# Fourth degree Casimir operator of the semisimple graded Lie algebra (Sp(2*N*); 2*N*)

Miroslav Bednář and Vladimir Sachl

*Institute of Physics, Czechoslovak Academy of Sciences, Prague, Czechoslovakia*
(Received 29 April 1977)

The Casimir operators of the graded Lie algebra (Sp(2$N$);2$N$) [denoted also by OSp(1|2$N$) in the literature] are discussed. A general method, according to which the higher degree Casimir operators of the graded Lie algebras, in our case of the (Sp(2$N$);2$N$), can be constructed is developed. It is shown that the third degree Casimir operator of this graded Lie algebra does not exist. The Casimir operator of the fourth degree is derived explicitly.

## I. INTRODUCTION

In recent years much attention has been devoted in the physical and mathematical literature to the graded Lie algebras (GLA's). By various methods in a number of papers the problem of classification of these GLA's (see Refs. 1–5) from a different point of view has been studied.

For physical applications of graded Lie algebras it is, of course, necessary to know, besides the classification of the GLA's the classification of their irreducible representations. However, the problem of classification (and consequently of the construction) of the irreducible representations of the GLA's has not yet been satisfactorily solved.

Only for the simplest GLA's: (Sp(2);2) and (SL(2) $\otimes$GL(1); $2\otimes\bar{2}$) (see Refs. 2 and 6), has the irreducible representations been classified and explicitly constructed. We here denote the GLA's according to the notation of Ref. 3.

A question of considerable importance for the solution of the up to date open problem of classification of the irreducible representations of the GLA's is the problem of constructing the complete set of independent Casimir operators of the GLA's (similar to the classical Lie algebras). Pais and Rittenberg[2] introduce this problem as one of those which has to be solved. The reason for this consists in the fact that the Casimir operators possess for the classification of the irreducible representations of the graded Lie algebras a nonsubstitutional role, contrary to Lie algebras in which representations may be classified either with the aid of the highest weight (which is the most general, and in literature the most applied approach) or by an equivalent way through the eigenvalues of the Casimir operators. A general definition of the Casimir operators has been given in Ref. 7 in which Backhouse (for the graded Lie algebras) has generalized the concepts well known from the representation theory of Lie algebras.

Nevertheless, to construct explicitly the Casimir operators of higher degree according to the approach given by Backhouse (as he himself emphasizes) is a very tedious problem.

In this way, the effort to find the simplest, and for practical purposes useful, method for the construction of the higher degree Casimir operators of the graded Lie algebras arises naturally. [For Lie algebras, e.g., SU($n$), O($2n+1$), O($2n$), and Sp($2n$) such a simplified method of the contruction of the higher degree Casimir operators has been developed by Okubo, see, e.g., Ref. 8.] At the same time, it is useful to search for expressions of higher degree Casimir operators of the graded algebras, such that for Okubo's type, higher degree Casimir operators of the Lie subalgebras will appear explicitly along with the eigenvalues which are possible to be found according to already worked out methods (see, e.g., Refs. 8 and 9).

The purpose of this paper is to discuss and derive the higher degree Casimir operators of one of the simplest semisimple graded Lie algebras, i.e., the algebra (Sp(2$N$);2$N$), which has been mentioned in the classical paper of Pais and Rittenberg.[2] The suggested method enables us to construct all the Casimir operators of the graded Lie algebra (Sp(2$N$);2$N$) (i.e., of the 2,4,6,..., 2$N$th degree) which include the (same degree, independent) Casimir operators of the Lie subalgebras Sp(2$N$).

It is shown that the third degree Casimir operators of the GLA's (Sp(2$N$);2$N$) do not exist. The Casimir operators of the fourth degree are then found explicitly.

As to the content of the paper: The structure of the graded Lie algebra (Sp(2$N$);2$N$) is recalled in the second section in detail. In Sec. III the higher degree Casimir operators of the algebra (Sp(2$N$); 2$N$) are discussed and some of them are given in explicit form. Useful mathematical identities are given in Appendix A. Appendix B contains the full derivation of the fourth degree Casimir operator.

## II. STRUCTURE OF THE SYMPLECTIC LIE ALGEBRA Sp(2*N*) AND GRADED LIE ALGEBRA (Sp(2*N*); 2*N*)

### A. Lie algebra Sp(2*N*)

The symplectic group Sp(2$N$) is formed by the linear transformations[10] (in the 2$N$-dimensional space) which leave the bilinear form

$$[x,y]=\sum_{i,j=-N} g_{ij}x_i y_j =\sum_{i=1}^{N}(x_i y_{-i}-x_{-i}y_i) \tag{1}$$

 0022-2488/78/1907-1487$1.00  

invariant.[8] Here the components $g_{ij}$ of the metric tensor are defined as

$$g_{ij}=\epsilon_i\delta_{i,-j}, \tag{2}$$

with

$$\epsilon_i=\begin{cases}1 \text{ for } i>0,\\ 0 \text{ for } i=0,\\ -1 \text{ for } i<0.\end{cases} \tag{3}$$

The generators[8] $X_{ij}$ $(i,j=-N,\ldots,-1,1,\ldots,N)$ (denoted as $X_j^i$ in Ref. 8) of its algebra fulfil the commutation relations

$$[X_{ij},X_{kl}]=\delta_{kj}X_{il}-\delta_{il}X_{kj}+\epsilon_i\epsilon_j\delta_{-lj}X_{k-i} +\epsilon_j\epsilon_k\delta_{-ik}X_{-jl}. \tag{4}$$

They may be represented in the coordinate representation by the differential operators as

$$X_{ij}=x_i\frac{\partial}{\partial x_j}-\epsilon_i\epsilon_j x_{-j}\frac{\partial}{\partial x_{-i}} \tag{5}$$

from which the following relations for the generators $X_{ij}$ follows:

$$\epsilon_i\epsilon_j X_{ij}=-X_{-j-i}, \tag{6}$$

$$\sum_{j=-N}^{N}X_{jj}=0. \tag{7}$$

The dimension of the algebra Sp($2N$) (the number of independent generators $X_{ij}$) is $N(2N+1)$. By using the differential form for the generators $X_{ij}$, Eq. (5), the equation which describes the transformation properties of the $2N$-dimensional vector $\mathbf{x}$ [with components $x_k$, $k=-N,\ldots,-1,1,\ldots,N$)] is given by

$$[X_{ij},x_k]=\delta_{kj}x_i-\epsilon_i\epsilon_j\delta_{-ik}x_{-j} \equiv (Q_{ij})_{lk}x_l. \tag{8}$$

Here the matrices $Q_{ij}$ consisting of the matrix elements

$$(Q_{ij})_{lk}=\delta_{il}\delta_{kj}-\epsilon_i\epsilon_j\delta_{-ik}\delta_{-jl} \tag{9}$$

represent the Sp($2N$) generators in the lowest, i.e., the $2N$-dimensional representation.

Denoting by $g$ the matrix whose matrix elements $g_{ij}$ are given by Eq. (2), we can write down the relation between the matrices $Q_{ij}$ and their transposed matrices $Q_{ij}^T$ in the following form,

$$Q_{ij}g=-gQ_{ij}^T$$

(from which it follows that the $gQ_{ij}$ is the symmetric matrix).

### B. The graded Lie algebra (Sp(2*N*);2*N*)

The graded Lie algebra (Sp($2N$);$2N$) is formed by the operators $X_{ij}$ and $V_i$ $(i,j=-N,\ldots,-1,1,\ldots,N)$. The $X_{ij}$ are at the same time the generators of the algebra Sp($2N$) and the $V_i$ components of the irreducible [w.r.t. Sp($2N$)] operator, which transforms under the lowest $2N$-dimensional representation of the group Sp($2N$). The transformation properties of the operators $V_i$ w.r.t $X_{ij}$ are therefore identical with those of $x_i$ (components of $\mathbf{x}$) in Eq. (8) and the equations

$$[X_{ij},V_k]=(Q_{ij})_{lk}V_l =\delta_{kj}V_i-\epsilon_i\epsilon_j\delta_{-ik}V_{-j}. \tag{11}$$

The GLA (Sp($2N$);$2N$), generated by the operators $X_{ij}$ and $V_i$, is defined besides Eqs. (4) and (11) by the "products" of two operators $V_k$ and $V_l$ which are represented by the anticommutators $\{V_k, V_l\}$. The algebra of operators $X_{ij}$ and $V_k$ is then closed if the relation

$$\{V_k,V_l\}=(gQ_{ij})_{kl}X_{ji} \tag{12}$$

is fulfilled.

By using Eqs. (2), (9), (11), and (12) the Jacobi identities

$$[V_j,\{V_k,V_l\}]+[V_k,\{V_l,V_j\}]+[V_l,\{V_j,V_k\}]=0, \tag{13}$$

$$[X_{ij},\{V_k,V_l\}]+\{[V_k,X_{ij}],V_l\}+\{[V_l,X_{ij}],V_k\}=0, \tag{14}$$

can be immmediately verified. As a result, by Eqs. (4), (11), and (12), the graded Lie algebra (Sp($2N$);$2N$) is defined.

Equation (12) by the use of Eqs. (2), (9), and (6) may be also rewritten into a simpler and useful form

$$\{V_k,V_l\}=2\epsilon_k X_{l-k}=2\epsilon_l X_{k-l}. \tag{15}$$

## III. CONSTRUCTION OF HIGHER INVARIANTS OF THE LIE ALGEBRA Sp(2*N*) AND GRADED ALGEBRA (Sp(2*N*); 2*N*)

### A. The construction of the irreducible tensors and invariants [w.r.t. the group Sp(2*N*)] with the aid of the operators $X_{ij}$ and $V_k$

It is well known[8] that with the aid of the generators $X_{ij}$ of the symplectic algebra Sp($2N$) it is possible to construct the irreducible tensors

$$T_{ij}^{(p)}=\sum_{i_1,i_2,\ldots,i_p=-N}^{N}X_{ii_1}X_{i_1i_2}X_{i_2i_3}\times\cdots X_{i_{p-1}i_p}X_{i_pj}, \tag{16}$$

which have the same transformation properties as the generators $X_{ij}$, i.e., they fulfil the commutation relations

$$[X_{ij},T_{kl}^{(p)}]=\delta_{kj}T_{il}^{(p)}-\delta_{il}T_{kj}^{(p)} +\epsilon_i\epsilon_j\delta_{-lj}T_{k-i}^{(p)}+\epsilon_j\epsilon_k\delta_{-ik}T_{-jl}^{(p)}. \tag{17}$$

Analogously, by using relations (4) and (11), we find that the operator $V_k'$, defined by the equation

$$V_k'=\sum_{j=-N}^{+N}X_{kj}V_j, \tag{18}$$

has the same transformation properties as the operator $V_k$, i.e.,

$$[X_{ij},V_k']=\delta_{kj}V_i'-\epsilon_i\epsilon_j\delta_{-ik}V_{-j}'. \tag{19}$$

The operator $V_k''$ also has the same transformation properties as the operators $V_k$ and $V_k'$, in case it is defined as

$$V_k''=\sum_{j=-N}^{+N}X_{kj}V_j'. \tag{20}$$

From the above-mentioned construction of operators $V_k'$ and $V_k''$ it is apparent how to further construct the operators of this kind.

In the frame of bilinear combinations of generators $V_k$, there is the following one,[11]

$$T'_{kl} = \epsilon_{-l} V_{-l} V_k. \tag{21}$$

This operator is remarkable by the fact that its transformation properties [w.r.t. Sp(2$N$)] are identical with the transformation properties of the generator $X_{kl}$, i.e., the following relations are valid,

$$[X_{ij}, T'_{kl}] = \delta_{kj} T'_{il} - \delta_{il} T'_{kj} + \epsilon_i \epsilon_j \delta_{-lj} T'_{k-i} + \epsilon_j \epsilon_k \delta_{-ik} T'_{-jl}. \tag{22}$$

If we have any two operators $T_{kl}$, $T'_{kl}$, which transform in the same way as the generators $X_{kl}$, then the expression $\sum_{k,l=-N}^{N} T_{kl} T'_{lk}$, constructed with the help of them, represents an invariant w.r.t. $X_{ij}$, i.e., we have

$$[X_{ij}, \sum_{k,l=-N}^{+N} T_{kl} T'_{lk}] = 0. \tag{23}$$

Similarly, if we have two vectors $V'_k$, $V''_k$ which transform [under the Sp(2$N$)] as the vector $x_k$, then the bilinear combination of operators $V'_k$ and $V''_k$, defined as

$$\sum_{k,k'=-N}^{+N} V'_k g_{kk'} V''_{k'} \tag{24}$$

also represents an Sp(2$N$)-invariant, so that the commutator is equal to zero,

$$[X_{ij}, \sum_{k,k'=-N}^{+N} V'_k g_{kk'} V''_{k'}] = 0. \tag{25}$$

### B. Construction of the invariants of the graded Lie algebra (Sp(2$N$); 2$N$)

As we have already mentioned in the Introduction, the construction of the second degree Casimir operators of GLA's has been discussed in a number of papers: In the case of the algebra (Sp(2$N$);2$N$) it is possible to write, with the aid of generators $X_{ij}$ and $V_i$, the Casimir operator of the second degree in the form

$$K_2 = X_{ij} X_{ji} + V_i g_{ij} V_j. \tag{26}$$

By using relations (11) and (12) it is easily found that $V_m$ commutes with the operator $K_2$,

$$[K_2, V_m] = 0.$$

The relation

$$[K_2, X_{ij}] = 0 \tag{28}$$

is evident, as the two terms in the $K_2$ are Sp(2$N$)-invariants w.r.t. the operators $X_{ij}$ separately.

In the next discussion we mention the Casimir operators of the third and fourth degree in details.

It is well known that the Casimir operator of the third degree, defined by the equation[8]

$$C_3 = X_{ij} X_{jk} X_{ki}, \tag{29}$$

is not independent for the symplectic algebra Sp(2$N$) [contrary to, e.g., the algebra SU($n$)] and it is expressed with the aid of the independent quadratic Casimir operator $C_2 = X_{ij} X_{ji}$ by the equation

$$C_3 = 2(N+1) C_2. \tag{30}$$

Thus, the question arises, whether the invariant of the third degree of the graded algebra (Sp(2$N$);2$N$) is possible to be constructed with the aid of the generators $X_{ij}$ and $V_k$. Using the following equations:

$$\mathrm{Sp}(Q_{i'j'}, Q_{ji}) = 2\delta_{ii'}\delta_{jj'} - 2\epsilon_{i'}\epsilon_{j'}\delta_{j-i'}\delta_{i-j'}, \tag{31}$$

$$g Q_{ij} g = Q_{ij}^T = Q_{ji}, \tag{32}$$

for the matrices $Q_{ij}$, defined by Eq. (9) (the matrix elements of $g$ are components of the metric tensor), from Eq. (12) the relation

$$V_k (g Q_{ij})_{kl} V_l = 2 X_{ji} \tag{33}$$

can be derived. This of course, means that the bilinear combination $V_k (gQ_{ij})_{kl} V_l$ of the generators $V_k$, which have the same transformation properties as the generator $X_{ji}$, is in the graded algebra directly proportional to this generator. Therefore, the Sp(2$N$) invariant of the third degree $V_k (gQ_{ij})_{kl} V_l X_{ij}$ is in the graded algebra (Sp(2$N$);2$N$) expressed with the help of the quadratic Casimir operator $C_2$ by the equation

$$V_k (gQ_{ij})_{kl} V_l X_{ij} = 2C_2 \equiv 2 X_{ji} X_{ij}. \tag{34}$$

The bilinear combination of operators $V_k$, defined by Eq. (21), represents the operator with the same transformation properties which has the generator $X_{kl}$. With the help of it, it is possible again to define the Sp(2$N$) invariant of the third degree,

$$T'_{kl} X_{lk}. \tag{35}$$

As, of course, the relation

$$\epsilon_{-l} V_{-l} V_k X_{lk} = X_{jk} X_{kj}, \tag{36}$$

is valid (see Appendix A) it is evident that the operator (35) as well the operator (34) is proportional to the quadratic Casimir operator $C_2$. It means that it is not possible to find, i.e., it does not exist, a third degree invariant of the graded algebra (Sp(2$N$);2$N$). From the three operators $V_k$, namely, it is not possible either to construct the Sp(2$N$) invariant.

Now, we start the discussion of the Casimir operator of the fourth degree of the graded algebra (Sp(2$N$);2$N$).

The Casimir operator of the fourth degree of the Lie subalgebra Sp(2$N$) is known from the paper in Ref. 8,

$$C_4 = X_{ij} X_{jj'} X_{j'i'} X_{i'i}. \tag{37}$$

For further discussion it is very advantageous to use the operator

$$T_{ij} = X_{ij'} X_{j'j}, \tag{38}$$

which has the same transformation properties as the operator $X_{ij}$. Then the Casimir operator $C_4$ of the Lie subalgebra Sp(2$N$) may be written down as

$$C_4 = T_{ij} T_{ji}. \tag{39}$$

As we have mentioned above, the operator $T'_{kl} = -\epsilon_l V_{-l} V_k$, bilinear in the $V_k$ [see Eq. (21)] has the same transformation properties as the operator $X_{kl}$. With the help of operators $T'_{kl}$ and $T_{kl}$ it is possible to construct two operators of the fourth degree (which contain two operators of the type $X_{ij}$ and two operators of the type $V_k$),

$$\epsilon_{-j} V_{-j} V_{j'} T_{jj'}, \quad T_{jj'} \epsilon_{-j} V_{-j} V_{j'}. \tag{40}$$

It has been mentioned before that it is possible to construct the Sp(2$N$)-invariant with the aid of the operators $V'_k$ defined by Eq. (18). The operator

$$V'_k g_{kk'} V'_{k'} = V'_k \epsilon_k V'_{-k} \tag{41}$$

is also a Sp(2$N$) invariant of the fourth degree [see Eq. (24)]. Of course, the operators, Eqs. (40) and (41), are not independent. Using Eqs. (4), (11), and (15) the following relations may be found:

$$V'_k g_{kk'} V'_{k'} = T_{jj'} \epsilon_{-j} V_{-j} V_{j'} + X_{kj} X_{jk}, \tag{42}$$

$$V'_k g_{kk'} V'_{k'} = \epsilon_{-j} V_{-j} V_{j'} T_{jj'} + X_{kj} X_{jk}, \tag{43}$$

in which besides the operators of the fourth degree quadratic Casimir operator $C_2$ of the algebra Sp(2$N$) appears.

It remains to construct the Sp(2$N$) invariant with the aid of the product of four operators $V_k$. The simplest possibility is to take such an operator in the form

$$\left(\sum_{j=-N}^{+N} \epsilon_{-j} V_{-j} V_j\right)\left(\sum_{j'=-N}^{+N} \epsilon_{-j'} V_{-j'} V_{j'}\right). \tag{44}$$

The other possibility is to take the operator in the following form,

$$T'_{kl} T'_{lk} = \epsilon_{-l} V_{-l} V_k \epsilon_{-k} V_{-k} V_l, \tag{45}$$

where $T'_{kl}$ is defined by Eq. (21). The operators (44) and (45) are certainly not independent as the following relation between them is valid,

$$\epsilon_{-l} V_{-l} V_k \epsilon_{-k} V_{-k} V_l = -(\epsilon_{-l} V_{-l} V_l)(\epsilon_k V_k V_{-k}) + 4 X_{jk} X_{kj} + 2(2N+1)\epsilon_{-l} V_{-l} V_l. \tag{46}$$

The Casimir operator of the fourth degree of the graded algebra (Sp(2$N$);2$N$) can be constructed with the help of the described Sp(2$N$)-invariants (39), (40), and (44). By using relations (4), (11), and (12) we can verify that the Casimir operator of the fourth degree is given as

$$\begin{aligned} T_{j'j} T_{jj'} &- (2N^2+5N+3) X_{kj} V_j g_{kk'} K_{k'j'} V_{j'} \\ &+ \tfrac{1}{2}(2N^2+5N+5)(\epsilon_{-j} V_{-j} V_{j'} T_{jj'} + T_{jj'} \epsilon_{-j'} V_{-j'} V_j) \\ &- \tfrac{1}{2}(\epsilon_{-j} V_{-j} V_j)(\epsilon_{-j'} V_{-j'} V_{j'}) = K_4. \end{aligned} \tag{47}$$

Namely, by using the above-mentioned equations, it can be proved that the following is valid,

$$[K_4, V_m] = 0 \quad (m = -N, \ldots, +N). \tag{48}$$

The proof of this statement is given in details in Appendix B.

## CONCLUSION

The simplest Casimir operators of the graded Lie algebra (Sp(2$N$);2$N$) were discussed. It was found that besides the quadratic Casimir operator, which is known from a number of papers, the Casimir operator of the fourth degree also exists. The third degree Casimir operator of the graded Lie algebra (Sp(2$N$);2$N$) does not exist. The position with the graded Lie algebra (Sp(2$N$); 2$N$) is in this respect very similar to that with the symplectic Lie algebra Sp(2$N$), where the independent Casimir operators are only of even degree. The Casimir operators of odd degree of the algebra Sp(2$N$)—contrary to the graded algebra (Sp(2$N$);2$N$)—exist, but are linearly dependent on the Casimir operators of lower degrees. The method used in this paper, for the derivation of Casimir operators of the fourth degree $K_4$ can be simply generalized and applied for the derivation of any higher Casimir operator, i.e., the operator of the sixth degree, eight degree, etc.

In a general case the Casimir operator of the 2$n$th degree of the GLA (Sp(2$N$);2$N$) will consist of the Sp(2$N$)-invariants formed by the polynomial 2$n$th degree in the generators $X_{ij}$ and $V_k$. The 2$n$th degree polynomial in the $X_{ij}$ is naturally the Sp(2$N$)-Casimir operator

$$T_{i_1 i_2} T_{i_2 i_3} \cdots T_{i_n i_1}.$$

Further, a contribution will come from all operators which are polynomials of the $(2n-2m)$th degree in $X_{ij}$ and of the 2$m$th degree polynomials in $V_k$ ($m = 1, 2, \ldots, n$) as

$$\begin{aligned} &T'_{i_1 i_2} T_{i_2 i_3} \cdots T_{i_n i_1} \\ &T_{i_1 i_2} T'_{i_2 i_3} \cdots T_{i_n i_1} \\ &\quad\vdots \\ &T_{i_1 i_2} T_{i_2 i_3} \cdots T'_{i_n i_1}, \\ &T'_{i_1 i_2} T'_{i_2 i_3} \cdots T_{i_n i_1}, \text{ etc.} \end{aligned}$$

Of course, the last operator will be the polynomial of the 2$n$th degree in $V_k$,

$$T'_{i_1 i_2} T'_{i_2 i_3} \cdots T'_{i_n i_1}.$$

The solution of the problem as to how the particular operators will contribute to the Casimir operators $K_{2n}$ may be found directly.

Obviously the number of independent Casimir operators of the graded algebra (Sp(2$N$);2$N$) is at least equal to the number of independent Casimir operators of the Lie subalgebra Sp(2$N$).

## APPENDIX A

### A. The derivation of relation (A1)

$$-X_{j'k} X_{kj'} = V_{j'} V_{-k} \epsilon_k X_{kj'}. \tag{A1}$$

If we carry out in the operator $V_{j'} V_{-k} \epsilon_k X_{kj'}$ the change of indices $k \to -j'$, $j' \to -k$, we can rewrite this operator as $V_{-k} V_{j'} \epsilon_{-j'} X_{-j',-k}$. If in addition we use Eq. (6), this operator may be rewritten in the form $V_{-k} V_{j'} \epsilon_k X_{kj'}$. Therefore, it is possible to write

$$\begin{aligned} V_{j'} V_{-k} \epsilon_k X_{kj'} &= \tfrac{1}{2}(V_{j'} V_{-k} \epsilon_k X_{kj'} + V_{-k} V_{j'} \epsilon_{-j'} X_{-j',-k}) \\ &= \tfrac{1}{2}(V_{j'} V_{-k} \epsilon_k X_{kj'} + V_{-k} V_{j'} \epsilon_k X_{kj'}) \\ &= \tfrac{1}{2}\epsilon_k \{V_{j'}, V_{-k}\} X_{kj'}. \end{aligned} \tag{A2}$$

Using relation (15) we get

$$\tfrac{1}{2}\epsilon_k \{V_{j'}, V_{-k}\} X_{kj'} = -X_{j'k} X_{kj'}.$$

If we use Eq. (A2) we obtain Eq. (A1). In a similar way it is possible to prove the following relations:

$$X_{j'k}X_{kj'} = V_{j'}V_{-k}\epsilon_{j'}X_{-j'-k},$$
$$X_{j'k}X_{kj'} = X_{j'k}V_k\epsilon_{-j'}V_{-j'} \tag{A3}$$

and a number of others.

### B. The deirvation of relation (46)

Using relations (11) and (15) we can derive the commutation relation

$$[\epsilon_{-j}V_{-j}V_j, V_m] = -4V_iX_{mi} - 2(2N+1)V_m.$$

With the aid of the invariant relation (45) it is possible to rewrite the above in the form

$$\begin{aligned}(\epsilon_{-l}V_{-l}V_k\epsilon_{-k}V_{-k}V_l) &= \epsilon_{-l}V_{-l}(V_lV_k\epsilon_{-k}V_{-k} + 4V_iX_{li} + 2(2N+1)V_l) \\ &= (\epsilon_{-l}V_{-l}V_l)(V_k\epsilon_{-k}V_{-k}) + 4X_{jk}X_{kj} + 2 \\ &\quad + 2(2N+1)\epsilon_{-l}V_{-l}V_l,\end{aligned}$$

which is just Eq. (46).

