpha \in \mathbb{N}^{n\nu}$, with $\Sigma_i\, \alpha_i = |\alpha| \leqslant 2$, then

$$\lim_{R\to\infty} \frac{\omega(L_3(f_R)^2)}{\Sigma_\nu^2 R^{2\nu+2}}$$
$$= \frac{2}{\nu+2}\Big(\rho + \int dx(\langle \rho^{(2)}\rangle(x) - \rho^2)\Big)$$
$$\times\Big(\langle \rho^{(1)}(p_1)^2\rangle + \int dx \langle \rho^2 p_{1,1}\, p_{2,1}\,(x)\rangle\Big)$$
$$= \lim_{R\to\infty}\Big\{2\,\frac{\omega(Q_1^2)}{\Sigma_\nu R^{\nu+2}}\,\frac{\omega(P_1^2)}{\Sigma_\nu R^{\nu}}\Big\}. \qquad (10)$$

Proof: The proof of the theorem is divided into two parts, corresponding to the different terms in the expression (8) of $L_3(f_R)^2$. That means that we take together the terms which form the k-particle part in one of the three different contributions in $L_3(f_R)^2$, namely $([L_3(f_R)]_1)^2$, $[L_3(f_R)]_1 \times [L_3(f_R)]_2$ (and $[L_3(f_R)]_2\,[L_3(f_R)]_1$), and $([L_3(f_R)]_2)^2$. So we handle successively $[A_1^2]_1$, $[A_1^2]_2$, $[A_1A_2]_2$ and $[A_2A_1]_2$, $[A_2^2]_2$, $[A_1A_2]_3$ and $[A_2A_1]_3$, $[A_2^2]_3$, and $[A_2^2]_4$, where for each observable $A\in\mathscr{A}$, $[A]_i$ denotes the i-particle component of A, $A_1 \equiv [L_3(f_R)]_1$, $A_2 \equiv [L_3(f_R)]_2$.

Now we look at the different contributions.

(1) $E_1(R) \equiv (\Sigma_\nu^2 R^{2\nu+2})^{-1}\omega([\,[L_3(f_R)]_1^2\,]_1)$.

Because of the fact that all $\langle \rho^{(1)} p^\beta\rangle$ are constants, $E_1(R)$ is a sum of contributions of the form

$$\frac{C}{\Sigma_\nu^2 R^{2\nu+2}}\int d^\nu x\, x_i\, x_j\, g_R(x),$$

or

$$\frac{C}{\Sigma_\nu^2 R^{2\nu+2}}\int d^\nu x\, x_i g_R(x),$$

where C is a constant and $\forall x\in\mathbb{R}^\nu, |g_R(x)| \leqslant f_{R+1}(x)$. So we have clearly $\lim_{R\to\infty} E_1(R) = 0$.

(2) $E_2(R) \equiv (\Sigma_\nu^2 \mathbb{R}^{2\nu+2})^{-1}\,\omega(\,[\,[L_3(f_R)]_1^2\,]_2)$.

This contribution contains only terms of the form

$$\frac{C}{\Sigma_\nu^2 R^{2\nu+2}}\int d^\nu x_1\, d^\nu x_2\, x_{1,i} x_{2,j}$$
$$\times \langle \rho^{(2)} p_1^{\beta_1} p_2^{\beta_2}\rangle(x_1,x_2) g_R^1(x_1)\, g_R^2(x_2),$$

where the g_R^l are as in remark II.2 with $i\neq j$, β_1 and $\beta_2 \in \{0,1\}$. Application of Lemma II.1 yields factorization:

$$\frac{C}{\Sigma_\nu^2 R^{2\nu+2}}\int d^\nu x_1\, x_{1,i} g_R^1(x_1)\langle \rho^{(1)} p^{\beta_1}\rangle(x_1)$$
$$\times \int d^\nu x_2\, x_{2,i}\, g_R^2(x_2)\langle \rho^{(1)} p^{\beta_2}\rangle(x_2).$$

Now there are two possibilities: (i) $|\beta_1|$ or $|\beta_2| = 1$, then by rotation invariance (remark II.2) the expression above is vanishing for each R, or (ii) $\beta_1 = \beta_2 = 0$. This means that the observable is simply the tensor product of two functions of the position operator. Therefore in each factor a momentum operator must have disappeared. This happens only by commutation with a position observable as in the following example:

$$q_1 f(q) p_2 f(q) \otimes q_2 f(q)^2 p_2$$
$$= p_2 q_1 f(q)^2 \otimes p_1 q_2 f(q)^2 + i q_1 f(q)\partial_2 f(q) \otimes p_1 q_2 f(q)^2$$
$$+ p_2 q_1 f(q)^2 \otimes 2 i q_2 f(q)\partial_1 f(q)$$
$$- 2 q_1 f(q)\partial_2 f(q) \otimes q_2 f(q)\partial_1 f(q).$$

The last term is of the type $\beta_1 = \beta_2 = 0$. We see that in this case the functions $g_R^1(x)$ and $g_r^2(x)$ are zero for $\|x\| \leqslant R$. Hence

$$\lim_{R\to\infty} \frac{c}{\Sigma_\nu^2 R^{2\nu+2}}\rho^2 \int d^\nu x_1\, d^\nu x_2\, x_{1,i} x_{2,j}\, g_R^1(x_2)\, g_R^2\,(x_2) = 0.$$

(3) $E_3 R \equiv (\Sigma_\nu^2 R^{2\nu+2})^{-1}$
$\times\, \omega([[L_3(f_R)]_2 [L_3(f_R)]_1]_2)$.

Each term in this two-particle observable contains a one-particle factor $q_i f_R(q)$ or $p_j f_R(q)$ or $\partial_j f_R(q)$, which have an expectation zero for each R. By remark II.2 and rotation invariance, $E_3(R) = 0$. Clearly the same argument holds for $\omega([[L_3(f_R)]_1\,[L_3(f_R)]_2]_2)$.

(4) $E_4(R) \equiv (\Sigma_\nu^2 R^{2\nu+2})^{-1}$
$\times\, \omega([[L_3(f_R)]_2\,[L_3(f_R)]_1]_3)$.

This observable contains terms of the type

$$q_i f_R(q) \otimes (p_j f_R(q) + i\,\partial_j f_R(q)) \otimes p^\beta q_k g_R(q),$$

with $|\beta| \leqslant 1$, β never equals $(0,\ldots 0,1,0\ldots,0)$, the k^{th} unit vector in $\mathbb{N}^\nu$. Because of remark II.2 we only have to consider the case where $\beta = 0$ and $g_R(q) = f_R(q)\partial_l f_R(q)$. Then we use Lemma II.1 and, dropping the terms which are trivially zero, the only contributions left in $E_4(R)$ are of the type

$$\frac{C}{\Sigma_\nu^2 R^{2\nu+2}}\rho\int d^\nu x_1\, d^\nu x_2\, x_{1,i}\langle \rho^{(2)} p_{2,j}\rangle_T(x_1,x_2)$$
$$\times f_R(x_1) f_R(x_2)\int d^\nu x_3\, x_{3,k}\, f_R(x_3)\,\partial_l f_R(x_3)$$

and of the type

$$\frac{C}{\Sigma_\nu^2 R^{2\nu+2}}\rho\int d^\nu x_1\, d^\nu x_2\, x_{1,i}\langle \rho^{(2)}\rangle_T(x_1,x_2) f_R(x_1)$$
$$\times \partial_j f_R(x_2)\int d^\nu x_3\, x_{3,k}\, f_R(x_3)\partial_l f_R(x_3).$$

Both expressions tend to zero when R is tending to infinity because

$$\int d^\nu x_3\, x_{3,k}\, f_R(x_3)\partial_l f_R(x_3) \simeq O(R^\nu)$$

and

$$\int d^\nu x_1\, d^\nu x_2\, x_{1,i}\, f_R(x_1)\partial_j f_R(x_2) S(x_1,x_2) \simeq O(R^{\nu+1}),$$

where $S(x_1,x_2) = S(x_1 - x_2, 0) \in L_1(\mathbb{R}^\nu)$. The last result is based on a dominated convergence argument. Of course, the observable $[[L_3(f_R)]_1\,[L_3(f_R)]_2]_3$ is treated in a completely analogous way.

Up to here we have found that all nonzero contributions must come from the part $[L_3(f_R)]_2^2$.

(5) $E_5(R) \equiv (\Sigma_\nu^2 R^{2\nu+2})^{-1}\,\omega(\,[\,[L_3(f_R)]_2^2\,]_2)$.

By Remark II.2 the remaining terms to consider are of the type

$$
\begin{aligned}
&q_i^2 f_R(q)^2 \otimes p_j^2 f_R(q)^2,\\
&q_i^2 f_R(q)^2 \otimes f_R(q)\partial_{jj}^2 f_R(q),\\
&q_i^2 f_R(q)^2 \otimes (\partial_j f_R(q))^2, \qquad (11)\\
&q_i f_R(q)\partial_j f_R(q) \otimes q_k f_R(q)\partial_l f_R(q),\\
&q_i f_R(q)\partial_j f_R(q) \otimes f_R(q)^2.
\end{aligned}
$$

We will calculate explicitly the expectation value of the observable (11), which gives a finite result, independent of i and j. It is easily seen that the other terms are of a lower order and hence vanish in the thermodynamic limit. The correct numerical coefficient is given by the product rule. One finds

$$
\lim_{R\to\infty} E_5(R) = \lim_{R\to\infty} \frac{2}{\Sigma_\nu^2 R^{2\nu+2}} \int d^\nu x_1\, d^\nu x_2 (x_{1,j})^2 f_R(x_1)^2 \times f_R(x_2)^2 \langle \rho^{(2)}(p_{2,j})^2\rangle(x_1,x_2).
$$

By Lemma II.1

$$
\lim_{R\to\infty} E_5(R) = 2\rho\langle \rho^{(1)}(p_j)^2\rangle \lim_{R\to\infty} \frac{1}{\Sigma_\nu R^\nu}\int d^\nu x_2 f_R(x_2)^2 \times \lim_{R\to\infty}\frac{1}{\Sigma_\nu R^{\nu+2}}\int d^\nu x_1 (x_{1,j})^2 f_R(x_1)^2
$$

or

$$
\lim_{R\to\infty} E_5(R) = [2/(\nu+2)]\,\rho\langle \rho^{(1)}(p_j)^2\rangle. \tag{12}
$$

(6) $E_6(R) \equiv (\Sigma_\nu^2 R^{2\nu+2})^{-1}\omega([[L_3(f_R)]_2^2]^3)$.

If we write this observable in the form (4), we obtain a large number of terms. But one proves without any problem that only the following four contribute:

$$
\left.\begin{aligned}
&q_1 f_R(q) \otimes p_2^2 f_R(q)^2 \otimes q_1 f_R(q),\\
&q_2 f_R(q) \otimes p_1^2 f_R(q)^2 \otimes q_2 f_R(q);
\end{aligned}\right\} \text{(a)}
$$

$$
\left.\begin{aligned}
&p_1 f_R(q) \otimes q_2^2 f_R(q)^2 \otimes p_1 f_R(q),\\
&p_2 f_R(q) \otimes q_1^2 f_R(q)^2 \otimes p_2 f_R(q).
\end{aligned}\right\} \text{(b)}
$$

The two terms of the type (a) give the same contribution a:

$$
a = \lim_R \frac{1}{\Sigma_\nu^2 R^{2\nu+2}} \int d^\nu x_1\, d^\nu x_2\, d^\nu x_3 \times x_{1,i} x_{3,i} f_R(x_1) f_R(x_2) f_R(x_3) \times \langle \rho^{(3)}(p_{2,j})^2\rangle(x_1,x_2,x_3).
$$

Again by the lemma and remark II.2 we get

$$
a = \lim_R \frac{1}{\Sigma_\nu R^{\nu+2}} \int d^\nu x_1\, d^\nu x_3 \times x_{1,i} x_{3,i} f_R(x_1) f_R(x_3)\langle \rho^{(2)}\rangle_T(x_1,x_3) \times \lim_R \frac{1}{\Sigma_\nu R^\nu}\int d^\nu x_2\, f_R(x_2)^2\langle \rho^{(1)}(p_j)^2\rangle.
$$

Using translation invariance, the substitution $x_2 = x + x_3$, and finally the dominated convergence theorem, the first integral becomes, after taking the limit $R\to\infty$,

$$
\frac{1}{\nu+2}\int d^\nu x \langle \rho^2\rangle(x,0).
$$

The second integral is trivial, and the two terms in (a) yield together

$$
\frac{2}{\nu+2}\langle \rho^{(1)}(p_j)^2\rangle \int d^\nu x \langle \rho^2\rangle_T(x,0). \tag{13}
$$

A similar calculation of the contributions (b) leads to

$$
\frac{2}{\nu+2}\rho\int d^\nu x\,\langle \rho^{(2)} p_{1,j} p_{2,j}\rangle_T(x,0). \tag{14}
$$

(7) $E_7(R) \equiv (\Sigma_\nu^2 R^{2\nu+2})^{-1}\omega([[L_3(f_R)]_2^2]_4)$.

This observable is a sum of terms of the form

$$
A_{ijkl} \equiv \frac{4!}{2!2!2!2!} q_i f_R(q) \otimes p_k f_R(q) \otimes q_j f_R(q) \otimes p_l f_R(q),
$$

or such an expression where one or both of the factors $p_k f_R(q)$ are replaced by $\partial_k f_R(q)$. It will be clear from the calculation below that the latter type gives a lower-order contribution, vanishing if R tends to infinity. Each one-particle factor in A_{ijkl} has zero expectation value, so, using the lemma and the rotation invariance of the state, we get

$$
\omega(A_{ijkl}) = \frac{1}{16}\int d^\nu x_1\, d^\nu x_3 \langle \rho^{(2)}\rangle_T(x_1,x_3) \times x_{1,i} x_{3,j} f_R(x_1) f_R(x_3) \times \int d^\nu x_2\, d^\nu x_4 \langle \rho^{(2)} p_{1,k} p_{2,l}\rangle_T(x_2,x_4) f_R(x_2) f_R(x_4).
$$

If we use once again translation and rotation invariance, the cluster condition, and dominated convergence, we get

$$
\lim_{R\to\infty} (\Sigma_\nu^2 R^{2\nu+2})^{-1}\omega(A_{ijkl}) = \frac{1}{16}\delta_{ij}\delta_{kl}\frac{1}{\nu+2}\int d^\nu x \langle \rho^{(2)}\rangle_T(x,0) \times \int d^\nu y \langle \rho^{(2)} p_{1,k} p_{2,k}\rangle.
$$

Due to the special form of the angular momentum L_3 we did not have to consider terms with $k=l$, but $i\neq j$. If one works out the product formula (3) for $[[L_3(f_R)]_2^2]_4$, one finds 32 terms such that $\delta_{ij}\delta_{kl}=1$. Hence the total four-particle contribution of $L_3(f_R)^2$ is

$$
\lim_{R\to\infty} E_7(R) = \frac{2}{\nu+2}\int d^\nu x \langle \rho^{(2)}\rangle_T(x,0) \times \int d^\nu y \langle \rho^{(2)} p_{1,i} p_{2,i}\rangle(y,0). \tag{15}
$$

Now we take together the nonvanishing contributions (12) to (15) and we find the expression

$$
\lim_{R\to\infty} [\omega(L_3(f_R)^2)/\Sigma_\nu^2 R^{2\nu+2}] = \frac{2}{\nu+2}\left\{\int\rho + \int d^\nu x \langle \rho^{(2)}\rangle_T(x,0)\right\}\left\{\langle \rho^{(1)}(p_1)^2\rangle + \int d^\nu x \langle \rho^{(2)} p_{1,1}\, p_{2,1}\rangle_T(x,0)\right\}.
$$

This is the first equality in expression (10). By application of the definition of the truncated correlation functions (9) and the rotation invariance of the state, we obtain

$$
\begin{aligned}
&\langle \rho^{(2)}\rangle_T = \langle \rho^{(2)}\rangle(x,0) - \rho^2,\\
&\langle \rho^{(2)} p_{1,1}\, p_{2,1}\rangle_T(x,0) = \langle \rho^2{}_{1,1}\, p_{2,1}\rangle(x,0).
\end{aligned}
$$

The second equality in expression (10) is a direct consequence of the following identities (see Ref. 3, Lemma 3.1):

$$\lim_{R\to\infty} \frac{\omega(P_i(f_R)^2) - \omega(P_i(f_R))^2}{\Sigma_\nu R^\nu}$$

$$= \langle \rho^{(1)}(p_i)^2\rangle + \int d^\nu x \langle \rho^{(2)} p_{1,i} p_{2,i}\rangle(x),$$

$$(\nu+2) \lim_{R\to\infty} \frac{\omega(Q_i(f_R)^2) - \omega(Q_i(f_R))^2}{\Sigma_\nu R^{\nu+2}}$$

$$= \rho + \int d^\nu x (\langle \rho^{(2)}\rangle(x) - \rho^2). \quad \blacksquare$$

Theorem III.2: If ω satisfies the conditions of Theorem III.1 and if ω is an equilibrium state for a two-particle interaction V satisfying (i) $x\in\mathbb{R}^\nu \to V(x)\in\mathbb{R}$ is differentiable a.e., (ii) $V(x) = V(-x)$ and rotation symmetric, and (iii) there exists an $\eta > 0$ such that

$$\int dx (1 + \|x\|^\eta)\|\nabla V(x)\| < \infty,$$

then

$$\lim_{R\to\infty} \frac{\omega(L_3(f_R)^2)}{2\omega(Q_1(f_R)^2)\Sigma_\nu R^\nu} = \rho kT. \tag{16}$$

Proof: By our previous result

$$\lim_{R\to\infty} \frac{\omega(L_3(f_R)^2)}{2\omega(Q_1(f_R)^2)\Sigma_\nu R^\nu}$$

$$= \lim_R \frac{\omega(P_1(f_R)^2)}{\Sigma_\nu R^\nu}$$

$$= \langle \rho^{(1)}(P_1)^2\rangle + \int dx\, \rho^2 \langle P_{1,1} P_{2,1}\rangle(x,0).$$

Under the conditions as stated in this theorem the value of the right-hand side (the bulk momentum density fluctuations) is ρkT (see Ref. 3, Theorem 3.2). $\blacksquare$

IV. DISCUSSION

To learn something more about the content of the L_1-clustering conditions it is interesting to investigate systems where a phase transition or a spontaneous symmetry breaking occurs. It is known that a $|x|^{-1}$ behavior of the correlation function $\langle \rho^{(2)}\rangle(x,0) - \rho^2$ (which is not integrable) is closely related to the breaking of a continuous symmetry (see Refs. 4 and 5), e.g., this is the case of a free Bose gas with condensate. There the density fluctuations are given by

$$\langle \rho^{(2)}\rangle(x,0) - \rho^2 = \rho_0(\pi\beta\|x\|)^{-1} + O(\|x\|^{-2}),$$

where ρ_0 denotes the density of the condensed phase. A first consequence of this weak clustering in the condensed phase is a divergence in the position fluctuations. A physical explanation of this phenomenon is given further. Also, the momentum fluctuations lose their classical behavior. One calculates that, for the free Bose gas,

$$\langle \rho^{(1)} p_1^2\rangle + \int dx \langle \rho^2 p_{1,1} p_{2,1}\rangle(x,0)$$

$$= \int dp\, p_1^2 (1 + (2\pi)^3 f_\beta(p)) f_\beta(p) = \rho_c kT, \tag{17}$$

where $f_\beta(p) = (2\pi)^{-3}(\exp(\beta p^2/2) - 1)^{-1}$ and ρ_c is the critical density given by

$$\rho_c = \int dp\, f_\beta(p). \tag{18}$$

For a free Bose gas in the condensation region, we have $\rho = \rho_c + \rho_0$. It means that if the condensate density is not zero, $\rho_c kT$ is strictly smaller than the classical value ρkT. The particles in the condensed phase (with momentum zero) do not contribute to the momentum fluctuations.

Another consequence of the $\|x\|^{-1}$ behavior of the density fluctuations is that $\langle L_3^2\rangle$ does not factorize because of the appearance of nonvanishing surface contributions. These terms depend on the type of cutoff functions f_R in the local approximations of the observables, which is equivalent to a particular choice of boundary conditions.

From a kinematical point of view, the appearance of quantized vortices is another way of violating the classical behavior. Equilibrium states for the ideal boson gas, which yield this specific quantum effect, are rigorously calculated in Ref. 2. As pointed out by Hein and Roepstorff,[6] one can obtain these states as limit Gibbs states, also by choosing appropriate boundary conditions. In their paper they study the structure of the vortices in elaborate examples. The particular type of boundary conditions needed causes a spontaneous breaking of the translational invariance and therefore these states actually do not fit into our treatment. In fact, without translational invariance it is no longer possible to prove factorization by the methods used.

Finally we discuss the relation between a divergence in the position or angular momentum fluctuations and an infinite compressibility. Thus we will obtain an explanation for the nonclassical behavior of the angular momentum of the boson gas with condensate, in terms of a thermodynamical quantity. Using the definition relation of the isothermal compressibility χ,

$$\rho + \int dx (\langle \rho^{(2)}\rangle(x) - \rho^2) = \rho^2 \beta^{-1}\chi,$$

formula (16) can be rewritten as

$$\lim_{R\to\infty} \frac{\omega(L_3(f_R)^2)}{\Sigma_\nu^2 R^{2\nu+2}} = \rho^3 \beta^{-2}\chi. \tag{19}$$

Starting from the thermodynamical definition of χ given by

$$\frac{1}{\chi} = -V\left(\frac{\partial P}{\partial V}\right)_{T,N},$$

where V is the volume of the system, P the pressure, T the absolute temperature, and N the number of particles, one calculates

$$\frac{1}{\chi} = \rho^2\left(\frac{\partial^2 f}{\partial p^2}\right)_{T,N},$$

where f denotes the free-energy density. The condensate of a free Bose gas has energy density and entropy density zero. Hence the free energy density is constant for the free Bose gas with condensate, because ρ_c is only a function of the temperature [see (17) and (18)]. Hence χ is infinite.

ACKNOWLEDGMENT

I would like to thank Professor A. Verbeure for introducing me to the problem and for many stimulating discussions.

[1] J. M. Blatt, S. T. Butler, and M. R. Schafroth, Phys. Rev. **100**, 481 (1955).
[2] J. T. Lewis and J. V. Pulé, Commun. Math. Phys. **45**, 115 (1975).
[3] M. Fannes, Ph. Martin, and A. Verbeure, J. Phys. A: Math. Gen. **16**, 4293 (1983).
[4] Ph. Martin, Nuovo Cimento B **64**, 302 (1982).
[5] M. Fannes, J. V. Pulé, and A. Verbeure, Lett. Math. Phys. **6**, 385 (1982).
[6] S. Hein and G. Roepstorff, Ann. Inst. H. Poincaré A **32**, 21 (1980).

Quantum statistical derivation of the hydrodynamics of a plasma model

Geoffrey L. Sewell
Department of Physics, Queen Mary College, London E1 4NS, England

(Received 5 November 1984; accepted for publication 1 March 1985)

We derive macroscopic electrohydrodynamical equations of Euler and Maxwell from the many-particle Schrödinger equation of the Jellium model, subject to viable initial conditions, together with certain simple assumptions of macroscopic regularity. The dynamics of the model becomes tractable, on a suitable macroscopic scale and in the limit where its size becomes infinite, because of simple scaling properties and the long range of the Coulomb forces. Our derivation of its phenomenological dynamics is obtained via that of the classical Vlasov equation.

I. INTRODUCTION

A prime objective of the quantum theory of macroscopic phenomena is the derivation of phenomenological laws of continuum mechanics, such as those of hydrodynamics of heat conduction, from the Schrödinger equation for large assemblies of particles. Since such phenomenological laws generally have radically different structures from the underlying quantum mechanical ones, it is evident that the passage from a microscopic to a macroscopic dynamical description is fraught with serious conceptual and mathematical problems. It is therefore not surprising that rigorous derivations of phenomenological continuum mechanical laws from the quantum theory of many-particle systems are very scanty. As positive examples, we cite the derivations, by Davies,[1] of Fourier's law of heat conduction of a somewhat rudimentary model and, by Narnhofer and Sewell,[2] of the Vlasov kinetic equation for a certain class of models, including a caricature of a system of gravitational fermions. At the classical level, there are also some derivations of heat conduction and hydrodynamical equations for certain special models.[3,4]

The object of the present article is to derive the hydrodynamics of the Jellium plasma model from its many-particle Schrödinger equation; a preliminary sketch of this work was given in a research note.[5] The model, which we denote by Σ_N, consists of N electrons in a box Δ_L, of side L, with uniform passive background of positive charge, which neutralizes that of the electrons. This model is of genuine physical interest, since the particles interact via Coulomb forces. We remark here that a quantum, rather than classical, treatment of Coulomb systems is highly desirable, since their very stability generally depends on both the Heisenberg and the Pauli principles.[6,7]

We shall be concerned here with the phenomenological dynamics of Σ_N in the limit where N and L tend to infinity in such a way that the particle number, namely

$$\bar{n} = N/L^3, \tag{1}$$

remains fixed and finite. The model has already been shown[8] to have good thermodynamical properties in this limit. Here we find that the large-scale dynamical properties of the model are quite tractable, because of the long-range and simple scaling properties of the Coulomb forces. The principal result we obtain is that, in the above limit and in a certain specified scaling, the local normalized electron density σ, the drift velocity u, and the internal electric field E, all classical functions of positions and time, evolve according to the following equations of Euler and Maxwell:

$$\frac{\partial\sigma}{\partial t} + \operatorname{div}(\sigma u) = 0, \tag{2a}$$

$$\frac{\partial u}{\partial t} + (u\cdot\nabla)u = E, \tag{2b}$$

and

$$\operatorname{div} E = \sigma - 1. \tag{2c}$$

These equations will be derived from the Schrödinger equation for a *pure* state, subject to certain initial conditions; we remark that it is quite unnecessary to introduce mixed states as the passage from the microscopic to the macroscopic description provides all the averaging we need.

In assessing the significance of the result given by Eqs. (2) it should be noted that neither viscosity nor pressure gradient appears there; the essential reason is that they are eliminated by the scaling and limiting procedures we employ. Consequently, the present work should not be regarded as more than a skeletal version of a quantum theory of plasma hydrodynamics, which needs to be supplemented by a finer analysis to accommodate pressure gradients and viscosity. On the positive side, Eqs. (2) do cover the classical oscillations of a plasma, since a linearized version of them yields the result that

$$\frac{\partial^2\sigma}{\partial t^2} + \sigma = 1.$$

Let us now outline our method for passing from the Schrödinger equation for the Jellium model Σ_N to the phenomenological laws given by (2). For simplicity, we treat the particles as spinless, and assume periodic boundary conditions, as these are convenient for a treatment of flow properties. Thus, at the microscopic level, the pure states of Σ_N are given by the normalized, antisymmetric wave functions $\Psi^{(N)}$ of the particle position vectors $X_1, X_2, \dots, X_N$, possessing the periodicity of the box Δ_L in each of these variables. The Hamiltonian for Σ_N takes the following standard form[9]:

$$H_N = \frac{\hbar^2}{2m}\sum_{j=1}^{N} \Delta_{X_j} + e^2 \sum_{\substack{j,k=1\\ j<k}}^{N} V^{(L)}(X_j - X_k), \tag{3}$$

where m and e are the electron mass and charge, respectively,

 0022-2488/85/092324-11$02.50

$$V^{(L)}(X)=\frac{4\pi}{L^3}\sum_{Q\neq 0}^{(L)}\frac{e^{iQ\cdot X}}{Q^2},\tag{4}$$

and the superscript (L) over Σ indicates that summation is taken over vectors $Q=(2\pi/L)(n_1,n_2,n_3)$, with the n's integers. Thus, $V^{(L)}(X)$ is a periodicized version of the difference between $|X|^{-1}$ and its spatial average over Δ_L. The time-dependent Schrödinger equation for Σ_N, with T the time variable, is

$$i\hbar\frac{\partial\Psi_T^{(N)}}{\partial T}=H_N\Psi_T^{(N)}.\tag{5}$$

We base our macroscopic description of Σ_N on scales of length and time given, respectively, by L, the side of the box Δ_L, and ω^{-1}, the inverse of the classical plasma frequency

$$\omega=(4\pi\bar{n}e^2/m)^{1/2}.\tag{6}$$

Correspondingly, we take the macroscopic scale for the momentum per particle to be $mL\omega$. We designate the hydrodynamical observables $A^{(N)}$ to be the intensive variables given by N^{-1} times the additive functions of the positions and momenta of the particles, on the scales we have just specified. These observables thus take the form

$$A^{(N)}=N^{-1}\sum_{j=1}^{N}A\left(\frac{X_j}{L},\frac{P_j}{mL\omega}\right),\quad\text{with } P_j=-i\hbar\nabla_{X_j},\tag{7}$$

where the operator function A is defined as the Fourier transform of a c-number function $\hat{A}(\xi,\eta)$ according to the formula

$$\begin{aligned}A(X_j/L,P_j/mL\omega)&=\sum_\eta{}'\int d\xi\,\hat{A}(\xi,\eta)\exp\left(\frac{i\xi\cdot P_j}{2mL\omega}\right)\\&\quad\times\exp\left(\frac{i\eta\cdot X_j}{L}\right)\exp\left(\frac{i\xi\cdot P_j}{2mL\omega}\right),\end{aligned}\tag{8}$$

the prime over Σ indicating that summation is taken over the values of η given by $2\pi(n_1,n_2,n_3)$, with the n's integers. Thus, η is a discrete variable. It will generally be assumed that $\hat{A}(\xi,\eta)$ is continuous in ξ, and that $\Sigma'_\eta\int d\xi\,|\hat{A}(\xi,\eta)|$ is finite, so that the right-hand side of (8) is well defined.

We center our considerations on the time-dependent expectation values of the hydrodynamical observables $A^{(N)}$ corresponding to the evolution of Σ_N from a certain class of initial states, which we shall presently specify. Here our principal objective is to show that these expectation values are governed by a classical distribution function of position and velocity, whose form is determined by a normalized particle density σ, a local drift velocity u, and an internal electric field E, satisfying the phenomenological equations (2). Now, in general, the time-dependent expectation value of $A^{(N)}$ for evolution from an initial state $\Psi^{(N)}$ is given, at a time t on the ω^{-1} scale, by

$$\begin{aligned}\langle A^{(N)}\rangle_t&=(\Psi^{(N)}_{\omega^{-1}t},A^{(N)}\Psi^{(N)}_{\omega^{-1}t}),\\&\quad\text{with } \Psi_0^{(N)}\equiv\Psi^{(N)},\end{aligned}\tag{9}$$

the t dependence of $\Psi^{(N)}_{\omega^{-1}t}$ being determined by the Schrödinger equation (5). In view of the indistinguishability of the electrons, it follows from Eqs. (7)–(9) that the time-dependent expectation values of the hydrodynamical observables are determined by the one-particle *characteristic function*

$$\begin{aligned}\mu_t^{(N,1)}(\xi,\eta)&=\left(\Psi^{(N)}_{\omega^{-1}t},\exp\left(\frac{i\xi\cdot P_1}{2mL\omega}\right)\exp\left(\frac{i\eta\cdot X_1}{L}\right)\right.\\&\quad\left.\times\exp\left(\frac{i\xi\cdot P_1}{2mL\omega}\right)\Psi^{(N)}_{\omega^{-1}t}\right),\end{aligned}\tag{10}$$

according to the formula

$$\langle A^{(N)}\rangle_t=\sum_\eta{}'\int d\xi\,\hat{A}(\xi,\eta)\mu_t^{(N,1)}(\xi,\eta).\tag{11}$$

The $\mu_t^{(N,1)}$ is, in fact, the Fourier transform of the one-particle Wigner distribution function, in the given scaling, and thus corresponds to the one-particle density matrix of the system. We note that it also follows from (7)–(9) that the time-dependent expectation values of the products of the hydrodynamical observables $A^{(N)}$ are determined by the n-particle characteristic functions $\mu_t^{(N,n)}$ defined by the formula

$$\begin{aligned}&\mu_t^{(N,n)}(\xi_1,\ldots,\xi_n;\eta_1,\ldots,\eta_n)\\&\quad=\left(\Psi^{(N)}_{\omega^{-1}t},\exp\left(\frac{i}{2}\sum_{j=1}^{n}\frac{\xi_j\cdot P_j}{mL\omega}\right)\exp\left(i\sum_{j=1}^{n}\frac{\eta_j\cdot X_j}{L}\right)\right.\\&\quad\quad\left.\times\exp\left(\frac{i}{2}\sum_{j=1}^{n}\frac{\xi_j\cdot P_j}{mL\omega}\right)\Psi^{(N)}_{\omega^{-1}t}\right).\end{aligned}\tag{12}$$

Our aim is to obtain the dynamical properties of the key function $\mu_t^{(N,1)}$, in the limit $N\to\infty$, subject to the following conditions (1)–(3), on the initial state $\Psi^{(N)}$—or more correctly, on the initial states $\{\Psi^{(N)}\}$ of a sequence of systems $\{\Sigma_N\}$. We shall show in the Appendix that these conditions are perfectly viable by constructing states which fulfill them.

(1) The expectation value of the kinetic energy $T^{(N)}$ of Σ_N, for the state $\Psi^{(N)}$, is less than some finite constant B times N, for N sufficiently large, i.e.,

$$(\Psi^{(N)},T^{(N)}\Psi^{(N)})<BN.\tag{13}$$

(2) The expectation value of the potential energy $V^{(N)}$ of Σ_N, for the state $\Psi^{(N)}$, is less than some finite constant C times $N^{5/3}$, for N sufficiently large, i.e.,

$$(\Psi^{(N)},V^{(N)}\Psi^{(N)})<CN^{5/3}.\tag{14}$$

This bound corresponds to the electrostatic energy of a continuous charge distribution for which the density is some smooth function of X/L. Thus, the bounds given by (13) and (14) correspond to an initial state in which the local densities of charge and kinetic energy are intensive variables, given by functions of X/L.

(3) The hydrodynamical observables $A^{(N)}$ become uncorrelated, for the state $\Psi^{(N)}$; in the limit $N\to\infty$, i.e.,

$$\begin{aligned}&(\Psi^{(N)},A_1^{(N)}\ldots A_n^{(N)}\Psi^{(N)})-\prod_{j=1}^{n}(\Psi^{(N)},A_j^{(N)}\Psi^{(N)})\\&\quad\to 0,\quad\text{as } N\to\infty.\end{aligned}$$

This is a characteristic property of a pure thermodynamic phase since it signifies that the global intensive variables are sharply defined, i.e., dispersion free, in the limit $N\to\infty$ (cf. Refs. 10). It is evident from Eqs. (7), (8), and (12) that this property is equivalent to the following one for the characteristic functions:

$$\begin{aligned}&\mu_t^{(N,n)}(\xi_1,\ldots,\xi_n;\eta_1,\ldots,\eta_n)-\prod_{1}^{n}\mu_t^{(N,1)}(\xi_j,\eta_j)\\&\quad\to 0,\quad\text{as } N\to\infty.\end{aligned}\tag{15}$$

The problem, then, is to obtain the properties of $\mu_t^{(n,1)}$ in the limit $N\to\infty$, subject to the initial conditions (1)–(3). Since the macroscopic length scale is L, it is natural to reformulate this problem by dint of a scale transformation $X\to x = X/L$, which maps Σ_N into a system Σ'_N of N particles in a unit cube Δ_1. This transformation is particularly suited to hydrodynamical purposes, since a system of N particles in a fixed and finite volume simulates a continuous distribution of matter when N tends to infinity. Under this scale transformation $\Psi^{(N)}$ and $\Psi^{(N)}_{\omega^{-1}t}$ are mapped into states $\psi^{(N)}$ and $\psi_t^{(N)}$, respectively, of Σ'_N, according to the formula

$$\psi_t^{(N)}(x_1,\dots,x_N) = L^{3N/2}\Psi^{(N)}_{\omega^{-1}t}(Lx_1,\dots,Lx_N), \quad \text{with } \psi_0^{(N)} \equiv \psi^{(N)}. \tag{16}$$

Furthermore, the macroscopically scaled momentum operator $P_j/mL\omega$ $(\equiv -i\hbar\nabla_{X_j}/mL\omega)$ is mapped into

$$p_j = -i\hbar_N\nabla_j\,(\equiv -i\hbar_N\nabla_{x_j}), \tag{17}$$

where

$$\hbar_N = \frac{\hbar}{mL^2\omega} \equiv \frac{\hbar}{m\omega}\left(\frac{n}{N}\right)^{2/3} \tag{18}$$

is a dimensionless, effective Planck constant. By Eqs. (16) and (18), the Schrödinger equation (5) for Σ_N transforms to the following one for Σ'_N:

$$i\hbar_N\frac{\partial\psi_t^{(N)}}{\partial t} = H'_N\psi_t^{(N)}, \tag{19}$$

where

$$H'_N = -\frac{1}{2}\hbar_N^2\sum_{j=1}^{N}\Delta_j + N^{-1}\sum_{\substack{j,k=1\\ j<k}}^{N} V(x_j - x_k) \tag{20}$$

and

$$V(x) = \sum_{q=0}{}' e^{iq\cdot x}/q^2, \tag{21}$$

the prime over Σ having the same significance as in Eq. (8). Hence, by (19) and (2), Σ'_N is a model of N particles of unit mass, for which the potential energy of two particles separated by x is $N^{-1}V(x)$ and the effective Planck constant is $\hbar_N$. Moreover, by Eqs. (12) and (16)–(18), the characteristic function $\mu_t^{(N,n)}$ of Σ_N may be expressed in terms of the Σ'_N variables by the formula

$$\mu_t^{(N,n)}(\xi_1,\dots,\xi_n;\eta_1,\dots,\eta_n) = (\psi_t^{(N)}, W^{(n)}(\xi_1,\dots,\xi_n;\eta_1,\dots,\eta_n)\psi_t^{(N)}), \tag{22}$$

where

$$W^{(n)}(\xi_1,\dots,\xi_n;\eta_1,\dots,\eta_n) = \exp\left(\frac{i}{2}\sum_1^n \xi_j\cdot p_j\right)\exp\left(i\sum_1^n \eta_j\cdot x_j\right) \times\exp\left(\frac{i}{2}\sum_1^n \xi_j\cdot p_j\right). \tag{23}$$

Equivalently, $\mu_t^{(N,n)}$ is given by the equation

$$\mu_t^{(N,n)}(\xi_1,\dots,\xi_n;\eta_1,\dots,\eta_n) = \int dx_1\cdots dx_N\,\psi_t^{(N)*}\left(x_1 - \frac{1}{2}\hbar_N\,\xi_1,\dots,x_n - \frac{1}{2}\hbar_N\,\xi_n,x_{n+1},\dots,x_N\right)\exp i\left(\sum_1^n \eta_j\cdot x_j\right) \times\psi_t^{(N)}(x_1 + \tfrac{1}{2}\hbar_N\xi_1,\dots,x_n + \tfrac{1}{2}\hbar_N\,\xi_n,x_{n+1},\dots,x_N). \tag{22'}$$

One sees from (22) that the $\mu_t^{(N,n)}$'s are characteristic functions for Σ'_N, with "normal" scaling, and from (16), (17), and (23) that the initial conditions (1)–(3) for Σ_N imply the following ones for Σ'_N.

(i) The expectation value of the total kinetic energy of Σ'_N, for the state $\psi^{(N)}$, is less than a constant times $N^{1/3}$, and hence the total kinetic energy per particle of Σ'_N tends to zero as $N\to\infty$, i.e.,

$$(\psi^{(N)},p_1^2\psi^{(N)})\to 0, \quad \text{as } N\to\infty. \tag{24}$$

(ii) The expectation value of the total potential energy $V'^{(N)}$ of Σ'_N, for the state $\psi^{(N)}$, is less than some constant C' times N, for N sufficiently large, i.e.,

$$(\psi^{(N)},V'^{(N)}\psi^{(N)}) < C'N. \tag{25}$$

(iii) The characteristic functions $\mu_t^{(N,n)}$, at $t = 0$, satisfy the factorization condition, still given by Eq. (15).

We have thus recast the problem into that of obtaining the dynamical properties of $\mu_t^{(N,1)}$, considered now as a characteristic function of Σ'_N, subject to the initial conditions (i)–(iii). This problem is crucially simplified by the fact that (a) the effective Planck constant $\hbar_N$ tends to zero as $N\to\infty$, by (19), and (b) the potential energy of Σ'_N is of the form $N^{-1}\Sigma V(x_j - x_k)$, by (2). In fact, for the modified version of Σ'_N in which V is a suitably regular potential, it follows immediately from the theory of Ref. 2 that, under initial conditions covered by (i)–(iii), the properties (a) and (b) lead to classical and mean field theoretic limits, respectively, as $N\to\infty$, such that $\mu_t^{(N,1)}$ converges to the Fourier transform of a probability distribution $f_t(x,v)$ (more properly a measure) satisfying the Vlasov equation

$$\frac{\partial f_t(x,v)}{\partial t} + \frac{v\cdot\partial f_t(x,v)}{dx} - \int dx'\,dv' \times f_t(x',v')\nabla V(x - x')\cdot\frac{\partial f_t(x,v)}{\partial v} = 0. \tag{26}$$

In the present case, however, $V(x)$ is the Coulomb potential, which is singular at $x = 0$. This singularity is an essential part of the model Σ'_N and cannot justifiably be removed by the introduction of a cutoff at some "small" fixed distance α, since the corresponding cutoff distances for the original model Σ_N would then be $L\alpha$ and so would tend to infinity with L. In order to extend our derivation of the Vlasov equation to the Coulomb system Σ'_N, we invoke a supplementary assumption to the effect that the repulsive character of the interelectronic forces acts so as to keep the particles apart and so render the Coulomb singularity harmless. Specifically, we assume the uniform boundedness, over finite time intervals, of the two-particle probability density

$$\rho_t^{(N,2)}(x_1,x_2) = \int dx_3 \cdots dx_N \, |\psi_t(x_1,\ldots,x_N)|^2. \tag{27}$$

That is, we assume that for any finite T, there is a constant $K_T(<\infty)$ such that

$$\rho_t^{(N,2)}(x_1,x_2) < K_T, \quad \text{for } 0 \leqslant t \leqslant T. \tag{28}$$

This is really a restrictive condition on the initial state $\psi^{(N)}$, which excludes situations where, at $t = 0$, a "few" particles approach one another at sufficiently high velocities to overcome the Coulomb repulsion and so cause a divergence in the two-particle density at some later time.

Thus, we shall derive the Vlasov equation (26) from the Schrödinger equation (19), subject to the initial conditions (i)–(iii) and the supplementary assumption (28). In order to pass from the Vlasov dynamics to the Euler–Maxwell equations (2), we shall then exploit the condition (i), which tells us that the initial mean kinetic energy per particle of Σ'_N tends to zero as $N \to \infty$. This is a special feature of the rescaled description of the plasma model, which ensues from the boundedness of the kinetic energy per particle of Σ_N [cf. (i)]; it has no counterpart, for example, in the model of Ref. 2. Its significance here is that it implies (cf. Sec. II) that, in the limit $N \to \infty$, the initial one-particle distribution function takes the form

$$f_0(x,v) = \sigma_0(x)\delta(v). \tag{29}$$

From this one can infer (cf. Sec. IV) that, in view of the conservative character of the Vlasov dynamics, the velocity distribution at each point x remains dispersion-free at all times, i.e., f_t takes the form

$$f(x,v) = \sigma_t(x)\delta(x - u_t(x)). \tag{30}$$

The functions σ_t and u_t may therefore be identified with the normalized electron density and drift velocity, respectively, at time t. It is now a simple matter to derive the phenomenological equations from the Vlasov equation (26) and the formula (30) for f_t. For, on inserting (30) into (26) and taking the zeroth and first moments, respectively, with respect to v, we obtain Eqs. (2a) and (2b), with

$$E = E_t(x) = -\int dx' \, \nabla V(x - x')\sigma_t(x'). \tag{31}$$

Equation (2c) then follows from this formula for E and the definition (21) of V.

Note: Since the phenomenological equations (2) are invariant under the scale transformation $x \to kx$, $\sigma \to \sigma$, $u \to ku$, $E \to kE$, they are applicable to the original system Σ_N, as well as the rescaled one Σ'_N, in the limit $N \to \infty$. Further, the essential reason why neither a pressure gradient ∇p nor a viscosity term proportional to $\nabla^2 u$ appears in (2b) is that these terms scale as L^{-1}, when $x \to Lx$, whereas $\partial u/\partial t$, $(u \cdot \nabla)u$, and E all scale as L. In other words, pressure gradient and viscosity terms are "scaled away" by the present treatment.

We shall present our treatment of the model as follows. In Sec. II, we shall prove that, under the conditions (i)–(iii) and (28), the characteristic functions $\mu_t^{(N,n)}$ converge pointwise to classical ones, corresponding to Fourier transforms of probability measures $m_t^{(n)}$, as $N \to \infty$. In Sec. III we shall derive the so-called Vlasov hierarchy of equations for $\{m_t^{(n)}\}$ from the Schrödinger equation and the boundedness assumption (28). In Sec. IV, we shall derive the one-particle Vlasov equation from this hierarchy, subject to an additional assumption that the Coulomb singularity does not lead to turbulent motion. In Sec. V, we shall pass from the Vlasov equation to the phenomenological ones (2) along the lines indicated above, subject to a further assumption of macroscopic regularity. We shall conclude, in Sec. VI, with some brief comments on the method and result obtained here.

II. THE CLASSICAL LIMIT

We shall prove the convergence of $\mu_t^{(N,n)}$ to a classical characteristic function in three stages. First, we shall show that the kinetic energy per particle of Σ'_N is uniformly bounded, with respect to both N and the time t. Second, we shall use this result to establish the uniform continuity of $\mu_t^{(N,n)}(\xi,\eta)$ with respect to ξ and hence, by the Arzela–Ascoli theorem, the pointwise convergence of this function over some subsequence of values of N. Third, we shall employ Bochner's theorem, allied to the convergence of $\hbar_N$ to zero as $N \to \infty$, to show that the limiting form of $\mu_t^{(N,n)}$ is classical.

A. Bound on the kinetic energy per particle

By (i), (ii), and the conservation of energy of Σ'_N, the time-dependent expectation value of the sum of the kinetic and potential energies $T'^{(N)}$ and $V'^{(N)}$ is majorized by N times some finite constant C'', i.e.,

$$(\psi_t^{(N)}, T'^{(N)}\psi_t^{(N)}) + (\psi_t^{(N)}, V'^{(N)}\psi_t^{(N)}) < C''N. \tag{32}$$

Further, it follows from the stability of neutral Coulomb systems that the ground energy level of the original model Σ_N exceeds $D_1 N$, with D_1 a finite constant (cf. Refs. 6 and 7). Therefore, since by Eqs. (3) and (20), $H'_N = H_N/mL^2\omega$, up to a unitary transformation, the ground energy level of Σ'_N exceeds $N^{1/3}$ times a finite constant D_2. Since the same stability argument would still be applicable if the particle mass were doubled, it follows that

$$\tfrac{1}{2}(\psi_t^{(N)}, T'^{(N)}\psi_t^{(N)}) + (\psi_t^{(N)}, V'^{(N)}\psi_t^{(N)}) > D_3 N^{1/3},$$

for some finite constant D_3. From this inequality and (32) it follows that

$$(\psi_t^{(N)}, T'^{(N)}\psi_t^{(N)}) < 2C''N - 2D_3 N^{1/3},$$

and on dividing this formula by N, we see that

$$(\psi_t^{(N)}, p_1^2 \psi_t^{(N)}) < D, \quad \text{a finite constant}, \tag{33}$$

for all t and sufficiently large N.

B. Convergence of $\mu_t^{(N,n)}$

Since the variables η are discrete, it follows from the Arzela–Ascoli theorem that the pointwise convergence of $\mu_t^{(N,n)}$, as $N \to \infty$ over some sequence of integers, will be guaranteed if we can prove that $\partial\mu_t^{(N,n)}/\partial t$ and $\partial\mu_t^{(N,n)}/\partial\xi_j$ are uniformly bounded over finite ranges of values of $t, \xi_1,\ldots,\xi_n, \eta_1,\ldots,\eta_n$. For simplicity we shall confine our proofs to the derivatives of $\mu_t^{(N,1)}$; the corresponding results for the higher-order characteristic functions may be obtained similarly.

Now by (22) and (23)

$$\frac{\partial \mu_t^{(N,1)}}{\partial \xi}(\xi,\eta) = -(i/2)(p_1\psi_t^{(N)}, W^{(1)}(\xi,\eta)\psi_t^{(N)}) + (i/2)(\psi_t^{(N)}, W^{(1)}(\xi,\eta)p_1\psi_t^{(N)}).$$

Hence, by the unitarity of $W^{(1)}$ and the inequality (33)

$$\left|\frac{\partial \mu_t^{(N,1)}}{\partial \xi}(\xi,\eta)\right| \leqslant \| p_1\psi_t^{(N)}\| = (\psi_t^{(N)}, p_1^2\psi_t^{(N)})^{1/2} < D^{1/2}, \tag{34}$$

which proves the uniform boundedness of $\partial\mu_t^{(N,1)}/\partial\xi$. Further, by the Schrödinger equation (19) and the formula (22′), as applied to $\mu_t^{(N,1)}$,

$$\frac{\partial \mu_t^{(N,1)}}{\partial t} - \eta\frac{\partial \mu_t^{(N,1)}}{\partial \xi} = -i\int dx_1\cdots dx_N\, \psi_t^{(N)*}\left(x_1 - \frac{1}{2}\hbar_N\xi, x_2,\ldots,x_N\right)\exp(i\eta\cdot x_1)$$
$$\times \psi_t^{(N)}(x_1 + \tfrac{1}{2}\hbar_N\xi_1, x_2,\ldots,x_N)\hbar_N^{-1}(V(x_1 - x_2 + \tfrac{1}{2}\hbar_N\xi) - V(x_1 - x_2 - \tfrac{1}{2}\hbar_N\xi)). \tag{35}$$

By the Schwartz inequality $|\int AB|^2 \leqslant (\int|A|^2)(\int|B|^2)$, the right-hand side of this equation is majorized by $(\Phi_t^{(N)}(\xi)\Phi_t^{(N)}(-\xi))^{1/2}$, where

$$\Phi_t^{(N)}(\xi) = \int dx_1\cdots dx_N|\psi_t^{(N)}(x_1,\ldots,x_N)|^2 \times|[V(x_1 - x_2 + \hbar_N\xi) - V(x_1-x_2)]/\hbar_N|$$
$$= \int dx_1\,dx_2\,\rho^{(N,2)}(x_1,x_2) \times|[V(x_1 - x_2 + \hbar_N\xi) - V(x_1,x_2)]/\hbar_N|,$$

by (27). Since $\hbar_N\to\infty$ as $N\to\infty$ and $V(x)\sim|x|^{-1}$ for small x, it follows easily from this formula for $\Phi_t^{(N)}$ and the assumption (28) of Sec. I that $\Phi_t^{(N)}(\xi)$ is uniformly bounded over finite ranges of values of t and ξ and all sufficiently large N. Hence, so too is the rhs of (35). Consequently, it follows from that equation and (34) that $\partial\mu_t^{(N,1)}/\partial\xi$ and $\partial\mu_t^{(N,1)}/\partial t$ are both uniformly bounded over finite ranges of t, ξ, and η and sufficiently large N. It follows then, from the Arzela–Ascoli theorem, that $\mu_t^{(N,1)}$ converges pointwise to a limit $\mu_t^{(1)}$, which is continuous in ξ and t, as N tends to infinity over some sequence of integers. Similarly,

$$\mu_t^{(N,n)}\to\mu_t^{(n)}, \quad \text{pointwise, for all } n\to N, \tag{36}$$

as $N\to\infty$ over some sequence of integers, $\mu_t^{(n)}$ being continuous in the ξ's and t.

C. Classical property of $\mu_t^{(n)}$

In view of the continuity property of $\mu_t^{(n)}$ that we have just observed, together with the fact that $\mu_t^{(n)}(0,\ldots,0;0,\ldots,0) = 1$, by (22) and (23), it follows from Bochner's theorem that in order to establish that $\mu_t^{(n)}$ is a classical characteristic function, it suffices to prove that

$$\sum_{r,s=1}^{l} \bar{c}_r c_s\,\mu_t^{(n)}(\xi_1^{(r)} - \xi_1^{(s)},\ldots,\xi_n^{(r)} - \xi_n^{(s)}; \eta_1^{(r)} - \eta_n^{(s)},\ldots,\eta_n^{(r)} - \eta_n^{(s)})\geqslant 0, \tag{37}$$

for arbitrary complex numbers c_r and values $(\xi_1^{(r)},\ldots,\xi_n^{(r)};\eta_1^{(r)},\ldots,\eta_n^{(r)})$ of $(\xi_1,\ldots,\xi_n;\eta_1,\ldots,\eta_n)$, for $r=1,\ldots,l$. To prove this, we note that since, by (17)

$$[x_j,p_k] = i\hbar_N\delta_{jk},$$

it follows that

$$W^{(n)}(\xi_1,\ldots,\xi_n;\eta_1,\ldots,\eta_n)W^{(n)}(\xi_1',\ldots,\xi_n';\eta_1',\ldots,\eta_n')$$
$$= W^{(n)}(\xi_1' - \xi_1,\ldots,\xi_n' - \xi_n;\eta_1' - \eta_1,\ldots,\eta_n' - \eta_n)\times\exp\left(i\frac{\hbar_N}{2}\sum_1^n(\xi_j'\cdot\eta_j - \xi_j\cdot\eta_j')\right). \tag{38}$$

Hence, by (22) and (23), the inequality

$$\left\|\sum_{r=1}^{l} c_l W^{(n)}(\xi_1^{(r)},\ldots,\xi_n^{(r)},\eta_1^{(r)},\ldots,\eta_n^{(r)})\psi_t^{(N)}\right\|^2\geqslant 0$$

reduces to the form

$$\sum_{r,s=1}^{l}\bar{c}_r c_s\,\mu_t^{(N,n)}(\xi_1^{(r)} - \xi_1^{(s)},\ldots,\xi_n^{(r)} - \xi_n^{(s)}; \eta_1^{(r)} - \eta_1^{(s)},\ldots,\eta_n^{(r)} - \eta_n^{(s)})\times\exp\left(i\frac{\hbar_N}{2}\sum_{j=1}^n(\xi_j^{(r)}\cdot\eta_j^{(s)} - \xi_j^{(s)}\cdot\eta_j^{(r)})\right)\geqslant 0.$$

Since by (18) and (23), $\hbar_N\to 0$ and $\mu_t^{(N,n)}\to\mu^{(n)}$ as $N\to\infty$, this last inequality reduces to the required form (37). We conclude therefore that $\mu_t^{(n)}$ is the characteristic function for a classical probability measure $m_t^{(n)}$, i.e.,

$$\mu_t^{(n)}(\xi_1,\ldots,\xi_n;\eta_1,\ldots,\eta_n) = \int\exp\left(i\sum_1^n(\xi_j\cdot v_j + \eta_j\cdot x_j)\right)\times dm_t^{(n)}(x_1,\ldots,x_n;v_n,\ldots,v_n). \tag{39}$$

Furthermore, since Eqs. (22)–(24) imply that

$$\mu_t^{(m+n)}(\xi_1,\ldots,\xi_n,0,\ldots,0;\eta_1,\ldots,\eta_n,0,\ldots,0) \equiv \mu_t^{(n)}(\xi_1,\ldots,\xi_n;\eta_1,\ldots,\eta_n),$$

it follows from (39) that $m_t^{(n)}$ is the restriction to $\Delta_1^n\times\mathbf{R}^{3n}$ of a probability measure m_t on $(\Delta_1\times\mathbf{R}^3)^\infty$, Δ_1 being the unit cube, as in Sec. I.

D. Bound on the two-particle density

We shall now show that the bound, given by (28), on the two-particle density of Σ_N' implies a corresponding bound for the infinite system.

For this purpose, we denote by $P_t^{(2)}$ the two-particle probability density induced by m_t, i.e.,

$$\int\phi(x_1,x_2)dP_t^{(2)}(x_1,x_2) = \int\phi(x_1,x_2)dm_t^{(2)}(x_1,x_2;v_1,v_2), \tag{40}$$

for bounded continuous functions ϕ. In particular, if the Fourier transform of ϕ, namely

$$\hat{\phi}(\eta_1,\eta_2)=\int dx_1\,dx_2 \exp - i(\eta_1\cdot x_1+\eta_2\cdot x_2)\phi(x_1,x_2), \tag{41}$$

is l^1 class, then it follows from (22), (27), (28), (36), and (39)–(41) that

$$\begin{aligned}\int \phi(x_1,x_2)dP_t^{(2)}(x_1,x_2) &= \sum_{\eta_1,\eta_2}{}' \hat{\phi}(\eta_1\cdot\eta_2)\mu_t^{(2)}(0,0;\eta_1,\eta_2)\\ &= \lim_{N\to\infty}\sum_{\eta_1,\eta_2}{}'\hat{\phi}(\eta_1,\eta_2)\mu_t^{(N,2)}(0,0;\eta_1,\eta_2)\\ &= \lim_{N\to\infty}\int \phi(x_1,x_2)\rho_t^{(N,2)}(x_1,x_2)dx_1\,dx_2\end{aligned}$$

and hence

$$\left|\int \phi(x_1,x_2)dP_t^{(2)}(x_1,x_2)\right| < K_T\int |\phi(x_1,x_2)|dx_1\,dx_2, \quad \text{for } 0\leqslant t\leqslant T. \tag{42}$$

Now this result has been obtained here for bounded continuous functions ϕ, whose Fourier transforms ϕ are l^1 class. Since these include the C^∞ class functions ϕ, it follows that (42) may be extended by continuity to all bounded continuous ones on $\Delta_1\times\Delta_1$. Consequently, $P_t^{(2)}$ is absolutely continuous with respect to the Lebesgue measure, and its density $\rho_t^{(2)}$ is majorized by K_T, i.e.,

$$dP_t^{(2)}(x_1,x_2)=\rho_t^{(2)}(x_1,x_2)dx_1\,dx_2 \tag{43}$$

and

$$\rho_t^{(2)}(x_1,x_2)<K_T \quad \text{(a.e.) for } 0\leqslant t\leqslant T. \tag{44}$$

Further, as the volume of Δ_1 is unity, it follows from this result that the one-particle probability density is also bounded by K_T, i.e.,

$$\rho_t^{(1)}(x)\equiv\int_{\Delta_1} dx'\,\rho_t^{(2)}(x,x')<K_T, \quad \text{for } 0\leqslant t\leqslant T. \tag{45}$$

E. The initial form of m

It follows immediately from Eqs. (15), (36), and (39) that the initial probability measure m_0 has the factorization property

$$dm_0^{(n)}(x_1,...,x_n;v_1,...,v_n)=\prod_{j=1}^n dm_0^{(1)}(x_j,v_j). \tag{46}$$

Hence, m_0 is completely determined by $m_0^{(1)}$. In order to show that this one-particle measure has the form given by (29) we observe that it follows from (34) that

$$\left|\frac{\partial\mu_0^{(N,1)}}{\partial\xi}\right|\leqslant(\psi^{(N)},p_1^2\psi^{(N)})^{1/2}.$$

Hence, by condition (i), $\partial\mu_0^{(N,1)}/\partial\xi\to 0$, uniformly with respect to ξ and η, as $N\to\infty$. Consequently, by (36), $\partial\mu_0^{(1)}/\partial\xi = 0$, i.e., $\mu_0^{(1)}$ is ξ independent. This implies that $m_0^{(1)}$ takes the form

$$dm_0^{(1)}(x,v)=dP^{(1)}(x)\delta(v)dv,$$

where $P^{(1)}$ is the one-particle spatial probability. Since, by (45) this corresponds to a density σ_0, say, it follows that

$$dm_0^{(1)}(x,v)=\sigma_0(x)\delta(v)dx\,dv, \tag{47}$$

which confirms the formula (29). We shall generally assume that the initial density σ_0 is strictly positive everywhere.

III. THE VLASOV HIERARCHY

We now aim to obtain the equation of motion for m_t. First we note that, for a modified version of Σ'_N, for which the potential V of Eq. (20) is suitably regular, it has already been established[2,11] that the limit measure m_t evolves according to the Vlasov hierarchy, namely

$$\begin{aligned}\frac{d}{dt}\int \phi^{(n)}(x_1,...,x_n;v_1,...,v_n)dm_t^{(n)}(x_1,...,x_n;v_1,...,v_n)\\ = \sum_{j=1}^n\int v_j\cdot\frac{\partial\phi^{(n)}}{\partial x_j}(x_1,...,x_n;v_1,...,v_n)\\ \times dm_t^{(n)}(x_1,...,x_n;v_1,...,v_n) - \sum_{j=1}^n\int \nabla V(x_j-x_{n+1})\\ \times\frac{\partial\phi^{(n)}}{\partial v_j}(x_1,...,x_n;v_1,...,v_n)\\ \times dm_t^{(n+1)}(x_1,...,x_{n+1};v_1,...,v_{n+1}),\end{aligned} \tag{48}$$

where the test functions $\phi^{(n)}$ are continuously differentiable, possess the periodicity of Δ_1 in the x's, and have compact support in the v's; we shall refer to these as the C_0^1-class functions. Our objective now is to derive the hierarchy (48) for the Coulomb system Σ'_N, subject to the uniform boundedness assumption on its two-particle density. We note here that, in view of the corresponding boundedness condition for the infinite system, it suffices, for the proof of (48), to establish the validity of that hierarchy for functions $\phi^{(n)}$, whose Fourier transforms, namely

$$\begin{aligned}\hat{\phi}^{(n)}(\xi_1,...,\xi_n;\eta_1,...,\eta_n)\\ =(2\pi)^{-3}\int dx_1\cdots dx_n\,dv_1\cdots dv_n\,\phi^{(n)}(x_1,...,x_n;v_1,...,v_n)\\ \times\exp i\sum_{j=1}^n(\xi_j\cdot x_j+\eta_j\cdot v_j),\end{aligned} \tag{49}$$

have compact support and are continuously differentiable with respect to the ξ's, since the extension to C_0^1-class functions can then be achieved by continuity.

Thus, we consider the time dependence of $\int\phi^{(n)}\,dm_t^{(n)}$ for functions $\phi^{(n)}$ whose Fourier transforms are continuously differentiable with respect to the ξ's and have compact support. Now, by (39) and (49),

$$\int\phi^{(n)}\,dm_t^{(n)}-\int\phi^{(n)}\,dm_0^{(n)}=\sum_\eta{}'\int d\xi\,\hat{\phi}^{(n)}(\mu_t^{(n)}-\mu_0^{(n)}),$$

where Σ'_η and $\int d\xi$ signify summation over $\eta_1,...,\eta_n$ and integration over $\xi_1,...,\xi_n$, respectively. Hence, by (36)

$$\begin{aligned}\int\phi^{(n)}\,dm_t^{(n)}-\int\phi^{(n)}\,dm_0^{(n)}\\ =\lim_{N\to\infty}\sum_\eta{}'\int d\xi\,\hat{\phi}^{(n)}(\mu_t^{(N,n)}-\mu_0^{(N,n)})\\ =\lim_{N\to\infty}\int_0^t ds\sum_\eta{}'\int d\xi\,\hat{\phi}^{(n)}\frac{\partial\mu_s^{(N,n)}}{\partial s}.\end{aligned}$$

On substituting into this equation the formula for $\partial\mu_s^{(N,n)}/\partial s$ obtained from the Schrödinger equation (19) and the definition (22′) of $\mu^{(N,n)}$, we find that

$$\int \phi^{(n)}\,dm_t - \int \phi^{(n)}\,dm_0 = -\lim_{N\to\infty}\int_0^t ds \sum_\eta{}' \int d\xi \sum_{j=1}^n \eta_j \cdot \frac{\partial\hat\phi^{(n)}}{\partial\xi_j}\mu_s^{(N,n)} \tag{50a}$$

$$+ \lim_{N\to\infty} N^{-1}\int_0^t ds \sum_\eta{}' \int d\xi\, \hat\phi^{(n)}(\xi_1,\ldots,\xi_n;\eta_1,\ldots,\eta_n)\int dx_1\cdot dx_N\, \psi_s^{(N)*}\Big(x_1 - \frac{1}{2}\hbar_N\,\xi_1,\ldots,x_n$$
$$-\frac{1}{2}\hbar_N\,\xi_n,x_{n+1},\ldots,x_N\Big)\psi_s^{(N)}\Big(x_1 + \frac{1}{2}\hbar_N\,\xi_1,\ldots,x_n + \frac{1}{2}\hbar_N\,\xi_n,x_{n+1},\ldots,x_N\Big)$$
$$\times\hbar_N^{-1}\sum_{j,k=1}^n\Big(V\Big(x_j - x_k - \frac{1}{2}\hbar_N(\xi_j-\xi_k)\Big) - V\Big(x_j - x_k + \frac{1}{2}\hbar_N(\xi_j-\xi_k)\Big)\Big) \tag{50b}$$

$$+ \lim_{N\to\infty}\Big\{\Big(1-\frac{n}{N}\Big)\int_0^t ds \sum_\eta{}' \int d\xi\, \hat\phi^{(n)}(\xi_1,\ldots,\xi_n;\eta_1,\ldots,\eta_n)\int dx_1\cdots dx_N\, \psi_s^{(N)*}\Big(x_1 - \frac{1}{2}\hbar_N\,\xi_1,\ldots,x_n$$
$$-\frac{1}{2}\hbar_N\,\xi_n,x_{n+1},\ldots,x_N\Big)\psi_s^{(N)}\Big(x_1 + \frac{1}{2}\hbar_N\,\xi_1,\ldots,x_n - \frac{1}{2}\hbar_N\,\xi_N,x_{n+1},\ldots,x_N\Big)\hbar_N^{-1}$$
$$\times\sum_{j=1}^n\Big(V\Big(x_j - x_{n+1} - \frac{1}{2}\hbar_N\,\xi_j\Big) - V\Big(x_j - x_{n+1} + \frac{1}{2}\hbar_N\,\xi_j\Big)\Big)\Big\}. \tag{50c}$$

Since $\hbar_N\to 0$ as $N\to\infty$, it follows easily from (36), (39), and (49) that

$$\text{Term (50a)} = -\int_0^t ds \sum_{j=1}^n \int v_j \cdot \frac{\partial\phi^{(n)}}{\partial x_j}\,dm_0^{(n)}. \tag{51a}$$

Further, by the argument used to prove the uniform boundedness of (35), it follows that the spatial integral in (50b) is also uniformly bounded, and therefore, in view of the factor N^{-1} in (49b),

$$\text{Term (50b)} = 0. \tag{51b}$$

The same argument may easily be employed to show that, if $V_g(x)$ is a continuously differentiable function that coincides with $V(x)$ for $|x|\geqslant g$ and whose gradient is less, in modulus, than c/g^2 for some constant c, then the whole term in curly brackets in (50c) is the limit, as $g\to 0$, of the corresponding term with V_g replacing V, the convergence to this limit being uniform in N. Hence,

Term (50c)

$$= \lim_{g\to 0}\lim_{N\to\infty}\Big(1-\frac{n}{N}\Big)\int_0^t ds$$
$$\times\sum_n{}'\int d\xi\,\hat\phi^{(n)}(\xi_1,\ldots,\xi_n;\eta_1,\ldots,\eta_n)$$
$$\times\int dx_1\cdots dx_N\,\psi_s^{(N)*}$$
$$\times(x_1 - \tfrac{1}{2}\hbar_N\,\xi_1,\ldots,x_n - \tfrac{1}{2}\hbar_N\,\xi_n,x_{n+1},\ldots,x_N)$$
$$\times\psi_s^{(N)}(x_1 + \tfrac{1}{2}\hbar_N\,\xi_1,\ldots,x_n + \tfrac{1}{2}\hbar_N\,\xi_n,x_{n+1},x_N)$$
$$\times\hbar_N^{-1}[V_g(x_j - x_{n+1} - \tfrac{1}{2}\hbar_N\,\xi_j)$$
$$- V_g(x_j - x_{n+1} + \tfrac{1}{2}\hbar_N\,\xi_j)].$$

Since V_g is bounded and continuously differentiable, it follows easily from (22′), (36), (39), and (49) that this last equation signifies that

Term (50c)

$$= \lim_{g\to 0}\int_0^t ds\sum_{j=1}^n\int -\nabla V_g(x_j - x_{n+1})\cdot\frac{\partial\phi^{(n)}}{\partial v_j}$$
$$\times dm_s^{(n+1)}(x_1,\ldots,x_{n+1};v_1,\ldots,v_{n+1}).$$

Hence, in view of our above specifications of V_g it follows from the uniform boundedness property (44) of the two-particle density that

Term (50c)

$$= \int_0^t ds\sum_{j=1}^n\int -\nabla V(x_j - x_{n+1})\cdot\frac{\partial\phi^{(n)}}{\partial v_j}\,dm_s^{(n+1)}. \tag{51c}$$

On using the formulas (51a)–(51c) for the terms (50a)–(50c), we see that

$$\int\phi^{(n)}\,dm_t^{(n)} - \int\phi^{(n)}\,dm_0^{(n)}$$
$$= \int_0^t ds\sum_{j=1}^n\Big(\Big(\int v_j\frac{\partial\phi^{(n)}}{\partial x_j}\,dm_s^{(n)}$$
$$- \int\nabla V(x_j - x_{n+1})\cdot\frac{\partial\phi}{\partial v_j}\,dm_s^{(n+1)}\Big),$$

which is equivalent to the Vlasov hierarchy (48).

