A general commutator equation in the algebra of spin operators[a)]

Dhirendra Kumar Jha

Department of Physics, Mithila University, Darbhanga 846004, India
(Received 21 March 1977)

An algebraic method of spin operators was developed by Jha and Valatin to solve (a) the Hamiltonian of the isotropic and anisotropic xy model in a one-dimensional lattice of N spin 1/2 particles and (b) the partition function of the Ising model in the absence of magnetic field in two dimensions. The pair of fermion operators used to explain BCS theory in superconductivity were shown to be related to a set of spin operators of Jha and Valatin in a very simple way. Onsager's Lie algebra for diagonalizing the partition function of the Ising model was found to be included within the said commutator algebra of the spin operators. The structure constants of the algebra are so simple as to allow the entire algebra to be casted in one general commutator equation. In the present paper, the author presents the proof of a general equation from which all sets of commutator relationships existing among the elements of the algebra directly follow.

I. INTRODUCTION

A. Set of spin operators

In an attempt to diagonalize the xy Hamiltonian (isotropic and anisotropic both) in one dimension, a set of spin operators $A_l^{xx}, A_l^{yy}, A_l^{xy}$, and A_l^{yx} was introduced by the author and Valatin.[1] These operators are defined as

$$A_l^{xx} = \tfrac{1}{2}\sum_{j=1}^{N} \sigma_j^x \sigma_{j+1}^z \sigma_{j+2}^z \cdots \sigma_{j+l-1}^z \sigma_{j+l}^x, \tag{1.1a}$$

$$A_l^{yy} = \tfrac{1}{2}\sum_{j=1}^{N} \sigma_j^y \sigma_{j+1}^z \sigma_{j+2}^z \cdots \sigma_{j+l-1}^z \sigma_{j+l}^y, \tag{1.1b}$$

$$A_l^{xy} = \tfrac{1}{2}\sum_{j=1}^{N} \sigma_j^x \sigma_{j+1}^z \sigma_{j+2}^z \cdots \sigma_{j+l-1}^z \sigma_{j+l}^y, \tag{1.1c}$$

$$A_l^{yx} = \tfrac{1}{2}\sum_{j=1}^{N} \sigma_j^y \sigma_{j+1}^z \sigma_{j+2}^z \cdots \sigma_{j+l-1}^z \sigma_{j+l}^x, \tag{1.1d}$$

where $\sigma_j^x, \sigma_j^y, \sigma_j^z$ are the Pauli operators defined at the lattice sites $j = 1, 2, 3, \ldots, N-1, N$ of a ring with periodicity conditions $\sigma_{j+N}^x = \sigma_j^x$, $\sigma_{j+N}^y = \sigma_j^y$, $\sigma_{j+N}^z = \sigma_j^z$. The set of operators defined in (1.1) could be briefly written as

$$A_l^{\alpha\beta} = \tfrac{1}{2}\sum_{j=1} \sigma_j^\alpha \sigma_{j+1}^z \sigma_{j+2}^z \cdots \sigma_{j+l-1}^z \sigma_{j+l}^\beta, \tag{1.2}$$

where α, β stand for any of the x, y indices. The mixed set of commutation and anticommutation rules of σ operators are given by

$$[\sigma_j^x, \sigma_{j'}^y] = 2i\delta_{jj'}\sigma_j^z, \quad [\sigma_j^y, \sigma_{j'}^z] = 2i\delta_{jj'}\sigma_j^x, \tag{1.3a}$$

$$[\sigma_j^z, \sigma_{j'}^x] = 2i\delta_{jj'}\sigma_j^y$$

so that

$$\sigma_j^x\sigma_j^y = i\sigma_j^z, \quad \sigma_j^y\sigma_j^z = i\sigma_j^x, \quad \sigma_j^z\sigma_j^x = i\sigma_j^y \tag{1.3b}$$

$$(\sigma_j^x)^2 = (\sigma_j^y)^2 = (\sigma_j^z)^2 = 1,$$

and

$$\{\sigma_j^x, \sigma_j^y\} = \{\sigma_j^y, \sigma_j^z\} = \{\sigma_j^z, \sigma_j^x\} = 0. \tag{1.3c}$$

By introducing the spin raising and lowering operators

$$\sigma_j^+ = \tfrac{1}{2}(\sigma_j^x + i\sigma_j^y), \quad \sigma_j^- = \tfrac{1}{2}(\sigma_j^x - i\sigma_j^y) \tag{1.4}$$

with the associated property

$$\sigma_j^z\sigma_j^+ = \sigma_j^+, \quad \sigma_j^+\sigma_j^z = -\sigma_j^+, \quad \sigma_j^z\sigma_j^- = -\sigma_j^-, \quad \sigma_j^-\sigma_j^z = \sigma_j^-, \tag{1.5a}$$

$$\sigma_j^+\sigma_j^- = \tfrac{1}{2}(1 + \sigma_j^z), \quad \sigma_j^-\sigma_j^+ = \tfrac{1}{2}(1 - \sigma_j^z), \tag{1.5b}$$

$$(\sigma_j^+)^2 = (\sigma_j^-)^2 = 0, \tag{1.5c}$$

the author defined an alternative set of operators A_l^{++}, A_l^{+-}, A_l^{-+}, and A_l^{--} which could be written in the abbreviated form as

$$A_l^{cd} = \sum_{j=1}^{N} \sigma_j^c \sigma_{j+1}^z \sigma_{j+2}^z \cdots \sigma_{j+l-1}^z \sigma_{j+l}^d, \tag{1.6}$$

where the symbols c, d stand for any of the $+, -$ indices. The set (1.6) has the following property:

$$(A_l^{+-})^\dagger = A_l^{+-}, \quad (A_l^{-+})^\dagger = A_l^{-+}, \quad (A_l^{++})^\dagger = A_l^{--} \tag{1.7}$$

The symbol $\dagger$ indicates the adjoint operator.

If we use the notations

$$A_l^{(xy)} = A_l^{xy} + A_l^{yx} \quad \text{and} \quad A_l^{\{yx\}} = A_l^{xy} - A_l^{yx}, \tag{1.8}$$

the two sets of operators (1.2) and (1.6) would be related through the equations

$$A_l^{xx} + A_l^{yy} = A_l^{+-} + A_l^{-+}, \tag{1.9a}$$

$$A_l^{xx} - A_l^{yy} = A_l^{++} + A_l^{--}, \tag{1.9b}$$

$$A_l^{(xy)} = -i(A_l^{++} - A_l^{--}), \tag{1.9c}$$

$$A_l^{\{xy\}} = i(A_l^{+-} - A_l^{-+}). \tag{1.9d}$$

The algebraic properties of these A_l operators were exploited by the author[1] to diagonalize the Hamiltonian of the xy model of a ring of N spin $\frac{1}{2}$ particles, to diagonalize the transfer matrix of a set of spin $\frac{1}{2}$, arranged on a rectangular lattice and interacting scalerly with nearest neighbours in the absence of a magnetic field.

[a)]This work was done by the author during his stay in the department of Physics at Queen Mary College, University of London.

 0022-2488/78/1911-2247$1.00

Recently, Felderhof[2] has used the A_l operators to diagonalize the transfer function of the Zero field free fermion model.

The number operator n_k and the pair operators b_k and b_k^+ used to explain the BCS theory[3,4] are also related to the A_l operators in a very simple way.

We define

$$n_k = a_k a_k^+, \quad b_k = a_k a_{-k}, \quad \text{and} \quad b_k^+ = a_{-k}^+ a_k^+ \tag{1.10}$$

where a_k, a_k^+ are fermion operators. If we further define

$$m_k = m_{-k} = 1 - n_k - n_{-k} \quad \text{and} \quad \overline{m}_k = -\overline{m}_{-k} = n_k - n_{-k}, \tag{1.11}$$

it is easily seen that

$$b_k = \frac{1}{2N}\sum_{l=1}^{2N} A_l^{++}\sin kl, \tag{1.12a}$$

$$b_k^\dagger = \frac{1}{2N}\sum_{l=1}^{2N} A_l^{--}\sin kl, \tag{1.12b}$$

$$m_k = \frac{1}{2N}\sum_{l=1}^{2N}(A_l^{+-} + A_l^{-+})\cos kl, \tag{1.12c}$$

$$\overline{m}_k = \frac{1}{2N}\sum_{l=1}^{2N} i(A_l^{+-} + A_l^{-+})\sin kl, \tag{1.12d}$$

with

$$k = (2\pi/N)\, r, \quad r = \tfrac{1}{2}, 1, \tfrac{3}{2}, 2, \cdots. \tag{1.12e}$$

The lie Algebra of Onsager[5] is generated by starting with the elements A_0^{xx} and A_1^{xx}. In this way he obtained the set of operators A_l^{xx} and $A_l^{(xy)}$ with the relationships:

$$[A_{l'}^{xx}, A_l^{xx}] = -iA_{l-l'}^{(xy)}, \tag{1.13a}$$

$$[A_{l'}^{xx}, A_l^{(xy)}] = 2i(A_{l+l'}^{xx} + A_{l'-l}^{xx}), \tag{1.13b}$$

$$[A_{l'}^{(xy)}, A_l^{(xy)}] = 0. \tag{1.13c}$$

The quantities A_l^{yy} and $A_l^{\{xy\}}$ derived in the more systematic treatment of the author's spin commutator algebra do not appear in Onsager's work. However, because of the relationship (1.19a), A_l^{yy} can be derived from the quantities A_l^{xx}. The relations (1.13a), (1.13b), (1.13c) then coincide with the equations (1.22a), (1.22d), (1.22f) respectively. The set $A_l^{\{xy\}}$ does not appear in Onsager's algebra because the sets A_l^{xy} and A_l^{yx} are not known separately in his work. So while Onsager's algebra contains only (3N-1) independent elements, ours contains (4N-2) independent elements. In fact, the additional set $A_l^{\{xy\}}$ of our algebra is needed to construct projection operators of the eigenstates.

It is further noticed that the pseudospin operators X_k, Y_k, and Z_k as introduced by Onsager obey the following relations:

$$[x_k, y_k] = -2iz_k, \quad [y_k, z_k] = -2ix_k, \quad [z_k, x_k] = -2iy_k. \tag{1.14}$$

These elements come into the spin commutor algebra as

$$x_k + iy_k = \frac{1}{2N}\sum_{l=1}^{2N} A_l^{xx} e^{ikl}, \tag{1.15a}$$

$$x_k - iy_k = \frac{1}{2N}\sum_{l=1}^{2N} A_l^{xx} e^{-ikl}, \tag{1.15b}$$

$$z_k = \frac{i}{2N}\sum_{l=1}^{2N} A_l^{(xy)}\sin kl, \tag{1.15c}$$

and can be identified with the fermion operators as

$$x_k = -m_k, \quad y_k = b_k + b_k^\dagger, \quad \text{and} \quad z_k = i(b_k^\dagger - b_k). \tag{1.16}$$

B. Properties of the A_l operators

Since the A_l operators have a useful periodicity of $2N$,

$$A_{l+2N}^{\alpha\beta} = A_l^{\alpha\beta}, \quad A_{l+2N}^{cd} = A_l^{cd} \tag{1.17}$$

so that the definitions (1.2) and (1.6) initially valid for any positive integer values of l ($l > 0$) could be extended to even negative integer values of l ($l < 0$) by writing

$$A_{-l}^{\alpha\beta} = A_{2N-l}^{\alpha\beta}, \quad A_0^{\alpha\beta} = A_{2N}^{\alpha\beta}, \tag{1.18a}$$

$$A_{-l}^{cd} = A_{2N-l}^{cd}, \quad A_0^{cd} = A_{2N}^{cd}. \tag{1.18b}$$

The following identities, however, hold for each l not equal to a multiple of N:

$$A_{-l}^{xx} = A_l^{yy}, \quad A_{-l}^{xy} = -A_l^{yx}, \tag{1.19a}$$

$$A_{-l}^{(xy)} = -A_l^{(xy)}, \quad A_{-l}^{\{xy\}} = -A_l^{\{xy\}}, \tag{1.19b}$$

$$A_{-l}^{+-} = A_l^{-+}, \quad A_{-l}^{-+} = A_l^{+-}, \tag{1.19c}$$

$$A_{-l}^{++} = -A_l^{++}, \quad A_{-l}^{--} = -A_l^{--}. \tag{1.19d}$$

For $l = 0$, the following identities are true:

$$A_0^{xx} = A_0^{yy} = -\tfrac{1}{2}\sum_{j=1}^{N}\sigma_j^z, \tag{1.20a}$$

$$A_0^{yx} = A_0^{xy} = iN/2, \tag{1.20b}$$

$$A_0^{(xy)} = 0, \quad A_0^{\{xy\}} = -iN, \tag{1.20c}$$

$$A_0^{+-} = -\sum_{j=1}^{N}\tfrac{1}{2}(1+\sigma_j^z), \quad A_0^{-+} = \sum_{j=1}^{N}\tfrac{1}{2}(1-\sigma_j^z), \tag{1.20d}$$

$$A_0^{++} = A_0^{--} = 0. \tag{1.20e}$$

For $l = N$,

$$A_N = -A_0 U, \quad U = \prod_{i=1}^{N}\sigma_i^z\,. \tag{1.21}$$

C. Number of independent elements in A_l

Although we have defined $2N$ operators for each of $A_l^{xx}, A_l^{yy}, A_l^{xy}$, and A_l^{yx}, the existence of relations (1.19) and (1.20) implies all of them are not linearly independent. A check on the number of independent operators shows that there are (N-1) $A_l^{(xy)}$, (N-1) $A_l^{\{xy\}}$, and $2N$ A_l^{xx} or $2N$ A_l^{yy} independent elements. Alternatively, we can say that there are (N-1) independent A_l^{++}, (N-1) independent A_l^{--} and $2N$ independent A_l^{+-} or A_l^{-+}. So altogether we have ($4N-2$) independent elements in this algebra.

D. Commutation algebra of A_l and its significance

The elements of the set A_l^{xx}, A_l^{yy}, A_l^{xy}, and A_l^{yx} generate the following algebra:

$$[A_{l'}^{xx}, A_l^{xx}] = -i(A_{l-l'}^{xy} + A_{l-l'}^{yx}) = -iA_{l-l'}^{(xy)}, \tag{1.22a}$$

$$[A_{l'}^{xx}, A_l^{yy}] = +i(A_{l+l'}^{xy} + A_{l+l'}^{yx}) = iA_{l+l'}^{(xy)}, \tag{1.22b}$$

$$[A_{l'}^{yy}, A_l^{yy}] = +i(A_{l-l'}^{xy} + A_{l-l'}^{yx}) = iA_{l-l'}^{(xy)}, \tag{1.22c}$$

$$[A_{l'}^{xx}, A_l^{xy}] = [A_{l'}^{xx}, A_l^{yx}] = i(A_{l+l'}^{xx} - A_{l-l'}^{yy}), \tag{1.22d}$$

$$[A_{l'}^{yy}, A_l^{xy}] = [A_{l'}^{yy}, A_l^{yx}] = -i(A_{l+l'}^{xx} - A_{l-l'}^{xx}), \tag{1.22e}$$

$$[A_{l'}^{xy}, A_l^{xy}] = [A_{l'}^{yx}, A_l^{yx}] = [A_{l'}^{xy}, A_l^{yx}] = 0. \tag{1.22f}$$

The elements of the alternative set $A_l^{++}, A_l^{--}, A_l^{+-}$, and A_l^{-+} have the corresponding algebra:

$$[A_{l'}^{++}, A_l^{++}] = [A_{l'}^{--}, A_l^{--}] = 0, \tag{1.23a}$$

$$[A_{l'}^{+-}, A_l^{+-}] = [A_{l'}^{-+}, A_l^{-+}] = [A_{l'}^{+-}, A_l^{-+}] = 0, \tag{1.23b}$$

$$[A_{l'}^{++}, A_l^{--}] = (A_{l+l'}^{+-} + A_{l+l'}^{-+}) - (A_{l-l'}^{+-} + A_{l-l'}^{-+}), \tag{1.23c}$$

$$[A_{l'}^{++}, A_l^{+-}] = [A_{l'}^{++}, A_l^{-+}] = A_{l+l'}^{++} - A_{l-l'}^{++}, \tag{1.23d}$$

$$[A_{l'}^{--}, A_l^{+-}] = [A_{l'}^{--}, A_l^{-+}] = -A_{l+l'}^{--} + A_{l-l'}^{--}. \tag{1.23e}$$

It is further seen that $A_{l'}^{\{xy\}}$ commutes with each elements of the set A_l^{xx}, A_l^{yy} and A_l^{xy}. Also,

$$[A_{l'}^{\{xy\}}, \tfrac{1}{2}(A_l^{+-} + A_l^{-+})] = 0, \tag{1.24a}$$

$$[A_{l'}^{\{xy\}}, A_l^{++}] = 0, \tag{1.24b}$$

$$[A_{l'}^{\{xy\}}, A_l^{--}] = 0. \tag{1.24c}$$

The introduction of the symmetric and the antisymmetric operators $A_l^{(xy)}$ and $A_l^{\{xy\}}$ reduce the algebra considerably.

II. GENERAL COMMUTATOR EQUATION FOR THE ELEMENTS OF THE ALGEBRA

The fact that there are only (4N-2) independent elements in the spin commutator algebra indicates that there exists considerable symmetry in structure constants of the algebra.

We introduce an operator $U^z(l,m)$ such that

$$U^z(l,m) = \sigma_l^z \sigma_{l+1}^z \cdots \sigma_{m-1}^z \sigma_m^z \quad \text{if } l < m, \qquad = \sigma_l^z \quad \text{if } l = m. \tag{2.1}$$

For $l,\ l' \geq 1$, the set of operators $A_l^{\alpha\beta}$ or $A_l^{\gamma\delta}$ could be written in the form:

$$A_{l'}^{\alpha\beta} = \tfrac{1}{2}\sum_{j'=1}^{N} \sigma_{j'}^{\alpha} U^z(j'+1, j'+l'-1)\sigma_{j'+l'}^{\beta},$$
$$A_l^{\gamma\delta} = \tfrac{1}{2}\sum_{j=1}^{N} \sigma_j^{\gamma} U^z(j+1,\ j+l-1)\sigma_{j+l}^{\delta}, \tag{2.2}$$

where $j, j' = 1, 2, \ldots, N$ are dummy indices and the symbols $\alpha, \beta, \gamma, \delta$ stand for any of the x, y indices. If we impose the conditions on l, l' to be $1 \leq l' \leq l-1$, then in this domain the entire relationships (1.22) could be put in one general commutator equation:

$$[A_{l'}^{\alpha\beta}, A_l^{\gamma\delta}] = c^{\beta\gamma} A_{l+l'}^{\alpha\delta} + c^{\alpha\delta} A_{l+l'}^{\gamma\beta} + a^{\alpha\gamma}\theta_{\beta} A_{l-l'}^{\bar{\beta}\delta} + a^{\beta\delta}\theta_{\alpha} A_{l-l'}^{\gamma\bar{\alpha}}, \tag{2.3a}$$

where the structure constants are defined as

$$c^{xy} = i, \quad c^{yx} = -i, \quad c^{xx} = c^{yy} = 0,$$
$$a^{xy} = a^{yx} = 0, \quad a^{xx} = a^{yy} = 1, \tag{2.3b}$$
$$\theta_x = -i, \quad \theta_y = i.$$

The indices with bar imply $\bar{x} = y$ and $\bar{y} = x$.

III. PROOF OF THE GENERAL COMMUTATOR EQUATION

In order to prove the general equation (2.3), we state two lemmas and first provide their proof as follows:

Lemma: For $l' \geq 1$,

$$\begin{aligned}
[A_{l'}^{\alpha\beta}, \sigma_j^{\gamma}] &= c^{\alpha\gamma}\{U^z(j, j+l'-1)\sigma_{j+l'}^{\beta}\} \\
&\quad - \sigma_j^{\gamma}\{\sigma_{j-1}^{\alpha} U^z(j, j+l'-2)\sigma_{j+l'-1}^{\beta} \\
&\quad + \sigma_{j-2}^{\alpha} U^z(j-1,\ j+l'-3)\sigma_{j+l'-2}^{\beta} \\
&\quad + \sigma_{j-3}^{\alpha} U^z(j-2, j+l'-4)\sigma_{j+l'-3}^{\beta} \\
&\quad + \cdots + \sigma_{j-l'+3}^{\alpha} U^z(j-l'+4, j+2)\sigma_{j+3}^{\beta} \\
&\quad + \sigma_{j-l'+2}^{\alpha} U^z(j-l'+3, j+1)\sigma_{j+2}^{\beta} \\
&\quad + \sigma_{j-l'+1}^{\alpha} U^z(j-l'+2, j)\sigma_{j+1}^{\beta}\} \\
&\quad + c^{\beta\gamma}\{\sigma_{j-l'}^{\alpha} U^z(j-l'+1, j)\}.
\end{aligned} \tag{3.1}$$

Proof: By definition

$$\begin{aligned}
[A_{l'}^{\alpha\beta}, \sigma_j^{\gamma}]
&= \tfrac{1}{2}\sum_{j'=1}^{N} [\sigma_{j'}^{\alpha}\sigma_{j'+1}^{z}\sigma_{j'+2}^{z}\cdots\sigma_{j'+l'-1}^{z}\sigma_{j'+l'}^{\beta}, \sigma_j^{\gamma}] \\
&= \tfrac{1}{2}\sum_{j'=1}^{N} \{[\sigma_{j'}^{\alpha}, \sigma_j^{\gamma}] U^z(j'+1, j'+l'-1)\sigma_{j'+l'}^{\beta} \\
&\quad + \sigma_{j'}^{\alpha}[\sigma_{j'+1}^{z}, \sigma_j^{\gamma}] U^z(j'+2, j'+l'-1)\sigma_{j'+l'}^{\beta} \\
&\quad + \sigma_{j'}^{\alpha} U^z(j'+1, j'+1)[\sigma_{j'+2}^{z}, \sigma_j^{\gamma}] U^z(j'+3, j'+l'-1)\sigma_{j'+l'}^{\beta} \\
&\quad + \cdots + \sigma_{j'}^{\alpha} U^z(j'+1, j'+l'-3)[\sigma_{j'+l'-2}^{z}, \sigma_j^{\gamma}]\sigma_{j'+l'-1}^{z}\sigma_{j'+l'}^{\beta} \\
&\quad + \sigma_{j'}^{\alpha} U^z(j'+1, j'+l'-2)[\sigma_{j'+l'-1}^{z}, \sigma_j^{\gamma}]\sigma_{j'+l'}^{\beta} \\
&\quad + \sigma_{j'}^{\alpha} U^z(j'+1, j'+l'-1)[\sigma_{j'+l'}^{z}, \sigma_j^{\gamma}]\}.
\end{aligned} \tag{3.2}$$

If we use the relations

$$\tfrac{1}{2}[\sigma_{j'}^{\alpha}, \sigma_j{}^{\gamma}] = c^{\alpha\gamma}\delta_{jj'}\sigma_j^z, \tag{3.3a}$$

and

$$\tfrac{1}{2}[\sigma_{j'}^{z}, \sigma_j^{\gamma}] = -\delta_{jj'}\sigma_j^{\gamma}\sigma_j^{z}, \tag{3.3b}$$

then (3.1) directly follows from (3.2). It would be noted that there are altogether $l'+1$ terms on the rhs of (3.1). Lemma 1 can be extended to hold in the doamin $1 \leq l' \leq l-1$, and we can write

$$[A_{l'}^{\alpha\beta}, \sigma^{\delta}_{j+l}]$$
$$= c^{\alpha\delta}\{U^z(j+l, j+l+l'-1)\sigma^{\beta}_{j+l+l'}\}$$
$$-\{\sigma^{\alpha}_{j+l-1}U^z(j+l, j+l+l'-2)\sigma^{\beta}_{j+l+l'-1}$$
$$+\sigma^{\alpha}_{j+l-2}U^z(j+l-1, j+l+l'-3)\sigma^{\beta}_{j+l+l'-2}$$
$$+\sigma^{\alpha}_{j+l-3}U^z(j+l-2, j+l+l'-4)\sigma^{\beta}_{j+l+l'-3}$$
$$+\cdots+\sigma^{\alpha}_{j+l-l'+3}U^z(j+l-l'+4, j+l+2)\sigma^{\beta}_{j+l+3}$$
$$+\sigma^{\alpha}_{j+l-l'+2}U^z(j+l-l'+3, j+l+1)\sigma^{\beta}_{j+l+2}$$
$$+\sigma^{\alpha}_{j+l-l'+1}U^z(j+l-l'+2, j+l)\sigma^{\beta}_{j+l+1}\}\sigma^{\delta}_{j+l}$$
$$+c^{\beta\delta}\{\sigma^{\alpha}_{j+l-l'}U^z(j+l-l'+1, j+l)\}. \qquad (3.4)$$

Lemma 2: For $1 \leqslant l' \leqslant l-1$

$$[A_{l'}^{\alpha\beta}, U^z(j+1, j+l-1)]$$
$$= C_1\sigma^{\alpha}_j\sigma^{\beta}_{j+l'}U^z(j+l', j+l-1)$$
$$+C_1\{\sigma^{\alpha}_{j-1}U^z(j,j)\sigma^{\beta}_{j+l'-1}U^z(j+l'-1, j+l-1)$$
$$+\sigma^{\alpha}_{j-2}U^z(j-1,j)\sigma^{\beta}_{j+l'-2}U^z(j+l'-2, j+l-1)$$
$$+\sigma^{\alpha}_{j-3}U^z(j-2,j)\sigma^{\beta}_{j+l'-3}U^z(j+l'-3, j+l-1)$$
$$+\cdots+\sigma^{\alpha}_{j-l'+3}U^z(j-l'+4,j)\sigma^{\beta}_{j+3}U^z(j+3, j+l-1)$$
$$+\sigma^{\alpha}_{j-l'+2}U^z(j-l'+3,j)\sigma^{\beta}_{j+2}U^z(j+2, j+l-1)$$
$$+\sigma^{\alpha}_{j-l'+1}U^z(j-l'+2,j)\sigma^{\beta}_{j+1}U^z(j+1, j+l-1)\}$$
$$-C_2\{U^z(j+1, j+l-1)\sigma^{\alpha}_{j+l-1}U^z(j+l, j+l+l'-2)$$
$$\times\sigma^{\beta}_{j+l+l'-1}+U^z(j+1, j+l-2)\sigma^{\alpha}_{j+l-2}U^z(j+l, j+l+l'-3)$$
$$\times\sigma^{\beta}_{j+l+l'-2}+U^z(j+1, j+l-3)\sigma^{\alpha}_{j+l-3}U^z(j+l, j+l+l'-4)$$
$$\times\sigma^{\beta}_{j+l+l'-2}+\cdots+U^z(j+1, j+l-l'+3)\sigma^{\alpha}_{j+l-l'+3}$$
$$\times U^z(j+l, j+l+2)\,\sigma^{\beta}_{j+l+3}+U^z(j+1, j+l-l'+2)$$
$$\times\sigma^{\alpha}_{j+l-l'+2}U^z(j+l, j+l+1)\sigma^{\beta}_{j+l+2}+U^z(j+1, j+l-l'+1)$$
$$\times\sigma^{\alpha}_{j+l-l'+1}U^z(j+l, j+l)\sigma^{\beta}_{j+l+1}\}$$
$$-C_2U^z(j+1, j+l-l')\sigma^{\alpha}_{j+l-l'}\sigma^{\beta}_{j+l} \qquad (3.5)$$

where $C=C_2=1$. (These have been introduced for easy reference.) Note that for a given l', we have only $2l'$ terms on the rhs of (3.5). There are l' terms for each of the coefficients C_1 and C_2.

Proof: We prove this lemma by induction. Assume l' to be fixed. We then work out for different $l \geqslant l'+1$. In the simplest case $l=l'+1$, we have by definition

$$[A_{l'}^{\alpha\beta}, U^z(j+1, j+l')]$$
$$=\left[\tfrac{1}{2}\sum_{j=1}^{N}\sigma^{\alpha}_{j'}U^z(j'+1, j'+l'-1)\sigma^{\beta}_{j'+l'}, U^z(j+1, j+l')\right]$$
$$=\tfrac{1}{2}\sum_{j'=1}^{N}[\sigma^{\alpha}_{j'}U^z(j'+1, j'+l'-1)\sigma^{\beta}_{j'+l'}, U^z(j+1, j+l')]$$
$$=\tfrac{1}{2}\sum_{j'=1}^{N}[\sigma^{\alpha}_{j'}, U^z(j+1, j+l')]U^z(j'+1, j'+l'-1)\sigma^{\beta}_{j'+l'}$$
$$+\tfrac{1}{2}\sum_{j'=1}^{N}\sigma^{\alpha}_{j'}U^z(j'+1, j'+l'-1)[\sigma^{\beta}_{j'+l'}, U^z(j+l, j+l')], \qquad (3.6)$$

since $[U^z(j'+1, j'+l'-1), U^z(j+1, j+l)]=0$. Further,

$$\tfrac{1}{2}\sum_{j'=1}^{N}[\sigma^{\alpha}_{j'}, U^z(j+1, j+l')]U^z(j'+1, j'+l'-1)\sigma^{\beta}_{j'+l'}$$
$$=\tfrac{1}{2}\sum_{j'=1}^{N}\{[\sigma^{\alpha}_{j'}, \sigma^z_{j+1}]U^z(j+2, j+l')U^z(j'+1, j'+l'-1)\sigma^{\beta}_{j'+l'}$$
$$+\sigma^z_{j+1}[\sigma^{\alpha}_{j'}, \sigma^z_{j+2}]U^z(j+3, j+l')U^z(j'+1, j'+l'-1)\sigma^{\beta}_{j'+l'}$$
$$+U^z(j+1, j+2)[\sigma^{\alpha}_{j'}, \sigma^z_{j+3}]U^z(j+4, j+l')$$
$$\times U^z(j'+1, j'+l'-1)\,\sigma^{\beta}_{j'+l'}+\cdots+U^z(j+1, j+l'-3)$$
$$\times[\sigma^{\alpha}_{j'}, \sigma^z_{j+l'-2}]U^z(j+l'-1, j+l')U^z(j'+1, j'+l'-1)$$
$$\times\sigma^{\beta}_{j'+l'}+U^z(j+1, j+l'-2)[\sigma^{\alpha}_{j}, \sigma^z_{j+l'-1}]\sigma^z_{j+l'}$$
$$\times U^z(j'+1, j'+l'-1)\,\sigma^{\beta}_{j'+l'}+U^z(j+1, j+l'-1)$$
$$\times[\sigma^{\alpha}_{j'}, \sigma^z_{j+l'}]U^z(j'+1, j'+l'-1)\sigma^{\beta}_{j'+l'}. \qquad (3.7)$$

If we use the relation

$$\tfrac{1}{2}[\sigma^{\alpha}_j, \sigma^z_j] = -\sigma^z_j\sigma^{\alpha}_j \qquad (3.8)$$

in (3.7), then all the l' terms for the co-efficients C_2 of (3.5) with $l=l'+1$ are reproduced. It should be noted that the terms in the coefficients C_2 of (3.5) occur in the reverse order to the terms of (3.7). Hence the first term of (3.7) is the last term of C_2 in (3.5) and vice versa.

We can similarly expand the terms of

$$\tfrac{1}{2}\sum_{j'=1}^{N}\sigma^{\alpha}_{j'}U^z(j'+1, j'+l'-1)[\sigma^{\beta}_{j'+l'}, U^z(j+1, j+l')].$$

Using the relation

$$\tfrac{1}{2}[\sigma^{\beta}_j, \sigma^z_j] = \sigma^{\beta}_j\sigma^z_j,$$

all the terms for the coefficients C_1 of (3.5) (with $l=l'+1$ are reproduced. This time the sequence of terms in the preceeding expression and those of C_1 in (3.5) remain the same.

Assuming (3.5) to be correct for l, we try for $l+1$ (l' being fixed) and thus evaluate $[A_{l'}^{\alpha\beta}, U^z(j+1, j+l)]$. By definition

$$[A_{l'}^{\alpha\beta}, U^z(j+1, j+l)] = [A_{l'}^{\alpha\beta}, U^z(j+1, j+l-1)\sigma^z_{j+l}]$$
$$=[A_{l'}^{\alpha\beta}, U^z(j+1, j+l-1)]\sigma^z_{j+l}+U^z(j+1, j+l-1)$$
$$\times[A_{l'}^{\alpha\beta}, \sigma^z_{j+l}]$$
$$=[A_{l'}^{\alpha\beta}, U^z(j+1, j+l-1)]\sigma^z_{j+l}+\tfrac{1}{2}U^z(j+1, j+l-1)$$
$$\times\sum_{j'=1}^{N}[\sigma^{\alpha}_{j'}U^z(j'+1, j'+l'-1)\sigma^{\beta}_{j'+l'}, \sigma^z_{j+l}]$$

$$= [A_{l'}^{\alpha\beta}, U^z(j+1, j+l-1)]\sigma_{j+l}^z + \tfrac{1}{2}U^z(j+1, j+l-1)$$
$$\times \sum_{j'=1}^{N} [\sigma_{j'}^{\alpha}, \sigma_{j+l}^z] U^z(j'+1, j'+l'-1)\sigma_{j'+l'}^{\beta}$$
$$+ \tfrac{1}{2}U^z(j+1, j+l-1) \sum_{j'=1}^{N} \sigma_{j'}^{\alpha} U^z(j'+1, j'+l'-1)$$
$$\times [\sigma_{j'+l'}^{\beta}, \sigma_{j+l}^z]$$
$$= [A_{l'}^{\alpha\beta}, U^z(j+1, j+l-1)]\sigma_{j+l}^z - U^z(j+1, j+l-1)$$
$$\times \sigma_{j+l}^z \sigma_{j+l}^{\alpha} U^z(j+l+1, j+l+l'-1)\sigma_{j+l+l'}^{\beta}$$
$$- U^z(j+1, j+l-1)\sigma_{j+l-l'}^{\alpha} U^z(j+l-l'+1, j+l-1)$$
$$\times \sigma_{j+l}^z \sigma_{j+l}^{\beta}, \tag{3.9}$$

where we have used the relation

$$\tfrac{1}{2}[\sigma_{j'}^{\alpha}, \sigma_j^z] = -\delta_{jj'}\sigma_j^z\sigma_j^{\alpha} \tag{3.10}$$

Assuming the validity of (3.5), we can rewrite Eq. (3.9) as

$$[A_{l'}^{\alpha\beta}, U^z(j+1, j+l)]$$
$$= C_1\sigma_j^{\alpha}\sigma_{j+l'}^{\beta} U^z(j+l', j+l)$$
$$+ C_1\{\sigma_{j-1}^{\alpha} U^z(j,j)\sigma_{j+l'-1}^{\beta} U^z(j+l'-1, j+l)$$
$$+ \sigma_{j-2}^{\alpha} U^z(j-1, j)\sigma_{j+l'-2}^{\beta} U^z(j+l'-2, j+l)$$
$$+ \sigma_{j-3}^{\alpha} U^z(j-2, j)\sigma_{j+l'-3}^{\beta} U^z(j+l'-3, j+l) + \cdots$$
$$+ \sigma_{j-l'+2}^{\alpha} U^z(j-l'+3, j)\sigma_{j+2}^{\beta} U^z(j+2, j+l)$$
$$+ \sigma_{j-l'+1}^{\alpha} U^z(j-l'+2, j)\sigma_{j+1}^{\beta} U^z(j+1, j+l)\}$$
$$- C_2\{U^z(j+1, j+l)\sigma_{j+l}^{\alpha} U^z(j+l+1, j+l+l'-1)\sigma_{j+l+l'}^{\beta}$$
$$+ U^z(j+1, j+l-1)\sigma_{j+l-1}^{\alpha} U^z(j+l+1, j+l+l'-2)$$
$$\times \sigma_{j+l+l'-1}^{\beta} + U^z(j+1, j+l-2)\sigma_{j+l-2}^{\alpha}$$
$$\times U^z(j+l+1, j+l+l'-3)\sigma_{j+l+l'-2}^{\beta} + \cdots$$
$$+ U^z(j+1, j+l-l'+2)\sigma_{j+l-l'+2}^{\alpha} U^z(j+l+1, j+l+1)$$
$$\times \sigma_{j+l+2}^{\beta}\}$$
$$- C_2 U^z(j+1, j+l-l'+1)\sigma_{j+l-l'+1}^{\alpha}\sigma_{j+l+1}^{\beta}$$
$$- C_2 U^z(j+1, j+l-l')\sigma_{j+l-l'}^{\alpha}\sigma_{j+l}^{\beta}\sigma_{j+l}^z$$
$$- C_2 U^z(j+1, j+l-1)\sigma_{j+l-l'}^{\alpha} U^z(j+l-l'+1, j+l-1)$$
$$\times \sigma_{j+l}^z\sigma_{j+l}^{\beta}. \tag{3.11}$$

As $C_2 = 1$, we have kept the second and third terms of (3.9) as the first and the last terms in the bracket of the coefficient C_2 of (3.11). All the other terms in (3.11) arise from $[A_{l'}^{\alpha\beta}, U^z(j+1, j+l-1)]\sigma_{j+l}^z$. The last two terms of (3.11) cancel since

$$\sigma_{j+l}^z\sigma_{j+l}^{\beta} = -\sigma_{j+l}^{\beta}\sigma_{j+l}^z$$

and

$$U^z(j+1, j+l-1)\sigma_{j+l-l'}^{\alpha} U^z(j+l-l'+1, j+l-1)$$
$$= U^z(j+1, j+l-l')\sigma_{j+l-l'}^{\alpha}.$$

We thus find only $2l'$ terms left in Eq. (3.11) which are exactly those obtainable from (3.5) provided l is replaced by $l+1$. Thus what is true for l is also true for $l+1$ in (3.5). Hence the proof.

Proof of the general equation (2.3): By definition

$$[A_{l'}^{\alpha\beta}, A_l^{\gamma\delta}]$$
$$= \left[A_{l'}^{\alpha\beta}, \tfrac{1}{2}\sum_{j=1}^{N}\sigma_j^{\gamma} U^z(j+1, j+l-1)\sigma_{j+l}^{\delta}\right]$$
$$= \tfrac{1}{2}\sum_{j=1}^{N}\{[A_{l'}^{\alpha\beta}, \sigma_j^{\gamma}]U^z(j+1, j+l-1)\sigma_{j+l}^{\delta}$$
$$+ \sigma_j^{\gamma}[A_{l'}^{\alpha\beta}, U^z(j+l, j+l-1)]\sigma_{j+l}^{\delta}$$
$$+ \sigma_j^{\gamma} U^z(j+l, j+l-1)[A_{l'}^{\alpha\beta}, \sigma_{j+l}^{\delta}]\}. \tag{3.12}$$

We can see through (3.1) and (3.4) that each of the commutator $[A_{l'}^{\alpha\beta}, \sigma_j^{\gamma}]$ and $[A_{l'}^{\alpha\beta}, \sigma_{j+l}^{\delta}]$ produces $l'+1$ terms. Similarly it can be seen through (3.5) that the commutator $[A_{l'}^{\alpha\beta}, U^z(j+1, j+l-1)]$ gives rise to $2l'$ terms. We can further check through equations (3.1), (3.4), and (3.5) that, within the sum of the rhs of (3.12), $(l'-1)$ terms of the first line together with the $(l'-1)$ terms of the third line cancel with, $(2l'-2)$ terms of the second line. The effective contributions are from the first and the $(l'+1)$th terms of $[A_{l'}^{\alpha\beta}, \sigma_j^{\gamma}]$, the first and $2l'$ th terms of $[A_{l'}^{\alpha\beta}, U^z(j+1, j+l-1)]$ and the first and the $(l'+1)$th terms of $[A_{l'}^{\alpha\beta}, \sigma_{j+l}^{\delta}]$. There are thus only six terms left within the summation of the rhs of (3.12). We thus write

$$[A_{l'}^{\alpha\beta}, A_l^{\gamma\delta}] = c^{\beta\gamma}A_{l+l'}^{\alpha\delta} + c^{\alpha\delta}A_{l+l'}^{\gamma\beta}$$
$$+ \tfrac{1}{2}c^{\alpha\gamma}\sum_{j=1}^{N}\sigma_j^z\sigma_{j+l'}^{\beta} U^z(j+l', j+l-1)\sigma_{j+l}^{\delta}$$
$$+ \tfrac{1}{2}c^{\beta\delta}\sum_{j=1}^{N}\sigma_j^{\gamma} U^z(j+1, j+l-l')\sigma_{j+l-l'}^{\alpha}\sigma_{j+l}^z$$
$$+ \tfrac{1}{2}\sum_{j=1}^{N}\sigma_j^{\gamma}\sigma_j^{\alpha}\sigma_{j+l'}^{\beta} U^z(j+l', j+l-1)\sigma_{j+l}^{\delta}$$
$$- \tfrac{1}{2}\sum_{j=1}^{N}\sigma_j^{\gamma} U^z(j+1, j+l-l')\sigma_{j+l-l'}^{\alpha}\sigma_{j+l}^{\beta}\sigma_{j+l}^{\delta}. \tag{3.13}$$

The third and the fifth terms in (3.13) are combined as follows: Since

$$c^{\alpha\gamma}\sigma_j^z = \tfrac{1}{2}(\sigma_j^{\alpha}\sigma_j^{\gamma} - \sigma_j^{\gamma}\sigma_j^{\alpha}),$$

$$\tfrac{1}{2}c^{\alpha\gamma}\sum_{j=1}^{N}\sigma_j^z\sigma_{j+l'}^{\beta} U^z(j+l', j+l-1)\sigma_{j+l}^{\delta}$$
$$+ \tfrac{1}{2}\sum_{j=1}^{N}\sigma_j^{\gamma}\sigma_j^{\alpha}\sigma_{j+l'}^{\beta} U^z(j+l', j+l-1)\sigma_{j+l}^{\delta}$$
$$= \tfrac{1}{2}\sum_{j=1}^{N}\tfrac{1}{2}(\sigma_j^{\alpha}\sigma_j^{\gamma} - \sigma_j^{\gamma}\sigma_j^{\alpha} + 2\sigma_j^{\gamma}\sigma_j^{\alpha})\sigma_{j+l'}^{\beta} U^z(j+l', j+l-1)\sigma_{j+l}^{\delta}$$

$$= \tfrac{1}{2}\sum_{j=1}^{N} \tfrac{1}{2}(\sigma_j^\alpha\sigma_j^\gamma + \sigma_j^\gamma\sigma_j^\alpha)\sigma_{j+l'}^\beta U^z(j+l', j+l-1)\sigma_{j+l}^\delta$$

$$= \tfrac{1}{2}a^{\alpha\gamma}\theta_\beta \sum_{j=1}^{N} \sigma_{j+l'}^{\bar\beta} U^z(j+l'+1, j+l-1)\sigma_{j+l}^\delta$$

$$= a^{\alpha\gamma}\theta_\beta A_{l-l'}^{\bar\beta\delta}, \qquad (3.14)$$

where we have defined

$$a^{\alpha\gamma} = \tfrac{1}{2}(\sigma^\alpha\sigma^\gamma + \sigma^\gamma\sigma^\alpha) = 1, \quad \text{if } \alpha = \gamma,$$
$$= 0, \quad \text{if } \alpha \neq \gamma, \qquad (3.15a)$$

$$\theta_\beta = -i, \quad \text{if } \beta = x,$$
$$= +i, \quad \text{if } \beta = y, \qquad (3.15b)$$

and

$$\bar\alpha, \bar\beta = y, \quad \text{if } \alpha, \beta = x,$$
$$= x, \quad \text{if } \alpha, \beta = y. \qquad (3.15c)$$

The fourth and the sixth terms in (3.13) can be similarly combined. They give the expression $a^{\beta\delta}\,\theta_\alpha A_{l-l'}^{\gamma\bar\alpha}$.

Equation (3.13) can, therefore, be written in the final form as

$$[A_{l'}^{\alpha\beta}, A_l^{\gamma\delta}] = c^{\beta\gamma}A_{l+l'}^{\alpha\delta} + c^{\alpha\delta}A_{l+l'}^{\gamma\beta} + a^{\alpha\gamma}\theta_\beta A_{l-l'}^{\bar\beta\delta} + a^{\beta\delta}\,\theta_\alpha A_{l-l'}^{\gamma\bar\delta},$$

which is Eq. (2.3).

The structure constants $c^{\beta\gamma}$ or $c^{\alpha\delta}$ are derived by comparing Eqs. (3.3a) with (1.4a), (1.4b), (1.4c), while $a^{\alpha\gamma}$ or $a^{\beta\delta}$ and θ_β are obtainable from (3.15). Hence the proof.

IV. DERIVATION OF THE COMMUTATOR EQUATIONS OF THE SET OF A_l OPERATORS

All the commutator equations (1.22a)–(1.22f) stem from the general equation (2.3). This derivation is restricted in the interval $1 \leq l \leq l'-1$. However, we can interchange the indices l and l' and exploit the relations (1.19) and (1.20) to show that they also hold good for $1 \leq l' \leq l-1$. The case $l = l'$ is trivial. Hence the commutator relations in (1.22) are valid for any $l, l' > 0$.

Example: Let $\alpha = \beta = \gamma = \delta = x$. We have from Eq. (2.3)

$$[A_{l'}^{xx}, A_l^{xx}] = -i(A_{l-l'}^{yx} + A_{l-l'}^{xy})$$
$$= -iA_{l-l'}^{(xy)} \quad \text{if } 1 \leq l \leq l'-1. \qquad (4.1)$$

If we interchange l and l' and use the relation

$$A_{l-l'}^{(xy)} = -A_{l'-l}^{(xy)},$$

then

$$[A_l^{xx}, A_{l'}^{xx}] = -iA_{l'-l}^{(xy)} = +iA_{l-l'}^{(xy)} \quad \text{if } 1 \leq l' \leq l-1,$$

or

$$[A_{l'}^{xx}, A_l^{xx}] = -iA_{l-l'}^{(xy)} \quad \text{if } 1 \leq l' \leq l-1. \qquad (4.2)$$

Hence, combining Eq. (4.1) and (4.2), we can write:

$$[A_{l'}^{xx}, A_l^{xx}] = -iA_{l-l'}^{(xy)}$$

for all $l, l' > 0$. (The case $l = l'$ is trivial and is excluded in these proofs). By suitable choice of α, β, γ, and δ, other relations in (1.22) can be easily derived.

The commutator equations in (1.23) are derivable from (1.22) if we further exploit the relations (1.9) and (1.19). For example,

$$[A_{l'}^{++}, A_l^{++}] = \tfrac{1}{4}[A_{l'}^{xx} - A_{l'}^{yy} + iA_{l'}^{(xy)}, A_l^{xx} - A_l^{yy} + iA_l^{(xy)}]. \qquad (4.3)$$

When the commutator on the rhs of (4.3) is expanded and the relations in (1.22) are used, the value is found to be zero. Hence, $[A_{l'}^{++}, A_l^{++}] = 0$.

V. DISCUSSION

The algebra of the A_l operators has been used by the author and Valatin[1] to diagonalize the symmetric and the asymmetric xy Hamiltonians in one dimension, to diagonalize the partition function of the Ising model in two dimensions in the absence of magnetic field. These solutions were also arrived at by Lieb, Schultz, and Maltis[6,7] through the use of fermion and fermion quasiparticle operators. Their method is however, beset with some difficulties. The translational symmetry of fermion operators are broken by the Jordon—Wigner transformation.[8] In higher dimensions even with nearest neighbor interactions, the Hamiltonian involves a polynomial of fermion operators roughly of order $2N$ for a system of N^2 spins. The translational symmetry can be retained only if we consider operators which are composed of an even number of fermion operators. It is possible that because of these fermion operators, their methods remain one dimensional in xy model or two dimensional in the Ising model, and an extension to higher dimensions does not seem possible. The present algebraic approach gets rid of the fermion operators. Felderhof[2] has used the A_l operators to diagonalize the transfer matrix of the zero field free-fermion model. The author is at present looking into the possibility of generalizing the commutator algebra of spin operators in two dimensions which may lead to the solutions of xy model in two dimensions and possibly the Ising model in three dimensions.

[1] D.K. Jha and J.G. Valatin, J. Phys. A 6, 1679 (1973).
[2] B.U. Felderhof, Physica 65, 421 (1973).
[3] J.G. Valatin, Nuovo Cimento 7, 843 (1958).
[4] N.N. Bogolubov, Nuovo Cimento 7, 794 (1958).
[5] L. Onsager, Phys. Rev. 65, 117 (1944).
[6] E. Lieb, T. Schultz, and D. Mattis, Ann. Phys. (N.Y.) 16, 407 (1961).
[7] T.D. Schultz, D.C. Maltis, and E.H. Lieb, Rev. Mod. Phys. 36, 856 (1964).
[8] P. Jordan and E. Wigner, Z. Physik 47, 631 (1928).

A nonlinear scalar field theory in isotropic homogeneous space–time

J. A. Okolowski

221 Irving Street, Chester, Pennsylvania 19013
(Received 22 May 1978)

A classical nonlinear scalar field theory in isotropic homogeneous space–time of uniform negative curvature is considered which admits a singularity-free spatially localized dynamically unstable solution. The associated field energy is obtained as a finite positive quantity only for suitably restricted values of a "size parameter" which measures the degree of spatial localization of the solution. The static flat space–time limit of the present field theory as well as a physically appropriate limitation on the magnitude of the field energy are discussed.

I. INTRODUCTION

With prescribed metric tensor components $g_{\mu\nu}$, a Lagrangian density of the generic form

$$L = \tfrac{1}{2} g^{\mu\nu}\Psi_{,\mu}\Psi_{,\nu} - f, \quad f = f(g^{\mu\nu}, \Psi), \tag{1.1}$$

leads to the field equation

$$(-g)^{-1/2}[g^{\mu\nu}(-g)^{1/2}\Psi_{,\mu}]_{,\nu} + \frac{\partial f}{\partial \Psi} = 0 \tag{1.2}$$

for the admissible real scalar field Ψ. In the presence of isotropic homogeneous space—time geometry characterized by the line element[1]

$$ds^2 = dt^2 - R^2[d\chi^2 + \sinh^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)], \quad R = R(t), \tag{1.3}$$

the field equation (1.2) takes the form

$$\frac{1}{R}\frac{\partial}{\partial t}\left(R^3\frac{\partial\Psi}{\partial t}\right) - \frac{1}{\sinh^2\chi}\frac{\partial}{\partial\chi}\left(\sinh^2 x\frac{\partial\Psi}{\partial\chi}\right) - \frac{1}{\sinh^2\chi\sin\theta}\left[\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Psi}{\partial\theta}\right) + \frac{1}{\sin\theta}\frac{\partial^2\Psi}{\partial\phi^2}\right] + R^2\frac{\partial f}{\partial\Psi} = 0, \tag{1.4}$$

which can be specialized to

$$\frac{1}{R}\frac{\partial}{\partial t}\left(R^3\frac{\partial\Psi}{\partial t}\right) - \frac{1}{\sinh^2\chi}\frac{\partial}{\partial\chi}\left(\sinh^2\chi\frac{\partial\Psi}{\partial\chi}\right) + R^2\frac{\partial f}{\partial\Psi} = 0 \tag{1.5}$$

when the scalar field is spherically symmetric. The latter field equation can be derived from the Lagrangian density

$$L = \tfrac{1}{2}\left(\frac{\partial\Psi}{\partial t}\right)^2 - \frac{1}{2R^2}\left(\frac{\partial\Psi}{\partial\chi}\right)^2 - f. \tag{1.6}$$

Consequently, the field energy (E) associated with a static solution of (1.5) is obtained by evaluating the functional

$$E[\Psi] = 4\pi R^3\int_0^\infty\left\{\frac{1}{2R^2}\left(\frac{\partial\Psi}{\partial\chi}\right)^2 + f\right\}\sinh^2\chi\, d\chi \tag{1.7}$$

at such a solution.

The purpose of the present work is to study the singularity-free, spatially localized, static solution of (1.5) for a specific classical nonlinear model field theory.

II. A SOLVABLE NONLINEAR FIELD THEORY

For a theory based on the Lagrangian density (1.6) with

$$f = -\lambda_0\Psi^2 - \lambda_1\Psi^4 - \lambda_2\Psi^6, \tag{2.1}$$

and time-dependent parameters

$$\lambda_0 \equiv \frac{\alpha_0}{2R^2}, \quad \lambda_1 \equiv \frac{\alpha_1}{4R^2}, \quad \lambda_2 \equiv \frac{\alpha_2}{6R^2}, \tag{2.2}$$

expressed in terms of the positive, dimensionless constants α_0, α_1, and α_2, the field equation (1.5) becomes

$$\frac{1}{R}\frac{\partial}{\partial t}\left(R^3\frac{\partial\Psi}{\partial t}\right) - \frac{1}{\sinh^2\chi}\frac{\partial}{\partial\chi}\left(\sinh^2\chi\frac{\partial\Psi}{\partial\chi}\right) - \alpha_0\Psi - \alpha_1\Psi^3 - \alpha_2\Psi^5 = 0. \tag{2.3}$$

It is easily verified that the singularity-free static solution of (2.3) is given by

$$\Psi_0(\chi) = \left(\frac{2\sigma}{\alpha_1}\right)^{1/2}(\sinh^2\chi + \sigma)^{-1/2}, \tag{2.4}$$

provided that $\alpha_0 = 1$ and that the constant dimensionless "size parameter" σ obeys the relation

$$\sigma = \left(1 + \frac{4\alpha_2}{3\alpha_1^2}\right)^{-1}. \tag{2.5}$$

To compute the field energy (rest mass), Eqs. (2.1) — (2.5) are substituted into (1.7) with the result

$$E = \frac{\pi R}{\alpha_1}Q(\sigma), \tag{2.6}$$

where

$$Q(\sigma) = \sigma^{1/2}(1-\sigma)^{-3/2}(1-2\sigma)\tan^{-1}\left(\frac{1-\sigma}{\sigma}\right)^{1/2} + \sigma(1-\sigma)^{-1}(4\sigma - 3). \tag{2.7}$$

A physically admissible, positive field energy exists only if $0 < \sigma < \sigma_0 \cong 0.142$, the quantity σ_0 being the root of $Q(\sigma) = 0$ on the interval $0 < \sigma < 1$. Therefore, (2.5) leads to the inequality

$$\alpha_2 > \tfrac{3}{4}\left(\frac{1-\sigma_0}{\sigma_0}\right)\alpha_1^2. \tag{2.8}$$

Furthermore, from (2.6) and (2.7) it follows that

$$E \lesssim (0.150)\frac{\pi R}{\alpha_1}, \tag{2.9}$$

the maximum value of the field energy occurring when $\sigma \cong 0.036$, hence,

$$\frac{\partial E}{\partial R} = \frac{E}{R} \lesssim (0.150)\frac{\pi}{\alpha_1}. \tag{2.10}$$

 0022-2488/78/1911-2253$1.00

Thus the requirement $\alpha_1 \gg 1$ ensures that

$$\frac{\partial E}{\partial R} = \frac{E}{R} \ll 1,$$

a condition necessitated by our tacit assumption that the field does not significantly alter the prescribed space — time geometry. In the limiting case of small particle sizes ($\sigma \ll \sigma_0$), the field energy

$$E = \frac{\pi^2}{4}\left(\frac{3}{\alpha_2}\right)^{1/2} R \tag{2.11}$$

exists as a quantity independent of α_1.

Under the coordinate transformation

$$r = r_0 \sinh\chi \tag{2.12}$$

with r_0 a positive constant having the dimensions of length,[2] the solution (2.4) is transformed according to

$$\Psi_0(\chi) = \hat{\Psi}_0(r) = \left(\frac{2\sigma r_0^2}{\alpha_1}\right)^{1/2}(r^2 + r_0^2\sigma)^{-1/2}, \tag{2.13}$$

which takes the limiting form

$$\Phi_0 = \Phi_0(r) = \left(\frac{2Z}{\alpha_1}\right)^{1/2}(r^2 + Z)^{-1/2} \tag{2.14}$$

by virtue of the limits $r_0 \to \infty$, $\sigma \to 0$ performed in such a way that the product $r_0^2\sigma$ remains finite and equal to an arbitrary positive constant Z. In the static, flat space — time limit ($r_0 \to \infty, R/r_0 \to 1$), it follows that (2.14) is the static, particlelike solution of the field equation

$$\frac{\partial^2\Phi}{\partial t^2} - \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\Phi}{\partial r}\right) = \frac{3\alpha_1^2}{4Z}\Phi^5, \tag{2.15}$$

in conformity with a previously investigated, Lorentz-invariant classical field theory[3] of which the present work is a generalization.

III. DYNAMICAL STABILITY

Let us now consider the dynamical stability of (2.4) when the perturbed solution is given by

$$\Psi(\chi, t) = \Psi_0(\chi) + \frac{\xi(\chi)}{\sinh\chi}\eta(t), \tag{3.1}$$

in which the function $\eta(t)$ satisfies the equation[4]

$$\ddot{\eta} + \frac{3\dot{R}}{R}\dot{\eta} + \frac{k^2}{R^2}\eta = 0, \tag{3.2}$$

together with the initial conditions

$$\eta(t_0) = 1, \quad \dot{\eta}(t_0) = 0. \tag{3.3}$$

Moreover, the dimensionless constant k appearing in (3.2) may be either purely real or purely imaginary, and

$$\frac{|\xi(\chi)|}{\sinh\chi} \ll |\Psi_0(\chi)|, \text{ for } \chi > 0.$$

The result of substituting (3.1) into (2.3) and retaining only terms linear in ξ is the eigenvalue equation for k^2 and ξ,

$$\xi''(\chi) + F(\chi)\,\xi(\chi) + k^2\xi(\chi) = 0, \tag{3.4}$$

where

$$F(\chi) = \frac{6\sigma}{(\sinh^2\chi + \sigma)} + \frac{15\sigma(1-\sigma)}{(\sinh^2\chi + \sigma)^2}. \tag{3.5}$$

Equation (3.4) must be supplemented with the appropriate boundary conditions for a singularity-free, localized perturbation

$$\xi(0) = 0, \quad \lim_{\chi\to\infty}\left[\frac{\xi(\chi)}{\sinh\chi}\right] = 0. \tag{3.6}$$

That the $k = 0$ solution of (3.4) with the boundary conditions (3.6) has at least one zero for $\chi > 0$ follows from considerations based on the function

$$G(\chi) \equiv \frac{6(3\sigma - 1)}{(\sinh^2\chi + \sigma)} + \frac{15\sigma(1-\sigma)}{(\sinh^2\chi + \sigma)^2} < F(\chi) \tag{3.7}$$

in combination with the differential equation

$$u''(\chi) + G(\chi)\,u(\chi) - 4u(\chi) = 0, \tag{3.8}$$

which admits the solution

$$u(\chi) = (\sinh^2\chi + \sigma)^{-3/2}\sinh\chi, \tag{3.9}$$

subject to the boundary conditions

$$u(0) = 0, \quad \lim_{\chi\to\infty} u(x) = 0. \tag{3.10}$$

Hence, we infer[5] the existence of an essentially unique solution of (3.4) with no nodes occurring for positive values of χ; the associated minimum, negative eigenvalue k^2 having -4 and $9 - 15/\sigma$ as upper and lower bounds, respectively. Because k is purely imaginary, the perturbation term in (3.1) increases with time[6] in a dynamically unstable manner.

To calculate the approximate minimum value of k^2 we employ the Rayleigh— Ritz procedure. For this purpose it is convenient to introduce the new independent and dependent variables

$$\rho \equiv \cos^{-1}\{\sigma^{1/2}(\sinh^2\chi + \sigma)^{-1/2}\cosh\chi\}, \tag{3.11}$$
$$0 \leq \rho < \rho_0 \equiv \cos^{-1}(\sigma)^{1/2}, \quad \omega(\rho) \equiv (\cos^2\rho - \sigma)^{1/2}\xi(\chi).$$

By means of the latter quantities (3.4) is transformed to the equation

$$\omega''(\rho) + \frac{\sigma(1-\sigma)(1+k^2)\omega(\rho)}{(\cos^2\rho - \sigma)^2} + \frac{6\sigma\omega(\rho)}{(\cos^2\rho - \sigma)} + 16\omega(\rho) = 0, \tag{3.12}$$

which leads to the variational principle

$$\delta\gamma^2 = 0, \quad \gamma^2 \equiv \int_0^{\rho_0}\left\{\omega'(\rho)^2 - \frac{6\sigma\omega(\rho)^2}{(\cos^2\rho - \sigma)} - 16\omega(\rho)^2\right\}d\rho, \tag{3.13}$$

contingent upon the normalization condition

$$\int_0^{\rho_0}\frac{\omega(\rho)^2 d\rho}{(\cos^2\rho - \sigma)^2} = 1, \tag{3.14}$$

the boundary conditions

$$\omega(0) = \omega(\rho_0) = 0, \tag{3.15}$$

and the definition

$$\gamma^2 \equiv \sigma(1-\sigma)(1+k^2). \tag{3.16}$$

We choose a trial function of the form

$$\omega(\rho) = \left(\frac{2}{\rho_0}\right)^{1/2}(\cos^2\rho - \sigma)\left(a\sin\frac{\pi\rho}{\rho_0} + b\sin\frac{2\pi\rho}{\rho_0}\right), \tag{3.17}$$

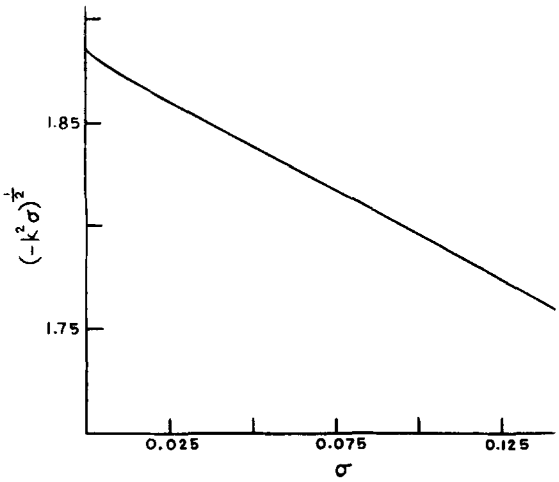

FIG. 1. The approximate value of $(-k^2\sigma)^{1/2}$ as a function of σ for the lowest eigenvalue k^2.

where a and b are variational parameters constrained by (3.14) to satisfy $a^2+b^2=1$. By minimizing the result of combining (3.17) with the definition part of (3.13), and then using (3.16), we obtain the relation shown in Fig. 1. Finally, we have

$$\lim_{\sigma\to 0}(-k^2\sigma)^{1/2} \cong 1.88. \tag{3.18}$$

[1]Units are chosen such that the constant of gravitation and the speed of light in a vacuum are both equal to unity.

[2]R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity*, (McGraw-Hill, New York, 1975), 2nd ed. p. 406.

[3]G. Rosen, J. Math. Phys. **6**, 1269 (1965).

[4]In what follows, we assume that $\dot{R}(t)$ is a continuous function of time for $t \geq t_0$.

[5]For example: R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. I, p. 458.

[6]G. Birkhoff and G.-C. Rota, *Ordinary Differential Equations*, (Blaisdell, Waltham, Mass., 1969), 2nd ed., pp. 39–41.

Energetically stable systems

Bart Demoen, Paul Vanheuverzwijn,[a)] and André Verbeure

Instituut voor Theoretische Fysica, K. U. Leuven, B-3030 Leuven, Belgium
(Received 5 April 1978)

For quantum systems as well as for classical continuous systems energetic stability is defined. It is proved that stability, supplemented with a cluster property, characterizes equilibrium states.

INTRODUCTION

Recently Pusz and Woronowicz[1] considered an equilibrium condition for quantum mechanical systems called passivity.

It is the aim of this paper to introduce what we call energetic stability which is weaker than passivity but nevertheless yields the same results as far as equilibrium is concerned.

In Sec. I we define energetic stability and prove that it is implied by passivity for general quantum systems. We prove that energetic stability also implies a correlation inequality which together with some cluster properties yields KMS states or equilibrium states.

In Sec. II we treat the continuous classical case. The setup of the classical dynamical system is inspired by the one given by Pulvirenti[2] where dynamical stability is discussed. We only give a formal discussion of the perturbed dynamics and start the rigorous discussion from the correlation inequality in order to avoid uninteresting technical conditions. The derivation is the classical analog of the one given in Ref. 1.

I. QUANTUM SYSTEMS

In Ref. 1 one starts from the pair $(\mathcal{A}, \alpha)$ where $\mathcal{A}$ is the C^*-algebra of observables and $\alpha = \{\alpha_t\}_{t\in R}$ is a strongly continuous one-parameter group of automorphisms of $\mathcal{A}$ describing the time evolution. If $x \in \mathcal{A}$ such that the limit

$$\lim_{t\to\infty} \frac{1}{t}[\alpha_t(x) - x] \equiv \delta(x) \tag{1}$$

exists, then $x \in \mathcal{D}(\delta)$, the domain of δ, the infinitesimal generator of the group α. Pusz and Woronowicz considered the system as no longer closed but acted upon through a perturbed dynamics α^h satisfying

$$\frac{d\alpha_t^h(x)}{dt} = \alpha_t^h[\delta(x)] + i\alpha_t^h([h_t, x]), \quad \alpha_0^h(x) = x, \tag{2}$$

where $\{h_t\}_{t\in[0,T]}$ is a differentiable family of self-adjoint elements of $\mathcal{A}$ such that $h_t = 0$ for $t \notin [0, T]$. A state ω on $\mathcal{A}$ is called passive if the work done by the external forces is positive.

In Ref. 3 the relation of this passivity condition with open thermodynamical stability is discussed.

The main idea of the present paper is to investigate the consequences of the energy change under the influence of a sudden (at time $t=0$) external local perturbation $h \in \mathcal{A}$ of the original dynamics. Then we let evolve the system in an isolated way, analogous to Ref. 4. In contradistinction with Ref. 1 where the perturbation h_t as a function of t is arbitrary, we consider its natural evolution, given by

$$h_t = \alpha_t(h), \quad t \geq 0. \tag{3}$$

The energy change in the time interval $[0, t]$ is then

$$\Delta E(t) = \int_0^t ds\, \omega\left[\alpha_s^h\left(\frac{d\,\alpha_s(h)}{ds}\right)\right], \tag{4}$$

where α_t^h is given by (2) [expression (4) is formally equivalent to Ref. 1, (1.8)].

Now we express the tendency of the system to evolve to a stable position and introduce the following notion:

Definition 1.1: A dynamical system $(\mathcal{A}, \alpha, \omega)$ is called *energetically stable* if for all $h = h^* \in \mathcal{A} \cap \mathcal{D}(\delta)$ and all $t \in [0, \epsilon]$, $\epsilon > 0$,

$$\Delta E(t) \leq 0. \tag{5}$$

Notice that this condition expresses the fact that the energy of the system does not increase, which is a natural thing to ask for a stable state. We speak about energetic stability because the system is considered to be thermally isolated (compare with Refs. 5–7). Note also that passive states[1] are defined by an inequality like (5) but with the opposite sign.

For $t \geq 0$ and any $h = h^* \in \mathcal{A} \cap \mathcal{D}(\delta)$ define

$$U_t = 1 + \sum_{n\geq 1} (-i)^n \int_0^{2t} ds_1 \int_0^{s_1} ds_2 \cdots \int_0^{s_{n-1}} ds_n\, \alpha_{s_1}(h) \cdots \alpha_{s_n}(h), \tag{6}$$

where the sum and integrals are in the uniform sense. Note that U_t satisfies:

(i) $U_t \in \mathcal{A} \cap \mathcal{D}(\delta)$, $U_t^* U_t = U_t U_t^* = 1$,

(ii) $\dfrac{dU_t}{dt} = -i\,\alpha_{2t}(h)\,U_t$, $U_{t=0} = 1$,

(iii) $\alpha_t^h(x) = U_t^* \alpha_t(x)\, U_t$, $x \in \mathcal{A}$.

Now we have,

Theorem 1.2: The dynamical system $(\mathcal{A}, \alpha, \omega)$ is energetically stable if and only if for all $t \geq 0$ and $h = h^* \in \mathcal{A} \cap \mathcal{D}(\delta)$

$$\omega\left(U_t^* \frac{\delta}{i} U_t\right) \geq 0, \tag{7}$$

where U_t is defined by (6).

Proof: As in Ref. 1, by partial integration we obtain

$$\Delta E(t) = \omega[\alpha_t^h \alpha_t(h)] - \omega(h) - \int_0^t ds\, \omega[\alpha_s^h \delta\alpha_s(h)].$$

[a)]Aspirant NFWO, Belgium.

 0022-2488/78/1911-2256$1.00

Hence

$$\Delta E(t) = \tfrac{1}{2}\{\omega[U_t^* \alpha_{2t}(h) U_t] - \omega(h)\}. \quad (8)$$

First suppose that the system is stable, then due to the continuity of $t \to \omega\{\alpha_t^*[d\alpha_t(h)/dt]\}$ we have $\omega[\delta(h)] \le 0$. As this inequality is also valid for h replaced by $-h$, $\omega[\delta(h)] = 0$, yielding the time invariance of the state. Using this invariance, we get

$$\omega[U_t^* \alpha_{2t}(h) U_t] - \omega(h) = \omega[I(t)] - 2\omega\left(U_t^* \frac{\delta}{i} U_t\right),$$

where

$$I(t) = U_t^*\left(\frac{2\delta}{i} U_t + [\alpha_{2t}(h), U_t]\right).$$

Using (ii), one derives

$$\frac{d\omega[I(t)]}{dt} = 0.$$

Hence $\omega[I(t)] = \omega[I(0)] = 0$, and (8) becomes

$$\Delta E(t) = -\omega\left(U_t^* \frac{\delta}{i} U_t\right).$$

Therefore, stability implies (7).

Conversely, suppose that (7) holds. Then the above computations can be done in the reverse order if we prove that the time invariance holds. But this is easy from the Taylor expansion of U_t,

$U_t = 1 - iht + O(t^2)$. ■

According to Ref. 1, Theorem 2.1, a system is passive if and only if $\omega(U^*(\delta/i)U) \ge 0$ for all $U \in U_0(\mathcal{A}) \cap D(\delta)$, where $U_0(\mathcal{A})$ is the connected unity component of the group of all unitaries of $\mathcal{A}$ in the norm topology. We have:

Corollary 1.3: A passive system is an energetically stable system. If the system is energetically stable, then for all $h = h^* \in \mathcal{A} \cap D(\delta)$,

$$\omega\left(h \frac{\delta}{i} h\right) \ge 0. \quad (9)$$

Proof: It is clear from (6) that for all $t \ge 0$, $h = h^* \in \mathcal{A} \cap D(\delta)$, and also $U_t \in U_0(\mathcal{A}) \cap D(\delta)$, therefore the first part of the corollary follows from Theorem 2.1 (Ref. 1). The last statement follows also from Theorem 2.1 (Ref. 1) using again the Taylor expansion of U_t. ■

As the inequality (9) is the only condition used in Ref. 1 to derive the KMS condition from passivity, we can formulate, among others, the following result.

Theorem 1.4: Let $(\mathcal{A}, \alpha, \omega)$ be an energetically stable system and let $\{\tau_g\}_{g\in G}$ be a locally compact amenable group of automorphisms of $\mathcal{A}$ commuting with α such that (weakly clustering)

$$\mathcal{M}_g\{\omega[x\tau_g(y)]\} = \omega(x)\,\omega(y), \quad x, y \in \mathcal{A},$$

where $\mathcal{M}$ is an invariant mean over the group G, then either

(i) ω is a KMS-state for some inverse temperature $\beta \ge 0$, or

(ii) ω is a ground state. ■

II. CONTINUOUS CLASSICAL SYSTEMS

Consider (K, ω) a Lebesgue probability space, where K stands for the phase space, ω for the state or measure. Suppose there exists a self-adjoint algebra $\mathcal{A}$ (s.a. with respect to complex conjugation), which is dense in $L_2(K, \omega)$ and on which a Poisson bracket is defined, i.e., a bilinear map $\{\cdot, \cdot\}$ of $\mathcal{A} \times \mathcal{A}$ into $L_1(K, \omega)$ such that:

(i) $\overline{\{f, g\}} = \{\bar{f}, \bar{g}\}$, $f, g \in \mathcal{A}$ ($\bar{\cdot}$ means conjugation),

(ii) $\{f, g\} = -\{g, f\}$, $f, g \in \mathcal{A}$,

(iii) $\{fg, h\} = f\{g, h\} + g\{f, h\}$, $f, g, h \in \mathcal{A}$,

(iv) the sesquilinear form $f, g \in \mathcal{A} \to \omega(\{\bar{f}, g\})$ defines an essentially self-adjoint operator L on $\mathcal{A}$,

$$\omega(\{\bar{f}, g\}) = (iLf, g).$$

$(\cdot, \cdot)$ denotes the scalar product on $L_2(K, \omega)$. There exists a subset $\mathcal{B}$ of $\mathcal{A}$ of essentially bounded functions such that $\mathcal{B}$ is a core for L. Furthermore we suppose that $\exp(-L^2/2)\mathcal{B} \subseteq \mathcal{B}$.

Finally we suppose that

(v) there exists an automorphism I for the algebraic structure of $\mathcal{A}$ and that $I^2 = 1$, $I\bar{f} = \overline{If}$, $I\{f, g\} = -\{If, Ig\}$, $I(\mathcal{B}) \subseteq \mathcal{B}$, and that I extends to a unitary operator on $L_2(K, \omega)$ (I = time reversal).

A system $(K, \omega, \mathcal{A}, \{\cdot, \cdot\})$ with the properties (i)–(v) is called a classical physical system.

Now we introduce the dynamics. Suppose we are given a derivation δ of $\mathcal{A}$ such that

(α) for all $f \in \mathcal{A}$, $\alpha_t(f) \equiv \sum_{n\ge 0}(t^n/n!)\,\delta^n f$ converges in $L_2(K, \omega)$ to a strongly continuous group $\alpha = (\alpha_t)_{t\in R}$ of * automorphisms of $\mathcal{A}$ and such that $\mathcal{B}$ is a core for δ.

(β) $\forall t \in R$, $\forall f, g \in \mathcal{A}: \alpha_t\{f, g\} = \{\alpha_t f, \alpha_t g\}$,

$\forall t \in R$, $I\alpha_t = \alpha_{-t} I$.

Although conceptually clearer, the condition that $\exp(-L^2/2)\mathcal{B} \subseteq \mathcal{B}$, might as well be replaced by $\exp(-H^2/2)\mathcal{B} \subseteq \mathcal{B}$, which is perhaps easier to check in applications. In the free one-particle case, for example, one can take the C_0^∞ functions for $\mathcal{B}$. That $\mathcal{B}$ is left invariant under $\exp(-H^2/2)$ is easily shown using the Paley–Wieners theorem.

A system $\{K, \omega, \mathcal{A}, \{\cdot, \cdot\}, \alpha\}$ with the properties (i)–(v), (α), and (β) is called a classical dynamical system. Such a system is called a KMS system if there exists a constant $\beta \ge 0$ such that for all $f, g \in \mathcal{A}: \omega(\{f, g\}) = -\beta\omega(f\delta g)$.

Modulo some other conditions on the dynamical system we could now proceed in the same way as in the quantum case introducing a perturbed dynamics. We will not do that but immediately define energetic stability as follows by a correlation inequality (cf. Corollary 1.3.).

Definition 2.1: A classical dynamical system $\{K, \omega, \mathcal{A}, \{\cdot, \cdot\}, \alpha\}$ is called energetically stable if for all $h \in \mathcal{A}$:

(i) $\omega[\delta(h)] = 0$ (time invariance), (10)

(ii) $\omega(\{\bar{h}, \delta(h)\}) \geq 0.$ (11)

Nevertheless let us sketch formally how we proceed from the perturbed dynamics. For $h=\bar{h}$, (2) becomes,

$$\frac{d}{dt}\alpha_t^h(f)=\alpha_t^h(\delta f+\{f, h_t\}), \quad \alpha_0^h(f)=f,$$

where $h_t=\alpha_t(h)$. A formal solution is given by

$$\alpha_t^h=\exp(2\delta+\{\cdot, h\})t\exp(-\delta t).$$

The energy change is then given by formula (4). Implying the stability as in Definition 1.1 yields first that

$$\omega[\delta(h)]\leq 0 \text{ for all } h=\bar{h}\in\mathcal{A},$$

hence $\omega[\delta(h)]=0$ (time invariance of ω) and then

$$\frac{d}{dt}\omega(\alpha_t^h[\delta(h_t)])\big|_{t=0}\leq 0,$$

implying $\omega(\{h, \delta(h)\})\geq 0$.

Now for arbitrary h we get

$$\omega(\{h, \delta(\bar{h})\})+\omega(\{\bar{h}, \delta(h)\})\geq 0. \qquad (*)$$

Using the time invariance of ω,

$$0=\frac{d}{dt}\omega(\alpha_t\{h, \bar{h}\})\big|_{t=0}=\frac{d}{dt}\omega(\{\alpha_t h, \alpha_t\bar{h}\})\big|_{t=0}$$

$$=\omega(\{\delta h, \bar{h}\})+\omega(\{h, \delta\bar{h}\}). \qquad (**)$$

Combining (*) and (**) we get the conditions of 2.1. Notice that for any KMS system (11) is satisfied; indeed

$$\omega(\{\bar{h}, \delta h\})=\beta\omega(|\delta h|^2)\geq 0.$$

Let us now proceed with stability as in Definition 2.1. The state is time invariant, hence there exists a self-adjoint operator H on $L_2(K, \omega)$ such that for $f\in\mathcal{A}$,

$$\alpha_t f=U_t f=\exp(itH)f, \quad \delta f=iHf.$$

Using (iv) the inequality (11) becomes, for all $h\in\mathcal{A}$,

$$(L h, Hh)\geq 0. \qquad (12)$$

Lemma 2.2: The operators L and H commute strongly if the system is stable.

Proof: The time invariance of ω yields

$$\omega(\{\alpha_t\bar{f}, g\})=\omega(\{\bar{f}, \alpha_{-t}g\}).$$

Using (iv) one gets

$$(L U_t f, g)=(U_t L f, g).$$

As $\mathcal{A}$ is dense in $L_2(K, \omega)$,

$$L U_t f=U_t L f, \quad f\in\mathcal{A}.$$

Again, as $\mathcal{A}$ is dense, L as well as $U_{-t}LU_t$ are essentially self-adjoint and $L=U_{-t}LU_t$ and the lemma follows. ■

As L and H strongly commute they have a common spectral decomposition written as,

$$H=\int_{R\times R}\epsilon\, dF(\epsilon, \lambda), \quad L=\int_{R\times R}\lambda\, dF(\epsilon, \lambda).$$

Lemma 2.3: For a stable system the joint spectrum $\sigma(H, L)$ is contained in $\{(\epsilon, \lambda)\in R^2 \mid \epsilon\lambda\geq 0\}$.

Proof: Take the bounded continuous function $\exp(-x^2/2)$, $x\in R$. Then from (iv) for all $h\in\mathcal{B}$ $\rightarrow \exp(-L^2/2)h\in\mathcal{B}$.

From (11),

$$(L\exp(-L^2/2)h, \ H\exp(-L^2/2)h)\geq 0.$$

By Lemma 2.2,

$$(L\exp(-L^2)h, \ Hh)\geq 0.$$

It is clear that $L\exp(-L^2)$ is a bounded operator. As $\mathcal{B}$ is a core for H, it follows that for all $\psi\in D(H)$

$$(L\exp(-L^2)\psi, H\psi)\geq 0.$$

The joint spectrum of $L\exp(-L^2)$ and H is given by $(\lambda\exp(-\lambda^2), \epsilon)\in R^2$ such that $\lambda\exp(-\lambda^2)\epsilon\geq 0$ or $\epsilon\lambda\geq 0$. ■

Lemma 2.4: For a stable system the spectrum $\sigma(H, L)$ is symmetric with respect to the origin, i.e., if $(\epsilon, \lambda)\in\sigma(H, L)$, then $(-\epsilon, -\lambda)\in\sigma(H, L)$.

Proof: From condition (v) on ω, it follows that $\omega\cdot I=\omega$. Therefore, for all $f, g\in\mathcal{A}$

$$\omega(\{f, g\})=-\omega(\{If, Ig\}).$$

By (iv),

$$(L\bar{f}, g)=-(LI\bar{f}, Ig).$$

Therefore,

$$ILI=-L.$$

Analogously, from condition (β) on the evolution,

$$(f, \alpha_t g)=(If, \alpha_{-t}Ig), \quad f, g\in\mathcal{B}.$$

After differentiation at $t=0$,

$$(f, Hg)=-(If, HIg).$$

As $\mathcal{B}$ is a core for H and IHI,

$$IHI=-H. \ ■$$

Now we are in a position to formulate the final result.

Theorem 2.5: Let $\{K, \omega, \mathcal{A}, \{\cdot, \cdot\}, \alpha\}$ be a stable dynamical system and let G be a locally compact amenable group τ of *-automorphisms τ_g, $g\in G$ of $\mathcal{A}$ commuting with α and such that $\tau_g\{f, h\}=\{\tau_g f, \tau_g h\}$, $f, h\in\mathcal{A}$. Assume that ω is weakly clustering for τ, then either:

(i) the system is a KMS system for some $\beta\geq 0$, or

(ii) the system is a ground state ($H=0$).

Proof: Under the condition of the theorem, we are in the situation of Proposition 4.2 of Ref. 1 and we get that the spectrum $\sigma(H, L)$ is additive. In view of Lemmas 2.3 and 2.4 the spectrum is on a line passing through the origin, i.e., $(\epsilon, \lambda)\in\sigma(H, L)$: $\lambda=\beta\epsilon$ with some finite $\beta\geq 0$ and $L=\beta H$, we get a KMS system, or the line coincides with $(\lambda, 0)$, $\lambda\in R$. The latter case corresponds to $H=0$. ■

Note that we defined ω to be a ground state if $\omega(f\delta g)=0$ for all $f, g\in\mathcal{A}$, or equivalently $H=0$. This is motivated by the KMS relation

$$\frac{1}{\beta}\omega_\beta(\{f,g\}) = -\omega(f\delta g)$$

and letting β tend to infinity.

Let us discuss here the example of the ground state of infinitely many free particles.

Consider for K the classical phase space of infinitely many particles: $x \in K$ is given by $x = (x_i | i \in N)$, $x_i = (q_i, p_i) \in R^{2\nu}$ satisfying the local finiteness condition and let $W(f)$, $f \in D(R^{2\nu})$, be the classical Weyl operators (for more details see Ref. 8). Take the Hamiltonian

$$H(x) = \sum_i p_i^2$$

inducing a quasi-free evolution

$$\alpha_t W(f) = W(\exp(tL)f),$$

where $Lf = \{f, p^2\}$.

Consider the state

$$\omega_{0,\rho}[W(f)] = \exp[\rho \int_{R^\nu} (\exp(if(q,0)) - 1)\, dq]$$

corresponding to the ground state of infinitely many free particles with density ρ. Then consider $(K, \omega_{0,\rho})$, the Lebesgue probability space induced by the state $\omega_{0,\rho}$. It is easily checked; this space satisfies all conditions of a physical system except (iv).

On the other hand, the state $\omega_{0,\rho}$ is clustering for the automorphism group of space translations. Furthermore $\omega_{0,\rho}$ satisfies the stability condition, i.e., for arbitrary $(\lambda_i)_{i=1,\ldots,n}$, λ_i complex numbers and $(f_i)_{i=1,\ldots,n}$, $f_i \in D(R^{2\nu})$ we have

$$\omega_{0,\rho}(\{\overline{\sum_k \lambda_k W(f_k)}, \{\sum_l \lambda_l W(f_l), H\}\})$$

$$= 2\rho \sum_{k,l} \bar{\lambda}_k \lambda_l \omega_{0,\rho}[W(f_l - f_k)] \times \int \left(\exp(if_l)\frac{\partial f_l}{\partial q}\exp(-if_k)\frac{\partial f_k}{\partial q}\right)(q,0)\, dq\,.$$

Notice on the other hand that

$$\omega_{0,\rho}[\sum_k \lambda_k W(f_k)\{\sum_l \lambda_l W(f_l), H\}] = 0$$

yielding $H = 0$.

[1]W. Pusz and S. L. Woronowicz, Commun. Math. Phys. 58, 273 (1978).
[2]M. Pulvirenti, J. Math. Phys. 18, 2099—103 (1977).
[3]A. Verbeure, in *Algebras of Operators and their Applications to Mathematical Physics*, Proceedings of the Marseille Conference 1977, edited by D. Kastler and D.W. Robinson, to appear.
[4]R. Haag, D. Kastler, and E.B. Trych-Pohlmeyer, Commun. Math. Phys. 38, 173 (1974).
[5]G. Sewell, Commun. Math. Phys. 55, 53 (1977).
[6]M. Fannes and A. Verbeure, J. Math. Phys. 19, 558 (1978).
[7]B. Demoen and A. Verbeure, "Homogeneous perturbed classical States and Equilibrium Conditions," preprint KUL-TF-77/021.
[8]J.V. Pulè and A. Verbeure, J. Math. Phys. 18, 862 (1977).

Relativistic quantum kinematics on stochastic phase spaces for massive particles[a)]

Eduard Prugovečki

Department of Mathematics, University of Toronto, Toronto, M5S 1A1, Canada
(Received 1 February 1978)

It is shown that to every Galilei-covariant nonrelativistic stochastic phase space representation of a system of massive particles, whose generator is rotationally invariant, corresponds a Poincaré-covariant relativistic representation sharing the same generator. The stochastic phase space probability densities of the two representations overlap in the limit of nonrelativistic velocities in the laboratory frame. The relativistic representations give rise to covariant and conserved probability currents at stochastic space–time points, in complete analogy with their nonrelativistic counterparts. This parallelism extends to the existence of a global representation of the proper Poincaré group in $L^2(\Gamma)$, which is reduced by each subspace of $L^2(\Gamma)$ spanned by the set of phase-space wavefunctions generated by some stochastic phase space representation of all pure states of the system.

1. INTRODUCTION

The concept of phase space has not, as yet, received as much attention in the context of quantum mechanics as it did in classical mechanics, due to the uncertainty principle which prohibits the simultaneous *infinitely* precise measurement of position and momentum. However, the concept of *stochastic value* for a set of noncommuting observables makes possible the introduction in quantum theory of the notion of probability distributions on *stochastic phase spaces* consisting of stochastic points which are not sharp but spread out to an extent that is in keeping with the uncertainty principle. The resulting theory (cf. Ref. 1 for a review) has a physical interpretation which is entirely consistent with the conventional one, and helps to bridge the conceptual gulf between classical and quantum statistical mechanics in the nonrelativistic context.

The concept of stochastic phase space was introduced also in relativistic quantum mechanics[2] in a manner that was consistent with the Newton–Wigner concept of localizability of relativistic particles, i.e., by assigning to one-particle states probability densities which in the limit of infinitely sharp position measurements coincide with those proposed by Newton and Wigner.[3] Naturally, since the Newton–Wigner probability density is not Lorentz-covariant, neither are the densities introduced in Ref. 2. In fact, no probability density in configuration space alone can be expected to be a relativistically covariant object, since configuration space volume is not a frame-independent entity. However, the local rest-frame volume[4] in the phase space Γ is Lorentz invariant,[4] and this opens the possibility of introducing into the quantum context stochastic phase spaces on which one can define covariant probability densities.

In purely physical terms, we carry out this task by setting up a *coherent array of elementary detectors*[5] $\mathcal{J}(\chi_{q,\mathbf{p}}^{(m)})$ at all points $(q,p)\in\mathbb{R}^4\times V_+^{(m)}$, where $q=(q^0,\mathbf{q})$ is a space–time point, and $p=(p^0,\mathbf{p})$ lies in the forward mass hyperboloid

$$V_+^{(m)}=\{(p^0,\mathbf{p})\,|\,p^0=(\mathbf{p}^2+m^2c^2)^{1/2}\}. \tag{1.1}$$

The procedure is, in operational terms, exactly the same as in the nonrelativistic case,[5] except that we use the proper Poincaré group $P_+^\uparrow$ instead of the Galilei group to relate the confidence functions of the elementary detectors in a state of motion in relation to the laboratory frame of reference to those of detectors that are stationary in that frame.

We begin with an origin-based elementary detector $\mathcal{J}(\chi_{0;\mathbf{0}}^{(m)})$ whose triggering signifies the presence at time $t=0$ of a particle of mass m within the volume Δ of the phase space $\Gamma=\mathbb{R}^6$ with the probability

$$\int_\Delta \chi_{0;\mathbf{0}}^{(m)}(\mathbf{x};\mathbf{k})\,d\mathbf{x}d\mathbf{k}. \tag{1.2}$$

Here we assume, as in the nonrelativistic case, that

$$\chi_{0;\mathbf{0}}^{(m)}(\mathbf{x},\mathbf{k})=\chi_0^{(m)}(\mathbf{x})\hat{\chi}_{\mathbf{0}}^{(m)}(\mathbf{k}), \tag{1.3}$$

$$\int_{\mathbb{R}^3}\chi_0^{(m)}(\mathbf{x})\,d\mathbf{x}=\int_{\mathbb{R}^3}\hat{\chi}_{\mathbf{0}}^{(m)}(\mathbf{k})\,d\mathbf{k}=1, \tag{1.4}$$

and since for any 4-vector $p\in V_+^{(m)}$ the zeroth component is a function of the 3-vector $\mathbf{p}$, we consistently suppress it in indexing of the confidence functions $\chi_{q;\mathbf{p}}^{(m)}$ with which we shall be dealing.

An equivalent but more compact way of describing the role played by this origin-based elementary detector would be to say that $\mathcal{J}(\chi_{0;\mathbf{0}}^{(m)})$ is supposed to detect at the instant $t=0$ the presence of a spinless particle of mass m at the origin-centered stochastic phase-space point $(0;\chi_0^{(m)})\times(\mathbf{0};\hat{\chi}_{\mathbf{0}}^{(m)})$ whose confidence function in phase space is $\chi_0^{(m)}(\mathbf{x})\hat{\chi}_{\mathbf{0}}^{(m)}(\mathbf{k})$.

By taking $\mathcal{J}(\chi_{0;\mathbf{0}}^{(m)})$ and translating it in space–time to the point q, we obtain the $(q;\mathbf{0})$-based detector $\mathcal{J}(\chi_{q;\mathbf{0}}^{(m)})$ which fulfils the function of detecting the presence at time $t=q^0/c$ of a particle at the $(\mathbf{q},\mathbf{0})$-centered stochastic phase-space point $(\mathbf{q},\chi_{\mathbf{q}}^{(m)})\times(\mathbf{0},\hat{\chi}_{\mathbf{p}}^{(m)})$, whose confidence function in phase space is

$$\chi_{q;\mathbf{0}}(\mathbf{x},\mathbf{k})=\chi_0^{(m)}(\mathbf{x}-\mathbf{q})\hat{\chi}_{\mathbf{0}}^{(m)}(\mathbf{k}). \tag{1.5}$$

[a)]Supported in part by a research grant from the National Research Council of Canada.

To obtain $\mathcal{I}(\chi_{q;\mathbf{p}}^{(m)})$ at the general point $(q;\mathbf{p})$ from $\mathbb{R}^4 \times V_+^{(m)}$, we consider an inertial frame that moves in relation to the laboratory frame at the speed $\mathbf{u}$ corresponding to the 3-momentum $\mathbf{p}$,

$$\mathbf{p} = m\mathbf{u}(1-\mathbf{u}^2/c^2)^{-1/2} \tag{1.6}$$

and whose origin coincides with that of the laboratory frame at both the laboratory time $t=0$ and its proper time $t'=0$. Thus, if we denote by Λ_p the pure Lorentz boost[6] from this new frame of reference to the laboratory frame, and write

$$x' = \Lambda_p^{-1}x, \quad k' = \Lambda_p^{-1}k, \tag{1.7}$$

we shall have[6]

$$x'^0 = \gamma\left(x^0 - \frac{\mathbf{u}\cdot\mathbf{x}}{c}\right), \quad \gamma = \left(1-\frac{\mathbf{u}^2}{c^2}\right)^{-1/2}, \tag{1.8}$$

$$\mathbf{x}' = \mathbf{x} - \mathbf{u}\left(x^0 - \frac{\gamma c^{-1}}{1+\gamma}\,\mathbf{x}\cdot\mathbf{u}\right)\gamma c^{-1}, \tag{1.9}$$

so that $x'=0$ for $x=0$, whereas $k=p$ for $\mathbf{k}'=0$ and $k_0' = mc$. By taking duplicates of the elementary detector $\mathcal{I}(\chi_{0;\mathbf{0}}^{(m)})$ and attaching them to various space points $\mathbf{q}'$ at various times $t'=q'^0/c$ in this new inertial frame of reference, we obtain $\mathcal{I}(\chi_{q';\mathbf{0}}^{(m)})$ by a procedure identical to the one previously carried out in the laboratory frame. However, in keeping with the special relativity principle, the observer who is at rest in the laboratory frame obviously has to identify $\mathcal{I}(\chi_{q';\mathbf{0}}^{(m)})$ with $\mathcal{I}(\chi_{q;\mathbf{p}}^{(m)})$. where $q=\Lambda_p^{-1}q'$. The characteristic function $\chi_{q;\mathbf{p}}^{(m)}(\mathbf{x},\mathbf{k})$ which he is going to obtain when the accuracy calibration[7,8] of this elementary detector is carried out in the laboratory frame has to satisfy the equation

$$\left(1-\frac{\mathbf{k}\cdot\mathbf{p}}{k^0p^0}\right)\chi_{q;\mathbf{p}}^{(m)}(\mathbf{x},\mathbf{k}) = \chi_{q';\mathbf{0}}^{(m)}(\mathbf{x}',\mathbf{k}'), \tag{1.10}$$

where $x=\Lambda_p x'$, with $x_0'=q_0'$, and $k=\Lambda_p k'$. The factor on the lhs of (1.10) represents the Jacobian implicit in the change of variables of integration due to the trasition from the rest frame of the detector to that of the observer,

$$d\mathbf{x}'d\mathbf{k}' = \left(1-\frac{\mathbf{k}\cdot\mathbf{u}}{ck^0}\right)d\mathbf{x}d\mathbf{k} = \left(1-\frac{\mathbf{k}\cdot\mathbf{p}}{k^0p^0}\right)d\mathbf{x}d\mathbf{k}, \tag{1.11}$$

and takes care of the normalization of $\chi_{q;\mathbf{p}}^{(m)}(\mathbf{x},\mathbf{k})$ with respect to $\Gamma=\mathbb{R}^6$. It should be noted that since simultaneity in the rest frame of the elementary detector does not imply in a relativistic context simultaneity in the laboratory frame, the observer stationed in this last frame has to carry out the accuracy calibration in such a manner that a positive reading of $\mathcal{I}(\chi_{q;\mathbf{p}}^{(m)})$ should signify the presence of a particle in the infinitesimal phase-space volume $d\mathbf{x}d\mathbf{k}$ around $(\mathbf{x},\mathbf{k})\in\Gamma$ with the "preponderancy" $\chi_{q;\mathbf{p}}^{(m)}(\mathbf{x},\mathbf{k})\,d\mathbf{x}d\mathbf{k}$ (cf. the Appendix) not at the instant q^0/c of the reading, but rather at the laboratory time x^0/c, where

$$x^0 = q^0 + c^{-1}\mathbf{u}\cdot(\mathbf{x}-\mathbf{q}). \tag{1.12}$$

With the resulting coherent array of elementary detectors the observer stationed in the laboratory frame can measure probability densities in the stochastic phase space

$$\Gamma_e^{(m)} = \{(\mathbf{q},\mathbf{p};\chi_{0,\mathbf{q},\mathbf{p}}^{(m)}\,|\,\mathbf{q},\mathbf{p}\in\mathbb{R}^3\} \tag{1.13}$$

at all instants $t=q^0/c$. Indeed, since $\chi_{q';\mathbf{0}}^{(m)}$ in (1.10) is independent of q_0', it follows from (1.5) and (1.9) that the confidence function measured at any fixed momentum $\mathbf{p}$ by an observer stationed at the point $\mathbf{q}$ in the laboratory frame is actually independent of q^0. In fact, we can easily compute from (1.9), (1.10), and (1.12) that

$$\chi_{q;\mathbf{p}}^{(m)}(\mathbf{x},\mathbf{k}) = \left(1-\frac{\mathbf{k}\cdot\mathbf{p}}{k^0p^0}\right)^{-1}\chi_{\mathbf{0}}^{(m)}\left[\mathbf{x}-\mathbf{q}+\frac{\mathbf{p}\cdot(\mathbf{x}-\mathbf{q})}{\mathbf{p}^2+2m^2c^2}\right]$$
$$\times\hat{\chi}_{\mathbf{0}}^{(m)}\left[\mathbf{k}-\frac{\mathbf{p}}{mc}\left(k^0-\frac{\mathbf{p}\cdot\mathbf{k}}{p^0+mc}\right)\right]. \tag{1.14}$$

In sec. 2 and 3 we show that we can associate with every nonrelativistic extremal stochastic phase-space representation of the quantum mechanics of a single spinless particle a corresponding relativistic representation having the same generator[5] $\tilde{e}_{\mathbf{0},\mathbf{0}}(\mathbf{k})$, and that this representation gives rise to a relativistically covariant probability amplitude $\psi_e(q;\mathbf{p})$ on $\Gamma_e^{(m)}$ for every pure quantum state if and only if $\tilde{e}_{\mathbf{0},\mathbf{0}}(\mathbf{k})$ is rotationally invariant; furthermore, the probability density associated with each such amplitude becomes approximately equal to its nonrelativistic counterpart of momentum $\mathbf{p}\approx m\mathbf{u}$ provided the confidence function $\chi_{q;\mathbf{p}}^{(m)}(\mathbf{x},\mathbf{k})$ has a sufficiently narrow spread in the momentum variable $\mathbf{k}$.

In Sec. 4 we show that a covariant and conserved probability current $j_e^\mu(q)$ can be associated with the probability density $|\psi_e(q;\mathbf{p})|^2$ by the method employed also in the nonrelativistic case.

In the concluding section we discuss the import of these results on the concept of localizability of a particle moving at relativistic velocities. We also show that, in complete analogy with the nonrelativistic case, the set of phase-space wavefunctions assigned to all pure states of the system by any stochastic phase space representation spans a closed subspace of $L^2(\Gamma)$ in which a globally defined representation of $P_+^\uparrow$ induces an irreducible representation.

Not to lose sight of the essential physical features of the problem under consideration, we concentrate throughout this paper on the case of extremal representations for pure states of a spinless single particle system, where the particle is assumed to have a nonzero rest mass. The generalizations to mixed states, to nonextremal representations, to many-particle free systems and the inclusion of spin are all a matter of routine, and do not warrant special consideration. The case of mass zero particles poses, however, special problems, and will be dealt with separately in its own terms in a future paper.

2. COVARIANT PROBABILITY AMPLITUDES ON STOCHASTIC PHASE SPACE

In nonrelativistic quantum mechanics for a single spinless particle, the extremal phase-space representations were obtained[5] by associating with the $(\mathbf{q},\mathbf{p})$-based elementary detector a normalized state vector $e_{\mathbf{q},\mathbf{p}}$ so that

$$\chi_{\mathbf{q}}(\mathbf{x}) = |e_{\mathbf{q},\mathbf{p}}(\mathbf{x})|^2, \quad \hat{\chi}_{\mathbf{p}}(\mathbf{k}) = |\tilde{e}_{\mathbf{q},\mathbf{p}}(\mathbf{k})|^2, \tag{2.1}$$

$$\tilde{e}_{\mathbf{q},\mathbf{p}}(\mathbf{k}) = h^{-3/2}\int \exp\left(-\frac{i}{\hbar}\mathbf{q}\cdot\mathbf{x}\right) e_{\mathbf{q},\mathbf{p}}(\mathbf{x})d\mathbf{x}, \tag{2.2}$$

where $\tilde{e}_{\mathbf{q},\mathbf{p}}$ was related to $\tilde{e}_{\mathbf{0},\mathbf{0}}$ by a ray representation $U(\mathbf{q},\mathbf{p})$ of the group of pure Galilean transformations (i.e., coordinate translations and velocity boosts),

$$\tilde{e}_{\mathbf{q},\mathbf{p}} = U(\mathbf{q},\mathbf{p})\tilde{e}_{\mathbf{0},\mathbf{0}}. \tag{2.3}$$

The probability distribution on the stochastic phase space

$$\Gamma_e = \{(\mathbf{q},\chi_{\mathbf{q}})\times(\mathbf{p},\chi_{\mathbf{p}}) \mid \mathbf{q},\mathbf{p}\in\mathbb{R}^3\} \tag{2.4}$$

associated with the density operator ρ was

$$\rho(\mathbf{q},\mathbf{p};e) = h^{-3}\langle e_{\mathbf{q},\mathbf{p}} \mid \rho e_{\mathbf{q},\mathbf{p}}\rangle \tag{2.5}$$

where $\langle\cdot\mid\cdot\rangle$ denotes the inner product in $L^2(\mathbb{R}^3)$.

The existence of a density operator $g_{\mathbf{0},\mathbf{0}}$ that generates a nonrelativistic stochastic phase space representation [and which, in an extremal case,[5] equals $|e_{\mathbf{0},\mathbf{0}}\rangle\langle e_{\mathbf{0},\mathbf{0}}|$ where $e_{\mathbf{0},\mathbf{0}}$ gives rise to an $e_{\mathbf{q},\mathbf{p}}$ that satisfies (2.1), (2.3) and (2.5)] turns out[7] to be a necessary consequence of the conditions of covariance under the Galilei group that were imposed on the stochastic phase space probability density $\rho(\mathbf{q},\mathbf{p};e)$. The existence of the generator $g_{\mathbf{0},\mathbf{0}}$ in turn implied marginality conditions that related $\rho(\mathbf{q},\mathbf{p};e)$ to the conventional probability densities $\langle\mathbf{q}\mid\rho\mid\mathbf{q}\rangle$ and $\langle\mathbf{p}\mid\rho\mid\mathbf{p}\rangle$ in position and momentum, respectively. For such a result to hold it was crucial that the confidence functions $\chi_{\mathbf{q}}$ and $\hat{\chi}_{\mathbf{p}}$ at the point $(\mathbf{q},\mathbf{p})\in\Gamma$ were the outcome of rigid translations of the confidence function $\chi_{\mathbf{0}}$ and $\hat{\chi}_{\mathbf{0}}$ at the origin:

$$\chi_{\mathbf{q}}(\mathbf{x}) = \chi_{\mathbf{0}}(\mathbf{x}-\mathbf{q}), \quad \hat{\chi}_{\mathbf{p}}(\mathbf{k}) = \hat{\chi}_{\mathbf{0}}(\mathbf{k}-\mathbf{p}). \tag{2.6}$$

The principal aim of the present section is to carry through as far as possible the same program in the relativistic context. Thus, we shall assign to the state-vectors of a spinless particle probability densities $\rho_e(q;\mathbf{p})$ which transform as scalar quantities under the action of $\mathcal{P}_+^\uparrow$, and which (in a sense to be made more precise later) are well approximated by their nonrelativistic counterparts $\rho(\mathbf{q},\mathbf{p};e)$ when the measurements are performed with the subset of elementary detectors from the coherent array described in the Introduction that move at nonrelativistic speeds in relation to the laboratory frame. It should be realized already at this stage, however, that since the confidence functions of this array of detectors are not congruent,[8] no marginality conditions can be expected to hold in a strict sense.[8] Indeed, by (1.14)

$$\chi_{0;\mathbf{p}}^{(m)}(\mathbf{x},\mathbf{k}) = \left(1-\frac{\mathbf{k}\cdot\mathbf{u}}{k^0 c}\right)^{-1}\chi_{0;\mathbf{0}}^{(m)}\left(\mathbf{x}+\frac{\mathbf{u}(\mathbf{u}\cdot\mathbf{x})}{2c^2-\mathbf{u}^2},\ \Lambda_p^{-1}\mathbf{k}\right), \tag{2.7}$$

and according to (1.6) and (1.9),

$$\Lambda_p^{-1}\mathbf{k} = \mathbf{k}-\mathbf{p}\left(\gamma-\frac{\gamma}{1+\gamma}\frac{\mathbf{u}\cdot\mathbf{k}}{mc^2}\right) = \mathbf{k}-\mathbf{p}+o\left(\frac{\mathbf{u}^2}{c^2}\right), \tag{2.8}$$

so that there is congruency in the sense of (2.6) to an approximate degree if and only if $\mathbf{u}^2/c^2 \ll 1$ and $k^0 \approx mc$.

Consequently, as a leading principle we adopt the requirement that the relativistic formalism on each stochastic phase space (1.13) should be approximated by its nonrelativistic counterpart in (2.4), where the correspondence between these two stochastic phase spaces should result from the usage of one and the same coherent array of elementary detectors in both cases. The distinction between the two cases lies therefore exclusively in the assumption as to which group of transformations—Poincaré or Galilei—applied to the array of stationary elementary detectors will reproduce the correct confidence functions of those elementary detectors that are in a state of motion in the laboratory frame. This means, however, that in its own rest frame each elementary detector $\mathcal{G}(\chi_{q;\mathbf{p}}^{(m)})$ had undergone one and the same accuracy calibration regardless of whether it was intended for use by relativistic or nonrelativistic observers—the difference between these two types of observers reflecting only the differences in their perceptions of this calibration once they move to some other inertial frame. Thus, if primed variables refer to the rest frame of $\mathcal{G}(\chi_{q;\mathbf{p}}^{(m)})$, then the confidence function associated with this elementary detector assumes in that frame the form of a product of the two functions appearing in (2.1), namely,

$$\chi_{q';\mathbf{0}}^{(m)}(\mathbf{x}',\mathbf{k}') = |e_{\mathbf{q}',\mathbf{0}}(\mathbf{x}')|^2\,|\tilde{e}_{\mathbf{q}',\mathbf{0}}(\mathbf{k}')|^2 \tag{2.9}$$

where $\tilde{e}_{\mathbf{q}',\mathbf{0}}(\mathbf{k}')$, and therefore also

$$e_{\mathbf{q}',\mathbf{0}}(\mathbf{x}') = h^{-3/2}\int\exp\left(\frac{i}{\hbar}\mathbf{q}'\cdot\mathbf{k}'\right)\tilde{e}_{\mathbf{q}',\mathbf{0}}(\mathbf{k}')d\mathbf{k}'$$

are normalized elements of $L^2(\mathbb{R}^3)$, and where according to (2.3)

$$e_{\mathbf{q},\mathbf{0}}(\mathbf{x}') = e_{\mathbf{0},\mathbf{0}}(\mathbf{x}'-\mathbf{q}'). \tag{2.10}$$

Hence, by (1.7) and (1.10) we can write

$$\chi_{q;\mathbf{p}}^{(m)}(\mathbf{x},\mathbf{k}) = \left(1-\frac{\mathbf{k}\cdot\mathbf{p}}{k^0p^0}\right)^{-1}|e_{\mathbf{0},\mathbf{0}}(\Lambda_p^{-1}x-\Lambda_p^{-1}q)|^2\,|\tilde{e}_{\mathbf{0},\mathbf{0}}(\Lambda_p^{-1}\mathbf{k})|^2 \tag{2.11}$$

provided we set $e_{\mathbf{0},\mathbf{0}}(x) = e_{\mathbf{0},\mathbf{0}}(\mathbf{x})$ for all $x^0\in\mathbb{R}^1$.

In the relativistic case, the pure states of a spinless particle can be described in the momentum representation by normalized functions $\hat{\psi}(\mathbf{k})$ from the space $L_\mu^2(\mathbb{R}^3)$ with inner product

$$(\hat{\psi}_1\mid\hat{\psi}_2) = \int\hat{\psi}_1^*(\mathbf{k})\hat{\psi}_2(\mathbf{k})\,d\mu(\mathbf{k}), \quad d\mu(\mathbf{k}) = (\mathbf{k}^2+m^2c^2)^{-1/2}d\mathbf{k}. \tag{2.12}$$

In $L_\mu^2(\mathbb{R}^3)$ the representation of the restricted Poincaré group $\mathcal{P}_+^\uparrow$ is

$$[\hat{U}(a,\Lambda)\psi]^\wedge(\mathbf{k}) = \exp\left(\frac{i}{\hbar}k\cdot a\right)\hat{\psi}(\Lambda^{-1}\mathbf{k}), \tag{2.13}$$

and therefore the time evolution of each state of the system is described as follows:

$$\hat{U}_t\hat{\psi}(\mathbf{k}) = \exp\left(-\frac{i}{\hbar}x^0k^0\right)\hat{\psi}(\mathbf{k}), \quad x^0 = ct. \tag{2.14}$$

Our aim is to assign to $\hat{U}_t\hat{\psi}$ an amplitude $\psi_e(q;\mathbf{p})$, $q^0 = ct$, that transforms as a scalar under the action of the Poincaré group, i.e.,

$$[\hat{U}(a,\Lambda)\psi]_e(q;\mathbf{p}) = \psi_e(\Lambda^{-1}(q-a);\Lambda^{-1}\mathbf{p}) \tag{2.15}$$

for all $\Lambda\in\mathcal{P}_+^\uparrow$, and is such that

$$\rho_e(q;\mathbf{p}) = |\psi_e(q;\mathbf{p})|^2 \tag{2.16}$$

is a probability density, which for small velocities $\mathbf{u}$, i.e., for $\mathbf{p}\approx m\mathbf{u}$, and $\hat{\chi}^{(m)}_{q;\mathbf{p}}(\mathbf{x},\mathbf{k})$ sharply peaked around $\mathbf{p}$ in the variable $\mathbf{k}$, $\rho_e(q;\mathbf{p})$ becomes approximately equal to (2.5). More precisely, we require that

$$\rho_e(q;\mathbf{p}) = h^{-3}\,|\langle e_{\mathbf{q},\mathbf{p}}|\hat{U}_t\hat{\psi}\rangle|^2 + o\left(\frac{\mathbf{u}^2}{c^2}\right), \tag{2.17}$$

$$\langle e_{\mathbf{q},\mathbf{p}}|\psi\rangle = (mc)^{1/2}\int \tilde{e}^*_{\mathbf{q},\mathbf{p}}(\mathbf{k})\hat{\psi}(\mathbf{k})\,d\mathbf{k}. \tag{2.18}$$

The appearance of $(mc)^{1/2}$ in (2.18) is due to the fact that, according to (2.12), in the relativistic case $|\hat{\psi}(\mathbf{k})|^2$ is a probability density not with respect to $d\mathbf{k}$ but rather with respect to

$$(m^2c^2+\mathbf{k}^2)^{-1/2}d\mathbf{k} = (mc)^{-1}\left[1+o\left(\frac{\mathbf{v}^2}{c^2}\right)\right]\,d\mathbf{k}, \tag{2.19}$$

where $\mathbf{v}$ denotes the 3-velocity associated with $\mathbf{k}$.

At $\mathbf{p}=\mathbf{0}$ in the laboratory frame, we can satisfy (2.17) if we adopt

$$\psi_e(ct,\mathbf{q};\mathbf{0}) = (mch^{-3})^{1/2}\int \tilde{e}^*_{\mathbf{q},\mathbf{0}}(\mathbf{k})(\hat{U}_t\hat{\psi})(\mathbf{k})\,d\mu(\mathbf{k}) \tag{2.20}$$

as the sought-after probability amplitude—the choice of the measure $d\mu(\mathbf{k})$ on momentum space being dictated by the requirement of relativistic invariance. Since all inertial frames of reference should be equivalent with regard to measurements performed in identical manner in relation to each one of them, (2.20) should hold also in the rest-frame of the detector $\mathcal{G}(\chi^{(m)}_{q;\mathbf{p}})$, i.e., we should have

$$\psi'_e(ct',\mathbf{q};0) = h^{-3/2}\int \hat{e}^*_{q';\mathbf{0}}(\mathbf{k}')\hat{\psi}(\mathbf{k}')\,d\mu(\mathbf{k}'), \tag{2.21}$$

where we have set by definition

$$\hat{e}_{q';\mathbf{0}}(\mathbf{k}') = (mc)^{1/2}\exp\left(\frac{i}{\hbar}q'_0k'_0\right)\tilde{e}_{\mathbf{q}',\mathbf{0}}(\mathbf{k}'), \tag{2.22}$$

and $\hat{\psi}'(\mathbf{k}')$ was obtained from $\hat{\psi}(\mathbf{k}')$ by applying to it $\hat{U}(0,\Lambda_p^{-1})$ in accordance to (2.13),

$$\hat{\psi}'(\mathbf{k}') = \hat{\psi}(\Lambda_p\mathbf{k}') = \hat{\psi}(\mathbf{k}). \tag{2.23}$$

The probability density $\rho_e(q;\mathbf{p})$ measured with $\mathcal{G}(\chi^{(m)}_{q;\mathbf{p}})$ in the laboratory frame equals the density $\rho'_e(q';\mathbf{0})$ measured with $\mathcal{G}(\chi^{(m)}_{q;\mathbf{p}})$ in its rest frame at $q'=\Lambda_p^{-1}q$. This is consistent with the choice

$$\psi_e(q;\mathbf{p}) = \psi'_e(q';\mathbf{0}),\quad q=\Lambda_pq', \tag{2.24}$$

which, combined with (2.21) and the definition

$$\hat{e}_{q;\mathbf{p}}(\mathbf{k}) = \hat{e}_{q';\mathbf{0}}(\Lambda_p^{-1}\mathbf{k}), \tag{2.25}$$

leads to the expression

$$\psi_e(q;\mathbf{p}) = h^{-3/2}\int \hat{e}^*_{q;\mathbf{p}}(\mathbf{k})\hat{\psi}(\mathbf{k})\,d\mu(\mathbf{k}) \tag{2.26}$$

upon noting that $d\mu(\mathbf{k}') = d\mu(\mathbf{k})$.

According to (2.2) and (2.3),

$$\tilde{e}_{\mathbf{q},\mathbf{0}}(\mathbf{k}) = \exp\left(-\frac{i}{\hbar}\mathbf{q}\cdot\mathbf{k}\right)\tilde{e}_{\mathbf{0},\mathbf{0}}(\mathbf{k}). \tag{2.27}$$

Let us insert this expression into (2.22) and use the result in (2.25). Since $q'\cdot k' = q\cdot k$, we obtain

$$\hat{e}_{q;\mathbf{p}}(\mathbf{k}) = (mc)^{1/2}\exp\left(\frac{i}{\hbar}q\cdot k\right)\tilde{e}_{\mathbf{0},\mathbf{0}}(\Lambda_p^{-1}\mathbf{k}), \tag{2.28}$$

which in view of (2.13), is equivalent to

$$\hat{e}_{q;\mathbf{p}} = (mc)^{1/2}U(q,\Lambda_p)\tilde{e}_{\mathbf{0},\mathbf{0}}. \tag{2.29}$$

The functions $\psi_e(\mathbf{q};\mathbf{p})$ defined in (2.26) are our candidates for probability amplitudes on the stochastic phase space $\Gamma^{(m)}_e$ introduced in (1.13). Since our ultimate goal is a relativistically covariant theory, we have to establish when these functions transform as scalars under the action of $\mathcal{P}^{\uparrow}_{+}$.

Theorem 2.1: The linear mapping

$$U_e(q^0):\hat{\psi}(\mathbf{k})\mapsto\psi_e(q;\mathbf{p}) = h^{-3/2}\langle\hat{e}_{q;\mathbf{p}}|\psi\rangle \tag{2.30}$$

has as a range a linear space $\mathcal{M}_e$ which does not depend on q^0. In this space the irreducible representation $\hat{U}(a,\Lambda)$ of $\mathcal{P}^{\uparrow}_{+}$, defined in $L^2_\mu(\mathbb{R}^3)$ by (2.13), induces a representation $U_e(a,\Lambda)$,

$$U_e(a,\Lambda)\psi_e(q;\mathbf{p}) = h^{-3/2}\langle\hat{e}_{q;\mathbf{p}}|\hat{U}(a,\Lambda)\hat{\psi}\rangle, \tag{2.31}$$

under which $\psi_e(q;\mathbf{p})$ behaves as a scalar quantity,

$$U_e(a,\Lambda)\psi_e(q;\mathbf{p}) = \psi_e[\Lambda^{-1}(q-a);\Lambda^{-1}\mathbf{p}], \tag{2.32}$$

if and only if $\tilde{e}_{\mathbf{0},\mathbf{0}}(\mathbf{k})$ is rotationally invariant in $\mathbf{k}\in\mathbb{R}^3$.

Proof: Let us denote by $\mathcal{M}_e$ the range of $U_e(0)$. To see that $U_e(q^0)$ has the same range $\mathcal{M}_e$ for all $q^0\in\mathbb{R}^1$, note that if $\hat{U}_t$ denotes the time-evolution operator defined in (2.14), then by (2.26) and (2.28),

$$U_e(ct)\hat{\psi} = U_e(0)\hat{U}_t\hat{\psi}, \tag{2.33}$$

whenever $\hat{\psi}\in L^2_\mu(\mathbb{R}^3)$.

To verify (2.32), we observe that

$$\hat{U}^{-1}(a,\Lambda)\hat{U}(q,\Lambda_p) = \hat{U}(\Lambda^{-1}(q-a),\Lambda^{-1}\Lambda_p). \tag{2.34}$$

The relation (2.32) then follows from (2.29), (2.31) and (2.34) if and only if

$$\hat{U}(0,\Lambda^{-1}\Lambda_p)\hat{e}_{0;\mathbf{0}} = \hat{U}(0,\Lambda_{p'})\hat{e}_{0;\mathbf{0}},\quad p'=\Lambda^{-1}p, \tag{2.35}$$

for every proper Lorentz transformation $\Lambda\in L^{\uparrow}_{+}$.

The equality (2.37) is certainly true if Λ is a boost in the direction of $\mathbf{p}$, since then we have

$$\Lambda^{-1}\Lambda_p = \Lambda_{p'},\quad p'=\Lambda^{-1}p. \tag{2.36}$$

This truism becomes evident as soon as it is recalled that, in accordance with (1.6)—(1.9), Λ_p can be viewed as a pure Lorentz boost that takes us from the rest frame of a classical particle of rest mass m to a frame in which the same particle has 4-momentum p. Consequently, if Λ_p is followed by another boost in the same direction, then the net result will be a boost to a frame of reference in which the same particle has 4-momentum $p'=\Lambda^{-1}p$, and therefore (2.36) holds true.

If, however, Λ^{-1} is not a boost in the direction of $\mathbf{p}$, then $\Lambda^{-1}\Lambda_p$ is not a pure Lorentz boost, but it can be reduced to the form $\Lambda_{p'}\cdot R$, where R is a Euclidean rotation.[6] The rotational invariance of $\tilde{e}_{\mathbf{0};\mathbf{0}}(\mathbf{k})$ ensures that (2.35) stays true, and therefore provides a sufficient condition for (2.32). Conversely, this rotational invariance is also a necessary condition, as can be immediately seen by setting in (2.35) $\mathbf{p}=\mathbf{0}$ and taking $\Lambda \in SO(3)$ to be an arbitrary Euclidean rotation. Q.E.D.

3. RELATIVISTICALLY COVARIANT CONTINUOUS RESOLUTIONS OF THE IDENTITY

The interpretation of $\psi_e(q;\mathbf{p})$ as a probability amplitude that gives rise to a probability density $\rho_e(q;\mathbf{p})$ on $\Gamma_e^{(m)}$ in accordance with (2.16) is consistent if and only if

$$\int |\psi_e(q^0,\mathbf{q};\mathbf{p})|^2\, d\mathbf{q}d\mathbf{p} \equiv 1 \tag{3.1}$$

for all $q^0 \in \mathbb{R}^1$, and for an arbitrary normalized element $\hat{\psi}(\mathbf{k})$ of $L^2_\mu(\mathbb{R}^3)$. But this criterion can be restated in the form of a condition on the mapping $U(q^0)$ defined in (2.30), namely that $U(q^0)$ should be an isometry between $L^2_\mu(\mathbb{R}^3)$ and a closed subspace of $L^2(\Gamma)$. Since this subspace would have to be the range of $U(q^0)$, and therefore it would coincide with the set $\mathcal{M}_e$ defined in Theorem 2.1, the condition (3.1) can be restated as a request that $U(q^0)$ provide a unitary transformation of $L^2_\mu(\mathbb{R}^3)$ onto the closed subspace $\mathcal{M}_e$ of $L^2(\Gamma)$.

Theorem 3.1: If the functions $\psi_e(q;\mathbf{p})$ assigned to each $\hat{\psi}(\mathbf{k}) \in L_\mu(\mathbb{R}^3)$ in accordance with (2.26)—(2.28) at a fixed choice of generator $\tilde{e}_{\mathbf{0},\mathbf{0}}(\mathbf{k})$ transform as scalars under the action of $L^\uparrow_+$, then

$$\int_{\mathbb{R}^6} \phi_e^*(q;\mathbf{p})\psi_e(q;\mathbf{p})\, d\mathbf{q}d\mathbf{p} = \int_{\mathbb{R}^3} \hat{\phi}^*(\mathbf{k})\hat{\psi}(\mathbf{k})\, d\mu(\mathbf{k}) \tag{3.2}$$

for arbitrary $\hat{\phi}(\mathbf{k}),\ \hat{\psi}(\mathbf{k}) \in L_\mu(\mathbb{R}^3)$.

Proof: Using the unitarity property of the Fourier—Plancherel transform in the $\mathbf{q}$-variable that is obtained when (2.28) is inserted in the expression (2.26) for ψ_e as well as in its counterpart for ϕ_e, we deduce that

$$\int d\mathbf{q} \int d\mathbf{p}\, \phi_e^*(q;\mathbf{p})\psi_e(q;\mathbf{p})$$

$$= mc \int d\mathbf{p} \int d\mu(\mathbf{k}) k_0^{-1} |e_{\mathbf{0},\mathbf{0}}(\Lambda_p^{-1}\mathbf{k})|^2 \hat{\phi}^*(\mathbf{k})\hat{\psi}(\mathbf{k}) \tag{3.3}$$

at each fixed value $q^0 \in \mathbb{R}^1$. Reversing by Fubini's theorem the orders of integration in $\mathbf{p}$ and $\mathbf{k}$, we observe that the rhs of (3.3) is equal to the rhs of (3.2) if and only if

$$\int |\tilde{e}_{\mathbf{0},\mathbf{0}}(\Lambda_p^{-1}\mathbf{k})|^2\, d\mathbf{p} = (mc)^{-1}k_0 \tag{3.4}$$

for almost all $\mathbf{k} \in \mathbb{R}^3$.

To establish that (3.4) is indeed true, let us recall that by Theorem 2.1 the functions $\psi_e(q;\mathbf{p})$ transform as scalars if and only if the generator of the representation is rotationally invariant, i.e., if and only if it can be expressed as a function of $|\mathbf{k}|$, or, equivalently, of mck^0:

$$\tilde{e}_{\mathbf{0},\mathbf{0}}(\mathbf{k}) = e(mck^0), \quad k^0 = (\mathbf{k}^2 + m^2c^2)^{1/2}. \tag{3.5}$$

However, in that case we have

$$\tilde{e}_{\mathbf{0},\mathbf{0}}(\Lambda_p^{-1}\mathbf{k}) = e(mc(\Lambda_p^{-1}k)^0) = e(p \cdot k) \tag{3.6}$$

and therefore, by reversing the roles of k and p

$$\tilde{e}_{\mathbf{0},\mathbf{0}}(\Lambda_p^{-1}\mathbf{k}) = \tilde{e}_{\mathbf{0},\mathbf{0}}(\mathbf{p}'), \quad \mathbf{p}' = \Lambda_k^{-1}\mathbf{p}. \tag{3.7}$$

Introducing $\mathbf{p}'$ as a new variable of integration and recalling that $p_0^{-1}d\mathbf{p}$ is invariant, we get

$$\int |\tilde{e}_{\mathbf{0},\mathbf{0}}(\Lambda_p^{-1}\mathbf{k})|^2\, d\mathbf{p} = \int |\tilde{e}_{\mathbf{0},\mathbf{0}}(\mathbf{p}')|^2 \frac{p_0}{p_0'}\, d\mathbf{p}'. \tag{3.8}$$

To complete the computation, we have to express p_0 in terms of p'. Hence, we decompose $\mathbf{p}$ into its orthogonal projection $p_\parallel$ along $\mathbf{k}$, and a component $\mathbf{p}_\perp$ orthogonal to $\mathbf{k}$,

$$p_\parallel = |\mathbf{k}|^{-1}\mathbf{k}\cdot\mathbf{p}, \quad \mathbf{p}_\perp = \mathbf{p} - p_\parallel |\mathbf{k}|^{-1}\mathbf{k}, \tag{3.9}$$

and submit $\mathbf{p}'$ to the same decomposition. Since $p' = \Lambda_k^{-1}p$, we have, in accordance with (1.6)—(1.9),

$$p_0 = \gamma\left(p_0' - \frac{\mathbf{v}\cdot\mathbf{p}'}{c}\right) = \gamma\left(p_0' - \frac{|\mathbf{v}|}{c}p_\parallel'\right) \tag{3.10}$$

$$\mathbf{k} = m\gamma\mathbf{v}, \quad \gamma = \left(1 - \frac{\mathbf{v}^2}{c^2}\right)^{-1/2} = (mc)^{-1}k^0. \tag{3.11}$$

Inserting the expression (3.10) for p_0 into (3.8), we arrive at the conclusion that (3.4) is satisfied if and only if

$$\int_{\mathbb{R}^2} dp_\perp' \int_{-\infty}^{+\infty} dp_\parallel' \frac{p_\parallel'}{p_0'} |e(mcp_0')|^2 = 0 \tag{3.12}$$

for almost all values of $\mathbf{k}$ on the unit sphere in $\mathbb{R}^3$. That this is indeed so follows from the fact that

$$p_0' = [(\mathbf{p}_\perp')^2 + (p_\parallel')^2 + m^2c^2]^{1/2} \tag{3.13}$$

is an even function of $p_\parallel'$, which makes the integrand of the $p_\parallel'$-integral in (3.12) an odd function. Q.E.D.