## APPENDIX B: Derivation of the Casimir operator of the fourth degree $K_4$

Let us consider the superposition of the operators

$$\begin{aligned}I = AT_{j'j}T_{jj'} &+ B(\epsilon_{-j}V_{-j}V_{j'}T_{jj'} + T_{j'j}\epsilon_{-j'}V_{-j'}V_j) \\ &+ C(\epsilon_{-j}V_{-j}V_j)(\epsilon_{-j'}V_{-j'}V_{j'}),\end{aligned}$$

which are defined by the Eqs. (39), (40), and (44). Then

$$\begin{aligned}[I, V_m] = AT_{j'j}[T_{jj'}, V_m] &+ A[T_{j'j}, V_m]T_{jj'} \\ &+ B\epsilon_{-j}V_{-j}V_{j'}[T_{jj'}, V_m] + B[\epsilon_{-j}V_{-j}V_{j'}, V_m]T_{jj'} \\ &+ BT_{j'j}[\epsilon_{-j'}V_{-j'}V_j, V_m] + B[T_{j'j}, V_m]\epsilon_{-j'}V_{-j'}V_j \\ &+ C(\epsilon_{-j}V_{-j}V_j)[\epsilon_{-j'}V_{-j'}V_{j'}, V_m] \\ &+ C[\epsilon_{-j}V_{-j}V_j, V_m]\epsilon_{-j}V_{-j}V_j.\end{aligned}$$

By using relations (4), (11), and (15) with Eq. (6) the separate terms in the commutator $[I, V_m]$ can be written down in the form

$$\begin{aligned}[T_{j'j}, V_m]T_{jj'} = 4V_{j'}X_{mj}T_{jj'} &- (6N+3)V_{-l}T_{m-l} \\ &+ V_{-l}X_{m-l}(2N+2)2N + 3V_mT_{kk} \\ &+ V_{j'}T_{mj'},\end{aligned}$$

$$\begin{aligned}[\epsilon_{-j}V_{-j}V_{j'}, V_m]T_{jj'} = &-4V_{j'}X_{mj}T_{jj'} - 2V_mT_{kk} \\ &+ 2(2N+1)V_{-j}T_{m-j} + 2(2N+2)V_{-j}X_{m-j},\end{aligned}$$

$$\begin{aligned}\epsilon_{-j}V_{-j}V_{j'}[T_{jj'}, V_m] = 2(\epsilon_jV_{-j}V_j)V_{j'}X_{mj'} &- 2V_mT_{kk} \\ &+ (2N+1)(V_{-j}\epsilon_jV_j)V_m - 4V_jT_{mj} \\ &+ 4(N+1)V_jX_{mj},\end{aligned}$$

$$\begin{aligned}&(V_{-j}\epsilon_{-j}V_j)[V_k\epsilon_kV_{-k}, V_m] \\ &\quad = (V_{-j}\epsilon_{-j}V_j)(-4V_kX_{mk} - 2(2N+1)V_m),\end{aligned}$$

$$\begin{aligned}&[T_{j'j}, V_m]\epsilon_{-j'}V_{-j'}V_j \\ &\quad = -2V_{j'}X_{mj'}(V_{-j}\epsilon_{-j}V_j) - (2N+1)V_m(V_{-j}\epsilon_{-j}V_j) \\ &\qquad + 4V_jT_{mj} + 4(N-1)V_jX_{mj} + 4N(2N+1)V_m + 2T_{kk}V_m,\end{aligned}$$

$$\begin{aligned}&[V_k\epsilon_kV_{-k}, V_m](V_j\epsilon_jV_{-j}) \\ &\quad = (-4V_kX_{mk} - 2(2N+1)V_m)(V_j\epsilon_jV_{-j}),\end{aligned}$$

etc.

By using these equations we may simply find that, under the assumption $A=B=1$, $C=-\frac{1}{2}$, it is possible to eliminate in the commutator $[I, V_m]$ all terms which contain the products of four generators of both types, $X_{ij}$ and $V_k$. Namely, we can write:

$$\begin{aligned}\omega_1 &= [T_{j'j}, V_m]T_{jj'} + [\epsilon_{-j}V_{-j}V_{j'}, V_m]T_{jj'} \\ &= -2NV_jT_{mj} + V_mT_{kk} + (2N+2)^2V_jX_{mj},\end{aligned}$$

$$\begin{aligned}\omega_2 &= T_{j'j}[T_{jj'}, V_m] + T_{j'j}[\epsilon_{-j'}V_{-j'}V_j, V_m] \\ &= 2NT_{mj}V_j - T_{kk}V_m + 2(2N+2)X_{mj}V_j,\end{aligned}$$

$$\begin{aligned}\omega_3 &= \epsilon_{-j}V_{-j}V_{j'}[T_{jj'}, V_m] - \tfrac{1}{2}V_{k'}\epsilon_{k'}V_{-k'}[V_k\epsilon_kV_{-k}, V_m] \\ &= -2V_mT_{kk} - 4V_jT_{mj} + 4(N+1)V_jX_{mj},\end{aligned}$$

$$\begin{aligned}\omega_4 &= [T_{j'j}, V_m]\epsilon_{-j'}V_{-j'}V_j - \tfrac{1}{2}[V_k\epsilon_kV_{-k'}, V_m]V_j\epsilon_jV_{-j} \\ &= 4V_jT_{mj} + 4(N-1)V_jX_{mj} + 2T_{kk}V_m + 4N(2N+1)V_m.\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{i=1}^{4}\omega_i = 2N[T_{mj}, V_j] &+ [T_{kk}, V_m] \\ &+ (4N^2+16N+4)V_jX_{mj} + (4N+4)X_{mj}V_j \\ &+ 4N(2N+1)V_m.\end{aligned}$$

Finally, by using the relations

$$[T_{mj}, V_j] = (2N+1)X_{mj}V_j - V_jX_{mj},$$
$$[T_{kk}, V_m] = 2(2V_kX_{mk} + (2N+1)V_m)$$

it is possible to write the sum of $\omega_i$, $\sum_{i=1}^{4}\omega_i$, in the form

$$\sum_{i=1}^{4}\omega_i = (2N^2+5N+3)[T_{kk}, V_m].$$

Therefore, for the case of $A=B=1$, $C=-\frac{1}{2}$, the operator

$$K_4 = I - (2N^2+5N+3)T_{kk}$$

fulfils the relation

$$[K_4, V_m] = 0.$$

Using Eqs. (42) and (43) we can write the operator $T_{kk}$ in the form

$$\begin{aligned}T_{kk} = V'_k g_{kk'} V'_{k'} &- \tfrac{1}{2}(T_{jj'}\epsilon_{-j}V_{-j}V_{j'} \\ &+ \epsilon_{-j}V_{-j}V_{j'}T_{jj'})\end{aligned}$$

and gain expression (47) for the Casimir operator $K_4$.


[1]P. G. O. Freund and I. Kaplansky, J. Math. Phys. **17**, 228 (1976).
[2]A. Pais and V. Rittenberg, J. Math. Phys. **16**, 2062 (1975).
[3]W. Nahm *et al.*, Phys. Lett. B **61**, 383 (1976).
[4]M. Scheunert *et al.*, J. Math. Phys. **17**, 1626, 1640 (1976).
[5]V. G. Kac, Funct. Anal. Appl. **9**, 91 (1975).
[6]M. Scheunert *et al.*, J. Math. Phys. **18**, 146, 155 (1977).
[7]N. B. Backhouse, J. Math. Phys. **18**, 239 (1977).
[8]A. M. Perelomov and V. S. Popov, Nucl. Phys. **3**, 1127 (1965) (in Russian).
[9]C. O. Nwachuku and M. A. Rashid, J. Math. Phys. **17**, 1611 (1976) and preprint IC/75/144, ICTP, Trieste.

[10]G. Racah, "Group Theory and Spectroscopy," lecture notes, Princeton (1951).

[11]Besides the operators $T'_{kl}$ it is possible to construct, in the frame of bilinear combinations of generators $V_k$, two other irreducible [under the Sp($2N$)] tensors: the invariant operator $g_{ij}V_k g_{kk'}V_{k'}$(scalar) and the antisymmetric tensor

$$R_{ij} = V_i V_j - V_j V_i - (g_{ij}/N)\, V_k g_{kk'} V_{k'},$$

which fulfils the conditions

$$R_{ij} = -R_{ji} \quad \text{and} \quad g_{ij}R_{ij} = 0.$$

The tensor $T'_{kl}$ corresponds to the Young tableau □□, while the remaining two tensors correspond to the tableau $\begin{smallmatrix}\square\\\square\end{smallmatrix}$, which appears in the decomposition of the direct product

$$\square \otimes \square = \square\square \oplus \begin{smallmatrix}\square\\\square\end{smallmatrix}\,.$$

# A note on representation of para-Fermi algebra

R. Jagannathan

*Department of Physics, St. Joseph's College, Tiruchirapalli-620 002, India*

R. Vasudevan

*MATSCIENCE, Institute of Mathematical Sciences, Madras-600 020, India*
(Received 5 July 1977)

Representation of para-Fermi algebras is obtained utilizing the operators of a *single* Fermi field. In analogy with Kalnay's para-Grassman algebras, para-Clifford algebras are defined in terms of Fermi operators.

## I. INTRODUCTION

Green[1] in 1953 gave the representation of the para-Fermi and para-bose operators by forming linear combinations of several different commuting Fermi and anticommuting Bose fields respectively. Ramakrishnan and his co-workers have obtained representations of para-Fermi rings employing the elements of the generalized Clifford algebra.[2,3] Para-Fermi fields of higher orders, $p>1$, may be related to higher spins.[4] Hence there is continued interest in such representations.

Recently Kalnay[5,6] realized the representations of para-Fermi operators employing boson operators and suitable boson vector spaces. It will be our attempt in this note to obtain in Sec. 2 the representations of para-Fermi operators in terms of the usual anticommuting operators belonging to a *single* Fermi field. In Sec. 3 we show that some results similar to those of Kalnay[5,6] can also be obtained using Fermi operators and we construct para-Clifford algebras in analogy with Kalnay's para-Grassman algebras.[6]

## 2. REPRESENTATION OF PARA-FERMI ALGEBRA

Para-Fermi operators of order $p$, $\{A_i^{(p)} \mid i=1,2,\ldots,n\}$, and their Hermitian conjugates obey the relations

$$[[A_i^{(p)\dagger}, A_j^{(p)}], A_k^{(p)\dagger}] = 2\delta_{jk}A_i^{(p)\dagger}, \tag{1}$$

$$[[A_i^{(p)}, A_j^{(p)}], A_k^{(p)}] = [[A_i^{(p)\dagger}, A_j^{(p)\dagger}], A_k^{(p)\dagger}] = 0, \tag{2}$$

$$(A_i^{(p)})^{p+1} = 0, \tag{3}$$

$$\forall\, i,j,k = 1,2,\ldots,n.$$

Green's ansatz of constructing the para-Fermi operators of order $p$ is to start with $p$ different Fermi fields which commute with each other but anticommute among themselves. Thus, letting $\{(a_j^{(l)} \mid j=1,2,\ldots,n) \mid l=0,1,\ldots,p-1\}$ and their Hermitian conjugates denote the set of $p$ commuting sets of Fermi operators, we have to form the linear combinations as

$$A_j^{(p)} = \sum_{l=0}^{p-1} a_j^{(l)}, \quad j=1,2,\ldots,n. \tag{4}$$

Then it can be easily seen[1] that $\{A_j^{(p)}\}$ satisfy the relations (1)–(3). We want to point out that in the above construction (4) the original $\{a_j^{(l)}\}$ belonging to different $l$'s commute with each other since they represent disparate Fermi fields.

Let us now turn to our efforts to constructing para-Fermi operators $\{A_j^{(p)}\}$ from the basic operators $\{a_j \mid j=1,2,\ldots,np\}$ belonging to a *single* Fermi field.

It is asserted that $\{A_j^{(p)}\}$, given by

$$A_j^{(p)} = \sum_{l=0}^{p-1}\left\{\left(\prod_{r=1}^{ln}[a_r, a_r^\dagger]\right)a_{ln+j}\right\}, \quad j=1,2,\ldots,n, \tag{5}$$

will satisfy all the conditions required of para-Fermi operators of order $p$ [Eqs. (1)–(3)].

To this end let us proceed as follows. Defining

$$\alpha_j = (a_j^\dagger + a_j), \quad \beta_j = i(a_j^\dagger - a_j), \quad j=1,2,\ldots,np, \tag{6}$$

it is easy to see that

$$\begin{aligned}&\{\alpha_j, \beta_k\} = 0,\\ &\{\alpha_j, \alpha_k\} = \{\beta_j, \beta_k\} = 2\delta_{jk},\\ &\quad j,k = 1,2,\ldots,np.\end{aligned} \tag{7}$$

Now let us define the Hermitian operators

$$P_{ln+j} = i^{ln(2ln+1)}\left(\prod_{r=1}^{ln}\alpha_r\beta_r\right)\alpha_{ln+j}, \tag{8}$$

$$Q_{ln+j} = i^{ln(2ln+1)}\left(\prod_{r=1}^{ln}\alpha_r\beta_r\right)\beta_{ln+j}, \tag{9}$$

$$j=1,2,\ldots,n, \quad l=0,1,\ldots,p-1.$$

Then it follows immediately that

$$\begin{aligned}&\{P_{ln+j}, Q_{ln+k}\} = 0,\\ &\{P_{ln+j}, P_{ln+k}\} = \{Q_{ln+j}, Q_{ln+k}\} = 2\delta_{jk},\\ &\forall\, l=0,1,\ldots,p-1 \text{ and } \forall\, j,k=1,2,\ldots,n,\end{aligned} \tag{10}$$

and

$$\begin{aligned}[P_{ln+j}, P_{mn+k}] &= [P_{ln+j}, Q_{mn+k}]\\ &= [Q_{ln+j}, Q_{mn+k}] = 0\\ &\forall\, j,k=1,2,\ldots,n \text{ and } l\neq m.\end{aligned} \tag{11}$$

It is now easy to derive from the $\{P\}$ and $\{Q\}$ operators, $a_j^{(l)}$ operators of (4) which constitute a set of $p$ commuting Fermi fields. Let us define

$$a_j^{(l)} = \tfrac{1}{2}(P_{ln+j} + iQ_{ln+j}), \tag{12}$$

 0022-2488/78/1907-1493$1.00  

$$a_j^{(l)\dagger}=\tfrac{1}{2}(P_{ln+j}-iQ_{ln+j}),\qquad \forall\ j=1,2,\ldots,n \text{ and } l=0,1,\ldots,p-1. \tag{13}$$

Now we have all the necessary ingredients to make use of Green's ansatz (4). Therefore, we write

$$\begin{aligned}A_j^{(p)}&=\sum_{l=0}^{p-1}a_j^{(l)}=\tfrac{1}{2}\sum_{l=0}^{p-1}(P_{ln+j}+iQ_{ln+j})\\&=\tfrac{1}{2}\sum_{l=0}^{p-1}\left\{\left(i^{ln(2ln+1)}\prod_{r=1}^{ln}\alpha_r\beta_r\right)(\alpha_{ln+j}+i\beta_{ln+j})\right\}\\&=\sum_{l=0}^{p-1}\left\{\left(\prod_{r=1}^{ln}[a_r,a_r^\dagger]\right)a_{ln+j}\right\},\quad j=1,2,\ldots,n.\end{aligned}\tag{14}$$

This may be written out explicitly if need be as

$$\begin{aligned}A_j^{(p)}=\{&a_j+[a_1,a_1^\dagger][a_2,a_2^\dagger]\cdots[a_n,a_n^\dagger]a_{n+j}\\&+[a_1,a_1^\dagger][a_2,a_2^\dagger]\cdots[a_{2n},a_{2n}^\dagger]a_{2n+j}\\&+\cdots+[a_1,a_1^\dagger][a_2,a_2^\dagger]\cdots[a_{(p-1)n},a_{(p-1)n}^\dagger]a_{(p-1)n+j}\},\end{aligned}\tag{15}$$

$\forall\ j=1,2,\ldots,n.$

For example, if $n=3$ and $p=2$, we have to take six operators $\{a_j\mid j=1,2,\ldots,6\}$ and their Hermitian conjugates of the basic Fermi field and construct the three operators of the para-Fermi field of order 2 as

$$\begin{aligned}A_1^{(2)}&=a_1+[a_1,a_1^\dagger][a_2,a_2^\dagger][a_3,a_3^\dagger]a_4,\\A_2^{(2)}&=a_2+[a_1,a_1^\dagger][a_2,a_2^\dagger][a_3,a_3^\dagger]a_5,\\A_3^{(2)}&=a_3+[a_1,a_1^\dagger][a_2,a_2^\dagger][a_3,a_3^\dagger]a_6.\end{aligned}\tag{16}$$

The Hermitian conjugates of the operators (14)–(16) are easily obtained. It can be directly checked that $A_j^{(p)}$, $A_j^{(p)\dagger}$, etc. given by (14) satisfy all the requirements [(1)–(3)] characterizing the para-Fermi field of order $p$.

In the boson description of fermions[5,6] the para-Fermi operators constructed from the boson operators have to operate on the $p$-boson subspace of the boson state vector space. It is emphasized here that our $A_j^{(p)}$ operators operate on the usual entire fermion state vector space and are constructed out of operators belonging to a *single* Fermi field.

## 3. PARA-CLIFFORD ALGEBRA

Kalnay[6] has shown that if

$$g_i=\sum_{r,s=1}^{m}G_{irs}b_r^\dagger b_s,\quad i=1,2,\ldots,n,\tag{17}$$

where $\{G_i\}$ is any $m$-dimensional representation of the ordinary Grassman algebra and $b_r^\dagger$, $b_s$, etc. are the usual boson operators, then $\{g_i\}$ generates an algebra with the commutation relations

$$[[g_i,g_j],g_k]=0,\quad \forall i,j,k=1,2,\ldots,n.\tag{18}$$

The $\{g_i\}$ algebra has been called the para-Grassman algebra by Kalnay.[6]

In a similar vein let us define

$$\epsilon_i=\sum_{r,s=1}^{m}M_{irs}a_r^\dagger a_s,\quad i=1,2,\ldots,n,\tag{19}$$

where $\{M_i\}$ is any $m$-dimensional representation of a given algebra and $a_r^\dagger$, $a_s$, etc. are the usual Fermi operators. It is easy to see that

$$\begin{aligned}\epsilon_i\epsilon_j\epsilon_k=&(M_iM_jM_k)_{rs}a_r^\dagger a_s\\&-[(M_iM_j)_{rs}M_{kr's'}+(M_jM_k)_{rs}M_{ir's'}\\&+(M_iM_k)_{rs}M_{jr's'}]a_r^\dagger a_{r'}^\dagger a_sa_{s'}\\&-M_{irs}M_{jr's'}M_{kr''s''}a_r^\dagger a_{r'}^\dagger a_{r''}^\dagger a_sa_{s'}a_{s''},\end{aligned}\tag{20}$$

$\forall\ i,j,k=1,2,\ldots,n.$

Hence it follows that

$$[[\epsilon_i,\epsilon_j],\epsilon_k]=([[M_i,M_j],M_k])_{rs}a_r^\dagger a_s,\qquad \forall\ i,j,k=1,2,\ldots,n.\tag{21}$$

In (20) and (21) summation over repeated indices is assumed.

Taking $\{M_i\}$ to be the representations of the usual Grassman algebra (21) shows that $\{\epsilon_i\}$ generates a para-Grassman algebra.[6] In contradistinction to Kalnay's[6] representations we note that the above $\{\epsilon_i\}$ involve Fermi operators instead of Boson operators.

If $\{M_i\}$ denote the representations of the usual Clifford algebra satisfying

$$\{M_i,M_j\}=2\delta_{ij},\quad \forall i,j=1,2,\ldots,n,\tag{22}$$

then $\{\epsilon_i\}$ satisfy the following relationships:

$$\tfrac{1}{4}[[\epsilon_i,\epsilon_j],\epsilon_k]=\epsilon_i\delta_{jk}-\epsilon_j\delta_{ik},\qquad \forall\ i,j,k=1,2,\ldots,n.\tag{23}$$

It is to be noted that (23) is a relation met with in the case of the Kemmer algebra.

We may call the $\epsilon$ algebra the para-Clifford algebra in analogy with the para-Grassman algebra derived by Kalnay.[6]

## ACKNOWLEDGMENT

The authors express their thanks to Professor Alladi Ramakrishnan for his kind encouragement.

[1]H.S. Green, Phys. Rev. 90, 270 (1953).
[2]Alladi Ramakrishnan, R. Vasudevan, P.S. Chandrasekharan, and N.R. Ranganathan, J. Math. Anal. Appl. 28, 108 (1969).
[3]Alladi Ramakrishnan, R. Vasudevan, and P.S. Chandrasekharan, J. Math. Anal. Appl. 35, 249 (1971).
[4]S. Kamefuchi and Y. Takahashi, Prog. Theor. Phys. Suppl. 37, 244 (1966).
[5]A.J. Kalnay, Prog. Theor. Phys. 54, 1848 (1975).
[6]A.J. Kalnay, Rep. Math. Phys. (Warsaw) 9, 9 (1976).

# Regge pole behavior in $\phi^3$ field theory

M. C. Bergère[a)] and C. Gilain

*CEN-Saclay, B. P. n° 2, 91190 Gif-Sur-Yvette, France*
(Received 30 November 1977)

We consider the Feynman amplitudes for all the essentially and crossed planar graphs of the four-point vertex function in $\phi^3$ field theory, and we evaluate their behavior at high energy (large $s$, fixed $t$). We compute the coefficients of all logarithms for the dominant amplitudes which behave in $s^{-1}$ (up to logarithms of $s$). This computation is performed by using the Bogoliubov–Parasiuk–Hepp $R$ operation and the Mellin transform of the Feynman amplitudes. The geometrical structure of the coefficients is such that all logarithms of $s$ of all dominant amplitudes can be summed to give the well-known Regge behavior with signature $+$. The Regge trajectory verifies an equation which may be solved explicitly in the lowest order approximation; the residue is found to be the ratio of two functions of $t$, the upper one being factorized into two vertex functions expressed as infinite series and the lower one providing a ghost killing factor.

## INTRODUCTION

One of the many striking facts when one deals with interaction between hadrons is the relevance of the Regge picture: the scattering amplitude at high energy and small angle is well reproduced by a Regge behavior, in particular the energy dependence is a $t$ dependent power of $s$. It is clear that any theory of strong interactions will have to reproduce this fact.

On the other hand, recent progress in quantum field theory, especially in gauge field theories, indicates that field theory is likely to provide us an underlying strong interaction theory. Consequently, the importance of deriving Regge behavior from field theory is quite obvious.

Such a proof exists in potential theory,[1] and, of course, it is not a new problem in field theory. Many papers have already appeared on reggeization, in scalar and gauge field theories, and it is necessary to make clear to the reader what we mean really by reggeization. We mean by reggeization the property of the two-body scattering amplitude at large energy $s$, for a given transfer $t$, to behave as a sum of terms of the form $\beta(t)s^{\alpha(t)}$. We shall not consider here the problem of the materialization of the Regge trajectories.[2]

Let us now discuss how Regge behavior may be exhibited from field theory.

One method consists in using the Bethe–Salpeter structure of the four-point function in the $t$ channel.[3] Assuming at large $s$ a factorization property in the $t$ channel, a Regge behavior may be proved quite generally for the solution of the Bethe–Salpeter equation.[4]

A second method exploits the large $s$ behavior of Feynman graphs. Many papers[5] deal with the large $s$ behavior of the graphs contributing to the low orders of the perturbation series, and then examine the coefficients of their logarithms of $s$, in order to exhibit (or not) the beginning of an exponential series.

In another set of papers, the authors select a subclass of diagrams (ladder graphs, for instance) and perform the summation of the leading logarithms of $s$, or if possible of more powers of logarithms.[6]

Several other papers determine the leading power of $s$ and the maximal power of logarithm of $s$ for the most general graph contributing to the amplitude.[7] Precise rules are given by Zav'yalov[8] for planar convergent graphs and by Zav'yalov and Stepanov[9] for planar divergent graphs. Unfortunately, no rules are given to find the coefficient of the logarithms (at least beyond the leading power of logarithm). Efremov *et al.*[10] in a series of papers have described a general procedure in order to perform the infinite sum of logarithms of $s$.

Our work lies in the same spirit than the papers of Zav'yalov, Stepanov, and Efremov. For all graphs of $\lambda\phi^3$ which are not susceptible to contribute to a possible Regge cut but only to a possible Regge pole, we prove the three following points:

(1) We prove the existence of a class of dominant amplitudes which behave in $s^{-1}$ up to logarithms of $s$. We characterize all graphs of this class (Sec. 2).

(2) We determine the coefficients of all powers of logarithms of $s$ for any graph of the above class. These coefficients are expressed in terms of subgraphs and reduced graphs (Sec. 3).

(3) We sum the result (2) over all graphs of the above class. The geometrical structure of the coefficients of all powers of logarithms of $s$ explains their exponentiation. The functions $\beta(t)$ and $\alpha(t)$ are found as infinite series (Sec. 4).

Our technique is based on the use of the multiple Mellin transform of a Feynman amplitude and of Bogoliubov–Parasiuk–Hepp $R$ operator, and is a generalization of Ref. 11.

The rest of this introduction is devoted to a classification of the graphs contributing to the amplitude and also to definitions and notations.

The connected 4-external legs, Green function $G^c_{(4)}(p_i, m, g)$ may be expressed in terms of the invariants $p_i^2$ and of the Mandelstam variables

$$s=(p_1+p_2)^2,\quad t=(p_1+p_3)^2,\quad u=(p_1+p_4)^2, \tag{1.1}$$

where the external momentum $p_i$ are ingoing. We re-

[a)]Physique Théorique CNRS.

 0022-2488/78/1907-1495$1.00  

call that these variables are dependent since

$$s+t+u=\sum_{i=1}^{4} p_i^2. \tag{1.2}$$

It is convenient to classify the different Feynman graphs contributing to $G^c_{(4)}(p_i, m, g)$ in four different classes depending upon their topology.

*Definition*: Given a connected Feynman graph $G$ with $n$ vertices, $l$ internal lines, and $L$ independent loops, we define:

—A one-tree subgraph is a connected tree subgraph connecting all the $n$ vertices of $G$. (Such a one-tree may be obtained by cutting $L$ lines of $G$ with the condition that each cutted line decreases the number of loops by one.)

—A two-tree is a two-connected tree subgraph obtained by cutting one line to one of the above one-tree. (One of its connected part may be an isolated vertex.) A two-tree partitions the external momentum into two parts $I_1$ and $I_2$ and

$$\sum_{i\in I_1} p_i = -\sum_{j\in I_2} p_j. \tag{1.3}$$

Given a one-tree constructed from a graph $G$ contributing to $G^c_{(4)}(p_i, m, g)$, this one-tree is called an $s$(resp. $t,u$)-one-tree if all the two-trees which are constructed from it are such that the different squares

$$\left(\sum_{i\in I_1} p_i\right)^2 = \left(\sum_{j\in I_2} p_j\right)^2 \tag{1.4}$$

which are not 0 or $p_i^2$ $(i=1,2,3,4)$ are $s$ (resp. $t,u$). It is clear that, for any graph $G$ in $\phi^3$, there exists at least one two-tree such that the corresponding square is either $s$, $t$, or $u$. The Feynman graphs with 4-external legs such that all their one-trees are $t$-one-trees contribute to $G^1_{(4)}(t, p_i^2, m, g)$. The Feynman graphs which have at least one $s$-one-tree but no $u$-one-trees contribute to $G^2_{(4)}(s, t, p_i^2, m, g)$. The graphs contributing to $G^1_{(4)}$ and $G^2_{(4)}$ are called essentially planar (all planar graphs are essentially planar but many nonplanar graphs are also essentially planar as it is shown on Fig. 1 where the striped kernels are nonplanar; such nonplanar kernels are self-energy graphs or three-external legs graphs). The Feynman graphs which have at least one $u$-one-tree but no $s$-one-trees contribute to $G^3_{(4)}(t, u, p_i^2, m, g)$ and are called crossed planar. Finally the remaining graphs which have at least one $s$- and one $u$-one-tree contribute to $G^4_{(4)}(s, t, u, p_i^2, m, g)$.

The graphs of $G^1_{(4)}(t, p_i^2, m, g)$ contribute as a constant to the large $s$, fixed $t$, behavior of the amplitude. They are taken apart from the following evolution of this paper and are considered only in the conclusion.

The graphs contributing to $G^i_{(4)}$ for $i=2,3,4$ are of the form given by Fig. 2, where each black dot represents a graph which contributes to the complete

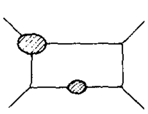


FIG. 1.