IV. THE VLASOV DYNAMICS

The Vlasov hierarchy (48) is closely related to the single-particle Vlasov equation, namely

$$\frac{d}{dt}\int\phi(x,v)dm_t^{(1)}(x,v)$$
$$= \int v\cdot\frac{\partial\phi}{\partial x}(x,v)dm_t^{(1)}(x,v) - \int\nabla V(x,x')\cdot\frac{\partial\phi}{\partial v}(x,v)$$
$$\times dm_t^{(1)}(x,v)dm_t^{(1)}(x',v'). \tag{52}$$

For if $m_t^{(1)}$ is a solution of this latter equation, with initial value $m_0^{(1)}$, then it follows easily from (48) that the Vlasov hierarchy has a solution

$$dm_t^{(n)}(x_1,\ldots,x_n;v_1,\ldots,v_n) = \prod_{j=1}^n dm_t^{(1)}(x_j,v_j), \tag{53}$$

which satisfies the initial condition (46). Moreover, it follows from a slight adaptation of the treatment of the Vlasov equation (52) by Illner and Neunzert[12] that the latter equation does indeed have a solution, *though there is no proof of its* *uniqueness*. What we can say, then, is that the hierarchy (48) has a solution possessing the factorization property (53), with $m_t^{(1)}$ satisfying the one-particle Vlasov equation. Here we remark that a stronger statement could be made if, instead of the singular Coulomb potential V, one had a suitably regular one in the Vlasov hierarchy; for in that case it would follow from known results[2,11,13,14] that the solutions of both the Vlasov equation (52) and the hierarchy (52) were unique and related by the formula (53).

Thus, as we do not have general uniqueness theorems for the Coulomb systems, we need further assumptions of a physical nature in order to obtain solutions of the Vlasov dynamics. Our key assumption is that, as in the case of a regular potential, the factorization property (53) is preserved at all times. Physically, this is equivalent to assuming that the Coulomb singularity does not lead to turbulence, since the property (53) is the condition for the macroscopic observables $A^{(N)}$ of the model Σ_N to be dispersion-free at all times, in the limit $N\to\infty$, i.e.,

$$\langle A_1^{(N)}\cdots A_n^{(N)}\rangle_t - \prod_{j=1}^{n}\langle A_j^{(N)}\rangle_t \to 0, \quad \text{as } N\to\infty.$$

Assuming, then, the factorization property (53), it follows from Eq. (48), for $n=1$, that, in view of the boundedness of the two-particle density, $m_t^{(1)}$ satisfies the Vlasov equation (52). This, in turn, is the same as the Liouville equation

$$\frac{d}{dt}\int \phi(x,v)dm_t^{(1)}(x,v) = \int\left(v\cdot\frac{\partial\phi(x,v)}{\partial x} + E_t(x)\cdot\frac{\partial\phi(x,v)}{\partial v}\right)dm_t^{(1)}(x,v), \tag{54}$$

where the electric field $E_t(x)$ is defined by the self-consistency condition

$$E_t(x) = -\int \nabla V(x-x')dm_t^{(1)}(x',v'). \tag{55}$$

In a standard way,[14] the solution of the Liouville equation (54) can be expressed in terms of that of the one-particle equations

$$\frac{dx_t}{dt} = v_t, \quad \frac{dv_t}{dt} = E_t(x_t),$$
$$\text{with } x_0 = x \quad \text{and } v_0 = v, \tag{56}$$

provided that the latter equations have a unique solution. In fact, they will have it if the electric field $E_t(x)$ is sufficiently regular, e.g., if its spatial derivatives are uniformly bounded over finite time intervals. We shall assume that $E_t(x)$ satisfies this latter condition of macroscopic regularity—one which evidently excludes the occurrence of shock waves. Under this condition, then, (56) has a unique solution

$$x_t = X_t(x,v), \quad v_t = V_t(x,v), \tag{57}$$

the transformation $x,v\to x_t,v_t$ being canonical. Correspondingly, the solution of (54) takes the form[14]

$$\int \phi(x,v)dm_t^{(1)}(x,v) = \int \phi\,[X_t(x,v),V_t(x,v)]\,dm_0^{(1)}(x,v). \tag{58}$$

In view of the initial condition (47), this solution may be reexpressed as

$$\int \phi(x,v)dm_t^{(1)}(x,v) = \int \phi\,[\overline{X}_t(x),\overline{V}_t(x)]\,\sigma_0(x)dx, \tag{59}$$

where

$$\overline{X}_t(x) \equiv X_t(x,0) \quad \text{and } \overline{V}_t(x) \equiv V_t(x,0). \tag{60}$$

Equations (53) and (59) constitute the solution of the Vlasov hierarchy.

V. PHENOMENOLOGICAL EQUATIONS

As a first step towards deriving the phenomenological equations (2) from the Vlasov dynamics, we observe that, by (56) and (60), $\overline{X}_t(x)$ and $\overline{V}_t(x)$ correspond to the position and velocity of a "fluid particle" initially at rest at x, in the Lagrangian description of hydrodynamics (cf. Lamb[15]). To pass from this to a Eulerian description, we have first to establish that $\overline{X}_t(x)$ is an invertible function of x, i.e., that the Jacobian $\partial\overline{X}_t(x)/\partial x$ of the transformation from x to $\overline{X}_t(x)$ does not vanish. We shall now show that the nonvanishing of this Jacobian follows from the boundedness of the time-dependent density [cf. (45)] and the assumption, made in Sec. II, that the initial density $\sigma_0(x)$ is strictly positive. Thus we note that if τ is the volume of an infinitesimal fluid particle at x, at time $t=0$, then its volume at time t, when the particle has migrated to $\overline{X}_t(x)$, is $\tau|\partial\overline{X}_t(x)/\partial x|$. Consequently, as one can infer from (59), the density of the fluid at $\overline{X}_t(x)$ is $\sigma_0(x)|\partial\overline{X}_t(x)/\partial x|^{-1}$. Hence in view of the boundedness of the time-dependent density and the nonvanishing of $\sigma_0(x)$, it follows that the Jacobian $\partial\overline{X}_t(x)/\partial x$ cannot vanish, and consequently that the transformation $x\to\overline{X}_t(x)$ is invertible.

It therefore follows that the formula (59) for $m_t^{(1)}$ may be expressed in the form

$$\int \phi(x,v)dm_t^{(1)}(x,v) = \int \phi(x,u_t(x))\sigma_t(x)dx, \tag{61}$$

where

$$\sigma_t(x) = \sigma_0(\overline{X}_t^{-1}(x))|\partial\overline{X}_t^{-1}(x)/\partial x| \tag{62}$$

and

$$u_t(x) = \overline{V}_t(\overline{X}_t^{-1}(x)). \tag{63}$$

Equation (61) may also be written in the form

$$dm_t^{(1)}(x) = \sigma_t(x)\delta(v-u_t(x))dx\,dv, \tag{61'}$$

which signifies that $\sigma_t(x)$ is the normalized density and $u_t(x)$ the hydrodynamical drift velocity at the point x, at time t. It now remains for us to show that the evolution of σ_t, u_t, and the electric field E_t is governed by Eqs. (2).

To obtain (2a), we note, by (62),

$$\int \sigma_t(x)\phi(x)dx = \int \sigma_0(x)\phi(\overline{X}_t(x))dx, \tag{64}$$

for any bounded continuous function ϕ. Choosing ϕ to be continuously differentiable, it follows that

$$\frac{d}{dt}\int \sigma_t(x)\phi(x)dx = \int \sigma_0(x)\frac{\partial\overline{X}_t(x)}{\partial t}\cdot\nabla\phi(\overline{X}_t(x))$$
$$= \int \sigma_0(x)\overline{V}_t(x)\cdot\nabla\phi(\overline{X}_t(x)),$$

by (56) and (60).

Hence, using (64) again,

$$\frac{d}{dt}\int \sigma_t(x)\phi(x)dx - \int \sigma_t(x)u_t(x)\cdot\nabla\phi(x)dx = 0, \tag{65}$$

which is the "weak" form of Eq. (2a), i.e., it is the formula obtained by integrating (2a) against C^1-class functions ϕ. It reduces precisely to (2a) under the further assumptions that $\sigma_t(x)$ is C^1 class with respect to x and t and that $u_t(x)$ is C^2 with respect to x.

To prove (2b), we observe that, by (63),

$$u_t(\overline{X}_t(x)) = \overline{V}_t(x),$$

and therefore, assuming differentiability of $u_t(x)$ with respect to x and t,

$$\frac{\partial u_t}{\partial t}(\overline{X}_t(x)) + \frac{\partial \overline{X}_t(x)}{\partial t}\cdot\nabla u_t(\overline{X}_t(x)) = \frac{\partial \overline{V}_t(x)}{\partial t}.$$

By (56) and (60), this reduces to the form

$$\frac{\partial u_t}{\partial t}(\overline{X}_t(x)) + \overline{V}_t(x)\cdot\nabla u_t(\overline{X}_t(x)) = E_t(\overline{X}_t(x)).$$

Hence, as $\overline{X}_t$ is invertible and $u_t(x) = \overline{V}_t(\overline{X}_t^{-1}(x))$, by (62),

$$\frac{\partial u_t(x)}{\partial t} + (u_t(x)\cdot\nabla)u_t(x) = E_t(x),$$

which is the required equation (2b).

To prove (2c), we employ Eqs. (55) and (61) to express E_t in the form

$$E_t(x) = -\int \nabla V(x-x')\sigma_t(x')dx'.$$

Hence, by (21) and the normalization of σ_t,

$$\operatorname{div} E_t(x) = \sigma_t(x) - 1,$$

which is Eq. (2c).

This completes the derivation of the phenomenological equations (2).

VI. CONCLUDING REMARKS

By exploiting the long range and the scaling properties of the Coulomb interactions, we have derived the phenomenological equations (2) of the Jellium model from its Schrödinger equation and the initial conditions (1)–(3). The additional assumptions on which our derivation depends are that (a) the interelectronic repulsion keeps the two-particle density bounded [cf. (28)] and thereby tames the Coulomb singularity, and (b) the macroscopic dynamics is free from turbulence and sufficiently regular for the internal electric field to be continuously differentiable (cf. Secs. IV and V). In our view, such assumptions are indispensable, since the problem of obtaining mathematical control on the effect of the Coulomb singularity on multiple electron scattering processes is fantastically complicated.

As regards the main result obtained here, we emphasize again that this represents the macroscopic continuum mechanics of the model on the largest possible length scale. As explained near the end of Sec. I, this description is too coarse to accommodate either viscous stresses or pressure gradients. Evidently, a finer treatment is needed in order to represent these effects. Hence, the present theory should be regarded as providing just a skeletal version of a derivation of plasma electrohydrodynamics from quantum theory.

ACKNOWLEDGMENTS

Part of this work was done during visits to the Universities of Vienna and Rome. The author would like to acknowledge a number of useful discussions on the subject matter of this article with colleagues in Rome and Vienna, especially H. Narnhofer, A. Pflug, E. Presutti, M. Pulvirenti, and W. Thirring.

The author would also like to thank the Science and Engineering Research Council for financial support, which made his university visits possible.

APPENDIX: CONSTRUCTIONS

In order to construct a class of states $\Psi^{(N)}$ of Σ_N, which sastisfy conditions (1)–(3), we start by dividing the box Δ_L into N' equal cubic cells $\{C_J\}$, whose sides are of length l, independent of N, and whose centers are at the respective points $\{X_J\}$. Thus

$$N' = (L/l)^3 = N/\bar{n}l^3. \tag{A1}$$

We distribute the particles of Σ_N among the cells, so that the number n_J in C_J is of the form

$$n_J = \nu(X_J/L), \tag{A2}$$

where ν is some smooth function. For each cell, C_J, we construct a state Ψ_J, given by a normalized, antisymmetric function of the positions of the n_J particles there, which satisfies the following conditions.

(A) First, Ψ_J and its derivatives vanish on the boundary of C_J. The total kinetic and potential energies of the particles in the cell, interacting via the two-body potential $e^2V^{(L)}$, are therefore given unambiguously by the formulas

$$T_J = \frac{\hbar^2}{2m}n_J\int dX_1\cdots dX_{n_J}|\nabla_{X_1}\Psi_J(X_1,\ldots,X_{n_J})|^2 \tag{A3}$$

and

$$V_J = \frac{1}{2}n_J(n_J-1)e^2\int dX_1\cdots dX_{n_J}V^{(L)}(X_1-X_2) \times|\Psi_J(X_1,\ldots,X_{n_J})|^2, \tag{A4}$$

all integrations being taken over C_J.

(B) Both T_J and V_J are less than some finite constant B_1 for every one of the cells C_J.

(C) The one-particle probability density for the state Ψ_J, namely

$$P_J(X) = \int dX_2\cdots dX_{n_J}|\Psi_J(X,X_2,\ldots,X_{n_J})|^2, \tag{A5}$$

is bounded, uniformly with respect to J, i.e.,

$$P_J(X) < B_2,$$

where B_2 is a finite constant, independent of J.

Having constructed the cellular states Ψ_J according to these specifications, we define $\Psi^{(N)}$ to be the state given by their antisymmetrized product, i.e.,

$$\Psi^{(N)} = A\sum_{\text{antisymm}}\prod_J\Psi_J, \quad \text{with } A = \left[\left(\prod_J n_J!\right)\,(N!)^{-1}\right]^{1/2}. \tag{A6}$$

We shall now show that it follows from this definition and the properties (A)–(C) of the Ψ_J's that $\Psi^{(N)}$ satisfies the conditions (1)–(3).

Proof of (1): It follows from (A6) and the property (A) that the expectation value of the kinetic energy of Σ_N, for the state $\Psi^{(N)}$, is $\Sigma_J T_J$; and, by (B) and Eq. (A1), this is less than $N'B_1 \equiv NB/\bar{n}l^3$. Thus (1) is satisfied, with $B = B_1/\bar{n}l^3$.

Proof of (2): Since, by Eq. (3), the interactions in Σ_N are due to the pair potential $e^2V^{(L)}$, it follows from Eqs. (A2) and (A4)–(A6) that the expectation value of the potential energy of this system, for the state $\Psi^{(N)}$, is

$$\sum_J V_J + \sum_{J\neq J'} e^2 v(X_J/L)v(X_{J'}/L)\int_{C_J} dX \int_{C_{J'}} dX' \times V^{(L)}(X-X')P_J(X)P_{J'}(X'). \tag{A7}$$

By property (B) and Eq. (A1), the first of these sums is less than $B_2N' \equiv B_2N/\bar{n}l^3$. It therefore remains for us to prove that the second sum is less than some constant times $N^{2/3}$, for large N. Since $V^{(L)}(X-X')$ is the difference between $|X-X'|^{-1}$ and its average over Δ_L, we can majorize this second sum by

$$e^2 \sum_{J\neq J'} v(X_J/L)v(X_{J'}/L) \times \int_{C_J} dX \int_{C_{J'}} dX'\, P_J(X)P_J(X')/|X-X'|. \tag{A8}$$

By property (C), both $P_J(X)$ and $P_{J'}(X')$ are less than a constant B_2; consequently, it is a simple matter to show that, to adequate accuracy, we may replace the double integral in (A8) by $|X_J - X_{J'}|^{-1}$, thereby reducing that expression to

$$e^2 \sum_{J\neq J'} v(X_J/L)v(X_{J'}/L)/|X_J - X_{J'}|.$$

Likewise, by standard arguments, we may replace this last expression by

$$\frac{e^2}{l^6}\int_{\Delta_L} dX \int_{\Delta_L} dX'\, v(X/L)v(X'/L)/|X-X'|.$$

On putting $X = Lx$ and $X' = Lx'$, this reduces to

$$\frac{e^2L^5}{l^6}\int_{\Delta_1} dx \int_{\Delta_1} dx'\, v(x)v(x')/|x-x'|,$$

Δ_1 being the unit box, as previously. The double integral in this expression is a finite constant, since v has been specified to be a smooth function. From this we conclude that the second sum in (A7) is majorized by a constant times L^5, and therefore, as $L = (N/\bar{n})^{1/3}$, by a constant times $N^{5/3}$, as required.

Proof of (3): We shall start by showing that, in order to prove (3), it suffices to establish the formula (15) for the case when the ξ's are all zero. Thus, we note that it follows from (12) that

$$\frac{\partial\mu_0^{(N,1)}}{\partial\xi}(\xi,\nu) = \frac{i}{2mL\omega}(\Psi^{(N)},(P_1U + UP_1)\Psi^{(N)}),$$

where

$$U = \exp\left(\frac{i\xi\cdot P_1}{2mL\omega}\right)\exp\left(\frac{i\eta\cdot X_1}{L}\right)\exp\left(\frac{i\xi\cdot P_1}{2mL\omega}\right),$$

and, therefore, as this operator is unitary,

$$\left|\frac{\partial\mu_0^{(N,1)}}{\partial\xi}\right|^2 \leqslant \frac{(\Psi^{(N)},P_1^2\Psi^{(N)})}{(mL\omega)^2},$$

which is just the kinetic energy per particle of Σ_N, divided by $\frac{1}{2}mL^2\omega^2$. Hence, it follows from (1) that $\partial\mu_0^{(n,1)}/\partial\xi \to 0$ as L, and thus N, tends to infinity, the convergence being uniform with respect to ξ and η. Consequently,

$$\mu_0^{(N,1)}(\xi,\eta) - \mu_0^{(N,1)}(0,\eta) \to 0, \quad \text{as } N\to\infty,$$

and likewise

$$\mu_0^{(N,n)}(\xi_1,\ldots,\xi_n;\eta_1,\ldots,\eta_n) - \mu_0^{(N,n)}(0,\ldots,0;\eta_1,\ldots,\eta_n) \to 0, \quad \text{as } N\to\infty.$$

Therefore, as (12) implies that the $\mu^{(N,n)}$'s are uniformly bounded, the proof of (3) reduces to that of Eq. (15) for the case where the ξ's are all zero. This means that we need only to prove that

$$\mu_t^{(N,n)}(0,\ldots,0;\eta_1,\ldots,\eta_n) - \prod_1^n \mu_0^{(N,1)}(0,\eta_j) \to 0, \quad \text{as } N\to\infty. \tag{A9}$$

We shall restrict our explicit proof to the case $n = 2$; the general proof follows a similar pattern. By Eq. (12),

$$\mu_0^{(N,2)}(0,0;\eta,\eta') - \mu_0^{(N,1)}(0,\eta)\mu_0^{(N,1)}(0,\eta') = \int_{\Delta_L} dX \int_{\Delta_L} dX'(P^{(N)}(X,X') - P^{(N,1)}(X)P^{(N,1)}(X')) \times \exp i(\eta\cdot X + \eta'\cdot X')/L, \tag{A10}$$

where $P^{(N,1)}$ and $P^{(N,2)}$ are the one- and two-particle probability densities for the state $\Psi^{(N)}$, i.e.,

$$P^{(N,1)}(X) = \int dX_2\cdots dX_N\,|\Psi^{(N)}(X,X_2,\ldots,X_N)|^2 \tag{A11}$$

and

$$P^{(N,2)}(X,X') = \int dX_3\cdots dX_N\,|\Psi^{(N)}(X,X',X_3,\ldots,X_N)|^2. \tag{A12}$$

Defining the one- and two-particle probability densities $P_J^{(1)}$ and $P_J^{(2)}$ for the cellular state Ψ_J in an analogous way, it follows from (A6) that

$$P^{(N,1)}(X) = A^2\frac{(N-1)!}{(n_J-1)!\Pi_{J'\neq J}n_{J'}!}P_J^{(1)}(X) = \frac{n_J}{N}P_J^{(1)}(X) \quad \text{if } X \text{ lies in } C_J,$$

$$P^{(N,2)}(X,X') = A^2\frac{(N-2)!}{(n_J-2)!\Pi_{J'\neq J}n_{J'}!}P_J^{(2)}(X,X') = \frac{n_J(n_J-1)}{N(N-1)}P_J^{(2)}(X,X') \quad \text{if } X,X' \text{ lie in } C_J,$$

and

$$P^{(N,2)}(X,X') = \frac{A^2(N-2)!}{(n_J-1)!(n_{J'}-1)!\Pi_{J''\neq J,J'}n_{J''}!}P_J^{(1)}(X)P_{J'}^{(1)}(X') = \frac{n_Jn_{J'}}{N(N-1)}P_J^{(1)}(X)P_{J'}^{(1)}(X')$$

if X,X' lie in different cells $C_J,C_{J'}$, respectively.

It now follows from Eqs. (A10) and these formulas for $P^{(N,1)}$ and $P^{(N,2)}$ that

$$|\mu_0^{(N,2)}(0,0;\eta,\eta') - \mu_0^{(N,1)}(0,\eta)\mu_0^{(N,1)}(0,\eta')|$$

$$<\sum_{J,J'}\int_{C_J} dX \int_{C_{J'}} dX'|P^{(N,2)}(X,X') - P^{(N,1)}(X)\times P^{(N,1)}(X')|$$

$$<\sum_J \int_{C_J} dX \int_{C_J} dX' \left[\frac{n_J(n_J-1)}{N(N-1)} P_J^{(2)}(X,X') + \frac{n_J^2}{N^2} P_J^{(1)}(X)P_J^{(1)}(X')\right] + \sum \int_{C_J} dX \int_{C_{J'}} dX' \frac{n_J n_{J'}}{N^2(N-1)} \times P_J^{(1)}(X)P_{J'}^{(1)}(X'),$$

and therefore, as $P_J^{(1)}$ and $P_J^{(2)}$ are the one- and two-particle probability densities for C_J,

$$|\mu_0^{(N,2)}(0,0;\eta,\eta') - \mu_0^{(N,1)}(0,\eta)\mu_0^{(N,1)}(0,\eta')| <\sum_J \left[\frac{n_J(n_J-1)}{N(N-1)} + \frac{n_J^2}{N^2}\right] + \sum_{J' \neq J} \frac{n_J n_{J'}}{N^2(N-1)}. \quad \text{(A13)}$$

Since $\Sigma_J n_J = N$, the second sum in this expression is $O(N^{-1})$. Further, as $n_J \leqslant N$, the first sum $\leqslant 2\Sigma_J n_J^2/N^2$, and since $n_J = \nu(X_J/L)$, with ν a smooth function, this bound may be approximated, to adequate accuracy, by $2N^{-2}l^{-3}\int_{\Delta_L} dx(\nu(X/L))^2$, which in turn is equal to $2N^{-2}(L/l)^3\int_{\Delta_1} dx(\nu(x))^2$ and therefore is $O(N^{-1})$. Consequently, the rhs of (A14) is $O(N^{-1})$ and therefore that inequality implies that (A9) is valid for $N=2$, as required.

[1]E. B. Davies, J. Stat. Phys. **18**, 161 (1978).
[2]H. Narnhofer and G. L. Sewell, Commun. Math. Phys. **79**, 9 (1981).
[3]A. Galves, C. Kipnis, C. Marchioro, and E. Presutti, Commun. Math. Phys. **81**, 127 (1982).
[4]C. Boldrighini, R. L. Dobrushin, and Y. M. Sukhov, J. Stat. Phys. **31**, 577 (1983).
[5]G. L. Sewell, Phys. Lett. A **97**, 35 (1983).
[6]F. J. Dyson and A. Lenard, J. Math. Phys. **8**, 423 (1967); **9**, 698 (1968); E. H. Lieb and W. Thirring, Phys. Rev. Lett. **35**, 689 (1975).
[7]Admittedly, the stability of the Jellium model does not depend on quantum mechanics, but the methods we use here can be extended to more complex Coulomb systems, which would be unstable but for quantum effects.
[8]E. H. Lieb and H. Narnhofer, J. Stat. Phys. **12**, 291 (1975).
[9]D. Bohm and D. Pines, Phys. Rev. **92**, 609 (1953).
[10]D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969); K. Hepp, Helv. Phys. Acta **45**, 237 (1972).
[11]H. Spohn, Math. Meth. Appl. Sci. **3**, 445 (1981).
[12]R. Illner and H. Neunzert, Math. Meth. Appl. Sci. **1**, 530 (1979). Although the treatment of these authors was based on the assumption that $m_0^{(1)}$ corresponded to a uniformly bounded density $f_0(x,v)$, their proof of the existence of a solution goes through under the assumption of our uniform boundedness condition (44) and the initial zero velocity property (47).
[13]W. Braun and K. Hepp, Commun. Math. Phys. **56**, 101 (1977).
[14]H. Neunzert, "Neuere qualitative und numerische methoden in der plasmaphysik," in Lecture Notes, Paderborn, 1975; H. Neunzert, Fluid Dyn. Trans. **9**, 229 (1978).
[15]H. Lamb, *Hydrodynamics* (Cambridge U.P., Cambridge, England, 1959).

On the asymptotics of distributions with support in cone

S. Berceanu and A. Gheorghe
Institute for Physics and Nuclear Engineering, Bucharest, P. O. Box MG-6, Romania

(Received 14 January 1985; accepted for publication 15 March 1985)

The asymptotic behaviors of tempered distributions with support in a convex, closed cone are classified by means of group theory. The notion of regularly varying distribution is introduced. An Abelian–Tauberian theorem for regularly varying tempered distributions, which generalizes the one-dimensional Abelian–Tauberian theorem of Hardy–Littlewood–Karamata and the many-dimensional extension due to Vladimirov, is proved. Applications to n-point functions are also presented.

I. INTRODUCTION

In the actual stage of quantum physics development, the study of the asymptotic behavior of distributions plays a special part. On the one hand, many of the rigorous results in quantum field theory concern the asymptotics of distributions (Green functions, propagators, n-point functions, ..., see, e.g., Ref. 1) and on the other hand, different collision models for elementary particles predict similar high-energy behaviors (see, e.g., Ref. 2). In this context, a program for testing the compatibility of the scaling behavior in inclusive collision processes with the principles of quantum field theory was initiated.[3] A powerful tool in this approach are the Tauberian theorems. Such theorems are useful in the investigation of high-energy multiparticle phenomenology.[4] Moreover, it has been proved that there is a one-to-one Abelian–Tauberian-type correspondence between the scaling behavior in the Bjorken variables of form factors in deep inelastic lepton–hadron collision,[5] light cone dominance,[6] and the power-type asymptotic behavior of the spectral function in the Jost–Lehmann–Dyson representation.[3,7–12]

In this paper, the asymptotic behaviors of tempered distributions with support in a given cone are classified by means of the group theory. It is shown that every type of asymptotic behavior is effectively determined by a multiplier of the group of automorphisms of the support cone. The automodel asymptotics[7,10–12] corresponds to the particular dilatation group of automorphisms. The distributions considered are called regularly varying distributions and their properties are studied in Sec. II. An Abelian–Tauberian theorem for regularly varying distributions is proved in Sec. III. This theorem is a generalization of the classical one-dimensional Abelian–Tauberian theorem of Hardy–Littlewood–Karamata[13,14] and of Vladimirov's many-dimensional Tauberian theorem.[15] The theorem establishes the conditions of equivalence between the asymptotic behavior of regularly varying distributions, their Laplace transforms, and the regularly varying distributions obtained by the action of the Riemann–Liouville operator. Finally, in Sec. IV, we digress a little on the applications of previous results to Lorentz invariant tempered distributions with support in the N-point future cone.

II. REGULARLY VARYING TEMPERED DISTRIBUTIONS

In this section, we introduce spaces of regularly varying distributions and classify them using group theory notions.

A. General notations and definitions

We start with some notations and definitions. For details see Refs. 16 and 17.

Let $\Gamma\subset\mathbf{R}^n$ be a convex, closed cone, with vertex at 0. The Γ cone is supposed to be a sharp one, i.e., $C=\operatorname{int}\Gamma^*\neq\Phi$, where

$$\Gamma^*=\{y|y\in\mathbf{R}^n,\ (y,\xi)\geqslant 0,\ \xi\in\Gamma\}\tag{2.1}$$

is the dual of the Γ cone, and (,) denotes the scalar product of the Euclidean real n-dimensional space $\mathbf{R}^n$.

Let $\mathscr{S}'(\Gamma)$ denote the linear complex distribution space from $\mathscr{S}'(\mathbf{R}^n)$ with support in the cone Γ. It is important to point out that $\mathscr{S}'(\Gamma)$ is a complex topological algebra of tempered distributions with respect to the convolution product $*$ (see Ref. 16, p. 96). The Laplace transform establishes an algebraical and topological isomorphism of the algebra $\mathscr{S}'(\Gamma)$ onto the complex algebra $H(C)$ of holomorphic functions in the tube $T^C=\mathbf{R}^n+iC\subset\mathbf{C}^n$ (see Ref. 16, p. 161). The Laplace transform of $g\in\mathscr{S}'(\Gamma)$ is defined by the relation

$$L\{g\}(z)=F\{g(\cdot)e^{-(y,\cdot)}\}(x),\quad z=x+iy\in T^C,\tag{2.2}$$

where F designates the Fourier transform, and is defined by the convention

$$\langle F\{f\},\varphi\rangle=\langle f,F\{\varphi\}\rangle,\quad f\in\mathscr{S}'(\mathbf{R}^n),\quad \varphi\in\mathscr{S}(\mathbf{R}^n),$$
$$F\{\varphi\}(\xi)=\int\varphi(x)e^{i(\xi,x)}\,dx,\quad \varphi\in\mathscr{S}(\mathbf{R}^n).\tag{2.3}$$

If $\mu\in\mathscr{S}'(\Gamma)$ is a measure, then its Laplace transform is

$$L\{\mu\}(z)=\int_\Gamma e^{i(z,\xi)}\mu(d\xi),\quad z\in T^C.\tag{2.4}$$

The distribution $e\in\mathscr{S}'(\Gamma)$ is called a *unity* if there exists $e'\in\mathscr{S}'(\Gamma)$ such that $e*e'=e'*e=\delta_\Gamma$, where δ_Γ is the Dirac distribution of the Γ cone.

B. The Riemann–Liouville operator

Definition 1: For every unity $e\in\mathscr{S}'(\Gamma)$, we call a *Riemann–Liouville operator* the linear continuous operator Ω_e: $\mathscr{S}'(\Gamma)\to\mathscr{S}'(\Gamma)$ defined by

$$\Omega_e f=e*f,\quad f\in\mathscr{S}'(\Gamma).\tag{2.5}$$

We now present some examples which we will take again further in the paper.

Let $\Theta_\Gamma\in\mathscr{S}'(\Gamma)$ be the usual characteristic function of the cone Γ,

$$\Theta_\Gamma(\xi) = \begin{cases} 1, & \xi\in\Gamma, \\ 0, & \xi\in\mathbb{R}^n \setminus \Gamma. \end{cases} \tag{2.6}$$

The Laplace transform of the Θ_Γ function is the Cauchy–Szegö nucleus of the cone Γ (see Ref. 16, p. 140)

$$\mathscr{K}_C(z) = \int_\Gamma e^{i(z,\xi)} d\xi, \quad z\in T^C. \tag{2.7}$$

Imposing on the cone Γ the regularity condition, Θ_Γ is a unity. Therefore, the distribution

$$\Theta_\Gamma^\alpha = L^{-1}\{\mathscr{K}_C^\alpha\}, \quad \alpha\in\mathbb{R}, \tag{2.8}$$

is also a unity.

Consequently, $\Omega_{\Theta_\Gamma^\alpha}$ is a fractional integration (resp. derivation) operator for $\alpha > 0$ (resp. $\alpha < 0$).

Example 1: In the case of the semiright line cone $\Gamma = [0,\infty) = \mathbb{R}_+$,

$$\Omega_{\Theta_{\mathbb{R}_+}^\alpha} = Y_{\alpha^*}, \quad \alpha\in\mathbb{R}, \tag{2.9}$$

where the unity $Y_\alpha \in \mathscr{S}'(\mathbb{R}_+)$ is defined[14,15] as

$$Y_\alpha(\xi) = \begin{cases} [\xi^{\alpha-1}/\Gamma(\alpha)]\Theta(\xi), & \xi\in\mathbb{R}, \quad \alpha > 0, \\ Y'_{\alpha+1}(\xi), & \alpha \leqslant 0. \end{cases} \tag{2.10}$$

The $\Omega_{\Theta_{\mathbb{R}_+}^\alpha}$ is the classical Riemann–Liouville operator.[18]

Example 2: Let V^+ be the Minkowski cone

$$V^+ = \{x | x = (x^0, x^1, \dots, x^{n-1}) \in \mathbb{R}^n, \quad x^0 > 0,$$
$$x^2 = (x^0)^2 - (x^1)^2 - \cdots - (x^{n-1})^2 > 0\}. \tag{2.11}$$

Lemma 1: The distribution $\Theta_{V^+}^\alpha$ has the expression

$$\Theta_{V^+}^\alpha = k_n^\alpha Z_{n\alpha}, \tag{2.12}$$

where

$$k_n = (4\pi)^{(n-1)/2}\Gamma(n/2)/\Gamma(1/2). \tag{2.13}$$

Proof: In formula (2.12), Z denotes the Riesz–Schwartz distribution (see Ref. 18 and Ref. 17, pp. 50 and 178). The explicit values of the Z distributions can be extracted from the paper[19] of Méthée as

$$Z_l = \begin{cases} \mathrm{Pf}(x^2)^{(l-n)/2}/H_n(l) \\ \quad l \notin \{n-2, n-4, \dots\} \cup \{0, -2, -4, \dots\}, & (2.14a) \\ H^{(n-l-2)/2}/(2^{l-1}\pi^{(n-2)/2}\Gamma(l/2)), \\ \quad l\in\{n-2, n-4, \dots\} \setminus \{0, -2, -4, \dots\}, & (2.14b) \\ \square^{(-l/2)}\delta, \quad l = 0, -2, -4, \dots, & (2.14c) \end{cases}$$

$$H_n(l) = 2^{l-1}\pi^{(n-2)/2}\Gamma(l/2)\Gamma((l+2-n)/2), \tag{2.14d}$$

where the meanings of Hadamard's symbol "Pf" and of the distribution H^k [which corresponds in the present situation to $\Theta(x^0)\delta^{(k)}(x^2)$ in usual notation (see, e.g., Ref. 20, p. 347] are stated precisely in Ref. 19.

To prove Eq. (2.12), we observe that Z_l ($\Theta_{V^+}^\alpha$, respectively) is a homogeneous distribution of degree $n-l$ [$n(\alpha-1)$, respectively] and therefore $\Theta_{V^+}^\alpha = c_{\alpha,n} Z_{n\alpha}$. The convolution properties [cf. Ref. 16, Chap. II, Eq. (5.2)]

$$\Theta_{V^+}^\alpha * \Theta_{V^+}^\beta = \Theta_{V^+}^{\alpha+\beta} \tag{2.15}$$

and [cf. Ref. 17, Eq. (VI. 5.19)]

$$Z_\alpha * Z_\beta = Z_{\alpha+\beta} \tag{2.16}$$

impose to the constant $c_{\alpha,n}$ the form $c_{\alpha,n} = k_n^\alpha$. The value (2.13) of the normalization constant k_n is deduced from Eq. (2.14a), which for $l = n$ reads

$$\Theta_{V^+} = k_n Z_n. \tag{2.17}$$

We remember that a general prescription of construction for Riemann–Liouville operators for different cones is available (see, e.g., Ref. 21).

C. The group *Aut* Γ

In order to classify the asymptotic behaviors of the distributions from $\mathscr{S}'(\Gamma)$, we now consider the group $Aut\,\Gamma$ of diffeomorphisms of the space $\mathbb{R}^n$ which invariates the Γ cone: $g\Gamma \subset \Gamma$ for every $g\in Aut\,\Gamma$. The group $Aut\,T^C$ of analytic automorphisms of the tube T^C is formed by the affine transformations $\alpha: T^C \to T^C$

$$\alpha(z) = Az + a, \quad z\in T^C, \tag{2.18}$$

where $A\in Aut\,\Gamma$, and a is a real vector (Theorem I, Chap. I, Ref. 22). It follows that $Aut\,\Gamma$ is a Lie subgroup of the general group $\mathrm{GL}(n,\mathbb{R})$ of real nonsingular $n\times n$ matrices, identified with the group of linear inversible transformations of the space $\mathbb{R}^n$.

Let us consider a connected Lie subgroup G of the group $Aut\,\Gamma$. Evidently, $G\subset \mathrm{GL}(n,\mathbb{R})$. A multiplier χ of the group G is a homomorphism of G into the multiplicative group of different from zero complex numbers $\mathbb{C}^* = \mathbb{C}\setminus\{0\}$ (see, e.g., Ref. 23), i.e.,

$$\chi(AB) = \chi(A)\chi(B), \quad A,B\in G. \tag{2.19}$$

If the function χ is bounded, the denomination of character is also used.

In order to give a group characterization of the asymptotics of distributions, a decomposition of the group G is needed.

Lemma 2: Let G be a connected Lie subgroup of $\mathrm{GL}(n,\mathbb{R})$. Then there exists an isomorphism

$$G \approx G''' \otimes (G' \otimes G''), \tag{2.20}$$

where the group G''' is given by $G''' = \chi^{-1}(1)$ and G' (resp. G'') is an Abelian noncompact (resp. compact) subgroup of G. Moreover, the isomorphism (2.20) is realized by continuous functions.

Proof: Because G''' is an invariant subgroup of the group G, G decomposes in the semidirect product

$$G \approx G''' \otimes G/G'''. \tag{2.21}$$

But $G''' \supset [G,G]$, where $[G,G]$ denotes the commutator of the group G, which is an invariant subgroup of the groups G and G''', and the inclusion $G/G''' \subset G/[G,G]$ follows. Moreover, the group $G/[G,G]$ is an Abelian one, and also the factor group G/G'''.

The group G/G''' being connected and Abelian, admits a decomposition of the announced form, where the groups G' and G'' are of the type $\mathbb{R}_+^{*p}$ (direct product of p groups of dilatation $\mathbb{R}_+^* = \mathbb{R}_+\setminus\{0\}$) and, respectively, a torus T^q, where p and q are positive integers [e.g., Theorem (9.4) from Ref. 23].

The canonical projection $\pi_0: G \to G'''$ is real and analytic (cf. Ref. 24, pp. 123 and 115) and so also is the canonical projection $\pi: G \to G'$.

Lemma 2 allows us to introduce a notion of limit, which is very important for our approach to the problem of asymptotics of distributions.

Definition 2: Let $\pi: G \to G'$ be the canonical projection and let the mapping $\omega = \omega' \circ \pi$, $\omega: G \to \mathbb{R}_+^{*p}$, where $\omega': G' \to \mathbb{R}_+^{*p}$, $\omega(A) = (\lambda_1, \dots, \lambda_p) \in \mathbb{R}_+^{*p}$, $A \in G$, and $\lambda_i = \lambda_i(A)$, $i = 1, \dots, p$, as in Lemma 2. For a function $f: G \to \mathbb{C}$, we introduce the notation

$$\lim_A f(A) = \lim_{\substack{\lambda_i \to \infty \\ 1 < i < p}} f(A(\lambda_1, \dots, \lambda_p)), \quad A \in G. \tag{2.22}$$

D. Regularly varying distributions

To every distribution $f \in \mathscr{S}'(\Gamma)$, matrix $A \in G$, and multiplier χ of G, let us associate a distribution $T_\chi(A) f \in \mathscr{S}'(\Gamma)$ by the formulas

$$T_\chi(A) f = |\det A| \chi(A) f_A, \tag{2.23a}$$

$$\langle f_A, \varphi \rangle = \langle f, \varphi_A \rangle, \quad \varphi \in \mathscr{S}(\mathbb{R}^n), \tag{2.23b}$$

$$\varphi_A(\xi) = |\det A|^{-1} \varphi(A^{-1}\xi), \quad \xi \in \Gamma. \tag{2.23c}$$

Note that if the distribution $f \in \mathscr{S}'(\Gamma)$ is a local integrable function, then the relations (2.23) can be explicated in the form

$$\langle T_\chi(A) f, \varphi \rangle = \chi(A) |\det A| \int_{\mathbb{R}^n} f(A\xi) \varphi(\xi) d\xi. \tag{2.24}$$

Therefore, the relations (2.23) can be *formally* written as

$$(T_\chi(A) f)(\cdot) = |\det A|_\chi(A) f(A \cdot), \tag{2.25}$$

and if $|\det A| = 1$, $\chi(A) = 1$, we also use the notation

$$(T_\chi(A) f)(\cdot) = f_A(\cdot) = f(A \cdot). \tag{2.26}$$

Remark 1: Equations (2.23) imply that the mapping $A \to T_\chi(A^{-1})$, $A \in G$ is a representation of the group G on $\mathscr{S}'(\Gamma)$,

$$T_\chi(AB) = T_\chi(B) T_\chi(A), \quad A, B \in G. \tag{2.27}$$

If the multiplier χ is continuous, this representation is continuous. In this paper, all the multipliers are supposed to be continuous.

After this preparation, we are ready to introduce the definition of regularly varying tempered distributions.

Definition 3: Let χ be a continuous multiplier of the connected Lie subgroup $G \subset Aut\,\Gamma$. The distribution $f \in \mathscr{S}'(\Gamma)$ is called a *regularly varying distribution of type* χ, if there exists $g \in \mathscr{S}'(\Gamma)$ and there exists the limit

$$\lim_A T_\chi(A) f = g \neq 0, \quad \text{in} \quad \mathscr{S}'(\Gamma). \tag{2.28}$$

Remark 2: The tempered distributions with support in a cone are asymptotically classified by the continuous multipliers of the Lie groups of transformations which invariate the support cone. These groups are effectively classified in Ref. 22.

Example 3: The case when G is the dilatation group D_Γ of the cone Γ,

$$D_\Gamma = \{\hat{\lambda} \,|\, \hat{\lambda} \in \mathrm{GL}(n, \mathbb{R});\ \lambda \in \mathbb{R}_+^*,\ \hat{\lambda}\xi = \lambda\xi,\ \xi \in \Gamma\}, \tag{2.29}$$

with the multiplier $\chi(\hat{\lambda}) = \lambda^{-\alpha - n}$, corresponds to distributions f with "quasi-asymptotics of order α" in the sense of Ref. 25:

$$\lim_{k \to \infty} \langle k^{-\alpha} f(k \cdot), \varphi(\cdot) \rangle = \langle g(\cdot), \varphi(\cdot) \rangle, \quad f, g \in \mathscr{S}'(\Gamma). \tag{2.30}$$

For the one-dimensional cone $\Gamma = \mathbb{R}_+$ it is easily seen that the limiting distribution g has the expression $g = Y_{\alpha+1}$ (see Ref. 26). Sufficient conditions for a function with quasi-asymptotics to have usual asymptotics are known (Lemma 4, Corolar Landau from Ref. 26). A necessary and sufficient condition for the existence of quasi-asymptotics is presented in Ref. 25. In particular, it follows from the structure theorem of tempered distributions that if the tempered distribution f has quasi-asymptotics of order α, then his growth index at ∞ (cf. Ref. 17, p. 241) is equal to $\alpha + nN$, where N is the order of the primitive of f for which a continuous function in the cone Γ is obtained by integration.

Also note that if a regularly varying function is substituted to k^α in Eq. (2.30), then the limiting distribution g is not affected, nor is the order of growth at infinity.[8,11]

Other examples will be presented in the next sections.

Remark 3: Remark 1 and the properties of the multipliers imply that if $f \in \mathscr{S}'(\Gamma)$, then

$$\lim_A T_\chi(AB) f = \lim_A T_\chi(A) f, \tag{2.31}$$

where $B \in G$ is fixed. In other words, the limiting distribution $g \in \mathscr{S}'(\Gamma)$ from Eq. (2.28) has the property of covariance

$$T_\chi(A) g = g, \quad A \in G. \tag{2.32}$$

We will present necessary and sufficient conditions for the existence of regularly varying tempered distributions in the next section.

III. CHARACTERIZATION OF REGULARLY VARYING DISTRIBUTIONS

Let us consider a tempered distribution $f \in \mathscr{S}'(\Gamma)$ and a continuous multiplier χ of the connected Lie subgroup $G \subset Aut\,\Gamma$. We now advance the following three hypotheses: (i) there exists $g \in \mathscr{S}'(\Gamma)$ and there exists the limit

$$\lim_A T_\chi(A) f = g \neq 0, \quad \text{in } \mathscr{S}'(\Gamma); \tag{3.1}$$

(ii) there exist two continuous multipliers χ' and χ'' of the group G, the unity $e \in \mathscr{S}'(\Gamma)$ invariant to T_χ, and the function H such that

$$\lim_A T_{\chi''}(A) \Omega_e f = H \neq 0, \quad \text{in } \mathscr{C}(\Gamma); \tag{3.2}$$

and (iii) there exist the distribution h and the limit

$$\lim_A L\{T_\chi(A) f\}(iy) = h(y) \neq 0, \quad y \in C. \tag{3.3}$$

Here $\mathscr{C}(\Gamma)$ denotes the complex topological linear space of continuous functions with support in the Γ cone (relatively to the topology of uniform convergence[17]).

Now we are ready to state the main results of this paper.

Theorem 1: (a) The hypotheses (i) and (ii) are equivalent, and the hypothesis (i) implies hypothesis (iii).

(b) If f is a positive measure, then the hypotheses (i), (ii), and (iii) are equivalent.

(c) If the hypotheses (i)–(iii) are fulfilled, then

$$\chi'' = \chi\chi', \tag{3.4a}$$
$$H = \Omega_e g, \tag{3.4b}$$
$$T_\chi(A)g = g, \tag{3.4c}$$
$$T_{\chi''}(A)H = H, \tag{3.4d}$$
$$h(A'z) = \chi(A)h(z), \tag{3.4e}$$
$$h(z) = L\{g\}(iz), \tag{3.4f}$$

where $A \in G$, $z \in T^C$ and A' denotes the transpose of the matrix A.

Proof: (a) From the structure theorem of distributions from $\mathscr{S}'(\Gamma)$ (see Ref. 17, p. 239, and Refs. 27 and 28) it follows that there exist a continuous function $f_0 \in \mathscr{C}(\Gamma)$ and a natural number α such that

$$f = \Theta_\Gamma^{-\alpha} * f_0. \tag{3.5}$$

Let now χ'' be a continuous multipler. Since

$$(f*g)_A = |\det A| f_A * g_A, \quad f,g \in \mathscr{S}'(\Gamma), \quad A \in G, \tag{3.6}$$

it follows from the definition (2.23) that

$$T_{\chi''}(A)(\Theta_\Gamma^\alpha * f) = |\det A|^2 \chi''(A)(\Theta_\Gamma^\alpha)_A * f_A. \tag{3.7}$$

Suppose now the decomposition (3.4a) for the multiplier χ'' and choose χ' as in Theorem 1

$$T_{\chi'}(A)\Theta_\Gamma^\alpha = \Theta_\Gamma^\alpha, \quad A \in G. \tag{3.8}$$

In fact, from the property of the Laplace transform

$$L\{f_A\}(z) = |\det A|^{-1} L\{f\}(A^{-1\prime}z), \quad z \in T^C, \tag{3.9}$$

applied to the Cauchy–Szegö nucleus (2.8), it results that

$$\mathscr{K}_C(A'z) = |\det A|^{-1}\mathscr{K}_C(z), \quad z \in T^C, \quad A \in G. \tag{3.10}$$

With definition (2.8), we get

$$(\Theta_\Gamma^\alpha)_A = |\det A|^{\alpha-1}\Theta_\Gamma^\alpha. \tag{3.11}$$

Therefore, for getting Eq. (3.8) satisfied, we fix the multiplier χ' to the value

$$\chi'(A) = |\det A|^{-\alpha}, \tag{3.12}$$

and relation (3.7) reads

$$T_{\chi''}(A)(\Theta_\Gamma^\alpha * f) = \Theta_\Gamma^\alpha * T_\chi(A)f. \tag{3.13}$$

Now, taking, in both sides of Eq. (3.13), the limit in the sense of Definition 2 and taking into account the continuity of the convolution in the topological algebra $\mathscr{S}'(\Gamma)$ and the continuity of the multipliers, the implication (i) $\Rightarrow$ (ii) is proved, and by similar arguments, the converse implication is also proved. The multiplier χ' is given by Eq. (3.12) and the function H is expressed as in Eq. (3.4b), where, in agreement with Eq. (3.5),

$$e = \Theta_\Gamma^\alpha. \tag{3.14}$$

The implication (i) $\Rightarrow$ (iii) follows immediately from the continuity of the Laplace transform for regular cones. In fact, from hypothesis (i) we get more than hypothesis (iii), namely we find that (iii′) there exists the limit

$$\lim_A L\{T_\chi(A)f\}(z) = L\{g\}(z), \tag{3.15}$$

$$L\{g\}(z) = h(-iz), \quad z \in T^C, \quad A \in G.$$

The statement (a) of Theorem 1 has been proved. The assertion (b) follows from (a) and the (extended) continuity for positive measures and their Laplace transforms (cf. Refs. 13 and 29).

To end the proof of the theorem, we combine the relations (2.23), (3.9), (3.15), and (2.31), and we get

$$\begin{aligned} h(B'z) &= \lim_A L\{T_\chi(A)\}(iB'z) \\ &= \lim_A \chi(A)|\det A| L\{f_A\}(iB'z) \\ &= \lim_A \chi(A) L\{f\}(i(A')^{-1}B'z) \\ &= \lim_A \chi(A)\chi(AB^{-1})|\det(AB^{-1})| L\{f_{AB^{-1}}\}(iz) \\ &= \lim_{AB^{-1}} \chi(B) L\{T_\chi(AB^{-1})f\}(iz) = \chi(B)h(z). \end{aligned}$$

Remark 4: We consider again the case of the one-dimensional cone $\Gamma = \mathbb{R}_+$ from example 1 and the multiplier $\chi(\lambda) = \lambda^{-\alpha}$, $\lambda \in \mathbb{R}_+^*$. The distribution Y, defined by Eq. (2.10), verifies the relations

$$\begin{gathered} \Theta_\Gamma = Y_1, \quad Y_\alpha * Y_\beta = Y_{\alpha+\beta}, \quad \alpha,\beta \in \mathbb{R}, \\ L\{Y_\alpha\}(z) = (-iz)^\alpha, \quad z = x + iy \in \mathbb{C}, \quad y > 0. \end{gathered} \tag{3.16}$$

If $f = \mu'$ and the distribution $\mu \in \mathscr{S}'(\Gamma)$ is a positive measure, then Theorem 1 establishes the equivalence of the relations

$$\begin{aligned} &\text{(i)}'' \quad \lim_{\lambda\to\infty} \lambda^{1-\alpha} f_\lambda = Y_\alpha; \\ &\text{(ii)}'' \quad \lim_{\lambda\to\infty} \lambda^{-\alpha}\int_0^{\lambda x} f(\xi)d\xi = Y_{\alpha+1}(x), \quad x \in \mathbb{R}_+; \\ &\text{(iii)}'' \quad \lim_{\lambda\to\infty} \lambda^{1-\alpha}\int_0^\infty e^{-y\xi} f_\lambda(\xi)d\xi = y^{-\alpha}, \quad y \in \mathbb{R}_+^*. \end{aligned} \tag{3.17}$$

This equivalence is in fact the classical Abelian–Tauberian theorem of Hardy–Littlewood–Karamata. The implication (i)″ $\Rightarrow$ (ii)″ is due to Karamata[30] (see also Ref. 13, p. 272 and Ref. 14, Theorem 2.2). The implication (ii)″ $\Rightarrow$ (iii)″ is an Abelian-type theorem and the implication (iii)″ $\Rightarrow$ (ii)″ is the Tauberian theorem, the latter two implications being consequences of the extended continuity (Ref. 13, p. 410 and Ref. 14). In the framework of the usual asymptotics, the implication (ii)″ $\Rightarrow$ (i)″ results from the monotony of f [Ref. 13, p. 421 and Ref. 14, Theorem (2.4)].

Remark 5: If $G = D_\Gamma$ as in example 3, then Theorem 1 implies, in particular, Vladimirov's theorem for tempered positive measures.[11,15]

Remark 6: The implication (iii) $\Rightarrow$ (i) is true only under additional conditions. Such a condition is the existence of a $y_0 \in C$ for which the functions $L\{T_\chi(A)f\}(z)$ are bounded in $\mathbb{R}^n + i\mathbb{R}_+^* y_0$. In this case, the Laplace transform passes in the limit $\rho y_0 \to 0$, $\rho > 0$ (in the sense of tempered distributions) into the Fourier transform F and the assertion follows from the isomorphism F: $\mathscr{S}'(\Gamma) \to \mathscr{S}'(\Gamma)$ (cf. Ref. 17, p. 251). Other sufficient conditions for the validity of the implication (iii) $\Rightarrow$ (i) in the quasi-asymptotics case are presented in Refs. 25 and 26.

We also emphasize that if the condition (iii) is replaced by the condition (iii)′, then (iii)′ $\Rightarrow$ (i) as a consequence of the algebraical and the topological isomorphism of $\mathscr{S}'(\Gamma)$ and

$H(C)$. In fact, the conditions (i), (ii), and (iii)′ are equivalent.

The relationship between the functions h and H appearing in Theorem 1 is given by the following Proposition.

Proposition 1: Let us suppose that the group G acts effectively on the Γ cone. Under the conditions of Theorem 1, the following integral equation is satisfied:

$$h(-iz)L\{e\}(z) = \int_S \Delta(z,\sigma)\,\widetilde{H}(\sigma)d\sigma,$$
$$\Delta(z,\sigma) = \int_G e^{i(z,A\sigma)}\chi''(A^{-1})|\det A|^{-1}\,dA, \tag{3.18}$$

where $z\in T^C$, dA is the Haar measure on G, $d\sigma$ is the measure on the orbit space, $S=\Gamma/G$, and $\widetilde{H}=H|_S$.

Proof: Applying the Laplace transform to both sides of Eq. (3.4b) and using Eq. (3.4f), we get

$$\begin{aligned}&\int_\Gamma e^{i(z,\xi)}H(\xi)d\xi\\ &\quad = L\{H\}(z) = L\{e*g\}(z) = L\{e\}(z)\cdot L\{g\}(z)\\ &\quad = L\{e\}(z)h(-iz),\ z\in T^C.\end{aligned} \tag{3.19}$$

Now Eq. (3.18) is a consequence of the effectiveness of the action of G on Γ, of the decomposition $d\xi = dA\otimes d\sigma$, $A\in G$, $\sigma\in S$, and of Eq. (3.4d) explicated in the form

$$H(\sigma A) = \chi''(A^{-1})|\det A|^{-1}H(A),\quad A\in G. \tag{3.20}$$

Remark 7: From the definition (2.8) of the function Θ^α_Γ and Eq. (3.19), it is easily seen that the integral equation (3.18) has for

$$g = c\Theta^p_\Gamma,\quad e=\Theta^q_\Gamma \tag{3.21}$$

the unique solution

$$H = c\Theta^{p+q}_\Gamma. \tag{3.22}$$

Example 4: To illustrate the content of Proposition 1, we consider as group G the product of dilatation groups D_{Γ_k} and as cone Γ the topological product of cones Γ_k, $k=1,\ldots,N$:

$$\Gamma = \bigotimes_{l=1}^N \Gamma_l,\quad \Gamma_l\subset\mathbf{R}^{n_l},\quad l=1,\ldots,N, \tag{3.23a}$$

$$G = \{A\,|A = \bigotimes_{l=1}^N \lambda_l \mathbf{1}_{n_l},\ \lambda_l\in\mathbf{R}^*_+,\ l=1,\ldots,N\}, \tag{3.23b}$$

$$\chi(A) = \prod_{l=1}^N \lambda_l^{-\beta_l}, \tag{3.23c}$$

where the numbers β_l are fixed and $\mathbf{1}_n$ represents the unit $n\times n$ matrix.

Now, combining Eqs. (3.4a), (3.4b), (3.12), (3.14), and (3.23), the integral equation (3.18) gets the form

$$h(-iz)\mathscr{K}^\alpha(z) = \int_S d\sigma\,\widetilde{H}(\sigma)\prod_{l=1}^N \frac{\Gamma(\beta_l+\alpha n_l)}{(-iz_l,\sigma_l)^{\beta_l+\alpha n_l}}, \tag{3.24}$$

where $z=(z_1,\ldots,z_N)$, $z_k\in T^{C_k}$, $C_k = \operatorname{int}\Gamma^*_k$, $S = \bigotimes_{k=1}^N S_k$, S_k is the unit sphere in R^{n_k} and $\mathscr{K}(z) = \Pi_{l=1}^N\,\mathscr{K}(z_l)$.

References 10, 15, and 25 deal with the case $N=1$.

IV. APPLICATIONS TO N-POINT DISTRIBUTIONS

(1) The N-particle distributions are tempered distributions with support in the product of N Minkowski cones $\overline{V}^+$ (2.12) (see, e.g., Ref. 31)

$$\overline{V}^+_N = \bigotimes_{k=1}^N \overline{V}^+. \tag{4.1}$$

The Minkowski scalar product is denoted by

$$xy = x^0y^0 - x^1y^1 - \cdots - x^{n-1}y^{n-1}, \tag{4.2}$$

where $x,y\in\mathbf{R}^n$ and the same notation is maintained if $x,y\in\mathbf{C}^n$.

The N-point cone $\overline{V}^+_N$ is a convex, closed cone, with the vertex at 0, self-dual (i.e., $\Gamma=\Gamma^*$), regular, with the Cauchy–Szegö nucleus

$$\mathscr{K}_{V^+_N}(z) = \prod_{l=1}^N \mathscr{K}_{V^+}(z_l), \tag{4.3}$$

$$\mathscr{K}_{V^+}(z_l) = k_n(-z_l^2)^{-n/2}, \tag{4.4}$$

where $z=(z_1,\ldots,z_N)\in T^{V^+_N}$ and the value of the constant k_n is given by Eq. (2.13). The latter relation follows from Eqs. (2.7) and (2.17) and the relation [see, e.g., Eq. (VIII, 7.37) from Ref. 17]

$$L\{Z_l\}(z) = (-z^2)^{-l/2},\quad z\in T^{V^+}. \tag{4.5}$$

The group of analytic automorphisms $Aut\,\overline{V}^+_N$ is formed by all matrices $A = \otimes^N_{l=1}\lambda\Lambda$, where Λ belongs to the restricted Lorentz group $L^{\uparrow}_+$ (see Ref. 31). A N-point distribution $f\in\mathscr{S}'(\overline{V}^+_N)$ is Lorentz invariant if $f_\Lambda = f$ for every $\Lambda\in L^\uparrow_+$ [see Eqs. (2.24) and (2.27)], and Λ acts as follows: $\Lambda(x_1,\ldots,x_N) = (\Lambda x_1,\ldots,\Lambda x_N)$, $x_1,\ldots,x_N\in\mathbf{R}^n$. The structure of the Lorentz invariant distributions is presented in Ref. 19.

We remember now that the Lorentz covariant distributions are linear combinations of Lorentz invariant distributions with standard Lorentz covariant polynomial coefficients.[28] Then it follows that the asymptotic behavior of basical distributions in quantum field theory (e.g., Wightman distributions, Steinmann distributions, Green functions, propagators[1,31]) is determined by the asymptotic behavior of Lorentz invariant distributions from $\mathscr{S}'(\overline{V}^+_N)$. Moreover, if these distributions are regularly varying, then Theorem 1 allows the testing of the compatibility of N-point functions behavior at high energy with the principles of quantum field theory.

We emphasize that if the distribution $f\in\mathscr{S}'(\overline{V}^+_N)$ is Lorentz invariant, then the functions h,g, and H appearing in Theorem 1 are also Lorentz invariant.

(2) We take again example 4 of Sec. III, where $\Gamma_k = \overline{V}^+$ and $Aut\,\Gamma$ is the direct product of dilatation groups, an important case for conformal quantum field theory (see, e.g., Ref. 32). Let us consider that f is given by the tensorial product

$$f = \bigotimes_{l=1}^N f_l\in\mathscr{S}'(\overline{V}^+_N) \tag{4.6}$$

of two-point Wightman distributions,[31] $f_l\in\mathscr{S}'(\overline{V}^+)$, $l=1,\ldots,N$. Taking into account that the two-point Wightman distributions are positive measures,[31] Theorem 1 asserts the equivalence of hypotheses (i)–(iii).

Let us now suppose f_l, $l=1,\ldots,N$, to be regularly varying distributions of type χ, with the multiplier χ given by Eq. (3.23c). From Eq. (3.4c), we get that the limiting distribution g has the form

$$g = \mathop{\otimes}_{l=1}^{N} g_l, \tag{4.7}$$

and every g_l is a homogeneous distribution of degree $\beta_l - n$. So we take in Eq. (4.7)

$$g_l = a\,(\Theta_{\overline{V}^+}/k_n)^{\beta_l/n} = a\,Z_{\beta_l}, \tag{4.8}$$

where a is a constant.

Because the primitive of first order of positive measure is a function from $\mathscr{C}(\overline{V}_N^+)$, the unity e (3.14) in Theorem 1 may be taken to be

$$e = \Theta_{\overline{V}^+}. \tag{4.9}$$

Then, Eqs. (3.21) and (3.22) imply that the unique solution of Eq. (3.24) is

$$H(z) = \prod_{l=1}^{N} H_l(z_l), \quad z_l \in T^{V^+}, \quad l = 1,...,N, \tag{4.10}$$

$$H_l = a k_n{}^{-\beta_l/n}\,\Theta_{\overline{V}^+}{}^{1+\beta_l/n} = a k_n\, Z_{\beta_l+n}. \tag{4.11}$$

Now we use Eqs. (4.7), (4.8), and (3.4f) and we find

$$h(z) = \prod_{l=1}^{N} h_l(z_l), \tag{4.12}$$

$$h_l(z_l) = a(-z_l^2)^{-\beta_l/2}, \quad z_l \in T^{V^+}, \quad l = 1,...,N. \tag{4.13}$$

Remembering the explicit formulas (2.14) of the Riesz–Schwartz Z distribution, it is seen that not every asymptotic behavior, determined by β_l, is acceptable for the distribution (4.6), product of two-point distributions.

For $n = 4$, the restrictions $\beta_l = 0$ or $\beta_l \geqslant 2$ follow. With Eq. (2.14), the relations (4.8) and (4.10) may be written explicitly as

$$H_l(\eta) = \begin{cases} a\,\Theta_{\overline{V}^+}(\eta), & \beta_l = 0, \\ a\,8\pi^{-1/2}\dfrac{\Gamma((\beta_l+5)/2)}{\Gamma(\beta_l+4)\Gamma(1+\beta_l/2)}\,\mathrm{Pf}(\eta^2)^{\beta_l/2}, \\ \beta_l \geqslant 2; \end{cases} \tag{4.14}$$

$$g_l(\eta) = \begin{cases} a\,\delta(\eta), & \beta_l = 0, \\ (2\pi)^{-1} a\,H(\eta), & \beta_l = 2, \\ \dfrac{a\,\mathrm{Pf}(\eta^2)^{(\beta_l-4)/2}}{\pi 2^{\beta_l-1}\Gamma(\beta_l/2)\Gamma(\beta_l/2-1)}, & \beta_l > 2, \ \eta \in \overline{V}^+. \end{cases} \tag{4.15}$$

If $N = 1$, Eqs. (4.13)–(4.15) express the case of quasiasymptotics of order $\beta - 4$ (see Refs. 11 and 33).

(3) The asymptotic behavior of regularly varying multiparticle distributions is effectively determined by the function h from Theorem 1.

The function h may be chosen such that

$$h = Ph_0,\ P,\ h_0 \in H(C), \tag{4.16}$$
$$P(A'z) = \chi(A)P(z), \quad h_0(A'z) = h_0(z), \quad z \in T^C, \quad A \in G,$$

where P is a polynomial without zeros in T^C.