The physical significance of the preceding two theorems is that they establish the possibility of constructing for every nonrelativistic extremal stochastic phase space representation of a one-particle system (or, more generally, a system of any number of noninteracting massive particles) with a rotationally-invariant generator a covariant relativistic representation, which shares the same generator, and therefore for which the respective probability densities at low laboratory-frame velocities approximate each other in the sense of (2.17). Computationally, the method of relating these two representations to each other is very straightforward: the generator $\tilde{e}_{\mathbf{0},\mathbf{0}}(\mathbf{k})$ of the nonrelativistic representation is rewritten in the form (3.5) as a function of mck^0, and then by (2.28) and (3.6),

$$\hat{e}_{q;\mathbf{p}}(\mathbf{k}) = (mc)^{1/2} \exp\left(\frac{i}{\hbar}q\cdot k\right) e(p\cdot k). \tag{3.14}$$

As an important class of examples, let us consider the optimal stochastic phase space representations. In the nonrelativistic case, these representations can be constructed[1,5,8] from vectors $e_{\mathbf{q},\mathbf{p}}$ which represent coherent states and in the momentum representation assume the form

$$\tilde{e}^{(s)}_{\mathbf{q},\mathbf{p}}(\mathbf{k}) = (\pi\hbar s^{-2})^{-3/4} \exp\left[-\frac{s^2}{2\hbar}(\mathbf{k}-\mathbf{p})^2 - \frac{i}{\hbar}\mathbf{q}\cdot\mathbf{k}\right]. \tag{3.15}$$

In (3.15) the parameter s plays the role of instrument characteristic[8] for the class of optimally accurate coherent arrays of elementary detectors, and is proportional to the imprecision of position measurements. The relativistic counterpart $\hat{e}^{(s)}_{q;\mathbf{p}}(\mathbf{k})$ of (3.15) is then easily computed by the above-described method: we set in (3.15) $\mathbf{q}=\mathbf{p}=\mathbf{0}$, and then rewrite the common generator in the form (3.5), thus obtaining from (3.15)

$\hat{e}^{(s)}_{q;\mathbf{p}}(\mathbf{k})$

$$= (mcs^3)^{1/2}(\pi\hbar)^{-3/4}$$

$$\times \exp\left[-\frac{s^2}{2m^2c^2\hbar}(k\cdot p)^2 + \frac{i}{\hbar}q\cdot k + \frac{m^2c^2s^2}{2\hbar}\right] . \quad (3.16)$$

From a purely mathematical point of view, the significance of Theorem 3.1 lies in its displaying a whole new family of continuous resolutions of the identity[9] in $L^2_\mu(\mathbb{R}^3)$. Indeed, denoting in general by $|f)(f|$ the projector from $L^2_\mu(\mathbb{R}^3)$ onto the normalized vector $f \in L^2_\mu(\mathbb{R}^3)$, we can write

$$h^{-3}\int_{\mathbb{R}^6}|\hat{e}_{q;\mathbf{p}})d\mathbf{q}d\mathbf{p}(\hat{e}_{q;\mathbf{p}}| = \mathbb{1} . \quad (3.17)$$

This is in fact only a restatement of (3.2), as seen when both sides of (3.17) are applied to $\hat{\psi}$, and afterwards the inner product with $\hat{\phi}$ is taken.

In the next section we shall see that in the limit of sharp position measurements the probability amplitude $\psi_e(x;\mathbf{p})$ becomes related to the conventional configuration representation[10]

$$\psi(x) = h^{-3/2}2^{-1/2}\int \exp\left(-\frac{i}{\hbar}x\cdot k\right)\hat{\psi}(\mathbf{k})\,d\mu(\mathbf{k}) \quad (3.18)$$

of $\hat{\psi}(\mathbf{k})$, in terms of which the inner product (2.12) assumes the form

$$(\hat{\psi}_1|\hat{\psi}_2) = i\hbar\int_{\mathbb{R}^3}\psi_1^*(x)\frac{\overleftrightarrow{\partial}}{\partial x^0}\psi_2(x)\,d\mathbf{x}. \quad (3.19)$$

It is therefore noteworthy that $\hat{e}_{q;\mathbf{p}}$ provides a continuous resolution of the identity also in the context of (3.19)—in the precise sense that is implicit in the following theorem.

Theorem 3.2: For any rotationally-invariant generator $\tilde{e}_{\mathbf{0},\mathbf{0}}(\mathbf{k})$ belonging to $L^2_\mu(\mathbb{R}^3)$, and for all $\hat{\phi}$, $\hat{\psi} \in L^2_\mu(\mathbb{R}^3)$

$$(\hat{\phi}|\hat{\psi}) = N_e^{-1}\int \phi_e^*(q;\mathbf{p})\frac{\overleftrightarrow{\partial}}{\partial q^0}\psi_e(q;\mathbf{p})\,d\mathbf{q}\frac{d\mathbf{p}}{p^0}, \quad (3.20)$$

$$N_e = \frac{2mc}{i\hbar}\int \hat{\chi}_{\mathbf{0}}^{(m)}(\mathbf{k})\,d\mu(\mathbf{k}) = \frac{2mc}{i\hbar}\|\tilde{e}_{\mathbf{0},\mathbf{0}}\|^2_\mu. \quad (3.21)$$

Proof: Combining (2.26) and (2.28), we obtain

$$i\hbar\int \phi_e^*(q;\mathbf{p})\frac{\overleftrightarrow{\partial}}{\partial q^0}\psi_e(q;\mathbf{p})\,d\mathbf{q}\frac{d\mathbf{p}}{p^0}$$

$$= 2mc\int\frac{d\mathbf{p}}{p^0}\int\frac{d\mathbf{k}}{k^0}\hat{\phi}^*(\mathbf{k})\hat{\psi}(\mathbf{k})|\tilde{e}_{\mathbf{0},\mathbf{0}}(\Lambda_p^{-1}\mathbf{k})|^2 \quad (3.22)$$

by using the unitarity property of the Fourier—Plancherel transform in $\mathbf{q}\in\mathbb{R}^3$. Applying (3.7) and the invariance of $p_0^{-1}d\mathbf{p}$ under Lorentz transformations, we deduce that

$$\int\frac{d\mathbf{p}}{p_0}|\tilde{e}_{\mathbf{0},\mathbf{0}}(\Lambda_p^{-1}\mathbf{k})|^2 = \int\frac{d\mathbf{p}'}{p_0'}|\tilde{e}_{\mathbf{0},\mathbf{0}}(\mathbf{p}')|^2 = \|\tilde{e}_{\mathbf{0},\mathbf{0}}\|^2_\mu. \quad (3.23)$$

After the orders of integration in $\mathbf{p}$ and $\mathbf{k}$ are reversed on the rhs of (3.22) by Fubini's theorem, (3.20) is obtained as a consequence of (3.23). Q.E.D.

The normalization of $\tilde{e}_{\mathbf{0},\mathbf{0}}(\mathbf{k})$ could have been adjusted so as to make N_e^{-1} equal to the conventional factor $i\hbar$ that appears in (3.19). However, there would be no physical motivation behind such a choice of normalization since the quantity $\psi_e^*(\overleftrightarrow{\partial}/\partial q^0)\psi_e$ is not positive definite on phase space, and therefore it cannot be interpreted as a probability density.[10]

4. COVARIANT PROBABILITY DENSITY CURRENTS AT STOCHASTIC CONFIGURATION POINTS

The main conclusion we can draw from the preceding considerations is that on any rotationally invariant stochastic phase space $\Gamma_e^{(m)}$ defined in (1.13) one can introduce a probability density $\rho_e(q;\mathbf{p})$ that transforms as a scalar under the action of the proper Lorentz group. We still have to establish that the properties of this density are consistent with those of the conventional probability density $|\hat{\psi}(\mathbf{k})|^2$ with respect to the relativistically invariant element of volume in momentum space.

In the nonrelativistic case, this consistency was established by verifying that the proper marginality conditions in $\mathbf{q}$ and $\mathbf{p}$ were satisfied.[7,8] However, as explained in the Introduction, since relationships of the type (2.6) do not hold in the relativistic case [except to an approximate degree when $\mathbf{u}^2 \ll c^2$ and for $\chi^{(m)}_{0;\mathbf{p}}(\mathbf{k})$ sharply peaked around $\mathbf{p}$] we cannot possibly hope to obtain relativistic counterparts of these marginality conditions. Indeed, for example, we have the relation

$$\int|\psi_e(q;\mathbf{p})|^2\,d\mathbf{q} = \int\frac{mc}{k^0}|\hat{e}_{0;\mathbf{p}}(\mathbf{k})|^2|\hat{\psi}(\mathbf{k})|^2\,d\mu(\mathbf{k}) \quad (4.1)$$

which, as predicted, assumes approximately the form of a bona fide marginality property only when the leading contribution to the integral on its rhs comes from the region where $k^0 \approx mc$.

We can investigate, however, what happens in the limit of infinitely sharp momentum measurements by following the type of procedure used in the nonrelativistic context,[11] namely by considering $\rho_e(q;\mathbf{p})$ for the optimal case, when it is derivable from the amplitude

$$\psi^{(s)}(q;\mathbf{p}) = h^{-3/2}(\hat{e}^{(s)}_{q;\mathbf{p}}|\hat{\psi}) \quad (4.2)$$

with $\hat{e}^{(s)}_{q;\mathbf{p}}$ defined in (3.16), and then trying to determine what takes place as $s \to +\infty$.

When $\mathbf{p}=\mathbf{0}$ the expression (4.2) becomes equal to the nonrelativistic inner product of $(mck_0^{-2})^{1/2}\hat{U}_t\hat{\psi}(\mathbf{k})$ with $\tilde{e}_{\mathbf{q},\mathbf{0}}(\mathbf{k})$ in (3.15). Therefore, the nonrelativistic treatment[11] applies when $\hat{\psi}(\mathbf{k})$ belongs to $L^1_\mu \cap L^2_\mu$ and is continuous. Thus we get

$$\lim_{s\to+0}(\pi\hbar s^{-2})^{3/4}\psi^{(s)}(q;\mathbf{0}) = (2mc)^{-1/2}\psi(q), \quad (4.3)$$

$$\lim_{s\to\infty}(\pi\hbar s^2)^{3/4}\psi^{(s)}(q;\mathbf{0}) = (mc)^{-1/2}\exp\left(-\frac{i}{\hbar}mcq^0\right)\hat{\psi}(\mathbf{0}). \quad (4.4)$$

Using (2.32) and the fact that for $\psi(q)$ defined in (3.18), we have

$$(\hat{U}(a,\Lambda)\psi)(q) = \psi(\Lambda^{-1}(q-a)), \quad (4.5)$$

we can immediately extend (4.3) and (4.4) to arbitrary values of q and $\mathbf{p}$ by replacing in these relations $\hat{\psi}$ with

$\hat{\psi}' = \hat{U}(0, \Lambda_p^{-1})\hat{\psi}$:

$$\psi(q) = \lim_{s\to +0} \left(\frac{m^2c^2h^3}{2s^6}\right)^{1/4} \psi^{(s)}(q;\mathbf{p}), \quad (4.6)$$

$$\hat{\psi}(\mathbf{p}) = \exp\left(\frac{i}{\hbar} q\cdot p\right) \lim_{s\to +\infty} \left(\frac{m^2c^2h^3s^6}{8}\right)^{1/4} \psi^{(s)}(q;\mathbf{p}). \quad (4.7)$$

Hence we can state the following theorem, which confirms the consistency of the present probability-amplitude interpretation of $\psi^{(s)}(q;\mathbf{p})$ with the probability amplitude interpretation of $\hat{\psi}(\mathbf{p})$.

Theorem 4.1: Let the probability amplitude $\hat{\psi}(\mathbf{k})$ in momentum space be a continuous function belonging to $L_\mu^2(\mathbf{R}^3) \cap L_\mu^1(\mathbf{R}^3)$, and let $\psi(x)$ be the scalar quantity defined in (3.18). If

$$\rho^{(s)}(q;\mathbf{p}) = |\psi^{(s)}(q;\mathbf{p})|^2 \quad (4.8)$$

is the probability density in the optimal relativistic stochastic phase space $\Gamma_{e,s}^{(m)}$ defined by (1.13) and (1.14), where

$$\chi_0^{(m,s)}(\mathbf{x}) = (\pi\hbar s^2)^{-3/2} \exp\left(-\frac{\mathbf{x}^2}{\hbar s^2}\right), \quad (4.9)$$

$$\hat{\chi}_0^{(m,s)}(\mathbf{k}) = (\pi\hbar s^{-2})^{-3/2} \exp\left(-\frac{s^2}{\hbar}\mathbf{k}^2\right), \quad (4.10)$$

and if $\psi^{(s)}(q;\mathbf{p})$ is related to $\hat{\psi}(\mathbf{k})$ by (2.26) and (3.16), then

$$|\psi(q)|^2 = 2mc \lim_{s\to +0} (\pi\hbar s^{-2})^{3/2} \rho^{(s)}(q;\mathbf{p}), \quad (4.11)$$

$$|\hat{\psi}(\mathbf{p})|^2 = mc \lim_{s\to +\infty} (\pi\hbar s^2)^{3/2} \rho^{(s)}(q;\mathbf{p}). \quad (4.12)$$

It is significant to note that although the function $\psi(x)$ in (3.18) cannot be interpreted[10] as a probability amplitude since its L^2-norm is not state-independent, its square being in fact

$$\int |\psi(x)|^2 d\mathbf{x} = \tfrac{1}{2}\int |\hat{\psi}(\mathbf{k})|^2 k_0^{-2} d\mathbf{k}, \quad (4.13)$$

it is nevertheless the limit of bona fide probability amplitudes on $\Gamma_{e,s}^{(m)}$ multiplied by the renormalization factors appearing in (4.6). The physical reason as to why $|\psi(q)|^2$ itself cannot represent a probability density on configuration space, whereas $|\psi^{(s)}(q;\mathbf{p})|^2$ does represent a probability density of phase space, is very simple: On one hand both these quantities are scalars, but on the other hand $d\mathbf{q}\,d\mathbf{p}$ is in a local-frame sense[4] an invariant under $L_+^\uparrow$, whereas $d\mathbf{q}$ itself is not (cf. the Appendix for details).

The preceding theorem suggests that one should introduce a probability current $j_e^\nu(q^0, \mathbf{q})$ at the stochastic configuration point $(\mathbf{q}, \chi_q^{(m)})$ in a manner analogous to that used in the nonrelativistic case.[12] Hence we set

$$j_e^\nu(q) = \int p^\nu \rho_e(q;\mathbf{p})\, d\mu(\mathbf{p}) \quad (4.14)$$

where the integration in the present instance is performed with respect to $d\mu(\mathbf{p})$ since in the relativistic case $p_0^{-1}d\mathbf{p}$ and not $d\mathbf{p}$ is the invariant momentum-space measure. We note, however, that

$$j_e^0(q) = \int \rho_e(q;\mathbf{p})\, d\mathbf{p}, \quad (4.15)$$

so that the zeroth component of this 4-current is related to the probability density $\rho_e(q;\mathbf{p})$ by exactly the same formula that holds in the nonrelativistic context.[12] Furthermore, if $j_e^\nu(q)$ corresponds to a wavepacket $\hat{\psi}(\mathbf{k})$ whose support is concentrated in some neighborhood of the origin where the last term in (2.8) can be neglected, and if the confidence function $\hat{\chi}_0(\mathbf{k})$ [related by (2.1) to the common generator of the two representation] also shares this feature, then (2.17) can be applied, and we thus infer that $j_e^0(q)$ approximately equals its nonrelativistic counterpart defined by Eq. (5.2) in Ref. 12. On the other hand, under the same circumstances, the space components $j_e^\nu(q)$, $\nu = 1,2,3$, of the relativistic 4-current become approximately equal to their nonrelativistic counterparts defined in Eq. (5.3) of Ref. 3 multiplied by the factor c^{-1}—the presence of this factor being required by the relativistic version of the equation of continuity that is derived below, which contains a derivative with respect to q^0 instead of a derivative with respect to the time variable t.

Theorem 4.2: In any relativistically covariant and extremal stochastic phase space representation with generator $e(mck^0)$ the 4-current (4.14) transforms as a 4-vector,

$$j_e'^\kappa(q) = \Lambda^\kappa{}_\nu\, j_e^\nu(\Lambda^{-1}q), \quad (4.16)$$

$$j_e'^\nu(q) = \int p^\nu |\psi_e'(q;\mathbf{p})|^2 d\mu(\mathbf{p}), \quad \psi' = U(0,\Lambda)\psi, \quad (4.17)$$

under proper Lorentz transformations. If in addition $e(mck^0)$ is a real function, then this current is conserved:

$$\frac{\partial}{\partial q^\nu} j_e^\nu(q) = 0. \quad (4.18)$$

Proof: The transformation law (4.16) is an immediate consequence of (2.32) and of the invariance of $d\mu(\mathbf{p})$ under $\mathcal{P}_+^\uparrow$:

$$j_e'^\kappa(q)$$

$$= \int p^\kappa |\psi(\Lambda^{-1}q;\Lambda^{-1}\mathbf{p})|^2 d\mu(\mathbf{p}) \quad (4.19)$$

$$= \int (\Lambda p')^\kappa |\psi_e(\Lambda^{-1}q;\mathbf{p}')|^2 d\mu(\mathbf{p}').$$

The verify (4.18), we express the lhs of that equation in the form

$$2\hbar^{-1}\mathrm{Im}\int d\mu(\mathbf{p})\psi_e^*(q;\mathbf{p})\int d\mu(\mathbf{k})(p\cdot k)\hat{e}_{q;\mathbf{p}}^*(\mathbf{k})\hat{\psi}(\mathbf{k}). \quad (4.20)$$

Upon introducing the wavefunctions $\hat{\psi}_q = \hat{U}(q, I)\hat{\psi}$, and then interchanging in (4.20) the orders of integration, we get

$$\frac{\partial}{\partial q^\nu} j_e^\nu(q) = \mathrm{Im}\int d\mu(\mathbf{k})\hat{\psi}_q(\mathbf{k})\int d\mu(\mathbf{k}')B(\mathbf{k},\mathbf{k}')\hat{\psi}_q^*(\mathbf{k}'), \quad (4.21)$$

$$B(\mathbf{k},\mathbf{k}') = 4\pi mch^{-4}\int (p\cdot k)e(p\cdot k)e(p\cdot k')d\mu(\mathbf{p}), \quad (4.22)$$

where, in writing (4.22), we have taken into account the reality condition imposed in the function $e(mck^0)$.

Since $d\mu(\mathbf{p})$ is an invariant under any $\Lambda \in L_+^\uparrow$, we have

$$B(\Lambda\mathbf{k}, \Lambda\mathbf{k}') = B(\mathbf{k},\mathbf{k}'). \quad (4.23)$$

Hence, as a function of the 4-momenta k and k', $B(\mathbf{k},\mathbf{k}')$ depends de facto only on $k\cdot k'$. Consequently, in addition to being real,

$$B(\mathbf{k},\mathbf{k}') = B(\mathbf{k}',\mathbf{k}), \tag{4.24}$$

and therefore $B(\mathbf{k},\mathbf{k}')$ is the kernel of a symmetric integral operator. But this implies that the integral on the rhs of (4.21) is real, and therefore that (4.18) is true. Q. E. D.

The optimal stochastic phase-space representations, for which $\hat{e}_{q;\mathbf{p}}$ is given by (3.16), satisfy all the conditions of the above theorem, and therefore give rise to the covariant and conserved probability currents

$$j^{(s)}_{\nu}(q) = \int \frac{p_\nu}{p_0} \rho^{(s)}(q;\mathbf{p})\, d\mathbf{p}. \tag{4.25}$$

It might be expected that, just in the nonrelativistic case,[12] in the limit $s \to +0$ the above currents go over into the conventional configuration space current $J_\nu(q)$, which in case of the Klein–Gordon equation has the components[10] (modulo at will chosen factors)

$$J^\nu(q) = i\hbar\psi^*(q)\frac{\overleftrightarrow{\partial}}{\partial q_\nu}\psi(q). \tag{4.26}$$

That this cannot be so—and that consequently the remarkable parallelism between the nonrelativistic and relativistic cases that was very much in evidence until now breaks down at this point—follows from the fact that $j^{(s)}_0(q) \geqslant 0$ at every space–time point q, whereas that is by no means true of $J_0(q)$, which is an indefinite real function.[10] Furthermore, although $\rho^{(s)}(q;\mathbf{p})$ multiplied by an appropriate renormalization factor approximates $|\psi(q)|^2$ in accordance to (4.11), that is not true of $j^{(s)}_0(q)$ in relation to $J_0(x)$ regardless of what kind of renormalization factor is employed.

To prove that, we shall demonstrate that the difference between $J_0(q)$ and $j^{(s)}_0(q)$ approaches a well-defined but nonzero limit as $s \to +0$. Indeed, reexpressing the rhs of (4.26) in the momentum representation, we get for $\hat{\psi} \in L^2_\mu \cap L^1_\mu$

$$J_\nu(q) = \frac{1}{2h^3}\int \psi^*_q(\mathbf{k}')\psi_q(\mathbf{k})\frac{k'_\nu + k_\nu}{k'_0 k_0}\, d\mathbf{k}'d\mathbf{k}. \tag{4.27}$$

On the other hand, by (3.16)

$$j^{(s)}_\nu(q) = \frac{mc}{h^3}(\pi\hbar s^{-2})^{-3/2}\exp\left(\frac{m^2c^2s^2}{\hbar}\right)\int\frac{d\mathbf{k}'d\mathbf{k}}{k'_0k_0}\hat{\psi}^*_q(\mathbf{k}')\hat{\psi}_q(\mathbf{k}) \tag{4.28}$$
$$\times\int d\mathbf{p}\frac{p_\nu}{p_0}\exp\left\{-\frac{s^2}{2m^2c^2\hbar}[(k'\cdot p)^2 + (k\cdot p)^2]\right\}.$$

Using the same technique as in deriving (3.4), we obtain

$$(\pi\hbar s^{-2})^{-3/2}\int\exp\left[-\frac{s^2}{m^2c^2\hbar}(k\cdot p)^2 + \frac{mc^2s^2}{\hbar}\right]d\mathbf{p} = \frac{k_0}{mc}, \tag{4.29}$$

and consequently

$$J_0(q) - j^{(s)}_0(q) = \frac{2^{1/2}mcs^3}{h^{3/2}}\exp\left(\frac{m^2c^2s^2}{\hbar}\right)\int\frac{d\mathbf{k}'d\mathbf{k}}{k'_0k_0}\hat{\psi}^*_q(\mathbf{k}')\hat{\psi}_q(\mathbf{k}) \tag{4.30}$$
$$\times\int d\mathbf{p}\left\{\exp\left[-\frac{s^2(k\cdot p)^2}{2m^2c^2\hbar}\right] - \exp\left[-\frac{s^2(k\cdot p)^2}{2m^2c^2\hbar}\right]\right\}^2.$$

After substituting $s\mathbf{p}$ in place of $\mathbf{p}$ as variable of integration, the limit $s \to +0$ can be easily taken, and we obtain

$$\lim_{s\to+0}[J_0(q) - j^{(s)}_0(q)] = \int d\mu(\mathbf{k}')\hat{\psi}^*(\mathbf{k}')\int d\mu(\mathbf{k})\exp\left[\frac{i}{\hbar}q\cdot(\mathbf{k}'-\mathbf{k})\right]D(\mathbf{k}',\mathbf{k})\hat{\psi}(\mathbf{k}), \tag{4.31}$$

where $D(\mathbf{k}',\mathbf{k})$ is the real and symmetric kernel of an integral operator in $L^2_\mu(R^3)$. From (4.30) it follows that

$$D(\mathbf{k}',\mathbf{k}) = \frac{mc}{4(\pi\hbar)^{3/2}}\int\left\{\exp\left[-\frac{(k'\cdot l)^2}{2\hbar m^2c^2}\right] - \exp\left[-\frac{(k\cdot l)^2}{2\hbar m^2c^2}\right]\right\}^2 d\mathbf{l}, \tag{4.32}$$

where l is on the light cone, i.e., $l^0 = |\mathbf{l}|$.

The integral operator in question is not positive definite despite the fact that $D(\mathbf{k}',\mathbf{k}) \geqslant 0$. In fact, the rhs of (4.30) must at some points assume negative values since both $J_0(q)$ and $j^{(s)}_0(q)$ are normalized to the value one in configuration space, so that the integral in $\mathbf{q}\in\mathbb{R}^3$ of both sides of (4.30) must yield zero for any $q^0\in\mathbb{R}^1$. Clearly, this observation can be then extended to (4.31) in the process of taking the limit $s \to +0$.

In physical terms, the fact that the limit (4.31) does not equal zero has a simple interpretation: $j^{(s)}_0(q)$ is a bona fide probability density, and that feature is being retained in the limit $s \to +0$, whereas, as generally agreed,[10] $J_0(q)$ can be consistently interpreted not as a probability density of a one-particle system, but rather as a charge density in a many-body version of the theory. Therefore we shall call $J_\nu(q)$ a "charge current," to distinguish it from the probability currents $j^\nu_e(q)$ and the limit

$$j_\nu(q) = \lim_{s\to+0} j^{(s)}_\nu(q). \tag{4.33}$$

We can also introduce a charge current $J^\nu_e(q)$ at the stochastic configuration point $(\mathbf{q}, \chi^{(m)}_q)$ by setting in analogy with (4.26)

$$J^\nu_e(q) = N_e^{-1}\int\psi^*_e(q;p)\frac{\overleftrightarrow{\partial}}{\partial q_\nu}\psi_e(q;\mathbf{p})\, d\mu(\mathbf{p}), \tag{4.34}$$

where the normalization constant N_e is the one provided by (3.21), and is chosen with (3.20) in mind.

Theorem 4.3: The charge current (4.34) is covariant and conserved, and

$$J^\nu(q) = \lim_{s\to+0} J^\nu_{e^{(s)}}(q). \tag{4.35}$$

Proof: The fact that $J^\nu_e(q)$ transforms as a 4-vector under $(a,\Lambda)\in\mathcal{P}^\uparrow_+$ is an immediate consequence of (2.32). Since (2.26) and (2.28) imply that $\psi_e(q;\mathbf{p})$ satisfies the

Klein—Gordon equation in the variable q, it follows that $J_e^\nu(q)$ is conserved.

To establish (4.35), we use (2.26), (3.14) and (3.21) to derive that

$$J_e^\nu(q)=\frac{1}{2h^3}\int \hat{\psi}_q^*(\mathbf{k}')\hat{\psi}_q(\mathbf{k})\frac{k'^\nu+k^\nu}{k_0'k_0}[1-F_e(\mathbf{k},\mathbf{k}')]\,d\mathbf{k}d\mathbf{k}', \tag{4.36}$$

$$1-F_e(\mathbf{k},\mathbf{k}')=\|e\|_\mu^{-2}\int e^*(p\cdot k)e(p\cdot k')\,d\mu(\mathbf{p}), \tag{4.37}$$

where $\|e\|_\mu$ denotes the L^2_μ-norm of $\tilde{e}_{\mathbf{0},\mathbf{0}}(\Lambda_k^{-1}\mathbf{p})=e(p\cdot k)$. For the optimal case presented in (3.16) we have

$$F_{e^{(s)}}(\mathbf{k},\mathbf{k}')=\left[\int \exp\left(\frac{s^2p_0^2}{\hbar}\right)d\mu(\mathbf{p})\right]^{-1}$$
$$\times\int\left\{\exp\left[-\frac{s^2(k\cdot p)^2}{2m^2c^2\hbar}\right]-\exp\left[-\frac{s^2(k'\cdot p)^2}{2m^2c^2\hbar}\right]\right\}^2$$
$$\times d\mu(\mathbf{p}). \tag{4.38}$$

Since the integral in (4.37) is an inner product, by employing the Schwartz—Cauchy inequality we get that $|F_e(\mathbf{k},\mathbf{k}')|\le 2$ for all $\mathbf{k},\mathbf{k}'\in\mathbb{R}^3$, and for any generator $e(mck_0)$. The integrand of the second integral in (4.38) is nonnegative and approaches zero pointwise as $s\to+0$. Hence, by Fatou's lemma, we conclude that $F_{e^{(s)}}(\mathbf{k},\mathbf{k}')\to 0$ as $s\to+0$ for all $\mathbf{k},\mathbf{k}'\in\mathbb{R}^3$. Comparing (4.36) and (4.27) we see that, due to Lebesgue's bounded convergence theorem, (4.35) indeed follows. Q.E.D.

5. CONCLUSIONS

The preceding considerations show that a relativistically covariant formalism on stochastic phase space is feasible, but that the notion of nonsharp point localizability in configuration space that it implicitly gives rise to does not lend itself to a limiting procedure that would yield a covariant formalism of sharp point localizability. The inescapable conclusion is that the concept of *sharp* point localizability in configuration space is in the quantum context at odds with the principles of relativistic covariance. The inconsistency stems from the lack of invariance of the volume element $d\mathbf{q}$ under the action of pure Lorentz transformations, and it points out the essential role that phase space should be assigned in relativistic quantum mechanics on account of the covariance of the probability densities in (4.8).

This role can be realized once the notion of nonsharp simultaneous measurement of position and momentum is introduced in quantum mechanics after the notion of probability space is generalized to that of stochastic probability space.[1] The fact that the resulting covariant formalism cannot lead to the existence of a covariant set of position operators, although it is totally in keeping with the existence of a 4-momentum set of operators P^ν, $\nu=0,\ldots,3$, that transform covariantly under Lorentz transformations, becomes obvious once it is recalled that for any free particle $\hat{P}^0$ is a function of $\hat{P}$, whereas the derivation of space—time densities from a spectral measure in space—time would necessitate the existence a set of independent and covariantly transforming space—time observables Q^ν, $\nu=0,1,2,3$, and therefore the association with each quantum state of probability measures that are finite over space—time.

The conclusion that in relativistic quantum mechanics only localizability at spread-out (but not at sharp) stochastic configuration-space points is meaningful is in total agreement with such well-known general features of positive-energy solutions $\psi(x)$ in (3.18) as the impossibility of localizing at any instant the probability amplitudes[2,10]

$$\Psi(x)=h^{-3/2}\int \exp\left(-\frac{i}{\hbar}x\cdot k\right)\hat{\psi}(\mathbf{k})k_0^{-1/2}\,d\mathbf{k} \tag{5.1}$$

with respect to the Newton—Wigner operator

$$\mathbf{X}=i\hbar\left(\nabla_{\mathbf{k}}-\frac{\mathbf{k}}{2k^0}\right) \tag{5.2}$$

by setting them equal to zero together with their time derivatives outside arbitrarily small regions[13]—whereas the inclusion of negative-energy solutions into the picture gives rise to Zitterbewegung (cf. Ref. 13, p. 65). The existence of the Newton—Wigner operators themselves does not contradict this conclusion. On the contrary, the sixtuple $\mathbf{X}$, $\mathbf{P}$ consisting of these operators and of the space part $\mathbf{P}$ of the relativistic 4-momentum operator provide an irreducible representation of the canonical commutation relations on $L^2_\mu(\mathbb{R}^3)$, and the unitary operator relating this representation to the nonrelativistic Schrödinger representation on $L^2(\mathbb{R}^3)$ supplies a transformation that injects the nonrelativistic formalism on $L^2(\mathbb{R}^3)$ into the Hilbert space $L^2_\mu(\mathbb{R}^3)$ [cf. Eqs. (2.7)—(2.10) in Ref. 2] *without* taking into account the requirements of relativistic covariance. Indeed, the formalism of Sec. 2 in Ref. 2 could be duplicated with $\phi^{(s)}_{\mathbf{q},\mathbf{p}}$ of Eq. (2.3) in Ref. 2 replaced by $\hat{e}_{q;\mathbf{p}}$ introduced in (2.28), thus attaching to each $\hat{\psi}(k)\in L^2_\mu(\mathbb{R}^3)$ the function

$$\Psi_e(q;\mathbf{p})=h^{-3/2}\int \hat{e}^*_{q;\mathbf{p}}(\mathbf{k})\hat{\psi}(\mathbf{k})k_0^{-1/2}\,d\mathbf{k} \tag{5.3}$$

in accordance with (5.1) and with Eq. (2.15) of Ref. 2. This arrangement would indeed provide a continuous resolution of the identity, since it is easily verified that

$$\int_{\mathbb{R}^6}|\Psi_e(q;\mathbf{p})|^2\,d\mathbf{q}d\mathbf{p}=mc\int_{\mathbb{R}^3}|\hat{\psi}(\mathbf{k})|^2\,d\mu(\mathbf{k}). \tag{5.4}$$

However, $\Psi_e(q;\mathbf{p})$ does not transform covariantly under the action of $P_+^\uparrow$. Instead, $\Psi_e(q;\mathbf{p})$ provides an expression which assumes the form of a nonrelativistic wavefunction in the nonrelativistic region (i.e., when $\mathbf{p}\approx m\mathbf{u}$ and when $|\tilde{e}_{\mathbf{0},\mathbf{0}}(\mathbf{k})|^2$ is sharply peaked around the value $\mathbf{k}=\mathbf{0}$). In fact, by (2.8),

$$\Psi_e(q;\mathbf{p})\approx h^{-3/2}\int \exp\left(-\frac{i}{\hbar}q\cdot k\right)\ \tilde{e}^*_{\mathbf{0},0}(\mathbf{k}-\mathbf{p})\hat{\psi}(\mathbf{k})\,d\mathbf{k}$$
$$=h^{-3/2}\int \exp\ \frac{i}{\hbar}\mathbf{q}\cdot\mathbf{k}\ \ \tilde{e}^*_{\mathbf{0},0}(\mathbf{k}-\mathbf{p})(U_t\psi)\hat{}(\mathbf{k})\,d\mathbf{k}, \tag{5.5}$$

and the above expression is easily recognized as being equal to the nonrelativistic probability amplitude (2.7) of Ref. 5.

The first part of Theorem 2.1 can be interpreted to mean that the operator $U_e(q^0)$ defined in (2.30) supplies a unitary mapping of $L^2_\mu(\mathbb{R}^3)$ onto a closed subspace $\mathcal{M}_e$ of $L^2(\Gamma)$. Thus, the entire theory dealing with the embedding into $L^2(\Gamma)$ of the wavefunction originating with extremal stochastic phase space representations (developed in Ref. 5) can be duplicated in the present relativistic context. Its general features, such as the existence of reproducing integral kernels in terms of which the orthogonal projectors $\mathbf{P}_e$ onto each subspace $\mathcal{M}_e$ can be expressed as integral operators, remain totally unchanged in the present context.

An essential difference emerges, however, as soon as it is realized that, in accordance with (2.32), in the present context, we are faced with a global representation of the Poincaré rather than of the Galilei group. As an immediate consequence, the time variable cannot be treated independently from the space variables, as was the case in the nonrelativistic context. Thus, to every element $\psi(\mathbf{q},\mathbf{p})$ of $L^2(\Gamma)$ we have to assign a unique function $\psi(q;\mathbf{p})$ of the space–time variable q in such a manner that the mapping

$$U(a,\Lambda):\ \psi(q;\mathbf{p})\mapsto\psi'(q;\mathbf{p})=\psi(\Lambda^{-1}(q-a);\ \Lambda^{-1}\mathbf{p}), \tag{5.6}$$

which obviously provides a global representation of $\mathcal{P}_+^{\uparrow}$ on $L^2(\Gamma)$, leaves each $\mathcal{M}_e$ invariant, inducing in each $\mathcal{M}_e$ the irreducible representation $U_e(a,\Lambda)$ defined by (2.32).

By analogy with Eq. (3.23) in Ref. 5, let us introduce in $L^2(\Gamma)$ the self-adjoint operators

$$P_\nu=-i\hbar\frac{\partial}{\partial q^\nu},\quad \nu=1,2,3, \tag{5.7}$$

$$P_0=(P_1^2+P_2^2+P_3^2+m^2c^2)^{1/2}. \tag{5.8}$$

It is easily verified by using (2.26) and (2.28) that the restrictions of P_ν, $\nu=0,\ldots,3$, to each $\mathcal{M}_e$ coincide with $U_e(0)\hat{P}_\nu U_e^{-1}(0)$, where the operators $\hat{P}_\nu$,

$$(\hat{P}_\nu\hat{\psi})(\mathbf{k})=k_\nu\hat{\psi}(\mathbf{k}),\quad \nu=0,\ldots,3, \tag{5.9}$$

are the components of the 4-momentum in $L^2_\mu(\mathbb{R}^3)$. After setting, by definition,

$$\psi(q;\mathbf{p})=\exp\left(-\frac{i}{\hbar}P_0q^0\right)\ \psi(\mathbf{q},\mathbf{p}), \tag{5.10}$$

for all $\psi(\mathbf{q},\mathbf{p})\in L^2(\Gamma)$, we arrive at the earlier mentioned extension of each element of $L^2(\Gamma)$ to a function on $\mathbb{R}^4\times V_+^{(m)}$. This function is a solution of the Klein–Gordon equation in the space–time variables q:

$$\left(\Box_q+\frac{m^2c^2}{\hbar^2}\right)\ \psi(q;\mathbf{p})=0. \tag{5.11}$$

Since the time evolution in (5.10) leaves $\mathcal{M}_e$ invariant and coincides there with the time evolution determined by $U_e(0)\hat{U}_tU_e(0)$, where $\hat{U}_t$ is defined in (2.14), we can immediately conclude by comparing (2.32) with (5.6) that $U_e\ (a,\Lambda)$ coincides with the restriction of $U(a,\Lambda)$ to $\mathcal{M}_e$.

Finally, a comment of an epistemological nature concerning the treatment of the time variable in the present stochastic phase space formalism. It might seem that since all position measurements considered in this formalism are of spread-out rather than sharp stochastic values of the position of the particle, the same should apply to the time variable, whose measurement involves the measurement of position. However, this measurement of position relates to that of a clock mechanism, and can be made as precise as desired. Furthermore, the idealized limit leading to an infinite precision in the measurement of time can be always taken in principle, since there is no uncertainty principle forbidding that.[14] This possibility is reflected by the mathematical fact that in both nonrelativistic and relativistic quantum theories the time variable is treated purely as a parameter.

ACKNOWLEDGMENT

I am indebted to Dr. S. T. Ali for several enlightening observations concerning the kinematics of Lorentz transformations.

APPENDIX: THE PRINCIPLES OF RELATIVITY, UNCERTAINTY AND COMPLEMENTARITY IN MEASUREMENTS WITH COHERENT ARRAYS OF DETECTORS

In accordance with Bohr's ideas[15] on the role of the apparatus in quantum measurement, an elementary detector $\mathcal{G}(\chi_{q;\mathbf{p}}^{(m)})$ could be viewed as a classical device labeled at the instant $t=q^0/c$ by a well-defined position vector $\mathbf{q}$ and velocity vector $\mathbf{u}$ in the laboratory frame, and consequently related to the location $\mathbf{q}'$ in its own rest-frame. By (1.6)–(1.9) we have

$$q'_{\parallel}=\gamma(q_{\parallel}-\beta q_0),\quad q'_0=\gamma(q_0-\beta q_{\parallel}), \tag{A1}$$

$$\mathbf{q}'_{\perp}=\mathbf{q}_{\perp},\quad \beta=\frac{|\mathbf{u}|}{c}=\frac{|\mathbf{p}|}{p^0}. \tag{A2}$$

The relativististically invariant nature of the stochastic phase-space probabilities derivable from the density $|\psi_e(q;\mathbf{p})|^2$ and measured with a coherent array of such detectors is exemplified in the following kind of measurement: the would-be observer sets up in phase space a random sample of elementary detectors that constitute a subset of a coherent array, and concentrates on some elementary detector $\mathcal{G}(\chi_{q;\mathbf{p}}^{(m)})=\mathcal{G}(\chi_{q';0}^{(m)})$ at the detector's proper time $t'=q'_0/c$ (and therefore laboratory time $t=q_0/c$). Then he observes how many detectors $\mathcal{G}(\chi_{q'_0,\mathbf{q}';\hat{\mathbf{p}}'}^{(m)})$ with values of $\hat{\mathbf{p}}'$ within an infinitesimal volume $d\mathbf{p}$ around $\hat{\mathbf{p}}'=\mathbf{0}$ have been triggered at that instant t' within an infinitesimal region of volume $d\mathbf{q}'$ around the value $\mathbf{q}'$. Since

$$\hat{p}_{\parallel}=\gamma(\hat{p}'_{\parallel}+\beta\hat{p}'_0),\quad \hat{\mathbf{p}}_{\perp}=\hat{\mathbf{p}}'_{\perp}, \tag{A3}$$

in the laboratory frame the observer will measure

$$d\mathbf{p}=\left[\gamma\left(1+\beta\frac{\hat{p}'_{\parallel}}{\hat{p}'_0}\right)\right]_{\hat{p}'_{\parallel}=0}dp'_{\parallel}d\mathbf{p}'_{\perp}=\gamma d\mathbf{p}'. \tag{A4}$$

By measuring distances in the laboratory frame at the instant t, in accordance with (A1) he will get $dq_{\parallel}=\gamma^{-1}dq'_{\parallel}$, and therefore

$$d\mathbf{q}d\mathbf{p}=(\gamma^{-1}d\mathbf{q}')(\gamma d\mathbf{p}')=d\mathbf{q}'d\mathbf{p}'. \tag{A5}$$

Thus the proportion of affirmative readings he will obtain if the elementary detectors were triggered by an ensemble of free particles in one and the same state $\psi_e(q;\mathbf{p}) = \psi'_e(q';\mathbf{p}')$ is given by the expression

$$|\psi(q;\mathbf{p})|^2 d\mathbf{q}d\mathbf{p} = |\psi'(q';\mathbf{0})|^2 d\mathbf{q}'d\mathbf{p}' \quad \text{(A6)}$$

for the probability over the infinitesimal region $d\mathbf{q}d\mathbf{p}$ in $\Gamma_e^{(m)}$. The manifest covariance of (A6) is a consequence of the fact that the expression on the rhs of (A6) relates to a specific frame[4]—namely the rest frame of $\mathcal{G}(\chi^{(m)}_{q;\mathbf{p}})$—and is due to the phenomenon of relativistic length contraction that had led to (A5).

The uncertainty principle is reflected in the basic properties of the confidence functions $\chi^{(m)}_{q'}(\mathbf{x}')$ and $\hat{\chi}^{(m)}_{\mathbf{0}}(\mathbf{k}')$ of each $\mathcal{G}(\chi^{(m)}_{q;\mathbf{p}})$ in its own rest frame, namely that the product of their spreads[7] for any conjugate components of position and momentum is not smaller than $\hbar$.

The complementarity principle is reflected by the form $(\mathbf{q}', \chi^{(m)}_{q'}) \times (\mathbf{p}', \hat{\chi}^{(m)}_{\mathbf{0}})$ of a stochastic phase space point in the rest frame of the detector $\mathcal{G}(\chi^{(m)}_{q';\mathbf{0}})$ used in its measurement, and is implicit in the interpretation[8,11] of the confidence functions $\chi^{(m)}_{q'}(\mathbf{x}')$ and $\chi^{(m)}_{\mathbf{0}}(\mathbf{k}')$. Indeed, the expression $\chi^{(m)}_{q'}(\mathbf{x}')\chi^{(m)}_{\mathbf{0}}(\mathbf{k}')d\mathbf{x}'d\mathbf{k}'$ is *not* interpreted as the probability of being able to detect a quantum particle within the phase-space volume $d\mathbf{x}'d\mathbf{k}'$ if $\mathcal{G}(\chi^{(m)}_{q';\mathbf{0}})$ had been triggered (that would violate the uncertainty principle!), but in a different, operational sense, related to the accuracy calibration procedure of $\mathcal{G}(\chi^{(m)}_{q';\mathbf{0}})$ (cf. the Appendix of Ref. 11). Therefore, we shall refer to this expression as a *preponderancy*. This interpretation amounts to considering $\chi^{(m)}_{q'}(\mathbf{x}')d\mathbf{x}'$ by itself, and $\hat{\chi}^{(m)}_{\mathbf{p}'}(\mathbf{k}')d\mathbf{k}$ also by itself—but not their product—as probabilities related to the distribution in either configuration space or in momentum space of systems prepared under identical conditions with sharp values of position or momentum, respectively, after those systems have triggered identical replicas of $\mathcal{G}(\chi^{(m)}_{q';\mathbf{0}})$ under identical kinematical circumstances in the rest frame of those detectors. If that same procedure is considered in the laboratory frame, we again have to refer to $\chi^{(m)}_{q;\mathbf{p}}(\mathbf{x},\mathbf{k})d\mathbf{x}d\mathbf{k}$ as a preponderancy (rather than a probability), i.e., interpret it in the following sense: The expression

$$d\mathbf{k}\int \chi^{(m)}_{q;\mathbf{p}}(\mathbf{x},\mathbf{k})\,d\mathbf{x} = d\mathbf{k}'\int \chi^{(m)}_{q';\mathbf{0}}(\mathbf{x}',\mathbf{k}')\,d\mathbf{x}' \quad \text{(A7)}$$

represents the probability that $\mathcal{G}(\chi^{(m)}_{q;\mathbf{p}})$ had been triggered at the laboratory time $t = q^0/c$ by a particle whose laboratory-frame momentum was within the infinitesimal volume $d\mathbf{k}$ around the value $\mathbf{k}$, whereas the expression

$$d\mathbf{x}\int \chi^{(m)}_{q;\mathbf{p}}(\mathbf{x},\mathbf{k})\,d\mathbf{k} = d\mathbf{x}'\int \chi^{(m)}_{q';\mathbf{0}}(\mathbf{x}',\mathbf{k}')\,d\mathbf{k}' \quad \text{(A8)}$$

is the probability that $\mathcal{G}(\chi^{(m)}_{q;\mathbf{p}})$ had been trigered by a particle that at the laboratory time x^0/c, with x^0 given in (1.12), was within the infinitesimal volume $d\mathbf{x}$ around the laboratory-frame position $\mathbf{x}$. We note that the complementarity principle is implicit in the either-or nature of these statements.

The formula (A8) makes it clear why in the relativistic context $|\psi(q)|^2$ does not have the properties of a conserved probability density in configuration space although at any fixed value of $\mathbf{p}$ the probability densities $|\psi^{(s)}(q;\mathbf{p})|^2$, when renormalized in accordance with (4.6), do approach $|\psi(q)|^2$: the different choices of $\mathbf{p}$ impose different rest-frames for the elementary detectors used in the measurement of q, and therefore, despite the formal covariance of $|\psi(q)|^2$, the lack of covariance of $d\mathbf{q}$ is going to lead to conflicting values when the candidates $|\psi'(q')|^2 d\mathbf{q}'$ for probabilities are transferred from each one of these rest frames to the laboratory frame.

[1] S. T. Ali and E. Prugovečki, Int. J. Theor. Phys. **16**, 689 (1977).
[2] E. Prugovečki, J. Phys. A **9**, 1851 (1976).
[3] T. D. Newton and E. P. Wigner, Rev. Mod. Phys. **21**, 400 (1949).
[4] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), Sec. 22.6.
[5] S. T. Ali and E. Prugovečki, Physica **89** A, (1977) (In press).
[6] J. D. Jackson, *Classical Electrodynamics* (Wiley, New York 1975).
[7] S. T. Ali and E. Prugovečki, J. Math. Phys. **18**, 219 (1977).
[8] E. Prugovečki, J. Math. Phys. **17**, 517 (1976).
[9] E. B. Davies, *Quantum Theory of Open Systems*, (Academic, New York, 1976).
[10] S. S. Schweber, *Relativistic Quantum Field Theory* (Row—Peterson, Elmsford, N.Y., 1961).
[11] E. Prugovečki, J. Math. Phys. **17**, 1673 (1976).
[12] E. Prugovečki, Ann. Phys. (N.Y.) **110**, 102 (1978).
[13] E. Corinaldesi and F. Strocchi, *Relativistic Wave Mechanics*, (North-Holland, Amsterdam, 1963).
[14] Y. Aharonov and D. Bohm, Phys. Rev. **122**, 1649 (1961).
[15] N. Bohr, *Atomic Physics and Human Knowledge*, (Science, New York, 1961).

Stochastic phase space kinematics of the photon[a)]

Eduard Prugovečki

Department of Mathematics, University of Toronto, Toronto, Canada, M5S 1A1

A stochastic phase space description of photon states is obtained by attaching to each circular polarization mode a probability amplitude at stochastic phase space points that are frontally localized. The ensuing formalism gives rise to a probability density at such points that transforms as a scalar under proper Lorentz transformations. There also exist conserved covariant currents associated with these probability densities. The wavefunctions corresponding to all extremal stochastic phase space representations can be embedded in a single Hilbert space, and give rise there to irreducible representations of the proper Poincaré group $P_+^\uparrow$.

1. INTRODUCTION

It has been shown in a previous publication[1] (referred to hereafter as I) that the relativistic quantum theory of massive free particles can be embedded in a covariant manner in $L^2(\Gamma)$ space, and that the resulting formalism gives rise to covariant probability densities on stochastic phase space and to conserved and covariant probability currents on stochastic space— time. The question arises as to what extent these results can be extended to the case of particles of zero mass. In this note we shall study this problem in the context of the most important of zero-mass particles, namely the photon, but the technique itself is of general validity.

The difference in approach to stochastic phase-space representations in this paper as opposed to that in I stems from the lack of a rest frame for classical zero-mass particles. In quantum-mechanical terms, this means that for mass-zero particles there is no concept equivalent to that of an elementary detector *stochastically* at rest in relation to the detected quantum particle — a concept which played such an essential role in I in relating the relativistic to the nonrelativistic treatment of massive particles. Thus, a covariant theory of stochastic phase space for zero-mass particles can be developed exclusively by isolating and then translating into the phase-space formalism only those features of the relativistically covariant theory for massive particles which are shared by the zero-mass case, rather than by trying to achieve approximate concurrence at low laboratory-frame particle speeds between such a theory and nonrelativistic theories.

This kind of procedure cannot consist of something as straightforward as letting $m \to +0$ in the formula for the massive case. When such an approach is applied, for example, to the basic formula I(2.26) for the probability amplitude in stochastic phase space it yields an everywhere zero result. A renormalization procedure eliminating the presence of mass-dependent multiplicative factors from the expression I(2.28) for $\hat{e}_{q;p}(\mathbf{k})$ would not lead to any meaningful covariant expressions either. This becomes evident as soon as it is recalled that by Theorem 2.1 in I, the covariance of the probability amplitudes imposes rotational invariance on the generator $\tilde{e}_{0,0}(\mathbf{k})$ of the representation, and that in turn leads to I(3.6). However, in the zero-mass case we have

$$\int_{\mathbb{R}^3} |e(p\cdot k)|^2\, d\mathbf{p} \geq \tfrac{1}{2}\int_{\mathbb{R}^3} |e(p\cdot k)|^2 \left(1-\frac{p_\parallel}{p_0}\right) d\mathbf{p} = \tfrac{1}{2}\int_{\mathbb{R}^2} d\mathbf{p}_\perp \int_0^\infty |e(\lambda k^0)|^2\, d\lambda, \tag{1.1}$$

$$p_\parallel = \mathbf{p}\cdot\mathbf{k}\,|k|^{-1}, \quad \mathbf{p}_\perp = \mathbf{p} - p_\parallel \mathbf{k}\,|k|^{-1}. \tag{1.2}$$

The inequality (1.1) implies that any function $\hat{e}_{q;p}(k)$ of the form I(3.14) either vanishes almost everywhere or is not integrable over $\mathbb{R}^3$ in either the variables $\mathbf{p}$ or the variables $\mathbf{k}$. This invalidates the approach used in I(Theorem 3.1) for proving I(3.2). Further considerations show that the very kinematics of zero-mass particles interferes with any attempt to bypass this difficulty with any renormalization procedure based on, say, the introduction of momentum cutoffs. Indeed, any such procedure gives rise in the expression for the cutoff version of the integral in the lhs of (1.1) to terms containing not only the factor k_0 but also k_0^{-1}, so that no zero-mass counterpart of the relation I(3.4) (which played an essential role in the case of massive particles) can be found.

Thus, the approach to a covariant stochastic phase-space kinematics for a massless particle, such as the photon, requires a rather drastic reevaluation of the methods used in I, and it has to take into account the very specific behavior under Lorentz transformations of lightlike vectors. In mathematical terms, the solution turns out to lie in the formula

$$k_0' = \gamma\left(1 - \frac{|\mathbf{u}|}{c}\right) k_0 \tag{1.3}$$

describing the transformation of the zero component $k^0 = |\mathbf{k}|$ of such a vector under pure Lorentz boosts in the direction of $\mathbf{k}$. In physical terms, the procedure consists in adopting for prototypes elementary detectors which, in contradistinction to the nonzero-mass case, are of necessity in a state of motion with respect to the photon they are designed to detect. Hence, a different orientation of such a detector in relation to the laboratory frame has to be adopted for every stochastic value $(\mathbf{p}, \hat{\chi}_\mathbf{p})$ of some standard absolute value of the measured momentum, which for convenience we will choose to be centered at $|\mathbf{p}| = 1$. The protype detector for each direction ω of motion is then submitted to pure Lorentz boosts in the direction ω. After this procedure is re-

[a)]Supported in part by a grant from the National Research Council of Canada.

peated at all space—time points, it results in a coherent array of elementary detectors for each particular direction of motion. The stochastic phase space points themselves are only frontally localized, as opposed to the globally localized stochastic points used for massive particles.

In Sec. 2 we describe this procedure in detail for photons. Thus we arrive at zero-mass counterparts of the formulas (2.26)—(2.28) in I for wavefunctions on stochastic phase space. Then in Sec. 3 we prove that these formulas supply a continuous resolution of the identity. This establishes that these wavefunctions are indeed probability amplitudes on stochastic phase spaces. In Sec. 4 we prove that the resulting probability densities transform covariantly under proper Lorentz transformation, and that they give rise to covariant and conserved probability currents.

In the concluding section we consider the behavior of these expressions in the limit of sharp frontal stochastic values of position and momentum, respectively, and show that such limits exist and behave covariantly.

2. STOCHASTIC PHASE SPACES CONSISTING OF FRONTALLY LOCALIZED POINTS

In this section we shall construct stochastic phase spaces for the photon, on which probability densities that are covariant with respect to the proper Poincaré group $\mathcal{P}_+^\uparrow$ can be later assigned to every state of the photon. To achieve that, we shall follow as far as possible the kinematical principles employed in the case of particles of nonzero mass. Thus, we first set up a coherent array of elementary detectors for the photon.

In the nonzero-mass case, the starting point of the construction was the origin-based elementary detector, which served as the prototype to which operational procedures in keeping with the kinematics of the Poincaré group were applied in order to obtain the remaining detectors in the coherent array.[1] This detector was located at the origin of the laboratory frame, and it detected particles that in relation to that frame were at rest in the stochastic sense [i.e., whose momentum had the stochastic value $(\mathbf{p}, \hat{\chi}_\mathbf{p})$ centered at $\mathbf{p} = 0$]. However, since there is no rest frame for the photon treated as classical particle, we have to reconcile ourselves to adopting as a prototype an origin-based detector which determines some nonzero stochastic value of momentum. We take that to be a detector $\mathcal{D}(\chi_{0;\omega})$, set up to detect at the instant $t=0$ and at the stochastic phase-space point $(0,\omega;\chi_{0;\omega})$, a photon that is moving in the given direction ω. Since this direction of motion is determined sharply, by the uncertainty principle the confidence function of the determined stochastic phase-space value must have the form

$$\chi_{0,\omega}(\mathbf{x},\mathbf{k}) = \delta^2(\mathbf{k}_\perp)\chi_{0,1}^{(\omega)}(x_\parallel, k_\parallel), \tag{2.1}$$

$$k_\parallel = \mathbf{k}\cdot\omega, \quad \mathbf{k}_\perp = \mathbf{k} - k_\parallel \omega. \tag{2.2}$$

We shall refer to stochastic points $(\mathbf{q},\mathbf{p};\chi_{\mathbf{q},\mathbf{p}})$ whose confidence functions $\chi_{\mathbf{q},\mathbf{p}}(\mathbf{x},\mathbf{k})$ do not depend on $\mathbf{x}_\perp$ as *frontally localized*. The need of introducing elementary detectors whose confidence functions are frontally localized is suggested by previous work on the "front localizability" of the photon,[2,3] and is confirmed by later observations of a kinematical nature.

To obtain a coherent array of elementary detectors for photons moving in the direction ω, let us introduce (by analogy with the pure Lorentz transformation Λ_p used in the massive case) the pure Lorentz boost $\Lambda_{\omega,r}$ for which, by definition,

$$\Lambda_{\omega,r}\omega = r\omega, \quad \omega^0 = |\omega|, \tag{2.3}$$

at a fixed real number $r>0$. Thus $\Lambda_{\omega,r}$ boosts a zero-mass classical particle having 4-momentum ω into one with 4-momentum $p = r\omega$.

Let $\mathcal{L}_1^{(\omega)}$ denote the laboratory frame of reference and $\mathcal{L}_r^{(\omega)}$ the inertial frame of reference moving in relation to $\mathcal{L}_1^{(\omega)}$ in the direction ω at the speed $\mathbf{u}$ in such a manner that at the laboratory time $t=0$ and proper time $t'=0$ its origin coincides with the origin of the laboratory frame, and its Cartesian axes are so oriented that $\Lambda_{\omega,r}$ represents the pure Lorentz transformation from $\mathcal{L}_r^{(\omega)}$ to $\mathcal{L}_1^{(\omega)}$. If k_ν are the coordinates of a lightlike vector in $\mathcal{L}_1^{(\omega)}$, then the coordinates of the same vector equal[4]

$$k_0' = \gamma\left(k_0 - \frac{\mathbf{u}\cdot\mathbf{k}}{c}\right), \quad \gamma = \left(1 - \frac{\mathbf{u}^2}{c^2}\right)^{-1/2}, \tag{2.4}$$

$$k_\parallel' = \gamma\left(k_\parallel - \frac{k_0|\mathbf{u}|}{c}\right), \quad \mathbf{k}_\perp' = \mathbf{k}_\perp, \tag{2.5}$$

with respect to $\mathcal{L}_r^{(\omega)}$. To relate $\mathbf{u}$ to r and ω we note that (2.4) and (2.5) represent $\Lambda_{\omega,r}^{-1}$, so that by (2.3)

$$r = \gamma\left(\omega^0 + \frac{|\mathbf{u}|}{c}\right) = \left(\frac{c+|\mathbf{u}|}{c-|\mathbf{u}|}\right)^{1/2}, \tag{2.6}$$

and consequently

$$\mathbf{u} = c\,\frac{r^2-1}{r^2+1}\omega. \tag{2.7}$$

A duplicate of the prototype elementary detector placed at the origin of $\mathcal{L}_r^{(\omega)}$ at $t=0$ represents $\mathcal{D}(\chi_{0;\mathbf{p}})$ for $\mathbf{p} = r\omega$. More generally, $\mathcal{D}(\chi_{\mathbf{q};\mathbf{p}})$ is obtained by placing such a prototype at the point $\mathbf{q}'$ of $\mathcal{L}_r^{(\omega)}$ at the instant $t' = q_0'/c$, where $q' = \Lambda_{\omega,r}^{-1}q$. A *coherent array of elementary detectors for the direction* ω is obtained by carrying out this procedure for all $\mathbf{p} = r\omega$, where r is real and different from zero.

The family of frontally localized stochastic phase space points resulting from this procedure is

$$\mathcal{G}_e^{(\omega)} = \{(\mathbf{q},\mathbf{p};\chi_{\mathbf{q};\mathbf{p}}) \mid \mathbf{q}\in\mathbb{R}^3, \ \mathbf{p} = r\omega, \ 0 < r < +\infty\}, \tag{2.8}$$

where, if we denote by $\Lambda_{\omega,r}^{-1}(\mathbf{x}-\mathbf{q})$ and $\Lambda_{\omega,r}^{-1}\mathbf{k}$ the space components of the 4-vectors obtained by applying $\Lambda_{\omega,r}^{-1}$ to the spacelike vector $x-q$ with $(\Lambda_{\omega,r}x)^0 = (\Lambda_{\omega,r}q)^0$ and to the lightlike vector k, respectively, we have

$$\frac{2}{1+r^2}\chi_{q;p}(\mathbf{x},\mathbf{k}) = \chi_{0,\omega}(\Lambda_{\omega,r}^{-1}(\mathbf{x}-\mathbf{q}), \ \Lambda_{\omega,r}^{-1}\mathbf{k}). \tag{2.9}$$

As in the nonzero-mass case, the factor $2(1+r^2)^{-1}$ that appears in (2.9) is the Jacobian due to the change from the variables of integration $x_\parallel'$ and $k_\parallel'$ to the variables $x_\parallel$ and $k_\parallel$, at fixed $\mathbf{x}_\perp' = \mathbf{x}_\perp$ and $\mathbf{k}_\perp' = \mathbf{k}_\perp = 0$. Its presence takes care of the correct front normalization of $\chi_{q;p}(\mathbf{x},\mathbf{k})$.

The set $\mathcal{G}_e^{(\omega)}$ in (2.8) does not as yet constitute a stochastic phase space since not all choices of $\mathbf{p}\in\mathbb{R}^3$

have been realized. By repeating the procedure for every vector ω on the unit sphere in $\mathbb{R}^3$ we do arrive, however, at the stochastic phase space

$$\Gamma_e^{(0)} = \bigcup_{|\omega|=1} \mathcal{G}_e^{(\omega)}, \tag{2.10}$$

which contains frontally localized stochastic points $(\mathbf{q},\mathbf{p};\chi_{\mathbf{q};\mathbf{p}})$ centered at all values $(\mathbf{q},\mathbf{p})\in\Gamma$ except those with $|\mathbf{p}|=0$. The lack of elementary detectors for stochastic values of momentum centered at zero is, however, completely in keeping with the nonexistence of rest frames for zero-mass classical particles, and therefore with the lack of frames which would be stochastically at rest with respect to zero-mass quantum particles (such as the photon).

Following the pattern established in I, we shall attach now to every point of $\Gamma_e^{(0)}$ and to every photon state an expression for a potential probability amplitude.

Let us denote by $\mathcal{J}_Q$ the conventional Hilbert space of the photon,[3] consisting of vector-valued functions $\mathbf{f}(\mathbf{k})$ that satisfy the gauge condition

$$\mathbf{k}\cdot\mathbf{f}(\mathbf{k})=0, \tag{2.11}$$

and in which the inner product equals

$$(\mathbf{f}|\mathbf{g})=\int \mathbf{f}^*(\mathbf{k})\cdot\mathbf{g}(\mathbf{k})k_0^{-1}\,d\mathbf{k}, \quad k_0=|\mathbf{k}|. \tag{2.12}$$

Let $\xi(\mathbf{k})$ and $\eta(\mathbf{k})$ be any two continuous functions of $\mathbf{k}$, assuming values in the unit sphere in $\mathbb{R}^3$ and such that ξ,η and $|\mathbf{k}|^{-1}\mathbf{k}$ are the unit vectors of a right-hand Cartesian coordinate system in $\mathbb{R}^3$. Then

$$f^{(\pm)}(\mathbf{k})=2^{-1/2}[\xi(\mathbf{k})\pm i\eta(\mathbf{k})]\cdot\mathbf{f}(\mathbf{k}) \tag{2.13}$$

represents[5] the right and left circular polarization amplitudes, respectively. Accordingly, we shall assume that every elementary detector $\mathcal{G}(\chi_{q;p})$ consists of two parts $\mathcal{G}^{(+)}(\chi_{q;p})$ and $\mathcal{G}^{(-)}(\chi_{q;p})$ that detect photons of right and left polarization, respectively.

Following the guidelines set up by the case of massive particles treated in I, we postulate the existence of a stochastic phase space representation generator

$$\tilde{e}_{\omega}(\mathbf{k})=\delta^2(\mathbf{k}_\perp)\tilde{e}_{\omega,1}(k_\parallel) \tag{2.14}$$

associated with the prototype elementary detector $\mathcal{G}(\chi_{0;\omega})$. Then, by analogy with I(2.9), we shall have

$$\chi_{0,1}^{(\omega)}(x_\parallel,k_\parallel)=|e_{\omega,1}(x_\parallel)|^2|\tilde{e}_{\omega,1}(k_\parallel)|^2, \tag{2.15}$$

$$\int_{-\infty}^{+\infty}|\tilde{e}_{\omega,1}(\mathbf{k}_\parallel)|^2\,dk_\parallel=1, \tag{2.16}$$

$$e_{\omega,1}(x_\parallel)=h^{-1/2}\int_{-\infty}^{+\infty}\exp\left(\frac{i}{\hbar}x_\parallel k_\parallel\right)\tilde{e}_{\omega,1}(k_\parallel)dk_\parallel. \tag{2.17}$$

In accordance with I(2.28), we introduce the two families

$$\mathbf{e}_{q;p}^{(\pm)}(\mathbf{k})=N_e\exp\left(\frac{i}{\hbar}q\cdot k\right)\tilde{e}_{\omega}(\Lambda_{\omega,p_0}^{-1}\mathbf{k})[\xi(\mathbf{k})\pm i\eta(\mathbf{k})] \tag{2.18}$$

of vector-valued pseudofunctions (namely, de facto, δ measures in the variables $\mathbf{k}_\perp$), in which N_e is a normalization constant to be determined in the next section. By analogy with I(2.26), we propose

$$f_e^{(\pm)}(\mathbf{q};\mathbf{p})=h^{-1/2}\,\mathrm{l.i.m.}\int_{\mathbb{R}^3}\mathbf{e}_{q;p}^{(\pm)*}(\mathbf{k})\cdot\mathbf{f}(\mathbf{k})k_0^{-1}\,d\mathbf{k}, \tag{2.19}$$

as candidates for probability amplitudes measured with $\mathcal{G}^{(\pm)}(\chi_{q;p})$.

The implicit limit in the mean with respect to $dq_\parallel$ is required by virtue of the fact that $\tilde{e}_{\omega,1}(\mathbf{k})$ contains a δ-function in $\mathbf{k}_\perp$. Thus, (2.19) has to be interpreted in terms of the limit

$$f_e^{(\pm)}(q;\mathbf{p})=\left(\frac{2}{h}\right)^{1/2}\mathrm{l.i.m.}_{S\downarrow\{0\}}\frac{N_e}{|S|}\int_S d\mathbf{k}_\perp \times\int_{-\infty}^{+\infty}d\mathbf{k}_\parallel k_0^{-1}\exp\left(-\frac{i}{\hbar}q\cdot k\right)\tilde{e}^*_{\omega,p_0}(\mathbf{k})f^{(\pm)}(\mathbf{k}), \tag{2.20}$$

where, by (2.5)–(2.7) and (2.18)

$$\tilde{e}_{\omega,r}(\mathbf{k})=\tilde{e}_{\omega,1}[\Lambda_{\omega,r}^{-1}k)_\parallel]=\tilde{e}_{\omega,1}\left(r\frac{k_\parallel-k_0}{2}+\frac{k_\parallel+k_0}{2r}\right), \tag{2.21}$$

and the limit is to be taken over any monotonically decreasing sequence $S_1\supset S_2\supset\cdots$ of Borel sets in $\mathbb{R}^2$ which have in common only the zero vector (and whose Lebesgue measures are denoted by $|S_n|$).

3. THE EXISTENCE OF STOCHASTIC PHASE SPACE PROBABILITY DENSITIES FOR THE PHOTON

The interpretation of the expressions in (2.19) as right and left polarization probability amplitudes on the stochastic phase space $\Gamma_e^{(0)}$ defined in (2.8) and (2.10) can be consistent if and only if the two functions

$$\rho_e^{(\pm)}(q;\mathbf{p})=|f_e^{(\pm)}(q;\mathbf{p})|^2 \tag{3.1}$$

add up to a relative probability density on $\Gamma=\mathbb{R}^6$, i.e., if

$$\lim_{S\uparrow\mathbb{R}^2}\frac{1}{|S|}\int_{\mathbb{R}^3}d\mathbf{p}\int_S^{+\infty}d\mathbf{q}_\perp\int_{-\infty}^{+\infty}d\mathbf{q}_\parallel[\rho_e^{(+)}(q;\mathbf{p})+\rho_e^{(-)}(q;\mathbf{p})=1, \tag{3.2}$$

for all $q^0\in\mathbb{R}^1$ and all photon states represented by a normalized element $\mathbf{f}(\mathbf{k})$ of $\mathcal{J}_0$. In (3.2), the limit can be taken on any monotonically increasing sequence $S_1\subset S_2\subset\cdots$ of Borel sets in $\mathbb{R}^2$ whose union equals $\mathbb{R}^2$. The need to renormalize the resulting integral in $\mathbf{q}_\perp$ over S by dividing it with the Lebesgue measure $|S|$ of S is a consequence of (2.1) and (2.14), which make $\rho_e^{(\pm)}(q;\mathbf{p})$ independent of

$$\mathbf{q}_\perp=\mathbf{q}-(\mathbf{q}\cdot\omega)\omega, \tag{3.3}$$

and therefore interpretable only as relative rather than absolute probability densities.

We shall prove now a theorem which, in view of the fact that

$$|\mathbf{f}(\mathbf{k})|^2=|\mathbf{f}^{(+)}(\mathbf{k})|^2+|\mathbf{f}^{(-)}(\mathbf{k})|^2, \tag{3.4}$$

establishes that (3.2) is satisfied provided the function $\tilde{e}_{\omega,1}(\mathbf{k}_\parallel)$ in (2.14) is ω-independent and N_e in (2.18) is appropriately chosen in relation to $\tilde{e}_{\omega,1}(\mathbf{k}_\parallel)$.

Theorem 3.1: suppose there is a continuous complex function $e(\lambda)$, $\lambda\in\mathbb{R}^1$, that vanishes for $\lambda\leq 0$ and is such that

$$\tilde{e}_{\omega,1}(k_\parallel)=\theta(k_\parallel)e(k_\parallel)=e(k_\parallel) \tag{3.5}$$

for all vectors ω on the unit sphere in $\mathbb{R}^3$, and furthermore that $e(\lambda)$ approaches zero as $\lambda \downarrow 0$ so rapidly that

$$\int_0^\infty \lambda^{-4} |e(\lambda)|^2 d\lambda < +\infty. \tag{3.6}$$

Then, for arbitrary f(k) and g(k) from the space $\mathcal{J}_0$ of momentum-representation photon wavefunctions, and for all $q^0 \in \mathbb{R}^1$,

$$\lim_{S \uparrow \mathbb{R}^2} \frac{1}{|S|} \int_{\mathbb{R}^6} \Theta_S^{(\omega)}(\mathbf{q}) f_e^{(\pm)*}(q;\mathbf{p}) g_e^{(\pm)}(q;\mathbf{p})\, d\mathbf{q} d\mathbf{p} = \int_{\mathbb{R}^3} f^{(\pm)*}(\mathbf{k}) g^{(\pm)}(\mathbf{k}) k_0^{-1}\, d\mathbf{k}, \tag{3.7}$$

where $\omega = p_0^{-1}\mathbf{p}, \Theta_S^{(\omega)}(\mathbf{q}) = 1$ if $\mathbf{q}_\perp \in S$ and $\Theta_S^{(\omega)}(\mathbf{q}) = 0$ otherwise, whereas N_e in (2.18) is chosen as follows:

$$N_e = \left(2\int_0^\infty |e(\lambda)|^2 \lambda^{-4} d\lambda\right)^{-1/2}. \tag{3.8}$$

Proof: Let us chose f and g so that $f^{(\pm)}$ and $g^{(\pm)}$ are continuous, as well as integrable and square-integrable with respect to $k_0^{-1}d\mathbf{k}$ for any unit vector ω. Then by (2.20), (2.21) and (3.5),

$$f_e^{(\pm)}(q;\mathbf{p}) = \left(\frac{2}{h}\right)^{1/2} N_e \int_0^\infty \exp\left[\frac{i}{\hbar}(q_\parallel - q_0)k^0\right] \times e^*\left(\frac{k_0}{p_0}\right) f^{(\pm)}(k^0\omega) k_0^{-1}\, dk_0, \tag{3.9}$$

and a similar expression holds true for $g_e^{(\pm)}(q;\mathbf{p})$. Consequently, expressing p in spherical coordinates, so that $d\mathbf{p} = r^2 dr d^2\omega$, and then performing in (3.7) the q integration first by writing $d\mathbf{q} = dq_\parallel d\mathbf{q}_\perp$ at each fixed value of ω, we obtain upon taking advantage of the unitarity property of the ensuing Fourier transform in the variable $q_\parallel$

$$\lim_{S \uparrow \mathbb{R}^2} \frac{1}{|S|} \int \Theta_S^{(\omega)}(\mathbf{q}) f_e^{(\pm)*}(q;\mathbf{p}) g_e^{(+)}(q;\mathbf{p})\, d\mathbf{q} d\mathbf{p} = 2N_e^2 \int d\omega \int_0^\infty r^2\, dr \int_0^\infty dk^0 \times k_0^{-2} f^{(\pm)*}(k^0\omega) g^{(\pm)}(k^0\omega) \left| e\left(\frac{k^0}{r}\right)\right|^2. \tag{3.10}$$

Reversing the order of integration in r and k^0, and noting that

$$\int_0^\infty \left| e\left(\frac{k^0}{r}\right)\right|^2 r^2\, dr = k_0^3 \int_0^\infty |e(\lambda)|^2 \lambda^{-4}\, d\lambda, \tag{3.11}$$

we obtain that the right-hand sides of (3.7) and (3.10) are equal if N_e is chosen as in (3.8). Since the set of all elements f(k) and g(k) of $\mathcal{J}_0$ that satisfy the conditions posed at the beginning of the proof is dense in $\mathcal{J}_0$, the conclusion immediately extends to arbitrary $\mathbf{f}, \mathbf{g} \in \mathcal{J}_0$.

Q.E.D.

Comparing (3.7) with I(3.2), we see that the essential mathematical difference between the massive and massless case resides in the non-square-integrability in space directions orthogonal to p of the mass-zero probability amplitudes. Hence, instead of the resolution of identity I(3.17) we now have

$$\lim_{S \uparrow \mathbb{R}^2} |S|^{-1} \int_{\mathbb{R}^6} d\mathbf{q} d\mathbf{p} \Theta_S^{(\omega)}(\mathbf{q}) \times \{ |\mathbf{e}_{q;\mathbf{p}}^{(+)})(\mathbf{e}_{q;\mathbf{p}}^{(+)}| + |\mathbf{e}_{q;\mathbf{p}}^{(-)})(\mathbf{e}_{q;\mathbf{p}}^{(-)}| \} = \mathbb{1}. \tag{3.12}$$

The result is completely in keeping with such a peculiar feature of the photon state-space $\mathcal{J}_0$ as the lack of Newton—Wigner position operators for the photon and the existence instead of operators for "front" localization.[2] It has, however, the unpleasant consequence that the functions $f_e^{(\pm)}(q;\mathbf{p})$ cannot be embedded in $L^2(\Gamma)$.

To formulate a substitute for $L^2(\Gamma)$ into which the photon probability amplitudes $f_e^{(\pm)}(q;\mathbf{p})$ on $\Gamma_e^{(0)}$ can be embedded, let us introduce on Γ the measure

$$d\sigma(\mathbf{q},\mathbf{p}) = p_0^2 dp_0 d\omega dq_\parallel d\mu_\delta(\mathbf{q}_\perp), \tag{3.13}$$

where the δ measure $d\mu_\delta(\mathbf{q}_\perp) = \delta^2(\mathbf{q}_\perp)d\mathbf{q}_\perp$ is chosen for the sake of convenience (any other finite measure on $\mathbb{R}^2$ would do). Then $f_e^{(+)} \oplus f_e^{(-)}$ could be regarded as an element in the space

$$\mathcal{J}(\Gamma) = \hat{L}_\sigma^2(\Gamma) \oplus \hat{L}_\sigma^2(\Gamma), \tag{3.14}$$

where $\hat{L}_\sigma^2(\mathbb{R}^6)$, as opposed to $L_\sigma^2(\mathbb{R}^6)$, is assumed to consist of functions $f(\mathbf{q},\mathbf{p})$ which are not only square-integrable on $\mathbb{R}^6$ with respect to the measure σ, but also $\mathbf{q}_\perp$-independent at each fixed nonzero value $\mathbf{p} \in \mathbb{R}^3$.

Theorem 3.2: The mapping

$$U_e(q_0) : \mathbf{f}(\mathbf{k}) \mapsto f_e^{(+)}(q,\mathbf{p}) \oplus f_e^{(-)}(q,\mathbf{p}) \tag{3.15}$$

is a unitary operator that maps the photon state space $\mathcal{J}_0$ onto a closed subspace U_e of $\mathcal{J}(\Gamma)$ in (3.14) which is q_0-independent.

Proof: Upon introducing the time-evolution operator $\hat{U}_t$ for a free photon,

$$(\hat{U}_t \mathbf{f})(\mathbf{k}) = \exp\left(-\frac{i}{\hbar} ctk_0\right)\mathbf{f}(\mathbf{k}), \tag{3.16}$$

we see that in accordance with (2.20)

$$U_e(q_0)\mathbf{f}(\mathbf{k}) = U_e(0)[(\hat{U}_t \mathbf{f})(\mathbf{k})] \tag{3.17}$$

Hence the range of $U_e(q_0)$ coincides at all $q_0 \in \mathbb{R}^1$ with the range of $U_e(0)$, which we denote by $\mathcal{M}_e$.

According to (2.12) and (2.13)

$$(\mathbf{f}|\mathbf{g}) = \int [f^{(+)*}(\mathbf{k}) g^{(+)}(\mathbf{k}) + f^{(-)*}(\mathbf{k}) g^{(-)}(\mathbf{k})] k_0^{-1}\, d\mathbf{k}. \tag{3.18}$$

Combining this result with (3.7) we get

$$(\mathbf{f}|\mathbf{g}) = \sum_{\alpha=\pm} \int f_e^{(\alpha)*}(q;\mathbf{p}) g_e^{(\alpha)}(q;\mathbf{p}) d\sigma(\mathbf{q},\mathbf{p}). \tag{3.19}$$

Hence $U_e(q^0)$ provides an isometrical mapping of $\mathcal{J}_0$ into $\mathcal{J}(\Gamma)$ at all $q^0 \in \mathbb{R}^1$, and consequently its range must coincide with a closed subspace of $\mathcal{J}(\Gamma)$. Q.E.D.

4. RELATIVISTIC COVARIANCE AND CURRENTS

The gauge condition (2.11) is not relativistically invariant. It corresponds, however, to the relativistically invariant Lorentz condition

$$k^\nu f_\nu(\mathbf{k}) = 0, \tag{4.1}$$

for the choice $f_0(\mathbf{k}) \equiv 0$. The functions $f_\nu(\mathbf{k})$, $\nu = 0, \cdots, 3$, transform as a 4-vector under proper Lorentz transformations. In fact, the family of operators $\hat{U}(a,\Lambda)$,

$$[\hat{U}(a,\Lambda)f]_\mu(\mathbf{k}) = \exp\left(-\frac{i}{\hbar}a\cdot k\right)\Lambda_\mu{}^\nu(\Lambda^{-1}\mathbf{k}) , \qquad (4.2)$$

provide an irreducible representaion of $\mathcal{P}_+^\uparrow$.

To have the amplitudes $f^{(\pm)}(\mathbf{k})$ transform covariantly, we have to extend $\xi(\mathbf{k})$ and $\eta(\mathbf{k})$ to 4-vectors. The invariance of $f^{(\pm)}(\mathbf{k})$, under the gauge transformation

$$f_\nu(\mathbf{k}) \to f_\nu(\mathbf{k}) + k_\nu\tilde{\phi}(\mathbf{k}) \qquad (4.3)$$

can be retained for arbitrary choices of the scalar function $\tilde{\phi}(\mathbf{k})$ if and only if we set $\xi_0(\mathbf{k}) = \eta_0(\mathbf{k}) = 0$. If we agree that under the proper Lorentz transformation Λ the 4-vector functions $\xi(\mathbf{k})$ and $\eta(\mathbf{k})$ should transform into

$$\xi'_\mu = \Lambda_\mu{}^\nu\xi_\nu(\Lambda^{-1}\mathbf{k}), \quad \eta'_\mu(\mathbf{k}) = \Lambda_\mu{}^\nu\eta_\nu(\Lambda^{-1}\mathbf{k}), \qquad (4.4)$$

then the quantities

$$f_\pm(\mathbf{k}) = 2^{-1/2}[f^\nu(\mathbf{k})\xi_\nu(\mathbf{k}) \pm if^\nu(\mathbf{k})\eta_\nu(\mathbf{k})] \qquad (4.5)$$

transform as scalars:

$$[\hat{U}(a,\Lambda)f]_\pm(\mathbf{k}) = \exp\left(-\frac{i}{\hbar}a\cdot k\right)f_\pm(\Lambda^{-1}\mathbf{k}). \qquad (4.6)$$

Naturally, if Λ is not a Euclidean rotation then $\xi'_0 \neq 0$ and $\eta'_0 \neq 0$, and $f_\pm(\mathbf{k})$ will not generally coincide any longer with the circular polarization probability amplitudes in the frame of reference to which Λ takes us. Nevertheless, the probability density $\rho(\mathbf{k})$ for detecting a photon of 3-momentum $\mathbf{k}$ still equals the sum of $|f_+|^2$ and $|f_-|^2$,

$$\rho(\mathbf{k}) = f_\nu^*(\mathbf{k})f^\nu(\mathbf{k}) = |f_+(\mathbf{k})|^2 + |f_-(\mathbf{k})|^2, \qquad (4.7)$$

and the inner product in $\mathcal{J}_0$ can be still expressed in a form analogous to (3.18):

$$(f|g) = \int f_\nu^*(\mathbf{k})g^\nu(\mathbf{k})k_0^{-1}\,d\mathbf{k} = \int (f_+^*g_+ + f_-^*g_-)k_0^{-1}\,d\mathbf{k}. \qquad (4.8)$$

Hence we shall retain the space $\mathcal{J}(\Gamma)$ in (3.14) in which each photon momentum-space wavefunction $f(\mathbf{k})$ has the representatives

$$f_e(q;\mathbf{p}) = (U_e(q^0)f)(\mathbf{k}) = f_e^{(+)} \oplus f_e^{(-)} \qquad (4.9)$$

at each $q^0 \in \mathbb{R}^1$ in the original laboratory frame. The Lorentz covariance of the theory will be established by proving that both components $f_{e\pm}(q,\mathbf{p})$ of $f_e(q;\mathbf{p})$ transforms as scalar under $\mathcal{P}_+^\uparrow$ as a result of the application of $\hat{U}(a,\Lambda)$ to $f_\pm(\mathbf{k})$ in accordance with (4.6).

Theorem 4.1: The unitary image

$$U_e(a,\Lambda) = U_e(0)\hat{U}(a,\Lambda)U_e^{-1}(0) \qquad (4.10)$$

of the representation $\hat{U}(a,\Lambda)$ of $\mathcal{P}_+^\uparrow$ acting on the closed subspace $\mathcal{M}_e$ of $\mathcal{J}(\Gamma)$ constitutes an irreducible unitary representation in $\mathcal{M}_e$ under which any of the functions $f_e(q;\mathbf{p})$ in (4.9) transform as scalar quantities:

$$(U_e(a,\Lambda)f_e)(q;\mathbf{p}) = f(\Lambda^{-1}(q-a);\Lambda^{-1}\mathbf{p}). \qquad (4.11)$$

Proof: Let Λ^{-1} be any proper Lorentz transformation, that applied to the laboratory frame takes us to an inertial frame moving at speed μ with respect to that frame. Upon setting

$$q' = \Lambda^{-1}q, \quad p' = \Lambda^{-1}p, \quad \boldsymbol{\omega}' = \frac{\mathbf{p}'}{|\mathbf{p}'|}, \qquad (4.12)$$

the lightlike 4-vectors p' and p become related by (2.4) and (2.5), and we note that

$$\mathbf{p}' = p_0\Lambda^{-1}\left(\frac{\mathbf{p}}{p^0}\right) = p_0\Lambda^{-1}\boldsymbol{\omega}. \qquad (4.13)$$

Consequently, we deduce from (2.4) that

$$\boldsymbol{\omega}' = \frac{p_0}{p'_0}\Lambda^{-1}\boldsymbol{\omega}, \quad \frac{p'_0}{p_0} = \gamma\left(1 - \frac{\mathbf{u}\cdot\boldsymbol{\omega}}{c}\right). \qquad (4.14)$$

Consider now a function $f(\mathbf{k}) \in \mathcal{J}_0$ that is continuous in $\mathbf{k} \in \mathbb{R}^3$ and integrable with respect to $k_0^{-1}dk_0$ for every $\boldsymbol{\omega}$ on the unit sphere in $\mathbb{R}^3$. Then so is $f(\Lambda^{-1}\mathbf{k})$, and therefore by (3.9)

$$f'_\pm(q;\mathbf{p}) = [U_e(0,\Lambda)f_e](q;\mathbf{p}) \qquad (4.15)$$

$$= \left(\frac{2}{h}\right)^{1/2}N_e\int_0^\infty \exp\left(\frac{i}{\hbar}q\cdot k\right)e^*\left(\frac{k_0}{p_0}\right)f_\pm[\Lambda^{-1}(k^0\boldsymbol{\omega})]\frac{dk^0}{k^0}. \qquad (4.16)$$

Setting $k' = \Lambda^{-1}k$ and noting that

$$\Lambda^{-1}(k_0\boldsymbol{\omega}) = k_0\Lambda^{-1}\boldsymbol{\omega} = k'_0\boldsymbol{\omega}' \qquad (4.17)$$

we can rewrite (4.16) in the form

$$f'_\pm(q;\mathbf{p}) = \left(\frac{2}{h}\right)^{1/2}N_e\int_0^\infty \exp\left(\frac{i}{\hbar}q'\cdot k'\right)e^*\left(\frac{k_0}{p_0}\right)f_\pm(k'_0\boldsymbol{\omega})\frac{dk_0}{k^0}. \qquad (4.18)$$

On the other hand, we obtain from (2.9) and (4.14) that

$$k'_0 = \gamma\left(1 - \frac{\mathbf{u}\cdot\boldsymbol{\omega}}{c}\right)k_0 = \frac{p'_0}{p_0}k_0. \qquad (4.19)$$

This leads to the equalities

$$\frac{k_0}{p_0} = \frac{k'_0}{p'_0}, \quad \frac{dk_0}{k_0} = \frac{dk'_0}{k'_0}, \qquad (4.20)$$

which, substituted in (4.18), yield

$$f'_\pm(q;\mathbf{p}) = f_\pm(q';\mathbf{p}'). \qquad (4.21)$$

Thus (4.12) is established when $a=0$ for the class of functions $f(\mathbf{k})$ satisfying the conditions stipulated at the beginning of the proof. Since this class is dense in $\mathcal{J}_0$, and $\hat{U}(0,\Lambda)$ as well as $U_e(0)$ are unitary, the result immediately extends to all of $\mathcal{M}_e$.

The proof for $U_e(a,I)$, $a \in \mathbb{R}^4$, is trivial, and since

$$\hat{U}(a,\Lambda) = \hat{U}(a,I)\hat{U}(0,\Lambda), \qquad (4.22)$$

(4.11) immediately follows.

The unitary and irreducibility of $U_e(a,\Lambda)$ in $\mathcal{M}_e$ is an immediate consequence of the corresponding properties of $\hat{U}(a,\Lambda)$ in $\mathcal{J}_0$, and of the unitarity of $U_e(0)$.

Q.E.D.

The above theorem establishes that the functions $f_\pm(q;\mathbf{p}) \in \mathcal{M}_e$ transform covariantly although the measure σ is not left invariant by Lorentz transformations. But, this is understandable, since the choice of σ was one of expediency, meant to secure the square-integrability of functions that do not depend on $\mathbf{q}_\perp$ at any fixed choice of $\mathbf{p}$. Any other choice would have had, however, to run into the same problem since it would have had to involve a finite measure in $\mathbf{q}_\perp \in \mathbb{R}^2$. Thus a totally covariant formulation has to fall back to the renormalization procedure inherent in (3.2) and (3.7).

The key corollary of Theorem 4.1 is that the (gauge-invariant) probability density

$$\rho_e(q;\mathbf{p}) = |f_e^{(+)}(q;\mathbf{p})|^2 + |f_e^{(-)}(q;\mathbf{p})|^2 \tag{4.23}$$

of detecting a photon (of arbitrary polarization) at the stochastic phase space point $(\mathbf{q},\mathbf{p};\chi_{q;p})$ transforms as a scalar under $(a,\Lambda)\in \mathcal{P}_+^\uparrow$:

$$\rho_e'(q;\mathbf{p}) = \sum_{\alpha=\pm} |U(a,\Lambda) f_e^{(\alpha)}(q;\mathbf{p})|^2 = \rho_e(\Lambda^{-1}(q-a);\Lambda^{-1}\mathbf{p}). \tag{4.24}$$

Consequently, $\rho(q;\mathbf{p})$ gives rise to a covariant probability current.

Theorem 4.2: The probability current

$$j_e^\nu(q) = \int_{\mathbb{R}^3} p^\nu \rho_e(q;\mathbf{p}) \frac{d\mathbf{p}}{p^0} \tag{4.25}$$

transforms as a vector under $(a,\Lambda)\in \mathcal{P}_+^\uparrow$:

$$j_e^\mu(q) \mapsto j_e'^\mu(q) = \Lambda^\mu{}_\nu j_e^\nu(\Lambda^{-1}(q-a)). \tag{4.26}$$

If $j_e^\nu(q)$ corresponds to a photon state whose momentum space wavefunction $\mathbf{f}(\mathbf{k})$ is continuous and integrable in $\mathbb{R}^3$ with respect to the measure $k_0^{-1}d\mathbf{k}$, then $j_e^\nu(q)$ is differentiable and

$$\frac{\partial}{\partial q^\nu} j_e^\nu(q) = 0. \tag{4.27}$$

Proof: The covariance feature (4.26) is an immediate consequence of (4.12) and of the invariance of $p_0^{-1}d\mathbf{p}$ under each $\Lambda\in L_+^\uparrow$

To prove (4.27), note that by (4.23) and (4.25)

$$\frac{\partial}{\partial q^\nu} j_e^\nu(q) = 2\,\mathrm{Re}\sum_{\alpha=\pm}\int f_e^{(\alpha)*}(q;\mathbf{p}) \times\left(\frac{\partial}{\partial q^0} + \frac{\mathbf{p}}{p^0}\cdot\nabla_{\mathbf{q}}\right) f_e^{(\alpha)}(q;\mathbf{p})\, d\mathbf{p}. \tag{4.28}$$

However, under the conditions imposed on $\mathbf{f}(\mathbf{k})$ in the theorem (3.9) is applicable, and consequently,

$$\frac{\partial}{\partial q^0} f_e^{(\pm)}(q;\mathbf{p}) = -\frac{\partial}{\partial q_\parallel} f_e^{(\pm)}(q;\mathbf{p}) = -\omega\cdot\nabla_{\mathbf{q}} f_e^{(\pm)}(q;\mathbf{p}). \tag{4.29}$$

Hence, the integrand in (4.28) equals zero, and (4.27) follows Q.E.D.

Although the formal appearance of the formulae I-(4.14) and (4.25) for the current in the massive and the zero-mass case is exactly the same, there are essential physical differences between these two cases. The current I(4.14) has a time component which is a probability density at a bona fide stochastic configuration point when the particle is in a state of slow (nonrelativistic) motion in relation to the laboratory frame. This is not the case, however, with the probability current (4.25) for zero-mass particles, since no such states exist for those particles. Consequently, the configuration and momentum dependence of the confidence function of a stochastic phase space point are now always inseparably intertwined. Thus, for zero-mass particles the value of $j_e^0(q)$ at any space–time point q has to be viewed as a superposition of probabilities at all stochastic phase space points centered at points $(q;\mathbf{p})$ that share the same space–time part q, rather than as a probability at a single stochastic configuration space value.

5. DISCUSSION

It might appear strange that the stochastic phase space approach permits the existence of probability currents for the photon, whereas neither of the two conventional approaches— namely configuration space and momentum space representations—allow for similar possibilities [cf. Ref. 5, Sec. 4, for a discussion of the reasons behind the impossibility of constructing currents from vector potential $A_\nu(x)$]. This phenomenon can be taken, however, to be a natural mathematical expression of a fundamental physical characteristic of zero-mass particles: their localizability in space–time is inseparably tied in with their direction of motion,[2,3] and consequently any formalism that treats position separately from momentum cannot possibly supply a consistent expression for a covariant current whose timelike component would be a probability density.

In the case of massive particles some insights were achieved by considering the probability as well as the charge current at spread-out stochastic configuration points originating with optimal representations, and then studying their behavior as we went to the limit of sharp configuration points. In a sense, we shall adopt the same tactic in the present context, but first we have to dispense with a technical difficulty: Since there is no nonrelativistic theory of the photon, there is no physical basis for a definition of optimality for stochastic phase-space points at which a photon might be located. Hence we shall have to let ourselves be quided only by formal analogies with the relativistic case of massive particles.

We shall transfer the exponential behavior characterizing the generators of optimal stochastic phase space representations for massive particles to the photon case by considering in (3.5) generators of the form

$$e^{(s)}(k_\parallel) = (\pi\hbar^{-2})^{-1/4}\exp\left[-\frac{s^2}{2\hbar}(1-k_\parallel)^2\right]\zeta^{(s)}(k_\parallel), \tag{5.1}$$

where $\zeta^{(s)}(k_\parallel)$ is a smooth function which vanishes for $k_\parallel<\epsilon(s)$, where $\epsilon(s)<1$ is some positive number so that (3.6) is satisfied. To be more specific, we shall choose $\epsilon(s)>0$ so small that $2\epsilon(s)<1$, and then set $\zeta^{(s)}(k_\parallel) = c(s)$ for $k_\parallel>2\epsilon(s)$, where the magnitude of the positive constant $c(s)$ is chosen so that the normalization condition

$$\int |e^{(s)}(\lambda)|^2 = 1 \tag{5.2}$$

implicit in (2.16) is satisfied.

In letting $s\to+\infty$ we shall take $\epsilon(s)$ independent of s, so that on account of the well-known δ-behavior of the exponential part of $e^{(s)}(\lambda)$ we have $c(s)\to 1$ and $N_{e(s)}\to 2^{-1/2}$. Consequently, for functions $f^{(\pm)}(k_0\omega)$ that are continuous and integrable in $k_0^{-1}dk_0$ we get, by applying (3.9),

$$\lim_{s\to+\infty}(2\hbar^3 s^{-2})^{-1/4} f_{e(s)}^{(\pm)}(q;\mathbf{p}) = \exp\left(-\frac{i}{\hbar}q\cdot p\right) f^{(\pm)}(\mathbf{p}). \tag{5.3}$$

The above relation represents the zero-mass counterpart of I(4.7). It shows that the probability densities

(4.23) at stochastic phase space points corresponding in accordance with (2.1) and (2.15) to the generators (5.1), after being appropriately renormalized, indeed approach the conventional density (4.7) at sharp momentum values, i.e.,

$$\lim_{s\to\infty}(2h^3 s^{-2})^{-1/2}\rho_{e(s)}(q;\mathbf{p}) = |\mathbf{f}(\mathbf{k})|^2 \quad (5.4)$$

in the gauge (2.11).

On the other hand, in letting $s \to +0$, we shall also let $\epsilon(s) \to +0$, so that

$$f_{(0)}^{(\pm)}(q;\mathbf{p}) = \lim_{s\to 0}[2^{1/2}c(s)N_{e(s)}]^{-1} f_{e(s)}^{(\pm)}(q;\mathbf{p}) \quad (5.5)$$

does exist for the same choice of $f^{(\pm)}(\mathbf{k})$ as above. In fact, we easily compute from (3.9) that

$$f_{(0)}^{(\pm)}(q;\mathbf{p}) = h^{-1/2}\int_0^\infty \exp\left(-\frac{i}{\hbar}q\cdot k\right) f^{(\pm)}\left(\frac{k^0}{p^0}\mathbf{p}\right) k_0^{-1}\,dk_0. \quad (5.6)$$

Thus, we see that there is no counterpart to I(4.6), since

$$\mathbf{A}(\mathbf{q}) = h^{-3/2}\int \exp\left(-\frac{i}{\hbar}q\cdot k\right)[f^{(+)}(\mathbf{k})\xi(\mathbf{k}) + f^{(-)}(\mathbf{k})\eta(\mathbf{k})]\frac{d\mathbf{k}}{k^0} \quad (5.7)$$

has no components proportional to (5.6). As a consequence, the current $j_{e(s)}^{\nu}(q)$ is not algebraically related to $\mathbf{A}(\mathbf{q})$ in the limit $s \to +0$. As a matter of fact, the limit itself,

$$\begin{aligned} j_0^{(\nu)}(q) &= \lim_{s\to+\infty}[2c^2(s)N_{e(s)}^2]^{-1} j_{e(s)}^{\nu}(q) \\ &= \int \frac{p_\nu}{p_0}[|f_{(0)}^{(+)}(q;\mathbf{p})|^2 + |f_{(0)}^{(-)}(q;\mathbf{p})|^2]d\mathbf{p}, \end{aligned} \quad (5.8)$$

although formally covariant due to the covariance of (5.6), is actually represented by a divergent integral since $f_{(0)}^{(\pm)}(q;\mathbf{p})$ in (5.6) is easily seen to be independent of p_0.

Thus, we have to conclude that although the stochastic phase-space theory of zero-mass particles shares some common features with its nonzero-mass counterpart, there are also essential differences, and procedures that work if second case fail in the first, and vice versa. The concept of stochastic phase space itself has to be modified by the introduction of frontally localized points (which on the other hand cannot be related to covariant densities in the massive case), and emerges as being even more crucial to a covariant treatment of localizability of zero-mass particles than it was in the case with massive particles.

[1]E. Prugovečki, J. Math. Phys. 19, 2260 (1978), referred to as I.
[2]R. Acharya, and R. Sudarshan, J. Math. Phys. 1, 532 (1960).
[3]E. Prugovečki, J. Phys. A 9, 1851 (1976).
[4]J.D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), Sec. 11.3.
[5]V.B. Berestetskiĭ, E.M. Lifschitz, and L.P. Pitaevskiĭ, *Relativistic Quantum Theory*, translated by J.B. Sykes and J.S. Bell (Pergamon, Oxford, 1971).

Representations of the Weyl Lie algebra as models of elementary particles

Håkan Snellman

Department of Theoretical Physics, Royal Institute of Technology, s—100 44 Stockholm, Sweden
(Received 16 May 1978)

Some irreducible representations of the 11-parameter Weyl Lie algebra are suggested as models of elementary particles.

1. INTRODUCTION

Hadron spectroscopy data indicate that hadrons have internal structure. In any theory of particles with structure it seems inevitable to introduce an internal manifold to describe this structure of the particle. For the irreducible representations (irreps) of the Poincaré group $\mathcal{P}_+^\uparrow = T^4 \otimes SO_0(3,1)$ the absence of such a manifold corresponds to the particle being essentially a point particle. The spin degree of freedom is attached to the particle in an abstract way. However, Wigner has shown[1] how one can associate with any $m>0$, positive-energy irrep of $\mathcal{P}_+^\uparrow$ a system of differential equations for the Casimir operators on a manifold M^5 derived from a manifold $M^8 = \{\xi^\mu, p^\mu\}$ restricted by the subsidiary conditions $\{p^2 = m^2,\ \xi^2 = -1,\ p\xi = 0\}$. In the rest system of the particle this means that the coordinates $\bar{\xi}$ can be thought of as describing some spatial extension of the particle in the form of a sphere of unit radius. For integer spin reps the spin operators are then the ordinary rotation generators on the sphere.[2]

It is tempting to think of the sphere as representing somehow the internal structure of the particle. In this work we shall take this idea seriously and use an internal closed space to represent the coordinates of particle matter, the motion of the center of energy of this matter being described by the 4-momentum p^μ. That this is a natural description of composite objects is heuristically digressed upon in Sec. 5. Furthermore, it turns out that there is a consistent division of the relativistic motion of such a composite object in that the internal motion can be described essentially nonrelativistically while the center of energy motion is still relativistic. This is technically brought out in the three models of particles described below. These models describe families of particles as irreps of the 11-parameter Weyl Lie algebra $\mathcal{W} = (t^4 \oplus SO(3,1)) \oplus \mathbb{R}_+$. Each one of these representations is integrable on the Poincaré subalgebra, so that global Poincaré transformations can be applied to the particles. In Sec. 2 we describe bosons and in Sec. 3 fermions. Section 4 describes a model composed of two spin-$\frac{1}{2}$ constituents (e.g., "charmonium") and the resulting energy levels.

2. IRREDUCIBLE REPRESENTATIONS OF $\mathcal{W}$ WITH INTEGER SPIN

The Weyl Lie algebra $\mathcal{W}$ is spanned by the vectors P^μ, $M^{\mu\nu}$, and D, corresponding to translations, Lorentz transformations and dilatations. Besides the ordinary commutation relations between P^μ and $M^{\mu\nu}$, these vectors satisfy the relations

$$[D, P^\mu] = -P^\mu, \tag{2.1}$$

$$[D, M^{\mu\nu}] = 0. \tag{2.2}$$

Let $\mathcal{H} = L^2(\mathbb{R}^3, d^3p/(\bar{p}^2 + m_0^2)^{1/2}) \otimes L^2(S, d^3x)$. The manifold S is the ball in $\mathbb{R}^3$ with radius a. In $\mathcal{H}$ we can define a rep $T(\mathcal{W})$ of $\mathcal{W}$ as follows. Let $\psi(\bar{p}, \bar{x}) \in S_0$, the subspace of $\mathcal{H}$ obtained by restricting the ψ:s to belong to $\mathcal{S}(\mathbb{R}^3)$ with respect to $\bar{p}$ and to C^∞ and vanish together with all derivatives at 0 and on ∂S with respect to $\bar{x}$. Then define the following skew-symmetric operators on S_0:

$$T(P^\mu)\psi(\bar{p}, \bar{x}) = ip^\mu \Delta_x \psi(\bar{p}, \bar{x}), \tag{2.3a}$$

$$T(\bar{J})\psi(\bar{p}, \bar{x}) = \left(\bar{p} \times \frac{\partial}{\partial \bar{p}} - \bar{x} \times \frac{\partial}{\partial \bar{x}}\right)\psi(\bar{p}, \bar{x}), \tag{2.3b}$$

$$T(\bar{K})\psi(\bar{p}, \bar{x}) = \left[p_0 \frac{\partial}{\partial \bar{p}} - \frac{1}{p_0 + m_0}\,\bar{p} \times \left(\bar{x} \times \frac{\partial}{\partial \bar{x}}\right)\right]\psi(\bar{p}, \bar{x}), \tag{2.3c}$$

$$T(D)\psi(\bar{p}, \bar{x}) = \left(\tfrac{1}{2}\bar{x} \cdot \frac{\partial}{\partial \bar{x}} + \tfrac{3}{4}\right)\psi(\bar{p}, \bar{x}). \tag{2.3d}$$

Here $p_0 = (\bar{p}^2 + m_0^2)^{1/2}$. Let $T_J(W^2)$ and $T_J(P^2)$ be the Poincaré Casimir invariants. Then $-T_J(W^2/P^2) = \bar{S}^2$, where $\bar{S} = -i\bar{x} \times \partial/\partial\bar{x}$, is invariant with respect to $T(\mathcal{W})$, and has eigenvalues $J(J+1)$. We can then decompose $\mathcal{H}$ and T into reps T_J on $\mathcal{H}_J$, with $\mathcal{H} = \oplus_J \mathcal{H}_J$. For any J, $S_0 \cap \mathcal{H}_J$ is a dense invariant domain in $\mathcal{H}_J$ for the rep $T_J(\mathcal{W})$ defined as above. A basis S_π^J of $\mathcal{H}_J$ on which $M^2 = -P_\mu P^\mu$ is diagonal is given by

$$S_\pi^J = \{\psi_{\bar{n}, k_J, J}(\bar{p}, r, \theta, \varphi) = \Phi_{\bar{n}}(\bar{p}) J_{J+1/2}(k_J r)/\sqrt{r}\; Y^J_{m_J}(\theta, \varphi);\ -J < m_J < J,\ k_J \in F_J,\ \bar{n} \in Z^3\}.$$

Here $F_J = \{y; J_{J+1/2}(ya) = 0\}$ with $J_{J+1/2}$ being a Bessel function of the first kind corresponding to angular momentum J, and $\Phi_{\bar{n}}(\bar{p})$ are Hermite functions on $\mathbb{R}^3$. The radial functions in this basis derive from the Sturm-Liouville expansion of $L^2([0, a], r dr)$ with respect to solutions of

$$\left(\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} - \frac{(J + \frac{1}{2})^2}{r^2}\right) f_{k_J}(r) = -k_J^2 f_{k_J}(r) \tag{2.4}$$

with boundary conditions $f_{k_J}(a) = f_{k_J}(0) = 0$. The eigenvalues of M^2 on S_π^J are $M^2 = m_0^2 k_J^4$, $k_J \in F_J$. As is shown in the Appendix each rep $T_J(\mathcal{W})$ is Schur-irreducible in $\mathcal{H}_J$ in the sense that every bounded operator A that commutes strongly with $T_J(P^\mu)$, $T_J(M^{\mu\nu})$ and weakly with $T_J(D)$ is a multiple of the identity operator.

Each irrep $T_J(\mathcal{W})$ is integrable on the Poincaré subalgebra by construction. However, the operator $iT(D)$ has defect indices $(0,1)$ and is not s.a. in $\mathcal{H}_J$ and can

 0022-2488/78/1911-2278$1.00

therefore not be integrated. The physical interpretation of these irreps of $\mathcal{W}$ can be in terms of a ball, the radial excitation modes of which correspond to the excited states of different masses. The spectrum of masses in $\mathcal{H}$ is shown in Fig. 1 for different values of the spin J of the particle.

When space reflexion is defined by the operation $\bar{r} \to -\bar{r}$, there is a natural grouping of states with respect to odd and even values of J. The fixed parity, fixed radial excitation modes in $\mathcal{H}$ will then respect the $\Delta J = 2$ rule.

Finally we mention that the irreps of ω presented in Ref. 3 are closely related to these ones. If the spin space there is chosen as the $(2S+1)$-dimensional carrier space of a D^s rep of SO(3), then we obtain a system of particles with masses $m = m_0 k$, $k = 1, 2, \cdots$ and spin S. However, the interpretation of the model is then different, since the internal coordinates have no direct relation to each other.

3. IRREPS WITH HALF-INTEGER SPIN

Half-integer spin irreps can be obtained by replacing $\bar{L} = -i\bar{x}\times\partial/\partial\bar{x}$ by $\bar{S} = -i\bar{x}\times\partial/\partial\bar{x} + \frac{1}{2}\bar{\sigma}$ in the generators $T(\bar{J})$ and $T(\bar{K})$ in Sec. 2. Here $\bar{\sigma}$ are the Pauli matrices. Let the rep space be $\mathcal{H} = L^2(R^3, d^3p/p_0) \otimes C_2$. Reducing this rep on the Poincaré algebra, there is now a degeneracy in the mass spectrum with respect to spin. This degeneracy can be removed by introducing an extra term in the expression for $T(P^\mu)$, giving rise to the following rep of $\mathcal{W}$ in $\mathcal{H}$:

$$T(P^\mu)\psi_{m_1}(\bar{p},\bar{x}) = -ip^\mu(-\Delta\cdot\mathbb{1} + 2\alpha r^{-2}\bar{L}\circ\bar{\sigma})\psi_{m_1}(\bar{p},\bar{x}), \quad (3.1)$$

$$T(\bar{J})\psi_{m_1}(\bar{p},\bar{x}) = \left(\bar{p}\times\frac{\partial}{\partial\bar{p}} - i\bar{S}\right)\psi_{m_1}(\bar{p},\bar{x}), \quad (3.2)$$

$$T(\bar{K})\psi_{m_1}(\bar{p},\bar{x}) = \left(p^0\frac{\partial}{\partial\bar{p}} - \frac{1}{p_0+m_0}\right)(\bar{p}\times iS)\psi_{m_1}(\bar{p},\bar{x}), \quad (3.3)$$

$$T(D)\psi_{m_1}(\bar{p},\bar{x}) = \tfrac{1}{4}\left(\bar{x}\frac{\partial}{\partial\bar{x}} + \frac{\partial}{\partial\bar{x}}\bar{x}\right)\psi_{m_1}(\bar{p},\bar{x}) \quad (3.4)$$

for functions in $\mathcal{H}$ which have the same restrictions with respect to $\bar{x}$ and $\bar{p}$ as S_0 in Sec. 2.

The invariants are now $\bar{\sigma}^2$, $\bar{L}^2$, and $(\bar{L}+\frac{1}{2}\bar{\sigma})^2$ with eigenvalues $1, l(l+1)$, and $J(J+1)$ respectively in $\mathcal{H}$. The rep $T(\mathcal{W})$ is reduced by the decomposition $\mathcal{H} = \oplus_{l,J}\mathcal{H}_{l,J}$ with $J = 1 \pm \frac{1}{2}$. For the further decomposition

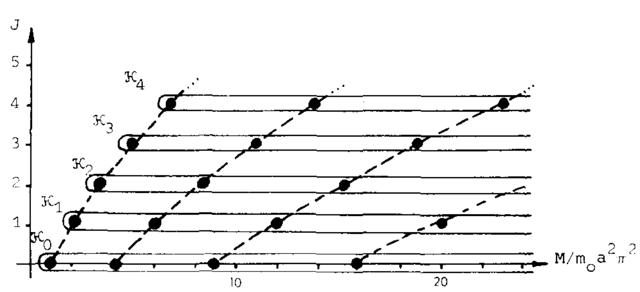

FIG. 1. Irreducible representations of $\mathcal{W}$ with integer spin.

of each $\mathcal{H}_{l,J}$ into irreducible Poincaré reps we get the following eigenvalue equations for the radial functions:

$$\left(\frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} - r^{-2}(l^2 + l + \alpha l)\right)\psi_{l+1/2}(r)$$
$$= -k^2_{J,l}(\alpha)\Psi_{l+1/2}(r), \quad (3.5)$$

$$\left(\frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} - r^{-2}[l^2 + l - \alpha(l+1)]\right)\psi_{l-1/2}(r)$$
$$= -k^2_{J,l}(\alpha)\Psi_{l-1/2}(r). \quad (3.6)$$

For the eigenvalues of the mass operator in $\mathcal{H}_{l,J}$ with respect to the same boundary conditions as in Sec. 1, we then have

$$M_{l,J} = m_0 a^2 j^2_{\nu_\pm,n}$$

where

$$\nu_+ = [(l+\tfrac{1}{2})^2 + \alpha l]^{1/2} \qquad \text{for } J = l + \tfrac{1}{2},$$
$$\nu_- = [(l+\tfrac{1}{2})^2 - \alpha(l+1)]^{1/2} \quad \text{for } J = l - \tfrac{1}{2}, \quad l > 1.$$

and $j_{\nu,n}$ is the nth zero of the Bessel function $J_{\nu+1/2}$. The splitting of the levels for $\alpha > 0$ is shown in Fig. 2.

4. A MODEL WITH CONSTITUENTS

As a simple illustration of possible generalizations of the models above, we shall consider a model with "constituents" confined to the ball S (e.g., "charmonium"). We can think of $\bar{x}$ above as the coordinate for the distance between the "constituents," each of which we take to have spin $\frac{1}{2}$. We can then construct a rep $T(\mathcal{W})$ of $\mathcal{W}$ on the space $\mathcal{H} = L^2(\mathbb{R}^3, d^3p/p^0)\otimes L^2(S, d^3x)\otimes C_2 \otimes C_2$. In $\mathcal{H}$ the generators of $T(\mathcal{W})$ are defined by

$$T(P^\mu)\Psi_{m_1m_2}(\bar{p},\bar{x})$$
$$= -ip^\mu[-\Delta_x + \bar{x}^{-2}(\alpha + 2\beta\bar{L}\cdot\bar{S} + 4\gamma\bar{s}_1\circ\bar{s}_2)]\Psi_{m_1m_2}(\bar{p},\bar{x}), \quad (4.1)$$

$$T(\bar{J})\Psi_{m_1m_2}(\bar{p},\bar{x})$$
$$= \left(\bar{p}\times\frac{\partial}{\partial\bar{p}} - i\bar{L} - i\bar{S}\right)\Psi_{m_1m_2}(\bar{p},\bar{x}), \quad (4.2)$$

$$T(\bar{K})\Psi_{m_1m_2}(\bar{p},\bar{x})$$
$$= \left[p^0\frac{\partial}{\partial\bar{p}} - \frac{1}{p^0+m_0}\bar{p}\times(i\bar{L}+i\bar{S})\right]\Psi_{m_1m_2}(\bar{p},\bar{x}), \quad (4.3)$$

$$T(D)\Psi_{m_1m_2}(\bar{p},\bar{x})$$
$$= \tfrac{1}{4}\left(\bar{x}\frac{\partial}{\partial\bar{x}} + \frac{\partial}{\partial\bar{x}}\bar{x}\right)\Psi_{m_1m_2}(\bar{p},\bar{x}). \quad (4.4)$$

Here $\bar{S} = \bar{s}_1 + \bar{s}_2$ and $\bar{L} = -i\bar{x}\times\partial/\partial\bar{x}$. When both β and γ are different from zero, there is no degeneracy with respect to the decomposition of $\mathcal{H}$ into Poincaré algebra reps. To decompose $\mathcal{H}$ into irreps of $\mathcal{W}$, defined by the eigenvalues of $(\bar{L}+\bar{S})^2 = J(J+1)$, $\bar{L}^2 = l(l+1)$, and $\bar{S}^2 = S(S+1)$, we decompose $L^2(S, d^3x)$ as follows:

$$L^2(S, d^3x) = \sum_{l=0}^{\infty} L^2_l([0,a], r^2dr)\otimes D^l(\Omega, d\Omega), \quad (4.5)$$

where $\Omega = \partial S$.

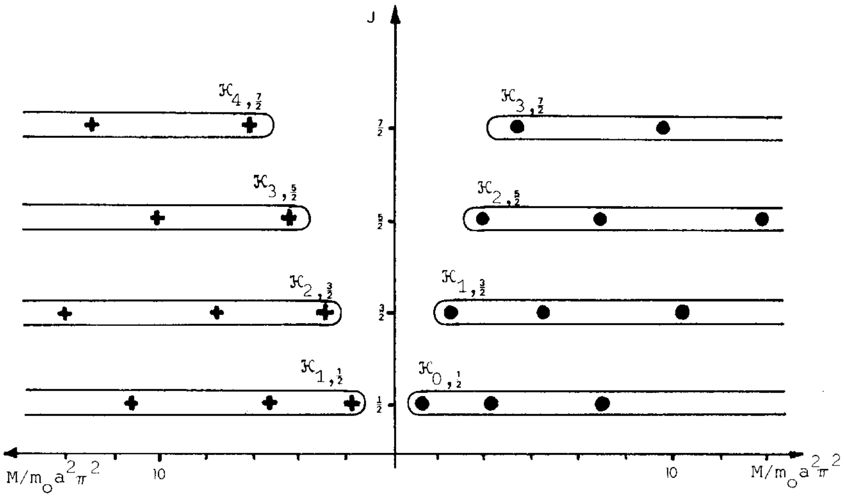

FIG. 2. Irreducible representations of W with half-integer spin.

Then

$$H = \underset{J,l,S}{\oplus} H_{J,l,S},$$

where

$$H_{J,l,S} = L^2(R^3, d^3p/p^0) \otimes D^{J,l,S}(\Omega, d\Omega) \otimes L^2_l([0,a], r^2dr).$$

$H_{J,l,S}$ carries an irrep $T_{J,l,S}(W)$ of W. This rep is again Schur irreducible in the same sense as in Sec. 2.

The mass spectrum, again with respect to the same boundary conditions as in Sec. 1, is given by

$$M = m_0 a^2 j^2_{\nu,n}, \tag{4.6}$$

where $j_{\nu,n}$ is the nth zero of the Bessel function $J_{\nu+1/2}$ with index $\nu = [(1+\frac{1}{2})^2 - f]^{1/2}$, where f is the eigenvalue of the operator

$$F = \alpha + \beta(\bar{J}^2 - \bar{L}^2 - \bar{S}^2) + 2\gamma(\bar{S}^2 - 3/2).$$

Here $\bar{J} = \bar{L} + \bar{S}$ and $\bar{L} = -i\bar{x} \times \partial/\partial x$.

For β and γ small, we have

$$M \approx m_0 a^2 \Big(j^2_{\nu_0,n} - \frac{1}{\nu_0} j_{\nu_0,n} \frac{dj_{\nu_0,n}}{d\nu} \{\beta[J(J+1) - l(l+1) - S(S+1)] + 2\gamma(S(S+1) - \tfrac{3}{2})\}\Big)$$

with $\nu_0 = [(1+\frac{1}{2})^2 - \alpha]^{1/2}$.

The mass spectrum is shown in Fig. 3. It corresponds by construction to the levels obtained by current nonrelativistic models.[4]

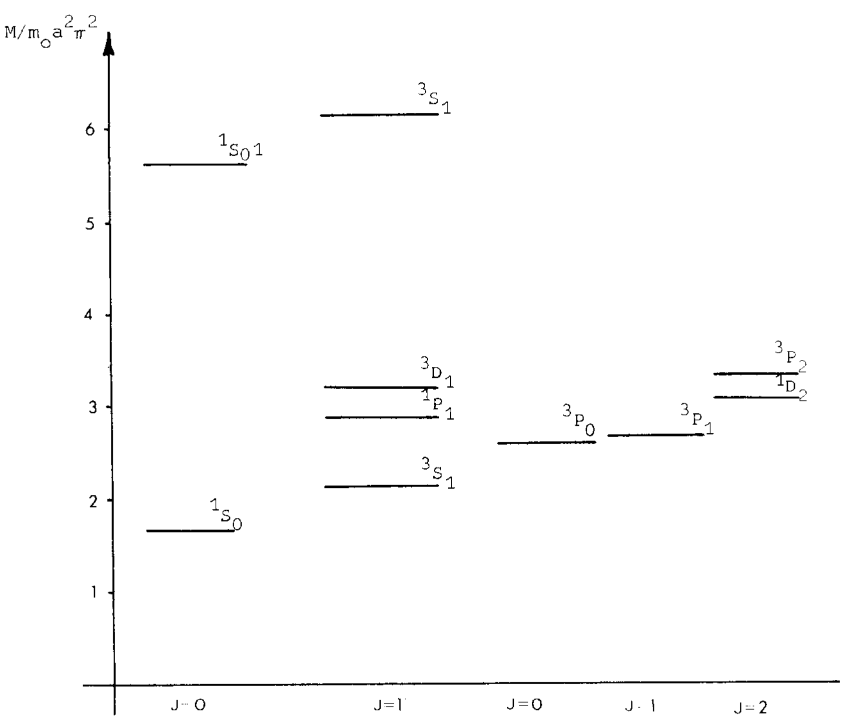

FIG. 3. Mass spectrum for the composite model.

Again, the relative inner parity π can be defined by the properties under the reflexion $\bar{x} \to -\bar{x}$. The spin-statistics postulate forbids singlet states with odd parity and triplet states with even parity when the constituent particles are identical. This is of course not the case for the particle—antiparticle pair.

5. DISCUSSION

The models of hadronlike systems considered in this work can be seen as variations of one and the same theme: realizations of Weyl Lie algebra reps that are Poincaré integrable. To some extent these models represent a new type of closed systems in relativistic particle physics that seem to have a rich potential structure. Earlier suggestions along the same lines are found in Ref. 3 and 5. The so-called dynamical groups are a somewhat different approach to the same idea and suffer from various difficulties that are not present for Lie algebra reps.

The models describe the particle states with different mass but the same spin as radial excitations of one and the same object within one irrep of $\mathcal{W}$. The particles have an inner space S in which the matter is confined. In order to illustrate the possible physical interpretation of this, we consider two free particles with mass m described by the energy operator

$$P^0\psi(\bar{r}_1,\bar{r}_2)=[(m^2-\Delta_1)^{1/2}+(m^2-\Delta_2)^{1/2}]\psi(\bar{r}_1,\bar{r}_2). \tag{5.1}$$

In the nonrelativistic limit we can write

$$P^0\psi(\bar{R},\bar{r})\approx\left(2m-\frac{1}{4m}\Delta_R-\frac{1}{m}\Delta_r\right)\psi(\bar{R},\bar{r}), \tag{5.2}$$

where $\bar{R}=\frac{1}{2}(\bar{r}_1+\bar{r}_2)$ and $\bar{r}=\bar{r}_1-\bar{r}_2$.

Introducing a strong central potential between the particles, it is reasonable to change (5.2) into

$$P^0\psi(\bar{R},\bar{r})=\left(2m-\frac{1}{2M}\Delta_R-\frac{1}{m}\Delta_r+V(r)\right)\psi(\bar{R},\bar{r}), \tag{5.3}$$

where M is the mass operator of the bound system, i.e., for $\psi(\bar{R},\bar{r})=\Phi(\bar{R})\chi(\bar{r})$ we have

$$\left(2m-\frac{1}{m}\Delta_r+V(r)\right)\chi(\bar{r})=M\chi(\bar{r}). \tag{5.4}$$

Then (5.2) becomes

$$P^0\psi(\bar{R},\bar{r})=\left(M-\frac{1}{2M}\Delta_R\right)\psi(\bar{R},\bar{r}), \tag{5.5}$$

which is the nonrelativistic limit of

$$P^0\psi(\bar{R},\bar{r})=(M^2-\Delta_R)^{1/2}\psi(\bar{R},\bar{r}). \tag{5.6}$$

Consistent with (5.6) is

$$\bar{P}\psi(\bar{R},\bar{r})=-i\nabla_R\psi(\bar{R},\bar{r}). \tag{5.7}$$

Let

$$M\chi_i(\bar{r})\equiv mH\chi_i(\bar{r})=mh_i\chi_i(\bar{r}). \tag{5.8}$$

Then we can accordingly scale $\bar{P}$ on each irrep of the Poincaré group, so that Eqs. (5.6) and (5.7) after Fourier transformation read

$$P^0\psi_i(\bar{p},\bar{r})=h_i(m^2+\bar{p}^2)^{1/2}\psi_i(\bar{p},\bar{r}), \tag{5.6'}$$

$$\bar{P}\psi_i(\bar{p},\bar{r})=h_i\bar{p}\psi_i(\bar{p},\bar{r}). \tag{5.7'}$$

Identifying H with

$$H=-\frac{1}{m^2}\Delta_r+\frac{1}{m}V(r)+2$$

and defining $\bar{x}=m\bar{r}$, we have

$$H=-\Delta_x+\frac{1}{m}V\left(\frac{|\bar{x}|}{m}\right)+2.$$

If we take the confining potential to be

$$V(r)=\begin{cases}-2m & \text{for } r<a/m,\\ +\infty & \text{for } r\geq a/m,\end{cases}$$

then

$$\begin{cases}H\chi_i(\bar{x})=-\Delta_x\chi(\bar{x}) & \text{for } x\leq a,\\ H\chi_i(\bar{x})=0 & \text{for } x>a.\end{cases}$$

This is equivalent to $H=-\Delta_x$ in $L^2(S,d^3x)$ in accordance with the models used above.

Finally we remark that, from the point of view of representation theory for $\mathcal{W}$, obviously any two rotationally invariant symmetrical differential operators A and B in R^3, with $[A,B]=-B$ can realize a rep of $\mathcal{W}$ if we put $P^\mu=p^\mu B$ and $D=A$ and make the necessary changes in boundary conditions that make A self-adjoint. Our choice has been guided by the physical interpretation above in terms of a potential problem. The confining potential, represented here by a closed compact space S, might perhaps be an approximate description of the effects of some strong nonlinearity around the hadron matter that introduces a curved Riemann metric of closed type, or simply the infinite approximation to a finite but extremely deep potential, similar to the infinite square-well potential in nuclear physics.

ACKNOWLEDGMENTS

This work has been made with financial support from the Swedish Atomic Research Council (AFR), which is gratefully acknowledged. The author wishes to thank Professor M. Flato for several useful discussions and G. Tengstrand for suggesting the method of calculating the integral in the Appendix.