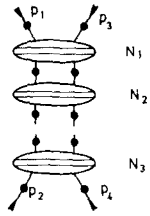



FIG. 2.

propagator and where each kernel $N_i$ represents a graph two-line irreducible in the $t$ channel. If, for a given graph $G$ represented by Fig. 2, at least one of the subgraphs $N_i$ has both $s$- and $u$-one-trees, then it is clear by the cutting rule of internal lines that $G$ contributes to $G^4_{(4)}(s, t, u, p_i^2, m, g)$. Such graphs are susceptible to generate Regge cuts[6] and are beyond the scope of this paper. If, for a given graph $G$ represented by Fig. 2, no subgraph $N_i$ has both $s$- and $u$-one-trees but $q$ subgraphs $N_i$ have at least one $u$-one-tree, then, if $q$ is odd, $G$ is a crossed planar graph and contributes to $G^3_{(4)}(t, u, p_i^2, m, g)$; if $q$ is even, then $G$ contributes to $G^2_{(4)}(s, t, p_i^2, m, g)$.

We now define a Feynman amplitude by its Schwinger-integral representation

$$I_G^\epsilon = (-g)^n i^{-\omega(G)/2}\int_0^\infty \prod_{a=1}^{l} d\alpha_a \exp\left(-i\exp(-i\epsilon)\sum_{a=1}^{l}\alpha_a m^2\right)$$
$$\times R\left\{\frac{\exp\{i\exp(-i\epsilon)[k_i(\epsilon)\,d_{ij}(\alpha)\,k_j(\epsilon)]\}}{P_G^2(\alpha)}\right\}. \tag{1.5}$$

For $\epsilon=\pi/2$, we obtain the amplitude in Euclidean space; in Minkowski space $I_G$ is defined as the limit $\epsilon\to 0$ of $I_G^\epsilon$ and is known to be a distribution. The external momentum $k(\epsilon)$ are defined as

$$k(\epsilon) = (p_0\exp(i\epsilon), \vec{p}) \tag{1.6}$$

and scalar products are defined in the Minkowski metric $(+---)$ (with this definition the metric in Euclidean space is $----$). The above representation is an application of Wick's rotation. In this paper we purposely omit writing the $\epsilon$ dependence of $I_G^\epsilon$ in most cases. The superficial degree of divergence of $G$ is

$$\omega(G) = 4L(G) - 2l(G), \tag{1.7}$$

where $L(G)$ and $l(G)$ are the number of independent loops and the number of internal lines of the graph $G$. The functions $d_{ij}(\alpha)$ and $P_G(\alpha)$ are characteristic of the topology of the graph. The operator[12] $R$ is a subtraction operator which acts directly upon the variables $\alpha$ and ensures the ultraviolet convergence. We define

$$R = \prod_{\varphi\subseteq G}(1 - \mathcal{T}_\varphi^{-2l(\varphi)}), \tag{1.8}$$

where the operators $\mathcal{T}$ are generalized Taylor operators and the product runs over the $(2^l - 1)$ subgraphs of $G$.

The generalized Taylor operators $\mathcal{T}$ are defined as follows: Given a function $f(x)$ such that $x^{-\nu}f(x)$ is infinitely differentiable for $\nu$ complex, then

$$\mathcal{T}_x^n f(x) = x^{-\lambda-\epsilon}T^{n+\lambda}\{x^{\lambda+\epsilon}f(x)\}. \tag{1.9}$$

This definition is $\lambda$ independent provided that $\lambda \geq -E'(\nu)$, where $E'(\nu)$ is the integer part of Re$\nu$ and $E'(\nu) \geq$ Re$\nu$;

$\epsilon = E'(\nu) - \nu$. This definition is generalizable to the case of several variables $\alpha$

$$\mathcal{T}^n_\varphi f(\alpha) = [\mathcal{T}^n_\rho f(\alpha)|_{\alpha_a \to \rho^2 \alpha_a,\ \forall\ a\in\varphi}]_{\rho=1} \ . \tag{1.10}$$

The method used in this paper to obtain the asymptotic behavior of $I_G$ at large $s$, fixed $t$, is a generalization of the method used in Ref. 11 for the behavior of $I_G$ when external momentum are scaled to infinity. We calculate the Mellin transform of the Feynman amplitude in regards to the variable which becomes large. Two cases may appear. Either, the integrand of the Mellin transform expressed in the variable $\alpha$ has a "simutaneous Taylor expansion" in every Hepp's sector defined as an ordering of the variable $\alpha$

$$0 \leqslant \alpha_{i_1} \leqslant \alpha_{i_2} \leqslant \cdots \leqslant \alpha_{i_l} \tag{1.11}$$

(there are $l!$ sectors); then, the operator[11] $R$ defines an analytic continuation of the Mellin transform beyond the first singularity and extracts the residue at the first pole. This residue is closely related to the coefficients of all powers of logarithms for the leading power. In Ref. 11 for instance, we have given geometrical rules in terms of subgraphs and reduced graphs to describe the coefficients of the powers of logarithms. The geometry of these coefficients is such that a summation of all logarithms of the leading power can be performed.[13] Or, the integrand of the Mellin transform does not have a "simultaneous Taylor expansion" in every Hepp's sector,[14] which is the case in this paper. It is then necessary to split the $\alpha$ integrand of the Mellin transform into several parts, each of them having a "simultaneous Taylor expansion" in every Hepp's sector. This leads to a multiple Mellin transform which is analytic in a tube, the real part of which is a convex polyhedron. The asymptotic behavior is then determined by an extremal point of the polyhedron. This discussion is performed in Sec. 2; it is found that, in $\phi^3$ field theory, although the multiple Mellin transform is unavoidable for nondominant amplitudes of $G^2_{(4)}$ (behavior in $s^{-p}\log^X s$, $p > 1$), the dominant amplitudes (behavior in $s^{-1}\log^X s$) may be treated by a single Mellin transform even if the integrand does not have a "simultaneous Taylor expansion" in every Hepp's sector. In Sec. 3, we extract for a dominant amplitude of $G^2_{(4)}$ and $G^3_{(4)}$ the coefficients of the logarithms. In Sec. 4, we sum all the logarithms of $s$ for the power $s^{-1}$ and for all dominant amplitudes contributing to $G^2_{(4)}$ and $G^3_{(4)}$. Finally in Sec. 5, we give the lowest order contribution to the Regge trajectory found in Sec. 4.

## 2. ESTIMATION OF THE LEADING POWER IN $s$ FOR A GRAPH CONTRIBUTING TO $G^2_{(4)}$

To any graph contributing to $G^2_{(4)}$, there exists a corresponding graph contributing to $G^3_{(4)}$, and the estimation in $s$ of the first amplitude is the same as the estimation in $u$ of the second one. Consequently, we restrict our discussion to a graph of $G^2_{(4)}$.

The quadratic form $[k_i(\epsilon)\, d_{ij}(\alpha)\, k_j(\epsilon)]$ which enter in (1.5) can be written at $\epsilon = 0$ as a sum over the invariants $s$, $t$, and $p_i^2$ as

$$sA_s(\alpha) + tA_t(\alpha) + \sum_{i=1}^{4} p_i^2 A_i(\alpha). \tag{2.1}$$

The function $A_s(\alpha)$ is the ratio of two polynomials $N_s(\alpha)$ over $P_G(\alpha)$ and is homogeneous in all $\alpha$'s of degree 1 since $N_s(\alpha)$ is homogeneous in all $\alpha$'s of degree $[L(G)+1]$ and $P_G(\alpha)$ of degree $L(G)$.

We define an $s$ cut as a set of lines $(i_1, \ldots, i_p)$ such that if these lines are cut, the graph $G$ becomes two-connected with one connected part containing the external legs $p_1$ and $p_2$, the other connected part containing the external legs $p_3$ and $p_4$, and such that no subset of $(i_1, \ldots, i_p)$ has the same property. An $s$ cut defines two-connected subgraphs $G_L$ and $G_R$; any one-tree of $G_L$ union any one-tree of $G_R$ defines an $s$-two-tree of $G$. We have

$$A_s(\alpha) = N_s(\alpha)/P_G(\alpha), \tag{2.2a}$$

$$N_s(\alpha) = \sum_{\{s \text{ cuts}\}} \prod_{\{a\in s \text{ cut}\}} \alpha_a P_{G_L}(\alpha) P_{G_R}(\alpha). \tag{2.2b}$$

Given a subgraph $\varphi$ with $\chi_\varphi$ connected parts and an $s$ cut $c$, this $s$ cut $c$ split $\varphi$ into two subgraphs $\varphi_L$ and $\varphi_R$ with respectively $\chi_{\varphi_L}$ and $\chi_{\varphi_R}$ connected parts (some of them being eventually reduced to single vertices). From the topological relation

$$l(\varphi) + \chi_\varphi = n(\varphi) + L(\varphi).$$

where $n(\varphi)$ is the number of vertices of $\varphi$, it is easy to show that when all $\alpha$ variables corresponding to lines of $\varphi$ vanish like $\rho$, the expression

$$A_c(\alpha) = \prod_{\{a\in c\}} \alpha_a P_{G_L}(\alpha) P_{G_R}(\alpha)/P_G(\alpha) \tag{2.3}$$

vanishes like $\rho^{y_c(\varphi)}$, where

$$y_c(\varphi) = \chi_{\varphi_L} + \chi_{\varphi_R} - \chi_\varphi. \tag{2.4}$$

Consequently, $A_s(\alpha)$ vanishes in the same condition like $\rho^{y(\varphi)}$

$$y(\varphi) = \inf_{\{c\}} y_c(\varphi). \tag{2.5}$$

A subgraph $\varphi$ is said to be essential if $y(\varphi) \geqslant 1$ (otherwise it is a nonessential subgraph). An essential subgraph $\varphi$ is such that the reduced subgraph $[G/\varphi]$, where $\varphi$ is shrunk into $\chi_\varphi$ points has an $s$-independent Feynman amplitude.

*Example*: We consider the graph of Fig. 3:

$$P_G(\alpha) = (\alpha_1 + \alpha_3 + \alpha_5)(\alpha_2 + \alpha_4 + \alpha_7) + \alpha_6(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_7),$$

$$N_s(\alpha) = \alpha_1\alpha_3(\alpha_2 + \alpha_4 + \alpha_6 + \alpha_7) + \alpha_2\alpha_4(\alpha_1 + \alpha_3 + \alpha_5 + \alpha_6) + \alpha_1\alpha_4\alpha_6 + \alpha_2\alpha_3\alpha_6.$$

The $s$ cuts are $\{13\}$, $\{146\}$, $\{236\}$, $\{24\}$. The subgraph $\{123\}$ is an essential subgraph with $y(\varphi) = 1$; the subgraph $\{567\}$ is nonessential.

Let us show on this graph the property that $[A_s(\alpha)]^x$ which appear in the Mellin transform of the amplitude does not have a "simultaneous Taylor expansion" in every sector. We choose, for instance, the sector

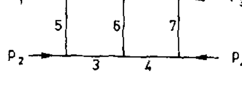



FIG. 3.

$0 \le \alpha_1 \le \alpha_2 \le \alpha_3 \le \alpha_4 \le \alpha_5 \le \alpha_6 \le \alpha_7.$

Because $P_G(\alpha)$ does have a "simultaneous Taylor expansion" in every sector, there exists a largest monomial for the above sector which is $\alpha_6\alpha_7$. On the other hand, $\alpha_2\alpha_4\alpha_6$ dominates in the part of $N_s(\alpha)$ which is $[\alpha_2\alpha_4(\alpha_1+\alpha_3+\alpha_5+\alpha_6)+\alpha_1\alpha_4\alpha_6+\alpha_2\alpha_3\alpha_6]$ and $\alpha_1\alpha_3\alpha_7$ dominates in the part $\alpha_1\alpha_3(\alpha_2+\alpha_4+\alpha_6+\alpha_7)$. Nevertheless, $\alpha_2\alpha_4\alpha_6$ and $\alpha_1\alpha_3\alpha_7$ cannot be compared. If we perform the change of the variables $\alpha$ into the sector variables $\beta$ defined by

$$\alpha_i = \prod_{j=i}^{7} \beta_j^2, \quad d\alpha_i = \prod_{j=i+1}^{7} \beta_j^2\, 2\beta_i d\beta_i,$$

we obtain for $[A_s(\alpha)]^x$ a behavior of the type

$$[A_s(\alpha)]^x \sim (\beta_2^2\beta_3^2\beta_4^4\beta_5^4\beta_6^2\beta_7^2)^x(\beta_1^2\beta_3^2+\beta_6^2)^x.$$

The function $(\beta_1^2\beta_3^2+\beta_6^2)^x$ does not have a "simultaneous Taylor expansion" around $\beta_i=0$ if $x$ is not a nonnegative integer.

This example clearly demonstrates that there exist graphs such that the corresponding $A_s(\alpha)$ does not have the required "simultaneous Taylor expansion" needed for the extraction of the residues from the single Mellin transform. Among all possible partitions of $A_s(\alpha)$ into parts which do have the needed Taylor property, the most natural partition we can think of is the partition into $s$ cuts. For each $s$ cut, the contribution (2.3) to $A_s(\alpha)$ clearly has the required Taylor property.

We may now apply the results of Ref. 14. There, we have proved the following theorem:

Given a convergent Feynman amplitude $I_G\{a_k\}$ in Euclidean space, where $\{a_k\}$ is the set of invariants (square masses and $(\sum p_i)^2$ built from the external momenta $p_i$), then, if we scale a subset $\{a_m\}$ of the invariants by $\lambda$, $I_G[\lambda\{a_m\},\{a_n\}]$ has an asymptotic expansion for large $\lambda$, of the type.

$$I_G[\lambda\{a_m\},\{a_n\}] = \sum_{p=\Omega}^{-\infty} \lambda^p \sum_{q=0}^{q_{\max}(p)} \log^q\lambda\ g_{pq}[\{a_m\},\{a_n\}], \tag{2.6}$$

where $p$ runs over the rational values of a decreasing arithmetic progression with $\Omega$ as leading power, and $q$, for a given $p$, runs over a finite number of nonnegative integer values.

This theorem is proved by using a multiple Mellin representation of the convergent Feynman amplitude. The extension of this theorem to divergent Feynman amplitudes does not present any theoretical difficulties. Moreover, the theorem is valid in some cases of asymptotic expansion for a Minkowskian Feynman amplitude, namely, when the asymptotic expansion of the Feynman amplitude is determined by end-point singularities ($\alpha$'s $\sim 0$) as is the case in Euclidean space, excluding pinch singularities [only the graphs contributing to $G^4_{(4)}(s,t,u,p_i^2,m,g)$ are susceptible to develop pinch singularities].

In Appendix A, we generalize the method used in Ref. 14 and apply it to determine the large $s$ behavior of the graphs contributing to $G^2_{(4)}(s,t,p_i^2,m,g)$ taking into account the existence of the unique connected divergent subgraph of $\phi^3$ (Fig. 4). We note that the graphs of Fig. 4 are always disjoint and their union is always nonessential.

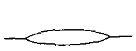


FIG. 4

Let us now discuss how we determine the leading power behavior $\Omega$. First, we remind the reader how it works with the single Mellin transform $M(x)$. The function $M(x)$ is analytic in $x$ in a band $\Omega < \mathrm{Re}x < \Omega'$; in fact, $M(x)$ is found to be a meromorphic function in $x$ with multiple poles, the first of which on the left is at $\mathrm{Re}x=\Omega$. The extension of this picture to multiple Mellin transform goes as follows: The function $M(x_1,\ldots,x_n)$ is analytic in a tube the real part of which is a convex polyhedron $\mathcal{P}$; then $\Omega$ is now obtained as

$$\Omega = \inf_{\mathcal{P}} \sum_{1}^{n} x_i. \tag{2.7}$$

Of course, this inf is obtained for points on the border of $\mathcal{P}$. In practice, the situation is more complicated as we can see from Appendix A. We first decompose the domain of integration upon the variables $\alpha$ into Hepp's sectors; then, if the graph contains divergences, the subtraction operator $R$ is decomposed into contributions which are attached to equivalence classes of nested subgraphs (see Ref. 15). Then, for each Hepp's sector and for each equivalence class of nested subgraphs we obtain a finite sum of terms, each defining a convex polyhedron $\mathcal{P}$ which determines for that contribution an $\Omega_{\mathcal{P}}$. The leading behavior $\Omega$ is then the largest $\Omega_{\mathcal{P}}$ over all polyhedron $\mathcal{P}$.

We now state the results of Appendix A. Our intention is not to give for the most general graph of $\phi^3$ contributing to $G^2_{(4)}$ a rule to obtain $\Omega$; this was done by Zavyalov and Stepanov[9] using naive power counting (although we have no counterexample to this rule at the present time, we have not been able to justify it). On the other hand, we prove that for any graph of $\phi^3$ contributing to $G^2_{(4)}$

$$\Omega \le -1 \tag{2.8}$$

and $\Omega=-1$, if and only if at least one of the kernel $N_i$ described in Fig. 2 is a single rung $\gamma_i$ (see Fig. 5). Consequently, all graphs contributing to $G^2_{(4)}$ such that no kernels $N_i$ is a single rung behaves for large $s$ as $s^\Omega \log^x s$ with $\Omega$ strictly less that $(-1)$ and are nondominant by a power of $s$ in regards to the graphs such that at least one kernel $N_i$ is a single rung $\gamma_i$ which behaves as $s^{-1}\log^x s$. Assuming that summation of logarithms for nondominant graphs still gives a nondominant contribution with regard to the summation of logarithms for dominant graphs, we concentrate in the next section on the graphs which have at least one kernel $N_i$ equal to a single rung.

## 3. ASYMPTOTIC EXPANSION OF FEYNMAN AMPLITUDES RELATED TO DOMINANT ESSENTIALLY AND CROSSED PLANAR GRAPHS

We now consider a graph as described in Fig. 2 and where each kernel $N_j$ represents a subgraphs such that

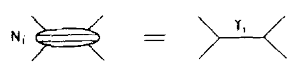



FIG. 5.

only $s$- and $t$-one-trees may be constructed from it. We define $I$ as the nonempty subset of indices $j$ such that the kernels $N_j$, $j \in I$, are single rungs $\gamma_j$. Then, the function $A_s(\alpha)$ defined in (2.2a) is given by

$$A_s(\alpha, \gamma) = \frac{\prod_{j \in I} \gamma_j \prod_{j \notin I} N_j(\alpha) \prod_t P_t(\alpha)}{P_G(\alpha, \gamma)}, \tag{3.1}$$

where we associate a variable $\gamma_j$ with the rungs $N_j$, $j \in I$. The functions $N_j(\alpha)$ associated with the kernels $N_j$, $j \notin I$, can be found by using the expressions (2.3) for each $s$ cut $c$. Each polynomial $P_t(\alpha)$ corresponds to a self-energy graph represented by a black dot in Fig. 2. Taking into account the subtractions to be performed over the logarithmically divergent subgraphs described in Fig. 4, the renormalized Feynman amplitude for the graph $G$ is

$$I_G(s) = (-g)^n (i)^{-\omega(G)/2} \int_0^\infty \Pi d\alpha \, \Pi d\gamma \exp[-i(\textstyle\sum \alpha + \sum \gamma) m^2] \times R \left\{ \frac{\exp\{i[sA_s(\alpha, \gamma) + tA_t(\alpha, \gamma) + \sum_{i=1}^4 p_i^2 A_i(\alpha, \gamma)]\}}{P_G^2(\alpha, \gamma)} \right\}, \tag{3.2}$$

where we omit mentioning the fact that $I_G(s)$ is really a limit $\epsilon \to 0$ of $I_G^\epsilon(s)$. The above amplitude is real in the Euclidean region. The subtraction operator $R$ in this case reduces to

$$R = \prod_\varphi (1 - \mathcal{T}_\varphi^{-4}) = 1 + \sum_{\mathcal{N}} \prod_{\varphi \in \mathcal{N}} (-\mathcal{T}_\varphi^{-4}), \tag{3.3}$$

where the subgraphs $\varphi$ are the logarithmically divergent subgraphs. In (3.3) we sum over all nests $\mathcal{N}$ of divergent subgraphs $\varphi$. In Appendix A we prove that limit $s \to \infty \, I_G(s) \sim s^{-1} \log^x s$ and on the other hand $I_G(s=0)$ is finite $\neq 0$. Consequently, the single Mellin transform

$$M_G(x) = \int_0^\infty ds \, s^{-x-1} I_G(s) \tag{3.4}$$

exists and is analytic for $-1 < \text{Re}x < 0$. In this region we may interchange the integrals over $s$ and over $\alpha$ and $\gamma$. We may also interchange the integral over $s$ and the subtraction operator $R$ because $A_s(\alpha, \gamma)$ does not vanish when all variables $\alpha_a$ corresponding to all divergent subgraphs vanish:

$$\int_0^\infty ds \, s^{-x-1} R\{\exp[isA_s(\alpha, \gamma) \cdot\}$$
$$= R\{\int_0^\infty ds \, s^{-x-1} \exp[isA_s(\alpha, \gamma)] \cdot\}. \tag{3.5}$$

If we remember that $s$ in $\exp[isA_s(\alpha, \gamma)]$ has a small positive imaginary part, we have

$$\int_0^\infty ds \, s^{-x-1} \exp[isA_s(\alpha, \gamma)]$$
$$= \Gamma(-x) \exp(-i\pi x)[iA_s(\alpha, \gamma)]^x \tag{3.6}$$

where $i^x$ is $\exp(i\pi x/2)$ and where we insist on the $[+iA_s]$ for the homogeneity reason and reality of the $(\alpha - \gamma)$ integrals in the Euclidean region. The Mellin transform $M_G(x)$ is found to be

$$M_G(x) = (-g)^n (i)^{-\omega(G)/2} \Gamma(-x) \exp(-i\pi x) \int_0^\infty \Pi d\alpha \, \Pi d\gamma \times \exp[-i(\textstyle\sum \alpha + \sum \gamma) m^2] \times R \left\{ \frac{[iA_s(\alpha, \gamma)]^x \exp\{i[tA_t(\alpha, \gamma) + \sum_{i=1}^4 p_i^2 A_i(\alpha, \gamma)]\}}{P_G^2(\alpha, \gamma)} \right\} \tag{3.7}$$

for $-1 < \text{Re}x < 0$.

Now, we know, from Appendix A, that the subgraphs responsible for the multiple poles at $x = -1$ are the single rungs $\gamma_j$ union any disconnected logarithmically divergent subgraphs. We call these subgraphs leading. On the other hand, if we define $\bar{M}_G(x)$ as the right-hand side of (3.7) but where the $R$ operator is defined at $x = -1 - \eta$ with $\eta$ small positive, then $\bar{M}_G(x)$ is analytic for $a_G < \text{Re}x < 0$ with $a_G < -1$ (Appendix B). The fact that the subtraction operator is defined at $x = -1 - \eta$ is such that not only the logarithmically divergent subgraphs, but also the leading subgraphs are subtracted. The $R$ operator in $\bar{M}_G(x)$ may also be written as a sum over all nests of divergent and leading subgraphs. In the region $-1 < \text{Re}x < 0$, we may compare $M_G(x)$ and $\bar{M}_G(x)$; the difference between the subtraction operators $R$ is a sum over all nests, each containing at least one leading subgraph. We group together the nests which have the same minimal leading subgraph. If this minimal leading subgraph $L$ contains a logarithmically divergent piece $T$ (itself union of several divergent subgraphs), it is always possible to associate by pair the nests of the corresponding group in order to form

$$\cdots (-\mathcal{T}_L^{-2l(L)})(1 - \mathcal{T}_T^{-2l(T)}) \cdots, \tag{3.8}$$

and this is easily proved to be zero. Consequently, in the difference between $M_G(x)$ and $\bar{M}_G(x)$, there remain only those groups where the minimal leading subgraph is a union of single rungs. Given a union $J$ of $\nu(J)$ single rungs of $G$, we obtain

$$\sum_J \{[\sum_{\mathcal{N}_J} \prod_{\substack{\varphi \in \mathcal{N}_J \\ \varphi \neq J}} (-\mathcal{T}_\varphi^{-2\rho(\varphi)})](-\mathcal{T}_J^{-2\nu(J)})\}, \tag{3.9}$$

where $\mathcal{N}_J$ are all possible nests of leading subgraphs with $J$ as minimal element. Now, we note that

$$A_s(\alpha, \gamma) = \prod_{\gamma_j \in J} \gamma_j \prod_{\gamma_j \notin J} \gamma_j \prod_{j \notin I} N_j(\alpha) \prod_t P_t(\alpha) / P_G(\alpha, \gamma), \tag{3.10}$$

and, consequently,

$$(-\mathcal{T}_J^{-2\nu(J)}) \left\{ \frac{[A_s(\alpha, \gamma)]^x \exp\{i[tA_t(\alpha, \gamma) + \sum_{i=1}^4 p_i^2 A_i(\alpha, \gamma)]_G\}}{P_G^2(\alpha)} \right\}$$
$$= -\prod_{\gamma_j \in J} \gamma_j^x \cdot \frac{[\prod_{\gamma_j \notin J} \gamma_j \prod_{j \notin I} N_j(\alpha) \prod_t P_t(\alpha)]^x}{P_{G/J}^{2+x}(\alpha)} \times \exp\{i[tA_t(\alpha, \gamma) + \sum_{i=1}^4 p_i^2 A_i(\alpha, \gamma)]_{G/J}\}, \tag{3.11}$$

where the graph $G/J$ is the reduced graph obtained from $G$ by shrinking the single rungs of $J$ into $\nu(J)$ points. We observe a complete factorization of the $\alpha$ integrals into the single rungs on one part and the reduced graph $G/J$ on the other part. Moreover, the sum over all nests $\mathcal{N}_J$ reconstructs the operator $R$ for the graph $G/J$. We define

$$\bar{M}_{[G/J]}(x, t, p_i^2, m, g) = (-g)^{n-2\nu(J)} (i)^{-[\omega(G)/2+\nu(J)]} \times \int_0^\infty \Pi d\alpha \prod_{j \notin J} d\gamma_j \exp[-i(\textstyle\sum \alpha + \sum_{j \notin J} \gamma_j) m^2]$$

$$\cdot R\left\{\frac{[i^{1-\nu(J)}\Pi_{j\notin J}\gamma_j\Pi_{j\notin I}N_j(\alpha)\Pi_t P_t(\alpha)]^x}{[P_{G/J}(\alpha)]^{x+2}}\right.$$

$$\left.\times\exp\{i[tA_t(\alpha,\gamma)+\sum_{i=1}^{4}p_i^2A_i(\alpha)]_{G/J}\}\right\}, \qquad (3.12)$$

where the operator $R$ subtracts the logarithmically divergent subgraphs and the single rungs of $G/J$. The function $\overline{M}_{[G/J]}(x)$ is shown in Appendix B to be analytic for $a'_{G/J} < \mathrm{Re}x < 0$ with $a'_{G/J} < -1$. $\overline{M}_{[G/J]}(x)$ is real in the Euclidean region. If we integrate over the variables $\gamma_j$ for $j \in J$ (with a small positive $\epsilon$), we find that

$$M_G(x) = \sum_J (-g)^{2\nu(J)} m^{-2\nu(J)(x+1)}$$

$$\times[\Gamma(x+1)]^{\nu(J)}\,\Gamma(-x)\exp(-i\pi x)$$

$$\times\,\overline{M}_{[G/J]}(x,t,p_i^2,m,g). \qquad (3.13)$$

Only the term with $J$ empty does not contribute to the pole at $x=-1$. We use now the inverse Mellin transform (with a positive $\epsilon$)

$$I_G(s) = \frac{1}{2i\pi}\int_{\sigma-i\infty}^{\sigma+i\infty} dx\, s^x\; M_G(x), \quad -1<\sigma<0. \qquad (3.14)$$

The presence of the functions $\Gamma$ makes this integral absolutely convergent; if we push the contour towards the left beyond $x=-1$, we use the Cauchy theorem around $x=-1$ to obtain $I_G^{as}(s)$ and we neglect a background integral $\sim s^{\sigma<-1}$.