We now present another example. Let $G = \mathrm{SO}(1,1)^\uparrow \otimes \mathrm{SO}(n-2) \otimes D_{\overline{V}^+}$ and for $A \in G$ we choose the parametrization

$$\begin{pmatrix} \cosh\Theta & \sinh\Theta \\ \sinh\Theta & \cosh\Theta \end{pmatrix} \otimes \lambda \mathbb{1}_{n-2}. \tag{4.17}$$

If

$$h(z) = [-i(z^0 - z^1)]^\mu[-i(z^0 + z^1)]^\nu(-z^2)^{-l/2}, \quad z \in T^{V^+}, \tag{4.18}$$

then

$$\chi(A) = \lambda^{\mu+\nu-l} e^{\Theta(\mu-\nu)}. \tag{4.19}$$

It can be verified by direct calculation that this character gives indeed the asymptotic behavior of the distribution $L^{-1}\{h\}(\xi)$. In fact, starting with formula (4.5) and using the fractional derivative from Sec. II (see, e.g., Ref. 34, Eqs. (8) and (77), Chap. XIII), it is found that the distribution whose Laplace transform is the function (4.18) has the expression

$$L^{-1}\{h\}(\xi) = \frac{u^{-\mu}v^{-\nu}}{\Gamma(1-\mu)\Gamma(1-\nu)}\,\frac{(-\xi'^2)^{(l-n)/2}}{H_n(l)} \times {}_3F_2(1, 1, (n-l)/2; 1-\mu, 1-\nu; uv/\xi'^2), \tag{4.20}$$

where $u = \xi^0 - \xi^1$, $v = \xi^0 + \xi^1$, $(\xi')^2 = (\xi^2)^2 + \cdots + (\xi^{n-1})^2$, $\xi \in \overline{V}^+$, and the calculation was performed for type function distributions for $\mathrm{Re}\,\mu$, $\mathrm{Re}\,\nu < 0$.

If $n = 4$, $G = \mathrm{SO}(1,1)^\uparrow \otimes \mathrm{SO}(2)$, we may take

$$h(z) = (z^0 - z^1)^\alpha (z^0 + z^1)^\beta h_0((z^2)^2 + (z^3)^2), \quad z \in V^+, \tag{4.21}$$

where the constants $\alpha, \beta \in \mathbb{R}$ specify the multiplier χ.

Note that the asymptotic behavior determined by Eq. (4.21) is compatible with the uniparticle cross sections of inclusive reactions at high energy predicted by the local quantum field theory[35,36] and also by the phenomenology of elementary particles (see, e.g., Ref. 37).

In fact, Eqs. (4.16) determine classes of asymptotic behaviors which are compatible with the bahaviors of the multiparticle amplitudes at high energies.

V. CONCLUSIONS

In this paper we have introduced the regularly varying distributions as tempered distributions with asymptotic behavior determined by the multipliers of the group of automorphisms of the support cone. A Tauberian–Abelian theorem for regularly varying distributions which generalizes the classical Hardy–Littlewood–Karamata theorem and also Vladimirov's theorem has been proved. This theorem establishes the equivalence of asymptotic behavior of the regularly varying distributions, of their Laplace transforms, and of the regularly varying distributions which result after the action of the Riemann–Liouville operator. The covariance properties of the limiting distributions and their Laplace transforms, which are connected by an integral equation, have also been established. The results have been applied to multiparticle distributions which are the product of regularly varying two-point distributions with support in a product of Minkowski cones.

The results of this paper can be extended to spaces of hyperfunctions,[38] and the asymptotic behaviors may be refined by introducing slowly varying functions.[14,29]

[1]C. Itzykson and J. B. Zuber, *Quantum Field Theory* (McGraw–Hill, New York, 1980).

[2]S. Berceanu, Rev. Roum. Phys. **20**, 831 (1975).
[3]N. N. Bogoliubov, V. S. Vladimirov, and A. N. Tavkhelidze, Theor. Math. Phys. **12**, 305 (1972).
[4]S. Berceanu, "On Tauberian theorems and inclusive processes," JINR, Dubna Communication E2-8542, 1975.
[5]D. J. Bjorken, Phys. Rev. **179**, 1547 (1969).
[6]R. Jackiw, R. Royen, and G. B. West, Phys. Rev. D **2**, 2473 (1970); H. Leutwyler and J. Stern, Nucl. Phys. B **20**, 77 (1970); R. Brandt and S. Preparata, Nucl. Phys. B **27**, 541 (1971); R. Gatto and P. Menotti, Phys. Rev. D **5**, 1493 (1972).
[7]B. I. Zavialov, Theor. Math. Phys. **17**, 178 (1973).
[8]M. Maag, Commun. Math. Phys. **38**, 225 (1974); J. Kühn, Nuovo Cimento A **23**, 420 (1974).
[9]E. Brüning and P. Stichel, Commun. Math. Phys. **36**, 137 (1974); "Asymptotics and light cone singularities in quantum field theory," in *Lecture Notes in Physics No. 39* (Springer, Berlin, 1975), p. 72.
[10]B. I. Zavialov, Theor. Math. Phys. **33**, 310 (1977).
[11]V. S. Vladimirov and B. I. Zavialov, Theor. Math. Phys. **40**, 155 (1979).
[12]V. S. Vladimirov and B. I. Zavialov, Theor. Math. Phys. **50**, 163 (1979).
[13]W. Feller, *An Introduction to Probability Theory and its Applications* (Wiley, New York, 1966), Vol. II, 2nd ed.
[14]E. Seneta, "Regularly varying functions," *Lecture Notes in Mathematics, No. 508* (Springer, Berlin, 1976).
[15]V. S. Vladimirov, Izv. Akad. Nauk. SSSR, Ser. Math. **40**, 1084 (1976).
[16]V. S. Vladimirov, *Generalized Functions in Mathematical Physics* (Nauka, Moscow, 1976) (in Russian).
[17]L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1966).
[18]M. Riesz, Acta Math. **81**, 1 (1949).
[19]P. D. Méthée, Comm. Math. Helv. **28**, 225 (1954).
[20]V. S. Vladimirov, *Methods of the Theory of Functions of Several Complex Variables* (Nauka, Moscow, 1964) (in Russian).
[21]S. G. Gindikin, Usp. Math. Nauk **19**, 3 (1964).
[22]I. I. Pyatetski-Shapiro, *Automorphic Functions and the Geometry of Classical Domains* (Gordon and Breach, New York, 1969).
[23]E. Hewitt and K. A. Ross, *Abstract Harmonic Analysis* (Springer, Berlin, 1963), Vol. I.
[24]S. Helgasson, *Differential Geometry, Lie Groups and Symmetric Spaces* (Academic, New York, 1978).
[25]I. M. Drojinov and B. I. Zavialov, Mat. Sb. **108**, 78 (1979).
[26]I. M. Drojinov and B. I. Zavialov, Mat. Sb. **102**, 372 (1977).
[27]V. S. Vladimirov, Siberian Math. J. **9**, 1238 (1968).
[28]J. Bros, H. Epstein, and V. Glaser, Commun. Math. Phys. **6**, 77 (1967).
[29]A. L. Iakimov, Mat. Sb. **115**, 463 (1981).
[30]J. Karamata, Mathematica (Cluj) **4**, 38 (1930).
[31]R. Jost, *The Generalized Theory of Quantized Fields* (Am. Math. Soc., Providence, RI, 1965).
[32]E. M. Sotkov and P. R. Zaikov, Rep. Math. Phys. **19**, 335 (1984).
[33]K. A. Bukin, Theor. Math. Phys. **40**, 28 (1979).
[34]H. Batteman and A. Erdely, *Tables of Integral Transforms* (McGraw–Hill, New York 1954), Vol. II.
[35]H. Araki and R. Haag, Commun. Math. Phys. **4**, 77 (1966).
[36]A. A. Logunov, M. A. Mestvirisvili, and A. V. Petrov, Theor. Math. Phys. **32**, 3 (1977).
[37]21st International Conference on High Energy Physics, Journal de Physique C3-1982.
[38]M. Sato, T. Kawai, and M. Kashiwara, "Hyperfunctions and pseudodifferential equations," in *Lecture Notes in Mathematics No. 287* (Springer, Berlin, 1973).

Convergent expansions for excited glueball masses in 2 + 1 strongly coupled lattice gauge theories

Michael O'Carroll[a)]
Harvard University, Lyman Laboratory of Physics, Cambridge, Massachusetts 02138

(Received 31 May 1984; accepted for publication 1 November 1984)

It is known that the mass spectrum of a strongly coupled ($\beta = 2/g^2$ small) 2 + 1 Wilson action lattice gauge theory contains a mass $m_0 \sim -4 \ln \beta$ and two excited masses $m_1, m_2 \sim -6 \ln \beta$ and that $m_0 + 4 \ln \beta$ has a convergent expansion in powers of β. We show that m_1, m_2 admit convergent expansions of the form $-6 \ln \beta + r(\beta)$, where $r(\beta)$ is analytic at $\beta = 0$. Furthermore, a finite lattice algorithm is given for determining c_n, the nth $\beta = 0$ Taylor coefficient of $r(\beta)$. Here, c_n only depends on a finite number of $\beta = 0$ Taylor series coefficients of the plaquette–plaquette, plaquette–double plaquette, and double plaquette–double plaquette truncated correlation functions at a finite number of points. For the gauge group Z_2, by duality, m_1, m_2 map to bound states of the low-temperature Ising model; a possible relation between an increasing number of bound states and roughening is discussed.

I. INTRODUCTION AND RESULTS

In the early work of Ref. 1 and more recently in Refs. 2 and 3 there has been much effort devoted to finding the mass spectrum of lattice gauge theories. In Ref. 4, for a Wilson action strongly coupled ($\beta = 2/g^2$ small) lattice gauge theory it is shown that particles exist and in Ref. 5 convergent expansions are obtained for the masses. In the 2 + 1 lattice gauge theory with a gauge group representation with real character it is shown[4] that there is a lowest mass m_0, associated with the truncated plaquette–plaquette (p–p) correlation function (cf), which admits the representation

$$m_0 = -4 \ln \beta + r(\beta),$$

where $r(\beta)$ is analytic at $\beta = 0$. Furthermore, in Ref. 6 a Z_4 symmetry for zero momentum states is deduced and exploited to show that there are masses m_1 and m_2 associated with the trivial and nontrivial real representations of Z_4, respectively. The asymptotic forms of m_1 and m_2 are $-6 \ln \beta$ and $|m_1 - m_2| = O(\beta)$. Here, m_0 is also associated with the identity representation; m_1 and m_2 are associated with the truncated double plaquette–double plaquette (dp–dp) cf.

Using the β analyticity and decay rate of various truncated cf's given in Ref. 6 we show that the methods of Ref. 5 apply to give a convergent expansion for m_2. The masses m_0 and $m_1 > m_0$ occur as simple poles on the imaginary axis of the complex energy plane of the momentum space truncated p–p cf. There is necessarily a simple zero ρ between these poles. A convergent expansion for m_1 is obtained from an implicit equation which is a perturbation of an implicit equation for ρ. We show that the masses m_2, m_1 (as well as ρ) admit expansions of the form

$$-6 \ln \beta + r(\beta),$$

where $r(\beta)$ is analytic at $\beta = 0$. Furthermore, a finite lattice algorithm is given for c_n, the nth $\beta = 0$ Taylor series coefficient of $r(\beta)$. Here, c_n depends on only a finite number of $\beta = 0$ Taylor series coefficients of the p–p, dp–dp, and p–dp cf's at a finite number of points.

In Sec. II we introduce notation and deduce the expansion for m_2. In Sec. III we obtain the expansion for m_1 (and ρ). Some concluding remarks are made in Sec. IV. In an appendix we give bounds on decay rates of various cf's used in Secs. II and III.

II. NOTATION AND EXPANSION FOR m_2

We consider a Wilson action 2 + 1 lattice gauge theory with Boltzmann factor given formally by $\exp\{\beta \sum_p \chi(g_p)\}$; χ is a real character of an irreducible unitary representation of the gauge group, p denotes the plaquettes of Z^3, and g_p is the ordered product of group elements around the border of P. We let $x = (x_0, x_1, x_2) = (x_0, \mathbf{x})$ denote points of Z^3; $|\beta|$ will be taken small throughout. For a gauge invariant function ϕ that depends on a finite number of bond variables we denote the thermodynamic limit expectations by $\langle \phi \rangle$. The existence, β analyticity, and translational invariance of $\langle \cdot \rangle$ follow from the polymer expansion of Ref. 7. For ϕ, ψ, which are gauge invariant functions of a finite number of variables in the time zero $x_0 = 0$ plane, we let $\phi(x)$, $\psi(x)$ denote the translation by x. The relation of the expectations to quantum field theory is given by the Feynman–Kac formula

$$\langle \bar{\phi} \psi(x_0, \mathbf{x}) \rangle = \langle \bar{\phi} \psi(-x_0, \mathbf{x}) \rangle$$
$$= (\phi^v, e^{-H|x_0|} e^{i\mathbf{P}\cdot\mathbf{x}} \psi^v)_{\mathscr{H}},$$

where the left side is used to construct the Hilbert space $\mathscr{H}$, energy-momentum operators H, $\mathbf{P}$, and the Hilbert space vectors ϕ^v, ψ^v on the right using the construction of Refs. 7 and 8.

With ϕ, ψ as above let $G_{\phi\psi}(x)$ denote the truncated cf

$$G_{\phi\psi}(x) = \langle \bar{\phi}(0) \psi(x) \rangle - \langle \bar{\phi}(0) \rangle \langle \psi(x) \rangle,$$

which decays exponentially in the distance between the supports of ϕ and ψ. Also, let

$$\hat{G}_{\phi\psi}(x_0) = \sum_{\mathbf{x}} G_{\phi\psi}(x_0, \mathbf{x})$$

and

$$\tilde{G}_{\phi\psi}(p_0) = \sum_{x_0} \hat{G}_{\phi\psi}(x_0) e^{ip_0 x_0},$$

[a)] On leave from Departamento de Física (ICEx), Universidade Federal de Minas Gerais, 30.000, Belo Horizonte, MG, Brazil.

 0022-2488/85/092342-04$02.50

which is the Fourier transform at zero space momentum. Here, $\hat{G}_{\phi\psi}(x_0)$ will also be considered as a matrix operator on $l_2(Z)$ with matrix elements $G_{\phi\psi}(x_0;y_0)=G_{\phi\psi}(x_0-y_0)$ and similarly for $G_{\phi\psi}(x)$ in $l_2(Z^3)$.

Let $R_{\mathbf{a}}$ denote a rotation of $\pi/2$ about an axis $\mathbf{a}$ parallel to the x_0 (time) direction. Then, $I, R_{\mathbf{a}}, R_{\mathbf{a}}^2, R_{\mathbf{a}}^3$ make up the elements of the group Z_4. Defining

$$P_{\mathbf{a}}^{(1)}=\tfrac{1}{4}(I+R_{\mathbf{a}}+R_{\mathbf{a}}^2+R_{\mathbf{a}}^3),$$
$$P_{\mathbf{a}}^{(2)}=\tfrac{1}{4}(I-R_{\mathbf{a}}+R_{\mathbf{a}}^2-R_{\mathbf{a}}^3),$$
$$P_{\mathbf{a}}^{(3)}=\tfrac{1}{4}(I+iR_{\mathbf{a}}-R_{\mathbf{a}}^2-iR_{\mathbf{a}}^3),$$
$$P_{\mathbf{a}}^{(4)}=\tfrac{1}{4}(I-iR_{\mathbf{a}}-R_{\mathbf{a}}^2+iR_{\mathbf{a}}^3),$$

we have $P_{\mathbf{a}}^{(i)}P_{\mathbf{a}}^{(j)}=\delta_{ij}P_{\mathbf{a}}^{(i)}$ and $\Sigma_i P_{\mathbf{a}}^{(i)}=1$. It is shown[6] that

$$\hat{G}_{P_{\mathbf{a}}^{(i)}\phi\psi}(x_0)=\hat{G}_{\phi,P_{\mathbf{b}}^{(i)}\psi}(x_0),$$

so that $\hat{G}_{\phi\psi}(x_0)=0$ if $\phi=P_{\mathbf{a}}^{(i)}\phi$, $\psi=P_{\mathbf{b}}^{(j)}\psi$, with $i\neq j$, which provides us with a selection rule on zero momentum states.

Let $\chi_h(g)=\chi(g_w)$, where g_w is the six-sided rectangular loop (double plaquette) in the time-zero plane located at the origin with long axis along x_1. Let $\chi_1=P_0^{(1)}\chi_h$ and $\chi_2=P_0^{(2)}\chi_h$. Denote the single plaquette function by χ.

The expansion for m_2 is obtained in a manner completely analogous to m_0 in Ref. 5, but using Lemma A.1 for $\hat{G}_{\chi_2\chi_2}(x_0)$ falloff and an expansion of $\tilde{\Gamma}_{\chi_2\chi_2}(p_0,\beta)$, $\tilde{\Gamma}_{\chi_2\chi_2}(p_0,\beta)=-\tilde{G}_{\chi_2\chi_2}(p_0,\beta)^{-1}$, with the terms up to and including order β^6 explicited. We have, letting

$$\hat{\Gamma}_{22S}(x_0,\beta)=\hat{\Gamma}_{22}(x_0,\beta)-\sum_{m=0}^{6}\frac{\beta^m}{m!}\frac{\partial^m\hat{\Gamma}}{\partial\beta^m}(x_0,\beta=0),$$

$$\tilde{\Gamma}_{22}(P_0,\beta)=2+\sum_{m=1}^{6}\gamma_m\beta^m-\frac{c_7}{6!}(e^{-ip_0}+e^{ip_0})\beta^6$$
$$+\hat{\Gamma}_{22S}(x_0=0,\beta)+\sum_{n=1}^{\infty}\hat{\Gamma}_{22S}(x_0=n,\beta)$$
$$\times(e^{-ip_0n}+e^{ip_0n}).$$

Introduce the auxiliary complex variable w and function $H(w,\beta)$ such that

$$H(w=2-(c_7/6!)\beta^6e^{-ip_0},\beta)=\tilde{\Gamma}_{22}(p_0,\beta).$$

By the analytic implicit function theorem[9] $H(w(\beta),\beta)=0$, $w(\beta)$ analytic, $w(0)=0$, and $w(\beta)=2-(c_7/6!)\beta^6e^{m_2(\beta)}$, or

$$m_2(\beta)=-6\ln\beta+\ln(6!/c_7)+\ln(2-w(\beta))$$
$$\equiv-6\ln\beta+r(\beta).$$

The argument for determining $c_n=(1/n!)(d^n r/d\beta^n)(\beta=0)$ goes as in Ref. 5 and reduces to the determination of $(d^m\hat{\Gamma}_{22}/d\beta^m)(x_0,\beta=0)$ which, by the Neumann series expansion for $\hat{\Gamma}_{22}$ in Lemma A.1, reduces to the determination of $(d^l G_{22}/d\beta^l)(x,\beta=0)$ for a finite number of x and l.

III. EXPANSION FOR m_1

In Ref. 6 it is shown that $\tilde{G}_{\chi\chi}(p_0)$ has two simple poles, at $p_0=\mathrm{i}m_0$ and $p_0=\mathrm{i}m_1$. Thus $\tilde{\Gamma}_{\chi\chi}(p_0)=-\tilde{G}_{\chi\chi}(p_0)^{-1}$ has simple zeroes at $p_0=\mathrm{i}m_0$, $p_0=\mathrm{i}m_1$. By the spectral representation of $\tilde{G}_{\chi\chi}(p_0)$, $\tilde{G}_{\chi\chi}(p_0)$ also has a simple zero at $p_0=i\rho$, $m_0<\rho<m_1$. To determine m_1 we first introduce

$$\tilde{F}_{\chi_1\chi_1}(p_0)=\tilde{G}_{\chi_1\chi_1}(p_0)+\tilde{G}_{\chi_1\chi}(p_0)\tilde{\Gamma}_{\chi\chi}(p_0)\tilde{G}_{\chi\chi_1}(p_0),$$

which subtracts out the vacuum and m_0, m_1 particle poles from $\tilde{G}_{\chi_1\chi_1}$. However, there is still a simple pole of $\tilde{\Gamma}_{\chi\chi}(p_0)$ at $p_0=i\rho$ coming from the zero at $p_0=i\rho$ of $\tilde{G}_{\chi\chi}(p_0)$.

We remark that using the estimates in Lemma A.2 on $\hat{F}_{\chi_1\chi_1}(x_0)$ and $\hat{\Phi}_{\chi_1\chi_1}(x_0)$, the convolution inverse of $-\hat{F}_{\chi_1\chi_1}(x_0)$, we can obtain directly a convergent expansion for ρ as the zero $\tilde{\Phi}_{\chi_1\chi_1}(p_0=i\rho)$ using the same method as in Sec. II. Furthermore, this is the only zero of $\tilde{\Phi}_{\chi_1\chi_1}(p_0)$. An indirect method for finding ρ was employed in Ref. 6.

We now derive an implicit equation for m_1. Let

$$\tilde{L}_{\chi\chi_1}(p_0)=\tilde{\Gamma}_{\chi\chi}(p_0)\tilde{G}_{\chi\chi_1}(p_0)\tilde{\Phi}_{\chi_1\chi_1}(p_0)\equiv\tilde{L}(p_0)=\tilde{L}_{\chi_1\chi}(p_0).$$

By the bounds of Lemma A.3 this is analytic up to $\operatorname{Im}p_0=-7(1-\epsilon)\ln\beta$; i.e., the singularities in $\tilde{\Gamma}_{\chi\chi}(p_0)$ and $\tilde{G}_{\chi\chi_1}(p_0)$ at $p_0=im_0$, $p_0=i\rho$, and $p_0=im_1$ are canceled. Define $\tilde{M}(p_0)$ by

$$\tilde{\Gamma}_{\chi\chi}(p_0)=\tilde{L}_{\chi\chi_1}(p_0)\tilde{F}_{\chi_1\chi_1}(p_0)\,\tilde{L}_{\chi_1\chi}(p_0)+\tilde{M}(p_0).$$

From the estimates of Lemma A.4 $\tilde{M}(p_0)$ is analytic on $0<\operatorname{Im}p_0<-7(1-\epsilon)\ln\beta$. Furthermore, it is shown[6] that $L_{\chi_1\chi}(p_0=i\rho)\neq0$ so that the only singularity of $\tilde{\Gamma}_{\chi\chi}(p_0)$ is $p_0=i\rho$. We are interested in the zero of $\tilde{\Gamma}_{\chi\chi}(p_0)$ at $p_0=im_1$, $m_1\sim-6\ln\beta$. Write

$$\tilde{\Gamma}_{\chi\chi}(p_0)=\tilde{F}_{\chi_1\chi_1}(p_0)\left[\tilde{L}^2(p_0)-\tilde{M}(p_0)\tilde{\Phi}_{\chi_1\chi_1}(p_0)\right].$$

Thus the zero of $\tilde{\Gamma}_{\chi\chi}(p_0)$ at $p_0=im_1\sim-i6\ln\beta$ is given by

$$\tilde{\Phi}_{\chi_1\chi_1}-\tilde{L}^2/\tilde{M}=0,$$

which is the implicit equation we now solve. From the bounds on $\hat{M}$, $\hat{L}$, and $\hat{\Phi}_{\chi_1\chi_1}$ we can write

$$\tilde{M}=-1+\sum_{i=1}^{6}v_i\beta^i+\sum_{i=4}^{6}\alpha_i\beta^i(e^{-ip_0}+e^{ip_0})+\tilde{M}_7,$$

$$\tilde{L}=\sum_{i=1}^{7}\gamma_i\beta^i+\tilde{L}_8,$$

$$\tilde{\Phi}_{\chi_1\chi_1}=-4+\sum_{i=1}^{6}\phi_i\beta^i+\rho\beta^6(e^{-ip_0}+e^{ip_0})+\tilde{\Phi}_{\chi_1\chi_1 7},$$

where the coefficients of β^i can be determined explicitly.

In the above $\tilde{M}_7$, $\tilde{\Phi}_{\chi_1\chi_1 7}(\tilde{L}_8)$ are the $\beta=0$ Taylor expansions of $\tilde{M}$, $\tilde{\Phi}_{\chi_1\chi_1}(\tilde{L})$ with the terms up to and including order $\beta^6(\beta^7)$ subtracted. To be more explicit

$$\tilde{M}_7=\hat{M}_7(x_0=0,\beta)+\sum_{n=1}^{\infty}\hat{M}_7(x_0=n,\beta)(e^{-ip_0n}+e^{ip_0n}),$$

$$\tilde{\Phi}_7=\hat{\Phi}_7(x_0=0,\beta)+\sum_{n=1}^{\infty}\hat{\Phi}_7(x_0=n,\beta)(e^{-p_0n}+e^{ip_0n}),$$

$$\tilde{L}_8=\hat{L}_8(x_0=0,\beta)+\sum_{n=1}^{\infty}\hat{L}_8(x_0=n,\beta)(e^{-ip_0n}+e^{ip_0n}),$$

where the subscript l means the Taylor series with the terms up to but not including l subtracted.

We obtain the correct singular and constant term for the mass m_1 by transforming to the variables w,β, where

$$w=-1+(c_7\beta^6/2\cdot6!)e^{-ip_0},\quad c_7=2\cdot6!/d^6.$$

We let $\tilde{M}',\tilde{\Phi}',\tilde{L}'$ denote $\tilde{M},\tilde{\Phi},\tilde{L}$, respectively, written in terms of the w,β variables. Now observe that we can write

$$\tilde{\Phi}' = 2w + g(w,\beta), \quad g(0,0) = 0, \quad \frac{\partial g}{\partial w}(0,0) = 0,$$

$$\beta^2 \tilde{M}' = (c_4/4!)(2\cdot 6!/c_7) + h(w,\beta), \quad h(0,0) = 0.$$

Furthermore, $\tilde{L}'$, g,h are jointly analytic and $(\beta^2\tilde{M}')^{-1}$ has a Taylor expansion beginning with a constant. Let

$$F(w,\beta) \equiv \tilde{\Phi}' - \beta^2 \tilde{L}'^2 (\beta^2 \tilde{M}')^{-1} = 2w + g(w,\beta) - \beta^2 \tilde{L}'^2(\beta^2\tilde{M}')^{-1}.$$

As $F(w,\beta)$ is jointly analytic, $F(0,0) = 0$, $(\partial F/\partial w)(0,0) = 2$, the analytic implicit function theorem applies and gives $w(\beta)$, where $w(\beta)$ is analytic, $F(w(\beta),\beta) = 0$, and $w(0) = 0$. Thus

$$m_1 = -6\ln\beta + 6\ln d + \ln(1 + w(\beta)).$$

Arguing as in Ref. 5, c_n, the nth $\beta = 0$ Taylor coefficient of $w(\beta)$, only depends on $(d^m\Phi/d\beta^m)(x;y,\beta=0)$, $(d^m M/d\beta^m)(x;y,\beta=0)$, and $(d^m L/d\beta^m)(x;y,\beta=0)$ for a finite number of x,y, and m, which by the Neumann expansions for Φ, M, and L only depends on $(d^r/d\beta^r) \times G_{\chi\chi}(x,\beta=0)$, $(d^s/d\beta^s) G_{\chi_1\chi_1}(x,\beta=0)$, and $(d^t/d\beta^t) \times G_{\chi\chi_1}(x,\beta=0)$ for a finite number of r,s,t, and x. The determination of the above cf $\beta = 0$ derivatives can be reduced to a finite lattice problem by expressing $G_{\chi\chi}$, $G_{\chi_1\chi_1}$, and $G_{\chi\chi_1}$ as appropriately differentiated logs of partition functions and using the polymer expansion (see Ref. 7).

IV. CONCLUDING REMARKS

We have obtained convergent expansions for the bound states below the order of magnitude of $-7\ln\beta$. The question of how many bound states there are below the two-particle continuum at the order of magnitude of $-8\ln\beta$ is left unanswered. For example, do the three or more plaquette functions give rise to new bound states? Also, we have not determined whether the bound states become more or less tightly bound as β increases; for example, do m_1/m_0, m_2/m_0 increase or decrease with β? This could be determined by calculating the coefficients of the series. Do the bound states disappear into the continuum for some β? It would be interesting to have a spectral interpretation of the crossover region. By duality,[10] for the special case of the gauge group Z_2, the masses m_0, m_1, m_2 are the one particle and the two bound states of the quantum field theory associated with the 2 + 1 low-temperature Ising model (see Ref. 8). The appearance or disappearance of bound states as the critical temperature is approached from below could shed light on possible roughening effects in the low-temperature Ising model with plus boundary conditions. A detailed knowledge of the low-temperature bound state spectrum would also allow improved scaling function approximates (sfa) (see Ref. 11). In Ref. 11, sfa were given assuming no bound states. The location of the bound state spectrum will affect the deviation from (and corrections to) Orstein–Zernike behavior.

ACKNOWLEDGMENTS

I thank Professor Arthur Jaffe for his hospitality at Harvard where this work was carried out.

This research was partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

APPENDIX: DECAY PROPERTIES

We give bounds, obtained in Secs. III and IV of Ref. 6, on G_{22}, Γ_{22}, F_{11}, Φ_{11}, L, and M used in the arguments of Secs. II and III. We have abbreviated $\chi_2\chi_2$, $\chi_1\chi_1$ by 22 and 11, respectively. Here, c, c', c_1, c_2,... will denote strictly positive constants. The cf's that appear below are analytic in β and all Neumann series converge in l_2 operator norm. Here, d is the dimension of the representation of the gauge group.

Lemma A.1:

(a) $\hat{G}_{22}(x_0 = 0) = \frac{1}{2} + O(\beta)$.

(b) $\hat{G}_{22}(x_0 = 1) = c_7/4\cdot 6!\,\beta^6 + O(\beta^7)$,

$c_7 = 2\cdot 6!/d^6$.

(c) $|\hat{G}_{22}(x_0)| \leqslant c|c'\beta|^{6|x_0|}$.

(d) $\hat{G}_{22} = \hat{P}_2[1 + \hat{P}_2^{-1}(\hat{G}_{22} - \hat{P}_2)]$,

$\hat{P}_2(x_0;y_0) = \hat{G}_{22}(x_0;y_0)\delta_{x_0,y_0}$.

(e) $$\hat{\Gamma}_{22} = -\hat{G}_{22}^{-1} = [1 + \hat{P}_2^{-1}(\hat{G}_{22} - \hat{P}_2)]^{-1}\hat{P}_2^{-1} = \sum_{n=0}^{\infty}(-1)^n[\hat{P}_2^{-1}(\hat{G}_{22} - \hat{P}_2)]^n\hat{P}_2^{-1}.$$

(f) $$|\hat{\Gamma}_{22}(x_0)| \leqslant \begin{cases} c_2|c'\beta|^{7|x_0|}, & |x_0| \neq 1, \\ c_2|c'\beta|^{6|x_0|}, & |x_0| = 1. \end{cases}$$

Lemma A.2:

(a) $\hat{F}_{11}(x_0 = 0) = \frac{1}{4} + O(\beta)$.

(b) $\hat{F}_{11}(x_0 = 1) = (c_7/6!4)\beta^6 + O(\beta^7)$.

(c) $$\hat{F}_{11} = \hat{G}_{11} + \hat{G}_{1\chi}\hat{\Gamma}_{\chi\chi}\hat{G}_{\chi 1} = \hat{G}_{11}[1 - \hat{\Gamma}_{11}\hat{G}_{\chi_1\chi}\hat{\Gamma}_{\chi\chi}\hat{G}_{\chi\chi_1}],$$

where

$$\hat{\Gamma}_{\chi\chi} = -\hat{G}_{\chi\chi}^{-1} = [1 + \hat{P}^{-1}(\hat{G}_{\chi\chi} - \hat{P})]^{-1}\hat{P},$$

$$\hat{P}(x_0,y_0) = G_{\chi\chi}(x_0,y_0)\delta_{x_0,y_0}.$$

(d) $$\hat{\Phi}_{11} = -\hat{F}_{11}^{-1} = [1 - \hat{\Gamma}_{11}\hat{G}_{\chi_1\chi}\hat{\Gamma}_{\chi\chi}\hat{G}_{\chi\chi_1}]^{-1}\hat{\Gamma}_{11}, \quad \hat{\Gamma}_{11} = \hat{G}_{11}^{-1},$$

$\hat{\Gamma}_{11}$ defined as for $\hat{\Gamma}_{\chi\chi}$ of (c).

(e) $\hat{\Phi}_{11}(x_0 = 0) = -4 + O(\beta)$.

(f) $\hat{\Phi}_{11}(x_0 = 1) = (c_7/6!)\beta^6 + O(\beta^7)$.

(g) $$|\Phi_{11}(x_0)| \leqslant \begin{cases} c_2|c'\beta|^{6|x_0|}, & |x_0| = 1, \\ c_2|c'\beta|^{7|x_0|}, & |x_0| \neq 1. \end{cases}$$

Lemma A.3:

(a) $\hat{L}(x_0 = 0) = (4/d)\beta + O(\beta^2)$.

(b) $\hat{L}(x_0 = 1) = c\beta^8 + O(\beta^9)$,

(c) $$|\hat{L}(x_0)| \leqslant \begin{cases} c|c'\beta|^8, & |x_0| = 1, \\ c|c'\beta|^{7|x_0|}, & |x_0| \neq 1. \end{cases}$$

Lemma A.4:

(a) $\hat{M}(x_0 = 0) = -1 + O(\beta)$.

(b) $\hat{M}(x_0 = 1) = (c_4/4!)\beta^4 + O(\beta^5)$,

$c_4 = 4\cdot 4!/d^4$.

(c) $$|\hat{M}(x_0)| \leqslant \begin{cases} c|c'\beta|^{4|x_0|}, & |x_0| = 1. \\ c|c'\beta|^{7|x_0|}, & |x_0| \neq 1. \end{cases}$$

[1] J. Kogut, D. K. Sinclair, and L. Susskind, Nucl. Phys. B **114**, 199 (1976).
[2] G. Münster, Nucl. Phys. B **190**, 439 (1981).
[3] B. Berg and A. Billoire, Nucl. Phys. B **221**, 109 (1983).
[4] R. Schor, Nucl. Phys. B **222**, 71 (1983).
[5] M. O'Carroll and G. Braga, J. Math. Phys. **25**, 2741 (1984).
[6] R. Schor, Commun. Math. Phys. **92**, 369 (1984).
[7] E. Seiler, "Gauge Theories as a Problem of Constructive Quantum Field Theory and Statistical Mechanics," in *Lecture Notes in Physics*, Vol. 159 (Springer, New York, 1982).
[8] R. Schor, Commun. Math. Phys. **59**, 213 (1978).
[9] E. Hille, *Analytic Function Theory* (Ginn, Boston, 1962). Vols. I and II.
[10] R. Balian, J. M. Drouffe, and C. Itzykson, Phys. Rev. D **11**, 2098 (1975).
[11] M. Fisher and H. Tarko, Phys. Rev. B **11**, 1217 (1975).

The vector form of the neutrino equation and the photon neutrino duality

Jouko Mickelsson[a)]
Research Institute for Theoretical Physics, University of Helsinki, Siltavuorenpenger 20 C, SF-00170 Helsinki, Finland

(Received 22 August 1984; accepted for publication 22 February 1985)

A new form of the vector wave equation for a neutrino is derived that makes the relation to the massive case more transparent than that given by Reifler. Some new aspects of the photon neutrino duality are discussed.

I. INTRODUCTION

In a recent article[1] Reifler gave a vectorial form of the two-component neutrino equation (Weyl equation) in an Abelian external potential. His idea was to use the 2–1 Cartan map from the spinor space to the space of complex three-vectors $F=(F_1,F_2,F_3)$ with $F^2=F_1^2+F_2^2+F_3^2=0$. The wave equation can be written completely in terms of F and the external vector potential. The advantage of the vectorial form as compared to the standard way of writing the neutrino wave equation lies in the fact that the first one is *holonomic*: in a curved space-time it is not necessary to make a choice of an orthonormal tetrad field in order to be able to write down the wave equation (spinors are defined only with respect to a space-time tetrad); furthermore, not in every space-time does there exist a global tetrad field.

The vectorial form of the massive Dirac equation was studied extensively by Kofink[2] but the equivalence with the spinor formulation was first established by Takabayasi[3] in great detail, motivated by the desire to give a magnetohydrodynamical interpretation of the Dirac field. Since then the vectorial form (in the massive case) has been studied by several authors from various aspects.[4]

In Ref. 5, I showed that the Takabayasi equations can be understood and derived in a simple way using the (natural) action of the group GL (2, $\mathbb{C}$) on Dirac spinors. In this paper I want to show that the massless case (two-component Weyl equation) can be handled in a similar way. The resulting equations will differ in form from those in Ref. 1, but of course they are equivalent. The profit of writing the wave equation in my way, Eq. (15), is that both the massive and massless equations are formally the same: the main difference (except for the value of the mass parameter) lies in the characteristics of the "polarization tensor" $P_{\mu\nu}$, which is the main bilinear quantity constructed of the Dirac (or Weyl) field. In the massive case at least one of the Lorentz invariants $P_{\mu\nu}P^{\mu\nu}$, $P_{\mu\nu}{*P}^{\mu\nu}$ has to be nonzero, whereas in the massless case both vanish and $F=E+iH$, $E_k=P_{0k}$, $H_k=-\frac{1}{2}\epsilon_{ijk}P^{ij}$ is the complex null vector used in Ref. 1.

The considerations of the present paper throw some light on the problem of the photon neutrino duality. It is an old idea to construct photons from neutrino pairs.[6] This was revived again in a recent article by Luther and Schotte.[7] In Ref. 7 it was shown how to construct photon creation and annihilation operators from pairs of fermion operators. The inverse construction is also possible. In fact the first construction of this type was given by Jordan, but it was criticized by Pryce because of the lack of rotational invariance. (Problems in combining the correct commutation relations with rotational invariance were not completely solved in the later papers.[6]) This difficulty has been avoided in Ref. 7. Since the vectorial equations for $P_{\mu\nu}$ are very similar to the Maxwell equations (the difference is in a source term which is a nonlinear function of P) one could somehow think of the neutrino as a "twisted" photon. This possibility will be discussed in Sec. III in the framework of functional integration, in the spirit of Witten's fermionization of the soliton field in the Wess–Zumino model.[8]

II. THE VECTORIAL FORM OF THE NEUTRINO EQUATION

Consider the Dirac equation

$$\gamma^\nu(\partial_\nu+ieA_\nu)\psi+im_0\psi=0 \tag{1}$$

in the Minkowski space; metric $g=\mathrm{diag}\,(+1,-1,-1,-1)$. I shall use the representation

$$\gamma_0=\begin{bmatrix}0 & \sigma_0\\ \sigma_0 & 0\end{bmatrix},\quad \gamma_k=\begin{bmatrix}0 & -\sigma_k\\ \sigma_k & 0\end{bmatrix}\quad (k=1,2,3)$$

where $\sigma_0=1$ and

$$\sigma_1=\begin{bmatrix}0&1\\1&0\end{bmatrix},\quad \sigma_2=\begin{bmatrix}0&-i\\i&0\end{bmatrix},\quad \sigma_3=\begin{bmatrix}1&0\\0&-1\end{bmatrix}.$$

As in Ref. 5, I shall replace (1) by an equivalent 2×2-matrix equation; the new matrix-valued Dirac field is related to the vector components $\psi_1,\ldots,\psi_4$ through

$$\psi=\begin{bmatrix}\psi_1 & -\bar\psi_4\\ \psi_2 & \bar\psi_3\end{bmatrix}.$$

In this notation the Dirac equation is

$$\sigma^\nu\partial_\nu\psi+ie\,\partial^\nu\sigma_3A_\nu+im_0\dot\psi\sigma_3=0, \tag{2}$$

where

$$\dot a=\begin{bmatrix}\bar\delta & -\bar\gamma\\ -\bar\beta & \bar\alpha\end{bmatrix},\quad \text{for } a=\begin{bmatrix}\alpha&\beta\\ \gamma&\delta\end{bmatrix}. \tag{3}$$

In particular, $\dot a=a^{*-1}$ if $\det a=1$.

In the case the mass $m_0=0$, Eq. (2) decouples into two independent two-component spinor equations (Weyl spinor equations). Let us project out the equation determined by the first column of the matrix equation (2); the projection operator for this is the right multiplication by the matrix ω

[a)] Permanent address: Department of Mathematics, University of Jyväskylä, Seminaarinkatu 15, SF-40100 Jyväskylä 10, Finland.

$=\frac{1}{2}(\sigma_1+i\sigma_2)$. The resulting two-component equation for (ψ_1,ψ_2) is

$$\sigma^\nu\,\partial_\nu\psi\,\omega-ie\sigma^\nu\psi\omega A_\nu=0\,. \tag{4}$$

Since (ψ_3,ψ_4) do not contribute to (4), we can choose, for example,

$$\psi=\begin{bmatrix}\psi_1 & -\bar\psi_2\\ \psi_2 & \bar\psi_1\end{bmatrix}. \tag{5}$$

However, in order to be better able to compare with the massive case let us keep the general form of ψ. Let us assume now that

$$\det\psi=\psi_1\,\bar\psi_3+\psi_2\,\bar\psi_4\neq 0\,. \tag{6}$$

We can multiply (4) from the right by $\det\psi\cdot\psi^{-1}$. Define

$$P:=\det\psi\cdot\psi\omega\psi^{-1}=P_{\mu\nu}\,\sigma^{\mu\nu}\,, \tag{7}$$

where $\sigma^{0k}=-\sigma^{k0}=\frac{1}{2}\sigma^k$, $\sigma^{kl}=-\frac{i}{2}\epsilon_{klj}\,\sigma^j$ ($\epsilon_{123}=+1$, ϵ antisymmetric). Taking the Hermitian and anti-Hermitian parts of

$$\det\psi\cdot\sigma^\nu(\partial_\nu\psi)\omega\,\psi^{-1}-ie\sigma^\nu A_\nu P=0\,, \tag{8}$$

one gets a pair of vectorial equations

$$\partial^\nu P_{\nu\mu}-2eA^{\nu*}P_{\nu\mu}=-\operatorname{Re}\operatorname{tr}P\,\partial_\mu\psi\,\psi^{-1}\,, \tag{9a}$$

$$\partial^{\nu*}P_{\nu\mu}+2e\,A^\nu P_{\nu\mu}=\operatorname{Im}\operatorname{tr}P\,\partial_\mu\,\psi\,\psi^{-1}\,. \tag{9b}$$

The dual $^*P$ is defined as

$$^*P_{\mu\nu}=\tfrac{1}{2}\,\epsilon_{\mu\nu\alpha\beta}\,P^{\alpha\beta}\,, \tag{10}$$

with $\epsilon_{0123}=+1$, ϵ antisymmetric. The field ψ defines an orthonormal tetrad system $\{u,l,m,n\}$ by

$$u^\nu\,\sigma_\nu=|\det\psi|^{-1}\cdot\psi\sigma_0\psi^*,\ldots,$$
$$n^\nu\,\sigma_\nu=|\det\psi|^{-1}\cdot\psi\sigma_3\psi^*\,. \tag{11}$$

From (7) and (11), it follows that

$$P_{\mu\nu}=(al_\mu+bm_\mu)(-u_\nu+n_\nu)-(al_\nu+bm_\nu)(-u_\mu+n_\mu)$$

or

$$P=(al+bm)\wedge k\,, \tag{12}$$

where $k_\nu=-u_\nu+n_\nu$ is a null vector and $a+ib=\det\psi$. Denoting by Λ the matrix with columns $\{u,l,m,n\}$,

$$\partial_\mu\psi\psi^{-1}=\partial_\mu\,\Lambda_{\alpha k}\,\Lambda_\beta{}^k\,\sigma^{\alpha\beta}+\tfrac{1}{2}\,\partial_\mu\,\ln\det\psi\,,$$

and Eqs. (9) can also be written as

$$\partial^\nu P_{\nu\mu}-2e\,A^\nu\,{}^*P_{\nu\mu}=(al+bm)\cdot\partial_\mu k\,, \tag{13a}$$

$$\partial^\nu\,{}^*P_{\nu\mu}+2e\,A^\nu P_{\nu\mu}=(am-bl)\cdot\partial_\mu k\,. \tag{13b}$$

Note that

$$^*P=(am-bl)\wedge(-u+n) \tag{14}$$

so that the duality transformation generates the group of rotations in the (a,b) plane. By (13), the potential A^ν can be interpreted as the gauge potential associated to this U(1) group. Thus we can write

$$D^\nu P_{\nu\mu}=(al+bm)\cdot\partial_\mu k\,, \tag{15a}$$

$$D^\nu\,{}^*P_{\nu\mu}=(am-bl)\cdot\partial_\mu k\,, \tag{15b}$$

where $D^\nu=\partial^\nu-2eA^\nu{}^*$. The field $P_{\mu\nu}$ and the vectors $al+bm$ and $am-bl$ are invariant with respect to (1) simultaneous rotations in the (a,b) and (l,m) planes; (2) the hyperbolic rotations,

$$u'=u\cosh\xi+n\sinh\xi\,,$$
$$n'=u\sinh\xi+n\cosh\xi\,,$$

combined with the scaling $(a',b')=e^{-\xi}(a,b)$; and (3) a two-parameter family of Lorentz transformations generated by $\sigma_{31}+\sigma_{01}=\frac{1}{2}\,\omega$ and $\sigma_{23}+\sigma_{20}=(i/2)\,\omega$. The maximal compact subgroup of this four-parameter group is just the U(1) gauge group. These symmetries arise because we are interested only in the first column of the (2×2)-matrix field ψ; the second column can be chosen arbitrarily subject only to the condition $\det\psi\neq 0$. A simple choice is that given by Eq. 5. In this case $a=|\psi_1|^2+|\psi_2|^2$, $b=0$, and $u=(1,0,0,0)$. By the definition, Eq. (7),

$$P=\begin{bmatrix}-\psi_1\psi_2 & \psi_1^2\\ -\psi_2^2 & \psi_1\psi_2\end{bmatrix}. \tag{16}$$

The map $(\psi_1,\psi_2)\to P$ is called the Cartan map and it is a 2–1 map from the space of two-component spinors to the space of antisymmetric tensor fields $P_{\mu\nu}$ satisfying

$$P_{\mu\nu}\,P^{\mu\nu}=P_{\mu\nu}\,{}^*P^{\mu\nu}=0\,. \tag{17}$$

If one writes

$$P=(E_k+iH_k)\sigma^k\,, \tag{18}$$

then P satisfies (17) if and only if

$$\mathbf{E}\perp\mathbf{H}\ \text{ and }\ \|\mathbf{E}\|=\|\mathbf{H}\|\,. \tag{19}$$

Any P subject to the condition (17) can be written as

$$P=h\wedge k\,, \tag{20}$$

where h is spacelike and k is a light vector orthogonal to h; the representation (20) is not unique, since clearly $h'=\rho h$, $k'=\rho^{-1}k$ is an acceptable choice for any real $\rho\neq 0$. Also $h'=h+tk$, $k'=k$ defines the same P. These symmetries correspond to the transformations (2) and (3) above.

Reifler wrote his neutrino equation entirely in terms of $P_{\mu\nu}$ and the vector j_μ,

$$j_0:=\|\mathbf{E}\|,\quad \mathbf{j}:=\|\mathbf{E}\|^{-1}\mathbf{E}\times\mathbf{H}\,, \tag{21}$$

defined by P. Using the symmetries described above, the apparent dependence of my equations (15a) and (15b) on k and $h=al+bm$ is completely artificial; of course, Eqs. (15a) and (15b) must be equivalent to (2.2) in Ref. 1:

$$D^0\mathbf{F}=i\mathbf{D}\times\mathbf{F}-(\mathbf{DF})\cdot\mathbf{j}\,j_0^{-1}\,,$$

where $\mathbf{F}=\mathbf{E}+i\mathbf{H}$.

The extra degrees of freedom can be avoided either by the choice of the second column in ψ [as in (5)] or directly by a specific choice of a, b, l, m, and k. A simple choice is

$$k=j,\ \ h=(0,\mathbf{E}/\|\mathbf{E}\|)=al+bm\,,$$
$$h^\perp=am-bl=(0,\mathbf{H}/\|\mathbf{H}\|)\,. \tag{22}$$

Using this choice it is easy to see the equivalence of (15) with (2.2).

It is interesting to compare the system (15) with the corresponding equations

$$\partial^\nu P_{\nu\mu}=2m_0\,v_\mu+2ea\,A_\mu+al\cdot\partial_\mu m+bu\cdot\partial_\mu n\,, \tag{23a}$$

$$\partial^\nu {}^*P_{\nu\mu} = 2ebA_\mu + bl\cdot\partial_\mu m - au\cdot\partial_\mu n\,, \tag{23b}$$

in the massive case $m_0 \neq 0$; here

$$v_\mu = \sqrt{a^2+b^2}\,u_\mu\,,\quad P = \det \psi\cdot\psi(i/2)\sigma_3\psi^{-1}\,. \tag{24}$$

Except for the mass term $2m_0 v_\mu$ there are two important differences: (1) the vector potential A_μ is (massive case) the gauge potential for rotations in the (l, m) plane, and (2) the tensor P is nondegenerate, $\det P \neq 0$. However, the first difference is only superficial (and is related to the second one). Namely, in the massless case the effect of a duality rotation on P by an angle ϕ in the (a, b) plane is the same as a rotation by the angle $-\phi$ in the (l, m) plane; this is due to the special degenerate form of P, as defined in (7) or in (12). Using this observation, both the massless equations (13) and the massive equations (23) can be written in a uniform manner:

$$D^\nu P_{\nu\mu} = 2m_0 v_\mu + \langle P, D_\mu \Lambda\Lambda^{-1}\rangle\,, \tag{25a}$$

$$D^\nu {}^*P_{\nu\mu} = \langle {}^*P, D_\mu\Lambda\Lambda^{-1}\rangle\,, \tag{25b}$$

where $\langle\cdot\,,\cdot\rangle$ is the Killing form, of so(3, 1), $\langle A, B\rangle = A_{\mu\nu} \times B^{\mu\nu}$; there is a 1–1 correspondence between antisymmetric tensors and elements of $\mathrm{so}(3,1)\cong\mathrm{sl}(2,\mathbb{C})$ by $A_{\mu\nu} \rightarrow A = A_{\mu\nu}\,\sigma^{\mu\nu}\in\mathrm{sl}(2,\mathbb{C})$. In the massive case $D^\nu P = \partial^\nu P$ whereas in the massless case the vector potential A_μ drops out from the right-hand side. In the degenerate case, Eq. (12), one has to put $m_0 = 0$; otherwise Eqs. (25) do not have nontrivial solutions (a two-component neutrino equation cannot have a mass term).

III. A GENERALIZATION AND THE PHOTON NEUTRINO DUALITY

The considerations above can be extended also to the case of electroweak interactions. The starting point is now the equation

$$\sigma^\nu\,\partial_\nu\psi + i\sigma^\nu\psi B_\nu + im_0\dot\psi f = 0\,, \tag{26}$$

where $B_\nu = B_\nu^{(k)}\,\sigma_k$ is a vector potential taking values in the Lie algebra $\mathrm{su}(2)\oplus\mathrm{u}(1)\cong\mathrm{u}(2)$ and $f = f_1\sigma_1 + f_2\sigma_2 + f_3\sigma_3$ is a Higgs field which transforms under the representation $f\rightarrow kfk^{-1}$ of the gauge group U(2), $\psi\rightarrow\psi k^{-1}$ and $B_\nu \rightarrow kB_\nu k^{-1} + i\,\partial_\nu k k^{-1}$. Let us now define the SU(2) invariant field

$$P = (i/2)\det\psi\cdot\psi f\psi^{-1}\,. \tag{27}$$

We can also define the SO(3) covariant derivative for vector fields by setting

$$\begin{aligned} D_\mu l &= \partial_\mu l + 2B_\mu^{(3)} m - 2B_\mu^{(2)} n\,,\\ D_\mu m &= \partial_\mu m + 2B_\mu^{(1)} n - 2B_\mu^{(3)} l\,,\\ D_\mu n &= \partial_\mu n + 2B_\mu^{(2)} l - 2B_\mu^{(1)} m\,, \end{aligned} \tag{28}$$

and $D_\mu u = 0$, the gauge group being the group of rotations in the three-space spanned by $\{l, m, n\}$. By (27) the U(1) part in the whole gauge group U(2) acts as duality rotations of $P_{\mu\nu}$ and so we set

$$D_\mu P_{\alpha\beta} = D_\mu P_{\alpha\beta} - 4B_\mu^{(0)}\,{}^*P_{\alpha\beta}\,. \tag{29}$$

(All coupling constants have been absorbed in the definition of B_ν.)

By multiplying (26) by $f\psi^*$ one gets, after some algebra, the vector form

$$D^\nu P_{\nu\mu} = 2m_0 v_\mu\,\mathrm{Re} f^2 + \langle P, D_\mu\Lambda\Lambda^{-1}\rangle\,, \tag{30a}$$

$$D^\nu {}^*P_{\nu\mu} = 2m_0 v_\mu\,\mathrm{Im} f^2 + \langle {}^*P, D_\mu\Lambda\Lambda^{-1}\rangle\,, \tag{30b}$$

where $f^2 = f_1^2 + f_2^2 + f_3^2$. In particular, Eq. (30) is the system of equations for the Dirac electron if $f = \sigma_3$ and the neutrino equation is obtained for $f = \sigma_1 + i\sigma_2$. In the general case, $P_{\mu\nu}$ must be treated as independent fields with respect to the tetrad $\Lambda = \{u, l, m, n\}$. This kind of transformation from the spinorial to the vectorial form was used in Ref. 9 for a unification of electroweak interactions with gravity.

One could make objections against using only vectorial fields to describe a spin-$\frac{1}{2}$ particle since the fermionic properties of the system are not apparent. However, as Witten has shown[8] there is a way around this problem. The idea in Ref. 8 was to add an anomaly term I to the SU(3) nonlinear σ-model Lagrangian $L = \mathrm{const}\times\int\partial^\nu U\,\partial_\nu U^*\,d^4x$ such that $\exp iI(U)$ produces a factor -1 in the quantum mechanical action when $U(\mathbf{x}, t)$ represents a soliton which is rotated by an angle 2π when the time t goes from $-\infty$ to $+\infty$. Thus the boson field U behaves like a fermion. Explicitly, $U(\mathbf{x}, t) = V(t)U_0(\mathbf{x})V(t)^{-1}$, where

$$V(t) = \begin{bmatrix} e^{it/2} & 0\\ 0 & e^{-it/2}\end{bmatrix},\ \text{time } 0 \leqslant t \leqslant 2\pi\,,$$

$$U_0(\mathbf{x}) = \exp\ if(r)\,(\mathbf{x}/r)\cdot\boldsymbol{\sigma}\,,$$

and $f(r)$ is a smooth function of $r = \|\mathbf{x}\|$ rising monotonically from 0 to 2π as r goes from 0 to ∞; the 2×2 matrix $U(\mathbf{x}, t)$ is embedded in SU(3) in the obvious way. As was noted in Ref. 8 one can quantize the σ field U as a fermion also in the case of SU(2) using the fact that $\pi_4(\mathrm{SU}(2)) = Z_2 = \{\pm 1\}$; one puts $(1/\pi)I = \pi_4(U)$ and gets again the factor -1 in the action since $\pi_4(U) = -1$ for the 2π-rotated soliton. The idea that π_4 could be responsible for the fermionization of a Bose field was investigated already by Finkelstein and Rubinstein in Ref. 10.

Now our P (in the massless case) is parametrized by the action of the group $\mathbb{R}_+\times\mathrm{SO}(3)$: any complex null vector can be written uniquely in the form $\lambda R\,(7, i, 0)$, where $\lambda\in\mathbb{R}_+$ and $R\in\mathrm{SO}(3)$. The fourth homotopy group of $\mathbb{R}_+\times\mathrm{SO}(3)$ is the same as that of SU(2). Starting from a bosonic Lagrangian for the (radiative) photon field $P_{\mu\nu}$ (e.g., the Maxwell Lagrangian) one can quantize a solitonic configuration [i.e., a $P_{\mu\nu}$ configuration which is parametrized by a SO(3) soliton] as a fermion by adding again $I(P)$ to the Langrangian. In this way one can think of the "twisted" radiative Maxwell field to represent a neutrino.

After completing this work I have learned about a new paper by Reifler[11] in which he constructs a vectorial equation for the four-component Dirac field, which is different from the earlier forms.[2–5] Instead of the tensor field $P = \det\psi\cdot\psi(i/2)f\psi^{-1}$ he uses the Yang–Mills triplet

$$\mathbf{F}_j = i\det\psi\cdot\psi\sigma_j\psi^{-1}\,,\quad j = 1, 2, 3\,. \tag{31}$$

The components of $\mathbf{F} = (\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3)$ are not independent:

$$\mathbf{F}_j\cdot\mathbf{F}_k := \mathrm{tr}\,\tfrac{1}{2}\,\mathbf{F}_j\,\mathbf{F}_k = -(\det\psi)^2\delta_{jk}\,. \tag{32}$$

The SO(3) gauge transformations in this picture are rotations of the components of F and the SO(2) duality rotations are a multiplication by a phase. In my notation,

$$\tfrac{1}{2}F_1 = am\wedge n + bu\wedge l\,,$$
$$\tfrac{1}{2}F_2 = an\wedge l + bu\wedge m\,, \qquad (33)$$
$$\tfrac{1}{2}F_3 = al\wedge m + bu\wedge n\,,$$

and one can see that the rotations in the three-dimensional space spanned by $\{l, m, n\}$ correspond precisely to the **F** rotations, and the (a, b)-duality rotations correspond to phase rotations via

$$F_j = (F_j)_{\mu\nu}\,\sigma^{\mu\nu} = (\mathbf{E}_j + i\mathbf{H}_j)\cdot\boldsymbol{\sigma}\in \mathbb{C}^3\,. \qquad (34)$$

Reifler also discusses the electroweak interactions; his approach is very close to my ideas in Ref. 9 but the choice of basic vector fields in Ref. 11 is different from Ref. 9.

ACKNOWLEDGMENT

I thank C. Montonen for calling to my attention Ref. 7 and for telling me about the history of the problem.

[1]F. Reifler, J. Math. Phys. **25**, 1088 (1984). Some special cases were studied by M. Flato and P. Hillion in Phys. Rev. D **1**, 1667 (1970).

[2]W. Kofink, Ann. Phys. **38**, 421, 565 (1940).

[3]T. Takabayasi, Suppl. Prog. Theor. Phys. **4**, 1 (1957).

[4]J. des Cloizeaux, Saclay preprint, 1982; F. A. Kaempffer, Phys. Rev. D **23**, 918 (1981); S. G. Mikhov and D. Ts. Stoyanov, preprint E2-12865, Dubna, 1979; Y. Takahashi, J. Math. Phys. **24**, 1783 (1983); H. Hestenes, J. Math Phys. **14**, 893 (1973); R. Boudet, C. R. Acad. Sci. (Paris) **272**, 767 (1971); **278**, 1063 (1974).

[5]J. Mickelsson, Lett. Math. Phys. **6**, 221 (1982).

[6]L. de Broglie, C. R. Acad. Sci. (Paris) **195**, 862 (1932); **197**, 1377 (1933); **199**, 1165 (1934); P. Jordan, Z. Phys. **93**, 464 (1935); **98**, 759 (1936); **99**, 109 (1936); **102**, 243 (1936); **105**, 229 (1937); L. de Broglie and J. Winter, C. R. Acad. Sci. (Paris) **199**, 813 (1934); N. S. Nagendra Nath, Proc. Ind. Acad. Sci. **3**, 448 (1936); Sokolow, Phys. Z. Sowiet. **12**, 148 (1937); R. de L. Kronig, Physica **2**, 968 (1935); M. H. L. Pryce, Proc. R. Soc. London Ser. A **165**, 247 (1938). Renaissance of the neutrino theory of light, 1957–1972: K. M. Case, Phys. Rev. **106**, 1316 (1957); N. Rosen and P. Singer, Bull. Res. Counc. Israel F **8**, 51 (1959); I. M. Barbour, A. Bietti, and B. F. Touschek, Nuovo Cimento **28**, 452 (1963); D. K. Sen, Nuovo Cimento **31**, 660 (1964); B. Ferretti, and I. Venturi, Nuovo Cimento **35**, 644 (1964); B. Jouvet, Nuovo Cimento **38**, 951 (1965); P. Bandyopadhyay, Nuovo Cimento **38**, 1912 (1965); W. A. Perkins, Phys. Rev. B **137**, 1291 (1965); Phys. Rev. D **5**, 1375 (1972); V. S. Berezinskii, Zh. Eksp. Teor. Fiz. **51**, 1374 (1966) [Sov. Phys. JETP **24**, 927 (1967)].

[7]A. Luther and K. D. Schotte, Nucl. Phys. B **242**, 407 (1984).

[8]E. Witten, Nucl. Phys. B **223**, 433 (1983).

[9]J. Mickelsson, Czech. J. Phys. B **32**, 521 (1982).

[10]D. Finkelstein and J. Rubinstein, J. Math. Phys. **9**, 1762 (1968).

[11]F. Reifler, J. Math. Phys. **26**, 542 (1985).

Probability density function of the single eigenvalue outside the semicircle using the exact Fourier transform

Nazakat Ullah[a)]
Max-Planck-Institut für Kernphysik, 69 Heidelberg, West Germany

(Received 25 May 1984; accepted for publication 9 May 1985)

It is shown that the probability density function far outside the semicircle is closely related to the form factor of N one-dimensional nucleons. The exact Fourier transform of the probability density function for the Gaussian unitary ensemble is given. Using this transform it is shown that the probability density function $P(x)$ far outside the semicircle is a Gaussian function multiplied by powers of X.

I. INTRODUCTION

The random-matrix ensembles were introduced by Wigner[1] to study the statistical properties of the compound-nucleus level widths and their positions. It was soon established that the dominant form of the single eigenvalue probability density function is a semicircle. Since then various other matrix ensembles[2] have been introduced to study many other properties of the many-body systems. In these studies one is not only interested in the behavior of the distribution of the single eigenvalue in the central region but also in its behavior outside the semicircle.

In the present work we would like to show that the probability density function far outside the semicircle is closely related to the form factor of N one-dimensional nucleons. This approach is like the one used by Mehta and Gaudin[3] to find the semicircular distribution using the known density of the states of the one-dimensional nuclear system. However, in the present case the one-dimensional form factor is not known. We, therefore, use the exact Fourier transform of the single-eigenvalue probability density function to find its behavior far outside the semicircle. To keep the formulation simple we shall consider here only the Gaussian unitary ensemble (GUE).

We describe the formulation in Sec. II and the concluding remarks in Sec. III.

II. FORMULATION

The single eigenvalue probability density function $P(x)$ using the δ function technique is given by[4]

$$P(x)=\frac{1}{N}\int \mathrm{tr}[\delta(x-H)]\,P(H)dH, \tag{1}$$

where $P(H)$ denotes the distribution of Hamiltonian matrix elements and tr denotes the trace of the operator.

We now consider the Gaussian unitary ensemble (GUE) for which expression (1) can be written as[1]

$$P(x)=\frac{1}{N}\int \delta(x-t)\sum_{m=0}^{N-1}\phi_m^2(t)\,dt, \tag{2}$$

where the ϕ_m are normalized harmonic oscillator wave functions and N is the dimension of the Hamiltonian matrix.

Using the Fourier transform of the δ function we can write

$$P(x)=\frac{1}{N}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}dk\,\exp(ikx)\times\sum_{m=0}^{N-1}\langle\phi_m|\exp(-ikt)|\phi_m\rangle, \tag{3}$$

where the $\langle\ \rangle$ sign denotes the matrix element in the basis of the harmonic oscillator wave functions.

As mentioned in the Introduction, we are interested in the large-x behavior of $P(x)$. From expression (2) it is obvious that if x is large, then most of the contribution will come from small values of k. For small values of k we must expand $\exp(-ikt)$ in terms of spherical Bessel functions. This is the same expansion that one uses in the calculation of form factors of nuclei and is known as the long-wavelength approximation.

Thus the problem of finding $P(x)$ for large values of x is reduced to finding the Fourier transform of the form factor of N one-dimensional nucleons. In three dimensions one knows that this form factor is Gaussian and has small oscillations as the value of the momentum transfer increases. However, no simple expression is known for the form factor in the one-dimensional case. We would now like to show that one can write the exact Fourier transform of $P(x)$ and use it to find the behavior of $P(x)$ for large values of x far outside the semicircle.

From expression (3), the Fourier transform $g(\alpha)$ of $P(x)$ is given by[5]

$$g(\alpha)=\frac{1}{\sqrt{2\pi}}\frac{1}{N}\exp\left(-\frac{\alpha^2}{4}\right)\sum_{m=0}^{N-1}L_m\left(\frac{\alpha^2}{2}\right). \tag{4}$$

The sum over Laguerre polynomials L_m can be written in terms of the associated Laguerre polynomial[5] and $g(\alpha)$ can therefore be written as

$$g(\alpha)=\frac{1}{\sqrt{2\pi}}\frac{1}{N}\exp\left(-\frac{\alpha^2}{4}\right)L_{N-1}^{(1)}\left(\frac{\alpha^2}{2}\right). \tag{5}$$

Expression (6) gives the exact Fourier transform of the single probability density function.

We now shall use the exact Fourier transform to find the behavior of $P(x)$. Using the explicit expression[5] for $L_{N-1}^{(1)}(x)$ it can be shown that $\exp(-\alpha^2/4)(1/N)L_{N-1}^{(1)}(\alpha^2/2)$ can be written as

[a)] Present and permanent address: Tata Institute of Fundamental Research, Bombay 400 005, India.

 0022-2488/85/092350-02$02.50

$$\exp(-\alpha^2/4)L_{N-1}^{(1)}(\alpha^2/2) = \frac{2J_1(\sqrt{2N}\alpha)}{\sqrt{2N}\alpha} + \frac{1}{96}\exp\left(-\frac{N\alpha^2}{2}\right)[\alpha^4 + \cdots], \quad (6)$$

where $J_1(x)$ is the Bessel function.

On taking the Fourier transform of expression (6) we find that the Bessel function part gives the semicircular distribution, while the second term gives the behavior of $P(x)$ for large values of x. Using expressions (5) and (6) and the Fourier transform, $P(x)$ is given by

$$P(x) = \frac{1}{\pi N}\sqrt{2N - x^2} + \frac{9\sqrt{6\pi}}{N^2} \times \exp\left(-\frac{3x^2}{N}\right)\left[\frac{1}{4} - \frac{9}{2N}x^2 + \frac{3x^4}{N^2}\right], \quad (7)$$

where the square root function is taken to be zero for $x^2 > 2N$, which is outside the semicircle.