APPENDIX

Let p^2 and W^2 be the Poincaré Casimir invariants. By construction $T_J(W^2/P^2)$ has eigenvalue $-J(J+1)$ in $\mathcal{H}_J$ and the spin is invariant under $T_J(D)$. In this case J therefore characterizes the rep. Let A be an arbitrary bounded s.a. operator on $\mathcal{H}_J$. We decompose $\mathcal{H}_J$ with respect to Poincaré algebra irreps $\mathcal{H}_{J,m}$ which are integrable due to construction. A therefore commutes strongly with $T_J(P^\mu)$ and $T_J(M^{\mu\nu})$ on each $\mathcal{H}_{J,m}$. Writing $T_J(P^2)=\Sigma_{m\in \mathrm{Sp}\,P^2}m^2E_m$, where E_m is a projector on $\mathcal{H}_{J,m}$ we have

$$A=\sum\nolimits_{m\in\mathrm{Sp}P^2}a_mE_m,\quad \mathrm{Sp}P^2=\text{Spectrum of } T_J(P^2) \tag{A1}$$

in $\mathcal{H}_J$. Let $\Psi\in S_0\cap\mathcal{H}_J$ have the decomposition

$$\psi=\sum_m E_m\psi=\sum_m\psi_me_m, \tag{A2}$$

where e_m is an eigenvector of $T_J(P^2)$ with eigenvalue m^2. Now, we can write, according to hypothesis

$$0=(\varphi,[AT_J(D)-T_J(D)A]\psi)=(A\varphi,T_J(D)\psi)$$
$$-(\varphi,T_J(D)A\Psi) \qquad \text{(A3)}$$

for φ and Ψ belonging to $S_0\cap\mathcal{H}_J$. Inserting (A1) and (A2), we get

$$0=\sum_{m,m'}(a_m-a_{m'})\varphi_m^*\psi_{m'}(e_m,T_J(D)e_{m'}). \qquad \text{(A4)}$$

Since φ and ψ are arbitrary and $S_0\cap\mathcal{H}_J$ is dense in $\mathcal{H}_J$, we have

$$(a_m-a_{m'})(e_m,T_J(D)e_{m'})=0. \qquad \text{(A5)}$$

Below we shall show that $(e_m,T_J(D)e_{m'})\neq 0$. Then (A5)

$$\Longrightarrow a_m=a_{m'}, \quad \Longrightarrow A=aI.$$

Relative to the basis S_r^J we have

$$(e_m,T_J(D)e_{m'})=\delta_{m_J,m'_J}\delta_{\bar{n},\bar{n}'}d_{k_J,k'_J}, \qquad \text{(A6)}$$

where

$$d_{k_J,k'_J}=-i\tfrac{1}{2}J\delta_{k_Jk'_J}N^2+i\tfrac{1}{2}k'_J\int_0^a dr\,r^2J_{J+1/2}(k_Jr)$$
$$\times J_{J+3/2}(k'_Jr).$$

Put

$$I(x)=\int_0^a dr\,rJ_{J+1/2}(k_Jr)J_{J+1/2}(xk'_Jr).$$

Then

$$\frac{\partial I}{\partial x}(1)=-k'_J\int_0^a dr\,r^2J_{J+1/2}(k_Jr)J_{J+3/2}(k'_Jr)+(J+\tfrac{1}{2})N^2\delta_{k,k'}.$$

But

$$I(x)=-\frac{k_Ja}{k_J^2-x^2k_J'^2}J_{J+1/2}(xk'_Ja)J'_{J+1/2}(k_Ja);$$

hence

$$\frac{\partial I}{\partial x}(1)=-\frac{k_Jak'_Ja}{k_J^2-k_J'^2}J'_{J+1/2}(k'_Ja)J'_{J+1/2}(k_Ja), \quad k_J\neq k'_J.$$

Comparing the two expressions we find

$$k'_J\int_0^a dr\,r^2J_{J+1/2}(k_Jr)J_{J+3/2}(k'_Jr)$$
$$=(J+\tfrac{1}{2})N^2\delta_{k_J,k'_J}+\frac{k_Jk'_Ja^2}{k_J^2-k_J'^2}J'_{J+1/2}(k'_Ja)J'_{J+1/2}(k_Ja).$$

Now $J'_{J+1/2}(k'_Ja)\neq 0$ when $J_{J+1/2}(k_Ja)=0$. This shows that $d_{k_J,k'_J}\neq 0$. QED

[1]E. P. Wigner, in *Theoretical Physics* (IAEA, Vienna, 1963), p. 59.

[2]O. Olsson, "Några relativistiskt kovarianta rörelseekvationer," TRITA-TFY-74-17.

[3]M. Flato and H. Snellman, J. Math. Phys. 15, 857 (1974).

[4]J. D. Jackson, "New particle spectroscopy," TH. 2351-CERN, 29 July 1977.

[5]M. Flato and D. Sternheimer, Commun. Math. Phys. 12, 296 (1969).

An exact solution to Einstein's equations with a stiff equation of state

Paul S. Wesson

St. John's College, Cambridge University, England
(Received 19 January 1978)

A solution to the equations of general relativity is given which is spherically-symmetric and nonstatic with an inhomogeneous density profile ρ and a pressure p given by the stiff equation of state $p = \rho c^2$. The solution may be of use in representing collapsed astrophysical systems or the early stages of an inhomogeneous cosmology.

Solutions to Einstein's equations with the "stiff" equation of state $p = \rho c^2$ have been considered in several contexts, mostly of an astrophysical nature.[1–9] Zeldovitch showed that such an equation of state is consistent with a limiting velocity of sound waves equal to the velocity of light c in such a medium, and numerical work on spherically-symmetric, nonstatic $p = \rho c^2$ solutions has been carried out by Miller[7] and by Henriksen and Wesson.[8] One reason for studying such solutions is that they are believed to be relevant to collapsed astrophysical systems.[6,10–12] There has also been a recent upswing in interest in inhomogeneous cosmological models, and the $p = \rho c^2$ equation of state is expected to be relevant in this connection (a review of anisotropic cosmology has been given by MacCallum,[13] and subsequent work has been reported in Refs. 14–18). A method for generating cosmological solutions of the type treated below has been developed by Wesson.[19] In the next three paragraphs (a) the spherically-symmetric equations of general relativity are put into a form suitable for analysis, (b) a two-parameter family of solutions is obtained possessing the noted stiff equation of state, and (c) a particularly simple solution with only one assignable constant is isolated.

Following Podurets,[20] a spherically-symmetric metric of the following form is taken,

$$ds^2 = c^2 e^{\sigma} dt^2 - e^{\omega} dR^2 - r^2(d\theta^2 + \sin^2\theta \, d\phi^2) \quad (1)$$

Here, R is comoving but $r = r(R, T)$ is not. The metric coefficients σ, ω, r and the parameters of the matter p, ρ, m form six unknowns to be obtained from the field equations. The latter are taken in the form[20,21]

$$2Gm/c^2 r = 1 + e^{-\sigma}\dot{r}^2/c^2 - e^{-\omega} r'^2, \quad (2a)$$

$$m' = 4\pi\rho r^2 r' \quad (2b)$$

$$\dot{m} = -4\pi p r^2 \dot{r}/c^2, \quad (2c)$$

$$\sigma' = -2p'/(p + \rho c^2), \quad (2d)$$

$$\dot{\omega} = -2\dot{\rho} c^2/(p + \rho c^2) - 4\dot{r}/r. \quad (2e)$$

In the above (′) means $\partial/\partial R$ and (·) means $\partial/\partial t$. The mass $m(R, t)$ within some coordinate distance R is defined implicity by these equations. The condition $p(R, t) = \rho(R, t)c^2$ removes one of the six unknowns, leaving five unknowns to be obtained from the five equations (2). To put (2) into a form suitable for solving under the $p = \rho c^2$ assumption, one can assume the existence of a constant t_0 and define three dimensionless functions as follows:

$$\frac{8\pi G\rho r^2}{c^2} \equiv \eta\left(\frac{t}{t_0}\right), \quad \frac{2Gm}{c^2 r} \equiv M\left(\frac{t}{t_0}\right), \quad r \equiv RS\left(\frac{t}{t_0}\right). \quad (3)$$

With (3), (2b) gives $M = \eta$ immediately. Using (3) in (2) generally and eliminating M in favor of η gives

$$\eta = 1 + e^{-\sigma} R^2 \dot{S}^2/c^2 - e^{-\omega} S^2, \quad (4a)$$

$$\dot{\eta}/\eta = -2\dot{S}/S, \quad (4b)$$

$$\sigma' = 2/R, \quad (4c)$$

$$\dot{\omega} = -2\dot{S}/S - \dot{\eta}/\eta. \quad (4d)$$

These four equations have to be solved for σ, ω, S, and η.

The solution family is now easily obtained. Equation (4c) gives $e^{\sigma} = (R/ct)^2$ where part of an arbitrary function $\exp(t/t_0)$ has been put equal to unity. Equation (4d) gives $e^{\omega} = 1$, choosing an arbitrary function $\exp[\omega(R/ct_0)] = \eta_0^2$, where η_0 is a constant. The latter constant is defined by the integral of (4b), which gives $\eta = (\eta_0/S)^2$. With the specified choices of the arbitrary functions of integration involved, the remaining equation to be solved is (4a) in the form

$$S^2(1 + \dot{S}^2 t^2 - S^2) - \eta_0^2 = 0. \quad (5)$$

Solving this equation and collecting together the other parameters gives the full solution family for the metric coefficients e^{σ}, e^{ω}, S and the matter properties p, ρ, m as defined by (3) as follows:

$$e^{\sigma} = (R/ct)^2 \quad e^{\omega} = 1, \quad r \equiv RS, \quad (6a)$$

$$(S^4 - S^2 + \eta_0^2)^{1/2} + S^2 - \tfrac{1}{2} = (t/T_0)^{\pm 2}, \quad (6b)$$

$$p = \rho c^2 = c^4\eta_0^2/8\pi G R^2 S^4, \quad m = c^2\eta_0^2 R/2GS. \quad (6c)$$

The solution (6) is a two-parameter family depending on the two constants T_0 (defined in place of t_0 above) and η_0. It represents an evolving [expanding or contracting, depending on the sign choice in (6b)] spherically-symmetric configuration of matter with an inhomogeneous ($\propto R^{-2}$) density and pressure profile.

For the special choice $\eta_0 = \frac{1}{2}$, the solution (6) takes a particularly simple form which, because it is likely to be of more immediate use than (6) in astrophysical contexts, can be stated in metric form as follows:

 0022-2488/78/1911-2283$1.00

$$ds^2 = (R/t)^2\,dt^2 - dR^2 - R^2S^2(d\theta^2 + \sin^2\theta\,d\phi^2), \tag{7a}$$

$$2S^2 = 1 + (t/T_0)^{\pm 2}, \tag{7b}$$

$$p = \rho c^2 = c^4/32\pi G R^2 S^4, \qquad m = c^2R/8GS. \tag{7c}$$

This solution, which has only one assignable constant (T_0) and whose validity (like that of the full two-parameter family) may be checked by resubstitution into (2), is of astrophysical interest. In particular, it might be used in situations involving white holes,[10] lagging cores of a big-bang cosmology,[11,12] other condensed objects like neutron stars or the central regions of quasars, and the early stages of inhomogeneous cosmological models.[3,5] It is hoped that the family of solutions (6) or the simple form (7) will find application in one or more of these contexts.

ACKNOWLEDGMENTS

Thanks for comments are due to R.N. Henriksen, R. Geroch, and I. Roxburgh. This work was supported by a Royal Society research fellowship.

[1] K. Gödel, Rev. Mod. Phys. **21**, 447–50 (1949).
[2] Y.B. Zeldovitch, Sov. Phys.-JETP **14**, 1143–7 (1961).
[3] Y.B. Zeldovitch, Mon. Not. R. Astron. Soc. **160**, 1P–3P (1972).
[4] L.C. Shepley, Ph.D. dissertation, Princeton University (University Microfilms, Ann Arbor, Michigan, 1965).
[5] E.P. Liang, Mon. Not. R. Astron. Soc. **171**, 551–3 (1975).
[6] M.P. Ryan and L.C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U.P., Princeton, N.J., 1975).
[7] B.D. Miller, Astrophys. J. **208**, 275–85 (1976).
[8] R.N. Henriksen and P.S. Wesson, Astrophys. Space Sci. **53**, 429–44 (1978).
[9] G.V. Bicknell and R.N. Henriksen, Astrophys. J. **219**, 1043–57 (1978).
[10] D. Eardley, Phys. Rev. Lett. **33**, 442–4 (1974).
[11] I.D. Novikov, Sov. Astron. AJ **8**, 853–63 (1965).
[12] Y. Ne'eman, Astrophys. J. **141**, 1303–5 (1965).
[13] M.A.H. MacCallum, in *Cargese Lectures in Physics* (Gordon and Breach, London, 1973), Vol. 6, pp. 61–174.
[14] J. Silk, Astrophys. J. **193**, 525–7 (1974).
[15] J. Silk, Astron. Astrophys. **59**, 53–8 (1977).
[16] P. Szekeres, Commun. Math. Phys. **41**, 55–64 (1975).
[17] W.B. Bonner and N. Tomimura, Mon. Not. R. Astron. Soc. **175**, 85–93 (1976).
[18] E.P. Liang, Astrophys. J. **204**, 235–50 (1976).
[19] P.S. Wesson, "A New Look at the Cosmological Principle," Astron. Astrophys. (to be published).
[20] M.A. Podurets, Sov. Astron. AJ **8**, 19–22 (1964).
[21] C.W. Misner and D.H. Sharp, Phys. Rev. **136**, B571–B576 (1964).

The conserved densities of the Korteweg–De Vries equation

Mark J. McGuinness

Department of Physics, University of Canterbury, New Zealand
(Received 19 October 1977; revised manuscript received 2 March 1978)

The conserved densities of the Korteweg–de Vries equation are identified as energy densities associated with higher order equations generated from the KdV equation and governing its solutions.

1. INTRODUCTION

Miura, Gardner, and Kruskal[1] have discovered an infinite number of conserved densities for the Korteweg–de Vries (KdV) equation in the course of their extensive work on this equation. Noether's theorem[2] provides a natural way of associating a conserved quantity of a system with an infinitesimal transformation on that system. Since in many cases there is a biunique correspondence between conservation law and transformation,[3] the conservation law can often be labeled or identified by its associated transformation. For example, a conserved density can be identified as an energy density if it is associated with time translation, an angular momentum density if it is associated with rotation, or a charge density if it is associated with a phase or gauge transformation.

Steudel[4] has shown that extended Bäcklund transformations are associated with the infinite set of conservation laws for the KdV equation, via Noether's theorem. The purpose of this paper is to give an alternative simpler identification of these conservation laws by associating them with infinitesimal time translations on new (integrodifferential) equations of motion. These equations are higher-order KdV equations, and their solution sets contain the solution set of the KdV equation. That is, solutions to the KdV equation must also be solutions to these higher-order equations, so that solutions to the KdV equation must obey any conservation laws for the higher-order equations.

Hence this paper identifies each of the infinite number of conserved quantities for the KdV equation, as a conserved energy density for each of the infinite number of higher-order (integrodifferential) KdV equations, which must be obeyed by all solutions to the KdV equation.

It is hoped that a similar approach will yield a general technique for explaining via Noether's theorem the infinite sets of conservation laws associated with other nonlinear equations, in terms of higher order equations which also govern these systems.

A brief summary of Noether's theorem and a generalized Noether's theorem will be given in Sec. 2. The main result of this paper follows, and a corollary that the energy density of the nth "generalized" KdV equation, discovered by Lenard (see Gardner, Greene, Kruskal and Miura[5]), is identical to the nth polynomial conserved density of the KdV equation, is given in Sec. 4.

2. NOETHER'S THEOREM AND GENERALIZATION

Noether's theorem can be applied to a system which can be described by the Euler–Lagrange equation of motion

$$E(\mathcal{L})=0, \tag{2.1}$$

where

$$E\equiv\sum_{a=0}^{\infty}(-1)^a d_{\mu_1}\cdots d_{\mu_a}\left(\frac{\partial}{\partial\phi_{\mu_1,\ldots,\mu_a}}\right) \tag{2.2}$$

with the sum being over different combinations of $\{\mu_1,\cdots,\mu_a\}$, and the Lagrangian density is

$$\mathcal{L}=\mathcal{L}(x,\phi,\phi_\mu,\phi_{\mu\nu},\cdots),$$

where

$$d_{\mu_1}\equiv\frac{d}{dx^{\mu_1}},\quad \phi_{\mu_1}\equiv\frac{\partial\phi}{\partial x^{\mu_1}}(x).$$

Noether's theorem states that if the action integral

$$\mathcal{I}=\int_v \mathcal{L}\,dx \tag{2.3}$$

is invariant under the infinitesimal transformation

$$x^1=x+\delta x,\quad \phi^1(x)=\phi(x)+\bar{\delta}\phi(x), \tag{2.4}$$

then the following relation holds,

$$-\bar{\delta}\phi E(\mathcal{L})=d_\mu[\pi^\mu(\mathcal{L})\bar{\delta}\phi+\mathcal{L}\,\delta x^\mu+G^\mu], \tag{2.5}$$

where

$$\delta\mathcal{I}=\int_v d_\mu G^\mu dx$$

under the transformation (2.4), and

$$\pi^\nu(\mathcal{L})\equiv\sum_{a=0}^{\infty}\sum_{b=0}^{\infty}(-1)^b d_{\mu_1}\cdots d_{\mu_b}\left(\frac{\partial\mathcal{L}}{\partial\phi_{\mu_1,\ldots,\mu_a,\nu}}\right)\times d_{\mu_{b+1}}\cdots d_{\mu_a}. \tag{2.6}$$

Note that ν is summed over as a repeated index, and the $(\mu_1,\cdots,\mu_a)$ are summed over combinations.

The relation (2.5) yields the conservation law for solutions to (2.1)

$$d_\mu[\pi^\mu(\mathcal{L})\bar{\delta}\phi+\mathcal{L}\,\delta x^\mu+G^\mu]=0. \tag{2.7}$$

Noether's theorem has been generalized in such a way that the existence of a Lagrangian density is not necessary in order to associate a conservation law of a system of equations with an infinitesimal transformation on that system. This has been done in a series of papers by Rosen.[6] If the equation of motion for the system is

$$F(x,\phi,\phi_\mu,\cdots)=0 \tag{2.8}$$

and if the relation

$$-\bar{\delta}\phi F = d_\mu J^\mu - K \tag{2.9}$$

holds, where $K(x,\phi,\phi_\mu,\cdots)$ is zero for solutions and is linearly independent of F, then the generalized Noether's theorem associates the field variation $\bar{\delta}\phi$ with the conservation law

$$d_\mu J^\mu = 0 \tag{2.10}$$

for solutions of the equations of motion (2.8).

Note the similarity of relation (2.9) to relation (2.5). Rosen has shown that if a Lagrangian density $\mathcal{L}$ exists for the system, the generalized Noether's theorem is equivalent to Noether's theorem, so that either method can be used to associate a transformation with a conservation law. It should be noted that the relations (2.5) and (2.9) have to hold for all values of the field variables, not just for solutions.

3. THE ENERGY DENSITIES OF THE HIGHER ORDER KdV EQUATIONS

The higher-order KdV equations are generated from the KdV equation

$$\phi_{xt} + \phi_1\phi_2 + \phi_4 = 0, \tag{3.1}$$

where

$$\phi_1 \equiv \phi_x \equiv \frac{\partial\phi}{\partial x}(x,t), \quad \phi_t \equiv \frac{\partial\phi}{\partial t}(x,t),$$

by operating on it n times in succession with

$$H \equiv d_x^2 + \tfrac{2}{3}\phi_1 + \tfrac{1}{3}\phi_2 \int^x dx, \tag{3.2}$$

where the lower limit of the integral is such that ϕ and its derivatives vanish on that boundary, giving

$$H^n(\phi_{xt} + \phi_1\phi_2 + \phi_4) = 0, \quad n = 0,1,\cdots. \tag{3.3}$$

This operator first appeared in a recursion formula due to Gardner, Greene, Kruskal, and Miura[5] for the conserved densities of the KdV equation

$$H d_x A_n = d_x A_{n+1}, \quad A_1 = \phi_x. \tag{3.4}$$

Their operator is equal to H within a constant number if their KdV equation is transformed to the form used in this paper. In the same paper, H is a generator for a class of evolution equations for the time independent Schrodinger equation which leave the discrete eigenvalues invariant in time, and which possess the same set of conserved densities as the KdV equation (the "generalized" KdV equations)

$$u_t + H^n u_x = 0, \quad n = 0,1,\cdots, \tag{3.5}$$

where $u \equiv \phi_x$. Note that these equations are partial differential,[5] while equations (3.3) are integrodifferential. However, equations (3.3) are solved by solutions to the KdV equation, and Eq. (3.5) are not. Hence, a system with a KdV equation of motion will be governed by Eq. (3.3), and by their conservation laws.

The energy vector of the nth equation (3.3) is J_n^μ such that

$$d_\mu J_n^\mu = -\phi_t H^n(\phi_{xt} + \phi_1\phi_2 + \phi_4) + K_n, \tag{3.6}$$

where K_n is linearly independent of the nth equation (3.3). J_n^μ is an energy vector because the generalized Noether's theorem [Eq. (2.9)] associates it with the transformation

$$\bar{\delta}\phi = -\epsilon\phi_t. \tag{3.7}$$

That is,

$$\delta t = -\epsilon, \quad \delta x = 0, \quad \delta\phi \equiv \phi^1(x^1) - \phi(x) = 0$$

which is an infinitesimal time translation (ϵ is some infinitesimal parameter).

If K_n is zero, at least for solutions to the nth equation (3.3), then J_n^t is a conserved energy density for that equation.

It will be shown that

$$-\phi_t H^n(\phi_{xt} + \phi_1\phi_2 + \phi_4) = d_t A_{n+3} + d_x X_{n+3}, \tag{3.8}$$

where the A_{n+3} are the conserved densities for the KdV equation generated by formula (3.4). Clearly, A_{n+3} will then be identified as a conserved energy density for the nth equation (3.3). Note that Eq. (3.8) must be proved for all values of the dependent variable ϕ; this variable is not required to be a solution of any of the Eqs. (3.3).

Equation (3.8) will be proved in two parts. In part (a) it will be shown that

$$\phi_t H^{n+1}(\phi_{xt}) = d_x F_{n+1}, \tag{3.9}$$

where F is equivalent to a function of ϕ and its derivatives, in a sense that will be defined in part (a).

In part (b) it will be shown that

$$-\phi_t H^n(\phi_1\phi_2 + \phi_4) = d_t A_{n+3} + d_x X'_{n+3}, \tag{3.10}$$

where X'_{n+3} is a function of ϕ and its derivatives.

This will be done using the conventional Noether's theorem, deriving a Lagrangian density for the lhs and hence the (conserved) energy density, treating

$$H^n(\phi_1\phi_2 + \phi_4) = 0 \tag{3.11}$$

as Euler–Lagrange equation for this purpose.

A. Proof of the integrodifferential part

Equation (3.9) will be proved by induction, assuming

$$\phi_t H^i(\phi_{xt}) = d_x F_i \quad \text{for } i = 0,1,2,\cdots,n. \tag{3.12}$$

Hence

$$\phi_t H^{n+1}(\phi_{xt}) = \phi_t H\left(\frac{d_x F_n}{\phi_t}\right). \tag{3.13}$$

The rhs can be rearranged as

$$d_x\left[d_x^2 F_n - 2\phi_{xt}\frac{d_x F_n}{\phi_t} + \tfrac{1}{3}\left(\int\phi_2\phi_t\right)\int\frac{d_x F_n}{\phi_t}\right] + \frac{d_x F_n}{\phi_t}\int\frac{d_x F_1}{\phi_t} \tag{3.14}$$

noting that

$$\int\frac{d_x F_1}{\phi_t} = \phi_{2t} + \tfrac{2}{3}\phi_1\phi_t - \tfrac{1}{3}\int\phi_2\phi_t. \tag{3.15}$$

Lemma

$$\frac{d_x F_{k+1}}{\phi_t}\int\frac{d_x F_i}{\phi_t} = d_x Q_{i,k} + \frac{d_x F_k}{\phi_t}\int\frac{d_x F_{i+1}}{\phi_t}, \quad \forall\ i \le n,\ k < n. \tag{3.16}$$

Proof: From Eq. (3.9)

$$\frac{d_xF_{k+1}}{\phi_t}\int\frac{d_xF_i}{\phi_t}=H\left(\frac{d_xF_k}{\phi_t}\right)\int\frac{d_xF_i}{\phi_t} \tag{3.17}$$

and the rhs can be rearranged as

$$d_x\left[\frac{d_x^2F_k}{\phi_t}\int\frac{d_xF_i}{\phi_t}-\phi_{xt}\frac{d_xF_k}{\phi_t^2}\int\frac{d_xF_i}{\phi_t}-\frac{d_xF_i}{\phi_t}\frac{d_xF_k}{\phi_t}+\tfrac{1}{3}\left(\int\frac{d_xF_k}{\phi_t}\right)\left(\int\phi_2\int\frac{d_xF_i}{\phi_t}\right)\right]+\left[\left(d_x^2+\tfrac{2}{3}\phi_1-\tfrac{1}{3}\int\phi_2\right)\int\frac{d_xF_i}{\phi_t}\right]\frac{d_xF_k}{\phi_t} \tag{3.18}$$

which equals

$$d_xQ_{i,k}+\frac{d_xF_k}{\phi_t}\int\frac{d_xF_{i+1}}{\phi_t},$$

where

$$Q_{i,k}\equiv\frac{d_x^2F_k}{\phi_t}\int\frac{d_xF_i}{\phi_t}-\phi_{xt}\frac{d_xF_k}{\phi_t^2}\int\frac{d_xF_i}{\phi_t}-\frac{d_xF_i}{\phi_t}\frac{d_xF_k}{\phi_t}+\tfrac{1}{3}\left(\int\frac{d_xF_k}{\phi_t}\right)\left(\int\phi_2\int\frac{d_xF_i}{\phi_t}\right). \tag{3.19}$$

Applying the lemma to the last term in expression (3.14) we have

$$\phi_tH^{n+1}(\phi_{xt})=d_xQ_{0,n}+d_xQ_{1,n-1}+\frac{d_xF_{n-1}}{\phi_t}\frac{d_xF_2}{\phi_t}. \tag{3.20}$$

Repeatedly applying the lemma to the nondivergence term

$$\phi_tH^{n+1}(\phi_{xt})=\sum_{i=0}^{k-1}d_xQ_{i,n-i}+\frac{d_xF_k}{\phi_t}\int\frac{d_xF_k}{\phi_t}\quad\text{if } n=2k-1 \tag{3.21}$$

or

$$\phi_tH^{n+1}(\phi_{xt})=\sum_{i=0}^{k-1}d_xQ_{i,n-i}+\frac{d_xF_{k+1}}{\phi_t}\int\frac{d_xF_k}{\phi_t}\quad\text{if } n=2k \tag{3.22}$$

so that if $n=2k-1$

$$\phi_tH^{n+1}(\phi_{xt})=d_x\left[\sum_{i=0}^{k-1}Q_{i,n-i}+\tfrac{1}{2}\left(\int\frac{d_xF_k}{\phi_t}\right)^2\right]\equiv d_xF_{n+1} \tag{3.23}$$

and if $n=2k$

$$\phi_tH^{n+1}(\phi_{xt})=d_x\left[\sum_{i=0}^{k-1}Q_{i,n-i}+\tfrac{1}{2}Q_{k,k+1}+\tfrac{1}{2}\left[\left(\int\frac{d_xF_k}{\phi_t}\right)\left(\int\frac{d_xF_{k+1}}{\phi_t}\right)\right]\right]\equiv d_xF_{n+1}. \tag{3.24}$$

Since

$$\phi_tH^0(\phi_{xt})=\phi_t\phi_{xt}=d_x(\tfrac{1}{2}\phi_t^2)$$

the induction process is started and Eq. (3.9) is proved for any n.

It will be noted that $Q_{i,k}$ contains the term

$$\left(\int\frac{d_xF_k}{\phi_t}\right)\left(\int\phi_2\int\frac{d_xF_i}{\phi_t}\right), \tag{3.25}$$

where the integrals are taken from $-\infty$ to x. To obtain a conserved density from a conservation equation

$$d_tT+d_xX=0 \tag{3.26}$$

the equation is integrated over x from $-\infty$ to $+\infty$, and the assumption that the field variables and their derivatives are zero at $x=\pm\infty$ is used to obtain

$$d_t\int_{-\infty}^{\infty}T\,dx=0, \tag{3.27}$$

hence the name "conserved density" for T. If, however, X contains a term like (3.25), Eq. (3.27) will not hold in general, since integrals of field variables do not necessarily vanish at infinity. One integral which appears in all such terms in F_n is

$$\int_{-\infty}^{x}\frac{d_xF_i}{\phi_t}\,dx. \tag{3.28}$$

It will be shown that

$$\frac{d_xF_i}{\phi_t} \tag{3.29}$$

is an x-divergence for solutions to equations (3.3) for $n\leq i$. Then integral (3.28) will be equivalent to a localized function of ϕ and its derivatives (where "equivalent" means "equal to, for solutions," when applied to a density or a flux in a conservation equation, as in Steudel's definition[3]), which will disappear when evaluated at the boundaries under the usual assumptions about the behavior of ϕ and its derivatives on these boundaries. Hence the integral terms in F_i will be acceptable flux terms in a conservation equation, yielding a conserved quantity on integration of the equation over x.

To show that expression (3.29) is an x-divergence for solutions, Eq. (3.12) should be rearranged as

$$H^i(\phi_{xt})=\frac{d_xF_i}{\phi_t},\quad i=0,\ldots,n \tag{3.30}$$

and Eq. (3.4) implies that

$$H^i(d_xA_0)=d_xA_i. \tag{3.31}$$

Hence

$$H^i\left(\frac{d_xF_0}{\phi_t}+d_xA_2\right)=\frac{d_xF_i}{\phi_t}+d_xA_{i+2}, \tag{3.32}$$

and since

$$\frac{d_xF_0}{\phi_t}+d_xA_2=\phi_{xt}+\phi_1\phi_2+\phi_4, \tag{3.33}$$

the lhs of Eq. (3.32) is zero for solutions to Eq. (3.3) for all $n\leq i$, and

$$\frac{d_xF_i}{\phi_t}=d_x(-A_{i+2}),\quad i=0,\ldots,n \tag{3.34}$$

for solutions. The work of Gardner, Greene, Kruskal, and Miura[5] shows A_{i+2} to be a polynomial in ϕ_x and its x derivatives, hence the integral (3.28) is equivalent to a localized function.

B. Proof of the partial differential part

To prove Eq. (3.10), the extensive literature on the KdV equation and the "generalized" KdV equations can be drawn upon to obtain a Lagrangian density for Eq. (3.11).

Equation (3.4) can be rewritten as

$$H^n(\phi_1\phi_2+\phi_4)=d_xA_{n+2} \tag{3.35}$$

and Miura[7] points out that

$$\frac{\delta}{\delta u}\int_{-\infty}^{\infty} A_m\,dx = A_{m-1}, \quad u\equiv\phi_x \tag{3.36}$$

within a constant number, where $\delta/\delta u$ is called the gradient or the variational derivative of the functional $\int A_m dx$, and is defined by

$$\frac{d}{d\alpha}F[u(\alpha)] = \int \frac{\delta F}{\delta u}\frac{d}{d\alpha}u(\alpha)\,dx, \tag{3.37}$$

where

$$F = \int G(u)\,dx$$

and the limits of integration are such that u and its derivatives are zero on the boundaries.

It is noted by Kruskal, Miura, Gardner, and Zabusky,[8] and can easily be shown from the definition, that

$$\frac{\delta}{\delta u}\int A_m(u)\,dx = E_u[A_m(u)], \tag{3.38}$$

where

$$E_u \equiv \sum_{n=0}(-1)^n d_{\mu_1}\cdots d_{\mu_n}\left(\frac{\partial}{\partial u_{\mu_1,\ldots,\mu_n}}\right) \tag{3.39}$$

which is the familiar Euler–Lagrange operator. Hence from Eqs. (3.35), (3.36), and (3.38)

$$H^n(\phi_1\phi_2+\phi_4) = d_x\{E_u[A_{n+3}(u)]\}. \tag{3.40}$$

A straightforward calculation shows that, as is noted by Miura[7]

$$d_x E_u[F(u)] = -E_\phi[F(\phi_x)] \tag{3.41}$$

where $u\equiv\phi_x$ and $F(\phi_x)$ is $F(u)$ with every u replaced by a ϕ_x. Then from Eqs. (3.40) and (3.41)

$$H^n(\phi_1\phi_2+\phi_4) = -E_\phi[A_{n+3}(\phi_x)]. \tag{3.42}$$

That is, $-A_{n+3}(\phi_x)$ is a Lagrangian density yielding $H^n(\phi_1\phi_2+\phi_4)$ as an Euler–Lagrange equation.

To use Noether's theorem to find the energy density from this Lagrangian density, the variation of the action integral $\mathcal{J}$ must be determined under the infinitesimal time translation.

$$\delta t=\epsilon,\quad \delta x=0,\quad \delta\phi=0,\quad \bar{\delta}\phi=-\epsilon\phi_t. \tag{3.43}$$

The variation of $\mathcal{J}$ can be written

$$\delta\mathcal{J} = \int\int \delta\mathcal{L}\,dxdt \tag{3.44}$$

$$= \int\int\left(\sum \frac{\partial\mathcal{L}}{\partial\phi_{\mu_1,\ldots,\mu_a}}\delta\phi_{\mu_1,\ldots,\mu_a} + \frac{\partial\mathcal{L}}{\partial x^\mu}\delta x^\mu\right)dxdt \tag{3.45}$$

$$= \int\int \frac{\partial\mathcal{L}}{\partial t}\epsilon\;\; dxdt \tag{3.46}$$

$$=0, \tag{3.47}$$

since

$$\mathcal{L} = -A_{n+3}(\phi_x) \tag{3.48}$$

which has no explicit t dependence. Inserting results (3.47) and (3.48) into Noether's relation (2.5) gives Eq. (3.10),

$$-\phi_t H^n(\phi_1\phi_2+\phi_4) = d_t(A_{n+3}) + d_x(X'_{n+3}), \tag{3.49}$$

where

$$X'_{n+3} = -\pi^x(A_{n+3})\phi_t, \tag{3.50}$$

and noting that

$$\pi^t[A_{n+3}(\phi_x)]\phi_t = 0. \tag{3.51}$$

This completes the proof of Eq. (3.8), which is obtained by simply adding left- and right-hand sides of Eqs. (3.9) and (3.10).

4. COROLLARY FOR GENERALIZED KdV EQUATIONS

These results also lead to the conclusion that the conserved energy density associated with each "generalized" KdV equation (3.5) is identical to the nth polynomial conserved density for the KdV equation. This follows from the equation

$$E_\phi[-\tfrac{1}{2}\phi_x\phi_t - A_{n+3}(\phi_x)] = \phi_{xt} + H^n(\phi_1\phi_2+\phi_4) \tag{4.1}$$

and since A_{n+3} is independent of ϕ_t, the conserved energy density of the nth "generalized" KdV equation is just A_{n+3}.

This result is closely paralleled by Gardner's work[9] on the KdV equation as a Hamiltonian system. He shows that the Hamiltonian functional yielding the nth "generalized" KdV equation is

$$\int_0^{2\pi} A_{n+3}\,dx \tag{4.2}$$

for periodic solutions with period 2π. In theoretical mechanics the Hamiltonian is often defined in such a way that it is equal to the energy functional in the Lagrangian formalism.[10] The equation (4.1) shows that Gardner's Hamiltonians are indeed equal to the (conserved) energy integrals of the "generalized" KdV equations which they generate.

ACKNOWLEDGMENT

I am indebted to Professor W.L. Jones for his guidance in this work.

[1]R.M. Miura *et al.*, J. Math. Phys. **9**, 1204 (1968).
[2]E. Noether, Kgl. Ges. Wiss. Nachrichten (Göttingen) Math. Phys. Klasse (1918), p. 235; J. Rosen, Int. J. Theor. Phys. **4**, 287 (1971); E.L. Hill, Rev. Mod. Phys. **23**, 253 (1951); U.E. Schroder, Fortschr. Phys. **16**, 357 (1968).
[3]H. Steudel, Ann. Phys. **32**, 214 (1975).
[4]H. Steudel, Ann. Phys. **32**, 445 (1975).
[5]C.S. Gardner *et al.*, Comm. P. Ap. Math. **27**, 97 (1974).
[6]J. Rosen, Ann. Phys. (N.Y.) **69**, 349 (1972); Ann. Phys. (N.Y.) **82**, 54 (1974); Ann. Phys. (N.Y.) **82**, 70 (1974).
[7]R.M. Miura, Siam Rev. **18**, 447 (1976).
[8]M.D. Kruskal *et al.*, J. Math. Phys. **11**, 954 (1970).
[9]C.S. Gardner, J. Math. Phys. **12**, 1548 (1971).
[10]E.J. Saletan and A.H. Cromer, *Theoretical Mechanics* (Wiley, New York, 1971), pp. 177–8, 287–8.

Bogoliubov transformations, propagators, and the Hawking effect[a)]

A. S. Lapedes

D.A.M.T.P., University of Cambridge, Silver Street, Cambridge, England and University of California, [b)] *Department of Physics, Santa Barbara, California 93106*

(Received 25 January 1978)

The thermal spectrum of radiation, seen by suitable observers using "Unruh particle detectors," in de Sitter spacetime is recovered using Bogoliubov transformation techniques. Previous attempts by other authors at calculating particle production in de Sitter spacetime, prior to the discovery of thermal radiation using propagator techniques failed. The discrepancy between these previous mode mixing calculations, and the calculations presented here, are traced to "de Sitter invariant" versus "observer dependent" formalisms. One consequence is that the initial vacuum state chosen for the quantum field is not unique.

I. INTRODUCTION

In this paper we analyze the relation of Bogoliubov transformation calculations of the Hawking effect with alternative propagator techniques, using de Sitter spacetime as an illustration. The conclusions should also apply to the known class of geometries in which an observer following the time like trajectory of a Killing vector of the spacetime detects a thermal spectrum of radiation, namely the Kerr–Newman–de Sitter[1] family and the Taub–NUT type geometries.[2] Specifically, we shall show that, for de Sitter spacetime, for which the Hawking effect has been derived using only propagator techniques,[1] an alternative calculation is possible using a Bogoliubov transformation between two distinct sets of modes in which the quantum test field φ is expanded.

Bogoliubov or "mode mixing" calculations in de Sitter space have been utilized extensively by various authors[3–5] prior to the Gibbons–Hawking "path integral" calculation; however, none of these authors predicted a thermal spectrum of radiation. This apparent discrepancy is a result of the previous author's attempts to construct a de Sitter group invariant method of calculating particle creation, in which the simple group structure of de Sitter space was used to select a de Sitter group invariant vacuum. In this paper we will use a "mode mixing" method of calculating particle production, but our results will not be de Sitter invariant; instead they will depend on the state of motion of the particle detector an observer uses. The thermal spectrum of radiation seen by the Gibbons–Hawking class of observers is correctly reproduced.

One result of the calculation is that the initial set of modes, and hence associated initial vacuum states, are arbitrary to a large degree. The possible initial vacuum states differ by an infinite number of particles (i.e., one "vacuum state" appears as a many-particle state relative to another choice of "vacuum state"); however, the "particles" are effectively redshifted out by the de Sitter event horizon and are not detected by a late time observer. It is essential to use a final vacuum state for the late time observer defined through use of positive frequency modes with respect to the locally timelike Killing vector. These modes correctly describe particles detected by an observer using an Unruh detector[1] who follows the trajectory of the timelike Killing vector. The phrase "observer dependence" will be used in this paper to refer to the necessity of adapting the quantum field theory discussion to the state of motion of an observer and his particle detector. The failure of previous authors to use such a formalism in similar calculations in de Sitter spacetime is the reason for the apparent discrepancy of their results with those of Gibbons and Hawking.[1]

In the following section we briefly review the relevant properties of de Sitter spacetime.

II. DE SITTER SPACE–THE GLOBAL PROPERTIES[6]

De Sitter space is an exact solution of the vacuum Einstein equations including the "cosmological term" and has constant positive curvature with a ten-parameter isometry group. It has topology $R\times S^3$ and can be written in a chart which covers the whole space (apart from trivial polar coordinate singularities) as:

$$ds^2 = -dt'^2 + R^2\cosh^2 t'/R\, d\Omega^2{}_3, \tag{1}$$

where

$$d\Omega^2{}_3 = d\chi^2 + \sin^2 x\,(d\theta^2 + \sin^2\theta d\varphi^2).$$

Constant t' sections are three spheres, where t' takes on all values from -00 to $+00$. χ ranges from 0 to π, and due to the homogeneity of the spacetime an observer can always be taken to be at the origin, $\chi=0$. θ and φ are the usual two-sphere polar and azimuthal angles. De Sitter space can also be written in a chart which does not cover the whole space:

$$ds^2 = -\left(1-\frac{r^2}{R^2}\right)dt^2 + \frac{dr^2}{1-r^2/R^2} + r^2(d\theta^2+\sin^2\theta d\varphi^2). \tag{2}$$

There is a coordinate singularity at $r=R$ which represents an event horizon for an observer stationed at $r=0$, following the trajectory of the Killing vector ∂_t.

[a)]Supported in part by the National Science Foundation.
[b)]Present address.

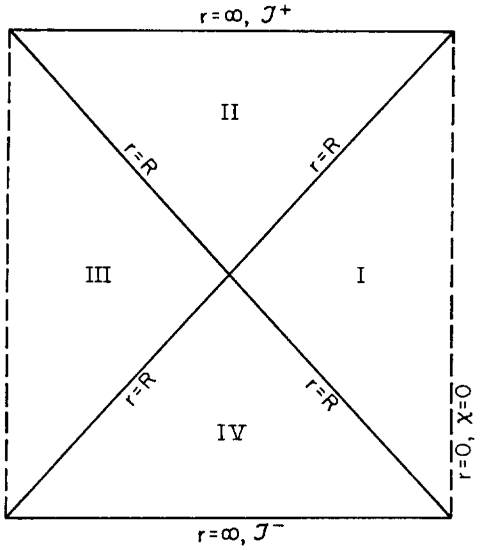

FIG. I.

∂_t is normalized to unity at $r=0$. The Penrose diagram of de Sitter space is shown in Fig. 1.

A convenient representation of de Sitter space is by a hyperbola embedded in a fictitious five-dimensional space with Lorentzian signature. A pseudo-spherical coordinate transformation yields Eq. (1):

$$v=R\sinh t'/R, \tag{3a}$$

$$w=R\cosh t'/R\cos\chi, \tag{3b}$$

$$x=R\cosh t'/R\ \sin\chi\cos\theta, \tag{3c}$$

$$y=R\cosh t'/R\sin\chi\sin\theta\cos\varphi, \tag{3d}$$

$$z=R\cosh t'/R\ \sin\chi\sin\theta\sin\varphi, \tag{3e}$$

$$ds^2=-dv^2+dw^2+dx^2+dy^2+dz^2.$$

Taking pseudo-spherical coordinates about a different axis in the five-dimensional Minkowski space yields Eq. (2):

$$v=(R^2-r^2)^{1/2}\sinh t/R, \tag{4a}$$

$$w=(R^2-r^2)^{1/2}\cosh t/R, \tag{4b}$$

$$x=r\sin\theta\cos\varphi, \tag{4c}$$

$$y=r\sin\theta\sin\varphi, \tag{4d}$$

$$z=r\cos\theta. \tag{4e}$$

Comparison of Eqs. (3) and (4) gives the transformation between the "globally good" coordinates t', χ and the "static" coordinates t, r:

$$\tanh t/R=(1/\cos\chi')\tanh t'/R \tag{5a}$$

$$r=R\sin\chi\cosh t'/R. \tag{5b}$$

III. CALCULATIONS

Unruh,[7] in a gedanken experiment involving a model particle detector in which a quantized field in a box is excited to higher energy levels if a particle is detected, has demonstrated that the correct time dependence of an unlocalized mode describing a "particle"/"antiparticle" is $\exp(\pm i\omega t)$, where t is the proper time along the detector/observer world line. In other words, the definition of a mode describing a particle is logically taken to be that mode which will excite a particle detector, and Unruh has shown that this mode is necessarily positive frequency with respect to the observer's proper time. In de Sitter spacetime, an observer stationed at the origin and at rest with respect to the timelike Killing vector ∂_t, can therefore define a particle mode, $\varphi_\omega^{\text{static}}$, as a solution of the conformally invariant (for simplicity) field equation:

$$(1/\sqrt{-g})\,\partial_a(\sqrt{-g}\,g^{ab}\partial b\varphi)+\tfrac{1}{6}R\varphi=0\,, \tag{6}$$

written in the "static" coordinates of Eq. (2), which is positive frequency with respect to the time t. We can write the normalized solution as

$$\phi^{\text{static}}=\frac{\exp(-i\omega t)}{\sqrt{\omega}}\,\frac{f_{\omega lm}(r)}{r}\,Y_{lm}(\theta,\phi). \tag{7}$$

Because the coordinate system defined in Eq. (2) does not cover the whole space but only region I of the figure, the mode φ^s, together with its complex conjugate, does not define a complete set of modes and it is necessary to define modes in the other half of the spacetime. The modes in this half, region III of Fig. 1, will be functionally equivalent to the modes of region I, but will only cover region III, and be identically zero in reigion I. We can therefore expand the field ϕ in a complete set of modes:

$$\phi=\sum_\omega {}_R a_{\omega\, R}\phi_\omega+{}_L a^{+}_{\omega\, L}\phi_\omega+\text{h.c.}, \tag{8}$$

where "h.c." stands for "Hermitian conjugate," the subscripts "L" and "R" stand for left and right respectively (referring to regions III and I respectively), the sum over the subscript ω is a shorthand for the appropriate sum over ω, l, m (since the spacetime is static and spherically symmetric the sum over l and m will rarely be of interest), and the association of the modes ϕ and $\bar\phi$ with the associated annihilation and creation operators and ${}_R a_\omega$, ${}_L a^{+}_\omega$, is determined by requiring the operators to have the correct commutation relations.[8] In the rest of this chapter we will be concerned with what an observer stationed at $r=0$ detects, and will only be interested in the mode ${}_R\phi_\omega$, which does not cover the whole spacetime. The "out vacuum" will be defined with respect to this mode and is the state $|0_+\rangle$ defined by $a_\omega|0_+\rangle=0$. The state $|0_+\rangle$ is a "no particle" state for the observer at $r=0$.

It is now necessary to define an alternative set of modes which cover the whole spacetime and have an "in vacuum" associated with them. We can then calculate the Bogoluibov transformation between the two sets of modes and obtain the particle creation rate. A complete set of modes which cover the whole spacetime can be defined by writing the wave equation (6) in the globally good coordinates of Eq. (1). Such modes can be written as

$$\phi_{\text{good}}=AF_N(t')Y^{(3)}_{Nlm}(\chi',\theta',\phi'), \qquad A=\text{normalization constant}, \tag{9}$$

where $Y^{(3)}_{Nlm}$ represents the hyperspherical harmonics on a 3-sphere and $F_N(t')$ is a particular solution of the second order equation that results from separation of variables in Eq. (6):

$$\cosh^{-3}\frac{t'}{R}\frac{d}{dt'}\quad\cosh^{3}\frac{t'}{R}\quad\frac{dF_N}{dt'}+\frac{N(N+2)}{R^2\cosh^2 t'/R}F_N$$

$$+\frac{2}{R^2}F_N=0, \tag{10a}$$

$$\phi=F_N(t')Y^{(3)}_{Nlm}(\theta,\phi,\chi), \tag{10b}$$

$$N=\text{integer}\in(0,\infty), \tag{10c}$$

$$l=\text{integer}\in(0,N), \tag{10d}$$

$$m=\text{integer}\in(-l,+l), \tag{10e}$$

$$Y^{(3)}_{Nlm}\,|\,\alpha\,|\,(\sin\chi)^l C^{l+1}_{N-l}(\cos\chi)Y^{(2)}_{lm}(\theta,\varphi), \tag{10f}$$

where $Y^{(2)}_{lm}$ is the usual two-dimensional spherical harmonic and $c^{l+1}_{N-l}(\cos\chi)$ is a Gegenbauer polynomial. φ can be expanded in terms of the above modes as:

$$\phi=\sum_{N=0}^{\infty}\sum_{l=0}^{N}\sum_{m=-l}^{l}(a_{Nlm}F_N(t')Y^{(3)}_{Nlm}+\text{h.c.}), \tag{11}$$

where F_N is a particular combination of the two linearly independent solutions to Eq. (10). A vacuum state, $|0_-\rangle$, will be defined when a particular F_N has been chosen, by the condition $a_{Nlm}|0_-\rangle=0$.

We are now faced with the typical problem first analyzed by Parker,[9] in the context of particle creation in time-dependent cosmological solutions, that no clear cut prescription exists for choosing a $F_N(t')$. Following the lead of Parker *et al.*,[9] one might impose the condition that F_N diagonalize the Hamiltonian at some particular time, say $t'=0$. Thus F_N would look "positive frequency" at $t'=0$: $\dot{F}_N|_{t'=0}=-i\omega_N F_N|_{t'=0}$, where if

$$\omega_N=\left(\frac{N(N+2)}{R^2}+\frac{2}{R^2}\right)^{1/2}$$

the Hamiltonian will be diagonal. At any later time, t', F_N will be a mixture of the positive and negative modes at that time, and hence the vacuum state at $t'=0$ will be a many-particle state relative to the vacuum defined by Hamiltonian diagonalization at time t'. The heart of the calculation in this paper is to show that an observer at late times never sees these "particles" (essentially due to redshifting at the horizon), and therefore it does not matter which of the vacuum states defined by instantaneous diagonalization at various times is chosen to be the initial vacuum for the field, φ. The method will consist of showing that an observer at late times will see a thermal spectrum if (for convenience) the field is chosen to be in a vacuum state defined by instantaneous diagonalization at time $t'=0$. It will then be shown that a Bogoluibov transformation to a positive frequency mode at another time, t', will not affect the thermal spectrum observed at late times.

We now show that an observer using the modes (7) detects a thermal spectrum of radiation at late times if the field is in the vacuum state defined by instantaneous diagonalization at $t'=0$. Since the modes (9) are complete, one can expand $\varphi^{\text{static}}_\omega$ as

$$\phi^{\text{static}}_\omega=\sum_{Nlm}(\alpha_{\omega Nlm}F_N Y^{(3)}_{Nlm}+\beta_{\omega Nlm}\bar{F}_N\bar{Y}^{(3)}_{Nlm}) \tag{12}$$

so that

$$\alpha_{\omega Nlm}=\langle F_N Y^{(3)}_{Nlm},\varphi^{\text{static}}_\omega\rangle \tag{13}$$

and

$$\beta_{\omega Nlm}=\langle\bar{F}_N\bar{Y}^{(3)}_{NlM},\phi^{\text{static}}_\omega\rangle=\alpha_{-\omega Nlm}, \tag{14}$$

where $\langle\ ,\ \rangle$ represents the usual Klein—Gordon inner product. To derive the thermal spectrum, it is only necessary to show that $|\alpha_{\omega Nlm}|^2=\exp(2\pi\omega R)|\beta\omega Nl\boldsymbol{m}|^2$ for wavepackets at late times that are peaked about a frequency ω, since the thermal spectrum is then an immediate consequence by now familiar manipulations.[10] We have

$$\phi^{\text{static}}_\omega\sim\frac{P_\omega}{r}\exp(-i\omega u)Y_{lm}(\theta,\varphi), \tag{15a}$$

where the retarded time u has been introduced:

$$u=t-\frac{R}{2}\log\left[\frac{1-r/R}{1+r/R}\right]; \tag{15b}$$

and thus from Eqs. (5a) and (5b) we obtain u and r re x' and t':

$$u=-\frac{R}{2}\log\left[\frac{\cos\chi-\tanh t'/R}{\cos\chi+\tanh t'/R}\right]\left[\frac{1-\sin\chi\cosh t'/R}{1+\sin\chi\cosh t'/R}\right], \tag{16a}$$

$$r=R\sin\chi\cosh t'/R. \tag{16b}$$

The inner products of Eqs. (13) and (14) will be taken on the $t'=0$ spacelike Cauchy surface, and since late times are of interest one can expand $\varphi^{\text{static}}_\omega$ using the above, about $t'=0$ and about $\chi=\pi/2$ (where the horizon intersects the $t'=0$ surface):

$$\phi^{\text{static}}_\omega\sim\frac{P_\omega}{R}\exp\left\{+\frac{i\omega R}{2}\log\left[\frac{\pi/2-\chi-t'/R}{\pi/2-\chi+t'/R}\right]\left(\frac{\pi/2-\chi}{2}\right)^2\right\} \tag{17a}$$

so that

$$\dot{\phi}^{\text{static}}_\omega\Big|_{t'=0}=\frac{i\omega R}{\chi-\pi/a}\ \phi^{\text{static}}_\omega\Big|_{t'=0}. \tag{17b}$$

We now form the inner product (13). Using Eq. (10f), the relation $\dot{F}|_{t'=0}=-i\omega' F|_{t'=0}$ and retaining only the nontrivial t',χ dependence we have (over-all real or pure phase factors are not essential to this calculation)

$$\alpha_{\omega Nlm}=\frac{1}{2i}\int[\bar{F}_N\bar{Y}_{Nlm}\dot{\phi}^{\text{static}}_\omega$$
$$-\phi^{\text{static}}_\omega\bar{Y}_{Nlm}\dot{\bar{F}}_{Nlm}]\sin^2\chi\sin\theta\, d\theta\, d\varphi\, d\chi$$
$$\sim\int_0^{\pi/2}\left[\sin\chi^{l+2}C^{l+1}_{N-l}(\cos\chi)\left\{\frac{-i\omega R}{\pi/2-\chi'}(\pi/2-\chi')^{i\omega R}\right.\right.$$
$$\left.+i\omega'(\pi/2-\chi')^{i\omega R}\right\}d\chi. \tag{18}$$

The $\sin\chi^{l+2}C^{l+1}_{N-l}(\cos\chi)$ can be expanded into a series of polynomials in $\exp(\pm i\chi)$ using the binomial theorem and the expansion[11]

$$C^{l+1}_{N-l}(\cos\chi)=\sum_{j=0}^{N-l}A_{Njl}\cos(N-l-2j)\chi \tag{19}$$

(the form of the A_{Njl} will not be needed here), so that

$$\sin\chi^{l+2}C^{l+1}_{N-l}(\cos\chi)=\sum_{j>0}[a_j\exp(ij\chi)+b_j\exp(-ij\chi)], \tag{20}$$

where for given N and l the maximum value of j is $N+2$. The values of a_j and b_j will not be needed, only the observation that with the above expansion the equation for $\alpha_{\omega Nlm}$, (18), has the form of a finite sum of

Fourier transforms. The typical transform, call it T, looks like

$$T = \int_0^{\pi/2} \exp(ij\lambda)(\pi/2 - \lambda)^v \, d\lambda, \begin{cases} v = i\omega k \text{ or } i\omega R - 1, \\ j = \pm \text{ integer}, \end{cases} \tag{21}$$

The dependence of T on j, for very large j, will not be affected if the lower limit is extended to minus infinity. Extending the integral and introducing the integration variable, $s = \pi/2 - \chi$, leads to

$$T \sim \int_0^{\infty} \exp(\tilde{j}(\pi/2 - s))(s - \pi/2)^v ds, \quad \tilde{j} = ij, \tag{22}$$

which, by Ref. 11, can be integrated to

$$T = \Gamma(1 + v)(ij)^{-v-1}\exp(+ij\pi/2). \tag{23}$$

Using the above, $\alpha_{\omega N l m}$ can be written

$$\alpha_{\omega N l m} \sim \sum_{j>0} a_j(i\omega R)\exp(-ij\pi/2)(ij)^{-i\omega R}\Gamma(i\omega R)[1 + \omega'/j] + \sum_{j>0} b_j(i\omega R)\exp(ij\pi/2)(-ij)^{-i\omega R}\Gamma(i\omega R)[1 - \omega'/j]. \tag{24}$$

For wavepackets peaked at late retarded times $u = 2\pi n/\epsilon$ (n = large positive integer, ϵ = small positive number) and at frequency $k\epsilon$ (k = positive integer) we should not be considering $\alpha_{\omega N l m}$ but rather $\alpha_{kn, N l m}$, where

$$\alpha_{kn, N l m} = \frac{1}{\sqrt{\epsilon}} \int_{k\epsilon}^{(k+1)\epsilon} \exp(2\pi i n\omega/l)\alpha_{\omega N l m} d\omega. \tag{25}$$

This results from using the peaking factor $\exp(2\pi i n\omega/\epsilon)$ to sum the continuum modes $\varphi_\omega^{\text{static}}$ [Eq. (12)] into wavepackets, in exactly the same fashion used in Hawking's original paper (c.f. Hawking, Ref. 10). After performing the integration one sees that only exponentially large positive j [and hence N, see Eq. (19)] contribute to $\alpha_{kn, N l m}$ in analogy to the exponentially large frequency contribution of Hawking's calculation. Thus for packets at late times $\alpha_{\omega N l m}$ is dominated by a term like

$$\alpha_{\omega N l m} \sim \alpha_j(i\omega R) \exp(-ji\pi/2)(ij)^{-i\omega R}\Gamma(i\omega R)[1 + \omega'/j], \tag{26}$$

where j is exponentially large and positive. The terms with b_j as a factor cannot contribute at late times unless j is exponentially large and negative, which can never be the case in the summation in Eq. (24). By Eq. (14) $\bar{\beta}_{\omega N l m} = \alpha_{-\omega N l m}$, and, using (26), we see that

$$|\beta_{\omega N l m}|^2 = \exp(-2\pi\omega R)|\alpha_{\omega N l m}|^2. \tag{27}$$

One must choose $(ij)^{-i\omega R} = \exp(+\pi\omega R/2)j^{-i\omega R}$ in the above, and not $\exp(-3\pi\omega R/2)j^{-i\omega R}$, because expression (26) results from Laplace transforming a function of s that is zero for large negative values of s. This implies the transform is analytic in the lower half j plane so that the branch cut in the function $(z)^{-i\omega R}$ should be taken just above the negative real z axis. Hence $(ij)^{-i\omega R}$ is determined by a counterclockwise rotation so that $(ij)^{-i\omega R} = (j\exp(i\pi/2))^{-i\omega R}$, not $(j\exp(-3\pi i/2))^{-i\omega R}$. Thus we obtain expression (27). This is the typical thermal result, indicating that at late times an observer who detects particles that are positive frequency with respect to the Killing time, l, will see a thermal spectrum of radiation at temperature $T = \kappa/2\pi = 1/2\pi R$, if the state of the field is chosen to be the vacuum state defined by instantaneous diagonalization on the Cauchy surface $l' = 0$.

We now show that it was not necessary to choose the F_N in Eq. (11) to diagonalize the Hamiltonian at $l' = 0$. The general solution $F'_N(l)$ to Eq. (10a) can be written as a linear combination of the mode diagonalizing the Hamiltonian $l' = 0$ and its complex conjugate:

$$F'_N = A_N F_N + B_N \bar{F}_N, \tag{28}$$

where A_N and B_N are constants. One can expand φ in modes with the new time dependence

$$\varphi = \sum_{N l m} a'_{N l m} F'_N Y_{N l m} + (\text{h.c.}), \tag{29}$$

and define a new vacuum, $|0'\rangle$ state by

$$a'_{N l m}|0'\rangle = 0. \tag{30}$$

Combining (28), (29), and (11) leads to

$$a'_{N l m} = \bar{A}_N a_{N l m} - (-1)^m \bar{B}_N a^*_{N l m}, \tag{31a}$$

$$a'^*_{N l m} = A_N a^*_{N l m} - (-1)^m B_N a_{N l m}, \tag{31b}$$

so upon evaluating the expectation value of the number operator $a'^*_{N l m} a'_{N l m}$ in the old vacuum state, one sees that the old vacuum state contains $|B_N|^2$ particles per mode relative to the new vacuum state.

To show that a thermal spectrum is still observed when the vacuum state is defined relative to the mode expansion (29) by Eq. (30), we expand $\varphi_\omega^{\text{static}}$ in analogy to Eq. (12):

$$\varphi_\omega^{\text{static}} = \sum_{N l m} \alpha'_{\omega N l m} F'_N Y_{N l m} + \beta'_{\omega N l m} \bar{F}'_N \bar{Y}_{N l m}. \tag{32}$$

If the initial vacuum state is defined by (30), then an observer at late times will detect a thermal spectrum if

$$|\beta'_{\omega N l m}|^2 = \exp(-2\pi\omega R)|\alpha'_{\omega N l m}|^2. \tag{33}$$

Substituting (28) in (32) results in

$$\beta'_{\omega N l m} = (-1)^m B_N \alpha_{\omega N l - m} - A_N \beta_{\omega N l m}, \tag{34a}$$

$$\alpha'_{\omega N l m} = A_N \alpha_{\omega N l - m} - \bar{B}_N \beta_{\omega N l m}(-1)^m. \tag{34b}$$

If B_N tends to zero for the exponentially large N that predominate in packets peaked at late retarded times [see the remark preceeding Eq. (20)], then Eqs. (34a), (34b) immediately yield (33), given the previously proved relation (27).

It therefore remains to show that B_N tends to zero at large N. Clearly some type of restriction on F'_N is necessary to prove this. Imposing the reasonable requirement that F'_N be "positive frequency" at at least some time, say l'_1, and also diagonalize the Hamiltonian at that time implies

$$\dot{F}'_N\big|_{t'_1} = -i\omega'_N F'_N\big|_{t'_1}, \tag{35}$$

or, using (28),

$$\dot{F}'_N\big|_{t'_1} = [A_N \dot{F}_N + B_N \dot{\bar{F}}_N]\big|_{t'_1}, \qquad = -i\omega_N[A_N F_N + B_N \bar{F}_N]\big|_{t'_1} \tag{36}$$

where if $\omega'_N = i\sqrt{N(N+2)}/R\cosh l'/R$, cross terms are eliminated in the Hamiltonian. Since extremely high N are of interest, we shall use the limit $\omega_N \to N/R$. Equation (36) implies

$$F'_N = A_N \left[F_N - \frac{\dot{F}_N(t'_1) + [iN/R\cosh(t'/R)]\,F_N(t'_1)}{\dot{\bar{F}}_N(t'_1) + [iN/R\cosh(t'/R)]\,\bar{F}_N(t'_1)}\,\bar{F}_N \right], \quad (37)$$

which combined with the condition

$$-2i = F'_N \dot{\bar{F}}'_N - \dot{F}'_N \bar{F}'_N \quad (38)$$

yields

$$|B_N|^2 = \frac{|K_N|^2}{1 - |K_N|^2} \cdot \begin{bmatrix} N \text{ independent} \\ \text{factor} \end{bmatrix}, \quad (39)$$

where

$$|K_N|^2 = \left| \frac{\dot{F}_N(t'_1) + [iN/R\cosh(t'/R)]\,F_N(t'_1)}{\dot{\bar{F}}_N(t'_1) + [iN/R\cosh(t'/R)]\,\bar{F}_N(t'_1)} \right|^2 . \quad (40)$$

Equation (38) results from imposing the usual commutation relations on the a'_{Nlm}, a'^{+}_{Nlm} of Eq. (29) at time $t' = t'_1$.

It only remains to check that $|K_N|^2 \to 0$ for high N, which implies $B_N \to 0$, and thus the desired result, (33), follows. The normalized modes, F_N, which are "positive frequency" and diagonalize the Hamiltonian at $t' = 0$ are solutions of the time dependent equation (10a). It is not difficult to verify that for large N

$$F_N = \frac{1}{\sqrt{N/R}} \cos(\tau/R)\exp(-i(N/R)\tau), \quad \tan\tau/R = \sinh t'/R \quad (41)$$

is just such a mode. Substituting (41) in (40) yields

$$\lim_{N\to\infty} |K_N|^2 = \lim_{N\to\infty} \Bigg|$$

$$= \lim_{N\to\infty} \left| \frac{-\sin\tau_1/\sqrt{NR}\,\cosh(t'_1/R)}{-\sin\tau_1/\sqrt{NR} + 2i\sqrt{N/R}\cos\tau_1/\cosh(t'_1/R)} \times \exp(-2iN\tau_1/R) \right|^2 \to 0. \quad (42)$$

Hence we see that an observer at late times sees a thermal spectrum of radiation if the initial vacuum state is *any* one of the possible "vacuum" states defined by instantaneous diagonalization on an arbitrary spacelike hypersurface. Each state is a "many particle" state relative to another; however, an observer at late times never detects these "particles."

IV. CONCLUSIONS

Using de Sitter space—time as an example (for simplicity), we have recovered the thermal spectrum of radiation seen by an observer at late times, which had previously only been calculated using propagator techniques. The "mode mixing" calculation used here differs from that of other authors by being explicitly observer dependent and not de Sitter invariant. It was shown that, although the usual problem of defining a vacuum in time-dependent situations was present, it did not affect the observations at late times of an observer following the trajectory of the Killing vector ∂_t. Any one of the infinite number of "vacuum" states defined by instantaneous diagonalization on some spacelike Cauchy surface is a permissible initial state leading to observations of a thermal spectrum at late times. In fact, the situation is slightly more general. We have seen that it is the high N behavior of B_N which determines whether a thermal spectrum will result, if the initial "vacuum" state, $|0'\rangle$, is defined via Eqs. (28), (29), and (30). Any state, $|0'\rangle$, even if it contains an infinite number of "particles," will lead to a thermal spectrum so long as the spectrum of "particles," $|B_N|^2$, falls off rapidly enough at high N.

ACKNOWLEDGMENTS

I would like to thank S.W. Hawking for suggesting this problem and helpful discussions during the course of its solution.

[1] G. W. Gibbons and S.W. Hawking, Phys. Rev. D **15**, 2738 (1977).
[2] A. S. Lapedes, Comm. Math. Phys. 51, 121 (1976).
[3] O. Nachtmann, Comm. Math. Phys. 6, 1 (1967).
[4] N. Chernikikov *et al.*, Ann. Inst. Henri Poincaré, 60(2), 109 (1968).
[5] M. Gutzwiller, Helv. Phys. Acta 29, 313 (1956).
[6] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U. P., Cambridge, 1973).
[7] W. Unruh, Phys. Rev. D **14**, 870 (1976).
[8] Cf. Ref. 7.
[9] L. Parker and S. A. Fulling, Phys. Rev. D **9**, 341 (1974).
[10] S. W. Hawking, Comm. Math. Phys. 43, 199 (1975).
[11] A. Erdelyi, *Bateman Manuscript Project* (McGraw-Hill, New York, (953).

Geometrical spacetime perturbation theory: Regular first-order structures[a)]

Robert H. Gowdy

Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742
(Received 5 April 1978)

Spacetime perturbation theory is formulated in a coordinate independent way by regarding a family of spacetimes as a $(4+n)$-dimensional manifold with a particular standard connection and deriving analogs of the Gauss–Weingarten equations to describe the imbedding of each spacetime in the family.

I. INTRODUCTION

There is an increasing effort to use spacetime perturbation theory to calculate the gravitational radiation from realistic sources.[1] Furthermore, spacetime perturbation theory has always been used to analyze the stability of solutions to Einstein's equations[2] and plays a central role in some theories of galaxy formation.[3] Thus, it is important to have an efficient way to understand and organize perturbation calculations in general relativity. This paper provides a finite-dimensional geometrical framework for these calculations, a framework with several important advantages:

(1) There is no need to single out a "background spacetime metric".

(2) All quantities are tensorial.

(3) Variations of tensor fields are accomplished by means of an operation that commutes with the spacetime metric so that one can raise or lower tensor indexes either before or after a variation.

(4) Gauge conditions are separated from coordinate conditions.

(5) Multiparameter perturbations are easily accomodated.

The proposed approach is just the natural differential geometry of a $(4+n)$-dimensional manifold which is locally an n-parameter family of spacetimes. I call such a manifold a "spacetime $4+n$ deformation". Section II describes the local structures on such a deformation. These structures are essentially those introduced by Geroch in his earlier work on spacetime limits.[4] The new idea in this paper is the use of a particular "standard" deformation connection, which is presented in Sec. III. In this approach to spacetime perturbation theory, the role of the spacetime metric tensor variation is played by the second fundamental tensor of spacetime. This second fundamental tensor describes the imbedding of a spacetime in a given deformation. It is related to the curvature of the deformation connection by a set of Gauss—Codazzi—Mainardi equations which are derived in Sec. IV. Section V sketches the application of this geometrical framework to regular perturbations of gravity and fields coupled to gravity.

My sign conventions and notation are taken primarily from the text by Misner, Thorne, and Wheeler.[5] Some conventions peculiar to this paper are: A *tensor field* on a manifold $\mathcal{M}$ means a C^∞ cross section of a tensor bundle over $\mathcal{M}$. A *surface* Σ in $\mathcal{M}$ means a C^∞ immersion $\Sigma: M \to \mathcal{M}$ where the dimension of M is smaller than that of $\mathcal{M}$. The tangent space to $\mathcal{M}$ at the point $P \in \mathcal{M}$ is denoted by $T_P(\mathcal{M})$ or just T_P and the corresponding cotangent space is denoted by T_P^*. The derivatives of a map $\phi: A \to B$ which connects two manifolds A and B are defined as usual[6] and denoted by $\phi_*: T_a \to T_b$, $\phi^*; T_b^* \to T_a^*$, where $b = \phi(a)$.

The pointwise action of derivations and second-rank tensor fields on form and vector fields will usually be written operator style without explicit parentheses or evaluation points. The convention implicit in such expressions is that each operator acts on everything to its right. Occasionally, extra parentheses will be used to turn off this rule as in the expression $\nabla_U HV = (\nabla_U H)V + H\nabla_U V$.

In the component expressions, Greek indexes range from zero to three, lower case Latin indexes from zero to $n+3$, and upper case Latin indexes in parentheses range from one to n. Upper and lower case Latin indexes are related to each other by the rule: $(A) = a$ corresponds to $a = A + 3$ as in $V^{(2)} = V^5$. The summation convention is used separately on each index type.

II. LOCAL DEFORMATION STRUCTURE

A. Deformation without gauge

A $4+n$ *deformation* is a $(4+n)$-dimensional manifold $\mathcal{M}$ with a *deformation tensor field* γ and a set of linearly independent *deformation form fields* $\theta^{(A)}$ such that:

(1) γ assigns the linear map $\gamma_P: T_P^* \to T_P$ to each $P \in \mathcal{M}$.

(2) For any 1-forms α, β, $\langle\alpha, \gamma\beta\rangle = \langle\beta, \gamma\alpha\rangle$. (2.1)

(3) For any $P \in \mathcal{M}$ there exists a surface Σ through P such that $\Sigma_* T_p = \gamma T_P^*$, where $P = \Sigma(p)$. (2.2)

(4) The maximum dimension of γT_P^* is four.

(5) For all (A),

$$\gamma\theta^{(A)} = 0. \tag{2.3}$$

(6) The deformation forms are closed:

[a)]Supported in part by the National Science Foundation: Grant PHY 76-20029 and in part by the Alfred P. Sloan Foundation.

 0022-2488/78/1911-2294$1.00

$$d\theta^{(A)}=0. \tag{2.4}$$

The surface Σ is the *integral surface* of γ through the point P. Note that the deformation forms are only required to be linearly independent as fields and may be linearly dependent at a particular evaluation point. This freedom in the definition will become important in later papers which discuss singular perturbations.

From these definitions, straightforward computation shows that

$$\Sigma^*\theta^{(A)}=0 \tag{2.5}$$

and, if $\mathcal{U}$ is a contractible open set in $\mathcal{M}$, then there exist functions $\lambda^{(A)}:\mathcal{U}\to R$ such that

$$\theta^{(A)}\Big|_{\mathcal{U}}=d\lambda^{(A)}. \tag{2.6}$$

The integral surfaces of γ in the set $\mathcal{U}$ are then the level surfaces of these functions. The functions $\lambda^{(A)}$ will be called *perturbation parameters*.

Another direct consequence of the definitions is that there exists a unique symmetric tensor field g which assigns to each $P\in\mathcal{M}$ the linear map $g_P:\gamma T_P^*\otimes\gamma T_P^*\to R$, which is defined for any vectors A and B in γT_P^* by

$$g_P(A, B):=\langle\gamma_P^{-1}A, B\rangle, \tag{2.7}$$

where

$$\gamma_P^{-1}A:=\{\alpha\,|A=\gamma\alpha\}. \tag{2.8}$$

If the tensor g_P has Lorentz signature $(-+++)$, then P is a *spacetime point* of the deformation $(\mathcal{M},\gamma,\theta)$. If Σ is an integral surface of γ and every point of $\Sigma(M)$ is a spacetime point, then a spacetime metric is induced on M and the surface Σ is called a *spacetime* in the deformation $(\mathcal{M},\gamma,\theta)$.

An important feature of the definitions given here is that they permit points where the dimension of γT_P^* is less than four. Such points will be called *critical points* and will play a central role in later discussions of singular perturbation theory.

B. Gauge and coordinates

In spacetime perturbation theory, the word "gauge" is used to mean a set of supplementary conditions that determine the response of spacetime coordinates to spacetime metric tensor variations. The corresponding coordinate-free structure on a spacetime $4+n$ deformation must supply a way to match up points on neighboring spacetimes. Equivalently, it must supply a way to project a neighborhood of a given spacetime onto that spacetime. The infinitesimal version of such a structure is a projection tensor field. A *gauge* on a $4+n$ deformation $(\mathcal{M},\gamma,\theta)$ is defined to be a tensor field H which assigns to each $P\in\mathcal{M}$ a linear map $H_P: T_P\to\gamma T_P^*$ which is onto and satisfies

$$H^2=H, \tag{2.9}$$

$$H\gamma=\gamma. \tag{2.10}$$

A related tensor is the *identification gauge*

$$\iota:=1-H, \tag{2.11}$$

which satisfies

$$\iota^2=\iota, \tag{2.12}$$

$$\iota\gamma=0. \tag{2.13}$$

A surface $\Xi:N\to\mathcal{M}$ such that

$$\Xi_*T_p=\iota T_P,\ \text{where}\quad P=\Xi(p) \tag{2.14}$$

is an *identification surface* through the point P. A gauge that admits such a surface is *integrable* at P.

The action of H and ι on a 1-form field will be defined by

$$\langle H\alpha, V\rangle:=\langle\alpha, HV\rangle \tag{2.15}$$

for any vector $V\in T_P$. From this definition and Eqs. (2.10) and (2.13) it follows that

$$\gamma H=\gamma, \tag{2.16}$$

$$\gamma\iota=0, \tag{2.17}$$

$$H\theta^{(A)}=0, \tag{2.18}$$

$$\iota\theta^{(A)}=\theta^{(A)}. \tag{2.19}$$

With the gauge concept introduced in a coordinate-independent way, it is now safe to retreat from abstract notation and introduce coordinates. A set of coordinate functions $\{x^a\}$ on an open set $\mathcal{U}$ of spacetime points in a deformation is an *aligned chart* with respect to a gauge H if

$$dx^{(A)}=\theta^{(A)}\big|_{\mathcal{U}} \tag{2.20}$$

and

$$H\left(\frac{\partial}{\partial x^{(A)}}\right)=0. \tag{2.21}$$

From Eq. (2.6), the coordinates $x^{(A)}$ are just the perturbation parameters $\lambda^{(A)}$ that label spacetimes. The level surfaces of the coordinates x^μ are just the identification surfaces. Thus, the existence of a chart aligned with H implies that the gauge H is integrable. In such an aligned chart, the only nonzero components of the gauge tensors are

$$H^\mu{}_\nu=\delta^\mu_\nu \tag{2.22}$$

and

$$\iota^{(A)}{}_{(B)}=\delta^{(A)}_{(B)}. \tag{2.23}$$

III. DEFORMATION CONNECTION

A. Definition and some restrictions

A connection ∇ on a manifold $\mathcal{M}$ is defined to be a tensorial assigment of vector-field derivations to vector fields.[7] The derivation assigned to the vector field V is ∇_V. As usual, ∇ is torsion-free if, for any vector fields A and B,

$$\nabla_A B - \nabla_B A = [A, B]. \tag{3.1}$$

Also as usual, the action of ∇ on a covariant tensor field T is defined by

$$(\nabla_V T)(A, B, \cdots) := \nabla_V T(A, B, \cdots) - T(\nabla_V A, B, \cdots) - T(A, \nabla_V B, \cdots) - \cdots. \tag{3.2}$$

A similar definition then extends ∇ to contravariant and mixed tensor fields.

A connection ∇ on a $4+n$ deformation $(\mathcal{M}, \gamma, \theta)$ is *deformation compatible* if, for any vector field V,

$$\nabla_V \theta^{(A)} = 0 \tag{3.3}$$

and

$$(\nabla_V \gamma) = 0. \tag{3.4}$$

Equation (3.2) may be used to produce an alternative form of Eq. (3.4):

$$[\nabla_V, \gamma] = 0. \tag{3.5}$$

Because $HB \in \gamma T_P^*$, Eq. (3.5) yields the following useful result:

Propositition 1: Let $(\mathcal{M}, \gamma, \theta)$ be a $4+n$ deformation with gauge H and deformation-compatible connection ∇. For any vector fields B and V and any form field α,

$$\iota \nabla_V HB = 0 \tag{3.6}$$

and

$$H \nabla_V \iota\alpha = 0. \tag{3.7}$$

A surface Ξ: $N \to \mathcal{M}$ is *flat* in the connection V if there is a frame field e on N such that the vectors $\Xi_* e_j$ satisfy

$$\nabla_{\Xi_* e_k} \Xi_* e_j = 0.$$

A deformation connection ∇ will be called *gauge flat* with respect to a gauge H if H is integrable and its identification surfaces are flat in ∇.

Proposition 2: If a $4+n$ deformation $(\mathcal{M}, \gamma, \theta)$ has a connection ∇ which is gauge-flat with respect to a gauge H and $\{x^a\}$ is a chart aligned with H, then

$$\nabla_{\partial/\partial x^{(C)}} (\partial/\partial x^{(A)}) = 0. \tag{3.8}$$

Proof: Use Eq. (2.20) and then expand the vectors $\partial/\partial x^{(A)}$ in the covariantly constant basis. See the Appendix for details.

B. Second fundamental tensor

The gauge H and the deformation connection ∇ may be used to define a *second fundamental tensor* h. For any vector fields U, V define

$$h_U V := H[\nabla_{HV}, H]U. \tag{3.9a}$$

Because of Proposition 1, this definition is equivalent to

$$h_U V = [\nabla_{HV}, H]U. \tag{3.9b}$$

Equation (3.9a) is just the definition that is used in Riemannian geometries, stated in terms of vector fields and tangent projections.[8] Equation (3.9b) does not have a Riemannian counterpart and arises from the fact that ∇ is compatible with the degenerate tensor γ instead of a Riemannian metric. Because its definition is the usual one, h retains several of its familiar properties.[8]

Proposition 3: The second fundamental tensor h of a $4+n$ deformation $(\mathcal{M}, \gamma, \theta)$ with gauge H and deformation-compatible connection ∇ has the properties

$$Hh_U = h_U H = h_U, \tag{3.10}$$

$$h_{\iota U} = h_U, \tag{3.11}$$

$$\langle \alpha, h_U \gamma \beta \rangle + \langle \beta, h_U \gamma \alpha \rangle = (\pounds_{\iota U} \gamma)(H\alpha, H\beta) \tag{3.12}$$

for any vector field U and form fields α and β.

Proof: Equations (3.10) and (3.11) are easy consequences of Proposition 1. The detailed verification of Eq. (3.12) is given in the Appendix.

A familiar property of h that does not survive automatically in a deformation geometry is its symmetry. This property must be imposed as an additional condition on the connection. A deformation connection ∇ will be called *gauge-normal* with respect to a gauge H if the corresponding second fundamental tensor satisfies

$$\langle \alpha, h_U \gamma \beta \rangle = \langle \beta, h_U \gamma \alpha \rangle \tag{3.13}$$

for any 1-forms α and β. If ∇ is both gauge-normal and deformation-compatible, then Eqs. (3.12) and (3.13) yield a useful expression for h: For any 1-form fields α and β, and vector field U,

$$\langle \alpha, h_U \gamma \beta \rangle = \tfrac{1}{2} (\pounds_{\iota U} \gamma)(H\alpha, H\beta). \tag{3.14}$$

The Lie derivative $\pounds_{\iota U}\gamma$ is essentially the usual variational derivative of the inverse spacetime metric tensor.[9] Thus the second fundamental tensor h_U is the familiar first-order spacetime metric variation except for a factor of $-\frac{1}{2}$.

Equation (3.14), together with the other results of Proposition 3, determine the behavior of the second fundamental tensor under a gauge transformation.

Proposition 4: Let ∇ and ∇' be torsion-free, deformation-compatible connections. Let ∇ be gauge-normal with respect to a gauge H and let ∇' be gauge-normal with respect to a gauge H'. The corresponding second fundamental forms h and h' are then related by

$$a := H - H',$$

$$\langle \alpha, (h'_U - h_U)\gamma\beta\rangle = -\tfrac{1}{2}\{\langle \alpha\, \nabla_{\gamma\beta} a)U\rangle + \langle \beta, (\nabla_{\gamma\alpha} a)U\rangle\} \quad (3.15)$$

for any 1-forms α and β and any vector U.

Proof: Write the Lie derivative in Eq. (3.14) in terms of the connection ∇, note that $aU \in \gamma T^*_p$ so that $HaU = aU$, and recall that both H and H' reduce to the identity on γT^*_p so that $aH = 0$. Equation (3.15) then follows by a straightforward calculation which is given in the Appendix.

C. Standard deformation connection

A connection on a $4+n$ deformation will be called *standard* with respect to an integrable gauge H if it is

(1) torsion-free,

(2) deformation-compatible,

(3) gauge-normal with respect to H,

(4) gauge-flat with respect to H.

The virtue of this standard connection is that it is simple, exists where it is needed, and is unique where it exists. Thus, even if one prefers some other connection, the standard connection is a reasonable starting point.

Theorem 1: Let Σ be a spacetime in a $4+n$ deformation $(\mathcal{M}, \gamma, \theta)$ and let H be an integrable gauge. There is then a neighborhood U of Σ within which the corresponding standard connection exists and is unique. This connection acts on vector fields A and B according to

$$\nabla_A B = \Box_{HA} HB + \flat_{\iota A}\iota B + H[\iota A, HB] - h_A B - h_B A, \quad (3.16)$$

where $\Box = H\nabla H$ is the torsion-free, metric-compatible connection induced on the integral surfaces of Σ, and $\flat = \iota\nabla\iota$ is the flat, torsion-free connection induced on the identification surfaces of H. In an aligned chart $\{x^a\}$ with $e_a := \partial/\partial x^a$, the connection coefficients

$$\Gamma^a{}_{bc} := \left\langle dx^a, \nabla_{e_c} e_b \right\rangle \quad (3.17)$$

are given by

$$\Gamma^{(A)}{}_{bc} = 0, \quad \Gamma^a{}_{(B)(C)} = 0, \quad (3.18)$$

$$\Gamma^\alpha{}_{\beta(C)} = \Gamma^\alpha{}_{(C)\beta} = -h^\alpha{}_{\beta(C)} := -\langle dx^\alpha, h_{e_{(C)}} e_\beta\rangle, \quad (3.19)$$

$$\Gamma^\alpha{}_{\beta\gamma} = \tfrac{1}{2}\gamma^{\alpha\rho}(g_{\rho\beta,\gamma} + g_{\rho\gamma,\beta} - g_{\beta\gamma,\rho}). \quad (3.20)$$

For this connection, the second fundamental form satisfies

$$[\nabla_V, H]U = (\nabla_V H)U = h_U V, \quad (3.21)$$

$$(\nabla_{\iota A} h)_{\iota B} = (\nabla_{\iota B} h)_{\iota A} \quad (3.22)$$

for all vector fields U, V, A, B.

Proof: Equation (3.16) is obtained by using $1 = H + \iota$ to decompose the vector fields A and B. The neighborhood U consists of aligned charts built on an atlas of charts on Σ. In each aligned chart, Eq. (3.16), Proposition 2, and the standard connection requirements yield Eqs. (3.18)–(3.20) uniquely. Equations (3.21), (3.22) then follow by direct computation. The details of the proof are given in the appendix.

IV. DEFORMATION CURVATURE

A. Definition and identities

The curvature $\mathcal{R}$ of a connection ∇ is defined as usual: For any vector fields A, B, C

$$\mathcal{R}(A, B)C := ([\nabla_A, \nabla_B] - \nabla_{[A,B]})C. \quad (4.1)$$

The symmetries

$$\mathcal{R}(A,B) = -\mathcal{R}(B,A), \quad (4.2)$$

$$\mathcal{R}(A,B)C + \mathcal{R}(B,C)A + \mathcal{R}(C,A)B = 0 \quad (4.3)$$

and the Bianchi identities

$$(\nabla_C\mathcal{R})(A,B) + (\nabla_A\mathcal{R})(B,C) + (\nabla_B\mathcal{R})(C,A) = 0 \quad (4.4)$$

follow from this definition and the assumption that ∇ is torsion-free. The further assumption that ∇ is deformation-compatible yields the symmetry

$$\langle \alpha, \mathcal{R}\gamma\beta\rangle = -\langle \beta, \mathcal{R}\gamma\alpha\rangle \quad (4.5)$$

for any 1-forms α, β and the conditions

$$\langle \theta^{(A)}, \mathcal{R}\, C\rangle = 0 \quad (4.6)$$

for any vector field C.

If ∇ is a torsion-free connection on a $4+n$ deformation with a gauge H, then the curvature of ∇ is related to the second fundamental tensor by the Codazzi identity which may be stated in the form

$$[\mathcal{R}(HA, HB), H]C = (\nabla_{HA} h)_C HB - (\nabla_{HB} h)_C HA \quad (4.7)$$

for any vector fields A, B, C. The proof of this identity is well known.[8] However, for completeness and for the convenience of those who are not familiar with the notation used here, the details are included in the Appendix.

With a standard deformation connection, a full decomposition of the deformation curvature in terms of the second fundamental tensor and the spacetime curvature tensor becomes possible. The spacetime curvature ${}^s\mathcal{R}$ is defined by

$${}^s\mathcal{R}(HA, HB)HC = ([\Box_{HA}, \Box_{HB}] - \Box_{[HA, HB]})HC, \quad (4.8)$$

where

$$\Box_{HA} = H\nabla_{HA} H.$$

The identities which express this decomposition are analogous to the Gauss, Codazzi–Mainardi, and Ricci equations of Riemannian Gauss–Weingarten surface imbedding theory.[8] However, because of the degeneracy of deformation geometry, only the Codazzi–Mainardi equation [Eq. (4.7) above] takes its familiar Riemannian form.

Theorem 2: If ∇ is a standard deformation connection with respect to an integrable gauge H, then

$$\iota \mathcal{R}(A,B)C=0, \tag{4.9}$$

$$\mathcal{R}(\iota A,\iota B)\iota C=0, \tag{4.10}$$

$$\mathcal{R}(HA,HB)HC={}^{g}\mathcal{R}(HA,HB)HC, \tag{4.11}$$

$$\mathcal{R}(HA,HB)\iota C=(\nabla_{HB}h)_C HA-(\nabla_{HA}h)_C HB, \tag{4.12}$$

$$\mathcal{R}(HA,\iota B)\iota C=(\nabla_{\iota B}h)_C HA-h_C h_B HA, \tag{4.13}$$

$$\mathcal{R}(\iota A,\iota B)HC=[h_B,h_A]C, \tag{4.14}$$

$$\langle\sigma,\mathcal{R}(\gamma\alpha,\iota B)\gamma\beta\rangle=\langle\sigma,(\nabla_{\gamma\beta}h)_B\gamma\alpha\rangle-\langle\beta,(\nabla_{\gamma\sigma}h)_B\gamma\alpha\rangle \tag{4.15}$$

for any vector fields A, B, C and form fields α,β,σ.

Proof: Equations (4.9)–(4.11) follow directly from the standard connection requirements. Equation (4.12) is an alternative version of the Codazzi identity [Eq. (4.7). The remaining results are derived just as in ordinary Riemannian geometry, with the use of Eq. (3.22) to simplify Eq. (4.14). The details of the proof are given in the Appendix.

B. Index notation

The abstract notation that has been used up to this point has the advantage of keeping the geometrical concepts front and center. However, it does not handle tensor contractions and higher-order covariant derivatives as well as the traditional index notation. Let e_a be a frame-field with θ^a its dual coframe field and define the components

$$R^a{}_{bcd}:=\langle\theta^a,R(e_c,e_d)e_b\rangle,$$

$$H^a{}_b:=\langle\theta^a,He_b\rangle,$$

$$\iota^a{}_b:=\langle\theta^a,\iota e_b\rangle,$$

$$h^a{}_{bs}:=\langle\theta^a,h_{e_s}e_b\rangle,$$

$$h_s:=h^a{}_{as},$$

$$\gamma^{ab}:=\langle\theta^a,\gamma\theta^b\rangle,$$

$$H_{ab}:=\langle\gamma^{-1}He_a,He_b\rangle=g_{ab}.$$

In this notation, one raises indexes with γ^{ab}. The degeneracy of the geometry prevents one from lowering them at all. The standard deformation-covariant derivative is denoted by a dot as in

$$T^{ab\cdots}{}_{cd\ldots r\cdot k}:=(\nabla_{e_k}T)(\theta^a,\theta^b,\ldots,e_c,e_d,\ldots,e_r).$$

A shortcoming of the index notation is that projections of high-rank tensors produce bewildering arrays of indexes. This defect can be remedied by adopting the convention that an index which is followed by the symbol $\uparrow$ has been projected with $H^a{}_b$ while an index which is follwed by the symbol $\downarrow$ has been projected with ι. For example,

$$V_{a\uparrow}:=V_rH^r{}_a,\quad V_{a\downarrow}:=V_r\iota^r{}_a,\quad V^{a\uparrow}:H^a{}_rV^r.$$

I adopt the convention that a deformation-covariant derivative which is indicated by a dot is to be performed before any indicated projections. For example, the definition of the second fundamental tensor becomes.

$$h^a{}_{bc}:=H^{a\uparrow}{}_{c.b\uparrow}=H^a{}_rH^r{}_{c.s}H^s{}_b.$$

When a covariant derivation is indicated by a vertical bar, then all indicated projections are to be carried out both before and after the derivative. For example, the spacetime-covariant derivative of a tensor field $T^a{}_b$ is defined to be

$$T^a{}_{b;c}:=T^{a\uparrow}{}_{b\uparrow|c\uparrow}=H^a{}_mH^n{}_b(H^m{}_jH^k{}_nT^j{}_k)_{.d}H^d{}_c.$$

In terms of this index notation, Theorem 2 yields the following decomposition of the deformation curvature tensor:

$$R^{k\downarrow}{}_{cab}=0,\quad R^k{}_{c\downarrow a\downarrow b\downarrow}=0, \tag{4.16}$$

$$R^k{}_{c\uparrow a\uparrow b\uparrow}={}^{g}R^k{}_{cab}, \tag{4.17}$$

$$R^k{}_{c\downarrow a\uparrow b\uparrow}=h^k{}_{a\uparrow c\cdot b\uparrow}-h^k{}_{b\uparrow c\cdot a\uparrow}=h^k{}_{ac|b\uparrow}-h^k{}_{bc|a\uparrow}, \tag{4.18}$$

$$R^k{}_{c\downarrow a\uparrow b\downarrow}=h^k{}_{a\uparrow c\cdot b\downarrow}-h^k{}_{jc}h^j{}_{ab}=h^k{}_{ac|b\downarrow}-h^k{}_{jc}h^j{}_{ab}, \tag{4.19}$$

$$R^k{}_{c\uparrow a\downarrow b\downarrow}=h^k{}_{jb}h^j{}_{ca}-h^k{}_{ja}h^j{}_{cb}, \tag{4.20}$$

$$R^{kc}{}_{a\uparrow b\downarrow}=h^k{}_{ab}{}^{\cdot c}-h^c{}_{ab}{}^{\cdot k}=h^k{}_{ab}{}^{|c}-h^c{}_{ab}{}^{|k}. \tag{4.21}$$

Notice that the projection indicators are omitted from the twice-projected derivatives in Eqs. (4.18), (4.19), (4.21). This omission causes no ambiguity because one can use the projection identities [Eqs. (3.10), (3.11)] of Proposition 3 to show that only one combination of projections,

$$h^k{}_{ac|b}=h^{k\uparrow}{}_{a\uparrow c\downarrow|b},$$

is not zero.

The deformation Ricci tensor $R_{cb}:=R^a{}_{cab}$ has the decomposition

$$R_{c\uparrow b\uparrow}={}^{g}R_{cb}, \tag{4.22}$$

$$R_{c\downarrow b\uparrow}=h_{c\cdot b\uparrow}-h^a{}_{b\uparrow c\cdot a}=h_{c|b\uparrow}-h^a{}_{bc|a}, \tag{4.23}$$

$$R^c{}_{b\downarrow}=h_b{}^{\cdot c}-h^c{}_{ab}{}^{\cdot a}=h_b{}^{|c}-h^c{}_{ab}|_a, \tag{4.24}$$

$$R_{c\downarrow b\downarrow}=h_{c\cdot b\downarrow}-h^a{}_{jc}h^j{}_{ab}=h_{c|b\downarrow}-h^a{}_{jc}h^j{}_{ab}, \tag{4.25}$$

which follows directly from Equations (4.16)–(4.21) above. The deformation scalar curvature $R:=\gamma^{cb}R_{cb}$ is just the ordinary spacetime scalar curvature ${}^{g}R$ because of the degeneracy of γ.

The index notation may be used to contract the Bianchi identities. One contraction yields

$$R_{bd\cdot e} - R_{be\cdot d} + R^{a}{}_{bde\cdot a} = 0 \tag{4.26}$$

and a second contraction by means of the deformation tensor yields

$$R^{b}{}_{d\cdot b} - R_{\cdot d} + \gamma^{be} R^{a}{}_{bde\cdot a} = 0. \tag{4.27}$$

An inspection of Equations (4.16–21) reveals that

$$\gamma^{be} R^{a}{}_{bde} = R^{a}{}_{d}. \tag{4.28}$$

Thus, even in this non-Riemannian geometry, Eq. (4.27), the twice contracted Bianchi identity, takes the familiar form

$$2G^{b}{}_{d\cdot b} = 0 \tag{4.29}$$

where

$$G^{b}{}_{d} := R^{b}{}_{d} - \tfrac{1}{2} R \delta^{b}{}_{d} \tag{4.30}$$

is the geometrically natural definition of the deformation *Einstein tensor.*

IV. DEFORMATION FIELD EQUATIONS AND FIRST ORDER PERTURBATIONS

A. Gauge conditions to complete the geometry

The gauge tensor H on a deformation is arbitrary so far. To fix the geometry of a deformation, one must impose further conditions which determine H. For a given fiducial gauge H', the gauge H is determined by the tensor $a := H - H'$. Proposition 4 relates this tensor to the second fundamental forms h and h' corresponding to H and H'. In index notation, the relation is

$$h'^{ab}{}_{s} = h^{ab}{}_{s} - a^{(a}{}_{s}{}^{\cdot b)} \tag{5.1}$$

where parentheses are being used to denote symmetrization. This relation suggests that a can be fixed by imposing conditions on h. There is just a one-parameter family of conditions that are linear in h, involve only first derivatives of h, and have the appropriate number of nontrivial components to determine a: For a constant k,

$$(h^{rj}{}_{s} + k\gamma^{rj} h_{s})_{\cdot j} = 0. \tag{5.2}$$

Equation (5.1), the Ricci identity, and Eq. (4.9) (which causes a term to vanish) then yield

$$a^{r}{}_{s}{}^{\cdot j}{}_{j} + (1+2k) a^{j}{}_{s\cdot j}{}^{r} + R^{r}{}_{j} a^{j}{}_{s} = -(h'^{rj}{}_{s} + k\gamma^{rj} h'_{s})_{\cdot j}. \tag{5.3}$$

The choice $k = -\frac{1}{2}$ makes this system manifestly hyperbolic on each spacetime in a deformation. In a chart aligned with either H or H' the system then takes the spacetime-covariant form

$$a^{\rho}{}_{(S)}{}^{;\kappa}{}_{\kappa} + R^{\rho}{}_{\kappa} a^{\kappa}{}_{(S)} = -(h'^{\rho\kappa}{}_{(S)} - \tfrac{1}{2}\gamma^{\rho\kappa} h'_{(S)})_{;\kappa}, \tag{5.4}$$

where the components $a^{\rho}{}_{(S)}$ are treated as components of a collection of spacetime vector fields. If the components $a^{\rho}{}_{(S)}$ and $a^{\rho}{}_{(S);\kappa}$ are given as initial data on a spacelike hypersurface in each spacetime of a deformation, then the tensor field a and therefore the gauge H are determined uniquely throughout the Cauchy development of each hypersurface.[10]

The choice $k = -\frac{1}{2}$ in Eq. (5.2) makes it convenient to define the trace-reversed second fundamental tensor

$$\bar{h}^{r}{}_{js} := h^{r}{}_{js} - \tfrac{1}{2} H^{r}{}_{j} h_{s} \tag{5.5}$$

so that Eq. (5.2) becomes

$$\bar{h}^{rj}{}_{s\cdot j} = 0. \tag{5.6}$$

In conventional treatments of spacetime perturbation theory, the gauge defined by this condition is called the "Hilbert gauge" or sometimes the "Lorentz gauge."

Because the gauge concept has been divorced from coordinates and given a strictly geometrical meaning, one can regard the Hilbert gauge as an object that is every bit as respectable as the normal vector to a surface. The fact that there are many other gauges should not be any more significant than the fact that there are many different vectors that point out of a surface. Thus, I propose to adopt the Hilbert gauge condition, Eq. (5.6), as defining the "normal projection tensor" for a spacetime in a $4+n$ deformation. It should, of course, be remembered that this tensor is not determined locally as it would be in a Riemannian geometry but propagates according to a hyperbolic wave equation in each spacetime.

B. Field equations on a spacetime deformation

The deformation tensor γ of a spacetime deformation is required to induce spacetime metrics which obey Einstein's equations in each spacetime. If the spacetime deformation supports nongravitational fields, then these are required to obey the usual spacetime-covariant field equations in each spacetime. To cast these requirements into a deformation-covariant form, notice that the standard deformation connection reduces to the usual spacetime connection whenever it acts on spacetime-tangent vector and tensor fields. Thus, require the contravariant form of each field to be tangent to spacetimes, write each spacetime-covariant field equation in contravariant form, and then just replace all semicolons by dots. The resulting deformation-covariant field equations are then equivalent to the original spacetime covariant field equations. This section gives several examples of this procedure.

If each spacetime in a deformation supports a Klein—Gordon field ϕ that satisfies the field equation

$$\phi^{;\mu}{}_{\mu} = 0, \tag{5.7}$$

then regard ϕ as a function on the deformation and rewrite the field equations as

$$\phi_{\cdot ab}\gamma^{ab} = 0 \quad \text{or} \quad \phi^{\cdot a}{}_{a} = 0. \tag{5.8}$$

The degeneracy of the deformation tensor γ then ensures that Eq. (5.8) is equivalent to Eq. (5.7).

If each spacetime in a deformation contains a test particle and the particle world lines are deformed continuously as one passes from a given spacetime to its neighbors, then these world lines form an $n+1$ surface in the deformation. On this surface there is a spacetime-tangent vector field u which satisfies

$$u^{\alpha}u^{\beta}{}_{;\alpha}=0 \tag{5.9}$$

within each spacetime. The deformation-covariant version of this requirement is therefore

$$u^{a}u^{b}{}_{\circ a}=0. \tag{5.10}$$

If each spacetime in a deformation supports an electromagnetic field, then there is a Maxwell field tensor $F^{mn}=-F^{nm}$ which is tangent to spacetimes and obeys the deformation-covariant field equations

$$F^{mn}{}_{\circ n}=4\pi J^{m} \tag{5.11}$$

$$F^{mn\cdot a}+F^{na\cdot m}+F^{am\cdot n}=0, \tag{5.12}$$

where the vector field J is automatically tangent to spacetimes and obeys the conservation law

$$J^{a}{}_{\cdot a}=0 \tag{5.13}$$

by virtue of Eq. (5.11). Thus, J represents the usual conserved electric current within each spacetime where one finds

$$J^{\alpha}{}_{;\alpha}=0. \tag{5.14}$$

The procedure for obtaining the deformation-covariant form of Einstein's equations is exactly the same as the procedure for nongravitational equations. If each spacetime in a deformation obeys Einstein's equations, ${}^{s}G^{\mu\nu}=8\pi T^{\mu\nu}$, then the deformation obeys the system

$$G^{mn}=8\pi T^{mn}. \tag{5.15}$$

Equations (4.22) and (4.30), together with the degeneracy of the deformation tensor that must be used to raise indexes, guarantee that the deformation Einstein tensor G^{mn} is tangent to spacetimes and induces the usual spacetime Einstein tensor ${}^{s}G^{\mu\nu}$ on each. Thus, T^{mn} is automatically tangent to spacetimes and it only remains to identify it with the usual spacetime stress—energy tensor to recover the spacetime Einstein equations. The deformation Bianchi identity given by Eq. (4.29) yields a deformation-covariant version

$$T^{mn}{}_{\circ n}=0 \tag{5.16}$$

of the stress—energy conservation law.

C. Perturbation equations from Lie derivatives

Once a system of field equations is in deformation-covariant form, its response to spacetime fluctuations can be obtained by differentiating both the field variables and the field equations with respect to the perturbation parameter. For example, the response of a scalar field ϕ is described by the functions

$$\partial\phi/\partial x^{(A)}=\phi_{,(A)},\quad \text{where } x^{(A)}=\lambda^{(A)},$$

and the field equations which determine the evolution of these functions can be obtained by differentiating Eq. (5.8) The conventional approach to perturbation theory corresponds to taking the derivative $\partial(\phi^{\cdot a}{}_{a})/\partial x^{(A)}$. A more geometrical way to express this procedure is to say that one chooses a set of vector fields $U_{(A)}$ which satisfy $\langle\theta^{(B)}, U_{(A)}\rangle=\delta^{(B)}_{(A)}$ and uses the Lie-derived field equations

$$\pounds_{U(A)}(\phi^{\cdot a}{}_{a})=0 \tag{5.17}$$

to evolve the Lie derivatives $\pounds_{U(A)}\phi$ of the field.

In order to obtain the perturbation equations in a useful form, one must reexpress them as restrictions on the Lie derivatives of the field. Thus, one must move the Lie derivative inside of the covariant derivatives that appear in Eq. (5.17). This need to interchange two different types of derivatives characterizes the usual approach to perturbation theory and makes for lengthy computations. Ordinarily one performs the interchange by expressing everything in terms of coordinates. A more geometrical statement of the usual procedure is that one expresses the covariant derivatives in terms of the Lie derivatives, changes the order of Lie differentiation, and then expresses the resulting equations in terms of covariant derivatives of $\pounds_{U(A)}\phi$.

D. Nongravitational perturbation equations from covariant derivatives

This paper proposes to simplify spacetime perturbation theory by using deformation-covariant derivatives instead of Lie derivatives to obtain the response of a field to spacetime fluctuations. In this approach one chooses a gauge tensor H with its corresponding identification gauge tensor $\imath$ and represents the response of a scalar field ϕ to a spacetime fluctuation by the form field $\nabla_{\imath}\phi=\phi_{\circ r}\imath^{r}{}_{a}\theta^{a}$. One takes the perturbation equation to be the projected covariant derivative

$$\phi^{\cdot a}{}_{as\imath}=0 \tag{5.18}$$

of the deformation-covariant field equation. In order to reexpress this equation in terms of the form-field $\phi_{\cdot s\imath}$ one first uses the Ricci identity to change the order of differentiation and obtain

$$\phi_{\cdot s\imath}{}^{a}{}_{a}-\phi_{\cdot m}R^{m}{}_{s\imath}=0. \tag{5.19}$$

Next, one uses Eq. (3.21) in the form

$$\imath^{a}{}_{s\circ b}=-h^{a}{}_{bs} \tag{5.20}$$

to change the order of differentiation and projection so that Eq. (5.19) becomes

$$\phi_{|s\imath}{}^{a}{}_{a}+2\phi_{\cdot da}h^{da}{}_{s}+\phi_{\cdot d}(h^{da}{}_{s\cdot a}-R^{d}{}_{s\imath})=0.$$

The deformation compatibility of the connection and the projection identities satisfied by γ^{ab}, $h^{ab}{}_{s}$, and $R^{d}{}_{s}$ make it possible to insert spacetime projections into this expression without changing anything. The perturbation equation then becomes

$$\phi_{|s\downarrow}{}^{a\uparrow}{}_{a\uparrow} + 2\phi_{|d\uparrow a\uparrow}h^{da}{}_{s} + \phi_{|d\uparrow}(h^{da}{}_{s\cdot a} - R^{d\uparrow}{}_{s\downarrow}) = 0.$$

Now use Eq. (4.24) to evaluate the Ricci tensor term and notice that the result can be written in the form

$$\phi_{|s\downarrow}{}^{a\uparrow}{}_{a\uparrow} + 2\phi_{|d\uparrow a\uparrow}h^{da}{}_{s} + 2\phi_{|d\uparrow}\bar{h}^{da}{}_{s\cdot a} = 0,$$

which the Hilbert gauge condition simplifies further to just

$$\phi_{|s\downarrow}{}^{a\uparrow}{}_{a\uparrow} + 2\phi_{|d\uparrow a\uparrow}h^{da}{}_{s} = 0. \tag{5.21}$$

To see that my notation is not concealing any unpleasant complications, use Theorem 1 to write Eq. (5.21) in an aligned coordinate system:

$$[\phi_{,(S)}]^{;\alpha}{}_{\alpha} + 2\phi_{;\delta\alpha}h^{\delta\alpha}{}_{(S)} = 0. \tag{5.22}$$

Here, the spacetime-covariant derivative treats $\phi_{,(S)}$ as a collection of scalar fields. This result is not quite the same as the perturbation equation that one obtains from the conventional Lie derivative variation procedure. The conventional procedure yields an equation in which the second fundamental form that appears in Eq. (5.22) is replaced by its trace reverse. The two versions of the perturbation equation agree when the scalar field equation is satisfied, but not otherwise.

The procedure which has been illustrated here for a simple scalar field equation can be applied to any set of nongravitational field equations. It has several advantages over the conventional approach:

(1) It is entirely geometrical and does not require the introduction of coordinates or basis vector fields.

(2) It saves one the trouble of varying connection coefficients and reexpressing the result in terms of covariant derivatives. All such computations have been "prepackaged" in the form of the Ricci identity and the Gauss—Codazzi equations of Theorem 2.

(3) Because the deformation connection is deformation-compatible, one may raise and contract indexes on tensors either before or after varying them without worrying about the introduction of additional terms.

(4) Because the Hilbert gauge condition can be regarded as a part of the deformation geometry, one can use it without compromising coordinate independence.

E. Gravitational perturbation equations from covariant derivatives

The procedure for obtaining a perturbation equation from Einstein's field equation is similar in general outline to the procedure for nongravitational equations. However, the presence of the curvature tensor in the field equation changes some of the details of the procedure. As before, the perturbation equation is taken to be the projected deformation-covariant derivative

$$G^{mn}{}_{\cdot s\downarrow} = 8T^{mn}{}_{\cdot s\downarrow}. \tag{5.23}$$

Also as before, the projected derivative is to be moved inside of the spacetime derivatives that are already present in the field equation. However, the Ricci identity cannot be used for this purpose because the spacetime derivatives in G^{mn} are all hidden. However, the contracted Bianchi identity in the form

$$G^{mn}{}_{\cdot s} = R^{m}{}_{s}{}^{\cdot n} - \gamma^{mn}R^{e}{}_{s\cdot e} + R^{am}{}_{s}{}^{n}{}_{\cdot a}$$

serves the same purpose and converts the perturbation equation to the form

$$R^{m}{}_{s\downarrow}{}^{\cdot n} - \gamma^{mn}R^{e}{}_{s\downarrow\cdot e} + R^{am}{}_{s\downarrow}{}^{n}{}_{\cdot a} = 8\pi T^{mn}{}_{\cdot s\downarrow}.$$

Equation(5.20) can then be used to change the order of differentiation and projection with the result

$$R^{m}{}_{s\downarrow}{}^{|n} + h^{rn}{}_{s}R^{m}{}_{r} - \gamma^{nm}R^{e}{}_{s\downarrow|e} - \gamma^{mn}h^{r}{}_{es}R^{e}{}_{r}$$
$$+ R^{am}{}_{s\downarrow}{}^{n}{}_{|a} + h^{r}{}_{as}R^{am}{}_{r}{}^{n} = 8\pi T^{mn}{}_{\cdot s\downarrow}.$$

Now use the Gauss—Codazzi equation (4.24) in the form

$$R^{am}{}_{s\downarrow}{}^{n} = -R^{amn}{}_{s\downarrow} = h^{mn}{}_{s}{}^{\cdot a} - h^{an}{}_{s}{}^{\cdot m}$$

together with the Hilbert gauge condition and the Ricci identity to obtain the perturbation equation in the form

$$h^{mn}{}_{s}{}^{\cdot a}{}_{a} + 2R^{m}{}_{a}{}^{n}{}_{b}h^{ab}{}_{s} = 8\pi\bar{T}^{mn}{}_{\cdot s\downarrow}, \tag{5.24}$$

where $\bar{T}^{mn}$ denotes the trace reverse of the stress—energy tensor.

The insertion of spacetime projections into Eq. (5.24) is a trivial operation that brings it into the more explicit form

$$h^{mn}{}_{s}{}^{|a\uparrow}{}_{a\uparrow} + 2R^{m\uparrow}{}_{a\uparrow}{}^{n\uparrow}{}_{b\uparrow}h^{ab}{}_{s} = 8\pi\bar{T}^{m\uparrow n\uparrow}{}_{|s\downarrow}. \tag{5.25}$$

In an aligned coordinate system, this equation becomes

$$h^{\mu\nu}{}_{(S)}{}^{;\alpha}{}_{\alpha} + 2R^{\mu}{}_{\alpha}{}^{\nu}{}_{\beta}h^{\alpha\beta}{}_{(S)} = 8\pi\bar{T}^{\mu\nu}{}_{,(S)}. \tag{5.26}$$

Notice the similarity between this equation and the perturbation equation for the scalar field (5.22).

If Eq. (5.26) is compared with the Lie derivative perturbation equation that one normally sees,[11] one finds that it differs in exactly the same way as the scalar field perturbation: the undifferentiated second fundamental form has had its trace reversed. This change makes Eq. (5.26) somewhat simpler than the conventional perturbation equation. If the stress—energy components are given as particular functions, then the simplification is a notational illusion because, by Theorem 1,

$$\bar{T}^{\mu\nu}{}_{,(S)} = \bar{T}^{\mu\nu}{}_{,(S)} - \bar{T}^{\rho\nu}h^{\mu}{}_{\rho(S)} - \bar{T}^{\mu\rho}h^{\nu}{}_{\rho(S)}.$$

When the additional terms which have been introduced by the deformation connection are taken into account, one may use the field equation to obtain the conventional perturbation equation from Eq. (5.26). If, however the stress–energy tensor is expressed in terms of nongravitational fields, then the simplification is real because the deformation-covariant derivative $\overline{T}^{\mu\nu}{}_{*(s)}$ can be computed more easily than $\overline{T}^{\mu\nu}{}_{,(s)}$.

APPENDIX

Proof of Proposition 2

Use the abbreviation $\nabla_{(C)} := \nabla_{\partial/\partial x(C)}$ and note that

$$\langle dx^{(B)}, \nabla_{(C)}\partial/\partial x^{(A)}\rangle = -\langle \nabla_{(C)}dx^{(B)}, \partial/\partial x^{(A)}\rangle$$

$$= -\langle \nabla_{(C)}\theta^{(B)}, \partial/\partial x^{(A)}\rangle = 0. \quad \text{(A1)}$$

Expand $\partial/\partial x^{(C)}$ in the covariantly constant basis vectors $e_{(K)}$,

$$\partial/\partial x^{(A)} = F_{(A)}{}^{(K)} e_{(K)},$$

so that

$$\langle dx^{(B)}, \nabla_{(C)}\partial/\partial x^{(A)}\rangle = F_{(A)}{}^{(K)}{}_{,(C)}\langle dx^{(B)}, e_{(K)}\rangle.$$

Equation (A1) and the linear independence of the forms $dx^{(B)}$ and the vectors $e_{(K)}$ then imply

$$F_{(A)}{}^{(K)}{}_{,(C)} = 0$$

so that the coefficients $F_{(A)}{}^{(K)}$ are constant on each identification surface. But then

$$\nabla_{(C)}\partial/\partial x^{(A)} = \nabla_{(C)}(F_{(A)}{}^{(K)} e_{(K)})$$

$$= F_{(A)}{}^{(K)}{}_{,(C)} e_{(K)} + F_{(A)}{}^{(K)}\nabla_{(C)} e_{(K)} = 0.$$

Derivation of Eq. (3.12)

$$(\pounds_{\iota U}\gamma)(H\alpha, H\beta)$$

$$= -\langle H\alpha, \nabla_{\gamma\beta}\iota U\rangle - \langle H\beta, \nabla_{\gamma\alpha}\iota U\rangle$$

$$= -\langle \alpha, H\nabla_{\gamma\beta}\iota U\rangle - \langle \beta, H\nabla_{\gamma\alpha}\iota U\rangle$$

$$= \langle \alpha, [\nabla_{\gamma\beta}, H]\iota U\rangle + \langle \beta, [\nabla_{\gamma\alpha}, H]\iota U\rangle$$

$$= \langle \alpha, h_U\gamma\beta\rangle + \langle \beta, h_U\gamma\alpha\rangle. \quad \text{(A2)}$$

Derivation of Eq. (3.15)

From Eqs. (3.14) and (A2)

$$2\langle \alpha, h'_U\gamma\beta\rangle$$

$$= -\langle H'\alpha, \nabla_{\gamma\beta}\iota' U\rangle - \langle H'\beta, \nabla_{\gamma\alpha}\iota' U\rangle$$

$$= \langle \iota'\nabla_{\gamma\beta}H'\alpha + \iota'\nabla_{\gamma\alpha}H'\beta, U\rangle$$

$$= \langle \iota\nabla_{\gamma\beta}H\alpha + \iota\nabla_{\gamma\alpha}H\beta + a\nabla_{\gamma\beta}H\alpha + a\nabla_{\gamma\alpha}H\beta$$

$$- \iota\nabla_{\gamma\beta}a\alpha - \iota\nabla_{\gamma\alpha}a\beta, U\rangle.$$

Here, the terms quadratic in a vanish because a annihilates spacetime-tangent vectors. With this expression and Eqs. (3.14) and (A2) again, find

$$2\langle \alpha, (h'_U - h_U)\gamma\beta\rangle$$

$$= \langle a\nabla_{\gamma\beta}H\alpha + a\nabla_{\gamma\alpha}H\beta - \iota\nabla_{\gamma\beta}a\alpha - \iota\nabla_{\gamma\alpha}a\beta, U\rangle$$

$$= \langle a(\nabla_{\gamma\beta}H)\alpha + aH_{\gamma\beta}\alpha + a(\nabla_{\gamma\alpha})H\beta + aH\nabla_{\gamma\alpha}\beta$$

$$- \iota(\nabla_{\gamma\beta}a)\alpha - \iota a\nabla_{\gamma\beta}\alpha - \iota(\nabla_{\gamma\alpha}a)\beta - \iota a\nabla_{\gamma\alpha}\beta, U\rangle. \quad \text{(A3)}$$

But

$$\langle a(\nabla_{\gamma\beta}H)\alpha, U\rangle = \langle \alpha, (\nabla_{\gamma\beta}H)aU\rangle = \langle \alpha, h_{aU}\gamma\beta\rangle$$

and, from Eq. (3.11) and the fact that aU is in $\gamma T_p{}^*$,

$$h_{aU} = h_{\iota aU} = 0$$

so that two of the terms in Eq. (A3) vanish and leave

$$\langle \alpha, 2(h'_U - h_U)\gamma\beta\rangle = \langle (aH - \iota a)(\nabla_{\gamma\beta}\alpha + \nabla_{\gamma\alpha}\beta)$$

$$- \iota(\nabla_{\gamma\beta}a)\alpha - \iota(\nabla_{\gamma\alpha}a)\beta, U\rangle$$

$$= \langle \nabla_{\gamma\beta}\alpha + \nabla_{\gamma\alpha}\beta, (Ha - a\iota)U\rangle - \langle \alpha, (\nabla_{\gamma\beta}a)\iota U\rangle$$

$$- \langle \beta, (\nabla_{\gamma\alpha}a)\iota U\rangle.$$

But $HaU = aU$ and, because H and H' must both reduce to the identity on $\gamma T_P{}^*$, $aH = 0$ so that $a\iota = a$ and therefore $(Ha - a\iota)U = 0$ which leaves

$$\langle \alpha, (h'_U - h_U)\gamma\beta\rangle = -\tfrac{1}{2}\{\langle \alpha, (\nabla_{\gamma\beta}a)\iota U\rangle + \langle \beta, (\nabla_{\gamma\alpha}a)iU\rangle\}$$

$$= -\tfrac{1}{2}\{\langle \alpha, (\nabla_{\gamma\beta}a\iota)U - a(\nabla_{\gamma\beta}\iota)U\rangle$$

$$+ \langle \beta, (\nabla_{\gamma\alpha}a\iota)U - a(\nabla_{\gamma\alpha}\iota)U\rangle\}$$

$$= -\tfrac{1}{2}\{\langle \alpha, (\nabla_{\gamma\beta}a)U + ah_U\gamma\beta\rangle$$

$$+ \langle \beta, (\nabla_{\gamma\alpha}a)U + ah_U\gamma\alpha\rangle\}$$

$$= -\tfrac{1}{2}\{\langle \alpha, (\nabla_{\gamma\beta}a)U\rangle + \langle \beta, (\nabla_{\gamma\alpha}a)U\rangle\}.$$

Proof of Theorem 1

First note that the identification surfaces can be used to construct an aligned chart on $\mathcal{M}$ from each chart on the spacetime Σ. Thus there is a neighborhood U of Σ which can be covered by such aligned charts. If Eqs. (3.18)–(3.20) hold on each such chart, then they define a standard deformation connection on U and imply Eq. (3.16). Thus the existence of the standard connection is proven by construction.

The nontrivial part of the proof is uniqueness: Assume that ∇ is a standard connection and show that it must satisfy Eqs. (3.16)–(3.20) in each aligned chart. It is convenient to establish Eq. (3.16) first. Use the decomposition $1 = H + \iota$ to obtain

$$\nabla_A B = \nabla_{HA}HB + \nabla_{HA}\iota B + \nabla_{\iota A}HB + \nabla_{\iota A}\iota B. \quad \text{(A4)}$$

From gauge flatness, $H\nabla_{\iota A}\iota B = 0$ and $\iota\nabla_{\iota A}\iota B = \mathcal{S}_{\iota A}\iota B$ so that $\nabla_{\iota A}\iota B = \not{b}_{\iota A}\iota B$. From Proposition 1 [Eq. (3.6)], $\nabla_{HA}HB$ is tangent to spacetimes so that $\nabla_{HA}HB = \sqcap_{HA}HB$.

The first and last terms of Eq. (A4) have now been brought into agreement with Eq. (3.16). Now consider the term:

$$\nabla_{HA}\iota B=\nabla_{HA}B-\nabla_{HA}HB=\nabla_{HA}B-[\nabla_{HA},\ H]B$$

$$-H\nabla_{HA}B=\iota\nabla_{HA}B-h_BA.$$

From Proposition 1 [Eq. (3.6)], $\iota\nabla_{HA}B=\iota\nabla_{HA}\iota B$ so that

$$\nabla_{HA}\iota B=\iota\nabla_{HA}\iota B-h_BA.$$

Because the torsion vanishes, $\nabla_{HA}\iota B=\nabla_{\iota B}HA+[HA,\ \iota B]$ and thus

$$\nabla_{HA}\iota B=\iota\nabla_{\iota B}HA+\iota[HA,\iota B]-h_BA.$$

Another use of Eq. (3.6) then gives

$$\nabla_{HA}\iota B=\iota[HA,\ \iota B]-h_BA.$$

Only one term of Eq. (A4) remains to be considered. The techniques that have just been used reduce it as follows:

$$\nabla_{\iota A}HB=H\nabla_{\iota A}HB=H\nabla_{HB}\iota A+H[\iota A,\ HB]$$

$$=-h_AB+H[iA,\ HB]$$

and Eq. (A4) yields Eq. (3.16). Now take the vector fields A and B in Eq. (3.16) to be the basis vectors $e_a:=\partial/\partial x^a$ and e_b of an aligned chart. Equations (3.18)—(3.20) then follow.

Derivation of Eq. (3.22)

The simplest procedure is to write out the components in an aligned chart:

$$h^{\delta}{}_{\alpha(A)\cdot(B)}=h^{\delta}{}_{\alpha(A),(B)}+h^{\rho}{}_{\alpha(A)}\Gamma^{\delta}{}_{\rho(B)}-h^{\delta}{}_{\rho(A)}\Gamma^{\rho}{}_{\alpha(B)}$$

$$=\tfrac{1}{2}(\gamma^{\delta\rho},{}_{(A)}g_{\rho\alpha}),{}_{(B)}-h^{\rho}{}_{\alpha(A)}h^{\delta}{}_{\rho(B)}$$

$$+h^{\delta}{}_{\rho(A)}h^{\rho}{}_{\alpha(B)}$$

$$=\tfrac{1}{2}\gamma^{\delta\rho},{}_{(A)(B)}g_{\rho\alpha}+\tfrac{1}{2}\gamma^{\delta\rho},{}_{(A)}g_{\rho\alpha,(B)}-h^{\rho}{}_{\alpha(A)}h^{\delta}{}_{\rho(B)}$$

$$+h^{\delta}{}_{\rho(A)}h^{\rho}{}_{\alpha(B)}$$

$$=\tfrac{1}{2}\ \gamma^{\delta\rho}{}_{,(A)(B)}g_{\rho\alpha}-2h^{\delta\rho}{}_{(A)}h_{\rho\alpha(B)}+h^{\delta}{}_{\rho(A)}\ h^{\rho}{}_{\alpha(B)}$$

$$-h^{\delta}{}_{\rho(B)}h^{\rho}{}_{\alpha(A)}$$

$$=\tfrac{1}{2}\gamma^{\delta\rho},{}_{(A)(B)}g_{\rho\alpha}-(h^{\delta}{}_{\rho(A)}h^{\rho}{}_{\alpha(B)}+h^{\delta}{}_{\rho(B)}h^{\rho}{}_{\alpha(A)}).$$

This expression is manifestly symmetric in (A) and (B) which establishes Eq. (3.22).

Derivation of Eq. (4.7) (*Codazzi identity*)

Equation (4.1) and the Jacobi identity for commutators imply that, for any vector fields J and K,

$$[\mathcal{R}(J,K),\ H=[[\nabla_J,\ \nabla_K],\ H]-[\nabla_{[J,K]},H]$$

$$=-[[\nabla_K,H],\nabla_J]-[[H,\nabla_J],\nabla_K]$$

$$-[\nabla_{[J,K]},H]$$

$$=\nabla_J[\nabla_K,H]-[\nabla_K,H]\nabla_J+[\nabla_J,H]\nabla_K$$

$$-\nabla_K[\nabla_J,H]-[\nabla_{[J,K]},H].$$

Now take $J=HA$ and $K=HB$ and use the definition of h [Eq. (3.9)] to obtain $[\mathcal{R}(HA,HB),H]C=\nabla_{HA}h_CHB-h_{\nabla_{HA}C}HB+h_{\nabla_{HB}C}HA-\nabla_{HB}h_CHA-h_C[HA,HB]$. Use vanishing torsion [Eq. (3.1)] to reexpress the last term:

$$[\mathcal{R}(HA,HB),H]C=(\nabla_{HA}h_CHB-h_{\nabla_{HA}}{}_CHB-h_C\nabla_{HA}HB)$$

$$-(\nabla_{HB}h_CHA-h_{\nabla_{HB}}{}_CHA-h_C\nabla_{HB}HA)$$

$$=(\nabla_{HA}h)_CHB-(\nabla_{HB}h)_CHA.$$

Proof of Theorem 2 (*Gauss, Codazzi, Mainardi identities*)

To eastablish Eq. (4.9), calculate

$$\langle\theta^{(A)},\ \iota\mathcal{R}(A,B)C\rangle=\langle\theta^{(A)},\ [\nabla_A,\nabla_B]C-\nabla_{[A,B]}C\rangle.$$

Deformation compatibility [Eq. (3.3)] makes it possible to bring all of the derivatives outside of the inner products

$$\langle\theta^{(A)},\ \iota\mathcal{R}(A,B)C\rangle=([\nabla_A,\nabla_B]-\nabla_{[A,B]})\ \ \langle\theta^{(A)},C\rangle$$

$$=([A,B]-[A,B])\langle\theta^{(A)},C\rangle=0.$$

Thus, Eq. (4.9) is established.

Equation (4.10) is just a statement of gauge flatness.

Equation (4.11) follows directly from Eqs. 3.16 and 4.8.

Equation (4.12) is obtained from Eq. (4.7) by replacing the vector C by ιC.

To establish Equation (4.13), write the definition

$$\mathcal{R}(HA,\iota B)\iota C=\nabla_{HA}\nabla_{\iota B}\iota C-\nabla_{\iota B}\nabla_{HA}\iota C-\nabla_{[HA,\iota B]}\iota C \tag{A5}$$

and then use the definition [Eq. (3.9)] of h together with Eqs. (3.1) and (3.5) and the consequence $[\nabla_{\iota V},\iota]=0$ of gauge flatness to commute all of the ι factors in (A5) to the left:

$$\nabla_{HA}\iota C=[\nabla_{HA},\iota]C+\iota\nabla_{HA}C=\iota\nabla_{HA}C-h_CHA,$$

$$\nabla_{\iota B}\nabla_{HA}\iota C=\nabla_{\iota B}\iota\nabla_{HA}C-\nabla_{\iota B}h_CHA$$

$$=\iota\nabla_{\iota B}\nabla_{HA}C-\nabla_{\iota B}h_CHA, \tag{A6}$$

$$\nabla_{HA}\nabla_{\iota B}\iota C=\nabla_{HA}\iota\nabla_{\iota B}C=\iota\nabla_{HA}\nabla_{\iota B}C-h_{\nabla_{\iota B}C}HA. \tag{A7}$$

The last term in (A5) must first be simplified by using vanishing torsion:

$$[HA, \imath B] = \nabla_{HA}\imath B - \nabla_{\imath B}HA = \imath\nabla_{HA}B - h_B HA$$

$$- \nabla_{\imath B}HA = \imath\nabla_{HA}B - Hh_B HA - H\nabla_{\imath B}HA.$$

The last term in (A5) then becomes

$$\nabla_{[HA, \imath B]}\imath C = \nabla_{\imath\nabla_{HA}B}\imath C - \nabla_{H(h_B HA + \nabla_{\imath B}HA)}\imath C$$

$$= \imath\nabla_{\imath\nabla_{HA}B}C - \imath\nabla_{H(h_B HA + \nabla_{\imath B}HA)}\imath C$$

$$+ h_C(h_B HA + \nabla_{\imath B}HA). \qquad \text{(A8)}$$

Now use Eq. (4.9) to obtain

$$\mathcal{R}(HA, \imath B)\imath C = H\mathcal{R}(HA, \imath B)\imath C. \qquad \text{(A9)}$$

Equations (A6)—(A8) yield

$$H\nabla_{\imath B}\nabla_{HA}\imath C = -H\nabla_{\imath B}h_C HA = -(1-\imath)\nabla_{\imath B}Hh_C HA$$

$$= -\nabla_{\imath B}h_C HA \quad \text{[use Eq. (3.6)]}, \quad H\nabla_{HA}\nabla_{\imath B}\imath C$$

$$= -h_{\nabla_{\imath B}C}HA,$$

$$H\nabla_{[HA, \imath B]}\imath C = h_C(h_B HA + \nabla_{\imath B}HA),$$

and Eqs. (A5) and (A9) then become

$$\mathcal{R}(HA, \imath B)\imath C = -h_{\nabla_{\imath B}C}HA + \nabla_{\imath B}h_C HA - h_C h_B HA - h_c\nabla_{\imath B}HA$$

$$= (\nabla_{\imath B}h)_C HA - h_C h_B HA,$$

which establishes Eq. (4.13).