The asymptotic part of the Feynman amplitude corresponding to a dominant essentially planar graph is found to be

$$I_G^{as}(s) = \sum_J \frac{(-g)^{2\nu(J)}}{[\nu(J)-1]!}\,\frac{\partial^{\nu(J)-1}}{\partial x^{\nu(J)-1}}$$

$$\times\{\Gamma(-x)\exp(-i\pi x)\,s^x[\Gamma(x+2)]^{\nu(J)}$$

$$\times m^{-2\nu(J)(1+x)}\,\overline{M}_{[G/J]}(x,t,p_i^2,m,g)\}_{x=-1}, \qquad (3.15)$$

where the sum over $J$ now exclude $J$ empty.

Of course, this result shows that $I_G^{as}(s) \sim s^{-1}\log^{r-1}s$, where $r$ is the total number of single rungs $\gamma$ in $G$.

For crossed planar graphs, we exchange $s$ and $u$, $p_2$ and $p_4$. From (1.2) and (1.6), $u\sim(-s)$ and the term $\exp(-i\pi x)$ is absent, and we get for crossed planar graphs

$$I_G^{as}(u\to -s) = \sum_J \frac{(-g)^{2\nu(J)}}{[\nu(J)-1]!}\,\frac{\partial^{\nu(J)-1}}{\partial x^{\nu(J)-1}}$$

$$\times\{\Gamma(-x)\,s^x[\Gamma(x+2)]^{\nu(J)}\,m^{-2\nu(J)(1+x)}$$

$$\times\overline{M}'_{[G/J]}(x,t,p_i^2,m,g)\}_{x=-1}, \qquad (3.16)$$

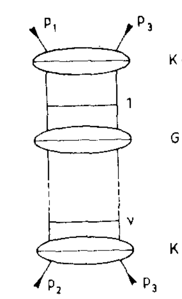



FIG. 6.

where $\overline{M}'$ is the function $\overline{M}$ after exchanging $p_2$ and $p_4$; we note that on the mass shell $\overline{M}'=\overline{M}$.

## 4. INFINITE SUMMATION OF THE LOGARITHMS OF $s$

In Sec. 3, we found the structure in power of logarithms of $s$ for all essentially and crossed planar dominant graphs of $G_{(4)}$. Now, we show that the geometrical structure of the coefficients allows us to sum $I_G^{as}(s)$ over all essentially and crossed planar dominant graphs. Of course, the sum over essentially and crossed planar graphs are related by the interchange $s \leftrightarrow u$, $p_2 \leftrightarrow p_4$ and, then, we consider only essentially planar dominant graphs.

Let us concentrate on all graphs which have at least $\nu$ rungs. We may represent such a graph $G$ by Fig. 6, where the kernels $K_1, K_2, G_1, G_2, \ldots, G_{\nu-1}$ represent a graph one particle irreducible in the $t$ channel, not necessarily connected. If the graph $G$ has $r$ rungs, there are, of course, $\binom{r}{\nu}$ different ways of representing $G$ by Fig. 6; but if we look at (3.15), each of the $\binom{r}{\nu}$ representation corresponds to a different contribution to $I_G^{as}(s)$. Moreover, the function $\overline{M}_{[G/J]}$ factorizes into $[\nu(J)+1]$ functions; we have

$$\overline{M}_{[G/J]}(x,t,p_i^2,m,g)$$

$$=\overline{M}_{\overline{K}_1}(x,t,p_1^2,p_3^2,m,g)\cdot\overline{M}_{\overline{K}_2}(x,t,p_2^2,p_4^2,m,g)$$

$$\cdot\prod_{i=1}^{\nu-1}\overline{M}_{\overline{G}_i}(x,t,m,g). \qquad (4.1)$$

In (4.1), we use

$$\overline{M}_{\overline{K}_1}(x,t,p_1^2,p_3^2,m,g)$$

$$=(-g)^{n(K_1)}\,i^{-\omega(K_1)/2}\int_0^\infty \Pi\, d\alpha\,\exp(-i\textstyle\sum\alpha m^2)$$

$$\times R\left\{\frac{[N_{K_1}(\alpha)]^x}{P_{\overline{K}_1}(\alpha)^{x+2}}\exp\{i[tA_t(\alpha)+p_1^2A_1(\alpha)+p_3^2A_3(\alpha)]\}\right\}, \qquad (4.2)$$

where the integral is attached to the vertex graph $\overline{K}_1$ given in Fig. 7 and where we do not distinguish anymore the rungs in $K_1$ from the other lines. The functions $A_t(\alpha)$, $A_1(\alpha)$, $A_3(\alpha)$ are characteristic of the graph $\overline{K}_1$ and $N_{K_1}(\alpha)$ is the product of the variables $\alpha$ attached to the rungs of $K_1$, by the product of the functions $N_j(\alpha)$ corresponding to the two particles irreducible kernels (in the $t$ channel) $N_j$ of $K_1$, and by the various $P_t(\alpha)$ corresponding to self-energies. Similarly $\overline{M}_{\overline{K}_2}$ is obtained by exchanging $p_1$ and $p_2$, $p_3$, and $p_4$. By convention, $M_Y(x)=1$.

Now, the functions $\overline{M}_{\overline{G}_i}$ are given by

$$\overline{M}_{\overline{G}_i}(x,t,m,g)=(-g)^{n(G_i)}\,i^{-\omega(G_i)/2}\int_0^\infty \Pi\,d\alpha\exp(-i\textstyle\sum\alpha m^2)$$

$$\times R\left\{\frac{[-iN_{G_i}(\alpha)]^x}{P_{\overline{G}_i}(\alpha)^{x+2}}\exp[itA_t(\alpha)]\right\}, \qquad (4.3)$$

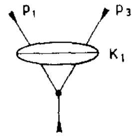



FIG. 7. The graph $\overline{K}_1$.

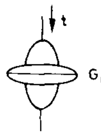


FIG. 8. The graph $\overline{G}_i$.

where the functions $N_{G_i}$, $P_{\overline{G}_i}$, and $A_t(\alpha)$ are similar to the above equivalent functions but are now attached to the two-point graph $\overline{G}_i$ given in Fig. 8.

Any choice of $K_1, K_2, G_1, G_2, \ldots, G_{\nu-1}$ corresponds, by Fig. 6, to a dominant essentially planar graph with at least $\nu$ rungs. When we sum the contribution (3.15) corresponding to $\nu$ rungs over all essentially planar dominant graphs with at least $\nu$ rungs, we obtain all possible one-particle irreducible graphs in the $t$ channel for $K_1, K_2, G_1, G_2, \ldots, G_{\nu-1}$. We define

$$B(x,t,m,g)=\sum_{\overline{G}}\Theta_{\overline{G}}\,\overline{M}_{\overline{G}}(x,t,m,g) \tag{4.4}$$

where we sum over all possible graph $\overline{G}$ with the weight $\Theta_{\overline{G}}$,

$$A_{13}(x,t,p_1^2,p_3^2,m,g)=\sum_{\overline{K}}\Theta_{\overline{K}}\overline{M}_{\overline{K}}(x,t,p_1^2,p_3^2,m,g) \tag{4.5}$$

and similarly $A_{24}(x,t,p_2^2,p_4^2,m,g)$. We obtain for the contribution to the infinite sum corresponding to the contraction of $\nu$ rungs

$$\frac{1}{(\nu-1)!}\frac{\partial^{\nu-1}}{\partial x^{\nu-1}}\left\{\left[\left(\frac{g}{m}\right)^2\Gamma(x+2)\,m^{-2x}B(x,t,m,g)\right]^{\nu-1}\times\left[\left(\frac{g}{m}\right)^2\Gamma(x+2)\,\Gamma(-x)\exp(-i\pi x)\left(\frac{s}{m^2}\right)^x A_{13}A_{24}\right]\right\}_{x=-1}. \tag{4.6}$$

The functions $A_{13}$, $A_{24}$, and $B$ diverge as well as the original perturbation theory. It remains now to sum expression (4.6) from $\nu=1$ to infinity, that is, to solve the Lagrange problem[16]

$$\sum_{\nu=0}^{\infty}\frac{1}{\nu!}\frac{\partial^\nu}{\partial x^\nu}\{f^\nu(x)g(x)\}_{x=-1}. \tag{4.7}$$

This sum is easily performed in the Mellin transform space:

$$\sum_{\nu=0}^{\infty}\frac{f^\nu(x)g(x)}{(x+1)^{\nu+1}}=\frac{g(x)}{x+1-f(x)}. \tag{4.8}$$

However, we must be careful about the required conditions for interchanging the Cauchy contour around $(-1)$ and the infinite sum over $\nu$. We choose for Cauchy contour, two straight lines $\sigma_{\rm I}+iz$ and $\sigma_{\rm II}+iz$ with $-2<\sigma_{\rm II}<-1<\sigma_{\rm I}<0$, and the real integration variable $z$ runs from $-\infty$ to $+\infty$. The rest of the contour is at $z=\pm\infty$ where the functions $|f(x)|$ and $|g(x)|$ vanish due to the functions $\Gamma(x+2)$ and $\Gamma(-x)$ (and $\epsilon>0$ when needed). Now, the interchange of the Cauchy contour and the infinite sum over $\nu$ is allowed if $|f(x)/(x+1)|<1$ along the contour. It must be said at this point that since the function $B(x,t,m,g)$ which enters the function $f(x)$ is defined by a divergent series, little can be said about the validity of the interchange. We prove in Sec. 5, at the lowest order approximation (see Fig. 10), that there exist two intervals, one where $\sigma_{\rm II}$ and one where $\sigma_{\rm I}$ may be chosen such that $|f(\sigma)|<|\sigma+1|$ and *a fortiori* $|f(x)|<|x+1|$. If we can give a sense to the infinite series which define $f(x)$ and if we prove that $|f(x)|<|x+1|$ along the Cauchy contour, then a theorem by Lagrange[16] states that the equation

$$f(x)=x+1 \tag{4.9}$$

has one root $x_0$ and only one inside the Cauchy contour.

In this case we obtain

$$\sum_{\nu=0}^{\infty}\frac{1}{\nu!}\frac{\partial^\nu}{\partial x^\nu}\{f^\nu(x)g(x)\}_{x=-1}=\frac{g(x_0)}{1-f'(x_0)}. \tag{4.10}$$

Given the function $x_0(t,m,g)$ solution of the equation

$$\left(\frac{g}{m}\right)^2\Gamma(x+2)\,m^{-2x}B(x,t,m,g)=x+1, \tag{4.11}$$

the large $s$ behavior for all essentially planar graphs is

$$\left(\frac{g}{m}\right)^2 A_{13}(x_0,t,p_i^2,m,g)\,A_{24}(x_0,t,p_i^2,m,g)\left(\frac{s}{m^2}\right)^{x_0}\times\exp(-i\pi x_0)\,\frac{\pi(1+x_0)}{\sin\pi(1+x_0)}\times\left\{1-\frac{\partial}{\partial x_0}\left[\left(\frac{g}{m}\right)^2\Gamma(x_0+2)\,m^{-2x_0}B(x_0,t,m,g)\right]\right\}^{-1}. \tag{4.12}$$

We note that on the mass shell $A_{13}$ is equal to $A_{24}$.

In the same manner, the large $s$ behavior for all crossed planar graph is obtained from (4.12) by omitting the term $\exp(-i\pi x_0)$. The large $s$ behavior for all essentially and crossed planar graphs is

$$\left(\frac{g}{m}\right)^2 A_{13}(x_0,t,p_i^2,m,g)\,A_{24}(x_0,t,p_i^2,m,g)\left(\frac{s}{m^2}\right)^{x_0}\times[1+\exp(-i\pi x_0)]\,\frac{\pi(1+x_0)}{\sin\pi(1+x_0)}\times\left\{1-\frac{\partial}{\partial x_0}\left[\left(\frac{g}{m}\right)^2\Gamma(x_0+2)\,m^{-2x_0}B(x_0,t,m,g)\right]\right\}^{-1}. \tag{4.13}$$

## 5. LOWEST ORDER CONTRIBUTION TO THE REGGE TRAJECTORY (SEE REF. 6)

We wish to explore Eq. (4.11) which gives the trajectory $x_0(t,m,g)$, in the lowest order approximation for the function $B(x,t,m,g)$, that is, the contribution of the graph of Fig. 9.

This approximation is certainly valid for small coupling constant $g$. Using the weight $\Theta_{\overline{G}}$ equal to $1/2$ and (4.3), (4.4), we get the equation

$$\frac{\pi}{2}\frac{g^2}{m^2}\int_0^1 d\alpha\left[1-\frac{t}{m^2}\,\alpha(1-\alpha)\right]^x=\sin\pi(x+1). \tag{5.1}$$

To explore this equation, let us first solve two of its approximations: the small $t$ dependence and the small $g$ dependence.

FIG. 9.

### A. The small $t$ dependence

We write the solution

$$x_0(t,m,g)=A(m,g)+t\beta(m,g)+O(t^2), \tag{5.2}$$

where $A(m,g)$ is the intercept and $\beta(m,g)$ the slope of the trajectory. An expansion of (5.1) around $t$ equal to zero gives

$$A(m,g)=-1+\frac{1}{\pi}\arcsin\left(\frac{\pi g^2}{2m}\right), \tag{5.3}$$

$$\beta(m,g)=\frac{-A(m,g)g^2}{12m^4}\left[1-\frac{\pi}{4}\left(\frac{g}{m}\right)^4\right]^{-1/2}. \tag{5.4}$$

Equation (5.3) shows that this solution has a meaning only for $|g|\leq m\sqrt{2/\pi}$, and when $|g|$ runs from 0 to $m\sqrt{2/\pi}$, the intercept goes from $-1$ to $-1/2$. The slope $\beta(m,g)$ for this range of $g$ goes from 0 to $+\infty$. The value $|g|=m\sqrt{2/\pi}$ is the limiting value for the Lagrange problem to have a solution, as is illustrated in Fig. 10 in the simple case $t=0$.

The equation $f(x)$ equal to $(x+1)$ has two roots $r_2=A(m,g)$ and $r_3$ symmetrical in regards to $-\frac{1}{2}$ and the equation $|f(x)|=|x+1|$ has four roots, $r_2$, $r_3$, $r_0$, and $r_1$ symmetrical of $r_2$ and $r_3$ with regard to $(-1)$. Consequently, the choice of $\sigma_{\text{I}}$ and $\sigma_{\text{II}}$ which justify (4.10) is given by $r_2<\sigma_{\text{I}}<r_3$ and $r_0<\sigma_{\text{II}}<r_1$. At $g$ equal to $m\sqrt{2/\pi}$, $r_2$ and $r_3$ (resp. $r_0$ and $r_1$) coincide at $x=-\frac{1}{2}$ (resp. $-\frac{3}{2}$).

Of course, for $g$ equal to $m\sqrt{2/\pi}$, the only contribution of Fig. 9 is not valid any more, and more graphs should be computed.

### B. The small $g$ dependence (see also Ref. 3)

We write the solution

$$x_0(t,m,g)=-1+a(t,m)g^2+O(g^4) \tag{5.5}$$

and

$$a(t,m)=\tfrac{1}{2}\int_0^1 d\alpha[m^2-t\alpha(1-\alpha)]^{-1}, \tag{5.6}$$

that is,

$$a(t,m)=[t(t-4m^2)]^{-1/2}\log\left\{\frac{(4m^2-t)^{1/2}+(-t)^{1/2}}{(4m^2-t)^{1/2}-(-t)^{1/2}}\right\} \quad \text{for } t\leq 0, \tag{5.7a}$$

$$a(t,m)=2[t(4m^2-t)]^{-1/2}\arctan[t^{1/2}(4m^2-t)^{-1/2}] \quad \text{for } 0\leq t\leq 4m^2. \tag{5.7b}$$

This approximation is valid for $g$ and $t$ small

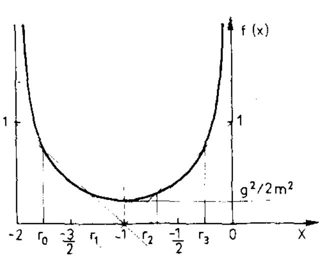



FIG. 10.

[$(x_0+1)$ small] and is certainly bad when $t$ is close to $4m^2$ since for any values of $t$, $m$, and $g\neq 0$ and, for the graph of Fig. 9, $x_0(t,m,g)<0$.

### C. Tabulation of the integral (5.1)

A complete study of the hypergeometric function defined by the integral (5.1) gives for Regge trajectories, the curves of Fig. 11 (see also Ref. 3). Again, the part of the curves where the slope becomes large should not be taken seriously and higher order contributions should be taken into account. At $g=0$, the trajectory becomes $x_0=-1\ \forall\ t\neq 4m^2$ and $-1\leq x_0\leq 0$ for $t=4m^2$. For $t>4m^2$, the trajectories are complex.

We also mention the intersection of the trajectories with $x_0(t,m,g)$ equal to $-1/2$; we get the following relation between $g/m$ and $t/m^2$:

$$\frac{\pi g^2}{mt^{1/2}}\operatorname{arg\,sinh}\left(\frac{t}{4m^2-t}\right)^{1/2}=1, \tag{5.8}$$

for $g/m\leq\sqrt{2/\pi}$ and consequently $0\leq t<4m^2$.

### D. Comments on the complete series which define the Regge trajectories

First, we insist again on the fact that $B(x,t,m,g)$ which enters, equation (4.11) is a divergent series and that anything which may be said here is applicable to a calculation up to a finite order. The left-hand side of Eq. (4.11) is positive or null so that $x_0(t,m,g)\geq -1$. As mentioned at the end of Appendix B, and as may be seen directly on the series $B(x,t,m,g)$, the graph of Fig. 9 becomes infinite at $x=0$. For a similar reason

$$-1\leq x_0(t,m,g)<0 \tag{5.9}$$

whatever the finite number of terms computed in $B(x,t,m,g)$ is. In fact, it may be wrong to interpret this as an absence of bounds states. First, we have seen in the low order approximation that the summation technique breaks down at a certain negative $x_0^{\max}(g/m)$ and nothing can be said above. A similar situation should be true also if we compute more terms of $B$. Also, we show in Appendix B that the poles of $[\overline{M}_G(x)/\Gamma(-x)]$ and of $\overline{M}_{G/J}(x)$ were spurious and did not occur originally in $[M_G(x)/\Gamma(-x)]$. Because of the presence of

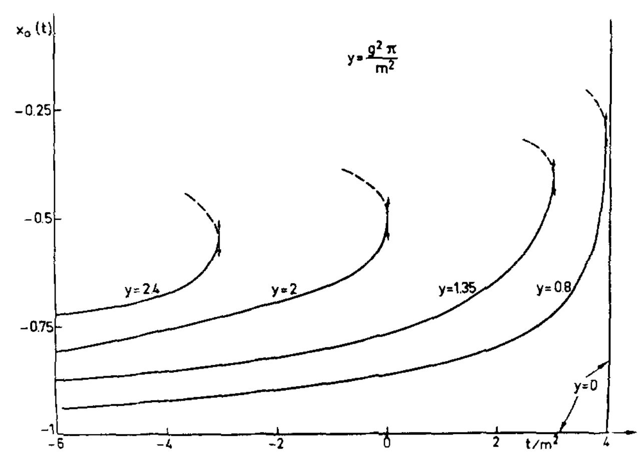



FIG. 11.

$\overline{M}_G(x)$ in the background terms, it may happen that this background at $x=0$ is comparable to the leading asymptotic behavior. This undesirable situation should be then a consequence of our desingularization technique, but at the present time it is the only technique we know which allows infinite summation as in Sec. 4. On the other hand, we may remind our assumption which says that what is negligible by a power of $s$ in perturbation remains negligible; the summation of all logarithms of the power $s^{-1}$ does not prevail upon the power $s^0$ (which has no sum of logarithms). We shall leave open this problem.

Let us finally mention that the limit $t \to -\infty$ of $B(x,t,m,g)$ is zero for $x<0$ and this implies

$$\lim_{t\to-\infty} x_0(t,m,g) = -1. \qquad (5.10)$$

This is so because the leading contribution to $B(x,t,m,g)$ when $t\to-\infty$ comes from the graph of Fig. 9 which behaves as $t^x$ when $x>-1$ and as $t^{-1}\log|t|$ when $x=-1$.

## 6. CONCLUSIONS

We first remind to the reader the validity of the result given in (4.13) and which partially describes the large $s$ behavior at fixed $t$ of the scattering amplitude in $\phi^3$ field theory. The class of graphs which is considered in this paper does not include the graphs which are susceptible to generate Regge cuts; these graphs might in addition develop a Regge pole contribution which is expected to complete our result by nonplanar corrections. We consider all graphs which generate only a Regge pole behavior and, among these, we characterize a class of dominant graphs ($\sim s^{-1}\log^x s$). We neglect the contribution of nondominant graphs and of the nonleading power of $s$ for the dominant graphs because it is negligible graph by graph by a power of $s$, and we assume that, when summing over all graphs, the infinite sum of logarithms of $s$ does not destroy this dominance (as it is also assumed for the scaling properties when we neglect the right-hand side of Callan–Symanzik equation). The technique of summation used in Sec. 4 is based on a theorem by Lagrange and is justified if there exists an interval between $(-1)$ and 0 where Eq. (4.11) has one root and only one; we have shown in Sec. 5 that such a root exists somewhere between $-1$ and 0 at the lowest order approximation, and we assume its existence for higher orders.

The Regge pole behavior obtained in (4.13) possesses the following characteristics:

—The Regge trajectory $x_0(t,m,g)$ is found to be the solution of an equation which contains an infinite series of Feynman-like contributions. The root $x_0(t,m,g)$ is $\geqslant -1$ and at the lowest order approximation increases with $t$ up to a certain negative $x_0^{\max}(g/m)$ where the procedure of computation breaks down. For a small value of the coupling constant $g$ ($g \ll m\sqrt{2/\pi}$) the intercept of the Regge trajectory is found to be $(-1+g^2/2m^2)$ and the slope $g^2/12m^4$. For $t\to-\infty$, $x_0\to-1$, and for $t>4m^2$ the trajectory is complex.

—The signature is positive because of the symmetry $s-u$ of the system.

—The residue factorizes into two vertex functions which are also infinite series of Feynman-like contributions. These functions are themselves dependent of $x_0(t,m,g)$.

—The pole structure appear under the form

$$\frac{1+x_0}{\sin\pi(1+x_0)}\left[1-\frac{\partial}{\partial x_0}\left\{\frac{g^2}{m^2}\,\Gamma(x_0+2)\,m^{-2x_0}\,B(x_0,t,m,g)\right\}\right]^{-1}. \qquad (6.1)$$

This form is infinite for $x_0=0,1,2,\cdots$ and remains finite for $x_0=-1$ as well as for $x_0=-2,-3,\cdots$ because of the "ghost killing factor" in the square bracket [ ] [however, we have shown that $x_0(t,m,g)\geqslant-1$]. The square bracket [ ] in (6.1) does not vanish in the lowest order approximation as long as $x_0<x_0^{\max}(g/m)$ where the summation procedure breaks down. We expect a similar property to hold at higher orders.

—The large $s$ behavior obtained in the paper is

$$\sum_{i=1}^{3} G^i_{(4)}(s) = G^1_{(4)}(t,m,g)\,s^0 + \left(\frac{1+x_0(t,m,g)}{\sin\pi[1+x_0(t,m,g)]}\right) \times\beta(t,p_i^2,m,g)\{1+\exp[-i\pi x_0(t,m,g)]\} \times\left(\frac{s}{m^2}\right)^{x_0(t,m,g)} + \cdots. \qquad (6.2)$$

By the optical theorem, the total cross section $\sigma_{\rm tot}(s)$ is given as

$$\sigma_{\rm tot}(s) \sim \frac{1}{s}\,{\rm Im}G_{(4)}(s,t=0)\big|_{p_i^2=m^2} \text{ for } s \text{ large}. \qquad (6.3)$$

Using the fact that the constant contribution $G^1_{(4)}(t,m,g)$ (Sec. 1) is real at $t=0$, our result, neglecting Regge cuts, gives

$$\sigma_{\rm tot}(s) \sim s^{x_0(t=0)-1} \qquad (6.4)$$

with an intercept which satisfies

$$-1 \leqslant x_0(t=0) < 0. \qquad (6.5)$$

We, of course, should not try to use the numerical values obtained here to describe any physical situation since the main point of this paper is in fact to prove that, in $\phi^3$ field theory, Regge trajectories can be constructed. It is clear that $\phi^3$ is not a relevant field theory for describing hadron physics. For instance, an intercept in $\phi^3$ around $(-1)$ is mainly due to the $s^{-1}$ behavior of the dominant graphs. In $\phi^4$ field theory, all graphs behave as $s^0$ up to logarithms of $s$, and we expect a higher intercept; moreover, $\phi^4$ is a strictly renormalizable field theory, and we know that the renormalization group plays an important role in the large $s$ behavior.[17] Also, it may be useful to investigate the action of group symmetries on the trajectories and, for instance, already, in $\phi^2\varphi$, we observe some splitting of the trajectories. These Lagrangians have to be understood as constructive tests before attacking the description of hadron physics from gauge field theories.

## ACKNOWLEDGMENTS

We wish to thank G. Cohen-Tannoudji for useful discussions. The part of this work relative to multiple Melling transform is an application of the recent results of C. de Calan on this subject and we wish to

express our gratitude to him. We also thank J. M. Drouffe for programming on the computer the integral (5.1).

## APPENDIX A: BEHAVIOR OF $I_G(s)$ WHEN $s \to \infty$ FOR AN ESSENTIALLY PLANAR GRAPH

A method to obtain bounds on a Feynman amplitude when some invariants become large is exposed in Refs. 11, 14. In Minkowski space $I_G(s)$ is defined as the limit $\epsilon \to 0$ of $I_G^\epsilon$ given in (1.5). Let us remind the reader of the main steps which transform $I_G(s)$ into a sum of expressions upon which bounds can be obtained.

### A. Sector decomposition

First we decompose the $\alpha$ integration domain into Hepp's sectors. Each sector is defined by an ordering of the $l$ variables $\alpha_a$. The union of the $l!$ sectors is the original $\alpha$-integration domain. Given a sector $\mathcal{S} = \{0 \leqslant \alpha_{a_1} \leqslant \alpha_{a_2} \leqslant \cdots \leqslant \alpha_{a_l}\}$, we perform the change of variables

$$\alpha_{a_i} = \beta_i^2 \beta_{i+1}^2 \cdots \beta_l^2 \tag{A1a}$$

with $0 \leqslant \beta_l < \infty$ and $0 \leqslant \beta_{i\neq l} \leqslant 1$,

$$d\alpha_{a_i} = 2\beta_i \beta_{i+1}^2 \cdots \beta_l^2 \, d\beta_i. \tag{A1b}$$

It is convenient to define the subgraphs $R_i = \{a_1, a_2, \ldots, a_i\}$ so that all variables $\alpha$ which are attached to $R_i$ are dilated by $\beta_i^2$ in the above change of variables.