III. CONCLUDING REMARKS

We have shown that the behavior of the single eigenvalue probability density function $P(x)$ far outside the semicircle is closely related to the form factor of N one-dimensional nucleons. The exact Fourier transform turns out to be a Gaussian multiplied by an associated Laguerre polynomial $L_{N-1}^{(1)}(x)$. Using the expansion of $L_{N-1}^{(1)}(x)$ it turns out that $P(x)$ far outside the semicircle falls as a Gaussian multiplied by some function of powers of x. This form of $P(x)$ is somewhat different than the one given by Bronk,[6] where it is shown to be exponentially falling. There is only one numerical study of $P(x)$ given in Wigner's unpublished article.[7] However, since the values far outside semicircle are quite small, it will be hard to say whether the fall of $P(x)$ far outside the semicircle is exponential or Gaussian multiplied by a power of x. Both the forms seem to be consistent with the one given in Wigner's article.

ACKNOWLEDGMENT

I would like to thank Professor Hans A. Weidenmüller for his warm hospitality during my stay at Max-Planck-Institut für Kernphysik.

[1]M. L. Mehta, *Random Matrices* (Academic, New York, 1967).

[2]S. F. Edwards and R. C. Jones, J. Phys. A **9**, 1595 (1976).

[3]M. L. Mehta and M. Gaudin, Nucl. Phys. **18**, 420 (1960).

[4]N. Ullah, J. Stat. Phys. **24**, 413 (1981).

[5]H. Bateman, *Higher Transcendental Functions*, edited by A. Erdelyi (McGraw–Hill, New York, 1953), Vol. 2, Chap. X; *Handbook of Mathematical Functions*, edited by M. Abramowitz and A. Stegun (Dover, New York, 1965), Chap. 22.

[6]B. V. Bronk, J. Math. Phys. **5**, 217 (1964).

[7]E. P. Wigner, "Distribution laws of the roots of a random Hermitean matrix," 1962. A reprint of this article can be found in C. E. Porter, *Statistical Theories of Spectra: Fluctuations* (Academic, New York, 1965), p. 446.

Statistical theory of scattering and propagation through a rough boundary–The Bethe–Salpeter equation and applications

K. Furutsu
Nakato 187-6, Musashi-Murayama, Tokyo 190-12, Japan

(Received 29 February 1984; accepted for publication 12 March 1985)

A systematic theory is given for a general class of scalar waves by introducing a surface Green's function, which is a 2×2 matrix function governed by boundary equations, transferred onto two reference boundary planes enclosing the real boundary inside. It is subjected to several symmetries, including a relation eventually leading to optical relations. Governing equations of statistical surface Green's functions of first and second orders are obtained unperturbatively in exactly the same way as for a random medium, on replacing the medium to a surface impedance. Two operator methods are introduced to obtain the surface impedance and integral equations of reflection–transmission coefficients exactly for a given boundary change. The space Green's functions outside the boundary are obtained by a simple continuation of the surface Green's functions, and scattering cross sections are obtained from their asymptotic expressions at large distances. Various quantities and equations associated with the incoherent waves are written exactly by the introduction of a scattering matrix, as contrasted with the conventional one for a coherent scatterer. A slightly rough boundary is investigated, with the cross sections for both reflected and transmitted waves, where obtained equations are significant also for higher-order effects, including multiple scattering. An application to boundary-value problems in layer transport is suggested.

I. INTRODUCTION

Theory of scattering by a rough surface has been investigated by many authors[1–14] in the last two decades, mostly for the two cases of either when the surface is slightly rough or changes over a large scale, and also for the case of randomly distributed bosses, as investigated in Refs. 7 and 8, and, more generally, in Ref. 14. When the surface is slightly random, the incoherent scattering cross sections have been obtained even for a boundary of finite dielectric constant of electromagnetic waves,[3–5,14] while, for a large-scale rough surface, the so-called tangent plane approximation has been conventionally employed as the basic means,[2,6] with results that are obviously contradictory with the power conservation, although much attention has not been paid to that point. To improve this aspect, a shadowing function has been introduced,[15–18] but the problem should not be solved with this function (even though partially done) since the basic defect remains unchanged. The transmitted waves were also investigated,[19] on taking into account the multiple reflection by use of a series-summation technique and the result of Ref. 2 (which does not satisfy the power conservation). On the other hand, a quite different method of approach has been tried, based on the extinction theorem (or the extended boundary condition) in Refs. 10–13 and others.[20]

In a previous paper,[14] the scattering by a one-side boundary (making only reflection without transmission) was investigated systematically by introducing a surface Green's function that is determined only by the boundary equation and, nevertheless, given in exactly the same form as the ordinary Green's function in a random medium. For example, the scattering cross sections were obtained for both slightly rough and large-scale surfaces, with a particular emphasis on the power conservation; an operator technique was introduced and extensively utilized to evaluate a surface impedance to be used in the boundary equation, exactly and in a compact form for given surface change; and an integral equation for the reflection matrix was also derived therefrom, which is bounded over the entire range of real values of the variable of Fourier transformation, as should be, but in contrast with those obtained according to the extinction theorem. Also, an equation for the second-order surface Green's function was obtained unperturbatively in a form of the Bethe–Salpeter (BS) equation, and was then continued to the space outside the boundary; the solution was written in terms of an incoherent scattering matrix, which enables various quantities and equations associated with the incoherent wave to be written exactly, including the scattering cross section and related optical relations.

In this paper, essentially the same method is applied to a two-side rough boundary which lets the waves also transmit through; a general class of scalar waves is considered, except in Appendix B where a modified version of several basic equations is shown for electromagnetic waves. A slightly random boundary is treated rather separately and in some detail because of a particular difficulty associated with the approximation, and cross sections are obtained to first order for both reflected and transmitted waves.

II. PRELIMINARIES AND SURFACE GREEN'S FUNCTION

In this paper, the rough surface is assumed to be plane on average, and the coordinate vector in three-dimensional space is denoted by $\hat{x} = (x_1, x_2, x_3) = (\boldsymbol{\rho}, z)$ in terms of the two-dimensional coordinate vector $\boldsymbol{\rho} = (x_1, x_2)$ and $z = x_3$, where the z axis is taken in the direction normal to the average boundary (Fig. 1). The scalar product of two space vectors

 0022-2488/85/092352-18$02.50

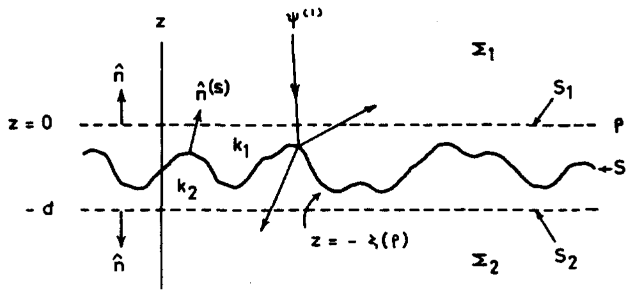


FIG. 1. Geometry and notations for Eqs. (2.10) and (2.17).

$\hat{a}=(\mathbf{a},a_z)$ and $\hat{b}=(\mathbf{b},b_z)$ will be denoted by $\hat{a}\cdot\hat{b} = \mathbf{a}\cdot\mathbf{b}+a_zb_z$, where $\mathbf{a}\cdot\mathbf{b}=a_1b_1+a_2b_2$.

A scalar wave function $\psi(\hat{x})e^{i\omega t}$, $\omega>0$, $t=$ time, is considered, giving the wave equation by

$$\left[-\left(\frac{\partial}{\partial z}\right)^2-\left(\frac{\partial}{\partial \boldsymbol{\rho}}\right)^2-k^2\right]\psi(\hat{x})=j(\hat{x}), \tag{2.1a}$$

where k may have an infinitesimal negative imaginary part whenever necessary, to mean a slight dissipation; $j(\hat{x})$ provides a source term. More often in the following, the $\boldsymbol{\rho}$ coordinates will be suppressed, expressing the wave equation (2.1a) by

$$\left[-\left(\frac{\partial}{\partial z}\right)^2-h^2\right]\psi(z)=j(z), \tag{2.1b}$$

in terms of an operator h, defined by

$$h=\left[k^2+\left(\frac{\partial}{\partial \boldsymbol{\rho}}\right)^2\right]^{1/2}, \tag{2.2}$$

where $\mathrm{Im}(h)<0$ in terms of the eigenvalues given by the Fourier transform, to be given by Eq. (2.7).

Hence, the solution of Eq. (2.1b) in free space, say $\psi^{(0)}(z)$, is given by

$$\psi^{(0)}(z)=\int dz'(2ih)^{-1}\exp\left[-ih\,|z-z'|\right]j(z'), \tag{2.3}$$

in the same way as when obtaining the solution in one-dimensional space. Here, the full coordinate expression $\psi^{(0)}(\boldsymbol{\rho},z)$ of $\psi^{(0)}(z)$ can be obtained from Eq. (2.3) by substitution of a corresponding expression $j(\boldsymbol{\rho},z')$ of $j(z')$, given by

$$j(\boldsymbol{\rho},z')=\int d\boldsymbol{\rho}'\,\delta(\boldsymbol{\rho}-\boldsymbol{\rho}')j(\boldsymbol{\rho}',z'), \tag{2.4a}$$

with the integral representation

$$\delta(\boldsymbol{\rho}-\boldsymbol{\rho}')=(2\pi)^{-2}\int_{-\infty}^{\infty}d\boldsymbol{\lambda}\exp[-i\boldsymbol{\lambda}\cdot(\boldsymbol{\rho}-\boldsymbol{\rho}')], \tag{2.4b}$$

hence,

$$\psi^{(0)}(\boldsymbol{\rho},z)=\int d\boldsymbol{\rho}'\int dz'\,g^{(0)}(\boldsymbol{\rho}-\boldsymbol{\rho}',z-z')j(\boldsymbol{\rho}',z'). \tag{2.5}$$

Here,

$$\begin{aligned} g^{(0)}&(\boldsymbol{\rho}-\boldsymbol{\rho}',z-z')\\ &=(2ih)^{-1}\exp[-ih\,|z-z'|]\delta(\boldsymbol{\rho}-\boldsymbol{\rho}')\\ &=(2\pi)^{-2}\int d\boldsymbol{\lambda}[2i\tilde{h}(\boldsymbol{\lambda})]^{-1}\\ &\quad\times\exp[-i\boldsymbol{\lambda}\cdot(\boldsymbol{\rho}-\boldsymbol{\rho}')-i\tilde{h}(\boldsymbol{\lambda})|z-z'|]\\ &=|4\pi(\hat{x}-\hat{x}')|^{-1}\exp[-ik\,|\hat{x}-\hat{x}'|], \end{aligned} \tag{2.6}$$

where

$$\tilde{h}(\boldsymbol{\lambda})=\begin{cases}(k^2-\boldsymbol{\lambda}^2)^{1/2}, & |\boldsymbol{\lambda}|\leqslant k,\\ -i(\boldsymbol{\lambda}^2-k^2)^{1/2}, & |\boldsymbol{\lambda}|>k.\end{cases} \tag{2.7}$$

As illustrated in Fig. 1, the entire rough boundary S is assumed to be described by $z=-\zeta(\boldsymbol{\rho})<0$ and also to be perfectly nondissipative for the time being; the unit vector normal to S will be denoted by $\hat{n}^{(S)}=(\mathbf{n}^{(S)},n_z^{(S)})$, $n_z^{(S)}>0$. Hence, the space is divided into two parts Σ_1 of $z>-\zeta$ and Σ_2 of $z<-\zeta$, with the propagation constants k_1 and k_2, respectively. Also, two reference boundary planes S_1 at $z=0$ and S_2 at $z=-d$ are introduced in such a way that S is completely involved in the space enclosed by S_1 and S_2. Here, the unit vectors directed outward normal to S_1 and S_2 are both denoted by $\hat{n}$, and the notation $\partial_n=\hat{n}\cdot\partial/\partial\hat{x}$ will often be used.

The boundary condition on S is assumed to be the continuity of

$$\psi|_S \quad\text{and}\quad \eta^{-1}\hat{n}^{(S)}\cdot\frac{\partial}{\partial\hat{x}}\psi|_S, \tag{2.8}$$

with some real constant η characterizing the medium, where η may be the density (sound waves), the dielectric constant ϵ, or the magnetic susceptibility $\mu=1$ (electromagnetic waves of vertical or horizontal polarization, respectively, in two-dimensional space). Consistently with the boundary condition, the power vector $\widehat{W}$ is defined by

$$\widehat{W}=(2i\eta)^{-1}\psi^*\left(\frac{\overleftarrow{\partial}}{\partial\hat{x}}-\frac{\overrightarrow{\partial}}{\partial\hat{x}}\right)\psi, \tag{2.9}$$

whose component normal to S is ensured to be continuous. Here, the boundary condition can be transferred onto the two reference planes S_1 and S_2, as will be shown in Sec. III, in the form

$$\begin{aligned} -\eta_1^{-1}\partial_n\psi_1&=B_{11}\psi_1+B_{12}\psi_2,\\ -\eta_2^{-1}\partial_n\psi_2&=B_{21}\psi_1+B_{22}\psi_2. \end{aligned} \tag{2.10}$$

Here, ψ_j and η_j denote ψ on S_j and η in space Σ_j, respectively; B_{ij}, $i,j=1,2$, are $\boldsymbol{\rho}$ operators (i.e., functions of $\boldsymbol{\rho}$ and/or $\partial/\partial\boldsymbol{\rho}$). Generally, any $\boldsymbol{\rho}$ operator, say Q, can be represented by a $\boldsymbol{\rho}$-coordinate matrix having matrix elements $Q(\boldsymbol{\rho}|\boldsymbol{\rho}')$, defined by

$$\begin{aligned} Q(\boldsymbol{\rho}|\boldsymbol{\rho}')&=Q\delta(\boldsymbol{\rho}-\boldsymbol{\rho}')\\ &=(2\pi)^{-2}\int_{-\infty}^{\infty}d\boldsymbol{\lambda}\,Q\exp[-i\boldsymbol{\lambda}\cdot(\boldsymbol{\rho}-\boldsymbol{\rho}')] \end{aligned} \tag{2.11a}$$

(where Q affects only $\boldsymbol{\rho}$ and not $\boldsymbol{\rho}'$). Hence, for any $f(\boldsymbol{\rho})$, it holds the rule

$$Qf(\rho) = \int d\rho'\, Q(\rho|\rho')f(\rho'). \qquad (2.11b)$$

Hereinafter, each B_{ij} will be understood to represent also a ρ matrix having the elements $B_{ij}(\rho|\rho')$; hence, Eq. (2.10) can be written explicitly as

$$-\eta_j^{-1}\,\partial_n \psi_j(\rho) = \sum_{i=1,2} \int d\rho'\, B_{ji}(\rho|\rho')\psi_i(\rho'). \qquad (2.12)$$

In addition, the B_{ij}'s will be referred to as the surface impedance, and will be obtained for given $\zeta(\rho)$ in Sec. III.

The boundary equation (2.10) can be written in a compact form, by introducing 2×2 matrices $\mathbf{B}$, $\boldsymbol{\eta}$ and a two-column vector $\boldsymbol{\psi}$, written by boldface letters and defined by the elements

$$\mathbf{B} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \quad \boldsymbol{\eta} = \begin{pmatrix} \eta_1 & 0 \\ 0 & \eta_2 \end{pmatrix}, \quad \boldsymbol{\psi} = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}, \qquad (2.13a)$$

as

$$-\boldsymbol{\eta}^{-1}\,\partial_n \boldsymbol{\psi} = \mathbf{B}\boldsymbol{\psi}, \quad -\boldsymbol{\eta}^{-1}\,\partial_n \boldsymbol{\psi}^* = \boldsymbol{\psi}^*\mathbf{B}^\dagger. \qquad (2.13b)$$

Here, the last is the complex-conjugate equation, and $\mathbf{B}^\dagger$ denotes the Hermitian-conjugate matrix of $\mathbf{B}$ with respect to both the coordinates and the indices, i.e., $B^\dagger_{ij}(\rho|\rho') = B^*_{ji}(\rho'|\rho)$.

Here, the total power emitted away from the boundaries $S_1 + S_2$ is given by the integrals

$$W \equiv \sum_{j=1,2} \int_{S_j} d\rho (2i\eta_j)^{-1} \psi_j^* (\overleftarrow{\partial}_n - \vec{\partial}_n)\psi_j \qquad (2.14a)$$

$$= \sum_{i,j=1,2} (2i)^{-1} \int d\rho\, d\rho'\, \psi_j^*(\rho)[B_{ji}(\rho|\rho') - B^\dagger_{ji}(\rho|\rho')]\,\psi_i(\rho'), \qquad (2.14b)$$

which should be zero in the present case, showing that $\mathbf{B}$ is a Hermitian matrix, subjected to the condition

$$\mathbf{B}^\dagger = \mathbf{B}. \qquad (2.15)$$

To obtain the general solution of homogeneous wave equation (2.1b) for $j(z) = 0$, we introduce an operator solution $\psi^{(1)}(z)$ [which should not be confused with the c-number wave function $\psi(z)$ in Eq. (2.1b)], given, in terms of the notations

$$h_1 = \left[k_1^2 + \left(\frac{\partial}{\partial\rho}\right)^2\right]^{1/2}, \quad h_2 = \left[k_2^2 + \left(\frac{\partial}{\partial\rho}\right)^2\right]^{1/2}, \qquad (2.16)$$

by the components $\psi_1^{(1)}(z)$, $z \geqslant 0$, and $\psi_2^{(1)}(z)$, $z \leqslant -d$, of the form

$$\psi_1^{(1)}(z) = [\exp(ih_1 z) - \exp(-ih_1 z)](2ih_1)^{-1}\eta_1 + \exp(-ih_1 z)g_{11}, \quad z \geqslant 0, \qquad (2.17a)$$

$$\psi_2^{(1)}(z) = \exp[ih_2(z+d)]g_{21}, \quad z \leqslant -d. \qquad (2.17b)$$

Here, g_{11} and g_{21} are ρ operators independent of z, and give the boundary values of $\psi^{(1)}(z)$ at $z = 0$ and $-d$, respectively; the operator $\psi^{(1)}(z)$ also can be regarded as a ρ matrix, whose matrix elements, say $\psi^{(1)}(\rho,z|\rho')$, are defined according to rule (2.11a). Now, substitution of $\psi^{(1)}(z)$ into Eq. (2.10) yields

$$(ih_1\eta_1^{-1} - B_{11})g_{11} - B_{12}g_{21} = 1, \qquad (2.18a)$$

$$-B_{21}g_{11} + (ih_2\eta_2^{-1} - B_{22})g_{21} = 0, \qquad (2.18b)$$

which provide g_{11} and g_{21} in terms of B_{ij}.

In the same way, by introducing another operator solution $\psi^{(2)}(z)$ of the form

$$\psi^{(2)}(z) = \exp(-ih_1 z)g_{12}, \quad z \geqslant 0, \qquad (2.19a)$$

$$= \{\exp[-ih_2(z+d)] - \exp[ih_2(z+d)]\} \times (2ih_2)^{-1}\eta_2 + \exp[ih_2(z+d)]g_{22}, \quad z \leqslant -d \qquad (2.19b)$$

(which represents waves incident from the downward direction), we find equations of g_{12} and g_{22} similar to Eq. (2.18).

The four equations thus obtained can be written by one 2×2 matrix equation, as

$$\begin{pmatrix} ih_1\eta_1^{-1} - B_{11} & -B_{12} \\ -B_{21} & ih_2\eta_2^{-1} - B_{22} \end{pmatrix} \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad (2.20)$$

hence,

$$(i\mathbf{h}\boldsymbol{\eta}^{-1} - \mathbf{B})\mathbf{g} = 1. \qquad (2.21a)$$

Here, the matrices $\mathbf{g}$, $\mathbf{h}$, and 1 are defined by the matrix elements g_{ij}, $h_i\delta_{ij}$, and δ_{ij}, respectively, and the solution of Eq. (2.21a) is given by

$$\mathbf{g} = (i\mathbf{h}\boldsymbol{\eta}^{-1} - \mathbf{B})^{-1}, \qquad (2.21b)$$

where $\mathbf{h}\boldsymbol{\eta}^{-1} = \boldsymbol{\eta}^{-1}\mathbf{h}$. Note that $\mathbf{g}$ is a 2×2 matrix ρ operator entirely independent of z; and it will hereinafter be referred to as the surface Green's function, in view of the form similar to the ordinary Green's function in a medium $\mathbf{B}$.

A reflection coefficient matrix operator $\mathbf{R}$ will be defined, according to

$$\mathbf{g} = (1 + \mathbf{R})(2i\mathbf{h})^{-1}\boldsymbol{\eta} = \begin{pmatrix} (1+R_{11})(2ih_1)^{-1}\eta_1 & R_{12}(2ih_2)^{-1}\eta_2 \\ R_{21}(2ih_1)^{-1}\eta_1 & (1+R_{22})(2ih_2)^{-1}\eta_2 \end{pmatrix}. \qquad (2.22)$$

Hence, $\psi^{(1)}(z)$ in Eq. (2.17) can be written as

$$\psi_1^{(1)}(z) = [\exp(ih_1 z) + \exp(-ih_1 z)R_{11}](2ih_1)^{-1}\eta_1, \quad z \geqslant 0,$$

$$\psi_2^{(1)}(z) = \exp[ih_2(z+d)]R_{21}(2ih_1)^{-1}\eta_1, \quad z \leqslant -d, \qquad (2.23)$$

and similar expressions are also written for $\psi^{(2)}(z)$. Here, from Eqs. (2.22) and (2.21b), $\mathbf{R}$ is given in terms of $\mathbf{B}$ by

$$\mathbf{R} = \mathbf{g}2i\mathbf{h}\boldsymbol{\eta}^{-1} - 1 = (i\mathbf{h}\boldsymbol{\eta}^{-1} - \mathbf{B})^{-1}(i\mathbf{h}\boldsymbol{\eta}^{-1} + \mathbf{B}), \qquad (2.24)$$

which has the same form as when the boundary is smooth. However, the order of the factors is important since they are generally not commutable. Conversely,

$$\mathbf{B} = i\mathbf{h}\boldsymbol{\eta}^{-1} - \mathbf{g}^{-1} = \mathbf{i}\mathbf{h}\boldsymbol{\eta}^{-1}(\mathbf{R} - 1)(\mathbf{R} + 1)^{-1}. \qquad (2.25)$$

The matrix elements of $\mathbf{g}$ are not quite independent of one another, subjected to a few relations. Here, we first apply the Green's theorem to the boundary space enclosed by $S_1 + S_2$ (where $j = 0$) to find that, for arbitrary solutions ψ' and ψ'' of wave equation (2.1) subjected to Eq. (2.8),

$$\int_{S_1+S_2} d\rho\, \eta^{-1}\psi'(\overleftarrow{\partial}_n - \vec{\partial}_n)\psi'' = 0, \qquad (2.26)$$

and then utilize boundary equation (2.12). Hence,

$$\sum_{i,j}\int d\rho\, d\rho'\, \psi_i'(\rho)[-B_{ji}(\rho'|\rho)+B_{ij}(\rho|\rho')]\psi_j''(\rho')=0, \tag{2.27}$$

which shows that

$$B_{ji}(\rho'|\rho)=B_{ij}(\rho|\rho'), \quad \text{or } \mathbf{B}^T=\mathbf{B}=\mathbf{B}^*, \tag{2.28}$$

where the superscript T denotes the transposed matrix with respect to both the coordinates and the indices, and the last equality holds only in the case of a nondissipative medium.

Hence, since $\mathbf{h}^T=\mathbf{h}$, it follows from Eq. (2.21) that

$$\mathbf{g}^T=\mathbf{g}, \tag{2.29}$$

and therefore, from the first equality of Eq. (2.24), also

$$\mathbf{h}\boldsymbol{\eta}^{-1}\mathbf{R}=(\mathbf{h}\boldsymbol{\eta}^{-1}\mathbf{R})^T=\mathbf{R}^T\mathbf{h}\boldsymbol{\eta}^{-1}, \tag{2.30}$$

showing that $\tilde{h}_1(\lambda)\eta_1^{-1}\tilde{R}_{12}(\lambda|\lambda')=\tilde{R}_{21}(\lambda'|\lambda)\tilde{h}_2(\lambda')\eta_2^{-1}$, where $\tilde{\mathbf{R}}$ is the Fourier transform of $\mathbf{R}$ to be defined by Eq. (2.35), while, from the last equality of Eq. (2.24),

$$\begin{aligned}\mathbf{R}^{-1}&=(i\mathbf{h}\boldsymbol{\eta}^{-1}+\mathbf{B})^{-1}(i\mathbf{h}\boldsymbol{\eta}^{-1}-\mathbf{B}),\\ \mathbf{R}^*&=(i\mathbf{h}^*\boldsymbol{\eta}^{-1}+\mathbf{B})^{-1}(i\mathbf{h}^*\boldsymbol{\eta}^{-1}-\mathbf{B}),\end{aligned} \tag{2.31}$$

where $\mathbf{B}^*=\mathbf{B}$ by Eq. (2.28), showing that $\mathbf{R}^{-1}=\mathbf{R}^*$ if all the processes are taking place in the optical range where $\mathbf{h}^*=\mathbf{h}$.

The Hermitian condition of $\mathbf{B}$ imposes another constraint on $\mathbf{g}$, which eventually leads to a relation ensuring the power conservation in the scattering, i.e., the optical relation. This can be simply derived from the first equality of Eq. (2.25), which should be invariant against the Hermitian conjugation, hence,

$$i\mathbf{h}\boldsymbol{\eta}^{-1}-\mathbf{g}^{-1}=-i\mathbf{h}^\dagger\boldsymbol{\eta}^{-1}-\mathbf{g}^{\dagger-1}, \tag{2.32}$$

which becomes, upon multiplication of the both sides to the right with $(2i)^{-1}\mathbf{g}$ and to the left with $\mathbf{g}^\dagger$, as

$$\mathbf{W}\equiv 2^{-1}\mathbf{g}^\dagger(\mathbf{h}^\dagger+\mathbf{h})\boldsymbol{\eta}^{-1}\mathbf{g}+(2i)^{-1}(\mathbf{g}-\mathbf{g}^\dagger)=0. \tag{2.33}$$

Here, the diagonal elements, say W_1 and W_2, are exactly the integrated powers over S_1+S_2 of $\psi^{(1)}(z)$ and $\psi^{(2)}(z)$, respectively, as may be directly shown by Eq. (2.14a) with Eqs. (2.17) and (2.19).

In the same way, from the last equality of (2.25), the corresponding constraint for $\mathbf{R}$ is found to be

$$\begin{aligned}&2^{-1}\mathbf{R}^\dagger(\mathbf{h}^\dagger+\mathbf{h})\boldsymbol{\eta}^{-1}\mathbf{R}-2^{-1}(\mathbf{h}^\dagger+\mathbf{h})\boldsymbol{\eta}^{-1}\\ &\quad+2^{-1}[\mathbf{R}^\dagger(\mathbf{h}^\dagger-\mathbf{h})\boldsymbol{\eta}^{-1}-(\mathbf{h}^\dagger-\mathbf{h})\boldsymbol{\eta}^{-1}\mathbf{R}]=0.\end{aligned} \tag{2.34}$$

Here, it may be remarked that matrix elements of the last term become zero in the optical range where $\mathbf{h}^\dagger=\mathbf{h}$, as is explicitly shown by the Fourier transform given, in terms of the notation

$$\tilde{\mathbf{R}}(\lambda|\lambda')=\int d\rho\, d\rho'\, \exp[i(\lambda\cdot\rho-\lambda'\cdot\rho')]\mathbf{R}(\rho|\rho'), \tag{2.35}$$

by

$$\begin{aligned}&2^{-1}[\tilde{\mathbf{R}}^\dagger(\lambda|\lambda')[\tilde{\mathbf{h}}^*(\lambda')-\tilde{\mathbf{h}}(\lambda')]\boldsymbol{\eta}^{-1}\\ &\quad-[\tilde{\mathbf{h}}^*(\lambda)-\tilde{\mathbf{h}}(\lambda)]\boldsymbol{\eta}^{-1}\tilde{\mathbf{R}}(\lambda|\lambda')].\end{aligned} \tag{2.36}$$

III. EVALUATION OF THE SURFACE GREEN'S FUNCTION

A. First Method

The surface impedance and the Green's function derived therefrom have been exactly obtained for given $\zeta(\rho)$ in Ref. 14 for a one-side boundary having a constant surface impedance, and, to the perturbative approximation of ζ, it was further generalized to surfaces of given refractive index to apply to electromagnetic wave scattering. Also, for the present case of a two-side boundary, the method essentially remains unchanged although the actual procedure is not quite the same, including an alternative method to be newly added based on the Green's theorem.

The unit vector normal to S is given by $\hat{n}^{(S)}=(\mathbf{n}^{(S)},n_z^{(S)})$, $n_z^{(S)}>0$, with

$$n_z^{(S)}=\left[1+\left(\frac{\partial\zeta}{\partial\rho}\right)^2\right]^{-1/2}, \quad \mathbf{n}^{(S)}=n_z^{(S)}\frac{\partial\zeta}{\partial\rho}. \tag{3.1}$$

We first observe that $\psi_j^{(1)}(z)$ in Eq. (2.23) is a solution of the homogeneous wave equation of (2.1) ($j=0$) over the entire space of k_j, and that, on S, the boundary value is $\psi_j^{(1)}(-\zeta)$ when ζ is a constant independent of ρ; even when $\zeta=\zeta(\rho)$, the boundary value can be given, on attaching an ordering symbol $\mathscr{N}$ to be defined by Eq. (3.4), in the same form, as

$$\psi^{(1)}|_S=\psi_1^{(1)}(-\zeta)=\psi_2^{(1)}(-\zeta). \tag{3.2}$$

Here, from Eq. (2.23),

$$\psi_1^{(1)}(-\zeta)=[\exp^{(N)}(-ih_1\zeta)+\exp^{(N)}(ih_1\zeta)R_{11}](2ih_1)^{-1}\eta_1, \tag{3.3a}$$

$$\psi_2^{(1)}(-\zeta)=\exp^{(N)}[ih_2\bar{\zeta}]R_{21}(2ih_1)^{-1}\eta_1, \tag{3.3b}$$

with $\bar{\zeta}=d-\zeta$, and

$$\begin{aligned}\exp^{(N)}(ih\zeta)&\equiv\mathscr{N}[\exp(ih\zeta)]\\ &=\sum_{n=0}^{\infty}(n!)^{-1}\zeta^n(ih)^n,\end{aligned} \tag{3.4}$$

which is ordered in such a way that the coefficients $[\zeta(\rho)]^n$ are always placed to the left of the operators $[ih]^n$. That is, the symbol $\mathscr{N}$ means that all the ζ variables (including its derivatives) involved in the referenced function are understood to be placed to the left of all the h and/or $\partial/\partial\rho$ if any, as may be well defined by the power series expansion with respect to h and/or $\partial/\partial\rho$ (Ref. 21). For example,

$$\mathscr{N}\left[h^l\zeta^m\left(\frac{\partial}{\partial\rho}\right)^n\right]=\mathscr{N}\left[\left(\frac{\partial}{\partial\rho}\right)^n h^l\zeta^m\right]=\zeta^m h^l\left(\frac{\partial}{\partial\rho}\right)^n.$$

Hence, for any functions f and g,

$$\begin{aligned}&\mathscr{N}[f]\equiv f^{(N)}, \quad \mathscr{N}[f+g]=f^{(N)}+g^{(N)},\\ &\mathscr{N}[fg]=\mathscr{N}[gf]=\mathscr{N}[g^{(N)}f^{(N)}]\neq g^{(N)}f^{(N)}.\end{aligned} \tag{3.5}$$

In Eq. (3.12), the symbol $\mathscr{N}$ will be redefined with a minor additional condition.

Thus, from Eqs. (3.2) and (3.3), we find a simple relation between R_{21} and R_{11}, as

$$\begin{aligned}R_{21}=\exp^{(N)}(ih_2\bar{\zeta})^{-1}[&\exp^{(N)}(-ih_1\zeta)\\ &+\exp^{(N)}(ih_1\zeta)R_{11}],\end{aligned} \tag{3.6}$$

and, by Eq. (2.22), it can be rewritten in terms of g_{21} and g_{11}, by

$$g_{21} = \exp^{(N)}(ih_2\bar{\zeta})^{-1}[-\sin^{(N)}(h_1\zeta)h_1^{-1}\eta_1 + \exp^{(N)}(ih_1\zeta)g_{11}]. \quad (3.7)$$

The situation is the same also of the second boundary condition, i.e., the continuity of

$$\eta^{-1}\hat{n}^{(S)}\cdot\frac{\partial}{\partial\hat{x}}\psi^{(1)}\Big|_S = \eta^{-1}\left[\mathbf{n}^{(S)}\cdot\frac{\partial}{\partial\boldsymbol{\rho}} + n_z^{(S)}\frac{\partial}{\partial z}\right]\psi^{(1)}(z)\Big|_S, \quad (3.8)$$

which, for $\psi_2^{(1)}(z)$ in Eq. (2.23), for example, becomes

$$\eta_2^{-1}\hat{n}^{(S)}\cdot\frac{\partial}{\partial\hat{x}}\psi_2^{(1)}\Big|_S = \eta_2^{-1}\mathscr{N}\left[\left(\mathbf{n}^{(S)}\cdot\frac{\partial}{\partial\boldsymbol{\rho}} + n_z^{(S)}ih_2\right)\times\exp^{(N)}(ih_2\bar{\zeta})\right]R_{21}(2ih_1)^{-1}\eta_1. \quad (3.9)$$

Here, the following equations can be made simple by the introduction of a surface impedance $B_S^{(1)}$ on the real boundary S, defined by

$$-\eta^{-1}\hat{n}^{(S)}\cdot\frac{\partial}{\partial\hat{x}}\psi^{(1)}\Big|_S = B_S^{(1)}\psi^{(1)}|_S. \quad (3.10)$$

Here, $B_S^{(1)}$ should be the same for both solutions $\psi_1^{(1)}(z)$ and $\psi_2^{(1)}(z)$, in consequence of the boundary condition of Eq. (2.8), and, for $\psi_2^{(1)}(z)$, substitution of Eqs. (3.3b) and (3.9) into Eq. (3.10) yields

$$-\eta_2^{-1}\mathscr{N}\left[\left(\mathbf{n}^{(S)}\cdot\frac{\partial}{\partial\boldsymbol{\rho}} + n_z^{(S)}ih_2\right)\exp^{(N)}(-ih_2\zeta)\right] = B_S^{(1)}\exp^{(N)}(-ih_2\zeta), \quad (3.11)$$

where a common factor has been deleted from both sides. In the same way, also for $\psi_1^{(1)}(z)$, we obtain a similar equation, and the result can be rewritten, as

$$\mathscr{N}\left[\left(\mathbf{n}^{(S)}\cdot\frac{\partial}{\partial\boldsymbol{\rho}} - n_z^{(S)}ih_1 + \eta_1 B_S^{(1)}\right)\exp^{(N)}(ih_1\zeta)\right]R_{11} = -\mathscr{N}\left[\left(\mathbf{n}^{(S)}\cdot\frac{\partial}{\partial\boldsymbol{\rho}} + n_z^{(S)}ih_1 + \eta_1 B_S^{(1)}\right)\times\exp^{(N)}(-ih_1\zeta)\right]. \quad (3.12)$$

Here, the operator $B_S^{(1)}$ should be to the left of all the factors, and, hereinafter, the symbol $\mathscr{N}$ will be redefined with this additional condition.

Hence, with the $B_S^{(1)}$ determined by Eq. (3.11), Eq. (3.12) provides the operator R_{11} for given $\zeta(\boldsymbol{\rho}) > 0$ explicitly; alternatively, the $\boldsymbol{\rho}$ matrix elements of R_{11}, as defined by Eq. (2.11a), can be obtained as a solution of a set of ordinary integral equations free from any operator (except $B_S^{(1)}$) and also from the symbol $\mathscr{N}$, by multiplication of both sides of Eq. (3.12) to the right with $\exp(-i\boldsymbol{\lambda}\cdot\boldsymbol{\rho})$ and subsequent use of the integral representation

$$R_{11}\exp(-i\boldsymbol{\lambda}\cdot\boldsymbol{\rho}) = \int d\boldsymbol{\rho}'\, R_{11}(\boldsymbol{\rho}|\boldsymbol{\rho}')\exp(-i\boldsymbol{\lambda}\cdot\boldsymbol{\rho}') = (2\pi)^{-2}\int d\boldsymbol{\lambda}'\exp(-i\boldsymbol{\lambda}'\cdot\boldsymbol{\rho})\widetilde{R}_{11}(\boldsymbol{\lambda}'|\boldsymbol{\lambda}), \quad (3.13)$$

where the Fourier transform $\widetilde{R}_{11}(\boldsymbol{\lambda}'|\boldsymbol{\lambda})$ is defined by Eq. (2.35). Hence, according to the definition of the symbol $\mathscr{N}$, Eq. (3.12) becomes written as

$$(2\pi)^{-2}\int_{-\infty}^{\infty} d\boldsymbol{\lambda}'[-i\mathbf{n}^{(S)}(\boldsymbol{\rho})\cdot\boldsymbol{\lambda}' - in_z^{(S)}(\boldsymbol{\rho})\tilde{h}_1(\boldsymbol{\lambda}') + \eta_1 B_S^{(1)}]\times\exp[i\zeta(\boldsymbol{\rho})\tilde{h}_1(\boldsymbol{\lambda}') - i\boldsymbol{\rho}\cdot\boldsymbol{\lambda}']\widetilde{R}_{11}(\boldsymbol{\lambda}'|\boldsymbol{\lambda}) = [i\mathbf{n}^{(S)}(\boldsymbol{\rho})\cdot\boldsymbol{\lambda} - i\eta_z^{(S)}(\boldsymbol{\rho})\tilde{h}_1(\boldsymbol{\lambda}) - \eta_1 B_S^{(1)}]\times\exp[-i\zeta(\boldsymbol{\rho})\tilde{h}_1(\boldsymbol{\lambda}) - i\boldsymbol{\rho}\cdot\boldsymbol{\lambda}], \quad (3.14a)$$

for given $\boldsymbol{\rho}$ operator $B_S^{(1)}$, where the latter is obtained, in terms of the matrix elements $B_S^{(1)}(\boldsymbol{\rho}|\boldsymbol{\rho}')$, as a solution of an integral equation, given by multiplication of Eq. (3.11) to the right with $\exp(-i\boldsymbol{\lambda}\cdot\boldsymbol{\rho})$ and subsequent manipulation similar to Eq. (3.13), as

$$\int_{-\infty}^{\infty} d\boldsymbol{\rho}'\, B_S^{(1)}(\boldsymbol{\rho}|\boldsymbol{\rho}')\exp[-i\zeta(\boldsymbol{\rho}')\tilde{h}_2(\boldsymbol{\lambda}) - i\boldsymbol{\rho}'\cdot\boldsymbol{\lambda}] = i\eta_2^{-1}[\mathbf{n}^{(S)}(\boldsymbol{\rho})\cdot\boldsymbol{\lambda} - n_z^{(S)}(\boldsymbol{\rho})\tilde{h}_2(\boldsymbol{\lambda})]\times\exp[-i\zeta(\boldsymbol{\rho})\tilde{h}_2(\boldsymbol{\lambda}) - i\boldsymbol{\rho}\cdot\boldsymbol{\lambda}]. \quad (3.14b)$$

Here, when ζ is a constant, for example, the substitution of $B_S^{(1)}(\boldsymbol{\rho}|\boldsymbol{\rho}') = B_S^{(1)}(\boldsymbol{\rho}-\boldsymbol{\rho}')$ yields $\widetilde{B}_S^{(1)}(\boldsymbol{\lambda}) = -i\eta_2^{-1}\tilde{h}_2(\boldsymbol{\lambda})$, as it should be. Also, it may be remarked that, when $B_S^{(1)}$ is a given constant, the integral equation (3.14a) for $\widetilde{R}_{11}$ agrees exactly with the corresponding equation (80) in Ref. 14, derived based on a different method. On the other hand, an integral equation of $\widetilde{R}_{21}$ is obtained from Eq. (3.6) with known solution $\widetilde{R}_{11}$ of Eq. (3.14), as

$$(2\pi)^{-2}\int d\boldsymbol{\lambda}'\exp[i\bar{\zeta}(\boldsymbol{\rho})\tilde{h}_2(\boldsymbol{\lambda}') - i\boldsymbol{\rho}\cdot\boldsymbol{\lambda}']\widetilde{R}_{21}(\boldsymbol{\lambda}'|\boldsymbol{\lambda}) = (2\pi)^{-2}\int d\boldsymbol{\lambda}'\exp[i\zeta(\boldsymbol{\rho})\tilde{h}_1(\boldsymbol{\lambda}') - i\boldsymbol{\rho}\cdot\boldsymbol{\lambda}']\widetilde{R}_{11}(\boldsymbol{\lambda}'|\boldsymbol{\lambda}) + \exp[-i\zeta(\boldsymbol{\rho})\tilde{h}_1(\boldsymbol{\lambda}) - i\boldsymbol{\rho}\cdot\boldsymbol{\lambda}]. \quad (3.15)$$

By using Eq. (2.24), Eq. (3.12) of R_{11} can be rewritten as an equation of g_{11}, as

$$\mathscr{N}\left[\left(\eta_1^{-1}\mathbf{n}^{(S)}\cdot\frac{\partial}{\partial\boldsymbol{\rho}} - \eta_1^{-1}n_z^{(S)}ih_1 + B_S^{(1)}\right)\exp^{(N)}(ih_1\zeta)\right]g_{11} = \mathscr{N}\left[\left(\eta_1^{-1}\mathbf{n}^{(S)}\cdot\frac{\partial}{\partial\boldsymbol{\rho}} + B_S^{(1)}\right)\sin^{(N)}(h_1\zeta)h_1^{-1}\eta_1 - n_z^{(S)}\cos^{(N)}(h_1\zeta)\right], \quad (3.16)$$

and g_{21} is given in terms of g_{11} by Eq. (3.7). So far the particular solution $\psi^{(1)}(z)$ has been considered, which represents waves incident from the k_1 side of S; however, the same is also true of the solution $\psi^{(2)}(z)$ for waves incident from the opposite side.

B. Alternative method

In the previous method, the two wave functions $\psi_1^{(1)}(z)$ and $\psi_2^{(1)}(z)$ of Eq. (2.23) were continued directly on the real boundary, and thereby R_{11} and R_{21} were obtained through $B_S^{(1)}$ given by Eq. (3.11). There exists an alternative method which enables R_{21} to be directly obtained and R_{11} to be from known R_{21}; in this method, the continuation of the wave functions is performed rather indirectly, with the aid of the Green's theorem.

By applying the Green's theorem to the space enclosed by S_1 and S (Fig. 1), we find, for arbitrary solutions ψ' and ψ'' of wave equation (2.1), the relation

$$[\psi',\psi''] \equiv \int_{z=0} d\boldsymbol{\rho}\, \eta_1^{-1} \psi'(\hat{x}) \left(\frac{\overleftarrow{\partial}}{\partial z} - \frac{\overrightarrow{\partial}}{\partial z}\right) \psi''(\hat{x}) \tag{3.17a}$$

$$= \int_{z=-\zeta} ds\, \eta_1^{-1} \psi'(\hat{x}) \left[\frac{\overleftarrow{\partial}}{\partial \hat{x}} \cdot \hat{n}^{(S)} - \hat{n}^{(S)} \cdot \frac{\overrightarrow{\partial}}{\partial \hat{x}} \right] \psi''(\hat{x}), \tag{3.17b}$$

where $ds = d\boldsymbol{\rho}/n_z^{(S)}$ is the two-dimensional element of surface S. Here, the relation can be written in a compact form by regarding $\psi'(\hat{x})$ and $\psi''(\hat{x})$ as components of $\boldsymbol{\rho}$ vectors $\psi'(z)$ and $\psi''(z)$ (which, more generally, may be $\boldsymbol{\rho}$ matrices), respectively, as [cf. Eq. (3.9)]

$$[\psi',\psi''] = \eta_1^{-1} \psi'(z) \left(\frac{\overleftarrow{\partial}}{\partial z} - \frac{\overrightarrow{\partial}}{\partial z}\right) \psi''(z) \Big|_{z=0} \tag{3.18a}$$

$$= \bar{\mathscr{N}} \left[\psi'(-\zeta) \frac{\overleftarrow{\partial}}{\partial \hat{x}} \cdot \frac{\hat{n}^{(S)}}{n_z^{(S)} \eta} \right] \mathscr{N}[\psi''(-\zeta)] - \bar{\mathscr{N}}[\psi'(-\zeta)] \mathscr{N} \left[\frac{\hat{n}^{(S)}}{n^{(S)} \eta} \cdot \frac{\overrightarrow{\partial}}{\partial \hat{x}} \psi''(-\zeta) \right]. \tag{3.18b}$$

Here, the $\boldsymbol{\rho}$ integration is involved as a natural consequence of the inner product; that is, for any $\boldsymbol{\rho}$ matrices A and B,

$$\int d\boldsymbol{\rho}\, A(\boldsymbol{\rho}'|\boldsymbol{\rho}) B(\boldsymbol{\rho}|\boldsymbol{\rho}'') = AB(\boldsymbol{\rho}'|\boldsymbol{\rho}'');$$

and the symbol $\bar{\mathscr{N}}$ denotes an ordering operator similar to $\mathscr{N}$; however, the ordering is inverse such that, for example,

$$\bar{\mathscr{N}}[\exp(ih\zeta)] \equiv \exp^{(\bar{N})}(ih\zeta) = \sum_{n=0}^{\infty} (n!)^{-1} (ih)^n \zeta^n, \tag{3.19}$$

where all the ζ variables are understood to be placed to the right of all the h and/or $\partial/\partial\boldsymbol{\rho}$ if any.

Here, we set $\psi''(z) = \psi_1^{(1)}(z)$ of Eq. (2.23) and

$$\psi'(z) = \psi^{(+)}(z) \equiv \exp(-ih_1 z), \tag{3.20}$$

hence, the substitution into Eq. (3.18a) simply yields

$$[\psi^{(+)},\psi^{(1)}] = -1. \tag{3.21}$$

For Eq. (3.18b), on the other hand, we observe that, by virtue of boundary condition (2.8), all the factors associated with $\psi_1^{(1)}(z)$ can be replaced with those of $\psi_2^{(1)}(z)$ given by Eq. (2.17b), and, since, by Eq. (3.1),

$$\frac{\hat{n}^{(S)}}{n_z^{(S)}} = (\boldsymbol{\zeta}',1), \quad \boldsymbol{\zeta}' = \frac{\partial \zeta}{\partial \boldsymbol{\rho}}, \tag{3.22}$$

also that

$$\mathscr{N} \left[\frac{\hat{n}^{(S)}}{n_z^{(S)} \eta} \cdot \frac{\overrightarrow{\partial}}{\partial \hat{x}} \psi_2^{(1)}(-\zeta) \right] = \mathscr{N} \left[\eta_2^{-1} \left(\boldsymbol{\zeta}' \cdot \frac{\partial}{\partial \boldsymbol{\rho}} + ih_2 \right) \exp^{(N)}(ih_2\bar{\zeta}) \right] g_{21}, \tag{3.23a}$$

$$\bar{\mathscr{N}} \left[\psi^{(+)}(-\zeta) \frac{\overleftarrow{\partial}}{\partial \hat{x}} \cdot \frac{\hat{n}^{(S)}}{n_z^{(S)} \eta} \right] = \bar{\mathscr{N}} \left[\eta_1^{-1} \exp^{(\bar{N})}(ih_1\zeta) \left(\frac{\overleftarrow{\partial}}{\partial \boldsymbol{\rho}} \cdot \boldsymbol{\zeta}' - ih_1 \right) \right], \tag{3.23b}$$

where we can set $\overleftarrow{\partial/\partial\boldsymbol{\rho}} = -\overrightarrow{\partial/\partial\boldsymbol{\rho}}$ by partial integration. Hence, with Eq. (3.21), the result can be written as

$$T^{(+)} g_{21} = 1, \tag{3.24a}$$

where

$$T^{(+)} = \eta_1^{-1} \bar{\mathscr{N}} \left[\exp^{(\bar{N})}(ih_1\zeta) \left(\frac{\partial}{\partial \boldsymbol{\rho}} \cdot \boldsymbol{\zeta}' + ih_1 \right) \right] \times \exp^{(N)}(ih_2\bar{\zeta}) + \eta_2^{-1} \exp^{(\bar{N})}(ih_1\zeta) \mathscr{N} \left[\left(\boldsymbol{\zeta}' \cdot \frac{\partial}{\partial \boldsymbol{\rho}} + ih_2 \right) \times \exp^{(N)}(ih_2\bar{\zeta}) \right]. \tag{3.24b}$$

In the same way, replacing $\psi^{(+)}(z)$ with

$$\psi^{(-)}(z) \equiv \exp(ih_1 z), \tag{3.25}$$

Eq. (3.18a) becomes, on using Eq. (2.23),

$$[\psi^{(-)},\psi^{(1)}] = (2ih_1)\eta_1^{-1} R_{11} (2ih_1)^{-1} \eta_1, \tag{3.26}$$

whereas Eq. (3.18b) becomes what would be obtained by replacing $h_1 \to -h_1$ in Eqs. (3.23a) and (3.23b). Thus, we obtain another equation

$$g_1^{(0)-1} R_{11} g_1^{(0)} = -T^{(-)} g_{21}, \tag{3.27a}$$

where

$$g_1^{(0)} = (2ih_1)^{-1} \eta_1, \tag{3.27b}$$

and $T^{(-)}$ is obtained from $T^{(+)}$ with $h_1 \to -h_1$. Here, with Eq. (3.24a),

$$g_{11} = (1 + R_{11}) g_1^{(0)} = g_1^{(0)} (T^{(+)} - T^{(-)}) g_{21}. \tag{3.28}$$

For a specific example of a slightly random boundary to be treated in the next section, it is straightforward to directly confirm the equivalence of the two methods. The second method is particularly convenient for electromagnetic waves and other waves of a multicomponent, in view of several conditions to be fulfilled on the real boundary for the first method.

C. Case of slightly random boundary

When the boundary changes in a sufficiently large scale compared with the wavelength, a new tangent plane method[14] can be utilized to solve Eqs. (3.14) and (3.15) under essentially the same condition of applicability as for the conventional,[2,6] and, in a separate paper, the solutions will be utilized as basic quantities to obtain the scattering cross sections consistent with both power conservation and multiple scattering, including shadow effect, as in the case of one-side boundary. Also, when the boundary is slightly random such that $|\partial\zeta/\partial\boldsymbol{\rho}| \ll 1$ and $|h_j\, \zeta| \ll 1, j = 1,2$ the previous equations become considerably simple, as follows.

To the first order of $\zeta(\boldsymbol{\rho})$, $\mathbf{n}^{(S)} = \boldsymbol{\zeta}' (= \partial\zeta/\partial\boldsymbol{\rho})$ and $n_z^{(S)} = 1$, and Eq. (3.11) is reduced to

$$-\eta_2^{-1}\left(ih_2+\zeta h_2^2+\zeta'\cdot\frac{\partial}{\partial\boldsymbol{\rho}}\right)=B_S^{(1)}(1-i\zeta h_2),$$

hence,

$$B_S^{(1)}=-\eta_2^{-1}\left[ih_2+(\zeta h_2-h_2\zeta)h_2+\zeta'\cdot\frac{\partial}{\partial\boldsymbol{\rho}}\right]. \tag{3.29}$$

In the same way, from Eq. (3.16), we obtain

$$\left[ih_1\eta_1^{-1}-B_S^{(1)}-\eta_1 B_S^{(1)}\zeta B_S^{(1)}\right.$$
$$\left.-\eta_1^{-1}\left(\zeta h_1^2+\zeta'\cdot\frac{\partial}{\partial\boldsymbol{\rho}}\right)\right]g_{11}=1, \tag{3.30}$$

and, upon the substitution of Eq. (3.29), can be written in the form

$$[i(h_1\eta_1^{-1}+h_2\eta_2^{-1})-b^{(11)}]\,g_{11}=1. \tag{3.31}$$

Here,

$$b^{(11)}=(\eta_1^{-1}-\eta_2^{-1})\zeta'\cdot\frac{\partial}{\partial\boldsymbol{\rho}}+\zeta(h_1^2\eta_1^{-1}-h_2^2\eta_2^{-1})$$
$$+(\eta_2-\eta_1)\eta_2^{-2}h_2\zeta h_2, \tag{3.32a}$$

where

$$h_1^2\eta_1^{-1}-h_2^2\eta_2^{-1}$$
$$=k_1^2\eta_1^{-1}-k_2^2\eta_2^{-1}+(\eta_1^{-1}-\eta_2^{-1})\left(\frac{\partial}{\partial\boldsymbol{\rho}}\right)^2, \tag{3.32b}$$

hence, Eq. (3.32a) can be rewritten also as

$$b^{(11)}=(\eta_2-\eta_1)\eta_2^{-2}h_2\zeta h_2$$
$$+(\eta_1^{-1}-\eta_2^{-1})\frac{\partial}{\partial\boldsymbol{\rho}}\cdot\zeta\frac{\partial}{\partial\boldsymbol{\rho}}$$
$$+(k_1^2\eta_1^{-1}-k_2^2\eta_2^{-1})\zeta, \tag{3.33}$$

each term of which is a Hermitian operator, except for the first term, which is not exactly Hermitian since $h^\dagger\neq h$ in the nonoptical range. Thus, from Eq. (3.31), we obtain

$$g_{11}=g_0+g_0b^{(11)}g_0, \tag{3.34}$$

where

$$g_0=(ih_1\eta_1^{-1}+ih_2\eta_2^{-1})^{-1}. \tag{3.35}$$

On the other hand, Eq. (3.7) for g_{21} becomes, on putting $d=0$ and therefore $\bar\zeta=-\zeta$,

$$(1-i\zeta h_2)g_{21}=-\eta_1\zeta+(1+i\zeta h_1)g_{11}, \tag{3.36}$$

which shows that, when $\zeta=0$,

$$g_{21}=g_{11}=g_{22}=g_{12}=g_0, \tag{3.37}$$

and therefore that the term of $-\eta_1\zeta$ can be replaced with $-\eta_1\zeta(ih_1\eta_1^{-1}+ih_2\eta_2^{-1})g_{11}$, in view of Eq. (3.35). Hence, on multiplying to the left with $ih_2^\dagger\eta_2^{-1}$, Eq. (3.36) can be rewritten, in terms of b_2, defined by

$$b_2=b_2^\dagger=(\eta_1-\eta_2)\eta_2^{-2}h_2^\dagger\zeta h_2, \tag{3.38}$$

also as

$$ih_2^\dagger\eta_2^{-1}(-g_{11}+g_{21})=b_2\,g_{21}, \tag{3.39}$$

which provides a basic equation to be used later. Alternatively, if h_2 were used instead of $h_2^\dagger$, we would obtain

$$ih_2\eta_2^{-1}(-g_{11}+g_{21})=b_2''\,g_{21}, \tag{3.40a}$$

where

$$b_2''=(\eta_1-\eta_2)\eta_2^{-2}h_2\zeta h_2=(h_2/h_2^\dagger)b_2 \tag{3.40b}$$

and is not Hermitian.

Thus, to the first order of ζ, use of Eq. (3.40) leads to

$$g_{21}=g_0+g_0b^{(21)}g_0, \tag{3.41a}$$

where

$$b^{(21)}=(\eta_1^{-1}-\eta_2^{-1})\left[-h_1\zeta h_2+\frac{\partial}{\partial\boldsymbol{\rho}}\cdot\zeta\frac{\partial}{\partial\boldsymbol{\rho}}\right]$$
$$+(k_1^2\eta_1^{-1}-k_2^2\eta_2^{-1})\zeta. \tag{3.41b}$$

Both Eqs. (3.34) and (3.41) are correct to the first order of ζ, but are not consistent with the power conservation; it is remarked, however, that the last aspect can be perfectly improved by constructing a set of equations for g_{11} and g_{21} consistent with the original equations.

Equation (3.31) can be rewritten, on using Eqs. (3.40), also as

$$ih_1\eta_1^{-1}g_{11}+ih_2\eta_2^{-1}g_{21}=1+b_1\,g_{11}, \tag{3.42}$$

in terms of the Hermitian operator b_1 defined by

$$b_1=(\eta_1^{-1}-\eta_2^{-1})\frac{\partial}{\partial\boldsymbol{\rho}}\cdot\zeta\frac{\partial}{\partial\boldsymbol{\rho}}+(k_1^2\eta_1^{-1}-k_2^2\eta_2^{-1})\zeta, \tag{3.43}$$

with the relation

$$b^{(11)}=b_1-b_2''. \tag{3.44}$$

Thus, Eqs. (3.39) and (3.42) provide a set of equations for the unknown g_{11} and g_{21}, and are consistent with the power conservation, as will be shown soon; there seems to be no other choice in this sense. Here, the following equation formulation becomes definitely simple by representing the two equations by one 2×2 matrix equation; that is, introducing two 2×2 matrix operators $\boldsymbol{\pi}$ and $\mathbf{b}$, defined by

$$\boldsymbol{\pi}=\begin{pmatrix} ih_1\eta_1^{-1} & ih_2\eta_2^{-1}\\ -ih_2^\dagger\eta_2^{-1} & ih_2^\dagger\eta_2^{-1}\end{pmatrix},\quad \mathbf{b}=\begin{pmatrix} b_1 & 0\\ 0 & b_2\end{pmatrix}, \tag{3.45}$$

and also two vectors $\vec{g}$ and $\vec{1}$, defined by

$$\vec{g}=\begin{pmatrix} g_{11}\\ g_{21}\end{pmatrix},\quad \vec{1}=\begin{pmatrix}1\\0\end{pmatrix}, \tag{3.46}$$

we obtain the basic equation

$$\boldsymbol{\pi}\vec{g}=\vec{1}+\mathbf{b}\vec{g}. \tag{3.47a}$$

Here, $\mathbf{b}=\mathbf{b}^\dagger$ is Hermitian, hence,

$$\vec{g}^\dagger\boldsymbol{\pi}^\dagger=\vec{1}+\vec{g}^\dagger\mathbf{b}, \tag{3.47b}$$

whereas $\boldsymbol{\pi}$ is not; nevertheless, it has the remarkable properties

$$\boldsymbol{\pi}-\boldsymbol{\pi}^\dagger=i(\mathbf{h}^\dagger+\mathbf{h})\boldsymbol{\eta}^{-1}, \tag{3.48a}$$

$$\boldsymbol{\pi}^{-1}=g_0\begin{pmatrix}1 & -h_2/h_2^\dagger\\ 1 & h_1\eta_2/h_2^\dagger\eta_1\end{pmatrix}. \tag{3.48b}$$

Note that, in virtue of Eq. (3.48a), the total power integrated over S_1+S_2 is given according to Eq. (2.33) by

$$W_1=(2i)^{-1}[\vec{g}^\dagger(\boldsymbol{\pi}-\boldsymbol{\pi}^\dagger)\vec{g}+\vec{1}\vec{g}-\vec{g}^\dagger\vec{1}],$$

which becomes exactly zero, in consequence of Eqs. (3.47a) and (3.47b). This convinces us that the solution of Eqs. (3.47) is significant over all the orders of ζ, in the sense of selected summation as when treating waves in a turbulent air, where the medium fluctuation is slight but still its accumulated effect can be quite large.[22–24] We shall return to the present case of a slightly random boundary at the end of Sec. VI (and Appendix C), where Eq. (3.47) will be utilized as the basic equation to obtain statistical equations of the same form as those in Sec. IV, and also resulting cross sections therefrom.

In the special case of $\eta_1 = \eta_2 = 1$, $b_2 = 0$ by Eq. (3.38), hence, it follows that $g_{21} = g_{11} = g_{22} = g_{12}$ and Eq. (3.42) becomes

$$[ih_1 + ih_2 - (k_1^2 - k_2^2)\zeta]\, g_{11} = 1, \tag{3.49}$$

and coincides with that by the original equation (3.31). Note that Eq. (3.49) remains unchanged against the exchange of k_1 and k_2 with the replacement of $\zeta \to -\zeta$, showing that $g_{22} = g_{11}$.

D. Scattering matrix of a small boss on homogeneous boundary

In Eq. (2.21a), let $\mathbf{B}$ be the surface impedance when a small boss is placed, say, on S_2 at $z = -d$. Then, if $\mathbf{B}_0$ is the impedance when the boundary is perfectly homogeneous, $\mathbf{b} = \mathbf{B} - \mathbf{B}_0$ differs from zero only over a small area of the boss. Therefore, in terms of the Green's function for the homogeneous boundary, say $\mathbf{g}_0$, given by

$$\mathbf{g}_0 = (i\mathbf{h}\boldsymbol{\eta}^{-1} - \mathbf{B}_0)^{-1}, \tag{3.50a}$$

Eq. (2.21a) becomes

$$\mathbf{g} = \mathbf{g}_0(1 + \mathbf{b}\mathbf{g}). \tag{3.50b}$$

Here, as $d \to 0$, all the elements of $\mathbf{g}_0$ tend to g_0 [Eq. (3.37)], implying that $\mathbf{g}_0^{-1}$ does not exist and therefore neither does $\mathbf{B}_0$, by Eq. (3.50a) (see Sec. III C), while g_0 has been investigated in detail in connection with the ground wave propagation over a flat earth.[14] Equation (3.50b) can be rewritten in the form

$$\mathbf{g} = \mathbf{g}_0 + \mathbf{g}_0\mathbf{T}_b\mathbf{g}_0, \tag{3.51}$$

in terms of a scattering matrix $\mathbf{T}_b$, defined by

$$\mathbf{T}_b\mathbf{g}_0 = \mathbf{b}\mathbf{g}, \tag{3.52}$$

which, by substitution of Eq. (3.51) into the right-hand side, gives a formal expression of $\mathbf{T}_b$ as

$$\mathbf{T}_b = (1 - \mathbf{b}\mathbf{g}_0)^{-1}\mathbf{b} \tag{3.53a}$$

$$= \mathbf{b} + \mathbf{b}\mathbf{g}_0\mathbf{b} + \mathbf{b}\mathbf{g}_0\mathbf{b}\mathbf{g}_0\mathbf{b} + \cdots. \tag{3.53b}$$

An explicit coordinate expression of Eq. (3.51) can be written as

$$\mathbf{g}(\boldsymbol{\rho}|\boldsymbol{\rho}') = \mathbf{g}_0(\boldsymbol{\rho} - \boldsymbol{\rho}') + \int d\boldsymbol{\rho}''d\boldsymbol{\rho}'''\mathbf{g}_0(\boldsymbol{\rho} - \boldsymbol{\rho}'') \times \mathbf{T}_b(\boldsymbol{\rho}'' - \boldsymbol{\rho}_b|\boldsymbol{\rho}''' - \boldsymbol{\rho}_b)\mathbf{g}_0(\boldsymbol{\rho}''' - \boldsymbol{\rho}'), \tag{3.54}$$

where $\boldsymbol{\rho}_b$ are the center coordinates of the boss. Here, as $|\boldsymbol{\rho}| \to \infty$, $\mathbf{g}_0(\boldsymbol{\rho})$ changes almost with the phase factor $\exp[-ik_j|\boldsymbol{\rho}|]$ of either $j = 1$ or 2; hence, at sufficiently large distances from both the points $\boldsymbol{\rho}$ and $\boldsymbol{\rho}'$, the factor $\mathbf{T}_b$ in Eq. (3.54) can be replaced by a scattering amplitude, given by the Fourier transform of $\mathbf{T}_b$ at the values of wave-number vectors of the incident and scattered waves, as in the case of a discrete scatterer in three-dimensional space. Here, evaluating the scattering amplitude is generally an involved task, particularly when it is a boss of complicated shape, but the amplitude may be a good experimental observable.

In the general case where many small bosses are randomly distributed over S_2, $\mathbf{b}$ is given by

$$\mathbf{b} = \sum_j \mathbf{b}_j, \tag{3.55}$$

where $\mathbf{b}_j$ denotes the contribution of the jth boss; the situation therefore becomes the same as when discrete scatterers are randomly distributed in space [see also Eqs. (4.13)–(4.15)].

IV. STATISTICAL SURFACE GREEN'S FUNCTIONS OF FIRST AND SECOND ORDERS

The procedure of deriving governing equations for the statistical surface Green's functions is basically the same as for deriving those for a random medium[14]; that is, on putting $\mathbf{B} = \mathbf{B}_0 + \mathbf{b}$ in Eq. (2.21a), we obtain

$$[i\mathbf{h}\boldsymbol{\eta}^{-1} - \mathbf{B}_0 - \mathbf{b}]\mathbf{g} = 1, \tag{4.1}$$

which has exactly the same form as the equation of the ordinary Green's function in a random medium with the random part $\mathbf{b}$.

Here, to find the first-order Green's function $\mathbf{G} = \langle \mathbf{g} \rangle$, we first introduce an effective 2×2 matrix operator $\mathbf{M}$ of $\mathbf{b}$ (mass operator), defined by [see Eqs. (4.12) and (4.13), for example]

$$\langle \mathbf{b}\mathbf{g} \rangle = \mathbf{M}\mathbf{G}, \tag{4.2}$$

or, explicitly, by

$$\sum_j \langle b_{ij}\, g_{jk} \rangle = \sum_j M_{ij}G_{jk}. \tag{4.3}$$

Hence, averaging Eq. (4.1) yields

$$[i\mathbf{h}\boldsymbol{\eta}^{-1} - \mathbf{B}_0 - \mathbf{M}]\mathbf{G} = 1, \tag{4.4a}$$

with the formal solution

$$\mathbf{G} = [i\mathbf{h}\boldsymbol{\eta}^{-1} - \mathbf{B}_0 - \mathbf{M}]^{-1}. \tag{4.4b}$$

In the same way, with $\mathbf{B}_0^* = \mathbf{B}_0$,

$$[-i\mathbf{h}^*\boldsymbol{\eta}^{-1} - \mathbf{B}_0 - \mathbf{M}^*]\mathbf{G}^* = 1. \tag{4.5}$$

The second-order Green's function is defined, in matrix form, by

$$\mathbf{G}(1;2) = \langle \mathbf{g}^*(1)\mathbf{g}(2) \rangle, \tag{4.6a}$$

or, writing explicitly, by

$$G_{ij;kl}(\boldsymbol{\rho}_1;\boldsymbol{\rho}_2|\boldsymbol{\rho}_1';\boldsymbol{\rho}_2') = \langle g_{ik}^*(\boldsymbol{\rho}_1|\boldsymbol{\rho}_1')g_{jl}(\boldsymbol{\rho}_2|\boldsymbol{\rho}_2') \rangle. \tag{4.6b}$$

Hereinafter, the coordinates having indices 1 and 2 will be used exclusively for quantities of the complex-conjugate wave function and the original wave function, respectively, in the manner of Eq. (4.6).