Equation (4.14) may be obtained from Eq. (4.13) by using the cyclic sum identity in the form

$$\mathcal{R}(\imath A, \imath B)HC = \mathcal{R}(HC, \imath B)\imath A - \mathcal{R}(HC, \imath A)\imath B.$$

Equation (4.15) may also be obtained from the cyclic sum identity in the form

$$\mathcal{R}(\gamma\alpha, \imath B)\gamma\beta - \mathcal{R}(\gamma\beta, \imath B) = \mathcal{R}(\gamma\alpha, \gamma\beta)\imath B.$$

Use the deformation compatibility of the connection [Eq. (3.5)] to put this identity into the form

$$\langle\sigma, \mathcal{R}(\gamma\alpha, \imath B)\gamma\beta\rangle + \langle\alpha, \mathcal{R}(\gamma\beta, \imath B)\gamma\sigma\rangle = \langle\sigma, \mathcal{R}(\gamma\alpha, \gamma\beta)\imath B\rangle. \qquad \text{(A10)}$$

Now define the quantities

$$x := \langle\sigma, \mathcal{R}(\gamma\alpha, \imath B)\gamma\beta\rangle, \quad u := \langle\sigma, \mathcal{R}(\gamma\alpha, \gamma\beta)\imath B\rangle,$$

$$y := \langle\alpha, \mathcal{R}(\gamma\beta, \imath B)\gamma\sigma\rangle, \quad v := \langle\alpha, \mathcal{R}(\gamma\beta, \gamma\sigma)\imath B\rangle,$$

$$z := \langle\beta, \mathcal{R}(\gamma\sigma, \imath B)\gamma\alpha\rangle, \quad w := \langle\beta, \mathcal{R}(\gamma\sigma, \gamma\alpha)\imath B\rangle,$$

and notice that a cyclic permutation of (α, β, σ) produces cyclic permutations of (x, y, z) and (u, v, w). Thus, (A10) becomes

$$x + y = u$$

and cyclic permutation yields

$$y + z = v \quad \text{and} \quad z + x = w.$$

Solve these three equations to obtain

$$x = \tfrac{1}{2}(u - v + w),$$

and then use Eqs. (4.5) and (4.12) to obtain Eq. (4.15).

[1]An idea of the extent of this effort can be obtained from this list of recent papers: W. H. Press, Phys. Rev. D **15**, 965—68 (1977); R. V. Wagoner and C. M. Will, Report No. OAP 449 (unpublished); R. Epstein and R. V. Wagoner, Astrophys. J. **197**, 717—23 (1975); R. V. Wagoner, Astrophys. J. Lett. **196**, L63—L65 (1975); R. J. Crowly and K. S. Thorne, Report No. OAP 450 (unpublished); K. S. Thorne and S. J. Kovacs, Astrophys. J. **200**, 245—62 (1975); J. Ehlers, A. Rosenblum, J. N. Goldberg and P. Havas, Astrophys. J. Lett. **208**, L77—L81 (1976); J. L. Anderson, Gen. Relativ. Gravit. **7**, 643—52 (1976); J. M. Bird, Ann. Phys. (N. Y.) **101**, 345—54 (1976); R. J. Adler and N. B. Zeks, Phys. Rev. D **12**, 3007—12 (1975); W. L. Burke, J. Math. Phys. **12**, 401—18 (1971).

[2]There is a large literature on the stability of gravitational collapse and black holes. See, for example, R. H. Price, Ph. D. dissertation, California Institute of Technology (1971); Phys. Rev. D **5**, 2419—38, 2439—54 (1972); C. V. Vishveshwara, Phys. Rev. D **1**, 2870—79 (1970); W. H. Press and S. A. Teukolsky, Nature **238**, 211—12 (1972); V. Moncrief, Phys. Rev. D **12**, 1526—37 (1975).

[3]References on the gravitational instability theory of galaxy formation may be found in B. J. T. Jones, Rev. Mod. Phys. **48**, 107—49 (1976).

[4]R. Geroch, Commun. Math. Phys. **13**, 180—93 (1969). Geroch's "limit space" is what I am calling a "regular spacetime 4 + 1 deformation."

[5]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[6]See, for example, S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U. P., Cambridge, 1973), p. 22.

[7]Hawking and Ellis, Ref. 6, p. 30.

[8]N. J. Hicks, *Note on Differential Geometry* (Van Nostrand, New York, 1965).

[9]For a discussion of spacetime perturbations in terms of such Lie derivatives on spacetime deformations, see J. M. Stewart and M. Walker, Proc. Roy. Soc. (Lond.) A **341**, 49—74 (1974).

[10]Hawking and Ellis, Ref. 6, pp. 233—43.

[11]Misner, Thorne and Wheeler, Ref. 5, p. 967.

Causality in homogeneous spaces

F. J. Chinea[a]

Department of Physics, Syracuse University, Syracuse, New York 13210
(Received 30 December 1977)

A simple proof is given of the microcausality of quantum fields defined on certain homogeneous pseudo-Riemannian spaces. The proof is group theoretic in nature and does not depend on the detailed form of the generalized Pauli–Jordan propagator. As an illustration, applications are given to de Sitter and anti-de Sitter spaces; in the latter case, it is shown that the commutator of any boson field vanishes for any pair of points in the space.

1. INTRODUCTION

It is of some interest to consider the quantization of fields in pseudo-Riemannian spaces, having in mind its possible application to gravitational and cosmological external field problems. The quantization procedure is free of ambiguities and permits a straightforward and invariant particle interpretation in the case of a space on which a group of isometries acts transitively; the quantization is formulated in such a case along group-theoretic lines,[1] and it has the additional mathematical interest of its applications to the representation theory of the corresponding group.

The causality properties of a quantum field defined in that context may be studied by considering the behavior of the generalized Pauli-Jordan propagator Δ, which is defined either from a differential equation point of view[2] or by group-theoretic considerations,[1] the latter aspect being related to the former one by means of harmonic analysis on the group of isometries.[3]

In the following we give a proof of the causality of Δ which depends on the group invariance of Δ and its antisymmetry, making it unnecessary to consider its explicit form in each particular case. Such a procedure may be applied to certain homogeneous spaces, of which examples are given.

2. MICROCAUSALITY IN HOMOGENEOUS PSEUDO-RIEMANNIAN SPACES

Let us consider a pseudo-Riemannian space M, a homogeneous space of a group of isometries G. Let Ω be an arbitrary point in M and H_Ω its isotropy group; we have $M \approx G/H_\Omega$. Let $\phi(x)$ be a free-scalar quantum field defined on M. The notion of causality we are going to consider is that of *microcausality*, that is to say,

$$[\phi(x), \phi(y)] = 0$$

for spacelike separation of x and y [timelike, spacelike, and null separations are defined by means of the square of the geodesic distance between x and y on M, denoted in what follows by $\Gamma(x,y)$]. On the other hand, in order to have the desired dynamics of the field, positivity of the scalar product, a particle interpretation in Fock space, etc., we require

$$[\phi(x), \phi(y)] = i\Delta(x,y),$$

where Δ is the generalized Pauli—Jordan commutator function. In order to implement microcausality we must therefore require

$$\Delta(x,y)=0,$$

for x,y spacelike. Δ satisfies the following properties:

(i) Group invariance:

$$\Delta(x,y) = \Delta(g.x, g.y) \text{ for any } g \in G$$

(where we denote by $g \cdot x$ the group action $G \times M \to M$).

(ii) Antisymmetry

$$\Delta(x,y) = -\Delta(y,x).$$

These two properties are enough to ensure microcausality in those homogeneous spaces which satisfy the hypothesis of the following lemma, independently of the fact that Δ is an eigendistribution of the Laplace—Beltrami operator on M.

Lemma: Let $f(x,y)$ be any group-invariant, odd function (distribution) defined on $M \approx G/H_\Omega$. Let us assume that any point $z \in M$ which is spacelike with respect to Ω satisfies $z = g \cdot \Omega$ for some $g \in G$ (always true by transitivity) and $z = h.g^{-1}\Omega$ for some $h \in H_\Omega$. Then $f(x,y) = 0$ for any pair of points x, y which are spacelike relative to each other.

Proof: Let $x, y \in M$ be spacelike. By transitivity, there exist $g_x, g_y \in G$ such that $x = g_x.\Omega$, $y = g_y.\Omega$. We have

$$f(x,y) = f(g_x \cdot \Omega, g_y.\Omega) = f(\Omega, g_x^{-1}.g_y.\Omega). \tag{2.1}$$

But Ω and $g_x^{-1}.g_y.\Omega$ are spacelike,

$$\Gamma(\Omega, g_x^{-1}.g_y.\Omega) = \Gamma(g_x.\Omega, g_y.\Omega) = \Gamma(x,y)$$

due to the invariance of Γ under the group of isometries G. Under the hypothesis of the lemma, we have $g_x^{-1}.g_y.\Omega = g.\Omega$ and $g_x^{-1}.g_y.\Omega = h.g^{-1}\Omega$ for some $g \in G$, $h \in H_\Omega$. Hence

$$\begin{aligned} f(\Omega, g_x^{-1}.g_y.\Omega) &= f(\Omega, h.g^{-1}\Omega) = f(h^{-1}, g^{-1}\Omega, g^{-1}\Omega) \\ &= f(\Omega, g^{-1}\Omega) = f(g.\Omega, \Omega) = f(g_x^{-1}g_y.\Omega, \Omega) = f(g_y.\Omega, g_x.\Omega) \\ &= f(y,x). \end{aligned} \tag{2.2}$$

Finally, (2.1) and (2.2) imply $f(x,y) = 0$ by the antisymmetry of $f(x,y)$.

Remark: Let us mention that the preceding lemma provides a simple proof of the causality of the usual Pauli-Jordan propagator in Minkowski space without any reference to its explicit functional form or the fact

[a]Present address: Departamento de Métodos Matemáticos, Facultad de Ciencias Físicas, Universidad de Madrid, Madrid-3, Spain

 0022-2488/78/1911-2305$1.00

that it is a solution to the Klein–Gordon equation. G is given in the present case by $\mathcal{L}_+^\uparrow \times \mathbb{T}_4$

3. APPLICATION TO DE SITTER SPACE

We may define de Sitter space M as the hyperboloid

$$\eta_{ab}x^a x^b = -1$$

embedded in a five-dimensional flat space S with normal hyperbolic metric: $\eta_{ab} = \mathrm{diag}(1,-1,-1,-1,-1)$ $(a,b,\cdots = 0,1,2,3,4)$.[4]

The isometry group $SO(4,1)\times \mathbb{T}_5$ of the embedding space S is restricted to $SO(4,1)$ on de Sitter space. The corresponding Killing vectors $\xi_{\{bc\}}$, labelled by the ten antisymmetric pairs of indices $\{bc\}$, have the following contravariant components in S,

$$\xi^a_{\{bc\}} = \delta^a_b x_c - \delta^a_c x_b.$$

The one-parameter subgroup generated by the Killing vector $\xi_{\{bc\}}$ will be denoted by $g_{\{bc\}}(t)$.

It is easily seen that the square of the geodesic distance between points x and y in M is

$$\Gamma(x,y) = (\mathrm{arccosh}\,|\eta_{ab}x^a y^b|)^2,$$

where x^a, y^b are the coordinates of x and y considered as points in S. Spacelike points x,y are therefore characterized by

$$|\eta_{ab}x^a y^b| < 1. \qquad (3.1)$$

Let us introduce the following parametrization of $z \in M$

$$z^0 = \sinh\alpha$$

$$z^1 = \cos\chi\cosh\alpha$$

$$z^2 = \sin\theta\cos\varphi\sin\chi\cosh\alpha$$

$$z^3 = \sin\theta\sin\varphi\sin\chi\cosh\alpha$$

$$z^4 = \cos\theta\sin\chi\cosh\alpha,$$

with $\alpha \in \mathbb{R}, \theta, \chi \in [0,\pi), \varphi \in [0, 2\pi)$. (3.1) is now written as

$$|\cos\chi\cosh\alpha| < 1. \qquad (3.2)$$

We choose Ω as the point $\Omega = (0,1,0,0,0)$ in the embedding space. Then, $z = g.\Omega$ with

$$g = g_{\{23\}}(\varphi).g_{\{24\}}(\pi/2-\theta).g_{\{12\}}(\chi).g_{\{01\}}(-\alpha).$$

The isotropy group of Ω, H_Ω, is generated by the one-parameter subgroups

$$g_{\{02\}}(t),\ g_{\{03\}}(t),\ g_{\{04\}}(t),\ g_{\{23\}}(t),\ g_{\{24\}}(t),\ g_{\{34\}}(t).$$

The hypothesis of the lemma is satisfied by taking

$$h = g_{\{23\}}(\varphi).g_{\{24\}}(\pi/2-\theta).g_{\{02\}}\left(\mathrm{arccosh}\left[\frac{\cos\chi + \cosh\alpha}{\cos\chi\cosh\alpha + 1}\right]\right) \qquad (3.3)$$

and hence $\Delta(x,y)$ vanishes for spacelike separation of its arguments. The solution (3.3) given for h depends on (3.2); for instance, no solution can be found for $\cos\chi\cosh\alpha + 1 = 0$, according to the fact that it describes one sheet of the light cone of Ω, where Δ is not supposed to vanish in general.

4. TRIVIAL NATURE OF BOSON FILEDS IN ANTI-DE SITTER SPACE

In analogy with the procedure followed in the preceding Sec. 3, we define anti-de Sitter space as the hyperboloid

$$\eta_{ab}x^a x^b = -1$$

in a five-dimensional space with metric given by

$$\eta_{ab} = \mathrm{diag}(1,1,-1,-1,-1)$$

We choose the parametrization

$$z^0 = \sinh\alpha\cos\beta$$

$$z^1 = \sinh\alpha\sin\beta$$

$$z^2 = \cosh\alpha\sin\theta\cos\varphi$$

$$z^3 = \cosh\alpha\sin\theta\sin\varphi$$

$$z^4 = \cosh\alpha\cos\theta,$$

with $\alpha \in \mathbb{R}$, $\theta, \chi \in [0,\pi)$, $\beta, \varphi \in [0,2\pi)$.

The group of isometries is $SO(3,2)$. The Killing vectors are again of the form

$$\xi^a_{\{bc\}} = \delta_b{}^a x_c - \delta_c{}^a x_b.$$

Take $\Omega = (0,0,0,0,1)$. The isotropy group H_Ω is generated by

$$g_{\{01\}}(t),\ g_{\{02\}}(t),\ g_{\{03\}}(t),\ g_{\{12\}}(t),\ g_{\{13\}}(t),\ g_{\{23\}}(t).$$

The generic point z is given by $z = g.\Omega$ with

$$g = g_{\{01\}}(-\beta).g_{\{23\}}(\varphi-\pi/2).g_{\{34\}}(-\theta).g_{\{04\}}(-\alpha)$$

We find that for *all* points $z \in M$, we can find $h \in H_\Omega$ such that $z = g.\Omega = h.g^{-1}\Omega$. There are three separate cases:

(i) $\cos\theta\cosh\alpha + 1 < 0$
h is then given by

$$h = g_{\{01\}}(-\beta).g_{\{23\}}(\varphi-\pi/2).g_{\{03\}}\left(\mathrm{arccosh}\left[-\frac{\cosh\alpha+\cos\theta}{\cos\theta\cosh\alpha+1}\right]\right).$$

(ii) $\cos\theta\cosh\alpha + 1 > 0$

$$h = g_{\{01\}}(\pi-\beta).g_{\{23\}}(\varphi+\pi/2).g_{\{03\}}\left(\mathrm{arccosh}\left[\frac{\cosh\alpha+\cos\theta}{\cos\theta\cosh\alpha+1}\right]\right).$$

(iii) $\cos\theta\cosh\alpha + 1 = 0$

$$h = g_{\{01\}}(-\beta).g_{\{23\}}(\varphi+\pi/2).g_{\{03\}}\left(\mathrm{arccosh}\left[-\frac{1+\cos^2\theta}{2\cos\theta}\right]\right).$$

We conclude that $\Delta(\Omega,z) = 0$ for any z, and therefore $\Delta(x,y) = 0$ for *any* pair of points x and y by group invariance. This conclusion is also valid for any antisymmetric, invariant function, and as a consequence the commutator of any boson field (not necessarily a scalar field) vanishes identically, leading to a trivial theory. This result should not be surprising, as there exist in this space closed timelike curves, which spoil causality.[4]

[1]F.J. Chinea, Ph.D. thesis, Syracuse University, 1977 (unpublished).
[2]J. Leray, *Hyperbolic Differential Equations*, Institute for Advanced Study mimeographed notes (Princeton, New Jersey, 1955).
[3]K. Maurin, *General Eigenfunction Expansions and Unitary Representations of Topological Groups* (PWN, Warszawa, 1968).
[4]S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space—Time* (Cambridge U.P., Cambridge, 1973).

Coherent states and lattice sums[a)]

M. Boon

Institute for Theoretical Physics, University of Nijmegen, The Netherlands

J. Zak

Department of Physics, Technion–Israel Institute of Technology, Haifa, Israel
(Received 20 February 1978)

The expansion of harmonic oscillator states in discrete coherent states on a von Neumann lattice leads to relationships between lattice sums and expansion coefficients of the Weierstrass σ function. It is shown that these relationships can be generalized to arbitrary lattices. Some interesting identities are obtained between infinite sums of different convergence rates.

I. INTRODUCTION: THE VON NEUMANN LATTICE

There has recently been revived interest in the von Neumann discrete set of coherent states.[1–4]. This set is defined as follows. Assume that $|\alpha\rangle$ is a normalized coherent state[5]

$$\hat{a}|\alpha\rangle=\alpha|\alpha\rangle, \tag{1}$$

where $\hat{a}$ is the annihilation operator

$$\hat{a}=\frac{1}{\lambda\sqrt{2}}\left(\hat{x}+i\frac{\lambda^2}{\hbar}\hat{p}\right) \tag{2}$$

and α is the eigenvalue

$$\alpha=\frac{1}{\lambda\sqrt{2}}\left(\bar{x}+\frac{\lambda^2}{\hbar}\bar{p}\right). \tag{3}$$

Here $\hat{x}$ and $\hat{p}$ are the coordinate and momentum operators, and $\bar{x}$ and $\bar{p}$ are their expectation values in the state $|\alpha\rangle$. The constant λ ($\lambda^2=\hbar/m\omega$) is associated with the harmonic oscillator for which the ground state is the coherent state $|0\rangle$ ($\alpha=0$).

The von Neumann discrete set of coherent states $|\alpha_{mn}\rangle$ is obtained by defining a lattice of points ($m,n=0,\pm1,\dots$),

$$\alpha_{mn}=\frac{1}{\lambda\sqrt{2}}\left(ma+i\frac{2\pi}{a}\lambda^2 n\right) \tag{4}$$

in the α plane, where a is an arbitrary constant and the unit cell area is π. The states $|\alpha_{mn}\rangle$ with α_{mn} in (4) form the von Neumann set. This set was shown to be complete[1–3] and it can be used as a basis in which an arbitrary vector $|f\rangle$ can be expanded.[2,4] For carrying out such expansions a biorthogonal set of states $\{|\tilde{\alpha}_{mn}\rangle\}$ was defined

$$\langle\tilde{\alpha}_{m'n'}|\alpha_{mn}\rangle=\delta_{m'm}\delta_{n'n}, \tag{5}$$

where the states $|\alpha_{00}\rangle$ and $\langle\tilde{\alpha}_{00}|$ are excluded. This latter exclusion is necessary because the von Neumann set $|\alpha_{mn}\rangle$ is overcomplete by just one state,[2,3] that is, if one member is removed from the set, the rest are still complete, but this is not true if two are removed. Without losing generality, one can remove the state $|\alpha_{00}\rangle$. In Refs. 2 and 4 it was shown that any state $|f\rangle$ can be expanded in a series

$$|f\rangle\sim\sum_{m,n}{}'|\alpha_{mn}\rangle\langle\tilde{\alpha}_{mn}|f\rangle, \tag{6}$$

where the prime excludes the state $|\alpha_{00}\rangle$. For a coherent state $|\alpha\rangle$ the coefficients $\langle\tilde{\alpha}_{mn}|\alpha\rangle$ are[2]

$$\langle\tilde{\alpha}_{mn}|\alpha\rangle=(-1)^{m+n+mn}\exp(-\tfrac{1}{2}|\alpha|^2)\frac{\alpha_{mn}\sigma(\alpha)}{\alpha(\alpha-\alpha_{mn})}\times\exp(-\nu\alpha^2). \tag{7}$$

Here $\sigma(\alpha)$ is the Weierstrass σ function[6]

$$\sigma(\alpha)=\alpha\prod_{m,n}{}'\left(1-\frac{\alpha}{\alpha_{mn}}\right)\exp\left(\frac{\alpha}{\alpha_{mn}}+\frac{\alpha^2}{2\alpha_{mn}^2}\right), \tag{8}$$

where again the prime excludes $\alpha_{00}=0$. The constant ν in (7) is expressed by means of the ζ function in the following general way[2] (the notations in Ref. 2 coincide with the ones used in the literature[6] after replacing $\omega_1\to2\omega_1^*$, $\omega_2\to2\omega_2^*$)

$$\nu=\frac{i}{S}\left(\zeta(\omega_1)\omega_2^*-\zeta(\omega_2)\omega_1^*\right). \tag{9}$$

The definition in (9) is given for a general lattice

$$\alpha_{mn}=2m\omega_1+2n\omega_2 \tag{10}$$

with $\mathrm{Im}(\omega_1/\omega_2)\neq0$. S in (9) is the unit cell area which for the von Neumann lattice equals π.

In Ref. 4, the harmonic oscillator states $|N\rangle$ were expanded in the set $|\alpha_{mn}\rangle$ and the coefficients $\langle\alpha_{mn}|N\rangle$ were found to be

$$\langle\tilde{\alpha}_{mn}|N\rangle=-(-1)^{m+n+mn}\frac{(N!)^{1/2}}{(\alpha_{mn})^N}\sum_{p=0}^{[N/2]}a_{2p}\alpha_{mn}^{2p}, \tag{11}$$

where $[N/2]$ equals $N/2$ for even N and $N/2-\frac{1}{2}$ for odd N, and a_p are the expansion coefficients of the even function

$$\frac{\sigma(\alpha)\exp(-\nu\alpha^2)}{\alpha}=\sum_{p=0}^{\infty}a_{2p}\alpha^{2p}, \tag{12}$$

$$a_p=\frac{1}{p!}\frac{d}{d\alpha^p}\left[\frac{\sigma(\alpha)\exp(-\nu\alpha^2)}{\alpha}\right]\Big|_{\alpha=0} \tag{12a}$$

In particular, it was shown in Ref. 4 that the following closure relation holds

$$\sum_{m,n}{}'\langle M|\alpha_{mn}\rangle\langle\tilde{\alpha}_{mn}|N\rangle=\delta_{MN}, \tag{13}$$

where $\langle\tilde{\alpha}_{mn}|N\rangle$ is given in (11) and the well-known coefficient $\langle M|\alpha_{mn}\rangle$ for the harmonic oscillator is[4,5]

$$\langle M|\alpha_{mn}\rangle=\exp(-\tfrac{1}{2}|\alpha_{mn}|^2)\frac{(\alpha_{mn})^M}{(M!)^{1/2}}. \tag{14}$$

[a)] Supported by a grant from the United States—Israel Binational Science Foundation (B.S.F.), Jerusalem, Israel.

It was noticed in Ref. 4 that the closure relation (13) together with the result (11) and the known expression (14) lead to some interesting identities. First by putting $N=0$ and $M=p$ in (13), (11), and (14) one finds

$$S_p^{(v)} \equiv \sum_{m,n}{}' (-1)^{m+n+mn} \alpha_{mn}^p \exp(-\tfrac{1}{2}|\alpha_{mn}|^2) = -\delta_{p0} \quad (p \geq 0), \qquad (15)$$

where the superscript v stands for von Neumann lattice. This identity holds for any positive p (for odd p, S_p is trivially zero). More generally, the relation (13) leads to the following additional relationships (see Ref. 4)

$$\sum_{p=0}^{N} a_{2p} S_{2p-2N}^{(v)} = -\delta_{N0} \quad \text{(for any } N\text{)}, \qquad (16)$$

where, in general, for any p

$$S_p^{(v)} = \sum_{m,n}{}' (-1)^{m+n+mn} \alpha_{mn}^p \exp(-\tfrac{1}{2}|\alpha_{mn}|^2). \qquad (17)$$

The identities in (15) and (16) are of some very general nature. While the identities in (15) were shown (see Ref. 2) to lead to connections between Θ functions, those in (16) give relationships between the expansion coefficients a_{2p} in (12a) and the sums (17). On the other hand, both (15) and (16) follow from the closure relation (13), which holds for the von Neumann lattice when the unit cell area $S=\pi$. Because of their general nature the validity of (15) and (16) could not be restricted to the von Neumann lattice and it should be possible to derive them for an arbitrary lattice.

In this paper a derivation of the identities (15) and (16) is given for a general lattice (10) and some of their consequences are discussed in more detail. In particular, some interesting relations are obtained between sums of quite different convergence rates.

II. GENERAL LATTICE

An alternative simple derivation exists for the identities (15) and (16) on a general lattice (10). First we notice that Eqs. (9) and (12) hold for a general lattice. In addition, there is a well-known theorem for analytic functions with only simple poles that equate the function to an expansion in terms of the poles and residues (see Ref. 6, p. 134). From this theorem the following expansion is obtained for the inverse of the function in (12) [the residues can be found according to formula (26) of Ref. 2],

$$\frac{\alpha^{M+1}}{\sigma(\alpha)\exp(-\nu\alpha^2)} = \sum_{mn}{}' (-1)^{m+n+mn} \frac{\alpha_{mn}^{M+1}}{\alpha - \alpha_{mn}} \times \exp\left(-\frac{\pi}{2S}|\alpha_{mn}|^2\right) \qquad (18)$$

for any $M \geq 0$. This expansion can also be rewritten as follows for $\alpha < |\alpha_{mn}|$,

$$\frac{\alpha^{M+1}}{\sigma(\alpha)\exp(-\nu\alpha^2)} = -\sum_{p=0}^{\infty} S_{M-2p}\alpha^{2p}, \qquad (19)$$

where the sums S_{M-2p} are simple generalizations of (17)

$$S_{M-2p} = \sum_{mn}{}' (-1)^{m+n+mn} \frac{\alpha_{mn}^M}{(\alpha_{mn})^{2p}} \exp\left(-\frac{\pi}{2S}|\alpha_{mn}|^2\right) \qquad (20)$$

(the superscript is omitted).

Let us now show that the information contained in Eqs. (12) and (18)–(20) is sufficient for rederiving the identities (15) and (16) for general lattices (10). First, the left-hand side of (18) equals δ_{M0} for $\alpha=0$. By comparing (18) for $\alpha=0$ with the definition in (20) we find the generalized relationship for (15),

$$S_M \equiv \sum{}' (-1)^{m+n+mn} \alpha_{mn}^M \exp\left(-\frac{\pi}{2S}|\alpha_{mn}|^2\right) = -\delta_{M0}. \qquad (15a)$$

For the von Neumann lattice ($S=\pi$), Eq. (15a) goes over into (15).

It is also easy to rederive Eqs. (16). For this let us multiply both sides of (19) for $M=0$ by the corresponding sides of Eq. (12). Since the left-hand sides of these two relations are the inverse of one another one arrives at the following result,

$$\sum_{p,p'=0}^{\infty} a_{2p} S_{-2p'} \alpha^{2(p+p')} = -1. \qquad (21)$$

By putting $p+p'=N$ and by keeping in mind that $S_{2p}=0$ for $p>0$ [See Eq. (15a)] we find from (21)

$$\sum_{p=0}^{N} a_{2p} S_{2p-2N} = -\delta_{N0}. \qquad (16a)$$

These relationships are clearly a generalization of (16) for the general lattice (10). For $N=0$, (16a) agrees with (15a) for $M=0$. For $N \neq 0$, the Eqs. (16a) give connecting formulas between the expansion coefficients a_{2p} of the function in (12) and the sums S_{2p-2N} as defined in (20). For later reference let us write down explicit expressions of the Eqs. (16a) for $N=1,2,3$, and 4:

$$a_2 = S_{-2}, \qquad (22)$$

$$a_4 = S_{-2}^2 + S_{-4}, \qquad (23)$$

$$a_6 = S_{-2}^3 + 2S_{-2}S_{-4} + S_{-6}, \qquad (24)$$

$$a_8 = S_{-2}^4 + 3S_{-2}^2 S_{-4} + 2S_{-2}S_{-6} + S_{-4}^2 + S_{-8}. \qquad (25)$$

It can now be shown that the relationships (22)–(25) [and in general (16a)] establish interesting connections between sums of apparently quite different nature. On one hand, the S_{-2p} are given by the infinite sums (20). On the other hand, the coefficients a_{2p} are in a simple way connected with the coefficients in the power series of $\sigma(\alpha)/\alpha$ (Ref. 7, p. 635),

$$\frac{\sigma(\alpha)}{\alpha} = 1 + a\alpha^4 + b\alpha^6 + c\alpha^8 + \cdots \qquad (26)$$

with

$$a = -\frac{g_2}{2^4 \cdot 3 \cdot 5}, \quad b = -\frac{g_3}{2^3 \cdot 3 \cdot 5 \cdot 7}, \quad c = -\frac{g_2^2}{2^9 \cdot 3^2 \cdot 5 \cdot 7}, \qquad (27)$$

and

$$g_2 = 60 \sum_{mn}{}' \frac{1}{\alpha_{mn}^4}, \qquad (28)$$

$$g_3 = 140 \sum_{mn}{}' \frac{1}{\alpha_{mn}^6}. \qquad (29)$$

From the Eqs. (26)–(29) and (12) it follows that the expansion coefficients a_{2p} can be expressed in the following way:

$$a_2 = -\nu, \qquad (30)$$

$$a_4 = -\tfrac{1}{4}\sum_{mn}{}' \frac{1}{\alpha_{mn}^4} + \tfrac{1}{2}\nu^2, \quad (31)$$

$$a_6 = -\tfrac{1}{6}\sum_{mn}{}' \frac{1}{\alpha_{mn}^6} + \frac{\nu}{4}\sum_{mn}{}' \frac{1}{\alpha_{mn}^4} - \tfrac{1}{6}\nu^3, \quad (32)$$

$$a_8 = -\frac{5}{224}\left(\sum_{mn}{}' \frac{1}{\alpha_{mn}^4}\right)^2 + \frac{\nu}{6}\sum_{mn}{}' \frac{1}{\alpha_{mn}^6} - \frac{\nu^2}{8}\sum_{mn}{}' \frac{1}{\alpha_{mn}^4} + \tfrac{1}{24}\nu^4, \quad (33)$$

where the prime in the sums excludes the term with $m = n = 0$. And finally, by comparing Eqs. (22)—(25) with (30)—(33) we find the following relations:

$$-\tfrac{1}{4}\sum_{mn}{}' \frac{1}{\alpha_{mn}^4} = \tfrac{1}{2}(S_{-2})^2 + S_{-4}, \quad (34)$$

$$-\tfrac{1}{6}\sum_{mn}{}' \frac{1}{\alpha_{mn}^6} = \tfrac{1}{3}(S_{-2})^3 + S_{-2}S_{-4} + S_{-6}, \quad (35)$$

$$S_{-8} = -\tfrac{13}{28}S_{-2}^4 - \tfrac{13}{7}S_{-2}^2 S_{-4} - S_{-2}S_{-6} - \tfrac{19}{14}S_{-4}^2. \quad (36)$$

Let us consider in more detail the results in (34)—(36). The first two lines express relations between sums of different convergence rates. While the sums on the left-hand side in (34), (35) converge rather slowly, those on the right hand contain the factor $\exp[-(\pi/2S)|\alpha_{mn}|^2]$ and converge, in general, much faster. The additional relation [Eq. (36)] connects sums S_{-2p} with different indices. In fact, by using Eqs. (16a) and the information on the coefficients a_{2p} one can express any S_{-2p} with $p > 3$ by means of S_{-2}, S_{-4}, and S_{-6}. (36) is the corresponding expressions for S_{-8}.

The Eqs. (34)—(36) simplify considerably in the case of a square lattice

$$\alpha_{mn} = 2\omega(m + in). \quad (37)$$

(The von Neumann square lattice corresponds to $\omega = \sqrt{\pi}/2$.) In the case of a square lattice, $\nu = 0$. This follows from the following considerations. First, from (9) we have

$$\nu = \frac{i\omega^*}{S}[\zeta(\omega) - \zeta(i\omega)].$$

Also, for $\omega = 1$, $\zeta(1) = \pi/4$ and $\zeta(i) = -i\pi/4$ (Ref. 7, p. 680), and $\zeta(\omega) = \omega\zeta(1)$, $\zeta(i\omega) = \omega\zeta(i)$ (Ref. 7, p. 631). Therefore, $\nu = 0$. As a consequence of this, the expansion coefficient a_{2p} in 12 will become

$$a_{2p} = \frac{1}{(2p)!}\left[\frac{d^{2p}}{d\alpha^{2p}}\left(\frac{\sigma(\alpha)}{\alpha}\right)\right]_{\alpha=0}. \quad (38)$$

The sums in (20) will assume the form

$$S_{-2p} = \frac{1}{(2\omega)^{2p}}\sum_{m,n}{}' (-1)^{m+n+mn}\frac{\exp[-(\pi/2)(m^2+n^2)]}{(m+in)^{2p}}. \quad (39)$$

From (39) it follows that $S_{-2} = 0$ [this also follows from Eqs. (22), (30), and $\nu = 0$). One can also check that $S_{-6} = 0$ in (39) [this is in agreement with (24) because for a square lattice $g_3 = 0$; See Ref. 7, p. 629]. The Eqs. (34) and (36) will turn into the following equalities [(35) has zeros on both sides]:

$$-\tfrac{1}{4}\sum_{m,n}{}' \frac{1}{(m+in)^4} = \sum_{m,n}{}' (-1)^{m+n+mn}\frac{\exp[-(\pi/2)(m^2+n^2)]}{(m+in)^4}, \quad (40)$$

$$\sum_{m,n}{}' (-1)^{m+n+mn}\frac{\exp[-(\pi/2)(m^2+n^2)]}{(m+in)^8} = -\tfrac{19}{14}\left(\sum_{m,n}{}' (-1)^{m+n+mn}\frac{\exp[-(\pi/2)(m^2+n^2)]}{(m+in)^4}\right)^2. \quad (41)$$

In (40) we have an equality of two infinite sums of very different convergence rates. Because of the factor $\exp[-(\pi/2)(m^2+n^2)]$ on the right-hand side of (40) the latter converges much faster than the sum on the left-hand side. The sum $\sum'_{m,n}[1/(m+in)^4]$ is in a simple way connected with g_2 [see Eq. (28)] which is tabulated (Ref. 7, p. 680) and is known up to ten digits. The right-hand side of (40) was calculated on a computer[8] and the number -0.787803005384747 was found. Up to ten known digits it coincides with the left-hand side of (40)!

Equation (41) is a consequence of (40) and the known recursion relations (Ref. 7. p. 636) for the expansion coefficients of $\sigma(\alpha)/\alpha$. More generally, (34) and (35) are new relations between infinite sums of different convergence rates while Eq. (36) is a consequence of the latter and the above recursion relations between the expansion coefficients of $\sigma(\alpha)/\alpha$.

Another relation that is worth mentioning is the one that follows from (9), (22), and (30),

$$\frac{i}{S}(\zeta(\omega_1)\omega_2^* - \zeta(\omega_2)\omega_1^*) = -S_{-2}. \quad (42)$$

Together with Lagrange relation (Ref. 6, p. 446)

$$\zeta(\omega_1)\omega_2 - \zeta(\omega_2)\omega_1 = \frac{i\pi}{2} \quad (43)$$

we find

$$\zeta(\omega_1)\left(\omega_2^* - \frac{\omega_2\omega_1^*}{\omega_1}\right) = iSS_{-2} - i\frac{\pi}{2}\frac{\omega_1^*}{\omega_1}, \quad (44)$$

$$\zeta(\omega_2)\left(\frac{\omega_1\omega_2^*}{\omega_2} - \omega_1^*\right) = iSS_{-2} - i\frac{\pi}{2}\frac{\omega_2^*}{\omega_2}.$$

These relations lead to an alternative way of finding the values of ζ functions at half periods by means of the sums S_{-2} [see Eq. (20)].

In conclusion, we would like to point out that the lattice sums considered in this paper [Eq. (20)] are of similar nature to the ones that were discussed in detail in a series of papers by Glasser.[9] These sums have recently attracted much attention[10] and they appear in various physical applications.[11–14] One final point: Equation (21) is perfectly symmetric under exchange of the S symbols and the a symbols, so that one can express the S_{-2p} in terms of the a_{2p} by making the substitution $S_{-2p} \longleftrightarrow a_{2p}$ in (22)—(25). Thus we can, if we wish, adopt the viewpoint that the new sums S_{-2p} are evaluated in terms of known quantities from elliptic function theory, in the spirit of Zucker's approach.

ACKNOWLEDGMENT

The authors would like to thank Professors H. Bacry and A. Grossmann at C.N.R.S. in Marseille for useful and profitable discussions.

[1]V. Bargmann, P. Butera, L. Girardello, and J.R. Klauder, Rep. Math. Phys. **2**, 221 (1971).
[2]A.M. Perelomov, Theor. Math. Phys. **6**, 156 (1971).
[3]H. Bacry, A. Grossmann, and J. Zak, Phys. Rev. B **12**, 1118 (1975).
[4]M. Boon and J. Zak, Phys. Rev. D **15** (1978).
[5]J.R. Klauder and E.C.G. Sudarshan, *Fundamentals of Quantum Optics* (Princeton U.P., Princeton, New Jersey, 1955).
[6]E.T. Whittaker and G.N. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, 1950).
[7]M. Abramowitz and J. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).
[8]The authors would like to thank Prof. A. Peres at the Physics Department of the Technion for verifying that only a few terms contribute to this sum.
[9]M.L. Glasser, J. Math. Phys. **14**, 409, 701 (1973); **15**, 188 (1974).
[10]A.N. Chaba and R.K. Pathria, J. Phys. A **10**, 1823 (1977).
[11]S. Greenspoon and R.K. Pathria, Phys. Rev. A **9**, 2103 (1974).
[12]A.L. Fetter, P.C. Hohenberg, and P. Pincus, Phys. Rev. **147**, 140 (1966).
[13]F.E. Harris and H.J. Monkhorst, Phys. Rev. B **2**, 4400 (1970).
[14]I.J. Zucker, J. Phys. A **7**, 1568 (1974).

The Bäcklund problem for the equation $\partial^2 z/\partial x^1 \partial x^2 = f(z)$[a]

W. F. Shadwick

Department of Mathematics, King's College, Strand, London WC2R 2LS

(Received 6 February 1978)

The Bäcklund problem for the equation $\partial^2 z/\partial x^1 \partial x^2 = f(z)$ is discussed for analytic functions f, using the procedure of Estabrook and Wahlquist, starting from a Lagrangian formulation. The condition $d^2 f/dz^2 = kf$, k constant, necessary for the existence of nontrivial Bäcklund maps when the space of new dependent variables is $\mathbb{R}$, is shown to be closely related to the structure of the Lie algebra SL(2,$\mathbb{R}$).

1. INTRODUCTION

The problem of determining Bäcklund maps[1] for the equation

$$\frac{\partial^2 z}{\partial x^1 \partial x^2} = f(z) \tag{1.1}$$

where f is an analytic function of z is discussed in this paper using the procedure of Estabrook and Wahlquist.[2] The starting point is a differential ideal of 2-forms naturally associated to (1.1) through the Lagrangian

$$L = \frac{\partial z}{\partial x^1}\frac{\partial z}{\partial x^2} + 2F(z) \tag{1.2}$$

where $dF/dz = f$, from which (1.1) is derived.

In the special case where the space of new dependent variables is $\mathbb{R}$, the well-known condition

$$\frac{d^2 f}{dz^2} = kf \tag{1.3}$$

where k is a constant,[3] appears as a condition for nontriviality of the Bäcklund map. For functions f which satisfy (1.3) it is shown that the equations given by the Estabrook–Wahlquist procedure can be integrated to give not only the underlying Lie algebra, which is SL(2, $\mathbb{R}$), but also the representation of the algebra appropriate to the function f. The usual form of the Bäcklund map is obtained in each case.

In the general case, the space of new dependent variables is $\mathbb{R}^n$ and the Bäcklund problem[1] is reduced to the problem of finding representations of an infinite-dimensional Lie algebra associated to the function f. Solutions to this problem in terms of finite-dimensional Lie algebras are discussed. It is shown that the condition (1.3) is closely related to the structure of SL(2, $\mathbb{R}$). The problem of determining the class of functions which may be associated to other finite-dimensional Lie algebras in this way is briefly considered.

2. THE GENERAL FORMALISM

The first part of this section comprises a summary of the relevant definitions and notation from the theory of jet bundles.[4] In the second part it is shown that the solutions of (1.1) can be characterized by a differential ideal of 2-forms naturally associated to the Lagrangian (1.2).

Let $C^\infty(\mathbb{R}^2, \mathbb{R}^n)$ denote the set of C^∞ maps from $\mathbb{R}^2$ to $\mathbb{R}^n$ and let $J^k(\mathbb{R}^2, \mathbb{R}^n)$ denote the k-jet bundle of these maps. For $k > l$ let π_l^k denote the natural projection

$$\pi_l^k : J^k(\mathbb{R}^2, \mathbb{R}^n) \to J^l(\mathbb{R}^2, \mathbb{R}^n).$$

The 0-jet bundle $J^0(\mathbb{R}^2, \mathbb{R}^n)$ is identified with $\mathbb{R}^2 \times \mathbb{R}^n$. If $x \in \mathbb{R}^2$ and $g \in C^\infty(\mathbb{R}^2, \mathbb{R}^n)$, let $j_x^k g$ and $j^k g$ denote the k-jet of g at x and the k-jet extension of g respectively. The source projection is the map

$$\alpha : J^k(\mathbb{R}^2, \mathbb{R}^n) \mapsto \mathbb{R}^2$$

by $j_x^k g \to x$.

Let x^a, $a = 1, 2$, be coordinates on $\mathbb{R}^2$ and let z and y^μ, $\mu = 1, 2, \ldots, n$, be coordinates on $\mathbb{R}$ and $\mathbb{R}^n$ respectively. The summation and range conventions on indices a, b, c and μ, ν will be used throughout. Standard coordinates on $J^k(\mathbb{R}^2, \mathbb{R})$ and $J^k(\mathbb{R}^2, \mathbb{R}^n)$ are $x^a, z, z_a, \ldots, z_{a_1 \cdots a_k}$ and $x^a, y^\mu, y^\mu_a, \ldots, y^\mu_{a_1 \cdots a_k}$, where if f is any function in the equivalence class $j_x^k f$,

$$z = f(x), \quad z_a = \frac{\partial f}{\partial x^a}(x), \ldots, z_{a_1 \cdots a_k} = \frac{\partial^k f(x)}{\partial x^{a_1}(x) \cdots \partial x^{a_k}}$$

and similarly for $y^\mu, y^\mu_a, \ldots, y^\mu_{a_1 \cdots a_k}$.

In these coordinates the 1-form $\theta = dz - z_a dx^a$ is by itself a basis for the contact module Ω_1 on $J^1(\mathbb{R}^2, \mathbb{R})$ and the 1-forms $\theta^\mu = dy^\mu - y^\mu_a dx^a$ comprise a basis for the contact module Ω_2 on $J^1(\mathbb{R}^2, \mathbb{R}^n)$.

Let V denote the subspace of vector fields tangent to $J^1(\mathbb{R}^2, \mathbb{R})$ such that $\alpha_* X = 0$ for $X \in V$. The vector fields $\partial/\partial z$, $\partial/\partial z_a$ form a basis for V in standard coordinates.

In the jet bundle formulation, the differential equation (1.1) is the subset Z of $J^2(\mathbb{R}^2, \mathbb{R})$ defined by $\Sigma = 0$ where Σ is the function given by $\Sigma = z_{12} - f(z)$. The Lagrangian (1.2) is regarded as a function on $J^1(\mathbb{R}^2, \mathbb{R})$ defined by

$$L = z_1 z_2 + 2F(z). \tag{2.1}$$

A solution of (1.1) is a map $g \in C^\infty(\mathbb{R}^2, \mathbb{R})$ such that

$$\Sigma \circ j^2 g = 0.$$

The solution of (1.1) may also be characterized by a differential ideal of 2-forms associated to L in the following way.

The Cartan form associated to L (Ref. 5) is the 2-form Θ defined by

$$\Theta := L\, dx^1 \wedge dx^2 + \frac{\partial L}{\partial z_1}\theta \wedge dx^2 - \frac{\partial L}{\partial z_2}\theta \wedge dx^1.$$

[a]Supported by the National Research Council of Canada.

 0022-2488/78/1911-2312$1.00

From (2.1) it follows that

$$\Theta = (2F(z) - z_1 z_2)\, dx^1 \wedge dx^2 + z_2\, dz \wedge dx^2 - z_1\, dz \wedge dx^1.$$

Let I be the module of 2-forms generated by $\{X \lrcorner d\Theta \,|\, X \in V\}$. This module has a basis consisting of τ^1, τ^2, and τ^3 where

$$\tau^1 := \frac{\partial}{\partial z_1} \lrcorner\, d\Theta = -\, dz \wedge dx^2 + z_1\, dx^1 \wedge dx^2,$$
$$\tau^2 := \frac{\partial}{\partial z_2} \lrcorner\, d\Theta = dz \wedge dx^1 + z_2\, dx^1 \wedge dx^2, \tag{2.2}$$
$$\tau^3 := \frac{\partial}{\partial z} \lrcorner\, d\Theta = dz_1 \wedge dx^1 - dz_2 \wedge dx^2 + 2f(z)\, dx^1 \wedge dx^2.$$

It is easily verified that I is a differential ideal. The ideal characterizes the Euler–Lagrange equation (1.1) for $\not{L}$, in that a map $g \in C^\infty(\mathbb{R}^2, \mathbb{R})$ is a solution of (1.1) if and only if $(j^1 g)^*(I) = 0$.[5]

3. THE BÄCKLUND PROBLEM

In this section, the Bäcklund problem[1] for Z is formulated and the procedure invented by Estabrook and Wahlquist[2] for solving this problem is briefly outlined.

Let ψ be a C^∞ map from $J^1(\mathbb{R}^2, \mathbb{R}) \times \mathbb{R}^n$ to $J^1(\mathbb{R}^2, \mathbb{R}^n)$ such that diagram (3.1) commutes

$$\begin{array}{ccc} J^1(\mathbb{R}^2, \mathbb{R}) \times \mathbb{R}^n & \xrightarrow{\psi} & J^1(\mathbb{R}^2, \mathbb{R}^n) \\ {\scriptstyle \alpha \times \mathrm{id}_{\mathbb{R}^n}} \downarrow & \swarrow {\scriptstyle \pi_0^1} & \\ \mathbb{R}^2 \times \mathbb{R}^n & & \end{array} \tag{3.1}$$

In this case ψ acts trivially on $\mathbb{R}^2$ and $\mathbb{R}^n$ and the map is completely determined once the coordinates y_a^μ are given as functions of x^b, z, z_b, and y^ν, say

$$y_a^\mu = \psi_a^\mu(x^b, z, z_b, y^\nu).$$

If g and h are maps $g \in C^\infty(\mathbb{R}^2, \mathbb{R})$ and $h \in C^\infty(\mathbb{R}^2, \mathbb{R}^n)$ respectively, consider the following maps from $\mathbb{R}^2$ to $J^1(\mathbb{R}^2, \mathbb{R}^n)$:

$$j^1 h : \mathbb{R}^2 \to J^1(\mathbb{R}^2, \mathbb{R}^n) \quad \text{by} \quad x \mapsto j^1 h,$$

and $\tilde{h} := \psi \circ (j^1 g \times h) \circ \Delta : \mathbb{R}^2 \to J^1(\mathbb{R}^2, \mathbb{R}^n)$ by $x \mapsto \psi(j^1 g, h(x))$ where Δ is the diagonal map $\Delta(x) = (x, x)$. From diagram (3.1) it follows that $j^1 h$ and $\tilde{h}$ agree iff they give the same prescription for y_a^μ, that is, iff

$$\frac{\partial h^\mu}{\partial x^a} = \psi_a^\mu\left(x^b, g(x), \frac{\partial g}{\partial x^b}, h^\nu(x)\right). \tag{3.2}$$

Equation (3.2) will be satisfied only if the map ψ satisfies an integrability condition. In local coordinates this condition can be written conveniently in terms of the vector fields $\tilde{\partial}_a$ defined on $J^2(\mathbb{R}^2, \mathbb{R}) \times \mathbb{R}^n$ by

$$\tilde{\partial}_a := \frac{\partial}{\partial x^a} + z_a \frac{\partial}{\partial z} + z_{ab} \frac{\partial}{\partial z_b} + \psi_a^\nu \frac{\partial}{\partial y^\nu}.$$

The integrability condition is[1]

$$\tilde{\partial}_{[a} \psi_{b]}^\mu = 0, \tag{3.3}$$

where square brackets denote antisymmetrization.

The map ψ is called an ordinary Bäcklund map for Z if the integrability condition (3.3) is satisfied on the subset $\tilde{Z} = Z \times \mathbb{R}^n \subset J^2(\mathbb{R}^2, \mathbb{R}) \times \mathbb{R}^n$. The problem of finding such maps is called the Bäcklund problem for Z.

In the case where the functions ψ_a^μ depend only on the y^ν, the map will be called trivial.

If $\psi : J^1(\mathbb{R}^2, \mathbb{R}) \times \mathbb{R}^n \to J^1(\mathbb{R}^2, \mathbb{R}^n)$ is given, a map $\psi^1 : J^2(\mathbb{R}^2, \mathbb{R}) \times \mathbb{R}^n \to J^2(\mathbb{R}^2, \mathbb{R}^n)$ called the first prolongation of ψ can be defined.[1] This map is required to be such that diagram (3.4) commutes.

$$\begin{array}{ccc} J^2(\mathbb{R}^2, \mathbb{R}) \times \mathbb{R}^n & \xrightarrow{\psi^1} & J^2(\mathbb{R}^2, \mathbb{R}^n) \\ {\scriptstyle \pi_1^2 \times \mathrm{id}_{\mathbb{R}^n}} \downarrow & & \downarrow {\scriptstyle \pi_1^2} \\ J^1(\mathbb{R}^2, \mathbb{R}) \times \mathbb{R}^n & \xrightarrow{\psi} & J^1(\mathbb{R}^2, \mathbb{R}^n) \end{array} \tag{3.4}$$

It remains only to specify the coordinates y_{ab}^μ and the appropriate choice is[1]

$$y_{ab}^\mu = \tilde{\partial}_{(a} \psi_{b)}^\mu,$$

where round brackets denote symmetrization. Successive prolongations ψ^s of ψ can be defined inductively. If for some s, the image of ψ^s restricted to $\tilde{Z}$ is a differential equation $Z' \subset J^{s-1}(\mathbb{R}^2, \mathbb{R}^n)$, the correspondence between Z and Z' is called the Bäcklund transformation determined by ψ.[1]

A very efficient technique for finding solutions to the Bäcklund problem has been developed by Estabrook and Wahlquist. This technique, which they call prolongation will be referred to as the "Estabrook–Wahlquist procedure" in the following and the word prolongation will be reserved for the process described above. Their procedure, as it applies here, may be described as follows.

Let ψ be a map, $\psi : J^1(\mathbb{R}^2, \mathbb{R}) \times \mathbb{R}^n \to J^1(\mathbb{R}^2, \mathbb{R}^n)$. The induced map of forms ψ^* pulls back the contact module Ω_2 on $J^1(\mathbb{R}^2, \mathbb{R}^n)$ to $\psi^*(\Omega_2)$ on $J^1(\mathbb{R}^2, \mathbb{R}) \times \mathbb{R}^n$. The 1-forms $\psi^* \theta^\mu = dy^\mu - \psi_a^\mu dx^a$ form a basis for $\psi^*(\Omega_2)$. Let I be the differential ideal with basis τ_1, τ_2, τ_3 given by (2.2) and denote by I' the module generated by I together with the 1-forms $\psi^* \theta^\mu$. The requirement that I' be a differential ideal imposes on the function ψ_a^μ a system of partial differential equations. These equations are sufficient to ensure that the integrability condition (3.3) for ψ is satisfied on $\tilde{Z}$. Thus a solution of this system of equations provides a solution of the Bäcklund problem for Z.

4. THE ONE-DIMENSIONAL CASE

In this section, the system of partial differential equations for the functions ψ_a^μ is derived. The equations are solved in the special case of C^∞ maps ψ, $\psi : J^1(\mathbb{R}^2, \mathbb{R}) \times \mathbb{R}^n \to J^1(\mathbb{R}^2, \mathbb{R}^n)$ with $n = 1$.[6] The general case where $n \geq 1$ is left to section 5. For simplicity, only maps which have the translational invariance of (1.1) are considered; thus $\partial \psi_a^\mu / \partial x^b = 0$ is assumed.

Since I is a differential ideal, I' will be a differential ideal iff there exist functions f_i^μ and 1-forms η_ν^μ such that

$$d(\psi^* \theta^\mu) = \sum_{i=1}^{3} f_i^\mu \tau^i + \eta_\nu^\mu \theta^\nu. \tag{4.1}$$

It follows from (4.1) that ψ_a^μ must satisfy

$$\frac{\partial \psi_a^\mu}{\partial z_b} = 0, \quad a \neq b, \qquad \frac{\partial \psi_1^\mu}{\partial z_1} + \frac{\partial \psi_2^\mu}{\partial z_2} = 0, \tag{4.2}$$

and

$$\psi_2^\nu \frac{\partial\psi_1^\mu}{\partial y^\nu} - \psi_1^\nu \frac{\partial\psi_2^\mu}{\partial y^\nu} = - z_2 \frac{\partial\psi_1^\mu}{\partial z} + z_1 \frac{\partial\psi_2^\mu}{\partial z} + f(z)\left(\frac{\partial\psi_2^\mu}{\partial z_2} - \frac{\partial\psi_1^\mu}{\partial z_1}\right).$$

It is easily verified that (4.2) is sufficient to ensure that the functions ψ_a^μ satisfy (3.3) when $z_{12} = f(z)$. Partial integration of (4.2) gives

$$\psi_1^\mu = A^\mu(z, y^\nu) + z_1 B^\mu(y^\nu) \tag{4.3}$$

and

$$\psi_2^\mu = C^\mu(z, y^\nu) - z_2 B^\mu(y^\nu),$$

where

$$B^\nu \frac{\partial A^\mu}{\partial y^\nu} - A^\nu \frac{\partial B^\mu}{\partial y^\nu} = \frac{\partial A^\mu}{\partial z}, \tag{4.4a}$$

$$C^\nu \frac{\partial B^\mu}{\partial y^\nu} - B^\nu \frac{\partial C^\mu}{\partial y^\nu} = \frac{\partial C^\mu}{\partial z}, \tag{4.4b}$$

and

$$A^\nu \frac{\partial C^\mu}{\partial y^\nu} - C^\nu \frac{\partial A^\mu}{\partial y^\nu} = 2f(z)B^\mu. \tag{4.4c}$$

Now specialize to the case of maps from $J^1(\mathbb{R}^2, \mathbb{R}) \times \mathbb{R}$ to $J^1(\mathbb{R}^2, \mathbb{R})$. Without loss of generality, $B^1(y^1) \neq 0$ since otherwise it follows from (4.3) and (4.4) that the functions ψ_a^1 would depend only on y^1 and the Bäcklund maps would be trivial. Let y be defined by $dy = dy^1/B^1$ and let A and C be defined by $A := A^1/B^1$ and $C := C^1/B^1$. From (4.4) it follows that A and C must satisfy

$$\frac{\partial A}{\partial y} = \frac{\partial A}{\partial z}, \tag{4.5a}$$

$$\frac{\partial C}{\partial y} = -\frac{\partial C}{\partial z}, \tag{4.5b}$$

and

$$A\frac{\partial C}{\partial y} - C\frac{\partial A}{\partial y} = 2f(z). \tag{4.5c}$$

If coordinates $u := z + y$ and $v := z - y$ are used it follows easily that

$$\frac{d^2A}{du^2} = \tfrac{1}{4} kA, \tag{4.6a}$$

$$\frac{d^2C}{dv^2} = \tfrac{1}{4} kC \tag{4.6b}$$

and

$$\frac{d^2f}{dz^2} = kf, \tag{4.6c}$$

where k is a constant. The condition (4.6c) has been derived in various ways.[3] It will be discussed further in section 5.

If a function f satisfying (4.6c) is given, then (4.6a) and (4.6b) can be integrated to give A and C as functions of y as well as of z. All solutions are of the form

$$A = a_0(z) + a_1(z)Y_1(y) + a_2(z)Y_2(y),$$

$$C = c_0(z) + c_1(z)Y_1(y) + c_2(z)Y_2(y),$$

where the functions Y_1 and Y_2 are such that d/dy, $Y_1 d/dy$, $Y_2 d/dy$ is a representation of the Lie algebra $SL(2, \mathbb{R})$, characterized by the constant k. This is in contrast to the usual situation which arises in the use of the Estabrook–Wahlquist procedure[2,7] in that

(1) the representation as well as the algebraic structure is determined.

(2) the structure constants of the algebra contain no "eigenvalue" parameter; the parameter (see Table I) arises as a constant of integration in A and C.

The Bäcklund map is given by

$$y_1 = \psi_1 = A(z + y) + z_1, \quad y_2 = \psi_2 = C(z - y) - z_2, \tag{4.7}$$

where $\psi_a := \psi_a^1/B^1$. The first prolongation of ψ is defined by $y_{ab} = \tilde{\partial}_{(a}\psi_{b)}$ and thus, in particular,

$$y_{12} = \tilde{\partial}_{(1}\psi_{2)} = \frac{1}{2}\left(A\frac{\partial C}{\partial y} + C\frac{\partial A}{\partial y}\right). \tag{4.8}$$

It follows from (4.5c) and (4.8) that

$$y_{12} = g(y)$$

where $d^2g/dy^2 = kg$. If Z' is the subset of $J^2(\mathbb{R}^2, \mathbb{R})$ defined by $y_{12} = g(y)$, then ψ determines a Bäcklund transformation between Z and Z'.

In Table I a summary of the results is given for various solutions of (4.6c). The usual form of the Bäcklund map is obtained in each case.[8]

5. THE GENERAL CASE

In this section, the general case is considered where ψ is a C^∞ map $\psi : J^1(\mathbb{R}^2, \mathbb{R}) \times \mathbb{R}^n \to J^1(\mathbb{R}^2, \mathbb{R}^n)$ and $n \geq 1$. In this case the solution of (4.4) involves an infinite-dimensional Lie algebra $\mathfrak{a}_f$ associated with the function f. The Bäcklund problem for Z is reduced to the problem of finding representations of $\mathfrak{a}_f$ or of its image under a Lie algebra homomorphism. It is shown that the existence of a homomorphism to a finite-dimensional Lie algebra $\mathfrak{a}$ requires that the function f have a certain decomposition (to be defined below) with respect to $\mathfrak{a}$.

It is convenient to define vector fields

$$A := A^\nu \frac{\partial}{\partial y^\nu}, \quad B := B^\nu \frac{\partial}{\partial y^\nu}, \quad \text{and} \quad C := C^\nu \frac{\partial}{\partial y^\nu}.$$

With these definitions (4.4) is equivalent to

$$[B, A] = \frac{\partial A}{\partial z}, \tag{5.1a}$$

$$[C, B] = \frac{\partial C}{\partial z}, \tag{5.1b}$$

$$[A, C] = 2f(z)B, \tag{5.1c}$$

where

$$[X, Y] := [X, Y]^\nu \frac{\partial}{\partial y^\nu} = \left(X^\mu \frac{\partial Y^\nu}{\partial y^\mu} - Y^\mu \frac{\partial X^\nu}{\partial y^\mu}\right)\frac{\partial}{\partial y^\nu}.$$

The solution to (5.1a) and (5.1b) is

$$A(z, y^\nu) = \exp(zB)A_0(y^\nu)\exp(-zB), \tag{5.2a}$$

$$C(z, y^\nu) = \exp(-zB)C_0(y^\nu)\exp(zB), \tag{5.2b}$$

where

$$A_0(y^\nu) := A(0, y^\nu) \quad \text{and} \quad C_0(y^\nu) := C(0, y^\nu).$$

In order to use these solutions in (5.1c), it is convenient to rewrite (5.2) in the form

TABLE I. Backlund maps for various solutions of (4.6c). Note that in each instance the entries for $f(z)=z$ are the linearized versions of those for $f(z)=\sin z$ and $f(z)=\sinh z$.

f	k	Defining equation for Z	Bäcklund map	Defining equation for Z'	Representation of SL(2, ℝ)
$\sin z$	-1	$z_{12}=\sin z$	$y_1=2a\sin\left(\frac{y+z}{2}\right)+z_1$ $y_2=2a^{-1}\sin\left(\frac{y-z}{2}\right)-z_2$	$y_{12}=\sin y$	$2\sin(\frac{1}{2}y)\frac{d}{dy}$, $2\cos(\frac{1}{2}y)\times\frac{d}{dy}$, $\frac{d}{dy}$
z	0	$z_{12}=z$	$y_1=a(y+z)+z_1$ $y_2=a^{-1}(y-z)-z_2$	$y_{12}=y$	$y\frac{d}{dy}$, $y^2\frac{d}{dy}$, $\frac{d}{dy}$
$\sinh z$	1	$z_{12}=\sinh z$	$y_1=2a\sinh\frac{y+z}{2}+z_1$ $y_2=2a^{-1}\sinh\frac{y-z}{2}-z_2$	$y_{12}=\sinh y$	$2\sinh(\frac{1}{2}y)\frac{d}{dy}$, $2\cosh(\frac{1}{2}y)$, $\times\frac{d}{dy}$, $\frac{d}{dy}$
e^{2z}	4	$z_{12}=e^{2z}$	$y_1=a\exp(z+y)+z_1$ $y_2=a^{-1}\exp(z-y)-z_2$	$y_{12}=0$	$e^y\frac{d}{dy}$, $e^{-y}\frac{d}{dy}$, $\frac{d}{dy}$.

$$A=A_0+\sum_{n=1}^{\infty}\frac{1}{n!}z^nA_n, \tag{5.3a}$$

$$C=C_0+\sum_{n=1}^{\infty}\frac{1}{n!}z^nC_n, \tag{5.3b}$$

where

$$A_n:=\left.\frac{\partial^nA}{\partial z^n}\right|_{z=0}=ad^nB(A_0)$$

and

$$C_n:=\left.\frac{\partial^nC}{\partial z^n}\right|_{z=0}=(-1)^nad^nB(C_0).$$

The Jacobi identity together with (5.1) yields

$$\left[\frac{\partial^nA}{\partial z^n},\frac{\partial^mC}{\partial z^m}\right]=2^{1-n-m}f^{(n+m)}_{(z)}B, \tag{5.4a}$$

$$n,m=0,1,2,\ldots,$$

so that in particular

$$[A_n,C_m]=2^{1-n-m}f^{(n+m)}_{(0)}\,B, \tag{5.4b}$$

where $f^{(k)}:=d^kf/dz^k$.

Let $\mathfrak{a}_f$ denote the Lie algebra generated by A_0, B, and C_0. The multiplication table for $\mathfrak{a}_f$ is given in part by (5.4b) and more of it can be deduced by using the Jacobi identity. Note that none of the brackets $[A_n,A_m]$ or $[C_n,C_m]$ are given explicitly, although, for example, $[B,[A_n,A_m]]$ and $[C_l,[A_n,A_m]]$ can be computed.

The vector fields A and C are given in terms of vectors from $\mathfrak{a}_f$ by (5.3). This is the usual situation arising from the application of the Estabrook–Wahlquist procedure.[2,7]

If a representation of $\mathfrak{a}_f$ or of its image under a Lie algebra homomorphism can be found then the Bäcklund problem for Z will be solved. If the representation is given by

$$A_0\to\mathcal{A}_0:=\mathcal{A}_0^\mu\frac{\partial}{\partial\eta^\mu},\quad B\to\mathcal{B}:=\mathcal{B}^\mu\frac{\partial}{\partial\eta^\mu},\quad C_0\to\mathcal{C}_0:=\mathcal{C}_0\frac{\partial}{\partial\eta^\mu},$$

then the Bäcklund map is given by

$$\eta_1^\mu=\mathcal{A}^\mu(z,\eta^\nu)+z_1\mathcal{B}^\mu(\eta^\nu),$$
$$\eta_2^\mu=\mathcal{C}^\mu(z,\eta^\nu)-z_2\mathcal{B}^\mu(\eta^\nu), \tag{5.5}$$

where

$$\mathcal{A}^\mu=\mathcal{A}_0^\mu+\sum_{n=1}\frac{1}{n!}z^n(ad^n\mathcal{B}(\mathcal{A}_0))^\mu$$

and

$$\mathcal{C}^\mu=\mathcal{C}_0^\mu+\sum_{n=1}^{\infty}\frac{1}{n!}(-z)^n(ad^n\mathcal{B}(\mathcal{C}_0))^\mu.$$

In particular, if a homomorphism h from $\mathfrak{a}_f$ to a finite-dimensional Lie algebra $\mathfrak{a}$ can be found, then by Ado's theorem,[9] a faithful finite-dimensional representation ϕ of $\mathfrak{a}$ exists. Then $\phi\circ h$ will be a representation of $\mathfrak{a}_f$ and hence provide a solution to the Bäcklund problem for Z. It is this case which will be considered here.

It follows from (5.4) and (5.5) that if $h(\mathfrak{a}_f)$ is Abelian, or if any of hB, hA_0 or hC_0 is the zero of $\mathfrak{a}$, the Bäcklund map will not depend on z, z_1, or z_2 and will therefore be trivial. Only homomorphisms h giving nontrivial Bäcklund maps will be considered.

If the function f satisfies (4.6c) the results of Sec. 4 show that a homomorphism $h:\mathfrak{a}_f\to\mathrm{SL}(2,\mathbb{R})$ exists. This is illustrated for $f(z)=\sin z$ in Example 5.1 below. Conversely, if $h:\mathfrak{a}_f\to\mathrm{SL}(2,\mathbb{R})$, then by using a one-dimensional representation of $\mathrm{SL}(2,\mathbb{R})$ in which hB is represented by $b(y)\,d/dy$ with $b(y)\neq0$, the arguments of Sec. 4 may be repeated to show that f must satisfy (4.6c) for some constant k. Thus the condition $d^2f/dz^2=kf$ is necessary and sufficient for the existence of a homomorphism $h:\mathfrak{a}_f\to\mathrm{SL}(2,\mathbb{R})$.

Example 5.1: Let

$$f(z)=\sin z=\sum_{n=0}^{\infty}\frac{z^{2n+1}}{(2n+1)!}(-1)^n.$$

From (5.4c) the multiplication table for $\mathfrak{a}_f$ is given in part by

$$[A_0, C_0] = 2f(0)B = 0,$$
$$[A_1, C_0] = [A_0, C_1] = f'(0)B = B,$$
$$[A_2, C_0] = [A_1, C_1] = [A_0, C_2] = \tfrac{1}{2}f''(0)B = 0$$

$$\vdots$$

Let Y_1, Y_2, Y_3 be a basis for SL(2, IR) with

$$[Y_1, Y_2] = -Y_3, \quad [Y_3, Y_1] = Y_2, \quad \text{and } [Y_2, Y_3] = Y_1.$$

Define h by $hA_0 = aY_1$, $hB = \frac{1}{2}Y_3$, $hC_0 = (1/a)Y_1$ and extend to a Lie algebra homomorphism. Thus for example $hA_1 = h[B, A_0] = [hB, hA_0] = (a/2)Y_2$.

Let

$$hA := hA_0 + \sum_{n=1}^{\infty} \frac{1}{n!} z^n hA_n$$

and

$$hC := hC_0 + \sum_{n=1}^{\infty} \frac{1}{n!} z^n hC_n.$$

It is easy to check that

$$hA = a\left(\cos\frac{z}{2} Y_1 + \sin\frac{z}{2} Y_2\right)$$

and

$$hC = \frac{1}{a}\left(\cos\frac{z}{2} Y_1 - \sin\frac{z}{2} Y_2\right).$$

If the representation of SL(2, IR) given for $f(z) = \sin z$ in Table I is used, the Bäcklund map is

$$y_1 = 2a \sin\frac{y+z}{2} + z_1, \qquad y_2 = 2a^{-1} \sin\frac{y-z}{2} - z_2.$$

Similarly, a homomorphism h can be found for any of the functions f in Table I by working backwards from the basis of SL(2, IR) given there.

If one associates the condition (4.6c) on the function f with SL(2, IR) it is natural to ask what conditions on f are implied by the existence of a homomorphism from $\mathfrak{a}_f$ to some other fixed finite-dimensional Lie algebra $\mathfrak{a}$. In Example 5.2 below, it is shown that if $\mathfrak{a} = \mathrm{SO}(3)$, the function f must satisfy

$$\frac{d^2 f}{dz^2} = -|k|^2 f$$

for some constant $k \neq 0$. Conversely it can be shown that for the functions $\sin|k|z$ and $\cos|k|z$ a homomorphism $h: \mathfrak{a}_f \to \mathrm{SO}(3)$ can be found.

In general the function f for which $h: \mathfrak{a}_f \to \mathfrak{a}$ exists must admit a decomposition with respect to $\mathfrak{a}$ in the following sense.

Let Y_i, $i = 1, 2, \ldots N$ be a basis for $\mathfrak{a}$ in which $hB = Y_1$ and let Γ^k_{ij} be the structure constants for this basis. Since the Y_i form a basis, hA can be rewritten as $hA = a^i(z)Y_i$ (using the summation and range conventions on indices i, j, k here and in the following). Similarly $hC = c^i(z)Y_i$. From (5.1) it follows that $\partial/\partial z\, hA = [hB, hA]$ and $\partial/\partial z\, hC = [hC, hB]$ so that

$$\frac{\partial a^i}{\partial z} = \Gamma^i_{1j} a^j$$

and

$$\frac{\partial c^i}{\partial z} = -\Gamma^i_{1j} c^j.$$

Thus

$$a^i(z) = (\exp z\Gamma_1)^i_j a^j(0)$$

and

$$c^i(z) = (\exp - z\Gamma_1)^i_j c^j(0)$$

where Γ_1 is the matrix with $(\Gamma_1)^i_j := \Gamma^i_{1j}$.

From (5.4a) it follows that the function f must have the decomposition

$$(\exp z\Gamma_1)^i_l a^l(0)(\exp - z\Gamma_1)^j_n c^n(0)\Gamma^k_{ij} = 2f(z)\delta^k_1. \tag{5.6}$$

It is clear from (5.6) that for a given algebra $\mathfrak{a}$, the class of functions may be quite small. This is illustrated in Example 5.3 where it is shown that for a certain nilpotent algebra the only possibility is that the function is constant.

Example 5.2: Let $\mathfrak{a} = \mathrm{SO}(3)$ and let $h: \mathfrak{a}_f \to \mathfrak{a}$ be the homomorphism. Let X_1, X_2, X_3 be the basis of SO(3) with

$$[X_1, X_2] = X_3, \quad [X_3, X_1] = X_2, \quad \text{and} \quad [X_2, X_3] = X_1.$$

Then $hA = a^i(z)X_i$, $hB = b^i X_i$ and $hC = c^i(z)X_i$.

It is convenient to use a vector notation with

$$\bar{a} = (a^1, a^2, a^3), \quad \bar{b} = (b^1, b^2, b^3), \quad \text{and} \quad \bar{c} = (c^1, c^2, c^3).$$

Since $\partial hA/\partial z = [hB, hA]$ it follows that

$$\frac{\partial \bar{a}}{\partial z} = \bar{b} \times \bar{a}$$

and similarly that

$$\frac{\partial \bar{c}}{\partial z} = \bar{c} \times \bar{b}.$$

From these equations it follows that

$$\bar{a} = \bar{a}_0 + \bar{a}_1 \cos|b|z + \bar{a}_2 \sin|b|z$$

and

$$\bar{c} = \bar{c}_0 + \bar{c}_1 \cos|b|z + c_2 \sin|b|z$$

where

$$\bar{b} \times \bar{a}_0 = \bar{b} \times \bar{c}_0 = 0.$$

Since

$$\left[\frac{\partial^2 hA}{\partial z}, hC\right] = \frac{1}{2} \frac{d^2 f}{dz^2} hB,$$

and

$$\bar{b} \times \bar{a}_0 = 0; \qquad \frac{d^2 f}{dz^2} = -4|b|^2 f.$$

Example 5.3: Let $\mathfrak{a}$ be the Lie algebra with basis $Y_1, \ldots, Y_5$, where $[Y_1, Y_2] = [Y_3, Y_4] = Y_5$ and all other brackets vanish. Let $hA_0 = a^i Y_i$, $hB = b^i Y_i$ and $hC_0 = c^i Y_i$. Then

$$hA_1 = (a^1 b^2 - a^2 b^1 + a^3 b^4 - a^4 b^3) Y_5 =: \hat{a} Y_5$$

so that $0 = hA_2 = hA_3 = \cdots$, and $hA = a^i Y_i + \hat{a} Y_5 z$. Similarly, $hC = c^i Y_i + \bar{c} Y_5 z$. It follows from $[hA, hC] = 2f(z)\, hE$ that $b_1 = b_2 = b_3 = b_4 = 0$, and that f is the constant function given by $f = \frac{1}{2}(b^5)^{-1}(a^1c^2 - a^2c^1 + a^3c^4 - a^4c^3)$.

ACKNOWLEDGMENT

Useful discussions with David Robinson and Felix Pirani are gratefully acknowledged.

[1]F. A. E. Pirani and D. C. Robinson, C. R. Acad. Sci. Paris **285** Ser. A, 581 (1977); F. A. E. Pirani, D. C. Robinson and W. F. Shadwick, "A local jet bundle formulation of Bäcklund transformations," preprint (1977).

[2]H. D. Wahlquist and F. B. Estabrook, J. Math. Phys. **16**, 1 (1975); F. B. Estabrook and H. D. Wahlquist, J. Math. Phys. **17**, 1293 (1976).

[3]See, for example, D. W. LcLaughlin and A. C. Scott, J. Math. Phys. **14**, 1817 (1973), J. P. Corones and F. Testa in *Bäcklund Transformations*, edited by R. Miura (Springer, Berlin, Heidelberg, New York, 1976); R. K. Dodd and R. K. Bullough, Proc. R. Soc. London Ser. A **351**, 499 (1976).

[4]C. Ehresmann, "Introduction a la théorie des structures infinitésimales et des pseudo-groupes de Lie" (Colloque international du C. N. R. S., Paris, 1953).

[5]R. Hermann, *Vector Bundles in Mathematical Physics* (Benjamin, New York, 1970), Vol. I.

[6]J. P. Corones, J. Math. Phys. **17**, 756 (1976); J. P. Corones and F. Testa, Ref. 3.

[7]H. C. Morris, J. Math. Phys. **17**, 1867 (1976); **18**, 285 (1977) (1977).

[8]See for example, D. W. McLaughlin and A. C. Scott, Ref. 3.

[9]N. Jacobson, *Lie algebras* (Interscience, New York, London, 1962), p. 202.

Path integrals with a periodic constraint: Entangled strings

Akira Inomata and Vijay A. Singh

Department of Physics, State University of New York at Albany, Albany, New York 12222
(Received 10 February 1978)

Path integrals with a periodic constraint $\int\dot{\theta}\,ds = \Theta + 2\pi n$ (n = integer) are studied. In particular, the path integral for a string entangled around a singular point in two dimensions is evaluated in polar coordinates. Applications are made for the entangled polymers with and without interactions, the Aharonov–Bohm effect, and the angular momentum projection of a spinning top.

1. INTRODUCTION

The study of path integrals in the presence of topological constraints is unquestionably important. It is known that statistical properties of polymers, such as the elasticity of rubber and the melting point of DNA, are considerably altered by entanglement of their polymeric constituents.[1] A path integral, if it describes the statistical function for an entangled polymer system, should be subjected to a constraint due to the entanglement. The soliton—soliton scattering problem may be another example for which the path integral approach is powerful. In order to select the transition between two correct momentum states, one must again insert a constraint into the path integral.[2] There are indeed a number of situations for which one has to deal with path integrals involving constraints. However, as the class of soluble path integrals is very limited, so is the class of soluble constraint problems. Therefore, any single example, if soluble, would be worth investigating. It could be a key to further extensions.

What we shall study in the present paper are path integrals with a simple periodic constraint of U(1). To visualize the constraint best, we consider an idealized flexible string which, having two fixed end points, is entangled around a singular point in two dimensions, and describe the probability function for its possible configurations in terms of the constrained path integral. This is basically the same problem as that Edwards has formulated for an ensemble of long polymer chains and solved by reducing the path integral to a differential equation.[3] However our problem differs from that of Edwards in the following two points. First, our main interest is in carrying out directly the path integration for the constrained system. Secondly, our path integral is set more general than his, so that the result may be useful to a wider class of problems. For the calculations, we utilize the polar coordinate formulation of path integrals,[4,5] thus demonstrating the usefulness of polar coordinates in the path integral evaluation.

In Sec. 2, we set up our problem in the path integral form mainly following Edwards.[3] In Sec. 3, we perform the polar path integration for the entangled string under the influence of an arbitrary central potential. The calculation in polar coordinates is not all trivial but far more advantageous than Cartesians. Section 4 is devoted to applications. As the first example, we deal with the model Edwards considered for entangled polymers and reproduce the result he obtained in a different manner. An immediate extension of the above is the second example which treats the path integral for extangled polymers with an interaction of the form $V(r) = ar^2 + b/r^2$. A possible application to the Aharonov—Bohm effect[6] is discussed as the third example; flux and charge quantizations are also considered in this context. The final example is a spinning top (a two-dimensional rotator) which involves the nonperiodic radial constraint as well as the periodic angular constraints.

2. PATH INTEGRAL FOR AN ENTANGLED STRING

In order to make it easier to visualize the topological constraint, we consider an idealized string commencing at $\mathbf{r}'$ and terminating at $\mathbf{r}''$ by stipulating that a possible configuration of the string is given by a path of the random walk from $\mathbf{r}'$ to $\mathbf{r}''$. The probability that such a random walk is completed in a time period τ is

$$P(\mathbf{r}'',\mathbf{r}';\tau) = (4\pi D\tau)^{-1}\exp[-(\mathbf{r}''-\mathbf{r}')^2/(4D\tau)], \tag{2.1}$$

where D is the diffusion coefficient. If the random motion takes place with an average speed v, the total length of the path will be $\sigma = v\tau$. Thus, by using this in (2.1), we obtain the probability of finding the string of length σ in a configuration that one end of it is at $\mathbf{r}'$ and the other at $\mathbf{r}''$.

$$P(\mathbf{r}'',\mathbf{r}';\sigma) = (\pi l\sigma)^{-1}\exp[-(\mathbf{r}''-\mathbf{r}')^2/(l\sigma)], \tag{2.2}$$

where $l = 4D/v$ is a constant having the dimension of length. In the Wiener representation, we have

$$P(\mathbf{r}'',\mathbf{r}';\sigma) = A\int \exp\left[-\frac{1}{l}\int_0^\sigma \dot{\mathbf{r}}^2(s)\,ds\right]\mathcal{D}\mathbf{r}, \tag{2.3}$$

where the integral is taken over all paths $\mathbf{r}(s)$ such that $\mathbf{r}(0) = \mathbf{r}'$ and $\mathbf{r}(\sigma) = \mathbf{r}''$, and the normalization factor is to be so chosen that

$$\int P(\mathbf{r}'',\mathbf{r}';\sigma)\,d\mathbf{r}'' = 1. \tag{2.4}$$

As usual, one can convert the probability function (2.3) into the propagator for a free particle of mass μ in quantum mechanics by replacement,

$$\sigma \to \tau, \quad l \to 2i\hbar/\mu, \tag{2.5}$$

and into the density matrix in statistical mechanics by

$$\sigma \to \beta, \quad l \to 2\hbar^2/\mu, \tag{2.6}$$

where $\beta = 1/(kT)$. The above consideration is applicable in a three-dimensional Euclidean space, but henceforth we shall confine ourselves to a two-dimensional plane where $\mathbf{r} = (r,\theta)$.

 0022-2488/78/1911-2318$1.00

To impose a periodic constraint on the string, we place a singular point in the plane. Since the space becomes doubly connected in the presence of the singularity, the configuration of the string extending from $\mathbf{r}'$ to $\mathbf{r}''$ without encircling the singular point is homotopically inequivalent to that in which the string encircles the singular point. The topologically different configurations can be classified by the number of turns around the singularity. Let the singular point be our coordinate origin and let Θ $(0 \leq \Theta \leq 2\pi)$ be the angle $\cos^{-1}(\mathbf{r}' \cdot \mathbf{r}'')$. Then we have

$$\int_0^{\sigma} \dot{\theta}\, ds = \Theta + 2\pi n, \tag{2.7}$$

where $n = 0, \pm 1, \pm 2, \ldots, n$ indicating the number of turns (n turns counterclockwise if $n \geq 0$ and $n+1$ turns clockwise if $n \leq -1$) around the singular point. The configuration with no entanglement corresponds to $n=0$ or $n=-1$.

Now we incorporate into the probability function (2.3) the constraint

$$\int \dot{\theta}\, ds = \phi \tag{2.8}$$

by writing theprobability function as

$$P_{\phi}(\mathbf{r}'', \mathbf{r}'; \sigma) = A \int \delta\left(\phi - \int_0^{\sigma} \dot{\theta}\, ds\right) \exp\left[-\frac{1}{l}\int_0^{\sigma} \dot{\mathbf{r}}^2\, ds\right] \mathcal{D}\mathbf{r}. \tag{2.9}$$

The constraint (2.8) selects a set of topologically equivalent configurations. Apparently the constrained probability function (2.9) satisfies the condition

$$P(\mathbf{r}'', \mathbf{r}'; \sigma) = \int_{-\infty}^{\infty} P_{\phi}\, d\phi. \tag{2.10}$$

Using the relation,

$$2\pi\delta(x) = \int_{-\infty}^{\infty} \exp(i\lambda x)\, d\lambda, \tag{2.11}$$

valid for any real number x, we express P_{ϕ} as

$$P_{\phi}(\mathbf{r}'', \mathbf{r}'; \sigma) = (2\pi)^{-1}\int_{-\infty}^{\infty} P_{\lambda}(\mathbf{r}'', \mathbf{r}'; \sigma) \exp(i\lambda\phi)\, d\lambda \tag{2.12}$$

with a path integral.

$$P_{\lambda}(\mathbf{r}'', \mathbf{r}'; \sigma) = A \int \exp\left[-\frac{1}{l}\int_0^{\sigma} (\dot{\mathbf{r}}^2 + i\lambda l \dot{\theta})\, ds\right] \mathcal{D}\mathbf{r}. \tag{2.13}$$

Thus, we are led to evaluate the path integral (2.13).

In comparison with the Feynman path integral,[7] the path integral (2.13) is seen to carry an effective Lagrangian of the form,

$$L = \tfrac{1}{2}\dot{\mathbf{r}} - U(\mathbf{r}, \dot{\mathbf{r}}), \tag{2.14}$$

with an effective scalar potential,

$$U(\mathbf{r}, \dot{\mathbf{r}}) = -\tfrac{1}{2} i\lambda l \dot{\theta}. \tag{2.15}$$

Edwards[3] has written (2.14) in an alternative way,

$$L = \tfrac{1}{2}\dot{\mathbf{r}}^2 + \tfrac{1}{4} i(\lambda l / g)\mathbf{A} \cdot \dot{\mathbf{r}}, \tag{2.16}$$

with an effective vector potential,

$$\mathbf{A} = \tfrac{1}{2} g(y\mathbf{i} - x\mathbf{j})/(x^2 + y^2), \tag{2.17}$$

where g is a magnetic charge. The latter form is very similar in structure to the Lagrangian for a charged particle of a unit mass $\mu = 1$ in a magnetic potential. By analogy with the Schrödinger equation for a charged particle, one finds a differential equation for $P_{\lambda}(\mathbf{r}, \mathbf{r}'; \sigma)$ of (2.13) to obey

$$\left[\frac{\partial}{\partial s} - \frac{l}{4}\left(\nabla - i\frac{\lambda}{2g}\mathbf{A}\right)^2\right] P_{\lambda}(\mathbf{r}, \mathbf{r}'; s - s') = \delta(\mathbf{r} - \mathbf{r}')\delta(s - s'). \tag{2.18}$$

Solving this equation in polar coordinates, Edwards[3] has obtained the constrained probability P_{ϕ} of (2.12). In the next section, we evaluate the path integral (2.13) directly without resort to the differential equation (2.18).

3. EVALUATION OF THE CONSTRAINED PATH INTEGRAL

To evaluate the path integral (2.13), we first write it in the standard time-division form,[7]

$$P_{\lambda}(\mathbf{r}'', \mathbf{r}'; \sigma) = \lim_{N\to\infty} A_N \int \exp\left[-\frac{2}{l}\sum_{j=1}^{N} S(\mathbf{r}_j, \mathbf{r}_{j-1})\right] \prod_{j=1}^{N-1} d\mathbf{r}_j, \tag{3.1}$$

where $\mathbf{r}_j = \mathbf{r}(s_j)$, $\mathbf{r}_0 = \mathbf{r}'$, $\mathbf{r}_N = \mathbf{r}''$, $s_j - s_{j-1} = \sigma/N$ and A_N is the normalization factor in the Nth approximation. The effective Lagrangian (2.14) may be used to define the action S in (3.1). However, in order to make the problem slightly more general, we modify the effective potential (2.15) in (2.14) as

$$U(\mathbf{r}, \dot{\mathbf{r}}) = -\tfrac{1}{2} i\lambda l \dot{\theta} + V(r) \tag{3.2}$$

by allowing the influence of a central force from the singular point at the origin on each small segment of the string. Then we attempt to carry out the integrations in (3.1) in polar coordinates.

The partial action in a small interval $\Delta s_j = s_j - s_{j-1} = \epsilon$ is approximated by

$$S(\mathbf{r}_j, \mathbf{r}_{j-1}) = \tfrac{1}{2}(\Delta\mathbf{r}_j)^2/\epsilon - \epsilon U(\Delta\theta_j/\epsilon, r_j), \tag{3.3}$$

where $\Delta\mathbf{r}_j = \mathbf{r}_j - \mathbf{r}_{j-1}$ and $\Delta\theta_j = \theta_j - \theta_{j-1}$. In polar coordinates,

$$S(\mathbf{r}_j, \mathbf{r}_{j-1}) = \tfrac{1}{2}(r_j^2 + r_{j-1}^2)/\epsilon - (r_j r_{j-1}/\epsilon)\cos(\theta_j - \theta_{j-1}) + \tfrac{1}{2} i\lambda l(\theta_j - \theta_{j-1}) - \epsilon V(r_j), \tag{3.4}$$

where (3.2) is explicitly used. In order to take account of contributions up to first order in ϵ, we utilize the following approximation formulas for small ϵ.

$$\cos(\Delta\theta) \approx \cos(\Delta\theta + a\epsilon) + a\epsilon\Delta\theta + \tfrac{1}{2}a^2\epsilon^2, \tag{3.5}$$

and

$$\exp\left(\frac{u}{\epsilon}\cos\theta\right) \approx \left(\frac{\epsilon}{2\pi u}\right)^{1/2} \sum_{m=-\infty}^{\infty} \exp\left(im\theta + \frac{u}{\epsilon} - \tfrac{1}{2}(m^2 - \tfrac{1}{4})\frac{\epsilon}{u}\right). \tag{3.6}$$

The last expression follows from the expansion formula,

$$\exp\left(\frac{u}{\epsilon}\cos\theta\right) = \sum_{m=-\infty}^{\infty} \exp(im\theta)\, I_m\left(\frac{u}{\epsilon}\right), \tag{3.7}$$

and the asymptotic form of the modified Bessel function for small ϵ,[8]

$$I_m\left(\frac{u}{\epsilon}\right) \approx \left(\frac{\epsilon}{2\pi u}\right)^{1/2} \exp\left(\frac{u}{\epsilon} - \tfrac{1}{2}(m^2 - \tfrac{1}{4})\frac{\epsilon}{u} + O(\epsilon^2)\right), \tag{3.8}$$

which is valid for $|\arg(u/\epsilon)| < \pi/2$. The exponential of (3.4) can be computed as

$$\exp\left(-\frac{2}{l}S(\mathbf{r}_j, \mathbf{r}_{j-1})\right) = \exp\left(-\frac{1}{\epsilon l}(r_j^2 + r_{j-1}^2) + \frac{2\epsilon}{l}V(r_j)\right) \exp\left[\frac{2r_j r_{j-1}}{l\epsilon} \times \cos\left(\theta_j - \theta_{j-1} + \frac{i\lambda l\epsilon}{2r_j r_{j-1}}\right) - \frac{\lambda^2 l\epsilon}{4r_j r_{j-1}}\right]$$

$$= \left(\frac{l\epsilon}{4\pi r_j r_{j-1}}\right)^{1/2} \exp\left[-\frac{1}{l\epsilon}(r_j^2 + r_{j-1}^2) + \frac{2\epsilon}{l}V(r_j)\right]$$
$$\times \sum_{m=-\infty}^{\infty} \exp\left(im(\theta_j - \theta_{j-1}) + \frac{2r_j r_{j-1}}{l\epsilon} - [(m+\lambda)^2 - \tfrac{1}{4}]\right.$$
$$\left.\times \frac{l\epsilon}{4r_j r_{j-1}}\right).$$

Use of (3.8) in this result enables us to put the integrand of (3.1) in the form

$$\exp\left(-\frac{2}{l}\sum_{j=1}^{N} S(\mathbf{r}_j, \mathbf{r}_{j-1})\right)$$
$$= \prod_{j=1}^{N}\left(\sum_{m_j=-\infty}^{\infty} \exp[im_j(\theta_j - \theta_{j-1})] R_{m_j+\lambda}(r_j, r_{j-1})\right), \qquad (3.9)$$

where

$$R_{\nu_j}(r_j\ r_{j-1}) = \exp[-(r_j^2 + r_{j-1}^2)/(l\epsilon)$$
$$+ (2\epsilon/l)V(r_j)] I_{|\nu_j|}(2r_j r_{j-1}/l\epsilon). \qquad (3.10)$$

After interchanging the multiplications and summations on the right-hand side of (3.9), we substitute the result into (3.1) to obtain

$$P_\lambda(\mathbf{r}'', \mathbf{r}'; \sigma) = \lim_{N\to\infty} A_N \sum_{m_1 m_2 \cdots m_N} \int \prod_{j=1}^{N} [\exp(im_j \Delta\theta_j)$$
$$\times R_{m_j+\lambda}(r_j, r_{j-1})] \prod_{j=1}^{N-1} (r_j dr_j d\theta_j). \qquad (3.11)$$

The angular integrations in (3.11) can easily be performed if the following orthogonality relation is employed,

$$\int_0^{2\pi} \exp[i(m' - m)\theta]\, d\theta = 2\pi\delta_{m'm}. \qquad (3.12)$$

Namely we get for the angular integrations

$$\int \prod_{j=1}^{N} \exp[im_j \Delta\theta_j] \prod_{j=1}^{N-1} d\theta_j = (2\pi)^{N-1} \prod_{j=1}^{N-1} \delta_{m_N m_j} \exp[im_N(\theta'' - \theta')]. \qquad (3.13)$$

Consequently the probability function (3.11) takes on the form

$$P_\lambda(\mathbf{r}'', \mathbf{r}'; \sigma) = \sum_{m=-\infty}^{\infty} \exp[im(\theta'' - \theta')] Q_{m+\lambda}(r'', r'; \sigma) \qquad (3.14)$$

with the radial probability function,

$$Q_\lambda(r'', r'; \sigma) = \lim_{N\to\infty} (2\pi)^{N-1} A_N \int \prod_{j=1}^{N} R_\lambda(r_j, r_{j-1}) \prod_{j=1}^{N-1} (r_j dr_j). \qquad (3.15)$$

which remains to be evaluated, contingent on specification of the interaction potential $V(r)$.

Now we turn ourselves to the constrained probability function P_ϕ of (2.9) which is given, upon substitution of (3.14), by

$$P_\phi = \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \exp[im(\theta'' - \theta') + i\lambda\phi] Q_{m+\lambda} d\lambda. \qquad (3.16)$$

Changing the variable λ to $\lambda - m$ yields

$$P_\phi = \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \exp[im(\theta'' - \theta' - \phi) + i\lambda\phi] Q_\lambda\, d\lambda. \qquad (3.17)$$

Use of the identity,

$$\sum_{m=-\infty}^{\infty} \exp(im\theta) = 2\pi \sum_{n=-\infty}^{\infty} \delta(\theta + 2\pi n), \qquad (3.18)$$

further reduces (3.17) to the form

$$P_\phi(\mathbf{r}'', \mathbf{r}'; \sigma) = \sum_n \delta(\theta'' - \theta' - \phi + 2\pi n)$$
$$\times \int_{-\infty}^{\infty} \exp(i\lambda\phi) Q_\lambda(r'', r'; \sigma)\, d\lambda. \qquad (3.19)$$

The delta function in this expression selects out the probability function for a particular set of topologically equivalent configurations of the string with $\phi = \theta'' - \theta' + 2\pi n$; namely,

$$P_n(r'', \theta''; r', \theta'; \sigma) = \int_{-\infty}^{\infty} \exp[i\lambda(\theta'' - \theta' + 2\pi n)]$$
$$\times Q_\lambda(r'', r'; \sigma)\, d\lambda. \qquad (3.20)$$

The probability that the string can take any configuration is given by

$$P(\mathbf{r}'', \mathbf{r}'; \sigma) = \int_{-\infty}^{\infty} P_\phi\, d\phi = \sum_n P_n \qquad (3.21)$$

which is written, with the aid of the Poisson formula, as

$$P(\mathbf{r}'', \mathbf{r}'; \sigma) = \sum_{m=-\infty}^{\infty} \exp[im(\theta'' - \theta')] Q_m(r'', r'; \sigma). \qquad (3.22)$$

4. APPLICATIONS

A. Entangled polymers without interactions

The result obtained in the previous section is of course immediately applicable to the Edwards model for the entanglement of polymer molecules in which no interactions between molecule chains are assumed. We identify the string with a flexible long molecule chain, and interprete the constant l as the size of each segmental molecule. The effective Lagrangian is

$$L = \tfrac{1}{2}(\dot{r}^2 + r^2\dot{\theta}^2) + \tfrac{1}{2} i\lambda l\dot{\theta}, \qquad (4.1)$$

and the radial function (3.10) for this system is

$$R_\lambda(r_j, r_{j-1}) = \exp[-(r_j^2 + r_{j-1}^2)/l\epsilon] I_{|\lambda|}(2r_j r_{j-1}/l\epsilon), \qquad (4.2)$$

with which we calculate the path integral,

$$Q_\lambda(r'', r') = \lim_{N\to\infty} (2\pi)^{N-1} A_N \exp[-(r'^2 + r''^2)/l\epsilon]$$
$$\times \int \exp[-2(r_1^2 + r_2^2 + \cdots + r_{N-1}^2)/l\epsilon$$
$$\times I_\lambda(2r_0 r_1/l\epsilon) \cdots I_\lambda(2r_{N-1} r_N/l\epsilon) \prod_{j=1}^{N-1} (r_j dr_j). \qquad (4.3)$$

In a previous paper,[5] the following has been derived,

$$\int_0^\infty \exp\left(i\alpha \sum_{j=1}^{N-1} r_j^2\right) \prod_{j=1}^{N} I_\nu(-i\beta r_{j-1} r_j) \prod_{j=1}^{N-1} (r_j dr_j)$$
$$= \prod_{j=1}^{N-1} (i/2\alpha_j) \exp\left\{-i\left[r_0^2 \sum_{j=1}^{N-1} (\beta_j^2/4\alpha_j) + r_N^2\beta^2/\alpha_N\right]\right\}$$
$$\times I_\nu(-i\beta_N r_0 r_N), \qquad (4.4)$$

valid for $\mathrm{Re}(\nu) > -1$ and $\mathrm{Re}(\alpha) > 0$. Here α_j and β_j are coefficients to be determined by solving the algebraic equations:

$$\alpha_1 = \alpha,\ \alpha_{j+1} = \alpha - \beta^2/(4\alpha_j), \quad \text{for } j \geq 1, \qquad (4.5a)$$

$$\beta_1 = \beta,\ \beta_{j+1} = \beta \prod_{k=1}^{j} (\beta/2\alpha_k), \quad \text{for } j \geq 1. \qquad (5.4b)$$

The multi-integral formula (4.4) enables us to write (4.3) as

$$Q_\lambda(r'', r'; \sigma) = \lim_{N\to\infty} (2\pi)^{N-1} A_N a_N \exp(if_N r_0^2 + ig_N r_N^2)$$
$$\times I_\lambda(-ia_N r_0 r_N), \qquad (4.6)$$

where

$$a_N = \prod_{j=1}^{N-1} (i/2\alpha_j). \tag{4.7a}$$

$$f_N = \tfrac{1}{2}\beta - \sum_{j=1}^{N-1} (\beta_j^2/4\alpha_j), \tag{4.7b}$$

$$g_N = \tfrac{1}{2}\beta - \beta^2/(4\alpha_N). \tag{4.7c}$$

For $\alpha=\beta$, (4.7a) and (4.7b) yield, respectively,

$$\alpha_j = \tfrac{1}{2}\alpha(j+1)/j, \tag{4.8a}$$

$$\beta_j = \alpha/j. \tag{4.8b}$$

Accordingly, with $\alpha = 2i/(l\epsilon)$ we find for (4.8)

$$a_N = 2(l\epsilon/2)^N (l\sigma)^{-1}, \tag{4.9a}$$

$$f_N = i(l\sigma)^{-1}, \tag{4.9b}$$

$$g_N = i(l\sigma)^{-1}. \tag{4.9c}$$

Thus we arrive at a simple expression,

$$Q_\lambda(r'', r';\sigma) = (\pi l\sigma)^{-1} \exp[-(r'^2 + r''^2)/l\sigma] I_\lambda(2r'r''/l\sigma), \tag{4.10}$$

where we have used the normalization factor $A_N = (\pi l\epsilon)^{-N}$. Substitution into (3.20) yields

$$P_n(r'', \theta''; r', \theta';\sigma) = (\pi l\sigma)^{-1} \exp[-(r'^2 + r''^2)/l\sigma] \times \int_{-\infty}^{\infty} \exp[i\lambda(\theta'' - \theta' + 2\pi n)] \times I_\lambda(2r'r''/l\sigma)\, d\lambda \tag{4.11}$$

which is identical with the result Edwards[3] obtained from the differential equation (2.18). The total probability function calculated by using (3.7) and (4.10) in (3.22) is

$$P(\mathbf{r}'', \mathbf{r}';\sigma) = (\pi l\sigma)^{-1} \exp[-(\mathbf{r}'' - \mathbf{r}')^2/l\sigma], \tag{4.12}$$

which coincides with (2.2) as expected.

B. Entangled polymers with interactions

To be more realistic with the polymer model, we would have to consider an interaction between two molecule chains entangled with each other. Such an interaction must include both long range attraction and short range repulsion. A reasonable choice for the potential is

$$V(r) = ar^2 + b/r^2, \tag{4.13}$$

where a and b are positive constants. Since the radial function (3.10) is expressible with the help of (3.8) as

$$R_\lambda(r_j, r_{j-1}) = \left(\frac{l\epsilon}{4\pi r_j r_{j-1}}\right)^{1/2} \exp\left[-\frac{(r_j - r_{j-1})^2}{l\epsilon} - (\lambda^2 - \tfrac{1}{4}) \times \frac{l\epsilon}{4r_j r_{j-1}} + \frac{2\epsilon}{l} V(r_j)\right], \tag{4.14}$$

the radial path integral (3.15) can be put in the form

$$Q_\lambda(r'', r';\sigma) = (\pi l)^{-1} \int \exp\left\{-\frac{2}{l}\int \left[\frac{1}{2}\dot{r}^2 - \frac{l^2}{8r^2}\left(\lambda^2 - \frac{1}{4}\right) - V(r)\right] ds\right\} \mathcal{D}r \tag{4.15}$$

or more explicitly,

$$Q_\lambda(r'', r';\sigma) = (\pi l)^{-1} \int \exp\left\{-\frac{2}{l}\int \left[\frac{1}{2}\dot{r}^2 - \frac{l^2}{8r^2}\left(\nu^2 - \frac{1}{4}\right) - ar^2\right] ds\right\} \mathcal{D}r \tag{4.16}$$

with $\nu = (\lambda^2 + 8b/l^2)^{1/2}$. In an earlier paper,[5] the following formula has been derived,[9]

$$\int \exp\left\{\frac{i}{\hbar}\int \left[\frac{1}{2}\mu\dot{r}^2 - \frac{1}{2\mu r^2}\left(\nu^2 - \frac{1}{4}\right) - \frac{1}{2}\mu\omega^2 r^2\right] dt\right\} \mathcal{D}r = -i(r'r'')^{1/2}\mu\omega \csc(\omega\tau) \exp[\tfrac{1}{2}(i\mu\omega/\hbar)(r'^2 + r''^2) \times \cot(\omega\tau)] I_\nu[-i(\mu\omega r'r''/\hbar)\csc(\omega\tau)], \tag{4.17}$$

for $\mathrm{Re}(\nu) > -1$. Replacing $i\mu \to -2/l$, $\mu \to 1$, and $\omega^2 \to 2a$ in this formula gives us the correct form of the right-hand side of (4.16). Inserting this result in (3.20), we get[10]

$$P_n(r'', \theta'', r', \theta';\sigma) = \int_{-\infty}^{\infty} (\pi l)^{-1}\sqrt{2a}\csc(\sqrt{2a}\sigma) \times \exp[-(\sqrt{2a}/l)(r'^2 + r''^2)\cot(\sqrt{2a}\sigma)] \times I_{\nu(\lambda)}[2(\sqrt{2a}/l)r'r''\csc(\sqrt{2a}\,\sigma)] \times \exp[i\lambda(\theta'' - \theta' + 2\pi n)]\, d\lambda, \tag{4.18}$$

where $\nu(\lambda) = (\lambda^2 + 8b/l^2)^{1/2}$.

For the case of the harmonic potential $V(r) = ar^2$, we put $b=0$ in (4.18) to get

$$P_n = \int_{-\infty}^{\infty} (\pi l)^{-1}\sqrt{2a}\csc(\sqrt{2a}\sigma)\exp[-(\sqrt{2a}/l)(r'^2 + r''^2) \times \cot(\sqrt{2a}\sigma)] I_\lambda[2(\sqrt{2a}/l)r'r''\csc(\sqrt{2a}\,\sigma)] \times \exp[i\lambda(\theta'' - \theta' + 2\pi n)]d\lambda. \tag{4.19}$$

For the case of the inverse-square potential $V(r) = b/r^2$, we set $a=0$ to obtain

$$P_n = \int_{-\infty}^{\infty} (\pi l\sigma)^{-1} \exp[-(r'^2 + r''^2)/(l\sigma)] I_{\nu(\lambda)}[2r'r''/l\sigma] \times \exp[i\lambda(\theta'' - \theta' + 2\pi n)]d\lambda. \tag{4.20}$$

The statistical properties of entangled polymers described by (4.18) and subsequent specialized cases will be discussed elsewhere.

C. The Aharonov-Bohm effect, flux and charge quantizations

Edwards[3] has exploited, in reducing the constrained path integral (2.13) to a different equation, the important fact that the constraint introduced in (2.13) behaves like a vector potential $\mathbf{A}$ of (2.17). An additional interesting fact is that the magnetic field counterpart of the potential vanishes,

$$\mathbf{B} = \nabla \times \mathbf{A} = 0. \tag{4.21}$$

If the singularity at the origin is taken as representing magnetic flux passing through the singular point and if the string is interpreted as the path of an electron, then we have a complete setup for the measurement of the Aharonov–Bohm effect.[6]

By including the reduced mass μ explicitly in both terms of (2.16) instead of setting $\mu=1$, and by making the replacements, $l \to 2i\hbar/\mu$, $\lambda \to 2eg/c\hbar$, $s \to t$, and $\sigma \to \tau$, we get

$$L = \tfrac{1}{2}\mu\dot{\mathbf{r}}^2 + \frac{e}{c}\mathbf{A}\cdot\dot{\mathbf{r}}. \tag{4.22}$$

By the same replacements, we obtain from (4.10)

$$Q_\lambda(r'', r'; \tau) = (\mu/2\pi i\hbar\tau)\exp[-(r'^2 + r''^2)\mu/2i\hbar\tau] \times \delta(\lambda - 2eg/\hbar c) I_\lambda(-ir'r''\mu/\hbar\tau). \tag{4.23}$$

Thus we can write down the constrained propagator of the electron directly from (3.19) and (3.20) as

$$K_\phi(\mathbf{r}'', \mathbf{r}'; \tau) = \sum_n \delta(\theta'' - \theta' - \phi + 2\pi n) K_n(r'', \theta''; r', \theta'; \tau), \tag{4.24}$$

where

$$K_n(r'', \theta''; r', \theta'; \tau) = (\mu/2\pi i\hbar\tau)\exp[i(r'^2 + r''^2)\mu/2\hbar\tau] \times \exp[2ieg(\theta'' - \theta' + 2\pi n)\hbar c] \times I_{(2eg/\hbar c)}[-ir'r''\mu/\hbar\tau]. \tag{4.25}$$

The nonvanishing phase difference between the propagators for two different values of n, say, $n = 0$ and $n = -1$,

$$K_0/K_{-1} = \exp(4\pi ieg/\hbar c), \tag{4.26}$$

suggests an observable effect of the vector potential $\mathbf{A}$ for which $\mathbf{B} = 0$ though.

Furthermore we notice that for $\mathbf{r}'' = \mathbf{r}'$,

$$\phi = \oint \dot{\theta}\, ds = (e/\hbar c)\oint \mathbf{A}\cdot d\mathbf{r} \tag{4.27}$$

describe the magnetic flux passing through the origin, measured in units of $\hbar c/e$. The delta function in (4.24) indicates that the flux must be quantized as[11]

$$\oint \mathbf{A}\cdot d\mathbf{r} = 2\pi n\hbar c/e \quad (n = \text{integer}). \tag{4.28}$$

The full propagator evaluated from (3.22) with (4.25) takes on the form

$$K(r'', \theta''; r', \theta'; \tau) = (\mu/2\pi i\hbar\tau)\exp[-(r'^2 + r''^2)\mu/2i\hbar\tau] \times \sum_{m=-\infty}^{\infty} \delta(m - 2eg/\hbar c) I_m(-ir'r''\mu/\hbar\tau) \times \exp[im(\theta'' - \theta')]. \tag{4.29}$$

It is remarkable that the delta function in (4.29) implies nothing but the charge quantization,[12]

$$eg = \tfrac{1}{2}m\hbar c \quad (m = \text{integer}). \tag{4.30}$$

A detailed account of this subject will be given in a forthcoming paper.

D. Spinning top

The final example is a top spinning about its symmetry axis, for which the effective Lagrangian is given by

$$L = \tfrac{1}{2}I\dot{\theta}^2 - \lambda\hbar\dot{\theta}, \tag{4.31}$$

where $I = r^2\mu$ is a constant. Imposing the radial constraint $\exp[-i\epsilon V(r_j)/\hbar] = \delta(r - r_j)$, we write the radial function (3.10) as

$$R_\lambda(r_j, r_{j-1}) = \delta(r - r_j)\exp[i\mu(r_j^2 + r_{j-1}{}^2)/2\hbar\epsilon] \times I_{|\lambda|}(-i\mu r_j r_{j-1}/\hbar\epsilon). \tag{4.32}$$

The path integral (3.15) for the radial propagator can be evaluated with the aid of the asymptotic formula (3.8) as

$$Q_\lambda(r;\tau) = \lim_{N\to\infty}(2\pi)^{N-1}A_N \int \prod_{j=1}^{N}\left\{\delta(r - r_j)\left(\frac{i\hbar\epsilon}{2\pi r_j r_{j-1}}\right)^{1/2} \times \exp\left[\frac{i(r_j - r_{j-1})^2}{2\hbar\epsilon} - \frac{i\hbar\epsilon}{2r_j r_{j-1}}\left(\lambda^2 - \frac{1}{4}\right)\right]\right\} \prod_{j=1}^{N-1}(r_j dr_j) = (4\pi^2 I)^{-1/2}\exp(-i\lambda^2\hbar\tau/2I), \tag{4.33}$$

where we have chosen the normalization factors

$$A_N = [2\pi i\hbar\epsilon\exp(-i\hbar\epsilon/4I)]^{-N/2}.$$

Substituting (4.33) into (3.19) and integrating over λ, we arrive at the constrained propagator,

$$K_\phi(\theta'', \theta'; \tau) = \sum_n \delta(\theta'' - \theta' - \phi + 2\pi n) K_n(\theta'', \theta'; \tau), \tag{4.34}$$

with the projected propagator

$$K_n(\theta'', \theta'; \tau) = (2\pi i\hbar\tau)^{-1/2}\exp[iI(\theta'' - \theta' + 2\pi n)^2/2\hbar\tau]. \tag{4.35}$$

The total propagator is also readily obtained by integrating (4.34) over ϕ as

$$K(\theta'', \theta'; \tau) = (2\pi i\hbar\tau)^{-1/2}\sum_n \exp[iI(\theta'' - \theta' + 2\pi n)^2/2\hbar\tau]. \tag{4.36}$$

Using the theta function,[13]

$$\vartheta_3\{z, \zeta\} = \sum_{m=-\infty}^{\infty}\exp(2miz + im^2\pi\zeta) \tag{4.37}$$

and its transformation formula,

$$\vartheta_3\{z, \zeta\} = (-i\zeta)^{-1/2}\exp(-iz^2/\pi\zeta)\vartheta_3\{z/\zeta, -1/\zeta\}, \tag{4.38}$$

we can also put (4.36) in the form

$$K(\theta'', \theta'; \tau) = (i/2\pi\hbar\tau)^{1/2}\exp[iI(\theta'' - \theta')^2/2\hbar\tau] \times \vartheta_3\{\pi I(\theta'' - \theta')/\hbar\tau,\ 2\pi I/\hbar\tau\}, \tag{4.39}$$

or, in the standard form,

$$K(\theta'', \theta'; \tau) = (2\pi)^{-1}\sum_{m=-\infty}^{\infty}\exp(i\hbar m^2\tau/2I) \times \exp[-im(\theta'' - \theta')]. \tag{4.40}$$

It may be instructive to point out that K_n in (4.40) is not the projected angular momentum propagator. The integral number m appearing in (4.40) is to be understood as the eigenvalue of the angular momentum M.

If we are interested in the projected angular momentum propagator, then we should insert $\delta(M - \lambda\hbar)$ into (4.33), identifying the running parameter λ with the angular momentum M divided by $\hbar$, as

$$Q_\lambda(r;\tau) = (2\pi)^{-1}\delta(M - \lambda\hbar)\exp(-i\lambda^2\hbar\tau/2I). \tag{4.41}$$

From this and (3.22) readily follows the propagator for a fixed momentum,

$$K_m(\theta'', \theta'; \tau) = (2\pi)^{-1}\sum_{m=-\infty}^{\infty}(M - m\hbar)\exp(im^2\hbar\tau/2I) \times \exp[im(\theta'' - \theta')], \tag{4.42}$$

which coincides with the one that has been considered by Callen and Gross.[2] The total propagator (4.40) can be recovered by integrating (4.42) over M, i.e., by

$$K(\theta'', \theta'; \tau) = \int K_M(\theta'', \theta'; \tau)\, dM. \tag{4.43}$$

5. DISCUSSIONS

In an effort to provide a general computational technique of the path integral with a periodic constraint of the U(1) type, we have evaluated the path integral for the string entangled around a singular point in two dimensions, and applied the results to such variegated examples as the polymer problems, the Aharonov–Bohm effect and the spinning top.

In application to the entangled polymers, we have succeeded in reproducing the result reached by Edwards[3] via the differential equation (2.18). Since our problem has been set up so as to be formally applicable to any central force potential, the evaluation of the probability junction of entangled polymers under the influence of the potential $V=ar^2+b/r^2$ is a straightforward, if not trivial, matter. Our considerations on polymers have been limited to the two-dimensional cases, but the extension to three dimensions is not too difficult. This and other related problems will be discussed elsewhere.

The path integral calculations on the Cartesian basis are severely restricted to the Gaussian class. The harmonic oscillator with $V(r)=ar^2$ is a typical soluble example. The use of polar coordinates has made it possible to treat the potential $V(r)=ar^2+b/r^2$ within the path integral framework.[5] Our calculations in the present paper show that the path integral can be performed for a system with a effective potential $U(\mathbf{r},\dot{\mathbf{r}})=ar^2+b/r^2+c\dot{\theta}$, and indicate that the use of polar coordinates is rather comprehensive in evaluating path integrals.

In general, it is not easy to handle a charge particle in a magnetic field **B** by the path integral method. For a uniform field, the path integral has been computed.[5,14] Since the periodic constraint behaves like a magnetic vector potential **A**, we have another soluble example with a vector potential. In this case, the potential is rotationless, so that the magnetic field counterpart **B** exists nowhere in the two-dimensional space but the singular point. We have exploited this particular situation for considering the Aharonov–Bohm effect. It is interesting that the flux quantization is built in the present path integral formulation. The flux quantization and the encircling of the electron path about a singular point in the multiply-connected space appear to be mutually related. It is also surprising that the formulation is consistent only if Dirac's condition for the charge quantization is satisfied. Here we have adopted the Aharonov–Bohm effect as a possible application of the constrained path integral, and put our emphasis on the computational aspect. The path integral formulation of the Aharonov–Bohm effect and its related problems would deserve further investigations. They would serve as means to study some questions of fundamental nature in quantization.

In dealing with a top rotating about its symmetry axis, we have demonstrated how the angular momentum projection can be introduced in the propagator. The number of entanglements corresponds to the number of rotations but cannot be understood as the angular quantum number. The entanglement number and the angular quantum number are in a sense complementary to each other. In our treatment of the Aharonov–Bohm effect, we have used the B-field concentrated at a point. This special distribution of the B-field may seem unrealistic but does not harm the generality of the argument. If we wish, we may replace at the price of simplicity the singular point by a finite circular extension in which $B=0$. For instance, the analysis applied to the spinning top can immediately be converted to the one appropriate for this purpose. The angular momentum quantization gives rise to the charge quantization, while the number of rotations corresponds to the quantized flux number.

ACKNOWLEDGMENTS

We wish to acknowledge the help of Christopher Gerry for checking some of the calculations. One of us (A.I.) would like to thank Professor S.F. Edwards and Professor G.J. Papadopoulos for useful discussions.

[1]J. Vinograd, J. Lebowitz, and R. Watson, J. Mol. Biol. **23**, 123 (1968); K. Freed, Adv. Chem. Phys. **22**, 1 (1972).
[2]C.G. Callan and D.J. Gross, Nucl. Phys. B **93**, 29 (1975).
[3]S.F. Edwards, Proc. Phys. Soc. London **91**, 513 (1967).
[4]S.F. Edwards and Y.V. Gulyaev, Proc. R. Soc. London A **279**, 229 (1964); S. Ozaki, Lectures at Kyushu University, 1955 (unpublished).
[5]D. Peak and A. Inomata, J. Math. Phys. **10**, 1422 (1969).
[6]Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).
[7]R.P. Feynman, Rev. Mod. Phys. **20**, 367 (1948); R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[8]Rigorously speaking, this asymptotic form does not hold for the Feynman path integral for which $|\arg z|=\pi/2$. However, it is operationally valid for all cases since the difficult can be avoided by introducing a real parameter in z to make $|\arg z|<\pi/2$ and by letting the parameter vanish after the integration.
[9]This formula has been derived in Ref. 5. See the formula (56) in Ref. 5 which erroneously contains the term $m\omega^2/r^2$ in the integrand.
[10]This result differs from the distribution function obtained from a different approach in F.W. Wiegel, Twente University of Technology Preprint (1977).
[11]See, i.e., J.J. Sakurai, *Advanced Quantum Mechanics* (Addison–Wesley, Reading, Mass., 1967), p. 16.
[12]Equation (4.30) is identical in form with Dirac's quantization condition. Although g has the dimension of a magnetic charge, there is no reason to interpret g as the magnetic charge of a single monopole. We have introduced g in (2.17) only for dimensional consistency with the magnetic potential interpretation of **A**. Our potential therefore differs dimensionally from the choice of Edwards in Ref. 3 for which $g=2$.
[13]E.T. Whittaker and G.N. Watson, *Modern Analysis* (Cambridge U.P., Cambridge, 1965), 4th ed., p. 475.
[14]M.L. Glasser, Phys. Rev. B **113**, 831 (1964); A. Inomata, Benet Laboratories Technical Report WVT-6718 (1967).

Conservation equations and the gravitational symplectic form

T. N. Palmer[a)]

Department of Astrophysics, South Parks Road, Oxford, England [b)]
(Received 22 February 1978)

By considering space–times whose metric is given by a perturbation expansion away from a background which admits a Killing vector, conservation equations, based on the energy–momentum tensor, are derived to first and second orders in the perturbation expansion. To the first order, equations are derived independently of the Einstein field equations, and describe secular changes in the energy–momentum distribution of the matter fields. To the second order, a gravitational energy–momentum contribution arises from the conservation equations which may be constructed from the symplectic inner product on the solution space of the linearized Einstein field equations. Considering a similar scheme based on the Bel–Robinson tensor, it is shown that whilst first order conservation equations can be formulated, the lack of a symplectic form for the perturbed Bel–Robinson tensor implies the nonexistence of second order conservation equations, except when the background is flat. The results are applied to perturbations of a stationary black hole, and simple expressions are found for the mass and angular momentum fluxes, through the event horizon, due to a gravitational perturbation. By considering a monochromatic wave, it is seen that the conservation of the gravitational symplectic form reduces, in suitable coordinates, to the Wronskian condition of Teukolsky and Press.

1. INTRODUCTION

Conservation laws play an important role in both classical and quantum physics, enabling one, for example, to investigate problems of interacting systems where a detailed knowledge of the interaction is either unknown, or is too complicated to have any practical value.

In Newtonian physics, the most important conserved properties of an isolated system are its mass, linear momentum, angular momentum, and energy. The first three of these quantities are multipole moments of the system. The latter, however, arises as a useful concept, purely because it satisfies a conservation equation.

The kinetic energy of a system is related to its ability to do work under the influence of some force. However, in order that energy be a conserved quantity, it must be generalized from merely being kinetic; in the case where the force derives from a potential, this is not difficult.

This process of generalization is central to this paper, where we shall examine conservation equations in general relativity.

Consider a space—time (M, g_{ab}) which admits a Killing vector k^a, and whose matter fields are described by the symmetric energy-momentum tensor, T^{ab}, which by virtue of Einstein's equation satisfies the divergence-free condition

$$\nabla_a T^{ab} = 0. \tag{1.1}$$

Using (1.1) and Killing's equation, then we have the well known integrable equation

$$\nabla_a (T^a{}_b k^b) = 0 \tag{1.2}$$

describing the conservation of the energy-momentum content of the matter fields, relative to the Killing vector.

When space—time does not admit a group of motions, then the process of generalization must include a contribution to the total energy-momentum content of the gravitational field itself. In the full nonlinear theory of general relativity, this process has proved difficult to formulate, and has led many investigators to reject the energy-momentum concept for a generic space—time. However, in a recent paper,[1] the author has shown that in fact these difficulties can be overcome, and meaningful equations for the conservation of the total energy-momentum content of a finite spacelike Euclidean 3-volume in an arbitrary space—time, can be given.

In this paper we consider a restrictive class of space—times, in order to investigate this process of generalization. Consider a perturbation of the nonempty space—time which admits a Killing vector k^a. The question we ask is whether the divergence-free condition (1.1) can be expressed as an integrable conservation equation, and if so, under what circumstances can one educe from it, for each order in the perturbation expansion, an expression for the gravitational radiation flux through an arbitrary hypersurface, and when will this flux be gauge invariant?

We shall see that to second order in the perturbation, this flux will naturally emerge in terms of the symplectic form on the solution space of the linearized Einstein equations. For this reason, in Sec. 2, we review some salient features on the existence and conservation of the symplectic form for second-order linear differential operators.

In Sec. 3 we set up a perturbation scheme, and consider first-and second-order perturbations of (1.1), showing in particular how the symplectic form naturally arises.

As an example where conservation equations cannot be formulated (due to the nonexistence of a symplectic form) we consider in Sec. 4, a similar scheme for the Bel—Robinson tensor[2,3] T_{abcd}, which in vacuum satisfies in an analogous fashion to (1.1), the restriction

[a)] Present address: Geophysical Fluid Dynamics Laboratory Meteorological Office, Bracknell, Berkshire, England.
[b)] Address for reprint requests.

 0022-2488/78/1911-2324$1.00

$$\nabla_a T^{abcd} = 0. \tag{1.3}$$

In Sec. 5 we apply the results of Sec. 3 to perturbations of a stationary black hole space—time, obtaining simple gauge invariant expressions for the energy-momentum flux density through the event horizon due to a purely gravitational perturbation, obtaining the results of Hawking and Hartle[4] as a special case. We also show that the conservation of the covariant gravitational symplectic form reduces in suitable coordinates to the Wronskian condition of Teukolsky and Press.[5]

Some concluding remarks are made in Sec. 6.

Throughout this paper we shall employ the Battelle sign and index conventions of Penrose.[6]

This work is based on part of the author's doctoral thesis at the University of Oxford.[7] Many details may be found therein.

2. THE SYMPLECTIC FORM

It is usual to obtain the energy-momentum tensor of a matter field by varying the action of the matter Lagrangian with respect to the metric tensor. In this section we shall discuss an alternative though equivalent method for constructing the energy-momentum of a linear field whose field equations are determined by a symmetric differential operator.

This technique for constructing an energy-momentum flux has been used by Ashtekhar and Magnon[8] in the context of quantization of classical matter fields in curved space—time, and by Friedman and Schutz[9] in the context of Lagranian perturbation theory of stationary fluids.

As mentioned in the introduction, our motivation for discussing it here is that it will emerges as a natural candidate for describing the energy-momentum of gravitational fluxes on curved space—time backgrounds, a regime where there is no immediate divergence-free energy—momentum tensor.

Let us start by considering a field ϕ^I, where I is a multiple index. (For example, if ϕ^I is a tensor field of valence r, then $|I| = r$.) Suppose I satisfies the field equation

$$\mathcal{D}^{IJ}\phi_J = 0 \tag{2.1}$$

(indices are raised and lowered with g^{IJ} and g_{IJ}) where $\mathcal{D}^{IJ}$ is a second-order linear differential operator. Explicitly we may put

$$\mathcal{D}^{IJ} = A^{IJab}\nabla_a\nabla_b + B^{IJa}\nabla_a + C^{IJ}. \tag{2.2}$$

To examine the conditions under which $\mathcal{D}^{IJ}$ is symmetric, we introduce the inner product

$$(\Phi^I, \Psi_I)_M = \int_M \Phi^I \Psi_I dv \tag{2.3}$$

on the space—time (M, g_{ab}), where Φ^I, Ψ_I are elements of the inner product space of tensor fields of valence $|I|$ with compact support.

Using (2.2) and (2.3), and Stokes' theorem, the condition that $\mathcal{D}^{IJ}$ be symmetric on this domain, i.e., that

$$(\Phi^I, \mathcal{D}_{IJ}\Psi^J)_M = (\mathcal{D}_{IJ}\Phi^J, \Psi^I)_M \tag{2.4}$$

imposes the restrictions

$$A_{IJ}{}^{ab} = A_{JI}{}^{ab}, \tag{2.5}$$

$$B_{IJ}{}^a = 2\nabla_b A_{IJ}{}^{ab} - B_{JI}{}^a, \tag{2.6}$$

$$C_{IJ} = C_{JI} - \nabla_a B_{JI}{}^a + \nabla_a\nabla_b A_{IJ}{}^{ab}, \tag{2.7}$$

where, without loss generality, we may also impose the condition

$$A_{IJ}{}^{ab} = A_{IJ}{}^{(ab)}, \tag{2.8}$$

since the term

$$\nabla_{[b}\nabla_{a]}\Phi^I \tag{2.9}$$

is expressable in terms of Φ^I and the Riemann tensor, and thus may be absorbed into $C_I{}^J\Phi_J$.

Therefore, considering two solutions ϕ_I and ψ_I of (2.1), with symmetric $\mathcal{D}_{IJ}$, we have for some compact 4-volume v (with boundary ∂v)

$$\begin{aligned} 0 &= \int_v (\phi^I \mathcal{D}_I{}^J \psi_J - \psi^I \mathcal{D}_I{}^J \phi_J) dv \\ &= \int_{\partial v} \{\phi^I A_{IJ}{}^{ab}\nabla_a \psi^J - \psi^I A_{IJ}{}^{ab}\nabla_a \phi^J \\ &\quad + \phi^I (B_{IJ}{}^a - B_{JI}{}^a)\Psi^J\} d\Sigma_a. \end{aligned} \tag{2.10}$$

Hence the sympletic form

$$\begin{aligned} \Omega(\phi^I, \psi^I) = \int_\Sigma (\phi^I A_{IJ}{}^{ab}\nabla_b \psi^J \\ - \psi^I A_{IJ}{}^{ab}\nabla_b \phi^J + 2\phi^I B_{[IJ]}{}^a \psi^J) d\Sigma_a \end{aligned} \tag{2.11}$$

is independent of the choice of Cauchy surface, assuming space—time to be globally hyperbolic, and that the boundary conditions on ϕ^I and ψ^I are sufficient to ensure that the surface integral at spacial infinity vanishes. This latter condition means physically that the energy-momentum of the field is finite on Σ.