Then, it is well known that, in the above change of variables, the Jacobian of the transformation is $2^l \prod_{i=1}^l \beta_i^{2l(R_i)-1}$,

$$\sum_{a=1}^{l} \alpha_a \to \sum_{i=1}^{l} \beta_i^2 \cdots \beta_l^2 \tag{A2}$$

and

$$P_G(\alpha) \to \prod_{i=1}^{l} \beta_i^{2L(R_i)} [1 + Q(\beta)] \tag{A3}$$

with $Q(\beta) \geqslant 0$ when $\beta \geqslant 0$ and $\beta_l$ independent.

We now transform the quadratic form $[sA_s(\alpha) + tA_t(\alpha) + \sum_{i=1}^4 p_i^2 A_i(\alpha)]$ into

$$s \sum_{c=1}^{q} \prod_{i=1}^{l} \beta_i^{2y_c(R_i)} \frac{[1 + \mathcal{D}_c(\beta)]}{1 + Q(\beta)} + \beta_l^2 \mathcal{E}(t, p_i^2, \beta). \tag{A4}$$

In (A4), we sum in the coefficient of $s$ over all $s$ cuts $c$; $y_c(R_j)$ is defined in (2.4); $\mathcal{D}_c(\beta) \geqslant 0$ when $\beta \geqslant 0$, and the functions $\mathcal{D}_c(\beta)$ and $\mathcal{E}(t, p_i^2, \beta)$ are $\beta_l$ independent. In the expression $[1 + \mathcal{D}_c(\beta)]$ the presence of 1 is due to the fact that for each $s$ cut $c$, the corresponding contribution to $A_s(\alpha)$ has a simultaneous Taylor series expansion.

### B. The $R$ operator and the absolute convergence

Next, we must tell what the renormalization operator $R$ becomes and how it acts. The $R$ operator acts upon $\alpha$ variables of each subgraph and does not recognize the subgraphs (except for the subgraphs $R_i$) when the integrand is expressed in the $\beta$ variables. Consequently, before performing the change of variables (A1), we must introduce new variables which allow the generalized Taylor operators $\mathcal{T}_\varphi$ to recognize its subgraph $\varphi$ when the change of variable $\alpha \to \beta$ is performed. The $R$ operator is expressed as a sum over all nests $\mathcal{N}$, of products of $\mathcal{T}$ operators. If we consider only one nest $\mathcal{N}$, the corresponding $\beta$ integrals diverge, but we know[15] how to construct for each sector $\mathcal{S}$, equivalent classes of nests $\Gamma$, such that for the sum over all nests in $\Gamma$, the corresponding $\beta$ integrals converge. We now resume here the main features of this construction.

(i) Each equivalent class $\Gamma$ is characterized by its maximal nest $\mathcal{G}$ and its minimal nest $\mathcal{K} \subseteq \mathcal{G}$.

(ii) Every nest $\mathcal{N}$ such that $\mathcal{K} \subseteq \mathcal{N} \subseteq \mathcal{G}$ belongs to $\Gamma$; each nest belongs only to one equivalent class $\Gamma$ and the sum over all equivalent classes reconstruct the sum over all nests.

(iii) The subgraphs of any nest $\mathcal{N}$ which belong to $\Gamma$ can be partitioned into the subgraphs of $\mathcal{K}$ and some subgraphs of $\mathcal{H} = \mathcal{G} - \mathcal{K}$. Consequently,

$$\sum_{\mathcal{N} \in \Gamma} \prod_{\varphi \in \mathcal{N}} (-\mathcal{T}_\varphi^{-2l(\varphi)}) = \prod_{\varphi \in \mathcal{K}} (-\mathcal{T}_\varphi^{-2l(\varphi)}) \prod_{\varphi \in \mathcal{H}} (1 - \mathcal{T}_\varphi^{-2l(\varphi)}). \tag{A5}$$

At this point we consider a given sector $\mathcal{S}$ and a given equivalent class of nests $\Gamma$.

In the construction of $\mathcal{K}$ and $\mathcal{H}$,[15] we define from the subgraphs $R^i$ the subnests $\mathcal{K}^i$ and $\mathcal{H}^i$ for $i = 1, \ldots, l$ such that $\cup \mathcal{K}^i = \mathcal{K}$ and $\cup \mathcal{H}^i = \mathcal{H}$, and we label the subgraphs of $\mathcal{K}^i$ and $\mathcal{H}^i$ by $K_j^i$ and $H_j^i$ for $j = 1, \ldots, r_i - 1$. Moreover, $K_j^i \subset H_j^i \subset K_{j+1}^i \subset \cdots$, and

$$H_j^i = K_{j+1}^i \cap (R^i \cup K_j^i) \tag{A6a}$$

$$\sum_{j=1}^{r_i - 1} [l(H_j^i) - l(K_j^i)] = l(R^i) \quad \text{for } i = 1, \ldots, l. \tag{A6b}$$

Let us remind to the reader that $\mathcal{H}^i$ is never empty.

We now define the new variables upon which the $\mathcal{T}$ operators act. Given a line $a \in K_j^i$ we dilate the variable $\alpha_a \to \alpha_a (\sigma_j^i)^2$, and given a line $a \in H_j^i$, we dilate $\alpha_a \to \alpha_a (\chi_j^i)^2$. Then, we perform the change of variables (A1).

*Theorem*: We denote by $(\mathcal{S}\Gamma)$ the transformation of a function $\mathcal{Z}(\alpha_a)$ into a function $\mathcal{Z}_{\mathcal{S}}{}^{\Gamma}(\beta_i, \sigma_j^i, \chi_j^i)$. Then, the function $\mathcal{Z}_{\mathcal{S}}{}^{\Gamma}$ is of the form $\mathcal{Z}_{\mathcal{S}}{}^{\Gamma}(\sigma_j^i/\beta_i, \beta_i \chi_j^i)$.

The proof is given in Ref. 15.

Consequently in the $(\mathcal{S}\Gamma)$ transformation,

$$P_G(\alpha) \xrightarrow{(\mathcal{S}\Gamma)} \prod_{K_j^i \in \mathcal{K}} \left(\frac{\sigma_j^i}{\beta_i}\right)^{2L(K_j^i)} \prod_{H_j^i \in \mathcal{H}} (\chi_j^i \beta_i)^{2L(H_j^i)} \times [1 + Q(\sigma/\beta, \beta\chi)] \tag{A7}$$

and the quadratic form $[sA_s(\alpha) + tA_t(\alpha) + \sum_{i=1}^4 p_i^2 A_i(\alpha)]$ becomes

$$\mathcal{A}\left(s, t, p_i^2, \frac{\sigma}{\beta}, \beta\chi\right) = s \sum_{c=1}^{q} \prod_{\substack{K_j^i \in \mathcal{K} \\ i \neq l}} \left(\frac{\sigma_j^i}{\beta_i}\right)^{2y_c(K_j^i)} \prod_{H_j^i \in \mathcal{H}} (\chi_j^i \beta_i)^{2y_c(H_j^i)} \frac{[1 + \mathcal{D}_c(\sigma/\beta, \beta\chi)]}{[1 + Q(\sigma/\beta, \beta\chi)]} + (\chi_1^l \beta_l)^2 \mathcal{E}(t, p_i^2, \sigma/\beta, \beta\chi), \tag{A8}$$

where $\mathcal{D}_c$ and $\mathcal{E}$ are $\beta_l\chi_1^l$ independent and where we use the fact that $K^l$ contains only the empty subgraph $\phi$ and $H^l=\{G\}$; by homogeneity we have $y_c(H_1^l=G)=1\ \forall\, c$.

We now apply the operators $T_{\sigma_j^i}^{-2l(K_j^i)}$ corresponding to subgraphs of $K$ by following the rules given in Ref. 12. We get a sum of terms of the form

$$\prod_{i=1}^{l} \beta_i^{\sum_{j=1}^{r_i-1}[4L(K_j^i)-4L(H_j^i)-a_j^i]} \prod_{H_j^i\in H} \left(1 - T_{\chi_j^i}^{\omega(H_j^i)}\right) \Lambda_{\{a_j^i\}}(\beta\chi) \tag{A9}$$

with $0\le a_j^i\le \omega(K_j^i)$ provided that all $K_j^i\in K$ are divergent subgraphs, otherwise we get zero. In (A9)

$$\Lambda_{\{a_j^i\}}(\beta\chi)=\prod_{K_j^i\in K}\left[\frac{1}{a_j^i!}\,\frac{\partial^{a_j^i}}{\partial(\sigma_j^i/\beta_i)^{a_j^i}}\right.$$
$$\left.\times\left\{\frac{\exp[i\mathcal{A}(s,t,p_i^2,\sigma/\beta,\beta\chi)]}{[1+Q(\sigma/\beta,\beta\chi)]^2}\right\}\right]_{\sigma_j^i=0}. \tag{A10}$$

It is important to note that $\Lambda_{\{a_j^i\}}(\beta\chi)$ has a Taylor expansion in the variables $\beta\chi$ around $\beta\chi=0$. Using the integral representation for the rest of the Taylor series relative to the elements of $H$, we finally transform (A9) into

$$\prod_{i=1}^{l}\beta_i^{p_i}\int_0^1 \prod_{\substack{H_j^i\in H\\ \omega(H_j^i)\ge 0}}\left[d\chi_j^i\,\frac{(1-\chi_j^i)^{\omega(H_j^i)}}{\omega(H_j^i)!}\left(\frac{\partial}{\partial(\chi_j^i\beta_i)}\right)^{\omega(H_j^i)+1}\right]$$
$$\times\Lambda_{\{a_j^i\}}(\beta\chi)\Big|_{\substack{\chi_j^i=1\\ \omega(H_j^i)<0}} \tag{A11}$$

with

$$p_i=\sum_{j=1}^{r_i-1}[4L(K_j^i)-4L(H_j^i)-a_j^i]+\sum_{H_j^i\ \mathrm{div}}[\omega(H_j^i)+1]. \tag{A12}$$

It is easy to prove that

$$2l(R_i)+p_i\ge\sum_{\substack{H_j^i\\ \omega(H_j^i)<0}}[-\omega(H_j^i)]+\sum_{\substack{H_j^i\\ \omega(H_j^i)\ge 0}}(1)>0. \tag{A13}$$

Thus, we have proved that for a given sector $\mathcal{S}$ and for a given equivalence class $\Gamma$, we obtain a sum of terms, each of which is of the form

$$\int_0^\infty d\beta_l\,\beta_l^{2l+p_l-1}\int_0^1\prod_{i=1}^{l-1}(d\beta_i\beta_i^{2l(R_i)+p_i-1})$$
$$\times\exp\Big(-i\sum_{i=1}^{l}\beta_i^2\cdots\beta_l^2 m^2\Big)$$
$$\times\int_0^1\prod_{\substack{H_j^i\in H\\ \omega(H_j^i)\ge 0}}\left[d\chi_j^i\,\frac{(1-\chi_j^i)^{\omega(H_j^i)}}{\omega(H_j^i)!}\left(\frac{\partial}{\partial(\chi_j^i\beta_i)}\right)^{\omega(H_j^i)+1}\right]$$
$$\times\Lambda_{\{a_j^i\}}(\beta\chi)\Big|_{\substack{\chi_j^i=1\\ \omega(H_j^i)<0.}} \tag{A14}$$

This achieves the proof of Bogoliubov and Parasiuk theorem which states the absolute convergence of the renormalized Feynman amplitudes. The expression (A14) is also the starting point if we wish to obtain a bound of $I_G(s)$ for large $s$.

### C. The behavior of $I_G(s)$ for large $s$

We describe now a generalization of the result given in Ref. 14 to the case where $G$ contains only logarithmically divergent subgraphs. Here, the numbers $a_j^i$ in (A10) are all null, and we calculate

$$\prod_{H_j^i\ \mathrm{div}}\frac{\partial}{\partial(\beta_i\chi_j^i)}\left\{\frac{\exp[i\mathcal{A}(s,t,p_i^2,0,\beta\chi)]}{[1+Q(0,\beta\chi)]^2}\right\}. \tag{A15}$$

Because the logarithmically divergent subgraphs $K_j^i$ are nonessential, $\mathcal{A}(s,t,p_i^2,0,\beta\chi)$ is $s$ dependent, but the sum over the $s$ cuts $c$ in (A8) is reduced to a sum over those cuts $c'$ which do not intersect the largest subgraph $K_j^i$.

The expression (A15) is a sum of terms of the form

$$s^{\sum_{c'}p_{c'}}(\beta_l\chi_1^l)^{2n}\prod_{i=1}^{l}(\beta_i\chi_{r_i-1}^i)^{2\sum_{c'}p_{c'}y_{c'}(H^i_{r_i-1})-\nu(H^i_{r_i-1})}$$
$$\times\phi(\beta\chi)\exp[i\mathcal{A}(s,t,p_i^2,0,\beta\chi)], \tag{A16}$$

where $p_{c'}$ and $n$ are nonnegative integers, $\phi(\beta\chi)$ is $\beta_l$ independent and has a simultaneous Taylor expansion in $\beta\chi$ around zero, and where the nonnegative integers $\nu(H_j^i)$ are null if $H_j^i$ is a convergent subgraph and are smaller or equal to $\inf[1, 2\sum_{c'}p_{c'}y_{c'}(H_j^i)]$ if $H_j^i$ is a divergent subgraph. In (A16) we have

$$\mathcal{A}(s,t,p_i^2,0,\beta\chi)=s\sum_{c'}\prod_{i=1}^{l}(\beta_i\chi_{r_i-1}^i)^{2y_{c'}(H^i_{r_i-1})}\left[\frac{1+\mathcal{D}_c(0,\beta\chi)}{1+Q(0,\beta\chi)}\right]$$
$$+(\beta_l\chi_1^l)^2\,\mathcal{E}(t,p_i^2,0,\beta\chi), \tag{A17}$$

where, of course, because of the property $K_j^i\subset H_j^i\subset K_{j+1}^i\subset\cdots$, we have only one possible variable $\chi_{r_i-1}^i$ in front of the square bracket [ ]. We now integrate over the variable $\beta_l$; using $p_l+2l=-\omega(G)>0$, we get

$$\Gamma(\gamma)\,s^{\sum_{c'}p_{c'}}\int_0^1\prod_{i\ne l}[d\beta_i\beta_i^{2i+p_i-1}]\prod_{H_j^i\ \mathrm{div}}d\chi_j^i$$
$$\times\prod_{i=1}^{l-1}(\beta_i\chi_{r_i-1}^i)^{2\sum_{c'}p_{c'}y_{c'}(H^i_{r_i-1})-\nu(H^i_{r_i-1})}$$
$$\times\phi(\beta\chi)\Big[im_{i_l}^2+i\sum_{j=1}^{l-1}\beta_j^2\cdots\beta_{l-1}^2m_{ij}^2-i\mathcal{A}(s,t,p_i^2,0,\beta\chi)/\beta_l^2\Big]^{-\gamma}, \tag{A18}$$

where

$$\gamma=n+\sum_{c'}p_{c'}-\omega(G)/2, \tag{A19}$$

and where all variables $\chi_j^i$ are set equal to one for convergent subgraphs $H_j^i$. We must note that with the definition (1.5) of a Feynman amplitude, the masses $m^2$ have a small negative imaginary part and the invariant $s$, $t$, $p_i^2$ a small positive imaginary part, so that at $\epsilon>0$ the square bracket [ ] in (A18) has a positive real part. At this step, we may use the following integral representation:

$$\Gamma(\gamma)\Big(\sum_{c'}A_{c'}+B\Big)^{-\gamma}=\prod_{c'}\left\{\frac{1}{2i\pi}\int_{p_{c'}-i\infty}^{p_{c'}+i\infty}dz_{c'}\right\}\prod_{c'}\Gamma(-z_{c'}+p_{c'})$$
$$\times\Gamma\Big[\sum_{c'}(z_{c'}-p_{c'})+\gamma\Big]\cdot\prod_{c'}A_{c'}^{z_{c'}-p_{c'}}B^{-[\sum_{c'}(z_{c'}-p_{c'})+\gamma]}, \tag{A20}$$

for $\mathrm{Re}\gamma > 0$, $\mathrm{Re}A_{c'} > 0$, $\mathrm{Re}B > 0$ and with $\mathrm{Re}Z_{c'} = \rho_{c'} < p_{c'}$, and $\sum_{c'}\rho_{c'} > \omega(G)/2 - n$. For $\epsilon > 0$, the above integral representation can be introduced in (A18) and the integrals over $z_{c'}$ and over the variables $\beta$ and $\chi$ may be interchanged in a region of $\rho_{c'}$ to be defined later on. For each term (A18), we define the multiple Mellin transform $M(z_{c'})$ by using (A20). We get

$$\prod_{c'}\left\{\frac{1}{2i\pi}\int_{\rho_{c'}-i\infty}^{\rho_{c'}+i\infty} dz_{c'}\right\} s^{\sum_{c'} z_{c'}} M(z_{c'}), \tag{A21}$$

with $M(z_{c'})$ given by

$$\begin{aligned} M(z_{c'}) = {} & \prod_{c'} \Gamma(-z_{c'} + p_{c'})\Gamma\left(\sum_{c'} z_{c'} + n - \frac{\omega(G)}{2}\right) \\ & \times \int_0^1 \prod_{i\neq l} [d\beta_i \beta_i^{2i+p_i-1}] \prod_{H_j^i \,\mathrm{div}} d\chi_j^i \\ & \cdot \prod_{i=1}^{l-1} (\beta_i \chi^i_{r_i-1})^{2\sum_{c'} p_{c'} y_{c'}(H^i_{r_i-1}) - \nu(H^i_{r_i-1})} \phi(\beta\chi) \\ & \cdot \prod_{c'} \left[-i \prod_{i=1}^{l-1} (\beta_i \chi^i_{r_i-1})^{2y_{c'}(H^i_{r_i-1})} \left\{\frac{1 + \mathcal{D}_c(0,\beta\chi)}{1 + Q(0,\beta\chi)}\right\}\right]^{z_{c'} - p_{c'}} \\ & \cdot \left[ i m^2_{i_l} + i \sum_{j=1}^{l-1} \beta_j^2 \cdots \beta_{l-1}^2 m^2_{ij} \right. \\ & \left. - i\mathcal{E}(t, p_i^2, 0, \beta\chi)\right]^{-\sum_{c'} z_{c'} + n - \omega(G)/2}. \end{aligned} \tag{A22}$$

It is clear that the multiple Mellin transform does have a simultaneous Taylor series expansion around $\beta_i$ and $\chi_j^i$ equal to zero, and, consequently, we may explore from (A22) the region of analyticity of $M(z_{c'})$. The integrals in $\chi^i_{r_i-1}$ converge if

$$2\sum_{c'} \rho_{c'} y_{c'}(H^i_{r_i-1}) > -1 + \nu(H^i_{r_i-1}) \quad \text{for } H^{i\neq l}_{r_i-1} \text{ div.} \tag{A23}$$

The integrals in $\beta_i$ converge if

$$2\sum_{c'} \rho_{c'} y_{c'}(H^i_{r_i-1}) > -(2i + p_i), \quad i \neq l. \tag{A24}$$

We note that (A24) is really a condition for $H^i_{r_i-1}$ convergent since it is automatically satisfied by (A23), the values of $\nu(H^i_{r_i-1})$ and (A13) for $H^i_{r_i-1}$ divergent. For each term (A18) the inequalities (A23), (A24) $\rho_{c'} < p_{c'}$ and $\sum_{c'}\rho_{c'} > [\omega(G)/2 - n]$ define a convex polyhedron inside which any point with coordinates $\rho_{c'}$, may be used to calculate the integrals (A22). The fact that such a polyhedron is nonempty justify the interchange of the $\beta$, $\chi$, and $z$ integrals.

To find the large $s$ behavior of a term of the type (A18) we must find the minimum $(\sum_{c'}\rho_{c'})$ over all points of the polyhedron. If $\Delta$ is such a minimum, then a term of the type (A18) behaves for $\epsilon > 0$ [14] as $s^{\Delta}$ up to logarithms of $s$. We do not prove here that this result remains valid for $\epsilon \to 0$ provided that the integrals (A22) exist. It remains to compare the different values of $\Delta$ obtained for the different terms of the type (A18), that is for the different values of $n$, $p_{c'}$, $\nu(H^i_{r_i-1})$ and for the different sectors and equivalent classes of nests. This is a difficult task for the most general graph which contributes to $G^2_{(4)}$, and at the present time we have not been able to justify (although we have no counterexample) the general rule given in Ref. 9. Anyhow, if

$$\Omega = \sup\{\Delta\}, \tag{A25}$$

then

$$|I_G(s)| < c s^{\Omega+\eta}, \quad \eta > 0. \tag{A26}$$

### D. Proof that $\Omega \leqslant -1$; graphs for $\Omega = -1$

The remaining part of this appendix proves that $\Omega \leqslant -1$ and determines the graphs such that $\Omega = -1$.

We consider a connected subgraph $\varphi$ of $G$ with $n(\varphi)$ vertices and $N(\varphi)$ external legs and an $s$ cut $c$ of $G$ which splits $\varphi$ into $\chi$ connected parts. Of course, $n(\varphi) \geqslant \chi$; moreover, $N(\varphi) \geqslant \chi$ because if there exists one connected part without an external leg of $\varphi$, then the $s$ cut of $G$ split $G$ into more than two connected parts (with $p_1$ and $p_2$ on one side, $p_3$ and $p_4$ on the other side) and $c$ is not an $s$ cut of $G$. Consequently,

$$n(\varphi) + N(\varphi) \geqslant 2\chi. \tag{A27}$$

Let us characterize the graphs satisfying the equality in (A27). From $n(\varphi) = N(\varphi) = \chi$, we get $l(\varphi) = \chi$ and then the number of independent loops is $L(\varphi) = 1$. If we call $l_c(\varphi)$ the number of lines of $\varphi$ cut by $c$ and $L_c(\varphi)$ the number of independent loops of $\varphi$ destroyed by $c$, using

$$l_c(\varphi) - L_c(\varphi) = \chi - 1 = y_c(\varphi), \tag{A28}$$

we see that either $L_c(\varphi) = 1$ and all lines of $\varphi$ are destroyed by $c$ (examples of such graphs are given in Figs. 12a, 12b) or $L_c(\varphi) = 0$ and all lines of $\varphi$ but one are destroyed by $c$. In this last case, since we have one loop and one line only which are not destroyed by $c$, we must have a subgraph $\varphi$ with tadpole as shown in Fig. 13.

We note that the subgraphs of Fig. 13 does not occur in $G$ and the subgraphs of Fig. 12a–12b occur only inside self-energy parts or 3-external legs vertices; otherwise, $G$ is not essentially planar. Because $[n(\varphi) + N(\varphi)]$ is equal to $[4 - \omega(\varphi)]$ which is even, we just proved that except for the subgraphs of Fig. 12a–12b, for all connected subgraphs $\varphi$ of $G$, we have

$$n(\varphi) + N(\varphi) \geqslant 2\chi + 2, \tag{A29}$$

that is,

$$\omega(\varphi)/2y_c(\varphi) \leqslant -1. \tag{A30}$$

For disconnected subgraphs $\varphi = \cup_i \varphi_i$, since

$$\omega(\varphi) = \sum_i \omega(\varphi_i), \tag{A31a}$$

$$y_c(\varphi) = \sum_i y_c(\varphi_i), \tag{A31b}$$

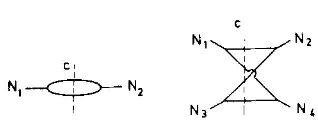



FIG. 12.

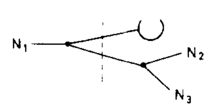



FIG. 13.

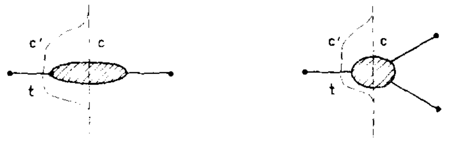



FIG. 14.

we have

$$\omega(\varphi)/2y_c(\varphi) \leqslant \sup_i[\omega(\varphi_i)/2y_c(\varphi_i)] \leqslant -1 \tag{A32a}$$

provided that any connected part $\varphi_i$ differs from a graph of the type given in Fig. 12 {since $\omega(\varphi_i) < 0$, $[\omega(\varphi_i)/y_c(\varphi_i)]$ is $-\infty$ for $y_c(\varphi_i) = 0$}. For a given nest $\mathcal{N}$, we call "bad cuts," those cuts (if any) which intersect a subgraph $\varphi \in \mathcal{N}$ in a way described in Fig. 12. For any "good cuts," (A32a) is valid. Since the graphs of Fig. 12 cannot be essential, for any subgraph we have

$$\omega(\varphi)/2y(\varphi) \leqslant -1. \tag{A32b}$$

With any subgraph $T$ of the type given in Fig. 12, we associate a line $t$ which is one of the external leg to the self energy which contains $T$ or the external leg to the three-external legs vertex which contains $T$ and on the same side of the $s$ cut $c$ (see Fig. 14).

Of course, $c'$ is an $s$ cut as well as $c$, and $t$ does not belong to a graph of the type given in Fig. 13 for the cut $c'$. Given a nest of subgraphs, it is always possible to find a good cut $c'$ which avoids the situation of Fig. 12 for any subgraph of the nest.

We consider a polyhedron defined by (A23), (A24), $\rho_{c'} < p_c$, and $\sum_{c'} \rho_{c'} > [\omega(G)/2 - n]$. Such a polyhedron is related to the large $s$ behavior of a term of the form (A18). When $n$ varies, we obtain nested polyhedrons $\mathcal{P}_0 \subseteq \mathcal{P}_1 \subseteq \cdots \subseteq \mathcal{P}_n \subseteq \cdots$ and consequently $\Delta(n=0) \geqslant \Delta(n=1) \geqslant \cdots$. Since we look for the sup$\{\Delta\}$, we keep in mind $\Delta(n=0)$. Similarly, in (A23) when the quantities $\nu(H^i_{r_i-1})$ vary, we obtain nested polyhedrons $\mathcal{P}_{\{\nu\}}$ with $\mathcal{P}_{\{0\}} \supseteq \cdots \supseteq \mathcal{P}_{\{\nu_{\max}\}}$ where

$$\nu^{\max}(H^i_{r_i-1}) = \inf[1, 2\sum_{c'} p_{c'} y_{c'}(H^i_{r_i-1})]. \tag{A33}$$

Consequently, the largest $\Delta$ is obtained for $\Delta(n=0, \{\nu\} = \{\nu^{\max}\})$. The same kind of arguments does not apply to the variations of $p_{c'}$, because $\nu^{\max}$ is $p_{c'}$ dependent and the various polyhedrons obtained, when $p_{c'}$, vary are not nested [the inf condition in (A33) prevents the polyhedrons to be empty].