Here, to find a governing equation for $\mathbf{G}(1;2)$, we introduce a quantity $\Delta\mathbf{b}$, defined by

$$\Delta\mathbf{b} = \mathbf{b} - \mathbf{M}, \quad \langle \Delta\mathbf{b}\mathbf{g} \rangle = 0, \tag{4.7}$$

and exhibit $\mathbf{g}$ in terms of $\mathbf{G}$, on using Eq. (4.1), by

$$\mathbf{g} = \mathbf{G}[1 + \Delta\mathbf{b}\mathbf{g}], \tag{4.8}$$

and, in the same way,

$$\mathbf{g}^* = \mathbf{G}^*[1 + \Delta\mathbf{b}^*\mathbf{g}^*], \quad \langle \Delta\mathbf{b}^*\mathbf{g}^* \rangle = 0. \tag{4.9}$$

Hence, by substituting expressions (4.8) and (4.9) of $\mathbf{g}(2)$ and $\mathbf{g}^*(1)$ in the right-hand side of (4.6a), we obtain, by virtue of the condition for $\Delta\mathbf{b}$ set in Eq. (4.7), an equation of the form

$$\mathbf{G}(1;2) = \mathbf{G}^*(1)\mathbf{G}(2)[1 + \mathbf{K}(1;2)\mathbf{G}(1;2)]. \tag{4.10}$$

Here, $\mathbf{K}(1;2)$ is defined, like $\mathbf{M}$ by Eq. (4.2), according to

$$\mathbf{K}(1;2)\mathbf{G}(1;2) = \langle \Delta\mathbf{b}^*(1)\Delta\mathbf{b}(2)\mathbf{g}^*(1)\mathbf{g}(2)\rangle, \tag{4.11}$$

which gives a source term of the incoherent part of $\mathbf{G}(1;2)$. Equation (4.10) for the second-order Green's function has a form of the Bethe–Salpeter equation, as this is generally true independently of the random quantity involved and also of its statistical property, e.g., of whether it is Gaussian or not.[22]

Here, to the first order of $\mathbf{b}$ [cf. Eq. (C8)],

$$\mathbf{M} = \langle \mathbf{bGb}\rangle, \quad \langle \mathbf{b}\rangle = 0, \tag{4.12a}$$

$$\mathbf{K}(1;2) = \langle \mathbf{b}^*(1)\mathbf{b}(2)\rangle, \quad (\mathbf{b}^* = \mathbf{b}), \tag{4.12b}$$

which have the same form as the corresponding quantities in a turbulent air, for example, in Refs. 22–24 and others. Here, when setting $d = 0$, however, a special consideration is necessary, and a modified version of the statistical equations is shown in detail in Appendix C.

In another typical case of randomly distributed bosses, the corresponding expressions can be obtained in exactly the same way as for random scatterers in space, with the result[14]

$$\mathbf{M}(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_2') = n \int_{-\infty}^{\infty} d\boldsymbol{\rho}_b \langle \mathbf{T}_b^M(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_b | \boldsymbol{\rho}_2' - \boldsymbol{\rho}_b)\rangle'. \tag{4.13}$$

Here, n is the density of bosses per unit area, and $\langle\cdots\rangle'$ means the statistical averaging over all possible characteristics of the bosses, e.g., their shapes, sizes, orientations, etc.; $\mathbf{T}_b^M$ is the scattering matrix defined by Eq. (3.53) for one boss, with the replacement of $\mathbf{g}_0$ to the statistical Green's function $\mathbf{G}$ that depends on $\mathbf{M}$. In matrix form, Eq. (4.13) can be written as

$$\mathbf{M}(2) = n \int d\boldsymbol{\rho}_b \langle \mathbf{T}_b^M(2)\rangle', \tag{4.14}$$

and also provides a means to determine the still unknown $\mathbf{M}$ self-consistently.[25,26] In most cases of not too large density of the bosses, however, $\mathbf{M}$ involved in $\mathbf{T}_b^M$ may be set equal to zero, yielding an explicit expression of $\mathbf{M}$.

In the same way,

$$\mathbf{K}(1;2) = n \int d\boldsymbol{\rho}_b \langle \mathbf{T}_b^{M*}(1)\mathbf{T}_b^M(2)\rangle'. \tag{4.15}$$

Here, the $\mathbf{M}$ and $\mathbf{K}$ given by Eqs. (4.14) and (4.15) strictly satisfy optical relation (4.25) to be given later, in spite of the fact that $\mathbf{M}$ involved in $\mathbf{T}_b^M$ is not Hermitian.[25]

A. Incoherent scattering matrix

We first introduce a scattering matrix $\mathbf{S}(1;2)$, defined by

$$\mathbf{K}(1;2)\mathbf{G}(1;2) = \mathbf{S}(1;2)\mathbf{G}^*(1)\mathbf{G}(2), \tag{4.16}$$

to write Eq. (4.10) as

$$\mathbf{G}(1;2) = \mathbf{G}^*(1)\mathbf{G}(2) + \mathbf{G}^*(1)\mathbf{G}(2)\mathbf{S}(1;2)\mathbf{G}^*(1)\mathbf{G}(2). \tag{4.17}$$

Here, the substitution into the left-hand side of (4.16) yields

$$\mathbf{B}(1;2) = \mathbf{K}(1;2)[1 + \mathbf{G}^*(1)\mathbf{G}(2)\mathbf{S}(1;2)], \tag{4.18}$$

which provides an equation to obtain the unknown $\mathbf{S}(1;2)$; hence [cf. Eq. (3.53)],

$$\mathbf{S}(1;2) = [1 - \mathbf{K}(1;2)\mathbf{G}^*(1)\mathbf{G}(2)]^{-1}\mathbf{K}(1;2) \tag{4.19a}$$

$$= \mathbf{K}(1;2)[1 - \mathbf{G}^*(1)\mathbf{G}(2)\mathbf{K}(1;2)]^{-1} \tag{4.19b}$$

$$= \mathbf{K}(1;2) + \mathbf{K}(1;2)\mathbf{G}^*(1)\mathbf{G}(2)\mathbf{K}(1;2) + \mathbf{K}(1;2)\mathbf{G}^*(1)\mathbf{G}(2)\mathbf{K}(1;2)\mathbf{G}^*(1)\mathbf{G}(2)\mathbf{K}(1;2) + \cdots, \tag{4.20}$$

each term of which gives the matrix of multiple scattering of definite number of times. The matrix elements of $\mathbf{S}$ will be denoted by $S_{ij;kl}$, $S_{ij;kl}(\boldsymbol{\rho}_1;\boldsymbol{\rho}_2|\boldsymbol{\rho}_1';\boldsymbol{\rho}_2')$, and $\mathbf{S}(\boldsymbol{\rho}_1;\boldsymbol{\rho}_2|\boldsymbol{\rho}_1';\boldsymbol{\rho}_2')$ in the same way as for $\mathbf{G}(1;2)$; for example, on S_1 (Fig. 1), Eq. (4.17) becomes

$$\begin{aligned} G_{11;11}(\boldsymbol{\rho}_1;\boldsymbol{\rho}_2|\boldsymbol{\rho}_1';\boldsymbol{\rho}_2') &= G_{11}^*(\boldsymbol{\rho}_1 - \boldsymbol{\rho}_1')G_{11}(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_2') \\ &\quad + \sum_{ijkl} \int d\boldsymbol{\rho}_i\, d\boldsymbol{\rho}_j\, d\boldsymbol{\rho}_k\, d\boldsymbol{\rho}_l \\ &\quad \times G_{1i}^*(\boldsymbol{\rho}_1 - \boldsymbol{\rho}_i)G_{1j}(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_j)S_{ij;kl}(\boldsymbol{\rho}_i;\boldsymbol{\rho}_j|\boldsymbol{\rho}_k;\boldsymbol{\rho}_l) \\ &\quad \times G_{k1}^*(\boldsymbol{\rho}_k - \boldsymbol{\rho}_1')G_{l1}(\boldsymbol{\rho}_l - \boldsymbol{\rho}_2'), \end{aligned} \tag{4.21}$$

which contains the terms of $S_{22;22}$ and others having index 2, implying that $G_{11;11}$ is necessarily affected by a multiple (incoherent) scattering on the back side of S.

B. Integrated optical relation

When the medium is not dissipative, $\mathbf{M}$ and $\mathbf{K}$ are not quite independent of one another, subjected to a constraint resulting from the Hermitian condition (2.15) for $\mathbf{B}$.

We first observe, on using Eqs. (4.9) and (4.11), that

$$\langle \Delta\mathbf{b}(2)\mathbf{g}^*(1)\mathbf{g}(2)\rangle = \mathbf{G}^*(1)\mathbf{K}(1;2)\mathbf{G}(1;2),$$

and therefore, by substitution of Eq. (4.7) into the left-hand side, that

$$\langle \mathbf{b}(2)\mathbf{g}^*(1)\mathbf{g}(2)\rangle = [\mathbf{M}(2) + \mathbf{G}^*(1)\mathbf{K}(1;2)]\mathbf{G}(1;2). \tag{4.22a}$$

In the same way,

$$\langle \mathbf{b}^*(1)\mathbf{g}^*(1)\mathbf{g}(2)\rangle = [\mathbf{M}^*(1) + \mathbf{G}(2)\mathbf{K}(1;2)]\mathbf{G}(1;2). \tag{4.22b}$$

Here, the following equations can be written in a compact form, by introduction of a 2×2 matrix operator $\boldsymbol{\delta}(1;2)$, defined by the matrix elements $\delta(\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2)\delta_{ij}$, in such a way that, for any $\mathbf{A}^*(1)$ and $\mathbf{B}(2)$, the product $\boldsymbol{\delta}(1;2)\mathbf{A}^*(1)\mathbf{B}(2)$ represents

$$\begin{aligned} \sum_{ij} \int d\boldsymbol{\rho}_1\, d\boldsymbol{\rho}_2\, \delta(\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2)\delta_{ij}\, A_{ik}^*(\boldsymbol{\rho}_1|\boldsymbol{\rho}_1')B_{jl}(\boldsymbol{\rho}_2|\boldsymbol{\rho}_2') &= \sum_j \int d\boldsymbol{\rho}\, A_{kj}^\dagger(\boldsymbol{\rho}_1'|\boldsymbol{\rho})B_{jl}(\boldsymbol{\rho}|\boldsymbol{\rho}_2') \\ &= (A^\dagger B)_{kl}(\boldsymbol{\rho}_1'|\boldsymbol{\rho}_2'). \end{aligned} \tag{4.23}$$

Hence, subtracting Eq. (4.22a) from Eq. (4.22b) and then multiplying the result to the left with $\boldsymbol{\delta}(1;2)$, the left-hand side becomes zero since, by virtue of $\mathbf{b}^\dagger = \mathbf{b}$,

$$\boldsymbol{\delta}(1;2)[\mathbf{b}^*(1) - \mathbf{b}(2)]\mathbf{g}^*(1)\mathbf{g}(2) = \mathbf{g}^\dagger(\mathbf{b}^\dagger - \mathbf{b})\mathbf{g} = 0, \tag{4.24}$$

showing by the right-hand side, that

$$\boldsymbol{\delta}(1;2)\{\mathbf{M}^*(1) - \mathbf{M}(2) + [\mathbf{G}(2) - \mathbf{G}^*(1)]\mathbf{K}(1;2)\} = 0, \tag{4.25}$$

which gives the constraint inherent between **M** and **K**.

The relation (4.25) is exactly fulfilled by **M** and **K** of Eq. (4.12) obtained to the first order, as is directly shown by

$$\boldsymbol{\delta}(1;2)[\mathbf{G}(2)-\mathbf{G}^*(1)]\mathbf{K}(1;2) = \langle \mathbf{b}^\dagger(\mathbf{G}-\mathbf{G}^\dagger)\mathbf{b}\rangle = \boldsymbol{\delta}(1;2)[\mathbf{M}(2)-\mathbf{M}^*(1)]. \quad (4.26)$$

Also, the BS equation (4.10) is ensured to satisfy the averaged version of Eq. (2.33),[14] i.e.,

$$\langle \mathbf{W}\rangle = \boldsymbol{\delta}(1;2)\{2^{-1}[\mathbf{h}^*(1)+\mathbf{h}(2)]\boldsymbol{\eta}^{-1}\mathbf{G}(1;2) + (2i)^{-1}[\mathbf{G}(2)-\mathbf{G}^*(1)]\} = 0. \quad (4.27)$$

Here, the meaning of the relation becomes more explicit when written in terms of the scattering matrix $\mathbf{S}(1;2)$; this can be achieved by substituting the expression

$$\mathbf{G}^*(1)-\mathbf{G}(2) = \{i[\mathbf{h}^*(1)+\mathbf{h}(2)]\boldsymbol{\eta}^{-1}+\mathbf{B}_0(1)-\mathbf{B}_0(2) + \mathbf{M}^*(1)-\mathbf{M}(2)\}\mathbf{G}^*(1)\mathbf{G}(2), \quad (4.28)$$

derived from Eqs. (4.4b) and (4.5), into Eq. (4.25), followed by the multiplication to the right with $(2i)^{-1}[1-\mathbf{G}^*(1)\mathbf{G}(2)\mathbf{K}(1;2)]^{-1}$. Whence,

$$\boldsymbol{\delta}(1;2)\{(2i)^{-1}[\mathbf{M}^*(1)-\mathbf{M}(2)] - (2\boldsymbol{\eta})^{-1}[\mathbf{h}^*(1)+\mathbf{h}(2)]\mathbf{G}^*(1)\mathbf{G}(2)\mathbf{S}(1;2)\} = 0, \quad (4.29)$$

in consequence of Eq. (4.19b). Here, the meaning of Eq. (4.29) becomes quite clear by the multiplication to the right with $\mathbf{G}^*(1)\mathbf{G}(2)$; hence, the first term means the total coherent power absorbed by the entire surface S_1+S_2 [as may be shown by substitution of $\langle\psi^{(1)}\rangle$ from Eq. (2.17) into Eq. (2.14b) and subsequent utilization of definition (4.2) for **M**], whereas the second term means ($-$) times the total incoherent power scattered by S_1+S_2, as given by the second term in the right-hand side of Eq. (4.17). Note that this is not the case of the original relation (4.25), whose meaning is not immediately clear.

There exists also a local optical relation which ensures the conservation of the power at every point on S_1 and S_2, in contrast with the total power integrated over the entire surfaces, and can be written in terms of $\mathbf{S}(1;2)$ in a form of showing an obvious meaning of each term, as has been shown in detail in Ref. 14 for a one-side boundary.

Generally, in terms of the scattering matrix $\mathbf{S}(1;2)$, various quantities and equations associated with the incoherent waves (including cross sections and optical relations) can be written exactly and also in such a way of enabling a straightforward physical interpretation (Sec. VI). Here, $\mathbf{S}(1;2)$ is obtained as a solution of the integral equation (4.18), for which a practical version will be given later by Eq. (6.27).

V. STATISTICAL GREEN'S FUNCTIONS IN SPACE

So far the Green's functions have been defined only on the reference boundaries S_1 and S_2, but the continuation into the spaces outside the boundaries is straightforward by replacing Eqs. (2.17) by a set of equations for the space Green's function, say $g_{ij}(z|z')$, according to

$$g_{11}(z|z') = \{\exp[-ih_1|z-z'|]-\exp[-ih_1(z+z')]\} \times(2ih_1)^{-1}\eta_1+\exp(-ih_1z)g_{11}\exp(-ih_1z'),$$
$$z,z'\geqslant 0, \quad (5.1a)$$

$$g_{21}(z|z') = \exp[ih_2(z+d)]g_{21}\exp[-ih_1z'],$$
$$z\leqslant -d, \quad z'\geqslant 0. \quad (5.1b)$$

Similar equations for $g_{22}(z|z')$ and $g_{12}(z|z')$ are also obtained from Eq. (2.19). Here, the g_{ij}'s are the same operators as defined in the previous sections, and therefore $g_{ij}(z|z')$ results as the solution of

$$\left[-\left(\frac{\partial}{\partial z}\right)^2-h_i^2\right]g_{ij}(z|z') = \eta_i\,\delta_{ij}\,\delta(z-z'), \quad (5.2a)$$

subjected to boundary condition (2.10) on S_1 and S_2; the proof is straightforward by reference to the solution of Eq. (2.3) in free space. The full-coordinate expression of Eq. (5.2a) is, of course,

$$\left[-\left(\frac{\partial}{\partial \hat{x}}\right)^2-k_i^2\right]g_{ij}(\hat{x}|\hat{x}') = \eta_i\,\delta_{ij}\,\delta(\hat{x}-\hat{x}'), \quad (5.2b)$$

and $i,j=1$ or 2 according as the points $\hat{x}$ and $\hat{x}'$ are in space Σ_1 or Σ_2, respectively.

Thus, the statistical Green's function of first order in space is obtained simply by replacing all the coefficient g_{ij}'s in Eqs. (5.1) with the statistical G_{ij}'s of Eq. (4.4). The situation is the same also for the statistical Green's function of second order, and most previous equations for the statistical surface Green's functions remain unchanged with the redefinition of $\mathbf{G}^*(1)$, $\mathbf{G}(2)$, and $\mathbf{G}(1;2)$ by those of the space coordinates $\hat{x}=(\boldsymbol{\rho},z)$. That is, $\mathbf{G}(1;2)$ now represents the $\hat{x}$-coordinate matrix having the elements $G_{ij;kl}(\hat{x}_1;\hat{x}_2|\hat{x}_1';\hat{x}_2')$, with the possible abbreviations $G_{ij;kl}(z_1;z_2|z_1';z_2')$ and $\mathbf{G}(z_1;z_2|z_1';z_2')$. Here, since $\mathbf{K}(1;2)$ and $\mathbf{S}(1;2)$ are $\boldsymbol{\rho}$-coordinate matrices different from zero only on S_1 and S_2, the $\hat{x}$-coordinate expression of $\mathbf{S}(1;2)$, for example, can be written, in terms of the surface δ_j function defined by

$$\delta_j(z) = \delta(z+d_j), \quad (5.3a)$$

with $d_1=0$ and $d_2=d$, as

$$S_{ij;kl}(\hat{x}_1;\hat{x}_2|\hat{x}_1';\hat{x}_2') = \delta_i(z_1)\delta_j(z_2)S_{ij;kl}(\boldsymbol{\rho}_1;\boldsymbol{\rho}_2|\boldsymbol{\rho}_1';\boldsymbol{\rho}_2')\delta_k(z_1')\delta_l(z_2'), \quad (5.3b)$$

where $\hat{x}_j=(\boldsymbol{\rho}_j,z_j)$. Hence, expression (4.21) for $G_{11;11}(1;2)$ given in terms of $\mathbf{S}(1;2)$, for example, holds as it is, simply with the replacement of all the $\boldsymbol{\rho}$ coordinates to the $\hat{x}$'s.

Here, the equation of the first-order Green's function, (5.1) (that is, with the replacement of each $g_{ij}\to G_{ij}$), can be written in several useful forms, by the introduction of several new 2×2 matrices as follows:

$$\mathbf{g}^{(0)} = (2i\mathbf{h})^{-1}\boldsymbol{\eta} = \begin{pmatrix} g_1^{(0)} & 0 \\ 0 & g_2^{(0)} \end{pmatrix} = \begin{pmatrix} (2ih_1)^{-1}\eta_1 & 0 \\ 0 & (2ih_2)^{-1}\eta_2 \end{pmatrix}, \quad (5.4a)$$

where, from Eq. (2.3), $g_j^{(0)}$ means the Green's function in a homogeneous medium of k_j except for the constant factor η_j [but, not to be confused with g_0 defined by Eq. (3.35) for the homogeneous boundary]; also two "attenuation" coefficient matrices **A** and $\overline{\mathbf{A}}$ for the coherent wave, defined by

$$\mathbf{G} = 2\mathbf{A}\mathbf{g}^{(0)} = 2\mathbf{g}^{(0)}\overline{\mathbf{A}}, \quad (5.4b)$$

and therefore given by

$$\mathbf{A} = i\mathbf{G}\mathbf{h}\boldsymbol{\eta}^{-1} = [\mathbf{h} + i\boldsymbol{\eta}(\mathbf{B}_0 + \mathbf{M})]^{-1}\mathbf{h},$$
$$\overline{\mathbf{A}} = i\boldsymbol{\eta}^{-1}\mathbf{h}\mathbf{G} = \mathbf{A}^T, \tag{5.4c}$$

where use has been made of Eq. (4.4b). Hence, from Eq. (2.24), the reflection coefficient matrix $\langle \mathbf{R} \rangle$ becomes

$$\langle \mathbf{R} \rangle = 2\mathbf{A} - 1 = (i\mathbf{h}\boldsymbol{\eta}^{-1} - \mathbf{B}_0 - \mathbf{M})^{-1} \times (i\mathbf{h}\boldsymbol{\eta}^{-1} + \mathbf{B}_0 + \mathbf{M}), \tag{5.4d}$$

or, alternatively,

$$\langle \mathbf{R} \rangle = \mathbf{g}^{(0)}\mathbf{T}, \tag{5.4e}$$

where $\mathbf{T}$ is a scattering matrix, defined by

$$\mathbf{T} = 2i\mathbf{h}\boldsymbol{\eta}^{-1}\langle \mathbf{R} \rangle = \mathbf{T}^T, \tag{5.4f}$$

which has symmetrical matrix elements, in contrast with $\langle \mathbf{R} \rangle$ [cf. Eq. (2.30)]. Finally, we introduce a diagonal 2×2 matrix function $\mathbf{g}^{(0)}(z-z')$ defined by the elements

$$g_j^{(0)}(z-z') = g_j^{(0)} \exp[-ih_j|z-z'|], \tag{5.5a}$$

whose boundary value on S_j at $z = -d_j$ will be particularly denoted by

$$g_j^{(0)}(0|z) = g_j^{(0)}(z|0) \equiv g_j^{(0)} \exp[-ih_j|z+d_j|]. \tag{5.5b}$$

Hence, from Eqs. (2.22) and (5.4e),

$$\mathbf{G} = (1 + \langle \mathbf{R} \rangle)\mathbf{g}^{(0)} = \mathbf{g}^{(0)} + \mathbf{g}^{(0)}\mathbf{T}\mathbf{g}^{(0)}, \tag{5.6}$$

and is continued to the ouside spaces as

$$\mathbf{G}(z|z') = \mathbf{g}^{(0)}(z-z') + \mathbf{g}^{(0)}(z|0)\mathbf{T}\mathbf{g}^{(0)}(0|z'), \tag{5.7}$$

which represents the averaged version of the whole set of equations (5.1). Note that $\mathbf{g}^{(0)}(z-z') = 0$ when the points z and z' are not in the same space of either k_1 or k_2. In the same way, the continued versions of Eq. (5.4b) are

$$\mathbf{G}(0|z) = 2\mathbf{A}\mathbf{g}^{(0)}(0|z), \tag{5.8a}$$
$$\mathbf{G}(z|0) = 2\mathbf{g}^{(0)}(z|0)\overline{\mathbf{A}}, \tag{5.8b}$$

which, together with $\mathbf{G}(z|z')$, are nondiagonal 2×2 matrices having the matrix elements also for the points z and z' existing in separate spaces, in contrast with $\mathbf{g}^{(0)}(0|z)$, $\mathbf{g}^{(0)}(z|0)$, and $\mathbf{g}^{(0)}(z-z')$. Here, it may be remarked that, from Eqs. (5.4)–(5.8), all the corresponding equations in Ref. 14 are reproduced by changing the boldface letter (representing a 2×2 matrix operator) to lightface letter for every quantity involved, with $\boldsymbol{\eta} = 1$.

The continuation of the second-order Green's function into the spaces is also straightforward by use of the expression of Eq (4.17) with Eqs. (5.7) and (5.8). Hence,

$$\mathbf{G}(z_1;z_2|z_1';z_2') = \mathbf{G}^*(z_1|z_1')\mathbf{G}(z_2|z_2') + \mathbf{g}^{(0)*}(z_1|0)\mathbf{g}^{(0)}(z_2|0) \times \boldsymbol{\sigma}^{(I)}(1;2)\mathbf{g}^{(0)*}(0|z_1')\mathbf{g}^{(0)}(0|z_2'), \tag{5.9}$$

where

$$\boldsymbol{\sigma}^{(I)}(1;2) = 2^4\overline{\mathbf{A}}^*(1)\overline{\mathbf{A}}(2)\mathbf{S}(1;2)\mathbf{A}^*(1)\mathbf{A}(2), \tag{5.10}$$

and is a $\boldsymbol{\rho}$-matrix version of incoherent scattering cross section, having the matrix elements $\sigma_{ij;kl}^{(I)}(\boldsymbol{\rho}_1;\boldsymbol{\rho}_2|\boldsymbol{\rho}_1';\boldsymbol{\rho}_2')$ independent of the z coordinate. Equation (5.9) can be written as an $\hat{x}$-matrix equation, by

$$\mathbf{G}(1;2) = \mathbf{G}^*(1)\mathbf{G}(2) + \mathbf{g}^{(0)*}(1)\mathbf{g}^{(0)}(2)\boldsymbol{\sigma}^I(1;2) \times \mathbf{g}^{(0)*}(1)\mathbf{g}^{(0)}(2), \tag{5.11}$$

and, on using Eq. (5.7), it can be rewritten further in the form

$$\mathbf{G}(1;2) = \mathbf{g}^{(0)*}(1)\mathbf{g}^{(0)}(2) + \mathbf{g}^{(0)*}(1)\mathbf{g}^{(0)}(2)\boldsymbol{\sigma}^{(T)}(1;2) \times \mathbf{g}^{(0)*}(1)\mathbf{g}^{(0)}(2), \tag{5.12a}$$

in terms of a total scattering matrix $\boldsymbol{\sigma}^{(T)}(1;2)$, defined by

$$\boldsymbol{\sigma}^{(T)}(1;2) = \mathbf{T}^*(1)\mathbf{T}(2) + \boldsymbol{\sigma}^{(I)}(1;2), \tag{5.12b}$$

on neglect of the interference terms.[27]

VI. SCATTERING CROSS SECTIONS

Equation (5.9) for the second-order Green's function becomes simple in an asymptotic region far enough from the boundaries S_1 and S_2, with the aid of a representation in terms of optical quantities. From Eq. (5.8a), we first observe that

$$G_{ij}(\boldsymbol{\rho}|\hat{x}) = 2\int d\boldsymbol{\rho}'\, A_{ij}(\boldsymbol{\rho}-\boldsymbol{\rho}')g_j^{(0)}(\boldsymbol{\rho}'-\hat{x}), \tag{6.1a}$$

in view of the translational invariance; here, in such a region where $|k_j(\boldsymbol{\rho}-\hat{x})| \gg 1$ and $|\boldsymbol{\rho}'-\boldsymbol{\rho}|, d \ll |\boldsymbol{\rho}-\hat{x}|$,

$$g_j^{(0)}(\boldsymbol{\rho}'-\hat{x}) = \exp[-ik_j\boldsymbol{\Omega}\cdot(\boldsymbol{\rho}'-\boldsymbol{\rho})]\, g_j^{(0)}(\boldsymbol{\rho}-\hat{x}), \tag{6.1b}$$

with

$$\hat{\Omega} = (\boldsymbol{\Omega}, \pm\Omega_z) = (\boldsymbol{\rho}-\hat{x})/|\boldsymbol{\rho}-\hat{x}|, \quad \Omega_z > 0. \tag{6.1c}$$

Hence, Eq. (6.1a) is asymptotically given in the form

$$G_{ij}(\boldsymbol{\rho}|\hat{x}) = 2A_{ij}(\hat{\Omega})g_j^{(0)}(\boldsymbol{\rho}-\hat{x}). \tag{6.2}$$

Here, in terms of the Fourier transform $\widetilde{A}_{ij}$ given according to Eq. (5.4c) by

$$\widetilde{A}_{ij}(\mathbf{u}) = i\widetilde{G}_{ij}(\mathbf{u})\tilde{h}_j(\mathbf{u})\eta_j^{-1} = \widetilde{\overline{A}}_{ij}(\mathbf{u}), \tag{6.3}$$

$A_{ij}(\hat{\Omega})$ is defined by

$$A_{ij}(\hat{\Omega}) = \widetilde{A}_{ij}(\mathbf{u})|_{\Omega} = \overline{A}_{ji}(\hat{\Omega}), \tag{6.4}$$

where $|_{\Omega}$ means to set

$$\mathbf{u} = k_j\boldsymbol{\Omega}, \quad \tilde{h}_j(\mathbf{u}) = k_j\Omega_z. \tag{6.5}$$

On the other hand, for given $\hat{\Omega}$, $\mathbf{u}$ is undetermined by the factor k_j and therefore, whenever confusing, the notation $\hat{\Omega}^{(j)}$ will be used for $\hat{\Omega}$ to mean that $\mathbf{u} = k_j\boldsymbol{\Omega}^{(j)}$. In Eq. (6.3), $\widetilde{G}_{ij}(\mathbf{u})$ is the solution of the Fourier transform of Eq. (4.4a), i.e.,

$$\sum_j [i\tilde{h}_i(\mathbf{u})\eta_i^{-1}\delta_{ij} - \widetilde{B}_{0,ij}(\mathbf{u}) - \widetilde{M}_{ij}(\mathbf{u})]\widetilde{G}_{jk}(\mathbf{u}) = \delta_{ik}. \tag{6.6}$$

Hence, for evaluation of the second term in Eq. (5.9), we can utilize the asymptotic expression [cf. Eq. (4.17)]

$$G_{cj}^*(\boldsymbol{\rho}-\mathbf{r}/2|\hat{x})G_{dj}(\boldsymbol{\rho}+\mathbf{r}/2|\hat{x}) = A_{cj}^*(\hat{\Omega})A_{dj}(\hat{\Omega})\exp[-ik_j\boldsymbol{\Omega}\cdot\mathbf{r}] \times |2g_j^{(0)}(\boldsymbol{\rho}-\hat{x})|^2, \tag{6.7a}$$

on using the relative coordinates $\boldsymbol{\rho}$ and $\mathbf{r}$ defined by

$$\boldsymbol{\rho}_1 = \boldsymbol{\rho} - \mathbf{r}/2, \quad \boldsymbol{\rho}_2 = \boldsymbol{\rho} + \mathbf{r}/2, \quad d\boldsymbol{\rho}_1\, d\boldsymbol{\rho}_2 = d\boldsymbol{\rho}\, d\mathbf{r}. \tag{6.7b}$$

The same is also true of another factor $G_{ia}^*(\hat{x}|\boldsymbol{\rho}-\mathbf{r}/2) \times G_{ib}(\hat{x}|\boldsymbol{\rho}+\mathbf{r}/2)$.

Hence, performing the two $\mathbf{r}$ integrations in the second term of Eq. (5.9), the result becomes written, when the point

$\hat{x}_1=\hat{x}_2=\hat{x}$ is in the space of k_i and the point $\hat{x}_1'=\hat{x}_2'=\hat{x}'$ is in the space of k_j, in the form

$$G_{ii;jj}(\hat{x};\hat{x}|\hat{x}';\hat{x}') = |G_{ij}(\hat{x}|\hat{x}')|^2 + \int d\boldsymbol{\rho}''\, d\boldsymbol{\rho}'''|\hat{x}-\boldsymbol{\rho}''|^{-2}\eta_i^2 \times \sigma^{(I)}_{ii;jj}(\hat{\Omega}\,|\boldsymbol{\rho}''-\boldsymbol{\rho}'''|\hat{\Omega}')|g_j^{(0)}(\boldsymbol{\rho}'''-\hat{x}')|^2. \quad (6.8)$$

Here, $\hat{\Omega}$ and $\hat{\Omega}'$ are the unit vectors in the directions of $\hat{x}-\boldsymbol{\rho}''$ and $\boldsymbol{\rho}'''-\hat{x}'$, respectively, (Fig. 2) and

$$\sigma^{(I)}_{ii;jj}(\hat{\Omega}\,|\boldsymbol{\rho}''-\boldsymbol{\rho}'''|\hat{\Omega}') = \sum_{abcd} 4\bar{A}^*_{ia}(\hat{\Omega})\bar{A}_{ib}(\hat{\Omega})S_{ab;cd}(\hat{\Omega}^{(i)}|\boldsymbol{\rho}''-\boldsymbol{\rho}'''|\hat{\Omega}^{(j)\prime}) \times A^*_{cj}(\hat{\Omega}')A_{dj}(\hat{\Omega}'), \quad (6.9)$$

where, in terms of the matrix elements $S_{ab;cd}(\boldsymbol{\rho}_1;\boldsymbol{\rho}_2|\boldsymbol{\rho}_1';\boldsymbol{\rho}_2')$ of $\mathbf{S}(1;2)$,

$$S_{ab;cd}(\hat{\Omega}^{(i)}|\boldsymbol{\rho}''-\boldsymbol{\rho}'''|\hat{\Omega}^{(j)\prime}) = (2\pi)^{-2}\int d\mathbf{r}''\, d\mathbf{r}'''\exp[i(k_i\boldsymbol{\Omega}\cdot\mathbf{r}''-k_j\boldsymbol{\Omega}'\cdot\mathbf{r}''')] \times S_{ab;cd}(\boldsymbol{\rho}''-\mathbf{r}''/2;\boldsymbol{\rho}''+\mathbf{r}''/2|\boldsymbol{\rho}'''-\mathbf{r}'''/2;\boldsymbol{\rho}'''+\mathbf{r}'''/2), \quad (6.10)$$

which is a short-range function of $\boldsymbol{\rho}''-\boldsymbol{\rho}'''$, appreciable only within an effective range of multiple scattering [Eq. (4.20)].

Here, in view of the translational invariance, the full Fourier transform of $S_{ab;cd}$ can be written in the form

$$\tilde{S}_{ab;cd}(\boldsymbol{\lambda}_1;\boldsymbol{\lambda}_2|\boldsymbol{\lambda}_1';\boldsymbol{\lambda}_2') = (2\pi)^2\delta(\boldsymbol{\lambda}-\boldsymbol{\lambda}')\tilde{S}_{ab;cd}(\mathbf{u}|\mathbf{u}')_\lambda, \quad (6.11)$$

upon the change of the variables, according to

$$\boldsymbol{\lambda}=\boldsymbol{\lambda}_2-\boldsymbol{\lambda}_1,\quad \mathbf{u}=(\boldsymbol{\lambda}_2+\boldsymbol{\lambda}_1)/2,\quad d\boldsymbol{\lambda}\,d\mathbf{u}=d\boldsymbol{\lambda}_1\,d\boldsymbol{\lambda}_2, \quad (6.12a)$$

where, with Eq. (6.7b),

$$-\boldsymbol{\lambda}_1\cdot\boldsymbol{\rho}_1+\boldsymbol{\lambda}_2\cdot\boldsymbol{\rho}_2=\boldsymbol{\lambda}\cdot\boldsymbol{\rho}+\mathbf{u}\cdot\mathbf{r}. \quad (6.12b)$$

Hence,

$$S_{ab;cd}(\hat{\Omega}^{(i)}|\boldsymbol{\rho}|\hat{\Omega}^{(j)\prime}) = (2\pi)^{-4}\int d\boldsymbol{\lambda}\,\tilde{S}_{ab;cd}(k_i\boldsymbol{\Omega}|k_j\boldsymbol{\Omega}')_\lambda e^{-i\boldsymbol{\lambda}\cdot\boldsymbol{\rho}}, \quad (6.13a)$$

and therefore

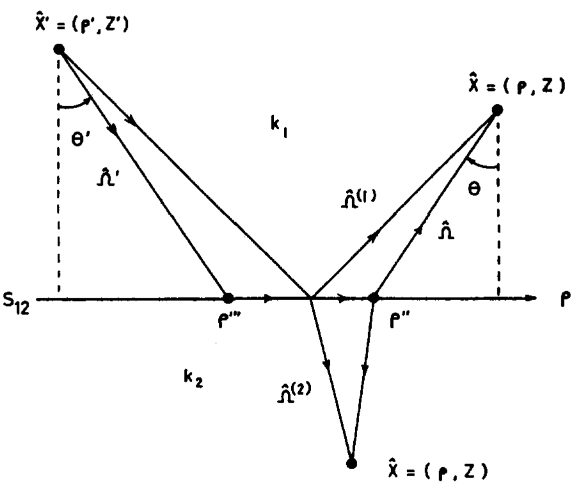


FIG. 2. Geometry and notations for Eqs. (6.8) and (6.18). $\hat{\Omega}^{(1)}$ and $\hat{\Omega}^{(2)}$ refer to the directions of specular reflection and transmission.

$$S_{ab;cd}(\hat{\Omega}^{(i)}|\hat{\Omega}^{(j)\prime}) \equiv \int d\boldsymbol{\rho}\, S_{ab;cd}(\hat{\Omega}^{(i)}|\boldsymbol{\rho}|\hat{\Omega}^{(j)\prime}) = (2\pi)^{-2}\tilde{S}_{ab;cd}(k_i\boldsymbol{\Omega}|k_j\boldsymbol{\Omega}')_\lambda|_{\lambda=0}. \quad (6.13b)$$

Here, in Eq. (6.8), when the point $\hat{x}'$ is sufficiently far from the boundary, the $\boldsymbol{\rho}'''$ integration can be performed by regarding the last factor to be constant. Hence, Eq. (6.8) becomes written, on using Eq. (6.13b), further, in the simple form

$$G_{ii;jj}(\hat{x};\hat{x}|\hat{x}';\hat{x}') = |G_{ij}(\hat{x}|\hat{x}')|^2 + \int d\boldsymbol{\rho}''|\hat{x}-\boldsymbol{\rho}''|^{-2}\eta_i^2 \times \sigma^{(I)}_{ii;jj}(\hat{\Omega}\,|\hat{\Omega}')|g_j^{(0)}(\boldsymbol{\rho}''-\hat{x}')|^2, \quad (6.14)$$

in terms of an incoherent scattering cross section per unit area $\sigma^{(I)}_{ii;jj}(\hat{\Omega}\,|\hat{\Omega}')$, defined by [cf. Eq. (5.10)]

$$\sigma^{(I)}_{ii;jj}(\hat{\Omega}\,|\hat{\Omega}') = \sum_{abcd} 4\bar{A}^*_{ia}(\hat{\Omega})\bar{A}_{ib}(\hat{\Omega}) \times S_{ab;cd}(\hat{\Omega}^{(i)}|\hat{\Omega}^{(j)\prime})A^*_{cj}(\hat{\Omega}')A_{dj}(\hat{\Omega}'). \quad (6.15)$$

Here, $\sigma^{(I)}_{11;11}$, for example, contains the terms of $S_{22;22}$ and others of index 2, meaning that it also contains an effect of multiple scattering over the back side of S.[28]

In the same way, also for the coherent term in Eq. (6.8), we obtain an expression of the form (Appendix A)

$$|G_{ij}(\hat{x}|\hat{x}')|^2 = |g_j^{(0)}(\hat{x}-\hat{x}')|^2\,\delta_{ij} + \int d\boldsymbol{\rho}''|\hat{x}-\boldsymbol{\rho}''|^{-2} \times \Omega_z^{(i)}|\langle R_{ij}(\hat{\Omega})\rangle|^2\,\delta(\hat{\Omega}^{(i)}-\hat{\Omega}^{(i)\prime}) \times |g_j^{(0)}(\boldsymbol{\rho}''-\hat{x}')|^2. \quad (6.16)$$

Here, $\hat{\Omega}^{(i)\prime}=\hat{\Omega}^{(i)}(\boldsymbol{\Omega}')$ means the unit vector in the direction of wave propagation when the incident wave of given $\mathbf{u}=k_j\boldsymbol{\Omega}'$ is observed in the medium of k_i (Fig. 2).[27]

Thus, the resultant asymptotic expression of the second-order Green's function is obtained, on substituting Eq. (6.16) into (6.14), in the form

$$G_{ii;jj}(\hat{x};\hat{x}|\hat{x}';\hat{x}') = |g_i^{(0)}(\hat{x}-\hat{x}')|^2\,\delta_{ij} + \int d\boldsymbol{\rho}''|\hat{x}-\boldsymbol{\rho}''|^{-2} \times \sigma^{(T)}_{ii;jj}(\hat{\Omega}\,|\hat{\Omega}')|g_j^{(0)}(\boldsymbol{\rho}''-\hat{x}')|^2. \quad (6.17)$$

Here,

$$\sigma^{(T)}_{ii;jj}(\hat{\Omega}\,|\hat{\Omega}') = \Omega_z^{(i)}|\langle R_{ij}(\hat{\Omega})\rangle|^2\,\delta(\hat{\Omega}^{(i)}-\hat{\Omega}^{(i)\prime}) + \eta_i^2\sigma^{(I)}_{ii;jj}(\hat{\Omega}\,|\hat{\Omega}'), \quad (6.18)$$

and provides the total scattering cross section per unit area, including the coherent one; it can also be regarded as a specific expression of the operator equation (5.12b), except for the factor η_i^2.

A. Optical relations

When the medium is nondissipative, as has been assumed, the cross section $\sigma^{(T)}(\hat{\Omega}\,|\hat{\Omega}')$ of Eq. (6.18) should be subjected to an optical relation that ensures the power conservation in the scattering. It is derived from the Fourier transform of relation (4.29), which can be written, on using Eq. (6.11), as

$$\tilde{\gamma}_{cd}(\mathbf{u}') = \sum_{jab} (2\pi)^{-2} \int_{|\mathbf{u}|<k_j} d\mathbf{u}\, \tilde{h}_j(\mathbf{u})\eta_j^{-1} \times \widetilde{G}^*_{ja}(u)\widetilde{G}_{jb}(\mathbf{u})\widetilde{S}_{ab;cd}(\mathbf{u}|\mathbf{u}')_\lambda|_{\lambda=0}, \tag{6.19}$$

where

$$\tilde{\gamma}_{cd}(\mathbf{u}) = (2i)^{-1}[\widetilde{M}^*_{dc}(\mathbf{u}) - \widetilde{M}_{cd}(\mathbf{u})], \tag{6.20a}$$

$$\gamma_{cd}(\hat{\Omega}) = \tilde{\gamma}_{cd}(\mathbf{u})|_{\Omega} = \gamma^*_{dc}(\hat{\Omega}), \tag{6.20b}$$

and $\widetilde{G}_{jb}$ is given by the solution of Eq. (6.6).[29]

Here, changing the variable of integration $\mathbf{u}$ to $\hat{\Omega}$ with Eq. (A8), Eq. (6.19) becomes written as

$$\gamma_{cd}(\hat{\Omega}') = \sum_{jab} k_j\eta_j \int_{2\pi} d\hat{\Omega}\, \overline{A}^*_{ja}(\hat{\Omega})A_{jb}(\hat{\Omega}) \times S_{ab;cd}(\hat{\Omega}|\hat{\Omega}'), \tag{6.21}$$

and, in terms of $\sigma^{(I)}_{ij;kl}(\hat{\Omega}|\hat{\Omega}')$ [given by an obvious generalization of Eq. (6.15) with respect to the indices], further as

$$4\sum_{cd} \gamma_{cd}(\hat{\Omega}')A^*_{ca}(\hat{\Omega}')A_{db}(\hat{\Omega}') = \sum_j k_j\eta_j \int_{2\pi} d\hat{\Omega}\, \sigma^{(I)}_{jj;ab}(\hat{\Omega}|\hat{\Omega}'), \tag{6.22}$$

which provides a fundamental relation inherent between the quantities of the coherent wave on the left-hand side and those of the incoherent wave on the right-hand side.

Also, for the coherent wave, there exists a similar relation, as has been given by Eq. (2.34) for the original reflection coefficient $\mathbf{R}$. That is, using Eqs. (4.4) and (4.5) leads to the relation

$$4\mathbf{A}^\dagger(2i)^{-1}(\mathbf{M}^\dagger - \mathbf{M})\mathbf{A} + \langle\mathbf{R}^\dagger\rangle(2\boldsymbol{\eta})^{-1}(\mathbf{h}^\dagger + \mathbf{h})\langle\mathbf{R}\rangle - (2\boldsymbol{\eta})^{-1}(\mathbf{h}^\dagger + \mathbf{h}) + \langle\mathbf{R}^\dagger\rangle(2\boldsymbol{\eta})^{-1}(\mathbf{h}^\dagger - \mathbf{h}) + (2\boldsymbol{\eta})^{-1}(\mathbf{h} - \mathbf{h}^\dagger)\langle\mathbf{R}\rangle = 0, \tag{6.23}$$

which, by the Fourier transformation, can be written in terms of $\tilde{\gamma}_{ij}(\mathbf{u})$ of Eq. (6.20a) as

$$\sum_{ij} 4\tilde{\gamma}_{ij}(\mathbf{u})\widetilde{A}^*_{ia}(\mathbf{u})\widetilde{A}_{jb}(\mathbf{u}) + \sum_j \tilde{h}_j(\mathbf{u})\eta_j^{-1}\langle\widetilde{R}^*_{ja}(\mathbf{u})\rangle\langle\widetilde{R}_{jb}(\mathbf{u})\rangle - \tilde{h}_a(\mathbf{u})\eta_a^{-1}\delta_{ab} = 0, \tag{6.24a}$$

within the optical region where $\tilde{h}^*_a = h_a$ and $\tilde{h}^*_b = \tilde{h}_b$, so that no contribution is made from the last two terms in Eq. (6.23). Hence, in terms of the $\hat{\Omega}$ variables, it can be written further as

$$\sum_{ij} 4\gamma_{ij}(\hat{\Omega})A^*_{ia}(\hat{\Omega})A_{jb}(\hat{\Omega}) = k_a\eta_a^{-1}\Omega_z^{(a)}\delta_{ab} - \sum_j k_j\eta_j^{-1}\Omega_z^{(j)}\langle R^*_{ja}(\hat{\Omega})\rangle\langle R_{jb}(\hat{\Omega})\rangle, \tag{6.24b}$$

where $\Omega_z^{(j)}$ is defined by Eq. (A10).

Hence, by combining relations (6.22) and (6.24b), the total cross section $\sigma^{(T)}$ of Eq. (6.18) is found to be subjected to the simple relation

$$\sum_j k_j\eta_j^{-1} \int_{2\pi} d\hat{\Omega}\, \sigma^{(T)}_{jj;ab}(\hat{\Omega}|\hat{\Omega}') = k_a\eta_a^{-1}\Omega_z^{(a)}\delta_{ab}. \tag{6.25}$$

In fact, the total power of the scattered waves becomes, on using the second term on the right-hand side of Eq. (6.17), as

$$\int d\boldsymbol{\rho}'' \sum_i k_i\eta_i^{-1} \int_{2\pi} d\hat{\Omega}\, \sigma^{(T)}_{ii;jj}(\hat{\Omega}|\hat{\Omega}')|g_j^{(0)}(\boldsymbol{\rho}'' - \hat{x}')|^2 = k_j\eta_j^{-1} \int_{2\pi} d\hat{\Omega}'|\boldsymbol{\rho}'' - \hat{x}'|^2|g_j^{(0)}(\boldsymbol{\rho}'' - \hat{x}')|^2, \tag{6.26}$$

in consequence of Eq. (6.25) and $d\boldsymbol{\rho}'' = \Omega_z'^{-1}|\boldsymbol{\rho}'' - \hat{x}'|^2 d\hat{\Omega}'$, giving exactly the total power of the incident wave.

Here, it may be remarked that, in both optical relations (6.22) and (6.25), the $\hat{\Omega}$ range of integration is limited strictly within the half-solid angle 2π, and this is a consequence of using Eq. (6.19) or, more originally, Eq. (4.29) that is given in terms of the incoherent scattering matrix $\mathbf{S}(1;2)$. On the other hand, the range of integration would not be limited within the 2π if use was made of the original relation of Eq. (4.25) given in terms of $\mathbf{K}(1;2)$, extending over an additional range of integration in the complex plane, as in the case of a one-side boundary.[14]

On the other hand, in order to obtain the incoherent scattering cross section given by Eq. (6.15), we must find the factor $S_{ab;cd}(\hat{\Omega}|\hat{\Omega}')$ from $\mathbf{S}(1;2)$, and the latter is the solution of integral equation (4.18). Here, the former can be directly obtained as the solution of an integral equation, given by rewriting the original equation as

$$S_{ab;cd}(\hat{\Omega}|\hat{\Omega}') = S^{(0)}_{ab;cd}(\hat{\Omega}|\hat{\Omega}') + \sum_{ijkl} \eta_i\eta_j \int (d\hat{\Omega}''\,\Omega_z''^{-1})_{ij} \times S^{(0)}_{ab;ij}(\hat{\Omega}|\hat{\Omega}'')\overline{A}^*_{ik}(\hat{\Omega}'')\overline{A}_{jl}(\hat{\Omega}'')S_{kl;cd}(\hat{\Omega}''|\hat{\Omega}'). \tag{6.27}$$

Here,

$$(d\hat{\Omega}\,\Omega_z^{-1})_{ij} = d\mathbf{u}[\tilde{h}^*_i(\mathbf{u})\tilde{h}_j(\mathbf{u})]^{-1} = k_ik_j^{-1}\,d\hat{\Omega}^{(i)}\Omega_z^{(j)-1} = k_jk_i^{-1}\,d\hat{\Omega}^{(j)}\,\Omega_z^{(i)-1}, \tag{6.28}$$

and

$$S^{(0)}_{ab;cd}(\hat{\Omega}|\hat{\Omega}') = (2\pi)^{-2}\widetilde{K}_{ab;cd}(\mathbf{u}|\mathbf{u}')_\lambda|_{\lambda=0;\Omega}, \tag{6.29}$$

where $\widetilde{K}_{ab;cd}(\mathbf{u}|\mathbf{u}')_\lambda$ is the Fourier transform of $\mathbf{K}(1;2)$, defined in the same way as $\widetilde{S}_{ab;cd}(\mathbf{u}|\mathbf{u}')_\lambda$ by Eq. (6.11) for $\mathbf{S}(1;2)$. The path of integration is again not limited within the half-solid angle 2π, unlike those in Eqs. (6.21) and (6.25). In terms of the angles of incident and reflected waves, say θ' and θ, defined by $\Omega_z' = \cos\theta'$ and $\Omega_z = \cos\theta$, respectively, it goes from 0 to $\pi/2 + i\infty$ via $\pi/2$.

Here, the first term $S^{(0)}_{ab;cd}(\hat{\Omega}|\hat{\Omega}')$ and the factor $\overline{A}^*_{ik}A_{jl}(\hat{\Omega}'')$ may be given by using $K_{ab;cd}$ and G_{ab} of appropriate approximation, but, in order to ensure the optical relation (6.21) for the solution $S_{ab;cd}(\hat{\Omega}|\hat{\Omega}')$, they should be consistent with the condition of Eq. (4.25), as it is exactly so in the typical cases of both slightly random surface [Eqs. (4.12)] and randomly distributed bosses [Eqs. (4.14) and (4.15)]. The condition can be rewritten in the present notations also as

$$\gamma_{cd}(\hat{\Omega}') = \sum_{ijab} \eta_i\eta_j \int (d\hat{\Omega}\,\Omega_z^{-1})_{ij}[k_i\eta_i^{-1}\Omega_z\delta_{ij} + \gamma_{ij}(\hat{\Omega})]\overline{A}^*_{ia}(\hat{\Omega})A_{jb}(\hat{\Omega})S^{(0)}_{ab;cd}(\hat{\Omega}|\hat{\Omega}'). \tag{6.30}$$

Here, it may be remarked that, if the term $\gamma_{ij}(\hat{\Omega})$ is negligible in the integrand of (6.30), $S^{(0)}_{ab;cd}(\hat{\Omega}|\hat{\Omega}')$ results in satisfying the same relation as Eq. (6.21), implying, therefore, that $S_{ab;cd} \sim S^{(0)}_{ab;cd}$. It implies also that, in Eq. (6.30), the range of integration for the term $\gamma_{ij}(\hat{\Omega})$ is not limited within 2π because of the original definition (6.20), as in Eq. (6.27), in contrast with the range of Eq. (6.21) limited strictly within 2π.

So far we have considered a general class of scalar waves, subjected to boundary condition (2.8), and therefore, for electromagnetic waves, the results are applicable only when the waves are scattered within the plane of incidence. Briefly shown in Appendix B is a generalized version of several basic equations to meet the three-dimensional scattering of electromagnetic waves.

B. Case of a slightly random boundary

Basic equations in this case were obtained in Sec. III C, where what is meant by Eq. (3.37) is that $i\mathbf{h}\boldsymbol{\eta}^{-1} - \mathbf{B}_0 = \mathbf{g}_0^{-1}$ ($d = 0$) does not exist, so that the basic equation of $\mathbf{g}$ cannot be given by Eq. (4.1) with a Hermitian $\mathbf{b}$. Nevertheless, this does not mean that all the equations of the statistical Green's functions obtained in Sec. IV need to be modified to meet the special situation; in Appendix C, the modified version of the equations is shown in detail, including a BS equation consistent (exactly) with the power conservation, which has the same form as Eq. (4.10), including $\mathbf{M}$ and $\mathbf{K}$ given by Eq. (4.12). Here, since the equations have been obtained to the first order of ζ, they may first be considered not to be significantly used to obtain a higher-order effect, like multiple scattering. But, it may be remarked that, as when treating waves in a turbulent air (Sec. III C), the solution of the present BS equation of ladder type [Eq. (C11)] is considered to be significant over all the orders of ζ in the sense of selected summation, so that the effect of multiple scattering is still demonstrated by the series of Eq. (4.20).

Here, when only the first term of the series is considered, an operator version of the cross sections is given by Eqs. (C19) and (C20) of a form similar to Eq. (5.10). Here, by definition (6.4), $A_{11}(\hat{\Omega})$ and $\overline{A}_{21}(\hat{\Omega})$ are obtained, on setting $k_j^2 = \epsilon_j k_0^2$, $j = 1,2$, in Eqs. (C17) and (C18) [with g_0 of Eq. (3.35)], as

$$A_{11}(\hat{\Omega}) = \Omega_z[\Omega_z + (\eta_1/\eta_2)(\epsilon_2/\epsilon_1 - \mathbf{\Omega}^2)^{1/2}]^{-1} = \overline{A}_{11}(\hat{\Omega}), \tag{6.31a}$$

$$\overline{A}_{21}(\hat{\Omega}) = \Omega_z[\Omega_z + (\eta_2/\eta_1)(\epsilon_1/\epsilon_2 - \mathbf{\Omega}^2)^{1/2}]^{-1}, \tag{6.31b}$$

where the angle $\hat{\Omega} = (\mathbf{\Omega}, \Omega_z)$ denotes $\hat{\Omega}^{(1)}$ and $\hat{\Omega}^{(2)}$ [Eq. (A10)] in the respective equations. In terms of a "power" spectrum density function of $\zeta(\boldsymbol{\rho})$, $W(\boldsymbol{\lambda})$, defined by

$$W(\boldsymbol{\lambda}) = 4(2\pi)^{-2}\int d\boldsymbol{\rho}\, \exp[i\boldsymbol{\lambda}\cdot(\boldsymbol{\rho} - \boldsymbol{\rho}')]\langle\zeta(\boldsymbol{\rho})\zeta(\boldsymbol{\rho}')\rangle, \tag{6.32}$$

use of $b^{(11)}$ and $b^{(12)}$ of Eqs. (3.33) and (3.41b) yields the cross section $\sigma^{(I)}_{jj;11}(\hat{\Omega}|\hat{\Omega}')$ of the form [cf. Eq. (6.14)]

$$\eta_j^2\sigma^{(I)}_{jj;11}(\hat{\Omega}|\hat{\Omega}') = k_0^4 W[k_0(\epsilon_1^{1/2}\mathbf{\Omega}' - \epsilon_j^{1/2}\mathbf{\Omega})] \times |\eta_j \overline{A}_{j1}(\hat{\Omega})[f_{j1}(\hat{\Omega}|\hat{\Omega}') + \epsilon_1\eta_1^{-1} - \epsilon_2\eta_2^{-1}]A_{11}(\hat{\Omega}')|^2. \tag{6.33}$$

Here,

$$f_{11}(\hat{\Omega}|\hat{\Omega}') = (\eta_1^{-1} - \eta_2^{-1})\epsilon_1[\eta_1\eta_2^{-1}(\epsilon_2/\epsilon_1 - \mathbf{\Omega}^2)^{1/2} \times (\epsilon_2/\epsilon_1 - \mathbf{\Omega}'^2)^{1/2} - \mathbf{\Omega}\cdot\mathbf{\Omega}'], \tag{6.34a}$$

$$f_{21}(\hat{\Omega}|\hat{\Omega}') = -(\eta_1^{-1} - \eta_2^{-1})(\epsilon_1\epsilon_2)^{1/2}[(\epsilon_1/\epsilon_2 - \mathbf{\Omega}^2)^{1/2} \times (\epsilon_2/\epsilon_1 - \mathbf{\Omega}'^2)^{1/2} + \mathbf{\Omega}\cdot\mathbf{\Omega}']. \tag{6.34b}$$

Here, for sound waves, $\eta \neq \epsilon$, while, for electromagnetic waves of vertical polarization (scattered within the plane of incidence) $\eta = \epsilon$, so that the last two terms in Eq. (6.33) are cancelled exactly. On the other hand, in the case of horizontal polarization, $\eta = \mu = 1$ and therefore there is entirely no contribution from all the terms of f_{j1}. In either case, the backscattering cross sections given by setting $\hat{\Omega}' = -\hat{\Omega}$ agree with those obtained by previous authors,[3,5,30] whereas, for the transmitted waves, no cross section has been shown so far in the literature.

In the special case of $\eta_2 = \infty$, Eq. (6.31a) gives $A_{11}(\hat{\Omega}) = 1$ independent of $\hat{\Omega}$, and this unphysical result[31] is a consequence of Eq. (C17) using the approximation of $\mathbf{G} = \boldsymbol{\pi}^{-1}$. If evaluated exactly by using Eq. (C4), A_{11} tends to zero as $\hat{\Omega}_z \to 0$ [Eq. (C22)]. On the other hand, when $|k_2| \sim \infty$ such that $|\partial/\partial\boldsymbol{\rho}| \ll |k_2|$, we observe that $h_2 \simeq k_2$ and therefore, from Eq. (3.11), $B_S^{(1)} = -ik_2\eta_2^{-1}n_z^{(s)}$. Hence, by the substitution into Eq. (3.30), the equation of g_{11} is obtained again in the form of Eq. (3.31), wherein, in the particular case of either $\eta_1 = \eta_2 = 1$ or $k_2\eta_2^{-1} \sim 0$, $b^{(11)}$ is exactly the same as those given by Eq. (3.33) at the same limits. This means that, in spite of the failure of fulfilling the condition $|h_2\zeta| \ll 1$ necessary to derive the perturbative results, the backscattering cross sections of electromagnetic waves for a perfectly conducting boundary coincide with those given by Eqs. (6.33) and (6.34) and also by previous authors, for both polarizations.

VII. SUMMARY AND DISCUSSIONS

Scattering and transmission of waves through a rough boundary were systematically investigated for both coherent and incoherent waves, based on the previous method[14] of using surface Green's function for a one-side boundary. Except in Appendix B for electromagnetic waves in the general case, the waves were assumed to be subjected to the boundary conditions of Eq. (2.8), which meet a wide class of scalar waves, including sound waves and electromagnetic waves of both horizontal and vertical polarizations in two-dimensional space. The boundary condition can be transferred onto two reference boundary planes, so chosen that the entire boundary is completely involved between the two planes (Fig. 1); hence, it can be given by Eq. (2.10) in terms of surface impedances B_{ij}. The surface Green's function $\mathbf{g}$ is a 2×2 matrix operator having the elements g_{ij} defined according to Eqs. (2.17) and (2.19), and is governed by Eq. (2.21), which has the same form as that of the ordinary Green's function in a random medium $\mathbf{B}$. Two operator methods were shown to enable the Green's function to be obtained exactly and in a compact form for given boundary change $\zeta(\boldsymbol{\rho})$. In one method, the continuation of wave functions on different sides of the boundary is performed directly on the real boundary, whereas, in the other method, it is done rather indirectly, based on the Green's theorem. The equation of the Green's function can be converted to a set of integral equations for the reflection and transmission coefficients, which is an ordinary integral equation entirely free from any operator. The Green's function is determined once the surface impedances are given, and vice versa. However, the former is a long-range function, whereas the latter is a short-range function of the order of correlation distance of $\zeta(\boldsymbol{\rho})$, as is clear in either case of slightly rough boundary or of random bosses, for example; therefore, the latter may be considered to be a fundamental quantity to describe the entire statistical system. Various equations of the statistical Green's functions can be obtained unperturbatively in the same way as those in a random medium, particularly with the Bethe–Salpeter equation (4.10) for the coherence function; most of the basic equations can be given by the same equations as those for a one-side boundary,[14] with the replacement of the boldface letter (representing 2×2 matrix operator) to the corresponding lightface letter for each symbol. Examples of the BS equation were shown for two typical cases of a slightly random boundary and an embossed boundary. The incoherent scattering matrix $\mathbf{S}(1;2)$ enables various quantities and equations associated with the incoherent wave to be written exactly, including scattering cross sections, optical relations, power flux density, etc. The scattering cross sections are obtained from asymptotic expressions of the second-order Green's function in a region sufficiently far from the boundary. The multiple scattering is necessarily involved [Eq. (4.20)], and the power conservation is accomplished only with this effect, leading to a constraint [Eq. (4.29)] which eventually leads to the optical relations [e.g., Eq. (6.25)]. There exists also a local optical relation established at every point on the reference boundaries.[14] Approximations were made only when obtaining the perturbative results. Here, when applying the conventional method of setting $d=0$ to a slightly rough boundary, a special treatment is necessary primarily because of the nonexistence of $\mathbf{B}_0$ in Eq. (4.1), resulting from the symmetry of Eq. (3.37); the details of basic equations were shown in Sec. III C [where Eq. (3.47a) plays the role of Eq. (4.1)], and resulting statistical equations therefrom were in Appendix C. Here, it may be remarked that the present perturbative equations, including equations of multiple scattering, are consistent with the power conservation and are considered to be significant over all the orders of ζ, in the sense of selected summations, as when treating a turbulent air. As for the cross sections by single scattering, no result has been shown for the transmitted waves in the literature, while, for the reflected waves, the cross sections obtained by previous authors are reproduced.

Another important aspect of the present theory is that the theory can be generalized to such a case where the medium in one or both spaces outside the boundary is also random, and further to the case of a random layer with two rough boundaries, to solve boundary value problems in random layer transport. Here, the conventional transport equation may be utilized for each random medium as a natural consequence of making an approximation (as has been known to be converted from each BS equation to a good approximation[32,25]). The present theory can directly provide an exact version of the conventional boundary condition for the transport equation, which has been given so far heuristically. More specifically, in the case of one rough boundary, say S_{12}, dividing a random medium $\mathbf{q}$ into two parts, say q_1 and q_2 of different kind, the BS equation of the mutual coherence function in the system of $\mathbf{q}$ plus S_{12}, say $\mathbf{G}^{(q+12)}(1;2)$, can be written in the form

$$\mathbf{G}^{(q+12)}(1;2)=\mathbf{G}^{*}(1)\mathbf{G}(2)[1+\{\mathbf{K}^{(12)}(1;2)+\mathbf{K}^{(q)}(1;2)\}\times\mathbf{G}^{(q+12)}(1;2)],\tag{7.1}$$

similar to Eq. (4.10). In fact, when the medium is homogeneous, then $\mathbf{K}^{(q)}(1;2)=0$ and the equation of $\mathbf{G}^{(q+12)}=\mathbf{G}^{(12)}$, say, is reduced to

$$\mathbf{G}^{(12)}(1;2)=\mathbf{G}^{*}(1)\mathbf{G}(2)[1+\mathbf{K}^{(12)}(1;2)\mathbf{G}^{(12)}(1;2)],\tag{7.2}$$

which is exactly the spatialy continued version of Eq. (4.10) with $\mathbf{K}(1;2)=\mathbf{K}^{(12)}(1;2)$ (Sec. V). Here, the first-order Green's function $\mathbf{G}$ is the same as that in Sec. V, being the solution of Eq. (5.2) subjected to Eq. (2.10) with $\mathbf{B}\rightarrow\mathbf{B}_0+\mathbf{M}$; here, when $\mathbf{q}\neq 0$, the equations slightly differ from the previous ones by the replacement of k_j to $k_j^{(M)}$ (of complex values), resulting from the medium fluctuation. A great advantage of Eq. (7.1) is that it can be rewritten, in terms of $\mathbf{G}^{(12)}(1;2)$, as

$$\mathbf{G}^{(q+12)}(1;2)=\mathbf{G}^{(12)}(1;2)[1+\mathbf{K}^{(q)}(1;2)\mathbf{G}^{(q+12)}(1;2)].\tag{7.3}$$

Here, the factor $\mathbf{G}^{(12)}(1;2)$ can be given by Eq. (5.12a), for example, in terms of known $\boldsymbol{\sigma}^{(T)}(1;2)$, which may be a good observable even though the original $\mathbf{K}^{(12)}(1;2)$ may not be. It therefore provides a direct means of solving the boundary-value problem of wave transport in the medium $\mathbf{q}$, in terms of $\mathbf{K}^{(q)}$ and $\mathbf{M}^{(q)}$ (corresponding to the boundary counterpart $\mathbf{M}^{(12)}=\mathbf{M}$). An alternative expression of $\mathbf{G}^{(q+12)}(1;2)$ corresponding to Eq. (5.12a) can be written in the form

$$\mathbf{G}^{(q+12)}(1;2)=\mathbf{G}^{(0q)}(1;2)+\mathbf{G}^{(0q)}(1;2)\boldsymbol{\sigma}^{(12/q)}(1;2)\mathbf{G}^{(0q)}(1;2).\tag{7.4}$$

Here, $\mathbf{G}^{(0q)}$ is the solution of Eq. (7.1) when $\mathbf{K}^{(12)}=0$ and the boundary is removed (no reflection), and is, therefore, a diagonal 2×2 matrix function whose elements are the solutions in a homogeneous random medium of each space; $\boldsymbol{\sigma}^{(12/q)}$ is an effective total scattering matrix of the boundary when affected by the medium $\mathbf{q}$, and can be written in terms of the unaffected $\boldsymbol{\sigma}^{(12)}=\boldsymbol{\sigma}^{(T)}$ [Eq. (5.12b)] and an incoherent scattering matrix of the medium, say $\mathbf{S}^{(0q)}(1;2)$, which enables $\mathbf{G}^{(0q)}(1;2)$ to be written in the form of Eq. (4.17). Also, for a random layer with two rough boundaries, the method remains essentially the same. Note that, in the BS equation (7.1), the boundary and the medium are now involved on exactly the same footing. The details will be treated in a different paper.[33]

APPENDIX A: DERIVATION OF EQ. (6.16)

From Eq. (5.7), it is straightforward to obtain

$$
\begin{aligned}
|G_{ij}(\hat{x}|\hat{x}')|^2 &= |g_i^{(0)}(\hat{x}-\hat{x}')|^2\,\delta_{ij} + \int d\boldsymbol{\rho}''\,d\boldsymbol{\rho}''' \int d\mathbf{r}''\,d\mathbf{r}''' \\
&\quad \times |g_i^{(0)}(\hat{x}-\boldsymbol{\rho}'')|^2 \exp[ik_i\boldsymbol{\Omega}\cdot\mathbf{r}'']\,T_{ij}^*[\boldsymbol{\rho}''-\boldsymbol{\rho}''' \\
&\quad -(\mathbf{r}''-\mathbf{r}''')/2]\,T_{ij}[\boldsymbol{\rho}''-\boldsymbol{\rho}'''+(\mathbf{r}''-\mathbf{r}''')/2] \\
&\quad \times \exp[-ik_j\boldsymbol{\Omega}'\cdot\mathbf{r}''']\,|g_j^{(0)}(\boldsymbol{\rho}'''-\hat{x}')|^2, \qquad \text{(A1)}
\end{aligned}
$$

on following the procedure for deriving Eq. (6.7a) and also neglecting the interference terms. Here, the integration can be effected in the same way as in Eqs. (6.8)–(6.10), with the replacement of the Fourier transform (6.11) to $(\mathbf{A}=\frac{1}{2})$

$$
\begin{aligned}
&\widetilde{T}_{ij}^*(\boldsymbol{\lambda}_1)\widetilde{T}_{ij}(\boldsymbol{\lambda}_2)(2\pi)^4\,\delta(\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_1')\delta(\boldsymbol{\lambda}_2-\boldsymbol{\lambda}_2') \\
&\quad \simeq |\widetilde{T}_{ij}(\mathbf{u})|^2(2\pi)^4\,\delta(\mathbf{u}-\mathbf{u}')\delta(\boldsymbol{\lambda}-\boldsymbol{\lambda}'), \qquad \text{(A2)}
\end{aligned}
$$

to the approximation of $\widetilde{T}_{ij}(\mathbf{u}+\boldsymbol{\lambda}/2)\simeq\widetilde{T}_{ij}(\mathbf{u})$ where $|\boldsymbol{\lambda}|<|\mathbf{u}|\sim k$, yielding a result of the form

$$
\begin{aligned}
&\int d\boldsymbol{\rho}''\,|\hat{x}-\boldsymbol{\rho}''|^{-2}|2^{-1}\eta_i T_{ij}(\hat{\Omega})|^2 \\
&\quad \times\delta(k_i\boldsymbol{\Omega}-k_j\boldsymbol{\Omega}')|g_j^{(0)}(\boldsymbol{\rho}''-\hat{x}')|^2. \qquad \text{(A3)}
\end{aligned}
$$

Here,

$$
\begin{aligned}
T_{ij}(\hat{\Omega}) &= T_{ji}(\hat{\Omega}) = \widetilde{T}_{ij}(\mathbf{u})|_{\Omega} \qquad \text{(A4)}\\
&= 2ik_i\eta_i^{-1}\Omega_z^{(i)}\langle R_{ij}(\hat{\Omega})\rangle, \qquad \text{(A5)}
\end{aligned}
$$

where

$$
\langle R_{ij}(\hat{\Omega})\rangle = \langle\widetilde{R}_{ij}(\mathbf{u})\rangle|_{\Omega} \neq \langle R_{ji}(\hat{\Omega})\rangle \quad (i\neq j), \qquad \text{(A6)}
$$

in terms of $\Omega_z^{(i)}$ to be defined by Eq. (A10).