Now if the background space—time admits an isotropy group, so there exists a Killing vector field k^a, then the Lie derivative, $\mathcal{L}_k$, with respect to k^a, will commute with $\mathcal{D}^{IJ}$ so that if ϕ^I is a solution of (2.1), then so is $\mathcal{L}_k\phi^I$. The quantity $\frac{1}{2}\Omega(\phi^I, \mathcal{L}_k\phi^I)$ is our candidate for the energy—momentum along k^a, of the field ϕ^I.

As a simple example, a massive scalar field, ϕ, satisfies

$$(\Box + m^2)\phi = 0. \tag{2.12}$$

The operator is symmetric with

$$A_{IJ}{}^{ab} = g^{ab}, \tag{2.13}$$

$$B_{IJ}{}^a = 0, \tag{2.14}$$

$$C_{IJ} = m^2. \tag{2.15}$$

The symplectic form is given by

$$\Omega^{s=0}(\phi, \mathcal{L}_k\phi) = \int_\Sigma (\phi\nabla^a \mathcal{L}_k\phi - \mathcal{L}_k\phi\nabla^a\phi) d\Sigma_a. \tag{2.16}$$

Using Stokes' theorem, and employing suitable boundary conditions, one may readily show that

$$\Omega^{s=0}(\phi, \mathcal{L}_k\phi) = 2\int_\Sigma T^a{}_b k^b d\Sigma_a, \tag{2.17}$$

where $T^a{}_b$ is the energy-momentum tensor for ϕ, given by

$$T^a{}_b = \nabla^a\phi\nabla_b\phi - \tfrac{1}{2}\delta^a{}_b(\nabla_c\phi\nabla^c\phi - m^2\phi^2). \tag{2.18}$$

3. CONSERVATION EQUATIONS BASED ON THE ENERGY-MOMENTUM TENSOR

Consider an arbitrary vector field k^a defined on an arbitrary 4-volume v, compact with boundary, in space—time. From (1.1) we have the identity

$$\nabla_a(T^a{}_b k^b) = -\tfrac{1}{2}T_{ab}\mathcal{L}_k g^{ab}. \tag{3.1}$$

Using the notation that for a derivation $\mathcal{D}$, and arbitrary valence tensor fields A^I, B^J,

$$A^I \overleftrightarrow{\mathcal{D}} B^J = A^I \mathcal{D} B^J - B^J \mathcal{D} A^I \tag{3.2}$$

then the right-hand side of (3.1) may be written in the form

$$-\tfrac{1}{2}T_{ab}\mathcal{L}_k g^{ab} = -\tfrac{1}{4}\mathcal{L}_k T + \tfrac{1}{4}g^{ab}\overleftrightarrow{\mathcal{L}_k}\, T_{ab}. \tag{3.3}$$

Defining the quantity $E(\partial v)$ by

$$E(\partial v) = \int_{\partial v} T^a{}_b k^b \, d\Sigma_a, \tag{3.4}$$

then using Stokes' theorem and (3.1) we have

$$E(\partial v) = -\tfrac{1}{2}\int_v T_{ab}\mathcal{L}_k g^{ab}\, dv \tag{3.5}$$

or from (3.3)

$$E(\partial v) = \tfrac{1}{4}\int_v (g^{ab}\overleftrightarrow{\mathcal{L}_k}\, T_{ab} - \mathcal{L}_k T) dv. \tag{3.6}$$

Let us now introduce a perturbation expansion. For any quantity $\hat{Q}$ defined on some space—time $(M, \hat{g}_{ab})$, we assume the expansion, for some suitable ϵ,

$$\hat{Q} = Q + \epsilon \underset{1}{Q} + \epsilon^2 \underset{2}{Q} + \cdots \tag{3.7}$$

about the background space—time (M, g_{ab}). For a geometric interpretation of this expansion see Ref. 10. In particular, the metric tensor, and the energy-momentum tensor are expanded as

$$\hat{g}_{ab} = g_{ab} + \epsilon h_{ab} + \epsilon^2 j_{ab} + \cdots \tag{3.8}$$

and

$$\hat{T}_{ab} = T_{ab} + \epsilon \underset{1}{T}_{ab} + \epsilon^2 \underset{2}{T}_{ab} + \cdots. \tag{3.9}$$

The vector field $\hat{k}^a$ is also given by

$$\hat{k}^a = k^a + \epsilon \underset{1}{k}^a + \epsilon^2 \underset{2}{k}^a + \cdots. \tag{3.10}$$

We shall assume in the following that k^a is a Killing vector field of the background metric. The vectors $\underset{1}{k}^a$, $\underset{2}{k}^a, \cdots$ shall be completely undefined.

By expanding the identity

$$\hat{g}_{ab} = \hat{g}^{cd}\hat{g}_{ca}\hat{g}_{db} \tag{3.11}$$

and equating coefficients of each power in ϵ, the expansion of the inverse metric to (3.8) may be found. The result is

$$\hat{g}^{ab} = g^{ab} - \epsilon h^{ab} - \epsilon^2 (j^{ab} - h^a{}_c h^{cb}) + \cdots, \tag{3.12}$$

where the indices on h_{ab} and j_{ab} have been raised with the background metric.

Let us now return to (3.5). Using (3.8), (3.9), (3.10), and (3.12) we have, retaining terms to O(ϵ),

$$\begin{aligned}\underset{1}{E}(\partial v) &= \int_{\partial v} (\underset{1}{T}^a{}_b + \tfrac{1}{2}T^a{}_b h)k^b \, d\Sigma_a \\ &= \tfrac{1}{2}\int_v T_{ab}\mathcal{L}_k h^{ab}\, dv,\end{aligned} \tag{3.13}$$

where $h = h_{ab}g^{ab}$. In deriving (3.13) we have used firstly the fact that k^a generates a group of motions in the background space—time, and secondly that the terms containing the undefined vector field $\underset{1}{k}^a$ cancel. Indeed, it is clear that this cancellation will occur to all orders in the perturbation expansion, so that (3.10) may be replaced by

$$\hat{k}^a = k^a. \tag{3.14}$$

Noticing that T_{ab} is dragged along k^a, then from (3.13) we have the O(ϵ) conservation law

$$\underset{1}{E}(\partial v) + \underset{1}{P}(\partial v) = 0, \tag{3.15}$$

where

$$\underset{1}{E}(\partial v) = \int_{\partial v} \underset{1}{\mathcal{T}}_a{}^b k^a n_b \, d^3x \tag{3.16}$$

and

$$\underset{1}{P}(\partial v) = -\tfrac{1}{2}\int_{\partial v} \mathcal{T}^{ab} h_{ab} k^c n_c \, d^3x, \tag{3.17}$$

where we use the tensor density

$$\mathcal{T}_{ab} = \sqrt{-g}\, T_{ab} \tag{3.18}$$

and the unit normal, n_a, to ∂v.

For a globally hyperbolic space—time and vector field $\partial/\partial v$ transversal to one of its Cauchy surfaces Σ,

(3.15) may be expressed as

$$\frac{d}{dv}(\underset{1}{E}(\Sigma) + \underset{1}{P}(\Sigma)) = 0. \tag{3.19}$$

This conservation equation describes secular changes in the energy-momentum distribution of the matter fields. Outside the support of the energy-momentum tensor, the integrands of E and P are identically zero. Clearly to O(ϵ) no energy-momentum is carried away from the matter fields by gravitational radiation.

To consider these radiative effects, we must retain terms to O(ϵ^2). However, before doing so it should be noted that although the integrands of E and P are not in general gauge invariant, an application of the Einstein—Klein theorem[11] will show that the sum $(E+P)$ is.

Now, from (3.6) we may put

$$\hat{E}(\partial v) = \tfrac{1}{4}\int_v \hat{A}\, \hat{d}v \tag{3.20}$$

where, using (3.14)

$$\hat{A} = \hat{g}^{ab}\mathcal{L}_k \hat{T}_{ab} - \mathcal{L}_k \hat{T}. \tag{3.21}$$

Hence, using (3.12)

$$A = 0, \tag{3.22}$$

$$\underset{1}{A} = 2\mathcal{L}_k (h^{ab}T_{ab}), \tag{3.23}$$

and

$$\begin{aligned}\underset{2}{A} = {}& \mathcal{L}_k\{2(j^{ab} - h^{ac}h_{cb})T_{ab} + h^{ab}\underset{1}{T}_{ab}\} \\ & - h^{ab}\overleftrightarrow{\mathcal{L}_k}\, \underset{1}{T}_{ab}.\end{aligned} \tag{3.24}$$

Now, using the first order field equations

$$\underset{1}{R}_{ab} - \tfrac{1}{2}g_{ab}\underset{1}{R} - \tfrac{1}{2}h_{ab}R = -\kappa \underset{1}{T}_{ab}, \tag{3.25}$$

where $\kappa = 8\pi G/c^2$ is the Einstein gravitational constant, then a little manipulation gives

$$h^{ab}\overleftrightarrow{\mathcal{L}}_k \underset{1}{T}_{ab} = -\frac{1}{\kappa}(h^{ab} - \tfrac{1}{2}hg^{ab})\overleftrightarrow{\mathcal{L}}_k \underset{1}{R}_{ab} + \tfrac{1}{2}h\overleftrightarrow{\mathcal{L}}_k(h^{ab}T_{ab}). \tag{3.26}$$

Furthermore since

$$\underset{2}{E}(\partial v) = \tfrac{1}{4}\int_v (\underset{2}{A} + \tfrac{1}{2}h\underset{1}{A})dv \tag{3.27}$$

then (3.23), (3.24), (3.26) and another small amount of manipulation gives

$$\underset{2}{E}(\partial v) = \tfrac{1}{4}\int_v \mathcal{L}_k Q\, dv + (1/4\kappa)\int_v (h^{ab} - \tfrac{1}{2}hg^{ab})\overleftrightarrow{\mathcal{L}}_k \underset{1}{R}_{ab}\, dv, \tag{3.28}$$

where

$$Q = (2j^{ab} - 2h^{ac}h_c{}^b + \tfrac{1}{2}hh^{ab})T_{ab} + h^{ab}\underset{1}{T}_{ab}. \tag{3.29}$$

Notice that the first integrand on the right-hand side of (3.28) is already in the form of a pure divergence.

Concentrate on the second integral on the right-hand side of (3.28). Using the well known result (see for example Ref. 12)

$$\underset{1}{R}_{ab} = \tfrac{1}{2}\Box h_{ab} - \nabla_d\nabla_{(a}k_{b)}{}^d \tag{3.30}$$

where

$$k_{ab} = h_{ab} - \tfrac{1}{2}hg_{ab}, \tag{3.31}$$

together with the identity

$$\nabla_d\nabla_a k_b{}^d = \nabla_{(d}\nabla_{a)}k_b{}^d + \tfrac{1}{2}R_{dab}{}^c k_c{}^d + \tfrac{1}{2}R_{da}k_b{}^d, \tag{3.32}$$

then, in the notation of Sec. 2 we may put

$$\underset{1}{R}_{ab} = \mathcal{D}_{ab}{}^{cd}k_{cd}, \tag{3.33}$$

where

$$\mathcal{D}_{ab}{}^{cd} = A_{ab}{}^{cd(ef)}\nabla_e\nabla_f + B_{ab}{}^{cde}\nabla_e + C_{ab}{}^{ef}, \tag{3.34}$$

$$A_{ab}{}^{cd(ef)} = \tfrac{1}{2}\delta_{(a}{}^c\delta_{b)}{}^d g^{ef} - \tfrac{1}{4}g_{ab}g^{cd}g^{ef} - \delta_{(a}{}^c\delta_{b)}{}^{(e}g^{f)d}, \tag{3.35}$$

$$B_{ab}{}^{cde} = 0, \tag{3.36}$$

and

$$C_{ab}{}^{cd} = \tfrac{1}{2}R^c{}_{(ab)}{}^d + \tfrac{1}{2}\delta^c{}_{(a}R_{b)}{}^d. \tag{3.37}$$

From (2.5), (2.6), and (2.7), the conditions for the symmetry of $\mathcal{D}_{ab}{}^{cd}$ are satisfied. As a consequence

$$(k^{ab}, \mathcal{D}_{ab}{}^{cd}\mathcal{L}_k k_{cd})_v - (\mathcal{D}^{cd}{}_{ab}k^{ab}, \mathcal{L}_k k_{cd})_v \tag{3.38}$$

may be expressed as an integral on ∂v.

From (3.28) and (2.10) we have the explicit $O(\epsilon^2)$ conservation law

$$\underset{2}{E}(\partial v) = \tfrac{1}{4}\int_{\partial v} Qk^d\, d\Sigma_d + \frac{1}{8\kappa}\int_{\partial v}\{h^{ab}\overleftrightarrow{\mathcal{L}}_k(\nabla^d h_{ab} - 2\nabla_a h_b{}^d) - h\overleftrightarrow{\mathcal{L}}\nabla^d h + h^{db}\overleftrightarrow{\mathcal{L}}_k\nabla_b h + h\overleftrightarrow{\mathcal{L}}_k\nabla_b h^{db}\}\, d\Sigma_d. \tag{3.39}$$

Outside the support of the energy–momentum tensor $\underset{2}{E}(\partial v) = Q = 0$. Only the second integral on the right-hand side of (3.39) is nonzero. This is our candidate for gravitational energy–momentum. Considerable light may be shed on this somewhat unwieldy expression by imposing a gauge condition. One suitable choice would be the de Donder gauge, $\nabla_a k^a{}_b = 0$. In terms of the normal, n_a, to ∂v, another convenient choice of gauge is

$$h^{ab}n_a = 0 \tag{3.40}$$

and it is this one we shall employ in this paper.

We may use the freedom in the choice of the normal n_a to ∂v in the perturbed space–time to set

$$\hat{n}_a = n_a, \tag{3.41}$$

so that the undefined covector $\underset{1}{n}_a$ is set equal to zero and also

$$\underset{1}{n}^a = -h^{ab}n_b + g^{ab}\underset{1}{n}_b = 0. \tag{3.42}$$

Now consider the symmetric tensor

$$\hat{K}_{ab} = \mathcal{L}_{\hat{n}}\hat{g}_{ab}, \tag{3.43}$$

where we extend n_a onto an open neighborhood of ∂v. Using (3.42) we have

$$\underset{1}{K}_{ab} = n^d(\nabla_d h_{ab} - 2\nabla_{(a}h_{b)d}). \tag{3.44}$$

Also, using (3.40)

$$\underset{1}{K} = -h^{ab}K_{ab} + g^{ab}\underset{1}{K}_{ab} = n^d\nabla_d h. \tag{3.45}$$

If k^a is transversal to ∂v, we may define the extension of n_a by

$$\mathcal{L}_k n_a = 0. \tag{3.46}$$

If k^a lies in ∂v, our results will be extension independent.

Defining the tensor density

$$\hat{\pi}_{ab} = (\hat{K}_{ab} - \hat{K}\hat{g}_{ab})\sqrt{-\hat{g}} \tag{3.47}$$

then from (3.43), (3.44), and (3.45)

$$h^{ab}\overleftrightarrow{\mathcal{L}}_k \underset{1}{\pi}_{ab} = \{h^{ab}\overleftrightarrow{\mathcal{L}}_k(\nabla^d h_{ab} - 2\nabla_a h_b{}^d) - h\overleftrightarrow{\mathcal{L}}_k\nabla^d h + h\overleftrightarrow{\mathcal{L}}_k\nabla_b h^{bd}\}\sqrt{-g}\, n_d. \tag{3.48}$$

Comparing (3.48) with (3.39), then in the gauge given by (3.40), the $O(\epsilon^2)$ conservation equation becomes

$$\underset{2}{E}(\partial v) = \tfrac{1}{4}\int_{\partial v} Qk^d\, d\Sigma_d + \frac{1}{8\kappa}\int_{\partial v} h^{ab}\overleftrightarrow{\mathcal{L}}_k \underset{1}{\pi}_{ab}d^3x. \tag{3.49}$$

In terms of a Cauchy surface Σ, suitable boundary conditions to ensure that the total energy–momentum content on Σ is finite, and a vector field $\partial/\partial v$, transversal to Σ, then (3.49) may be written as

$$\frac{d}{dv}\left[\underset{2}{E}(\partial v) - \tfrac{1}{4}\int_\Sigma Qk^d\, d\Sigma_d - \frac{1}{8\kappa}\int_\Sigma h^{ab}\overleftrightarrow{\mathcal{L}}_k \underset{1}{\pi}_{ab}d^3x\right] = 0. \tag{3.50}$$

Again, using the Einstein–Klein theorem, the expression in square brackets may be seen to be gauge invariant.

If, to $O(\epsilon^2)$, space–time is vacuum, then the quantity

$$\Omega^{s=2}(h_{ab}, \mathcal{L}_k h_{ab}) = -\frac{1}{4\kappa}\int_\Sigma h^{ab}\overleftrightarrow{\mathcal{L}}_k \underset{1}{\pi}_{ab}d^3x \tag{3.51}$$

is independent of the choice, Σ, of Cauchy surface.

Here, $\Omega^{s=2}$ is the spin-2 symplectic form. In terms of two solutions $\underset{A}{h}_{ab}$ and $\underset{B}{h}_{ab}$ of the linearized Einstein equations, it may be defined by

$$\Omega^{s=2}(\underset{A}{h}_{ab}, \underset{B}{h}_{ab}) = \frac{1}{4\kappa}\int_\Sigma (\underset{1}{\pi}_{ab}\underset{B}{h}^{ab} - \underset{A}{h}^{ab}\underset{1}{\pi}_{ab})d^3x \tag{3.52}$$

We have already seen that for a scalar field

$$\Omega^{s=0}(\phi, \pounds_k\phi) = 2\int_\Sigma \overset{s=0}{T^a}_b k^b\, d\Sigma_a \tag{3.53}$$

defines a conserved energy—momentum tensor. Similarly, putting

$$\Omega^{s=2}(h_{ab}, \pounds_k h_{ab}) = 2\int_\Sigma \overset{s=2}{T^a}_b k^b\, d\Sigma_a \tag{3.54}$$

defines a conserved energy-momentum tensor for the linearized gravitational field on a curved vacuum space—time.

Treating $h^{\alpha\beta}$, the projection of h^{ab} onto Σ, as generalized coordinates on the space of solutions to the linearized Einstein equations, then the canonical momenta are given by $\underset{1}{\pi}_{\alpha\beta}$.[13] Introducing the Hamiltonian density, $H_{(k)}$ on Σ, relative to the Killing vector k^a, we have

$$\pounds_k h^{\alpha\beta} = \partial H_{(k)}/\partial \underset{1}{\pi}_{\alpha\beta}, \tag{3.55}$$

$$\pounds_k \underset{1}{\pi}_{\alpha\beta} = -\partial H_{(k)}/\partial h^{\alpha\beta}. \tag{3.56}$$

Substituting (3.55) and (3.56) into (3.51) we have, using (3.40)

$$\Omega^{s=2}(h_{ab}, \pounds_k h_{ab}) = \frac{1}{4\kappa}\int_\Sigma H_{(k)}\, d^3x \tag{3.57}$$

thus relating $\Omega^{s=2}$ with a canonical formalism.

4. PERTURBATIONS OF THE BEL-ROBINSON TENSOR

Before going on to consider concrete examples where the conservation equations derived in the previous section apply, it is instructive to consider an example where integral conservation equations on space—times with Killing vectors cannot be extended when space—time is perturbed.

In vacuum, the Bel—Robinson tensor may be defined in terms of the Weyl spinor Ψ_{ABCD}[3] so that

$$T_{abcd} = \Psi_{ABCD}\overline{\Psi}_{A'B'C'D'}. \tag{4.1}$$

It may be seen from (4.1) that T_{abcd} is symmetric, trace-free and divergence-free [cf. (1.3)], and so if space—time admits the Killing vector k^a than

$$\int_{\partial v} T^a{}_{bcd} k^b k^c k^d\, d\Sigma_a = 0. \tag{4.2}$$

It may be commented that (4.2) is a special case of an equation holding when space—time admits a conformal Killing tensor of valence three. A Killing tensor need not be reducible to three Killing vectors, an example of which being the valence two irreducible Killing tensor of the Kerr space—time.[14]

In a similar fashion to the energy-momentum tensor we may put

$$T(\partial v) = \tfrac{3}{2}\int_v T^{ab}{}_{cd} k^c k^d \pounds_k g_{ab}\, dv$$

$$= \tfrac{3}{4}\int_v g^{ab}\overleftrightarrow{\pounds}_k (T_{abcd}k^c k^d)\, dv. \tag{4.4}$$

Perturbing away from the background space—time on which k^a is Killing, it is straightforward to derive an $O(\epsilon)$ conservation equation based on the Bel—Robinson tensor.

Putting

$$\hat{T}_{abcd} = T_{abcd} + \underset{1}{T}_{abcd} + \cdots \tag{4.5}$$

and

$$\mathcal{T}_{abcd} = \sqrt{-g}\, T_{abcd} \tag{4.6}$$

and using (3.8), (3.12), and (3.14), we have in an analogous fashion to (3.15), (3.16), and (3.17)

$$\underset{1}{T}(\partial v) + \underset{1}{V}(\partial v) = 0, \tag{4.7}$$

where

$$\underset{1}{T}(\partial v) = \int_{\partial v} \underset{1}{\mathcal{T}}^a{}_{bcd} k^b k^c k^d n_a\, d^3x \tag{4.8}$$

and

$$\underset{1}{V}(\partial v) = \int_{\partial v} \mathcal{T}_{abcd} h^{ab} k^c k^d k^e n_e\, d^3x. \tag{4.9}$$

However, the Bel—Robinson tensor is quadratic in the gravitational field variables, and insofar as it describes, some property of gravitational radiation, then conservation equations derived from it are only useful to $O(\epsilon^2)$ and higher. This is where the trouble begins.

As with the energy-momentum tensor, the crucial term to study in attempting to formulate an $O(\epsilon^2)$ conservation equation is

$$\int_v h^{ab}\overleftrightarrow{\pounds}_k \underset{1}{T}_{abcd} k^c k^d\, dv. \tag{4.10}$$

Writing

$$\underset{1}{T}_{abcd} k^c k^d = \mathcal{E}_{ab}{}^{cd} h_{cd}, \tag{4.11}$$

then the possibility of writing (4.10) as a divergence will depend on whether $\mathcal{E}_{ab}{}^{cd}$ is a symmetric differential operator.

In terms of the Riemann tensor and its left dual, then

$$T_{abcd}k^c k^d = (R_{aecf}R_b{}^e{}_d{}^f + {}^*R_{aecf}{}^*R_b{}^e{}_d{}^f)k^c k^d. \tag{4.12}$$

Further, putting

$$\mathcal{E}_I{}^J h_J = A_I{}^{Jab}\nabla_a\nabla_b h_J + B_I{}^{Ja}\nabla_a h_J + C_I{}^J h_J \tag{4.13}$$

we find from (4.12) that

$$B_I{}^{Ja} = 0 \tag{4.14}$$

so that, from (2.6), a necessary condition for the symmetry of $\mathcal{E}_I{}^J$ is that

$$\nabla_a A_I{}^{Jab} = 0. \tag{4.15}$$

It is easily shown that (4.15) is satisfied (trivially) if and only if

$$R_{abcd} = 0. \tag{4.16}$$

Hence, in general, the Bel—Robinson tensor does not yield conservation equations to $O(\epsilon^2)$ in the metric perturbation.

When (4.16) is satisfied, then the generalized Bel—Robinson tensor (for an arbitrary zero rest mass field $\phi_{AB\cdots D}$, satisfying $\partial^A{}_{A'}\phi_{AB\cdots D}=0$)

$$T_{ab\cdots de\cdots fg\cdots h}=\partial_{EE'}\cdots\partial_{FF'}\phi_{AB\cdots D}\times\partial_{GG'}\cdots\partial_{HH'}\phi_{A'B'\cdots D'} \tag{4.17}$$

is divergence-free on all its indices.

These conservation equations have been investigated by a number of authors (e.g. see Ref. 15). In the literature they are referred to as "zilch." They correspond to the fact that the number of quanta associated with each mode of a free field is a constant in time in flat space, a somewhat trivial property.

The nonexistence of conservation equations derived from the Bel—Robinson tensor under circumstances where they would be of dynamical interest make it a rather unsatisfactory object to be of importance in general relativistic situations. One might regard the existence of equations derived from it when space—time admits a conformal Killing tensor as being an unstable feature on the solution space of Einstein's equations.

5. BLACK HOLE PERTURBATIONS

We shall now apply some of the preceding formalism to a situation of some astrophysical interest, by taking the background metric of the perturbation expansion to describe a stationary vacuum space—time containing a black hole (characterized by the existence of a regular event horizon). It is now well known that such a space—time is uniquely described by the Kerr[13] metric which admits two commuting Killing vector fields. That Killing vector which is timelike at infinity and so corresponds to time translations is denoted by K^a. That Killing vector which is spacelike and corresponds to rotations about the axis of symmetry is denoted by $\tilde{K}^a$.

For simplicity, let us suppose that this space—time is perturbed by a purely gravitational disturbance. Furthermore let us take the bounding surface ∂v to comprise those regions of space—time in which one is most interested; at infinity and on the event horizon. Hence, we shall take

$$\partial v=\mathscr{I}^+\cup\mathscr{I}^-\cup\mathscr{H}^+\cup\mathscr{H}^-, \tag{5.1}$$

where, with a slight abuse of notation, $\mathscr{I}^+$ and $\mathscr{I}^-$ are, respectively, future and past asymptotic null hypersurfaces, $\mathscr{H}^+$ and $\mathscr{H}^-$ are, respectively, the future and past event horizons.

With this, the $O(\epsilon^2)$ conservation equation becomes

$$\frac{1}{8\kappa}\int_{\partial v}\underset{1}{\pi}_{ab}\overleftrightarrow{\mathcal{L}}_k h^{ab}\,d^3x=0, \tag{5.2}$$

where k^a may be equal to either K^a or $\tilde{K}^a$.

Let us concentrate on the gravitational flux through $\mathscr{H}^+$ which we can suppose is zero at some early (retarded) time, and decays away at some late (retarded) time.

Our contention is that the mass and angular momentum flux through $\mathscr{H}^+$ are, from (5.2), given by

$$\delta M=\frac{1}{8\kappa}\int_{H^+}\underset{1}{\pi}_{ab}\overleftrightarrow{\mathcal{L}}_K h^{ab}\,d^3x \tag{5.3}$$

and

$$\delta J=\frac{1}{8\kappa}\int_{H^+}\underset{1}{\pi}_{ab}\overleftrightarrow{\mathcal{L}}_{\tilde{K}} h^{ab}\,d^3x, \tag{5.4}$$

respectively.

Now on $\mathscr{H}^+$ we have

$$k^a\,d\Sigma_a=0 \tag{5.5}$$

so that, by an application of the identity

$$\int_\Sigma\mathcal{L}_g f^a\,d\Sigma_a=\int_\Sigma g^a\nabla_b f^b\,d\Sigma_a+2\int_{\partial\Sigma}g^a f^b\,d\Sigma_{ab} \tag{5.6}$$

for arbitrary g^a, f^a, and 3-surface Σ, it may easily be shown that

$$\tfrac{1}{2}\int_{H^+}\underset{1}{\pi}_{ab}\overleftrightarrow{\mathcal{L}}_k h^{ab}\,d^3x=\int_{H^+}\underset{1}{\pi}_{ab}\mathcal{L}_k h^{ab}d^3x, \tag{5.7}$$

giving

$$\delta M=\frac{1}{4\kappa}\int_{H^+}\underset{1}{\pi}_{ab}\mathcal{L}_K h^{ab}\,d^3x, \tag{5.8}$$

$$\delta J=\frac{1}{4\kappa}\int_{H^+}\underset{1}{\pi}_{ab}\mathcal{L}_{\tilde{K}} h^{ab}\,d^3x. \tag{5.9}$$

Before preceeding with these expressions, it should be noted that an interpretation of the expression

$$\frac{1}{4\kappa}\underset{1}{\pi}_{ab}\mathcal{L}_k h^{ab} \tag{5.10}$$

as a gravitational energy—momentum density only makes good physical sense if it is gauge invariant.

Let ξ^a generate an arbitrary gauge transformation. Then (5.10) transforms as[10]

$$\frac{1}{4\kappa}\underset{1}{\pi}_{ab}\mathcal{L}_k h^{ab}\mapsto\frac{1}{4\kappa}(\underset{1}{\pi}_{ab}+\mathcal{L}_\xi\pi_{ab})(\mathcal{L}_k h^{ab}-\mathcal{L}_\xi\mathcal{L}_k g^{ab}) \tag{5.11}$$

which is gauge invariant only if

$$\pi_{ab}=0. \tag{5.12}$$

From (3.43) and (3.47), (5.12) holds if

$$\mathcal{L}_l g_{ab}=0, \tag{5.13}$$

where the generator, l^a of $\mathscr{H}^+$ is an element of the Hawking—Hartle tetrad $\{l^a, n^a, m^a, \overline{m}^a\}$,[4] given by

$$l^a=K^a+\omega_+\tilde{K}^a, \tag{5.14}$$

ω_+ being a constant. From (5.14) we see that (5.13) and therefore (5.12) is satisfied, and we may therefore, with confidence, proceed with our analysis.

From (3.48), and the gauge condition (3.40) which now reads

$$h_{ab}l^b=0 \tag{5.15}$$

we have

$$\delta M=\frac{1}{4\kappa}\int_{H^+}(\mathcal{L}_K h^{ab}\mathcal{L}_l h_{ab}-\mathcal{L}_K h\mathcal{L}_l h)d\Sigma \tag{5.16}$$

$$\delta J=\frac{1}{4\kappa}\int_{H^+}(\mathcal{L}_{\tilde{K}} h^{ab}\mathcal{L}_l h_{ab}-\mathcal{L}_{\tilde{K}} h\mathcal{L}_l h)d\Sigma. \tag{5.17}$$

Now in terms of the null tetrad,

$$g^{ab} = 2l^{(a}n^{b)} - 2m^{(a}\overline{m}^{b)}. \tag{5.18}$$

Using the fact that on $\mathscr{H}^+$ the Newman– Penrose[16] quantities

$$\rho = \sigma = 0 \tag{5.19}$$

then a straightforward calculation gives

$$\delta M = \frac{1}{4\kappa}\int_{H^+}(\partial_\kappa h_{mm}\partial_l h_{\overline{m}\overline{m}} - \partial_\kappa h_{m\overline{m}}\partial_l h_{m\overline{m}})d\Sigma + (\text{complex conjugate}), \tag{5.20}$$

$$\delta J = \frac{1}{4\kappa}\int_{H^+}(\partial_{\tilde{\kappa}} h_{mm}\partial_l h_{\overline{m}\overline{m}} - \partial_{\tilde{\kappa}} k_{m\overline{m}}\partial_l h_{m\overline{m}})d\Sigma + (\text{complex conjugate}), \tag{5.21}$$

where

$$h_{mm} = h_{ab}m^a m^b, \tag{5.22}$$

etc.

In order to express (5.20) and (5.21) in terms of Newman– Penrose quantities, we shall use some work of Chrzanowski[17] who was able to relate the perturbed spin coefficients and perturbed curvature scalars to derivatives of the metric. (In fact we must generalize his work very slightly as we cannot, a priori, use any additional gauge freedom to set $h = 0$; full details are in Ref. 7). The results we need are that

$$\underset{1}{\rho} = \partial_l h_{m\overline{m}} \tag{5.23}$$

and

$$\underset{1}{\sigma} = \tfrac{1}{2}\partial_l h_{mm}. \tag{5.24}$$

Now the $O(\epsilon)$ Newman– Penrose equation for the rate of change of $\underset{1}{\rho}$ along l^a is, on

$$\partial_l \underset{1}{\rho} = 2\epsilon \underset{1}{\rho} \tag{5.25}$$

so $\underset{1}{\rho}$ increases exponentially in magnitude with time. This implies that the surface area of the horizon would be continuously increasing, which would not be compatible with the black hole returning to a stationary state when the perturbation died down. We must therefore choose the zero solution of (5.25). Hence, from (5.23) we have, on

$$\partial_l h_{m\overline{m}} = \partial_l h = 0 \tag{5.26}$$

so that

$$\delta M = \frac{1}{4\kappa}\int_{H^+}\pounds_\kappa h^{ab}\pounds_l h_{ab}\, d\Sigma, \tag{5.27}$$

$$\delta J = \frac{1}{4\kappa}\int_{H^+}\pounds_{\tilde{\kappa}} h^{ab}\pounds_l h_{ab}\, d\Sigma, \tag{5.28}$$

or, equivalently,

$$\delta M = \frac{1}{4\kappa}\int_{H^+}\partial_\kappa h_{mm}\partial_l h_{\overline{m}\overline{m}}\, d\Sigma + (\text{complex conjugate}), \tag{5.29}$$

$$\delta J = \frac{1}{4\kappa}\int_{H^+}\partial_{\tilde{\kappa}} h_{mm}\partial_l h_{\overline{m}\overline{m}}\, d\Sigma + (\text{complex conjugate}). \tag{5.30}$$

If the background metric is static, then on $\mathscr{H}^+$ we have

$$K^a = l^a \tag{5.31}$$

and

$$\delta M = \frac{1}{2\kappa}\int_{H^+}\partial_l h_{mm}\partial_l h_{\overline{m}\overline{m}}\, d\Sigma \tag{5.32}$$

$$= \frac{2}{\kappa}\int_{H^+}\underset{1}{\sigma}\underset{1}{\overline{\sigma}}\, d\Sigma, \tag{5.33}$$

using (5.24).

This is the famous result derived originally by Hawking and Hartle,[4] based on calculating the area increase of the horizon. Hawking and Hartle's method, however, does not extend to calculating separate mass and angular momentum fluxes for a rotating black hole.

Notice that from (5.29) and (5.30) we have the immediate result that

$$\delta M + \omega_+\delta J = \frac{2}{\kappa}\int_{H^+}\underset{1}{\sigma}\underset{1}{\overline{\sigma}}\, d\Sigma. \tag{5.34}$$

To express δM and δJ purely in terms of Newman–Penrose quantities, we may use the results of Hughston and Sommers[18] that the Killing vectors K^a and $\tilde{K}^a$ may be obtained from the irreducible Killing tensor of the Kerr space–time. Details of this may be found in Ref. 7.

For the present, let us derive coordinate expressions for δM and δJ.

Teukolsky[5,19] has shown that the curvature scalar ψ_0 may be written in an ingoing null coordinate system $\{v, r, \theta, \phi\}$ as

$$\psi_0 = \int d\omega \exp(-i\omega v)\sum_{lm}\exp(im\phi)S_{lm}(\theta)R_{lm}(r), \tag{5.35}$$

where S_{lm} and R_{lm} are solutions of ordinary differential equations.

The $O(\epsilon)$ Newman–Penrose equation for the change in the shear of $\mathscr{H}^+$ is given by

$$\partial_l \underset{1}{\sigma} = 2\epsilon\underset{1}{\sigma} + \psi_0. \tag{5.36}$$

Integrating this equation, using the boundary condition that the perturbation dies away at late times, we have, using (5.35)

$$\underset{1}{\sigma} = -\int d\omega\sum_{lm}\sigma_{lm}, \tag{5.37}$$

where

$$\sigma_{lm} = \frac{e^{-i\omega v}e^{im\phi}S_{lm}(\theta)R_{lm}(\phi)}{2\epsilon + i\omega - i\omega_+ m}. \tag{5.38}$$

Hence, from (5.24)

$$h_{mm} = 2i\int d\omega\sum_{lm}\frac{\sigma_{lm}}{\omega - \omega_+ m}. \tag{5.39}$$

Therefore, from (5.29) and (5.30) we have, using the orthogonality of the azimuthal functions, the results

$$\delta M = \frac{2}{\kappa}Re\int_{H^+}d\Sigma\int d\omega\int d\omega'\sum_{ll'm}\frac{\omega}{\omega - \omega_+ m}\sigma_{lm}(\omega)\overline{\sigma}_{l'm}(\omega') \tag{5.40}$$

$$\delta J = \frac{2}{\kappa}Re\int_{H^+}d\Sigma\int d\omega\int d\omega'\sum_{ll'm}\frac{m}{\omega - \omega_+ m}\sigma_{lm}(\omega)\overline{\sigma}_{l'm}(\omega'). \tag{5.41}$$

The expression for δJ has been obtained in a completely different way by Prior.[20,21]

Notice that for each mode, we have the immediate consequence

$$\delta M/\delta J = \omega/m. \tag{5.42}$$

So far our discussion has been confined to the event horizon. A similar analysis may be performed on one of the asymptotic null hypersurfaces. In terms of a suitable null tetrad on $\mathscr{I}^+$, and putting $K^a = n^a$, we have

$$\frac{1}{4\kappa}\int_{H+} \underset{1}{\pi}_{ab} \mathcal{L}_\kappa h^{ab}\, d^3x = \frac{1}{2\kappa}\int_{H+} \partial_n h_{mm} \partial_n h_{\bar{m}\bar{m}}\, d\Sigma . \tag{5.43}$$

Interchanging l^a and n^a in (5.23) and (5.24) gives

$$\delta M = \frac{2}{\kappa}\int_{\mathscr{I}+} \underset{1}{\lambda}\underset{1}{\bar{\lambda}}\, d\Sigma \tag{5.44}$$

which is the well known Bondi mass flux[22] expression at future null infinity.

For a monochromatic wave of frequency ω and azimuthal number m, it is now completely straightforward to show that the covariant conservation equations (5.2) reduce to the Wronskian condition of Teukolsky and Press, that

$$\left[\frac{\omega}{k(4\epsilon^2 + k^2)}|\psi_0|^2\right]_{r=r_+} = \left[\frac{1}{\omega^2}|\psi_4|^2 + \frac{1}{\omega^2}|\psi_0|^2\right]_{r=\infty}, \tag{5.45}$$

where

$$k = \omega - \omega_+ m. \tag{5.46}$$

At the time, Teukolsky and Press commented;

"Although not surprising physically, this result is surprising mathematically, because there is no known divergenceless T_{ab} for gravitational perturbation on the Kerr background."

In fact, at that time, one did exist.[23] However, we have shown in this paper that the symplectic form is a particularly simple way of representing this T_{ab}, and does reduce to (5.45) for a monochromatic wave.

As a final comment for this section, notice that whilst

$$|\psi_0|^2 = T_{abcd} l^a l^b l^c l^d \tag{5.47}$$

it would be misleading in view of our previous analysis, to suppose that the conservation equation (5.45) was in any way related to the Bel—Robinson tensor.

6. CONCLUSIONS AND OUTLOOK

In this paper, we have seen how conservation equations may be derived by perturbing the energy—momentum tensor of a space—time which admits a Killing vector, a gravitational contribution arising from the symplectic form on the space of solutions to the linearised Einstein equations. We have also seen how the method breaks down for the Bel—Robinson. The success of the methods depends on the self-adjointness of the field equations. In this respect the symplectic form unifies the concept of energy—momentum for different field theories.

In a forthcoming paper[24] the author will explore the formal similarity, provided by the symplectic form, between fluxes of gravitational energy—momentum in Newtonian theory and in Relativity theory. These two fluxes have very different physical interpretations, the former being inductive, the latter being radiative. Conditions under which the relativistic flux may be considered inductive will also be investigated.

For the present, however, let us conclude by considering extending the perturbation expansion to include terms of $O(\epsilon^3)$. The nonlinearity of the field equations now becomes apparent, manifested by the fact that

$$\underset{2}{R}_{ab} = \mathcal{D}_{ab}{}^{cd} j_{cd} + \mathcal{J}_{ab}[h_{ab}], \tag{6.1}$$

where $\mathcal{J}$ is a nonlinear differential functional of h_{ab}, $\mathcal{D}_{ab}{}^{cd}$ being given in Sec. 3. This latter term in (6.1) effectively destroys the elegance of our approach, and a certain amount of brute force and perserverence is necessary to obtain conservation equations to this order. For this reason we shall give no details here.

ACKNOWLEDGMENTS

The author would like to thank Dr. J.C. Miller and Dr. D.W. Sciama for their help and guidance, and the SRC for a research studentship.

[1] T. N. Palmer, "Covariant Conservation Equations and their Relation to the Energy-Momentum Concept in General Relativity" (to appear in Physical Review).
[2] L. Bel, C.R. Acad. Sci. Paris, **248**, 1297 (1959).
[3] R. Penrose, Ann. Phys. (N.Y.) **10**, 171 (1960).
[4] S.W. Hawking and J.B. Hartle, Commun. Math. Phys. **27**, 283 (1972).
[5] S.A. Teukolsky and W. Press, Ap.J. **193**, 443 (1974).
[6] R. Penrose, *Battelle Rencontres*, edited by C.M. DeWitt and J.A. Wheeler (Benjamin, New York, 1968).
[7] T.N. Palmer, D. Phil. thesis, Oxford University, 1977.
[8] A. Ashtekhar and A. Magnon, Proc. R. Soc. London Ser. A **364**, 375 (1975).
[9] J.L. Friedman and B.F. Schutz, "Lagrangian Perturbation Theory of Nonrelativistic Fluids" (preprint, University College, Cardiff, 1977).
[10] J.M. Stewart and M. Walker, Proc. R. Soc. London Ser. A **341**, 49 (1974).
[11] L. Landau and E. Lifshitz, *The Classical Theory of Fields* (Pergamon, London, 1962).
[12] R. Geroch, J. Math. Phys. **13**, 956 (1972).
[13] C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[14] M. Walker and R. Penrose, Commun. Math. Phys. **18**, 265 (1970).
[15] T. Morgan, J. Math. Phys. **5**, 1659 (1964).
[16] E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[17] P.L. Chrzanowski, Phys. Rev. D **13**, 806 (1976).
[18] L.P. Hughston and P. Sommers, Commun. Math. Phys. **32**, 147 (1973).
[19] S.A. Teukolsky, Ap.J. **185**, 635 (1973).
[20] C.R. Prior, Proc. R. Soc. London Ser. A **354**, 379 (1977).
[21] C.R. Prior, Proc. R. Soc. London Ser. A **355**, 1 (1977).
[22] R.K. Sachs, Proc. R. Soc. London Ser. A **270**, 103 (1962).
[23] B.F. Shutz, Ap.J. Suppl. **24**, 343 (1972).
[24] T.N. Palmer, "Momentum Transfer in Gravitation Theory" (to appear in Gen. Rel. Grav).

Constructing quantum fields in a Fock space using a new picture of quantum mechanics[a)]

M. O. Farrukh[b)]

International Centre for Theoretical Physics, Trieste, Italy
(Received 13 December 1977)

For any conventional nonrelativistic quantum theory of a finite number of degrees of freedom, we construct a picture which we call "the scattering picture," combining the "nice" properties of both the interaction and the Heisenberg pictures, and show that, in the absence of bound states, the theory could be formulated in terms of a free Hamiltonian and an effective potential. We generalize the equations thus derived to the relativistic case and show that, given a Poincaré invariant self-adjoint operator D densely defined on a Fock space, there exists an interacting field which is asymptotically free and has as the scattering matrix the nontrivial operator $S = e^{iD}$, provided that D annihilates the vacuum and the one-particle states. Crossing relations could easily be imposed on D, but, apart from a few comments, the problem of analyticity of S is left open.

INTRODUCTION

In a conventional nonrelativistic quantum theory (of a finite number of degrees of freedom) there exist three well known equivalent pictures: the Schrödinger picture, the Heisenberg picture, and the interaction picture. However, trying to meet the requirements of relativistic invariance, the Schrödinger picture was found to be a "bad" picture,[1] because it treats the time coordinate in a way essentially different from the way in which it treats the other spatial coordinates.[2] Also, the interaction picture was found not to exist.[2] So we were left with one picture, i.e., the Heisenberg picture. But all attempts to construct a nontrivial relativistic quantum theory in four-dimensional space—time in the Heisenberg picture were not successful.[3]

The problem of constructing a nontrivial quantum field theory has two aspects: the kinematical aspect, which is the construction of the underlying Hilbert space with the unitary representation of the Poincarè group, and the dynamical aspect, which is the construction of the quantum field itself as a local covariant operator-valued distribution. However, the two aspects are closely related to each other, to the extent that solving one of them might be equivalent to solving the other.

We think that this situation exists because the axioms of quantum field theory are too restrictive.[2] Trying to see what a general quantum field theory, not necessarily satisfying the Wightman axioms,[2] might look like, we had to first study the nonrelativistic analog of a quantum theory having a finite number of degrees of freedom. We have found that for such a theory one may construct a picture, which we call "the scattering picture," having the following properties:

(i) To every element φ in a subspace of the Hilbert space, called the subspace of scattering states, which in the absence of bound states is equal to the whole space, there corresponds another element in the Hilbert space denoted by φ_{in}.

(ii) To every Heisenberg operator $A(t)$ there corresponds a scattering picture operator denoted by $\hat{A}(t)$ satisfying $\hat{A}(t) = \exp(itH_0/\hbar)\hat{A}(0)\exp(-itH_0/\hbar)$, where H_0 is the free Hamiltonian.

(iii) If $\varphi \to \varphi_{\text{in}}$, $\psi \to \psi_{\text{in}}$ and $A(t) \to \hat{A}(t)$ in the above correspondence, then

$$\langle \varphi, A(t)\psi\rangle = \langle \varphi_{\text{in}}, \hat{A}(t)\psi_{\text{in}}\rangle.$$

(iv) There exists a unitary operator S such that for any bounded operator $\hat{A}(t)$ in the scattering picture there correspond two operators $A_{\text{in}}(t)$ and $A_{\text{out}}(t)$ satisfying the conditions:

(a) $A_{\text{in}}(t) = \exp(itH_0/\hbar)A_{\text{in}}(0)\exp(-itH_0/\hbar)$,

(b) $A_{\text{out}}(t) = \exp(itH_0/\hbar)A_{\text{out}}(0)\exp(-itH_0/\hbar)$,

(c) $A_{\text{out}}(t) = S^{-1}A_{\text{in}}(t)S$,

(d) $\text{w-lim}_{t\to-\infty}\{\hat{A}(t) - A_{\text{in}}(t)\} = 0$,

(e) $\text{w-lim}_{t\to+\infty}\{\hat{A}(t) - A_{\text{out}}(t)\} = 0$,

where "w-lim" means the weak limit.

(v) In the absence of bound states one may introduce an "effective potential $F(t)$ satisfying $F(t) = \exp(itH_0/\hbar)F\exp(-itH_0/\hbar)$ such that the integral $\int_{-\infty}^{\infty}F(t)\,dt = D$ exists, and such that the equation of motion of an observable in the scattering picture is given by $\hat{A}(t) = U^{-1}(t)A_0(t)U(t)$, where $U(t) = \exp(i\int_{-\infty}^{t}F(t')\,dt')$ and $A_0(t)$ is the equation of motion in the interaction picture, i.e., the "free" observable. The S matrix is given by $S = \exp(iD)$.

One may easily note that the scattering picture has the "mathematics" of the Heisenberg picture of a free theory, yet it contains all the information about scattering. The states are time independent as in the Heisenberg picture, and the observables transform under the free Hamiltonian as in the interaction picture. Hence, using the scattering picture instead of the Heisenberg picture, one can disentangle kinematics from dynamics, losing nothing, except of course the direct information about bound state, if there are any.

[a)] Work supported by the Lebanese Univesity Council.

[b)] Permanent address: Department of Physics, Faculty of Sciences, Lebanese University, Hadeth, Beirut, Lebanon.

 0022-2488/78/1911-2332$1.00

The relativistic version of the equation of motion of an observable in the scattering picture in the absence of bound states can be easily guessed. The underlying space is the Fock space with its "free" representation of the Poincaré group. If $\{G(x); x\in R^4\}$ is an integrable covariant self-adjoint operator-valued function over space–time, then the integral $\int_{y\in\Gamma_-(x)} G(y)\,d^4y$, where $\Gamma_-(x)$ is the past cone of the point $x\in\mathbb{R}^4$, exists and is a covariant self-adjoint operator-valued function of $x\in\mathbb{R}^4$, which converges strongly to zero as x goes to the infinite past along a timelike direction, and converges strongly to $D=\int_{y\in\mathbb{R}^4} G(y)\,d^4y$ as x goes to the infinite future along a timelike direction. The general equation of a field operator is $\varphi(x)=U^{-1}(x)\varphi_0(x)U(x)$, where $\varphi_0(x)$ is the free field and $U(x)=\exp(i\int_{y\in\Gamma_-(x)} G(y)\,d^4y)$. This is the equation of an asymptotically free covariant field, where $\varphi_{\rm in}=\varphi_0(x)$ and $\varphi_{\rm out}=S^{-1}\varphi_0(x)S$, and $S=\exp(iD)$.

One could safely say that in fact this is a theory of the phase matrix: For given a Poincaré invariant self-adjoint operator D which annihilates the vacuum and the one-particle states, one could easily construct the scattering matrix $S=\exp(iD)$ and an effective potential density $\{G(x);\ x\in\mathbb{R}^4\}$ such that $D=\int G(x)\,d^4x$, from which the interacting field could be derived. The field theory thus constructed satisfies all the Wightman axioms[2] except microcausality. If we try to satisfy this requirement too, then, as a consequence of Haag's theorem[2], we must take $D=0$. On the other hand, the S matrix thus constructed satisfies all the required axioms[4] except crossing and analyticity, which are consequences of microcausality.[5,6]

On could think that such a theory cannot be physical because it violates causality. Yet, it is believed that the S matrix could have macrocausal aspects without the fields being microcausal.[7] The various consequences of microcausality are discussed below.

Spin and Statistics: The correct relation between spin and statistics could be derived by assuming that the asymptotic fields are microcausal.[8,9] According to an argument due to Doebner,[10] it makes sense to talk about the commutativity of local asymptotic observables for a spacelike separation, because we only measure these observables. However, the spin of a particle during the interaction might not be a meaningful concept.

Crossing and PCT: It is not difficult at all to choose among the possible phase matrices those operators which satisfy the requirements of crossing and PCT, and regard the rest as being nonphysical.

Analyticity: This is the hardest part of the construction. It is not easy to tell what choice of the phase matrix would give an analytic S matrix. To a person who is interested in the theory itself, this would be the right point to tackle, while to a person who is interested in numerical results, microcausality is a tool by which the number of unknowns in the theory is reduced to a few numerical parameters, which can be fixed by comparison with experiments. It is not difficult at all to suggest for the D matrix schemes with this property, but this still leaves open the question of the analyticity of the S matrix.

I. THE SCATTERING PICTURE OF QUANTUM MECHANICS

A. The theory of scattering

This section is a revision of a work by Jauch,[11] presented in a slightly different notation and included for the sake of completeness.

Axiom I (quantum assumption): $\mathcal{H}$ is a complete Hilbert space and H_0 (the free Hamiltonian) and H (the total Hamiltonian) are self-adjoint operators, densely defined on $\mathcal{H}$, each having a dense set of analytic vectors.[12]

It follows directly from Axiom I, that because both H_0 and H have dense sets of analytic vectors, the series $\exp(itH_0/\hbar)=\sum_{n=0}^{\infty}(1/n!)\,(itH_0/\hbar)^n$ and the series $\exp(itH/\hbar)=\sum_{n=0}^{\infty}(1/n!)(itH/\hbar)^n$ converge strongly on their respective sets of analytic vectors for every $t\in\mathbb{R}$. Since $\exp(itH_0/\hbar)$ and $\exp(itH/\hbar)$ are bounded operators, they can be extended by continuity to all elements of $\mathcal{H}$.

Defining $V(t)=\exp(itH_0/\hbar)\exp(-itH/\hbar)$, we can write the equation satisfied by the wavefunction of any system in the interaction picture in the form

$$\psi_I(t)=V(t)\psi,\quad \psi_I(0)=\psi\in\mathcal{H}. \tag{1}$$

Denote by $\mathcal{H}_+$ and $\mathcal{H}_-$ the sets of vectors $\varphi\in\mathcal{H}$ for which the strong limits s-$\lim_{t\to+\infty} V(t)\varphi$ and s-$\lim_{t\to-\infty}V(t)\varphi$ exist, respectively. $\mathcal{H}_+$ and $\mathcal{H}_-$ are closed subspaces of $\mathcal{H}$. If $\varphi_{\rm in}=$ s-$\lim_{t\to-\infty}V(t)\varphi$, $\varphi\in\mathcal{H}_-$, and $\varphi_{\rm out}=$ s-$\lim_{t\to+\infty} V(t)\varphi$, $\varphi\in\mathcal{H}_+$, then the maps $W_-\varphi=\varphi_{\rm in}$ and $W_+\varphi=\varphi_{\rm out}$, called the wave operators, are unitary transformations from their domains $\mathcal{H}_-$ and $\mathcal{H}_+$ into their ranges $W_-(\mathcal{H}_-)=\mathcal{H}^{\rm in}$ and $W_+(\mathcal{H}_+)=\mathcal{H}^{\rm out}$, respectively.

If I_A denotes the identity map on an arbitrary set A, then the unitarity of W_- and W_+ can be expressed in the form

$$W_-^\dagger W_-=I_{\mathcal{H}_-},\quad W_+^\dagger W_+=I_{\mathcal{H}_+},$$
$$W_-W_-^\dagger=I_{\mathcal{H}^{\rm in}},\quad W_+W_+^\dagger=I_{\mathcal{H}^{\rm out}}. \tag{2}$$

In order to get more information, we must postulate something about the domains and ranges of the wave operator. On a physical basis, one can say that any free state may undergo some scattering process, and the outcomes of all scattering processes cover the full range of free states. Hence, we introduce the following axiom about the ranges of the wave operators.

Axiom II (asymptotic completeness):

$$\mathcal{H}^{\rm in}=\mathcal{H}^{\rm out}=\mathcal{H}.$$

Also, if something goes into the scattering region, something should come out of it, and nothing comes out of the scattering region unless something goes into it. Hence, we introduce the following axiom about the domains of the wave operators.

Axiom III (unitarity):

$$\mathcal{H}_+ = \mathcal{H}_- = \mathcal{H}^{sc}.$$

To the free Hamiltonian H_0, all elements of $\mathcal{H}$ represent free states, while to the total Hamiltonian H, the space $\mathcal{H}$ is decomposed into two orthogonal subspaces: the space $\mathcal{H}^{sc}$ of scattering states and the space $\mathcal{H}^b$ of bound states. These two spaces are invariant under H. On the other hand, H decomposes $\mathcal{H}$ into two orthogonal subspaces: the space $\mathcal{H}^c$ of continuous spectrum and the space $\mathcal{H}^d$ of discrete spectrum. Usually, axiom III is stengthened by requiring $\mathcal{H}^{sc} = \mathcal{H}^c$.

With the aid of Axioms II and III, Eq. (2) takes the simple form

$$W_\pm^\dagger W_\pm^\dagger = I_{\mathcal{H}^{sc}}, \quad W_\pm W_\pm^\dagger = I. \tag{3}$$

It follows directly from the definitions of the wave operators that

$$\exp(itH_0/\hbar)W_\pm = W_\pm \exp(itH/\hbar), \quad t \in \mathbb{R} \tag{4}$$

or, equivalently, if $\mathcal{D}_0$ and $\mathcal{D}$ are the domains of H_0 and H, respectively, then

$$H = W_\pm^\dagger H_0 W_\pm \ \text{ on } \mathcal{D} \cap \mathcal{H}^{sc} = W_\pm^\dagger(\mathcal{D}_0). \tag{5}$$

Recalling the definitions of φ_{in} and φ_{out} for any $\varphi \in \mathcal{H}^{sc}$ and defining the S matrix by the equation $S\varphi_{\text{in}} = \varphi_{\text{out}}$, it follows that $S = W_+ W_-^\dagger$. Also, if we write $U(t) = \exp(itH_0/\hbar)W_-^\dagger \exp(-itH_0/\hbar)$ defined on $\mathcal{H}$ and ranging in $\mathcal{H}(t) \subseteq \mathcal{H}$, then, for any $t \in \mathbb{R}$, $U(t)$ is a unitary transformation of $\mathcal{H}$ into $\mathcal{H}(t)$ with the property

$$\operatorname*{s-lim}_{t\to-\infty} U(t) = I, \quad \operatorname*{s-lim}_{t\to+\infty} U(t) = S. \tag{6}$$

B. The scattering picture

We derive in this section the scattering picture for a general quantum system satisfying the three axioms of the previous section. We first note that because $U(0) = W_-^\dagger$, we have $\mathcal{H}(0) = \mathcal{H}^{sc}$, and if E^{sc} is the projection operator on $\mathcal{H}^{sc}$, then $E(t) = \exp(itH_0/\hbar)\, E^{sc} \times \exp(-itH_0/\hbar)$ is the projection operator on $\mathcal{H}(t)$.

Let $\varphi \in \mathcal{H}$ and A be an operator (densely) defined on $\mathcal{H}$, and denote by the subscripts S, H, and I, the Schrödinger, Heisenberg, and interaction pictures, respectively. Now, for any $t \in \mathbb{R}$ we have the well-known equations of motion

$$A_s(t) = A, \quad \varphi_s(t) = \exp(-itH/\hbar)\varphi,$$
$$A_H(t) = \exp(itH/\hbar)A\exp(-itH/\hbar), \quad \varphi_H(t) = \varphi,$$
$$A_I(t) = \exp(itH_0/\hbar)A\exp(-itH_0/\hbar), \quad \varphi_I(t) = V(t)\varphi, \tag{7}$$

provided we fix $A_s(0) = A_H(0) = A_I(0) = A$ and $\varphi_S(0) = \varphi_H(0) = \varphi_I(0) = \varphi$. The equivalence of the three pictures is established by the relation

$$\langle \varphi_S(t), A_s(t)\psi_s(t)\rangle = \langle \varphi_H(t), A_H(t)\psi_H(t)\rangle = \langle \varphi_I(t), A_I(t)\,\psi_I(t)\rangle, \tag{8}$$

where $\langle\ ,\ \rangle$ denotes the scalar product in $\mathcal{H}$.

We shall write, for simplicity, $A(t)$ instead of $A_H(t)$, $A_0(t)$ instead of $A_I(t)$ and substitute for $\varphi_H(t)$ its value φ.

Let $\varphi, \psi \in \mathcal{H}^{sc}$, then $W_-\varphi = \varphi_{\text{in}}$, $W_-\psi = \psi_{\text{in}}$, or $\varphi = W_-^\dagger \varphi_{\text{in}}$, $\psi = W_-^\dagger \psi_{\text{in}}$,

and

$$\langle \varphi, A(t)\psi\rangle = \langle W_-^\dagger \varphi_{\text{in}}, \exp(itH/\hbar)A\exp(-itH/\hbar)W_-^\dagger \psi_{\text{in}}\rangle$$
$$= \langle \exp(-itH/\hbar)W_-^\dagger\varphi_{\text{in}}, A\exp(-itH/\hbar)W_-^\dagger\psi_{\text{in}}\rangle$$

[and using (4) we get]

$$= \langle W_-^\dagger \exp(-itH_0/\hbar)\varphi_{\text{in}}, AW_-^\dagger \exp(-itH_0/\hbar)\psi_{\text{in}}\rangle$$
$$= \langle W_-^\dagger \exp(-itH_0/\hbar)\varphi_{\text{in}}, E^{sc}AW_-^\dagger \exp(-itH_0/\hbar)\psi_{\text{in}}\rangle$$
$$+ \langle W_-^\dagger \exp(-itH_0/\hbar)\varphi_{\text{in}}, (I - E^{sc})AW_-^\dagger \times \exp(-itH_0/\hbar)\psi_{\text{in}}\rangle. \tag{9}$$

But $\exp(-itH_0/\hbar)\varphi_{\text{in}} \in \mathcal{H}$ and $E^{sc}AW_-^\dagger \exp(-itH_0/\hbar)\psi_{\text{in}} \in \mathcal{H}^{sc}$, hence, by unitarity of $W_-^\dagger$ (between $\mathcal{H}$ and $\mathcal{H}^{sc}$) we get

$$\langle W_-^\dagger \exp(-itH_0/\hbar)\varphi_{\text{in}}, E^{sc}AW_-^\dagger \exp(-itH_0/\hbar)\psi_{\text{in}}\rangle$$
$$= \langle \exp(-itH_0/\hbar)\varphi_{\text{in}}, W_-E^{sc}AW_-^\dagger \exp(-itH_0/\hbar)\psi_{\text{in}}\rangle$$
$$= \langle \varphi_{\text{in}}, \exp(itH_0/\hbar)U_{(0)}^\dagger E(0)A_0(0)U(0)\exp(-itH_0/\hbar)\psi_{\text{in}}\rangle$$
$$= \langle \varphi_{\text{in}}, U^\dagger(t)E(t)A_0(t)U(t)\psi_{\text{in}}\rangle. \tag{10}$$

Also, $W_-^\dagger \exp(-itH_0/\hbar)\varphi_{\text{in}} \in \mathcal{H}^{sc}$ and $(I - E^{sc})AW_-^\dagger \times \exp(-itH_0/\hbar)\psi_{\text{in}} \in \mathcal{H}^b = (\mathcal{H}^{sc})^\perp$, hence

$$\langle W_-^\dagger \exp(-itH_0/\hbar)\varphi_{\text{in}}, (I - E^{sc})AW_-^\dagger \exp(-itH_0/\hbar)\psi_{\text{in}}\rangle = 0. \tag{11}$$

Hence, from (9)–(11) we get

$$\langle \varphi, A(t)\psi\rangle = \langle \varphi_{\text{in}}, U^\dagger(t)E(t)A_0(t)U(t)\psi_{\text{in}}\rangle. \tag{12}$$

Using (12), we define the scattering picture operator

$$\hat{A}(t) = U^\dagger(t)E(t)A_0(t)U(t) \tag{13}$$

satisfying

$$\hat{A}(t) = \exp(itH_0/\hbar)\hat{A}(0)\exp(-itH_0/\hbar) \tag{14}$$

and the scattering picture state $\varphi_{\text{in}} = W_-\varphi$, whenever $\varphi \in \mathcal{H}^{sc}$. Hence, we have shown that to every Heisenberg operator $A(t)$, there corresponds a scattering picture operator $\hat{A}(t)$ such that for any $\varphi, \psi \in \mathcal{H}^{sc}$ we have

$$\langle \varphi, A(t)\psi\rangle = \langle \varphi_{\text{in}}, \hat{A}(t)\psi_{\text{in}}\rangle. \tag{15}$$

Going back to Eq. (13) we have

$$\hat{A}(t) = \exp(itH_0/\hbar)W_-E^{sc}AW_-^\dagger \exp(-itH_0/\hbar)$$
$$= W_- \exp(itH/\hbar)E^{sc}A\exp(-itH/\hbar)W_-^\dagger$$
$$= W_-E^{sc}\exp(itH/\hbar)A\exp(-itH/\hbar)E^{sc}W_-^\dagger$$
$$= W_-E^{sc}A(t)E^{sc}W_-^\dagger. \qquad (16)$$

Hence, $W_-: \mathcal{H}^{sc} \to \mathcal{H}$ defines a unitary equivalence between $E^{sc}A(t)E^{sc}$, the restriction of $A(t)$ as a quadratic form to $\mathcal{H}^{sc}\times\mathcal{H}^{sc}$, and $\hat{A}(t)$. The inverse of (16) is given by

$$E^{sc}A(t)E^{sc} = W_-^\dagger\hat{A}(t)W_-. \qquad (17)$$

Although Eq. (16) defines the scattering picture operator $\hat{A}(t)$ in terms of the Heisenberg operator $A(t)$, Eq. (17) tells us that $A(t)$ cannot be recovered from $\hat{A}(t)$ except on $\mathcal{H}^{sc}\times\mathcal{H}^{sc}$.

C. The S matrix

The S matrix in the interaction picture is defined as the unitary transformation which relates the asymptotic behavior of a scattering state at $t=+\infty$ to its asymptotic behavior at $t=-\infty$. This gives the expression $S = W_+W_-^\dagger$.

In the scattering picture, the states are time independent, so we should be able to derive the S matrix from the observables.

We shall prove that $\text{w-lim}_{t\to-\infty}\{\hat{A}(t)-A_0(t)\}=0$ and $\text{w-lim}_{t\to+\infty}\{\hat{A}(t)=S^{-1}A_0(t)S\}=0$, whenever A is a bounded operator.

To prove the first limit, we note that

$$\hat{A}(t) - A_0(t) = U^\dagger(t)E(t)A_0(t)U(t) - A_0(t)$$
$$= U^\dagger(t)E(t)A_0(t)\{U(t)-I\} + \{U^\dagger(t)E(t)-I\}A_0(t). \qquad (18)$$

Let $f, g\in\mathcal{H}$, then

$$\langle f, \{\hat{A}(t)-A_0(t)\}g\rangle = \langle f, U^\dagger(t)E(t)A_0(t)\{U(t)-I\}g\rangle$$
$$+\langle f, \{U^\dagger(t)E(t)-I\}A_0(t)g\rangle$$
$$=\langle A_0^\dagger(t)U(t)f, \{U(t)-I\}g\rangle$$
$$+\langle\{U(t)-I\}f, A_0(t)g\rangle. \qquad (19)$$

Hence

$$|\langle f, \{\hat{A}(t)-A_0(t)\}g\rangle|$$
$$\le |\langle A_0^\dagger(t)U(t)f, \{U(t)-I\}g\rangle|$$
$$+|\langle\{U(t)-I\}f, A_0(t)g\rangle|$$
$$\le \|A_0^\dagger(t)U(t)f\|\,\|\{U(t)-I\}g\| + \|\{U(t)-I\}f\|\,\|A_0(t)g\|$$
$$\le \|A^\dagger\|\,\|f\|\,\|\{U(t)-I\}g\| + \|\{U(t)-I\}f\|\,\|A\|\,\|g\|. \qquad (20)$$

But because $\text{s-lim}_{t\to-\infty}U(t)=I$, it follows from (20) that $\text{w-lim}_{t\to-\infty}\{\hat{A}(t)-A_0(t)\}=0$. The proof of the other limit follows the same steps. Hence, the S matrix derivable in this picture is defined on all $\mathcal{H}$ and is identical with the S matrix derived in the interaction picture. However, it is known[13] that if A is a bounded operator on $\mathcal{H}$, then the Heisenberg operator $A(t)$ converges weakly on $\mathcal{H}^{sc}\times\mathcal{H}^{sc}$ to an in operator as $t\to-\infty$ and to an out operator as $t\to+\infty$, related to each other by a different S matrix given by $\tilde{S}=W_+^\dagger$. If $\varphi, \psi\in\mathcal{H}^{sc}$, then

$$\langle\varphi, \tilde{S}\psi\rangle = \langle\varphi, W_-^\dagger W_+\psi\rangle = \langle W_-^\dagger\varphi_{in}, W_-^\dagger W_+W_-^\dagger\psi_{in}\rangle$$
$$= \langle W_-^\dagger\varphi_{in}, W_-^\dagger S\psi_{in}\rangle = \langle\varphi_{in}, S\psi_{in}\rangle, \qquad (21)$$

which is the correct formula for the transformation from the Heisenberg picture on $\mathcal{H}^{sc}$ to the scattering picture on $\mathcal{H}$.

D. Pure scattering and effective potentials

We study in this section the special case of pure scattering, which means the absence of bound states. Mathematically, this means $\mathcal{H}^{sc}=\mathcal{H}$ or equivalently $E^{sc}=I$. Hence, the wave operators $W_\pm$ become unitary operators on $\mathcal{H}$, and $U(t)$ becomes, for any $t\in\mathbb{R}$, a unitary transformation of $\mathcal{H}$ into itself. In this case, the scattering picture becomes equivalent to any other picture, and the time dependence of an operator in this picture takes the form

$$\hat{A}(t) = U^{-1}(t)A_0(t)U(t). \qquad (22)$$

To derive the other pictures from the scattering picture, we introduce the following table:

Picture	States	Observables
Scattering	φ	$A(t)$
Interaction	$\varphi_I(t)=U(t)\varphi$	$A_I(t)=U(t)A(t)U^{-1}(t)$
Heisenberg	$\varphi_H(t)=U(0)\varphi$	$A_H(t)=U(0)A(t)U^{-1}(0)$
Schrödinger	$\varphi_S(t)=U(0)e^{-itH/\hbar}\varphi$	$A_S(t)=U(0)A(0)U^{-1}(0)$

We now proceed to analyze the family $\{U(t): t\in\mathbb{R}\}$. Since $W_-^\dagger$ is a unitary operator, it follows that there exists a unique resolution of the identity on the interval $[0, 2\pi]$ such that[14]

$$W_-^\dagger = \int_0^{2\pi}\exp(i\lambda)dE_\lambda. \qquad (23)$$

Using this we define the bounded self-adjoint operator

$$A = \int_0^{2\pi}\lambda\, dE_\lambda. \qquad (24)$$

Setting $A(t)=\exp(itH_0/\hbar)A\exp(-itH_0/\hbar)$, it follows that $\exp[iA(t)] = \exp(itH_0/\hbar)\exp(iA)\exp(-itH_0/\hbar) = \exp(itH_0/\hbar)W_-^\dagger\exp(-itH_0/\hbar)=U(t)$. We also set $F(t)=(i/\hbar)[H_0, A(t)]$, from which it follows

$$A(t)-A(t_0) = \int_{t_0}^{t}F(t')\,dt'. \qquad (25)$$

We show in the Appendix to Sec. I that under mild mathematical restrictions, the strong limit of $A(t)$ as $t\to-\infty$ is zero. Hence $A(t)=\int_{-\infty}^{t}F(t')\,dt'$ and thus

$$U(t) = \exp(i\textstyle\int_{-\infty}^{t}F(t')\,dt'). \qquad (26)$$

It is because of (26) that we call $F(t)$ the effective potential. The left-hand side of (26) converges strongly to S as $t \to +\infty$. Hence, one would expect that the integral $D=\int_{-\infty}^{\infty} F(t)\,dt$ should exist. We call this quantity the phase matrix. However, the prescription $S \to D$ such that $\exp(iD)=S$ is by no means unique. For, if N is a nonbounded operator densely defined on $\mathcal{H}$ with a dense set of analytic vectors,[12] then the operator $U=\exp(iN)$ exists and is unitary. Hence, there exists a unique bounded self-adjoint operator B with spectrum in $[0, 2\pi]$ such that $U=\exp(iB)$. Clearly, $N \neq B$. Because of these considerations, we are led to think that the effective potential could be introduced as a fundamental object, satisfying the condition of integrability over the whole real line.

II. QUANTUM FIELDS IN A FOCK SPACE

A. The Fock space S matrix

If we consider all stable systems and all systems with lifetimes much longer than the duration of the interaction under consideration as elementary particles, while considering other systems as resonances, then the free states could be described using one single Fock space. Apart from any field-theoretic considerations, the overall transitions could be described using one single unitary operator (called the S matrix) invariant under the Poincaré group. The vacuum state and all one-particle states are eigenstates of the S matrix with eigenvalues on the unit circle $\{z : z \in \mathbb{C} \text{ and } |z|=1\}$. These multiplicative unimodular Poincaré invariant constants have no physical significance. Hence, it should be assumed that the vacuum state and all one-particle states are eigenstates of the S matrix with unit eigenvalue. This assumption is necessary for the subsequent discussion.

Since the physics of scattering, whether elastic or not, could be described using a single Fock space, we choose the relativistic Fock space as our framework and introduce the following axiom.

Axiom I (relativistic Fock spcae): $\mathcal{J}$ is a Fock space and U is a unitary representation of the Poincaré group $E(1, 3)$ on $\mathcal{J}$.

Notation: We use natural units in which $\hbar=c=1$. The 4-vector $x=(x_0, \mathbf{x})$ where $\mathbf{x}=(x_1, x_2, x_3)$. The inner product $xy=x_0y_0-\mathbf{x}\cdot\mathbf{y}$, where $\mathbf{x}\cdot\mathbf{y}=x_1y_1+x_2y_2+x_3y_3$. In particular, $x^2=xx=x_0^2-|\mathbf{x}|^2$. If p is the 4-momentum of a particle, then $p^T=(p_0, -\mathbf{p})$. We often write E_p for p_0. If f and g are functions of x (and other variables), then

$$f(x, -)\overleftrightarrow{\frac{\partial}{\partial x}}g(x, -)=f(x, -)\frac{\partial g}{\partial x}(x, -)-\frac{\partial f}{\partial x}(x, -)g(x, -).$$

Definition 1: $V_m^{+}=\{p : p \in \mathbb{R}^4,\ p_0>0 \text{ and } p^2=m^2\}$ is the mass shell of a particle of mass $m \geq 0$.

The Fock space $\mathcal{J}$ is not very suitable for the following discussion. We need a space with nicer properties. We choose a subspace $\mathcal{S}(\mathcal{J}) \subseteq \mathcal{J}$, called the subspace of good vectors, as follows.

On the space $\mathcal{S}_{\mathbb{C}}(\mathbb{R}^{3n})$ of good test functions[2] of $3n$-real variables, we define a sequence of norms as follows: for any $s \in \mathbb{N}$, $f \in \mathcal{S}_{\mathbb{C}}(\mathbb{R}^{3n})$

$$|f|_s=\sup_{\substack{x\in\mathbb{R}^{3n}\\ |\alpha|,\,|\beta|\leq s}}\left|(1+|\mathbf{x}_1|^2)^{\alpha_{1/2}}\cdots(1+|\mathbf{x}_n|^2)^{\alpha_{n/2}}\frac{\partial^{\beta_1^1}}{\partial x_1^1}\frac{\partial^{\beta_1^2}}{\partial x_1^2}\frac{\partial^{\beta_1^3}}{\partial x_1^3}\cdots\frac{\partial^{\beta_n^1}}{\partial x_n^1}\frac{\partial^{\beta_n^1}}{\partial x_n^2}\frac{\partial^{\beta_n^3}}{\partial x_n^3}f(x)\right|,$$

where

$$|\alpha|=\alpha_1+\alpha_2+\cdots+\alpha_n$$

and

$$|\beta|=\beta_1^1+\beta_1^2+\beta_1^3+\cdots+\beta_n^1+\beta_n^2+\beta_n^3.$$

Consider for simplicity the case of a single scalar neutral particle and define $\mathcal{J}^{(n)}$ as the subspace of n particles. Denote by E_n the projection operator on $\mathcal{J}^{(n)}$. $\mathcal{S}(\mathcal{J}^{(0)})=\mathcal{J}^{(0)}$ and $\mathcal{S}(\mathcal{J}^{(m)})$ for $n \in \mathbb{N}^*$ is the subspace of $\mathcal{J}^{(n)}$ consisting of vectors with good wave-functions in momentum space. Now, if $\psi \in \mathcal{J}^{(n)}$ with $\psi=\int_p^{(n)}\tilde{\psi}(p_1, p_2, \ldots, p_n)\,|p_1, p_2, \ldots, p_n\rangle$

where $\int_p^{(n)}$ denotes

$$\int\frac{d^3p_1}{2E_{p_1}}\int\frac{d^3p_1}{2E_{p_2}}\cdots\int\frac{d^3p_n}{2E_{p_n}},$$

then $\|\psi\|_s=|\tilde{\psi}|_s$.

Definition 2: $\mathcal{S}(\mathcal{J})$ is the set of vectors $f \in \mathcal{J}$ such that for any $r, s \in \mathbb{N}$, $\|f\|_{r,s}=\sum_{n=0}^{\infty} n!\,n^r\|E_nf\|_s^2<+\infty$. A sequence $\{f_n : n \in \mathbb{N}\}$ of elements of $\mathcal{S}(\mathcal{J})$ converges to zero in the $\mathcal{S}$ topology iff $\lim_{n\in\mathbb{N}}\|f_n\|_{r,s}=0$ for all $r, s \in \mathbb{N}$.

As we have mentioned in the Introduction, we are interested in the case where $S=\exp(iD)$ (which is always possible[14]) with D satisfying certain conditions. We now prove a theorem about a special class of self-adjoint operators on $\mathcal{J}$.

Theorem 1: Let D

(i) be a self-adjoint operator densely defined on $\mathcal{J}$ and having $\mathcal{S}(\mathcal{J})$ in its domain,

(ii) be commuting with $U(\Lambda, a)$ for every $(\Lambda, a) \in E(1, 3)$, and

(iii) have the space $\mathcal{J}^{(0)} \oplus \mathcal{J}^{(1)}$ as a null space,

then there exists a densely defined operator-valued function $\{G(x) : x \in \mathbb{R}^4\}$ such that

(a) $G^\dagger(x)=G(x)$ for every $x \in \mathbb{R}^4$,

(b) $U(\Lambda, a)\,G(x)\,U^{-1}(\Lambda, a)=G(\Lambda x+a)$, for every $(\Lambda, a) \in E(1, 3)$, and

(c) $\int G(x)\,d^4x=D$ strongly on $\mathcal{S}(\mathcal{J})$.

Proof: We shall prove the theorem for a theory of a single neutral scalar particle of mass $m>0$. The generalization to other cases should not be difficult.

Let a be the annihilation operator of $\mathcal{J}$. Since $\mathcal{S}(\mathcal{J})$ belong to the domain of D, there exists a double sequence of distributions $G^{(j,k)} \in \mathcal{S}'\{(V_m^\dagger)^{j+k}\}$; $(j, k) \in \mathbb{N}\times\mathbb{N}$ such that on $\mathcal{S}(\mathcal{J})$ we have

$$D=\sum_{j,k=0}^{\infty} \int_{p'}^{(j)} \int_{p}^{(k)} G^{(j,k)}(p_1', p_2', \ldots, p_j', p_1, p_2, \ldots, p_k)$$

$$\times a^\dagger(p_1')a^\dagger(p_2') \cdots \times a^\dagger(p_j')a(p_1)a(p_2) \ldots a(p_k), \qquad (27)$$

where

$$\int_p^{(n)} = \int \frac{d^3p_1}{2Ep_1} \cdots \int \frac{d^3p_n}{2E_{p_n}}.$$

Because of Bose statistics, the $G^{(j,k)}$ distribution can be taken to be symmetric in the first j and the last k $V_m^\dagger$ variables, without any loss of generality. From self-adjointness of D we have

$$G^{*(j,k)}(p_1', p_2', \ldots, p_j'; p_1, p_2, \ldots, p_k)$$

$$= G^{(k,j)}(p_1, p_2, \ldots, p_k; p_1', p_2', \ldots, p_j'). \qquad (28)$$

From translational invariance we get

$$G^{(j,k)}(p_1', \cdots p_j', p_j', p_1, \ldots, p_k)$$

$$=\delta^4(p_1' + \cdots + p_j' - p_1 - \cdots - p_k)F^{(j,k)}(p_1', \ldots, p_j'; p_1, \ldots, p_k)$$

and Lorentz invariance implies (29)

$$F^{(j,k)}(\Lambda p_1', \ldots, \Lambda p_j'; \Lambda p_1, \ldots, \Lambda p_k)$$

$$= F^{(j,k)}(p_1', \ldots, p_j'; \quad p_1, \ldots, p_k)$$

for any $\Lambda \in SO(1,3)$. (30)

But because $\mathcal{J}^{(0)} \oplus \mathcal{J}^{(1)}$ is a null space of D, it follows that

$$F^{(j,0)} = F^{(j,1)} = F^{(0,j)} = F^{(1,j)} = 0 \text{ for any } j \in \mathbb{N}. \qquad (31)$$

Hence from (27), (29), and (31) we get

$$D=\sum_{j,k=2}^{\infty} \overbrace{\int_{p'}^{(j)} \int_{p}^{(k)}} F^{(j,k)}(p_1', \ldots, p_j'; p_1, \ldots p_k)a^\dagger(p_1')$$

$$\times \cdots a^\dagger(p_j')a(p_1) \cdots a(p_k), \qquad (32)$$

where

$$\overbrace{\int_{p'}^{(j)} \int_{p}^{(k)}} \cdots = \int_{p'}^{(j)} \int_{p}^{(k)} \delta^{(4)}(p_1' + \cdots + p_j' - p_1 - \cdots - p_k).$$

The functions $F^{(j,k)}$ are well behaved square integrable functions (because D is an operator and not only a quadratic form).

Now, define the operator-valued function

$$G(x) = \sum_{j,k=2}^{\infty} \int_{p'}^{(j)} \int_{p}^{(k)} F^{(j,k)}(p_1', \ldots, p_j'; p_1, \ldots, p_k)$$

$$\times \exp[i(p_1' + \cdots + p_j' - p_1 - \cdots - \cdots - p_k)x]$$

$$\times a^\dagger(p_1') \cdots a^\dagger(p_j')a(p_1) \cdots a(p_k). \qquad (33)$$

This functions meets the requirements of the theorem. ∎

We have two comments on the above theorem. The first is that $G(x)$ is not unique, for if we add to $F^{(j,k)}$ in definition (33) a term of the form

$$(p_1' + \cdots + p_j' - p_1 - \cdots - p_k)^2 F'^{(j,k)}(p_1', \ldots, p_j'; p_1, \ldots, p_k),$$

where $F'^{(j,k)}$ satisfies conditions (28) and (30), then the new function $G'(x)$ also meets the requirements of the theorem. The second comment is about requiring $\mathcal{J}^{(0)} \oplus \mathcal{J}^{(1)}$ to be a null space of D. In general, by the hermiticity and Poincaré invariance of D, D could have in addition to expression (32) a term of the form $\alpha + \beta N$ where $N = \int_p a^\dagger(p)a(p)$; $\alpha, \beta \in \mathbb{R}$. Such a term cannot be obtained by the integration of any Poincaré invariant function, unless $\alpha = \beta = 0$, which means that $\mathcal{J}^{(0)} \oplus \mathcal{J}^{(1)}$ should be a null space of D.

B. The interacting fields

In our model of field theory we do not have bound states in the nonrelativistic sense, because, as we have explained before, we consider such states either as elementary, and hence include them in the free description, or as resonances, and hence they correspond to peaks in the scattering amplitudes. Hence, the Heisenberg picture and the scattering picture are, in our case, equivalent. However, we prefer the scattering picture for two reasons. First, the equation of motion of any observable is known in general, and could be directly be written in terms of an "effective potential density." Second, the S matrix derivable from the scattering picture has a direct experimental interpretation because it is identical with the S matrix of the interaction picture.

The fundamental object in our construction is the "effective potential density."

Definition 3: For any $x \in \mathbb{R}^4$, $\Gamma_-(x) = \{y : y \in \mathbb{R}^4, y_0 < x_0, (y-x)^2 > 0\}$ is the past cone of the point $x \in \mathbb{R}^4$.

We state without proof the following theorem.

Theorem 2: For any $n \in V_1^\dagger$ and $x \in \mathbb{R}^4$, the function

$\mathbb{R} \to \mathcal{P}(\mathbb{R}^4), \tau \to \Gamma_-(x + \tau n)$,
where $\mathcal{P}(\mathbb{R}^4)$ is the power set of $\mathbb{R}^4$, is increasing. Moreover,

$$\lim_{\tau\to-\infty} \Gamma_-(x + \tau n) = \emptyset \text{ and } \lim_{\tau\to+\infty} \Gamma_-(x + \tau n) = \mathbb{R}^4.$$

We now introduce the following axiom.

Axiom II: $\{G(x) : x \in \mathbb{R}^4\}$ is a densely defined operator-valued function called the effective potential density. It satisfies the following conditions:

(i) $G^\dagger(x) = G(x)$ for any $x \in \mathbb{R}^4$.

(ii) $U(\Lambda, a)G(x)U^{-1}(\Lambda, a) = G(\Lambda x + a)$, for any $(\Lambda, a) \in E(1,3)$.

(iii) $D = \int G(x) d^4x$, called the phase matrix, exists strongly on $\mathcal{J}$.

(iv) For any $f \in \mathcal{S}(\mathcal{J})$, $x \in \mathbb{R}^4$, and $n \in V_1^\dagger$, we have

$$\mathcal{S}-\lim_{\tau\to-\infty} U(x + \tau n)f = f \text{ and } \mathcal{S}-\lim_{\tau\to+\infty} U(x + \tau n)f$$

$= \exp(iD)f$,

where $\mathcal{S}$-lim denotes limit in the $\mathcal{S}$ topology and $U(x) = \exp\{i\int_{y\in\Gamma_-(x)} G(y)d^4y\}$.

We are now in a position to "define" the S matrix and the interacting field.

Definition 4: The scattering matrix $S=\exp(iD)$.

Definition 5: The interacting field $\varphi(x)=U^{-1}(x)\varphi_0(x)U(x)$, where φ_0 is a free field.

It can easily be verified that the S matrix defined above is unitary and Poincaré invariant, and the interactering field is Poincaré covariant. The family $\{U(x): x\in \mathbb{R}\}$ is a relativistic generalization of the family $\{U(t): t\in \mathbb{R}\}$ of Sec. I. To prove this statement, let $x=(t, \mathbf{x})$, then, in the nonrelativistic limit as $c\to\infty$, $\Gamma_-(x)\to\{y: y\in\mathbb{R}^4$ and $y_0<t\}$.

Hence,

$$U(x)=U(t, \mathbf{x})\to \exp\{i\int_{-\infty}^t dy_0\int_{y\in\mathbb{R}} G(y_0, y)d^3y\}$$
$$=\exp\{i\int_{-\infty}^t F(t')dt'\}=U(t),$$

where $F(t)=\int d^3x\, G(t, \mathbf{x})$.

We prove in the Appendix to sec. II the following important theorem.

Theorem 3: $\mathcal{J}^{(0)}\oplus\mathcal{J}^{(1)}$ is a null space of $G(x)$, for any $x\in\mathbb{R}^4$.

Definition 6: Let $f\in\mathcal{S}_{\mathbb{C}}(\mathbb{R}^3)$, then the smeared field $\varphi(f, t)=\int_{x_0=t} d^3x\, \varphi(x)f(\mathbf{x})$.

Theorem 4: For any $f\in\mathcal{S}_{\mathbb{C}}(\mathbb{R}^3)$, w-$\lim_{t\to-\infty}\{\varphi(f, t) - \varphi_0(f, t)\}=0$ and w-$\lim_{t\to+\infty}\{\varphi(f, t)-S^{-1}\varphi_0(f, t)S\}=0$ on $\mathcal{S}(\mathcal{J})\times\mathcal{S}(\mathcal{J})$.

Proof: Let $A\in\mathcal{S}(\mathcal{J}^{(m)})$, $B\in\mathcal{S}(\mathcal{J}^{(n)})$, and $f\in\mathcal{S}_{\mathbb{C}}(\mathbb{R}^3)$, then

$$\langle A, \{\varphi^{(+)}(f, t)-\varphi_0^{(+)}(f, t)\}B\rangle$$
$$=\int f(\mathbf{x})d^3x\int_{k'}^{(m)}A^*(k')\int_k^{(n)}B(k)\langle k'|$$
$$\times\exp(-i\mathbf{x}\cdot\mathbf{P})\{\varphi^{(+)}(t, 0)-\varphi_0^{(+)}(t, 0)\}\times\exp(i\mathbf{x}\cdot\mathbf{P})|k\rangle$$
$$=\int d^3xf(x)\int_{k'}^{(m)}A^*(k')\int_k^{(n)}B(k)\exp(i(\mathbf{k}-\mathbf{k}')\mathbf{x})\langle k'|\{\varphi^{(+)}(t, 0)$$
$$-\varphi_0^{(+)}(t, 0)\}|k\rangle$$
$$=\int_{k'}^{(m)}A^*(k')\int_k^{(n)}B(k)\tilde{f}(\mathbf{k}-\mathbf{k}')\langle k'|\{\varphi^{(+)}(t, 0)$$
$$-\varphi_0^{(+)}(t, 0)\}|k\rangle, \tag{34}$$

where $\varphi^{(+)}(x)=\int a(p)\exp(-ipx)$, the positive frequency component of the field. Equation (34) takes the form

$$\langle A, \{\varphi^{(+)}(f, t)-\varphi_0^{(+)}(f, t)\}B\rangle$$
$$=\int_{k'}^{(m)}A^*(k')\int_k^{(n)}B(k)\tilde{f}(\mathbf{k}-\mathbf{k}')\langle k'|U^{-1}(t, 0)\varphi_0^{(+)}(t, 0)\{U(t, 0)$$
$$-I\}|k\rangle+\int_{k'}^{(m)}A^*(k')\int_k^{(n)}B(k)\tilde{f}(\mathbf{k}-\mathbf{k}')\langle k'|\{U^{-1}(t, 0)-I\}$$
$$\times\varphi_0^{(+)}(t, 0)|k\rangle$$
$$=\int_{k'}^{(m)}A^*(k')\int_k^{(n)}B(k)\tilde{f}(\mathbf{k}-\mathbf{k}')\int\langle k'|U^{-1}(t, 0)a(p)\{U(t, 0)$$
$$-I\}|k\rangle\exp(-iE_pt)$$
$$+n\int_{k'}^{(m)}A^*(k')\int_k^{(n)}B(k)\tilde{f}(\mathbf{k}-\mathbf{k}')$$
$$\times\langle k'|\{U^{-1}(t, 0)-I\}|k\rangle'\exp(-iE_{k_1}t), \tag{35}$$

where $|k\rangle'=|k_2, \ldots, k_n\rangle$.

Now, because $|\exp(-iE_pt)|=|\exp(-iE_{k_1}t)|=1$ and there exists an $M\in\mathbb{R}$ such that $|\tilde{f}(k)|<M$ for all $k\in\mathbb{R}^3$, and because $\mathcal{S}$-$\lim_{t\to-\infty}\{U(t, 0)-I\}=0$ on $\mathcal{S}(\mathcal{J})$ and $\int_p a(p)$ is $\mathcal{S}$-continuous, it follows from **Eq.** (35) that w-$\lim_{t\to-\infty}\{\varphi^{(+)}(f, t)-\varphi_0^{(+)}(f, t)\}=0$ on $\mathcal{S}(\mathcal{J})\times\mathcal{S}(\mathcal{J})$. Now, since $\varphi^{(-)}(f, t)=[\varphi^{(+)}(f^*, t)]^\dagger$, we get w-$\lim_{t\to-\infty}$ $\{\varphi(f, t)-\varphi_0(f, t)\}=0$ on $\mathcal{S}(\mathcal{J})\times\mathcal{S}(\mathcal{J})$. The other limit could be proved in a similar way.■

It is well known that the free field expansion $\varphi_0(x) = \int_p\{a(p)\exp(-ipx)+a^\dagger(p)\exp(ipx)\}$ is invertible,

$$a(p)=i\int_{x_0=t}\exp(ipx)\,\overleftrightarrow{\frac{\partial}{\partial x_0}}\,\varphi_0(x)\,d^3x$$

and

$$a^\dagger(p)=-i\int_{x_0=t}\exp(-ipx)\,\overleftrightarrow{\frac{\partial}{\partial x_0}}\,\varphi_0(x)\,d^3x.$$

In general, one defines the time t annihilation and creation operators as follows:

$$a(p, t)=i\int_{x_0=t}\exp(ipx)\,\overleftrightarrow{\frac{\partial}{\partial x_0}}\,\varphi(x)\,d^3x$$

and

$$a^\dagger(p, t)=-i\int_{x_1=t}\exp(-ipx)\,\overleftrightarrow{\frac{\partial}{\partial x_0}}\,\varphi(x)d^3x$$

Theorem 5: The vacuum and the one-particle states are stable.

Proof: First we notice that if $|0\rangle$ is the vacuum state then $\varphi(x)|0\rangle=\varphi_0(x)|0\rangle$, for, by Theorem 3, $U(x)|0\rangle = |0\rangle$. Hence,

$$\varphi(x)|0\rangle=U^{-1}(x)\varphi_0(x)|0\rangle=U^{-1}(x)\int_p|p\rangle\exp(ipx)$$
$$=\int_p|p\rangle\exp(ipx)=\varphi_0(x)|0\rangle.$$

Now

$$a(p, t)|0\rangle=i\int_{x_0=t}\exp(ipx)\,\overleftrightarrow{\frac{\partial}{\partial x_0}}\,\varphi(x)|0\rangle$$
$$=\exp(ipx)$$

Hence, for any $t\in\mathbb{R}$, $a(p, t)$ annihilates the vacuum and the vacuum is stable. Also $a^\dagger(p, t)|0\rangle=a^\dagger(p)|0\rangle = |p\rangle$, and the creation of a particle from the vacuum state is time independent.■ ■

C. Crossing and PCT

We discuss how one can impose crossing relations on a theory of a spinless neutral particle. Assume that $\tilde{\Delta}_n$ is a symmetric function of n $\mathbb{R}^4$ variables with the properties:

(i) $\tilde{\Delta}_n^*(k_1, \ldots, k_n) = \tilde{\Delta}_n(-k_1, \ldots, -k_n)$,

(ii) $\tilde{\Delta}_n(\Lambda k_1, \ldots, \Lambda k_n) = \tilde{\Delta}_n(k_1, \ldots, k_n)$, $\Lambda \in SO(1,3)$.

Then the function

$$\Delta_n(x; y_1, \ldots, y_n) = \int d^4k_1 \cdots \int d^4k_n \exp[ik_1(y_1 - x)] \cdots \exp[ik_n(y_n - x)] \times \tilde{\Delta}_n(k_1, \ldots, k_n)$$

is symmetric in the y variables, with the properties:

(i) $\Delta_n^*(x; y_1, \ldots, y_n) = \Delta_n(x; y_1, \ldots, y_n)$,

(ii) $\Delta_n(\Lambda x + a; \Lambda y_1 + a, \ldots, \Lambda y_n + a) = \Delta_n(x; y_1, \ldots, y_n)$, $(\Lambda, a) \in E(1,3)$.

Consider the quadratic form

$$Q_n(x) = \int d^4y_1 \cdots \int d^4y_n \Delta_n(x; y_1, \ldots, y_n) : \varphi_0(y_1) \cdots \varphi_0(y_n): \tag{36}$$

We can write

$$\begin{aligned}\varphi_0(x) &= \int d^4p\, \theta(p_0)\delta(p^2 - m^2)\{a(p)\exp(-ipx) + a^\dagger(p) \times \exp(px)\} \\ &= \int d^4p\, \delta(p^2 - m^2)\{\theta(p^0)\, a(p) + \theta(-p^0)\, a^\dagger(-p)\} \times \exp(ipx)\}\end{aligned} \tag{37}$$

Substituting (37) in (36) we get

$$\begin{aligned}Q_n(x) &= \int d^4y_1 \cdots \int d^4y_n \int d^4k_1 \cdots \int d^4k_n \\ &\quad \times \exp[ik_1(y_1 - x)] \cdots \exp[ik_n(y_n - x)]\tilde{\Delta}_n(k_1, \ldots, k_n) \\ &\quad \times \int d^4p_1 \delta(p_1^2 - m^2) \cdots \int d^4p_n \delta(p_n^2 - m^2): \\ &\quad \times \{\theta(p_1^0)a(p_1) + \theta(-p_1^0)a^\dagger(-p_1)\} \cdots \\ &\quad \times \{\theta(p_n^0)a(p_n) + \theta(-p_n^0)a^\dagger(p_n)\}: \exp(-ip_1y_1) \cdots \\ &\quad \times \exp(-ip_ny_n) \\ &= \sum_{j=0}^{n} \int_{k'}^{(j)} \int_{k}^{(n-j)} C_n^j \tilde{\Delta}_n(-k_1', \ldots, -k_j', k_1, \ldots, k_{n-j}) \\ &\quad \times \exp[i(k_1' + \cdots + k_j' - k_1 - \cdots k_{n-j})x] \\ &\quad \times a^\dagger(k_1') \cdots a^\dagger(k_j') a(k_1) \cdots a(k_{n-j}).\end{aligned} \tag{38}$$

From (36) one gets the important formula

$$F^{(j,k)}(p_1', \ldots, p_j'; p_1, \ldots, p_k) = \frac{(j+k)!}{j!\,k!}\tilde{\Delta}_{j+k}(-p_1', \ldots, -p_j', p_1, \ldots, p_k). \tag{39}$$

Equation (39) connects different physical transitions. Although, upon integrating $Q_n(x)$ over $\mathbb{R}^4$, the contribution from the terms with less than two creation or annihilation operators vanishes, the form (36) is only locally integrable. This is a consequence of Theorem 3.

We define the "R product" $R\{:\varphi(y_1) \cdots \varphi(y_n):\}$ as the initial product, but with the terms of less than two creation or annihilation operators dropped. Hence we now define

$$G_n(x) = \int d^4y_1 \cdots \int d^4y_n \Delta_n(x; y_1, \ldots, y_n) R\{:\varphi(y_1) \cdots \varphi(y_n):\} \tag{40}$$

and the effective potential density takes the form $G(x) = \sum_{n=4}^{\infty} G_n(x)$.

To prove PCT invariance, all that is necessary then is $\tilde{\Delta}_n(-p_1, -p_2, \ldots, -p_n) = \tilde{\Delta}_n(p_1, p_2, \ldots, p_n)$. But this is a consequence of Poincaré invariance, because $\tilde{\Delta}_n$ should be formed by the scalar product of 4-momenta.

ACKNOWLEDGMENTS

The author would like to thank the Lebanese University Council for supporting this work and the National Council for Scientific Research in Lebanon for providing facilities. Thanks are also due to Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for its hospitality at the International Centre for Theoretical Physics, Trieste.

APPENDIX TO SEC. I.

A. The asymptotic limit of nonbounded operators

If A is a nonbounded operator and $\mathcal{D}_A$ is the set of all vectors $f \in \mathcal{H}$ such that $\|A_0(t)f\|$, $\|A_0(t)Sf\|$, and $\|A_0^\dagger(t)U(t)f\|$ are bounded for all $t \in \mathbb{R}$, then for every $f, g \in \mathcal{D}_A$,

$$\underset{t\to-\infty}{\text{w-lim}}\langle f, \{\hat{A}(t) - A_0(t)\}g\rangle = \underset{t\to+\infty}{\text{w-lim}}\langle f, \{\hat{A}(t) - S^{-1}A_0(t)S\}g\rangle = 0.$$

The proof follows the same steps, with the exception of the following:

set $M_f = \sup_{t\in\mathbb{R}} \|A_0^\dagger(t)U(t)f\|$ and $M_g = \sup_{t\in\mathbb{R}}\|A_0(t)\, g\|$, then

$$|\langle f, \{\hat{A}(t) - A_0(t)\}g\rangle| \leq M_f\|\{U(t) - I\}g\| + M_g\|\{U(t) - I\}f\|.$$

B. The integrability of the effective potential

Let $f \in \mathcal{H}$, then $\|U(t)f - f\|^2 = \langle U(t)f - f, U(t)f - f\rangle = 2\langle f, f\rangle - 2\,\mathrm{Re}\,\langle f, U(t)f\rangle$. Now $U(t) = \exp(itH_0/\hbar)W_-^\dagger \times \exp(-itH_0/\hbar) = \int_0^{2\pi}\exp(i\lambda)\, dE_\lambda(t)$, where $E_\lambda(t) = \exp(itH_0/\hbar)E_\lambda\exp(-itH_0/\hbar)$. Since s-$\lim_{t\to-\infty}U(t) = I$, it follows that $\lim_{t\to-\infty}\mathrm{Re}\langle f, U(t)f\rangle = \langle f, f\rangle$. But $\int_0^{2\pi} d\langle f, E_\lambda(t)f\rangle = \langle f, E_{2\pi}(t)f\rangle - \langle f, E_0(t)f\rangle = \langle f, f\rangle$. Hence

$$\lim_{t\to-\infty} \mathrm{Re}\int_0^{2\pi}[\exp(i\lambda)-1]\,d\langle f, E_\lambda(t)f\rangle=0$$

or

$$\lim_{t\to-\infty}\int_0^{2\pi}\sin^2\frac{\lambda}{2}\,d\langle f, E_\lambda(t)f\rangle=0.$$

(a) Assume that there exists a $\delta>0$ such that $E_{2\pi-\delta}=I$, then

$$0\leq\int_{2\pi-\delta}^{2\pi}\sin^2\frac{\lambda}{2}\,d\langle f, E_\lambda(t)f\rangle\leq\int_{2\pi-\delta}^{2\pi}d\langle f, E_\lambda(t)f\rangle$$

$$=\langle f, E_{2\pi}(t)f\rangle-\langle f, E_{2\pi-\delta}(t)f\rangle=0.$$

Hence

$$\lim_{t\to-\infty}\int_0^{2\pi-\delta}\sin^2\frac{\lambda}{2}\,d\langle f, E_\lambda(t)f\rangle=0.$$

But, if $0\leq\lambda\leq2\pi-\delta$, then $0\leq\lambda/(2\pi-\delta)\sin\delta/2\leq\sin\lambda/2$, hence

$$\lim_{t\to-\infty}\int_0^{2\pi-\delta}\lambda^2 d\langle f, E_\lambda(t)f\rangle$$

$$\leq\lim_{t\to-\infty}\int_0^{2\pi-\delta}\frac{(2\pi-\delta)^2}{\sin^2\delta/2}\sin^2\frac{\lambda}{2}\,d\langle f, E_\lambda(t)f\rangle=0.$$

But

$$\int_0^{2\pi-\delta}\lambda^2 d\langle f, E_\lambda(t)f\rangle=\|A(t)f\|^2;$$

hence

$$\text{s-lim}_{t\to-\infty}A(t)=0 \text{ and } U(t)=\exp\{i\int_{-\infty}^{t}F(t')\,dt'\}.$$

(b) [Weaker than (a)] Assume that for every $f\in\mathcal{H}$ and $\epsilon>0$ there exists $\delta>0$ and $L>0$ such that for every $t<-L$, $\langle f, E_{2\pi-\delta}(t)f\rangle>1-\epsilon$. Then, given $f\in\mathcal{H}$ and $\epsilon>0$, one can find two numbers $\delta>0$ and $L_1>0$ such that for any $t<-L_1$ we have $\langle f, E_{2\pi-\delta}(t)f\rangle>1-\epsilon/8\pi^2$. Also, one can find a number $L_2>0$ such that for any $t<-L_2$ we have

$$\int_0^{2\pi}\sin^2\frac{\lambda}{2}\,d\langle f, E_\lambda(t)f\rangle<\left(\frac{2\pi-\delta}{\sin\delta/2}\right)^{-2}\frac{\epsilon}{2}.$$

Take $L=\max(L_1, L_2)$. Hence, for every $t<-L$ we have

$$\int_0^{2\pi}\lambda^2 d\langle f, E_\lambda(t)f\rangle$$

$$=\int_0^{2\pi-\delta}\lambda^2 d\langle f, E_\lambda(t)f\rangle+\int_{2\pi-\delta}^{2\pi}\lambda^2 d\langle f, E_\lambda(t)f\rangle$$

$$\leq\int_0^{2\pi-\delta}\frac{(2\pi-\delta)^2}{\sin^2\delta/2}\sin^2\frac{\lambda}{2}\,d\langle f, E_\lambda(t)f\rangle$$

$$+\int_{2\pi-\delta}^{2\pi}4\pi^2 d\langle f, E_\lambda(t)f\rangle$$

$$\leq\frac{\epsilon}{2}+4\pi^2(1-\langle f, E_{2\pi-\delta}(t)f\rangle)\langle\frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon.$$

Hence, $\lim_{t\to-\infty}\int_0^{2\pi}\lambda^2 d\langle f, E_\lambda(t)f\rangle=0$ or $\text{s-lim}_{t\to-\infty}A(t)=0$.

APPENDIX TO SEC. II.

Proof of Theorem 3: Since $G(x)$ is strongly integrable, it follows that $G(x)G(y)$ is weakly integrable. Now, $\{(x, y): x\in\mathbb{R}^4 \text{ and } y\in\Gamma_-(x)\}\subseteq\mathbb{R}^4\times\mathbb{R}^4$. Hence $\int_{x\in\mathbb{R}^4}d^4x \times\int_{y\in\Gamma_-(x)}d^4y\,G(x)G(y)$ exists. But $H(x)=G(x) \times\int_{y\in\Gamma_-(x)}d^4y\,G(y)$ is a covariant function of $x\in\mathbb{R}^4$. Hence the integral $\int_{x\in\mathbb{R}^4}H(x)\,d^4x$ diverges if $G(x)$ does not annihilate $\mathcal{J}^{(0)}\oplus\mathcal{J}^{(1)}$, as can be seen explicitly. ∎

[1]P. A. M. Dirac, "Lectures on quantum field theory," Belfer Graduate School of Science, Yeshiva University, New York, 1966.

[2]R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964).

[3]J. Glimm and A. Jaffe, "Boson quantum field models," in *Mathematics of Contemporary Physics*, edited by R. F. Streater (Academic, New York, 1972).

[4]N. M. Queen and G. Violini, *Dispersion Theory in High Energy Physics* (Macmillan, London and Basingstoke, 1974).

[5]M. P. Stapp, "Analytic S-matrix Theory" in *High-Energy Physics and Elementary Particles* (IAEA, Vienna, 1965).

[6]N. N. Bogulyubov, B. V. Medvedev, and M. K. Polivanov, *Theory of Dispersion Relations* (Moscow State Technical Publishing House, 1958).

[7]H. D. Doebner and W. Lücke, J. Math. Phys. **18**, 138 (1977).

[8]S. Weinberg, Phys. Rev. **133**, B1318 (1964).

[9]S. Weinberg, Phys. Rev. **134**, B882 (1964).

[10]Private communication from Dr. H. D. Doebner, at the ICTP, Trieste, 1977.

[11]J. M. Jauch, Helv, Phys. Acta **31**, 127 (1958).

[12]E. Nelson, Ann. Math. **70**, 572 (1959).

[13]J. M. Jauch, Helv, Phys. Acta **31**, 661 (1958).

[14]N. J. Akhiezer and I. M. Glazman, *Theory of Linear Operators in Hilbert Space* (Ungar, New York, 1963), Vol. II.

Conditioning of states. II

Platon C. Deliyannis

Department of Mathematics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 18 October 1977)

An alternative axiomatic system describing the concept of conditioning (or preparation) of states in a quantum logic is proposed and its consequences developed. The main difference between this and the system proposed by Pool is our reliance on Mielnik's idea of transition probabilities, thereby avoiding some ad hoc hypotheses as well as the theory of Baer *-semigroups.

1. INTRODUCTION

The purpose of this paper is to study the change a state of a system has to undergo, given the occurence of an event, classically known as the conditioning of a state by the occurence of an event. In particular we wish to point out how the basic structure of the set $\mathcal{L}$ of all events can be recaptured, given the conditioning maps. We have used this point of view in Ref. 1 but in the restricted case of a "pure" conditioning. Pool has given, in Ref. 2, a discussion of this idea, and it is important to note the similarities and the differences. For the reader's convenience we have included as an appendix a list of Pool's axioms. The main tool used by Pool was the Baer *-semigroup of all operations he constructs using his axioms II.4 and II.5. These we shall not assume because their physical significance is, in the writer's opinion, not immediate. Instead, we shall make use of the concept of transition probability between two states, by assuming essentially an extension of Mielnik's axioms.[3]

In Sec. 2 we provide the basic hypotheses on $\mathcal{L}$ and derive the results we shall need later. Section 3 contains our discussion of the conditioning axioms, and Sec. 4 gives the characterization of the basic structural elements of $\mathcal{L}$ in terms of the conditioning maps. In Sec. 5 we establish that Axiom II.8 of Pool holds, from which the connection between semimodularity and pure conditionings follows. Finally, in Sec. 6 we discuss the consequences of the hypothesis that every state is a mixture of pure states, which can serve to replace some of our original axioms.

2. THE BACKGROUND HYPOTHESES

We shall assume that the set $\mathcal{L}$ of all events is an ortholattice (relative to $\leq$ and $'$) and that the set $\mathcal{M}$ of all states is quite full, i.e., that for any $A,B\in\mathcal{L}$, if $mB=1$ for all $m\in\mathcal{M}$ for which $mA=1$, then $A\leq B$. Write $\mathcal{P}$ for the set of all pure states.

Our next assumption is the existence of a support for each state [Axiom I.8(a) of Pool]: Given any $m\in\mathcal{M}$, the event $L_m=\inf\{A\mid mA=1\}$ exists and $m(L_m)=1$.

Propositon 1: The Jauch–Piron–Zierler (JPZ) Axiom holds: if $A=\inf\{A_i\}$ and $mA_i=1$ for all i, then $mA=1$ also.

Proof: Since $mA_i=1$, we have $L_m\leq A_i$ for all i, hence $L_m\leq A$; thus $mA\geq m(L_m)=1$.

Remark: For any $m,n\in\mathcal{M}$ we have $m(L_n)=0$ iff $L_m\perp L_n$ iff $n(L_m)=0$.

Next we assume Axiom I.9(a) of Pool: Given any $A\neq 0$, there exists a pure state m such that $mA=1$, i.e., $L_m\leq A$.

Now consider an event $A\neq 0$, and let $\mathcal{P}_A$ be a temporary notation for the set of all $m\in\mathcal{P}$ such that $mA=1$; by the above, this is just $\{m\in\mathcal{P}\mid L_m\leq A\}$ and is $\neq\emptyset$. Consider all possible subsets $\mathcal{S}$ of $\mathcal{P}_A$ for which $m,n\in\mathcal{S}$, $m\neq n$ implies $L_m\perp L_n$; such $\mathcal{S}$ exist (singleton sets for example). Hence by Zorn's Lemma we immediately see that there exists at least one such subset $\mathcal{S}_0$ of $\mathcal{P}_A$, maximal with respect to inclusion.

Proposition 2: Given an event $A\neq 0$ in $\mathcal{L}$ let $\{m_\alpha\}$ be a set of pure states as described above and let L_α be the support of m_α. Thus assume that (i) $m_\alpha(A)=1$ for each α, (ii) $L_\alpha\perp L_\beta$ for $\alpha\neq\beta$, and (iii) if $m\neq m_\alpha$ (all α), then L_m is not disjoint from all L_α. Then $A=\sup_\alpha\{L_\alpha\}$, and for any state m we have $mA=\sum_\alpha m(L_\alpha)$.

Proof: Since $L_\alpha\leq A$ for all α, we consider an upper bound B of $\{L_\alpha\}$, and assume that $A\wedge B\neq A$, i.e., that C, defined as $A\wedge(A\wedge B)'$, is $\neq 0$. Then there exists a pure state n with $nC=1$. Since $C\leq A$ we have $nA=1$, or $n\in\mathcal{P}_A$. On the other hand each $L_\alpha\leq A\wedge B$, hence $L_\alpha\perp L_n$ for all α, which contradicts hypothesis (iii). Therefore, $C=0$, or $A\wedge B=A$, i.e., $A\leq B$, which means $A=\sup\{L_\alpha\}$. Now consider any state m; if $mA=0$, then $m(L_\alpha)=0$ also and $mA=\sum_\alpha m(L_\alpha)$ holds. So let $mA\neq 0$. Since for any finite number of indices $\alpha_1,\alpha_2,\ldots,\alpha_k$ we have $\sum m(L_{\alpha_i})=m(\sum L_{\alpha_i})\leq 1$, we see that $\{\alpha\mid m(L_\alpha)\neq 0\}$ is countable, and we write it as $\{\alpha_1,\alpha_2\cdots\}$. Let $D=\sum_i L_{\alpha_i}(\leq A)$ and $E=A\wedge D'$. Then $mA=mD+mE$, $mD=\sum m(L_{\alpha_i})=\sum_\alpha m(L_\alpha)$. For $\alpha\neq\alpha_i$ we have $L_\alpha\perp L_{\alpha_i}$, hence $L_\alpha\perp D$, which implies $L_\alpha\leq E$. Also $\{L_\alpha\mid\alpha\neq\text{all }\alpha_i\}$ is maximal disjoint in E, because if $L_n\leq E$ is disjoint from all such L_α, L_n will be disjoint from all the L_α given originally, since $E\perp D=\sum L_{\alpha_i}$, which is impossible. Thus $E=\sup\{L_\alpha\mid\alpha\neq\text{all }\alpha_i\}$, and since $m(L_\alpha)=0$ for $\alpha\neq\alpha_i$ we have by JPZ that $mE=0$. Hence $mA=mD=\sum_\alpha m(L_\alpha)$.

Our major departure from Pool's system will now be discussed. We assume that to each pair of states m, $n\in\mathcal{M}$ a number $\langle m\mid n\rangle$ corresponds, which gives the probability of spontaneous transition. Specifically we assume that: $0\leq\langle m\mid n\rangle\leq 1$, $\langle m\mid n\rangle=\langle n\mid m\rangle$. Since L_n is the "cause" of all events occuring with certainty in the state n, it might appear reasonable to assume that $\langle m\mid n\rangle$ is the same as $m(L_n)$, i.e., to identify the occurrence of L_n with the system's being in state n. This, however, is not quite allowed. For example, it is not true in general that $L_{n_1}=L_{n_2}$ implies $n_1=n_2$, and in fact

 0022-2488/78/1911-2341$1.00

the classical Hilbert space model with $\langle m|n\rangle = \mathrm{Tr}(TS)$ (T,S the density operators representing m,n) violates this unless n_1, n_2 are pure. We shall therefore impose the condition $\langle m|n\rangle = m(L_n)$ only for pure states n and assume Mielnik's Axiom: For pure m,n we have $\langle m|n\rangle = 1$ iff $m=n$. This property justifies our identification of the occurrence of L_n with the system's transition to the state n (for n pure):

Proposition 3: For n_1, n_2 pure states we have $L_{n_1} = L_{n_2}$ iff $n_1 = n_2$.

Proof: Assume $L_{n_1} = L_{n_2}$; then $n_2(L_{n_1}) = n_2(L_{n_2}) = 1$, i.e., $\langle n_2|n_1\rangle = 1$ hence $n_1 = n_2$.

Note also that this proposition implies Mielnik's axiom: Suppose $\langle m|n\rangle = 1$; then $m(L_n) = 1$ hence $L_m \leq L_n$; by symmetry $L_n \leq L_m$ also, hence $L_n = L_m$, and so $n = m$.

We now summarize all our axioms.

Axiom 1: The set $\mathcal{L}$ of events is an ortholattice with respect to $\leq$ (implication) and $'$ (complementation) with maximum element I and minimum element 0.

Axiom 2: The set $\mathcal{M}$ of states is quite full: For any A, B, if $mB = 1$ for each state for which $mA = 1$, then $A \leq B$.

Axiom 3: For each $m \in \mathcal{M}$, the event $L_m = \inf\{A \mid mA = 1\}$ exists, and $m(L_m) = 1$.

Axiom 4: For each $A \neq 0$ there exists a pure state m such that $mA = 1$.

Axiom 5: To each $m, n \in \mathcal{M}$ there corresponds a number $\langle m|n\rangle$ such that:

(i) $0 \leq \langle m|n\rangle \leq 1$,

(ii) $\langle m|n\rangle = \langle n|m\rangle$,

(iii) For $n \in \mathcal{P}$ we have $\langle m|n\rangle = m(L_n)$,

(iv) For $m, n \in \mathcal{P}$ we have $\langle m|n\rangle = 1$ iff $m = n$.

Using Axioms 4 and 5 we can characterize the atoms of $\mathcal{L}$.

Proposition 4: The event A is an atom of $\mathcal{L}$ iff $A = L_m$ for some pure state m; such a state is unique.

Proof: Consider a pure state m, and let $B \leq L_m$, $C = L_m \wedge B' \neq 0$; then there exists a pure state n with $nC = 1$, hence $n(L_m) = 1$. But then $m = n$, hence $mC = 1$ and so $L_m \leq C$, which implies that $B = 0$. Thus L_m is an atom. Conversely, let A be an atom and consider a pure state m with $mA = 1$. Then $L_m \leq A$ and since $L_m \neq 0$ we have $A = L_m$. Uniqueness follows from proposition 3.

To see that L_m does not determine m when m is not pure, let $m_1, m_2, \cdots$ be pure states such that $\langle m_i|m_j\rangle = 0$ for $i \neq j$ and choosing $a_i > 0, \Sigma a_i = 1$, let $m = \Sigma a_i m_i$ be the mixture. We shall show that $L_m = \Sigma L_{m_i}$. Let $A = \Sigma L_{m_j}$ and note that $mB = 1$ implies $m_i B = 1$ or $L_{m_i} \leq B$, hence $A \leq B$. Thus $A \leq L_m$. On the other hand $mA = \Sigma_j m(L_{m_j}) = \Sigma_j \Sigma_i a_i m_i(L_{m_j})$, but $m_i(L_{m_j}) = \delta_{ij}$ and so $m(L_{m_j}) = a_j$ and $mA = \Sigma_j a_j = 1$, i.e., $L_m \leq A$. Finally note that changing the a_i changes m [since $m(L_{m_j}) = a_i$] but does not influence A.

A lemma, crucial for future results, will be established now.

Lemma 1: Let $a_1\langle m_1|n\rangle = a_2\langle m_2|n\rangle$ for all $n \in \mathcal{M}$. Then $a_1 = a_2$, and, if they are nonzero, we also have $m_1 = m_2$.

Proof: For pure n the hypothesis becomes $a_1 m_1(L_n) = a_2 m_2(L_n)$. Consider by Zorn a set of pure states $\{m_\alpha\}$ such that $\{L_{m_\alpha}\}$ are disjoint, and maximal relative to this property. Then $I = \sup\{L_{m_\alpha}\}$ and $\Sigma_\alpha m(L_{m_\alpha}) = 1$ by Proposition 2. Thus $a_1 = a_1 \sum_\alpha m_1(L_{m_\alpha}) = a_2 \sum_\alpha m_2(L_{m_\alpha}) = a_2$.

Now for $a_1, a_2 \neq 0$ we get $m_1(L_n) = m_2(L_n)$, and using a maximal disjoint family $\{L_{n_\alpha}\}$ all contained in A, we obtain as before $m_1 A = m_2 A$, for all A, hence $m_1 = m_2$.

3. THE CONDITIONING OF STATES

We shall now assume that the occurrence of an event A while the system is in a state m will produce a transition to a state, which we shall write as $m_{:A}$, and call "the state m conditioned by (the occurrence of) A." To include the case where it is impossible for A to occur in a given state m, we shall augment $\mathcal{M}$ by a formal state θ and assume the following:

Axiom 6: For any $A \in \mathcal{L}$, $m \in \mathcal{M} \cup \{\theta\}$ we have:

(i) $m_{:A} = \theta$ iff $m(A) = 0$; $\theta_{:A} = \theta$

(ii) $m_{:A:A} = m_{:A}$

(iii) $m_{:A} = m$ iff $m(A) = 1$

(iv) $\langle m|n\rangle = m(A)\,\langle m_{:A}|n\rangle$ for each $n \in \mathcal{M}$ with $n(A) = 1$.

The interpretation of (iv) is obvious: The transition from m to a state n in which A occurs with certainty is considered equivalent to the independent concatenation of the occurrence of A (hence the transition to $m_{:A}$) and the transition from $m_{:A}$ to n.

Remarks: (a) Such a map $m \to m_{:A}$ completely determines the event A, because if $m_{:A} = m_{:B}$ for all m, then $mA = 1$ iff $m = m_{:A}$ iff $m = m_{:B}$ iff $mB = 1$; but since $\mathcal{M}$ is quite full, this means $A = B$.

(b) Further, such a map $m \to m_{:A}$ is completely determined by its fixed points, because the set $\{m \mid m = m_{:A}\}$ is $\{\theta\} \cup \{m \mid mA = 1\}$, and the latter one determines A completely.

Proposition 5: For any $m, n \in \mathcal{M}$ we have $m(A)\langle m_{:A}|n\rangle = n(A)\langle n_{:A}|m\rangle$.

Proof: First note that we may assume not both $m(A)$, $n(A)$ to be zero. But if say, $m(A) = 0$, $n(A) \neq 0$, then $\langle m|n_{:A}\rangle = m(A)\langle m_{:A}|n_{:A}\rangle$ by (iv) of Axiom 6, since $n_{:A}(A) = 1$; thus $\langle m|n_{:A}\rangle = 0$ and the result holds. Finally let both $m(A), n(A) \neq 0$ [i.e., $m_{:A}(A) = n_{:A}(A) = 1$] and apply again the same axiom to obtain $m(A)\langle m_{:A}|n\rangle = m(A)\, n(A)\langle m_{:A}|n_{:A}\rangle = n(A)\,\langle m|n_{:A}\rangle = n(A)\langle n_{:A}|m\rangle$.

Corollary: A conditioning is uniquely determined by its effect on pure states.

Proof: Let $f_A(m), g_A(m)$ be the result of conditioning m by the event A, according to two conditionings, and assume $f_A(n) = g_A(n)$ for all pure n. Then we have $m(A) \times \langle f_A(m)|n\rangle = n(A)\,\langle m|f_A(n)\rangle = n(A)\,\langle m|g_A(n)\rangle = m(A) \times \langle g_A(m)|n\rangle$ for all pure states. Now if $mA = 0$, then $f_A(m) = \theta = g_A(m)$; if $mA \neq 0$, the argument in Lemma 1 applies to yield $f_A(m) = g_A(m)$.

We shall call a conditioning "pure" if $m_{:A}$ is (either θ or) pure for each pure state m. Note that there exists at most one pure conditioning on a given $\mathcal{L}$. In particular, as in Ref. 1 we have that if $m_{:A}$ is pure, it is the unique pure state in $\{m \mid mA=1\}$ which maximizes $\langle m \mid n\rangle$ under the constraint $nA=1$. First note that for $m_{:A}$ pure we have $\langle m \mid m_{:A}\rangle = m(A)\langle m_{:A} \mid m_{:A}\rangle = mA$, and so $\langle m \mid n\rangle = m(A)\ \langle m_{:A} \mid n\rangle \leqslant mA = \langle m \mid m_{:A}\rangle$, i.e., $m_{:A}$ maximizes $\langle m \mid n\rangle$. Also, if $\langle m \mid n\rangle = \langle m_{:A} \mid m\rangle$ for some pure n with $nA=1$, then again $\langle m \mid n\rangle = mA\ \langle m_{:A} \mid n\rangle$ implies $\langle m_{:A} \mid m\rangle = mA\ \langle m_{:A} \mid n\rangle$ or $m(A)=m(A)\langle m_{:A} \mid n\rangle$; since $mA \neq 0$ ($m_{:A}$ is pure) we have $\langle m_{:A} \mid n\rangle = 1$, and since $m_{:A}$, n are pure they are equal.

4. THE ALGEBRAIC STRUCTURE OF IN TERMS OF CONDITIONING

Characterizing A' in terms of the map $m \to m_{:A}$ is quite trivial.

Theorem 1: The map $m \to m_{:A'}$ is characterized by the property that its fixed points are the states mapped to θ by $m \to m_{:A}$.

Proof: Because $m(A')=1$ iff $mA=0$.

In case of successive conditionings it is convenient to omit the brackets; we shall write, e.g., $m_{:A:B}$ instead of $(m_{:A})_{:B}$.

Theorem 2: For any $A,B \in \mathcal{L}$ we have $B \leqslant A$ iff $m_{:B} = m_{:B:A}$.

Proof: Let $B \leqslant A$, and suppose $nB=1$; then $nA=1$ also, hence $n_{:A}=n$. But if $m_{:B} \neq \theta$, we have $m_{:B}(B)=1$, and so we obtain $m_{:B:A}=m_{:B}$; if $m_{:B}=\theta$ then $m_{:B:A}=\theta$ also. Conversely, if $m_{:B}=m_{:B:A}$ and we consider a state m with $mB=1$, we have $m_{:B}=m$, hence $m_{:A}=m_{:B:A}=m_{:B}=m$, i.e., $mA=1$ too. Thus $B \leqslant A$.

Proposition 6: If $B \leqslant A$, then $m_{:B}=m_{:A:B}$ for all m.

Proof: We have $m(B)\langle m_{:B} \mid n\rangle = n(B)\langle n_{:B} \mid m\rangle = n(B)\ n_{:B}(A) \times \langle n_{:B:A} \mid m\rangle$ since $n_{:B}(A)$ is either 1 or 0 (since $B \leqslant A$) and in either case the equality holds. But $n_{:B}(A)\ \langle n_{:B:A} \mid m\rangle = m(A)\ \langle m_{:A} \mid n_{:B}\rangle = m(A)\ m_{:A}(B)\ \langle m_{:A:B} \mid n\rangle$, and so we have for all m,n that $m(B)\ \langle m_{:B} \mid n\rangle = m(A)\ m_{:A}(B) \times \langle m_{:A:B} \mid n\rangle$. Thus by Lemma 1 we obtain $m(B) = m(A)m_{:A}(B)$, and if nonzero we have $m_{:B}=m_{:A:B}$. Now, if $m(B)=m(A)m_{:A}(B)=0$, we have $m_{:B}=\theta$ and either $m_{:A}=\theta$, or $m_{:A:B}=\theta$, which all lead to $m_{:B}=m_{:A:B}$ again.

Remark: It is not hard to see that there are several other ways of expressing the relation $B \leqslant A$ by means of conditioning: "$m_{:B}(A)=1$ for all m with $mB=1$" is one and "$m(B)=m(A)\ m_{:A}(B)$" is another; thus, in particular, Axiom II.6 of Pool holds. We prefer the way stated in Theorem 2 because it contains only the conditioning maps.

Theorem 3: For any $A,B \in \mathcal{L}$, we have $A \perp B$ iff $m_{:B:A} = \theta = m_{:A:B}$.

Proof: Since $A \perp B$ iff $A \leqslant B'$, assume the last and note that for $mA \neq 0$ we shall have $m_{:A}(B')=1$, or $m_{:A}(B)=0$. Thus either $mA=0$ or $m_{:A:B}=\theta$; in either case $m_{:A:B}=\theta$. Since $A \perp B$ is symmetric, we also have $m_{:B:A}=\theta$. Conversely, $m_{:A:B}=\theta$ implies $m_{:A}(B)=0$ or $m_{:A}(B')=1$, unless $m_{:A}=\theta$. Thus $m(A)=1$ will imply $m_{:A}=m \neq \theta$ and so $m(B')=1$ will hold, i.e., $A \leqslant B'$ will hold.