We define

$$\Delta' = \inf_{c'} \left\{ \sup_{\varphi} \left( \frac{\alpha(\varphi)}{2y_{c'}(\varphi)} \right) \right\}, \tag{A34}$$

where inf is taken over all "good cuts" $c'$ and sup runs over all graphs of the nest such that $y_{c'}(\varphi) > 0$, and over $G$. In (A34), $\alpha(\varphi)$ stands for $-(2i + p_i)$ corresponding to the subgraph $\varphi = H^i_{r_i-1}$. We note that

$$\Delta' \leqslant -1 \tag{A35}$$

since we have

$$\Delta \leqslant \left\{ \sup_{\varphi, G} \frac{\alpha(\varphi)}{2y_{c'}(\varphi)} \right\}$$

for any "good cut" $c'$. From (A13) we see that

$$-(2i + p_i) \leqslant \omega(H^i_{r_i-1}) \tag{A36}$$

then,

$$\frac{\alpha(\varphi)}{2y_{c'}(\varphi)} \leqslant -1 \;\; \forall \;\; \varphi \;\; (y_{c'}(\varphi) > 0), \tag{A37}$$

and this proves (A35). Now,

$$\Delta(n=0, \{\nu\} = \{\nu_{\max}\}) \leqslant \Delta'. \tag{A38}$$

*Proof of* (A38): We consider a "good cut" $\bar{c}$ and a graph $\bar{\varphi}$ such that

$$\Delta' = \alpha(\bar{\varphi})/2y_{\bar{c}}(\bar{\varphi}) = \sup_{\varphi}\{\alpha(\varphi)/2y_{\bar{c}}(\varphi)\}. \tag{A39}$$

Now, we consider the point $P$

$$P \begin{cases} \varphi_{c'} = -\eta^2 & \text{if } p_{c'} = 0 \text{ and } c' \neq \bar{c}, \\ \varphi_{c'} = +\eta & \text{if } p_{c'} > 0 \text{ and } c' \neq \bar{c}, \\ \varphi_{\bar{c}} = \Delta' + \eta, \end{cases} \tag{A40}$$

where $\eta$ is positive and small ($\eta^2 \ll \eta$). Clearly $\rho_{c'} < p_{c'}$ since $\Delta' < 0$. Moreover, (A23), (A24) are valid since $\Delta' \geqslant \alpha(\varphi)/2y_{\bar{c}}(\varphi)$ for any $\varphi$ such that $y_{\bar{c}}(\varphi) \neq 0$; for those graphs $H^i_{r_i-1}$ which are not cut by $\bar{c}$, the left-hand side of (A23), (A24) is strictly positive if at least one $p_{c'}$ is positive (and if that $c'$ cut $H^i_{r_i-1}$) while the right-hand side is negative or null (if all $p_{c'}$ are null for the cuts $c'$ which intersect $H^i_{r_i-1}$, the left-hand side is negative as close as we wish from zero and the right-hand side is less or equal to $-1$). Consequently, $P$ belongs to the polyhedron $\mathcal{P}_{\{\nu\max\}}$. On the other hand, $(\sum_{c'} \rho_{c'})$ is as close as wanted from the diagonal hyperplane $[\sum_{c'} \rho_{c'} = \Delta']$. This hyperplane crosses the polyhedron $\mathcal{P}_{\{\nu\max\}}$ or is tangent to it, and this proves (A38).

Since $\Omega$ is the sup over all possible $\Delta$ and since we have (A38) and (A35), this proves that $\Omega \leqslant -1$.

Finally, we determine the conditions for $\Omega$ to be $-1$. The expressions in (A38) and in (A35) must become strict equalities. (A38) is an equality if the hyperplane $[(\sum_{c'} \rho_{c'}) = \Delta']$ is really tangent to the polyhedron $\mathcal{P}_{\{\nu\max\}}$. Let $\eta$ in the coordinates of $P$ becomes 0; then, the point $P$ belongs to the boundary of the polyhedron defined by the edges

$$2\sum_{c'} \rho_{c'} y_{c'}(H^i_{r_i-1}) = 0 \tag{A41}$$

for the divergent subgraphs $H^i_{r_i-1}$ such that there exists $c'$ with $p_{c'} > 0$ and $y_{c'}(H^i_{r_i-1}) > 0$,

$$2\sum_{c'} \rho_{c'} y_{c'}(H^i_{r_i-1}) = -(2i + p_i) \tag{A42}$$

for the convergent subgraphs $H^i_{r_i-1}$ such that $2y_{\bar{c}}(H^i_{r_i-1}) = (2i + p_i)$,

$$-\rho_{c'} = 0 \tag{A43}$$

for those cuts $c' \neq \bar{c}$ such that $p_{c'} = 0$. We suppose that the graph $G$ is such that $\omega(G) < -2$ since the case $\omega(G) = -2$ is trivial. The condition for the graphs $H^i_{r_i-1}$ in (A42) is such that

$$1 = \frac{2i + p_i}{2y_{\bar{c}}(H^i_{r_i-1})} \geqslant \frac{-\omega(H^i_{r_i-1})}{2y_{\bar{c}}(H^i_{r_i-1})} \geqslant 1, \tag{A44}$$

so that $(2i + p_i) = -\omega(H^i_{r_i-1})$ which is the case for instance

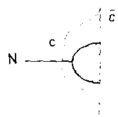


FIG. 15.

if $K^i$ is empty and if $H^i_{r_i-1}=R^i$. Moreover, these graphs should satisfy $2y_{\bar{c}}=-\omega$. The other "good cuts" $c'$ should be such that $y_{c'}\leq y_{\bar{c}}$ because of (A32a). A property of convex polyhedrons[14] is that if the diagonal hyperplane is tangent to the polyhedron in $P$, it may be generated by linear combinations of (A41)—(A43) with nonnegative coefficients. It is clear that this is only possible if all $y_{c'}=y_{\bar{c}}$ for all good cuts $c'$ which intersect the convergent subgraphs of (A42).

We now prove that a graph $\varphi$, such that

$$y_{c'}(\varphi)=y_{\bar{c}}(\varphi)\neq 0 \tag{A45}$$

for all "good cuts" $c'$ relative to a nest $\mathcal{N}$ which contains $\varphi$, and such that

$$\omega(\varphi)=-2y_{\bar{c}}(\varphi) \tag{A46}$$

is the union of single rungs (Fig. 5) and of logarithmically divergent subgraphs.

First, since $\bar{c}$ is a "good cut," for each connected part $\varphi_i$ of $\varphi$, we must have

$$\omega(\varphi_i)=-2y_{\bar{c}}(\varphi_i); \tag{A47}$$

otherwise, (A46) is not possible. Now the connected graphs which satisfy (A47), also satisfy

$$n(\varphi_i)+N(\varphi_i)=2\chi+2 \tag{A48}$$

where $n(\varphi_i)$, $N(\varphi_i)$, and $\chi$ are the number of vertices, external legs of $\varphi_i$ and the number of connected parts obtained after intersection of $\bar{c}$. Using again the inequalities $n(\varphi_i)\geq\chi$, $N(\varphi_i)\geq\chi$, we see that most connected parts should have only one external leg (Fig. 15). But if $\bar{c}$ is a good cut, clearly $c'$ is also a good cut; we have $y_{c'}<y_{\bar{c}}$ since $c'$ cut $\varphi_i$ into $(\chi-1)$ connected parts and, consequently, by (A45), one external leg connected parts are not allowed. We are left with the only other possibility: There are two external legs for each connected part $(N(\varphi)=2\chi)$. Then by (A48), $n(\varphi_i)=2$ and we obtain a single line. The only single line which satisfies (A45) is an essential single rung as shown in Fig. 5. Consequently, $\varphi$ is a union of single rungs, and of divergent subgraphs since these divergent pieces are not cut by the "good cuts" and do not destroy (A46); such graphs $\varphi$ are called leading. We proved in the same time that if, for a convergent nonleading subgraph $\sum$ and for a "good cut" $c$, we have $\omega(\sum)=-2y_c(\sum)$, then there exists a "good cut" $c'$ such that $y_{c'}(\sum)<y_c(\sum)$. For the leading subgraphs $\varphi$, we have for the "bad cuts" $c''$, $y_{c''}>y_{\bar{c}}$, and in the system (A42), (A43), we need Eq. (A43) for $c'=c''$ $(p_{c''}=0)$ in order to generate the tangent diagonal hyperplane in $P$; (A41) is then useless.

We just proved that, for those graphs which contain at least one single rung (Fig. 5) as essential subgraph, we have $\Omega=-1$.

## APPENDIX B

### A. The single Mellin transform $M_G(x)$

The single Mellin transform is defined in (3.7). The operators $\mathcal{T}$ which enter in $R$ are sensible to divergent subgraphs [by $P_G(\alpha,\gamma)$] and to essential subgraphs [by $A_s(\alpha,\gamma)$]. For an essential subgraph $\varphi$, we have a subtraction only if

$$\omega(\varphi)-E'[2y(\varphi)x]\geq 0 \tag{B1}$$

with $E'(a)$ defined as the smaller integer larger or equal to Re$a$, and where $x$ is taken between $-1$ and $0$.

By (A32b), the ratio $[\omega(\varphi)/2y(\varphi)]$ is always smaller or equal to $(-1)$. Consequently, for Re$x$ larger than $(-1)$, no essential subgraphs are subtracted. The same analysis as the one performed in Appendix A for the Feynman amplitude $I_G(s)$ can be written for $M_G(x)$ with $\exp[isA_s(\alpha,\gamma)]$ replaced by $[A_s(\alpha,\gamma)]^x$. We just mention here the technical differences. Since all subgraphs $K^i_j$ are nonessential, only the subgraphs $H^i_{r_i-1}$ might be essential because of the nested and alternance properties of the $K$'s and of the $H$'s. Then, we have

$$A_s(\beta_i\chi^i_j,\sigma^i_j/\beta_i)=(\beta_i\chi^i_{r_i-1})^{2y(H^i_{r_i-1})}\left[\sum_c\prod_{K^i_j}\left(\frac{\sigma^i_j}{\beta_i}\right)^{2y_c(K^i_j)}\right.$$
$$\left.\times\prod_{H^i_j}(\beta_i\chi^i_j)^{2y_c(H^i_j)-2y(H^i_j)}\left\{\frac{1+\mathcal{D}_c(\sigma/\beta,\beta\chi)}{1+Q(\sigma/\beta,\beta\chi)}\right\}\right]. \tag{B2}$$

Consequently, $p_i$ in (A12) has to be replaced by

$$p_i(x)=p_i+2y(H^i_{r_i-1})x. \tag{B3}$$

At $\sigma=0$, the part of the square bracket [ ] in (B2) which is not $Q$ or $\mathcal{D}_c$ is a function of $(\beta_i\chi^i_{r_i-1})$ only.

The equivalent equation to (A15) becomes

$$\Gamma(-x)\prod_{H^i_j\,\mathrm{div}}\frac{\partial}{\partial(\beta_i\chi^i_j)}$$
$$\times\left\{\frac{[(\beta_i\chi^i_{r_i-1})^{-2y(H^i_{r_i-1})}A_s(\beta_i\chi^i_j,0)]^x\exp(i(\beta_l\chi^l_1)^2\mathcal{E}(t,p^2_i,0,\beta\chi))}{(1+Q(0,\beta\chi))^2}\right\}. \tag{B4}$$

If we denote by $n$ the number of derivatives performed over the exponential and by $p_{c'}$ the number of derivatives performed in the square bracket [ ] of (B4) over the terms corresponding to an $s$ cut $c'$ which does not intersect any $K^i_j$, we get, up to multiplicative factors,

$$\Gamma(-x+\sum_{c'}p_{c'})(\beta_l\chi^l_1)^{2n}$$
$$\times\prod_{i=1}^{l-1}(\beta_i\chi^i_{r_i-1})^{2\Sigma_{c'}[y_{c'}(H^i_{r_i-1})-y(H^i_{r_i-1})]p_{c'}-\nu(H^i_{r_i-1})}$$
$$\times\phi(\beta\chi)[(\beta_i\chi^i_{r_i-1})^{-2y(H^i_{r_i-1})}A_s(\beta_i\chi^i_j,0)]^{x-\Sigma_{c'}p_{c'}}$$
$$\times\exp[i(\beta_l\chi^l_1)^2\mathcal{E}(t,p^2_i,0,\beta\chi)], \tag{B5}$$

where $\phi$ has a Taylor expansion in $\beta_i\chi^i_j$ and is $\beta_l$ independent and where $\nu(H^i_{r_i-1})$ satisfies the same inequality as in Appendix A since $y(H^i_{r_i-1})$ is null for a divergent subgraph. We now integrate over $\beta_l$ and obtain

$$\Gamma(-x+\sum_{c'}p_{c'})\,\Gamma\left(-\frac{\omega(G)}{2}+n+x\right)$$
$$\times\int_0^1\prod_{i=1}^{l-1}[d\beta_i\beta_i^{2i+p_i(x)-1}]\prod_{H^i_j\,\mathrm{div}}d\chi^i_j\,\phi(\beta\chi)$$

$$\times \prod_{i=1}^{l-1} (\beta_i \chi^i_{r_i-1})^{2\Sigma_{c'}[y_{c'}(H^i_{r_i-1})-y(H^i_{r_i-1})]\rho_{c'}-\nu(H^i_{r_i-1})}$$

$$\times\left\{\sum_{c'} \prod_{i=1}^{l-1} (\beta_i \chi^i_{r_i-1})^{2y_{c'}(H^i_{r_i-1})-2y(H^i_{r_i-1})} \frac{[1+\mathcal{D}_c(0,\beta\chi)]}{[1+Q(0,\beta\chi)]}\right\}^{x-\Sigma_{c'}\rho_{c'}}$$

$$\times[im^2_{i_i} - i\mathcal{E}(t,p^2_i,0,\beta\chi)]^{\omega(G)/2-n-x}. \tag{B6}$$

To evaluate the analyticity properties in $x$ of the expression (B6), it is convenient to use (A20) to transform the sum over $c'$ in the curly bracket $\{\}$ into a product. This leads exactly to the multiple Mellin transform (A22) if we write (B6) under the form

$$\prod_{c'}\left\{\frac{1}{2i\pi}\int_{\rho_{c'}-i\infty}^{\rho_{c'}+i\infty} dz_{c'}\right\} \delta(x-\sum_{c'} z_{c'}) M(z_{c'}). \tag{B7}$$

Since the single Mellin variable $x$ is the sum over all $s$ cuts $c'$ of the variables $z_{c'}$, it is easy from Appendix A to describe the analyticity properties of $M_G(x)$ in $x$. It was found there that a lower bound of analyticity in $x$ is given by $\Omega \le -1$ and equal to $(-1)$ for the graphs of Sec. 3. On the other hand, we may look for an upper bound. We consider the point $P:\{\rho_{c'}=0\}$ for all $s$ cuts $c'$. Such a point may be on the following edges of the polyhedron defined by (A23), (A24):

$$2\sum_{c'} \rho_{c'} y_{c'}(H^i_{r_i-1}) = 0 \tag{B8a}$$

for $H^i_{r_i-1}$ divergent and if $\exists\, c'$ such that $\rho_{c'} > 0$ and $y_{c'}(H^i_{r_i-1}) > 0$,

$$-\rho_{c'} = 0 \tag{B8b}$$

for those $s$ cuts $c'$ such that $\rho_{c'}=0$. It is impossible to generate a diagonal hyperplane tangent in $P$ at the polyhedron by linear combinations of (B8a) and (B8b) with nonnegative coefficients, except for the case where all $\rho_{c'}=0$ [(B8a) is then useless]. $P$ is then on the diagonal hyperplane $(\sum_{c'}\rho_{c'}=0)$ and that corresponds to the pole of $M_G(x)$ at $x=0$ which comes from the Euler function $\Gamma(-x)$. We just proved that $M_G(x)$ is analytic for $\Omega < \mathrm{Re}\,x < 0$, and in Sec. 3 for $-1 < \mathrm{Re}\,x < 0$. It is a consequence of Ref. 14 that $M_G(x)$ is meromorphic with poles at $\mathrm{Re}\,x = 0,1,2$ due to $\Gamma(-x)$ and $\mathrm{Re}\,x=\Omega$, $a_G < \Omega, \cdots$ due to the $\alpha$ integral. The purpose of Sec. 3 is to define the analytic continuation of $M_G(x)$ for $a_G < \mathrm{Re}\,x < -1$, and to extract the structure of the residue at $x=-1$.

### B. The function $\overline{M}_G(x)$

We define the function $\overline{M}_G(x)$ as the right-hand side of Eq. (3.7), where now the subtraction operator is defined for $a_G < \mathrm{Re}\,x < -1$.

The $R$ operator still subtracts once the logarithmically divergent subgraphs but it also subtracts once the leading subgraphs [essential subgraphs with $\omega(\varphi) = -2y(\varphi)$]. Such leading subgraphs were shown in Appendix A to be a nonempty union of single rungs with a union of divergent subgraphs. To find the analyticity properties of $\overline{M}_G(x)$ in $x$, we must reproduce Appendix A for this function. Let us give the main differences. The case $G$ leading is trivial, and we consider $G$ not leading.

First the elements of $\mathcal{K}$ are now divergent subgraphs and leading subgraphs. In the nest $\mathcal{K}\cup\mathcal{H}$, divergent subgraphs contain only divergent subgraphs and leading subgraphs contain only divergent and leading subgraphs. The largest $K^i$ being either divergent or leading all $H^i$'s except $H^i_{r_i-1}$ are divergent or leading. We have

$$y(K^i_j) \ge y(H^i_{j-1}) \ge y(K^i_{j-1}) \quad \forall\, i,j=1,2,\ldots,r_i-1. \tag{B9}$$

The function $A_s(\alpha)$ becomes now

$$A_s(\beta\chi,\sigma/\beta) = \prod_{K^i_j} (\sigma^i_j/\beta_i)^{2y(K^i_j)} \prod_{H^i_j} (\beta_i\chi^i_j)^{2y(H^i_j)}$$

$$\times\left[\sum_c \prod_{K^i_j} (\sigma^i_j/\beta_i)^{2y_c(K^i_j)-2y(K^i_j)} \times \prod_{H^i_j} (\beta_i\chi^i_j)^{2y_c(H^i_j)-2y(H^i_j)} \left\{\frac{1+\mathcal{D}_c(\sigma/\beta,\beta\chi)}{1+Q(\sigma/\beta,\beta\chi)}\right\}\right] \tag{B10}$$

and $p_i(x)$ becomes

$$p_i(x) = p'_i + 2\left[\sum_{H^i_j} y(H^i_j) - \sum_{K^i_j} y(K^i_j)\right] x \tag{B11}$$

with

$$p'_i = \sum_{j=1}^{r_i-1} [4L(K^i_j) - 4L(H^i_j)] + \sum_{H^i_j\ \mathrm{div}} (1) + \sum_{H^i_j\ \mathrm{leading}} (1). \tag{B12}$$

We define

$$V = \sum_{H^i_j} y(H^i_j) - \sum_{K^i_j} y(K^i_j) = y(H^i_{r_i-1}) + W. \tag{B13}$$

The quantity $W$ is $\le 0$ but $V$ may have all signs except if $H^i_{r_i-1}$ is leading where $V \ge 0$.

When we set $\sigma$ equal to zero in the square bracket [ ] in (B10), we obtain, except for the functions $\mathcal{D}_c$ and $Q$, only a dependence in $(\beta_i\chi^i_{r_i-1})$. The reason is the following: If $\varphi$ is divergent or leading, $y_c(\varphi)$ is equal to $y(\varphi)$ for all "good cuts" $c$. When we set $\sigma^i_{r_i-1}$ equal to zero all terms of the square bracket disappear except those corresponding to "good cuts" $c'$. Such a cut $c'$ do not intersect the divergent subgraphs of any $H^i_j$ and consequently $y_c(H^i_j)$ is equal to $y(H^i_j)$ and the variables $(\beta_i\chi^i_j)$ disappear. Then

$$\prod_{K^i_j} (\sigma^i_j/\beta_i)^{2y(K^i_j)} \prod_{H^i_j} (\beta_i\chi^i_j)^{2y(H^i_j)} A_s(\beta\chi,\sigma/\beta)\Big|_{\sigma^i_j=0}$$

$$= \left[\sum_{c'} \prod_{i\ne l} (\beta_i\chi^i_{r_i-1})^{2y_{c'}(H^i_{r_i-1})-2y(H^i_{r_i-1})} \left\{\frac{1+\mathcal{D}_c(0,\beta\chi)}{1+Q(0,\beta\chi)}\right\}\right]. \tag{B14}$$

After taking the derivatives $\partial/\partial\beta\chi$ on the divergent and leading subgraphs $H^i_j$, we get an expression similar to (B5) where now

$$0 \le \nu(H^i_{r_i-1}) \le \inf[1, 2\sum_{c'} (y_{c'}(H^i_{r_i-1}) - y(H^i_{r_i-1}))\rho_{c'}] \tag{B15}$$

for divergent and leading subgraphs $H^i_{r_i-1}$ and zero otherwise. The integration over $\beta_i$ gives back (B6) where now we have also integrations over the variables

$\chi_j^i$ corresponding to leading subgraphs and $p_i(x)$ is given by (B11). We then use a multiple Mellin representation of (B6), and we obtain "polyhedric conditions" for the integrals in (B6) to exist. We get, instead of (A22),

$$M(z_{c'}) = \prod_{c'} \Gamma(-z_{c'} + p_{c'})\, \Gamma\left(\sum_{c'} z_{c'} + n - \frac{\omega(G)}{2}\right)$$
$$\times \int_0^1 \prod_{i \neq l} [d\beta_i \beta_i^{2i+p_i(x)-1}] \prod_{H_j^i \left\{ \substack{\mathrm{div} \\ \mathrm{leading}} \right.} d\chi_j^i$$
$$\times \prod_{i=1}^{l-1} (\beta_i \chi_{r_i-1}^i)^{2\Sigma_{c'}[y_{c'}(H_{r_i-1}^i) - y(H_{r_i-1}^i)] z_{c'} - \nu(H_{r_i-1}^i)}$$
$$\times \phi(\beta\chi) \prod_{c'} \left[\frac{1 + D_c(0, \beta\chi)}{1 + Q(0, \beta\chi)}\right]^{z_{c'} - p_{c'}}$$
$$\times [im_{i_l}^2 + i \sum_{j=1}^{l-1} \beta_j^2 \cdots \beta_{l-1}^2 m_{i_j}^2$$
$$- i\mathcal{E}(t, p_i^2, 0, \beta\chi)]^{-[\Sigma_{c'} z_{c'} + n - \omega(G)/2]}. \tag{B16}$$

Absolute convergence for the variables $\chi_j^i$ requires

$$2\sum_{c'} [y_{c'}(H_{r_i-1}^i) - y(H_{r_i-1}^i)] \rho_{c'} > -1 + \nu(H_{r_i-1}^i), \tag{B17}$$

and absolute convergence for the variables $\beta_i$ means

$$2\sum_{c'} [y_{c'}(H_{r_i-1}^i) - y(H_{r_i-1}^i)] \rho_{c'} + 2\left[\sum_{H_j^i} y(H_j^i) - \sum_{K_j^i} y(K_j^i)\right] \left(\sum_{c'} \rho_{c'}\right) > -(2i + p_i') + \nu(H_{r_i-1}^i). \tag{B18}$$

In (B17), $H_{r_i-1}^i$ is either a divergent subgraph or a leading subgraph. The polyhedron of definition for the integral (B16) is given by (B17), (B18) and by $\rho_{c'} < p_{c'}$, $\sum_{c'} \rho_{c'} > \omega(G)/2 - n$. Again, we look for the hyperplane ($\sum_{c'} \rho_{c'} = \Delta$) which is tangent to the polyhedron and below, and we look for the polyhedron which gives the largest $\Delta$. Then, we take $n$ equal to zero and $\nu(H_{r_i-1}^i)$ equal to $\nu^{\max}(H_{r_i-1}^i)$ defined as the right-hand side of (B15). We note that if $H_{r_i-1}^i$ is divergent, (B18) is automatically verified by (B17) and (B11), (B12). It is convenient to rewrite (B17), (B18) under the form (with $\nu^{\max}$ replacing $\nu$)

$$2\sum_{c''} y_{c''}(H_{r_i-1}^i) \rho_{c''} > -1 + \nu^{\max}(H_{r_i-1}^i) \quad \text{if } H_{r_i-1}^i \text{ is div,} \tag{B19}$$

$$2\sum_{c''} [y_{c''}(H_{r_i-1}^i) - y(H_{r_i-1}^i)] \rho_{c''} > -1 + \nu^{\max}(H_{r_i-1}^i) \quad \text{if } H_{r_i-1}^i \text{ is leading,} \tag{B20}$$

where $c''$ are "bad cuts" which intersect divergent subgraphs of $H_{r_i-1}^i$,

$$2\sum_{c'} [y_{c'}(H_{r_i-1}^i) + W] \rho_{c'} > \omega(H_{r_i-1}^i) - 2W - \sum_{H_j^i \,\mathrm{div}} (1) - \sum_{H_j^i \,\mathrm{leading}} (1) + \nu^{\max}(H_{r_i-1}^i), \tag{B21}$$

for any convergent subgraph $H_{r_i-1}^i$, leading or not. In (B21), we exclude the case where $H_{r_i-1}^i$ is the only element of $\mathcal{H}^i$ and where

$$W = -y(K_{r_i-1}^i) = -y(H_{r_i-1}^i) \tag{B22}$$

since this case is already taken into account by (B20). In all other cases, $[y_{c'}(H_{r_i-1}^i) + W]$ is not necessarily positive, except for $H_{r_i-1}^i$ leading. We now prove that the right-hand side of (B21) is negative or null.

—If $H_{r_i-1}^i$ is leading $\omega(H_{r_i-1}^i)$ is $[-2y(H_{r_i-1}^i)]$ and $[\omega(H_{r_i-1}^i) - 2W] = -2V$ is negative or null; the remaining part is also trivially negative or null. We note that the right-hand side of (B21) is null in this case only if (B21) is equivalent to (B20) (with $\nu^{\max} = 1$); this being excluded from (B21), we conclude that if $H_{r_i-1}^i$ is leading, the right-hand side of (B21) is strictly negative.