Here, since $\mathbf{u}=k\,\boldsymbol{\Omega}$ and $\hat{\Omega}=(\boldsymbol{\Omega},\pm\Omega_z)$, $\Omega_z>0$, we get, on changing the variable $\mathbf{u}$ to $\hat{\Omega}$,

$$
\delta(\mathbf{u}) = \delta(\boldsymbol{\Omega})\left|\frac{\partial\boldsymbol{\Omega}}{\partial\mathbf{u}}\right| = \delta(\hat{\Omega})\left|\frac{\partial\hat{\Omega}}{\partial\mathbf{u}}\right|, \qquad \text{(A7)}
$$

with the Jacobians

$$
\left|\frac{\partial\boldsymbol{\Omega}}{\partial\mathbf{u}}\right| = k^{-2}, \quad \left|\frac{\partial\hat{\Omega}}{\partial\mathbf{u}}\right| = (k^2\Omega_z)^{-1}. \qquad \text{(A8)}
$$

Hence,

$$
\delta(k_i\boldsymbol{\Omega}-k_j\boldsymbol{\Omega}') = (k_i^2\Omega_z^{(i)})^{-1}\delta(\hat{\Omega}^{(i)}-\hat{\Omega}^{(i)\prime}), \qquad \text{(A9)}
$$

where $\hat{\Omega}^{(i)}=(\boldsymbol{\Omega}^{(i)},\pm\Omega_z^{(i)})$ is defined by

$$
\mathbf{u}=k_i\boldsymbol{\Omega}^{(i)}=k_j\boldsymbol{\Omega}^{(j)}, \quad \Omega_z^{(i)}=[1-(\boldsymbol{\Omega}^{(i)})^2]^{1/2}, \qquad \text{(A10)}
$$

and means the direction of wave propagation when a plane wave of given $\mathbf{u}=k_j\boldsymbol{\Omega}^{(j)}$ is observed in the medium of k_i.

Thus, from Eqs. (A5) and (A9),

$$
\begin{aligned}
&|2^{-1}\eta_i T_{ij}(\hat{\Omega})|^2\,\delta(k_i\boldsymbol{\Omega}-k_j\boldsymbol{\Omega}') \\
&\quad = \Omega_z^{(i)}|\langle R_{ij}(\hat{\Omega})\rangle|^2\,\delta(\hat{\Omega}^{(i)}-\hat{\Omega}^{(i)\prime}), \qquad \text{(A11)}
\end{aligned}
$$

which following substitution into Eq. (A3) yields Eq. (6.16).

APPENDIX B: ELECTROMAGNETIC WAVES IN GENERAL CASE

Even when a depolarization is possible in the scattering, basic equations can still be given in the same form with a redefinition of the notations, as follows. Let $\psi_j^{(h)}(z)$ and $\psi_j^{(v)}(z)$, $j=1,2$, be wave functions in space Σ_j, representing the waves of horizontal and vertical polarization, respectively, and also the $\epsilon_j^{(h)}=\mu_j$, whereas $\epsilon_j^{(v)}=\epsilon_j$. Then, the boundary condition of Eq. (2.10) is replaced with

$$
\begin{aligned}
-[\epsilon_1^{(a)}]^{-1}\,\partial_n\psi_1^{(a)} &= \sum_b [B_{11}^{(ab)}\psi_1^{(b)}+B_{12}^{(ab)}\psi_2^{(b)}], \\
-[\epsilon_2^{(a)}]^{-1}\,\partial_n\psi_2^{(a)} &= \sum_b [B_{21}^{(ab)}\psi_1^{(b)}+B_{22}^{(ab)}\psi_2^{(b)}],
\end{aligned} \qquad \text{(B1)}
$$

in terms of surface impedances $B_{ij}^{(ab)}$. Nevertheless, Eq. (B1) can be written, on suppressing the superscripts, in the form of Eq. (2.10), with the understanding that each B_{ij} now represents a 2×2 matrix with respect to the superscripts, having the matrix elements $B_{ij}^{(ab)}$. That is, with this redefinition, the equation of $\mathbf{g}$ is given by Eq. (2.21), and $\mathbf{R}$ is by Eq. (2.24), etc., by exactly the same equations as those for the scalar waves in Sec. II.

Here, the $\psi_j^{(a)}$'s are orthogonal in the sense that the total vertical (outward normally directed) power component integrated over S_j, say W_j, is given by

$$
W_j = \sum_a \int_{S_j} d\boldsymbol{\rho}\,[2i\,\epsilon_j^{(a)}]^{-1}[\psi_j^{(a)*}(\overleftarrow{\partial}_n-\vec{\partial}_n)\psi_j^{(a)}], \qquad \text{(B2)}
$$

without any cross terms by the components of different superscript. Hence, when the medium is not dissipative, the condition that W_1+W_2 be zero imposes the Hermitian condition (2.15), i.e.,

$$
B_{ij}^{(ab)*}(\boldsymbol{\rho}|\boldsymbol{\rho}') = B_{ji}^{(ba)}(\boldsymbol{\rho}'|\boldsymbol{\rho}), \qquad \text{(B3)}
$$

with respect to the subscripts, the superscripts, and the $\boldsymbol{\rho}$ coordinates. The same is also true of the relation (2.29) which is presently written as

$$
g_{ij}^{(ab)}(\boldsymbol{\rho}'|\boldsymbol{\rho}'') = g_{ji}^{(ba)}(\boldsymbol{\rho}''|\boldsymbol{\rho}'). \qquad \text{(B4)}
$$

The situation is the same also of the space Green's function continued into the outside spaces, and further of the statistical Green's functions; it holds the same equations as in Secs. IV and V, upon suppressing the superscripts. Hence, for example, the second-order space Green's function defined by

$$
\begin{aligned}
&G_{ij;kl}^{(ab;cd)}(\hat{x}_1;\hat{x}_2|\hat{x}_1';\hat{x}_2') \\
&\quad = \langle\, g_{ik}^{(ac)*}(\hat{x}_1|\hat{x}_1')g_{jl}^{(bd)}(\hat{x}_2|\hat{x}_2')\rangle \qquad \text{(B5)}
\end{aligned}
$$

satisfies the BS equation of the same form as Eq. (4.10), and can be given in terms of S(1;2) by Eq. (4.17). The asymptotic expression of Eq. (6.14) becomes

$$
\begin{aligned}
&G_{ii;jj}^{(ab;cd)}(\hat{x};\hat{x}|\hat{x}';\hat{x}') \\
&\quad = G_{ij}^{(ac)*}(\hat{x}|\hat{x}')G_{ij}^{(bd)}(\hat{x}|\hat{x}') + \int d\boldsymbol{\rho}''\,|\hat{x}-\boldsymbol{\rho}''|^{-2} \\
&\qquad \times\epsilon_i^{(a)}\epsilon_i^{(b)}\sigma_{ii;jj}^{(I,ab;cd)}(\hat{\Omega}\,|\hat{\Omega}')\epsilon_j^{(c)}\epsilon_j^{(d)}|4\pi(\boldsymbol{\rho}''-\hat{x}')|^{-2}, \qquad \text{(B6)}
\end{aligned}
$$

and the same is also true of the optical relations.

Thus, most of the statistical equations in Secs. IV–VI are reproduced with the understanding that each component η_j now represents a diagonal 2×2 matrix in a two-dimensional polarization space, having the elements $\epsilon_j^{(a)}\,\delta_{ab}$, and each element g_{ij} represents a 2×2 matrix operator in the same space, having the elements $g_{ij}^{(ab)}$. Hence, it becomes the same also of b_{ij} and other associated statistical quantities, e.g., M_{ij}, γ_{ij}, A_{ij}, $G_{ij;kl}(1;2)$, etc. The principle of deriving $B_{ij}^{(ab)}$ also remains unchanged, although the actual procedure becomes more involved, and we will not go further on this point in this paper.

APPENDIX C: BS EQUATION, SCATTERING CROSS SECTIONS, AND RELATED OPTICAL RELATIONS FOR THE SLIGHTLY RANDOM BOUNDARY

Since the basic equation (3.47a) differs from Eq. (4.1) only by the replacement of $i\mathbf{h}\boldsymbol{\eta}^{-1} - \mathbf{B}_0$ and 1 to $\boldsymbol{\pi}$ and $\vec{1}$, respectively, the procedure of obtaining the statistical Green's functions remains almost unchanged. The effective impedance $\mathbf{M}$ for the coherent wave, $\vec{G} = \langle \vec{g} \rangle$, is defined according to

$$\langle \mathbf{b}\vec{g} \rangle = \mathbf{M}\vec{G}, \quad \langle \mathbf{b} \rangle = 0, \tag{C1}$$

to write the averaged version of Eq. (3.47a) by

$$(\boldsymbol{\pi} - \mathbf{M})\vec{G} = \vec{1}. \tag{C2}$$

Hence,

$$\vec{G} = \mathbf{G}\vec{1}, \tag{C3}$$

where $\mathbf{G}(\neq \mathbf{G}^T)$ is a 2×2 first-order Green's function, defined by

$$(\boldsymbol{\pi} - \mathbf{M})\mathbf{G} = 1, \quad \mathbf{G} = (\boldsymbol{\pi} - \mathbf{M})^{-1}. \tag{C4}$$

Here, to find a specific expression of $\mathbf{M}$, we introduce the difference $\Delta\mathbf{b}$ defined by

$$\Delta\mathbf{b} = \mathbf{b} - \mathbf{M}, \quad \langle \Delta\mathbf{b}\vec{g} \rangle = 0, \tag{C5}$$

and rewrite Eq. (3.47a) in terms of $\mathbf{G}$ and $\Delta\mathbf{b}$ as

$$\vec{g} = \mathbf{G}[\vec{1} + \Delta\mathbf{b}\vec{g}], \tag{C6}$$

which, by the substitution into the left-hand side of Eq. (C1), shows that, to the second order of ζ,

$$\langle \mathbf{b}\mathbf{G}\Delta\mathbf{b}\vec{g} \rangle = \langle \mathbf{b}\mathbf{G}\mathbf{b} \rangle \langle \vec{g} \rangle, \tag{C7}$$

in consequence of the condition in Eq. (C5) (and Gaussian statistics). Hence,

$$\mathbf{M} = \langle \mathbf{b}\mathbf{G}\mathbf{b} \rangle. \tag{C8}$$

To obtain the second-order Green's function $\mathbf{G}(1;2)$ defined by

$$\langle \vec{g}^*(1)\vec{g}(2) \rangle = \mathbf{G}(1;2)\vec{1}(1)\vec{1}(2), \tag{C9}$$

so that

$$\langle g_{i1}^*(1)\, g_{j1}(2) \rangle = G_{ij;11}(1;2), \tag{C10}$$

we utilize Eq. (C6) for both $\vec{g}^*(1)$ and $\vec{g}(2)$ in Eq. (C9). Hence, an equation of $\mathbf{G}(1;2)$ is found again in the form of the BS equation

$$\mathbf{G}(1;2) = \mathbf{G}^*(1)\mathbf{G}(2)[1 + \mathbf{K}(1;2)\mathbf{G}(1;2)], \tag{C11}$$

similar to Eqs. (4.10) and (4.11), with

$$\mathbf{K}(1;2) = \langle \mathbf{b}^*(1)\mathbf{b}(2) \rangle. \tag{C12}$$

Hence, $\mathbf{M}$ and $\mathbf{K}$ are given formally by the same equation as (4.12).

More precisely, since $\mathbf{b}$ is a diagonal operator [Eq. (3.45)], the nonvanishing elements of $\mathbf{K}(1;2)$ are

$$\begin{pmatrix} K_{11;11} & K_{12;12} \\ K_{21;21} & K_{22;22} \end{pmatrix} = \begin{pmatrix} \langle b_1^*(1)b_1(2) \rangle & \langle b_1^*(1)b_2(2) \rangle \\ \langle b_2^*(1)b_1(2) \rangle & \langle b_2^*(1)b_2(2) \rangle \end{pmatrix}. \tag{C13}$$

To the approximation of $\mathbf{G} \simeq \boldsymbol{\pi}^{-1}$ [Eq. (C4)], using Eq. (3.48b) shows that

$$\mathbf{M} = \begin{pmatrix} \langle b_1 g_0 b_1 \rangle & -\langle b_1 g_0 b_2'' \rangle \\ \langle b_2 g_0 b_1 \rangle & \langle b_2 g_0 b_2' \rangle \end{pmatrix}, \tag{C14}$$

where b_2'' is defined by Eq. (3.40b) and

$$b_2' = (\eta_1 - \eta_2)\eta_1^{-1}\eta_2^{-1}h_1\zeta h_2 = (h_1\eta_2/h_2^\dagger\eta_1)b_2, \tag{C15}$$

with the relation [cf. Eq. (3.41b)]

$$b^{(21)} = b_1 + b_2'. \tag{C16}$$

The solution of the BS equation (C11) also can be given by the same equation as Eq. (4.17) in terms of the incoherent scattering matrix $\mathbf{S}(1;2)$ [defined by Eq. (4.18) and given by Eqs. (4.19) and (4.20)], and further by Eq. (5.11) in terms of $\boldsymbol{\sigma}^{(I)}(1;2)$ defined by Eq. (5.10), after the continuation into the outside spaces. Here, from Eq. (5.4c), $\mathbf{A}$ and $\overline{\mathbf{A}}(\neq \mathbf{A}^T)$ are given, to the approximation of $\mathbf{G} \simeq \boldsymbol{\pi}^{-1}$, by

$$\mathbf{A} = ig_0 \begin{pmatrix} h_1\eta_1^{-1} & -(h_2/h_2^\dagger)h_2\eta_2^{-1} \\ h_1\eta_1^{-1} & (h_2/h_2^\dagger)h_1\eta_1^{-1} \end{pmatrix}, \tag{C17}$$

$$\overline{\mathbf{A}} = ig_0 \begin{pmatrix} h_1\eta_1^{-1} & -(h_2/h_2^\dagger)h_1\eta_1^{-1} \\ h_2\eta_2^{-1} & (h_2/h_2^\dagger)h_1\eta_1^{-1} \end{pmatrix}. \tag{C18}$$

To the approximation of $\mathbf{S}(1;2) \simeq \mathbf{K}(1;2)$, use of Eq. (C13) with b_2' and b_2'' of Eqs. (C15) and (3.40b) leads to the result

$$\sigma_{11;11}^{(I)}(1;2) = 2^4 \overline{A}_{11}^*(1)\overline{A}_{11}(2)\langle b^{(11)*}(1)b^{(11)}(2) \rangle \times A_{11}^*(1)A_{11}(2), \tag{C19}$$

$$\sigma_{22;11}^{(I)}(1;2) = 2^4 \overline{A}_{21}^*(1)\overline{A}_{21}(2)\langle b^{(21)*}(1)b^{(21)}(2) \rangle \times A_{11}^*(1)A_{11}(2), \tag{C20}$$

as may be expected directly from Eq. (3.34) and (3.41), respectively. Here, the $\hat{x}$-coordinate version of Eq. (C10) is given by Eq. (6.14) in terms of the incoherent scattering cross section $\sigma_{ii;11}^{(I)}(\hat{\Omega}\,|\,\hat{\Omega}')$ given by Eq. (6.15). Hence, the cross sections of Eq. (6.33) are obtained by using Eqs. (C17)–(C20).

The situation remains unchanged also for the optical relations resulting from the Hermitian condition of $\mathbf{b}$. Hence, Eq. (4.25) holds as it is, and the same is also true of Eq. (4.29) given in terms of $\mathbf{S}(1;2)$, in consequence of Eq. (3.48a). Thus, we obtain Eqs. (6.19)–(6.21) as they are, with the particular indices $c = d = 1$. Here, to the approximation of Eqs. (C17)–(C20), the optical relation of Eq. (6.22) can be shown to be reduced to

$$4\sum_{cd} \gamma_{cd}(\hat{\Omega}')|A_{11}(\hat{\Omega}')|^2 = \sum_{j=1}^{2} k_j\eta_j \int_{2\pi} d\hat{\Omega}\, \sigma_{jj;11}^{(I)}(\hat{\Omega}\,|\,\hat{\Omega}'), \tag{C21}$$

with $\gamma_{cd}(\hat{\Omega})$ given by Eq. (6.20) with (C14) and $\sigma_{jj;11}^{(I)}(\hat{\Omega}\,|\,\hat{\Omega}')$ of Eq. (6.33).

Here, it may be remarked that, in the special case of $\eta_2 = \infty$ (as realized by a perfectly conducting boundary for electromagnetic waves of vertical polarization), $\boldsymbol{\pi}$ has the only nonvanishing element $\pi_{11} = ih_1\eta_1^{-1}$, which also becomes very small for the grazing waves for which $h_1 \sim 0$. Hence, Eq. (C17) gives $A_{11} = 1$ in that case, whereas, as $h_1 \to 0$, the exact A_{11} tends according to Eq. (5.4c) with (C14) (where $b_2 = b_2'' = 0$), to

$$A_{11} = -iM_{11}^{-1} h_1 \eta_1^{-1} \to 0, \quad \eta_2 = \infty, \tag{C22}$$

showing that effect of the term M_{11} is not negligible for $A_{11}(\hat{\Omega})$ in the neighborhood of grazing-wave incidence and/or scattering.

[1]S. O. Rice, Commun. Pure Appl. Math. **4**, 351 (1951).

[2]P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces* (Pergamon, New York, 1963).

[3]D. E. Barrick and W. H. Peake, Radio Sci. **3**, 865 (1968).

[4]W. H. Peake, D. E. Barrick, A. K. Fung, and H. L. Chan, IEEE Trans. Antennas Propag. **AP-18**, 716 (1970).

[5]J. W. Wright, IEEE Trans. Antennas Propag. **AP-14**, 749 (1966).

[6]F. G. Bass and I. M. Fuks, *Wave Scattering from Statistically Rough Surface, International Series in Natural Philosophy*, Vol. 93, translated and edited by C. B. Vesecky and J. F. Vesecky (Pergamon, New York, 1979).

[7]V. Twersky, J. Acoust. Soc. Am. **29**, 209 (1957).

[8]V. Twersky, J. Acoust. Soc. Am. **73**, 85 (1983).

[9]J. A. DeSanto, J. Math. Phys. **15**, 283 (1974).

[10]F. Toigo, A. Marvin, V. Celli, and N. R. Hill, Phys. Rev. B **15**, 5618 (1977).

[11]A. M. Marvin, Phys. Rev. B **22**, 5759 (1980).

[12]J. Shen and A. A. Maradudin, Phys. Rev. B **22**, 4234 (1980).

[13]M. Nieto-Vesperinas and N. Garcia, Opt. Acta **12**, 1651 (1981).

[14]K. Furutsu, Proc. Inst. Electr. Eng., Part F **130**, 601 (1983).

[15]B. G. Smith, IEEE Trans. Antennas Propag. **AP-15**, 668 (1967).

[16]M. I. Sancer, IEEE Trans. Antennas Propag. **AP-17**, 577 (1969).

[17]G. S. Brown, IEEE Trans. Antennas Propag. **AP-26**, 472 (1978).

[18]G. S. Brown, IEEE Trans. Antennas Propag. **AP-28**, 788 (1980).

[19]A. K. Fung, and H. J. Eom, IEEE Trans. Antennas Propag. **AP-29**, 463 (1981).

[20]A Fourier transform of Eq. (3.24a) is obtainable in a fashion similar to Eq. (3.14), and it has an analytical structure rather similar to those based on the extinction theorem.[10–13] In this connection, the author inclines to think that the present theory may have a mathematical basis equivalent to the latter, against his previous view.[14]

[21]With the expansion of Eq. (3.4), Eq. (3.3) is equivalent to the Taylor expansion of $\psi^{(1)}(-\zeta)$ with respect to $z = -\zeta(\boldsymbol{\rho})$ at each point $\boldsymbol{\rho}$. See also Ref. 14.

[22]K. Furutsu, J. Math. Phys. **21**, 2764 (1980).

[23]K. Furutsu, J. Opt. Soc. Am. **62**, 240 (1972).

[24]V. I. Tatarskii, *The Effects of the Turbulence Atmosphere on Wave Propagation*, reproduced by National Technical Information Service (1971). [English translation available from the U. S. Dept. of Commerce, Springfield, VA 22151).

[25]K. Furutsu, Radio Sci. **10**, 29 (1975).

[26]This method has been shown[22] to be equivalent to the coherent potential approximation, which has been successfully utilized in solid physics.

[27]In Eq. (5.12b), the interference terms are negligible since there is no such region where incident and reflected waves are both in the same direction, in contrast with the case of an isolated scatterer embedded in a homogeneous random medium, where the interference terms are definitely important, yielding a scattering cross section of negative value in the shadow direction.[33]

[28]The **M** and **K** given by Eqs. (4.14) and (4.15) for randomly distributed bosses have exactly the same form as those given in Ref. 14 for a one-side boundary, where the resulting incoherent scattering cross section is given by Eq. (182) with Eqs. (234)–(236) [similar to the present (6.15)]. The result differs from those given in Refs. 7 and 8 by the factor $|A(\hat{\Omega})|^2$, which is lacking in the latter.

[29]The physical dimension of $\gamma_{cd}(\hat{\Omega})$ defined by Eq. (6.20b) differs from $\gamma(\hat{\Omega})$ in Ref. 14 [Eq. (183)], by the factor k_0.

[30]The dependence of the cross section on $\partial^2\zeta/\partial\boldsymbol{\rho}^2$ was discussed in some detail,[14] showing that it becomes independent of $\partial^2\zeta/\partial\boldsymbol{\rho}^2$ when $|\epsilon| \gg 1$ in the case of vertical polarization, whereas, for the horizontal polarization, the same is true over the entire range of ϵ.

[31]In the grazing-wave direction $\Omega_z = 0$, no scattered wave can survive because of the dissipation by the multiple scattering.

[32]See, for example, Yu. N. Barabanenkov, Yu. A. Kravtsov, S. M. Rytov, and V. I. Tatarskii, Sov. Phys. Usp. **13**, 551 (1971).

[33]K. Furutsu, J. Opt. Soc. Am. A **2**, to be published (June, 1985).

Hamiltonian approach to the existence of magnetic surfaces

M. P. Bernardin
Department of Physics, University of Wisconsin, Madison, Wisconsin 53706
J. A. Tataronis
Department of Electrical Engineering, University of Wisconsin, Madison, Wisconsin 53706

(Received 20 September 1984; accepted for publication 26 April 1985).

A method is devised to investigate the existence of magnetic surfaces and magnetohydrodynamic (MHD) plasma equilibria in 3-D toroidal geometry. The key feature of this method is the utilization of a Hamiltonian formulation of the lines of force. Expanding the contravariant components of the magnetic field and scalar pressure in distance ρ from the magnetic axis, the 1-D Hamiltonian for the lines of force is written out explicitly. The Hamiltonian is then transformed to action-angle variables. It is shown that the action J corresponds to pressure in the equilibrium problem. Specifically, it is shown that if J is an invariant, then constant pressure and hence magnetic surfaces exist. A procedure of repeated canonical transformations is formulated and carried out to displace the coordinate dependence in the Hamiltonian to successively higher order in the expansion parameter, and thus make J an increasingly better adiabatic invariant. Arising in each successive canonical transformation is a series of potentially resonant denominators, i.e., denominators that may vanish. These potential resonances are identified, their significance explicated, and methods of handling them suggested.

I. INTRODUCTION

Magnetic field lines of force in three-dimensional space may be characterized by one or more of the following properties: (i) the field lines are closed, (ii) the field lines ergodically cover a two-dimensional "flux surface," and (iii) the field lines wander ergodically throughout a volume. In the interest of confining fusion plasmas magnetically, it is necessary to know the conditions for the magnetic field lines to lie on a dense set of magnetic flux surfaces. The plasma is governed by the magnetohydrodynamic (MHD) equilibrium equations

$$\nabla\cdot\mathbf{B} = 0, \tag{1.1}$$

$$\nabla\times\mathbf{B} = \mathbf{J}, \tag{1.2}$$

$$\mathbf{J}\times\mathbf{B} = \nabla P, \tag{1.3}$$

where $\mathbf{B}$ is the magnetic field, $\mathbf{J}$ is the current density, and P is the hydrostatic pressure. Eliminating the current density using Eq. (1.2), the equilibrium equations reduce to

$$\nabla\cdot\mathbf{B} = 0, \tag{1.4}$$

$$(\nabla\times\mathbf{B})\times\mathbf{B} = \nabla P. \tag{1.5}$$

Equations (1.4) and (1.5) are a set of four coupled, nonlinear, partial differential equations, with two real and two imaginary characteristics.[1] If a single-valued solution exists to this set of equations and if the pressure gradient vanishes almost nowhere, then the magnetic field lines lie on a dense set of magnetic flux surfaces.[2] (The term magnetic flux surfaces is synonymous with magnetic surfaces, flux surfaces, and constant pressure surfaces; all four expressions are used interchangeably.) It is well known that single-valued solutions do exist in configurations which possess either translational, rotational, or helical symmetry. However, in configurations which lack symmetry, there is evidence, based largely on the KAM theory of Hamiltonian systems, that periodic solutions generally do not exist, and therefore, that these configurations generally do not have a continuous distribution of flux surfaces.[3-5]

We propose a method of investigation into the existence of both exact and approximate solutions to these equations and, thus, the existence of exact and approximate surfaces in asymmetric toroidal geometry. The terminology "approximate solutions" is given a precise mathematical definition when we seek interpretation of our results in Sec. VI. Features of our proposed method of investigation include the following. First, we display an explicit representation of a Hamiltonian formulation of the lines of force for a magnetic field. Second, we formulate and apply a method of repeated canonical transformations to the Hamiltonian as a means of establishing exact or approximate flux surfaces. Third, we show explicitly the appearance of resonances in the magnetic field structure of 3-D toroidal plasma configurations and we interpret these resonances. The goal of this research is to discern criteria such that these magnetic fields possess minimal regions of island structure and stochasticity.

In this method of investigation, we utilize the analytic technique of expanding the quantities $\mathbf{B}$ and P in distance ρ from an arbitrary magnetic axis. The magnetic axis is a degenerate surface of zero enclosed volume. This technique was pioneered by Mercier,[6] and applied by Lortz and Nuhrenberg,[7] to study 3-D MHD equilibria of plasmas. These previous studies have inherent limitations in that (i) only a restricted domain of solution space is considered, (ii) the solutions that are generated are in terms of parameters over which there is little or no experimental control, and (iii) a prescription to calculate the magnetic surfaces to an arbitrarily high order is lacking. These limitations may be overcome by casting the problem into a Hamiltonian formulation. Since by Eq. (1.1), the magnetic field is divergence-free, a one-dimensional nonautonomous Hamiltonian can be constructed for the lines of force. By making the proper canonical transformation, the Hamiltonian becomes that of a nonlinear oscillator. The problem can now be addressed using the Lewis-Leach-Symon Hamiltonian treatment of the time-dependent oscillator.[8-13] Following Symon,[11] the Hamilton-

 0022-2488/85/092370-11$02.50

ian is first transformed to action-angle variables. Then a systematic procedure of repeated canonical transformations, similar to a generalized Poincaré–Von Zeipel scheme, is formulated and carried out. In the nth transformation, an F_2 generating function is constructed so as to eliminate the coordinate dependence in the Hamiltonian to order $(n-1)$. That is, the coordinate dependence in the Hamiltonian is displaced to higher and higher order in the expansion. From Hamilton's equations, the action J becomes an increasingly better adiabatic invariant with each canonical transformation. It is shown that J corresponds to pressure in the equilibrium problem, and, thus, that constant J surfaces are tantamount to constant pressure (and hence magnetic) surfaces in physical space.

Arising in each successive generating function is a series of potentially resonant denominators, i.e., denominators which may vanish. Vanishing denominators are inherent in the method of expansion in powers of ρ. Mathematically, their presence may preclude the convergence of the expansion series—this is the motivation for seeking approximate (series) solutions although the existence of exact solutions may still be considered. Topologically, their presence may correspond to the exhibition of new magnetic axes that encircle the general system of toroidal surfaces.[14] In order to design a magnetic confinement scheme possessing a large volume of nested surfaces, it is necessary to first identify which denominators are indeed resonant, and, second, to develop techniques to handle these resonances.

Before proceeding with the details of this calculation, we sketch a brief outline of the remainder of the paper. In Sec. II, the series expansions for the magnetic field components and the pressure are given. These expressions are formally substituted into the MHD equilibrium equations, Eqs. (1.4) and (1.5), and the lowest-order expansion coefficients are solved for explicitly. In addition, the Hamiltonian formulation of the field line equations is presented. In Sec. III, the lowest-order invariant is constructed. Furthermore, the significance of the invariant to the equilibrium problem is explained. In Sec. IV, the successive canonical transformation scheme is formulated. The scheme is carried out through fourth order (in ρ), then generalized to nth order. In Sec. V, the resonant denominators are identified and discussed. Finally, in Sec. VI, we summarize the results of the analysis, discuss the meaning of the solutions, and present some conclusions.

II. THE HAMILTONIAN FORMULATION OF THE FIELD LINE EQUATIONS

A. The power series expansions and equilibrium equations

The plasma confinement configuration under consideration is a three-dimensional torus without symmetry. The magnetic axis or degenerate surface is assumed to close upon itself in one toroidal revolution, but otherwise to have arbitrary curvature and torsion. The plasma is governed by the MHD equilibrium equations, Eqs. (1.4) and (1.5), subject to periodic boundary conditions. We work in Mercier coordinates, (ρ,θ,s) (see Refs. 6 and 7), where s is the distance measured along the magnetic axis, while ρ, θ are polar coordinates in an $s=\text{const}$ plane. The angle θ is measured from the normal to the magnetic axis. The contravariant components of the magnetic field and the pressure are now expanded in distance ρ from the magnetic axis:

$$B^{\rho}=\mathbf{B}\cdot\nabla\rho=a_1\rho+a_2\rho^2+a_3\rho^3+\cdots, \tag{2.1}$$

$$B^{\theta}=\mathbf{B}\cdot\nabla\theta=b_0+b_1\rho+b_2\rho^2+\cdots, \tag{2.2}$$

$$B^{s}=\mathbf{B}\cdot\nabla s=c_0+c_1\rho+c_2\rho^2+\cdots, \tag{2.3}$$

$$P=\eta_0+\eta_2\rho^2+\eta_3\rho^3+\cdots. \tag{2.4}$$

The coefficients in Eqs. (2.1)–(2.4) are, in general, periodic functions of θ and s. Note that $a_0=0$, $c_0=c_0(s)$, and $\eta_1=0$ because of the respective boundary conditions at the magnetic axis: the radial magnetic field is zero, the toroidal field is a function only of s, and the pressure is an extremum. Substituting the expressions in Eqs. (2.1)–(2.4) into Eqs. (1.4) and (1.5), the expansion coefficients are solved for order by order. The details of this part of the analysis are straightforward but quite lengthy to derive. To avoid unnecessary detail, we reserve this part of the calculation for another paper,[15] and display only those results necessary to meet the stated objectives of this paper. The solution of the lowest-order coefficients is

$$a_1(\theta,s)=-\tfrac{1}{2}c_0'(s)+b_{02c}\sin 2\theta-b_{02s}\cos 2\theta, \tag{2.5}$$

$$b_0(\theta,s)=\tfrac{1}{2}(j-2\tau c_0)+b_{02c}\cos 2\theta+b_{02s}\sin 2\theta, \tag{2.6}$$

$$c_0=c_0(s), \tag{2.7}$$

$$\eta_0=\text{const}. \tag{2.8}$$

In Eqs. (2.5) and (2.6), j is the current density on axis, τ is the torsion of the axis, and b_{02c}, b_{02s} are arbitrary periodic functions of s. The second-order pressure coefficient η_2 is governed by the following first-order, linear, partial differential equation:

$$b_0\frac{\partial\eta_2}{\partial\theta}+c_0\frac{\partial\eta_2}{\partial s}+2a_1\eta_2=0. \tag{2.9}$$

Using Fourier analysis, it can be shown that the general solution of Eq. (2.9) satisfying periodic boundary conditions is

$$\eta_2(\theta,s)=\eta_{20}(s)+\eta_{2c}(s)\cos 2\theta+\eta_{2s}(s)\sin 2\theta, \tag{2.10}$$

where the coefficients η_{20}, η_{2c}, and η_{2s} are governed by

$$\frac{d}{ds}\left(\frac{\eta_{20}}{c_0}\right)=2\frac{b_{02s}}{c_0}\frac{\eta_{2c}}{c_0}-2\frac{b_{02c}}{c_0}\frac{\eta_{2s}}{c_0}, \tag{2.11}$$

$$\frac{d}{ds}\left(\frac{\eta_{2c}}{c_0}\right)=2\frac{b_{02s}}{c_0}\frac{\eta_{20}}{c_0}+\left(2\tau-\frac{j}{c_0}\right)\frac{\eta_{2s}}{c_0}, \tag{2.12}$$

$$\frac{d}{ds}\left(\frac{\eta_{2s}}{c_0}\right)=-2\frac{b_{02c}}{c_0}\frac{\eta_{20}}{c_0}-\left(2\tau-\frac{j}{c_0}\right)\frac{\eta_{2c}}{c_0}. \tag{2.13}$$

In Eqs. (2.11)–(2.13), the functions b_{02c}, b_{02s}, c_0, τ, and j are arbitrary and are not determined until all of the expansion coefficients have been solved for. Thus, Eqs. (2.11)–(2.13) should be solved as functionals of b_{02c}, b_{02s}, c_0, τ, and j. Similar, but increasingly larger sets of equations govern η_3, η_4, etc. The general solution for η_{20}, η_{2c}, and η_{2s} from Eqs. (2.11)–(2.13) is quite unperspicuous. However, by casting the problem into a Hamiltonian formulation for the lines of force, general periodic solutions are obtainable for these coefficients.

B. The field line equations in Hamiltonian form

The objective of this section is to construct an explicit Hamiltonian for the magnetic field lines of force. The Hamiltonian formulation is presented in generalized coordinates x^1, x^2, x^3. At the end of the calculation, the generalized coordinates are specified in terms of ρ, θ, s.

Any arbitrary magnetic field has the property of being divergence-free, and thus is expressible in terms of a vector potential through

$$\mathbf{B} = \nabla \times \mathbf{A}. \tag{2.14}$$

The contravariant components of the magnetic field are written as

$$B^1 = \frac{1}{\sqrt{g}}\left(\frac{\partial A_3}{\partial x^2} - \frac{\partial A_2}{\partial x^3}\right), \tag{2.15}$$

$$B^2 = \frac{1}{\sqrt{g}}\left(\frac{\partial A_1}{\partial x^3} - \frac{\partial A_3}{\partial x^1}\right), \tag{2.16}$$

$$B^3 = \frac{1}{\sqrt{g}}\left(\frac{\partial A_2}{\partial x^1} - \frac{\partial A_1}{\partial x^2}\right), \tag{2.17}$$

where g is the determinant of the metric tensor. We simplify these expressions by choosing a gauge such that $A_2 = 0$. Then, we introduce functions p and H through the following definitions:

$$p \equiv -A_1, \tag{2.18}$$

$$H \equiv A_3. \tag{2.19}$$

With these definitions, Eqs. (2.15) and (2.17) become

$$p = \int dx^2 \sqrt{g} B^3 + \gamma(x^1, x^3), \tag{2.20}$$

$$H = \int dx^2 \sqrt{g} B^1 + \delta(x^1 x^3). \tag{2.21}$$

The free functions γ and δ, arising in Eqs. (2.20) and (2.21), are constrained by Eq. (2.16), that is,

$$\sqrt{g} B^2 + \frac{\partial H}{\partial x^1} + \frac{\partial p}{\partial x^3} = 0. \tag{2.22}$$

To complete the Hamiltonian formulation of the magnetic lines of force, we introduce the canonical coordinate, and show that p in Eq. (2.20) and H in Eq. (2.21) are the conjugate momentum and the Hamiltonian, respectively. To this end, two more expedients are needed. First, we define a coordinate transformation from x^1, x^2, x^3 to new coordinates q, p, t via the following equations:

$$q = x^1, \tag{2.23a}$$

$$p = p(x^1, x^2, x^3), \tag{2.23b}$$

$$t = x^3. \tag{2.23c}$$

Second, the equations for the lines of force are introduced. They are

$$\frac{dx^1}{B^1} = \frac{dx^2}{B^2} = \frac{dx^3}{B^3}. \tag{2.24}$$

Now, using Eqs. (2.23) and (2.24), we complete the calculation by showing that Hamilton's equations of motion are satisfied.

From Eq. (2.23),

$$\frac{dp}{dt} = \frac{\partial p}{\partial x^1}\frac{dx^1}{dt} + \frac{\partial p}{\partial x^2}\frac{dx^2}{dt} + \frac{\partial p}{\partial x^3}\frac{dx^3}{dt}. \tag{2.25}$$

We now eliminate $\partial p/\partial x^2$, $\partial p/\partial x^3$, dx^2/dt, and dx^3/dt from Eq. (2.25), using Eq. (2.20), (2.22), (2.24), and (2.23), respectively. The result is

$$\frac{dp}{dt} = \frac{\partial p}{\partial x^1}\frac{dx^1}{dt} - \frac{\partial H}{\partial x^1}. \tag{2.26}$$

Next, regard the Hamiltonian as a function of q, p, t, i.e., $H = H(q,p,t)$. Then,

$$\frac{\partial H}{\partial x^2} = \frac{\partial H}{\partial q}\frac{\partial q}{\partial x^2} + \frac{\partial H}{\partial p}\frac{\partial p}{\partial x^2} + \frac{\partial H}{\partial t}\frac{\partial t}{\partial x^2},$$

which reduces, by Eq. (2.23), to

$$\frac{\partial H}{\partial x^2} = \frac{\partial H}{\partial p}\frac{\partial p}{\partial x^2}. \tag{2.27}$$

Substituting into the left-hand side of Eq. (2.27) from Eq. (2.21), and into the right-hand side from Eq. (2.20), we see that

$$\sqrt{g} B^1 = \frac{\partial H}{\partial p}\sqrt{g} B^3. \tag{2.28}$$

The B^1 and B^3 are eliminated with Eq. (2.24). Identifying x^1 and x^3 with q and t, respectively, from Eq. (2.23), the first of Hamilton's equations is obtained as

$$\frac{dq}{dt} = \frac{\partial H}{\partial p}. \tag{2.29}$$

The second of Hamilton's equations will now be derived. Again, regarding $H = H(q,p,t)$, it is evident that

$$\frac{\partial H}{\partial x^1} = \frac{\partial H}{\partial q}\frac{\partial q}{\partial x^1} + \frac{\partial H}{\partial p}\frac{\partial p}{\partial x^1} + \frac{\partial H}{\partial t}\frac{\partial t}{\partial x^1},$$

which reduces, by Eq. (2.23), to

$$\frac{\partial H}{\partial x^1} = \frac{\partial H}{\partial q} + \frac{\partial H}{\partial p}\frac{\partial p}{\partial x^1}. \tag{2.30}$$

Substituting Eq. (2.30) into Eq. (2.26), and using Eq. (2.29), we have the second of Hamilton's equations,

$$\frac{dp}{dt} = -\frac{\partial H}{\partial q}. \tag{2.31}$$

In summary, the canonical coordinate, conjugate momentum, and Hamiltonian, for the magnetic field lines of force, are

$$q = x^1, \tag{2.32}$$

$$p = \int dx^2 \sqrt{g} B^3 + \gamma(x^1, x^3), \tag{2.33}$$

$$H = \int dx^2 \sqrt{g} B^1 + \delta(x^1, x^3), \tag{2.34}$$

where γ and δ are constrained by

$$\sqrt{g} B^2 + \frac{\partial H}{\partial x^1} + \frac{\partial p}{\partial x^3} = 0. \tag{2.35}$$

We now apply this Hamiltonian formulation to the expanded fields in Sec. II A. Choosing x^1, x^2, x^3 to be θ, ρ, s, Eqs. (2.32)–(2.35) reduce to

$$q=\theta, \tag{2.36}$$

$$p=\tfrac{1}{2}c_0\rho^2+\tfrac{1}{3}(c_1-c_0k\cos\theta)\rho^3+\cdots+(1/n)(c_{n-2}-c_{n-3}k\cos\theta)\rho^n+\cdots, \tag{2.37}$$

$$H=\tfrac{1}{2}b_0\rho^2+\tfrac{1}{3}(b_1-b_0k\cos\theta)\rho^3+\cdots+(1/n)(b_{n-2}-b_{n-3}k\cos\theta)\rho^n+\cdots. \tag{2.38}$$

To write the Hamiltonian as a function of q, p, t, it is necessary to invert Eq. (2.37) to write $\rho=\rho(q,p,t)$. Carrying this out, the Hamiltonian becomes

$$\begin{aligned}H(q,p,t)={}&\frac{b_0}{c_0}p+\frac{2\sqrt{2}}{3c_0^{3/2}}\left(b_1-\frac{c_1}{c_0}b_0\right)p^{3/2}\\&+\frac{1}{c_0^2}\left\{b_2-\tfrac{4}{3}\frac{c_1-\frac{1}{4}c_0k\cos q}{c_0}b_1\right.\\&+\left[\frac{4c_1(c_1-c_0k\cos q)}{3c_0^2}\right.\\&\left.\left.-\frac{c_2-c_1k\cos q}{c_0}\right]b_0\right\}p^2+\cdots.\end{aligned} \tag{2.39}$$

Hence, we have displayed an explicit representation of a Hamiltonian formulation of the lines of force. An exact invariant to the lowest-order Hamiltonian is now sought. It will be shown that this lowest-order invariant corresponds to the existence of approximate surfaces in the immediate vicinity of the magnetic axis.

III. CONSTRUCTION OF THE LOWEST-ORDER INVARIANT

The Hamiltonian in Eq. (2.39) can be transformed into a Hamiltonian descriptive of a nonlinear oscillator. The transformation, which takes the canonical coordinates (q,p,t) to new coordinates $(x,\bar{p},t)$, is accomplished through the following F_2 generating function:

$$F_2(q,\bar{p},t)=\tfrac{1}{2}\bar{p}^2\tan q. \tag{3.1}$$

The relationship between the old and new coordinates is established using Eq. (3.1):

$$p=\frac{\partial F_2}{\partial q}=\frac{1}{2}(\bar{p}\sec q)^2, \tag{3.2}$$

$$x=\frac{\partial F_2}{\partial \bar{p}}=\bar{p}\tan q. \tag{3.3}$$

With Eqs. (3.2) and (3.3), the new Hamiltonian can be written in the form

$$H(x,\bar{p},t)=H_2(x,\bar{p},t)+H_3(x,\bar{p},t)+H_4(x,\bar{p},t)+\cdots. \tag{3.4}$$

In Eq. (3.4), H_2 is a quadratic polynomial in x and $\bar{p}$, H_3 is a cubic polynomial, H_4 a quartic polynomial, etc. The explicit expressions for H_2, H_3, and H_4 are contained in Appendix A. In order to construct the lowest-order invariant, we write out the quadratic part of the Hamiltonian H_2 explicitly as

$$H_2(x,\bar{p},t)=\tfrac{1}{2}f(t)\bar{p}^2+g(t)\bar{p}x+\tfrac{1}{2}h(t)x^2. \tag{3.5}$$

The functions f, g, and h are obtained from Eqs. (A1) and (A2). This lowest-order Hamiltonian is precisely the lowest-order Hamiltonian treated by Symon.[11] He showed that this Hamiltonian (H_2) has the exact invariant

$$J=\frac{1}{2}\frac{x^2}{w^2}+\frac{1}{2}\left[w\bar{p}-(\dot{w}-gw)\frac{x}{f}\right]^2, \tag{3.6}$$

where w is any solution of

$$\ddot{w}-\frac{\dot{f}}{f}\dot{w}+\left(fh-g^2-\dot{g}+\frac{\dot{f}}{f}g\right)w=\frac{f^2}{w^3}. \tag{3.7}$$

For our purposes, w is chosen to be a periodic solution of Eq. (3.7) (see Ref. 16). To be complete, we express this exact invariant first in canonical coordinates q, p, t, and second, in Mercier coordinates ρ, θ, s. In the canonical coordinates, J is written as

$$\begin{aligned}J(q,p,t)=\Bigg[&\left(\frac{1}{w^2}+\frac{(\dot{w}-gw)^2}{f^2}\right)\sin^2 q\\&-2\frac{w(\dot{w}-gw)}{f}\cos q\sin q+w^2\cos^2 q\Bigg]p,\end{aligned}$$

while in Mercier coordinates, J to lowest order in ρ, is

$$\begin{aligned}J(\rho,\theta,s)=\frac{c_0}{4}\Bigg[&\left(w^2+\frac{1}{w^2}+\frac{(\dot{w}-gw)^2}{f^2}\right)\\&+\left(w^2-\frac{1}{w^2}-\frac{(\dot{w}-gw)^2}{f^2}\right)\cos 2\theta\\&-\frac{2w(\dot{w}-gw)}{f}\sin 2\theta\Bigg]\rho^2.\end{aligned} \tag{3.8}$$

It can be verified from Eq. (3.8) that if w is real, then $J=$ const surfaces are ellipses in physical space.

Before concluding this section, we elucidate the significance of the quantity J to the equilibrium problem. Specifically, we show that the invariance of J is correlated with the existence of constant pressure or, equivalently, magnetic surfaces.

The equation governing constant pressure surfaces is obtained by taking the scalar product of $\mathbf{B}$ with Eq. (1.5) as

$$\mathbf{B}\cdot\nabla P=0. \tag{3.9}$$

If a single-valued solution exists to this equation, then the magnetic field lines lie on a dense set of constant pressure surfaces. If J is an invariant for the field line Hamiltonian, then

$$\frac{dJ}{dt}=0. \tag{3.10}$$

In terms of physical space coordinates, the total time derivative d/dt is

$$\frac{d}{dt}=\frac{d\rho}{dt}\frac{\partial}{\partial\rho}+\frac{d\theta}{dt}\frac{\partial}{\partial\theta}+\frac{ds}{dt}\frac{\partial}{\partial s}. \tag{3.11}$$

From the field line equations, Eq. (2.24), we have

$$\frac{d\rho}{dt}=\frac{B^\rho}{B^s},\quad\frac{d\theta}{dt}=\frac{B^\theta}{B^s},\quad\frac{ds}{dt}=1.$$

Using these results in Eq. (3.11), it is evident that

$$\frac{d}{dt}=\frac{1}{B^s}\mathbf{B}\cdot\nabla. \tag{3.12}$$

Thus, from Eqs. (3.10) and (3.12), we see that the invariance of J, $dJ/dt=0$, implies that $\mathbf{B}\cdot\nabla J=0$. By comparison to Eq. (3.9), and noting that J is periodic in θ and s, it has been established that the invariant for the field line Hamiltonian corresponds to constant pressure surfaces in the equilibrium problem.

We note in passing that since $J(\rho, \theta, s)$ and $P(\rho, \theta, s)$ both satisfy the same equation ($\mathbf{B}\cdot\nabla J = \mathbf{B}\cdot\nabla P = 0$) the expansion coefficients of the pressure, η_i, can be obtained by setting $P = J$ and matching terms in equal powers of ρ. Hence, by comparing Eq. (2.4) to Eq. (3.8), we see that η_2 is given by

$$\eta_2(\theta, s) = \frac{c_0}{4}\Big[\Big(w^2 + \frac{1}{w^2} + \frac{(\dot{w} - gw)^2}{f^2}\Big) + \Big(w^2 - \frac{1}{w^2} - \frac{(\dot{w} - gw)^2}{f^2}\Big)\cos 2\theta - \frac{2w(\dot{w} - gw)}{f}\sin 2\theta\Big]. \tag{3.13}$$

Here in Sec. III, we have shown the existence of the invariant J, for the lowest-order field line Hamiltonian, and thus have demonstrated the existence of approximate surfaces located in the immediate vicinity of the magnetic axis. Mathematically speaking, we have established the existence of magnetic surfaces in the limit as $\rho \to 0$. As a first step toward our goal of establishing the existence of a large set of approximate flux surfaces, we seek to show that Eq. (3.9) is solvable order by order in the expansion. To accomplish this, the objective is to show that Eq. (3.10) is solvable order by order, and thus that J is an invariant to any desired order.

IV. INVARIANCE TO HIGHER ORDER

The purpose of this section is to construct an invariant of the expanded field line Hamiltonian, to any desired order in the expansion parameter. To this end, a method of repeated canonical transformations is formulated in order to displace the coordinate dependence in the Hamiltonian to higher and higher order. This procedure is similar to the Poincaré–Von Zeipel scheme, although this latter procedure is restricted to time-independent Hamiltonians.[17] The result of implementing the transformations scheme is that the momentum (in our case, it will be the action) becomes an increasingly better adiabatic invariant. The canonical transformation procedure is carried out explicitly through fourth order in ρ. A formal extension of the method to nth order is then made.

A. Invariance to third order in ρ

To find the invariant to third order in ρ, three successive canonical transformations are made. The purpose of the first two transformations is to produce the Hamiltonian in action-angle variables. The third transformation gives the third-order invariant.

We choose the first transformation so as to modify the quadratic portion of the expanded Hamiltonian from one corresponding to elliptic contours in phase space, to one corresponding to circular contours. In Appendix B, we show that the Hamiltonian

$$H_2(x, \bar{p}, t) = \tfrac{1}{2}f(t)\bar{p}^2 + g(t)\bar{p}x + \tfrac{1}{2}h(t)x^2$$

is transformed to

$$K_2(X, P, t) = (1/2\sigma^2)(X^2 + P^2)$$

by the following canonical transformation:

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 1/\sqrt{f}\,\sigma & 0 \\ (\sigma/2\sqrt{f})(2g - \dot{f}/f - 2\dot{\sigma}/\sigma) & \sqrt{f}\,\sigma \end{bmatrix}\begin{bmatrix} x \\ \bar{p} \end{bmatrix}. \tag{4.1}$$

In Eq. (4.1), σ is governed by

$$\ddot{\sigma} + \Big(fh - g^2 - \dot{g} + \frac{1}{2}\frac{\ddot{f}}{f} - \frac{3}{4}\frac{\dot{f}^2}{f^2} + \frac{\dot{f}}{f}g\Big)\sigma = \frac{1}{\sigma^3}. \tag{4.2}$$

Let Ω be defined to be

$$\Omega = 1/\sigma^2.$$

Then a generating function that yields this transformation is

$$F_2(x, P, t) = -\frac{1}{f}\Big(\frac{1}{2}g - \frac{\dot{f}}{f} + \frac{\dot{\Omega}}{\Omega}\Big)x^2 + \Big(\frac{\Omega}{f}\Big)^{1/2}xP.$$

Carrying out this canonical transformation, the Hamiltonian of Eq. (3.4) becomes

$$K(X, P, t) = K_2(X, P, t) + K_3(X, P, t) + K_4(X, P, t) + \cdots, \tag{4.3}$$

where K_2, the quadratic component of the new Hamiltonian, is

$$K_2(X, P, t) = \tfrac{1}{2}\Omega(t)(X^2 + P^2).$$

In Eq. (4.3), K_3 is a cubic polynomial in X and P, K_4 is a quartic polynomial, etc. We note in passing that Eq. (4.2) can be put in the form of Eq. (3.7) by setting $\sigma = f^{-1/2}w$; solutions of this equation are discussed in Ref. 16.

We now introduce a canonical transformation to action-angle variables through

$$X = (2J)^{1/2}\sin\gamma, \tag{4.4a}$$

$$P = (2J)^{1/2}\cos\gamma. \tag{4.4b}$$

The generating function for this transformation is

$$F_2(X, J) = (X/2)(2J - X^2)^{1/2} + J\sin^{-1}(X/\sqrt{2J}).$$

Noting that F_2 does not contain t explicitly, the new Hamiltonian is

$$K(\gamma, J, t) = K_2(J, t) + K_3(\gamma, J, t) + K_4(\gamma, J, t) + \cdots, \tag{4.5}$$

where

$$K_2(J, t) = \Omega(t)J,$$

$$K_3(\gamma, J, t) = [K_{33}(t)e^{3i\gamma} + K_{31}(t)e^{i\gamma} + \text{c.c.}](2J)^{3/2},$$

$$K_4(\gamma, J, t) = [K_{44}(t)e^{4i\gamma} + K_{42}(t)e^{2i\gamma} + K_{40}(t) + \text{c.c.}](2J)^2,$$

$$\vdots$$

$$K_n(\gamma, J, t) = \sum_l [K_{nl}(t)e^{il\gamma} + \text{c.c.}](2J)^{n/2}.$$

In the nth-order expression, the sum is over all integers l less than or equal to n and having the same parity (even or odd) as n. The c.c. stands for complex conjugate. The explicit expressions for K_{nl}, $n < 5$, are written out in Appendix A.

We restate the objective of this section. In view of the form of the Hamiltonian, Eq. (4.5), the objective is to eliminate the coordinate dependence γ in each successive order of the expanded Hamiltonian. This coordinate displacement is accomplished through suitably chosen successive canonical transformations. It follows from one of Hamilton's equations,

$$\frac{dJ}{dt} = -\frac{\partial K}{\partial \gamma},$$

that J then becomes an "invariant" to successively higher order.

To eliminate the coordinate dependence in $K_3(\gamma, J, t)$, we introduce a generating function which is the sum of two parts: the first part gives the identity transformation, while the second part contains the same harmonic content as K_3. Thus, the generating function W is written as

$$W(\gamma, \bar{J}, t) = \gamma\bar{J} + [W_3(t)e^{3i\gamma} + W_1(t)e^{i\gamma} + \text{c.c.}](2\bar{J})^{3/2}. \quad (4.6)$$

The relationship between the old and new coordinates is

$$J = \frac{\partial W}{\partial \gamma} = \bar{J} + [3iW_3 e^{3i\gamma} + iW_1 e^{i\gamma} + \text{c.c.}](2\bar{J})^{3/2}, \quad (4.7)$$

$$\bar{\gamma} = \frac{\partial W}{\partial \bar{J}} = \gamma + 3[W_3 e^{3i\gamma} + W_1 e^{i\gamma} + \text{c.c.}](2\bar{J})^{1/2}, \quad (4.8)$$

while the new Hamiltonian is found through

$$\bar{K}(\bar{\gamma}, \bar{J}, t) = K(\gamma, J, t)|_{\bar{\gamma}, \bar{J}} + \frac{\partial W(\gamma, \bar{J}, t)}{\partial t}\bigg|_{\bar{\gamma}, \bar{J}}. \quad (4.9)$$

The tedious part of this calculation is inverting Eq. (4.8) to obtain

$$\gamma = \gamma(\bar{\gamma}, \bar{J}, t). \quad (4.10)$$

This exercise is carried out in Appendix C. Substituting from Eqs. (4.7) and (4.10) into Eq. (4.9), the new Hamiltonian becomes

$$\bar{K}(\bar{\gamma}, \bar{J}, t) = \Omega\bar{J} + 2(\dot{\sigma}_1 + \Omega\sigma_2 + \beta_1)(2\bar{J})^{3/2} + \cdots, \quad (4.11)$$

where σ_k and β_k are defined by

$$\sigma_k = \text{Re}[(3i)^{k-1}W_3 e^{3i\bar{\gamma}} + i^{k-1}W_1 e^{i\bar{\gamma}}], \quad (4.12)$$

$$\beta_k = \text{Re}[(3i)^{k-1}K_{33}e^{3i\bar{\gamma}} + i^{k-1}K_{31}e^{i\bar{\gamma}}]. \quad (4.13)$$

In the above expressions, Re denotes the real part of the bracketed quantity.

Now, in order to eliminate the coordinate in the Hamiltonian at the order $\bar{J}^{3/2}$, the as yet unspecified functions W_1 and W_3 in the generating function must be chosen appropriately. In effect, we average the $\bar{J}^{3/2}$ term over the coordinate $\bar{\gamma}$. The coefficient of the $\bar{J}^{3/2}$ term is

$$\bar{K}_3(\bar{\gamma}, t) = 2\,\text{Re}\{[\dot{W}_3(t) + 3i\Omega(t)W_3(t) + K_{33}(t)]e^{3i\bar{\gamma}} + [\dot{W}_1(t) + i\Omega(t)W_1(t) + K_{31}(t)]e^{i\bar{\gamma}}\}. \quad (4.14)$$

It is obvious from Eq. (4.14) that the average of $\bar{K}_3(\bar{\gamma}, t)$ over $\bar{\gamma}$ is zero. We choose $W_3(t)$ and $W_1(t)$ such that

$$\dot{W}_3(t) + 3i\Omega(t)W_3(t) + K_{33}(t) = 0, \quad (4.15)$$

$$\dot{W}_1(t) + i\Omega(t)W_1(t) + K_{31}(t) = 0. \quad (4.16)$$

In Eqs. (4.15) and (4.16), Ω, K_{33}, and K_{31} are to be looked upon as known periodic functions of t. Equations (4.15) and (4.16) are special cases of the general equation

$$\dot{W}_l(t) + il\Omega(t)W_l(t) + K_{nl}(t) = 0. \quad (4.17)$$

Equation (4.17) arises in fourth- and higher-order calculations as well. Therefore, we develop periodic solutions for W_l and specialize the results to particular values of l as needed.

The first step is to decompose Ω into an average part plus a fluctuation part

$$\Omega = \bar{\Omega} + \tilde{\Omega}, \quad (4.18)$$

where $\bar{\Omega}$ is defined to be

$$\bar{\Omega} = \frac{1}{T}\int_0^T dt\, \Omega(t), \quad (4.19)$$

and $\tilde{\Omega}$ is the difference between Ω and $\bar{\Omega}$. In Eq. (4.19), T is the total arc length of the magnetic axis. We define $\hat{\Omega}$ to be

$$\hat{\Omega}(t) \equiv \int^t dt'\, \tilde{\Omega}(t').$$

With these definitions, Eq. (4.17) becomes

$$W_l(t) = \exp[-il(\bar{\Omega}t + \hat{\Omega})] \times \left\{C_{ll} - \int^t dt'\, K_{nl}(t')\exp[il(\bar{\Omega}t' + \hat{\Omega}(t'))]\right\}. \quad (4.20)$$

Here, C_{ll} is a constant of integration while K_{nl} and $e^{il\hat{\Omega}}$ are periodic functions of t with period T. Since the product of periodic functions is itself periodic, the product $K_{nl}e^{il\hat{\Omega}}$ may be expanded in a Fourier series. We have

$$K_{nl}(t)\exp[il\hat{\Omega}(t)] = \sum_m \lambda_{lm}\exp\left[i\frac{2\pi m}{T}t\right], \quad (4.21)$$

where λ_{lm} is

$$\lambda_{lm} = \frac{1}{T}\int_0^T dt\, K_{nl}(t)\exp[il\hat{\Omega}(t)]\exp\left[-i\frac{2\pi m}{T}t\right].$$

The summation in Eq. (4.21) is from negative to positive infinity. With Eq. (4.21), Eq. (4.20) becomes

$$W_l(t) = \exp[-il\hat{\Omega}(t)] \times \left\{C_{ll}e^{-il\bar{\Omega}t} - \sum_m \lambda_{lm}\frac{\exp[i(2\pi m/T)t]}{i(l\bar{\Omega} + 2\pi m/T)}\right\}. \quad (4.22)$$

We require that W_l be periodic, i.e., $W_l(t+T) = W_l(t)$. It is evident from Eq. (4.22) that periodicity requires that

$$C_{ll}(e^{il\bar{\Omega}T} - 1) = 0. \quad (4.23)$$

If $l\bar{\Omega}T \neq 2n\pi$ for all integers n, then $C_{ll} = 0$ and the periodic solution for W_l is

$$W_l(t) = i\exp[-il\hat{\Omega}(t)]\sum_m \lambda_{lm}\frac{\exp[i(2\pi m/T)t]}{l\bar{\Omega} + 2\pi m/T}. \quad (4.24a)$$

If $l\bar{\Omega}T = 2n\pi$ for some integer n, then a periodic solution for W_l exists if and only if λ_{l-n} is zero. The periodic solution is then

$$W_l(t) = i\exp[-il\hat{\Omega}(t)]{\sum_m}' \lambda_{lm}\frac{\exp[i(2\pi m/T)t]}{l\bar{\Omega} + 2\pi m/T}, \quad (4.24b)$$

where the primed summation excludes $m = -n$. For simplicity, C_{ll} has been set equal to zero in Eq. (4.24b).

The solution for the generating functions W_3 and W_1 are obtained from either Eq. (4.24a) or (4.24b) (depending on the value of the quantity $l\bar{\Omega}T$), with l set equal to 3 and 1, respectively. We note in passing that although Eq. (4.24a) is a valid solution for $l\bar{\Omega}T \neq 2n\pi$, there may exist terms in the summation that are arbitrarily large. This point will be examined more closely in Sec. V.