Theorem 4: The events A,B are compatible iff $m_{:A:B} = m_{:B:A}$ for all m. In such a case $A \wedge B$ produces the conditioning $m \to m_{:A:B}$.

Proof: Let $A=A_1+C, B=B_1+C$ with $C=A \wedge B$, and A_1, C, B_1 mutually orthogonal. Then $m_{:A_1:A}=m_{:A:A_1}$ for all m, and in particular we have $m_{:B:A_1:A}=m_{:B:A:A_1}$; since $B \perp A_1$ and $A_1 \leqslant A$, we find $m_{:B:A:A_1}=\theta$, or $m_{:B:A}(A_1)=0$. In case $m_{:B:A} \neq 0$, we have $1=m_{:B:A}(A) = m_{:B:A}(A_1+C) = m_{:B:A}(A_1)+m_{:B:A}(C)=m_{:B:A}(C)$. Thus $m_{:B:A}==m_{:B:A:C}$ and since $C \leqslant A, C \leqslant B$ we apply Proposition 6 to obtain $m_{:C}=m_{:B:A}$.

If $m_{:B:A}=\theta$, then $m_{:B:A:C}=\theta$ too, and the same arguments leads to $m_{:C}=m_{:B:A}$. Thus in all cases $m_{:C} = m_{:B:A}$ and since compatibility is symmetric we end up with $m_{:C}=m_{:A:B}$ also.

Conversely now, let $m_{:A:B}=m_{:B:A}$ and let $C=A \wedge B$. We must show that $A \wedge C' \perp B, B \wedge C' \perp A$; by symmetry it suffices to establish the first. So let $m(A \wedge C')=1$, i.e., $mA=1, mC=0$. Then $m_{:A}=m$, hence $m_{:A:B}=m_{:B}$, and so $m_{:B}=m_{:B:A}$. Suppose $m_{:B} \neq \theta$; then $m_{:B}(A)=1$ and since $m_{:B}(B)=1$ we have by JPZ that $m_{:B}(C)=1$; but $m(C)=m(B)m_{:B}(C)$, and so $0=m(C)=m(B) \neq 0$ gives a contradiction. Therefore, $m_{:B}=\theta$, i.e., $m(B)=0$, or $m(B')=1$. Thus $m(A \wedge C')=1$ implies $m(B')=1$, or $A \wedge C' \leqslant B'$, which is the desired conclusion.

5. CHARACTERIZATION OF CONDITIONINGS

We shall now calculate the support of $m_{:A}$ given any conditioning. Even though this will not characterize the map $m \to m_{:A}$ unless $m_{:A}$ is pure, the theorem below is quite useful. Note that is just Axiom II.8 of Pool.

Theorem 5: If $mA \neq 0$, then $L_{m_{:A}} = (L_m \vee A') \wedge A$.

Proof: Clearly $L_{m_{:A}} \leqslant A$ since $m_{:A}(A)=1$. We shall first show that $L_{m_{:A}} \leqslant L_m \vee A'$, by showing that $m_{:A}(L_m \vee A')=1$, or equivalently $m_{:A}(L'_m \wedge A)=0$. Note that $L'_m \wedge A \leqslant L'_m$, hence $m(L'_m \wedge A)=0$. Since $m(L'_m \wedge A) = m(A)m_{:A}(L'_m \wedge A)$, while $m(A) \neq 0$, we have the desired conclusion. So we have reached the half point: $L_{m:A} \leqslant L_m \vee A') \wedge A$. For the reverse, note that if it is false, then by orthomodularity there is a pure state n such that $n((L_m \vee A') \wedge A)=1, n(L_{m:A})=0$. Thus we have $n(A)=1$, $n(L_m \vee A')=1$. Also $n(L_{m:A})=0$ implies $n(L_{m:A'})=1$, hence $L_n \leqslant L'_{m:A}$, or $m_{:A}(L_n)=0$; but this means $\langle n \mid m_{:A}\rangle = 0$. Since $n=n_{:A}$ we use part (iv) of the conditioning axiom to obtain $\langle n_{:A} \mid m\rangle = m(A)\langle n_{:A} \mid m_{:A}\rangle = 0$, i.e., $\langle n \mid m\rangle = 0$. But again, as this means $m(L_n)=0$ or $m(L'_n)=1$ or $L_m \leqslant L'_n$, or $n(L_m)=0$, and since $n(A')=0$, we apply JPZ to obtain $n(L_m \vee A')=0$— a contradiction.

Corollary (Pool): If a pure conditioning exists for $\mathcal{L}$, then $\mathcal{L}$ is semimodular.

Proof: The argument given by Pool still holds. For the converse, i.e., that on a semimodular $\mathcal{L}$ every conditioning is pure, we need to have that if L_m is an atom of $\mathcal{L}$, then m is pure. This is essentially assumed by Pool in the form of Axiom I.9(b). It does not appear to follow from our axioms, but we can establish that it holds under another hypothesis which at the same time serves various other purposes. We discuss this in the next section

It is interesting to note that in case of a semimodular $\mathcal{L}$ we can always define a map $m \to m_{:A}$ from $\mathcal{P}$ to $\mathcal{P}$ satisfying the first three conditions of Axiom 6, motivated by the conclusion of Theorem 5. If $mA=0$, i.e., $L_m \leq A'$, we have $(L_m \vee A')\wedge A = A' \wedge A = 0$, so $m_{:A}$ is naturally defined as θ; for $mA \neq 0$, $(L_m \vee A')\wedge A$ is an atom, hence of the form L_n for a unique pure state n, which we define to be $m_{:A}$. Thus we have (i). Part (iii) is easy: If $mA=1$, then $L_m \leq A$, hence L_m, A, A' are mutually compatible, and so $L_{m_{:A}} = (L_m \vee A')\wedge A = (L_m \wedge A)\vee(A'\wedge A) = L_m$, and by uniqueness $m_{:A}=m$. On the other hand, if $m = m_{:A}$, then $L_m = (L_m \vee A')\wedge A \leq A$, hence $mA=1$. To verify $m_{:A:A} = m_{:A}$ we must calculate $\{[(L_m \vee A')\wedge A]\vee A'\}\wedge A$. Now $(L_m \vee A')\wedge A$ and A' are disjoint, hence form a modular pair; since $(L_m \vee A')\wedge A \leq A$, we have by semimodularity that we can distribute $\{[L_m \vee A')\wedge A]\vee A'\}\wedge A = \{[(L_m \vee A')\wedge A\}\vee(A'\wedge A) = (L_m \vee A')\wedge A \wedge A = (L_m \vee A')\wedge A$, which is precisely the relation $m_{:A:A} = m_{:A}$.

6. PURE GENERATION OF STATES

We say that a state m is a mixture of a family $\{m_x\}_{x\in X}$ of states, if X carries a positive measure μ such that $mA = \int_X m_x(A)d\mu(x)$ for all $A \in \mathcal{L}$; we write $m = \int_X m_x d\mu(x)$. The case of discrete mixtures is contained in this by considering a sequence of points $\{x_k\}$ in X and letting $\mu\{x_k\} = a_k$. We have used such mixtures in Sec. 2, which we write as $\Sigma a_k m_k$.

The hypothesis mentioned in the previous section is the following

(PGS): There exists a set X, and a σ-algebra of subsets such that to each $m \in M$ there corresponds a family $\{m_x\}_{x\in X}$ in $\mathcal{P}$ and a positive measure μ on X with $m = \int_X m_x d\mu(x)$.

This hypothesis has an obvious physical content, and is implicit in all applications. The first thing to observe is that PGS implies Axiom 4.

Proposition 7: Assume PGS; then, if $A \neq 0$, there exists a pure state m with $mA=1$.

Proof: By Gudder's argument[4] there exists a state n for which $nA=1$ (otherwise $\{n \mid nA=1\} = \phi = \{n \mid n0 = 1\}$, i.e., $A=0$ by the quite fullness of $\mathcal{M}$). Write m as $\int_X m_x d\mu(x)$; then $\int_X m_x(A)d\mu(x)=1$, and since $\mu X=1$, $m_x(A)\leq 1$, we have $m_x(A)=1$ a.e. in X; then any one of these m_x will do.

We shall now determine all $A \in \mathcal{L}$ which are supports of a given m, by using PGS. First we define a family of atoms $\{A_x\}_{x\in X}$ as being measurable in case for each $B \in \mathcal{L}$ the function $x \to m_x(B)$ is measurable, where m_x is the unique pure state having A_x as support.

Proposition 8: The event A is the support of the state m if there exists a measureable family $\{A_x\}$ of atoms of atoms in A and a (positive) measure μ on X such that: (i) $\mu E=0$ implies $A = \sup\{A_x \mid x \notin E\}$, (ii) $m = \int_X m_x d\mu(x)$.

Proof: Let $A = L_m$ and write $m = \int_X m_x d\mu(x)$ with $m_x \in \mathcal{P}$. Since $mA=1$, we have $m_x(A)=1$ a.e. relative to μ, and suppressing the exceptional m_x will not affect the equation $m = \int_X m_x d\mu(x)$. Thus we assume $m_x A = 1$ for all x, and writing A_x for the support of m_x, we have $A_x \leq A$. Now let $\mu E=0$ and suppose that $A_x \leq B$ for $x \notin E$. Then $A \wedge B \geq A_x$ for $x \notin E$ hence $m_x(A\wedge B)=1$ for $x \notin E$ and since $\mu E=0$ we have $m(A\wedge B)=1$, i.e., $A\wedge B \geq L_m = A$. Therefore, $B \geq A$ and so $A = \sup\{A_x \mid x\notin E\}$. For the converse let $A = \sup\{A_x \mid x \notin E\}$ for any E with $\mu E=0$, while $m = \int_X m_x d\mu(x) (A_x = L_m)$. Since $m_x A=1$ a.e., we have $mA=1$, or $L_m \leq A$. But if $mB=1$, then $m_x B=1$ except for $x\in E$ for some $E \subseteq X$ with $\mu E=0$. Thus $A_x \leq B$ for $x \notin E$, and so $A \leq B$ for any B with $mB=1$. Thus $A \leq L_m$, and $L_m = A$.

Of more immediate importance is the following;

Proposition 9: Let PGS hold, and let L_m be an atom; then m is pure. Thus Axiom II.9(b) of Pool will hold.

Proof: Again, write $m = \int_X m_x d\mu(x)$ and note that $m_x(L_m)=1$ a.e.; thus $0 \leq L_{m_x} \leq L_m$. But since L_m is an atom we have $L_{m_x} = L_m$ a.e.; however, there is a pure state n such that $L_m = L_n$, and so $L_{m_x} = L_n$, i.e., $m_x = n$ a.e. But this implies $m = \int_X n d\mu(x) = n\int_X d\mu(x) = n$, i.e., m is pure.

Corollary (Pool): If PGS holds and $\mathcal{L}$ is semimodular, then any conditioning is pure, hence unique.

Our final use of PGS will be to weaken the hypotheses on our transition probabilities, provided that the decompositions of a state m as $\int_X m_x d\mu(x)$ are "strongly" measurable, i.e., not only $x \to m_x(A)$ is measurable for all $A \in \mathcal{L}$, but also $x \to n(L_{m_x})$ is measurable for each $n \in \mathcal{M}$. In such a case it suffices to be given a Mielnik form $\langle\ |\ \rangle$ on $\mathcal{P}$ i.e., one which satisfies

(i) $0 \leq \langle m \mid n\rangle \leq 1$,

(ii) $\langle m \mid n\rangle = \langle n \mid m\rangle = n(L_m)$,

(iii) $\langle m \mid n\rangle = 1$ iff $m=n$.

[Because then, for arbitrary $m, n \in M$, one writes $m = \int_X m_x d\mu(x), n = \int_X n_y d\nu(y)$ and defines $\langle m \mid n\rangle$ to be $\int\int_{X\times X}\langle m_x \mid n_y\rangle d\mu(x)d\nu(y)$.] Existence is guaranteed by our assumptions and it is not hard to see that Axiom 5 is valid.

APPENDIX

We reproduce here for the reader's convenience the axioms employed by Pool,[2] stated in our notation.

Axioms I.1–I.7 amount to the assumption that $\mathcal{L}$ is an orthomodular set, closed under countable disjoint suprema, with quite a full set of states closed under countable mixtures. Two more axioms are assumed:

I.8(a) Given m, the $\inf\{A \mid mA=1\} = L_m$ exists and $m(L_m)=1$.

(b) If $A \neq 0$, there exists a state m such that $A = L_m$.

I.9(a) If $A \neq 0$, then there exists a pure state m with $mA=1$.

(b) The state m is pure iff there exists an event A such that $nA=1$ iff $n=m$.

The definition of a conditioning contains the following:

II.1 For each A the domain is $\{m \mid mA \neq 0\}$.

II.2 If $mA=1$, then $m_{:A}=m$.

II.3 $m_{:A}(A)=1$ for each m in the domain.

II.4 If for all m we have $m_{:A_1:A_2:\cdots:A_k}=m_{B_1:B_2:\cdots B_l}$, then $m_{:A_k:A_{k-1}:\cdots:A_1}=m_{:B_l:B_{l-1}:\cdots:B_1}$.

II.5 Given $A_1, A_2, \cdots, A_n$, there exists an A such that $\{m \mid mA=1\}$ is the complement of the domain of the composition of the maps $m \to m_{:A_i}$.

II.6 If $B \leq A$, then $m(B)=m(A)\ m_{:A}(B)$.

II.7 If A, B are compatible, then $m_{:A}(B)=m_{:A}(A \wedge B)$.

II.8 If $mA \neq 0$, then $L_{m_{:A}}=(L_m \vee A') \wedge A$.

[1]P. C. Deliyannis, J. Math. Phys. **17**, 653 (1976).
[2]J. C. T. Pool, Commun. Math. Phys. **9**, 118 (1968).
[3]B. Mielnik, Commun. Math. Phys. **9**, 55 (1968).
[4]S. P. Gudder, Pac. J. Math. **19**, 81 (1968).

A unified radon inversion formula[a)]

Stanley R. Deans[b)]

Donner Laboratory, Lawrence Berkeley Laboratory, Berkeley, California 94720

A Radon inversion formula which holds in spaces of even or odd dimension n is obtained for functions which admit to a certain general decomposition. The inversion formula which is one member of a Gegenbauer transform pair is used to generate some interesting definite integrals involving special functions. Legendre and Tchebycheff transform pairs are discussed as special cases of the general result.

1. INTRODUCTION

The work reported here is part of an ongoing effort[1,2] to obtain a better understanding of the integral equations which emerge in a natural way when studying the Radon transform on Euclidean space R^n. The theory of this transform is the foundation for an enormous number of diverse applications. These range from macroscopic to microscopic (astrophysics and molecular biology, for example) and include medical applications as an intermediate case. In all of these applications the central aim is to obtain certain information about the internal structure of an object (or collection of objects) either by passing some probe (such as x rays) through the object or by making use of the fact that the object itself is a self-emitting source, such as an organ in the body which contains a radioactive isotope or the interior of the earth when motions occur. Reviews of many of these applications may be found in the articles by Budinger and Gullberg,[3] Gordon and Herman,[4] and Brooks and DiChiro.[5]

The Radon transform is not new, having originated with a paper by Radon[6] in 1917. Since that time various authors have contributed to an understanding of many technical aspects of the transform.[7-12] Our current purpose is to supplement this understanding with emphasis focused on the integral equation and special function aspects. The earlier work along these lines[1,2] was primarily devoted to a study of even dimension. There were two main reasons for this. First, nearly all aplications thus far have been associated with $n=2$. Second, the inversion formula for even dimension is somewhat more challenging, involving a Hilbert transform which does not appear in the odd n case. By making use of some recent work by Durand, Fishbane, and Simmons[13] it was possible to do the Hilbert transform for the even case and obtain a rather convenient inversion formula for all even dimensions.[2]

In Sec. 3 we shall show that the corresponding result for odd n differs only by an overall multiplicative factor of -1. And in Sec. 4, we simply write down a unification formula. In retrospect, this unification appears reasonable and with hindsight it is possible to say, with some enthusiasm, that the formula (17) which holds for both even and odd n is just what one would expect since there ought not to be such fundamental differences between even and odd n as n gets large. In this regard we observe that it is especially important to have an expression like (17) when extending the theory to very large n.

In Sec. 5 we take a close look at the special case $n=3$, which leads to a Legendre transform pair. In Sec. 6 the $n=4$ special case is discussed. In many respects this is similar to the $n=2$ case[1] with Tchebycheff polynomials of the first kind replaced by Tchebycheff polynomials of the second kind. Finally, in Sec. 7 we generalize results which appear in the special cases and find a rather interesting definite integral formula which appears to have been overlooked in the major tabulations.

2. THE RADON TRANSFORM

Let $F(x)=F(x_1,\ldots,x_n)$ be a function of n real variables and let x be a vector in R^n. The Radon transform of F is given by[7]

$$f(\xi,p)=R\{F\}=\int F(x)\delta(p-\xi\cdot x)\,dx, \tag{1}$$

where ξ is a unit vector; p is a real number, $\xi\cdot x=\xi_1x_1+\cdots+\xi_nx_n$, $dx=dx_1\cdots dx_n$, δ is the Dirac δ function, and the integral is over the entire space. For our purposes here it is assumed that F is a rapidly decreasing C^∞ function.[14]

Suppose $F(x)$ can be decomposed in the form

$$F(x)=G_l(r)S_{lm}(\hat{x}), \tag{2}$$

where $\hat{x}=x/|x|$, $r=|x|$, and $S_{lm}(\hat{x})$ is a real generalized spherical harmonic. (For a full discussion of the S_{lm}, see Hochstadt.[15]) Then it follows that f also admits of a similar decomposition,

$$f(\xi,p)=g_l(p)S_{lm}(\xi), \tag{3}$$

and g_l is a rapidly decreasing C^∞ function. If g has negative argument, the defining equation is taken to be the symmetry condition

$$g_l^{(k)}(-p)=(-1)^{l+k}g_l^{(k)}(p), \tag{4}$$

where

$$g_l^{(k)}(p)=\left(\frac{d}{dp}\right)^k g_l(p). \tag{5}$$

The function G_l and g_l are related by[2] ($p\geq 0$)

$$g_l(p)=\frac{(4\pi)^\nu\Gamma(l+1)\Gamma(\nu)}{\Gamma(l+2\nu)}\int_p^\infty r^{2\nu}G_l(r)C_l^\nu\left(\frac{p}{r}\right)\times\left[1-\frac{p^2}{r^2}\right]^{\nu-1/2}dr, \tag{6}$$

[a)] This work was done with support from the U. S. Department of Energy.

[b)] National Science Foundation, Fellow, University of California, Berkeley, CA 1977–78. Permanent address: Department of Physics, University of South Florida, Tampa, Florida 33620.

 0022-2488/78/1911-2346$1.00

where $\nu=\frac{1}{2}(n-2)$ and the dimensionality n may be even or odd. The functions C_l^ν which appear in (6) are Gegenbauer polynomials of the first kind.[13]

3. THE INVERSION FORMULA FOR ODD DIMENSION

The general inversion formula may be written as an integration over a unit sphere in ξ space,[11]

$$F(x)=\int_\Omega f^*(\xi,\xi\cdot x)\,d\Omega_\xi, \tag{7}$$

where f^* is found from f by the equation

$$f^*(\xi,\xi\cdot x)=\Upsilon f(\xi,p). \tag{8}$$

For odd n the operator Υ is defined by

$$f^*(\xi,t)=\frac{(-1)^{(n-1)/2}}{2(2\pi)^{n-1}}\left(\frac{\partial}{\partial p}\right)^{n-1} f(\xi,p)\Bigg]_{p=t}. \tag{9}$$

(For even n the operator Υ involves a Hilbert transform. This case is discussed in detail in Ref. 2.)

For the decompositions (2) and (4) it is straightforward to obtain the equation

$$G_l(r)=\frac{(-1)^{(n-1)/2}M_l^\nu}{2(2\pi)^{n-1}}\int_{-1}^{+1} g_l^{(n-1)}(rt)C_l^\nu(t)(1-t^2)^{\nu-1/2}\,dt, \tag{10}$$

by use of the Hecke–Funk theorem.[15] In (10), the normalization factor M_l^ν is given by

$$M_l^\nu=\frac{(4\pi)^\nu\Gamma(l+1)\Gamma(\nu)}{\Gamma(l+2\nu)}. \tag{11}$$

By use of the symmetry properties of g_l and C_l^ν the limits of integration may be changed, $\int_{-1}^{+1}\to 2\int_0^1$, and by the change of variable $t\to t/r$ we have

$$G_l(r)=\frac{(-1)^{(n-1)/2}M_l^\nu}{(2\pi)^{n-1}r}\int_0^r g_l^{(n-1)}(t)C_l^\nu\left(\frac{t}{r}\right)\times\left(1-\frac{t^2}{r^2}\right)^{\nu-1/2}dt. \tag{12}$$

But the integral $\int_0^r$ can be modified by making use of $\int_0^r=\int_0^\infty-\int_r^\infty$. This yields

$$G_l(r)=I_l^\nu-\frac{(-1)^{(n-1)/2}M_l^\nu}{(2\pi)^{n-1}r}\int_r^\infty g_l^{(n-1)}(t)C_l^\nu\left(\frac{t}{r}\right)\left(1-\frac{t^2}{r^2}\right)^{\nu-1/2}\times dt, \tag{13}$$

where

$$I_l^\nu=\frac{(-1)^{(n-1)/2}M_l^\nu}{(2\pi)^{n-1}r}\int_0^\infty g_l^{(n-1)}(t)C_l^\nu\left(\frac{t}{r}\right)\left(1-\frac{t^2}{r^2}\right)^{\nu-1/2}dt. \tag{14}$$

In the Appendix it is shown that (14) vanishes. Thus,

$$G_l(r)=\frac{(-1)^{(n+1)/2}(-1)^{\nu-1/2}M_l^\nu}{(2\pi)^{n-1}r}\int_r^\infty g_l^{(n-1)}(t)C_l^\nu\left(\frac{t}{r}\right)\times\left(\frac{t^2}{r^2}-1\right)^{\nu-1/2}dt, \tag{15}$$

or, making use of $\nu=\frac{1}{2}(n-2)$ and (11),

$$G_l(r)=\frac{\Gamma(l+1)\Gamma(\nu)}{2\pi^{\nu+1}\Gamma(l+2\nu)r}\int_r^\infty g_l^{(n-1)}(t)C_l^\nu\left(\frac{t}{r}\right)\times\left(\frac{t^2}{r^2}-1\right)^{\nu-1/2}dt, \tag{16}$$

for odd n.

4. THE UNIFICATION

In Ref. 2 the result for even n differed from (16) by a negative sign. Hence, it is possible to write a unified formula, valid for even or odd n. In terms of $\nu=\frac{1}{2}(n-2)$,

$$G_l(r)=\frac{(-1)^{2\nu+1}\Gamma(l+1)\Gamma(\nu)}{2\pi^{\nu+1}\Gamma(l+2\nu)r}\int_r^\infty g_l^{(2\nu+1)}(t)C_l^\nu\left(\frac{t}{r}\right)\times\left[\frac{t^2}{r^2}-1\right]^{\nu-1/2}dt. \tag{17}$$

Equations (17) and (6) constitute a Gegenbauer transform pair.

Ludwig[11] also observed a fundamental connection between Radon and Gegenbauer transforms; however, in that work the inversion formula was not unified in the sense that it was necessary to treat the even and odd dimensional cases separately. Although the purpose here is to focus attention on results for the Radon transform, it may be useful to point out that the transform pair (6) and (17) is not the only possible Gegenbauer transform pair and general methods (not involving the Radon transform) exist for deriving such pairs. In particular, another pair was found by Higgins,[16] and Sneddon[17] has given a procedure utilizing the Mellin transform which yields inversion formulas for integral transform pairs of a general kind. Finally, by the very nature of the method used here to obtain the pair (6) and (17) there is a restriction on the degree l and order ν associated with C_l^ν; however, the Gegenbauer functions can be defined for general degree and order,[13] and fractional differentiation and integration[18,19] can be used to extend the transform to more general values of the indices, $l>0$ and Re $\nu>-\frac{1}{2}$.

5. SPECIAL CASE $n=3$

If $n=3$ (or $\nu=\frac{1}{2}$) the Gegenbauer transform pair becomes the Legendre transform pair. (Here x is a real variable and not a vector.)

$$g_l(s)=2\pi\int_s^\infty xG_l(x)P_l\left(\frac{s}{x}\right)dx, \tag{18}$$

$$G_l(x)=\frac{1}{2\pi x}\int_x^\infty g_l''(t)P_l\left(\frac{t}{x}\right)dt. \tag{19}$$

Note that if $2\pi xG_l(x)$ is replaced by $G(x)$ and if $g_l(s)$ is replaced by $g(s)$ we obtain the pair

$$g(s)=\int_s^\infty G(x)P_l\left(\frac{s}{x}\right)\,dx, \tag{20}$$

$$G(x)=\int_x^\infty g''(t)P_l\left(\frac{t}{x}\right)\,dt. \tag{21}$$

It is of interest to try a direct verification that (21) does indeed satisfy (20). When (21) is substituted into (20) we obtain

$$\int_s^\infty dx\,P_l\left(\frac{s}{x}\right)\int_x^\infty dt\,g''(t)P_l\left(\frac{t}{x}\right), \tag{22}$$

and by changing the order of integration (22) becomes $(0<s<t)$

$$\int_s^\infty dt\,g''(t)\int_s^t dx\,P_l\left(\frac{s}{x}\right)P_l\left(\frac{t}{x}\right). \tag{23}$$

We designate the integral on the right by $K_l^\nu(t,s)$ with $\nu=\frac{1}{2}$,

$$K_l^{1/2}(t,s)=\int_s^t P_l\left(\frac{s}{x}\right)P_l\left(\frac{t}{x}\right)\,dx=t-s. \tag{24}$$

After substituting (24) into (23) and doing one integration by parts we have

$$-\int_s^\infty g'(t)\,dt=g(s) \tag{25}$$

as required.

6. SPECIAL CASE $n=4$

In $n=4$, or equivalently $\nu=1$, the Gegenbauer transform pair becomes the Tchebycheff pair

$$g_l(s)=\frac{4\pi}{l+1}\int_s^\infty x^2G_l(x)U_l\left(\frac{s}{x}\right)\left(1-\frac{s^2}{x^2}\right)^{1/2}dx, \tag{26}$$

$$G_l(x)=\frac{-1}{2\pi^2(l+1)x}\int_x^\infty g_l'''(t)U_l\left(\frac{t}{x}\right)\left(\frac{t^2}{x^2}-1\right)^{1/2}dt. \tag{27}$$

If, as in the previous section, we attempt a direct verification that (27) satisfies (26) we find, after changing the order of integration, that we must evaluate the integral

$$K_l^1(t,s)=\int_s^t xU_l\left(\frac{s}{x}\right)U_l\left(\frac{t}{x}\right)\left(1-\frac{s^2}{x^2}\right)^{1/2}\times\left(\frac{t^2}{x^2}-1\right)^{1/2}dx. \tag{28}$$

The result is

$$K_l^1(t,s)=\frac{\pi}{4}(l+1)^2(t-s)^2. \tag{29}$$

After doing two integrations by parts the verification is immediate.

7. AN INTEGRAL FORMULA

If a direct verification that (17) satisfies (6) is attempted, as was done with the $n=3$ and $n=4$ special cases, we are led to the formula $(0<s\leq t)$

$$K_l^\nu(t,x)=\int_s^t x^{2\nu-1}C_l^\nu\left(\frac{s}{x}\right)C_l^\nu\left(\frac{t}{x}\right)\left[1-\frac{s^2}{x^2}\right]^{\nu-1/2}\times\left[\frac{t^2}{x^2}-1\right]^{\nu-1/2}dx$$

$$=\frac{\pi}{2^{2\nu-1}}\left[\frac{\Gamma(l+2\nu)}{\Gamma(l+1)\Gamma(\nu)}\right]^2\frac{(t-s)^{2\nu}}{\Gamma(2\nu+1)}. \tag{30}$$

A search through some of the more extensive sources for integral formulas[20,21] seems to indicate that this result has been overlooked in the tabulations.

Formulas (24), (29), and (30) follow from the requirement that K_l^ν yield the kernel for the Weyl fractional integral[22] which converts $g_l^{(2\nu+1)}$ to g_l. Once the general formula (30) is found by this "working backward" procedure of insisting that (17) satisfy (6), it is possible to check various special cases and directly verify the formula by double mathematical induction.[23]

ACKNOWLEDGMENT

It is a pleasure to acknowledge discussions with Professor T. F. Budinger and G. T. Gullberg along with helpful comments by the referee.

APPENDIX

We wish to show that the integral I_l^ν defined in (14) vanishes. Consider the integral

$$J=\int_0^\infty g_l^{(n-1)}(t)C_l^\nu\left(\frac{t}{r}\right)\left(1-\frac{t^2}{r^2}\right)^{\nu-1/2}dt. \tag{A1}$$

Now, for odd n $(3,5,7,\cdots)$, $C_l^\nu(x)(1-x^2)^{\nu-1/2}$ is a polynomial of degree $l+n-3$ in x. Let us designate this polynomial by $Q_{l+n-3}(x)$; then (A1) becomes

$$J=\int_0^\infty g_l^{(n-1)}(t)Q_{l+n-3}\left(\frac{t}{r}\right)\,dt. \tag{A2}$$

After doing $n-1$ integrations by parts this becomes

$$J=r^{1-n}\int_0^\infty g_l(t)Q_{l+n-3}^{(n-1)}\left(\frac{t}{r}\right)\,dt. \tag{A3}$$

Observe that the integrated part always vanishes since $g_l^{(k)}$ vanishes at ∞, $Q_{l+n-3}(-x)=(-1)^l Q_{l+n-3}(x)$, and the symmetry condition (4) holds.

Since the polynomial $Q_{l+n-3}^{(n-1)}$ is of degree $l-2$ we redesignate it by Q_{l-2} and observe that $Q_{l-2}(-x)=(-1)^lQ_{l-2}(x)$. If g_l in (A3) is replaced by (6), the integral J is proportional to

$$\int_0^\infty dt Q_{l-2}\left(\frac{t}{r}\right)\int_t^\infty x^{2\nu}G_l(x)C_l^\nu\left(\frac{t}{x}\right)\left(1-\frac{t^2}{x^2}\right)^{\nu-1/2}dx. \tag{A4}$$

If the order of integration is reversed, then

$$J \propto \int_0^\infty dx\, x^{2\nu} G_l(x) \int_0^x dt\, Q_{l-2}\left(\frac{t}{r}\right) C_l^\nu\left(\frac{t}{x}\right) \times \left[1 - \frac{t^2}{x^2}\right]^{\nu-1/2}. \tag{A5}$$

By the symmetry of the integrand the $\int_0^x$ integral can be written as $\frac{1}{2}\int_{-x}^{x}$ and by a change of variable $t \to tx$,

$$J \propto \int_0^\infty dx\, x^{2\nu+1} G_l(x) \int_{-1}^{+1} dt\, Q_{l-2}\left(\frac{xt}{r}\right) C_l^\nu(t)(1-t^2)^{\nu-1/2}. \tag{A6}$$

Since Q_{l-2} is a polynomial of degree $l-2$ in t it follows by the orthogonality property of the Gegenbauer polynomials that $J = 0$. Hence, I_l^ν defined in (14) vanishes.

[1]S.R. Deans, Lawrence Berkeley Laboratory Report LBL-5691, November 1977.
[2]S.R. Deans, "Gegenbauer Transforms via the Radon Transform," SIAM J. Math. Anal. (to be published).
[3]T.F. Budinger and G.T. Gullberg, IEEE Trans. Nucl. Sci. **NS-21**, 2 (1974).
[4]R. Gordon and G.T. Herman, Int. Rev. Cytol. **38**, 111 (1974).
[5]R.A. Brooks and G. DiChiro, Phys. Med. Biol. **21**, 689 (1976).
[6]J. Radon, Ber. Verh. Sächs. Akad. **69**, 262 (1917).
[7]I.M. Gel'fand, M.I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.
[8]F. John, Math. Ann. **109**, 488 (1934).
[9]F. John, *Plane Waves and Spherical Means Applied to Differential Equations* (Interscience, New York, 1955).
[10]S. Helgason, Acta Math. **113**, 153 (1965).
[11]D. Ludwig, Commun. Pure Appl. Math. **19**, 49 (1966).
[12]K.T. Smith, D.C. Solomon, and S.L. Wagner, Bull. Amer. Math. Soc. **83**, 1227 (1977).
[13]L. Durand, P.M. Fishband, and L.M. Simmons, Jr., J. Math. Phys. **17**, 1933 (1976).
[14]L. Schwartz, *Mathematics for the Physical Sciences* (Addison-Wesley, Reading, Mass., 1966).
[15]H. Hochstadt, *The Functions of Mathematical Physics* (Wiley, New York, 1971).
[16]T.P. Higgins, J. Soc. Indust. Appl. Math. **11**, 886 (1963).
[17]I.N. Sneddon, Glasgow Math. J. **9**, 67 (1968).
[18]A. Erdélyi, J. Soc. Indust. Appl. Math. **12**, 15 (1964).
[19]T.P. Higgins, J. Soc. Indust. Appl. Math. **12**, 601 (1964).
[20]I.S. Gradshteyn and I.W. Ryzhik, *Tables of Integrals Series and Products* (Academic, New York, 1965).
[21]D. Bierens de Haan, *Nouvelles Tables D'Intégrales Définies* (Hafner, New York, 1957).
[22]A. Erdélyi, W. Magnus, F. Oberhettinger, and F.G. Tricomi, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. 2, p. 181.
[23]S. Iyanaga and Y. Kawada, Eds., *Encyclopedic Dictionary of Mathematics* (M.I.T., Cambridge, Mass., 1977), Vol. 2, p. 918.

Electromagnetic and gravitational Hertz potentials

Jerzy F. Plebański[a)]

Centro de Investigación y de Estudios Avanzados del I.P.N., México 14, D.F., 14-740

Ivor Robinson

The University of Texas at Dallas, Richardson, Texas 75080
(Received 29 August 1977)

With generality of complex relativity, the classical theory of the electromagnetic Hertz potentials is outlined in terms of spinors and forms. Particularly interesting are $D(0,1)$ and $D(1,0)$ null Hertz potentials. Then, a new spinorial approach to heavens ($\mathcal{H}$ spaces) is proposed, which reveals in their structure the presence of the left *null* gravitational Hertz potential [of the type $D(0,2)$]. The relevant hints which follow from our results and concern the structure of the most general solutions of the Einstein vacuum equations (type $G \otimes G$), are discussed, in particular on the level of the linearized theory.

1. PRELIMINARIES AND THE FORMALISM USED

A fairly complete review of various approaches to electromagnetic Hertz potentials (which also includes the treatment employing differentials forms) can be found in the excellent article by Cohen and Kegels[1]; the older pertinent references can be also localized there.

Our aim is to propose a new approach to the dynamical equations of nonlinear general relativity, founded on an appropriately generalized notion of Hertz potentials. For this purpose, we will first outline in this section the spinorial description of the Riemannian geometry of a complex space—time. Then, the classical theory of Hertz potentials and the basic results of the theory of heavens established in Ref. 2 will be examined in the light of the spinorial formalism. This will lead in the subsequent sections to some general ideas about the gravitational Hertz potentials.

Thus, we will work in a complex (Riemannian) space—time which is a pair: a (complex) analytic differential manifold M_4 and the metric

$$g = -\tfrac{1}{2} g_{A\dot{B}} \overset{\otimes}{s} g^{A\dot{B}} \tag{1.1}$$

where, labelled by the spinorial indices ($A = 1, 2$; $\dot{B} = \dot{1}, \dot{2}$), $g_{A\dot{B}} \in \Lambda^1$ form the base of the cotangent space. The spinorial indices are manipulated by Levi-Civita's symbols according to the usual conventions, e.g., $\psi_A = \epsilon_{AB}\psi^B \longleftrightarrow \psi^A = \psi_B \epsilon^{BA}$. The gauge group of the theory $\mathcal{G} = \mathrm{SL}(2,\mathbb{C}) \times \mathrm{SL}(2,\mathbb{C})$ in an obvious symbolism. In real relativity, $\dot{\mathrm{SL}}$ transformations are complex conjugates of SL, $\dot{\mathrm{SL}} = (\mathrm{SL})^*$ and $\mathcal{G}$ maintains the hermicity of $g_{A\dot{B}}$—and thus the signature $(+++-)$ of the real metric over the real manifold M_4. In complex relativity, the two copies of the $\mathrm{SL}(2,\mathbb{C})$ group, SL and $\dot{\mathrm{SL}}$, remain independent.

In the space of the multiforms, $\Lambda = \oplus_{p=0}^{4} \Lambda^p$, we have two basic mappings: the external differential and the Hodge star

$$d: \Lambda^p \to \Lambda^{p+1}, \quad d^2 = 0, \tag{1.2a}$$

$$*: \Lambda^p \to \Lambda^{4-p}, \quad ** = id. \tag{1.2b}$$

[a)]On leave of absence from the University of Warsaw, Warsaw, Poland.

We will also employ the concept of the codifferential defined by

$$\delta \equiv -i{*}d{*}: \Lambda^p \to \Lambda^{p-1}, \quad \delta^2 = 0. \tag{1.3}$$

For our present purposes, it is also convenient to work with the concept of the Laplace—Beltrami operator

$$\Delta \equiv d\delta + \delta d: \Lambda^p \to \Lambda^p, \tag{1.4}$$

called subsequently the "Laplacian," and the associated operator

$$\blacktriangle \equiv -d\delta + \delta d: \Lambda^p \to \Lambda^p, \tag{1.5}$$

called subsequently the "anti-Laplacian." We have, of course,

$$d\delta = \tfrac{1}{2}(\Delta - \blacktriangle), \quad \delta d = \tfrac{1}{2}(\Delta + \blacktriangle) \tag{1.6}$$

and then

$$\Delta\blacktriangle - \blacktriangle\Delta = 0, \tag{1.7a}$$

$$\Delta^2 = (d\delta)^2 + (\delta d)^2 = \blacktriangle^2. \tag{1.7b}$$

The name "anti-Laplacian" is perhaps justified by the fact that:

$$\Delta d - d\Delta = 0, \quad \Delta\delta - \delta\Delta = 0, \quad \Delta * - *\Delta = 0, \tag{1.8a}$$

$$\blacktriangle d + d\blacktriangle = 0, \quad \blacktriangle\delta + \delta\blacktriangle = 0, \quad \blacktriangle * + *\blacktriangle = 0. \tag{1.8b}$$

The formalism which we use employs, with respect to Λ^p-valued spinors, the covariant differential D: $\Lambda^p \to \Lambda^{p+1}$ defined according to

$$DT^{A\dot{B}\cdots}_{C\dot{D}\cdots} \equiv dT^{A\dot{B}\cdots}_{C\dot{D}\cdots} + \Gamma^{A}{}_{S} \wedge T^{S\dot{B}\cdots}_{C\dot{D}\cdots} + \Gamma^{\dot{B}}{}_{\dot{S}} \wedge T^{A\dot{S}\cdots}_{C\dot{D}\cdots} - \Gamma^{S}{}_{C} \wedge T^{A\dot{B}\cdots}_{S\dot{D}\cdots} - \Gamma^{\dot{S}}{}_{\dot{D}} \wedge T^{A\dot{B}\cdots}_{C\dot{S}\cdots} + \cdots, \tag{1.9}$$

where $\Gamma_{AB} = \Gamma_{(AB)}$ and $\Gamma_{\dot{A}\dot{B}} = \Gamma_{(\dot{A}\dot{B})}$ are respectively the left and right connection 1-forms. The connection forms we understand as determined by the $g_{A\dot{B}}$ via the first structure equations:

$$Dg^{A\dot{B}} \equiv dg^{A\dot{B}} + \Gamma^{A}{}_{S} \wedge g^{S\dot{B}} + \Gamma^{\dot{B}}{}_{\dot{S}} \wedge g^{A\dot{S}} = 0. \tag{1.10}$$

Now the covariant derivatives of the connections, which are $\mathcal{G}$-tensors, determine at the same time the left and right curvature forms:

$$D\Gamma^{A}{}_{B} \equiv d\Gamma^{A}{}_{B} + \Gamma^{A}{}_{S} \wedge \Gamma^{S}{}_{B} = : R^{A}{}_{B}, \tag{1.11a}$$

$$D\Gamma^{\dot{A}}{}_{\dot{B}} \equiv d\Gamma^{\dot{A}}{}_{\dot{B}} + \Gamma^{\dot{A}}{}_{\dot{S}} \wedge \Gamma^{\dot{S}}{}_{\dot{B}} = : R^{\dot{A}}{}_{\dot{B}} \tag{1.11b}$$

The Bianchi identities, which are integrability conditions for (1.11), are then

$$DR^{A}{}_{B} = 0 = DR^{\dot{A}}{}_{\dot{B}}, \tag{1.12}$$

Observe that the general formulas

$$DDT^{A\dot{B}\cdots}_{C\dot{D}\cdots} = R^{A}{}_{S} \wedge T^{S\dot{B}\cdots}_{C\dot{D}\cdots} + R^{\dot{B}}{}_{\dot{S}} \wedge T^{A\dot{S}\cdots}_{C\dot{D}\cdots} - R^{S}{}_{C} \wedge T^{A\dot{B}\cdots}_{S\dot{D}\cdots} - R^{\dot{S}}{}_{\dot{D}} \wedge T^{A\dot{B}\cdots}_{C\dot{S}\cdots} + \cdots, \tag{1.13}$$

take in the present formalism the role of the Ricci formulas, and permit us to investigate conveniently the integrability conditions of any equations formulated by the use of the D operation.

The relation:

$$g^{A\dot{B}} \wedge g^{C\dot{D}} = \epsilon^{AC} S^{\dot{B}\dot{D}} + S^{AC} \epsilon^{\dot{B}\dot{D}} \tag{1.14}$$

defines us the objects,

$$S_{AB} = {*}\, S_{AB} = S_{(AB)} \in \Lambda^2, \tag{1.15a}$$

$$S_{\dot{A}\dot{B}} = -{*}\, S_{\dot{A}\dot{B}} = S_{(\dot{A}\dot{B})} \in \Lambda^2, \tag{1.15b}$$

which form a complete base of Λ^2. Knowing this, and using as a consequence of (1.10) $DD\, g_{A\dot{B}} = 0$ (i.e., $R^{A}{}_{S} \wedge g^{S\dot{B}} + R^{\dot{B}}{}_{\dot{S}} \wedge g^{A\dot{S}} = 0$), one easily shows that the curvature forms can be always represented as:

$$R_{AB} = -\tfrac{1}{2} C_{ABCD} S^{CD} + (R/24) S_{AB} + \tfrac{1}{2} C_{AB\dot{C}\dot{D}} S^{\dot{C}\dot{D}}, \tag{1.16a}$$

$$R_{\dot{A}\dot{B}} = -\tfrac{1}{2} C_{\dot{A}\dot{B}\dot{C}\dot{D}} S^{\dot{C}\dot{D}} + (R/24) S_{\dot{A}\dot{B}} + \tfrac{1}{2} C_{CD\dot{A}\dot{B}} S^{CD}, \tag{1.16b}$$

where the D(2,0) and D(0,2) objects, $C_{ABCD} = C_{(ABCD)}$ and $C_{\dot{A}\dot{B}\dot{C}\dot{D}} = C_{(\dot{A}\dot{B}\dot{C}\dot{D})}$, are the spinorial images of the self-dual and anti-self-dual parts of the conformal curvature tensor $(C_{\alpha\beta\gamma\delta} \pm {*}\, C_{\alpha\beta\gamma\delta})$; R is the scalar curvature; and $C_{AB\dot{C}\dot{D}} = C_{(AB)(\dot{C}\dot{D})}$ is a D(1,1) object which corresponds to the traceless part of the Ricci tensor.

Our formalism employs the concept of the spinorial gradient, $\partial_{A\dot{B}}$, and of the covariant spinorial gradient $\nabla_{A\dot{B}}$. Of course, $\partial_{A\dot{B}}$ can be thought of as the base of the tangent space, and for every $T_{K\dot{L}\cdots} \in \Lambda^0$ we have

$$dT_{K\dot{L}\cdots} = -\tfrac{1}{2} g^{A\dot{B}} \partial_{A\dot{B}} T_{K\dot{L}\cdots}. \tag{1.17}$$

The covariant gradient is then defined by a parallel formula:

$$DT_{K\dot{L}\cdots} = -\tfrac{1}{2} g^{A\dot{B}} \nabla_{A\dot{B}} T_{K\dot{L}\cdots}. \tag{1.18}$$

Using the operator $\nabla_{A\dot{B}}$, one easily shows that the Bianchi identities amount to

$$\nabla^{S}{}_{\dot{D}} C_{SABC} + \nabla_{(A}{}^{\dot{S}} C_{BC)\dot{D}\dot{S}} = 0, \tag{1.19a}$$

$$\nabla_{D}{}^{\dot{S}} C_{\dot{S}\dot{A}\dot{B}\dot{C}} + \nabla^{S}{}_{(\dot{A}} C_{|DS|\dot{B}\dot{C})} = 0, \tag{1.19b}$$

$$\nabla^{R\dot{S}} C_{AR\dot{B}\dot{S}} + \tfrac{1}{8} \nabla_{A\dot{B}} R = 0, \tag{1.19c}$$

The formalism succinctly outlined here, for the sake of completeness, is described more fully in Ref. 3. Further developments and applications of the formalism are discussed in the first section of Ref. 2 and the subsequent papers about heavens, Refs. 4, 5 and particulary 6.

We close this section by establishing our convection for the inner product of forms:

$$\alpha \,\lrcorner\, \beta : = {*}\,(\alpha \wedge {*}\,\beta) \tag{1.20}$$

which holds with the star so normalized that $** =$ identity.

2. THE MAXWELL EQUATIONS AND THE ELECTROMAGNETIC HERTZ POTENTIALS

In a real V_4, the 2-form of the electromagnetic field $f = \frac{1}{2} f_{\mu\nu} dX^\mu \wedge dX^\nu$ (in terms of the local components) has to fulfill the Maxwell (vacuum) equations

$$df = 0 = \delta f \tag{2.1}$$

and is supposed to be real. It can be always decomposed into the pure self-dual and anti-self-dual parts:

$$\omega \equiv f + {*}f = :\, 2 f_{AB} S^{AB}, \qquad \bar{\omega} \equiv f - {*}f = :\, 2 f_{\dot{A}\dot{B}} S^{\dot{A}\dot{B}}, \tag{2.2}$$

where f_{AB} and $f_{\dot{A}\dot{B}}$, the spinorial images of the electromagnetic field, can be thought of as objects of helicity $+\hbar$ and $-\hbar$ respectively. The objects ω and $\bar{\omega}$ then satisfy

$$*\,\omega = \omega, \quad d\omega = 0, \quad \delta\omega = 0, \tag{2.3a}$$

$$*\,\bar{\omega} = -\bar{\omega}, \quad d\bar{\omega} = 0, \quad \delta\bar{\omega} = 0, \tag{2.3b}$$

In a real V_4, $\bar{\omega}$ is complex conjugate of ω. In a complex V_4, ω and $\bar{\omega}$ became independent objects and (2.3a) and (2.3b) are respectively the left ("heavenly") and the right ("hellish") Maxwell equations. Notice that with ω and ϖ being of definite helicity, it is enough to assume that either the differential or codifferential of ω and ϖ vanish; the others then vanish automatically. Notice also that, as a consequence of (2.3), we have

$$\Delta\omega = 0, \quad \blacktriangle\omega = 0, \tag{2.4a}$$

$$\Delta\bar{\omega} = 0, \quad \blacktriangle\bar{\omega} = 0. \tag{2.4b}$$

We can now state the basic idea of the Hertz potentials as follows: Let Π, $\bar{\Pi} \in \Lambda^2$ be forms which have a definite helicity and are solutions of the harmonic equation:

$$*\Pi = \Pi, \quad \Delta\Pi = 0, \tag{2.5a}$$

$$*\bar{\Pi} = -\bar{\Pi}, \quad \Delta\bar{\Pi} = 0. \tag{2.5b}$$

Then, having such forms, we can construct solutions of the left and right Maxwell equations by writing

$$\omega : = +\,\delta d\bar{\Pi} = -\,d\delta\bar{\Pi} = \tfrac{1}{2}\blacktriangle\bar{\Pi}, \tag{2.6a}$$

$$\bar{\omega} : = +\,\delta d\Pi \equiv -\,d\delta\Pi = \tfrac{1}{2}\blacktriangle\Pi. \tag{2.6b}$$

Indeed, the differentials and codifferentials of both ω and ϖ then vanish as the consequence of $d^2 = 0 = \delta^2$ and we have $*\omega = \omega$ and $*\bar{\omega} = -\bar{\omega}$, as the consequence of $*\blacktriangle + \blacktriangle * = 0$ and the assumed helicities for $\bar{\Pi}$ and Π.

One can easily show that every left and right electromagnetic field can be *always* represented through the corresponding Hertz potentials and, moreover, that there remains a great deal of ambiguity with which these potentials are defined when ω and $\overline{\omega}$ are given. One easily shows that the formulas (2.5) and (2.6) stay unchanged with respect to such a gauge of Π's that

$$\Pi \to \Pi + \nu, \quad \Delta\nu = 0 = \blacktriangle\nu, \tag{2.7a}$$

$$\overline{\Pi} \to \overline{\Pi} + \overline{\nu}, \quad \Delta\overline{\nu} = 0 = \blacktriangle\overline{\nu}, \tag{2.7b}$$

i.e., the gauge forms ν, $\overline{\nu} \in \Lambda^2$ being the solutions of the inhomogenous left and right Maxwell equations,

$$*\nu = \nu, \quad d\nu = i*\delta\mu, \quad \delta\nu = d\mu, \tag{2.8a}$$

$$*\overline{\nu} = -\overline{\nu}, \quad d\overline{\nu} = -i*\delta\overline{\mu}, \quad \delta\overline{\nu} = d\overline{\mu}, \tag{2.8b}$$

where μ, $\overline{\mu} \in \Lambda^\circ$ are arbitrary harmonic functions:

$$\Delta\mu = 0 = \Delta\overline{\mu}. \tag{2.9}$$

It is of interest to describe all this in terms of Λ°-valued spinorial images and the spinorial covariant gradient. The left and right Maxwell equations, (2.3) can be seen to be equivalent to

$$\nabla^{S\dot{B}} f_{SA} = 0, \tag{2.10a}$$

$$\nabla^{A\dot{S}} f_{\dot{S}\dot{B}} = 0. \tag{2.10b}$$

Now, with the Π's of pure parities represented according to

$$\Pi = 2\Pi_{AB} S^{AB} \tag{2.11a}$$

$$\overline{\Pi} = 2\Pi_{\dot{A}\dot{B}} S^{\dot{A}\dot{B}} \tag{2.11b}$$

we can compute—by applying the spinorial Ricci formulas given in Ref. 6 as (2.7a)–(2.7b)—the Laplacians and anti-Laplacians of Π's. The result is

$$\Delta\Pi = -2S^{AB}\{(\Box + R/3)\Pi_{AB} + C_{ABCD}\Pi^{CD}\}, \tag{2.12a}$$

$$\Delta\overline{\Pi} = -2S^{\dot{A}\dot{B}}\{(\Box + R/3)\Pi_{\dot{A}\dot{B}} + C_{\dot{A}\dot{B}\dot{C}\dot{D}}\Pi^{\dot{C}\dot{D}}\}, \tag{2.12b}$$

and

$$\blacktriangle\Pi = 2\nabla^{R}{}_{\dot{A}}\nabla^{S}{}_{\dot{B}}\Pi_{RS}\cdot S^{\dot{A}\dot{B}}, \tag{2.13a}$$

$$\blacktriangle\overline{\Pi} = 2\nabla_{A}{}^{\dot{R}}\nabla_{B}{}^{\dot{S}}\Pi_{\dot{R}\dot{S}}\cdot S^{AB}, \tag{2.13b}$$

where we introduced the abbreviation

$$\Box \equiv -\tfrac{1}{2}\nabla_{N\dot{M}}\nabla^{N\dot{M}}: \Lambda^\circ \to \Lambda^\circ. \tag{2.14}$$

Consequently, Eqs. (2.10) are satisfied by

$$f_{AB} = \tfrac{1}{2}\nabla_{(A}{}^{\dot{R}}\nabla_{B)}{}^{\dot{S}}\Pi_{\dot{R}\dot{S}}, \tag{2.15a}$$

$$f_{\dot{A}\dot{B}} = \tfrac{1}{2}\nabla^{R}{}_{(\dot{A}}\nabla^{S}{}_{\dot{B})}\Pi_{RS} \tag{2.15b}$$

provided that the $D(0,1)$ and $D(1,0)$ Hertz potentials fulfill correspondingly

$$(\Box + R/3)\Pi_{AB} + C_{ABCD}\Pi^{CD} = 0, \tag{2.16a}$$

$$(\Box + R/3)\Pi_{\dot{A}\dot{B}} + C_{\dot{A}\dot{B}\dot{C}\dot{D}}\Pi^{\dot{C}\dot{D}} = 0. \tag{2.16b}$$

Now, if the gauge forms ν, $\overline{\nu}$ from (2.7) are represented according to

$$\nu = 2\nu_{AB} S^{AB}, \tag{2.17a}$$

$$\overline{\nu} = 2\nu_{\dot{A}\dot{B}} S^{\dot{A}\dot{B}}, \tag{2.17b}$$

then Eqs. (2.8) and (2.9) can be stated in the equivalent scalar form as

$$\nabla^{S}{}_{\dot{B}}\nu_{AS} = -\tfrac{1}{4}\nabla_{A\dot{B}}\mu, \quad \Box\mu = 0, \tag{2.18a}$$

$$\nabla_{A}{}^{\dot{S}}\nu_{\dot{B}\dot{S}} = -\tfrac{1}{4}\nabla_{A\dot{B}}\overline{\mu}, \quad \Box\overline{\mu} = 0. \tag{2.18b}$$

By using the freedom of ν-gauges, one can bring the Hertz potentials to various plausible or useful forms. For the reduction of the last to the Debye potentials, see Ref. 1; for some applications of the Debye potentials in the theory of Einstein-Maxwell equations see Ref. 7.

In heavens (in particular, in flat space—time) there exists a "homogeneous spinor" $K_{\dot{A}} \neq 0$, such that

$$DK_{\dot{A}} = 0. \tag{2.19}$$

[Indeed, according to (1.13), the integrability condition of (2.19) is $0 = -K_{\dot{B}} R^{\dot{B}}{}_{\dot{A}}$; but precisely in (strong) heavens $C_{\dot{A}\dot{B}\dot{C}\dot{D}} = C_{ABCD} = R = 0 \to R_{\dot{A}\dot{B}} = 0$]. Using then the freedom of $\overline{\nu}$ gauge one can bring the left [i.e., $D(0,1)$ Hertz potential] to the particularly simple and plausible form of

$$\Pi_{\dot{A}\dot{B}} = HK_{\dot{A}}K_{\dot{B}} \tag{2.20}$$

characterized by the algebraic condition

$$\overline{\Pi} \lrcorner \overline{\Pi} = 0. \tag{2.21}$$

Equation (2.16b) reduces then to the simple scalar equation:

$$\Box H = 0. \tag{2.22}$$

For flat space—time, this specialization of the Hertz potential was obtained many years ago by Penrose.[8] For the sake of completeness, we shall derive it again in the present notation. In flat space—time, there exists a frame such that

$$g^{A\dot{B}} = {}^\circ g^{A\dot{B}} \underset{\text{s.f.}}{=} dX^{A\dot{B}} \tag{2.23}$$

with $X^{A\dot{B}} \in \Lambda^\circ$ being (Cartesian) coordinates. We describe this as a "special frame" of "s.f." The flat tetrad ${}^\circ g_{A\dot{B}}$ induces, of course connections ${}^\circ\Gamma_{\dot{A}\dot{B}}$ and ${}^\circ\Gamma_{AB}$ such that

$${}^\circ D\,{}^\circ\Gamma_{AB} = 0 = {}^\circ D\,{}^\circ\Gamma_{\dot{A}\dot{B}} \tag{2.24}$$

and in our special frame ${}^\circ\Gamma_{AB} \underset{\text{s.f.}}{=} 0 \underset{\text{s.f.}}{=} {}^\circ\Gamma_{\dot{A}\dot{B}}$. Consequently, we have

$$\nabla_{A\dot{B}} = {}^\circ\nabla_{A\dot{B}} \underset{\text{s.f.}}{=} {}^\circ\partial_{A\dot{B}} \underset{\text{s.f.}}{=} -2\partial/\partial X^{A\dot{B}} \tag{2.25}$$

in the obvious notation, so that the left Maxwell equations amount to the simple

$$(\partial/\partial X^{A\dot{B}}) f^{A}{}_{C} = 0 \tag{2.26}$$

in our s.f. Let now Z_A be any two (out of the four) independent variables. We have then a simple lemma: The condition $\partial\psi^A/\partial Z^A = 0$ implies (and is implied by) the existence of such a ψ that $\psi_A = \partial\psi/\partial Z^A$. Therefore, equations $\partial f^{A}{}_{C}/\partial X^{A\dot{2}} = 0$ imply $f_{AB} = \partial\psi_B/\partial X^{A\dot{2}}$, but because f_{AB} is symmetric, $\partial\psi^A/\partial X^{A\dot{2}} = 0$, and by again applying the lemma we infer the existence of such a scalar H that

$$f_{AB} \underset{\text{s.f.}}{=} 2(\partial/\partial X^{A\dot{2}})(\partial/\partial X^{B\dot{2}})H. \tag{2.27}$$

Substituting in the remaining equations $(\partial/\partial X^{A\dot{1}}) f^{A}{}_{C} = 0$, we obtain

$$(\partial/\partial X^{A\dot{2}})({}^\circ\Box H) = 0, \tag{2.28}$$

where ${}^\circ\Box := -\tfrac{1}{2}{}^\circ\nabla_{A\dot{B}}{}^\circ\nabla^{A\dot{B}} \underset{\text{s.f.}}{=} -2(\partial/\partial X_{A\dot{B}})(\partial/\partial X^{A\dot{B}})$. Consequently, ${}^\circ\Box H = \alpha(X^{A\dot{1}})$, where the function of the two variables α is arbitrary. However, regauging H according to $H \to H + \beta_A(X^{B\dot{1}})X^{A\dot{2}}$ we leave (2.27) unchanged, while the arbitrary $\beta_A(X^{B\dot{1}})$ can always be so choosen that the regauged H fulfills already the homogenous equation:

$$^\circ\Box H = 0. \tag{2.29}$$

It is now clear that by introducing a constant spinor

$$K_{\dot{A}} \underset{\text{s.f.}}{=} (1,0) \to K^{\dot{A}} = (0,1), \quad {}^\circ DK_{\dot{A}} \underset{\text{s.f.}}{=} 0 \tag{2.30}$$

we can rewrite (2.27) in an *arbitrary* $\mathcal{G}$-frame precisely in the form (2.20).

Therefore, extending the argument given above on the right side and summing up, we can state the following: In flat (possibly complex) space–time, the Hertz potentials for the *most general* (vacuum) electromagnetic field can be always so gauged that

$$\Pi_{\dot{A}\dot{B}} = HK_{\dot{A}}K_{\dot{B}}, \quad {}^\circ DK_{\dot{A}} = 0, \quad {}^\circ\Box H = 0, \tag{2.31a}$$

$$\Pi_{AB} = \bar{H}K_AK_B, \quad {}^\circ DK_A = 0, \quad {}^\circ\Box\bar{H} = 0, \tag{2.31b}$$

with the homogeneous spinors $K_{\dot{A}} \neq 0 \neq K_A$ being otherwise arbitrary. The spinorial images of the electromagnetic field are then given by (2.15) [with $\nabla_{A\dot{B}}$ replaced by commuting ${}^\circ\nabla_{A\dot{B}}$], and, the final general solution of the Maxwell equations amounts to

$$\omega = -d\aleph, \quad \aleph := {}^\circ g^{A\dot{B}}{}^\circ\nabla_{A\dot{C}}(\Pi_{\dot{B}}{}^{\dot{C}}), \tag{2.32a}$$

$$\bar{\omega} = -d\bar{\aleph}, \quad \bar{\aleph} := {}^\circ g^{A\dot{B}}{}^\circ\nabla_{C\dot{B}}(\Pi_A{}^C). \tag{2.32b}$$

In the real case $\bar{\omega} = (\omega)^*$, and $\bar{\aleph} = (A_\mu + \check{A}_\mu)dX^\mu$, where the real A_μ are electric potentials $(f_{\mu\nu} = A_{\mu,\nu} - A_{\nu,\mu})$ and the pure imaginary $\check{A}_\mu$ are magnetic potentials $[\check{f}_{\mu\nu} \equiv (i/2\sqrt{-g})\epsilon^{\circ\circ\rho\sigma}_{\mu\nu}f_{\rho\sigma} = \check{A}_{\mu,\nu} - \check{A}_{\nu,\mu}]$.

Therefore, the integral varieties of the left and right Maxwell equations in the flat space are entirely determined by the integral varieties of the simple equations ${}^\circ DK_{\dot{A}} = 0 = {}^\circ\Box H$ and ${}^\circ DK_A = 0 = {}^\circ\Box\bar{H}$ correspondingly.

A similar calculation for $\mathcal{H}$-space leads to a special case of the electromagnetic Hertz potential given in Ref. 9.

3. A COVARIANT SPINORIZATION OF HEAVENS AND THE LEFT GRAVITATIONAL HERTZ POTENTIAL

If one intends to extend the idea of the Hertz potentials to the nonlinear dynamics of the Einsteinian gravity, it is natural to begin by examining the simple case of heavens ($\mathcal{H}$-spaces). These space–times (complex) which fulfill Einstein vacuum equations and have vanishing anti-self-dual part of the conformal curvature (i.e., they are "half-flat", $R_{\dot{A}\dot{B}} = 0$), were first encountered by Newman[10] in his study of the complexified asymptotics of gravitational radiation, and give rise to Penrose's concept of the nonlinear graviton.[11] The $\mathcal{H}$-spaces were parallelly studied in Refs. 2, 4, and 5, where the simple results of Ref. 2 (which provides an explicit construction of $\mathcal{H}$-spaces) were further developed. The results of our group were, however, described in a formalism which suffers some technical disadvantage: by freezing the hellish $\dot{S}L$ gauge and by working in such $\mathcal{G}$-frame where $\Gamma_{\dot{A}\dot{B}} = 0$, one loses insight into some structural properties which are residual (in $\mathcal{H}$-spaces) of the general structure of the $G \otimes G$ solutions. In this section we will be able to propose a new spinorial description of heavens with an unfrozen hellish gauge, where the standard results concerning heavens will became more revealing, particularly from the point of view of the nature and the structure of the gravitational Hertz potentials.

We begin by summarizing the basic results of Ref. 2. Thus, *every* (strong) heaven with $C_{\dot{A}\dot{B}\dot{C}\dot{D}} = C_{AB\dot{C}\dot{D}} = 0 = R$ can be described as follows: There exist a coordinate chart $\{X^\mu\} = \{u, v, x, y\}$ and the key function $\Theta = \Theta(u, v, x, y)$ such that in a special $\mathcal{G}$-frame the cotangent tetrad is given by

$$(g^{A\dot{B}}) = \sqrt{2}\begin{pmatrix} e^4, & e^2 \\ e^1, & -e^3 \end{pmatrix}$$

$$\underset{\text{s.f.}}{=} \sqrt{2}\begin{pmatrix} -dv, & dx - \Theta_{yy}du + \Theta_{xy}dv \\ du, & dy + \Theta_{xy}du - \Theta_{xx}dv \end{pmatrix}. \tag{3.1}$$

This formula is accompanied by the expression for the tangent tetrad:

$$(\partial_{A\dot{B}}) = -\sqrt{2}\begin{pmatrix} \partial_4, & \partial_2 \\ \partial, & -\partial_3 \end{pmatrix}$$

$$\underset{\text{s.f.}}{=} -\sqrt{2}\begin{pmatrix} -[\partial_v - \Theta_{xy}\partial_x + \Theta_{xx}\partial_y], & \partial_x \\ \partial_u + \Theta_{yy}\partial_x - \Theta_{xy}\partial_y, & \partial_y \end{pmatrix}. \tag{3.2}$$

The spinorial connection forms are then

$$(\Gamma_{AB}) \underset{\text{s.f.}}{=} \begin{pmatrix} \Theta_{xxy}du - \Theta_{xxx}dv, & \Theta_{xyy}du - \Theta_{xxy}dv \\ \text{idem}, & \Theta_{yyy}du - \Theta_{xyy}dv \end{pmatrix}$$

$$\Gamma_{\dot{A}\dot{B}} \underset{\text{s.f.}}{=} 0, \tag{3.3}$$

if it is assumed that the key function fulfills the second heavenly equation[2]:

$$\Theta_{xx}\Theta_{yy} - (\Theta_{xy})^2 + \Theta_{xu} + \Theta_{yv} = 0. \tag{3.4}$$

The only nontrivial curvature objects which accompany this "heavenly tetrad" are the components of C_{ABCD} given by

$$C_{1111} \underset{\text{s.f.}}{=} \Theta_{xxxx}, \quad C_{1112} \underset{\text{s.f.}}{=} \Theta_{xxxy},$$

$$C_{1122} \underset{\text{s.f.}}{=} \Theta_{xxyy}, \quad C_{1222} \underset{\text{s.f.}}{=} \Theta_{xyyy}, \tag{3.5}$$

$$C_{2222} \underset{\text{s.f.}}{=} \Theta_{yyyy}.$$

In order to provide a fully covariant description of these results, introduce in the $\mathcal{G}$-frame used above a homogeneous hellish spinor:

$$K_{\dot{A}} \underset{\text{s.f.}}{=} (1, 0) \to K^{\dot{A}} = (0, 1), \quad DK_{\dot{A}} \underset{\text{s.f.}}{=} 0 \tag{3.6}$$

[see comment after (2.19)]. Then, in the same frame of the spinorial gauge, we can introduce the *spinorial coordinates* labelled by the two indices:

$$({}^\circ g^{A\dot{B}}) \underset{\text{s.f.}}{=} (dX^{A\dot{B}}) =: \sqrt{2}\begin{pmatrix} -dv & dx \\ du & dy \end{pmatrix}. \tag{3.7}$$

We can now introduce—all the time in the same $\mathcal{G}$-frame—a new $D(0,2)$ object:

$$\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}} \underset{\text{s.f.}}{=}: 2\Theta K_{\dot{A}} K_{\dot{B}} K_{\dot{C}} K_{\dot{D}}, \tag{3.8}$$

which, with Θ treated as a scalar, is then defined in any $\mathcal{G}$-frame.

It is then a simple algebraic exercise to show that (3.1) and (3.2) can be equivalently rewritten in the form of

$$g^{A\dot{B}} \underset{\text{s.f.}}{=} dX^{A\dot{B}} - dX^{C\dot{D}}(\partial/\partial X_A{}^{\dot{R}})(\partial/\partial X^{C\dot{S}})(\Pi^{\dot{B}\dot{R}\dot{S}}{}_{\dot{D}}) \tag{3.9}$$

and

$$-\tfrac{1}{2}\partial_{A\dot{B}} \underset{\text{s.f.}}{=} (\partial/\partial X^{A\dot{B}}) + (\partial/\partial X^{A\dot{R}})(\partial/\partial X_C{}^{\dot{S}})(\Pi^{\dot{R}\dot{S}\dot{D}}{}_{\dot{B}}) \cdot \partial/\partial X^{C\dot{D}}. \tag{3.10}$$

At the same time, the formulas for the connections and the curvature assume the form:

$$\Gamma_{AB} \underset{\text{s.f.}}{=} dX^{C\dot{D}}(\partial/\partial X^{A\dot{R}})(\partial/\partial X^{B\dot{S}})(\partial/\partial X^{C\dot{T}})(\Pi^{\dot{R}\dot{S}\dot{T}}{}_{\dot{D}}),$$

$$\Gamma_{\dot{A}\dot{B}} \underset{\text{s.f.}}{=} 0, \tag{3.11}$$

and

$$\tfrac{1}{2}C_{ABCD} \underset{\text{s.f.}}{=} (\partial/\partial X^{A\dot{R}})(\partial/\partial X^{B\dot{S}})(\partial/\partial X^{C\dot{T}})(\partial/\partial X^{D\dot{U}})\Pi^{\dot{R}\dot{S}\dot{T}\dot{U}}. \tag{3.12}$$

The second heavenly equation can be then expressed in the terms of the object (3.8):

$$4(\Theta_{xu} + \Theta_{yv} + \Theta_{xx}\Theta_{yy} - \Theta_{xy}\Theta_{xy})K_{\dot{A}}K_{\dot{B}}K_{\dot{C}}K_{\dot{D}} \underset{\text{s.f.}}{=} -2(\partial^2/\partial X_{K\dot{L}}\partial X^{K\dot{L}})\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}} + 2(\partial^2\Pi_{\dot{P}\dot{Q}(\dot{A}\dot{B}}/\partial X^K{}_{\dot{P}}\partial X^L{}_{\dot{Q}}) \times (\partial^2\Pi_{\dot{C}\dot{D})\dot{R}\dot{S}}/\partial X_{K\dot{R}}\partial X_{L\dot{S}}) = 0. \tag{3.13}$$

The question arises whether formulas (3.9)–(3.13) can be written in a covariant form with respect to the complete gauge group, $\mathcal{G} = \mathrm{SL}\times\mathrm{S\dot{L}}$. An obvious manner of giving the positive answer to this question consists in using the both-sidedly flat tetrad from (3.7), defined through its values in our special frame. This tetrad [compare (2.23), (2.24), and (2.25)] induces the covariant differential, and through it the covariant gradient, ${}^\circ DT^{A\dot{B}\cdots} = -\tfrac{1}{2}{}^\circ g^{R\dot{S}}{}^\circ\nabla_{R\dot{S}}T^{A\dot{B}\cdots}$, for $T^{A\dot{B}\cdots} \in \Lambda^\circ$. Because in our s.f. both ${}^\circ\Gamma_{AB}$ and ${}^\circ\Gamma_{\dot{A}\dot{B}}$ vanish, it is clear that

$${}^\circ\nabla_{A\dot{B}} \underset{\text{s.f.}}{=} {}^\circ\partial_{A\dot{B}} \underset{\text{s.f.}}{=} -2\partial/\partial X^{A\dot{B}}. \tag{3.14}$$

It follows that we can now write our formulas which describe $\mathcal{H}$-space *in an arbitrary* $\mathcal{G}$*-frame* in the simple form of

$$\mathcal{H}\begin{cases} g^{A\dot{B}} = {}^\circ g^{A\dot{B}} + \tfrac{1}{4}{}^\circ g^{C\dot{D}}{}^\circ\nabla^A{}_{\dot{R}}{}^\circ\nabla_{C\dot{S}}(\Pi^{\dot{B}\dot{R}\dot{S}}{}_{\dot{D}}), & (3.15)\\ \partial_{A\dot{B}} = {}^\circ\partial_{A\dot{B}} - \tfrac{1}{4}{}^\circ\nabla_{A\dot{R}}{}^\circ\nabla^C{}_{\dot{S}}(\Pi^{\dot{R}\dot{S}\dot{D}}{}_{\dot{B}}){}^\circ\partial_{C\dot{D}}, & (3.16)\\ \Gamma_{AB} = {}^\circ\Gamma_{AB} - \tfrac{1}{8}{}^\circ g^{C\dot{D}}{}^\circ\nabla_{A\dot{R}}{}^\circ\nabla_{B\dot{S}}{}^\circ\nabla_{C\dot{T}}(\Pi^{\dot{R}\dot{S}\dot{T}}{}_{\dot{D}}), & (3.17)\\ \Gamma_{\dot{A}\dot{B}} = {}^\circ\Gamma_{\dot{A}\dot{B}} & (3.18)\\ C_{ABCD} = \tfrac{1}{8}{}^\circ\nabla_{A\dot{R}}{}^\circ\nabla_{B\dot{S}}{}^\circ\nabla_{C\dot{T}}{}^\circ\nabla_{D\dot{U}}(\Pi^{\dot{R}\dot{S}\dot{T}\dot{U}}), & (3.19)\\ C_{\dot{A}\dot{B}\dot{C}\dot{D}} = 0, & (3.20)\\ {}^\circ\Box\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}} + \tfrac{1}{8}{}^\circ\nabla_K{}^{\dot{P}}{}^\circ\nabla_L{}^{\dot{Q}}\Pi_{\dot{P}\dot{Q}(\dot{A}\dot{B}}{}^\circ\nabla^{K\dot{R}}{}^\circ\nabla^{L\dot{S}}\Pi_{\dot{C}\dot{D})\dot{R}\dot{S}} = 0, & (3.21)\end{cases}$$

where, of course, we denoted ${}^\circ\Box \equiv -\tfrac{1}{2}{}^\circ\nabla_{R\dot{S}}{}^\circ\nabla^{R\dot{S}} \underset{\text{s.f.}}{=} -2(\partial/\partial X^{R\dot{S}})(\partial/\partial X^{R\dot{S}})$. (At this point, we can mention in passing that if one defines the covariant codifferential $\tilde{D} \equiv -i{*}D{*}$, then the "covariant" Laplacian $D\tilde{D} + \tilde{D}D$: $\Lambda^\bullet \to \Lambda^\bullet$ coincides in Λ° with operation $\Box$; of course working with ${}^\circ D$ we obtain then ${}^\circ\Box$.)

We can add that by using the results of Ref. 2 concerning the S objects in our s.f., and then by generalizing them to an arbitrary $\mathcal{G}$-frame, we can accompany formulas (3.15)–(3.21) by

$$\begin{cases} S^{AB} = {}^\circ S^{AB} - \tfrac{1}{4}{}^\circ S^{\dot{R}\dot{S}}{}^\circ\nabla^A{}_{\dot{T}}{}^\circ\nabla^B{}_{\dot{U}}(\Pi_{\dot{R}\dot{S}}{}^{\dot{T}\dot{U}}), & (3.22)\\ S^{\dot{A}\dot{B}} = {}^\circ S^{\dot{A}\dot{B}} + \tfrac{1}{4}{}^\circ D({}^\circ g^R{}_{\dot{D}}{}^\circ\nabla_{R\dot{C}}\Pi^{\dot{A}\dot{B}\dot{C}\dot{D}}) \\ \quad = {}^\circ S^{\dot{A}\dot{B}} - \tfrac{1}{8}{}^\circ S^{\dot{C}\dot{D}}{}^\circ\Box\Pi^{\dot{A}\dot{B}}{}_{\dot{C}\dot{D}} + \tfrac{1}{8}{}^\circ S^{CD}{}^\circ\nabla_C{}^{\dot{R}}{}^\circ\nabla_D{}^{\dot{S}}\Pi^{\dot{A}\dot{B}}{}_{\dot{R}\dot{S}}. & (3.23)\end{cases}$$

To close the covariant description of heavens given by these formulas, we must, of course, list the information that

and

with

$$\mathcal{H}\begin{cases} {}^\circ D{}^\circ g^{A\dot{B}} = 0, \quad {}^\circ D{}^\circ\Gamma_{AB} = 0 = {}^\circ D{}^\circ\Gamma_{\dot{A}\dot{B}}, & (3.24)\\ \Pi_{\dot{A}\dot{B}\dot{C}\dot{D}} = 2\Theta K_{\dot{A}\dot{B}\dot{C}\dot{D}}\text{—is of the type N}, & (3.25)\\ {}^\circ DK_{\dot{A}} = 0 \longleftrightarrow {}^\circ\nabla_{A\dot{B}}K_{\dot{C}} = 0 \;\begin{cases}\text{defining a homogenous}\\ \text{spinor.}\end{cases} & (3.26)\end{cases}$$

We will now demonstrate a *theorem*: formulas)–(3.15)–(3.26), which describe covariantly (with respect to $\mathcal{G}$) the $\mathcal{H}$-spaces, remain valid if in all terms which contain the object $\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}$ (or the single spinor $K_{\dot{A}}$) we replace the objects and operations referring to the flat tetrad by these referring to the complete tetrad of $\mathcal{H}$-spaces according to the following scheme:

$${}^\circ g^{A\dot{B}} \to g^{AB}, \tag{3.27a}$$

$${}^\circ S_{AB} \to S_{AB} \tag{3.27b}$$

$${}^\circ S_{\dot{A}\dot{B}} \to S_{\dot{A}\dot{B}}, \tag{3.27c}$$

$${}^\circ\partial_{A\dot{B}} \to \partial_{AB}, \tag{3.27d}$$

$${}^\circ\nabla_{A\dot{B}} \to \nabla_{A\dot{B}}, \tag{3.27e}$$

$${}^\circ D \to D, \tag{3.27f}$$

$${}^\circ\Box \to \Box. \tag{3.27g}$$

We will prove this theorem in several steps. First we notice that because $\Gamma_{\dot{A}\dot{B}} = {}^\circ\Gamma_{\dot{A}\dot{B}}$, manifestly

$${}^\circ DK_{\dot{A}} = 0 \to DK_{\dot{A}} = 0 \longleftrightarrow \nabla_{A\dot{B}}K_{\dot{C}} = 0. \tag{3.28}$$

Then, remembering that (3.15) and (3.16) can be interpreted in the sense that

$$g^{A\dot{B}} = {}^\circ g^{A\dot{B}} + (\text{something}) \cdot K^{\dot{B}},$$

$$\partial_{A\dot{B}} = {}^\circ\partial_{A\dot{B}} + (\text{something}) \cdot K_{\dot{B}}. \tag{3.29}$$

we easily infer that

$$\begin{aligned} g^{A\dot{B}}(\text{diff. op})\Pi_{\dot{B}\dot{C}\dot{D}\dot{E}} &= {}^\circ g^{A\dot{B}}(\text{diff. op})\Pi_{\dot{B}\dot{C}\dot{D}\dot{E}}, \\ (\text{diff. op})\Pi^{\dot{B}\dot{C}\dot{D}\dot{E}}\partial_{A\dot{B}} &= (\text{diff. op})\Pi^{\dot{B}\dot{C}\dot{D}\dot{E}}{}^\circ\partial_{A\dot{B}}, \end{aligned} \tag{3.30}$$

where (diff. op) can be any differential operator constructed from the covariant gradients ${}^\circ\nabla_{A\dot{B}}$ and $\nabla_{A\dot{B}}$ with respect to which $K_{\dot{A}}$ is constant.

In particular, applying the mechanism discussed, we have

$$\Gamma_{AB} \underset{\text{s.f.}}{=} g^{C\dot{D}}(\partial/\partial X^{A\dot{R}})(\partial/\partial X^{B\dot{S}})(\partial/\partial X^{C\dot{T}})\Pi^{\dot{R}\dot{S}\dot{T}}{}_{\dot{D}}. \qquad (3.31)$$

Applying then the rule

$$-\tfrac{1}{2}g_{A\dot{B}}{}^{\mu}g^{C\dot{D}}{}_{\mu} = \delta_A{}^C\delta_{\dot{B}}{}^{\dot{D}} \qquad (3.32)$$

[where, of course, $g_{A\dot{B}} =: g_{A\dot{B}\mu}dX^{\mu}$ and local-coordinate indices are manipulated by the Riemannian metric $g_{\mu\nu}$], we easily infer that

$$-\tfrac{1}{2}\Gamma_{AB\mu}g^{\mu}{}_{C\dot{D}} \underset{\text{s.f.}}{=} (\partial/\partial X^{A\dot{R}})(\partial/\partial X^{B\dot{S}})(\partial/\partial X^{C\dot{T}})(\Pi^{\dot{R}\dot{S}\dot{T}}{}_{\dot{D}})$$

$$= (\text{something})\cdot K_{\dot{D}}. \qquad (3.33)$$

We can now prove a lemma that

$$\nabla_{A_1\dot{B}_1}\cdots\nabla_{A_k\dot{B}_k}\Pi^{\dot{B}_1\cdots\dot{B}_k\dot{C}_1\cdots\dot{C}_{4-k}}$$

$$= {}^{\circ}\nabla_{A_1\dot{B}_1}\cdots{}^{\circ}\nabla_{A_k\dot{B}_k}\Pi^{\dot{B}_1\cdots\dot{B}_k\dot{C}_1\cdots\dot{C}_{4-k}}, \qquad (3.34)$$

for $k = 1, 2, 3, 4$. We prove this by induction with respect to k.

For $k=1$, because of (3.29) we have

$$\nabla_{A\dot{B}}\Pi^{\dot{B}\dot{C}_1\dot{C}_2\dot{C}_3} \underset{\text{s.f.}}{=} {}^{\circ}\partial_{A\dot{B}}\Pi^{\dot{B}\dot{C}_1\dot{C}_2\dot{C}_3} \underset{\text{s.f.}}{=} {}^{\circ}\nabla_{A\dot{B}}\Pi^{\dot{B}\dot{C}_1\dot{C}_2\dot{C}_3}. \qquad (3.35)$$

We assume now (3.34) for some $k \geq 1$. Then

$$\nabla_{A_1\dot{B}_1}\cdots\nabla_{A_{k+1}\dot{B}_{k+1}}\Pi^{\dot{B}_1\cdots\dot{B}_{k+1}\dot{C}_1\cdots\dot{C}_{3-k}}$$

$$= \nabla_{A_1\dot{B}_1}{}^{\circ}\nabla_{A_2\dot{B}_2}\cdots{}^{\circ}\nabla_{A_{k+1}\dot{B}_{k+1}}\Pi^{\dot{B}_1\cdots\dot{B}_{k+1}\dot{C}_1\cdots\dot{C}_{3-k}}$$

$$\underset{\text{s.f.}}{=} \partial_{A_1\dot{B}_1}({}^{\circ}\nabla_{A_2\dot{B}_2}\cdots{}^{\circ}\nabla_{A_{k+1}\dot{B}_{k+1}}\Pi^{\dot{B}_1\cdots\dot{B}_{k+1}\dot{C}_1\cdots\dot{C}_{3-k}})$$

$$= \Gamma^{S}{}_{A_2\mu}g^{\mu}{}_{A_1\dot{B}_1}\cdot{}^{\circ}\nabla_{S\dot{B}_2}\cdots{}^{\circ}\nabla_{A_{k+1}\dot{B}_{k+1}}\Pi^{\dot{B}_1\cdots\dot{B}_{k+1}\dot{C}_1\cdots\dot{C}_{3-k}} + \cdots, \qquad (3.36)$$

where $(\cdots)$ denotes the terms with $\Gamma^{S}{}_{A_i\mu}g^{\mu}{}_{A_1\dot{B}_1}$, which correspondingly take care of all undotted indices. Because, however, of (3.33), all these terms contain the contraction $K_{\dot{B}_1}\Pi^{\dot{B}_1\cdots}$, and hence all vanish. Using thus in the term with $\partial_{A_1\dot{B}_1}$ the second of the properties (3.29), we can replace it by ${}^{\circ}\partial_{A_1\dot{B}_1} \underset{\text{s.f.}}{=} {}^{\circ}\nabla_{A_1\dot{B}_1}$. We have therefore

$$\nabla_{A_1\dot{B}_1}\cdots\nabla_{A_{k+1}\dot{B}_{k+1}}\Pi^{\dot{B}_1\cdots\dot{B}_{k+1}\dot{C}_1\cdots\dot{C}_{3-k}}$$

$$\underset{\text{s.f.}}{=} {}^{\circ}\nabla_{A_1\dot{B}_1}\cdots{}^{\circ}\nabla_{A_{k+1}\dot{B}_{k+1}}\Pi^{\dot{B}_1\cdots\dot{B}_{k+1}\dot{C}_1\cdots\dot{C}_{3-k}}. \qquad (3.37)$$

Being valid in an s.f., this covariant equation is valid in any frame. This concludes the inductive proof of (3.34).

We still must demonstrate—as the last lemma necessary for our purposes—that

$$\Box\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}} = {}^{\circ}\Box\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}. \qquad (3.38)$$

We again prove it in our s.f. First, we have

$$\Box\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}} = -\tfrac{1}{2}\nabla_{R\dot{S}}\nabla^{R\dot{S}}\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}} \underset{\text{s.f.}}{=} -\tfrac{1}{2}\nabla_{R\dot{S}}\partial^{R\dot{S}}\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}$$

$$= -\tfrac{1}{2}\partial_{R\dot{S}}\partial^{R\dot{S}}\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}} - \tfrac{1}{2}\Gamma^{R}{}_{K\mu}g^{\mu}{}_{R\dot{S}}\partial^{K\dot{S}}\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}, \qquad (3.39)$$

where for $\partial_{R\dot{S}}$ we can use (3.10), and in the last term we can apply (3.33). In doing so, it is relevant to remember a formal property of the tangent spinors

$$(\partial/\partial X^{R\dot{S}})\epsilon^{RT} = -\partial/\partial X_T{}^{\dot{S}}, \qquad (3.40)$$

which was first clearly encountered and discussed in Ref. 4 [this property is important also when rewriting (3.9) and (3.10) in the covariant form of (3.15) and (3.16); (3.40) explains the necessity of an additional change of sign in these formulas]. We obtain, therefore,

$$\Box\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}} \underset{\text{s.f.}}{=} -2\{\partial/\partial X^{R\dot{S}} + (\partial/\partial X^{R\dot{P}})(\partial/\partial X_K{}^{\dot{Q}})(\Pi^{\dot{P}\dot{Q}\dot{U}}{}_{\dot{S}})\partial/\partial X^{K\dot{U}}\}$$

$$\times\{\partial/\partial X_{R\dot{S}} - (\partial/\partial X_R{}^{\dot{M}})(\partial/\partial X_L{}^{\dot{N}})(\Pi^{\dot{M}\dot{N}\dot{U}\dot{S}})\partial/\partial X^{L\dot{V}}\}$$

$$\times\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}} + 2(\partial/\partial X_R{}^{\dot{P}})(\partial/\partial X^{K\dot{Q}})(\partial/\partial X^{R\dot{N}})(\Pi^{\dot{P}\dot{Q}\dot{N}}{}_{\dot{S}})$$

$$\times\{\partial/\partial X_{K\dot{S}} - (\partial/\partial X_K{}^{\dot{M}})(\partial/\partial X_L{}^{\dot{N}})(\Pi^{\dot{M}\dot{N}\dot{V}\dot{S}})\partial/\partial X^{L\dot{V}}\}$$

$$\times\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}. \qquad (3.41)$$

Executing here the differentiations, cancelling and dropping all terms which contain the contractions like $K_{\dot{S}}K^{\dot{S}}$, we easily find that

$$\Box\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}} \underset{\text{s.f.}}{=} -2\partial/\partial X_{R\dot{S}}(\partial/\partial X^{R\dot{S}})\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}$$

$$= -\tfrac{1}{2}{}^{\circ}\nabla_{R\dot{S}}{}^{\circ}\nabla^{R\dot{S}}\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}} = {}^{\circ}\Box\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}. \qquad (3.42)$$

Valid in an s.f., this covariant equation must also hold in any $\mathcal{G}$-frame, and thus (3.88) is proven. [Notice that already in Ref. 2 treating Θ as a scalar it was established that $\Box\Theta = {}^{\circ}\Box\Theta$ which, with a constant $K_{\dot{A}}$, is essentially equivalent to (3.38); the present proof is given in order to assure the completness of this paper.]

We shall see that the sequence of the lemmas demonstrated above proves our theorem; indeed, that we can replace ${}^{\circ}g_{A\dot{B}}$ and ${}^{\circ}\partial_{A\dot{B}}$ by $g_{A\dot{B}}$ and $\partial_{A\dot{B}}$ in the right-hand members of (3.15), (3.16), and (3.17) follows from (3.30). That we can replace ${}^{\circ}\nabla_{A\dot{B}}$ by $\nabla_{A\dot{B}}$ in all formulas beginning from (3.15) to (3.23) follows from the lemma (3.34); then, (3.38) guarantees that we can replace ${}^{\circ}\Box$ by $\Box$ when acting on $\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}$. As far as the operation ${}^{\circ}D$ is concerned, we have already (3.28); ${}^{\circ}D$ in (3.23) amounts to the operation on the object with pure dotted indices so that with $\Gamma_{\dot{A}\dot{B}} = {}^{\circ}\Gamma_{\dot{A}\dot{B}}$ it can be again replaced here by D. Eventually, the fact that we can replace in the terms with $\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}$ the S's by ${}^{\circ}S$'s in (3.22) and (3.23) is the consequence of the fact $S_{\dot{A}\dot{B}} = {}^{\circ}S_{\dot{A}\dot{B}} + (\text{something})\, K_{\dot{A}}K_{\dot{B}}$ and that ${}^{\circ}D({}^{\circ}g^{R}{}_{\dot{D}}{}^{\circ}\nabla_{R\dot{C}}\Pi^{\dot{A}\dot{B}\dot{C}\dot{D}}) = D(g^{R}{}_{\dot{D}}\nabla_{R\dot{C}}\Pi^{\dot{A}\dot{B}\dot{C}\dot{D}})$ according to our lemmas.

The replacements of objects and operators referring to ${}^{\circ}g_{A\dot{B}}$ by those referring to $g_{A\dot{B}}$ according to the scheme (3.27) in the formulas (3.15)–(3.16)—of course, only in the terms involving $\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}$ or $K_{\dot{A}}$—is therefore leading to valid formulas either if we execute these replacements only partially, i.e., in some of these formulas, or completely, eliminating the objects and operations referring to ${}^{\circ}g_{A\dot{B}}$ in all possible places. In the latter case, we obtain therefore a set of formulae describing $\mathcal{H}$-spaces in the form of

$$\mathcal{H}\begin{cases} {}^\circ g^{A\dot{B}} = g^{A\dot{B}} - \frac{1}{4} g^{C\dot{D}} \nabla^{A}{}_{\dot{R}} \nabla_{C\dot{S}} (\Pi^{\dot{B}\dot{R}\dot{S}}{}_{\dot{D}}), & (3.43)\\ {}^\circ\partial_{A\dot{B}} = \partial_{A\dot{B}} + \frac{1}{4}\nabla_{A\dot{R}}\nabla^{C}{}_{\dot{S}}(\Pi^{\dot{R}\dot{S}\dot{D}}{}_{\dot{B}})\partial_{C\dot{D}}, & (3.44)\\ {}^\circ\Gamma_{AB} = \Gamma_{AB} + \frac{1}{8} g^{C\dot{D}} \nabla_{A\dot{R}} \nabla_{B\dot{S}} \nabla_{C\dot{T}} (\Pi^{\dot{R}\dot{S}\dot{T}}{}_{\dot{D}}), & (3.45)\\ {}^\circ\Gamma_{\dot{A}\dot{B}} = \Gamma_{\dot{A}\dot{B}}, & (3.46)\\ C_{ABCD} = \frac{1}{8}\nabla_{A\dot{R}}\nabla_{B\dot{S}}\nabla_{C\dot{T}}\nabla_{D\dot{U}}(\Pi^{\dot{R}\dot{S}\dot{T}\dot{U}}), & (3.47)\\ C_{\dot{A}\dot{B}\dot{C}\dot{D}} = 0, & (3.48)\\ \Box\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}} = -\frac{1}{8}\nabla_{K}{}^{\dot{P}}\nabla_{L}{}^{\dot{Q}}\Pi_{\dot{P}\dot{Q}(\dot{A}\dot{B}}\cdot\nabla^{K\dot{R}}\nabla^{L\dot{S}}\Pi_{\dot{C}\dot{D})\dot{R}\dot{S}}, & (3.49)\\ {}^\circ S^{AB} = S^{AB} + \frac{1}{4} S^{\dot{R}\dot{S}} \nabla^{A}{}_{\dot{T}} \nabla^{B}{}_{\dot{U}} (\Pi_{\dot{R}\dot{S}}{}^{\dot{U}\dot{T}}), & (3.50)\\ {}^\circ S^{\dot{A}\dot{B}} = S^{\dot{A}\dot{B}} - \frac{1}{4} D(g^{R}{}_{\dot{D}}\nabla_{R\dot{C}}\Pi^{\dot{A}\dot{B}\dot{C}\dot{D}}) & \\ \quad = S^{\dot{A}\dot{B}} + \frac{1}{8}\Box\Pi^{\dot{A}\dot{B}}{}_{\dot{C}\dot{D}}\cdot S^{\dot{C}\dot{D}} - \frac{1}{8}S^{CD}\nabla_{C}{}^{\dot{R}}\nabla_{D}{}^{\dot{S}}(\Pi^{\dot{A}\dot{B}}{}_{\dot{R}\dot{S}}), & (3.51)\\ {}^\circ D{}^\circ g^{A\dot{B}} = 0, \quad {}^\circ D{}^\circ\Gamma_{AB} = 0 = {}^\circ D{}^\circ\Gamma_{\dot{A}\dot{B}}, & (3.52)\\ \Pi_{\dot{A}\dot{B}\dot{C}\dot{D}} = 2\Theta K_{\dot{A}}K_{\dot{B}}K_{\dot{C}}K_{\dot{D}} \text{—is of the type N}, & (3.53)\\ DK_{\dot{A}} = 0 \longleftrightarrow \nabla_{A\dot{B}}K_{\dot{C}} = 0 \quad \begin{cases}\text{defining a homogenous}\\ \text{spinor.}\end{cases} & (3.54) \end{cases}$$

It should be noted that in $\mathcal{H}$ spaces with $R_{\dot{A}\dot{B}} = 0$, according to (2.7) from Ref. 6, we have

$$\nabla_{(C}{}^{\dot{N}}\nabla_{D)\dot{N}}\psi^{\dot{B}_1\cdots\dot{B}_{2q}}_{A_1\cdots A_{2p}} = 4pC^{N}{}_{CD(A_1}\psi^{\dot{B}_1\cdots\dot{B}_{2q}}_{A_2\cdots A_{2p})N}, \tag{3.55a}$$

$$\nabla^{N(\dot{C}}\nabla_{N}{}^{\dot{D})}\psi^{\dot{B}_1\cdots\dot{B}_{2q}}_{A_1\cdots A_{2p}} = 0, \tag{3.55b}$$

and consequently the covariant gradients do commute on the objects of the type D(0,q); in particular they commute on $\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}$ which assures the correct symmetries in (3.43)–(3.54).

Our last set of formulas can be now interpreted as follows: In $\mathcal{H}$-spaces there exists the left gravitational Hertz potential $\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}$—of the type D(0,2)—which fulfills the wavelike equation (3.49); this equation can be seen to be the necessary (and sufficient) condition for ${}^\circ S_{\dot{A}\dot{B}}$ defined as the right-hand member of (3.51) to fulfill the crucial algebraic condition of $\mathcal{S}^{E}_{H}$ structures discussed in Ref. 23:

$${}^\circ S^{\dot{A}\dot{B}} \wedge {}^\circ S_{\dot{C}\dot{D}} = \dot{\rho}_0\, \delta^{\dot{A}}{}_{(\dot{C}}\delta^{\dot{B}}{}_{\dot{D})}, \quad \dot{\rho}_0 \neq 0. \tag{3.56}$$

In the terms of the graviational Hertz potential, the graviational field (of helicity $+2\hbar$) C_{ABCD} is given by (3.47) and fulfills the wave equation $\nabla^{S}{}_{\dot{A}}C_{SBCD} = 0$ [compare (1.19a)]. Moreover, the new tetrad constructed from the tetrad of the $\mathcal{H}$-space and the potential, either in the cotangent form (3.43) or the tangent form (3.44), is *flat*, i.e., induces the both-sidedly flat connections ${}^\circ D{}^\circ\Gamma_{AB} = 0 = {}^\circ D{}^\circ\Gamma_{\dot{A}\dot{B}}$. The potential which assures all these things is selected to be of the type N, with the quadruple Penrose spinor being proportional to the homogeneous spinor.

We should like to emphasize the striking analogy of this result with the description of the general left and right electromagnetic fields in flat space–time in the terms of the Hertz potentials in the null gauge, formulas (2.31) and (2.32). We also notice the analogy of our result concerning the theorem about the possibility of equivalently replacing objects and operators referred to ${}^\circ g_{A\dot{B}}$ and g_{AB}, with similar mechanisms which one encounters in the theory of the Kerr–Schild metrics[12,13] and, more generally, of the double K–S metrics (see Ref. 14 and then Ref. 15 for the general theory of the last metrics).

Suppose, for example, that

$$g = {}^\circ g + K \underset{s}{\otimes} K \in \Lambda^1 \otimes \Lambda^1,$$

where ${}^\circ g$ is flat and $K \in \Lambda^1$ is null (i.e., $K \lrcorner K = 0$) and geodesic with respect to ${}^\circ g$ or g. It is then null and geodesic with respect to the both, and its optical scalars are the same for both metrics. The key to these properties (and to the similar properties of the congruences of null strings for the double K–S metrics) is, of course, the *nullity* of the structure which modifies the basic metric. In $\mathcal{H}$-spaces too, the *nullity* of $\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}$, and the proportionality of the quadruple Penrose spinor to a homogeneous spinor, are the properties underlying our theorem.

4. COMPLEX GRAVITY IN LINEAR APPROXIMATION

The dynamical equations of general relativity have been studied in the linear approximation from many points of view. Our results on $\mathcal{H}$-spaces suggest yet another approach. $\mathcal{H}$-space with the tetrad oriented as in the previous section, provides the most general right-flat solution of Einstein's equations. Changing the orientation (by making a tetrad transformation of determinant minus one), we obtain the most general left-flat solution. From these solutions of the rigorous equations, we derive solutions to the equations of the linear approximation. For these equations, however, we can superimpose solutions. By this means, we recover the general solution of the linear approximation which Penrose[8] obtained from completely different considerations.

Following the program outlined above, we first construct the most general "left-flat" solution. The null tetrad transformation $e^1 \to e^2$, $e^2 \to e^1$, $e^3 \to e^3$, $e^4 \to e^4$ corresponds to the replacement of the dotted indices by undotted and vice-versa, i.e., we obtain the hellish tetrad from the formulas (3.15)–(3.26) formally by "conjugating" and treating the objects $g^{A\dot{B}}$, ${}^\circ g^{A\dot{B}}$, ${}^\circ\nabla_{A\dot{B}}$ (and ${}^\circ\Box$) as if they were "Hermitian". This leads to the following list of formulas:

$$\mathcal{H}:\begin{cases} g^{A\dot{B}} = {}^\circ g^{A\dot{B}} + \frac{1}{4}{}^\circ g^{D\dot{C}}{}^\circ\nabla_{R}{}^{\dot{B}}{}^\circ\nabla_{S\dot{C}}(\Pi^{ARS}{}_{D}), & (4.1)\\ \partial_{A\dot{B}} = {}^\circ\partial_{A\dot{B}} - \frac{1}{4}{}^\circ\nabla_{R\dot{B}}{}^\circ\nabla_{S}{}^{\dot{C}}(\Pi^{RSD}{}_{A}){}^\circ\partial_{D\dot{C}}, & (4.2)\\ \Gamma_{AB} = {}^\circ\Gamma_{AB}, & (4.3)\\ \Gamma_{\dot{A}\dot{B}} = {}^\circ\Gamma_{\dot{A}\dot{B}} - \frac{1}{8}{}^\circ g^{D\dot{C}}{}^\circ\nabla_{R\dot{A}}{}^\circ\nabla_{S\dot{B}}{}^\circ\nabla_{T\dot{C}}(\Pi^{RST}{}_{D}), & (4.4)\\ C_{ABCD} = 0, & (4.5)\\ C_{\dot{A}\dot{B}\dot{C}\dot{D}} = \frac{1}{8}{}^\circ\nabla_{R\dot{A}}{}^\circ\nabla_{S\dot{B}}{}^\circ\nabla_{T\dot{C}}{}^\circ\nabla_{U\dot{D}}(\Pi^{RSTU}), & (4.6)\\ {}^\circ\Box\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}} = -\frac{1}{8}{}^\circ\nabla^{P}{}_{\dot{K}}{}^\circ\nabla^{Q}{}_{\dot{L}}\Pi_{PQ(AB}\,{}^\circ\nabla^{R\dot{K}}{}^\circ\nabla^{S\dot{L}}\Pi_{CD)RS}, & (4.7)\\ S^{AB} = {}^\circ S^{AB} + \frac{1}{4}{}^\circ D({}^\circ g_{D}{}^{\dot{R}}{}^\circ\nabla_{C\dot{R}}\Pi^{ABCD}) & (4.8)\\ \quad = {}^\circ S^{AB} - \frac{1}{8}{}^\circ S^{CD}{}^\circ\Box\Pi^{AB}{}_{CD} + \frac{1}{8}{}^\circ S^{\dot{R}\dot{S}}{}^\circ\nabla^{C}{}_{\dot{R}}{}^\circ\nabla^{D}{}_{\dot{S}}(\Pi^{AB}{}_{CD}), & \\ S^{\dot{A}\dot{B}} = {}^\circ S^{\dot{A}\dot{B}} - \frac{1}{4}{}^\circ S^{RS}{}^\circ\nabla_{T}{}^{\dot{A}}{}^\circ\nabla_{U}{}^{\dot{B}}(\Pi_{RS}{}^{TU}), & (4.9)\\ {}^\circ D{}^\circ g_{A\dot{B}} = 0, \quad {}^\circ D{}^\circ\Gamma_{AB} = 0 = {}^\circ D{}^\circ\Gamma_{\dot{A}\dot{B}}, & (4.10)\\ \Pi_{ABCD} = 2\bar{\Theta}K_{A}K_{B}K_{C}K_{D} \text{ is of the type N} & (4.11)\\ {}^\circ DK_{A} = 0 \longleftrightarrow {}^\circ\nabla_{A\dot{B}}K_{C} = 0 \quad \begin{cases}\text{defining a homogenous}\\ \text{spinor}\end{cases} & (4.12) \end{cases}$$

It is self-evident that our theorem concerning the possibility of replacing objects and operators according to the scheme (3.27) in all terms containing Π_{ABCD} (or K_A) applies also the present collection of the formulae. Clearly, Π_{ABCD} plays the role of the right (null) gravitational Hertz potential for the space $\mathring{\mathcal{H}}$.

In the next step, we consider, in both sets of the formulas (3.15)–(3.26) [$\mathcal{H}$] and (4.1)–(4.12) [$\mathring{\mathcal{H}}$], the corresponding Hertz potentials as the quantities of the first order, $\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}} \to \delta\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}$, $\Pi_{ABCD} \to \delta\Pi_{ABCD}$ (i.e., $\Theta \to \delta\Theta$, $\bar{\Theta} \to \delta\bar{\Theta}$), where "δ" denotes the order in the parameter of smallness. Then, neglecting the terms of higher order and superposing linearly both structures, we obtain for the solutions of Einstein equations which are both-sidely general, but only infinitesimally deviate from the flatness ($\delta G \otimes \delta G$ solutions) the following collection of formulas:

$$\delta G\otimes\delta G:\left\{\begin{aligned}
& g^{A\dot{B}} = {}^\circ g^{A\dot{B}} + \tfrac{1}{4}{}^\circ g^{C\mathring{D}}\,{}^\circ\nabla^{A}{}_{\dot{R}}\,{}^\circ\nabla_{C\mathring{S}}(\delta\Pi^{\dot{B}\dot{R}\dot{S}}{}_{\dot{D}}) \\
&\qquad + \tfrac{1}{4}{}^\circ g^{D\mathring{C}}\,{}^\circ\nabla_{R}{}^{\dot{B}}\,{}^\circ\nabla_{S\mathring{C}}(\delta\Pi^{ARS}{}_{D}), && (4.13)\\
& \partial_{A\mathring{B}} = {}^\circ\partial_{A\dot{B}} - \tfrac{1}{4}{}^\circ\nabla_{A\dot{R}}\,{}^\circ\nabla^{C}{}_{\mathring{S}}(\delta\Pi^{\dot{R}\dot{S}\mathring{D}}{}_{\dot{B}})\,{}^\circ\partial_{C\dot{D}} \\
&\qquad - \tfrac{1}{4}{}^\circ\nabla_{R\dot{B}}\,{}^\circ\nabla_{S}{}^{\dot{C}}(\delta\Pi^{RSD}{}_{A})\,{}^\circ\partial_{D\dot{C}}, && (4.14)\\
& \Gamma_{AB} = {}^\circ\Gamma_{AB} - \tfrac{1}{8}{}^\circ g^{C\mathring{D}}\,{}^\circ\nabla_{A\mathring{R}}\,{}^\circ\nabla_{B\dot{S}}\,{}^\circ\nabla_{C\mathring{T}}(\delta\Pi^{\dot{R}\dot{S}\dot{T}}{}_{\mathring{D}}), && (4.15)\\
& \Gamma_{\mathring{A}\mathring{B}} = {}^\circ\Gamma_{\dot{A}\mathring{B}} - \tfrac{1}{8}{}^\circ g^{D\dot{C}}\,{}^\circ\nabla_{R\dot{A}}\,{}^\circ\nabla_{S\dot{B}}\,{}^\circ\nabla_{T\mathring{C}}(\delta\Pi^{RST}{}_{D}), && (4.16)\\
& C_{ABCD} = \tfrac{1}{8}{}^\circ\nabla_{A\mathring{R}}\,{}^\circ\nabla_{B\mathring{S}}\,{}^\circ\nabla_{C\dot{T}}\,{}^\circ\nabla_{D\mathring{U}}(\delta\Pi^{\mathring{R}\mathring{S}\dot{T}\mathring{U}}), && (4.17)\\
& C_{\mathring{A}\dot{B}\mathring{C}\mathring{D}} = \tfrac{1}{8}{}^\circ\nabla_{R\mathring{A}}\,{}^\circ\nabla_{S\dot{B}}\,{}^\circ\nabla_{T\dot{C}}\,{}^\circ\nabla_{U\dot{D}}(\delta\Pi^{RSTU}), && (4.18)\\
& {}^\circ\Box\delta\Pi_{\mathring{A}\mathring{B}\mathring{C}\mathring{D}} = 0, && (4.19)\\
& {}^\circ\Box\delta\Pi_{ABCD} = 0, && (4.20)\\
& S^{AB} = {}^\circ S^{AB} - \tfrac{1}{4}{}^\circ S^{\mathring{R}\mathring{S}}\,{}^\circ\nabla^{A}{}_{\mathring{T}}\,{}^\circ\nabla^{B}{}_{\dot{U}}(\delta\Pi_{\dot{R}\mathring{S}}{}^{\mathring{T}\mathring{U}}) \\
&\qquad + \tfrac{1}{4}{}^\circ D({}^\circ g_{D}{}^{\mathring{R}}\,{}^\circ\nabla_{C\mathring{R}}\delta\Pi^{ABCD}) \\
&\quad = {}^\circ S^{AB} - \tfrac{1}{8}{}^\circ S^{CD}\,{}^\circ\Box\delta\Pi^{AB}{}_{CD} + \tfrac{1}{8}{}^\circ S^{\dot{R}\mathring{S}}\{{}^\circ\nabla^{C}{}_{\mathring{R}}\,{}^\circ\nabla^{D}{}_{\mathring{S}} \\
&\qquad \times\delta\Pi^{AB}{}_{CD} - 2\,{}^\circ\nabla^{A}{}_{\mathring{T}}\,{}^\circ\nabla^{B}{}_{\mathring{U}\mathring{C}}\Pi_{\mathring{R}\mathring{S}}{}^{\mathring{T}\mathring{U}}\}, && (4.21)\\
& S^{\mathring{A}\dot{B}} = {}^\circ S^{\mathring{A}\dot{B}} - \tfrac{1}{4}{}^\circ S^{RS}\,{}^\circ\nabla_{T}{}^{\mathring{A}}\,{}^\circ\nabla_{U}{}^{\dot{B}}(\delta\Pi_{RS}{}^{TU}) \\
&\qquad + \tfrac{1}{4}{}^\circ D({}^\circ g^{R}{}_{\dot{D}}\,{}^\circ\nabla_{R\mathring{C}}\delta\Pi^{\mathring{A}\mathring{B}\mathring{C}\dot{D}}) \\
&\quad = {}^\circ S^{\mathring{A}\dot{B}} - \tfrac{1}{8}{}^\circ S^{\dot{C}\mathring{D}}\,{}^\circ\Box\delta\Pi^{\mathring{A}\mathring{B}}{}_{\mathring{C}\dot{D}} + \tfrac{1}{8}{}^\circ S^{RS}\{{}^\circ\nabla_{R}{}^{\mathring{C}}\,{}^\circ\nabla_{S}{}^{\dot{D}} \\
&\qquad \times\delta\Pi^{\mathring{A}\mathring{B}}{}_{\dot{C}\dot{D}} - 2\,{}^\circ\nabla_{T}{}^{\mathring{A}}\,{}^\circ\nabla_{U}{}^{\mathring{B}}\delta\Pi_{RS}{}^{TU}\}, && (4.22)\\
& {}^\circ D\,{}^\circ g_{A\dot{B}} = 0, \quad {}^\circ D\,{}^\circ\Gamma_{\mathring{A}\mathring{B}} = 0 = {}^\circ D\,{}^\circ\Gamma_{AB}, && (4.23)\\
& \delta\Pi_{\dot{A}\dot{B}\mathring{C}\mathring{D}} = 2\delta\Theta K_{\dot{A}}K_{\dot{B}}K_{\mathring{C}}K_{\mathring{D}} && (4.24)\\
& \delta\Pi_{ABCD} = 2\mathring{\delta}\bar{\Theta}K_A K_B K_C K_D \quad \text{are both of the type N,} && (4.25)\\
& {}^\circ DK_{\dot{A}} = 0 \longleftrightarrow {}^\circ\nabla_{A\dot{B}}K_{\mathring{C}} = 0 && (4.26)\\
& {}^\circ DK_{A} = 0 \longleftrightarrow {}^\circ\nabla_{A\dot{B}}K_{C} = 0 \quad \text{(defining two homogeneous spinors.)} && (4.27)
\end{aligned}\right.$$

Of course, from the point of view of the nonlinear theory, all formulas from this collection hold with an accuracy of the first order in δ. From the point of view of the equations of linearized (complex) gravity, these formulas provide the rigorous and most general solution which is determined by the integral variety of the simple equations:

$$ {}^\circ\Box\delta\Theta = 0, \quad {}^\circ DK_{\dot{A}} = 0, \qquad (4.28a)$$

$$ {}^\circ\Box\delta\bar{\Theta} = 0, \quad {}^\circ DK_{\dot{A}} = 0, \qquad (4.28b)$$

the same equations which determine the integral variety of the Maxwell equations. This is the general solution of Penrose.[8]

The analogy between our $\delta G\otimes\delta G$ formulas and the description of the electromagnetic field (in vacuum) by the D(0,1) and D(1,0) null Hertz potentials is striking. Clearly, $\delta\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}$ and $\delta\Pi_{ABCD}$ play the role of the D(0,2) and D(2,0) Hertz potentials for the equations of the linear approximation, the potentials which admit the null gauge, with the quadruple Penrose spinors proportional to homogeneous spinors.

It should be observed that if we assume (i) that ${}^\circ g_{A\dot{B}}$ is Hermitian and (ii) that $(\delta\Pi_{ABCD})^* = \delta\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}$, then $g_{A\dot{B}}$ from (4.13) is Hermitian; and the (approximate) Einsteinian metric induced by $\delta G\otimes\delta G$ is real. We notice also the interesting fact that all our $\delta G\otimes\delta G$ formulas permit us again to replace in them objects and operators according to the scheme (3.27) in all terms which involve $\delta\Pi_{\dot{A}\dot{B}\dot{C}\dot{D}}$, $\delta\Pi_{ABCD}$ and $K_{\dot{A}}$, K_A. This time this holds trivially because all $\delta G\otimes\delta G$ formulae are valid with precision up to $O(\delta^2)$.

5. FINAL REMARKS

The fundamental question arises concerning how our $\delta G\otimes\delta G$ structure generalizes within the complete nonlinear theory. For the (complex) space–times which are one-sidedly flat, i.e., heavens $[-]\otimes G$ and $G\otimes[-]$ we know the answer: One of (infinitesimal) Hertz potentials of the linear approximation goes to the zero limit, while the second potential becomes finite and fulfills a simple nonlinear equation with quadratic nonlinearity, maintaining from the linear approximation two crucial properties (i) its type N, (ii) the proportionality of the quadruple Penrose spinor to an homogeneous spinor.

In the general case of the solutions of the Einstein equations of the type $G\otimes G$, the present results seem to suggest strongly that it should be possible to describe entirely these solutions in the terms of some two Hertz potentials of the types D(0,2) and D(2,0) respectively. How this should be done can perhaps become more transparent when the basic results of this paper concerning the "spinorization" and "covariantization" of the $\mathcal{H}$-spaces, will be extended to the theory of $\mathcal{HH}$-spaces.

The last spaces, being the solutions of Einstein (empty space) equations of the type Deg.$\otimes$Gen., are entirely described in the terms of one function of four variables and some (gauge dependent) functions of the two variables. (See Refs. 9 and 16 for succinct resumées, and 17 for the complete proofs; Ref. 18 contains a spinorial description of $D\otimes G$ spaces but with the SL gauge completely "frozen" from the D side and partially restricted from the G side; Refs. 19 and 20 contain the generalization of the theory of $\mathcal{HH}$ spaces on the case of Einstein–Maxwell equations and then subsequent spinorial description of the results obtained; Ref. 21 contains comparision of the results of the theory of the type D solutions as stated in Ref. 22 with the theory of $\mathcal{HH}$ spaces). The $\mathcal{HH}$ equation—very similar to the second heavenly equation (3.4)—is the only condition on the function of four variables which determines these spaces. Some work in the direction of the "covariantization" of the description of $\mathcal{HH}$ spaces and the identification in these spaces of the corresponding Hertz potentials is now in progress (jointly with Dr. A. Garcia).

We also believe that some relevant hints concerning the structure of the $G^{\otimes}G$ solutions described by Hertz potentials can be obtained by using $\delta G^{\otimes}\delta G$ structure as the first order approximation in a *covariant* approximation procedure which would permit us to determine all pertinent quantities with the precision up to one order higher [up to $O(\delta^3)$]. Some work in this direction (jointly with Dr. S. Hacyan) is now in progress.

It can be also noticed, that the structure equations with the built-in Einstein equations $C_{AB\dot{C}\dot{D}}=0$, $R=-4\lambda$ can be stated together with Bianchi identities in the form of

$$*R_{AB}=R_{AB}, \quad DR_{AB}=0, \quad \tilde{D}R_{AB}=0, \tag{5.1a}$$

$$*R_{\dot{A}\dot{B}}=-R_{\dot{A}\dot{B}}, \quad DR_{\dot{A}\dot{B}}=0, \quad \tilde{D}R_{\dot{A}\dot{B}}=0, \tag{5.1b}$$

(where $\tilde{D}=-i*D*$ is the covariant codifferential), becoming this way very similar to Eqs. (2.3) for the Maxwellian field, which suggest the usefulness of the electromagnetic Hertz potentials. We also consider at the present time whether Eqs. (5.1) can be directly approached as the starting point in introducing the corresponding gravitational Hertz potentials.

[1]J.M. Cohen and L.S. Kegels, Phys. Rev. D **10**, 1070 (1974), Phys. Lett. A **47**, 261 (1974).
[2]J.F. Plebański, J. Math. Phys. **16**, 2395 (1975).
[3]J.F. Plebański, "Spinors, Tetrads and Forms," a monograph of Centro de Invest. y Est. Avanz, del I.P.N., 1975 (unpublished).
[4]J.D. Finley III and J.F. Plebański, J. Math. Phys. **17**, 585 (1976).
[5]C.P. Boyer and J.F. Plebański, J. Math. Phys. **18**, 1022–31 (1977).
[6]S. Hacyan and J.F. Plebański, J. Math. Phys. **17**, 2203 (1976).
[7]H. Stephani, J. Math. Phys. **15**, 9-14 (1974).
[8]R. Penrose, Proc. Roy. Soc. A **284**, 159 (1965).
[9]J.F. Plebański and I. Robinson, Phys. Rev. Lett. **37**, 493 (1976).
[10]E.T. Newman, General Relativity 7 Conference, Tel Aviv 1974; see also C.W. Fette, Allen I. Janis, and E.T. Newman, J. Math. Phys. **17**, 660 (1976); E.T. Newman, in *Proc of Asymptotic Structure of Space—Time Conference Cincinnati, June* 1976, edited by L. Witten (Plenum, New York, 1977); E.T. Newman, G.R. 8 Conference, Waterloo 1976.
[11]R. Penrose, "Nonlinear Graviton," First Award Gravity Foundation Essay (1975) and Gen. Rel. Grav. **7**, 31 (1976).
[12]R.P. Kerr and A. Schild, Atti del Convengo sulla relatività generale Firenze (1965), p. 173.
[13]G. Debney, R.P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).
[14]J.F. Plebański, Ann. Phys. (N.Y.) **90**, 196 (1975).
[15]J.F. Plebański and A. Schild, Nuovo Cimento B **35**, 35 (1976).
[16]J.F. Plebański and I. Robinson, in *Proc. of* V—e *Colloque Internationalle sur la Theoriê de Groupes en Physique," Montreal, July*, 1976, edited by R. Sharp (Academic, New York, 1977), p. 361.
[17]J.F. Plebański and I. Robinson, *Proc. of Asympt. Struct. of Space-Time, Conference, Cincinnati, June*, 1976, edited by L. Witten (Plenum, New York, 1977), pp. 364—406.
[18]J.D. Finley III and J. Plebański, J. Math. Phys. **17**, 2207 (1976).
[19]A. García, J.F. Plebański, and I. Robinson, Gen. Rel. Grav. **8**, 841 (1977).
[20]J.D. Finley III and J.F. Plebański, J. Math. Phys. **18**, 1662 (1977).
[21]A. García and J.F. Plebański, Nuovo Cimento B **40**, 224 (1977).
[22]J.F. Plebański and M. Demiański, Ann. Phys. (N.Y.) **98**, 98—127 (1976).
[23]J.F. Plebański, J. Math. Phys. **18**, 2511 (1977).

Quantum field theory Potts model

Florin Constantinescu[a] and Herbert M. Ruck[a]

Institute for Applied Mathematics, Johann Wolfgang Goethe University, Frankfurt am Main, West Germany
(Received 20 January 1978)

We consider a quantum field theory analog to the three states Potts model [R. B. Potts, Proc. Camb. Phil. Soc. **48**, 106 (1952)] in two dimensions. Our model can be interpreted as a neutral vector model with discrete gauge symmetry. We prove the existence of the thermodynamic limit by using the lattice approximation and correlation inequalities.

1. INTRODUCTION

There are several extensions of the Ising model. One of them has been given by Potts.[1] In the simplest form (three state Potts model) the spins are two-dimensional vectors of constant length which point in three given directions at 0°, 120°, and 240°. There is nearest neighbor interaction. Parallel spins interact with energy ϵ_0 whereas nonparallel spins interact with energy $\epsilon_1 > \epsilon_0$.

The classical Potts model, ferromagnetic in character, has a long standing tradition in statistical physics.[2] At low temperatures one expects at least three phases. This result seems to be implied by rigorous work of Gertsik and Dobrushin[3] and Pirogov and Sinai.[4]

We construct a Euclidean quantum field theory analog to the classical Potts model. Let $\mathbf{\Phi}(x)$ be a vector field with two components $\Phi_1(x)$, $\Phi_2(x)$. We introduce the Hamiltonian

$$H_0(x) = \tfrac{1}{2}\{[\nabla\mathbf{\Phi}(x)]^2 + m^2\mathbf{\Phi}^2(x)\} \tag{1}$$

and the formal Gaussian measure

$$\exp[-\textstyle\int H_0(x)\,dx] \prod_{x\in\mathbb{R}^2} d\Phi_1(x) \prod_{x\in\mathbb{R}^2} d\Phi_2(x). \tag{2}$$

The interaction is given by $P(\mathbf{\Phi}) + \boldsymbol{\omega}\cdot\mathbf{\Phi}$, where

$$P(\mathbf{\Phi}) = \lambda(\mathbf{\Phi}^2)^2 + \nu(\Phi_1^3 - 3\Phi_1\Phi_2^2) + \mu\mathbf{\Phi}^2 \tag{3}$$

and $\lambda, \nu, \mu, \boldsymbol{\omega}(\omega_1, \omega_2)$ are constants with $\lambda > 0$.

In polar coordinates

$$P(\rho, \theta) = \lambda\rho^4 + \nu\rho^3\cos 3\theta + \mu\rho^2. \tag{4}$$

$P(\mathbf{\Phi})$ has the $\mathbb{Z}_3$ symmetry of the Potts model. If $\nu = 0$, the model has rotational symmetry. The rotational symmetric case was studied by Fröhlich.[5]

Our model ($\nu \neq 0$) can be interpreted as a neutral model with discrete symmetry.

We consider the model defined by Eqs. (2) and (3) in the rigorous frame of constructive quantum field theory of Glimm and Jaffe.[6,7]

First we remark that the $\mathbb{Z}_3$ symmetry of the model is not destroyed by the Wick ordering. Indeed we have

$$:\Phi_1^3(x) - 3\Phi_1(x)\,\Phi_2^2(x): = \Phi_1^3(x) - 3\Phi_1(x)\,\Phi_2^2(x). \tag{5}$$

[a]Supported by the Deutsche Forschungsgemeinschaft.

The free measure Eq. (2) is a product of two Gaussian measures with mean zero and covariance $(-\Delta + m^2)^{-1}$. The cutoff interaction can be defined as usual. In this paper we introduce the lattice approximation[8] of the model and prove certain correlation inequalities (including the second GKS inequality). The existance of the thermodynamic limit for half-Dirichlet and Dirichlet boundary conditions follows by using the lattice approximation and correlation inequalitites.

In the case $m = 1$, $\mu < 0$, $\nu < 0$, $|\nu| \gg 1$, and $|\lambda/\mu| \ll 1$ the function $P(\Phi) + \frac{1}{2}m^2\Phi^2$ has three deep minima separated by high barriers. We expect that the model shows in this case symmetry breaking and at least three phases.

2. CORRELATION INEQUALITIES

We consider a family of random two-dimensional (spin) vectors $\mathbf{s}_j(s_j^2, s_j^2)$, $j = 1, 2, \ldots, N$, with joint probability distribution

$$\frac{1}{Z}\exp\left[\sum_{j=1}^N \mathbf{a}_j\cdot\mathbf{s}_j + \sum_{j=1}^N b_j((s_j^1)^3 - 3s_j^1(s_j^2)^2) + \sum_{j,k=1}^N \sum_{i=1}^2 J_{jk}^i s_j^i s_k^i\right] \prod_{j=1}^N d\rho_j(\mathbf{s}_j), \tag{6}$$

where

$$Z = \int_{\mathbb{R}^{2N}} \exp\left[\sum_{j=1}^N \mathbf{a}_j\cdot\mathbf{s}_j + \sum_{j=1}^N b_j((s_j^1)^3 - 3s_j^1(s_j^2)^2) + \sum_{j,k=1}^N \sum_{i=1}^2 J_{jk}^i s_j^i s_k^i\right] \prod_{j=1}^N d\rho_j(\mathbf{s}_j), \tag{7}$$

$\mathbf{a}_j(a_j^1, a_j^2)$, b_j, J_{jk}^i are real constants, and each $\rho_j(\mathbf{s})$ is a positive measure on $\mathbb{R}^2$ such that $\int_{\mathbb{R}^2}\exp(c|\mathbf{s}|^2)\,d\rho_j(\mathbf{s}) < \infty$ for all $c \in \mathbb{R}$. In the lattice approximation of the Euclidean quantum field Potts model with external field $J_{jk}^i \equiv J_{jk}$ (independent of i), $J_{jk} \geq 0$ for $j \neq k$ and

$$d\rho_j(\mathbf{s}) = \exp[-Q_j(\mathbf{s})]\,d\mathbf{s}, \tag{8}$$

where $Q_j(\mathbf{s})$ collects the terms $(\mathbf{\Phi}^2)^2$ and $\mathbf{\Phi}^2$. Further, we can assume $J_{jj} = 0$ by absorbing $J_{jj}\mathbf{s}_j^2$ into Q_j.

Let $\mathcal{L}$ be the family of functions[9] on $(\mathbb{R}^2)^N$ which (in polar coordinates) are multinomials with nonnegative coefficients in $\cos(\mathbf{m}\cdot\boldsymbol{\theta}) \equiv \cos(m_1\theta_1 + \cdots + m_N\theta_N)$ and in $\prod_{j=1}^N h_j(r_j)$, where each $h_j(r_j)$ is nonnegative, nondecreasing on $[0, \infty)$ and $O(\exp(cr^2))$ for some $c > 0$.

Theorem 1: Suppose $\mathbf{s}_j$, $j = 1, \ldots, N$ are random two-

 0022-2488/78/1911-2359$1.00

dimensional vectors whose joint probability distribution is given by (6) and (7) with the measure $\rho_j(\mathbf{s})$ being spherically symmetric for all $j=1,\ldots,N$.

If $\mathbf{a}_j=(a_j,0)\geq 0$, $b_j\geq 0$, $|J^2_{jk}|\leq J^1_{jk}$ $(j\neq k)$, and $J^2_{jj}<J^1_{jj}$ for all j,k, then for any $F,G\in\mathcal{L}$

$$\langle F(\mathbf{s}_1,\ldots,\mathbf{s}_N)\rangle\geq 0 \quad \text{(first GKS inequality)}, \tag{9a}$$

$$\langle F(\mathbf{s}_1,\ldots,\mathbf{s}_N)\,G(\mathbf{s}_1,\ldots,\mathbf{s}_N)\rangle \geq \langle F(\mathbf{s}_1,\ldots,\mathbf{s}_N)\rangle\langle G(\mathbf{s}_1,\ldots,\mathbf{s}_N)\rangle \quad \text{(second GKS inequality)}. \tag{9b}$$

Proof. Let $\mathcal{C}(K)$ be the algebra of real continuous functions on a compact set K with supremum norm. Let σ be a probability measure on K. We say that $S\subset\mathcal{C}(K)$ satisfies condition (Q3) (see Ref. 9, p. 311) if, for any finite family $f_1,\ldots,f_n$ of elements in S,

$$\int d\sigma(x)f_1(x)\cdots f_n(x)\geq 0.$$

Let $\langle f\rangle_h=Z_h^{-1}\int d\sigma(x)f(x)\exp[-h(x)]$ and $Z_h=\int d\sigma(x)\times\exp[-h(x)]$.