—If $H_{r_i-1}^i$ is not leading but convergent, $\nu^{\max}$ is null, $[W + y(K_{r_i-1}^i)]$ is positive or null, so that we must prove that $[\omega(H_{r_i-1}^i) + 2y(K_{r_i-1}^i)]$ is negative or null. The subgraph $K_{r_i-1}^i$ possesses $y(K_{r_i-1}^i)$ single rungs; each single rung has four adjacent lines. Let $q$ be the number of adjacent lines in $H_{r_i-1}^i$ and let $\varphi$ be the subgraph obtained from $H_{r_i-1}^i$ after cutting the $r$ rungs and the $q$ adjacent lines. Let $q'$ be the number of loops destroyed by the cutting of these lines; we have

$$\omega(\varphi) \leq 0, \tag{B23}$$

$$q' \leq E(q/2) \leq q/2. \tag{B24}$$

Now,

$$\omega(H_{r_i-1}^i) = \omega(\varphi) - 2y(K_{r_i-1}^i) - 2q - 4q'. \tag{B25}$$

This proves that $[\omega(H_{r_i-1}^i) + 2y(K_{r_i-1}^i)]$ is negative or null and then that the right-hand side of (B21) is negative or null. For this quantity to be null, we must have the following conditions: $H_{r_i-1}^i$ is a convergent nonleading subgraph and $\mathcal{H}^i = \{H_{r_i-1}^i\}$, $\omega(\varphi) = 0$ which implies $q > 0$ since $H_{r_i-1}^i$ is not leading, and the $q/2$ pair of adjacent lines form with the single rungs of $K_{r_i-1}^i$, $q/2 = q'$ independent loops. In this case

$$y(H_{r_i-1}^i) = y(K_{r_i-1}^i) - q/2 < y(K_{r_i-1}^i) \tag{B26}$$

and for all "good cuts" $c'$, the quantities $[y_{c'}(H_{r_i-1}^i) + W]$ are equal and strictly negative. We may now look for the hyperplane ($\sum_{c'} \rho_{c'} = \Delta$) tangent to the polyhedron and below. We first note that since $G$ is not leading, the inequality $2\sum_{c'} \rho_{c'} > \omega(G)$ is a special case of (B21) for the index $i = l$. We define

$$\Delta' = \inf_{c'} \left\{ \sup_i \left(\frac{\alpha(H_{r_i-1}^i)}{\beta_{c'}(H_{r_i-1}^i)}\right)\right\}, \tag{B27}$$

where inf is taken over all "good cuts" $c'$ and sup runs only over the indices $i$ such that $\beta_{c'}(H_{r_i-1}^i) > 0$. The quantities $\alpha(H_{r_i-1}^i)$ are the right-hand sides of the inequalities (B21) and $\beta_{c'}(H_{r_i-1}^i)$ is the corresponding coefficient of $\rho_{c'}$ on the left-hand side. Since $\beta_{c'}(H_{r_i-1}^i) > 0$, $\alpha(H_{r_i-1}^i)$ is strictly negative (at least for $H_{r_i-1}^i = G$, we have $\beta_{c'} > 0$). Then,

$$\alpha(H_{r_i-1}^i)/\beta_{c'}(H_{r_i-1}^i) = -1 + \left(\omega(H_{r_i-1}^i) + 2y_{c'}(H_{r_i-1}^i) - \sum_{H_j^i \,\mathrm{div}} (1) - \sum_{H_j^i \,\mathrm{leading}} (1) + \nu^{\max}(H_{r_i-1}^i)\right)\{2[y_{c'}(H_{r_i-1}^i) + W]\}^{-1} \tag{B28}$$

is less or equal to $(-1)$. It is equal to $(-1)$ if $\mathcal{H}^i = \{H^i_{r_i-1}\}$, $\mathcal{K}^i$ is empty or $K^i_{r_i-1}$ is a divergent subgraph, so that $W=0$, $H^i_{r_i-1}$, is leading and there exists a "bad cut" $c''$ such that $p_{c''}>0$, $y_{c''}(H^i_{r_i-1}) > y(H^i_{r_i-1})$. It is also equal to $(-1)$ if $\mathcal{H}^i=\{H^i_{r_i-1}\}$, $H^i_{r_i-1}$ is convergent nonleading, and if $\omega(H^i_{r_i-1}) = -2y_{c'}(H^i_{r_i-1})$.

Consequently,

$$\Delta' \leq -1. \tag{B29}$$

Now, we prove that $\Delta \leq \Delta'$. We define $\overline{\varphi}$ as one of the subgraph $H^i_{r_i-1}$, and the "good cut" $\overline{c}$, such that $\beta_{\overline{c}}(\overline{\varphi}) > 0$ and

$$\Delta' = \alpha(\overline{\varphi})/\beta_{\overline{c}}(\overline{\varphi}) \geq \alpha(\varphi)/\beta_{\overline{c}}(\varphi) \tag{B30}$$

for all $\varphi$ such that $\beta_{\overline{c}}(\varphi)>0$. We consider the point $P$ such that

$$\begin{aligned} \rho_{c'} &= -\eta^3 \quad \text{if } p_{c'}=0 \text{ and } c'\neq\overline{c},\\ \rho_{c'} &= +\eta^2 \quad \text{if } p_{c'}>0 \text{ and } c'\neq\overline{c},\\ \rho_{\overline{c}} &= \Delta'+\eta, \end{aligned} \tag{B31}$$

where $\eta$ is positive and small $(\eta^3 \ll \eta^2 \ll \eta)$. The point $P$ is as close as wanted to the hyperplane $(\sum_{c'}\rho_{c'} = \Delta')$. The point $P$ satisfies (B19) and (B20). For (B21), either $[y_{\overline{c}}+W]$ is negative, and since $\Delta'$ is also negative (B21) is valid; either $[y_{\overline{c}}+W]$ is positive and by (B30), (B21) is valid; or $[y_{\overline{c}}+W]$ is null and (B21) is valid because the right-hand side of (B21) is, then, a strictly negative integer. Consequently, the point $P$ is inside the polyhedron for $\eta$ small and positive, and

$$\Delta \leq \Delta'. \tag{B32}$$

It remains to show that when $\Delta'=-1$, $\Delta<\Delta'$.

We consider the point $\{\rho_{c'}=0 \text{ for } c''\neq\overline{c},\ \rho_{\overline{c}}=-1\}$. This point is on the following edges of the polyhedron defined by (B17), (B20), (B21):

$$2\sum_{c''} y_{c''}(H^i_{r_i-1})\,\rho_{c''}=0 \tag{B33a}$$

for $H^i_{r_i-1}$ divergent and if $\exists\, c''$ such that $p_{c''}>0$, $y_{c''}(H^i_{r_i-1})>0$,

$$2\sum_{c''}[y_{c''}(H^i_{r_i-1}) - y(H^i_{r_i-1})]\,\rho_{c''}=0 \tag{B33b}$$

for $H^i_{r_i-1}$ leading and if $\exists\, c''$ such that $p_{c''}>0$, $y_{c''}(H^i_{r_i-1}) > y(H^i_{r_i-1})$,

$$2\sum_{c'}[y_{c'}(H^i_{r_i-1})+W]\,\rho_{c'} = \omega(H^i_{r_i-1}) - 2W \tag{B33c}$$

for $H^i_{r_i-1}$ leading, if $\mathcal{H}^i=\{H^i_{r_i-1}\}$, $W\neq -y(H^i_{r_i-1})$, $\exists\, c''$ such that $p_{c''}>0$, $y_{c''}(H^i_{r_i-1}) > y(H^i_{r_i-1})$, and for $H^i_{r_i-1}$ convergent and nonleading, if $\mathcal{H}^i=\{H^i_{r_i-1}\}$; in both cases we must have $\omega(H^i_{r_i-1}) = -2y_{\overline{c}}(H^i_{r_i-1})$,

$$-\rho_{c'}=0 \tag{B33d}$$

for $p_{c'}=0$ and $c'\neq\overline{c}$. We must find a linear combination of (B33a)—(B33d) with nonnegative coefficients which generates a diagonal hyperplane. If (B33c) is used with $H^i_{r_i-1}$ leading, for all "good cuts" $y_{c'}=y_{\overline{c}}$, but for the "bad cuts" $c''$, $y_{c''}>y_{\overline{c}}$ and since $p_{c''}>0$, there is no way of obtaining a diagonal hyperplane. If (B33c) is used for $H^i_{r_i-1}$ convergent nonleading, we proved at the end of Appendix A that since $\omega(H^i_{r_i-1}) = -2y_{\overline{c}}(H^i_{r_i-1})$, there exists another "good cut" $c'$ such that $y_{c'}<y_{\overline{c}}$ and clearly again there is no way of generating a diagonal hyperplane. Consequently, if $\Delta'=-1$, $\Delta$ is strictly less than $\Delta'$. We have just proved that a lower bound of analyticity in $x$ for $\overline{M}_G(x)$ is given by a number $a_G<-1$. Let us look for an upper bound.

We consider the point $\{\rho_{c'}=0\ \forall\, c'\}$. This point might be on the edges of the polyhedron. First, it might be on the edges $\{-\rho_{c'}=-p_{c'}=0 \text{ for all } c'\}$. In this case we do generate a diagonal hyperplane by linear combinations of the edges with positive coefficients and this explains the pole at $x=0$ in $\Gamma(-x)$. The above point may be on the edges (B19), (B20), but it is impossible to generate a diagonal hyperplane (with positive coefficients) between them and with the edges $\{-\rho_{c'}=-p_{c'}=0\}$. On the contrary, the point may be on the edges (B21) when $H^i_{r_i-1}$ is not leading convergent and the right-hand side is zero; in this case the quantities $[y_{c'}+W]$ are negative and equal for all "bad cuts" $c''$, with the edges $\{-\rho_{c''}=-p_{c''}=0\}$ we may generate a diagonal hyperplane even if $y_{c''}>y_{c'}$.

We have proved in part B of this appendix that $\overline{M}_G(x)$ is analytic for $a_G<\operatorname{Re}x<0$ with $a_G<-1$ and that in some cases the pole at $x=0$ is not only due to the Euler function $\Gamma(-x)$ but also to the $\alpha$ integral.

### C. The function $\overline{M}_{G/J}(x)$

The function $\overline{M}_{G/J}(x)$ is defined in (3.12) and is found to be the product of several functions $M_{\overline{G}_i}$ and of two functions $\overline{M}_{\overline{K}_i}$ defined respectively in (4.3) and (4.2). Their peculiarity is that $[N_{G_i}/P_{\overline{G}_i}]$ and $[N_{K_i}/P_{\overline{K}_i}]$ are homogeneous in all variables $\alpha_a$ of degree respectively equal to $(-1)$ and zero. The operator $R$ subtracts the leading and the divergent subgraphs of the kernels $G_i$ or $K_i$. If we perform a decomposition of the $\alpha$ integrals into Hepp's sectors and into equivalent classes of nests, we obtain for the functions $[N_{G_i}/P_{\overline{G}_i}]$ and $[N_{K_i}/P_{\overline{K}_i}]$, two expressions similar to the expression (B10). For some graph $\varphi^i$ which are the union of a subgraph $\varphi'$ in the kernels $G_i$ or $K_i$ with some lines external to the kernels, the functions $y_c(\varphi)$ are given by

$$y_c(\varphi) = y_c(\varphi') - \delta(\varphi), \quad \forall\, c \tag{B34}$$

where $\delta(\varphi)$ is zero, one or two if $\varphi$ has zero, one or two more independent loops than $\varphi'$. Some coefficients $y_c(\varphi)$ may become negative. The graphs $K^i_j$ are necessarily leading or divergent and belong to the kernels $K_i$ or $G_i$; only $H^i_{r_i-1}$ may have external lines to the kernel.

It is easy to see that, since

$$H^i_{r_i-1} = R^i \cup K^i_{r_i-1} \tag{B35}$$

if $R^i = R'^i \cup \{l\}$, where $R'^i$ is a possibly empty subgraph of the kernel, we have

$$H^i_{r_i-1} = H'^i_{r_i-1} \cup \{l\}, \tag{B36}$$

where $H'^i_{r_i-1}$ belongs to the kernel and is $(R'^i \cup K^i_{r_i-1})$. Consequently, to any equivalent class $\Gamma$ with the nests $\mathcal{K}^i$, $\mathcal{H}^i$, we may associate an equivalent class $\Gamma'$ with the nests $\mathcal{K}'^i=\mathcal{K}^i$, $\mathcal{H}'^i=\{H'^i_j=H^i_j, H'^i_{r_i-1}\}$ built uniquely with graphs in the kernel. If now we look for the polyhedric conditions of analyticity, we get (B17), (B18), $\rho_{c'} < \rho_{c'}$ and $[1-\delta(G)](\sum_{c'}\rho_{c'}) > \omega(G)/2 - n$. The smallest polyhedron is given by $\nu = \nu^{\max}$ and $n=0$ and we obtain the conditions (B19), (B20), (B21).

From the entire graph $\overline{K}_i$, we get no condition since $\delta(G)=1$; from the entire graph $\overline{G}_i$, we get

$$\sum_{c'} \rho_{c'} < -\omega(\overline{G}_i)/2. \tag{B37}$$

Since $\omega(\overline{G}_i) \leq 0$ (it is zero for the graph of Fig. 9), (B37) contains $\sum_{c'}\rho_{c'} < 0$. From the graphs $H^i_{r_i-1}$ such that their part $H'^i_{r_i-1}$ in the kernel is empty, we get no condition since they have no loops. Now we prove that for all graphs $H^i_{r_i-1}$, the conditions (B19), (B20), (B21), $\sum_{c'}\rho_{c'} < 0$ define a polyhedron which completely contains the corresponding polyhedron obtained from the corresponding graphs $H'^i_{r_i-1}$ in the kernel. The conditions (B19) and (B20) are the same; we prove that the conditions (B21) for $H'^i_{r_i-1}$ leads automatically to the conditions (B21) for $H^i_{r_i-1}$. We have from (B34)

$$2\sum_{c'}[\nu_{c'}(H^i_{r_i-1}) + W]\rho_{c'} = 2\sum_{c'}[\nu_{c'}(H'^i_{r_i-1}) + W]\rho_{c'} - 2\delta(H^i_{r_i-1})\Big(\sum_{c'}\rho_{c'}\Big), \tag{B38}$$

and, since $(\sum_{c'}\rho_{c'})$ is negative or null,

$$2\sum_{c'}[\nu_{c'}(H^i_{r_i-1}) + W]\rho_{c'} \geq 2\sum_{c'}[\nu_{c'}(H'^i_{r_i-1}) + W]\rho_{c'}. \tag{B39}$$

On the other hand, $\omega(H'^i_{r_i-1}) \geq \omega(H^i_{r_i-1})$ since each loop of $H^i_{r_i-1}$ which is not in $H'^i_{r_i-1}$ contains independently two external legs to the kernel. We proved that the tangent diagonal hyperplane from below to the polyhedron described by (B19)–(B21) for the graphs $H^i_{r_i-1}$ is below or equal to the tangent diagonal hyperplane to the corresponding polyhedron for the graphs $H'^i_{r_i-1}$ which are all in the kernel. Thus we have for region of analyticity in $x$ of $\Gamma(-x)\overline{M}_{G_i}(x)$,

$$a_{\overline{G}_i} < \mathrm{Re}x < 0 \tag{B40}$$

with $a_{\overline{G}_i} < -1$. A similar relation holds for $\Gamma(-x)\, M_{\overline{K}_i}(x)$. The singularity at $\mathrm{Re}x=0$ might, in some cases, comes from the $\alpha$ integral. Finally, the region of analyticity in $x$ for $\overline{M}_{G/J}(x)$ is given by

$$a_{G/J} < \mathrm{Re}x < 0 \tag{B41}$$

with $a_{G/J} = \sup_{\overline{G}_i \overline{K}_i} (a_{\overline{G}_i}, a_{\overline{K}_i}) < -1$.

Let us mention that the singularity at $\mathrm{Re}x=0$ which comes from the $\alpha$-integral (as in Fig. 9 for instance), is present simultaneously in $\overline{M}_G(x)$ and in some of the functions $\overline{M}_{G/J}(x)$ and is spurious since it was absent from $M_G(x)$. This spurious singularity is due to our method of desingularization, but this method is at present the only one we know which allows the summation of Sec. 4.


[1] T. Regge, Nuovo Cimento **14**, 951 (1959); **18**, 947 (1960).

[2] M. Gell-Mann and M. L. Goldberger, Phys. Rev. Lett. **9**, 275 (1962), S. Mandelstam, Phys. Rev. **137**, 949 (1965).

[3] B.W. Lee and R.F. Sawyer, Phys. Rev. **127**, 2266 (1962).

[4] J. Dash, Nuovo Cimento **9**, 265 (1972), and private communication.

[5] H.T. Nieh and Y.P. Yao, Phys. Rev. Lett. **32**, 1074 (1974); B.M. McCoy and T.T. Wu, Phys. Rev. D **12**, 3257, (1975); **13**, 1076 (1976); C.Y. Lo and H. Cheng, Phys. Rev. D **13**, 1131 (1976).

[6] See, for example, R.J. Eden, P.V. Landshoff, D.I. Olive, and J.C. Polkinghorne, *The Analytic S-Matrix* (Cambridge U.P., Cambridge, 1966).

[7] P.G. Federbush and M.T. Grisaru, Ann. Phys. **22**, 263 (1963); J.D. Bjorken and T.T. Wu, Phys. Rev. **130**, 2566 (1963); G. Tiktopoulos, Phys. Rev. **131**, 480, 2373 (1963); T.L. Trueman and T. Yao, Phys. Rev. **132**, 2741 (1963); J.C. Polkinghorne, J. Math. Phys. **4**, 503, 1396 (1963); A.V. Efremov, Joint Inst. for Nuclear Research Preprint R-1242, R-1305 (1963). Zh. Eksp. Teor. Fiz. **47**, 1099 (1964).

[8] O.I. Zav'yalov, Zh. Eksp. Teor. Fiz. **47**, 1099 (1964) [Sov. Phys. JETP, 20, 736 (1965)].

[9] O.I. Zav'yalov and B.M. Stepanov, Yad. Fiz. **1**, 922 (1965) [Sov. J. Nucl. Phys. **1**, 658 (1965)].

[10] For example, A.V. Efremov, I.F. Ginzburg, and V.G. Serbo, JINR Preprint E2-4572 (1969).

[11] M.C. Bergère and Y.M.P. Lam, Comm. Math. Phys. **39**, 1 (1974); Berlin Preprint FUB HEP(May 1974).

[12] M.C. Bergère and J.B. Zuber, Comm. Math. Phys. **35**, 113 (1974).

[13] M.C. Bergère and C. Bervillier, to be published.

[14] M.C. Bergère, C. de Calan, and A.P.C. Malbouisson, "A theorem on asymptotic expansion of Feynman amplitudes," Ecole Polytechnique Preprint A2580477 (1977).

[15] M.C. Bergère and Y.M.P. Lam, J. Math. Phys. **17**, 1546 (1976).

[16] E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge U.P., Cambridge, 1958), 4th ed., pp. 132–33.

[17] K. Nishijima, T.N. Pham, and Z.F. Ezawa, Prog. Theor. Phys. **57**, 1013 (1977).

# A note on recoupling coefficients for SU(3)[a)]

D. J. Millener

*Brookhaven National Laboratory, Upton, New York 11973*

(Received 6 February 1978)

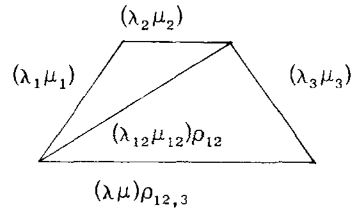
A 6-$(\lambda\mu)$ coefficient, denoted by $Z$ and different from the usual $U$ coefficient, associated with a specific recoupling of three irreducible representations of SU(3), is defined. A general 9-$(\lambda\mu)$ coefficient, analogous to the unitary 9-$J$ coefficient of the angular momentum Racah algebra, is then expressed in terms of the $Z$ coefficient and two $U$ coefficients. In this way problems associated with the existence of outer multiplicities in the products of irreducible representations of SU(3) are circumvented.

## 1. INTRODUCTION

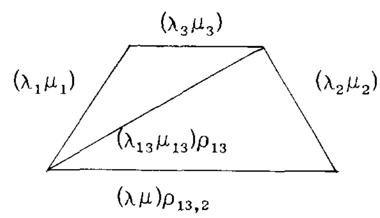
Classification according to SU(3) of the spatial part of wavefunctions for light nuclei provides a good basis for shell model calculations. The most efficient way to perform such calculations is to first make an expansion of the effective interaction into SU(3) irreducible tensors by appropriately coupling togther the creation and annihilation operators for particles in individual oscillator shells.[1-3] The evaluation of many-body matrix elements in terms of single shell matrix elements[4] is then reduced to an exercise in the use of the Wigner–Eckart theorem and recoupling techniques for the groups involved. The spherical tensor formalism employed in the Rochester–Oak–Ridge shell model code[4] carries over almost unchanged to the present problem, the only essentially new feature being the appearance of outer multiplicity labels for SU(3) couplings. Draayer and Akiyama have given algorithms for calculating Wigner and Racah coefficients for SU(3) in the most general case[5] and have provided computer codes for evaluating these coefficients.[6] Their results are sufficient for shell model calculations within a single major shell. For more than one shell 9-$(\lambda\mu)$ coefficients are required. It is the purpose of this note to describe a general method for calculating such 9-$(\lambda\mu)$ coefficients. In special cases 9-$(\lambda\mu)$ coefficients have been used previously by a number of authors.[7-10] The notation of Ref. 5 is adhered to throughout.

## 2. THE $Z$ AND 9-$(\lambda\mu)$ COEFFICIENTS

The $U$ and 9-$(\lambda\mu)$ coefficients are by definition a straightforward generalization of the corresponding coefficients for SU(2). However, the evaluation of the 9-$(\lambda\mu)$ coefficient as a sum over products of three 6-$(\lambda\mu)$ coefficients is not quite so straightforward. The reason is the nonexistence of a symmetry relation of the Wigner coefficients permitting the interchange of the two $(\lambda\mu)$'s in a product when there exists outer multiplicity in that product.[5] We define the $Z$ coefficients to be the elements of a unitary transformation that effects the following recoupling transformation:




$$= \sum_{(\lambda_{13}\mu_{13})\rho_{13}\rho_{13,2}} Z((\lambda_2\mu_2)(\lambda_1\mu_1)(\lambda\mu)(\lambda_3\mu_3), (\lambda_{12}\mu_{12})\rho_{12}\rho_{12,3}(\lambda_{13}\mu_{13})\rho_{13}\rho_{13,2})$$




(1)

The $Z$ coefficients may be calculated in the same way as the $U$ coefficients,[5] namely as the solution of a set of simultaneous equations obtained by fixing $\epsilon_{13}\Lambda_{13}$ and $\epsilon\Lambda$ at their highest weight values in the relation

$$\begin{aligned}
&\sum_{\rho_{13,2}} \langle(\lambda_{13}\mu_{13})\epsilon_{13}\Lambda_{13}(\lambda_2\mu_2)\epsilon_2\Lambda_2 \,\|\, (\lambda\mu)\epsilon\Lambda\rangle_{\rho_{13,2}} \\
&\quad\times Z((\lambda_2\mu_2)(\lambda_1\mu_1)(\lambda\mu)(\lambda_3\mu_3),\ (\lambda_{12}\mu_{12})\rho_{12}\rho_{12,3} \\
&\qquad\quad (\lambda_{13}\mu_{13})\rho_{13}\rho_{13,2}) \\
&\quad= \sum_{\epsilon_1\Lambda_1\Lambda_3\Lambda_{12}} \langle(\lambda_1\mu_1)\epsilon_1\Lambda_1(\lambda_3\mu_3)\epsilon_3\Lambda_3 \,\|\, (\lambda_{13}\mu_{13})\epsilon_{13}\Lambda_{13}\rangle_{\rho_{13}} \\
&\quad\times\langle(\lambda_1\mu_1)\epsilon_1\Lambda_1(\lambda_2\mu_2)\epsilon_2\Lambda_2 \,\|\, (\lambda_{12}\mu_{12})\epsilon_{12}\Lambda_{12}\rangle_{\rho_{12}} \\
&\quad\times\langle(\lambda_{12}\mu_{12})\epsilon_{12}\Lambda_{12}(\lambda_3\mu_3)\epsilon_3\Lambda_3 \,\|\, (\lambda\mu)\epsilon\Lambda\rangle_{\rho_{12,3}}(-)^{\Lambda_1+\Lambda-\Lambda_{12}-\Lambda_{13}} \\
&\quad\times U(\Lambda_2\Lambda_1\Lambda\Lambda_3, \Lambda_{12}\Lambda_{13}).
\end{aligned} \tag{2}$$

The 9-$(\lambda\mu)$ coefficient may now be expressed as

$$\begin{bmatrix}
(\lambda_1\mu_1) & (\lambda_2\mu_2) & (\lambda_{12}\mu_{12}) & \rho_{12} \\
(\lambda_3\mu_3) & (\lambda_4\mu_4) & (\lambda_{34}\mu_{34}) & \rho_{34} \\
(\lambda_{13}\mu_{13}) & (\lambda_{24}\mu_{24}) & (\lambda\mu) & \rho_{13,24} \\
\rho_{13} & \rho_{24} & \rho_{12,34} &
\end{bmatrix}$$

$$= \sum_{\substack{(\lambda_0\mu_0) \\ \rho_{13,2}\rho_{04}\rho_{12,3}}} U((\lambda_{13}\mu_{13})(\lambda_2\mu_2)(\lambda\mu)(\lambda_4\mu_4), (\lambda_0\mu_0)\rho_{13,2}\rho_{04}$$

$$\times(\lambda_{24}\mu_{24})\rho_{24}\rho_{13,24})$$

[a)]Research at Brookhaven supported by the U.S. Department of Energy under Contract No. EY-76-C-02-0016.

 0022-2488/78/1907-1513$1.00  

$$\times Z((\lambda_2\mu_2)(\lambda_1\mu_1)(\lambda_0\mu_0)(\lambda_3\mu_3),\ (\lambda_{12}\mu_{12})\rho_{12}\rho_{12,3} \times(\lambda_{13}\mu_{13})\rho_{13}\rho_{13,2})$$
$$\times U((\lambda_{12}\mu_{12})(\lambda_3\mu_3)(\lambda\mu)(\lambda_4\mu_4),(\lambda_0\mu_0)\rho_{12,3}\rho_{04}(\lambda_{34}\mu_{34}) \times\rho_{34}\rho_{12,34}). \qquad (3)$$

When $\rho_{12}^{\max}=1$ and $\rho_{13,2}^{\max}=1$ the $Z$ coefficient reduces to a $U$ coefficient with the same arguments times a phase factor $(-)^{\lambda_1+\mu_1+\lambda+\mu+\lambda_{12}+\mu_{12}+\lambda_{13}+\mu_{13}}$ and we recover from Eq. (3) the straightforward generalization of the corresponding expression for SU(2).