To complete the transformation from γ, J, t to $\bar{\gamma}, \bar{J}, t$, $\bar{K}(\bar{\gamma}, \bar{J}, t)$ is to be computed. With the solutions for W_3 and W_1 from Eq. (4.24), Eq. (4.11) becomes

$$\bar{K}(\bar{\gamma}, \bar{J}, t) = \Omega \bar{J} + 2(6\beta_1\sigma_2 + \chi_1 + K_{40})(2\bar{J})^2 + \cdots . \tag{4.25}$$

In Eq. (4.25), σ_2 and β_1 are obtained from Eqs. (4.12) and (4.13), respectively, while χ_1 is obtained from Eq. (4.26) by setting k equal to unity:

$$\chi_k = \mathrm{Re}[(2i)^{k-1}K_{44}e^{4i\bar{\gamma}} + i^{k-1}K_{42}e^{2i\bar{\gamma}}]. \tag{4.26}$$

It is apparent from Eq. (4.25) that

$$\frac{d\bar{J}}{dt} = 0 + \mathscr{O}(\bar{J}^2).$$

Since $\bar{J}$ is proportional to ρ^2, $\bar{J}$ is invariant through third order in ρ. $\bar{J}$ can be written in terms of γ, J, and t by inverting Eq. (4.7). The inversion, which is carried out in Appendix C, yields the result

$$\bar{J} = J - 2\bar{\sigma}_2(2J)^{3/2} + 12\bar{\sigma}_2^2(2J)^2 - 84\bar{\sigma}_2^3(2J)^{5/2} + \cdots, \tag{4.27}$$

where $\bar{\sigma}_2$ is defined to be

$$\bar{\sigma}_2 = \mathrm{Re}(3iW_3e^{3i\gamma} + iW_1e^{i\gamma}).$$

B. Invariance to fourth order in ρ

To identify the fourth-order invariant, a canonical transformation will be constructed in order to displace the coordinate dependence in the Hamiltonian to the next higher order in the expansion. The first step is to rewrite Eq. (4.25) so as to reveal the harmonic constituency in the coordinate. The result is

$$\bar{K}(\bar{\gamma}, \bar{J}, t) = \Omega\bar{J} + [\bar{K}_{46}e^{6i\bar{\gamma}} + \bar{K}_{44}e^{4i\bar{\gamma}} + \bar{K}_{42}e^{2i\bar{\gamma}} + \bar{K}_{40} + \text{c.c.}](2\bar{J})^2 + \mathscr{O}(\bar{J}^{5/2}), \tag{4.28}$$

where the functions $\bar{K}_{nl}(t)$ are written out explicitly in Appendix A. The canonical transformation to be constructed will take the old variables $\bar{\gamma}, \bar{J}, t$ into the new variables $\tilde{\gamma}, \tilde{J}, t$. The appropriate generating function should contain the second, fourth, and sixth harmonics in the coordinate to eliminate the coordinate dependence in the $\bar{J}^2$ term. Therefore, we write

$$\widetilde{W}(\bar{\gamma}, \tilde{J}, t) = \bar{\gamma}\tilde{J} + (W_6e^{6i\bar{\gamma}} + W_4e^{4i\bar{\gamma}} + W_2e^{2i\bar{\gamma}} + \text{c.c.})(2\tilde{J})^2. \tag{4.29}$$

The old variables are related to the new ones via the following equations:

$$\bar{J} = \frac{\partial\widetilde{W}}{\partial\bar{\gamma}} = \tilde{J} + (6iW_6e^{6i\bar{\gamma}} + 4iW_4e^{4i\bar{\gamma}} + 2iW_2e^{2i\bar{\gamma}} + \text{c.c.})(2\tilde{J})^2, \tag{4.30}$$

$$\tilde{\gamma} = \frac{\partial\widetilde{W}}{\partial\tilde{J}} = \bar{\gamma} + 4(W_6e^{6i\bar{\gamma}} + W_4e^{4i\bar{\gamma}} + W_2e^{2i\bar{\gamma}} + \text{c.c.})(2\tilde{J}). \tag{4.31}$$

The new Hamiltonian $\widetilde{K}$ is computed from

$$\widetilde{K}(\tilde{\gamma}, \tilde{J}, t) = \bar{K}(\bar{\gamma}, \bar{J}, t)|_{\bar{\gamma}, \bar{J}} + \left.\frac{\partial W(\bar{\gamma}, \tilde{J}, t)}{\partial t}\right|_{\bar{\gamma}, \tilde{J}}. \tag{4.32}$$

To evaluate the right-hand side of Eq. (4.32), $\bar{\gamma}$ and $\bar{J}$ must be written in terms of the new variables from Eqs. (4.30) and (4.31). The inversion procedure is similar to that used in Sec. IV A. Substituting these results into Eq. (4.32), we find

$$\begin{aligned}\widetilde{K}(\tilde{\gamma}, \tilde{J}, t) = \Omega\tilde{J} + 2\,\mathrm{Re}\{&(\dot{W}_6 + 6i\Omega W_6 + \bar{K}_{46})e^{6i\tilde{\gamma}} \\ &+ (\dot{W}_4 + 4i\Omega W_4 + \bar{K}_{44})e^{4i\tilde{\gamma}} \\ &+ (\dot{W}_2 + 2i\Omega W_2 + \bar{K}_{42}) \\ &\times e^{2i\tilde{\gamma}} + \bar{K}_{40}\}(2\tilde{J})^2 + \cdots .\end{aligned} \tag{4.33}$$

The coordinate dependence is eliminated from the $\tilde{J}^2$ term in Eq. (4.33) if

$$\dot{W}_6(t) + 6i\Omega(t)W_6(t) + \bar{K}_{46}(t) = 0, \tag{4.34}$$

$$\dot{W}_4(t) + 4i\Omega(t)W_4(t) + \bar{K}_{44}(t) = 0, \tag{4.35}$$

$$\dot{W}_2(t) + 2i\Omega(t)W_2(t) + \bar{K}_{42}(t) = 0. \tag{4.36}$$

The solution of Eqs. (4.34)–(4.36) are obtained from Eq. (4.24) with l set equal to 6, 4, and 2, respectively. To complete the transformation from $\bar{\gamma}, \bar{J}, t$ to $\tilde{\gamma}, \tilde{J}$, t, $\widetilde{K}(\tilde{\gamma}, \tilde{J}, t)$ is to be computed. Using Eq. (4.24), Eq. (4.33) becomes

$$\widetilde{K}(\tilde{\gamma}, \tilde{J}, t) = \Omega\tilde{J} + 2\bar{K}_{40}(t)(2\tilde{J})^2 + \mathscr{O}(\tilde{J}^{5/2}). \tag{4.37}$$

The action $\tilde{J}$ is a fourth-order invariant since

$$\frac{d\tilde{J}}{dt} = -\frac{\partial\widetilde{K}}{\partial\tilde{\gamma}} = 0 + \mathscr{O}(\tilde{J}^{5/2})$$

and $\tilde{J} \sim \rho^2$. Finally, $\tilde{J}$ is written in terms of $\bar{\gamma}$, $\bar{J}$, and t from Eq. (4.30) as

$$\tilde{J} = \bar{J} - 4\bar{\xi}_2(2\bar{J})^2 + 64\bar{\xi}_2^2(2\bar{J})^3 + \cdots, \tag{4.38}$$

where $\bar{\xi}_2$ is defined to be

$$\bar{\xi}_2 \equiv \mathrm{Re}[3iW_6e^{6i\bar{\gamma}} + 2iW_4e^{4i\bar{\gamma}} + iW_2e^{2i\bar{\gamma}}].$$

C. Generalization of invariance to nth order

The method for preceeding to the fifth and higher orders is now apparent. Let $K^{(n)}$ denote the nth new Hamiltonian, where $n \geqslant 1$. In this notation, $K^{(1)} = K(\gamma, J, t)$ as given in Eq. (4.5), $K^{(2)} = \bar{K}(\bar{\gamma}, \bar{J}, t)$ as given in Eq. (4.11), and $K^{(3)} = \widetilde{K}(\tilde{\gamma}, \tilde{J}, t)$ as given in Eq. (4.33). Let the corresponding canonical variables be denoted $J^{(n)}$ and $\gamma^{(n)}$. Now, suppose that we have succeeded in transforming away the γ dependence in all terms through the $n/2$ power of $J^{(n-1)}$. The Hamiltonian then has the form

$$\begin{aligned}K^{(n-1)} = \Omega J^{(n-1)} &+ \sum_{m=4} K_{mo}^{(n-1)}(t)[2J^{(n-1)}]^{m/2} \\ &+ \sum_l [K_{nl}^{(n-2)}(t)\exp(il\gamma^{(n-1)}) + \text{c.c.}] \\ &\times [2J^{(n-1)}]^{(n+1)/2} + \mathscr{O}[(J^{(n-1)})^{(n+2)/2}].\end{aligned} \tag{4.39}$$

The first sum in Eq. (4.39) is over all even integers m less than or equal to n, while the second sum is over all integers l less than or equal to n and having the same parity (even or odd) as n. A transformation to nth-order variables is made by means of the generating function

$$W = J^{(n)}\gamma^{(n-1)} + \sum_l [W_{nl}(t)\exp(il\gamma^{(n-1)}) + \text{c.c.}] \times (2J^{(n)})^{(n+1)/2}. \tag{4.40}$$

Solving for the new coordinates from Eq. (4.40), and substituting into Eq. (4.39), the new Hamiltonian becomes

$$\begin{aligned} K^{(n)} = \Omega J^{(n)} &+ \sum_{m=4} K^{(n)}_{mo}(t)(2J^{(n)})^{m/2} \\ &+ \sum_{l} [\dot{W}_{nl} + il\Omega W_{nl} \\ &+ K^{(n-1)}_{nl} + \text{c.c.}](2J^{(n)})^{(n+1)/2} \\ &+ \mathscr{O}[(J^{(n)})^{(n+2)/2}]. \end{aligned} \tag{4.41}$$

If l is not equal to zero, the corresponding term can be eliminated from the sum in Eq. (4.41) by requiring that W_{nl} satisfy the differential equation

$$\dot{W}_{nl} + il\Omega W_{nl} + K^{(n-1)}_{nl} = 0.$$

Elimination of the $l = 0$ terms would result in secular terms in the transformation that increase linearly with time.[11] Since l and n have the same parity, it is evident that the transformation given in Eq. (4.40) eliminates all terms of order n when n is odd, and eliminates all terms except that with $l = 0$ when n is even.

At this point, two primary objectives of this document have been met. An explicit representation of a Hamiltonian formulation of the lines of force has been displayed. In addition, a method of repeated canonical transformations to the Hamiltonian has been formulated and applied as a means of establishing approximate flux surfaces. Theoretically, the transformation procedure may be carried out to any desired order; however, in practice, its implementation may require computer assistance through an algebraic manipulator. To examine the existence of approximate flux surfaces, it is necessary to briefly discuss the generating functions and associated small denominators.

V. RESONANT DENOMINATORS

In the canonical transformations, each generating function contains potentially resonant denominators. The action is written in terms of these generating functions, and thus will also contain potentially resonant terms. In the generating function $W = W(\gamma, \bar{J}, t)$ written out in Eq. (4.6), resonances occur when the quantity $\bar{\Omega}T$ satisfies either of the following conditions:

$$\bar{\Omega}T = m2\pi, \tag{5.1a}$$

$$\bar{\Omega}T = m2\pi/3. \tag{5.1b}$$

In Eqs. (5.1a) and (5.1b), m is any integer. In the generating function $\widetilde{W} = \widetilde{W}(\bar{\gamma}, \tilde{J}, t)$ written in Eq. (4.29), resonances occur if

$$\bar{\Omega}T = m\pi, \tag{5.2a}$$

$$\bar{\Omega}T = m\pi/2, \tag{5.2b}$$

$$\bar{\Omega}T = m\pi/3. \tag{5.2c}$$

We can predict that the next-order generating function will have a vanishing denominator if

$$\bar{\Omega}T = m2\pi/l, \tag{5.3}$$

where $l = 1, 3, 5, 7$. Generalization to higher order is now apparent.

Vanishing denominators is inherent in the method of expanding in powers of ρ, but does not preclude the existence of toroidal surfaces. It is possible to have a configuration which exhibits several new magnetic axes that encircle the general system of toroidal surfaces. The appearance of the small denominators found in Eqs. (5.1) and (5.2) is associated with the fact that the lower-order term in the expansion of the pressure in powers of ρ has been taken to be quadratic; in the general case, the expansion can start with terms of still higher order. To achieve minimal regions of islands and stochastic field lines in a plasma confinement configuration, the equilibrium fields should be constructed such that the numerators corresponding to arbitrarily small denominators in the generating function vanish. In effect, such fields lead to optimal confinement configurations. The ability to attain these fields computationally and experimentally are matters yet to be investigated.

We briefly mention developments of expansion procedures to at least stave off the singularity to higher order in the expansion. The methods are part of the concept of renormalization. The resonances we find can be put off to higher order by renormalizing the pressure such that the expansion procedure begins at a higher power of ρ. This method has been applied to magnetic fields in a vacuum[14] (where a flux function was renormalized as opposed to the pressure). Another type of renormalization is that of accelerated convergence.[18] In this procedure, the expansions are developed in terms of special functions, instead of a single parameter, which permit more rapid convergence. In Hamiltonian systems, renormalization techniques yield solutions that converge in certain regions of phase space to actual solutions for motion taken over finite but long periods of time.[19] Furthermore, the solutions can in some cases closely approximate the motion within a prescribed course graining of the phase space for arbitrarily long times. This latter result is due to the actual convergence of certain so-called KAM series solutions. In summary, renormalization techniques may be applicable for treating the potential resonances appearing in the series solutions.

VI. DISCUSSION AND CONCLUSIONS

A method has been devised to investigate the existence of magnetic surfaces and MHD plasma equilibria in 3-D toroidal geometry. Expanding the contravariant components of the magnetic field and scalar pressure in distance ρ from the magnetic axis, a 1-D Hamiltonian for the lines of force was written out explicitly. A method of repeated canonical transformation was then formulated and applied to displace the coordinate dependence in the Hamiltonian to higher order in the expansion parameter. The result of this procedure was the identification of an adiabatic invariant, which was shown to correspond to pressure in the equilibrium problem. The invariant, contained in part in Eq. (4.38), is valid through fourth order in ρ. Arising in each successive canonical transformation was found a series of potentially resonant denominators. These potential resonances were identified, and their meaning discussed.

With the procedure outlined in Sec. IV C, an invariant to the Hamiltonian could in principle be obtained through any desired order in ρ. Each canonical transformation effectively determines the solution of another coefficient η_i in the pressure series expansion. An exact solution to the equilibrium equations [Eqs. (1.4) and (1.5)] using the expansion series [Eqs. (2.1)–(2.4)] entails, first, finding iterated solutions for a_n, b_n, c_n, and η_n, and, second, showing that each series converges. Establishing an exact solution appears to be a very arduous task. In principle, arriving at iterated solutions is possible although the complexity of their representations grows rapidly as n increases.[7,15] The existence of small denominators in the expansion coefficients however, places series convergence in jeopardy. Resonances will eventually occur in the series expansions if the quantity $\bar{\Omega}T/2\pi$ is a rational number [see Eq. (5.3)]. Only if $\bar{\Omega}T/2\pi$ is sufficiently irrational is there hope for series convergence.[4,20] From Sec. III A, we recall that $\bar{\Omega}T$ is defined to be

$$\bar{\Omega}T = \int_0^T dt\, \Omega(t),$$

where $\Omega(t) = \sigma(t)^{-2}$, and σ is governed by Eq. (4.2). An investigation into the evaluation of $\bar{\Omega}T/2\pi$ is currently under way, and the results will be reported in a later publication.

With series convergence in doubt, it behooves us to seek an alternative interpretation of these solutions. One possible interpretation is to classify them as asymptotic solutions. However, by the very definition of asymptotic solution, it is necessary to have already established the existence of exact solutions.[21] A more practical use of the results of our analysis would be in defining a concept of approximate solution. Some justification for this statement is that the static, ideal MHD equilibrium model, described by Eqs. (1.1)–(1.3), is, at best, a crude approximation of the experimental plasma. We define an approximate solution to be a solution for which given some fixed number M, each series, truncated at order N, has the property that the ratio of the sum of the $N+1$ through $N+M$ terms to the Nth term vanish in the limit as $\rho \to 0$. Investigation of approximate solutions using the results of this analysis are also currently under study. The utility of this approximate-solution concept must be evaluated by how well an experimental plasma, near MHD equilibrium, conforms to this approximate state.

What has been accomplished in a global sense by this analysis is an explicit identification of criteria (via the resonant denominators) necessary to construct approximate flux surfaces. Freedom to satisfy these criteria, and freedom to create optimal configurations possessing minimal regions of islands and stochasticity, lies in the specification of free functions contained in the expressions for the pressure and magnetic field components. [The explicit appearance of free functions is contained, for example, in Eqs. (2.5)–(2.7).] The specification of these free functions will ultimately determine idealized pressure profiles, as well as idealized outer boundaries. To determine these relationships, the canonical transformation procedure must be carried out to such an order that the addition of one more order has negligible effect on the equilibrium configuration. With each transformation, there exists an increasing amount of algebra. Application of an algebraic manipulator will be necessary. An alternative transformation procedure, utilizing Lie transforms, may serve to reduce the complexity of the transformation equations.

ACKNOWLEDGMENTS

One of the authors (J. T.) wishes to thank Los Alamos National Laboratory for warm hospitality during a portion of the present research. The authors benefited from a stimulating discussion with John Green on an earlier stage of this research.

This work was supported in part by the National Science Foundation under Contract No. ECS-8110038.

APPENDIX A: EXPLICIT REPRESENTATIONS OF THE HAMILTONIANS

In this appendix, we list the Hamiltonians that appear in the text. Explicit representations are made for the quadratic, cubic, and quartic parts.

The constituents of the first Hamitonian are the expansion coefficients of the mangetic field components [see Eq. (2.39)]. From the governing equilibrium equations, the expansion coefficients b_n and c_n have the following structure[15]:

$$b_n(\theta,s) = \sum_{l=l_0}^{n+2} b_{nlc}(s)\cos l\theta + b_{nls}(s)\sin l\theta,$$

$$c_n(\theta,s) = \sum_{l=l_0}^{n} c_{nlc}(s)\cos l\theta + c_{nls}(s)\sin l\theta.$$

In the summations, $l_0 = 0\,(1)$ if n is even (odd). For the zeroth harmonic $l_0 = 0$, we drop the subscript c. Thus $b_0(\theta,s)$ is written as

$$b_0(\theta,s) = b_{00}(s) + b_{02c}(s)\cos 2\theta + b_{02s}(s)\sin 2\theta.$$

The b_{nn+2c} and b_{nn+2s} coefficients are arbitrary periodic functions of s; they may be evaluated by specifying the shape of the outer boundary. The coefficients c_{nlc} and c_{nls} are functions of b_{nlc}, b_{nls} τ, k, j, and c_0.

The Hamiltonian $H(x,\bar{p},t)$ appears in Eq. (3.4). It can be written as

$$H(x,\bar{p},t) = H_2(x,\bar{p},t) + H_3(x,\bar{p},t) + H_4(x,\bar{p},t) + \cdots,$$

where

$$H_n(x,\bar{p},t) = \frac{1}{n} c_0^{-n/2} \sum_{l=0}^{n} h_{nl}(t)\bar{p}^l x^{n-l}. \tag{A1}$$

The functions h_{nl} are

$$h_{20} = b_{00} - b_{02c}, \tag{A2a}$$

$$h_{21} = 2b_{02s}, \tag{A2b}$$

$$h_{22} = b_{00} + b_{02c}, \tag{A2c}$$

$$h_{30} = b_{11s} - b_{13s},$$

$$h_{31} = -3b_{13c} + b_{11c} + 2k(b_{02c} - b_{00}),$$

$$h_{32} = 3b_{13s} + b_{11s} - 4kb_{02s},$$

$$h_{33} = b_{13c} + b_{11c} - 2k(b_{00} + b_{02c}),$$

$$h_{40} = b_{24c} - b_{22c} + b_{20} + (1/c_0)(b_{02c} - b_{00})(c_{20} - c_{22c}),$$

$$h_{41} = -4b_{24s} + 2b_{22s} + \tfrac{7}{3}kb_{13s} - \tfrac{7}{3}kb_{11s} + (2/c_0)[c_{22s}(b_{02c} - b_{00}) + b_{02s}(c_{22c} - c_{20})],$$

$$h_{42} = -3b_{24c} + 2b_{20} + 7kb_{13c} - \tfrac{7}{3}kb_{11c}$$

$$+\tfrac{14}{3}k^2(b_{00}-b_{02c})+(2/c_0)(c_{22c}b_{02c}$$
$$-2c_{22s}b_{02s}-c_{20}b_{00}),$$
$$h_{43}=4b_{24s}+2b_{22s}-7kb_{13s}-\tfrac{7}{3}kb_{11s}+\tfrac{28}{3}k^2b_{02s}$$
$$-(2/c_0)[c_{22s}(b_{02c}+b_{00})+b_{02s}(c_{20}+c_{22c})],$$
$$h_{44}=b_{24c}+b_{22c}+b_{20}-\tfrac{7}{3}k(b_{13c}+b_{11c})$$
$$-(1/c_0)(c_{20}+c_{22c})(b_{00}+b_{02c})+\tfrac{14}{3}k^2(b_{00}+b_{02c}).$$

The Hamiltonian $K(X,P,t)$ appears in Eq. (4.3). It is written as

$$K(X,P,t)=K_2(X,P,t)+K_3(X,P,t)+K_4(X,P,t)+\cdots,$$

where

$$K_n=\sum_{l=0}^{n}k_{nl}(t)P^lX^{n-l}.$$

The functions k_{nl} are

$$k_{20}=\Omega/2,\quad k_{21}=0,\quad k_{22}=\Omega/2,$$
$$k_{nl}=\Omega^{l-n/2}\sum_{m=0}^{n-l}\alpha_{nlm}h_{n\,n-m}h_{22}^{\frac{m}{2}-n/2}\times(2h_{21}-\dot h_{22}/h_{22}+\dot\Omega/\Omega)^{n-l-m},\quad n\geqslant 3, \tag{A3}$$

where the constants α_{nlm} have the following values:

$\alpha_{300}=-\frac{1}{24}$,	$\alpha_{301}=\frac{1}{4}$,	$\alpha_{302}=-\frac{1}{2}$,	$\alpha_{303}=\frac{1}{3}$,	
$\alpha_{310}=\frac{1}{4}$,	$\alpha_{311}=-1$,	$\alpha_{312}=1$,		
$\alpha_{320}=-\frac{1}{2}$,	$\alpha_{321}=1$,			
$\alpha_{330}=\frac{1}{3}$,				
$\alpha_{400}=\frac{1}{64}$,	$\alpha_{401}-\frac{1}{8}$,	$\alpha_{402}=\frac{3}{8}$,	$\alpha_{403}=-\frac{1}{2}$,	$\alpha_{404}=\frac{1}{4}$,
$\alpha_{410}=-\frac{1}{8}$,	$\alpha_{411}=\frac{3}{4}$,	$\alpha_{412}=-\frac{3}{2}$,	$\alpha_{413}=1$,	
$\alpha_{420}=\frac{3}{8}$,	$\alpha_{421}=-\frac{3}{8}$,	$\alpha_{422}=\frac{3}{2}$,		
$\alpha_{430}=-\frac{1}{2}$,	$\alpha_{431}=1$,			
$\alpha_{440}=\frac{1}{4}$.				

The Hamiltonian $K(\gamma,J,t)$ appears in Eq. (4.5). It is written as

$$K(\gamma,J,t)=K_2(J,t)+K_3(\gamma,J,t)+K_4(\gamma,J,t)+\cdots, \tag{A4}$$

where

$$K_2(J,t)=\Omega J, \tag{A5a}$$

$$K_3(\gamma,J,t)=\{\tfrac{1}{24}[k_{33}-3k_{31}+i(k_{30}-3k_{32})]e^{3i\gamma}+\tfrac{1}{8}[k_{33}+k_{31}-i(k_{30}+k_{32})]e^{i\gamma}+\text{c.c.}\}(2J)^{3/2}, \tag{A5b}$$

$$K_4(\gamma,J,t)=\{[\tfrac{1}{64}(k_{44}-6k_{42}+k_{40})+(i/16)\times(k_{41}-k_{43})]e^{4i\gamma}+[\tfrac{1}{8}(k_{44}-k_{40})-(i/8)(k_{43}+k_{41})]e^{2i\gamma}+\tfrac{3}{64}(k_{44}+2k_{42}+k_{40})+\text{c.c.}\}(2J)^2. \tag{A5c}$$

In Eqs. (A5b) and (A5c), c.c. stands for the complex conjugate of all preceding terms.

The Hamiltonian $\bar K(\bar\gamma,\bar J,t)$ appears in Eq. (4.28). It is written as

$$\bar K(\bar\gamma,\bar J,t)=\Omega\bar J+[\bar K_{46}(t)e^{6i\bar\gamma}+\bar K_{44}(t)e^{4i\bar\gamma}+\bar K_{42}(t)e^{2i\bar\gamma}+\bar K_{40}(t)+\text{c.c.}](2\bar J)^2+\cdots, \tag{A6}$$

where

$$\bar K_{46}=3[\mathrm{Re}\,K_{33}\,\mathrm{Re}\,3iW_3-\mathrm{Im}\,K_{33}\,\mathrm{Im}\,3iW_3+i(\mathrm{Re}\,K_{33}\,\mathrm{Im}\,3iW_3+\mathrm{Re}\,3iW_3\,\mathrm{Im}\,K_{33})],$$

$$\bar K_{44}=3[\mathrm{Re}\,K_{33}\,\mathrm{Re}\,iW_1+\mathrm{Re}\,K_{31}\,\mathrm{Re}\,3iW_3-\mathrm{Im}\,K_{33}\,\mathrm{Im}\,iW_1-\mathrm{Im}\,K_{31}\,\mathrm{Im}\,3iW_3+i(\mathrm{Re}\,K_{33}\,\mathrm{Im}\,iW_1+\mathrm{Re}\,K_{31}\,\mathrm{Im}\,3iW_3+\mathrm{Re}\,3iW_3\,\mathrm{Im}\,K_{31}+\mathrm{Re}\,iW_1\,\mathrm{Im}\,K_{33})]+K_{44},$$

$$\bar K_{42}=3[\mathrm{Re}\,K_{33}\,\mathrm{Re}\,iW_1+\mathrm{Re}\,K_{31}\,\mathrm{Re}\,3iW_3+\mathrm{Re}\,K_{31}\,\mathrm{Re}\,iW_1+\mathrm{Im}\,K_{33}\,\mathrm{Im}\,iW_1+\mathrm{Im}\,K_{31}\,\mathrm{Im}\,3iW_3-\mathrm{Im}\,K_{31}\,\mathrm{Im}\,iW_1-i(\mathrm{Re}\,K_{33}\,\mathrm{Im}\,iW_1+\mathrm{Re}\,3iW_3\,\mathrm{Im}\,K_{31}-\mathrm{Re}\,K_{31}\,\mathrm{Im}\,3iW_3-\mathrm{Re}\,iW_1\,\mathrm{Im}\,K_{33}-\mathrm{Re}\,K_{31}\,\mathrm{Im}\,iW_1-\mathrm{Re}\,iW_1\,\mathrm{Im}\,K_{31})]+K_{42},$$

$$\bar K_{40}=3[\mathrm{Re}\,K_{33}\,\mathrm{Re}\,3iW_3+\mathrm{Im}\,K_{33}\,\mathrm{Im}\,3iW_3+\mathrm{Re}\,K_{31}\,\mathrm{Re}\,iW_1+\mathrm{Im}\,K_{31}\,\mathrm{Im}\,iW_1]+K_{40}.$$

In the above expressions, K_{nl} contains the terms in $K_n(\gamma,J,t)$ multiplying $e^{il\gamma}$. Here, Re and Im denote the real and imaginary part, respectively.

Finally, $\tilde K(\tilde\gamma,\tilde J,t)$ is the Hamiltonian appearing in Eq. (4.37). It is written as

$$\tilde K(\tilde\gamma,\tilde J,t)=\Omega\tilde J+\bar K_{40}(t)(2\tilde J)^2+\cdots. \tag{A7}$$

APPENDIX B: TRANSFORMATION OF A QUADRATIC HAMILTONIAN

The objective of this appendix is to find a canonical transformation which takes the Hamiltonian

$$H(x,\bar p,t)=\tfrac{1}{2}f(t)\bar p^2+g(t)\bar p x+\tfrac{1}{2}h(t)x^2 \tag{B1}$$

into

$$K(X,P,t)=(1/2\sigma^2)(X^2+P^2). \tag{B2}$$

In Eq. (B2), σ is a function of time. Hamilton's equations, for the respective Hamiltonians, read

$$\begin{bmatrix} \dot{x} \\ \dot{\bar{p}} \end{bmatrix} = \begin{bmatrix} g & f \\ -h & -g \end{bmatrix} \begin{bmatrix} x \\ \bar{p} \end{bmatrix}, \tag{B3}$$

$$\begin{bmatrix} \dot{X} \\ \dot{P} \end{bmatrix} = \begin{bmatrix} 0 & 1/\sigma^2 \\ -1/\sigma^2 & 0 \end{bmatrix} \begin{bmatrix} X \\ P \end{bmatrix}. \tag{B4}$$

The transformation equation from $(x, \bar{p}, t)$ to (X, P, t) is written as

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} y_1 & y_2 \\ y_3 & y_4 \end{bmatrix} \begin{bmatrix} x \\ \bar{p} \end{bmatrix}. \tag{B5}$$

Taking the time derivative of Eq. (B5), and substituting from Eqs. (B3) and (B4), the governing equations for the transformation functions y_i are obtained. They are

$$\begin{bmatrix} \dot{y}_1 & \dot{y}_2 \\ \dot{y}_3 & \dot{y}_4 \end{bmatrix} = -\begin{bmatrix} y_1 & y_2 \\ y_3 & y_4 \end{bmatrix} \begin{bmatrix} g & f \\ -h & -f \end{bmatrix} + \begin{bmatrix} 0 & 1/\sigma^2 \\ -1/\sigma^2 & 0 \end{bmatrix} \begin{bmatrix} y_1 & y_2 \\ y_3 & y_4 \end{bmatrix}. \tag{B6}$$

A solution of this set of equations is

$$\begin{bmatrix} y_1 & y_2 \\ y_3 & y_4 \end{bmatrix} = \begin{bmatrix} f^{-1/2}\sigma^{-1} & \\ \sigma f^{-1/2}(g - \dot{f}/2f - \dot{\sigma}) & \end{bmatrix} \begin{bmatrix} & 0 \\ & \sqrt{f}\,\sigma \end{bmatrix}, \tag{B7}$$

where σ satisfies

$$\ddot{\sigma} + \left(fh - g^2 - \dot{g} + \frac{1}{2}\frac{\ddot{f}}{f} - \frac{3}{4}\frac{\dot{f}^2}{f^2} + \frac{\dot{f}}{f} g \right) \sigma = \frac{1}{\sigma^3}. \tag{B8}$$

Solutions to Eq. (B8) are discussed in Ref. 16. The transformation defined by Eqs. (B5) and (B7) is canonical as the Poisson bracket of X, P is unity.

APPENDIX C: INVERSION OF CANONICAL COORDINATES

In this appendix, we invert Eqs. (4.7) and (4.8) to obtain $\bar{J} = \bar{J}(\gamma, J, t)$ and $\gamma = \gamma(\bar{\gamma}, \bar{J}, t)$. Equation (4.7), which reads

$$J = \bar{J} + [3iW_3(t)e^{3i\gamma} + iW_1e^{i\gamma} + \text{c.c.}](2\bar{J})^{3/2}, \tag{C1}$$

is inverted as follows. First, $\bar{J}$ is expanded in half-powers of J beginning with the first power of J:

$$\bar{J} = J + \lambda_1(2J)^{3/2} + \lambda_2(2J)^2 + \lambda_3(2J)^{5/2} + \cdots. \tag{C2}$$

Then substituting Eq. (C2) into Eq. (C1) and collecting terms in equal powers of J, we have

$$\bar{J}(\gamma, J, t) = J - 2\bar{\sigma}_2(2J)^{3/2} + 12\bar{\sigma}_2{}^2(2J)^2 - 84\bar{\sigma}_2{}^3(2J)^{5/2} + \cdots, \tag{C3}$$

where $\bar{\sigma}_2$ is defined to be

$$\bar{\sigma}_2 = \text{Re}(3iW_3e^{3i\gamma} + iW_1e^{i\gamma}). \tag{C4}$$

Thus, the inversion of Eq. (4.7) has been completed. The inversion of Eq. (4.8) is somewhat more involved.

Equation (4.8) is written

$$\bar{\gamma} = \gamma + 3[W_3(t)e^{3i\gamma} + W_1e^{i\gamma} + \text{c.c.}](2\bar{J})^{1/2}. \tag{C5}$$

Now γ is expanded formally in powers of $(2\bar{J})^{1/2}$:

$$\gamma = \bar{\gamma} + \phi_1(2\bar{J})^{1/2} + \phi_2(2\bar{J}) + \phi_3(2\bar{J})^{3/2} + \cdots. \tag{C6}$$

In Eq. (C6) the ϕ_i's are functions of $\bar{\gamma}$ and t. Equation (C6) is substituted into Eq. (C5), with each of the exponentials expanded in a power series. Terms are then gathered in equal powers of $\bar{J}$. The results through order $\bar{J}^{3/2}$ are

$$\phi_1 = -6\sigma_1,$$

$$\phi_2 = -36\sigma_1\sigma_2,$$

$$\phi_3 = 108\sigma_1\sigma_3 + 216\sigma_1\sigma_2^2,$$

where $\sigma_k = \text{Re}[(3i)^{k-1}W_3e^{3i\bar{\gamma}} + i^{k-1}W_1e^{i\bar{\gamma}}]$.

[1] D. Lortz, Z. Naturforsch. **37a**, 892 (1982).
[2] M. D. Kruskal and R. M. Kulsrud, Phys. Fluids **1**, 265 (1958).
[3] H. Grad, Phys. Fluids **10**, 137 (1967).
[4] K. J. Whiteman, Rep. Prog. Phys. **40**, 1033 (1977).
[5] P. R. Garabedian, J. Anal. Math. **36**, 44 (1979).
[6] C. Mercier, Nucl. Fusion **4**, 213 (1964).
[7] D. Lortz and J. Nuhrenberg, Z. Naturforsch. **31a**, 1277 (1976).
[8] H. R. Lewis, Jr., Phys. Rev. Lett. **18**, 510 (1967).
[9] H. R. Lewis, Jr., J. Math. Phys. **9**, 1976 (1968).
[10] H. R. Lewis, Jr., Phys. Rev. **172**, 1313 (1968).
[11] K. R. Symon, J. Math. Phys. **11**, 1320 (1970).
[12] P. G. L. Leach, J. Math. Phys. **18**, 1902 (1977).
[13] H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 165 (1982).
[14] L. S. Solove'ev and V. D. Shafranov, *Reviews of Plasma Physics* (Consultants Bureau, New York, 1970), Vol. 5, p. 1.
[15] M. P. Bernardin and J. A. Tataronis, "Analytic representations of 3-D equilibrium magnetic fields," in preparation.
[16] A proof of the existence of a periodic solution to Eq. (3.7) is given in M. P. Bernardin and J. A. Tataronis, Phys. Rev. Lett. **54**, 2332 (1985).
[17] A. J. Lichtenberg and M. A. Lieberman, *Regular and Stochastic Motion* (Springer, New York, 1983), p. 74.
[18] N. N. Bogoljubov, J. A. Mitropoliskii, and A. M. Samoilenko, in *Methods of Accelerated Convergence in Nonlinear Mechanics* (Hindustan, Delhi, 1976).
[19] A. J. Lichtenberg and M. A. Lieberman, in Ref. 17, p. 77.
[20] E. T. Whittaker, *Analytical Dynamics* (Cambridge U. P., Cambridge, 1927), p. 437.
[21] See, for instance, C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw–Hill, New York, 1978), p. 89.

The plane rotations allowed for d-dimensional discrete lattices and their application to the Bravais lattices in three dimensions

Shoon K. Kim
Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122

(Received 11 February 1985; accepted for publication 24 May 1985)

It is shown that the plane rotations allowed for d-dimensional discrete lattices are independent of the dimension d and given by the n-fold rotations C_n with $n = 2, 3, 4$, and 6. The formalism is then applied to formulate the general analytical expressions of the Bravais lattices for three dimensions.

I. INTRODUCTION

Let $V^{(d)}$ be a d-dimensional vector space. Then it has been shown by the author[1] that a plane rotation in $V^{(d)}$ can be described by an infinitesimal plane rotation ω in a form which is free from the dimension d. Using this form, we shall first show that the allowed plane rotations for a discrete latice $L^{(d)}$ (or a group of vectors) in $V^{(d)}$ are limited to the n-fold rotations with $n = 2, 3, 4$, and 6. This is a well-known result for $d = 2$ or 3. The proof is based on the eigenvalue problem of a projective symmetry operator which defines a lattice vector on the plane of rotation from an arbitrary lattice vector in $L^{(d)}$. We shall then show that the eigenvalue problem plays the essential role in providing the general analytical expressions for the well-known 14 Bravais lattices in $V^{(3)}$ in a straightforward manner. There exist many methods of constructing the Bravais lattices; however, the geometrical methods discussed by many workers[2] and the matrix algebraic method introduced by Seitz[3] are far more involved than the simple nature of the concept warrants.

II. THE ALLOWED PLANE ROTATIONS FOR $L^{(d)}$

A plane rotation in a d-dimensional vector space $V^{(d)}$ is described by an infinitesimal plane rotation ω through an angle α as follows[1]:

$$\begin{aligned} R(\alpha\omega) &= \exp(\alpha\omega) \\ &= E + \omega \sin\alpha + \omega^2(1-\cos\alpha), \end{aligned} \tag{1}$$

where E is the unit tensor and ω is a skew-symmetric tensor which satisfies

$$\omega^3 + \omega = 0. \tag{2}$$

This is the reduced characteristic equation of ω with three distinct roots, so that one can introduce three projection operators.[1] For the present purpose we need only two projection operators defined by

$$-\omega^2, \quad 1+\omega^2, \tag{3}$$

which are mutually orthogonal. Let x be an arbitrary vector in $V^{(d)}$. Then the projections of x onto and normal to the ω plane are given by

$$x_\parallel = -\omega^2 \cdot x, \quad x_\perp = (1+\omega^2)\cdot x, \tag{4}$$

respectively. If one can express these projection operators in terms of the symmetry operations of a given lattice, one can find the lattice vectors lying on the ω plane or perpendicular to the ω plane from an arbitrary lattice vector.

Let $L^{(d)}$ be a discrete lattice in $V^{(d)}$, and let us write for a plane rotation which leaves $L^{(d)}$ invariant as follows:

$$C_n = \exp(2\pi\omega/n), \tag{5}$$

where n is simply a parameter to be determined. On account of the periodicity, however, one may assume $n > 1$, excluding the trivial case of $n = 1$. Then from (2) the symmetry operation of $L^{(d)}$ which is linear in the projection operator $-\omega^2$ is given by

$$\begin{aligned} P &\equiv (2E - C_n - C_n^{-1}) = -p_n\omega^2, \\ p_n &= 2[1-\cos(2\pi/n)]. \end{aligned} \tag{6}$$

It satisfies a projective eigenvalue problem

$$p^2 = p_n P, \tag{7}$$

where p_n is the nonzero eigenvalue of P and plays the essential role in the present argument. Let t be an arbitrary vector belonging to $L^{(d)}$. Then there exists another lattice vector given by $t_\parallel = Pt$ which is on the ω plane and satisfies

$$Pt_\parallel = p_n t_\parallel. \tag{8}$$

Since there exists a minimum lattice vector on the ω plane for a discrete lattice one concludes that the eigenvalue p_n is an integer. By definition of p_n given in (6), its values are limited to

$$p_6 = 1, \quad p_4 = 2, \quad p_3 = 3, \quad p_2 = 4, \tag{9}$$

corresponding to $n = 6, 4, 3$, and 2. Q.E.D.

Now according to Wulff's theorem[4,5] any proper or improper rotation in $V^{(d)}$ is given by a product of an even or odd number ($\leqslant d$) of reflections in $V^{(d)}$. Since a product of two reflections defines a plane rotation one can state that any proper rotation in $V^{(d)}$ can be given by a product of a number ($\leqslant d/2$) of plane rotations. The above result provides a certain limitation on the proper rotations allowed for $L^{(d)}$ as well as on the angle between two allowed hyperplanes of reflections in $L^{(d)}$. In the next section we shall show that the above formalism is also very effective in providing the general analytical expressions for the Bravais lattices in three dimensions.

III. THE ANALYTICAL EXPRESSIONS FOR THE BRAVAIS LATTICES IN $V^{(3)}$

Using the general formalism introduced in the previous section we shall now determine the possible symmetries and the lattice types allowed for $L^{(3)}$. For this purpose we intro-

 0022-2488/85/092381-02$02.50

duce a coordinate system on $L^{(3)}$. Let $C_n \in L^{(3)}$; then by the projective symmetry operation P given in (6) we can introduce two linearly independent lattice vectors $\mathbf{b}_1$, $\mathbf{b}_2$ on the ω plane. A lattice vector $\mathbf{b}_3$ normal to the plane is brought out by another projective symmetry operation H

$$H \equiv E + C_n + \cdots + C_n^{\,n-1} = n(1+\omega^2)\,, \tag{10}$$

which satisfies $H^2 = nH$. We may choose $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ to be the shortest in each direction and express a lattice vector $\mathbf{t} \in L^{(3)}$ by

$$\mathbf{t} = x\mathbf{b}_1 + y\mathbf{b}_2 + z\mathbf{b}_3 = (x, y, z) \quad (\text{Mod integers})\,. \tag{11}$$

The parallelepiped defined by $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ may be simply called the Bravais parallelepiped (BP). Then the lattice $L^{(3)}$ is completely characterized by additional lattice points, if any, on the faces or inside of the BP.

When $n > 2$, we may choose $\mathbf{b}_1$ to be one of the shortest on the ω plane and let $\mathbf{b}_2 = C_n \mathbf{b}_1$. Then the face $(\mathbf{b}_1, \mathbf{b}_2)$ becomes primitive since both $\mathbf{b}_1$ and $\mathbf{b}_2$ are the shortest on the plane. In this coordinate system, the rotation $C_n (n > 2)$ is represented by

$$C_n(x, y, z) = (-y, x + (2 - p_n)y, z)\,. \tag{12}$$

Then, from the requirement that $\mathbf{t} - C_n\mathbf{t}$ is also a lattice vector for any $\mathbf{t} \in L^{(3)}$ we obtain a general expression for $\mathbf{t}$ when $n > 2$,

$$\mathbf{t} = (-m/p_n, m/p_n, z) \quad (\text{Mod integers})\,, \tag{13}$$

where m is an integer bounded by $0 \leqslant |m| < p_n$ and z will be specified shortly. The symmetry of the x, y components in (13) combined with the inherent inversion symmetry i of any lattice $L^{(3)}$ show that there exists a binary axis of rotation C_2' about the lattice vector $\mathbf{b}_1 + \mathbf{b}_2$ and hence about $\mathbf{b}_1$ and $\mathbf{b}_2$ as well. When $n = 2$, the above argument fails so that there may or may not exist C_2' perpendicular to $C_2(\|\mathbf{b}_3)$. Thus, we arrive at the well-known conclusion that the point symmetry of $L^{(3)}$ is classified by

$$C_i, C_{2i}, D_{ni} = D_n \times C_i \quad (n = 2,3,4,6)\,, \quad O_i = O \times C_i\,. \tag{14}$$

The cubic system follows by augmenting the D_{2i} system with a C_3 axis of rotation. The notations for the improper point groups are introduced by the author for convenience in describing the extensions of the proper point groups.[6]

We shall next construct the lattice types belonging to $D_{ni}(n > 2)$ based on (13). By definition, the third basis vector $\mathbf{a}_3$ which forms a primitive unit cell with $\mathbf{b}_1$ and $\mathbf{b}_2$ is given by the lattice vector $\mathbf{t}$ with the smallest z component in (13). Thus, for $m = 0$ we have

$$\mathbf{a}_3 = \mathbf{b}_3\,, \tag{15a}$$

i.e., the BP is primitive. When $m \neq 0$, the z component cannot be zero since the face $(\mathbf{b}_1, \mathbf{b}_2)$ is primitive. Thus,

$$\mathbf{a}_3 = (-m/p_n, m/p_n, 1/p_n)\,, \quad 0 < |m| < p_n\,, \tag{15b}$$

p_n being a prime for $n > 2$. This means that all extra lattice points are inside of the BP and given by $l\mathbf{a}_3$, where l is an integer bounded by $0 < l < p_n$. Thus, the hexagonal lattice belonging to D_{6i} has only a primitive BP since $p_6 = 1$, while the tetragonal lattice belonging to D_{4i} has a body-centered BP with $m = 1, p_4 = 2$ in addition to the primitive BP. In the case of the lattice belonging to D_{3i} we have a double-centered hexagonal BP which contains two extra lattice points given by

$$\mathbf{a}_3 = (\tfrac{2}{3}, \tfrac{1}{3}, \tfrac{1}{3}), \quad 2\mathbf{a}_3 = (\tfrac{1}{3}, \tfrac{2}{3}, \tfrac{2}{3})\,, \tag{16}$$

corresponding to $m = 1$ and $p_3 = 3$. The lattice with $m = 2$ is equivalent to the above under the C_3 rotation about $\mathbf{b}_3$. The primitive rhombohedral lattice is defined by $\mathbf{a}_3$, $C_3\mathbf{a}_3$, $C_3^{\,2}\mathbf{a}_3$.

For the D_{2i} system we choose all three basis vectors in the directions of the three orthogonal C_2 axes. Then the possible lattice vectors are given by

$$\mathbf{t} = (m_1/2, m_2/2, m_3/2) \quad (\text{Mod integers})\,, \tag{17}$$

where $m_i = 0$ or ± 1 with the condition that $|m_1| + |m_2| + |m_3| \neq 1$, since each basis vector is the shortest in its own direction. Obviously, the condition should apply for the sums and differences of the coexisting $\mathbf{t}$'s for any one type of the BP. Thus, the only allowed lattice types are (i) primitive, (ii) single-face centered, (iii) all-face centered, and (iv) body centered. In the special case when the three basis vectors become equivalent, the D_{2i} system becomes the cubic system and thus we have only three types of the BP: primitive, all-face centered, and body centered. Now for the C_{2i} system the face (b_1, b_2) perpendicular to the C_2 axis can always be chosen to be primitive such that there remain only two lattice types: primitive and single-face centered. Finally, for the C_i system, the BP is always chosen to be primitive.

It should be noted here that any one of the lattice types given above cannot be obtained from the other belonging to the same symmetry system by means of a continuous deformation without going through the modification of the symmetry of the system during the process. Thus, we have constructed all the 14 Bravais lattices.

IV. CONCLUDING REMARKS

Based on the plane rotation expressed in a form (1) which is independent of the dimension d of the space we have shown that the allowed plane rotations for discrete lattices $L^{(d)}$ are given by the n-fold rotation C_n's limited to $n = 2$, 3, 4, and 6. The formalism is then used to construct the general analytical expression (15) for the Bravais lattices belonging to the D_{3i}, D_{4i}, and D_{6i} systems and (17) for the D_{2i} system. The present method of constructing the Bravais lattices is straightforward and much simpler than any other existing methods, geometrical[2] or matrix algebraic.[3] The construction of the general Bravais lattices for $L^{(d)}$ based on the present approach requires further investigation.[7]

[1]S. K. Kim, J. Math. Phys. **21**, 1299 (1980).

[2]See, for example, H. Hilton, *Mathematical Crystallography and the Theory of Groups of Movements* (Dover, New York, 1963).

[3]F. Seitz, Z. Kristallogr. **90**, 289 (1935).

[4]G. V. Wulff, Proc. Warsaw Nat. Hist. Soc. (1896, 1897).

[5]E. Cartan, *The Theory of Spinors* (MIT, Cambridge, MA, 1966).

[6]S. K. Kim, J. Math. Phys. **24**, 414, 419 (1983).

[7]For the existing theories for $L^{(d)}$, see R. L. E. Schwarzenberger, "N-dimensional crystallography," in *Putnam Research Notes in Math #41* (Pitman, San Francisco, 1980).

The continuity of dislocations

Iris Bloomer
Department of Physics, San Jose State University, San Jose, California 95192
John M. Charap
Department of Physics, Queen Mary College, University of London, London E1 4NS, England

(Received 22 January 1985; accepted for publication 2 May 1985)

Dislocation continuity is derived from the Bilby–Kondo theory of dislocations using exterior calculus. Dislocation density is represented by the torsion vector-valued two-form. Burgers vectors are associated with the vector part of the torsion while dislocation lines are associated with the two-form part. The exterior derivative of the torsion is shown to vanish when the crystal curvature vanishes. This implies two simultaneous continuity conditions: Burgers vector conservation and continuity of dislocation lines. On the other hand, dislocation continuity is violated when the curvature does not vanish. Since this can occur on grain boundaries it is inferred that grain boundaries are regions where crystal curvature is concentrated.

I. INTRODUCTION

Continuity is a well-established feature of dislocations. It involves both the dislocation line and the Burgers vector. That is, dislocation lines can never end in the crystal. They either form closed loops or end at the crystal boundary or grain boundaries. At the same time, Burgers vectors are conserved inside the crystal so that at a node the sum of ingoing Burgers vectors equals the sum of outgoing Burgers vectors.

Any viable theory of dislocations should account for continuity. Bilby[1] and, independently, Kondo[2] developed a theory for dislocations which associates dislocation density with the geometry of the crystal. According to their theory, a continuous distribution of dislocations can be represented by a geometric object, torsion. (Torsion can be related to a closure failure of a parallelogram constructed from basis vectors.) Expressing torsion as a tensor, Bilby and Smith[3] were able to account for conservation of Burgers vectors by manipulation of tensor indices.

In this paper both aspects of dislocation continuity are deduced from the Bilby–Kondo theory by using exterior calculus. Torsion is expressed as a vector-valued two-form. Burgers vectors are associated with the vector part of the torsion while dislocation lines are linked to the two-form part. It is shown that the exterior derivative of the torsion vector-valued two-form vanishes provided the total curvature of the crystal vanishes. Consequently, two simultaneous continuities are induced: the vector part of the torsion, i.e., Burgers vector, is conserved, and the two-form part is a continuous structure so that dislocation lines never end in the crystal.

Furthermore, at grain boundaries continuity can be violated, i.e., dislocation lines can end within the crystal at grain boundaries. In this case the exterior derivative of torsion should not vanish. This generates a geometrical interpretation of grain boundaries which is discussed further.

II. DEFINITIONS AND FRAMEWORK

This section follows the treatment on exterior calculus presented by Misner, Wheeler, and Thorne.[4] It is not intended as a complete account, but rather as a brief summary of some of the pertinent concepts used in the analysis.

The conventions used in this paper are such that all indices, upper and lower, run from one to three. Repeated indices are summed over. A comma denotes differentiation. Indices surrounded by square brackets indicate antisymmetric components.

At each point in the crystal we assume linearly independent basis vectors $\mathbf{e}_\alpha$ and corresponding basis one-forms $\mathbf{d}x^\alpha$. A general one-form $\mathbf{C}$ has components with respect to the basis one-form, i.e., $\mathbf{C} = x\, \mathbf{d}x$. A one-form can be visualized as a pattern of surfaces which do not necessarily mesh together.[4] The net number of these surfaces passing through a closed loop, then, will not be zero. This implies that $\mathbf{d}\boldsymbol{\sigma} \neq 0$, where $\mathbf{d}$ represents the exterior derivative, discussed below. This can be linked to the Stokes theorem.[4] Conversely, if the net number of one-form surfaces passing through a closed loop is zero, then $\mathbf{d}\boldsymbol{\sigma} = 0$.

The tensor product between basis vectors or basis one-forms or both can be constructed and used as a basis from which to build general tensors, e.g.,

$$\mathbf{S} = S^{\alpha\beta}{}_{\gamma\delta} = \mathbf{e}_\alpha \otimes \mathbf{e}_\beta \otimes \mathbf{d}x^\gamma \otimes \mathbf{d}x^\delta. \tag{1}$$

The wedge product between p one-forms, called a p-form, is constructed from the antisymmetrized tensor product of one-forms. A basis p-form is defined as

$$\mathbf{d}x^1 \wedge \mathbf{d}x^2 \wedge \cdots \wedge \mathbf{d}x^p = (1/p!)\delta^{\,1\,2\cdots p}_{\,\alpha_1\alpha_2\cdots\alpha_p}\, \mathbf{d}x^{\alpha_1} \otimes \mathbf{d}x^{\alpha_2} \otimes \cdots \otimes \mathbf{d}x^{\alpha_p}, \tag{2}$$

where $\delta^{\,1\,2\cdots p}_{\,\alpha_1\alpha_2\cdots\alpha_p}$ is the alternating tensor.

The highest value p can assume is equal to the dimension of the space under consideration; so that in three dimensions $p = 0, 1, 2, 3$. (A zero-form is a function.) A general p-form can be expressed in terms of its components along a basis p form; for example, a general two-form $\boldsymbol{\rho}$ can be written as

$$\boldsymbol{\rho} = \rho_{\alpha\beta}\mathbf{d}x^{\alpha} \wedge \mathbf{d}x^\beta. \tag{3}$$

A p-form can be envisioned as the intersection of p families of surfaces which form cells, each with a given orientation. For example, a two-form can be thought of as a tubular honeycomblike structure.[4] The cells of a general p-form do not necessarily mesh together. Consequently the net number

 0022-2488/85/092383-03$02.50

of cells emerging from a closed p-surface encompassing the p-form is nonzero. Cells have been created or destroyed within the p-surface. In this case the exterior derivative of the p-form is nonvanishing. On the other hand, a vanishing exterior derivative implies that the cells of the p-form blend smoothly into one another. No cells are created or destroyed within an encompassing p-surface.

Tensor-valued p-forms can be defined. They are p-forms which have a tensor associated with each cell. For example, torsion can be described as a vector-valued two-form. In this case a vector is attached to each of the tubes of the honeycomblike two-form structure.

The exterior derivative $\mathbf{d}$ operates on a p-form to create a $(p+1)$-form. For example, if f represents a zero-form, i.e., a function, then $\mathbf{d}f$ is the gradient one-form of the function

$$\mathbf{d}f = f_{,\alpha}\,\mathbf{d}x^\alpha. \tag{4}$$

Acting on a p form $\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} = \alpha_{\alpha\beta\gamma\ldots}\mathbf{d}x^\alpha \wedge \mathbf{d}x^\beta \wedge \cdots$, $\mathbf{d}$ produces the $(p+1)$-form $\mathbf{d}\boldsymbol{\alpha}$ such that

$$\begin{aligned}\mathbf{d}\boldsymbol{\alpha} &= \mathbf{d}\alpha_{\alpha\beta\gamma\ldots}\,\mathbf{d}x^\alpha \wedge \mathbf{d}x^\beta \wedge \cdots \\ &= \alpha_{\alpha\beta\gamma\ldots,\lambda}\,\mathbf{d}x^\lambda \wedge \mathbf{d}x^\alpha \wedge \mathbf{d}x^\beta \wedge \cdots.\end{aligned} \tag{5}$$

The exterior derivative acting twice on a p-form produces a vanishing form. For example, $\mathbf{d}(\mathbf{d}x^\alpha) = 0$.

Acting on a basis vector $\mathbf{e}_\alpha$, $\mathbf{d}$ produces the vector-valued one-form connection

$$\mathbf{d}\mathbf{e}_\alpha = \mathbf{e}_\beta \boldsymbol{\omega}^\beta{}_\alpha, \tag{6}$$

where

$$\boldsymbol{\omega}^\beta{}_\alpha = \Gamma^\beta{}_{\alpha\lambda}\,\mathbf{d}x^\lambda. \tag{7}$$

The components of $\boldsymbol{\omega}^\beta{}_\alpha$, namely $\Gamma^\alpha{}_{\beta\gamma}$, depend on the torsion components $C^\alpha{}_{\beta\gamma}$ as well as on the metric $g_{\alpha\beta}$; i.e.,

$$C^\alpha{}_{\beta\gamma} = \Gamma^\alpha{}_{\beta\gamma} - \Gamma^\alpha{}_{\gamma\beta} \tag{8}$$

and

$$\begin{aligned}\Gamma^\alpha{}_{\beta\gamma} &= \tfrac{1}{2}g^{\alpha\lambda}(g_{\lambda\beta,\gamma} + g_{\lambda\gamma,\beta} - g_{\beta\gamma,\lambda}) \\ &\quad + \tfrac{1}{2}(C^\alpha{}_{\beta\gamma} + C_{\beta\gamma}{}^\alpha + C_{\gamma\beta}{}^\alpha).\end{aligned} \tag{9}$$

III. CONDITION FOR WHICH THE EXTERIOR DERIVATIVE OF TORSION VANISHES

The density of a continuous distribution of dislocations can be represented by the torsion vector-valued two-form $\mathbf{C}$, where

$$\mathbf{C} = C^\alpha{}_{\beta\gamma}\mathbf{e}_\alpha\,\mathbf{d}x^\beta \wedge \mathbf{d}x^\gamma \tag{10}$$

and

$$C^\alpha{}_{\beta\gamma} = \Gamma^\alpha{}_{\beta\gamma} - \Gamma^\alpha{}_{\gamma\beta}. \tag{8}$$

The vector part of $\mathbf{C}$ is associated with Burgers vectors. The two-form part of $\mathbf{C}$ is associated with dislocation lines which run parallel to the tubes of this honeycomblike structure.

We now deduce the condition for which $\mathbf{dC} = 0$:

$$\mathbf{dC} = C^\alpha{}_{\beta\gamma,\lambda}\mathbf{e}_\alpha\,\mathbf{d}x^\lambda \wedge \mathbf{d}x^\beta \wedge \mathbf{d}x^\gamma + C^\alpha{}_{\beta\gamma}\,\mathbf{d}\mathbf{e}_\alpha \wedge \mathbf{d}x^\beta \wedge \mathbf{d}x^\gamma. \tag{11}$$

Inserting (6) into (11) gives

$$\mathbf{dC} = (C^\alpha{}_{\beta\gamma,\lambda} + C^\sigma{}_{\beta\gamma}\Gamma^\alpha{}_{\sigma\lambda})\mathbf{e}_\alpha\,\mathbf{d}x^\lambda \wedge \mathbf{d}x^\beta \wedge \mathbf{d}x^\gamma. \tag{12}$$

The components of $\mathbf{dC}$, namely $C^\alpha{}_{[\beta\gamma,\lambda]} + C^\sigma{}_{[\beta\gamma}\Gamma^\alpha{}_{/\sigma/\lambda]}$, vanish if the crystal has the property of absolute parallelism so that the total curvature $R^\alpha{}_{\beta\gamma\delta}$ vanishes. (The slashes around the subscript σ mean that this subscript is not included in the antisymmetrization.) This is because

$$C^\alpha{}_{[\beta\gamma,\lambda]} + C^\sigma{}_{[\beta\gamma}\Gamma^\alpha{}_{/\sigma/\lambda]} = -R^\alpha{}_{[\beta\gamma\lambda]}, \tag{13}$$

where

$$R^\alpha{}_{\beta\gamma\lambda} = \Gamma^\alpha{}_{\beta\lambda,\gamma} - \Gamma^\alpha{}_{\beta\gamma,\lambda} + \Gamma^\alpha{}_{\sigma\gamma}\Gamma^\sigma{}_{\beta\lambda} - \Gamma^\alpha{}_{\sigma\lambda}\Gamma^\sigma{}_{\beta\gamma}. \tag{14}$$

The identity (13) can be obtained by considering the vectorial torsion two-form $\boldsymbol{\Theta}^\mu$, where

$$\boldsymbol{\Theta}^\mu = \boldsymbol{\omega}^\mu{}_\nu \wedge \mathbf{d}x^\nu = -\tfrac{1}{2}C^\mu{}_{\alpha\beta}\,\mathbf{d}x^\alpha \wedge \mathbf{d}x^\beta. \tag{15}$$

As pointed out by Trautman[5]

$$\mathbf{d}\boldsymbol{\Theta}^\mu = \mathbf{R}^\mu{}_\gamma \wedge \mathbf{d}x^\gamma - \boldsymbol{\omega}^\mu{}_\gamma \wedge \boldsymbol{\Theta}^\gamma. \tag{16}$$

Here, $\mathbf{R}^\mu{}_\gamma$ is the tensorial curvature two-form

$$\mathbf{R}^\mu{}_\gamma = \mathbf{d}\boldsymbol{\omega}^\mu{}_\gamma + \boldsymbol{\omega}^\mu{}_\lambda \wedge \boldsymbol{\omega}^\lambda{}_\gamma = \tfrac{1}{2}R^\mu{}_{\gamma\alpha\beta}\,\mathbf{d}x^\alpha \wedge \mathbf{d}x^\beta. \tag{17}$$

Then from (15)–(17)

$$\begin{aligned}&-\tfrac{1}{2}C^\mu{}_{\alpha\beta,\gamma}\,\mathbf{d}x^\gamma \wedge \mathbf{d}x^\alpha \wedge \mathbf{d}x^\beta \\ &\quad = \tfrac{1}{2}R^\mu{}_{\alpha\beta\gamma}\,\mathbf{d}x^\alpha \wedge \mathbf{d}x^\beta \wedge \mathbf{d}x^\gamma \\ &\qquad + \tfrac{1}{2}C^\sigma{}_{\alpha\beta}\Gamma^\mu{}_{\sigma\gamma}\,\mathbf{d}x^\gamma \wedge \mathbf{d}x^\alpha \wedge \mathbf{d}x^\beta,\end{aligned} \tag{18}$$

which reduces to (13).

IV. DISCUSSION

When the curvature $R^\alpha{}_{\beta\gamma\delta}$ vanishes,

$$\mathbf{dC} = 0. \tag{19}$$

Since a vector and a two-form are associated with $\mathbf{C}$, the vanishing of its exterior derivative implies simultaneous conservation for both these elements. Therefore the Burgers vector is conserved so that at a node the ingoing Burgers vector is equal to the sum of outgoing Burgers vectors. Furthermore, the tubes of the two-form part of $\mathbf{C}$ must merge smoothly into one another so that the net number of tubes that pass through an arbitrary closed two-surface (e.g., the edges of a box) is zero. Since dislocation lines run along the tubes of the two-form, this implies that they must either end at the boundary of the crystal or form closed loops.

Continuity in other areas of physics can also be represented as the vanishing of the exterior derivative of a geometric object. For example, in electromagnetism, in any reference frame, magnetic field lines never end. This can be expressed as $\mathbf{dF} = 0$, where

$$\mathbf{F} = F_{ab}\,\mathbf{d}x^a \wedge \mathbf{d}x^b = A_{a,b}\,\mathbf{d}x^a \wedge \mathbf{d}x^b \tag{20}$$

is the Faraday two-form and A_a are the components of the electromagnetic potential one-form. (Latin indices run from 0, 1, 2, 3.)

On the other hand, electric lines can end on charges. This can be expressed as

$$\mathbf{d}{}^*\mathbf{F} = 4\pi{}^*\mathbf{J}, \tag{21}$$

where $^*\mathbf{F}$ is the Maxwell two-form dual to $\mathbf{F}$ and $^*\mathbf{J}$ is the charge three-form.

Similarly, in hydrodynamics, vortex lines cannot end in the fluid. They either end at the boundary or form closed loops. This can be expressed as

$$\mathbf{d}\boldsymbol{\omega} = 0, \tag{22}$$

where $\boldsymbol{\omega}$ represents the vorticity two-form

$$\boldsymbol{\omega} = \omega_{\mu\nu}\, \mathbf{d}x^{\mu} \wedge \mathbf{d}x^{\nu} = v_{\mu,\nu}\, \mathbf{d}x^{\mu} \wedge \mathbf{d}x^{\nu} \tag{23}$$

and v_{μ} represents the components of the velocity one-form.

Dislocation continuity can be violated when the crystal contains grains, i.e., dislocation lines can end on grain boundaries within the crystal. Accordingly, the exterior derivative of torsion should not vanish. The general equation

$$\mathbf{dC} = -R^{\alpha}{}_{\beta\gamma\delta}\, \mathbf{e}_{\alpha}\, \mathbf{d}x^{\beta} \wedge \mathbf{d}x^{\gamma} \wedge \mathbf{d}x^{\delta} \tag{24}$$

implies that grain boundaries can be represented by $R^{\alpha}{}_{[\beta\gamma\delta]}$.

Correspondingly, we can treat the crystal as divided into grains which are curvature-free except on their boundaries.

Regge calculus may prove useful in considerations involving grain boundaries. Regge calculus replaces an n-dimensional continuously curved, Riemannian manifold with n-dimensional blocks. All the curvature is concentrated on the skeletal (n-2)-dimensional hinges of these blocks.[4]

ACKNOWLEDGMENTS

We wish to thank Professor H. Morrison for the insightful discussions concerning dislocations. We are also very grateful to Dr. A. R. Forouhi for his critical review of the manuscript.

Funds in support of research were provided by the United States Department of Energy Laboratory Cooperative Program administered by Associated Western Universities, Inc. The work was performed at the Lawrence Berkeley Laboratory, which is operated by the University of California for the United States Department of Energy under Contract No. DE-AC07-76ET10723.

[1]B. A. Bilby, R. Bullough, and E. Smith, Proc. R. Soc. London Ser. A **231**, 263 (1955).

[2]K. Kondo, *Memoirs of the Unifying Study of the Basic Problems in Engineering by Means of Geometry* (Tokyo, 1955), Vols. I and II.

[3]B. A. Bilby and E. Smith, Proc. R. Soc. London Ser. A **236**, 481 (1956).

[4]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[5]A. Trautman, Bull. Acad. Pol. Sci. Ser. Sci. Math. Astron. Phys. **20**, 503 (1972).

Bloch oscillations in one-dimensional solids and solitary wave packets

Marshall Luban
Ames Laboratory, United States Department of Energy and Department of Physics, Iowa State University, Ames, Iowa 50011

(Received 29 October 1984; accepted for publication 10 May 1985)

We derive the exact solution of a single-band time-dependent Schrödinger equation for an electron in an idealized one-dimensional periodic solid in the presence of a constant uniform electric field. We show that all wave functions are necessarily periodic in time. This result is the fully quantum-mechanical analog of the well-known Bloch oscillations predicted by quasiclassical dynamics. Our method of solution consists of mapping the electron Schrödinger equation to the exactly solvable problem of a quantum planar rotor in the eikonal limit subject to an arbitrary angular and time-dependent external potential. The time periodicity of the electron wave functions is due to the fact that all of the rotor wave functions have the form of solitary wave packets.