We denote by $Q(S)$ the norm closure of the set of polynomials of elements of S and the identity $\mathbb{1}$ of $\mathcal{C}(K)$ with positive coefficients. Ginibre proves[9] that if S satisfies (Q3) and if $-h$, f, $g\in Q(S)$, then

$\langle f\rangle_h\geq 0$, first GKS inequality,

$\langle fg\rangle_h-\langle f\rangle_h\langle g\rangle_h\geq 0$, second GKS inequality.

Let $K=K_1\times K_2$ and $\sigma=\sigma_1\times\sigma_2$ be the product of two compact sets K_1 and K_2 and the corresponding probability measures. Let $S_1\subset\mathcal{C}(K_1)$, $S_2\subset\mathcal{C}(K_2)$ and let $S=S_1S_2\subset\mathcal{C}(K_1\times K_2)$ be the set of functions of the form $f(x_1,x_2)=f(x_1)f(x_2)$, where $f(x_1)\in S_1$, $f(x_2)\in S_2$. Then if S_1,S_2 both satisfy (Q3), the S also satisfy (Q3).[9]

The condition on K to be compact can be weakened in the case $K\subset\mathbb{R}^N$. The above results remain true if one replaces $\mathcal{C}(K)$ by the algebra of continuous functions of certain growth at infinity with the adequate norm. This remark is implicitly contained in Ref. 10.

Now we can take S_1 to be the set of functions $\cos(m_1\theta_1+\cdots+m_N\theta_N)$, where $m_1,\ldots,m_N$ are integers and S_2 the set of functions of the form $\prod_{j=1}^N h_j(r_j)$, where each $h_j(r)$ is nonnegative and nondecreasing on $[0,\infty)$ and $O(\exp(br^2))$ for some $b>0$. Both S_1 and S_2 satisfy (Q3) as proved in (Ref. 9 and Ref. 10, Lemmas 1 and 2, p. 232). Then $S=S_1S_2$ also satisfy (Q3).

The proof follows now by remarking that the functions

$$\mathbf{a}_j\cdot\mathbf{s}_j=a^1_jr_j\cos\theta_j,\quad a^1_j>0,$$

$$b_j((s^1_j)^3-3s^1_j(s^2_j)^2)=b_jr^3_j\cos3\theta_j,\quad b_j>0,$$

in (6) belong to S.

3. INFINITE VOLUME LIMIT

In this section we prove the existence of the infinite volume limit for half-Dirichlet and Dirichlet states for our model with external field. We will start by considering the interaction

$$P(\mathbf{\Phi})+\boldsymbol{\omega}\cdot\mathbf{\Phi} \tag{10}$$

with $\nu\leq 0$ and $\boldsymbol{\omega}\equiv(\omega_1,0)\leq 0$. For this case we have proved the GKS correlation inequalities in the lattice approximation. We remark that both half-Dirichlet and Dirichlet states have $\mathbb{Z}_3$ symmetry.

Before going to discuss the infinite volume limit we remark that Theorem 1 gives correlation inequalities for expectations of the form

$$\langle \textstyle\prod_i (s^1_i)^{n^1_i}\prod_{j,k}(\mathbf{s}_j\cdot\mathbf{s}_k)^{m_{jk}}\rangle. \tag{11}$$

Now consider the general expectation

$$\langle \prod_i (s^1_i)^{n^1_i}\prod_j (s^2_j)^{m^2_j}\rangle. \tag{12}$$

In (12) we pair $s^2_{j1}s^2_{j2}=\frac{1}{2}r_{j1}r_{j2}[\cos(\theta_{j1}-\theta_{j2})-\cos(\theta_{j1}+\theta_{j2})]$ and stay with cosine terms and possible $(\sin\theta)^k$, $k>0$, which cannot be paired any more. We write $\sin^2\theta=1-\cos^2\theta$ and we stay possible with $\sin\theta$. But

$$\langle \prod_i (s^1_i)^{n^1_i}s^2_j\rangle=0 \tag{13}$$

because of the symmetry $\rho_j(s^1,s^2)=\rho_j(s^1,-s^2)$ of the measure. It follows that a general expectation (12) can always be written as a linear combination of expectations (11). Now the Schwinger functions of our model (10) in the lattice approximation are linear combinations of (12) and therefore linear combinations of (11).

We can now state the following result:

Theorem 2: The quantum field Potts model in two dimensions, with half-Dirichlet (or Dirichlet) boundary conditions, has a unique thermodynamic limit.

Proof: Consider the case $\nu\leq 0$, $\omega=0$. The infinite volume (thermodynamic) limit is considered in the standard way. By lattice approximation and correlation inequalities we prove that the Schwinger functions

$$\langle \prod_n \Phi_1(g_n)\prod_{m,k}(\mathbf{\Phi}(g_m)\cdot\mathbf{\Phi}(g_k))\rangle \tag{14}$$

are nondecreasing in the regular cutoff and bounded above. This implies convergence. The rest of the Schwinger functions are linear combinations of (14) by the discussion above. The case $\nu\geq 0$, $\boldsymbol{\omega}=0$ is physically equivalent to $\nu\leq 0$ because the transformations $\nu\to-\nu$ represents a 180° rotation of the cordinate axes in the spin space.

The theorem is also true in the presence of an external field in the direction of a minimum of the potential $P(\mathbf{\Phi})$ which lowers this minimum.

Remarks: (1) Dirichlet boundary conditions are also allowed because of (5).

(2) The existence of correlation inequalities and the existence of the thermodynamic limit implied by them is a consequence of the model being "ferromagnetic" as can be seen from geometric considerations.

4. DISCUSSION AND PERSPECTIVES

We have proved GKS inequalities and the existence of the thermodynamic limit for the quantum field Potts model. The cluster expansion[6] in the "one phase region" can be mimicked giving an alternative existence theorem. A very similar cluster expansion is used in a recent work of Osterwalder and Seiler[11] on gauge field theories on the lattice.

Because of the $\mathbb{Z}_3$ symmetry the problem of phase transitions is expected to be simpler than in the usual $P(\Phi)_2$ model with at least sixth order polynomial interaction.[12,13]

The model is a good candidate for studying continuous systems with more than two phases and related problems as, for instance, the Gibbs phase rule.

The results of this paper can be easily generalized to the case of a $\mathbb{Z}_n$ symmetry group. The phase transition problem for certain values of the interaction parameters will be discussed in a subsequent paper.

ACKNOWLEDGMENTS

We are grateful to Professor A. Jaffe, Professor K. Osterwalder, Professor D. Ruelle, and Professor R. Weder for stimulating discussions.

[1]R.B. Potts, Proc. Camb. Phil. Soc. **48**, 106 (1952).

[2]J.P. Straley and M.E. Fisher, J. Phys. A **6**, 1310 (1973); J.P. Straley, J. Phys. A **7**, 2173 (1974); recent experimental papers: D. Mukamel, M.E. Fisher and E. Domany, Phys. Rev. Lett. **37**, 565 (1976); A. Aharony, K.A. Müller, and W. Berlinger, Phys. Rev. Lett. **38**, 33 (1977); M. Bretz, Phys. Rev. Lett. **38**, 501 (1977); E. Domany, M. Schick, and J.S. Walker, Phys. Rev. Lett. **38**, 1148 (1977).

[3]V.M. Gertsik and R.L. Dobrushin, Funk. Anal. **3**, 12 (1974).

[4]S.A. Pirogov and Ya.G. Sinai, Teor. Mat. Fis. **25**, 358 (1975); **26**, 61 (1976).

[5]J. Fröhlich, "Poetic phenomena in (two dimensional) quantum field theory: non-uniqueness of the vacuum, the solitons and all that," in *Les méthodes mathématique de la théorie quantique des champes, Marseille* 1975 (Editions du Centre National de la Recherche Scientifique, Paris, 1976).

[6]J. Glimm, A. Jaffe, and T. Spencer, "The particle structure of the weakly coupled $P(\Phi)_2$ model and other applications of the high temperature expansion, Parts I and II," in *Constructive Quantum Field Theory* Lecture Notes in Physics **25**, edited by G. Velo and A.S. Wightman (Springer-Verlag, Berlin, 1973).

[7]B. Simon, *The $P(\Phi)_2$ Euclidean (Quantum) Field Theory* (Princeton U.P., Princeton, N.J., 1974).

[8]F. Guerra, L. Rosen, and B. Simon, Ann. Math. **101**, 111 (1975).

[9]J. Ginibre, Comm. Math. Phys. **16**, 310 (1970).

[10]F. Dunlop and C.M. Newman, Commun. Math. Phys. **44**, 223 (1975).

[11]K. Osterwalder and E. Seiler, "Gauge field theories on the lattice," Harvard University preprint (1977).

[12]J. Glimm, A. Jaffe, and T. Spencer, Ann. Phys. (N.Y.) **101**, 610, 631 (1976).

[13]K. Gawedzki, "Existence of three phases for a $P(\Phi)_2$ model of quantum field," Warsaw University preprint (1977).

Polynomial irreducible tensors for point groups

J. Patera

Centre de Recherches Mathématiques, Université de Montréal, Montréal, Québec, Canada

R. T. Sharp

Physics Department, McGill University, Montreal, Quebec, Canada

P. Winternitz

Centre de Recherches Mathématiques, Université de Montréal, Montréal, Québec, Canada
(Received 8 March 1978)

Generating functions for (Γ_r,Γ_m) tensors for any point group G and any pair of its irreducible representations Γ_r and Γ_m are calculated explicitly. A (Γ_r,Γ_m) tensor transforms according to Γ_r, and its components are polynomials in another tensor transforming by Γ_m. Explicit integrity bases of (Γ_r,Γ_m) tensors are given for all pairs Γ_r and Γ_m for the groups C_n, D_n, T, and O, and for the same groups with reflections. A composition rule for extending the result to reducible representations is formulated. Point group tensors irreducible with respect to SO(3) are obtained, together with their generating functions.

I. INTRODUCTION

Whenever a mathematical description of a physical system like a molecule or a crystal with a nontrivial point or space group symmetry is attempted, it becomes necessary to decompose various quantities in terms of point group harmonics (PGH) corresponding to that particular group.[1] It is therefore important to know the PGH. So far only PGH of lower degrees are known explicitly. The present paper is devoted to the description of PGH of any degree.

Let us recall that PGH may be defined, for instance, as a system of homogeneous polynomials in Cartesian coordinates x, y, and z such that they span a space in which acts an irreducible representation Γ_r of a point group G. For the purpose of this article we somewhat generalize this definition. Thus by PGH we understand polynomials in coordinates $\alpha, \beta, \cdots$ of a general representation space R_m of a representation Γ_m of G. Therefore PGH in this extended sense are defined by a pair of representations, Γ_r and Γ_m, of G. For simplicity we refer to PGH as (Γ_r,Γ_m) hamonics or (Γ_r,Γ_m) tensors. In the present paper we shall not be concerned with their normalization and/or orthogonality.

Our problem thus is to describe effectively the (Γ_r,Γ_m) harmonics for any given point group G and for any pair Γ_r and Γ_m of its representations.

In a physical context the first particular harmonics were found by Bethe,[2] and von der Lage and Bethe.[3] Subsequently many authors devoted their attention to other particular cases of point or space group harmonics as well as to some aspects of the general problem (e.g., Refs. 4–6 and further references therein).

An exhaustive description of (Γ_r,Γ_m) tensors of an arbitrary degree exists only for the particular case when $\Gamma_m=\Gamma_v$, where Γ_v is the representation of G in the three-dimensional space spanned by x, y, z, and $\Gamma_r=\Gamma_1$ is the identity representative of G.[5] The (Γ_1,Γ_v) tensors are called invariants or scalars with respect to G.

Recently a step toward an efficient solution of the problem in all its generality was made by McLellan[6] who used a general method to solve a particular problem of integrity bases for polynomial functions of a symmetric second-order tensor which are invariant with respect to crystal point groups. The method described by McLellan is based on Burnside's[7] generalization of an earlier result of Molien.[8] It proceeds in two steps. First a generating function has to be found for each pair of representations Γ_r and Γ_m of every point group G. That allows one to determine how many copies of a representation Γ_r appear in the symmetrized tensor product $\{\Gamma_m\}^n$. The second step is then an explicit construction of homogeneous polynomials—components of (Γ_r,Γ_m) tensors—which form the so-called integrity basis for general (Γ_r,Γ_m) tensors.

The purpose of this article is to apply systematically the above procedure to all point groups and to their representations. Namely, we find and list all generating functions for the point groups $G=C_n, D_n, T, O, I$, and also $G\times P$, where P is the group of reflections. Then we construct an integrity basis for (Γ_r,Γ_m) tensors of all point groups but the icosahedral one. The latter group is of the least interest in solid state and molecular physics. (There are no crystals and few molecules with that symmetry). Practically, we solve the problem for all irreducible representations Γ_r, Γ_m and then supplement the procedure by composition rules which enable one to get the generating functions for reducible Γ_r and Γ_m from the irreducible ones.

In Sec. II we reproduce general properties of the generating functions. In Sec. III actual generating functions are calculated for all pairs Γ_r, Γ_m of irreducible representations of all point groups (a generalization to reducible representations Γ_r and Γ_m is given). Section IV contains the integrity bases in an explicit form for (Γ_r,Γ_m) tensors, for the point groups C_n, D_n, T, O. In Sec. V we give explicit integrity bases for the physically important case, where Γ_m is the represen-

[a)]Work supported in part by the National Research Council of Canada and by the Ministére de l'Education du Québec.

tation of C_n, D_n, T, and O which acts on the Cartesian coordinates (x,y,z). We treat similarly groups of the types $H[G$ and $G\times P$, which involve reflections. In Sec. VI we find the generating functions and polynomial bases for irreducible representations of SO(3) reduced to C_n, D_n, O, and T. Section VII contains some concluding remarks where, in particular, the relation of the present results to their analogs for continuous groups is pointed out.

II. PROPERTIES OF GENERATING FUNCTIONS

In this section, following Burnside,[7] we introduce first the generating function $B(\Gamma_r,\Gamma_m;\lambda)$ for tensors whose components transform irreducibly by a representation Γ_r of a point group G, and are polynomials in the components of another given tensor which transforms irreducibly by a representation Γ_m of G. Since we want to find such generating functions for all pairs of representations Γ_r, Γ_m of all point groups G, we first solve the problem for irreducible representations Γ_r and Γ_m and then formulate a composition rule which allows one to find $B(\Gamma_r,\Gamma_m\oplus\Gamma_{m'};\lambda,\lambda')$ from $B(\Gamma_r,\Gamma_m;\lambda)$ and $B(\Gamma_r,\Gamma_{m'};\lambda')$.

Let A_s be the matrix, in the representation Γ_m, of an element of G belonging to the class s. Consider the expression

$$\frac{1}{\det(1-\lambda A_s)}=\frac{1}{\Pi_i(1-\lambda\phi_{si})}$$

$$=\sum_{n=0}^{\infty}\lambda^n P_n(\phi_{si}); \qquad (1)$$

ϕ_{si} are the eigenvalues of the matrix A_s and $P_n(\phi_{si})$ is the sum of all products of powers of ϕ_{si} whose total degree is n. Thus $P_n(\phi_{si})$ is the character of the class s for the representation $\{\Gamma_m\}^n$ which is the symmetric part of the direct product $\Gamma_m\otimes\Gamma_m\otimes\cdots\otimes\Gamma_m$ of n copies of Γ_m.

A general expression for the generating function $B(\Gamma_r,\Gamma_m;\lambda)$ is an immediate consequence of (1) and the orthogonality of characters. One has

$$B(\Gamma_r,\Gamma_m;\lambda)=\frac{1}{N}\sum_s\frac{N_s\chi^*_{sr}}{\det(1-\lambda A_s)}, \qquad (2)$$

where N is the order of the group G, N_s is the order of the class s of G, and χ^*_{sr} is the complex conjugate of the character χ_{sr} of s in Γ_r. Obviously $\Sigma_s N_s=N$. Substitution of (1) into (2) transforms $B(\Gamma_r,\Gamma_m,\lambda)$ into a power series in λ,

$$B(\Gamma_r,\Gamma_m;\lambda)=\sum_{n=0}^{\infty}c_n\lambda^n. \qquad (3)$$

The tensors of type Γ_r of degree n in the components of a tensor of type Γ_m [for simplicity (Γ_r,Γ_m) tensors of nth degree] are "counted" by the coefficients c_n. Indeed, c_n is equal to the number of linearly independent (Γ_r,Γ_m) tensors of degree n. That is to say, a symmetrized direct product $\{\Gamma_m\}^n$ of Γ_m's contains the irreducible representation Γ_r exactly c_n times.

The generating function (2) contains more information than just the multiplicities of (Γ_r,Γ_m) tensors of given degree, or, the multiplicities of representation Γ_r in the decomposition of $\{\Gamma_m\}^n$. To see that, the finite summation in (2) has to be performed and the result expressed as a ratio of two polynomials in λ,

$$B(\Gamma_r,\Gamma_m;\lambda)=\frac{N_{rm}(\lambda)}{D_m(\lambda)}. \qquad (4)$$

It turns out that

$$N_{rm}(\lambda)=\sum_p k_p\lambda^p, \qquad D_m(\lambda)=\prod_q(1-\lambda^q), \qquad (5)$$

where the coefficients k_p are *positive integers* and the summation and multiplication variables run through finite sets of integers $p_1,p_2,\cdots$, and $q_1,q_2,\cdots$, respectively. For point groups the q's are distinct, except for the icosahedral group where q may take the value 3 twice. The denominator $D_m(\lambda)$ is common to all generating functions with the same Γ_m and G.

An integrity basis for invariants, i.e., (Γ_1,Γ_m) tensors (through this article Γ_1 stands for the identity representation of the corresponding point group G), consists of two types of basis tensors, $I^{(q)}(\Gamma_1,\Gamma_m)$ and $E^{(p)}(\Gamma_1,\Gamma_m)$; the superscript indicates the degree of I and E whenever it is useful. To each factor $(1-\lambda^q)$ of $D_m(\lambda)$ corresponds one invariant $I^{(q)}(\Gamma_1,\Gamma_m)$. All these tensors satisfy no polynomial relation. To each term $k_p\lambda^p$ of $N_{1m}(\lambda)$ correspond k_p linearly independent tensors $E^{(p)}(\Gamma_1,\Gamma_m)$ of degree p. Obviously, the trivial invariant $E^{(0)}(\Gamma_1,\Gamma_m)=1$ is always present in an integrity basis. Therefore, every generating function must contain a term equal to 1 in its numerator $N_{1m}(\lambda)$. A general invariant, (Γ_1,Γ_m) tensor, can be written as

$$P_0\{I^{(q)}(\Gamma_1,\Gamma_m)\}+\sum E^{(p)}(\Gamma_1,\Gamma_m)P[I^{(q)}(\Gamma_1,\Gamma_m)], \qquad (6)$$

where the summation extends over all $E^{(p)}(\Gamma_1,\Gamma_m)$'s; P_0 and P denote some polynomials in $I^{(q)}$'s.

An integrity basis for (Γ_r,Γ_m) tensors, $\Gamma_r\neq\Gamma_1$, consists of basis tensors $E^{(p)}(\Gamma_r,\Gamma_m)$ only. As before, each term of $N_{rm}(\lambda)$ implies the existence of k_p linearly independent $E^{(p)}(\Gamma_r,\Gamma_m)$'s of degree p. A general (Γ_r,Γ_m) tensor can be brought to the form.

$$\sum E^{(p)}(\Gamma_r,\Gamma_m)P(I^{(q)}(\Gamma_1,\Gamma_m)), \qquad (7)$$

where the summation runs through all the $E^{(p)}(\Gamma_r,\Gamma_m)$'s, and P stands for a polynomial in $I^{(q)}$'s.

Let us mention a dimensionality relation satisfied by the generating functions $B(\Gamma_r,\Gamma_m;\lambda)$, which provides a useful check. Suppose the dimension of Γ_a is f_a. Then

$$\sum_r f_r B(\Gamma_r,\Gamma_m;\lambda)=(1-\lambda)^{-f_m}.$$

The coefficient of λ^n in the expansion of $(1-\lambda)^{-f_m}$ equals $(f_m+n-1)!/n!(f_m-1)!$, the number of independent polynomials of degree n in f_m variables.

We conclude this section by indicating how to combine generating functions corresponding to irreducible representations Γ_r and Γ_m in order to obtain those corresponding to reducible representations. One has immediately

$$B(\Gamma_{r_1}\oplus\Gamma_{r_2},\Gamma_m;\lambda_1,\lambda_2) = B(\Gamma_{r_1},\Gamma_m;\lambda_1)\cdot B(\Gamma_{r_2},\Gamma_m;\lambda_2). \quad (8)$$

From the Clebsch–Gordan series

$$\Gamma_a\otimes\Gamma_b=\oplus_c \Gamma_c C_{ab}^c \quad (9)$$

one concludes that

$$B(\Gamma_r,\Gamma_{m_1}\oplus\Gamma_{m_2};\lambda_1,\lambda_2) = \sum_{r_1,r_2} B(\Gamma_{r_1},\Gamma_{m_1};\lambda_1)B(\Gamma_{r_2},\Gamma_{m_2};\lambda_2)\, C_{r_1r_2}^r \quad (10)$$

is the generating function for $(\Gamma_r,\Gamma_{m_1}\oplus\Gamma_{m_2})$ tensors. The coefficient of $\lambda_1^{n_1}\lambda_2^{n_2}$ in the expansion of $B(\Gamma_r,\Gamma_{m_1}\oplus\Gamma_{m_2};\lambda_1,\lambda_2)$ is the number of linearly independent $(\Gamma_r,\Gamma_{m_1}\oplus\Gamma_{m_2})$ tensors of degree n_1 and n_2 respectively in components of tensors transforming by Γ_{m_1} and Γ_{m_2}. For examples of (10) see Sec. V.

In case $\Gamma_r=\Gamma_1$ is the identity representation, Eq. (10) gives the Molien function

$$B(\Gamma_1,\Gamma_{m_1}\oplus\Gamma_{m_2};\lambda_1,\lambda_2) = \sum_r B(\Gamma_r,\Gamma_{m_1};\lambda_1)B(\Gamma_r^*,\Gamma_{m_2};\lambda_2) \quad (11)$$

for invariants (scalars). By Γ_r^* we denote the representation complex conjugate to Γ_r.

III. GENERATING FUNCTIONS FOR IRREDUCIBLE REPRESENTATIONS OF POINT GROUPS

In this section we present the results of our computation of generating functions $B(\Gamma_r,\Gamma_m;\lambda)$ for all pairs of irreducible representations Γ_r and Γ_m of all finite subgroups of O(3). First we obtain the results for the groups C_n (nth order axis), D_n (nth order vertical axis and n second order horizontal axes), T (tetrahedral), O (octahedral), I (icosahedral). Subsequently, we extend the results to the groups $C_n\times P$, $D_n\times P$, $T\times P$, $O\times P$, *and* $I\times P$, where P is the group of reflections. $H[G$ is isomorphic to G and hence need not be considered separately for the purpose of calculating generating functions or integrity bases for irreducible representations.

Our task is to obtain each generating function $B(\Gamma_r,\Gamma_m;\lambda)$ in the form (4) starting from (2). For that it is necessary to perform explicitly the summation in (2). The vanishing of $B(\Gamma_r,\Gamma_m;\lambda)$ implies that there are no (Γ_r,Γ_m) tensors; we list only nonzero generating functions below.

A. The group C_n

The group C_n of rotations about an nth order axis has n classes and n irreducible representations each of dimension one. The polynomial det $(1-\lambda A_s)$, required in (2), corresponding to the sth class and mth irreducible representation is

$$\det(1-\lambda A_s)=1-\lambda\exp\{2\pi i(m-1)(s-1)/n\}. \quad (12)$$

To describe the generating function $B(\Gamma_r,\Gamma_m;\lambda)$ we introduce an integer

$$a=a(m)=n/\mathrm{HCF}(n,m-1) \quad (13)$$

where the denominator is the highest common factor (HCF) of n and $m-1$. Then for each

$$r=p(m-1)_{\mathrm{mod}\,n}+1 \quad (p=0,1,\ldots,a-1) \quad (14)$$

the generation function for (Γ_r,Γ_m) tensors is

$$B(\Gamma_r,\Gamma_m;\lambda)=\frac{\lambda^p}{1-\lambda^a}. \quad (15)$$

If r is not of the form (14) there is no (Γ_r,Γ_m) tensor. In deriving (15) from (2) we made use of the summation formula

$$\frac{1}{n}\sum_{s=0}^{n-1}\sum_{q=0}^{\infty}\lambda^q\exp[2\pi is(qm-q-r+1)/n]=\frac{\lambda^p}{1-\lambda^a} \quad (16)$$

which holds for r given by (14); otherwise the sum (16) vanishes; a in (15) is given by (13).

B. The group D_n

It is convenient to consider separately the cases of n even and n odd.

The group D_n, n odd, has $(n+3)/2$ irreducible representations, the first two Γ_1 and Γ_2 of dimension 1, the remaining $(n-1)/2$ of dimension 2. The polynomials $\det(1-\lambda A_s)$ are shown in Table I. Substituting that into (2), one gets the generating functions

$$B(\Gamma_1,\Gamma_1;\lambda)=\frac{1}{1-\lambda}, \quad (17)$$

$$B(\Gamma_1,\Gamma_2;\lambda)=\frac{1}{1-\lambda^2}, \quad (18)$$

$$B(\Gamma_2,\Gamma_2;\lambda)=\frac{\lambda}{1-\lambda^2}. \quad (19)$$

For the remaining generating functions we first define two numbers a and p related to m and r, respectively,

$$a=n/\mathrm{HCF}(n,m-2), \quad (20)$$

where HCF(x,y) is the highest common factor of x and y and

$$r=2+n/2-|p(m-2)_{\mathrm{mod}\,n}-n/2|, \quad 1\leqslant p\leqslant(a-1)/2. \quad (21)$$

Then

$$B(\Gamma_1,\Gamma_m;\lambda)=\frac{1}{(1-\lambda^2)(1-\lambda^a)}, \quad (22)$$

$$B(\Gamma_{r_2},\Gamma_m;\lambda)=\frac{\lambda^a}{(1-\lambda^2)(1-\lambda^a)}, \quad m=3,4,\ldots,(n+3)/2, \quad (23)$$

$$B(\Gamma_r,\Gamma_m;\lambda)=\frac{\lambda^p+\lambda^{a-p}}{(1-\lambda^2)(1-\lambda^a)}, \quad p=1,2,\ldots,a-1. \quad (24)$$

The group D_n, n even, has $n/2+3$ irreducible representations, the first four being of dimension 1, the other $n/2-1$

TABLE I. The group D_n, n odd. Factors N_s and $\det(1-\lambda A_s)$ required in (2). The characters are equal to coefficients of λ taken with opposite sign.

N_s	1	2	n
Class / Representation	C_0	C_s $(s=1,2,\ldots,\frac{n-1}{2})$	A_1
Γ_1	$1-\lambda$	$1-\lambda$	$1-\lambda$
Γ_2	$1-\lambda$	$1-\lambda$	$1+\lambda$
Γ_m $(m=3,4,\ldots,\frac{n+3}{2})$	$(1-\lambda)^2$	$1-2\lambda\cos[2\pi s(m-2)/n]+\lambda^2$	$1-\lambda^2$

of dimension 2. The polynomials $\det(1-\lambda A_s)$ are again summarized in Table II. Substituting them into (2), one gets the generating functions

$$B(\Gamma_1,\Gamma_1;\lambda)=\frac{1}{1-\lambda},, \tag{25}$$

$$B(\Gamma_1,\Gamma_m;\lambda)=\frac{1}{1-\lambda^2},\qquad m=2,3,4, \tag{26}$$

$$B(\Gamma_m,\Gamma_m;\lambda)=\frac{\lambda}{1-\lambda^2}. \tag{27}$$

For $5<m<n/2+3$ we define a and p by

$$a=n/\mathrm{HCF}(n,m-4), \tag{28}$$

$$r=4+n/2-|p(m-4)-n/2|,$$

$$\begin{aligned}&1\leqslant p\leqslant(a-1)/2, && a \text{ odd},\\ &1\leqslant p\leqslant a/2-1, && a \text{ even}.\end{aligned} \tag{29}$$

Then

$$B(\Gamma_1,\Gamma_m;\lambda)=\frac{1}{(1-\lambda^2)(1-\lambda^a)}, \tag{30}$$

$$B(\Gamma_2,\Gamma_m;\lambda)=\frac{\lambda^a}{(1-\lambda^2)(1-\lambda^a)}, \tag{31}$$

$$B(\Gamma_3,\Gamma_m;\lambda)=B(\Gamma_4;\Gamma_m,\lambda)=\begin{cases}0, & a \text{ odd}\\ \dfrac{\lambda^{a/2}}{(1-\lambda^2)(1-\lambda^a)}, & a \text{ even},\end{cases} \tag{32}$$

TABLE II. The group D_n, n even. Factors N_s and $\det(1-\lambda A_s)$ required in (2). The characters are equal to coefficients of λ with opposite sign.

N_s	1	2	1	$n/2$	$n/2$
Class / Representation	Co	C_s $(s=1,2,\ldots,n/2-1)$	$C_{n/2}$	A_1	A_2
Γ_1	$1-\lambda$	$1-\lambda$	$1-\lambda$	$1-\lambda$	$1-\lambda$
Γ_2	$1-\lambda$	$1-\lambda$	$1-\lambda$	$1+\lambda$	$1+\lambda$
Γ_3	$1-\lambda$	$1-(-1)^s\lambda$	$1-(-1)^{n/2}\lambda$	$1-\lambda$	$1+\lambda$
Γ_4	$1-\lambda$	$1-(-1)^s\lambda$	$1-(-1)^{n/2}\lambda$	$1+\lambda$	$1-\lambda$
Γ_m $(m=5,6,\ldots,n/2+3)$	$(1-\lambda)^2$	$1-2\lambda\cos[2\pi(m-4)s/n]+\lambda^2$	$1-2(-1)^m\lambda+\lambda^2$	$1-\lambda^2$	$1-\lambda^2$

TABLE III. Tetrahedral group T. Table of factors det $(1-\lambda A_s)$; $\omega=\exp(i2\pi/3)$.

N_s	1	3	4	4
Class / Representation	E	C_2	C_3	C_3^2
Γ_1	$1-\lambda$	$1-\lambda$	$1-\lambda$	$1-\lambda$
Γ_2	$1-\lambda$	$1-\lambda$	$1-\omega\lambda$	$1-\omega^2\lambda$
Γ_3	$1-\lambda$	$1-\lambda$	$1-\omega^2\lambda$	$1-\omega\lambda$
Γ_4	$(1-\lambda)^3$	$(1+\lambda)(1-\lambda^2)$	$1-\lambda^3$	$1-\lambda^3$

$$B(\Gamma_r,\Gamma_m;\lambda)=(\lambda^p+\lambda^{a-p})/(1-\lambda^2)(1-\lambda^a). \quad (33)$$

C. The groups T, O, and I

The summation (2) becomes a straightforward simplification of polynomial rational expressions as soon as one substitutes for N_s and χ_{rs} the corresponding values and for det $(1-\lambda A_s)$ the corresponding polynomial in λ. For the tetrahedral, octahedral and icosahedral groups N_s and det $(1-\lambda A_s)$ are summarized in Tables III, IV, and V. Also the characters χ_{rs} for each irreducible representation are contained in these tables because they are the coefficients of the linear term in det$(1-\lambda A_s)$ taken with opposite sign.

The generating function in the form (4) is given if the exponents q_1, q_2 of the denominator polynomial $D(\lambda)$ are specified, and if for every irreducible representation Γ_r one has the exponents $p_1, p_2,\cdots$ and the coefficients k_{pr} of the numerator polynomial $N_r(\lambda)$. For the three groups under consideration these quantities are found in Tables VI, VII, and VIII. In particular, the intersection of a row Γ_m and the last column contains the exponents $q_1,q_2,\cdots$ of the denominator $D(\lambda)$. One notices repetition of the power 3 for the representation Γ_5 of the icosahedral group (Table VIII). The exponents and coefficients of a numerator $N_r(\lambda)$ of $B(\Gamma_r,\Gamma_m;\lambda)$ are found at the intersection of the column Γ_r and the row Γ_m. More precisely, an entry a^b indicates the presence of the term $b\lambda^a$ in the numerator $N_r(\lambda)$.

D. Groups with reflections

Consider now the group $G\times P$, where P is the reflection group, and G is any finite rotation group. Then to each element α of G correspond two elements α and α' of $G\times P$, where α' is α multiplied by the reflection. To each irreducible representation Γ_r of G correspond two irreducible representations Γ_r^e and Γ_r^o of $G\times P$, called even and odd respec-

TABLE IV. Octahedral group O. Table of factors det $(1-\lambda A_s)$.

N_s	1	8	3	6	6
Class / Representation	E	C_3	C_2	C_2	C_4
Γ_1	$1-\lambda$	$1-\lambda$	$1-\lambda$	$1-\lambda$	$1-\lambda$
Γ_2	$1-\lambda$	$1-\lambda$	$1-\lambda$	$1+\lambda$	$1+\lambda$
Γ_3	$(1-\lambda)^2$	$1+\lambda+\lambda^2$	$(1-\lambda)^2$	$1-\lambda^2$	$1-\lambda^2$
Γ_4	$(1-\lambda)^3$	$1-\lambda^3$	$(1-\lambda)(1+\lambda)^2$	$(1-\lambda)(1+\lambda)^2$	$(1-\lambda)(1+\lambda^2)$
Γ_5	$(1-\lambda)^3$	$1-\lambda^3$	$(1-\lambda)(1+\lambda)^2$	$(1-\lambda)^2(1+\lambda)$	$(1+\lambda)(1+\lambda^2)$

TABLE V. Icosahedral group I. Table of factors det $(1-\lambda A_s)$; $\omega=(1+\sqrt{5})/2$, $\bar{\omega}=(1-\sqrt{5})/2$.

N_s	1	15	20	12	12
Repres. \ Class	C_1	C_4	C_5	C_2	C_3
Γ_1	$1-\lambda$	$1-\lambda$	$1-\lambda$	$1-\lambda$	$1-\lambda$
Γ_2	$(1-\lambda)^3$	$(1-\lambda)(1+\lambda)^2$	$1-\lambda^3$	$(1-\lambda)(1+\omega\lambda+\lambda^2)$	$(1-\lambda)(1+\bar{\omega}\lambda+\lambda^2)$
Γ_3	$(1-\lambda)^3$	$(1-\lambda)(1+\lambda)^2$	$1-\lambda^3$	$(1-\lambda)(1+\bar{\omega}\lambda+\lambda^2)$	$(1-\lambda)(1+\omega\lambda+\lambda^2)$
Γ_4	$(1-\lambda)^4$	$(1-\lambda^2)^2$	$(1-\lambda)(1-\lambda^3)$	$1+\lambda+\lambda^2+\lambda^3+\lambda^4$	$1+\lambda+\lambda^2+\lambda^3+\lambda^4$
Γ_5	$(1-\lambda)^5$	$(1-\lambda)^3(1+\lambda)^2$	$(1-\lambda)(1+\lambda+\lambda^2)^2$	$1-\lambda^5$	$1-\lambda^5$

TABLE VI. Exponents and coefficients of generating functions for irreducible representations of the tetrahedral group. For notation see the text.

Γ_m \ Γ_r	Numerator				Denominator
	Γ_1	Γ_2	Γ_3	Γ_4	
Γ_1	0	-	-	-	1
Γ_2	0	1	2	-	3
Γ_3	0	2	1	-	3
Γ_4	0,6	2,4	2,4	$1,2,3^2,4,5$	2,3,4

TABLE VII. Exponents and coefficients of generating functions for irreducible representations of the octahedral group. For notation see the text.

Γ_m \ Γ_r	Numerator					Denominator
	Γ_1	Γ_2	Γ_3	Γ_4	Γ_5	
Γ_1	0	-	-	-	-	1
Γ_2	0	1	-	-	-	2
Γ_3	0	3	1,2	-	-	2,3
Γ_4	0,9	3,6	2,4,5,7	1,3,4,5,6,8	2,3,4,5,6,7	2,4,6
Γ_5	0	6	2,4	3,4,5.	1,2,3	2,3,4

TABLE VIII. Exponents and coefficients of generating functions for irreducible representations of the icosahedral group. For notation see the text.

$\Gamma_m \backslash \Gamma_r$	Numerator					Denominator
	Γ_1	Γ_2	Γ_3	Γ_4	Γ_5	
Γ_1	0	-	-	-	-	1
Γ_2	0,15	1,5,6,9,10,14	3,5,7,8,10,12	3,4,6,7,8,9,11,12	2,4,5,6,7,8,9,10,11,13	2,6,10
Γ_3	0,15	3,5,7,8,10,12	1,5,6,9,10,14	3,4,6,7,8,9,11,12	2,4,5,6,7,8,9,10,11,13	2,6,10
Γ_4	0,10	$3,4,5^2,6,7$	$3,4,5^2,6,7$	1,2,3,4,6,7,8,9	$2,3,4^2,5^2,6^2,7,8$	2,3,4,5
Γ_5	$0,5,6^2,7,12$	$3,4^2,5^4,6^4,7^4,8^2,9$	$3,4^2,5^4,6^4,7^4,8^2,9$	$2,3^3,4^3,5^3,6^4,7^3,8^3,9^3,10$	$1,2^2,3^2,4^4,5^4,6^4,7^4,8^4,9^2,10^2,11$	2,3,3,4,5

tively. For Γ_r^e the matrices representing α and α' are just those of the corresponding representation Γ_r of G; for Γ_r^o the matrix of α is just that of Γ_r, while the matrix of α' is reversed in sign.

Hence we find the following rules for the generating functions for $G\times P$ tensors expressed in terms of those for G:

$$B(\Gamma_r^e,\Gamma_m^e;\lambda)=B(\Gamma_r,\Gamma_m;\lambda), \tag{34}$$

$$B(\Gamma_r^o,\Gamma_m^e;\lambda)=0, \tag{35}$$

$$B(\Gamma_r^e,\Gamma_m^o;\lambda)=\tfrac{1}{2}[B(\Gamma_r,\Gamma_m;\lambda)+B(\Gamma_r,\Gamma_m;-\lambda)], \tag{36}$$

$$B(\Gamma_r^o,\Gamma_m^o;\lambda)=\tfrac{1}{2}[B(\Gamma_r,\Gamma_m;\lambda)-B(\Gamma_r,\Gamma_m;-\lambda)]. \tag{37}$$

Incidentally, the group D_{2n} is isomorphic to $D_n\times P$ with

TABLE IX. Generating elements for the irreducible representations of D_n, T, and O of dimension greater than one.

Group	Representation	Representation generating matrices
D_n $(n=3,5,\ldots)$	Γ_a $(a=3,4,\ldots,\frac{n+3}{2})$	$\begin{pmatrix}0&1\\1&0\end{pmatrix}$, $\begin{pmatrix}\exp[2\pi i(a-2)/n]&0\\0&\exp[-2\pi i(a-2)/n]\end{pmatrix}$
D_n $(n=2,4,\ldots)$	Γ_a $(a=5,6,\ldots,\frac{n}{2}+3)$	$\begin{pmatrix}0&1\\1&0\end{pmatrix}$, $\begin{pmatrix}\exp[2\pi i(a-4)/n]&0\\0&\exp[-2\pi i(a-4)/n]\end{pmatrix}$
T	Γ_4	$\begin{pmatrix}1&0&0\\0&-1&0\\0&0&-1\end{pmatrix}$, $\begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix}$
O	Γ_3	$\begin{pmatrix}-1&0\\0&1\end{pmatrix}$, $\frac{1}{2}\begin{pmatrix}-1&-\sqrt{3}\\\sqrt{3}&-1\end{pmatrix}$
	Γ_4	$\begin{pmatrix}1&0&0\\0&0&1\\0&-1&0\end{pmatrix}$, $\begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix}$
	Γ_5	$\begin{pmatrix}-1&0&0\\0&0&-1\\0&1&0\end{pmatrix}$, $\begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix}$

the rotation π about the $2n$th order axis playing the role of the reflection, if n is odd.

E. An example

Let us consider a specific example. Suppose we want to write explicitly the generating function $B(\Gamma_3,\Gamma_5;\lambda)$ for the (Γ_3,Γ_5) tensors of the icosahedral group in the form (4). At the intersection of column Γ_3 and row Γ_5 of Table VIII we find the entry 3, 4^2, 5^4, 6^4, 7^4, 8^2, 9 which is a short notation for the numerator $N_{35}(\lambda)=\lambda^3+2\lambda^4+4\lambda^5+4\lambda^6+4\lambda^7+2\lambda^8+\lambda^9$. The intersection of the last column and the same row contains 2, 3, 3, 4, 5 which are the powers of λ in the factors $(1-\lambda^a)$ of the denominator $D(\lambda)$. Thus finally, one has

$$B(\Gamma_3;\Gamma_5,\lambda) = \frac{\lambda^3+2\lambda^4+4\lambda^5+4\lambda^6+4\lambda^7+2\lambda^8+\lambda^9}{(1-\lambda^2)(1-\lambda^3)(1-\lambda^3)(1-\lambda^4)(1-\lambda^5)}. \quad (38)$$

Let us illustrate, on the same example, the information about the structure of the (Γ_3,Γ_5) polynomials contained in (38). In this particular case we are interested in tensors whose components span a Γ_3 representation space and are polynomials in components of another tensor which transforms as Γ_5.

The powers of λ present in the numerator of (33) simply the existence of (Γ_3,Γ_5) tensors of degrees 3, 4, 5, 6, 7, 8, and 9. The coefficient k_p of a power λ^p specifies that there are precisely k_p independent (Γ_3,Γ_5) tensors of the same degree p. Thus we get one such tensor of degrees 3 and 9, two of degrees 4 and 8, and four of degrees 5, 6, and 7. Furthermore, only these (Γ_3,Γ_5) tensors form the integrity basis for all (Γ_3,Γ_5) tensors. That is to say, any (Γ_3,Γ_5) tensor can be written as a sum of terms each depending linearly on one element of the (Γ_3,Γ_5) integrity basis multiplied by a polynomial of invariants, (Γ_1,Γ_5) tensors, transforming according to the identity representation Γ_1 of the group I. An integrity basis for these (Γ_1,Γ_5) tensors, consisting of one invariant of degree 2, 4, and 5 and two invariants of degree 3, is implied by the powers of λ of the denominator of (38).

In order to determine how many independent (Γ_3,Γ_5) tensors of degree, say 5, there are, it suffices to expand (38) into a power series (assuming $\lambda < 1$) and compute the coefficient of λ^5 in the expansion. One has from (40)

$$\begin{aligned} B(\Gamma_3,\Gamma_5;\lambda) &= (\lambda^3+2\lambda^4+4\lambda^5+4\lambda^6+4\lambda^7+2\lambda^8+\lambda^9) \\ &\quad \times(1+\lambda^2+2\lambda^3+2\lambda^4+\cdots) \\ &= \lambda^3+2\lambda^4+5\lambda^5+8\lambda^6+\cdots. \end{aligned} \quad (39)$$

Hence there are 5 different (Γ_3,Γ_5) tensors of degree 5. Four of them are just the elements of the (Γ_3,Γ_5) integrity basis and the fifth one is a product of the third degree (Γ_3,Γ_5) tensor with an invariant $[(\Gamma_3,\Gamma_5)$ tensor$]$ of degree 2.

IV. INTEGRITY BASES FOR IRREDUCIBLE REPRESENTATIONS OF POINT GROUPS

In this section we construct the integrity bases of (Γ_r,Γ_m) tensors for the groups C_n, D_n, T, and O, that is, we find the basis tensors $I^{(q)}(\Gamma_m)$, $E^{(p)}(\Gamma_r,\Gamma_m)$ for the irreducible representations of the above groups. An integrity basis for invariants consists of denominator invariants $I(\Gamma_m)$ and possibly also the numerator invariants $E(\Gamma_1,\Gamma_m)$. An integrity basis for (Γ_r,Γ_m) tensors different from invariants, i.e., $\Gamma_r\neq\Gamma_1$, consists only of $E(\Gamma_r,\Gamma_m)$ of corresponding degrees. Whenever the representation Γ_m is of dimension one, the integrity basis for (Γ_r,Γ_m) tensors consists of at most one tensor.

The generating function $B(\Gamma_r,\Gamma_m;\lambda)$ provides the number and the degree of the $E^{(p)}(\Gamma_r,\Gamma_m)$ and $I^{(q)}(\Gamma_1,\Gamma_m)$. To find an elementary tensor of type Γ_r and degree p in the components of a tensor of type Γ_m, assume that its components are arbitrary polynomials of degree p; then impose that its components transform by Γ_r when components on which it depends are transformed by Γ_m. For this purpose it suffices to apply only the generator elements of the group G. The ambiguity which arises when there is more than one tensor of a given degree is resolved by considerations of simplicity, and, of course, linear independence.

The components of a Γ_m type tensor are denoted by $\alpha,\beta,\gamma,\cdots$ and correspond to the rows $1,2,\cdots$ of the generating matrices in the Appendix. Separately in the subsequent section we consider the case when Γ_m is the representation, in general reducible, of Γ_v contained in the three-dimensional representation of O(3). There we use x, y, and z instead of α, β, and γ.

A. C_n tensors

The irreducible tensors of C_n representations have just one component. So we have to consider powers of that component.

For Γ_m, $2\leqslant m\leqslant n$, we find

$$\begin{aligned} &I(\Gamma_m)=\alpha^a, && a=n/\mathrm{HCF}(n,m-1), \\ &E(\Gamma_r,\Gamma_m)=\alpha^p, && p=1,2,\ldots,a-1, \text{ and } r=p(m-1)_{\mathrm{mod}\ n}, \end{aligned} \quad (40)$$

and trivially

$$I(\Gamma_1)=\alpha.$$

B. D_n tensors

First consider the case n odd. The integrity bases for the invariants follow from the generating functions (17), (18), and (22). One gets

$$\begin{aligned} &I(\Gamma_1)=\alpha && I(\Gamma_2)=\alpha^2, \\ &I^{(2)}(\Gamma_m)=\alpha\beta, && I^{(a)}(\Gamma_m)=\alpha^a+\beta^a, \quad m=3,4,\ldots,\frac{n+3}{2}. \end{aligned} \quad (41)$$

here a is given in (20) as a function of m. The remaining integrity bases are

$$\begin{aligned} &E(\Gamma_2,\Gamma_2)=\alpha, && E(\Gamma_2,\Gamma_m)=\alpha^a-\beta^a, \\ &E^{(p)}(\Gamma_r,\Gamma_m)=\begin{pmatrix}\alpha^p\\ \beta^p\end{pmatrix}, && E^{(1-p)}(\Gamma_r,\Gamma_m)=\begin{pmatrix}\beta^{a-p}\\ \alpha^{a-p}\end{pmatrix}, \end{aligned} \quad (42)$$

for $1\leqslant p(m-2)_{\mathrm{mod}\ n}\leqslant(n-1)/2$ and

$$E^{(p)}(\Gamma_r,\Gamma_m)=\begin{pmatrix}\beta^p\\ \alpha^p\end{pmatrix},$$

$$E^{(a-p)}(\Gamma_r,\Gamma_m)=\begin{pmatrix}\alpha^{a-p}\\ \beta^{a-p}\end{pmatrix}$$

for $(n+1)/2\leqslant p(m-2)_{\text{mod }n}\leqslant n-1$,

where the range of r and m is the same as in (41), r and p are related by the conditions (20), (21); otherwise there are no (Γ_r,Γ_m) tensors.

Suppose now that n is even. The integrity bases for invariants are:

$$I(\Gamma_1)=\alpha,\quad I(\Gamma_j)=\alpha^2,\quad j=2,3,4,$$
$$I^{(2)}(\Gamma_m)=\alpha\beta,\quad I^{(a)}(\Gamma_m)=\alpha^a+\beta^a, \tag{43}$$

where $m=5,6,...,(n+6)/2$, and a is given by (28).

The integrity bases for other tensors are then

$$E(\Gamma_j,\Gamma_j)=\alpha,\quad j=2,3,4,$$
$$E(\Gamma_2,\Gamma_m)=\alpha^a-\beta^a, \tag{44}$$
$$E^{(p)}(\Gamma_r,\Gamma_m)=\begin{pmatrix}\alpha^p\\ \beta^p\end{pmatrix},\quad E^{(a-p)}(\Gamma_r,\Gamma_m)=\begin{pmatrix}\beta^{a-p}\\ \alpha^{a-p}\end{pmatrix}.$$

for $1\leqslant p(m-4)_{\text{mod }n}\leqslant(n-2)/2$ and

$$E^{(p)}(\Gamma_r,\Gamma_m)=\begin{pmatrix}\beta^p\\ \alpha^p\end{pmatrix},\quad E^{(a-p)}(\Gamma_r,\Gamma_m)=\begin{pmatrix}\alpha^{a-p}\\ \beta^{a-p}\end{pmatrix}$$

for $(n+2)/2\leqslant p(m-4)_{\text{mod }n}\leqslant n-1$.

Here r and p are related by (29) and a is given in (28). The following integrity bases have no analog in the case n odd; they exist only for a even,

$$E(\Gamma_3,\Gamma_m)=\alpha^{a/2}-\beta^{a/2},$$
$$E(\Gamma_4,\Gamma_m)=\alpha^{a/2}+\beta^{a/2}. \tag{45}$$

C. Tetrahedral tensors

First we list the invariants $E(\Gamma_1,\Gamma_m)$ and $I(\Gamma_m)$ for all four irreducible representations of T. Their degrees are shown respectively in the first and last column of Table VI. We find

$$I(\Gamma_1)=\alpha,\quad I(\Gamma_2)=\alpha^3,\quad I(\Gamma_3)=\alpha^3,$$
$$I^{(2)}(\Gamma_4)=\alpha^2+\beta^2+\gamma^2,\quad I^{(3)}(\Gamma_4)=\alpha\beta\gamma, \tag{46}$$
$$I^{(4)}(\Gamma_4)=\alpha^4+\beta^4+\gamma^4,$$
$$E^{(6)}(\Gamma_1,\Gamma_4)=(\alpha^2-\beta^2)(\beta^2-\gamma^2)(\gamma^2-\alpha^2);$$

$E^{(6)}(\Gamma_1,\Gamma_4)$ is linearly independent of polynomials in $I^{(2)}(\Gamma_4)$, $I^{(3)}(\Gamma_4)$, $I^{(4)}(\Gamma_4)$, but its square is a polynomial in the lower degree scalars.

Corresponding to the intersection of column Γ_2 and row Γ_3 of Table VI we find,

$$E(\Gamma_2,\Gamma_2)=\alpha,\quad E(\Gamma_3,\Gamma_3)=\alpha,$$
$$E(\Gamma_3,\Gamma_2)=\alpha^2,\quad E(\Gamma_2,\Gamma_3)=\alpha^2. \tag{47}$$

The degrees of tensors of type Γ_2, Γ_3, and Γ_4, built from one of types Γ_4 are given in the last row of Table VI. One finds after some computation:

$$E^{(2)}(\Gamma_2,\Gamma_4)=\alpha^2+\omega\beta^2+\omega^2\gamma^2,$$
$$E^{(4)}(\Gamma_2,\Gamma_4)=\alpha^4+\omega\beta^4+\omega^2\gamma^4,$$
$$E^{(2)}(\Gamma_3,\Gamma_4)=\alpha^2+\omega^2\beta^2+\omega\gamma^2, \tag{48}$$
$$E^{(4)}(\Gamma_3,\Gamma_4)=\alpha^4+\omega^2\beta^4+\omega\gamma^4,$$

and

$$E^{(1)}(\Gamma_4,\Gamma_4)=\begin{pmatrix}\alpha\\ \beta\\ \gamma\end{pmatrix},\quad E^{(2)}(\Gamma_4,\Gamma_4)=\begin{pmatrix}\beta\gamma\\ \gamma\alpha\\ \alpha\beta\end{pmatrix},$$
$$E_a^{(3)}(\Gamma_4,\Gamma_4)=\begin{pmatrix}\alpha^3\\ \beta^3\\ \gamma^3\end{pmatrix},\quad E_b^{(3)}(\Gamma_4,\Gamma_4)=\begin{pmatrix}\alpha\beta^2\\ \beta\gamma^2\\ \gamma\alpha^2\end{pmatrix}, \tag{49}$$
$$E^{(4)}(\Gamma_4,\Gamma_4)=\begin{pmatrix}\gamma\beta^3\\ \alpha\gamma^3\\ \beta\alpha^3\end{pmatrix},\quad E^{(5)}(\Gamma_4,\Gamma_4)=\begin{pmatrix}\alpha\beta^4\\ \beta\gamma^4\\ \gamma\alpha^4\end{pmatrix}.$$

The linearly independent third degree tensors are distinguished by the subscripts a and b.

D. Octahedral tensors

The existence and degree of an elementary invariant $E(\Gamma_1,\Gamma_m)$ or $I(\Gamma_m)$ of the octahedral group is shown in the first or last column of Table VII. We find

$$I(\Gamma_1)=\alpha,\quad I(\Gamma_2)=\alpha^2,\quad I^{(2)}(\Gamma_3)=\alpha^2+\beta^2,$$
$$I^{(3)}(\Gamma_3)=3\alpha^2\beta-\beta^3,\quad I^{(2)}(\Gamma_4)=\alpha^2+\beta^2+\gamma^2,$$
$$I^{(4)}(\Gamma_4)=\alpha^4+\beta^4+\gamma^4,\quad I^{(6)}(\Gamma_4)=\alpha^6+\beta^6+\gamma^6,$$
$$E^{(9)}(\Gamma_1,\Gamma_4)=\alpha\beta\gamma(\alpha^2-\beta^2)(\beta^2-\gamma^2)(\gamma^2-\alpha^2), \tag{50}$$
$$I^{(2)}(\Gamma_5)=\alpha^2+\beta^2+\gamma^2,\quad I^{(3)}(\Gamma_5)=\alpha\beta\gamma,$$
$$I^{(4)}(\Gamma_5)=\alpha^4+\beta^4+\gamma^4.$$

The ninth degree invariant in the bases of Γ_4 should be used at most linearly because its square can be expressed as a polynomial in lower degree invariants.

For tensors of other types we find the following integrity bases:

$$E^{(1)}(\Gamma_2,\Gamma_2)=\alpha, \tag{51}$$
$$E^{(3)}(\Gamma_2,\Gamma_3)=3\alpha\beta^2-\alpha^3,$$
$$E^{(1)}(\Gamma_3,\Gamma_3)=\begin{pmatrix}\alpha\\ \beta\end{pmatrix}, \tag{52}$$
$$E^{(2)}(\Gamma_3,\Gamma_3)=\begin{pmatrix}2\alpha\beta\\ \alpha^2-\beta^2\end{pmatrix}.$$
$$E^{(3)}(\Gamma_2,\Gamma_4)=\alpha\beta\gamma,$$
$$E^{(6)}(\Gamma_2,\Gamma_4)=(\alpha^2-\beta^2)(\beta^2-\gamma^2)(\gamma^2-\alpha^2)$$
$$E^{(2)}(\Gamma_3,\Gamma_4)=\begin{pmatrix}\sqrt{3}(\beta^2-\gamma^2)\\ \beta^2+\gamma^2-2\alpha^2\end{pmatrix},$$
$$E^{(4)}(\Gamma_3,\Gamma_4)=\begin{pmatrix}\sqrt{3}(\beta^4-\gamma^4)\\ \beta^4+\gamma^4-2\alpha^4\end{pmatrix}, \tag{53}$$

$$E^{(5)}(\Gamma_3,\Gamma_4)=\alpha\beta\gamma\begin{pmatrix}\beta^2+\gamma^2-2\alpha^2\\ \sqrt{3}(\gamma^2-\beta^2)\end{pmatrix},$$

$$E^{(7)}(\Gamma_3,\Gamma_4)=\alpha\beta\gamma\begin{pmatrix}\beta^4+\gamma^4-2\alpha^4\\ \sqrt{3}(\gamma^4-\beta^4)\end{pmatrix},$$

$$E^{(1)}(\Gamma_4,\Gamma_4)=\begin{pmatrix}\alpha\\ \beta\\ \gamma\end{pmatrix},\quad E^{(3)}(\Gamma_4,\Gamma_4)=\begin{pmatrix}\alpha^3\\ \beta^3\\ \gamma^3\end{pmatrix},$$

$$E^{(4)}(\Gamma_4,\Gamma_4)=\begin{pmatrix}(\beta^2-\gamma^2)\beta\gamma\\ (\gamma^2-\alpha^2)\gamma\alpha\\ (\alpha^2-\beta^2)\alpha\beta\end{pmatrix},$$

$$E^{(5)}(\Gamma_4,\Gamma_4)=\begin{pmatrix}\alpha^5\\ \beta^5\\ \gamma^5\end{pmatrix},$$

$$E^{(6)}(\Gamma_4,\Gamma_4)=\begin{pmatrix}(\beta^4-\gamma^4)\beta\gamma\\ (\gamma^4-\alpha^4)\gamma\alpha\\ (\alpha^4-\beta^4)\alpha\beta\end{pmatrix},$$

$$E^{(8)}(\Gamma_4,\Gamma_4)=\begin{pmatrix}(\beta^6-\gamma^6)\beta\gamma\\ (\gamma^6-\alpha^6)\gamma\alpha\\ (\alpha^6-\beta^6)\alpha\beta\end{pmatrix},$$

$$E^{(2)}(\Gamma_5,\Gamma_4)=\begin{pmatrix}\beta\gamma\\ \gamma\alpha\\ \alpha\beta\end{pmatrix},\quad E^{(3)}(\Gamma_5,\Gamma_4)=\begin{pmatrix}\alpha(\beta^2-\gamma^2)\\ \beta(\gamma^2-\alpha^2)\\ \gamma(\alpha^2-\beta^2)\end{pmatrix},$$

$$E^{(4)}(\Gamma_5,\Gamma_4)=\begin{pmatrix}\alpha^2\beta\gamma\\ \beta^2\gamma\alpha\\ \gamma^2\alpha\beta\end{pmatrix},\quad E^{(5)}(\Gamma_5,\Gamma_4)=\begin{pmatrix}\alpha(\beta^4-\gamma^4)\\ \beta(\gamma^4-\alpha^4)\\ \gamma(\alpha^4-\beta^4)\end{pmatrix},$$

$$E^{(6)}(\Gamma_5,\Gamma_4)=\begin{pmatrix}\alpha^4\beta\gamma\\ \beta^4\gamma\alpha\\ \gamma^4\alpha\beta\end{pmatrix},\quad E^{(7)}(\Gamma_5,\Gamma_4)=\begin{pmatrix}\alpha(\beta^6-\gamma^6)\\ \beta(\gamma^6-\alpha^6)\\ \gamma(\alpha^6-\beta^6)\end{pmatrix},$$

and finally, from a Γ_5 tensor the integrity bases for (Γ_2,Γ_5), (Γ_3,Γ_5), (Γ_4,Γ_5), and (Γ_5,Γ_5) tensors can be chosen as

$$E^{(6)}(\Gamma_2,\Gamma_5)=(\alpha^2-\beta^2)(\beta^2-\gamma^2)(\gamma^2-\alpha^2),$$

$$E^{(2)}(\Gamma_3,\Gamma_5)=\begin{pmatrix}\sqrt{3}(\beta^2-\gamma^2)\\ \beta^2+\gamma^2-2\alpha^2\end{pmatrix},$$

$$E^{(4)}(\Gamma_3,\Gamma_5)=\begin{pmatrix}\sqrt{3}(\beta^4-\gamma^4)\\ \beta^4+\gamma^4-2\alpha^4\end{pmatrix},$$

$$E^{(3)}(\Gamma_4,\Gamma_5)=\begin{pmatrix}\alpha(\beta^2-\gamma^2)\\ \beta(\gamma^2-\alpha^2)\\ \gamma(\alpha^2-\beta^2)\end{pmatrix}, \tag{54}$$

$$E^{(4)}(\Gamma_4,\Gamma_5)=\begin{pmatrix}\beta\gamma(\beta^2-\gamma^2)\\ \gamma\alpha(\gamma^2-\alpha^2)\\ \alpha\beta(\alpha^2-\beta^2)\end{pmatrix},$$

$$E^{(5)}(\Gamma_4,\Gamma_5)=\begin{pmatrix}\alpha^3(\beta^2-\gamma^2)\\ \beta^3(\gamma^2-\alpha^2)\\ \gamma^3(\alpha^2-\beta^2)\end{pmatrix},$$

$$E^{(1)}(\Gamma_5,\Gamma_5)=\begin{pmatrix}\alpha\\ \beta\\ \gamma\end{pmatrix},\quad E^{(3)}(\Gamma_5,\Gamma_5)=\begin{pmatrix}\beta\gamma\\ \gamma\alpha\\ \alpha\beta\end{pmatrix},$$

$$E^{(3)}(\Gamma_5,\Gamma_5)=\begin{pmatrix}\alpha^3\\ \beta^3\\ \gamma^3\end{pmatrix}.$$

E. Integrity bases for groups with reflections

Integrity bases for representation spaces of a point group $G\times P$, which includes the reflections P, are easily obtained from those corresponding to G only. Recalling the notation Γ^e and Γ^o for representations of $G\times P$ corresponding to a representation Γ of G (cf sec. II D) we have

$$\begin{aligned}&E^{(k)}(\Gamma_r^e,\Gamma_m^e)=E^{(k)}(\Gamma_r,\Gamma_m),\quad E^{(k)}(\Gamma_r^o,\Gamma_m^e)=0,\\ &E^{(2k)}(\Gamma_r^e,\Gamma_m^o)=E^{(2k)}(\Gamma_r,\Gamma_m),\\ &E^{(2k+1)}(\Gamma_r^e,\Gamma_m^o)=0,\\ &E^{(2k+1)}(\Gamma_r^o,\Gamma_m^o)=E^{(2k+1)}(\Gamma_r,\Gamma_m),\\ &E^{(2k)}(\Gamma_r^o,\Gamma_m^o)=0.\end{aligned} \tag{55}$$

For simplicity of notation we did not distinguish the invariants $I^{(k)}(\Gamma_1,\Gamma_m)$ from $E^{(k)}(\Gamma_1,\Gamma_m)$ in (55).

As for the generating functions, the integrity bases of $H[G$ are also just those of G.

V. HOMOGENEOUS (x,y,z) INTEGRITY BASES FOR REPRESENTATIONS OF POINT GROUPS

In this section we reduce the three-dimensional $L=1$ representation of O(3) to representations Γ_v of a point group G, and use it as Γ_m, in the formalism of Secs. III and IV in order to find the generating functions and integrity bases. When Γ_v is irreducible the results coincide with those found previously (the variables α,β,γ should be replaced by x,y,z). Therefore, we here solve the problem only when Γ_v is reducible, i.e., for the groups C_n, D_n, $C_n[C_{2n}$, $C_n[D_n$, $D_n[D_{2n}$, $C_n\times P$, and $D_n\times P$.

The group $H[G$ is derived[5] from a rotational group G which has H as its subgroup of index 2. The elements of $H[G$ are those of H together with the elements JR where R belongs to G but not to H and J is the reflection of x, y, and z. As an abstract group $H[G$ is isomorphic to G; but the representation Γ_v is not the same for $H[G$ as for G.

The three-dimensional representation of O(3) contains the following representations Γ_v of point groups:

$\Gamma_v=\Gamma_1\oplus\Gamma_2\oplus\Gamma_n$	for C_n
$\Gamma_2\oplus\Gamma_3$	for D_n, n odd
$\Gamma_2\oplus\Gamma_5$	for D_n, n even
Γ_4	for T,
Γ_5	for O,
Γ_2	for I,
$\Gamma_n\oplus\Gamma_{n+1}\oplus\Gamma_{n+2}$	for $C_n[C_{2n}$,
$\Gamma_1\oplus\Gamma_3$	for $C_n[D_n$, n odd,
$\Gamma_1\oplus\Gamma_5$	for $C_n[D_n$, n even,
$\Gamma_4\oplus\Gamma_{n+3}$	for $D_n[D_{2n}$,
Γ_4	for $T[O$. (56)

The corresponding Γ_v for the groups $G\times P$, where $G=C_n$, D_n, T, 0, I, is the same as in (56) except that every Γ_a should be replaced by Γ_a^o (cf. Sec. III D).

A. The generating functions

The generating functions $B(\Gamma_r,\Gamma_v;\lambda)$, where Γ_v is given by (56), are read from the Tables 6,7,8 for the groups T,O,I and $T[0$, because the representation Γ_v is irreducible. For the groups of type $T\times P$, $0\times P$, $I\times P$ the generating functions are obtained from Tables VI, VII, and VIII in conjunction with Eq. (36) and (37). For the groups C_n, D_n, $C_n[C_{2n}$, $C_n[D_n$, $C_n[D_{2n}$, $C_n\times P$, and $D_n\times P$ for which Γ_v is reducible, the generating functions need to be calculated with the help of (10). Using (10) and (15), we get for C_n

$$B(\Gamma_r,\Gamma_1\oplus\Gamma_2\oplus\Gamma_n;\lambda_1,\lambda_2,\lambda_n)=\frac{1}{(1-\lambda_1)(1-\lambda_2^n)(1-\lambda_n^n)}\times\left\{\sum_{p=0}^{n-r}\lambda_2^{r+p-1}\lambda_n^p+\sum_{p=n-r+1}^{n-1}\lambda_2^{r+p-n-1}\lambda_n^p\right\} \quad (57)$$

In the case $r=1$ the second sum of (57) vanishes. Similarly for D_n, n odd, we get from (10), (19)–(24).

$$B(\Gamma_r,\Gamma_2\oplus\Gamma_3;\lambda_2,\lambda_3)=\frac{1+\lambda_2}{1-\lambda_2^2}B(\Gamma_r,\Gamma_3;\lambda_3)=\frac{(\lambda_3^{r-2}+\lambda_3^{n-r+2})(1-\lambda_2)}{(1-\lambda_2^2)(1-\lambda_3^2)(1-\lambda_3^n)}, \quad (58)$$

with $r=3,4,\ldots,(n+3)/2$, and

$$B(\Gamma_1,\Gamma_2\oplus\Gamma_3;\lambda_2,\lambda_3)=\frac{1+\lambda_2\lambda_3^n}{(1-\lambda_2^2)(1-\lambda_3^2)(1-\lambda_3^n)}, \quad (59)$$

$$B(\Gamma_2,\Gamma_2\oplus\Gamma_3;\lambda_2,\lambda_3)=\frac{\lambda_2+\lambda_3^n}{(1-\lambda_2^2)(1-\lambda_3^2)(1-\lambda_3^n)}.$$

For D_n, n even, one has

$$B(\Gamma_1,\Gamma_2\oplus\Gamma_5;\lambda_2,\lambda_5)=\frac{1+\lambda_2\lambda_5^n}{(1-\lambda_2^2)(1-\lambda_5^2)(1-\lambda_5^n)},$$

$$B(\Gamma_2,\Gamma_2\oplus\Gamma_5;\lambda_2,\lambda_5)=\frac{\lambda_2+\lambda_5^n}{(1-\lambda_2^2)(1-\lambda_5^2)(1-\lambda_5^n)}, \quad (60)$$

$$B(\Gamma_j,\Gamma_2\oplus\Gamma_5;\lambda_2,\lambda_5)=\frac{\lambda_5^{n/2}+\lambda_2\lambda_5^{n/2}}{(1-\lambda_2^2)(1-\lambda_5^2)(1-\lambda_5^n)},$$

$j=3,4,$

$$B(\Gamma_r,\Gamma_2\oplus\Gamma_5;\lambda_2,\lambda_5)=\frac{\lambda_5^{r-4}+\lambda_2\lambda_5^{n-r+4}}{(1-\lambda_2^2)(1-\lambda_5^2)(1-\lambda_5^n)},$$

where $r=5,6,\ldots,n/2+3$.

For $C_n[C_{2n}$ we have to distinguish n even and odd. For n odd we find

$$B(\Gamma_r,\Gamma_n\oplus\Gamma_{n+1}\oplus\Gamma_{n+2};\lambda_n,\lambda_{n+1},\lambda_{n+2})=D^{-1}\left[\sum_{p=0}^{n-r}\lambda_n^p\lambda_{n+2}^{r-1+p}+\sum_{p=n-r+1}^{n-1}\lambda_n^p\lambda_{n+2}^{p+r-n-1}\right]$$

for $1\leqslant r\leqslant n$, r odd

$$=D^{-1}\left[\sum_{p=0}^{2n-r}\lambda_n^p\lambda_{n+2}^{r-n-1+p}+\sum_{p=2n-r+1}^{n-1}\lambda_n^p\lambda_{n+2}^{r-2n-1+p}\right]$$

for $n+2\leqslant r\leqslant 2n-1$, r odd

$$=D^{-1}\left[\lambda_{n+1}\sum_{p=0}^{2n-r}\lambda_n^p\lambda_{n+2}^{r-n-1+p}+\sum_{p=2n-r+1}^{n-1}\lambda_n^p\lambda_{n+2}^{r-2n-1+p}\right]$$

for $2\leqslant r\leqslant n-1$, r even

$$=D^{-1}\left[\lambda_{n+1}\sum_{p=0}^{n-r}\lambda_n^p\lambda_{n+2}^{r-1+p}+\sum_{p=n-r+1}^{n-1}\lambda_n^p\lambda_{n+2}^{p+r-n-1}\right],$$

for $n+1\leqslant r\leqslant 2n$, r even, (61)

where

$$D=(1-\lambda_{n+1}^2)(1-\lambda_n^n)(1-\lambda_{n+2}^n). \quad (62)$$

For $C_n[C_{2n}$, n even, we have

$$B(\Gamma_r,\Gamma_n\oplus\Gamma_{n+1}\oplus\Gamma_{n+2};\lambda_n,\lambda_{n+1},\lambda_{n+2})=D^{-1}\left[\sum_{p=0}^{2n-r}\lambda_n^p\lambda_{n+2}^{r+p-1}+\sum_{p=2n-r+1}^{2n-1}\lambda_n^p\lambda_{n+2}^{r-1+p-2n}+\lambda_{n+1}\sum_{p=0}^{n-r}\lambda_n^p\lambda_{n+2}^{r-1+p+n}+\lambda_{n+1}\sum_{p=n-r+1}^{2n-1}\lambda_n^p\lambda_{n+2}^{r-1+p-n}\right] \quad (63)$$

for $1\leqslant r\leqslant n$

$$=D^{-1}\left[\sum_{p=0}^{2n-r}\lambda_n^p\lambda_{n+2}^{r+p-1}+\sum_{p=2n-r+1}^{2n-1}\lambda_n^p\lambda_{n+2}^{r-1+p-2n}+\lambda_{n+1}\sum_{p=0}^{3n-r}\lambda_n^p\lambda_{n+2}^{r-1+p-n}+\lambda_{n+1}\sum_{p=3n-r+1}^{2n-1}\lambda_n^p\lambda_{n+2}^{r-1+p-3n}\right],$$

for $n+1\leqslant r\leqslant 2n$,

where

$$D=(1-\lambda_{n+1}^2)(1-\lambda_n^{2n})(1-\lambda_{n+2}^{2n}). \quad (64)$$

For the group $C_n[D_n$ with n odd we find

$$B(\Gamma_1,\Gamma_1\oplus\Gamma_3;\lambda_1,\lambda_3)=[(1-\lambda_1)(1-\lambda_3^2)(1-\lambda_3^n)]^{-1}\equiv D^{-1}, \quad (65)$$

$$B(\Gamma_2,\Gamma_1\oplus\Gamma_3;\lambda_1,\lambda_3)=D^{-1}\lambda_3^n,$$

$$B(\Gamma_r,\Gamma_1\oplus\Gamma_3;\lambda_1,\lambda_3)=D^{-1}(\lambda_3^{r-2}+\lambda_3^{n-r+2}),$$

for $3\leqslant r\leqslant(n+3)/2$.

When n is even, we obtain

$$B(\Gamma_1,\Gamma_1\oplus\Gamma_5;\lambda_1,\lambda_5)=[(1-\lambda_1)(1-\lambda_5^2)(1-\lambda_5^n)]^{-1}\equiv D^{-1},$$

$$B(\Gamma_2,\Gamma_1\oplus\Gamma_5;\lambda_1,\lambda_5)=D^{-1}\lambda_5^n, \quad (66)$$

$$B(\Gamma_3,\Gamma_1\oplus\Gamma_5;\lambda_1,\lambda_5)=B(\Gamma_4,\Gamma_1\oplus\Gamma_5;\lambda_1,\lambda_5)=D^{-1}\lambda_5^{n/2},$$

$$B(\Gamma_r,\Gamma_1\oplus\Gamma_5;\lambda_1,\lambda_5)=D^{-1}(\lambda_5^{r-4}+\lambda_5^{n-r+4}),$$

for $5\leqslant r\leqslant n/2+3$.

For the group $D_n[D_{2n}$, n odd, we get

$$B(\Gamma_1,\Gamma_4\oplus\Gamma_{n+3};\lambda_4,\lambda_{n+3})=[(1-\lambda_4^2)(1-\lambda_{n+3}^2)(1-\lambda_{n+3}^n)]^{-1}\equiv D^{-1},$$

$$B(\Gamma_2,\Gamma_4\oplus\Gamma_{n+3};\lambda_4,\lambda_{n+3})=D^{-1}\lambda_{n+3}^n,$$

$$B(\Gamma_3,\Gamma_4\oplus\Gamma_{n+3};\lambda_4,\lambda_{n+3})=D^{-1}\lambda_{n+3}^n,$$

$$B(\Gamma_4,\Gamma_4\oplus\Gamma_{n+3};\lambda_4,\lambda_{n+3})=D^{-1}\lambda_4, \quad (67)$$

$$B(\Gamma_r,\Gamma_4\oplus\Gamma_{n+3};\lambda_4,\lambda_{n+3})$$
$$=D^{-1}(\lambda_{n+3}^{n-r+4}+\lambda_{n+3}^{r-4}) \quad \text{for } 6\leqslant r\leqslant n+3,\ r \text{ even},$$
$$=D^{-1}\lambda_4(\lambda_{n+3}^{n-r+4}+\lambda_{n+3}^{r-4}) \quad \text{for } 5\leqslant r\leqslant n+2,\ r \text{ odd}.$$

For the group $D_n[D_{2n}$ with n even has

$$B(\Gamma_1,\Gamma_4\oplus\Gamma_{n+3};\lambda_4,\lambda_{n+3})=D^{-1}(1+\lambda_4\lambda_{n+3}^n),$$

$$B(\Gamma_2,\Gamma_4\oplus\Gamma_{n+3};\lambda_4,\lambda_{n+3})=D^{-1}(\lambda_{n+3}^{2n}+\lambda_4\lambda_{n+3}^n),$$

$$B(\Gamma_3,\Gamma_4\oplus\Gamma_{n+3};\lambda_4,\lambda_{n+3})=D^{-1}(\lambda_{n+3}^n+\lambda_4\lambda_{n+3}^{2n}),$$

$$B(\Gamma_4,\Gamma_4\oplus\Gamma_{n+3};\lambda_4,\lambda_{n+3})=D^{-1}(\lambda_{n+3}^n+\lambda_4), \quad (68)$$

$$B(\Gamma_r,\Gamma_4\oplus\Gamma_{n+3};\lambda_4,\lambda_{n+3})$$
$$=D^{-1}(\lambda_{n+3}^{r-4}+\lambda_{n+3}^{2n-r+4}+\lambda_4\lambda_{n+3}^{n-r+4}+\lambda_4\lambda_{n+3}^{n+r-4}),$$

for $6\leqslant r\leqslant n+2, r$ even,

$$=D^{-1}(\lambda_{n+3}^{n+4-r}+\lambda_{n+3}^{n-4+r}+\lambda_4\lambda_{n+3}^{r-4}+\lambda_4\lambda_{n+3}^{2n-r+4}),$$

for $5\leqslant r\leqslant n+3$, r odd,

where $D=(1-\lambda_4^2)(1-\lambda_{n+3}^2)(1-\lambda_{n+3}^{2n})$. (69)

For the group $C_n\times P$ we obtain by repeated use of (34)–(37):

$$B(\Gamma_r^e,\Gamma_1^o\oplus\Gamma_2^o\oplus\Gamma_n^o;\lambda_1,\lambda_2,\lambda_n)$$
$$=\tfrac12[B(\Gamma_r,\Gamma_1\oplus\Gamma_2\oplus\Gamma_n;\lambda_1,\lambda_2,\lambda_n)$$
$$+B(\Gamma_r,\Gamma_1\oplus\Gamma_2\oplus\Gamma_n;-\lambda_1,-\lambda_2,-\lambda_n)], \quad (70)$$

$$B(\Gamma_r^o,\Gamma_1^o\oplus\Gamma_2^o\oplus\Gamma_n^0;\lambda_1,\lambda_2,\lambda_n)$$
$$=\tfrac12[B(\Gamma_r,\Gamma_1\oplus\Gamma_2\oplus\Gamma_n;\lambda_1,\lambda_2,\lambda_n)$$
$$-B(\Gamma_r,\Gamma_1\oplus\Gamma_2\oplus\Gamma_n;-\lambda_1,-\lambda_2,-\lambda_n)].$$

Here $r=1,2,\ldots,n$; on the right of (70) are the generating functions (57) of C_n

Finally, for $D_n\times P$ we have

$$B(\Gamma_r^e,\Gamma_2^o\oplus\Gamma_k^o;\lambda_2,\lambda_k)$$
$$=\tfrac12[B(\Gamma_r;\Gamma_2\oplus\Gamma_k;\lambda_2,\lambda_k)$$
$$+B(\Gamma_r,\Gamma_2\oplus\Gamma_k;-\lambda_2-\lambda_k)], \quad (71)$$

$$B(\Gamma_r^o,\Gamma_2^o\oplus\Gamma_k^o;\lambda_2,\lambda_k)$$
$$=\tfrac12[B(\Gamma_r,\Gamma_2\oplus\Gamma_k;\lambda_2,\lambda_k)$$
$$-B(\Gamma_r,\Gamma_2\oplus\Gamma_k;-\lambda_2,-\lambda_k)],$$
$$r=1,2,\ldots,n/2+3(k-1)/4,$$

where $k=3$ and 5 for n odd and even, respectively. On the right of (71) are the D_n generating functions (58)–(60).

B. The integrity bases

We choose the nth order axis of C_n as the z direction. Then the coordinates corresponding to Γ_1, Γ_2, and Γ_n are respectively, z, $x_+=x+iy$, and $x_-=x-iy$. Integrity bases are then read off the generating function (15). The invariants corresponding to the denominator of (57) are

$$I^{(1)}(\Gamma_v)=z,\quad I_+^{(n)}(\Gamma_v)=x_+^n,\quad I_-^{(n)}(\Gamma_v)=x_-^n, \quad (72)$$

while those corresponding to the numerator are, for $r=1$,

$$E^{(2p)}(\Gamma_1,\Gamma_v)=(x^2+y^2)^p=x_+^p x_-^p,\quad p=0,1,\ldots,n-1, \quad (73)$$

and for $r>1$,

$$E^{(r+2p-1)}(\Gamma_r,\Gamma_v)=x_+^p x_-^{r+p-1},\quad p=0,1,\ldots,n-r, \quad (74)$$

$$E^{(r+2k-1-n)}(\Gamma_r,\Gamma_v)=x_-^k x_+^{r+k-n-1},$$
$$k=n-r+1,n-r+2,\ldots,n-1.$$

The situation for C_n is summarized by saying that $x_+^a x_-^b z^c$ transforms by Γ_r, where $r=(a-b)_{\mathrm{mod}\,n}+1$. Similarly for $C_n[C_{2n}$ the results are summarized by saying that $x_+^a x_-^b z^c$ transforms as Γ_r, where $r=1+[a(n+1)+b(n-1)+cn]_{\mathrm{mod}\,2n}$, and x_+, x_-, z transforms by the representations Γ_{n+2}, Γ_n, Γ_{n+1} respectively. The integrity bases are found directly from the generating functions (61)–(64). We omit the great number of trivial formulas.

For D_n we choose the nth order axis as the z direction and one of the 2nd order axes as the x direction.

For n odd, the representation Γ_2 transforms z, while Γ_3 acts in the xy plane. From (59) we read off the invariants

$$I_z^{(2)}(\Gamma_v)=z^2,\quad I_x^{(2)}(\Gamma_3)=x^2+y^2=x_+x_-, \quad (75)$$

$$I^{(n)}(\Gamma_v)=x_+^n+x_-^n,\quad E^{(n+1)}(\Gamma_1,\Gamma_v)=z(x_+^n-x_-^n),$$

and for the other tensors from (58) and (59),

$$E(\Gamma_2,\Gamma_v)=z, \quad E^{(n)}(\Gamma_2,\Gamma_n)=x_+^n-x_-^n,$$

$$E^{(r-2)}(\Gamma_r,\Gamma_v)=\begin{pmatrix}x_+^{r-2}\\ x_-^{r-2}\end{pmatrix},$$

$$E^{(n-r+2)}(\Gamma_r,\Gamma_v)=\begin{pmatrix}x_-^{n-r+2}\\ x_+^{n-r+2}\end{pmatrix}, \tag{76}$$

$$E^{(r-1)}(\Gamma_r,\Gamma_v)=\begin{pmatrix}z\,x_+^{r-2}\\ -z\,x_-^{r-2}\end{pmatrix},$$

$$E^{(n-r+3)}(\Gamma_r,\Gamma_v)=\begin{pmatrix}z\,x_-^{n-r+2}\\ -z\,x_+^{n-2+2}\end{pmatrix},$$

for $r=3,4,\ldots,(n+3)/2$. For n even the representation Γ_2 acts on z while Γ_5 transforms the (x,y) plane. From (60) we conclude

$$I_z^{(2)}(\Gamma_v)=z^2, \quad I_x^{(2)}(\Gamma_v)=x^2+y^2=x_+x_-,$$

$$I^{(n)}(\Gamma_v)=x_+^n+x_-^n, \quad E^{(n+1)}(\Gamma_1,\Gamma_v)=z(x_+^n-x_-^n),$$

$$E^{(n/2)}(\Gamma_3,\Gamma_v)=x_+^{n/2}+x_-^{(n/2)},$$

$$E^{(n/2)}(\Gamma_4,\Gamma_v)=x_+^{n/2}-x_-^{n/2},$$

$$E^{(n/2+1)}(\Gamma_3,\Gamma_v)=z(x_+^{n/2}-x_-^{n/2}), \tag{77}$$

$$E^{(n/2+1)}(\Gamma_4,\Gamma_v)=z(x_+^{n/2}+x_-^{n/2}),$$

$$E^{(2)}(\Gamma_2,\Gamma_v)=z, \quad E^{(n)}(\Gamma_2,\Gamma_v)=x_+^n-x_-^n,$$

$$E^{(r-4)}(\Gamma_r,\Gamma_v)=\begin{pmatrix}x_+^{r-4}\\ x_-^{r-4}\end{pmatrix},$$

$$E^{(n-r+5)}(\Gamma_r,\Gamma_v)=z\begin{pmatrix}x_-^{n-r+4}\\ -x_+^{n-r+4}\end{pmatrix},$$

$r=5,6,\ldots,n/2+3$.

In the case of $C_n\,[D_n$ the z coordinate transforms by Γ_1 and x,y transform by Γ_3 for n odd and by Γ_5 for n even. The integrity bases are for n odd:

$$I^{(1)}\Gamma_v=z, \quad I^{(2)}(\Gamma_v)=x_+x_-=x^2+y^2,$$
$$I^{(n)}(\Gamma_v)=x_+^n-x_-^n,$$

$$E(\Gamma_2,\Gamma_v)=x_+^n+x_-^n, \quad E^{(r-2)}(\Gamma_r,\Gamma_v)=\begin{pmatrix}x_+^{r-2}\\ (-x_-)^{r-2}\end{pmatrix}, \tag{78}$$

$$E^{(n-r+2)}(\Gamma_r,\Gamma_v)=\begin{pmatrix}x_-^{n-r+2}\\ (-x_+)^{n-r+2}\end{pmatrix},$$

for $r=3,4,\ldots,(n+3)/2$,

and for n even:

$$I^{(1)}(\Gamma_v)=z, \quad I^{(2)}(\Gamma_v)=x_+x_-, \quad I^{(n)}=x_+^n+x_-^n,$$

$$E^{(n)}(\Gamma_2,\Gamma_v)=x_+^n-x_-^n,$$
$$E(\Gamma_3,\Gamma_v)=x_+^{n/2}+(-x_-)^{n/2}, \tag{79}$$
$$E(\Gamma_4,\Gamma_v)=x_+^{n/2}-(-x_-)^{n/2},$$

$$E^{(r-4)}(\Gamma_r,\Gamma_v)=\begin{pmatrix}x_+^{r-4}\\ (-x_-)^{r-4}\end{pmatrix},$$

$$E^{(n-r+4)}(\Gamma_r,\Gamma_v)=\begin{pmatrix}x_-^{n=r+4}\\ (-x_+)^{n-r+4}\end{pmatrix},$$

for $r=5,6,\ldots,n/2+3$.

The integrity bases for $D_n\,[D_{2n}$, n odd, are

$$I^{(2)}(\Gamma_v)=z^2, \quad I^{(2)}(\Gamma_v)=x_+x_-,$$
$$I^{(n)}(\Gamma_v)=x_+^n+x_-^n,$$

$$E^{(n)}(\Gamma_2,\Gamma_v)=x_+^n-x_-^n,$$
$$E^{(n+1)}(\Gamma_3,\Gamma_v)=z(x_+^n-x_-^n), \quad E^{(1)}(\Gamma_4,\Gamma_5)=z, \tag{80}$$

$$E^{(r-4)}(\Gamma_r,\Gamma_v)=\begin{pmatrix}x_+^{r-4}\\ x_-^{r-4}\end{pmatrix},$$

$$E^{(n-r+4)}(\Gamma_r,\Gamma_v)=\begin{pmatrix}x_-^{n-r+4}\\ x_+^{n-r+4}\end{pmatrix},$$

where r is even and $6\leqslant r\leqslant n+3$; for r odd and $5\leqslant r\leqslant n+2$ we have

$$E^{(r-3)}(\Gamma_r,\Gamma_v)=z\begin{pmatrix}x_+^{r-4}\\ -x_-^{r-4}\end{pmatrix},$$

$$E^{(n-r+3)}(\Gamma_r,\Gamma_v)=z\begin{pmatrix}x_-^{n-r+4}\\ -x_+^{n-r+4}\end{pmatrix}. \tag{81}$$

When n is even, the integrity bases are

$$I^{(2)}(\Gamma_v)=z^2, \quad I^{(2)}(\Gamma_v)=x_+x_-, \quad I^{(2n)}(\Gamma_v)=x_+^{2n}+x_-^{2n},$$

$$E^{(n+1)}(\Gamma_1,\Gamma_v)=z(x_+^n-x_-^n),$$
$$E^{(2n)}(\Gamma_2,\Gamma_v)=x_+^{2n}-x_-^{2n},$$
$$E^{(n+1)}(\Gamma_2,\Gamma_v)=z(x_+^n+x_-^n),$$
$$E^{(n)}(\Gamma_3,\Gamma_v)=x_+^n+x_-^n, \tag{82}$$
$$E^{(2n+1)}(\Gamma_3,\Gamma_v)=z(x_+^{2n}-x_-^{2n}),$$
$$E^{(n)}(\Gamma_4,\Gamma_v)=x_+^n-x_-^n,$$
$$E^{(2n+1)}(\Gamma_4,\Gamma_v)=z(x_+^{2n}+x_-^{2n}).$$

Furthermore for r even and such that $6\leqslant r\leqslant n+2$ we have

$$E^{(r-4)}(\Gamma_r,\Gamma_v)=\begin{pmatrix}x_+^{r-4}\\ x_-^{r-4}\end{pmatrix},$$

$$E^{(2n-r+4)}(\Gamma_r,\Gamma_v)=\begin{pmatrix}x_-^{2n-r+4}\\ x_+^{2n-r+4}\end{pmatrix},$$

$$\tag{83}$$

$$E^{(n+r-3)}(\Gamma_r,\Gamma_v)=z\begin{pmatrix}x_+^{n+r-4}\\ -x_-^{n+r-4}\end{pmatrix},$$

$$E^{(n-r+5)}(\Gamma_r,\Gamma_v)=z\begin{pmatrix}x_-^{n-r+4}\\ -x_+^{n-r+4}\end{pmatrix},$$

and for r odd such that $5\leqslant r\leqslant n+3$ we get

$$E^{(r-3)}(\Gamma_r,\Gamma_v)=z\begin{pmatrix} x_+^{r-4} \\ -x_-^{r-4}\end{pmatrix},$$

$$E^{(n+r-4)}(\Gamma_r,\Gamma_v)=\begin{pmatrix} x_+^{n+r-4} \\ x_-^{n+r-4}\end{pmatrix},$$

$$E^{(n-r+4)}(\Gamma_r,\Gamma_v)=\begin{pmatrix} x_-^{n-r+4} \\ x_+^{n-r+4}\end{pmatrix},$$

$$E^{(2n-r+5)}(\Gamma_r,\Gamma_v)=z\begin{pmatrix} x_-^{2n-r+4} \\ -x_+^{2n-r+4}\end{pmatrix}. \tag{84}$$

Finally, for the groups $C_n\times P$ and $D_n\times P$ the integrity bases follow immediately from those of C_n and D_n and (55) where we put $\Gamma_m^o=\Gamma_v$.

VI. POINT GROUP POLYNOMIAL BASES FOR IRREDUCIBLE REPRESENTATIONS OF SO(3)

The (Γ_r,Γ_v) tensors constructed in the preceding section transform irreducibly by the representations Γ_r of G, and their components are homogeneous polynomials in x,y,z, coordinates of a vector from a three 3-dimensional space R_3. In this section we want to relate these (Γ_r,Γ_v) tensors to irreducible representations of the rotation group SO(3), using the fact that the three-dimensional space on which Γ_v acts is spanned by x,y,z.

Polynomials in x,y,z of degree L ($L>1$) span a $\frac{1}{2}(L+1)(L+2)$-dimensional space $\{R_3\}^L$ which is reducible with respect to the group SO(3). More precisely,

$$\{R_3\}^L=R_{2L+1}\oplus R_{2L-3}\oplus R_{2L-7}\oplus\cdots, \tag{85}$$

where R_d is a space of dimension d irreducible with respect to SO(3).

Our task is to identify among the (Γ_r,Γ_v) tensors of degree L those which belong to R_{2L+1}. The methods of Secs III–V are readily adapted to this problem. First we find the generating function $C(\Gamma_r,\Gamma_1;\lambda)$ for the multiplicities of Γ_r in the SO(3) representations L acting in the space R_{2L+1}, and then we construct the basis tensors.

In order to find $C(\Gamma_r,\Gamma_v;\lambda)$ it suffices to notice that R_{2L+1} is the only subspace on the right side of (85) which does not contain polynomials with the SO(3) scalar $r^2=x^2+y^2+z^2$ as a factor. Hence $C(\Gamma_r,\Gamma_v;\lambda)$ is obtained from $B(\Gamma_r,\Gamma_v;\lambda)$ by eliminating the factor $(1-\lambda^2)^{-1}$ responsible for the scalar r^2 [an SO(3)-scalar is also an G-scalar]. Thus,

$$C(\Gamma_r,\Gamma_v;\lambda)=(1-\lambda^2)B(\Gamma_r,\Gamma_v;\lambda). \tag{86}$$

An explicit form of $C(\Gamma_r,\Gamma_v;\lambda)$ for the groups T, O, and I is found from (56) and Tables V, VI, and VII respectively; for C_n it suffices to put $\lambda_1=\lambda_2=\lambda_n=\lambda$ in (57), and for D_n to put $\lambda_2=\lambda_3=\lambda$ in (58), (59) or $\lambda_2=\lambda_5=\lambda$ in (60). In the expansion of (86),

$$C(\Gamma_r,\Gamma_v;\lambda)\ =\sum_{L=0}^{\infty} N_{rL}\lambda^L, \tag{87}$$

the coefficient N_{rL} of λ^L is the multiplicity of Γ_r in the SO(3) representation L.

The tensors implied by the generating function $C(\Gamma_r,\Gamma_v;\lambda)$ provide polynomial bases for irreducible representations of SO(3) reduced according to G. The integrity basis is conveniently constructed from those of Sec. V. The construction proceeds in two steps. First the SO(3) scalar r^2 is eliminated from the integrity basis; the remaining tensors are then in one-to-one correspondence with the required integrity basis. However, the tensors generated by the integrity basis may contain admixtures of the lower subspaces of (85). The second step consists of projecting out the unwanted part of each tensor. For T, O, I the scalar $I^{(2)}(\Gamma_v)=r^2$ is straightforwardly dropped from the integrity basis. For D_n, we drop $I_z^{(2)}(\Gamma_v)=z^2$. For C_n we want to eliminate even powers of z. This may be done by replacing $I^{(1)}(\Gamma_v)=z$ by $E^{(1)}=(\Gamma_1,\Gamma_v)=z$; the difference is that now z may appear only linearly.

The tensors generated by the modified integrity basis are in general linear combinations of terms belonging to the different subspaces of (85). So far we have insured only that each contains a nonzero part from R_{2L+1}. In order to project out the admixture of lower subspaces we use the traceless boson operator technique introducd by Lohe.[9] In our context this means making the substitution

$$\bar{r}\to\bar{r}'=\bar{r}-r^2(2N+3)^{-1}\bar{\nabla}, \tag{88}$$

where $N=\bar{r}\cdot\bar{\nabla}$ is the operator whose eigenvalue is the degree of the polynomial on which it acts. The components of $\bar{r}'$ mutually commute. Polynomials of degree L in x',y',z' satisfy Laplace's equation and transform irreducibly by the representation L of SO(3).

VII. CONCLUDING REMARKS

The generating functions serve here as an economic first step in constructing the integrity bases They are also of interest in their own right; they show the number of elements of the integrity basis, their degrees, and show which invariants are of type I and which of type E. We have calculated the generating functions for all pairs of irreducible representations of all finite subgroups of O(3), and given composition rules for constructing generating functions for reducible representations.

There are a number of possibilities for extending this work, apart from direct applications; let us mention some of them.

The problems studied here for point groups are interesting also for double groups, space groups, and discrete groups in general.

Suppose G is a continuous group with a finite subgroup G'. It is of interest to know G' tensors whose components are polynomials in the components of an irreducible G tensor. An expansion of the corresponding generating function provides the multiplicities of G' tensors of each degree. Further, it would be interesting to construct generating functions $D(\Gamma_r;\mu_1,\ldots,\mu_r)$ where the coefficient of $\mu_1^{a_1},\ldots,\mu_r^{a_r}$ in the expansion of D is the multiplicity of the irreducible representation Γ_r or G' in the irreducible representation $(a_1,...,a_r)$ of G.

Generating functions and integrity bases can be found and have proven useful for representations of continuous groups. Although the methods are general, they have been applied only to a few particular problems.[10]

ACKNOWLEDGMENTS

We thank P. Desmier, L. Michel, and H. Zassenhaus for helpful discussions and J. McKay for kindly reading the manuscript.

APPENDIX

The construction of the integrity bases for $(\Gamma_n \Gamma_m)$ tensors in explicit form requires a definite choice of the generator matrices for each irreducible representaiton. In order to make the paper self-contained we list them in Table IX for representations of D_n, T, and O of dimension greater than unity; for a one-dimensional representation the matrix is just the character.

[1]M. Lax *Symmetry Principles in Solid State and Molecular Physics* (Wiley–Interscience, New York, 1974); L.S. Lomont, *Applications of Finite Groups* (Academic New York, 1959); W.A. Wooster, *Crystal Physics* (Cambridge U.P., Cambridge, 1949).
[2]H. Bethe, Ann. Phys. **3**, 133 (1929).
[3]F.C. von der Lage and H. Bethe, Phys. Rev. **71**, 612 (1947).
[4]D.G. Bell, Rev. Mod. Phys. **26**, 311 (1954); S.L. Altmann, Proc. Cambridge Philos. Soc. **53**, 343 (1957); N.V. Cohan, Proc. Cambridge Philos. Soc. **54**, 28 (1958); W. Döring, Ann. Phys. **1**, 102 (1958); W. Döring and G. Simon, Ann. Phys. **5**, 373 (1960); S.L. Altmann and C.J. Bradley, Phil. Trans. R.S. London, Ser. A **255**, 199 (1962); S.L. Altmann, Rev. Mod. Phys. **35**, 641 (1963); G.F. Smith, M.M. Smith, and R.S. Rivlin, Arch. Ration. Mech. Anal. **12**, 93 (1963); G.F. Smith and R.S. Rivlin, Arch. Ration. Mech. Anal. **15**, 169 (1964); S.L. Altmann and A.P. Cracknell, Rev. Mod. Phys. **37**, 19 (1965); J. Killingbeck, J. Phys. C **5**, 2497 (1972); V. Kopský, J. Phys. C **8**, 3251 (1975); R.M. Nunes-Mendes, Trans. Amer. Math. Soc. **204**, 161 (1975), J. Patera and P. Winternitz, J. Chem. Phys. **65**, 2725 (1975); R.P. Bickerstaff and B.G. Wybourne, J. Phys. A **9**, 1051 (1976); M.W. Jarić and J.L. Birman, J. Math. Phys. **18**, 1456, 1459, 2085E (1977); N.J.A. Sloane, Amer. Math. Month. **84**, 82 (1977); V. Kopský and D.G. Sannikov, J. Phys. C **10**, 4347 (1977).
[5]B. Meyer, Can. J. Math. **6**, 135 (1954).
[6]A.G. McLellan, J. Phys. C **7**, 3326 (1974).
[7]W. Burnside, *Theory of Groups of Finite Order* (Dover, New York, 1955), 2nd ed.
[8]T. Molien, Sitzungber. König. Preuss. Akad. Wiss. 1152 (1897)– G.A. Miller, H.F. Blichfeldt, and L.E. Dickson, *Theory and Application of Finite Groups* (Wiley, New York, 1916, 2nd ed. (1938).
[9]M.A. Lohe, Ph.D. thesis University of Adelaide, 1974; see also M.A. Lohe and C.A. Hurst, J. Math. Phys. **12**, 1882 (1971).
[10]B.R. Judd, W. Miller Jr., J. Patera, and P. Winternitz, J. Math. Phys **15**, 1787 (1974); R.T. Sharp, J. Math. Phys. **16**, 2050 (1975); W. McKay, J. Patera, and R.T. Sharp. Comput. Phys. Commun. **10**, 1 (1975); C. Quesne, J. Math. Phys. **17**, 1452 (1976); **18**, 1210 (1977); R. Gaskell, A. Peccia, and R.T. Sharp, J. Math. Phys. **19**, 727 (1978).

Generalized Hamilton–Jacobi theories

Richard Baumeister

Department of Mathematics, Arizona State University, Tempe, Arizona 85281
(Received 28 February 1978)

The generalized momenta of a dynamical system with n degrees of freedom may be given a Clebsch representation in terms of independent scalar functions and gradients thereof. The canonical equations imply certain differential relations which must be satisfied by these scalars. If the dynamical system is nonrelativistic, the differential relations are shown to be a generalization of the classical Hamilton–Jacobi theory. Similar results are obtained if the dynamical system is relativistic. A multiple integral variational principle, whose Euler–Lagrange equations imply the equations of motion and an equation of continuity, is formulated. As an example, it is shown that the Einstein field equations for an incoherent matter field and the geodesic equations of motion may be derived from a single fourfold variational principle. The usual energy-momentum tensor for such a matter field emerges as a by-product of the variational principle.

1. INTRODUCTION

We shall initially be concerned with a nonrelativistic holonomic dynamical system with n degrees of freedom, defined by a Hamiltonian function $H(t,x^h,p_h)$, $h=1,\cdots,n$, where t denotes the time, x^h the local generalized coordinates, and p_h the generalized momenta of the system. In the parameter dependent Hamilton–Jacobi theory of Carathéodory,[1] it is assumed that a canonical momentum field p_h is the gradient of a scalar function $S(t,x^h)$. The function S must then satisfy the first order partial differential equation known as the "Hamilton–Jacobi equation." Conversely, if S is a solution of the Hamilton-Jacobi equation, the gradient of S will define a canonical momentum field.

The assumption that a momentum field p_h be a gradient is a rather severe restriction, in that it implies that the curl of p_h vanish identically. Recently Rund[2] has developed a generalized parameter dependent Hamilton–Jacobi theory which does not suffer from this drawback. In Rund's theory the field p_h is given a so-called *Clebsch representation* in terms of n independent scalar functions which are referred to as *the Clebsch potentials*. This implies that the components $p_1,\cdots,p_n$ are independent functions of x^h. If p_h is canonical, it is shown that the potentials must satisfy a system of ordinary differential equations which, because they possess a structure identical with that of the Hamiltonian canonical equations, are called the *associated canonical equations*. In addition, the potentials must satisfy a Hamilton–Jacobi type partial differential equation which is referred to as *the generalized Hamilton–Jacobi equation*. Conversely, if a given set of Clebsch potentials satisfy the generalized Hamilton–Jacobi equation and the associated canonical equations, then the potentials (via the Clebsch representation) define a canonical field.

In this paper we shall show that the generalized Hamilton–Jacobi theory of Rund is valid, irrespective of any assumptions concerning the dependence or otherwise of $p_1,\cdots,p_n$. This is accomplished by considering the nature of the Pfaffian[3] $\omega=p_h dx^h$. According to the solution[4] of the "Problem of Pfaff," ω may be reduced to a Pfaffian expression which involves s independent functions, where $1\leqslant s\leqslant n$. It follows that p_h may be given a Clebsch representation in terms of s independent functions. The number s, which depends upon the nature of the field p_h, shall be called *the character* of p_h. If the character of p_h is one, p_h has the gradient representation of the ordinary Hamilton–Jacobi theory, whereas if the character of p_h is n, p_h must have the representation assumed by Rund. The generalized Hamilton–Jacobi theory presented here encompasses both of the above theories, since it is valid for fields of arbitrary character.

A field p_h does not determine a unique set of Clebsch potentials, and any two sets of potentials which represent the same field are said to be related by means of a *Clebsch gauge transformation*.[5] The relationship between Clebsch gauge transformations and canonical transformations of the Clebsch potentials is investigated in Sec. 4. When the character of p_h is odd, it is shown that any such canonical transformation coincides with a Clebsch gauge transformation. If the character of p_h is even, a canonical transformation induces a representation of p_h in terms of a dependent set of potentials. In both cases a particularly simple form of the generalized theory occurs if the integrals of the associated canonical equations are chosen as Clebsch potentials.

Next, we shall consider a relativistic holonomic dynamical system with n degrees of freedom. Since the time t is no longer a preferred parameter, the canonical formalism which describes this system should be based upon a parameter invariant variational principle. Accordingly, we shall develop a generalized Hamilton–Jacobi theory within the framework of the so-called *homogeneous* canonical formalism.[6] A canonical momentum field with character s is given a Clebsch representation in terms of s independent Clebsch potentials. It is shown that the potentials must satisfy an associated system of parameter invariant canonical equations, together with a generalized Hamilton–Jacobi equation, and conversely, these equations are sufficient to ensure that the potentials define a canonical field. Also, in contrast with the parameter dependent case, the associated parameter invariant canonical equations are shown to imply the corresponding generalized Hamilton–Jacobi equation.

An important application of the generalized Hamilton–Jacobi theories is that they allow one to obtain the single integral canonical equations of motion from a multiple integral variational principle. This is accomplished in

 0022-2488/78/1911-2377$1.00

Sec. 6, for the parameter invariant formalism, by constructing a Lagrange density $\mathscr{L}$ in which the Clebsch potentials are treated as field variables. The Lagrange density must also, for dimensional reasons, contain a density function ν as a linear factor, and ν is likewise treated as a field variable. It is shown that the Euler–Lagrange equations which result from varying the Clebsch potentials and the density ν, imply the parameter invariant canonical equations and an equation of continuity for ν.

For certain physical fields, the procedure outlined above allows us to derive the field equations and the particle equations of motion from a single, n-fold variational principle. An example of this phenomenon is given at the end of Sec. 6, where the Einstein field equations and the geodesic equations of motion for an incoherent matter field are derived from a 4-fold variational principle.[7]

2. THE CLEBSCH REPRESENTATION OF A COVARIANT VECTOR FIELD

We shall let x^h, $h=1,\cdots,n$, denote the local coordinates of an n-dimensional differentiable manifold X_n. If p_h is a class C^2 covariant vector field on X_n, we may construct the invariant 1-form

$$\omega=p_h dx^h. \tag{2.1}$$

Now according to the solution of the classical "Problem of Pfaff,"[4] ω has one of two possible Pfaffian reductions. Either ω may be written as

$$\omega=d\psi+P_1 dQ^1+\cdots+P_m dQ^m, \tag{2.2}$$

where $(\psi,Q^1,\cdots,Q^m,P_1,\cdots,P_m)$ are $(2m+1)$ independent functions of x^h, or ω is expressible in terms of merely $2m$ independent functions of x^h, namely

$$\omega=P_1 dQ^1+\cdots+P_m dQ^m. \tag{2.3}$$

If ω has the reduction (2.2), the vector field p_h is said to have *character* $(2m+1)$ whereas if p_h has the reduction (2.3) the *character*[8] of p_h is $2m$. The character s, which depends upon the nature of the given field, may be shown[4] to be equal to the rank of the $n\times(n+1)$ augmented matrix

$$\left(\frac{\partial p_h}{\partial x^l}-\frac{\partial p_l}{\partial x^h},\,p_k\right)=(\omega_{lh},p_k),$$

where, for the sake of brevity, we have put

$$\omega_{lh}\equiv\frac{\partial p_h}{\partial x^l}-\frac{\partial p_l}{\partial x^h}. \tag{2.4}$$

It follows that $s\leqslant n$.

It will prove convenient to introduce the notation

$$\epsilon=\begin{cases}0, & \text{if the character of } p_h \text{ is even,}\\ 1, & \text{if the character of } p_h \text{ is odd.}\end{cases} \tag{2.5}$$

Then, if we let the index α run from 1 to m, the representations (2.2) and (2.3) may be combined into the concise form[9]

$$\omega=\epsilon d\psi+P_\alpha dQ^\alpha. \tag{2.6}$$

Upon comparing (2.6) with our original expression (2.1) for ω, we obtain

$$p_h=\epsilon\frac{\partial\psi}{\partial x^h}+P_\alpha\frac{\partial Q^\alpha}{\partial x^h}, \tag{2.7}$$

which gives a representation of p_h in terms of the $(2m+\epsilon)$ scalar functions $(\epsilon\psi,Q^\alpha,P_\alpha)$. In the special case of $n=3$, $m=1$, and $\epsilon=1$, (2.7) reduces to

$$\mathbf{p}=\nabla\psi+P\nabla Q$$

which is the representation usually attributed to Clebsch.[10] For this reason, (2.7) will be called a *Clebsch representation* of p_h, and the scalars $(\epsilon\psi,Q^\alpha,P_\alpha)$ shall be referred to as the *Clebsch potentials.*

It is obvious that a given vector field does not determine a unique Clebsch representation. Following Rund,[5] we shall call a transformation of the form

$$\epsilon\bar\psi=\epsilon\bar\psi(\epsilon\psi,Q^\beta,P_\beta),\quad \bar Q^\alpha=\bar Q^\alpha(Q^\beta,P_\beta),$$
$$\bar P_\alpha=\bar P_\alpha(Q^\beta,P_\beta), \tag{2.8}$$

a *Clebsch gauge transformation if*

$$\epsilon\frac{\partial\psi}{\partial x^h}+P_\alpha\frac{\partial Q^\alpha}{\partial x^h}=\epsilon\frac{\partial\bar\psi}{\partial x^h}+\bar P_\alpha\frac{\partial\bar Q^\alpha}{\partial x^h}.$$

That is, the potentials $(\epsilon\psi,Q^\alpha,P_\alpha)$ and $(\epsilon\bar\psi,\bar Q^\alpha,\bar P_\alpha)$ give rise to equivalent representations of the same vector field. Quantities which are invariant under Clebsch gauge transformations shall be called *Clebsch gauge invariants.*

In this paper we shall identify the manifold X_n with the configuration space of a holonomic, dynamical system with n degrees of freedom. The vector field p_h is identified with the canonical momentum of this dynamical system. If the character of p_h is one, the representation (2.7) reduces to

$$p_h=\frac{\partial\psi}{\partial x^h}, \tag{2.9}$$

which is just the representation for fields of momenta in the ordinary, parameter dependent Hamilton–Jacobi theory of Carathéodory.[1] If the character of p_h is n, (2.7) becomes

$$p_h=\epsilon\frac{\partial\psi}{\partial x^h}+\sum_{\alpha=1}^{(n-\epsilon)/2}P_\alpha\frac{\partial Q^\alpha}{\partial x^h} \tag{2.10}$$

which is the representation assumed by Rund[2] in his parameter dependent generalized Hamilton–Jacobi theory. However, an arbitrary momentum field will have character s, $1\leqslant s\leqslant n$.

3. A GENERALIZED HAMILTON–JACOBI THEORY FOR NONRELATIVISTIC DYNAMICAL SYSTEMS

We shall now consider the $(n+1)$-dimensional differentiable manifold $Y_{n+1}\equiv R\times X_n$ of the variables (t,x^h), where t denotes the time and the x^h represent the generalized coordinates of a nonrelativistic, holonomic, dynamical system with n degrees of freedom. This dynamical system is described by a given Hamiltonian function $H(t, x^l, p_l)$ where p_l represents the components of a momentum vector. For a given momentum field $p_h(t, x^l)$, we may define a conjugate velocity field $v^h(t, x^l)$ by means of

$$v^h(t,x^l) \equiv \frac{\partial H}{\partial p_h}[t, x^l, p_l(t,x^k)]. \qquad (3.1)$$

We shall assume that

$$\det\left(\frac{\partial v^h}{\partial p_k}\right) = \det\left(\frac{\partial^2 H}{\partial p_h \partial p_k}\right) \neq 0, \qquad (3.2)$$

for all possible arguments (t, x^h, p_h), so that (3.1) establishes a one-to-one correspondence between the velocity and momentum fields. In particular, (3.2) implies that H is not homogeneous of degree one in the variables p_h; hence the canonical formalism based upon the assumption (3.2) is described as "the nonhomogeneous theory".[11]

In our analysis we shall frequently be concerned with the differentiation of functions of the form

$$F = F[t, x^l, f_A(t, x^k)], A: 1, \cdots, M,$$

where the f_A represent M differentiable functions, and F is also differentiable. In such cases it is of crucial importance to distinguish between the partial derivative and the so-called "total derivative." If $x^h = x^h(t)$ represents a curve C in X_n, the total derivative of F with respect to t along C is defined by

$$\frac{dF}{dt} \equiv \frac{\partial F}{\partial t} + \frac{\partial F}{\partial x^l}\dot{x}^l + \sum_{A=1}^{M} \frac{\partial F}{\partial f_A} \times \left(\frac{\partial f_A}{\partial t} + \frac{\partial f_A}{\partial x^l}\dot{x}^l\right), \qquad (3.3)$$

where $\dot{x}^l \equiv dx^l/dt$. Similarly, the relations

$$f_A = f_A(t, x^l)$$

define a surface S of dimension $(n+1)$ in the space of the variables (t, x^h, f_A), and the total derivative of F with respect to x^h along S is defined by

$$\frac{dF}{dx^h} \equiv \frac{\partial F}{\partial x^h} + \sum_{A=1}^{M} \frac{\partial F}{\partial f_A}\frac{\partial f_A}{\partial x^h}. \qquad (3.4)$$

A curve $x^h = x^h(t)$ in X_n is said to be a *streamline* of the velocity field (3.1) if the functions $x^h(t)$ satisfy the equations

$$\dot{x}^h = v^h \equiv \frac{\partial H}{\partial p_h}. \qquad (3.5)$$

Furthermore, the field $p_h(t, x^l)$ is said to be *canonical* if

$$\frac{dp_h}{dt} = \frac{-\partial H}{\partial x^h}, \qquad (3.6)$$

along every streamline of v^h, that is if (3.6) holds as a result of (3.5). The $2n$ differential equations

$$\dot{x}^h = \frac{\partial H}{\partial p_h}, \quad \frac{dp_h}{dt} = -\frac{\partial H}{\partial x^h} \qquad (3.7)$$

are called *the canonical equations*. We shall assume that the temporal evolution of our dynamical system is determined by (3.7). As is well known, this is equivalent to Hamilton's principle as applied to the Lagrange function

$$L(t, x^h, \dot{x}^h) = -H + p_h \dot{x}^h, \qquad (3.8)$$

where it is understood that (3.2) has been used in order to express the p_h in terms of $\dot{x}^h$ in the right-hand side of (3.8).

Let us suppose that we are given a canonical field $p_h(t, x^l)$. Then (3.6), when taken in conjunction with (3.5), yields

$$\frac{\partial p_h}{\partial t} + \frac{\partial p_h}{\partial x^l}\dot{x}^l - \frac{\partial p_l}{\partial x^h}\dot{x}^l = -\frac{\partial H}{\partial x^h} - \frac{\partial H}{\partial p_l}\frac{\partial p_l}{\partial x^h},$$

which, by virtue of the notation introduced in (2.4) and (3.4), may be rewritten as

$$\frac{\partial p_h}{\partial t} + \omega_{lh}\dot{x}^l = -\frac{dH}{dx^h}. \qquad (3.9)$$

We will presently show that (3.9) gives rise to some interesting identities which involve the Clebsch potentials of p_h.

First let us note that since p_h may depend explicitly upon t, the Clebsch potentials $(\epsilon\psi, Q^\alpha, P_\alpha)$ of p_h may also have explicit t dependence. Next, the representation (2.7) implies that

$$\frac{\partial p_h}{\partial t} = \frac{\partial}{\partial x^h}\left(\epsilon\frac{\partial \psi}{\partial t} + P_\alpha \frac{\partial Q^\alpha}{\partial t}\right) + \frac{\partial P_\alpha}{\partial t}\frac{\partial Q^\alpha}{\partial x^h} - \frac{\partial Q^\alpha}{\partial t}\frac{\partial P_\alpha}{\partial x^h}, \qquad (3.10)$$

whereas (2.7) and (2.4) yield

$$\omega_{lh}\dot{x}^l = \frac{\partial Q^\alpha}{\partial x^h}\left(\frac{dP_\alpha}{dt} - \frac{\partial P_\alpha}{\partial t}\right) - \frac{\partial P_\alpha}{\partial x^h} \times \left(\frac{dQ^\alpha}{dt} - \frac{\partial Q^\alpha}{\partial t}\right). \qquad (3.11)$$

Substitution of (3.10) and (3.11) into (3.9) gives

$$\frac{\partial}{\partial x^h}\left(\epsilon\frac{\partial \psi}{\partial t} + P_\alpha \frac{\partial Q^\alpha}{\partial t}\right) + \frac{dP_\alpha}{dt}$$

$$\times \frac{\partial Q^\alpha}{\partial x^h} - \frac{dQ^\alpha}{dt}\frac{\partial P_\alpha}{\partial x^h} = -\frac{dH}{dx^h}. \quad (3.12)$$

The form of the above relation suggests that we define a function $T(t,x^l)$ by

$$T(t,x^l) \equiv H\left(t,x^l,\epsilon\frac{\partial\psi}{\partial x^l} + P_\alpha\frac{\partial Q^\alpha}{\partial x^l}\right) + \epsilon\frac{\partial\psi}{\partial t} + P_\alpha\frac{\partial Q^\alpha}{\partial t}. \quad (3.13)$$

Then (3.12) becomes

$$\frac{dT}{dx^h} = \frac{dQ^\alpha}{dt}\frac{\partial P_\alpha}{\partial x^h} - \frac{dP_\alpha}{dt}\frac{\partial Q^\alpha}{\partial x^h}. \quad (3.14)$$

We shall now take advantage of the fact that the Clebsch potentials in the representation (2.7) are independent functions of x^h. Because of this, it is possible to solve the $(2m+\epsilon)$ equations

$$\epsilon\psi = \epsilon\psi(t,x^h), \quad Q^\alpha = Q^\alpha(t,x^h), \quad P_\alpha = P_\alpha(t,x^h)$$

and obtain, upon relabeling if necessary, the variables $(x^1,\dots,x^{2m+\epsilon})$ in terms of the potentials $(\epsilon\psi,Q^\alpha,P_\alpha)$. In fact, there will exist a one-to-one correspondence between the coordinates $(t,x^1,\dots,x^n)$ and the variables $(t,\epsilon\psi,Q^1,\dots,Q^m,P_1,\dots,P_m,x^{2m+1+\epsilon},\dots,x^n)$. This correspondence allows us to define a function Φ by means of

$$\Phi(t,\epsilon\psi,Q^\alpha,P_\alpha,x^{2m+1+\epsilon},\dots,x^n) \equiv T(t,x^l). \quad (3.15)$$

Substitution of (3.15) into (3.14) then gives, for $h=1,\dots,2m+\epsilon$,

$$\epsilon\frac{\partial\Phi}{\partial\psi}\frac{\partial\psi}{\partial x^h} + \frac{\partial\Phi}{\partial Q^\alpha}\frac{\partial Q^\alpha}{\partial x^h} + \frac{\partial\Phi}{\partial P_\alpha}\frac{\partial P_\alpha}{\partial x^h} = \frac{dQ^\alpha}{dt}\frac{\partial P_\alpha}{\partial x^h} - \frac{dP_\alpha}{dt}\frac{\partial Q^\alpha}{\partial x^h}, \quad (3.16)$$

together with

$$\frac{\partial\Phi}{\partial x^h} + \epsilon\frac{\partial\Phi}{\partial\psi}\frac{\partial\psi}{\partial x^h} + \frac{\partial\Phi}{\partial Q^\alpha}\frac{\partial Q^\alpha}{\partial x^h} + \frac{\partial\Phi}{\partial P_\alpha}\frac{\partial P_\alpha}{\partial x^h} = \frac{dQ^\alpha}{dt}\frac{\partial P_\alpha}{\partial x^h} - \frac{dP_\alpha}{dt}\frac{\partial Q^\alpha}{\partial x^h}, \quad (3.17)$$

for $h=2m+\epsilon+1,\dots,n$.

The identity (3.16) implies that, for $h=1,\dots,2m+\epsilon$,

$$\epsilon\frac{\partial\Phi}{\partial\psi}\frac{\partial\psi}{\partial x^h} + \left(\frac{\partial\Phi}{\partial P_\alpha} - \frac{dQ^\alpha}{dt}\right)\frac{\partial P_\alpha}{\partial x^h} + \left(\frac{\partial\Phi}{\partial Q^\alpha} + \frac{dP_\alpha}{dt}\right)\frac{\partial Q^\alpha}{\partial x^h} = 0.$$

Since the functions $(\epsilon\psi,Q^\alpha,P_\alpha)$ are independent, we may conclude that

$$\epsilon\frac{\partial\Phi}{\partial\psi} = 0,$$

together with

$$\frac{dQ^\alpha}{dt} = \frac{\partial\Phi}{\partial P_\alpha}, \quad \frac{dP_\alpha}{dt} = \frac{-\partial\Phi}{\partial Q^\alpha}. \quad (3.18)$$

As a result of the above, (3.17) reduces to

$$\frac{\partial\Phi}{\partial x^h} = 0, \quad h = 2m+1+\epsilon,\dots,n. \quad (3.19)$$

Thus the definition (3.15) may be replaced by

$$\Phi(t,Q^\alpha,P_\alpha) = T(t,x^l). \quad (3.20)$$

Because the relations contained in (3.18) possess the same structure as the canonical equations (3.7), (3.18) shall be referred to as the *associated canonical equations*. The function $\Phi(t,Q^\alpha,P_\alpha)$ which occurs in (3.18) shall be called a *superpotential*. We have shown that a superpotential may be defined by (3.13) and (3.20) for any canonical field with the representation (2.7).

The relations (3.13) and (3.20) imply that

$$\Phi(t,Q^\alpha,P_\alpha) = H\left(t,x^l,\epsilon\frac{\partial\psi}{\partial x^l} + P_\alpha\frac{\partial Q^\alpha}{\partial x^l}\right) + \epsilon\frac{\partial\psi}{\partial t} + P_\alpha\frac{\partial Q^\alpha}{\partial t}, \quad (3.21)$$

which, for a given function Φ, may be interpreted as a first order partial differential equation in the variables $(\epsilon\psi,Q^\alpha,P_\alpha)$. In particular, if p_h has the representation (2.9) and if Φ is zero, (3.21) reduces to

$$\frac{\partial\psi}{\partial t} + H\left(t,x^l,\frac{\partial\psi}{\partial x^l}\right) = 0, \quad (3.22)$$

which is just the usual Hamilton–Jacobi equation for the so-called principal function ψ. For this reason, (3.21) shall be referred to as the *generalized Hamilton–Jacobi equation*.

In order to examine the relationship between the generalized Hamilton–Jacobi equation and the associated canonical equations, let us take the total derivative of the expression

$$H\left(t,x^l,\epsilon\frac{\partial\psi}{\partial x^l} + P_\alpha\frac{\partial Q^\alpha}{\partial x^l}\right) + \epsilon\frac{\partial\psi}{\partial t} + P_\alpha\frac{\partial Q^\alpha}{\partial t} - \Phi(t,Q^\beta,P_\beta)$$

with respect to x^h. This gives, after some manipulation,

$$\frac{d}{dx^h}\Big[H\Big(t,x^l,\epsilon\frac{\partial\psi}{\partial x^l}+P_\alpha\frac{\partial Q^\alpha}{\partial x^l}\Big)+\epsilon\frac{\partial\psi}{\partial t}+P_\alpha\frac{\partial Q^\alpha}{\partial t}-\Phi(t,Q^\beta,P_\beta)\Big]$$
$$=\frac{dp_h}{dt}+\frac{\partial H}{\partial x^h}+\frac{\partial P_\alpha}{\partial x^h}\Big(\frac{dQ^\alpha}{dt}-\frac{\partial\Phi}{\partial P_\alpha}\Big)-\frac{\partial Q^\alpha}{\partial x^h}\Big(\frac{dP_\alpha}{dt}+\frac{\partial\Phi}{\partial Q^\alpha}\Big). \quad (3.23)$$

From the above identity it is clear that the associated canonical equations (3.18) and the generalized Hamilton–Jacobi equation (3.21) are sufficient to imply that the momentum field p_h is canonical.

4. CANONICAL AND GAUGE TRANSFORMATIONS

As was mentioned in Sec. 2, a vector field $p_h(x^l)$ does not determine a unique set of Clebsch potentials. In fact, the totality of all differentiable transformations of the form (2.8) such that

$$\epsilon\frac{\partial\psi}{\partial x^h}+P_\alpha\frac{\partial Q^\alpha}{\partial x^h}=\epsilon\frac{\partial\bar\psi}{\partial x^h}+\bar P_\alpha\frac{\partial\bar Q^\alpha}{\partial x^h}, \quad (4.1)$$

constitutes an infinite group, the elements of which are called *Clebsch gauge transformations*. It can be shown[5] that the associated canonical equations (3.18) and the generalized Hamilton–Jacobi equation (3.21) are invariant under such transformations. Thus the results of Sec. 3 are gauge independent.

Let us consider a transformation of the form

$$\bar Q^\alpha=\bar Q^\alpha(t,Q^\beta,P_\beta),\quad \bar P_\alpha=\bar P_\alpha(t,Q^\beta,P_\beta). \quad (4.2)$$

We shall assume that there exist functions $\chi(t,Q^\beta,\bar Q_\beta)$ and $\bar\Phi(t,\bar Q^\alpha,\bar P_\alpha)$ which satisfy

$$d\chi=P_\alpha dQ^\alpha-\bar P_\alpha d\bar Q^\alpha-(\Phi-\bar\Phi)dt\ . \quad (4.3)$$

As is well known,[12] (4.3) represents a canonical transformation between the entities (Q^α,P_α,Φ) and ($\bar Q^\alpha,\bar P_\alpha,\bar\Phi$). We shall make use of two important properties of such transformations. First, the transformation (4.2) is invertible, in fact

$$\Big(\frac{\partial(\bar Q^\alpha,\bar P_\beta)}{\partial(Q^\mu,P_\nu)}\Big)^2=1.$$

It thus follows that the independence of the functions $Q^\alpha(t,x^h),P_\alpha(t,x^h)$ implies that of the potentials $\bar Q^\alpha(t,x^h),\bar P_\alpha(t,x^h)$. Secondly, the associated canonical equations (3.18) imply the equations

$$\frac{d\bar Q^\alpha}{dt}=\frac{\partial\bar\Phi}{\partial\bar P_\alpha}\quad\frac{d\bar P_\alpha}{dt}\quad-\frac{\partial\bar\Phi}{\partial\bar Q^\alpha}, \quad (4.4)$$

that is, the associated canonical equations are structurally invariant under (4.2).