I. INTRODUCTION

One of the remarkable results in Bloch's seminal article[1] of 1928 on the quantum theory of electrons in solids is that a weak static, uniform electric field can give rise to periodic electron oscillations, frequently referred to as "Bloch oscillations." The simplest system to consider is that of an idealized one-dimensional solid of lattice spacing a. The period of the oscillations is given by

$$\tau = h/(e\mathscr{E}a), \tag{1}$$

where h is Planck's constant, $-e$ is the charge of the electron, and $\mathscr{E}$ is the magnitude of the electric field. A simple qualitative explanation of this seemingly enigmatic phenomenon is given in most textbooks on solid-state physics.[2] The effect of the one-dimensional periodic electron–lattice potential is to replace the parabolic energy spectrum of a free electron by an infinite sequence of nondegenerate energy bands $E_l(k)$, labeled by the band index $l\,(=1,2,3,\ldots)$, where each function $E_l(k)$ is periodic in the wave vector k with period $2\pi/a$. Because of the energy gap separating successive bands, an electron initially occupying a given band l can be expected to remain in that band if one introduces a sufficiently weak,[3] static uniform field $\mathscr{E} = \mathscr{E}\hat{x}$. In the formalism of "quasiclassical dynamics,"[2] the equation of motion of the electron is given by $d(\hbar k)/dt = -e\mathscr{E}$. The energy of the electron at time t is then given by $E_l[k_0 - (e\mathscr{E}t/\hbar)]$ and its velocity by $v(t) = (1/\hbar)E_l'[k_0 - (e\mathscr{E}t/\hbar)]$, where k_0 is the initial wave vector. These are periodic functions of time with period τ given by (1). In addition, $v(t)$ is alternately positive and negative, suggesting a periodic spatial motion of the electron. In recent years with the advent of semiconductor superlattices, which provide a controlled one-dimensional periodic environment, the subject of Bloch oscillations has been transformed to one of great theoretical and technological importance in semiconductor physics. A specific theoretical proposal for the experimental observation of Bloch oscillations in semiconductor superlattices was made by Esaki and Tsu,[4] whereby the oscillations would give rise to a negative differential conductance. Confirming experimental observations were reported several years later by Esaki and Chang.[5]

In this article we provide the exact solution of the single-band time-dependent Schrödinger equation for an electron subject to an arbitrary spatially periodic one-dimensional potential in the presence of a uniform electric field. The basic picture of quasiclassical dynamics is corroborated in that all wave functions are indeed found to be periodic in time with period τ. Our method of solution consists of mapping this problem, formulated in Sec. II, to another exactly solvable problem, which is treated in Sec. III, that of a time-dependent Schrödinger equation

$$i\hbar\frac{\partial\Psi}{\partial t} = \left[-i\hbar\omega_0\frac{\partial}{\partial\phi} + V(\phi,t)\right]\Psi, \tag{2}$$

where $\omega_0 = 2\pi/\tau$. This equation describes a planar "pseudorotor" subject to an arbitrary angular and time-dependent external potential.[6] The appellation pseudorotor is appropriate because the Hamiltonian includes a term which is only *linear* in $L_z = -i\hbar\,\partial/\partial\phi$, the component of the angular momentum operator normal to the plane of the rotor. Now the key property of the solutions of (2) is that they satisfy the identity

$$|\Psi(\phi,t)|^2 = |\Psi(\phi - \omega_0 t,0)|^2. \tag{3}$$

That is, Ψ describes a solitary wave packet, one which preserves its shape given *any*[7] initial shape, and rotates with uniform angular velocity ω_0. This is the quantum analog of the result that the angular velocity of the classical pseudorotor is a constant of the motion with value ω_0 (see Sec. III below). The absence of any spatial spreading for *all* solutions of (2) is tied to the fact that, as discussed in Sec. III, Eq. (2) can be viewed as the Schrödinger equation for a *bona fide* rotor in the eikonal limit, that is, a rotor with an infinitely large moment of inertia and thus a constant angular velocity. The classical and quantum versions of the pseudorotor share the property that there is only one possible outcome for a position measurement which follows an earlier such measurement. The notion of the eikonal limit for a single nonrelativistic particle is discussed in some detail in Sec. IV.

In Sec. V A we apply the results derived for the pseudorotor to the problem of the single-band Schrödinger equation for an electron in an arbitrary, spatially periodic potential with an external electric field. The time periodicity of all electron wave functions is a direct consequence of the periodic solitary wave character of the pseudorotor wave functions. A number of special cases are considered in detail. For any nonzero value of $\mathscr{E}$, interband transitions which are strictly excluded in the single-band treatment are in principle open and have the potentiality of modifying the Bloch oscillations in some as yet unknown fashion. That is, the phenomenon of Bloch oscillations should be regarded then as an idealized phenomenon, valid only in an asymptotic sense for $\mathscr{E}\to 0$ and which will require ammendment in some form for any nonzero value of $\mathscr{E}$. (Analogous remarks have been made by Avron and Zak[8] concerning the related controversial problem of the existence of Stark ladders[9,10] for the one-dimensional solid in a uniform $\mathscr{E}$ field.) An exact formulation of the problem, involving all bands, is given in Sec. V B in terms of an infinite system of coupled linear partial differential equations of first order. This formulation, which utilizes a complete set of Wannier functions as basis states, may provide a potentially useful alternative to the traditional formulation based on the crystal momentum representation.[11] The subject of interband transitions has proved to be one of the most difficult yet intriguing questions of traditional solid-state physics.

Finally, in the Appendix we provide an alternate method for solving the single-band Schrödinger equation for the special case of nearest-neighbor overlap. The mathematical problem consists of solving an infinite system of first-order differential equations of tridiagonal form. The method employed in the Appendix consists of converting the set of differential equations to a system of coupled algebraic equations and solving the latter using a generating function method.

II. FORMULATION

We shall adopt as our one-electron Hamiltonian

$$H(x,t) = H_0(x,t) + e\mathscr{E}x\,, \tag{4}$$

where we require that H_0 be a Hermitian operator and spatially periodic with period a, $H_0(x+a,t) = H_0(x,t)$. For any problem of physical relevance H_0 can be taken to be time independent, of the form $H_0 = -(\hbar^2/2M)(\partial^2/\partial x^2) + V_L(x)$, where $V_L(x) = V_L(x+a)$ is the static electron–lattice interaction and M is the electron mass, but the more general form (4) can be treated with no additional difficulty. Now the solutions of the time-dependent Schrödinger equation $i\hbar\,\partial\psi/\partial t = H\psi$ can be expanded in terms of a complete set of orthonormal functions termed Wannier functions. Associated with each value of the band index l one defines a function $\phi_l(x)$. A doubly infinite set of functions $\{|n,l\rangle \equiv \phi_l(x-na)\}$ is called a set of Wannier functions if

$$\langle n,l|n',l'\rangle = \int_{-\infty}^{\infty} dx\,\phi_l^*(x-na)\phi_{l'}(x-n'a) = \delta_{ll'}\,\delta_{nn'}\,. \tag{5}$$

The single-band approximation consists of replacing the Hamiltonian (4) by a new operator $\widetilde{H}_l$ defined as

$$\widetilde{H}_l = \sum_{m,n=-\infty}^{\infty} \langle m,l|H|n,l\rangle |m,l\rangle\langle n,l|\,, \tag{6}$$

i.e., $\widetilde{H}_l$ consists of that portion of the spectral representation of H corresponding to a single band l. In the following we suppose that all wave functions of the system for an initial time $t=0$ are expandable exclusively in terms of the Wannier functions for the single band l. Then the Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = \widetilde{H}_l\psi \tag{7}$$

insures that the wave functions for any later time can also be so expanded, that is,

$$\psi(x,t) = \sum_{n=-\infty}^{\infty} f_n(t)\phi_l(x-na)\,. \tag{8}$$

Henceforth we shall suppress the band index. Because of the spatial periodicity of H_0 its matrix elements formed with the Wannier functions are of the Toeplitz form

$$\langle n|H_0|m\rangle = \langle n-m|H_0|0\rangle = \langle 0|H_0|m-n\rangle\,. \tag{9}$$

Furthermore, the matrix elements of the operator x can be expressed as

$$\begin{aligned}\langle n|x|m\rangle &= na\,\delta_{m,n} + \langle n-m|x|0\rangle\\ &= na\,\delta_{m,n} + \langle 0|x|m-n\rangle\,.\end{aligned} \tag{10}$$

Alternately, we may rewrite (9) and (10) as

$$\langle n|H|m\rangle = ne\mathscr{E}a\,\delta_{n,m} + \langle 0|H|m-n\rangle\,. \tag{11}$$

The time-dependent Schrödinger equation (7) is then easily shown to be equivalent to the infinite system of coupled homogeneous equations

$$i\hbar\frac{df_n}{dt} - ne\mathscr{E}a f_n = \sum_{n'} \langle 0|H|n'\rangle f_{n'+n}\,. \tag{12}$$

These equations serve as the starting point for the remainder of this work.

To solve these equations we define the generating function

$$\Psi(\phi,t) = (2\pi)^{-1/2} \sum_{n=-\infty}^{\infty} f_n(t)e^{in\phi} \quad (0<\phi<2\pi)\,. \tag{13}$$

Multiplying (12) by $e^{in\phi}$ and summing over all integers n one obtains

$$i\hbar\frac{\partial\Psi}{\partial t}(\phi,t) = [\omega_0 L_z + V(\phi,t)]\Psi(\phi,t)\,, \tag{2'}$$

where $L_z = -i\hbar\,\partial/\partial\phi$,

$$V(\phi,t) = \sum_{n=-\infty}^{\infty} e^{-in\phi}\langle 0|H|n\rangle\,, \tag{14}$$

and $\omega_0 = 2\pi/\tau$ with τ defined by (1). Note carefully that $V(\phi,t)$ is a *real* periodic function of ϕ with period 2π. The time dependence of V originates in any time dependence assumed for H_0.

The generating function (13) is thus seen to satisfy the time-dependent Schrödinger equation describing the pseudorotor defined in Sec. I. Now the key point is that (2) can be solved in closed form [see Eq. (24)], giving $\Psi(\phi,t)$ in terms of $\Psi(\phi,0)$, because $V(\phi,t)$ is a *local* potential. The details are provided in the following section. Specifying initial values $f_n(0)$ provides $\Psi(\phi,0)$, whereas the inverse relations

$$f_n(t) = (2\pi)^{-1/2} \int_0^{2\pi} d\phi\, e^{-in\phi} \Psi(\phi,t) \tag{15}$$

yield the values of $f_n(t)$ for any later time.

III. PLANAR PSEUDOROTOR

In this section we first examine the classical planar pseudorotor. This is followed by the derivation of the solution of the time-dependent Schrödinger equation (2) of the quantum pseudorotor.

Consider the time-dependent classical Hamiltonian

$$H(\phi,L_z,t) = \omega_0 L_z + V(\phi,t), \tag{16}$$

where ω_0 is an arbitrary constant, V is an arbitrary real function of its variables subject only to the requirement that $V(\phi + 2\pi,t) = V(\phi,t)$, and ϕ and L_z are standard canonically conjugate variables. To appreciate the specialized character of (16) note that the angular velocity of the classical system is a constant of the motion ω_0; one of the pair of Hamilton equations is $\dot{\phi} = \partial H/\partial L_z = \omega_0$. However, because of the term $V(\phi,t)$ the angular momentum L_z is not a conserved quantity. Now the term $\omega_0 L_z$ of (16) can be expressed as $(L_z + I\omega_0)^2/(2I) - I\omega_0^2/2$, when $I \to \infty$, enabling us to interpret (16) as the Hamiltonian for a *bona fide* rotor but with arbitrarily large moment of inertia, having a constant angular velocity $\dot{\phi} = \omega_0 + (L_z/I) \to \omega_0$ in this limit. One might expect then that there will be a close connection between the properties of the classical and quantum versions of this anomalous rotor.

A formal hint of this is provided by the following. The Hamilton–Jacobi equation for the system (16) is given by

$$\frac{\partial S}{\partial t} + \omega_0 \frac{\partial S}{\partial \phi} + V(\phi,t) = 0, \tag{17}$$

where $S(\phi,t)$ is Hamilton's principal function. Introducing the function

$$\tilde{S}(\phi,t) = -i\hbar \ln \Psi(\phi,t) \tag{18}$$

and using (2), one finds that $\tilde{S}$ also satisfies the classical equation of (17). This anomalous circumstance is, of course, due to the fact that the Hamiltonian (16) is linear in L_z. The physical consequences of this mathematical equivalence will become clear in the next paragraph.

To solve (17) we use Lagrange's method.[12] The subsidiary equations are

$$dt = (1/\omega_0) d\phi = -dS/V(\phi,t) \tag{19}$$

with independent solutions

$$u_1 = \phi - \omega_0 t = C_1, \tag{20}$$

$$u_2 = S + \int_0^t dt'\, V(C_1 + \omega_0 t', t') = C_2, \tag{21}$$

where C_1 and C_2 are arbitrary integration constants. The general solution of (17) can then be written as

$$S(\phi,t) = F(\phi - \omega_0 t) - \int_0^t dt'\, V(\phi - \omega_0 t + \omega_0 t', t'), \tag{22}$$

where $F(y)$ is any periodic complex function of the variable y with period 2π. Alternately we may rewrite (22) as

$$S(\phi,t) = S(\phi - \omega_0 t, 0) - \int_0^t dt'\, V(\phi - \omega_0 t', t - t'), \tag{23}$$

and the wave function as

$$\Psi(\phi,t) = \Psi(\phi - \omega_0 t, 0) \times \exp\left[-\frac{i}{\hbar} \int_0^t dt'\, V(\phi - \omega_0 t', t - t')\right]. \tag{24}$$

Recalling that V is real, it follows that

$$|\Psi(\phi,t)|^2 = |\Psi(\phi - \omega_0 t, 0)|^2, \tag{3}$$

and in particular $|\Psi|^2$ is independent of V. This confirms the statement made in the Introduction that Ψ describes a solitary wave packet, one which preserves any given initial shape, rotating with uniform angular velocity ω_0.

To emphasize the similarity between the classical and quantum versions of the pseudorotor, suppose in particular that a measurement performed at the instant $t = 0$ revealed that the quantum pseudorotor was at some angle ϕ_0. Then we have $|\Psi(\phi,t)|^2 = \delta(\phi - \omega_0 t - \phi_0)$ for all $t > 0$. In this instance the angular momentum as calculated for both the classical and quantum versions of (16) is given by

$$L_z(t) = L_z(0) - \int_0^t d\tau \frac{\partial}{\partial \phi_0} V(\phi_0 + \omega_0 \tau, \tau). \tag{25}$$

Before closing this section we examine the question of whether $\Psi(\phi,t)$ can be periodic in time with period τ. The quantity $|\Psi(\phi,t)|$ does in fact exhibit this property, as is easily seen using (3) and the fact that $\Psi(\phi,0)$ is periodic in ϕ with period 2π. In general this is *not* the case for the phase of Ψ, denoted by ξ, where $\Psi = |\Psi| e^{i\xi}$. Referring to (24) we have

$$\xi(\phi,t) = \xi(\phi - \omega_0 t, 0) - \frac{1}{\hbar} \int_0^t dt'\, V(\phi - \omega_0 t', t - t'). \tag{26}$$

Considering the times t and $t + \tau$ one finds that

$$\xi(\phi,t+\tau) - \xi(\phi,t) = -\frac{1}{\hbar} \int_0^t dt'\, V(\phi + \omega_0 t', t + t'). \tag{27}$$

The right-hand side (rhs) of (27) will in general vary with t so that ξ, and therefore Ψ, will not be periodic in time with period τ. An obvious exception[13] is the set of all time-independent potentials $V(\phi)$. For such potentials the rhs of (27) will vanish if we impose the single constraint[14]

$$\int_0^{2\pi} d\Theta\, V(\Theta) = 0. \tag{28}$$

IV. EIKONAL LIMIT

The material of this section, albeit of peripheral interest to the problem of Bloch oscillations, is presented in order to provide a deeper understanding of the conditions for the existence of only solitary wave-packet solutions of the Schrödinger equation.

We consider the one-particle nonrelativistic Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi + V\psi, \tag{29}$$

with V being an arbitrary real function of $\mathbf{r}$ and t. For our purposes it will be convenient to express ψ as

$$\psi(\mathbf{r},t) = F(\mathbf{r};t) \exp[(i/\hbar) S(\mathbf{r};t)], \tag{30}$$

where F and S are real functions of their arguments. Defining the quantity

$$\rho = F^2, \tag{31}$$

the Schrödinger equation is equivalent to the pair of equations

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \left(\rho \frac{1}{m} \nabla S\right) = 0, \tag{32}$$

$$\frac{\partial S}{\partial t} + \frac{1}{2m}(\nabla S)^2 + V - \frac{\hbar^2}{2m}\frac{1}{F}\nabla^2 F = 0. \tag{33}$$

Writing the phase S of the wave function as

$$S(\mathbf{r},t) = m\mathbf{v}_0 \cdot \mathbf{r} - E_0 t + \tilde{S}(\mathbf{r},t), \tag{34}$$

where $\mathbf{v}_0$ is a constant vector and $E_0 = \frac{1}{2}mv_0^2$, these equations become

$$\frac{\partial \rho}{\partial t} + \mathbf{v}_0 \cdot \nabla \rho + \nabla \cdot \left(\rho \frac{1}{m} \nabla \tilde{S}\right) = 0, \tag{35}$$

$$\frac{\partial \tilde{S}}{\partial t} + \mathbf{v}_0 \cdot \nabla \tilde{S} + V + (2m)^{-1}\left[(\nabla \tilde{S})^2 - \frac{\hbar^2}{F}\nabla^2 F\right] = 0. \tag{36}$$

Suppose now that in the large mass limit ($m \to \infty$, whereas $\hbar$ and $\mathbf{v}_0$ remain constant)

$$\lim_{m\to\infty} (1/m)\nabla \tilde{S} = 0. \tag{37}$$

Equations (35) and (36) reduce to

$$\frac{\partial \rho}{\partial t} + v_0 \frac{\partial \rho}{\partial z} = 0, \tag{38}$$

$$\frac{\partial \tilde{S}}{\partial t} + v_0 \frac{\partial \tilde{S}}{\partial z} + V = 0, \tag{39}$$

where we have chosen the direction of $\mathbf{v}_0$ as defining the z axis. We shall refer to this particular limit as the eikonal limit, for in essence Eq. (39) is the primary equation of the eikonal approximation[15] of scattering theory. The term involving $\hbar^2$ in (36) no longer appears in (39). Nevertheless, the eikonal limit is quite distinct from the classical limit ($\hbar \to 0$). The pair of equations (35) and (36), with $\hbar$ set equal to zero, are quite distinct from (38) and (39). Inspecting (17) and (23) it is evident that the general solutions of (38) and (39) are given by

$$S(x,y,z;t) = S(x,y,z - v_0 t;0) - \int_0^t dt'\, V(x,y,z - v_0 t';t - t'), \tag{40}$$

and

$$\rho(x,y,z;t) = \rho(x,y,z - v_0 t;0). \tag{41}$$

Thus in the eikonal limit the probability density behaves as a solitary wave and does not disperse in space with the passage of time. In particular, if the particle is found to be at a point $\mathbf{r}_0$ at $t = 0$ then a later position measurement will necessarily show that the particle is located at the point $\mathbf{r}_0 + \mathbf{v}_0 t$. In summary, in the eikonal limit *all* solutions of the Schrödinger equation (29) behave as solitary waves. For certain selected potentials V it is possible to find solitary wave solutions of (29), but these constitute an extremely small subclass of all possible solutions for that potential.[7]

V. ELECTRON WAVE FUNCTIONS

A. Solution of the single-band problem

The amplitude $f_n(t)$ has been found in preceding sections to be given by Eqs. (15) and (24). In the present section we shall suppose, first, that the one-electron Hamiltonian (4), and thus also V in (24), is time independent and, second, that $\langle 0|H|0\rangle = 0$ so that (28) is satisfied.[14] These assumptions result in $\Psi(\phi,t)$, and hence $f_n(t)$ and the wave function $\psi(x,t)$ of (8), being periodic in t with period τ. An explicit expression for $f_n(t)$ in terms of the set of initial values $\{f_n(0)\}$ is obtained upon combining (13), (15), and (24), the result being given by

$$f_n(t) = (2\pi)^{-1} \sum_{l=-\infty}^{\infty} f_l(0) e^{-il\omega_0 t} \int_0^{2\pi} d\phi \times \exp - i\left[(n-l)\phi + \frac{1}{\hbar}\int_0^t dt'\, V(\phi - \omega_0 t')\right]. \tag{42}$$

In the remainder of this section we shall consider a case of special interest, where $f_n(0) = \delta_{n,0}$, so that (42) reduces to

$$f_n(t) = (2\pi)^{-1} \int_0^{2\pi} d\phi \times \exp - i\left[n\phi + \frac{1}{\hbar}\int_0^t dt'\, V(\phi - \omega_0 t')\right]. \tag{43}$$

If the electric field is turned off, i.e., $\omega_0 \to 0$, Eq. (43) reduces to

$$f_n(t) = (2\pi)^{-1} \int_0^{2\pi} d\phi \exp - i\left[n\phi + \frac{1}{\hbar} V(\phi) t\right], \tag{44}$$

and, by contrast to (43), f_n is no longer a periodic function of t. Of particular interest is the behavior of (44) for large times. Suppose that the matrix elements $\langle 0|H|n\rangle$ ($n \neq 0$) are pure imaginary. The potential function $V(\phi)$ of (14) is then a real, odd function of ϕ. Suppose further that $V(\phi)$ has a single extremum within $(0,\pi)$ for an angle ϕ_0 and that $V''(\phi_0)$ is nonzero. Using the method of stationary phase one finds that

$$f_n(t) \sim [(2/\pi)(\hbar/t|V''(\phi_0)|)]^{1/2} \times \cos[n\phi_0 + (1/\hbar)V(\phi_0)t + (\pi/4)\mathrm{sgn}\, V''(\phi_0)]. \tag{45}$$

Thus, in the absence of an electric field, $f_n(t)$ exhibits damped oscillatory behavior for large t with an amplitude which decreases to zero as $t^{-1/2}$. This result for the single-band Schrödinger equation is the analog of the usual dispersion of a quantum wave packet for a free particle.

Returning to the result, Eq. (43), in the presence of an electric field, we specialize to the choice

$$V(\phi) = 2V_1 \sin\phi \quad (V_1^* = V_1), \tag{46}$$

which corresponds to the selection

$$\langle 0|H|n\rangle = iV_1(\delta_{n,1} - \delta_{n,-1}) \tag{47}$$

in (14), i.e., only nearest-neighbor overlap of Wannier functions. Equation (43) can then be rewritten as

$$f_n(t) = (2\pi)^{-1} e^{-(1/2)in\omega_0 t} \times \int_0^{2\pi} d\phi \exp - i[n\phi + \alpha(t)\sin\phi], \tag{48}$$

where $\alpha(t) = (4V_1/\hbar\omega_0)\sin(\frac{1}{2}\omega_0 t)$. Now the factor $\exp(-i\alpha$

$\times \sin\phi$) will be recognized as a generating function of Bessel functions of integral order

$$\exp(-i\alpha \sin\phi) = \sum_{m=-\infty}^{\infty} (-1)^m e^{im\phi} J_m(\alpha). \tag{49}$$

We thus arrive at the final result

$$f_n(t) = (-1)^n e^{-(1/2)in\omega_0 t} J_n((4V_1/\hbar\omega_0)\sin \tfrac{1}{2}\omega_0 t). \tag{50}$$

Noting that $J_{-n}(x) = (-1)^n J_n(x) = J_n(-x)$ one verifies at once that $f_{-n} = (-1)^n f_n$ and that f_n is periodic in t with period τ. In particular, $f_n(P\tau) = \delta_{n,0}$ for any integer value of P.

If the electric field is reduced to zero, (50) becomes

$$f_n(t) = (-1)^n J_n(2V_1 t/\hbar), \tag{51}$$

an expression valid for all t. For an arbitrary fixed value of t considered in the limit $n\to\infty$, Eq. (51) reduces to

$$f_n(t) \sim (-1)^n (1/n!)(V_1 t/\hbar)^n \quad (n\to\infty). \tag{52}$$

Another limiting case is that of fixed n and $t\to\infty$ for which Eq. (51) adopts the asymptotic form

$$f_n(t) \sim (\pi V_1 t/\hbar)^{-1/2} \cos[t(2V_1/\hbar) + \tfrac{1}{2}n\pi - \tfrac{1}{4}\pi]. \tag{53}$$

This result of course agrees with (45) for the present choice (46) for $V(\phi)$.

B. Formulation of the exact problem

In this subsection we drop the single-band restriction of Sec. II and consider the exact time-dependent Schrödinger equation based on the Hamiltonian H of Eq. (4). Employing the complete set of Wannier functions, the wave function can be expanded as

$$\psi(x,t) = \sum_{l=1}^{\infty} \sum_{n=-\infty}^{\infty} f_{n,l}(t)|n,l\rangle. \tag{54}$$

Using the orthonormality property, Eq. (5), of the Wannier functions, one can immediately generalize (11) to read

$$\langle n,l|H|n',l'\rangle = ne\mathscr{E}a\,\delta_{ll'}\,\delta_{nn'} + \langle 0,l|H|n'-n,l'\rangle. \tag{55}$$

For each band l we define a generating function

$$\Psi_l(\phi,t) = (2\pi)^{-1/2} \sum_{n=-\infty}^{\infty} f_{n,l}(t) e^{in\phi}. \tag{56}$$

It is then a simple exercise to verify that the exact time-dependent Schrödinger equation based on the Hamiltonian (4) is equivalent to the following system of coupled partial differential equations:

$$\frac{\partial \Psi_l}{\partial\phi} - \frac{1}{\omega_0}\frac{\partial \Psi_l}{\partial t} = \frac{1}{\hbar\omega_0} \sum_{l'=1}^{\infty} V_{ll'}(\phi)\Psi_{l'}(\phi,t), \tag{57}$$

where

$$V_{ll'}(\phi) = \sum_{n=-\infty}^{\infty} e^{-in\phi}\langle 0,l|H|n,l'\rangle. \tag{58}$$

As imposing as this system of equations may appear, it may prove to be a useful starting point for the treatment of the exact problem. In particular, it is likely that progress can be made by adopting an approximation such as retaining nonzero choices only for $V_{ll}(\phi)$ and $V_{l,l\pm 1}(\phi)$. We will pursue this matter in a later publication. The conventional approach to interband transitions, based on the crystal momentum representation,[11] has a long and venerable history, yet some of the underlying issues (e.g., Stark ladders) remain a topic of considerable controversy.[8–10] A new approach, based on Eq. (57), is worth careful study, and hopefully it may provide new insights.

By including interband coupling terms, the notion of strictly periodic oscillations (i.e., without damping) must break down. This is also suggested by the above formalism. In particular, Ψ_l can no longer have the form of a solitary wave. To support this statement, note that if Ψ_l were a solitary wave then $\int_0^{2\pi} d\phi\,|\Psi_l|^2$ would be time independent, whereas only the weaker normalization condition

$$\sum_{l=1}^{\infty} \int_0^{2\pi} d\phi\,|\Psi_l|^2 = 1 \tag{59}$$

is in effect.

VI. SUMMARY

In this article we have derived the exact solution of the single-band time-dependent Schrödinger equation for an electron in a spatially periodic one-dimensional potential in the presence of a uniform electric field. All solutions of this restricted Schrödinger equation are periodic in time with period given by (1). Our result provides a fully quantum-mechanical derivation of the phenomenon of Bloch oscillations which is traditionally derived using the formalism of quasiclassical dynamics. What makes this problem solvable is the fact that the restricted Schrödinger equation, in essence the infinite system of coupled equations (12), can be mapped to the exact problem of a quantum planar rotor considered in the eikonal limit. In this particular limit all eigenfunctions of the rotor have the form of solitary wave packets. An immediate consequence of this unique form is that all of the electron wave functions for the original problem are periodic in time. Once one removes the restriction to a single band, the electron wave functions are no longer strictly periodic in time. The mapping procedure gives rise to the imposing system of coupled partial differential equations (57). A fully quantum-mechanical treatment of the resulting modifications of Bloch oscillations remains as an interesting open problem in mathematical physics.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge several helpful suggestions of Pnina H. Lotan-Luban concerning the manuscript.

Ames Laboratory is operated for the United States Department of Energy by Iowa State University under Contract No. W-7405-Eng-82. This work was supported by the Director for Energy Research, Office of Basic Energy Sciences.

APPENDIX: LAPLACE TRANSFORM METHOD

The general result for $f_n(t)$, Eq. (42), was derived by exploiting a mapping of the system of coupled equations (12), which are equivalent to the single-band Schrödinger equation, to the problem of a pseudorotor in an external potential. For the special case where one includes only nearest-neighbor overlap of Wannier functions, i.e., the matrix elements $\langle 0|H|n\rangle$ are given by (47), we can obtain the closed-form solution, Eq. (50), using an alternate method based on Laplace transforms.

Our starting point is the reduced version of the system of equations of (12)

$$\frac{df_n}{dt} = -in\omega_0 f_n + \frac{V_1}{\hbar}(f_{n+1} - f_{n-1}), \tag{A1}$$

$$f_n(0) = \delta_{n,0}. \tag{A2}$$

Multiplying (A1) by e^{-st} and integrating over all positive times and employing (A2), one obtains

$$sF_n = \delta_{n,0} - in\omega_0 F_n - (V_1/\hbar)(F_{n+1} - F_{n-1}), \tag{A3}$$

where $F_n(s)$ denotes the Laplace transform of $f_n(t)$. To extract an explicit expression for F_n from (A3) we define the generating function

$$G(y;s) = \sum_{n=-\infty}^{\infty} F_n(s) y^n. \tag{A4}$$

More specifically, we show that the function G satisfies a simple, solvable first-order differential equation in the independent variable y. Once G is determined, the coefficient of y^n of its Laurent expansion is the desired quantity $F_n(s)$. Finally, calculation of the inverse Laplace transform of F_n yields the site amplitude $f_n(t)$.

Multiplying (A3) by y^n and summing over all integer values of n, one obtains the inhomogeneous differential equation

$$-i\omega_0 y \frac{dG}{dy} = -1 + \left[s + \frac{V_1}{\hbar}(y - y^{-1})\right] G. \tag{A5}$$

The solution of (A5) is given by

$$G = \frac{-i}{\omega_0} G_H \int dy (y G_H)^{-1}, \tag{A6}$$

where

$$G_H = y^{i(s/\omega_0)} \exp[i(V_1/\hbar\omega_0)(y + y^{-1})] \tag{A7}$$

is the homogeneous solution of (A5). Now the exponential term in (A7) is precisely a generating function of Bessel functions of integer order, namely

$$\exp[\tfrac{1}{2} iz(y + y^{-1})] = \sum_{n=-\infty}^{\infty} (iy)^n J_n(z). \tag{A8}$$

Combining (A6)–(A8), one obtains the final result for G

$$G(y;s) = \sum_{n=-\infty}^{\infty} (iy)^n \times \left\{ \sum_{m=-\infty}^{\infty} (s - im\omega_0)^{-1} J_m(\beta) J_{n+m}(\beta) \right\}, \tag{A9}$$

where

$$\beta = 2V_1/(\hbar\omega_0). \tag{A10}$$

Comparison of (A4) and (A9) provides an explicit formula for $F_n(s)$, namely

$$F_n(s) = i^n \sum_{m=-\infty}^{\infty} (s - im\omega_0)^{-1} J_m(\beta) J_{n+m}(\beta). \tag{A11}$$

Because of the very simple s dependence in (A11), the inverse transform $f_n(t)$ is immediate with the result

$$f_n(t) = i^n \sum_{m=-\infty}^{\infty} J_m(\beta) J_{n+m}(\beta) e^{im\omega_0 t}. \tag{A12}$$

The imposing-looking series (A13) can be written in closed form using Graf's addition formula[16]

$$\sum_{m=-\infty}^{\infty} J_m(\alpha) J_{n+m}(\beta) e^{im\gamma} = e^{in\sigma} J_n(\tau), \tag{A13}$$

where the quantities $\alpha, \beta, \gamma, \sigma, \tau$ are any complex numbers subject to the single relation

$$\beta = \alpha e^{-i\gamma} + \tau e^{i\sigma}. \tag{A14}$$

The final result is identical to (50).

[1]F. Bloch, Z. Phys. **52**, 555 (1928).

[2]See, for example, J. Ziman, *Principles of the Theory of Solids* (Cambridge U. P., Cambridge, England, 1972), 2nd ed., Chap. 6; N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders College, Philadelphia, 1976), Chap. 12.

[3]In principle any nonzero static electric field $\mathscr{E}$ gives rise to interband transitions (Zener breakdown); however, the numerical value of the transition probability plummets to zero for $\mathscr{E} \to 0$ (see Ref. 2).

[4]L. Esaki and R. Tsu, IBM J. Res. Dev. **14**, 61 (1970).

[5]L. Esaki and L. L. Chang, Phys. Rev. Lett. **33**, 495 (1974).

[6]We remark that a planar pseudorotor subject to time-periodic impulsive forces serves as the basis of an exactly solvable model of a one-dimensional system of electrons subject to a sum of potentials which are separately periodic but with incommensurate periods. See D. R. Grempel, S. Fishman, and R. E. Prange, Phys. Rev. Lett. **49**, 833 (1982).

[7]A linear harmonic oscillator is well known to support solitary wave packets, but only for very special initial shapes. See E. Schrödinger, Naturwissenschaften **14**, 664 (1926); E. H. Kennard, Z. Phys. **44**, 326 (1927); I. R. Senitzky, Phys. Rev. **95**, 1115 (1954); J. Plebański, Phys. Rev. **101**, 1825 (1956). The special case treated by Schrödinger is reproduced in L. I. Schiff, *Quantum Mechanics* (McGraw–Hill, New York, 1968), 3rd ed., p. 74–76. Also see S. M. Roy and V. Singh, J. Phys. A **14**, 2927 (1981) and references cited therein.

[8]J. Avron and J. Zak, Phys. Rev. B **9**, 658 (1974); J. Avron, Phys. Rev. Lett. **37**, 1568 (1976).

[9]G. H. Wannier, Phys. Rev. **117**, 432 (1960); Rev. Mod. Phys. **34**, 645 (1962); Phys. Rev. **181**, 1364 (1969).

[10]J. Zak, Phys. Rev. Lett. **20**, 1477 (1968); Phys. Rev. **181**, 1366 (1969); Phys. Lett. A **76**, 287 (1980).

[11]The following articles continue to provide an excellent exposition of the subject: E. N. Adams, Phys. Rev. **107**, 698 (1957); E. O. Kane, J. Phys. Chem. Solids **12**, 181 (1959).

[12]See, for example, E. T. Copson, *Partial Differential Equations* (Cambridge U. P., Cambridge, England, 1975), Chap. 1.

[13]More generally, the rhs of (24) will vanish for all potentials $V(\phi,t)$ which are periodic in ϕ and t with periods 2π and τ, respectively, and for which $\int_0^\tau dt \int_0^{2\pi} d\phi \, e^{in(\phi - \omega_0 t)} V(\phi,t) = 0$.

[14]In essence the constraint (28) is superfluous. It is trivial to show that if one waives this constraint for any time-independent potential $V(\phi)$, the quantities $f_n(t+\tau)$ and $f_n(t)$ differ only by a constant phase factor which is independent of n. Thus all physical properties derived from the electron wave functions (8) are left unaffected.

[15]See, for example, pp. 339–342 of Schiff's textbook cited in Ref. 7.

[16]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U. P., Cambridge, England, 1944), 2nd ed., Sec. 11.3.

The adsorption of simple particles on a $2\times N$ lattice space

J. L. Hock
Department of Electrical Engineering and Computer Science, Marquette University, Milwaukee, Wisconsin 53223

R. B. McQuistan
Department of Physics and Laboratory for Surface Science, University of Wisconsin—Milwaukee, Milwaukee, Wisconsin 53201

(Received 30 January 1985; accepted for publication 31 May 1985)

Utilizing a 15-term recursion relation that describes exactly the composite nearest neighbor degeneracy for simple indistinguishable particles on a $2\times N$ lattice, the adsorption isotherm is calculated and it is shown analytically that coverage as a function of the gas phase pressure exhibits no discontinuity, i.e., that no phase transition can occur.

I. INTRODUCTION

In a recent article,[1] we uitlized a set theoretic argument to develop a recursion that yields exactly the composite nearest neighbor degeneracy for simple, indistinguishable particles distributed on a rectangular $2\times N$ lattice space (see Fig. 1). Specifically, we have shown that $A[N,q,n_{11},n_{00}]$, the number of unique ways q indistinguishable particles can be arranged on a rectangular $2\times N$ lattice to form n_{11} occupied nearest neighbor pairs *and* n_{00} vacant nearest neighbor pairs (as well as n_{01}, the number of mixed nearest neighbor pairs), satisfies the recursion

$$\begin{aligned}
A&[N+3,q+3,n_{11}+4,n_{00}+4]\\
&=A[N+2,q+3,n_{11}+4,n_{00}+1]\\
&\quad+A[N+2,q+2,n_{11}+4,n_{00}+4]\\
&\quad+A[N+2,q+2,n_{11}+3,n_{00}+3]\\
&\quad+A[N+1,q+2,n_{11}+4,n_{00}+1]\\
&\quad-A[N+1,q+2,n_{11}+3,n_{00}]\\
&\quad+A[N+2,q+1,n_{11}+1,n_{00}+4]\\
&\quad+A[N+1,q+1,n_{11}+3,n_{00}+3]\\
&\quad-A[N+1,q+1,n_{11}+1,n_{00}+1]\\
&\quad+A[N+1,q,n_{11}+1,n_{00}+4]\\
&\quad-A[N+1,q,n_{11},n_{00}+3]\\
&\quad-A[N,q,n_{11}+3,n_{00}+3]\\
&\quad+(3)A[N,q,n_{11}+2,n_{00}+2]\\
&\quad-(3)A[N,q,n_{11}+1,n_{00}+1]\\
&\quad+A[N,q,n_{11},n_{00}], \qquad (1)
\end{aligned}$$

where the nonzero, initial values for A are shown in Table I.

The purpose of the present paper is to exploit this relationship to calculate the exact adsorption isotherm and to determine if a phase transition in the coverage is possible for such a system. The present calculation is concerned with a system in which each site has three nearest neighbors as opposed to a one-dimensional system where there are two nearest neighbors per site.

II. CALCULATION OF THE VARIOUS PARTITION FUNCTIONS

We first form the polynomials

$$f_{N,q}(x,y)=\sum_{n_{11},n_{00}} A[N,q,n_{11},n_{00}]x^{n_{11}}y^{n_{00}}, \qquad (2)$$

where the sum is over all permissible values of n_{11} and n_{00}, and where

$$x\equiv\exp[-V_{11}/kT], \quad y\equiv\exp[-V_{00}/kT],$$

in which V_{11} and V_{00} are the interaction potentials for occupied and vacant nearest neighbor pairs, respectively. Substituting Eq. (1) into Eq. (2) yields the canonical partition function (generating function)

$$\begin{aligned}
f_{N+3,q+3}&(x,y)\\
&=y^3f_{N+2,q+3}(x,y)+[1+xy]f_{N+2,q+2}(x,y)\\
&\quad+x^3f_{N+2,q+1}(x,y)+y^3[1-xy]f_{N+1,q+2}(x,y)\\
&\quad+xy[1-x^2y^2]f_{N+1,q+1}(x,y)\\
&\quad+x^3[1-xy]f_{N+1,q}(x,y)-xy[1-xy]^3f_{N,q}(x,y)
\end{aligned} \qquad (3)$$

with the initial conditions

$$\begin{aligned}
&f_{N,0}=y^3f_{N-1,0}, \quad N\geqslant 2,\\
&f_{N,1}=y^3f_{N-1,1}+2f_{N-1,0}, \quad N\geqslant 3,\\
&f_{N,2}=y^3f_{N-1,2}+[1+xy]f_{N-1,1}\\
&\qquad\quad+y^3[1-xy]f_{N-2,1}+xyf_{N-2,0}, \quad N\geqslant 3,\\
&f_{1,0}=y, \quad f_{2,1}=4y^2, \qquad (4)\\
&f_{2,2}=2[1+2xy], \quad f_{3,1}=2y^4[1+2y],\\
&f_{33}=2[1+2xy^2+2xy+3x^2y^2+2x^2y],\\
&f_{34}=4x^3+4x^2+2yx^4+4yx^3+yx^2,\\
&f_{35}=2x^4[1+2x], \quad f_{36}=x^7.
\end{aligned}$$

In Eqs. (3) and (4) we adopt the convention that $f_{N,q}(x,y)=0$ if $q<0$ or if $q>2N$.

The grand canonical partition (bivariate generating) function is now written

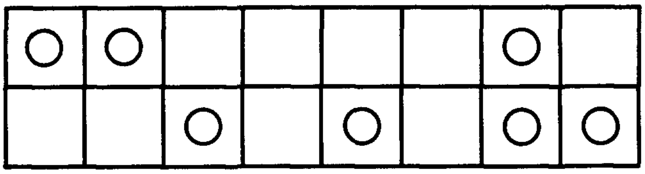


FIG. 1. An arrangement of seven particles on a 2×8 lattice, giving rise to three occupied nearest neighbors, 13 mixed neighbors, and six vacant nearest neighbors.

$$g_N(x,y,z)\equiv\sum_{q=0}^{2N} f_{N,q}(x,y)z^q\,, \tag{5}$$

where

$$z\equiv m\exp[\,\mu/kT\,]\,,$$

in which μ is the chemical potential, m is the adsorbed particle partition function

$$m=m(x)m(y)m(z)\exp[\,-V_0/kT\,]\,,$$

and V_0 is the interaction potential between the particle and the surface. It should be mentioned that in the Langmuir model for adsorption, $m(x)$, $m(y)$, and $m(z)$ are single-particle harmonic oscillator partition functions.

Now $g_N(x,y,z)$ can be found by substituting Eq. (3) into Eq. (5):

$$\begin{aligned} g_N(x,y,z) &= [\,y^3+z(1+xy)+x^3z^2]g_{N-1}(x,y,z)\\ &\quad+[\,y^3(1-xy)z+xy(1-x^2y^2)z^2\\ &\quad+x^3z^3(1-xy)]g_{N-2}(x,y,z)\\ &\quad-[xy(1-xy)^3z^3]g_{N-3}(x,y,z)\,, \end{aligned} \tag{6}$$

where the initial conditions are

$$g_1(x,y,z)=y+2z+xz^2\,, \tag{7a}$$

$$g_2(x,y,z)=y^4+4y^2z+2z^2(1+2xy)+4x^2z^3+x^4z^4\,, \tag{7b}$$

$$\begin{aligned} g_3(x,y,z) &= y^7+2y^4[1+2y]z\\ &\quad+[4y^3+4y^2+2xy^4+4xy^3+xy^2]z^2\\ &\quad+2[1+2xy^2+2xy+3x^2y^2+2x^2y]z^3\\ &\quad+[4x^3+4x^2+2x^4y+4x^3y+x^2y]z^4\\ &\quad+2x^4[1+2x]z^5+x^7z^6\,. \end{aligned} \tag{7c}$$

To obtain an explicit relation for $g_N(x,y,z)$, the grand canonical partition function, we first form the polynomials

$$\begin{aligned} h(x,y,z,\eta) &\equiv \sum_{N+1}^{\infty} g_N(x,y,z)\eta^N\\ &= \eta\left\{\frac{a_0+a_1\eta+a_2\eta^2}{b_0+b_1\eta+b_2\eta^2+b_3\eta^3}\right\}=\eta\,\frac{r(\eta)}{s(\eta)}\,, \end{aligned} \tag{8}$$

in which

$$a_0=y+2z+xz^2\,, \tag{9a}$$

$$\begin{aligned} a_1 &= z[\,y^2(4-x)-y(1+2y^2)]+z^2[\,y(2x-x^3)-xy^3]\\ &\quad+z^3[x^2(4-y)-x(1+2x^2)]\,, \end{aligned} \tag{9b}$$

$$\begin{aligned} a_3 &= z^3[4xy^2+4x^2y+2x^2y^2-4x^2y^3-4x^3y^2-4xy-x^3y\\ &\quad-xy^3+x^4y^2+x^2y^4+2x^3y^3]\,, \end{aligned} \tag{9c}$$

TABLE I. The initial conditions for the recursion expressed in Eq. (1).

N	q	n_{11}	n_{00}	A
1	0	0	1	1
1	1	0	0	2
1	2	1	0	1
2	0	0	4	1
2	1	0	2	4
2	2	0	0	2
2	2	1	1	4
2	3	2	0	4
2	4	4	0	1
3	0	0	7	1
3	1	0	4	2
3	1	0	5	4
3	2	0	2	4
3	2	0	3	4
3	2	1	2	1
3	2	1	3	4
3	2	1	4	2
3	3	0	0	2
3	3	1	1	4
3	3	1	2	4
3	3	2	1	4
3	3	2	2	6
3	4	2	0	4
3	4	2	1	1
3	4	3	0	4
3	4	3	1	4
3	4	4	1	2
3	5	4	0	2
3	5	5	0	4
3	6	7	0	1

and

$$b_0=1\,, \tag{10a}$$

$$b_1=-[\,y^3+z(1+xy)+z^2x^3]\,, \tag{10b}$$

$$b_2=-[z(1-xy)][\,y^3+xyz(1+xy)+z^2x^3]\,, \tag{10c}$$

$$b_3=xyz^3(1-xy)^3\,, \tag{10d}$$

and $r(\eta)$ is a quadratic function of η, while $s(\eta)$ is a cubic function of η. From

$$g_N(x,y,z)=\frac{1}{N!}\frac{\partial^N h}{\partial\eta^N}\bigg|_{\eta=0} \tag{11}$$

and using a partial fraction expansion of $h(x,y,z,\eta)$ we obtain

$$g_N(x,y,z)=\sum_{j=1}^{3}k_jR_j^N\,, \tag{12}$$

where the k_j's are given by

$$k_j=r(R_j)/s'(R_j) \tag{13}$$

and the R_j's are the reciprocals of the roots of the cubic [see Eq. (8)]

$$s(\eta)=b_0+b_1\eta+b_2\eta^2+b_3\eta^3=0\,. \tag{14}$$

If η_1 is the smallest root of $s(\eta)$, then, as $N\to\infty$, Eq. (12) becomes

$$g_N(x,y,z)\simeq k_1R_1^N\,, \tag{15}$$

where $R_1\equiv\eta_1^{-1}$.

This explicit expression for the grand canonical partition function can now be used to treat adsorption and the question of whether or not a phase transition occurs.

III. ADSORPTION

We first determine the expectation of θ, the coverage, i.e.,

$$\langle \theta \rangle_N = \langle q \rangle_N / 2N , \tag{16}$$

where

$$\langle q \rangle_N \equiv \left\{ \sum_{q=0}^{2N} q f_{N,q}(x,y) z^q \right\} \div \left\{ \sum_{q=0}^{2N} f_{N,q}(x,y) z^q \right\} , \tag{17}$$

so that

$$\sum_{q=0}^{2N} q f_{N,q}(x,y) z^q = 2N \langle \theta \rangle_N g_N(x,y,z) . \tag{18}$$

Utilizing Eqs. (6), (7), and (15) and assuming that $y \equiv 1$ and that, as $N \to \infty$,

$$\langle \theta \rangle_N = \langle \theta \rangle_{N-1} = \langle \theta \rangle_{N-2} = \cdots = \langle \theta \rangle ,$$

we find that

$$\langle \theta \rangle = \frac{3 - [3 + 2z(1+x) + x^3 z^2]\eta_1 - z(1-x)[2 + xz(1+x)]\eta_1^2}{4 - 2[1 + (1+x)z + x^3 z^2]\eta_1 - 2x(1-x)^3 z^3 \eta_1^3} . \tag{19}$$

To facilitate the graphical representation of Eq. (19), we first show that $\langle \theta \rangle = \frac{1}{2}$ when $z = x^{-3/2}$. To do so we set $\langle \theta \rangle = \frac{1}{2}$ in Eq. (19), utilize Eq. (14) for η_1, and obtain

$$(1 - x^3 z^2)\eta_1 [1 + z(1-x)\eta_1] = 0 . \tag{20}$$

Thus, either

$$1 - x^3 z^2 = 0 \tag{21a}$$

and/or

$$\eta_1 = 0 \tag{21b}$$

and/or

$$[1 + z(1-x)\eta_1] = 0 . \tag{21c}$$

But $\eta_1 = 0$ cannot satisfy Eq. (14). Furthermore, it can be shown that

$$\eta_1 = 1/z(x-1)$$

cannot be a root of Eq. (14). It follows that $\langle \theta \rangle = \frac{1}{2}$ at $z = x^{-3/2}$, i.e., at $u \equiv z x^{3/2} = 1$. The exponent ($\frac{3}{2}$) indicates a coordination number of 3 for a $2 \times N$ lattice space (see Ref. 2). In Fig. 2, we plot $\langle \theta \rangle$ vs $\ln u$ for selected values of x.

If we define

$$\gamma \equiv \langle \theta \rangle - \tfrac{1}{2}$$

and express γ in terms of u and x, we obtain

$$\gamma(u,x) = \frac{1 - [2 + x^{-3/2}(1+x)u]\eta_1 - x^{-3/2}(1-x)[2 + x^{-1/2}(1+x)]u\eta_1^2 + (1-x)^3 x^{-7/2} u^3 \eta_1^3}{4 - 2[1 + (1+x)x^{-3/2}u + u^2]\eta_1 - 2(1-x)x^{-7/2}u^3\eta^3} . \tag{22}$$

Using Eq. (10) and Eq. (14) one can show that

$$\eta(1/u) = u^2 \eta(u) . \tag{23}$$

Combining Eqs. (22), (23), and (14), it follows that

$$\gamma(u,x) = -\gamma(1/u,x) \tag{24a}$$

or equivalently that

$$\gamma[\ln u,x] = -\gamma[-\ln u,x] . \tag{24b}$$

Thus we see that γ is an antisymmetric function of $\ln u$. An examination of Eq. (22) shows that for all $x > 0$, γ has a nonzero first derivative and thus exhibits an inflection point at $\ln u = 0$ (i.e., $u = 1$). This implies that the maximum slope of γ (and hence $\langle \theta \rangle$) vs u occurs at $u = 1$.

The question of whether the system under discussion can exhibit a phase transition of coverage versus gas phase pressure resolves itself into a determination of $(d\langle \theta \rangle / du)_{u=1}$, the slope of the curve $\langle \theta \rangle$ vs u at $u = 1$, i.e., at $\langle \theta \rangle = \frac{1}{2}$.

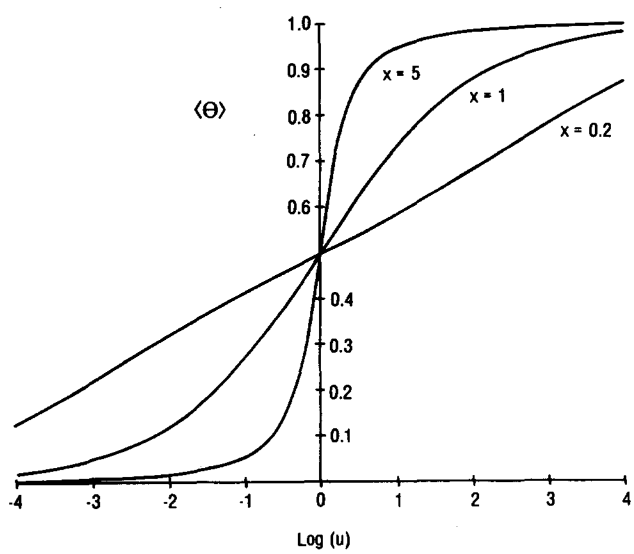


FIG. 2. The coverage as a function of $\ln u \equiv \ln(zx^{3/2})$, for various values of x.

We now show that there can be no singularity in $(d\langle\theta\rangle/du)_{u=1}$ and that it can never be negative.

First, we determine $(d\eta_j/du)|_{u=1}$ by considering Eq. (14) with $u = zx^{3/2}$ in the coefficients, we obtain

$$\left.\frac{d\eta_j}{du}\right|_{u=1} = -\eta_j . \tag{25}$$

It follows from Eq. (19) that

$$\left.\frac{d\langle\theta\rangle}{du}\right|_{u=1} = \frac{\eta_1 + x^{-3/2}(1-x)\eta_1^2}{3 - 2[2 + x^{-1/2} + x^{-3/2}]\eta_1 - x^{-2}(1-x)(1+x^{1/2})^2\eta_1} . \tag{26}$$

As $x \to \pm\infty$, for $u = zx^{3/2} = 1$, Eq. (14) becomes

$$(\eta - 1)^2 = 0 , \tag{27}$$

i.e., $\eta \to 1$. Thus, we can assume η to be expressed as a power series in x^{-1}. When this is done, it is seen that the numerator and denominator cannot be negative and that the denominator cannot vanish.

IV. CONCLUSION

On the basis of a recursion that yields exactly the composite nearest neighbor degeneracy for a $2\times N$ lattice space, we have determined the canonical, grand-canonical, and super-grand-canonical partition functions. Utilizing these partition functions we have determined the ensemble average coverage as a function of the interaction potentials and the gas phase pressure. In addition we have shown that this system does not exhibit a first-order phase transition in the coverage as a function of the gas phase pressure.

ACKNOWLEDGMENT

The research reported here has been supported in part by the Air Force Office of Scientific Research under Grant No. AFOSR81-0192.

[1] R. B. McQuistan and J. Hock, "An exact recursion for the composite nearest neighbor degeneracy for a $2\times N$ lattice space," J. Math. Phys. **25**, 261 (1984).

[2] T. L. Hill, *Statistical Thermodynamics* (Addison–Wesley, Reading, MA, 1960).

Fibonacci graphs possessing identical matching polynomials

Sherif El-Basil
Faculty of Pharmacy, Kasr El-Aini Street, Cairo, Egypt
Pavel Křivka
Department of Mathematics, Higher School of Chemical Technology, 53210 Pardubice, Czechoslovakia
Nenad Trinajstić
The Rugjer Bošković Institute, P.O.B. 1016, 41001 Zagreb, Croatia, Yugoslavia

(Received 10 October 1985; accepted for publication 22 February 1985)

A sequence of Fibonacci graphs is defined. A special case of Fibonacci graphs, i.e., those with identical matching polynomials, is discussed and conditions under which they appear are given. The appearance of Fibonacci graphs with identical matching polynomials may have some implications in statistical physics.

I. INTRODUCTION

The matching polynomial of a graph $G,\alpha(G)$, is a combinatorial mathematical structure defined by[1]

$$\alpha(G)=\alpha(G;x)=\sum_{k=0}^{[N/2]}(-1)^k p(G;k)x^{N-2k}, \tag{1}$$

where N is the number of vertices in G and $p(G,k)$ is the number of ways in which one can select k nonadjacent edges in G. By definition $p(G,0)=1$ for all G.

Recently, the matching polynomial has found many applications in diverse branches of science; viz., statistical physics,[2–5] thermodynamics,[6] pure mathematics,[1,7–9] molecular physics, and chemistry.[10–14] This warrants further studies on the structure of the matching polynomial.

In this paper we wish to present a sequence of families of graphs, all of which contain cycles and also possess the following properties: (1) each family has identical matching polynomial, and (2) each sequence of graphs consisting of one graph from each family represents what we will call a sequence of Fibonacci graphs.

II. FIBONACCI GRAPHS WITH IDENTICAL MATCHING POLYNOMIALS

Let L_i denote the matching polynomial of a linear chain of i vertices (or its characteristic polynomial since both polynomials are identical for chains). The recurrence relation[15]

$$L_i=xL_{i-1}-L_{i-2} \tag{2}$$

can be used to calculate the matching polynomials of linear chains starting with $L_0=1$ and $L_1=X$.

Let H be a graph with $|H|$ vertices, and let v be a fixed vertex of H. Then $G_{n,j}(H,v)$ denotes a graph with $n+j+|H|$ vertices obtained in the following way: Take a graph H and a cycle of $n+j$ vertices and connect any n vertices out of these $n+j$ with the vertex v. The graphs $G_{n,j}(H,v)$ have a particularly interesting feature: The matching polynomial for fixed n and j is always the same regardless of the positions of the n vertices which we choose to connect with v.

Definition: Let $G_1, G_2,\dots,G_n$ be a sequence of graphs. If, for the k matchings, $p(G_i,k)$'s, the recursion

$$p(G_i,k)+p(G_{i+1},k)=p(G_{i+2},k+1) \tag{3}$$

holds, we call the sequence $G_1, G_2,\dots,G_n$ a sequence of Fibonacci graphs. Hosoya[16] seems to be among the first who studied such sequences. He discussed the graphical aspects of the Fibonacci numbers in terms of his topological index.

Proposition 1: Let H be a graph, $v\in V(H)$ and $G\in G_{n,j}(H,v)$. Then

$$\alpha(G)=\alpha(H)\cdot(L_{n+j}-L_{n+j-2})-n\cdot\alpha(\overline{H})L_{n+j-1}, \tag{4}$$

where $\overline{H}$ is a graph obtained from H by omitting vertex v and adjacent edges.

Proof: We will make use of Heilbronner's formula[17]

$$\alpha(G)=\alpha(G-e)-\alpha(G-(e)), \tag{5}$$

where $G-e$ is a graph G without an edge e and $G-(e)$ is a graph G without an edge e and all edges adjacent to it. By means of this formula we can perform a simple and straightforward calculation of $\alpha(G)$. By removing an edge e, connecting the vertex v to the perimeter of G, we obtain

$$\alpha(G)=\alpha(G_1)-L_{n+j-1}\cdot\alpha(\overline{H}),$$

where $G_1\in G_{n-1,j+1}(H,v)$.

Repeating this calculation $(n-1)$ times we obtain

$$\alpha(G)=\alpha(G_{n-1})-(n-1)L_{n+j-1}\cdot\alpha(\overline{H}), \tag{6}$$

where $G_{n-1}\in G_{1,n+j-1}(H,v)$.

Finally,

$$\alpha(G_{n-1})=\alpha(H)\cdot\alpha(C_{n+j})-L_{n+j-1}\cdot\alpha(\overline{H}), \tag{7}$$

where the symbol C stands for a cycle. After using the formula[15,17] $\alpha(C_i)=L_i-L_{i-2}$, Eq. (7) transforms into

$$\alpha(G_{n-1})=\alpha(H)\cdot(L_{n+j}-L_{n+j-2})-L_{n+j-1}\cdot\alpha(\overline{H})$$

Q.E.D.

Corollary: Let $G_{n,j}$ belong to $G_{n,j}(H,v)$ for $H=v$. Then

$$\alpha(G_{n,j})=L_{n+j+1}-nL_{n+j-1}-L_{n+j-3}. \tag{8}$$

Proof: If $H=v$, then $\alpha(H)=x$ and we have

$$\alpha(G_{n,j})=x\cdot(L_{n+j}-L_{n+j-2})-nL_{n+j-1}.$$

On substituting $xL_{i-1}=L_i+L_{i-2}$ [see Eq. (2)] into (8) we obtain the required relationship

$$\alpha(G_{n,j})=L_{n+j+1}+L_{n+j-1}$$
$$-L_{n+j-1}-L_{n+j-3}-nL_{n+j-1}.$$

Example: We will calculate the first several matching polynomials for $H = v$, $n = 3$:

$$\alpha(G_{3,0}) = x^4 - 6x^2 + 3,$$
$$\alpha(G_{3,1}) = x^5 - 7x^3 + 8x,$$
$$\alpha(G_{3,2}) = x^6 - 8x^4 + 14x^2 - 3,$$
$$\alpha(G_{3,3}) = x^7 - 9x^5 + 21x^3 - 11x,$$
$$\alpha(G_{3,4}) = x^8 - 10x^6 + 29x^4 - 25x^2 + 3,$$
$$\alpha(G_{3,5}) = x^9 - 11x^7 + 38x^5 - 46x^3 + 14x,$$
$$\alpha(G_{3,6}) = x^{10} - 12x^8 + 48x^6 - 75x^4 + 39x^2 - 3.$$

From the above we see that the coefficients fulfill the Fibonacci recursion (3), i.e., these graphs form a sequence of Fibonacci graphs.

The following is again generally true.

Proposition 2: Let G_1 be a graph, and v_1 and v_2 two isolated vertices. Let $e = \{i,i+1\}$ be an edge of G_1, let G_2 be a graph obtained by removing the edge e from G_1 and inserting edges $\{i,v_1\}$, $\{v_1,i+1\}$, and let G_3 be a graph obtained by removing the edge e from G_1 and inserting a triplet of edges $\{i,v_1\}$, $\{v_1,v_2\}$, $\{v_2,i+1\}$. Then the graphs G_1, G_2, G_3 form a sequence of Fibonacci graphs.

Proof: Let M_1 be the set of all (unordered) k-tuples of nonadjacent edges in G_1 [the cardinality of M_1 is $p(G_1,k)$], let M_2 be the set of all $(k+1)$-tuples of nonadjacent edges in G_2 [the cardinality of M_2 *is* $p(G_2,k)$], and let M_3 be the set of all $(k+1)$-tuples of nonadjacent edges in G_3 [the cardinality of M_3 is $p(G_3,k)$]. Let us split the set M_3 into five subsets: M_{31} (with those elements of M_3 which contain edges $\{i,v_1\}$ and $\{v_2,i+1\}$), M_{32} (with those elements of M_3 which contain edge $\{v_1,v_2\}$), M_{33} (with those elements of M_3 which contain edge $\{i,v_1\}$ and not edge $\{v_2,i+1\}$), M_{34} (with those elements of M_3 which contain edge $\{v_2,i+1\}$ and not edge $\{i,v_1\}$), and finally M_{35} (with the remaining elements of M_3).

Let us take a $(k+1)$-tuple from M_{31}. This does not contain any edge adjacent to i or $i+1$ other than edges $\{i,v_1\}$ and $\{v_2,i+1\}$, so that by removing edges $\{i,v_1\}$ and $\{v_2,i+1\}$, and adding edge $\{i,i+1\}$ we obtain an element of M_1. Conversely, let us take an element from M_1 which contains the edge $\{i,i+1\}$. By removing it and adding edges $\{i,v_1\}$ and $\{v_2,i+1\}$, we produce an element of M_{31}. Similarly, by removing edge $\{v_1,v_2\}$ from an element of M_{32}, we obtain an element of M_1 which does not contain edge $\{i,i+1\}$ and vice versa, so that there is a 1–1 correspondence between $M_{31} \cup M_{32}$ and M_1. By an analogous procedure it is easy to show that there is also a 1–1 correspondence between $M_{33} \cup M_{34} \cup M_{35}$ and M_2. Q.E.D.

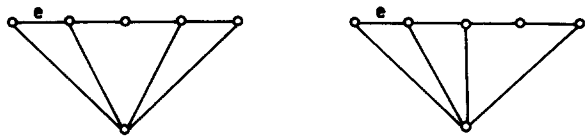
From Propositions 1 and 2 the following corollary follows by computation.

Corollary: Let H be a graph, $v \in V(H)$ and $G_1 \in G_{n,i}(H,v)$; $i = 1,...,k$, where $k \geq 3$ and n are arbitrary integers. Then the graphs G_i form a sequence of Fibonacci graphs.

III. REMARK OF THE FIBONACCI GRAPHS WITHOUT IDENTICAL MATCHING POLYNOMIALS

We now ask what will happen if we consider linear chains instead of centered cycles? The family of graphs thus obtained does not have the same matching polynomial (see the example below), but again forms a sequence of Fibonacci graphs (see Proposition 3).

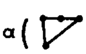
Example: Two graphs G_1 and G_2 obtained by connecting a vertex with four vertices of L_5 have different matching polynomials:



G_1 G_2

$\alpha(G_1) = \alpha($  $) - \alpha($  $)$

$= \alpha($  $) - (L_4 - x^2)$,

$\alpha(G_2) = \alpha($  $) - \alpha($  $)$

$= \alpha($  $) - (L_4 - L_2)$.

Proposition 3: Let H be a graph, $v \in V(H)$, and v_1 and v_2 two isolated vertices. Let $G_1 = H$, G_2 be obtained from H adding an edge $\{v,v_1\}$, and let G_3 be obtained from H by adding edges $\{v,v_1\}$ and $\{v_1,v_2\}$. Then the graphs G_1, G_2, G_3 form a sequence of Fibonacci graphs.

Proof: Let us take a $(k+1)$-tuple of nonadjacent edges from G_3 containing the edge $\{v_1,v_2\} = e$. By removing the edge e we obtain a k-tuple of nonadjacent edges, none of which is adjacent to v_1, i.e., a k-tuple of nonadjacent edges from G_1 [conversely taking a k-tuple of nonadjacent edges from G_1 and adding e to it, we obtain a $(k+1)$-tuple from G_3].

Let us now take a $(k+1)$-tuple of nonadjacent edges from G_3 that does not contain e. But this is also a $(k+1)$-tuple of nonadjacent edges from G_2. Thus, the set of all $(k+1)$-tuples of nonadjacent edges from G_3 is of the same cardinality as the union of the set of all k-tuples of nonadjacent edges from G_1 and $(k+1)$-tuples of nonadjacent edges from G_2.

IV. CONCLUSION

If we define in a certain intuitive sense the subtraction of a subgraph from a graph, we can sum up Proposition 2, its Corollary, and Proposition 3 in the following way: Whenever we have a sequence of graphs $G_1, G_2,...,G_n$ such that there exists for any three graphs G_i, G_{i+1}, G_{i+2} a common subgraph (it can be G_i), whereby subtraction of the subgraph leads to a 3-tuple of either centered cycles (their lengths differing by 1) or linear chains (their lengths also differing by 1), then the sequence $G_1, G_2,...,G_n$ is a sequence of Fibonacci graphs.

We also point out that the recursion formula for the Fibonacci numbers is related to a recursion formula recently introduced[18] for a topological function $\sigma(G)$, defined as the cardinality of the graph topology of G, and equal to the number of stable sets of G. In addition, the function $\sigma(G)$ is quite sensitive to the details of graph structure, particularly to the extents of branching and cyclization. In this respect it is similar to Hosoya's topological index,[6] which is equal to the number of stable edge sets of G; i.e., it is σ of the line graphs of G.

ACKNOWLEDGMENTS

S. El-Basil and P. Křivka would like to thank the theoretical Chemistry Group, Department of Physical Chemistry, The Rugjer Bošković Institute, Zagreb for financial support and generous hospitality which enabled the authors to meet in Zagreb for the final preparation of the manuscript.

[1]C. D. Godsil and I. J. Gutman, Graph Theory **5**, 137 (1981).
[2]O. J. Heilmann and E. H. Lieb, Phys. Rev. Lett. **24**, 1412 (1970).
[3]H. Kunz, Phys. Lett. A **32**, 311 (1970).
[4]C. Gruber and H. Kunz, Commun. Math. Phys. **22**, 133 (1971).
[5]O. J. Heilmann and E. H. Lieb, Commun. Math. Phys. **25**, 190 (1972).
[6]H. Hosoya, Bull. Chem. Soc. Jpn. **44**, 2332 (1971).
[7]I. Gutman, Publ. Math. Inst. (Belgrade) **22**, 63 (1977).
[8]E. J. Farrell, J. Comb. Theory Ser. B**27**, 75 (1979).
[9]D. Cvetković, M. Doob, and H. Sachs, *Spectra of Graphs* (Academic, New York, 1980), p. 355.
[10]I. Gutman, M. Milun, and N. Trinajstić, J. Am. Chem. Soc. **99**, 1692 (1977).
[11]I. Gutman and H. Hosoya, Theoret. Chim. Acta **48**, 279 (1978).
[12]B. Mohar and N. Trinajstić, J. Comput. Chem. **3**, 28 (1982).
[13]N. Trinajstić, *Chemical Graph Theory* (CRC, Boca Raton, FL, 1983), Vol. II, Chap. I.
[14]A. T. Balaban and I. Tomescu, Croat. Chem. Acta **57**, 391 (1984).
[15]A. J. Schwenk, Lect. Notes Math. **406**,153 (1977).
[16]H. Hosoya, Fibonacci **11**, 255 (1973).
[17]E. Heilbronner, Helv. Chim. Acta **36**, 170 (1953).
[18]R. E. Merrifield and H. E. Simmons, Theoret. Chim. Acta **55**, 55 (1980).