Now the functional dependence of the generator χ in (4.3) implies that

$$\frac{\partial\chi}{\partial Q^\alpha}=P_\alpha,\quad\frac{\partial\chi}{\partial\bar Q^\alpha}=-\bar P_\alpha,\quad\frac{\partial\chi}{\partial t}+\Phi=\bar\Phi. \quad (4.5)$$

We shall use the first two members in the above in order to rewrite the representation (2.7) as

$$p_h=\epsilon\frac{\partial\psi}{\partial x^h}+P_\alpha\frac{\partial Q^\alpha}{\partial x^h}=\epsilon\frac{\partial\psi}{\partial x^h}+\frac{d\chi}{dx^h}+\bar P_\alpha\frac{\partial\bar Q^\alpha}{\partial x^h}.$$

If we define

$$\tilde\chi(t,x^h)\equiv\chi(t,Q^\alpha(t,x^h),\bar Q^\alpha(t,x^h)),$$

and put

$$W(t,x^h)\equiv\epsilon\psi(t,x^h)+\tilde\chi(t,x^h), \quad (4.6)$$

the above becomes

$$p_h=\frac{\partial W}{\partial x^h}+\bar P_\alpha\frac{\partial\bar Q^\alpha}{\partial x^h}. \quad (4.7)$$

If the character of p_h is odd, then, by virtue of the independence of the potentials (ψ,Q^α,P_α), the functions ($W,\bar Q^\alpha,\bar P_\alpha$) will be independent functions of position. Hence, (4.7) is a Clebsch representation and the canonical transformation (4.2) is, in actuality, just a gauge transformation. When the character of p_h is even, the functions ($W,\bar Q^\alpha,\bar P_\alpha$) will not form an independent set. In this case, we shall call (4.7) a *dependent Clebsch representation* of p_h.

Let us now assume that p_h is a canonical field. Thus the associated canonical equations (3.18) are satisfied, which in turn imply the transformed equations (4.4). If we substitute the representation (4.7) into the relation (3.9) and repeat the analysis which led to (3.14), we obtain

$$\frac{d}{dx^h}\Big[\frac{\partial W}{\partial t}+\bar P_\alpha\frac{\partial\bar Q^\alpha}{\partial t}+H\Big(t,x^k,\frac{\partial W}{\partial x^k}+\bar P_\alpha\frac{\partial\bar Q^\alpha}{\partial x^k}\Big)\Big]=\frac{d\bar\Phi}{dx^h}.$$

This implies that

$$\frac{\partial W}{\partial t}+\bar P_\alpha\frac{\partial\bar Q^\alpha}{\partial t}+H\Big(t,x^h,\frac{\partial W}{\partial x^h}+\bar P_\alpha\frac{\partial\bar Q^\alpha}{\partial x^h}\Big)=\bar\Phi(t,\bar Q^\alpha,\bar P_\alpha) \quad (4.8)$$

where, as usual, an arbitrary function of t has been absorbed by the Hamiltonian. Conversely, the equations (4.4) and (4.8) are sufficient to guarantee that p_h is canonical.

An important point to observe is that the function $\bar{\Phi}$ is essentially arbitrary. In fact, as we shall see shortly, it may be taken to be identically zero. This is in contrast with the original superpotential Φ defined by (3.15) which, by construction, cannot vanish identically.

In order to show that Φ may be zero, it is sufficient to observe that the associated canonical equations (3.18) possess $2m$ independent integrals $\hat{Q}^\alpha,\hat{P}_\alpha$, and that the solutions Q^α,P_α of (3.18) may be written as[13]

$$Q^\alpha = Q^\alpha(t,\hat{Q}^\beta,\hat{P}_\beta), \quad P_\alpha = P_\alpha(t,\hat{Q}^\beta,\hat{P}_\beta). \quad (4.9)$$

By definition,

$$\frac{d\hat{Q}^\alpha}{dt} = 0, \quad \frac{d\hat{P}_\alpha}{dt} = 0 \quad (4.10)$$

as a result of (3.18), so that (4.9) may be viewed as a transformation between the entities (Q^α,P_α,Φ) and ($\hat{Q}^\alpha,\hat{P}_\alpha,0$) which preserves the structure of the associated canonical equations (3.18). From the theory of canonical transformations[14] it follows that there exists a function $\chi(t,Q^\alpha,\bar{Q}^\alpha)$ such that

$$d\chi = P_\alpha dQ^\alpha - b\hat{P}_\alpha d\hat{Q}^\alpha - \Phi dt, \quad (4.11)$$

where b is a nonzero constant. We may, without loss of generality, absorb the constant b into the variables $\hat{P}_\alpha$; thus (4.11) reduces to (4.3) with $\bar{\Phi}=0$. As a result, Eq. (4.8) becomes

$$\frac{\partial W}{\partial t} + \hat{P}_\alpha \frac{\partial \hat{Q}^\alpha}{\partial t} + H\left(t, x^h, \frac{\partial W}{\partial x^h} + \hat{P}_\alpha \frac{\partial \hat{Q}^\alpha}{\partial x^h}\right) = 0. \quad (4.12)$$

It is interesting to note that the generator χ must, as a consequence of (4.5) and (4.11), satisfy the Hamilton–Jacobi type equation[15]

$$\frac{\partial \chi}{\partial t} + \Phi\left(t, Q^\alpha, \frac{\partial \chi}{\partial Q^\alpha}\right) = 0. \quad (4.13)$$

In summary: We have shown that the associated canonical equations (3.18) and the generalized Hamilton–Jacobi equation (3.21) reduce to (4.10) and (4.12) respectively if p_h is given the representation

$$p_h = \frac{\partial W}{\partial x^h} + \hat{P}_\alpha \frac{\partial \hat{Q}^\alpha}{\partial x^h}. \quad (4.14)$$

The representation (4.14) is a Clebsch representation when the character of p_h is odd, otherwise it is a dependent Clebsch representation.[16]

5. A GENERALIZED HAMILTON–JACOBI THEORY FOR RELATIVISTIC DYNAMICAL SYSTEMS

In a relativistic dynamical system the time t is no longer a preferred parameter. Instead t is taken to be one of the coordinates of the underlying manifold, say $t = x^n$. Accordingly, it follows that a relativistic canonical formalism should be based upon a parameter-invariant action integral of the form

$$I = \int_C L(x^h, \dot{x}^h) d\tau, \quad (5.1)$$

where τ is an *arbitrary* parameter, C is a curve of the form $x^h = x^h(\tau)$, and $\dot{x}^h \equiv dx^h/d\tau$. We shall assume that (5.1) is invariant under arbitrary class C^1 parameter transformations of the form $\tau = \tau(\sigma)$, and that the Lagrangian L is invariant under the arbitrary class C^2 coordinate transformations $\bar{x}^h = \bar{x}^h(x^l)$. These assumptions are sufficient to ensure that the equations of motion, i.e., the Euler–Lagrange equations

$$E_h(L) \equiv \frac{d}{d\tau}\left(\frac{\partial L}{\partial \dot{x}^h}\right) - \frac{\partial L}{\partial x^h} = 0 \quad (5.2)$$

which arise from (5.1) are invariant under the above mentioned parameter and coordinate transformations.

We shall now give a brief description of the canonical formalism that we will use to describe our dynamical system. This formalism was originally formulated by Rund, and the reader is referred to Ref. 17 for further details.

Let us denote the n components of a relativistic momentum vector by y_h. We shall assume that our dynamical system is described by a Hamiltonian function $H(x^l, y_l)$ which is numerically equal to the Lagrangian ($L x^l, \dot{x}^l$) in (5.1).[18] The parameter invariance of (5.1) implies that H is homogeneous of degree one in the variables y_l, hence this particular formalism is referred to as the *homogeneous* theory. A velocity field $v^h(x^l)$ conjugate to $y_h(x^l)$ is defined by

$$v^h(x^l) \equiv H(x^l, y_l(x^k)) \frac{\partial H}{\partial y_h}(x^l, y_l(x^k)). \quad (5.3)$$

It is assumed that

$$\det\left(\frac{\partial v^h}{\partial y_l}\right) \neq 0, \quad (5.4)$$

for all possible values of x^h and y_h; thus (5.3) may be solved for the variables y_h so as to yield $y_h = y_h(x^l, v^l)$.

If the curve $x^h = x^h(\tau)$ is a streamline of $v^h(x^l)$, that is,

$$\dot{x}^h = v^h(x^l), \quad (5.5)$$

the equality of L and H together with the assumption (5.4) imply the functional relation

$$H(x^l, y_l(x^k, \dot{x}^k)) = L(x^l, \dot{x}^l). \quad (5.6)$$

It may then be shown that L and H satisfy the identities

$$\frac{\partial L}{\partial \dot{x}^h}=\frac{y_h}{H},\quad \frac{\partial L}{\partial x^h}=-\frac{\partial H}{\partial x^h}$$

(it is tacitly assumed that H does not vanish) which allows us to express the equations of motion (5.2) as

$$\frac{d}{d\tau}\left(\frac{y_h}{H}\right)=-\frac{\partial H}{\partial x^h}. \qquad (5.7)$$

The system

$$\dot{x}^h=H\frac{\partial H}{\partial y_h},\quad \frac{d}{d\tau}\left(\frac{y_h}{H}\right)=-\frac{\partial H}{\partial x^h}, \qquad (5.8)$$

which results from (5.3), (5.5), and (5.7), represents the canonical equations of the homogeneous theory. Because of the homogeneity of H, the n equations in the second member of (5.8) satisfy the identity

$$\left[\frac{d}{d\tau}\left(\frac{y_h}{H}\right)+\frac{\partial H}{\partial x^h}\right]\dot{x}^h=0. \qquad (5.9)$$

Finally, let us assume that the field $y_h(x^l)$ satisfies the relation

$$y_h(x^l)=H(x^l,y_l(x^k))\frac{\partial S}{\partial x^h}, \qquad (5.10)$$

where S is a class C^2 scalar function of position. The homogeneity of H leads to the result

$$H\left(x^l,\frac{\partial S}{\partial x^l}\right)=1, \qquad (5.11)$$

which in turn implies that y_h is canonical, i.e., y_h satisfies (5.7) as a result of (5.3) and (5.5). For this reason Eq. (5.11) is called the Hamilton–Jacobi equation for the homogeneous theory.

We shall now show that a generalized Hamilton–Jacobi theory results if the field $y_h(x^l)$ is given the Clebsch representation (2.7).

To this end, let us consider an arbitrary momentum field $y_h(x^l)$ with character c. If y_h is canonical, the relation (5.7) implies that

$$\frac{\partial y_h}{\partial x^l}\dot{x}^l-\frac{y_h}{H}\frac{dH}{d\tau}=-H\frac{\partial H}{\partial x^h},$$

or

$$\left(\frac{\partial y_h}{\partial x^l}-\frac{\partial y_l}{\partial x^h}\right)\dot{x}^l-\frac{y_h}{H}\frac{dH}{d\tau}$$

$$=-H\frac{\partial H}{\partial x^h}-\frac{\partial y_l}{\partial x^h}\dot{x}^l.$$

However, (5.3) and (5.5) yield

$$\frac{\partial y_l}{\partial x^h}\dot{x}^l=H\frac{\partial y_l}{\partial x^h}\frac{\partial H}{\partial y_l},$$

so that the above becomes

$$\left(\frac{\partial y_h}{\partial x^l}-\frac{\partial y_l}{\partial x^h}\right)\dot{x}^l-\frac{y_h}{H}\frac{dH}{d\tau}=-H\frac{dH}{dx^h}. \qquad (5.12)$$

The representation (2.7) allows us to put

$$\left(\frac{\partial y_h}{\partial x^l}-\frac{\partial y_l}{\partial x^h}\right)\dot{x}^l=\frac{dP_\alpha}{d\tau}\frac{\partial Q^\alpha}{\partial x^h}-\frac{dQ^\alpha}{d\tau}\frac{\partial P_\alpha}{\partial x^h};$$

thus (5.12) may be put in the form

$$\left(\epsilon\frac{\partial\psi}{\partial x^h}+P_\alpha\frac{\partial Q^\alpha}{\partial x^h}\right)\frac{1}{H^2}\frac{dH}{d\tau}+\frac{1}{H}\frac{\partial P_\alpha}{\partial x^h}\frac{dQ^\alpha}{d\tau}$$

$$-\frac{1}{H}\frac{\partial Q^\alpha}{\partial x^h}\frac{dP_\alpha}{d\tau}=\frac{dH}{dx^h}. \qquad (5.13)$$

As in the case of the development which led to (3.14), we may invert the ($2m+\epsilon$) equations

$$\epsilon\psi=\epsilon\psi(x^h),\quad Q^\alpha=Q^\alpha(x^h),\quad P_\alpha=P_\alpha(x^h)$$

and obtain, upon relabeling if necessary, the coordinates $(x^1,\cdots,x^{2m+\epsilon})$ as functions of the potentials $(\epsilon\psi,Q^\alpha,P_\alpha)$. This allows us to define a function Φ by

$$\Phi(\epsilon\psi,Q^\alpha,P_\alpha,x^{2m+\epsilon+1},\cdots,x^n)\equiv H(x^h,y_h(x^l)). \qquad (5.14)$$

Substitution of (5.14) into the right-hand side of (5.13) then leads, by virtue of the independence of ($\epsilon\psi,Q^\alpha,P_\alpha$), directly to the relations

$$\epsilon\frac{d\Phi}{d\tau}=\epsilon\Phi^2\frac{\partial\Phi}{\partial\psi},\quad \frac{d}{d\tau}\left(\frac{P_\alpha}{\Phi}\right)$$

$$=-\frac{\partial\Phi}{\partial Q^\alpha},\quad \frac{dQ^\alpha}{d\tau}=\Phi\frac{\partial\Phi}{\partial P^\alpha}, \qquad (5.15)$$

and

$$\frac{\partial\Phi}{\partial x^h}=0,\quad h=2m+1+\epsilon,\cdots,n\ .$$

We may conclude from the last relation that Φ is of the form

$$\Phi=\Phi(\epsilon\psi,Q^\alpha,P_\alpha). \qquad (5.16)$$

We have shown that if y_h is a canonical field, a function Φ, defined by (5.14) and (5.16), satisfies the relations (5.15). Because the last two members of (5.15) have the same structure as the canonical equations (5.8), (5.15) shall be referred to as the *associated canonical equations* (for the homogeneous theory). As in Sec. 3, the function Φ shall be called a *superpotential*. It is interesting to note that a fundamental difference between the canonical formalism presented in this section and the formalism given in Sec. 3 is exhibited by the role of the potential ψ in the respective associated canonical equations.

We shall now use (2.7) and (5.16) in order to rewrite the relation (5.14) as

$$H\left(x^h, \epsilon \frac{\partial \psi}{\partial x^h} + P_\alpha \frac{\partial Q^\alpha}{\partial x^h}\right) = \Phi(\epsilon\psi, Q^\alpha, P_\alpha). \tag{5.17}$$

For a given function Φ, (5.17) represents a first order partial differential equation in the variables ($\epsilon\psi, Q^\alpha, P_\alpha$). In particular, if y_h is given by (5.10), the equation (5.17) becomes

$$H\left(x^h, H(x^l, y_l) \frac{\partial S}{\partial x^h}\right) = \Phi,$$

or, because of the homogeneity of H,

$$H\left(x^h, \frac{\partial S}{\partial x^h}\right) = \frac{\Phi}{H} = 1,$$

which is the ordinary Hamilton–Jacobi equation (5.11) for the homogeneous theory. For this reason, (5.17) shall be referred to as the *generalized Hamilton–Jacobi* equation.

Let us now examine the consistency of the associated canonical equations. With the help of (5.15) we have that

$$\frac{d\Phi}{d\tau} = \epsilon \frac{\partial \Phi}{\partial \psi} \frac{d\psi}{d\tau} + \frac{\partial \Phi}{\partial Q^\alpha} \frac{dQ^\alpha}{d\tau} + \frac{\partial \Phi}{\partial P_\alpha} \frac{dP_\alpha}{d\tau}$$

$$= \frac{d\Phi}{d\tau} \left[\frac{\epsilon}{\Phi^2} \frac{d\psi}{d\tau} + \frac{P_\alpha}{\Phi} \frac{\partial \Phi}{\partial P_\alpha} \right].$$

Substitution of the third member of (5.15) into the above relation gives

$$\frac{d\Phi}{d\tau} = \frac{d\Phi}{d\tau} \left(\frac{\epsilon}{\Phi^2} \frac{d\psi}{d\tau} + \frac{P_\alpha}{\Phi^2} \frac{dQ^\alpha}{d\tau} \right).$$

Hence, if $d\Phi/d\tau$ does not vanish, we have

$$\Phi^2 = \epsilon \frac{d\psi}{d\tau} + P_\alpha \frac{dQ^\alpha}{d\tau}$$

$$= \left(\epsilon \frac{\partial \psi}{\partial x^h} + P_\alpha \frac{\partial Q^\alpha}{\partial x^h} \right) \dot{x}^h$$

$$= y_h \dot{x}^h. \tag{5.18}$$

However, the homogeneity of H together with (5.3) and (5.5) imply that

$$\dot{x}^h y_h = H y_h \frac{\partial H}{\partial y_h} = H^2, \tag{5.19}$$

so that the relation (5.18) becomes

$$H^2 = \Phi^2.$$

But since the replacement of Φ by $-\Phi$ does not affect the associated canonical equations (5.15), we may assume that the above implies $H = \Phi$. Thus, provided that $d\Phi/d\tau \neq 0$, the associated canonical equations imply the generalized Hamilton–Jacobi equation of the homogeneous theory.

Finally, we shall examine the relationship between the canonical equations (5.8) and the associated canonical equations (5.15). Differentiation of the generalized Hamilton–Jacobi expression ($H - \Phi$) with respect to x^h gives

$$\frac{d}{dx^h}(H - \Phi) = \frac{\partial H}{\partial x^h} + \frac{\partial H}{\partial y_l} \frac{\partial y_l}{\partial x^h} - \epsilon \frac{\partial \Phi}{\partial \psi} \frac{\partial \psi}{\partial x^h} - \frac{\partial \Phi}{\partial Q^\alpha} \frac{\partial Q^\alpha}{\partial x^h} - \frac{\partial \Phi}{\partial P_\alpha} \frac{\partial P_\alpha}{\partial x^h}. \tag{5.20}$$

The relations (5.3) and (5.5) allow us to obtain

$$\frac{\partial H}{\partial y_l} \frac{\partial y_l}{\partial x^h} = \frac{\dot{x}^l}{H} \left(\epsilon \frac{\partial^2 \psi}{\partial x^l \partial x^h} + \frac{\partial P_\alpha}{\partial x^h} \times \frac{\partial Q^\alpha}{\partial x^l} + P_\alpha \frac{\partial^2 Q^\alpha}{\partial x^l \partial x^h} \right)$$

$$= \frac{1}{H} \left(\frac{dy_h}{d\tau} + \frac{\partial P_\alpha}{\partial x^h} \frac{dQ^\alpha}{d\tau} - \frac{\partial Q^\alpha}{\partial x^h} \frac{dP_\alpha}{d\tau} \right).$$

Substitution of the above into (5.20) gives, after a little rearrangement,

$$\frac{d}{dx^h}(H - \Phi) = \frac{\partial H}{\partial x^h} + \frac{d}{d\tau}\left(\frac{y_h}{H}\right) + \frac{\partial P_\alpha}{\partial x^h} \left(\frac{1}{H} \frac{dQ^\alpha}{d\tau} - \frac{\partial \Phi}{\partial P_\alpha} \right) - \frac{\partial Q_\alpha}{\partial x^h} \left[\frac{d}{d\tau}\left(\frac{P_\alpha}{H}\right) + \frac{\partial \Phi}{\partial Q^\alpha} \right] + \epsilon \frac{\partial \psi}{\partial x^h} \left(\frac{1}{H^2} \frac{dH}{d\tau} - \frac{\partial \Phi}{\partial \psi} \right). \tag{5.21}$$

Now (5.21) is simply an identity which holds along any streamline of the velocity field v_h. If we assume, however, that the Clebsch potentials satisfy the associated canonical equations (5.15), then, from our previous work in this section, $H = \Phi$, and (5.21) reduces to

$$\frac{\tau}{d\tau}\left(\frac{y_h}{H}\right) + \frac{\partial H}{\partial x^h} = 0.$$

Thus, the associated canonical equations imply the canonical equations of the homogeneous theory.

6. THE ASSOCIATED MULTIPLE INTEGRAL VARIATIONAL PRINCIPLE

In this section we shall show how the equations of motion (5.7) of the homogeneous theory result from an n-fold multiple integral variational principle.[19] This is an important application of our generalized Hamilton–Jacobi theory, since it gives rise to a method whereby field equations and particle equations of motion may be obtained from a single variational principle.

Let us suppose that we are given a Lagrange density

$$\mathscr{L}=\mathscr{L}(x^h;\Theta^A;\Theta^A{}_{,h}),\quad A=1,\cdots,N$$

where the $\Theta^A(x^h)$ represent N differentiable field functions and $\Theta^A{}_{,h}\equiv\partial\Theta^A/\partial x^h$. If G denotes a domain in X_n, we may form the integral

$$I\equiv\int_G \mathscr{L}dx^1\wedge\cdots\wedge dx^n$$

which defines an n-fold variational principle. The corresponding Euler–Lagrange equations

$$E_{\Theta_A}(\mathscr{L})\equiv\frac{d}{dx^h}\left(\frac{\partial\mathscr{L}}{\partial\Theta^A{}_{,h}}\right)-\frac{\partial\mathscr{L}}{\partial\Theta^A}=0, \qquad (6.1)$$

represent field equations for the functions $\Theta^A(x^h)$.

The basic idea is to form a Lagrange density $\mathscr{L}$ in which the Clebsch potentials are treated as field variables.[20] The Lagrangian $\mathscr{L}$ is chosen so that the associated canonical equations (5.15) will be a consequence of the Euler–Lagrange equations which arise from $\mathscr{L}$. The results of Sec. 5 will then imply that the equations of motion (5.7) are satisfied.

To this end, let us define a function $\widetilde{H}$ by

$$\widetilde{H}(x^h;P_\alpha;\epsilon\psi_{,h};Q^\alpha{}_{,h})\equiv H(x^h;\epsilon\psi_{,h}+P_\alpha Q^\alpha{}_{,h}). \qquad (6.2)$$

If $\nu(x^h)$ denotes a density function and $\Phi=\Phi(\epsilon\psi,Q^\alpha,P_\alpha)$, the quantity $\nu(\widetilde{H}-\Phi)$ has the dimensions of a Lagrange density. We shall show that if $\mathscr{L}$ is defined by

$$\mathscr{L}(x^h;\nu;\epsilon\psi;Q^\alpha;P_\alpha;\epsilon\psi_{,h};Q^\alpha{}_{,h})\equiv\nu(\widetilde{H}-\Phi), \qquad (6.3)$$

then the field equations

$$E_\nu(\mathscr{L})=E_{Q^\alpha}(\mathscr{L})=E_{P_\alpha}(\mathscr{L})=E_\psi(\mathscr{L})=0 \qquad (6.4)$$

imply the equations of motion (5.7).

First let us note that the following identities are an immediate consequence of the definitions (6.2) and (5.3):

$$\frac{\partial\widetilde{H}}{\partial\psi_{,h}}=\frac{\epsilon v^h}{H},\quad \frac{\partial\widetilde{H}}{\partial P_\alpha}=\frac{\partial Q^\alpha}{\partial x^h}\frac{v^h}{H},$$

$$\frac{\partial H}{\partial Q^\alpha{}_{,h}}=\frac{v^h P_\alpha}{H}. \qquad (6.5)$$

Furthermore, since ν is a density function, the quantity

$$j^h(x^l)\equiv\nu v^h(x^l) \qquad (6.6)$$

is a current density. If we next substitute from (6.3), (6.5), and (6.6) into the Euler–Lagrange expression in the left-hand side of (6.1), we find that

$$E_\nu(\mathscr{L})=-(H-\Phi),$$

$$E_{P_\alpha}(\mathscr{L})=-\frac{\nu}{H}\left(\frac{\partial Q^\alpha}{\partial x^h}v^h-\frac{\partial\Phi}{\partial P_\alpha}\right),$$

$$E_{Q^\alpha}(\mathscr{L})=\frac{P_\alpha}{H}\frac{\partial j^l}{\partial x^l}+\nu\left[\frac{d}{dx^h}\left(\frac{P_\alpha}{H}\right)v^h+\frac{\partial\Phi}{\partial Q^\alpha}\right],$$

$$E_\psi(\mathscr{L})=\frac{\epsilon}{H}\frac{\partial j^l}{\partial x^l}-\epsilon\nu\left(\frac{1}{H^2}\frac{dH}{dx^h}v^h-\frac{\partial\Phi}{\partial\psi}\right). \qquad (6.7)$$

It would be easy enough to show from (6.7) that the field equations (6.4) imply the vanishing of the associated canonical equations (5.15), which in turn implies that the equations of motion (5.7) are satisfied. However, we may show this explicitly by rewriting the identity (5.21) in terms of the Euler–Lagrange expressions $E_\nu(\mathscr{L})$, $E_{P_\alpha}(\mathscr{L})$, $E_{Q^\alpha}(\mathscr{L})$, and $E_\nu(\mathscr{L})$. With the help of (5.5), the identity (5.21) becomes

$$\frac{d}{dx^h}(H-\Phi)=\frac{\partial H}{\partial x^h}+\frac{d}{d\tau}\left(\frac{y_h}{H}\right)+\frac{\partial P_\alpha}{\partial x^h}\times\left(\frac{1}{H}\frac{\partial Q^\alpha}{\partial x^l}v^l-\frac{\partial\Phi}{\partial P_\alpha}\right)-\frac{\partial Q^\alpha}{\partial x^h}\times\left[\frac{d}{dx^l}\left(\frac{P_\alpha}{H}\right)v^l+\frac{\partial\Phi}{\partial Q^\alpha}\right]+\epsilon\frac{\partial\psi}{\partial x^h}\times\left(\frac{1}{H^2}\frac{dH}{dx^l}v^l-\frac{\partial\Phi}{\partial\psi}\right).$$

If we substitute from (6.7), the above may be rewritten as

$$\frac{d}{dx^h}[-E_\nu(\mathscr{L})]=\frac{\partial H}{\partial x^h}+\frac{d}{d\tau}\left(\frac{y_h}{H}\right)-\frac{1}{\nu}\left(\frac{\partial P_\alpha}{\partial x^h}E_{P_\alpha}(\mathscr{L})+\frac{\partial Q^\alpha}{\partial x^h}E_{Q^\alpha}(\mathscr{L})+\frac{\partial\psi}{\partial x^h}E_\psi(\mathscr{L})\right)+\frac{y_h}{\nu H}\frac{\partial j^l}{\partial x^l}. \qquad (6.8)$$

Thus, the field equations (6.4) imply that

$$\frac{\partial H}{\partial x^h}+\frac{d}{d\tau}\left(\frac{y_h}{H}\right)+\frac{y_h}{\nu H}\frac{\partial j^l}{\partial x^l}=0. \qquad (6.9)$$

Upon multiplying the above relation by $\dot{x}^h$ and making use of the identities (5.9) and (5.19), we obtain

$$\frac{\partial j^l}{\partial x^l}=0, \qquad (6.10)$$

which has the obvious interpretation of an equation of continuity. Substitution of (6.10) into (6.9) then yields the equations of motion (5.7).

As an example of the above complex of ideas, let us consider a general relativistic matter field with proper density ν which is defined on a Riemannian manifold X_4, of the variables ($x^1,x^2,x^3,x^4=t$). The metric $g_{hk}(x^l)$ of X_4, which is assumed to have the signature ($-,-,-,+$), must satisfy the Einstein field equations

$$G^{hk}=\kappa T^{hk}, \qquad (6.11)$$

where G^{hk} is the Einstein tensor, κ is a constant, and T^{hk} is the energy-momentum tensor of our matter field.

We shall demand that the Hamiltonian H which describes our dynamical system satisfy the following three conditions:

(i) H is a scalar with respect to arbitrary class C^2 coordinate transformations;

(ii) H is homogeneous of degree one in the variable y_h;

(iii) H depends on the variables x^h only through the metric tensor, i.e.,

$$H(x^l,y_l)=H(g_{hk}(x^l),y_l).$$

The first two assumptions are in keeping with the parameter invariant formalism of Sec. 5, whereas assumption (iii) is reasonable since our matter field is incoherent, that is, it lacks an internal structure. Together, (i)–(iii) are sufficient to imply[21] that H is of the form

$$H(x^l,y_l)=k(g^{hl}y_hy_l)^{1/2}, \qquad (6.12)$$

where k is a nonzero constant, and g^{hl} is the inverse of g_{hl}.

We shall next consider the equations of motion which result from (6.12). According to (5.3), a velocity field $v^h(x^l)$ which is conjugate to $y_h(x^l)$ is given by

$$v^h=k^2g^{hl}y_l. \qquad (6.13)$$

We may solve (6.13) for y_l in terms of $\dot{x}^h$ along an arbitrary streamline of v^h and obtain

$$y_l=\frac{1}{k^2}g_{hl}\dot{x}^h.$$

Substitution of this result into the definition (5.6) then gives

$$L(x^l,\dot{x}^l)=\frac{1}{k}(g_{hl}\dot{x}^h\dot{x}^l)^{1/2}. \qquad (6.14)$$

A well-known calculation then yields the Euler–Lagrange equations

$$\frac{D}{D\tau}\left(\frac{\dot{x}^h}{L}\right)=0, \qquad (6.15)$$

where $D/D\tau$ denotes the process of absolute covariant differentiation. In particular, if we replace τ by the so-called proper time s, which is defined by

$$ds\equiv L(x^h,dx^h), \qquad (6.16)$$

then

$$L\left(x^h,\frac{dx^h}{ds}\right)=1, \qquad (6.17)$$

and (6.15) reduces to

$$\frac{D}{Ds}\left(\frac{dx^h}{ds}\right)=0 \qquad (6.18)$$

which are the geodesic equations of X_4.

Let us, in analogy with (6.3), define a Lagrange density $\mathscr{L}_0$ by

$$\mathscr{L}_0=\sqrt{g}\,\nu(\tilde{H}-\Phi), \qquad (6.19)$$

where $g=|\det(g_{hk})|$, $\Phi=\Phi(\epsilon\psi,Q^\alpha,P_\alpha)$ and $\tilde{H}=H(g_{hk}(x^l),\epsilon\psi_{,l}+P_\alpha Q^\alpha{}_{,l})$, that is,

$$\tilde{H}=k\,[g^{hl}(\epsilon\psi_{,h}+P_\alpha Q^\alpha{}_{,h})(\epsilon\psi_{,l}+P_\beta Q^\beta{}_{,l})]^{1/2}.$$

The Lagrangian $\mathscr{L}_0$ depends upon the metric tensor g_{hk} in addition to the variables ($\nu,\epsilon\psi,Q^\alpha,P_\alpha,\epsilon\psi_{,h},Q^\alpha{}_{,h}$). However, this does not affect the analysis which led to the result (6.8), namely, that the field equations

$$E_\nu(\mathscr{L}_0)=E_{Q^\alpha}(\mathscr{L}_0)=E_{P_\alpha}(\mathscr{L}_0)=E_\psi(\mathscr{L}_0)=0, \qquad (6.20)$$

imply the canonical equations (5.7), as well as the equation of continuity (6.10). We have seen that the canonical equations are given by (6.15), whereas the factor of $\sqrt{g}$ in (6.19) allows us to rewrite (6.10) as

$$\frac{\partial}{\partial x^h}(\sqrt{g}\,\nu v^h)=0. \qquad (6.21)$$

In order to obtain the Einstein field equations (6.11), we shall consider the Lagrangian

$$\mathscr{L}\equiv\sqrt{g}\,R-\mathscr{L}_0=\sqrt{g}\,[R-\nu(\tilde{H}-\Phi)]. \qquad (6.22)$$

Since the added term does not depend upon the Clebsch potentials or the scalar ν, the equations (6.20) are equivalent to

$$E_\nu(\mathscr{L})=E_\psi(\mathscr{L})=E_{Q^\alpha}(\mathscr{L})=E_{P_\alpha}(\mathscr{L})=0. \qquad (6.23)$$

Thus the field equations (6.23) imply the equations of motion (6.15) and the equation of continuity (6.21).

We must yet consider the equations which result from the variation of the metric tensor g_{rs}, namely, the equations

$$E_{g_{rs}}(\mathscr{L})\equiv E_{rs}(\mathscr{L})=0. \qquad (6.24)$$

It is well known that

$$E_{rs}(\sqrt{g}\,R)=\sqrt{g}\,G^{rs}. \tag{6.25}$$

Furthermore, from (6.19) we obtain

$$E_{rs}(\mathscr{L}_0)=\left(\frac{\partial\sqrt{g}}{\partial g_{rs}}\right)\nu(\widetilde{H}-\Phi)+\sqrt{g}\,\nu\frac{\partial\widetilde{H}}{\partial g_{rs}}. \tag{6.26}$$

It is easily verified that

$$\frac{\partial\widetilde{H}}{\partial g_{rs}}=-\frac{k^2}{2}\frac{v^r v^s}{L}, \tag{6.27}$$

and that $\widetilde{H}=\Phi$ as a result of the field equation $E_\nu(\mathscr{L})=0$. Thus (6.26) becomes

$$E_{rs}(\mathscr{L}_0)=\frac{\sqrt{g}\,k^2}{2}\frac{v^r v^s}{L}. \tag{6.28}$$

Upon comparison of (6.22) with (6.25), (6.17) and (6.28), we see that the field equations (6.24) imply that

$$G^{rs}=\frac{k^2\nu}{2}v^r v^s.$$

If we now identify the constant k with $\sqrt{2\kappa}$, the above reduces to the Einstein equations

$$G^{rs}=\kappa\nu v^r v^s=\kappa T^{rs}, \tag{6.29}$$

where

$$T^{rs}=\nu v^r v^s. \tag{6.30}$$

It is interesting that the energy-momentum tensor (6.30), which is usually postulated on physical grounds, arises as a by-product of our analysis. Furthermore, it is straightforward to show that the covariant divergence of T^{rs} vanishes as a result of the field equations (6.23).[22]

ACKNOWLEDGMENT

It is a pleasure to be able to thank Dr. Hanno Rund of the University of Arizona for his advice and encouragement.

[1]C. Carathéodory, *Variationsrechnung und partielle Differential-Gleichungen erster Ordnung* (Teubner, Leipzig and Berlin, 1935; translation, Holden-Day, San Francisco, 1967), Sec. 239.

[2]H. Rund, Arch. Ration. Mech. Anal. **65**, 305 (1977).

[3]Latin indices $h,k,l,\cdots$ shall run from 1 to n; summation over repeated suffixes is implied.

[4]Ref. 1, Sec. 140.

[5]Ref. 2, Sec. 4.

[6]H. Rund, *The Hamilton–Jacobi Theory in the Calculus of Variations* (Van Nostrand, New York, 1966; Krieger reprint, New York, 1973), Ch. 3.

[7]Most of the results in this paper are contained in: R. Baumeister, "Applications of Clebsch Potentials to Variational Principles in the Theory of Physical Fields," Ph.D. thesis, The University of Arizona, 1977.

[8]The "character" of p_h is identical with the "Pfaffian class" of the 1-form $\omega=p_h dx^h$.

[9]Greek indices $\alpha,\beta,\cdots$ shall run from 1 to m, summation over repeated indices being implied.

[10]A. Clebsch, J. reine angew. Math **56**, 1 (1859). (In this paper, Clebsch also obtains representations for fields with character n.)

[11]Ref. 6, Ch. 2.

[12]E.C.G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974), pp. 30-66.

[13]Ref. 1, Sec. 21.

[14]Ref. 12, pp. 41–44; also H. Rund, Tydskrif vir Natuurwetenskappe, June–Sept. 1969, pp. 181–90.

[15]This fact is also noted by H. Rund in Ref. 2, p. 319.

[16]It should be stressed that Φ vanishes only for a special gauge, and that the results obtained when $\Phi=0$ are *gauge dependent*.

[17]Ref. 6, Ch. 3.

[18]It follows that H is *not* to be identified with the energy of our dynamical system.

[19]An analogous "associated multiple integral variational principle" for nonrelativistic dynamical systems may be found in Ref. 2, pp. 319–23.

[20]Ref. 10, p. 9.

[21]This result is a special case of a theorem established by H. Rund (personal communication). A proof may be found in Ref. 7, pp. 134–40.

[22]This example is generalized in variational principles formulated by the author (Ref. 7, pp. 131–61) and H.Rund [*The Significance of Nonlinearity in the Natural Sciences*, edited by B. Kursunoglu, A Perlmutter, and L.F. Scott (Plenum, New York, 1977), p. 121–43].

Presymplectic manifolds and the Dirac–Bergmann theory of constraints

Mark J. Gotay

Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland, College Park, Maryland

James M. Nester

Department of Physics, University of Saskatchewan, Saskatoon, Saskatchewan

George Hinds

7304 Holly Avenue, Takoma Park, Md. 20012

(Received 7 March 1978)

We present an algorithm which enables us to state necessary and sufficient conditions for the solvability of generalized Hamilton-type equations of the form $\iota(X)\omega = \alpha$ on a presymplectic manifold (M,ω) where α is a closed 1-form. The algorithm is phrased in the context of global infinite-dimensional symplectic geometry, and generalizes as well as improves upon the local Dirac–Bergmann theory of constraints. The relation between our algorithm and the geometric constraint theory of Śniatycki, Tulczyjew, and Lichnerowicz is discussed.

I. INTRODUCTION

It is generally recognized[1–4] that classically, a physical system can be described in terms of a symplectic manifold, that is, a manifold M together with a nondegenerate closed 2-form ω. Physically, M is the phase space of the system while ω is essentially a generalization of the Poisson bracket.

The manifold M and the symplectic form ω are kinematical in nature; the dynamics of the system is determined by specifying a real-valued function H on phase space, the Hamiltonian. One then solves the Hamilton equations

$$\iota(X)\omega = dH, \tag{1.1}$$

thereby obtaining the dynamical trajectories of the system in phase space (i.e., the integral curves of the vector field X). The fact that ω is nondegenerate assures us that Eq. (1.1) has a unique solution; indeed, the nondegeneracy of ω means that the linear map $\flat: TM \to T^*M$ defined by $\flat(X) := \iota(X)\omega$ is an isomorphism. Thus for any H we can solve (1.1) uniquely: $X = \flat^{-1}(dH)$. Once X has been determined, one appeals to the standard results of differential equation theory in order to integrate X.

We want to consider in detail the case when ω is degenerate, in which case (M,ω) is said to be a *presymplectic* manifold. This situation usually arises when the system is constrained in some manner, and often when M is infinite-dimensional. When (M,ω) is degenerate, the Hamilton equations (1.1) may or may not possess solutions (and, in general, even if solutions exist they will not be unique) depending on whether or not dH is in the range of $\flat$. In the former case, the equations (1.1) possess nonunique solutions, the nonuniqueness being characterized by $\ker\omega$. It is the latter case which is the most interesting, for then (1.1) as it stands possesses no globally defined solutions. In order to "solve" the Hamilton equations, then, one must "modify" M, the equations (1.1), or both. We have developed an algorithm which enables us to produce and solve such a "modified" problem in both the finite- and infinite-dimensional cases. More precisely, we find whether or not there exists a submanifold N of M along which the equations (1.1) hold; if such a submanifold exists, we give a *constructive* method for finding it. Moreover, we show that this submanifold is *unique* in the sense that it contains any other submanifold along which (1.1) is satisfied (Sec. IV).

This work grew out of an attempt to globalize the Dirac–Bergmann theory of constraints,[5,6] first published circa 1950. In these papers, an algorithm was developed by Dirac, Bergmann, and his collaborators for dealing with Lagrangian systems which could not be put into canonical form in the usual manner owing to the fact that the momenta are not all independent functions of the velocities. This algorithm was nicely summarized by Dirac,[7] who showed that such systems could be put into a modified canonical form with the motion restricted to a "constraint" submanifold. Requiring the equations of motion to be consistent on this submanifold led to a sequence of further constraint submanifolds which either terminated or restricted the system to such an extent that no solution of the original variational problem could be found. He showed further that a modified Poisson bracket could be defined in such a way that certain constraints could be effectively eliminated, the remaining variables falling (in principle) into two classes: (i) those whose time development from given initial conditions is completely arbitrary, and (ii) those whose evolution is well defined by canonical equations of motion.

The point of developing this algorithm was not pedagogical, for several classical systems exist which display the above-mentioned feature; notably electromagnetism and gravity. Insofar as it is felt to be necessary to cast these theories into canonical form for the purpose of quantization, the Dirac–Bergmann algorithm provides, in principle, a method for doing this and, at the same time, for identifying the "physical observables" or "true degrees of freedom." In fact, Dirac applied his technique to general relativity[8] and electromagnetism[7] and showed that it was effective in isolating an

 0022-2488/78/1911-2388$1.00

appropriate set of variables with which to describe the motion.

While our algorithm is related to the Dirac–Bergmann method, there are several important differences in both the method and the results.

First, although the Dirac–Bergmann algorithm is clear in an algebraic sense, it is hard to gain an adequate geometric picture of what is taking place. Thus, we have chosen to phrase our discussion in *global* terms using the language of infinite-dimensional symplectic geometry. This manifestly coordinate-invariant language is eminently suited to both the algebraic and geometric aspects of the problem. To this end, much work has been done in recent years,[9–11] but the bulk of this has been mostly concerned with translating Dirac's concepts into the modern mathematical idiom and with symplectically reinterpreting the results of his algorithm. No one seems to have successfully globalized the *algorithm* itself. In Sec. III, we give a brief overview of this "geometric theory of constraints."

Secondly, as Dirac himself noticed,[12] his algorithm is ambiguous in the following sense (to be elaborated upon later): One is not certain whether or not the first-class secondary constraints should be included in the Hamiltonian. Put another way, Dirac is unable to show that the motions generated by the first-class secondary constraints are physically irrelevant (gauge) and hence cannot identify those observables which correspond to "true" degrees of freedom. Actually, this is not so much a problem with the Dirac–Bergmann *algorithm* per se as it is with its physical interpretation. The physical interpretation, in turn, is obscured by Dirac's nongeometric formulation of the constraint algorithm. In Sec. V, we show that our geometric algorithm not only globalizes (and thus substantiates) Dirac's results, but moreover that, strictly speaking, the Hamiltonian should *not* in general contain all the first-class secondary constraints.[13] This uncertainty concerning the first-class secondary constraints is fairly subtle, and we shall not consider it in depth in this paper. This question, and the related issue of the physical intepretation of our geometric algorithm will be discussed from another, more fundamental point of view in a companion paper.[14]

Lastly, our algorithm is applicable in situations considerably more general than those considered by Dirac. Specifically, the Dirac–Bergmann algorithm can only be applied when the degenerate manifold M is actually a "primary constraint submanifold" of some symplectic manifold W. The algorithm we propose does not require the *a priori* existence of such a nondegenerate manifold W. Physically, this may be of considerable importance in the case of an infinite number of degrees of freedom where ω may be degenerate even if there are no constraints.[15,16] The *a priori* presymplectic case is also of physical interest from the point of view of the quantization problem. Normally, when one quantizes a constrained system, one relies upon Śniatycki's theorem[9,11] to eliminate the second-class constraints from the theory. However, Śniatycki's theorem fails in the presymplectic case,[17] leading one to question whether or not such systems are actually quantizable.

After Dirac, a number of people approached the constraint problem from various viewpoints,[18–20] but no completely satisfactory analysis of the three above-mentioned aspects of the theory was forthcoming. (This paper amends an attempt made by one of us several years ago.[21]) In fact, there have been a number of papers[19,20] which challenge the validity of the Dirac–Bergmann algorithm on theoretical grounds. As our algorithm generalizes the Dirac–Bergmann theory, this approach would seem to verify the correctness of the latter, since our derivation is from a completely different (viz., geometrically rigorous) point of view. Moreover, it is not difficult to show that although several of the issues raised by these authors are of importance for the elucidation of the theory, their objections are without content (Sec. V).

Section II provides a very brief introduction to symplectic geometry and its application to Hamiltonian systems in an infinite-dimensional setting.[22] A more comprehensive treatment of these topics is given in the texts by Abraham and Marsden,[1] Souriau,[2] Chernoff and Marsden,[16] and Godbillon.[4] For some of the more advanced notions and applications, one should consult the lecture notes of both Woodhouse[3] and Weinstein.[23] The infinite-dimensional techniques used throughout this paper are clearly and comprehensively explained in the books by Marsden,[15] Chernoff and Marsden,[16] Lang.[24] In general, we shall try to keep our notation and terminology[25] consistent with that of Refs. 1, 16, 23, and 24.

Section III reviews the basic notions and tools of geometric constraint theory which are necessary for the presentation of the algorithm in Sec. IV and the correspondence with the Dirac–Bergmann theory detailed in Sec. V. Finally, we apply the algorithm to electromagnetism in Sec. VI as an example of the calculational techniques involved in the theory.

II. SYMPLECTIC GEOMETRY AND HAMILTONIAN MECHANICS[1–4,15,16,23,26]

Let M be a manifold modelled on a Banach space E, and suppose that ω is a closed 2-form on M. Then (M,ω) is said to be a *strong symplectic* manifold if the linear map $\flat: TM \to T^*M$ defined by $\flat(X) \equiv X^\flat := \iota(X)\omega$ is an isomorphism. However, it may happen that ω will be injective but not surjective, in which case (M,ω) is called a *weak symplectic* manifold, ω being *weakly* nondegenerate. Generically, $\flat$ will be neither injective nor surjective and ω is then *degenerate*. When E is finite-dimensional, there is of course no distinction between weak and strong symplectic forms. For brevity, strongly symplectic manifolds will often be referred to simply as *symplectic*, while weakly nondegenerate and degenerate forms will be dubbed *presymplectic*.

The simplest example of a weak symplectic manifold is the cotangent bundle T^*Q of any Banach manifold Q. In fact, on T^*Q there exists a canonical 1-form θ defined by

$$\langle v|\theta\rangle = \langle \pi_* v | \tau v\rangle$$

where $v \in TT^*Q$, and $\pi: T^*Q \to Q$, $\tau: TT^*Q \to T^*Q$ are the bundle projections. This 1-form defines the weak symplectic structure as follows: $\Omega = -d\theta$.

Locally, we can find a chart $U \subset F$, where F is the model space for Q, such that on U,[25]

$$\theta(x,\sigma)\cdot(a \oplus \pi)=\langle a|\sigma\rangle$$

and

$$\Omega(x,\sigma)\cdot(a \oplus \pi, b \oplus \tau)=\langle a|\tau\rangle-\langle b|\pi\rangle. \tag{2.1a}$$

If F is finite-dimensional, this is the same as saying that there exist coordinates (q^i,p_i) on U such that

$$\theta|U=p_i dq^i$$

and

$$\Omega \mid U=dq^i \wedge dp_i. \tag{2.1b}$$

The weak nondegeneracy of Ω follows from the above formulas after a simple calculation. In fact, when F is reflexive, Ω is strongly nondegenerate.[16]

However, not every strongly symplectic manifold (M,ω) is a cotangent bundle nor is ω always exact [e.g., (S^2,ω) where ω is a volume on S^2. Then ω cannot be exact, and S^2 is of course not a cotangent bundle] although locally both statements are true. That a strong symplectic manifold is locally a cotangent bundle follows from a normal form theorem, called Darboux's theorem,[27] which states that a chart always exists in which ω is constant. In such a chart ω must always have the form (2.1a) or (2.1b). However, this result need not hold in the presymplectic case.[28] This normal form theorem shows that strongly symplectic geometries are "flat"—this should be compared with the corresponding theorem in Riemannian geometry.

Another contrast with Riemannian geometry can be obtained by examining the infinitesimal automorphisms of a strong symplectic structure (i.e., the *locally Hamiltonian* vector fields). These are vector fields X such that

$$L_X\omega=0. \tag{2.2}$$

As ω is closed, it is clear that X will be a locally Hamiltonian vector field iff $d\iota(X)\omega=0$. Since ω is strongly nondegenerate, the map $\flat$ will have an inverse $\#$ and consequently we see that if α is a closed 1-form then $\alpha^{\#}$ will be a locally Hamiltonian vector field. As there are many closed forms on any manifold, there will exist many infinitesimal symplectic automorphisms. By way of contrast, in Riemannian geometry the existence of Killing vector fields is the exception rather than the rule.

Physically, the weak and strong symplectic manifolds one almost always encounters are cotangent bundles. This comes about as follows: One describes a physical system by specifying a manifold Q called configuration space and a function L, the Lagrangian, on velocity phase space TQ. One then casts the theory into canonical form by "changing variables" from (q^i,v^i) to (q^i,p_i) and replacing L by the Hamiltonian H via $H(q,p)=p_i v^i - L(q,v)$. Mathematically, this transition takes the form of a map $FL:TQ \to T^*Q$ which is called the *Legendre transformation* or the *fiber derivative*[1] and is defined by

$$\langle z \mid FL(w)\rangle := \frac{d}{dt} L(w+tz)\Big|_{t=0}, \tag{2.3}$$

where $z,w \in TQ$. The Hamiltonian is defined via

$$H \circ FL(w) := \langle w|FL(w)\rangle - L(w). \tag{2.4}$$

[This is provided that (2.4) does in fact define a single-valued function H on $FL(TQ)$. For further discussion regarding this point, see Ref. 43.] Additionally, in the finite-dimensional case, the canonical momenta are "defined" by

$$p_i \circ FL(w)=\partial L/\partial v^i(w). \tag{2.5}$$

One major advantage of changing a theory into Hamiltonian form is that T^*Q canonically carries a (weak) symplectic structure whereas TQ does not.

It is the weak symplectic structure on T^*Q which gives rise to the elegant simplicity of the Hamiltonian formalism. For example, the Hamilton equations (1.1), when written in terms of local Darboux coordinates [i.e., canonical coordinates for which (2.1b) holds] are simply

$$\frac{dq^i}{dt} \equiv X[q^i]=\frac{\partial H}{\partial p_i},$$

$$\frac{dp_i}{dt} \equiv X[p_i]=-\frac{\partial H}{\partial q^i}.$$

Similarly, one can use Ω (provided Ω is strongly symplectic)[26] to define the Poisson bracket of two functions f,g as follows

$$\{f,g\} := \Omega(\xi_f,\xi_g) \tag{2.6}$$

where $\xi_f := df^{\#}$. In a Darboux chart, $\{f,g\}$ reduces to the usual expression. The symplectic analog of a canonical transformation is a diffeomorphism $\zeta: T^*Q \to T^*Q$ such that $\zeta^*\Omega=\Omega$.

There do, however, exist physically interesting systems whose phase spaces are not cotangent bundles and whose symplectic forms are not exact. An example of such a system was given by Souriau,[2] who investigated the dynamics of a freely spinning massive particle in Minkowski spacetime from a symplectic viewpoint (in this example, $M=R^6 \times S^2$). Systems of this type do not possess configuration manifolds and consequently do not admit Lagrangian formulations (at least in the usual sense).

With this in mind, it is apparent that from a geometric viewpoint the Hamiltonian formulation of classical physics is of primary importance while the Lagrangian formalism is an alternative construction applicable only in special cases.

Turning now to the presymplectic case, we recall that a presymplectic manifold is obtained by relaxing the assumption that $\flat$ be bijective. Presymplectic manifolds arise quite frequently in physics, in particular when the Legendre transformation (2.3) is degenerate. This means that FL is no longer a local diffeomorphism, but merely an into map, the range of which defines a submanifold M of T^*Q. In more familiar terms, FL will fail to be a local diffeomorphism when the matrix

$$\left(\frac{\partial^2 L}{\partial v_i \partial v_j}\right)$$

is not invertible.

This is the starting point of the Dirac–Bergmann constraint theory, in which M is called the *primary constraint* submanifold. The *primary constraints* are a collection of functions on T^*Q which locally define M as a submanifold of T^*Q. One particular set of primary constraints are those relations (2.5) (or combinations thereof) which do not define the momenta p_i as independent functions of the velocities v^i. Geometrically, M will inherit a presymplectic structure from T^*Q by pulling Ω back to M via the inclusion $j:M \to T^*Q$. We are thus faced with the problem of determining the dynamics of a physical system on the presymplectic phase space $(M, j^*\Omega)$ where the Hamiltonian H is given by (2.4).

The presymplectic phase spaces of the above discussion are rather special in that they are *naturally* submanifolds of weakly nondegenerate manifolds. But this is not always the case, even in physics, as was shown by Künzle[29] who obtained genuinely *pre*symplectic phase spaces for spinning particles in curved spacetimes.

Thus, from both a mathematical and physical standpoint, there is considerable justification in considering presymplectic geometry in its own right. The physical issue that one is then confronted with is the following: A system is described by a presymplectic phase space (M,ω) and a Hamiltonian H on M. What does one mean by "consistent equations of motion" on M, and how does one obtain and solve such equations? The algorithm we propose will select a certain submanifold N of M upon which we can define and solve "consistent equations of motion." Before we can proceed to discuss the algorithm however, we must first examine the properties of such submanifolds.

III. GEOMETRIC CONSTRAINT THEORY[3,6,7,9–11,23]

We would like to have a classification scheme for submanifolds of presymplectic manifolds which is at the same time mathematically convenient and physically meaningful. Dirac[7] first developed a local classification of submanifolds of strongly symplectic manifolds which Śniatycki and Tulczyjew later globalized as the "geometric theory of constraints."[9,10] This classification is of the utmost importance insofar as the physical interpretation of the algorithm is concerned.[14] We briefly review this classification (generalized to the presymplectic case) following Śniatycki, Tulczyjew, and Lichnerowicz.[11]

Let N be a submanifold of the presymplectic manifold (M,ω) with inclusion j. The manifold N is called a *constraint* submanifold, and the triple (M,ω,N) is called a *canonical system*. We define the *symplectic complement* $TN^\perp$ of $\underline{TN}$ in TM to be

$$TN^\perp = \{Z \in TM|N \text{ such that } \omega|N(X,Z)=0 \text{ for all } X \in \underline{TN}\}.$$

For our purposes this is not the most convenient characterization of $TN^\perp$. We prefer that given by the following.

Proposition 1: $TN^\perp = \{Z \in TM|N \text{ such that } j^*[\iota(Z)\omega]=0\}$.

Proof: Let $Z \in TN^\perp$. Then for any $W \in TN$,

$$0=\omega|N(j_*W,Z)=j^*\langle j_*W|\iota(Z)\omega\rangle=\langle W|j^*[\iota(Z)\omega]\rangle.$$

As this is true for all $W \in TN$, it follows that $j^*[\iota(Z)\omega]=0$. Conversely, if $j^*[\iota(Z)\omega]=0$, then the equality is established by reversing the above calculation. Q.E.D.

If S is a subspace of a Banach space E, we define $S^\vdash \subset E^*$, the *annihilator* of S, to be the set of all $\beta \in E^*$ such that $\langle v|\beta\rangle=0$ for all $v \in S$. Similarly, if A is a subspace of E^*, we define $A^\vdash \subset E$ to be the collection of all $v \in E$ such that $\langle v|\lambda\rangle=0$ for all $\lambda \in A$. If E is reflexive, then it is possible to show[30] that $(A^\vdash)^\vdash=\bar{A}$. We shall say that ω is *topologically closed* provided the map $\flat$ is a closed map, i.e., $\flat$ maps closed sets into closed sets. We note that if ω is strongly nondegenerate, then it is necessarily topologically closed. We can now prove the following important fact[31]:

Proposition 2: If M is reflexive and ω is topologically closed, then

$$(TN^\perp)^\vdash=\underline{TN}^\flat.$$

Proof: With obvious shorthand notation,

$$W \in (\underline{TN}^\flat)^\vdash \Leftrightarrow \omega/N(W|\underline{TN})=0 \Leftrightarrow W \in TN^\perp$$

thereby proving that $(\underline{TN}^\flat)^\vdash=TN^\perp$. As ω is topologically closed, $\underline{TN}^\flat$ is closed in T^*M. The desired result follows from the above by taking $A=\underline{TN}^\flat$. Q.E.D.

The constraint submanifold N is said to be

(i) *isotropic* if $\underline{TN} \subset TN^\perp$,

(ii) *coisotropic* or *first-class* if $TN^\perp \subset \underline{TN}$,

(iii) *weakly symplectic* or *second class* if $\underline{TN} \cap TN^\perp=\{0\}$, and

(iv) *Lagrangian* if $\underline{TN}=TN^\perp$.

Clearly, $\underline{TN} \cap TN^\perp=\underline{\ker\omega_N}$, where $\ker\omega_N$ is the set of all $W \in TN$ such that $\iota(W)\omega_N=0$. If N does not happen to fall into any of these categories, then N is said to be a *mixed* constraint manifold.

Locally, a first-class constraint submanifold can be described by the vanishing of a collection of functions A such that for all $f \in A$, $W[f]|N=0$ for all $W \in TN^\perp$. If (M,ω) happens to be strongly nondegenerate, this is easily seen to be equivalent to Dirac's requirement that A be in involution, i.e., $\{f,g\}|N=0$ for all $f,g \in A$.

The functions $f \in A$ of the preceding paragraph are called *first-class constraint functions*. More generally, any function f (resp. 1-form γ) on M such that $f|N=0$ (resp. $j^*\gamma=0$) is called a *constraint function* (resp. *constraint form*), and any function g (resp. 1-form σ) on M such that $W[g]|N=0$ (resp. $\langle W|\sigma\rangle|N=0$) for all $W \in TN^\perp$ is said to be *first class*. Functions (resp. forms) which are not first-class are called *second class*. A second-class constraint submanifold, then, can be locally described by second-class constraint functions. In general, a mixed or isotropic constraint submanifold will require both first- as well as second-class constraint functions for its local description.

As an example of a second-class constraint submanifold, let $C \subset Q$, where Q is configuration space. Then T^*C is a weakly symplectic submanifold of (T^*Q,Ω), hence it is second class. Furthermore, the constraint submanifold $\pi^{-1}(C)\subset T^*Q$ is first class. The former is an example of a *holonomic* constraint.

We have discussed some simple properties of submanifolds of presymplectic manifolds in the above, but we have not yet indicated their origin. It is to this task that we now turn our attention.

IV. THE CONSTRAINT ALGORITHM

We begin by taking a presymplectic manifold (M_1,ω_1) to be the phase space of some physical system. Let H_1 be the Hamiltonian of the system. We inquire as to under what conditions and by what methods we can solve the canonical equations of motion $\iota(X)\omega_1=dH_1$. Actually, we can be somewhat more general[32] and write the Hamilton equations as

$$\iota(X)\omega_1=\alpha_1, \tag{4.1}$$

where α_1 is a closed 1-form, the *Hamiltonian form*. Locally, as α_1 is closed, we can always find a Hamiltonian function corresponding to α_1. As was mentioned in the Introduction, if α_1 is in the range of $\flat: TM_1\rightarrow T^*M_1$, then Eq. (4.1) is consistent as it stands and can be solved directly for X.[33]

In the generic case, however, this will not be so. But there may exist points of M_1 (such points being assumed to form a submanifold M_2 of M_1),[34] for which $\alpha_1|M_2$ is in the range of $\flat|M_2$. We are thus led to try and solve Eq. (4.1) restricted[35] to M_2, i.e.,

$$(\iota(X)\omega_1-\alpha_1)\circ j_2=0, \tag{4.2}$$

where j_2: $M_2\rightarrow M_1$ is the inclusion.

Equation (4.2) evidently possesses solutions, but this is not the whole story. Physically, we must demand that the motion of the system be constrained to lie in M_2, if this concept is to have any meaning. Thus, the locally Hamiltonian vector field X appearing in (4.2) must be tangent to M_2, that is, X must be of the form $X=j_{2*}\hat{X}$ with $\hat{X}\in TM_2$, or else the equations of motion will try to evolve the system off M_2.

This requirement will not necessarily be satisfied, forcing us to further restrict (4.1) to the submanifold M_3 of M_2 defined by

$$M_3:=\{m\in M_2 \text{ such that } \alpha_1(m)\in \underline{TM_2}^{\flat}\}.$$

We must now ensure that the solution to (4.1) restricted to M_3 is in fact tangent to M_3; this will in general necessitate yet further restrictions.

It is now clear how the algorithm must proceed. We generate a string of submanifolds

$$\cdots\rightarrow M_3 \xrightarrow{j_3} M_2 \xrightarrow{j_2} M_1$$

defined as follows

$$M_{l+1}:=\{m\in M_l \text{ such that } \alpha_1(m)\in \underline{TM}_l^{\flat}\}.$$

Once the constraint algorithm so defined is set into motion, only one of three distinct possibilities may occur.[36] They are:

Case 1: There exists a K such that $M_K=\phi$,

Case 2: Eventually, the algorithm produces a submanifold $M_K\neq\phi$ such that $\dim M_K=0$, and

Case 3: There exists a K such that $M_K=M_{K+1}$ with $\dim M_K\neq 0$.

In Case 1, $M_K=\phi$ means that the Hamilton equations (4.1) have no solutions at all in any sense. In principle, this means that (M_1,ω_1,α_1) does not accurately describe the dynamics of any system.

The second possibility results in a constraint submanifold which consists of isolated points. The equations (4.1) are consistent, but the only possible solution is $X=0$ and there is no dynamics.

For Case 3, we have a constraint submanifold M_K and completely consistent equations at motion on M_K of the form

$$(\iota(X)\omega_1-\alpha_1)|M_K=0. \tag{4.3}$$

It is this submanifold M_K (the *final* constraint submanifold) which corresponds to the submanifold N discussed in Sec. III.

If the algorithm terminates, then *by construction* we are assured that at least one solution X to the canonical equations exists and furthermore that this solution is tangent to M_K. We note that X need not be unique, for we can add to it any element of $\ker\omega_1\cap\underline{TM}_K$. In addition, it is obvious, again by construction, that the final constraint submanifold is *unique* in the following sense: if N is any other submanifold along which the equations (4.1) are satisfied, then $N\subset M_K$.

The algorithm we have proposed provides a geometrically intuitive and conceptually simple method for defining and solving consistent equations of motion on a presymplectic manifold. The algorithm is of very general applicability, requiring only that the phase spaces involved be Banach manifolds.

For many purposes, the algorithm as presented above is too "abstract." More precisely, it is somewhat difficult to use in practice, the calculation of the constraint submanifolds occasionally being a rather formidable task. In addition, the present form of the algorithm is too awkward to be useful for comparison with the Dirac–Bergmann theory. Consequently, we will now recast the algorithm into a form which is more tractable in these regards.

We begin by recharacterizing the constraint submanifold M_2. We can typify the inconsistency of Eq. (4.1) as follows: Consider the set $TM_1^{\perp}$ of vector fields characterized as in Proposition 1. If Eq. (4.1) is to be solvable, then $W\in TM_1^{\perp}$ implies that the left-hand side of (4.1) vanishes and consequently it follows that $\langle W|\alpha_1\rangle$ vanishes. On the other hand, if $W\in TM_1^{\perp}$ implies that $\langle W|\alpha_1\rangle=0$, then $\alpha_1\in(TM_1^{\perp})^{\perp}$. If ω_1 is topologically closed and if M_1 is reflexive, then by Proposition 2 we have that $\alpha_1\in(TM_1^{\flat})$. Thus, the points of M_1 where (4.1) is inconsistent are exactly those points for which $\langle W|\alpha_1\rangle$ is nonzero. Subject to the above asumptions, then, M_2 can alternatively characterized as follows

$$M_2:=\{m\in M_1 \text{ such that } \langle TM_1^{\perp}|\alpha_1\rangle(m)=0\}$$

with obvious shorthand notation. The consistency conditions $\langle TM_1^{\perp}|\alpha_1\rangle=0$ are called, after Dirac and Bergmann, *secondary constraints*.

Returning to the problem of solving (4.2), the demand that the solution X be tangent to M_2 leads to further consistency conditions (*tertiary* constraints) as follows: If there exists an X tangent to M_2 such that (4.2) holds, then for $W\in TM_1$,

$$\begin{aligned}0&=[\iota(W)\iota(X)\omega_1-\iota(W)\alpha_1]\circ j_2\\&=-j_2{}^*\langle j_{2*}\hat{X}\,|\,\iota(W)\omega_1\rangle-\langle W\,|\,\alpha_1\rangle\circ j_2\\&=-\langle X\,|\,j_2{}^*[\iota(W)\omega_1]\rangle-\langle W\,|\,\alpha_1\rangle\circ j_2\end{aligned}$$

where $\hat{X}\in TM_2$ with $X=j_{2*}\hat{X}$. Consequently, consistency of (4.2) demands that if W is such that $j_2{}^*[\iota(W)\omega_1]=0$ (i.e., $W\in TM_2{}^\perp$), then $\langle W|\alpha_1\rangle\circ j_2=0$. This, again, may not always be the case and we must correspondingly restrict the equation (4.2) to those points of M_2 where $\langle TM_2^\perp|\alpha_1\rangle=0$.

The algorithm then proceeds as before, generating a sequence of submanifolds

$$\cdots\to M_3\xrightarrow{j_3}M_2\xrightarrow{j_2}M_1$$

defined as follows

$$M_{l+1}:=\{m\in M_l \text{ such that } \langle TM_l^\perp|\alpha_1\rangle(m)=0\},$$

where

$$TM_l^\perp=\{W\in TM_1 \text{ such that } k_l{}^*[\iota(W)\omega_1]=0\}$$

for $l\geqslant 1$ with $k_l:=j_2\circ j_3\circ\cdots\circ j_l$ and $k_1:=\mathrm{id}|M_1$. The constraint functions on M_{l-1} which define M_l are called *l-ary constraints* and are always of the form $\langle TM_{l-1}{}^\perp|\alpha_1\rangle=0$. Sometimes, for convenience, all l-ary constraints are (for $l\geqslant 2$) simply called *secondary*.

If the algorithm terminates, we are faced with the same three possibilities as before. In the second or third case, we now explicitly show that (4.3) possesses solutions. We note that, as the algorithm terminates with M_K, $\langle TM_K{}^\perp|\alpha_1\rangle=0$.

Theorem: The canonical equations

$$(\iota(X)\omega_1=\alpha_1)|M_K$$

possess solutions tangent to M_K iff

$$\langle TM_K^\perp|\alpha_1\rangle=0.$$

Proof: $\Rightarrow$ Let $X\in TM_K$ be a solution, and suppose that $W\in TM_K^\perp$. Then

$$\begin{aligned}\langle W\,|\,\alpha_1\rangle&=\langle W,k_K{}^*X|\omega_1\rangle\circ k_K\\&=-k_K{}^*\langle k_{K*}X|\iota(W)\omega_1\rangle\\&=-\langle X|k_K{}^*[\iota(W)\omega_1]\rangle\quad(\text{as } X\in TM_K)\\&=0\end{aligned}$$

by Proposition 1.

$\Leftarrow$ Suppose $W\in TM_K^\perp$. Then $\langle W|\alpha_1\rangle=0$, so that $\alpha_1|M_K\in(TM_K^\perp)^\vdash$. But by Proposition 2, $(TM_K^\perp)^\vdash=\underline{TM}_K^\flat$. Thus, $\alpha_1|M_K\in\underline{TM}_K^\flat$, that is, there exists an $X\in\underline{TM_K}$ such that $[\iota(X)\omega_1=\alpha_1]|M_K$. Q.E.D.

It is of interest to note that the above theorem is actually independent of the constraint algorithm. In fact, if N is *any* submanifold of a presymplectic manifold (M,ω), then the equations $(\iota(X)\omega-\alpha)|N=0$ possess solutions tangent to N iff $\langle TN^\perp|\alpha\rangle=0$.

We now turn to the uniqueness of the final constraint manifold M_K. For suppose there exists some other submanifold N along which the equations (4.1) are satisfied, that is, let $X=j_*\hat{X}$, $\hat{X}\in TN$ be such that

$$[\iota(X)\omega_1-\alpha_1]|N=0,$$

where $j:N\to M$ is the inclusion. Then if $W\in TM_1^\perp$, we have from the above that $\langle W|\alpha_1\rangle\circ j=0$, so that $N\subset M_2$. Let $\tilde{j}_2:N\to M_2$ be the inclusion; then $j=j_2\circ\tilde{j}_2$. For $Y\in TM_2^\perp$,

$$\begin{aligned}0&=[\iota(Y)\iota(X)\omega_1-\iota(Y)\alpha_1]\circ j\\&=-\langle\hat{X}\,|\,j^*[\iota(Y)\omega_1]\rangle-\langle Y|\alpha_1\rangle\circ j\end{aligned}$$

Now $j^*[\iota(Y)\omega_1]=\tilde{j}_2{}^*\circ j_2{}^*[\iota(Y)\omega_1]=0$ as $Y\in TM_2^\perp$, so $\langle Y|\alpha_1\rangle\circ j=0$, and thus $N\subset M_3$. Continuing in this fashion, we see that $N\subset M_K$.

This version of the algorithm, while perhaps not quite as intuitive as the earlier construction, is still geometrically natural and much better suited to calculation. However, it is important to bear in mind that this version can be used only when the model space for M_1 is reflexive and ω_1 is topologically closed; otherwise one might obtain spurious results.

The canonical system (M_1,ω_1,M_K) and the equations of motion (4.3) are the end results of the constraint algorithm. The further development of the theory (Dirac brackets, the reduced phase space, quantization) follows from the geometric constraint formalism of Śniatycki, Tulczyjew and Lichnerowicz. But now we must turn to a thorough investigation of our geometric algorithm vis-a-vis the Dirac–Bergmann theory.

V. RELATION TO THE DIRAC–BERGMANN THEORY OF CONSTRAINTS[5–7,14]

We now compare the constraint algorithm presented in the last section with the Dirac–Bergmann theory, and show that ours does in fact generalize the latter. We also contrast our our method with similar algorithms presented by Shanmugadhasan, Kundt, and Hinds and point out that these algorithms disagree with ours and consequently with the Dirac–Bergmann theory as well.

We first briefly sketch the Dirac–Bergmann algorithm, displaying the correspondence between their techniques and our more geometric ones.

We start with a Lagrangian L and a reflexive configuration space Q. Changing to canonical form via the fiber derivative FL, we find that the motion of the system is constrained to the submanifold $M_1:=FL\,(TQ)$ of the strongly symplectic manifold T^*Q. Locally, on some neighborhood U, we can describe $U_1:=M_1\cap U$ by a set of primary constraints $\{\phi^A\}$. Using these, Dirac argues that the Hamiltonian on U should be of the form

$$h=\bar{H}_1+u_A\phi^A, \tag{5.1}$$

where $\bar{H}_1$ is any extension to U of the Hamiltonian H_1 induced on M_1 by FL and the u_A are yet to be determined Lagrange multipliers.[37]

Translating into symplectic terms, Dirac then searches for solutions to

$$(\iota(X)\Omega - dh)|U_1 = 0, \tag{5.2}$$

where Ω is the canonical symplectic form on T^*Q. As Ω is nondegenerate, solutions X certainly exist, but Dirac notes that the constraints ϕ^A must be preserved, that is,

$X[\phi^A]|U_1 = 0$. Geometrically, this means that X must be tangent to U_1. In terms of the Poisson bracket associated with Ω via (2.6), this requirement translates into a set of conditions

$$\dot{\phi}^A|U_1 = 0, \tag{5.3}$$

where

$$\dot{\phi}^A = \{\phi^A, \overline{H}_1\} + u_B\{\phi^A, \phi^B\}. \tag{5.4}$$

The vanishing of the expressions (5.4) by virtue of (5.3) will, in general, give some information about the u_A and will also give a number of additional constraints. To see this, consider all possible linear combinations of (5.3). Some of these linear combinations will be satisfied trivially, others will fix some of the Lagrange multipliers u_B, and the remaining ones will be independent of the u_B.

These latter conditions take the form $f^\alpha_A \dot{\phi}^A$ where

$$f^\alpha_A\{\phi^A, \phi^B\}|U_1 = 0$$

by (5.4), thus yielding

$$f^\alpha_A\{\phi^A, \overline{H}_1\}|U_1 = 0\ .$$

In general, of course, these last equations will not be satisfied except on a local submanifold U_2 of U_1. These conditions are therefore secondary constraints.

Denoting the quantities $f^\alpha_A\{\phi^A, \overline{H}_1\}$ by ζ^α, we see that the preservation of these secondary constraints requires that

$$\dot{\zeta}^\alpha|U_2 = 0,$$

where

$$\dot{\zeta}^\alpha = \{\zeta^\alpha, \overline{H}_1\} + u_B\{\zeta^\alpha, \phi^B\}.$$

As before, the linear combinations of the above conditions which are independent of the u_B, i.e., those linear combinations $g^a_\alpha \dot{\zeta}^\alpha$ such that

$$g^a_\alpha\{\zeta^\alpha, \phi^B\}|U_2 = 0, \tag{5.5}$$

will yield tertiary constraints

$$g^a_\alpha\{\zeta^\alpha, \overline{H}_1\}|U_2 = 0. \tag{5.6}$$

One then iterates this procedure, arriving at some final local constraint submanifold U_K (if the problem is solvable) and a solution X to the equations of motion of the form

$$\iota(X)\Omega = d\overline{H}_1 + u_\mu d\chi^\mu + u_i d\xi^i \tag{5.7}$$

restricted to U_K, where the χ^μ are first-class primary constraints (the Lagrange multipliers u_μ being arbitrary) and the ξ^i are second-class primary constraints (the u_i being fixed).[38]

Furthermore, it was shown that the first-class *primary* constraints are generating functions of motions (i.e., gauge transformations) which leave the physical state invariant (this is, of course, related to the fact that the u_μ are arbitrary). This led Dirac to conjecture that the first-class *secondary* constraints may also generate physically irrelevant motions and hence they should also (for the sake of completeness) be included in the Hamiltonian.[39] Dirac therefore proposed adjoining the first-class secondary constraints ψ^a with arbitrary multipliers λ_a to h thereby obtaining the "extended" Hamiltonian

$$h_E = \overline{H}_1 + u_\mu\chi^\mu + u_i\xi^i + \lambda_a\psi^a. \tag{5.8}$$

Thus, Dirac reasoned that the solutions of

$$(\iota(X)\Omega - dh_E)|U_K = 0 \tag{5.9}$$

would give the most general evolution of the system.

This, then, is the essence of the Dirac–Bergmann theory. With regard to our construction, the first important fact is that each Dirac–Bergmann local constraint submanifold U_l is an open submanifold of the M_l produced by our algorithm. To see this, consider the l th step of the Dirac–Bergmann algorithm, and let ζ^α be (at most) l-ary constraints. Define, as Ω is strongly nondegenerate, the vector field Y^a on U_l by

$$\iota(Y^a)\Omega = g^a_\alpha d\zeta^\alpha. \tag{5.10}$$

Using (2.6), Eqs. (5.5) become

$$0 = g^a_\alpha\{\zeta^\alpha, \phi^B\}|U_l$$

$$= -\iota(Y^a)d\phi^B|U_l$$

and consequently $Y^a \in \underline{TU_1}$, as the ϕ^B are primary constraints. Thus, if $j_1: M_1 \rightarrow T^*Q$ is the inclusion,

$$(j_1 \circ k_l)^*[\iota(Y^a)\Omega] = 0$$

by (5.10), so that $Y^a \in TU_l^{\perp\perp}$ by Proposition 1. Consequently,[38]

$$Y^a \in TU_l^{\perp\perp} \cap \underline{TU_1} = \underline{TU_l^\perp}.$$

Similarly, one can show that every vector field $Y \in \underline{TM_l^\perp}$ induces a condition of the form (5.6). Consequently, the same equations which define the local submanifold U_l also locally generate the constraint submanifold M_l.

Therefore, it is clear that the Dirac–Bergmann algorithm is just a local version of our algorithm. Even so, the algorithm we have presented has one significant advantage over the Dirac–Bergmann method in that it is of considerably more general applicability. It is apparent how crucially the Dirac–Bergmann algorithm depends upon the existence of the primary constraints. Our geometric algorithm, by way of contrast, requires only M_1 and its presymplectic structure for its utilization. The manifold M_1 *never* need be a primary constraint submanifold of some other strongly nondegenerate manifold.

But one important difference yet remains. Dirac solved the equations of motion on T^*Q along M_K, whereas we have done so on M_1 along M_K. We now show that we can lift our equations of motion (4.3) to T^*Q obtaining the equations (5.7) and thereby *proving* the formal equivalence of the two algorithms, and thus substantiating the Dirac–Bergmann procedure.

To find the analog of (4.3) on T^*Q, we write

$$\iota(X)\Omega-\alpha_0=\beta_0 \tag{5.11}$$

along M_K, where X is some solution of (4.3) and α_0 is any 1-form on T^*Q such that $\alpha_1=j_1{}^*\alpha_0$. As X solves (4.3), pulling (5.11) back to M_l gives $j_l{}^*\beta_0=0$ so that β_0 is a primary constraint form. Locally, β_0 can be decomposed (nonuniquely) in the form

$$\beta_0=l_\mu d\chi^\mu+g_i d\xi^i.$$

Thus, (5.11) becomes locally

$$\iota(X)\Omega-\alpha_0=l_\mu d\chi^\mu+g_i d\xi^i. \tag{5.12}$$

Now, (4.3) only determines X up to vector fields in $TM_K\cap\underline{\ker\omega_1}=\underline{\ker\omega_K}\cap\underline{\ker\omega_1}$. Letting $Y\in\underline{\ker\omega_K}\cap\underline{\ker\omega_1}$, we see that $X-Y$ must satisfy (5.11) as well, and since $\iota(Y)\Omega$ is a first-class primary constraint form, it can be locally expressed as $f_\mu d\chi^\mu$. Substituting into (5.12), we can write along M_K

$$\iota(X)\Omega=\alpha_0+(l_\mu+f_\mu)d\chi^\mu+g_i d\xi^i. \tag{5.13}$$

From this we see that the second-class piece $g_i d\xi^i$ of (5.12) is insensitive to the choice of X. Moreover, the first-class part $l_\mu d\chi^\mu$ is uniquely determined only for fixed X. Consequently, as X is not unique, the functions l_μ are arbitrary; on the other hand, the g_i are independent of the choice of X and hence are completely determined. Thus, we have reproduced Dirac's result (5.7).

It remains to discuss the "extended" equations of motion (5.9). We notice that nowhere in (5.13) do secondary constraints appear, nor is there any *a priori* reason why they should, at least from the geometric arguments presented above.

The ultimate resolution of this problem depends upon whether or not the first-class secondary constraints generate gauge transformations.[40] This, in turn, depends crucially upon one's definition of "physical state" and "gauge transformation." In other words, how "gauge" the first-class secondary constraints are depends upon the *physical interpretation* of the algorithm and consequently is not strictly amenable to proof.[41]

For example, in the "orthodox" interpretation of the algorithm,[14] all the first-class secondary constraints ψ^a are *assumed* to be gauge. In this case, one could append these constraints to the Hamiltonian as in (5.8) without changing the physical content of the theory; however, in practice one may not always want to do this. The reason is that one may have fixed a gauge (either inadvertently or by design) in the Hamiltonian; some of the ψ^a will then generate physically irrelevant motions that will not respect the gauge condition. If one wishes to retain this choice of gauge in the description of the system, then one cannot attach these constraints to the Hamiltonian. On the other hand, there may be certain other ψ^a which will generate gauge transformations which do not break the gauge; these can be included without reservation in the Hamiltonian—in fact, they are "already there" in some sense (for an example, see Sec. VI). Thus, from the standpoint of the usual interpretation of the algorithm, one in general does not need, or perhaps want, to append the first-class secondary constraints to the Hamiltonian: Some of the ψ^a will break the gauge choice, and those that do not are already present in the Hamiltonian.

There do exist other "unorthodox" interpretations of the algorithm in which certain of the first-class secondary constraints are *not* gauge. Consequently, these constraints certainly cannot be included in the Hamiltonian. The remaining ψ^a which do generate physically irrelevant motions may or may not be attached to the Hamiltonian as discussed above.

For a more detailed presentation of these points and examples thereof, consult Ref. 14.

In 1965, Hinds[21] presented an algorithm which, like ours, was stated in geometric language. Rather than consider this algorithm in detail, we merely point out the major differences between Hinds' approach and ours. Basically, the crux of the matter is that, at the l th step of the algorithm, Hinds attempts to solve the equation ($\omega_l:=k_l{}^*\omega_1$ etc.)

$$\iota(X)\omega_l=\alpha_l \tag{5.14}$$

in contrast to our equation

$$[\iota(X)\omega_1]\circ k_l=\alpha_1\circ k_l. \tag{5.15}$$

The conditions for the existence of solutions to an equation of the type (5.14) are less restrictive than those required for Eq. (5.15). To see this, note that the sets of vector fields which generate Hinds algorithm are $\ker\omega_l$, whereas ours are $TM_l^\perp$, and $\underline{\ker\omega}_l\subset TM_l^\perp$. The upshot of this is that after the $l=2$ step, Hinds' algorithm and ours diverge: The constraint submanifolds M_l for $l\geqslant 2$ are no longer the same in both algorithms. If one attempts to reproduce the Dirac–Bergmann results using Hinds's scheme, one obtains

$$h_E=\overline{H}_1+u_\mu\chi^\mu+u_i\xi^i+\lambda_a\psi^a+\lambda_\Delta\theta^\Delta,$$

where the coefficients λ_Δ of the second-class secondary constraints θ^a do not necessarily vanish.

A simple example which illustrates the above is the following: Take $TQ=TR^4$ with coordinates $\{q^i,v^i\}$ with Lagrangian

$$L(q,v)=\tfrac{1}{2}m(v^1)^2-\tfrac{1}{2}k(q^1)^2$$
$$-b(q^3q^4)+\tfrac{1}{2}c(v^4-aq^2)^2.$$

A somewhat different scheme was proposed by Shanmugadhasan[19] to rectify an alleged oversight in the Dirac–Bergmann theory. Shanmugadhasan for the most part works on velocity phase space and deals directly with the Lagrange equations. He claims that the Dirac–Bergmann theory overlooks certain subsidiary conditions arising from the degeneracy of the Hessian matrix $(\partial^2L/\partial v^i\partial v^j)$; this of course is not the case as these subsidiary conditions are none other than primary constraints (Sec. II). Furthermore, Shanmugadhasan completely ignores the possibility that secondary constraints might occur in the theory, and of course it is these which really form the core of the problem. In fact, Shanmugadhasan's method cannot cope with the perfectly consistent (if somewhat strange) Lagrangian given above.

Kundt[20] also quarrels with the Dirac–Bergmann algorithm and has offered his own interpretation of their theory[17] which, curiously enough, requires all the primary constraints to be first class. Kundt's theory fails for the Proca field.

VI. AN EXAMPLE: ELECTROMAGNETISM[7,16]

The Maxwell theory provides a nice illustration of both the geometric calculations involved in the algorithm and the application of modern infinite-dimensional techniques to symplectic geometry. Throughout this section, we shall closely follow the notation of Chernoff and Marsden.[16] We shall also sacrifice mathematical rigor (i.e., we shall ignore certain infinite-dimensional technicalities) in favor of a clearer exposition.

The $3+1$ decomposed Maxwell Lagrangian can be written as

$$L(A,\dot{A})=\tfrac{1}{2}\int_{R^3}[(\vec{\nabla}A_\perp)^2+2(\vec{\nabla}A_\perp)\cdot\dot{\vec{A}}+\dot{A}^2-(\vec{\nabla}\times\vec{A})^2]d\mu, \qquad (6.1)$$

where the vector potential is decomposed $A=(A_\perp,\vec{A})$, R^3 denotes a constant-time Cauchy surface in Minkowski spacetime, and μ is some measure on R^3.

We must first decide on a choice for velocity phase space TQ. The configuration space should be some Hilbert space of all 4-vectors $(A_\perp,\vec{A})$. As L contains at most first spatial derivatives of A, an appropriate choice for Q is

$$Q=H^1_\perp\oplus\vec{H}^1$$

with the obvious notational shorthand, where H^1 is the first Sobolev space on R^3. Velocity phase space, that is, the manifold of all $(A,\dot{A})$ is then

$$TQ=Q\oplus(L^2_\perp\oplus\vec{L}^2) \qquad (6.2)$$

as no spatial derivatives of $\dot{A}$ appear in L. The measure μ can then be taken to be the ordinary L^2 measure on R^3. We note that Q is reflexive, so that the symplectic form Ω on T^*Q is strongly nondegenerate and hence topologically closed.

To translate into the Hamiltonian language, we must calculate the fiber derivative FL. By definition, $FL|Q=\mathrm{id}|Q$ so

$$FL(A,\dot{A})\cdot(A,\dot{B})=(A,\dot{D}L(A,\dot{A})\cdot\dot{B}), \qquad (6.3)$$

where $\dot{D}$ denotes the Frechét derivative along the fiber. An easy calculation shows that

$$\dot{D}L(A,\dot{A})\cdot\dot{B}=\int_{R^3}[\dot{\vec{A}}\cdot\dot{\vec{B}}+(\vec{\nabla}A_\perp)\cdot\dot{\vec{B}}]d\mu. \qquad (6.4)$$

If we define the natural pairing $\langle\,|\,\rangle: TQ\times T^*Q\to R$ by

$$\langle(A,\dot{A})\,|\,(A,\pi)\rangle=\int_{R^3}[\dot{\vec{A}}\cdot\vec{\pi}+\dot{A}_\perp\pi_\perp]d\mu, \qquad (6.5)$$

where $(A,\pi)\in T^*Q$, then (6.3) becomes, using (6.4)

$$FL(A,\dot{A})=(A,\dot{\vec{A}}+\vec{\nabla}A_\perp). \qquad (6.6)$$

Defining the "canonical field momentum" $\vec{\pi}$ by

$$\vec{\pi}:=\dot{\vec{A}}+\vec{\nabla}A_\perp, \qquad (6.7)$$

it is suggestive that $\pi_\perp$ does not appear in (6.6). In fact, if one defines the projection $\mathrm{pr}^2_\perp$ on the second factor by

$$\mathrm{pr}^2_\perp(A,\pi)=\pi_\perp,$$

then it follows that

$$\mathrm{pr}^2_\perp\circ FL(A,\dot{A})=0. \qquad (6.8)$$

Thus, $\pi_\perp=0$ is a primary constraint. The primary constraint submanifold M_1 of T^*Q is then

$$M_1=Q\oplus\vec{L}^{2*}. \qquad (6.9)$$

We now apply the algorithm. The strong symplectic form Ω on T^*Q is given by (2.1a),

$$\Omega(a\oplus\pi,b\oplus\tau)=\langle a|\tau\rangle-\langle b|\pi\rangle \qquad (6.10)$$

with $a\oplus\pi$, $b\oplus\tau\in T(T^*Q)$.[25] If j_1 is the inclusion of M_1 into T^*Q, we have $\omega_1=j_1{}^*\Omega$. Consequently, as Ω is topologically closed, ω_1 is also. This, combined with the fact that M_1 is reflexive, allows us to use the second version of the algorithm presented in Sec. IV.

The first thing we must do is calculate $TM_1^\perp=\ker\omega_1$. That is, we search for vectors $b\oplus\tau\in TM_1$ which annihilate all other vectors $a\oplus\pi$ in TM_1. Using (6.5) and (6.10), we find that $b\oplus\tau\in TM_1^\perp$ iff

$$b\oplus\tau=(b_\perp,\vec{0})\oplus 0. \qquad (6.11a)$$

In other words,

$$TM_1^\perp=H^1_\perp\vec{0}. \qquad (6.11b)$$

The Hamiltonian H_1 induced on M_1 by FL is, according to (2.5), (6.1), and (6.6)

$$H_1(A,\pi)=\int_{R^3}[\tfrac{1}{2}\vec{\pi}^2-(\vec{\nabla}A_\perp)\cdot\vec{\pi}+\tfrac{1}{2}(\vec{\nabla}\times\vec{A})^2]d\mu. \qquad (6.12)$$

Consequently, if $b\oplus\tau\in TM_1$,

$$dH_1(A,\pi)\cdot(b\oplus\tau)=\int_{R^3}[\vec{\pi}\cdot\vec{\tau}+b_\perp(\vec{\nabla}\cdot\vec{\pi})+A_\perp(\vec{\nabla}\cdot\vec{\tau})+(\vec{\nabla}\times\vec{A})\cdot(\vec{\nabla}\times\vec{b})]d\mu. \qquad (6.13)$$

To continue with the algorithm, it is necessary to make sure that the primary constraint (6.8) is preserved. Thus, we demand that $\langle TM_1^\perp|dH_1\rangle=0$. Letting $b\oplus\tau\in TM_1^\perp$, we have from (6.11a) upon substitution into (6.13)

$$dH_1(A,\pi)\cdot(b\oplus\tau)=\int_{R^3}b_\perp(\vec{\nabla}\cdot\vec{\pi})d\mu.$$

This expression will be zero provided

$$\vec{\nabla}\cdot\vec{\pi}=0, \qquad (6.14)$$

as $b_\perp$ is arbitrary. We thus pick up a secondary constraint, M_2 being the submanifold of M_1 along which (6.14) is satisfied.

Pursuing the algorithm, we must now find $\underline{TM}_2^\perp$. For $a\oplus\pi$ in $\underline{TM}_2$ and $b\oplus\tau\in TM_1$,

$$\omega_1(a\oplus\pi,b\oplus\tau)=\int_{R^3}[\vec{\tau}\cdot\vec{a}-\vec{\pi}\vec{b}]d\mu \qquad (6.15)$$

by (6.9). In general, the right-hand side of (6.15) will vanish iff $\vec{\tau}=\vec{0}$ and $\vec{b}=\vec{\nabla}g$ for some function g, making use of (6.14) and an integration by parts. Consequently,

$$TM_2^\perp=\{b\oplus 0\in TM_1 \text{ such that } \vec{b}=\vec{\nabla}g,\ g\in H^1\}. \qquad (6.16)$$

At this point, the algorithm terminates. To see this, let $b\oplus 0$ be as in (6.16). Substitution into (6.13) gives

$$dH_1(A,\pi)\cdot(b\oplus 0)=\int_{R^3}[b_\perp(\vec{\nabla}\cdot\vec{\pi})+(\vec{\nabla}\times\vec{A})\cdot(\vec{\nabla}\times\vec{\nabla}g)]d\mu.$$

The first term vanishes by (6.14) and the second also as curl(grad)$=0$. Consequently, $\langle TM_2^{\perp} | dH_1 \rangle = 0$ and M_2 is the final constraint submanifold.

Thus the Maxwell canonical system is (M_1, ω_1, M_2). We now investigate the nature of this final constraint submanifold M_2. First of all, we claim that $TM_2^{\perp} \subset \underline{TM_2}$. Indeed, if $b \oplus \tau \in TM_2^{\perp}$, then $\tau_\perp = 0$ so that $b \oplus \tau \in \underline{TM_1}$ and moreover, $\vec{\tau} = \vec{0}$ so that (6.14) is satisfied. Furthermore, $TM_2^{\perp} \neq \underline{TM_2}$. This is easily understood, as $b \oplus (0, \vec{\tau})$ with $\vec{\nabla} \cdot \vec{\tau} = 0$ is a member of $\underline{TM_2}$ but not of $TM_2^{\perp}$ unless $\vec{\tau} = 0$. Hence, M_2 is strictly coisotropic, and the canonical system (M_1, ω_1, M_2) is first-class. In particular, the constraint $\vec{\nabla} \cdot \vec{\pi} = 0$ is first-class.

The basic theorem of Sec. IV assures us that solutions to Hamilton's equations

$$(\iota(X)\omega_1 - dH_1) \circ j_2 = 0 \tag{6.17}$$

exist. To find these solutions, write $X = a \oplus \sigma$, and let $b \oplus \tau \in TM_1$ be arbitrary, $(A, \pi) \in M_2$. The equations of motion can then be written

$$\omega_1(a \oplus \sigma, b \oplus \tau)|(A, \pi) = dH_1(A, \pi) \cdot (b \oplus \tau). \tag{6.18}$$

Using (6.13) and (6.15), the above becomes

$$\int_{R^3} [\vec{\tau} \cdot \vec{a} - \vec{\sigma} \cdot \vec{b}] d\mu = \int_{R^3} [\vec{\tau} \cdot \vec{\pi} + (b_\perp)(\vec{\nabla} \cdot \vec{\pi}) + A_\perp (\vec{\nabla} \cdot \vec{\tau}) + (\vec{\nabla} \times \vec{A}) \cdot (\vec{\nabla} \times b)] d\mu. \tag{6.19}$$

As $(A, \pi) \in M_2$, the second term on the right-hand side of (6.19) drops out. After a rearrangement of the last term and an integration by parts, the right-hand side becomes

$$\int_{R^3} [\vec{\tau} \cdot (\vec{\pi} - \vec{\nabla} A_\perp) + (\vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - \Delta \vec{A}) \cdot \vec{b}] d\mu.$$

Comparing the left-hand side of (6.19) with this, we obtain

$$\begin{aligned} \frac{d\vec{A}}{dt} &:= \vec{a} = \vec{\pi} - \vec{\nabla} A_\perp, \\ \frac{d\vec{\pi}}{dt} &:= \vec{\sigma} = \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - \Delta \vec{A}, \\ \frac{dA_\perp}{dt} &:= a_\perp = \text{undetermined}. \end{aligned} \tag{6.20}$$

These are, of course, just Maxwell's equations. Performing a transverse-longitudinal decomposition of $\vec{A}, \vec{\pi}$ we obtain

$$\begin{aligned} \frac{dA_\perp}{dt} &= \text{undetermined} \\ \frac{d\vec{A}_L}{dt} &= -\nabla A_\perp, \\ \frac{d\vec{A}_T}{dt} &= \vec{\pi}_T, \\ \vec{\pi}_L &= 0, \\ \frac{d\vec{\pi}_T}{dt} &= -\Delta \vec{A}_T. \end{aligned} \tag{6.21}$$

Consequently, these equations determine $\vec{A}_T, \vec{\pi}_T$ uniquely from given initial data, but the evolution of $A_\perp$ and $\vec{A}_L$ is arbitrary.

Let us compare the equations of motion (6.21) and the known gauge freedom of the electromagnetic field with the predictions of the algorithm. In particular, (6.17) shows that the Hamiltonian vector field X is unique only up to elements of $\ker\omega_1 \cap \underline{TM_2} = \ker\omega_1$. Consequently, vector fields in $\ker\omega_1$ necessarily generate gauge transformations; if $V \in \ker\omega_1$, then V is of the form $(V_\perp, \vec{0}) \oplus 0$ and its effect is to generate arbitrary changes in the evolution of $A_\perp$. This is clearly consistent with the field equations. Turning now to the first-class secondary constraint (6.14), we wonder if it is the generator of physically irrelevant motions. From the geometric point of view, we are really asking whether or not the vector fields in $\underline{\ker\omega_2} = TM_2^{\perp}$ are gauge vector fields. If $W = (0, -\vec{\nabla} g) \oplus 0$, then

$$\iota(W)\omega_1 \cdot (b \oplus \tau) = -\int_{R^3} \vec{\tau} \cdot (\vec{\nabla} g) d\mu.$$

Demanding that $X - W$ satisfy (6.17) as well as X has the effect of replacing the second of equations (6.21) by

$$\frac{d\vec{A}_L}{dt} = -\vec{\nabla} A_\perp - \vec{\nabla} g$$

and leaving the others invariant. As $A_\perp$ is arbitrary to begin with, it is evident that this equation is completely equivalent to (6.21). The addition of $-\vec{\nabla} g$ to the right-hand side of this equation has no physical effect whatsoever. Thus, $\underline{\ker\omega_2}$ consists of gauge vector fields.[42]

From another standpoint, rather than writing

$$\iota(X - W)\omega_1 = dH_1$$

along M_2, we can put

$$\iota(X)\omega_1 = dH_1 + \iota(W)\omega_1.$$

Effectively, we are adding a term $g(\vec{\nabla} \cdot \vec{\tau})$ to the right-hand side of (6.13). In terms of the Hamiltonian itself, we are replacing $-(\vec{\nabla} A_\perp) \cdot \vec{\pi}$ by $-[\vec{\nabla}(A_\perp + g)] \cdot \vec{\pi}$. An integration by parts finally gives

$$dH_1 + \iota(W)\omega_1 = d[H_1 + \int_{R^3} g(\vec{\nabla} \cdot \vec{\pi}) d\mu].$$

The function whose differential appears on the right-hand side of this equation is none other than the pullback to M_1 to Dirac's extended Hamiltonian (5.8). With respect to the discussion in the last section, the above arguments show that for ordinary electromagnetism, one *can* add the first-class secondary constraints to the Hamiltonian since (i) these constraints are gauge, and (ii) no choice of gauge has been fixed in the Lagrangian (6.1). Notice also that we know (i) to be true regardless of the physical interpretation of the algorithm; in fact, we have not really physically interpreted the algorithm at all. As may be expected, this is due to the fact that the Maxwell theory is so "simple."

In the generic case, result (i) above will *not* be independent of the physical interpretation of the algorithm. Neither will (ii) be the case in general. One need not look far or long for a Lagrangian which has both of these problems, for consider

$$L(A, \dot{A}) = \int_{R^3} \left[\tfrac{1}{2}(\partial_\mu A^\nu)(\partial^\mu A_\nu) - A_\mu \partial^\mu \phi - \frac{\lambda}{2}\phi^2 \right] d\mu.$$

Is this Lagrangian to be regarded as electromagnetism in the Lorentz gauge, or is it an entirely different (massless, divergence-free, spin 1) field? This particular Lagrangian is discussed further in Ref. 14.

One should compare the above calculation with that given by Dirac.[7] Although this is not really a "working physicist" type calculation, these rigorous infinite-dimensional techniques are capable of rapidly producing results—in fact, they are indispensible when one discusses purely presymplectic systems. In finite dimensions, this geometric formalism is every bit as convenient to use as are the standard techniques.

VII. CONCLUSION

The algorithm we have presented completely solves, from a mathematical point of view, the problem of constrained symplectic systems (in both the finite- and infinite-dimensional cases). Even more significantly, it allows us to solve the Hamilton equations in the hitherto untreated presymplectic case. Combined with the geometric constraint theory of Śniatycki, Tulczyjew, and Lichnerowicz, it furnishes a powerful physical tool.

In addition to generalizing the Dirac–Bergmann theory of constraints, the algorithm has the advantage of being a global, manifestly coordinate-free theory. The algorithm is presented in a mathematically rigorous fashion which we feel is geometrically natural, intuitive, and useful from a practical (calculational) standpoint.

The algorithm provides insight into the old "controversy" of whether or not first-class secondary constraints really generate gauge transformations. It can be shown[14] that the algorithm cannot actually *prove* that all such constraints will beget physically irrelevant motions; nonetheless, equipped with a suitable physical interpretation, this algorithm furnishes a superior framework for discussing such questions. Consequently, these techniques may be of great value for the consideration of theories whose gauge properties at this time are poorly understood.

Our algorithm can also be adapted[43,44] to the Lagrangian case. Here, the Dirac–Bergmann formalism cannot be applied at all, and other proposed schemes have met with only limited success.[18,19] From the standpoint of this paper, the Lagrangian case can be regarded as a specific example of a presymplectic manifold $(TQ, FL^*\Omega)$, where Ω is the canonical symplectic structure on T^*Q and hence can be dealt with by the algorithm presented here. In this way the formal equivalence of the Hamiltonian and Lagrangian formalisms can be established even in the degenerate case.[43,45]

Since this algorithm enables us to treat *a priori* presymplectic systems as well as ordinary constrained symplectic systems, this work may engender motivation for inquiring as to how to quantize such presymplectic systems,[17] perhaps from the viewpoint of the geometric Kostant–Souriau quantization program.[46]

ACKNOWLEDGMENTS

The authors would particularly like to express their appreciation to Professor R. Gowdy, Professor R. Skinner, Professor C. Misner, Professor M. Marshall, and Professor P. Green of their respective institutions who provided many key insights and encouraging conversations. We would also like to thank both the referee and members of the Maryland Relativity Group, P. Yasskin and J. Isenberg for many helpful ideas and comments. This work forms part of a doctoral dissertation submitted to the Graduate School of the University of Maryland by one of us (M.J.G.).

[1]R. Abraham and J. Marsden, *Foundations of Mechanics* (Benjamin, New York, 1967).
[2]J.M. Souriau, *Structures des Systemes Dynamiques* (Dunod, Paris, 1970).
[3]N.M.J. Woodhouse, *Lectures on Geometric Quantization*, unpublished notes from a course given at Oxford University, 1975.
[4]C. Godbillon, *Géométrie Différentielle et Méchanique Analytique* (Hermann, Paris, 1969).
[5]P.G. Bergmann, Helv. Phys. Acta Suppl. IV **79** (1956) and references contained therein; P.A.M. Dirac, Can. J. Math. **2**, 129 (1950); Proc. Roy. Soc. (London) **A246**, 326 (1958).
[6]E.C.G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974); A. Hanson, T. Regge, and C. Teitelboim, Accademia Nazionale dei Lincei, Rome, #22 (1976).
[7]P.A.M. Dirac, *Lectures on Quantum Mechanics*, Belfer Graduate School of Science Monograph Series #2, (1964). The Dirac–Bergmann theory of constraints will be briefly reviewed at the beginning of Section V.
[8]P.A.M. Dirac, Proc. Roy. Soc. (London) A **246**, 333 (1958); Phys. Rev. **114**, 924 (1959).
[9]J. Śniatycki, "Topics in Canonical Dynamics I: Geometric Theory of Constraints," Proceedings of "Journées Relativistes 1970," Caen, 119; Ann. Inst. H. Poincaré **20**, 365 (1974).
[10]W.M. Tulczyjew, Symposia Math. **14**, 247 (1974).
[11]A. Lichnerowicz, C.R. Acad. Sci. Paris A **280**, 523 (1975).
[12]Ref. 7, pp. 23–4.
[13]See Section VI for an example in which one *is* allowed to "append" the first-class secondary constraints to the Hamiltonian.
[14]J.M. Nester and M.J. Gotay, "The Definition of a Physical State and the Problem of First-Class Secondary Constraints" (in preparation).
[15]J. Marsden, *Applications of Global Analysis in Mathematical Physics*, Lecture Notes #3 (Publish or Perish, Boston, 1974).
[16]P. Chernoff and J. Marsden, *Properties of Infinite-Dimensional Hamiltonian Systems*, Lecture Notes in Mathematics #425 (Springer, Berlin, 1974).
[17]M.J. Gotay and J.M. Nester, "The Quantization of Presymplectic Systems" (work in progress).
[18]H.P. Künzle, Ann. Inst. H. Poincaré A **11**, 393 (1969).
[19]S. Shanmugadhasan, J. Math. Phys. **14**, 677 (1973).
[20]W. Kundt, *Canonical Quantization of Gauge-Invariant Field Theories*, Springer Tracts in Modern Physics, **40**, 107 (1966).
[21]G. Hinds, "Foliations and the Dirac Theory of Constraints," Ph.D. thesis, Univ. of Maryland, 1965.
[22]Throughout this paper we assume that all physical systems under consideration are time-independent.
[23]A. Weinstein, *Lectures on Symplectic Manifolds*, AMS Conference Board of the Mathematical Sciences Regional Conference Series in Mathematics #29 (1977).
[24]S. Lang, *Differential Manifolds* (Addison-Wesley, Reading, Mass., 1972).
[25]Herein we establish our notation and terminology. All manifolds appearing in this paper are assumed to be C^∞-Banach manifolds. In particular, if N is a submanifold of a Banach manifold M, then N is assumed to be a Banach manifold in its own right, and hence must be *closed* as a subset of M. If $N \subset M$, then $\bar{N}$ denotes the topological closure of N in M. A manifold is said to be reflexive if its model space is reflexive. Sign and normalization conventions for the tensor algebras (differential forms and vector fields) are those of Ref. 24. We denote by the same symbol TM both the tangent bundle of M and the space of all vector fields tangent to M, etc. We designate the natural pairing $TM \times T^*M \to R$ by $\langle \mid \rangle$. The symbol d denotes the exterior derivative, L_X the Lie derivative along a vector field X, ι the interior product, and D the Frechét derivative. If γ is a p-form, and $X_1, \ldots, X_p$ are vector fields, then we have the following equivalence class of notations

$\iota(X_1)\cdots\iota(X_p)\gamma \equiv \langle X_1,\ldots,X_p|\gamma\rangle \equiv \gamma(X_p,\ldots,X_1)$.

We define $\ker\gamma = \{ Y\in TM$ such that $\iota(Y)\gamma = 0\}$. The symbol "$|N$" denotes "restriction to the submanifold N." If $N\subset M$ with inclusion j and γ is a form on M, we denote by either $\gamma|N$ or $\gamma\circ j$ the restriction of γ to points of N, and by γ_N the pullback $j^*\gamma$ of γ to N. If $S\subset TN$, we put $\underline{S} := j_* S\subset TM$. We now briefly explain how one calculates locally, following Refs. 1 and 16. If $U\subset E$ is a chart on a manifold Q, then $T^*U = U\times E^*$ is a chart on T^*Q, and a point $m\in T^*Q$ has the local representation $m=(x,\sigma)$ where $x\in U$, $\sigma\in E^*$. A chart on $T(T^*Q)$ is $T(T^*U) = (U\times E^*)\oplus(E\times E^*)$. Thus a tangent vector X to T^*Q has the local representation $X(m) = (x,\sigma)\oplus(a,\pi)$ where $a\in E$ and $\pi\in E^*$. We will often suppress the base point (x,σ) and simply write this as $X =: a\oplus\pi$. Thus, for example, if α is a 1-form on T^*Q, the interior product $\iota(X)\alpha(m)$ is written locally $\alpha(x,\sigma)\cdot(a\oplus\pi)$.

[26] R. Hermann, *Lie Algebras and Quantum Mechanics* (Benjamin, New York, 1970).

[27] A. Weinstein, Adv. Math. **6**, 329 (1971).

[28] J. Marsden, Proc. Am. Math. Soc. **32**, 590 (1972).

[29] H.P. Künzle, J. Math. Phys. **13**, 739 (1972).

[30] This can be seen as follows: Let S be a subspace of E; we claim that $(S^\perp)^\perp = \bar{S}$. For if $\beta\in S^\perp$, then $\beta\in\bar{S}^\perp$ by continuity, and consequently $\bar{S}\subset(S^\perp)^\perp$. Conversely, suppose that $\bar{v}\in(S^\perp)^\perp - \bar{S}$. As $\bar{v}\notin\bar{S}$, by the Hahn-Banach theorem there exists a $\beta\in E^*$ such that $\langle v|\beta\rangle = 0$ for all $v\in S$ with $\langle\bar{v}|\beta\rangle\neq 0$. But this contradicts the fact that $\bar{v}\in(S^\perp)^\perp$. Thus $(S^\perp)^\perp\subset\bar{S}$ and the claim is proven. However, we see that the corresponding result $(\Lambda^\perp)^\perp = \bar{\Lambda}$ for $\Lambda\subset E^*$ will not be true in general as $\Lambda^\perp$ is defined as a subspace of E, not E^{**}, and consequently we cannot apply the Hahn-Banach theorem as before. On the other hand, if E is reflexive so that $E = E^{**}$, then the above arguments apply and in this case, $(\Lambda^\perp)^\perp = \bar{\Lambda}$.

[31] An interesting corollary to this result is the following: Let M be reflexive, and suppose that ω is topologically closed. Then weak nondegeneracy is equivalent to strong nondegeneracy.

[32] Insofar as the algorithm *itself* is concerned, the equations (4.1) can be completely general, i.e., ω_1 can be *any* 2-tensor and α_1 *any* 1-form. It is not known to the authors whether or not there is any physical significance in replacing the exact 1-form dH_1 by α_1, i.e., are there any physical systems which require for their description a Hamiltonian 1-form rather than a Hamiltonian function?

[33] As was mentioned in the Introduction, we must then integrate X. In finite dimensions, the equations defining the integral curves of X are ordinary differential equations and there is no problem in obtaining unique (local) solutions. In the infinite-dimensional case, these will be partial differential equations, and the situation is correspondingly much more complicated.

[34] We assume that all of the M_l appearing in the algorithm are in fact Banach submanifolds. In practice, of course, this assumption will not generally be valid. In such case, one must resort to standard tricks, e.g., cut the manifold into pieces where everything is nice and work locally. For a nice discussion of how to proceed see Hinds (Ref. 21) and Śniatycki (Ref. 10).

[35] In this context, we must emphasize that we are restricting (4.1) to M_2. Equation (4.2) is *not* the pullback of (4.1).

[36] A fourth possibility arises in the infinite-dimensional case where the algorithm may not terminate at all. In this situation the final constraint submanifold can be taken to be the intersection of all the submanifolds M_l.

[37] This is contrary to Kundt (Ref. 20) who views the u_A appearing in (5.1) as determined from the outset.

[38] Here, "first-" and "second-class" are defined as in Sec. III but with respect to the Ω-symplectic complement of TM_l in $T(T^*Q)$, which we denote by $TM_l^{\perp\perp}$. It is trivial to establish that $\underline{TM}_l^{\perp} = TM_l^{\perp\perp}\cap\underline{TM}_1$.

[39] Dirac, *op. cit.* Ref. 6, pp. 25–6.

[40] In the following discussion a number of results will be quoted without proof or extensive explanation. For further details, consult Ref. 14.

[41] This is in contradistinction to Dirac's intimation that a proof of the gaugeness of the first-class secondary constraints should exist (Ref. 12). In fact, Dirac's claim that one can generate first-class secondary constraints by taking Poisson brackets of first-class primary constraints is incorrect, as the Poisson bracket of two first-class primary constraints is necessarily another primary constraint.

[42] There is a much easier way to tell whether or not a given vectorfield is gauge than the type of argument presented here. This technique, which we call the "gauge vector field algorithm" is capable of generating all the gauge vector fields in a particular theory (or at least all those vector fields which preserve any choices of gauge made in the Hamiltonian). For details, see Ref. 14.

[43] M.J. Gotay and J.M. Nester, "Presymplectic Lagrangian Systems I: The Constraint Algorithm and the Equivalence Theorem," Univ. of Maryland preprint PP-78-172 (1978); submitted for publication to Ann. Inst. H. Poincaré.

[44] J.M. Nester and M.J. Gotay, "Presymplectic Lagrangian Systems II: The Second-Order Equation Problem" (in preparation).

[45] J. Śniatycki, *Lagrangian Systems, Canonical Systems, and Legendre Transformations*, Proc. Biennial Seminar of the Canadian Math. Cong. **2**, 125 (1972).

[46] D. Simms and N.M.J. Woodhouse, *Lectures on Geometric Quantization*, Lecture Notes in Physics #53 (Springer, Berlin, 1976).

A classification of four-dimensional Lie superalgebras

Nigel Backhouse

Department of Applied Mathematics and Theoretical Physics, The University, P. O. Box 147, Liverpool L69 3BX, England

(Received 13 December 1977; revised manuscript received 7 March 1978)

A classification is made of all Lie superalgebras (graded Lie algebras) of maximum dimension four.

1. INTRODUCTION

In the rapidly growing field of Lie superalgebras (LS), the major effort has been directed towards establishing a theory for some suitable definition of simplicity. The simple LS have now been classified by several authors independently, Refs. 1–7, and a start has been made on their representation theory, Refs. 8–16. Disappointingly, perhaps, for only one class of simple LS, the orthosymplectic series, is it true that all finite-dimensional representations are completely reducible, see Refs. 10, 12–15. Other departures from ordinary Lie theory include the fact that Lie's theorem is not valid; that Cartan's criterion for simplicity only works in one direction; that there is no obvious analog of Levi's theorem; and that there can exist zero divisors in the enveloping algebra. The emphasis in the literature on the class of simple LS is hardly surprising in view of the important role played in applications by their ordinary Lie theoretic analog. However, nonsimple LS are very plentiful and as yet have received little attention. In the present paper we give a classification of real LS, which are not Lie algebras (LA), up to dimension four, so providing another source of examples which exhibit some of the similarities and differences between LA and LS. We have already classified LS up to dimension three in Ref. 17. We note that the smallest simple LS, which is not a LA, the so-called di-spin algebra considered in Ref. 9, has dimension five. It follows that we do not encounter any simple LS in our classification and that all of those we classify are solvable.

We recall that a LS $\mathbf{L}=\mathbf{L}_0 \oplus \mathbf{L}_1$ consists of an even part $\mathbf{L}_0$, which is an ordinary LA, and an odd part $\mathbf{L}_1$, which in particular is an $\mathbf{L}_0$-module by restriction of the adjoint representation. We denote the elements of $\mathbf{L}_0$ (resp. $\mathbf{L}_1$) by Latin

TABLE I. Trivial algebras

Type	L	L′	Relations	Comments
(0,1)	A	$A_{1,1}$	$[\alpha,\alpha]=0$	Abelian
(1,1)	B	$A_{2,1}$	$[a,\alpha]=\alpha$	
(2,1)	C^1_p	$A_{3,3}, A_{3,4}$ $(p=\pm 1)$ $A^p_{3,5}$ $(p\neq\pm 1)$	$[a,b]=b$, $[a,\alpha]=p\alpha$	$p\neq 0$
(1,2)	C^2_p	$A_{3,3}, A_{3,4}$ $(p=\pm 1)$ $A^p_{3,5}$ $(p\neq\pm 1)$	$[a,\alpha]=\alpha, [a,\beta]=p\beta$	$0<\lvert p\rvert\leqslant 1$
	C^3	$A_{3,1}$	$[a,\beta]=\alpha$	Nilpotent
	C^4	$A_{3,2}$	$[a,\alpha]=\alpha$, $[a,\beta]=\alpha+\beta$	
	C^5_p	$A_{3,6}$ $(p=0)$ $A^p_{3,}$[illegible] $(p\neq 0)$	$[a,\alpha]=p\alpha-\beta$, $[a,\beta]=\alpha+p\beta$	$p\geqslant 0$
(3,1)	D^1	$A_{4,3}$	$[b,c]=a$, $[b,\alpha]=\alpha$	
	D^2_p	$A^p_{4,2}$	$[a,c]=a$, $[b,c]=a+b$, $[c,\alpha]=q\alpha$	$pq\neq 0$
	D^3_{pq}	$A^{p-q}_{4,5}$	$[a,c]=pa-b$, $[b,c]=a+pb$, $[c,\alpha]=q\alpha$	$q\neq 0$
(2,2)	D^5	$A^0_{4,9}$	$[a,\alpha]=\alpha$, $[a,\beta]=\beta$, $[b,\beta]=\alpha$	
	D^6	$A_{4,12}$	$[a,\alpha]=\alpha$, $[a,\beta]=\beta$, $[b,\alpha]=-\beta$, $[b,\beta]=\alpha$	
	D^1_{pq}	$A^{pq}_{4,5}$	$[a,b]=b$, $[a,\alpha]=p\alpha$, $[a,\beta]=q\beta$	$pq\neq 0, p\geqslant q$
	D^8_p	$A^{1/p}_{4,2}$	$[a,b]=b$, $[a,\alpha]=p\alpha$, $[a,\beta]=\alpha+p\beta$	$p\neq 0$
	D^9_{pq}	$A^{1/q,p/q}_{4,6}$	$[a,b]=b$, $[a,\alpha]=p\alpha-q\beta$, $[a,\beta]=q\alpha+p\beta$	$q>0$
	D^{10}_p	$A^p_{4,9}$	$[a,b]=b,[a,\alpha]=(p+1)\alpha$, $[a,\beta]=p\beta$, $[b,\beta]=\alpha$	
(1,3)	D^{11}_{pq}	$A^{pq}_{4,5}$	$[a,\alpha]=\alpha$, $[a,\beta]=p\beta$, $[a,\gamma]=q\gamma$	$0<\lvert p\rvert\leqslant\lvert q\rvert\leqslant 1$
	D^{12}	$A_{4,3}$	$[a,\alpha]=\alpha$, $[a,\gamma]=\beta$	
	D^{13}_p	$A^p_{4,2}$	$[a,\alpha]=p\alpha$, $[a,\beta]=\beta$, $[a,\gamma]=b+\gamma$	$p\neq 0$
	D^{14}_{pq}	$A^{pq}_{4,6}$	$[a,\alpha]=p\alpha$, $[a,\beta]=q\beta-\gamma$, $[a,\gamma]=\beta+q\gamma$	$p\neq 0, q\geqslant 0$
	D^{15}	$A_{4,1}$	$[a,\beta]=\alpha$, $[a,\gamma]=\beta$	Nilpotent
	D^{16}	$A_{4,4}$	$[a,\alpha]=\alpha$, $[a,\beta]=\alpha+\beta$, $[a,\gamma]=\beta+\gamma$	

 0022-2488/78/1911-2400$1.00

TABLE II. Nontrivial algebras.

Type	L	Relations	Comments
(1,1)	$(A_{1,1}+A)$	$[a,\alpha]=a$	Nilpotent
(2,1)	$(C^1_{1/2})$	$[a,b]=b$, $[a,\alpha]=\frac{1}{2}\alpha$, $[\alpha,\alpha]=b$	
(1,2)	$(A_{1,1}+2A)^1$	$[\alpha,\alpha]=a$, $[\beta,\beta]=a$	Nilpotent
	$(A_{1,1}+2A)^2$	$[\alpha,\alpha]=a$, $[\beta,\beta]=-a$	Nilpotent
(3,1)	$(A_{3,1}+A)$	$[b,c]=a$, $[\alpha,\alpha]=a$	Nilpotent
	$(D^3_{p,-1/2})$	$[a,b]=b$, $[a,c]=pc$, $[a,\alpha]=\frac{1}{2}\alpha$, $[\alpha,\alpha]=b$	$p\neq 0$
	$(D^2_{-1/2})^1$	$[a,b]=b$, $[a,c]=b+c$, $[a,\alpha]=\frac{1}{2}\alpha$, $[\alpha,\alpha]=b$	
	$(D^2_{-1/2})^2$	$[a,b]=b$, $[a,c]=-b+c$, $[a,\alpha]=\frac{1}{2}\alpha$, $[\alpha,\alpha]=b$	
(2,2)	$(D^7_{1/2\,1/2})^1$	$[a,b]=b$, $[a,\alpha]=\frac{1}{2}\alpha$, $[a,\beta]=\frac{1}{2}\beta$, $[\alpha,\alpha]=b$, $[\beta,\beta]=b$	
	$(D^7_{1/2\,1/2})^2$	$[a,b]=b$, $[a,\alpha]=\frac{1}{2}\alpha$, $[a,\beta]=\frac{1}{2}\beta$, $[\alpha,\alpha]=b$, $[\beta,\beta]=-b$	
	$(D^7_{1/2\,1/2})^3$	$[a,b]=b$, $[a,\alpha]=\frac{1}{2}\alpha$, $[a,\beta]=\frac{1}{2}\beta$, $[\alpha,\alpha]=b$	
	$(D^7_{1-p,p})$	$[a,b]=b$, $[a,\alpha]=p\alpha$, $[a,\beta]=(1-p)\beta$, $[\alpha,\beta]=b$	$p\leqslant\frac{1}{2}$
	$(D^8_{1/2})$	$[a,b]=b$, $[a,\alpha]=\frac{1}{2}\alpha$, $[a,\beta]=\alpha+\frac{1}{2}\beta$, $[\beta,\beta]=b$	
	$(D^9_{1/2,p})$	$[a,b]=b$, $[a,\alpha]=\frac{1}{2}\alpha-p\beta$, $[a,\beta]=p\alpha+\frac{1}{2}\beta$, $[\alpha,\alpha]=b$, $[\beta,\beta]=b$	$p>0$
	$(D^{10}_0)^1$	$[a,b]=b$, $[a,\alpha]=\alpha$, $[b,\beta]=\alpha$, $[\beta,\beta]=a$, $[\alpha,\beta]=-\frac{1}{2}b$	
	$(D^{10}_0)^2$	$[a,b]=b$, $[a,\alpha]=\alpha$, $[b,\beta]=\alpha$, $[\beta,\beta]=-a$, $[\alpha,\beta]=\frac{1}{2}b$	
	$(2A_{1,1}+2A)^1$	$[\alpha,\alpha]=a$, $[\beta,\beta]=b$	Nilpotent
	$(2A_{1,1}+2A)^2$	$[\alpha,\alpha]=a$, $[\beta,\beta]=b$, $[\alpha,\beta]=a$	Nilpotent
	$(2A_{1,1}+2A)^3_p$	$[\alpha,\alpha]=a$, $[\beta,\beta]=b$, $[\alpha,\beta]=p(a+b)$	$p>0$ Nilpotent
	$(2A_{1,1}+2A)^4_p$	$[\alpha,\alpha]=a$, $[\beta,\beta]=b$, $[\alpha,\beta]=p(a-b)$	$p>0$ Nilpotent
	(C^1_1+A)	$[a,b]=b$, $[a,\alpha]=\alpha$, $[\alpha,\beta]=b$	
	$(C^1_{1/2}+A)$	$[a,b]=b$, $[a,\alpha]=\frac{1}{2}\alpha$, $[\alpha,\alpha]=b$	
	$(C^2_{-1}+A)$	$[a,\alpha]=\alpha$, $[a,\beta]=-\beta$, $[\alpha,\beta]=b$	Jordan–Wigner quantization, see Ref. 9.
	(C^3+A)	$[a,\beta]=\alpha$, $[\beta,\beta]=b$	Nilpotent
	(C^5_0+A)	$[a,\alpha]=-\beta$, $[a,\beta]=\alpha$, $[\alpha,\alpha]=b$, $[\beta,\beta]=b$	
(1,3)	$(A_{1,1}+3A)^1$	$[\alpha,\alpha]=a$, $[\beta,\beta]=a$, $[\gamma,\gamma]=a$	Nilpotent
	$(A_{1,1}+3A)^2$	$[\alpha,\alpha]=a$, $[\beta,\beta]=a$, $[\gamma,\gamma]=-a$	Nilpotent

letters (resp. Greek letters) taken from the beginning of the alphabet. **L** possesses a bilinear bracket multiplication [,] which satisfies the following commutativity and Jacobi conditions:

$$[a,b]=-[b,a], \tag{1}$$

$$[a,\alpha]=-[\alpha,a], \tag{2}$$

$$[\alpha,\beta]=[\beta,\alpha], \tag{3}$$

for all $a,b\in\mathbf{L}_0$, $\alpha,\beta\in\mathbf{L}_1$, and

$$[[a,b],c]+[[b,c],a]+[[c,a],b]=0, \tag{4}$$

$$[[a,b],\alpha]+[[b,\alpha],a]+[[\alpha,a],b]=0, \tag{5}$$

$$[[a,\alpha],\beta]+[[\alpha,\beta],a]-[[\beta,a],\alpha]=0, \tag{6}$$

$$[[\alpha,\beta],\gamma]+[[\beta,\gamma],\alpha]+[[\gamma,\alpha],\beta]=0, \tag{7}$$

for all $a,b,c\in\mathbf{L}_0$ and $\alpha,\beta,\gamma\in\mathbf{L}_1$. Note that Eq. (6) differs from the others in that the cyclic symmetry of its terms is broken by a minus sign.

We say that $\mathbf{L}=\mathbf{L}_0\oplus\mathbf{L}_1$ and $\mathbf{L}'=\mathbf{L}_0'\oplus\mathbf{L}_1'$ are equivalent if there are isomorphisms $\mathbf{L}_0\longleftrightarrow\mathbf{L}_0'$ and $\mathbf{L}_1\longleftrightarrow\mathbf{L}_1'$ which preserve the bracket multiplication. We can ask the question: given a LA $\mathbf{L}_0$ and an $\mathbf{L}_0$-module **M**, how many inequivalent LS $\mathbf{L}_0=\mathbf{L}_0\oplus\mathbf{L}_1$ can we construct, where $\mathbf{L}_1$ and **M** are equivalent as $\mathbf{L}_0$-modules. Answering this question in low dimensional cases is the basis for our classification. We have found it convenient, both for the carrying out of our computations and for the tabular presentation of our results, to distinguish two types of LS: We say that **L** is trivial if $[\mathbf{L}_1,\mathbf{L}_1]=\{0\}$; otherwise **L** is nontrivial. The point is that a nontrivial algebra can be trivialized simply by putting to zero all of its anticommutators. The reverse process produces nontrivial algebras from the set of trivial algebras, which are more easily classified. It is also worth noting that the structure constants of a trivial LS, **L** say, can be interpreted as the structure constants of an associated LA, **L**′ say, provided that we replace the zero anticommutators of **L** by zero commutators of **L**′. However, under this correspondence, inequivalent LS can lead to equivalent LA.

2. TABULATIONS

Here we tabulate into families of equivalence classes the real indecomposable LS of maximum dimension four, which are not LA. The trivial and nontrivial algebras are tabulated separately and according to dimension structure: We say that $\mathbf{L}=\mathbf{L}_0\oplus\mathbf{L}_1$ is an (m,n) algebra if $\dim\mathbf{L}_0$ (resp. $\mathbf{L}_1$) is m (resp. n). For the labeling of the trivial algebras, the letters A,B,C,D with integral superscripts i and real subscripts p,q, denote the equivalence classes of algebras of dimension d, where $d=1$ for A, $d=2$ for B, $d=3$ for C, $d=4$ for D. The superscript i is omitted whenever its range is just the integer one. We also give the symbols, according to Ref. 18, of the associated LA. For the nontrivial algebras, we add to the bracketed symbol for the corresponding trivial algebra, where necessary, an integral superscript and a real subscript. With a single exception, all zero commutation/anticommutation relations are omitted.

[1]V.G. Kac, Ad. Math. **26**, 8–96 (1977).
[2]A. Pais and V. Rittenberg, J. Math. Phys. **16**, 2062–73 (1975).
[3]P.G.O. Freund and I. Kaplansky, J. Math. Phys. **17**, 228–31 (1976).
[4]D.Z. Djokovic, J. Pure Appl. Alg. **7**, 217–30 (1976).
[5]W. Nahm and M. Scheunert, J. Math. Phys. **17**, 868–79 (1976).
[6]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1626–39 (1976).
[7]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1640–4 (1976).
[8]L. Ross, Trans. Amer. Math. Soc. **120**, 17–23 (1965).
[9]L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573–603 (1975).
[10]L. Corwin, "Finite-dimensional representations of semisimple graded Lie algebras," Preprint.
[11]G. Hochschild, Ill. J. Math. **20**, 107–23 (1976).
[12]D.Z. Djokovic and G. Hochschild, Ill. J. Math. **20**, 134–43 (1976).
[13]D.Z. Djokovic, J. Pure Appl. Alg. **9**, 25–38 (1976).
[14]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146–54 (1977).
[15]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155–62 (1977).
[16]N.B. Backhouse, "Some Aspects of Graded Lie Algebras" in Proc. 5th Int. Colloq. Group Theoretical Methods in Physics, Montreal, 1976 (to be published).
[17]N.B. Backhouse, "On the Construction of Graded Lie Algebras" in Proc. 6th Int. Colloq. Group Theoretical Methods in Physics, Tübingen, 1977 (to be published).
[18]J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 986–94 (1976).