## 3. APPLICATIONS

The 9-$(\lambda\mu)$ coefficient typically appears when the matrix element of a coupled tensor operator acting on a two-component system is required. In Eq. (4) the tensor operators $R^{(\lambda_r\mu_r)}$ (1) and $S^{(\lambda_s\mu_s)}$ (2) operate on the first and second parts of the system respectively. The double- and triple-barred matrix elements are reduced matrix elements with respect to O(3) and SU(3),

$$\langle(\lambda_1\mu_1)(\lambda_2\mu_2);(\lambda\mu)\rho\kappa L\,\|\,[R^{(\lambda_r\mu_r)}(1)\times S^{(\lambda_s\mu_s)}(2)]^{(\lambda_t\mu_t)\rho_t\kappa_t L_t}\,\|$$
$$\times\,|(\lambda_1'\mu_1')(\lambda_2'\mu_2');(\lambda'\mu')\rho'\kappa'L'\rangle$$
$$=\sum_{\tilde\rho}\langle(\lambda'\mu')\kappa'L'(\lambda_t\mu_t)\kappa_tL_t\,\|\,(\lambda\mu)\kappa L\rangle_{\tilde\rho}\langle(\lambda_1\mu_1)$$
$$\times(\lambda_2\mu_2);(\lambda\mu)\rho\,|\,\|\,[R^{(\lambda_r\mu_r)}(1)\times S^{(\lambda_s\mu_s)}(2)]^{(\lambda_t\mu_t)\rho_t}\,\|$$
$$\times\,|(\lambda_1'\mu_1')(\lambda_2'\mu_2');(\lambda'\mu')\rho'\rangle_{\tilde\rho}, \qquad (4a)$$

$$\langle(\lambda_1\mu_1)(\lambda_2\mu_2);(\lambda\mu)\rho\,|\,\|\,[R^{(\lambda_r\mu_r)}(1)\times S^{(\lambda_s\mu_s)}(2)]^{(\lambda_t\mu_t)\rho_t}\|$$
$$\times\,|(\lambda_1'\mu_1')(\lambda_2'\mu_2');(\lambda'\mu')\rho'\rangle_{\tilde\rho}$$
$$=\sum_{\rho_1\rho_2}\begin{bmatrix}(\lambda_1'\mu_1') & (\lambda_r\mu_r) & (\lambda_1\mu_1) & \rho_1\\ (\lambda_2'\mu_2') & (\lambda_s\mu_s) & (\lambda_2\mu_2) & \rho_2\\ (\lambda'\mu') & (\lambda_t\mu_t) & (\lambda\mu) & \tilde\rho\\ \rho' & \rho_t & \rho & \end{bmatrix}$$
$$\times\langle(\lambda_1\mu_1)\,|\,\|R^{(\lambda_r\mu_r)}(1)\,\|\,|(\lambda_1'\mu_1')\rangle_{\rho_1}\langle(\lambda_2\mu_2)\,|\,\|S^{(\lambda_s\mu_s)}(2)\|$$
$$\times\,|(\lambda_2'\mu_2')\rangle_{\rho_2}. \qquad (4b)$$

The 9-$(\lambda\mu)$ coefficient reduces to a $U$ coefficient if $(\lambda_r\mu_r)\equiv(00)$ and to a $Z$ coefficient if $(\lambda_s\mu_s)\equiv(00)$. If the operator is a one-body operator which operates in two shells $(\lambda_r\mu_r)$ and $(\lambda_s\mu_s)$ are of the form $(\lambda 0)$ or $(0\mu)$ and the $Z$ reduces to a $U$. An example is Hecht's explicit construction of spurious states of center of mass motion.[7] The 9-$(\lambda\mu)$ coefficients also occur in the calculation of multinucleon spectroscopic amplitudes from SU(3) shell model wavefunctions.[8,9] Here also, at least under the assumption of cluster transfer, the $Z$ reduces to a $U$. For this case Hecht and Braunschweig express the 9-$(\lambda\mu)$ coefficient in terms of three $U$ coefficients in Appendix B of their paper.[9] In a very similar context 9-$(\lambda\mu)$ coefficients occur when shell model wavefunctions are related to cluster model wavefunctions.[10]

## 4. SUMMARY

The results of Draayer and Akiyama[5] have been extended to include those recoupling coefficients required for SU(3) shell model calculations in a multi-shell basis. The results are valid for arbitrary outer multiplicities in the couplings involved.

Computer codes to evaluate the $Z$ and 9-$(\lambda\mu)$ coefficients have been written. They are compatible with the routines of Akiyama and Draayer[6] and are available on request.


[1] K.T. Hecht, Ann. Rev. Nucl. Sci. **23**, 123 (1973).
[2] R.D. Ratna Raju, J.P. Draayer, and K.T. Hecht, Nucl. Phys. A **202**, 433 (1973).
[3] J.P. Draayer, Nucl. Phys. A **216**, 457 (1973).
[4] J.B. French, E.C. Halbert, J.B. McGrory, and S.S.M. Wong, Adv. Nucl. Phys., edited by M. Baranger and E. Vogt, **3**, 193 (1969).
[5] J.P. Draayer and Y. Akiyama, J. Math. Phys. **14**, 1904 (1973).
[6] Y. Akiyama and J.P. Draayer, Comput. Phys. Commun. **5**, 405 (1973).
[7] K.T. Hecht, Nucl. Phys. A **170**, 34 (1971).
[8] N. Anyas-Weiss *et al.*, Phys. Rep. C **12**, 201 (1974).
[9] K.T. Hecht and D. Braunschweig, Nucl. Phys. A **244**, 365 (1975).
[10] K.T. Hecht, Phys. Rev. C **16**, 2401 (1977).

# Solution of the almost-Killing equation and conformal almost-Killing equation in the Kerr spacetime[a)]

Clifford Henry Taubes[b)]

*Department of Physics, Harvard University, Cambridge, Massachusetts 02138*
(Received 16 January 1978)

Four linearly independent classes of vector solutions to the generalized almost-Killing equation in the Kerr spacetime are presented in terms of Teukolsky's radial and angular functions. The vector solutions which are asymptotic to the ten Minkowski-space Killing vectors are given by way of example.

## I. INTRODUCTION

This paper derives solutions of the vector equation

$$\nabla_\alpha \nabla^\alpha \xi_\beta + \nabla_\alpha \nabla_\beta \xi^\alpha - c\nabla_\beta \nabla^\alpha \xi_\alpha = 0 \tag{1.1}$$

(for constant $c$) in the Kerr spacetime, that is with the covariant derivative $\nabla_\alpha$ taken in the Kerr metric. When $c=2$, Eq. (1.1) reduces to Maxwell's equations for a source-free, test electromagnetic field in the Kerr background; $\xi_\alpha$ is the vector potential of this field. As is well known, the equation in this case admits to a remarkable decoupling of components and of variables first found by Teukolsky[1,2] and recently codified with great clarity by Chandrasekhar.[3] A main result of this paper is to show that a large class of solutions of (1.1) for general $c$ can be expressed in terms of essentially the same radial and angular functions that solve the electromagnetic Teukolsky equation and the scalar wave equation.

Equation (1.1) is not just an *ad hoc* generalization of the test Maxwell equation, but rather has definite physical interest of its own. When $c=0$, the equation is called the "almost-Killing equation" (AKE), and when $c=\frac{1}{2}$ it is the "conformal almost-Killing equation" (CAKE); these kinds of equations have been investigated by York[4] and others as a means for generating "natural" vector fields in an asymptotically flat spacetime, in terms of the symmetries of the spacetime at asymptotically flat spatial infinity. In applying this formalism to the specific case of the Kerr metric, as is done here, one hopes to make progress towards elucidating the very special "hidden symmetries" of the Kerr metric which have been noted by so many investigators. For example, one might hope to find a new coordinate system, in which the hidden symmetries become more manifest. Before proceeding, we should indicate how the AKE and CAKE can arise in this context:

The generator of an exact isometry of course satisfies Killing's equation

$$\nabla_\alpha \xi_\beta + \nabla_\beta \xi_\alpha = 0. \tag{1.2}$$

Since this equation has ten independent components, but only four unknowns, it has (in a general spacetime) no solution. In the Kerr spacetime, however, there are two linearly independent solutions, corresponding to the symmetries of time stationarity and axisymmetry. Most of the useful Kerr coordinate systems (e.g., the Boyer—Lindquist[5] system) adopt appropriate coordinates $t$ and $\phi$ such that $\partial/\partial t$ and $\partial/\partial\phi$ are Killing vectors. Since the Kerr metric is asymptotically flat, however, there are *further* Killing vectors of Minkowski space which are *asymptotically* Killing vectors of the Kerr spacetime, in the sense that the equation

$$\nabla_\beta \xi_\alpha + \nabla_\alpha \xi_\beta = O(1/r^2) \tag{1.3}$$

holds (where $r$ is the Boyer—Lindquist radial coordinate). York's AKE is obtained by acting on Eq. (1.2) with an additional $\nabla^\alpha$. The resulting equation has four components for its four unknowns, so it is in general solvable. Any solution of Killing's equation is also a solution of the AKE; and, generally, any asymptotic Killing vector is asymptotic to a solution of the AKE. Therefore, the AKE gives a natural way of extending symmetries (whether approximate or exact) from infinity to the entire spacetime. One program for finding "natural" Kerr coordinates might be to find four mutually commuting, linearly independent almost-Killing vectors of this sort, and then use their integral curves as a coordinate grid.

A generalization which extends the commutator algebra of the Killing vectors, is to also include "conformal Killing vectors" satisfying

$$\nabla_\beta \xi_\alpha + \nabla_\alpha \xi_\beta = \tfrac{1}{2} g_{\alpha\beta} \nabla_\gamma \xi^\gamma, \tag{1.4}$$

since any Killing vector is also a solution of (1.4). The equation derived by acting on (1.4) with $\nabla^\alpha$ (which is the CAKE) gives *a priori* just as natural an extension of symmetries from infinity. Evidently, any constant on the right-hand side could also be viewed as not "unnatural," so we are led to the general equation (1.1), which we now proceed to solve.

Section II consists of preliminaries and the introduction of the functions in terms of which our solutions will be expressed. In Sec. III, four linearly independent solutions are derived. Section IV consists of a presentation of those solutions to the AKE which are asymptotically Killing vectors of the Kerr space-time in the sense of Eq. (1.3).

## II. FORMALISM AND TEUKOLSKY FUNCTIONS

In Boyer—Lindquist coordinates with $c=G=1$ the Kerr metric is

$$ds^2 = -(1-2Mr/\Sigma)\,dt^2 - (4Mar\sin^2\theta/\Sigma)\,dt\,d\Phi + \Sigma/\Delta\,dr^2 + \Sigma\,d\theta^2 + \sin^2\theta(r^2+a^2+2Ma^2r\sin^2\theta/\Sigma)\,d\phi^2. \tag{2.1}$$

a)Supported in part by the National Science Foundation (PHY-76-14852).

b)National Science Foundation Pre-Doctoral Fellow.

Here $M$ is the mass of the black hole, $aM$ is its angular momentum ($0 \le a < M$) oriented in the $\theta = 0$ direction. We have also

$$\Sigma \equiv r^2 + a^2\cos^2\theta, \quad \Delta \equiv r^2 - 2Mr + a^2. \tag{2.2}$$

The derivations will use the standard Newman–Penrose formalism,[6] with Kinnersley's[7] null tetrad. The latter has $[t, r, \theta, \phi]$ components

$$\begin{aligned} l^\mu &= [(r^2+a^2)/\Delta, 1, 0, a/\Delta], \\ n^\mu &= [(r^2+a^2), -\Delta, 0, a]/2\Sigma, \\ m^\mu &= [ia\sin\theta, 0, 1, i/\sin\theta]/\sqrt{2}(r + ia\cos\theta). \end{aligned} \tag{2.3}$$

Given a vector field $\xi^\alpha$, one can form a field tensor

$$F_{\mu\nu} = \nabla_\nu \xi_\mu - \nabla_\mu \xi_\nu \tag{2.4}$$

and then project to get the components

$$\begin{aligned} \Phi_1 &= F_{\mu\nu} l^\mu m^\nu, \\ \Phi_0 &= \tfrac{1}{2} F_{\mu\nu}(l^\mu n^\nu + \bar{m}^\mu m^\nu), \\ \Phi_{-1} &= F_{\mu\nu}\bar{m}^\mu n^\nu. \end{aligned} \tag{2.5}$$

### A. Homogeneous functions: $h$, $Z$, $g$, $X$

We suppress the spheroidal-harmonic indices $n$ and $m$ which should label the separated solution, and write Teukolsky's solutions for $\Phi_1$ and $\Phi_{-1}$ which satisfy Maxwell's equations [Eq. (1.1) with $c = 2$]:

$$\begin{aligned} \Phi_1 &= h_1(r) Z_1(\theta, \phi, t), \\ \Phi_{-1} &= \Delta h_{-1}(r) Z_{-1}(\theta, \phi, t)/2(r - ia\cos\theta)^2. \end{aligned} \tag{2.6}$$

The functions $Z_1(\theta, \phi, t)$ and $Z_{-1}(\theta, \phi, t)$ have the following decompositions:

$$\begin{aligned} Z_1(\theta, \phi, t) &= \exp(-i\omega t)\exp(im\phi) S_1(\theta), \\ Z_{-1}(\theta, \phi, t) &= \exp(-i\omega t)\exp(im\phi) S_{-1}(\theta). \end{aligned} \tag{2.7}$$

The radial functions $h_1$ and $h_{-1}$ and the angular functions $S_1$ and $S_{-1}$ are governed by the equations

$$\frac{1}{\Delta}\frac{d}{dr}\left(\Delta^2 \frac{d}{dr}(h_{\pm 1})\right) + \frac{1}{\Delta}[K^2 \pm 2i(r - M)K]h_{\pm 1} + (\pm 4i\omega r - \lambda_1)h_{\pm 1} = 0, \tag{2.8}$$

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta \frac{d}{d\theta}(S_{\pm 1})\right) + (a^2\omega^2\cos^2\theta \mp 2a\omega\cos\theta)S_{\pm 1} - \frac{1}{\sin^2\theta}(m^2 + 1 \pm 2m\cos\theta)S_{\pm 1} + (\lambda_1 + 2 - a^2\omega^2 + 2a\omega m)S_{\pm 1} = 0, \tag{2.9}$$

where we have defined

$$K = (r^2 + a^2)\omega - am.$$

(In Teukolsky's notation $R_1 = h_1$, $R_{-1} \sim \Delta h_{-1}$.) The separation constant $\lambda_1$ is a characteristic value of the angular equation determined by the regularity conditions on $S_1$ and $S_{-1}$ at $\theta = 0$ and $\theta = \pi$. It follows from these equations that $S_1(\theta) = S_{-1}(\pi - \theta)$.

One may write these equations in a more concise form that is easier to manipulate by defining a pair of radial and angular operators which are closely related to the directional derivatives along the tetrad vectors:

$$\begin{aligned} D &= \frac{\partial}{\partial r} - \frac{ik}{\Delta}, \quad D^+ = \frac{\partial}{\partial r} + \frac{ik}{\Delta}, \\ \eth &= -\left[\frac{\partial}{\partial\theta} - \frac{m}{\sin\theta} + a\omega\sin\theta\right], \\ \eth^+ &= -\left[\frac{\partial}{\partial\theta} + \frac{m}{\sin\theta} - a\omega\sin\theta\right]. \end{aligned} \tag{2.10}$$

Here the notation is very similar to that used by Chandrasekhar[3] and in fact the derivation will be essentially by the same method he used in deriving the vector potential solutions to the vacuum Maxwell equations in the Kerr spacetime. If one considers the directional derivative of a quantity with $t$ and $\phi$ dependence $\exp(-i\omega t)\exp(im\phi)$, then these operators have the following definitions:

$$\begin{aligned} D &= l^\mu \nabla_\mu, \quad D^+ = -(2\Sigma/\Delta) n^\mu \nabla_\mu, \\ \eth &= \sqrt{2}(r + ia\cos\theta) m^\mu \nabla_\mu, \\ \eth^+ &= \sqrt{2}(r - ia\cos\theta)\bar{m}^\mu \nabla_\mu. \end{aligned} \tag{2.11}$$

In terms of these operators, the equations for the radial functions $h_{\pm 1}$ and the angular functions $S_{\pm 1}$ or $Z_{\pm 1}(\theta, \phi, t)$ are

$$\begin{aligned} (DD^+\Delta + 2i\omega r)h_1 &= (\lambda_1 + 2)h_1, \\ (D^+D\Delta - 2i\omega r)h_{-1} &= (\lambda_1 + 2)h_{-1}, \\ [\eth(\eth^+ - \cot\theta) - 2a\omega\cos\theta]Z_1 &= -(\lambda_1 + 2)Z_1, \\ [\eth^+(\eth - \cot\theta) + 2a\omega\cos\theta]Z_{-1} &= -(\lambda_1 + 2)Z_{-1}. \end{aligned} \tag{2.12}$$

Turn now to the scalar wave equation in the Kerr background, which Carter[8] first showed to be separable. The solution can be written as

$$\eta = h_0(r) Z_0(\theta, \phi, t), \tag{2.13}$$

where

$$Z_0(\theta, \phi, t) = \exp(im\phi)\exp(-i\omega t) S_0(\theta). \tag{2.14}$$

The radial function $h_0(r)$ and the angular function $S_0(\theta)$ are governed by the equations

$$\frac{d}{dr}\left(\Delta\frac{d}{dr}(h_0)\right) + \left(\frac{K^2}{\Delta} - \sigma_0^2\right)h_0 = 0, \tag{2.15}$$

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d}{d\theta}(S_0)\right) + \left(a^2\omega^2\cos^2\theta - \frac{m^2}{\sin^2\theta}\right)S_0 + (\sigma_0^2 - a^2\omega^2 + 2a\omega m)S_0 = 0. \tag{2.16}$$

The equations for the radial and angular functions may also be expressed in concise form using the operators defined previously from the tetrad directional derivatives:

$$(D^+\Delta D + 2i\omega r)h_0 = \sigma_0^2 h_0 \tag{2.17}$$

or equivalently

$$(D\Delta D^+ - 2i\omega r)h_0 = \sigma_0^2 h_0, \tag{2.18}$$

$$[(\eth^+ - \cot\theta)\eth - 2a\omega\cos\theta]Z_0 = -\sigma_0^2 Z_0, \tag{2.19}$$

or equivalently

$$[(\eth - \cot\theta)\eth^+ + 2a\omega\cos\theta]Z_0 = -\sigma_0^2 Z_0. \tag{2.20}$$

The separation constant $\sigma_0^2$ is determined by the regularity conditions on $S_0$.

The functions $h_1, h_{-1}, Z_1, Z_{-1}$ satisfy the following ladder relations expressable in terms of these operators[9–11]:

$$\mathcal{D}\mathcal{D}\Delta h_{-1}=\sigma_1^2 h_1,\quad \mathcal{D}^+\mathcal{D}^+\Delta h_1=\sigma_1^2 h_{-1}, \tag{2.21}$$

$$\eth(\eth-\cot\theta)Z_{-1}=\sigma_1^2 Z_1,\quad \eth^+(\eth^+-\cot\theta)Z_1=\sigma_{-1}^2 Z_{-1}. \tag{2.22}$$

Functions which we will need for the AKE and CAKE solutions are the intermediate functions obtained by operating on $h_{\pm1}$ and $Z_{\pm1}$ only once. We define

$$g_1=\frac{1}{\sigma_1}\mathcal{D}^+\Delta h_1,\quad g_{-1}=\frac{1}{\sigma_1}\mathcal{D}\Delta h_{-1}. \tag{2.23}$$

It follows from the ladder relations, Eq. (2.21) that

$$\mathcal{D}^+ g_1=\sigma_1 h_{-1},\quad \mathcal{D}g_{-1}=\sigma_1 h_1, \tag{2.24}$$

$$\mathcal{D}^+\Delta\mathcal{D}g_{-1}=\sigma_1^2 g_1,\quad \mathcal{D}\Delta\mathcal{D}^+ g_1=\sigma_1^2 g_{-1}, \tag{2.25}$$

and that $g_1$ satisfies the following inhomogeneous equation,

$$\mathcal{D}^+\Delta\mathcal{D}g_1+(2i\omega r-\lambda_1-2)g_1=-2i\omega h_1\Delta/\sigma_1, \tag{2.26}$$

where we have defined the constant

$$\sigma_1^2=[(\lambda_1+2)^2-4\omega^2a^2+4am\omega]^{1/2}. \tag{2.27}$$

The complex conjugate equation is satisfied by $g_{-1}$. One can make the analogous definitions for the angular functions and easily derive the analogous equations, viz.,

$$X_1=-\frac{1}{\sigma_1}(\eth^+-\cot\theta)Z_1,\quad X_{-1}=\frac{1}{\sigma_1}(\eth-\cot\theta)Z_{-1}. \tag{2.28}$$

Then from Eq. (2.22) one gets

$$\eth X_{-1}=\sigma_1 Z_1,\quad \eth^+X_1=-\sigma_1 Z_{-1}, \tag{2.29}$$

and

$$(\eth^+-\cot\theta)\eth X_{-1}=-\sigma_1^2 X_1,\quad (\eth-\cot\theta)\eth^+X_1=-\sigma_1^2 X_{-1}, \tag{2.30}$$

and

$$X_{-1}(\pi-\theta)=X_1(\theta).$$

### B. Case of zero frequency: Additional functions $f$ and $Q$

All these new functions are a bit confusing. To see what is going on, we can consider the limit of time-independent solutions, where $\omega=0$. (This is physically the most interesting case since one will generally want time-independent almost-Killing vectors.) When $\omega=0$, the functions $Z_{\pm1}$ are just the spin weight $\pm1$ spherical harmonics. They are generated from the ordinary scalar spherical harmonics, denoted $Y_0^{nm}$ (which in the limit that $\omega=0$, $Z_0$ becomes) by the following expressions using just the operators $\eth$ and $\eth^+$ that were defined in Eq. (2.10) (to avoid confusion, the zero frequency $Z_{\pm1}$ will be denoted $Y_{\pm1}^{nm}$):

$$Y_1^{nm}=[n(n+1)]^{-1/2}\eth Y_0^{nm},\quad Y_{-1}^{nm}=-[n(n+1)]^{-1/2}\eth^+Y_0^{nm}. \tag{2.31}$$

When $\omega=0$, the constants $\sigma_0^2$, $\lambda_1^2$, and $\sigma_1^2$ are identical and are equal to, for the $(n,m)$ harmonic, $n(n+1)$. The operator $\eth$ acts as a spin weight raising operator, while $\eth^+$ acts as a spin weight lowering operator, $Y_{\pm1}^{nm}$, $Y_0^{nm}$ also have the following relations:

$$[n(n+1)]^{-1/2}(\eth^+-\cot\theta)Y_1^{nm}=-Y_0^{nm},$$

$$[n(n+1)]^{-1/2}(\eth-\cot\theta)Y_{-1}^{nm}=Y_0^{nm}. \tag{2.32}$$

From these expressions, it becomes clear that in the time independent case, the functions $X_{\pm1}$ are both equal to $Y_0^{nm}$. For nonzero $\omega$, this is not the case. Similarly, in the time dependent case Eqs. (2.25) and (2.17) show that both $g_1$ and $g_{-1}$ are equal to $h_0$. For $\omega=0$, the operators $\mathcal{D}$ and $\mathcal{D}^+$ act as respectively spin weight raising and lowering operators on the radial functions. For $\omega=0$, the function $h_0(r)$ is essentially a polynomial in $r$.[12] One fundamental solution is

$$h_0^{nm}=(r-r_-)^{iam/\delta}(r-r_+)^{-iam/\delta}$$

$${}_2F_1(n+1;\,-n;\,1+2iam\delta;\,(r-r_-)/\delta),$$

where

$$\delta=2(M^2-a^2)^{1/2},\quad r_\pm=M\pm(M^2-a^2)^{1/2}$$

and ${}_2F_1$ is a hypergeometric function, a polynomial in $r$ of degree $n$ since $n$ is an integer. The other fundamental solution is the complex conjugate. Therefore, $h_{\pm1}$ are essentially, for $\omega=0$, polynomials in $r$ of degree $n-1$ since for $\omega=0$

$$h_\pm^{nm}=[n(n+1)]^{-1/2}\left(\frac{d}{dr}\mp\frac{iam}{\Delta}\right)h_0^{mn}. \tag{2.33}$$

When $\omega$ is not equal to zero, the function obtained when operating on $h_0(r)$ by $\mathcal{D}$ is no longer equal to the spin-1 radial function $h_1(r)$ as in Eq. (2.33) for $\omega=0$. The function generated in this way will be denoted

$$f_1=\frac{1}{\sigma_0}\mathcal{D}h_0(r). \tag{2.34}$$

Likewise, we define

$$f_{-1}=\frac{1}{\sigma_0}\mathcal{D}^+h_0(r). \tag{2.35}$$

From Eq. (2.17) one can easily derive the following ladderlike relations which $f_{\pm1}$ and $h_0$ satisfy:

$$\mathcal{D}^+\Delta f_1=\sigma_0 h_0-2i\omega r h_0/\sigma_0,$$

$$\mathcal{D}\Delta f_{-1}=\sigma_0 h_0+2i\omega r h_0/\sigma_0. \tag{2.36}$$

These two new functions satisfy the inhomogeneous equations

$$\mathcal{D}\mathcal{D}^+\Delta f_1+2i\omega r f_1-\sigma_0^2 f_1=-2i\omega h_0/\sigma_0,$$

$$\mathcal{D}^+\mathcal{D}\,\Delta f_{-1}-2i\omega r f_{-1}-\sigma_0^2 f_{-1}=2i\omega h_0/\sigma_0. \tag{2.37}$$

The analogous angular functions can be defined from the scalar angular function $Z_0$ when $\omega$ is not zero by using the operators $\eth$ and $\eth^+$. We define the functions $Q_1$ and $Q_{-1}$ which in the time independent limit become $Y_1$ and $Y_{-1}$ respectively through the relations

$$Q_1=\frac{1}{\sigma_0}\eth Z_0,\quad Q_{-1}=-\frac{1}{\sigma_0}\eth^+Z_0. \tag{2.38}$$

These functions satisfy inhomogeneous angular equations derived from the equations for $Z_0$, Eqs. (2.19) and (2.20):

$$[\eth(\eth^* - \cot\theta) - 2a\omega\cos\theta + \sigma_0^2]Q_1 = 2a\omega\sin\theta Z_0/\sigma_0,$$
$$[\eth^*(\eth - \cot\theta) + 2a\omega\cos\theta + \sigma_0^2]Q_{-1} = 2a\omega\sin\theta Z_0/\sigma_0. \quad (2.39)$$

The function $Z_0$ can be obtained from $Q_1$ (or $Q_{-1}$) by the operation of the appropriate lowering (raising) operator in exact analogy to the relations between the radial functions, $h_0, f_1, f_{-1}$ in Eq. (2.36):

$$(\eth^* - \cot\theta)Q_1 = -\sigma_0 Z_0 + 2a\omega\cos\theta Z_0/\sigma_0,$$
$$(\eth - \cot\theta)Q_{-1} = \sigma_0 Z_0 + 2a\omega\cos\theta Z_0/\sigma_0. \quad (2.40)$$

The constants $\sigma_0^2$ and $\lambda_1 + 2$ differ in lowest order in $\omega$ by a term proportional to $a\omega$.[13] The equations for $h_1$ and $f_1$ are similar, the difference being the inhomogeneous term in the equation for $f_1$ with its explicit $\omega$ dependence and the substitution of $\sigma_0^2$ for $\lambda_1 + 2$ in the equation for $f_1$ which also yields a difference which is to lowest order proportional to $a\omega$. Thus to zeroth order in $\omega$, $f_1$, and $h_1$ are identical, they differ in terms proportional to $\omega$. This difference is even present in the Minkowski space analogs of the functions $f_1$ $(=\mathcal{D} h_0/\sigma_0)$ and $h_1$ and is due to the inhomogeneous term in the equation for $f_1$ which remains even when $a = M = 0$.

### C. Inhomogeneous functions: $j(r)$, $w(r)$, $s(r)$, $A(\theta,\phi,t)$, $T(\theta,\phi,t)$, $B(\theta,\phi,t)$

Below, we will find four classes of solutions to our master almost-Killing equation (1.1). Three of these classes are expressible solely in terms of the functions defined so far, that is, explicitly in terms of solutions to the scalar wave equation and Teukolsky's equation. For the fourth class it seems, unfortunately, necessary to define certain additional functions, both radial and the analogous angular ones, which are defined as solutions of the inhomogeneous radial equations whose homogeneous solutions are the scalar wave equation radial function $h_0$ and the analogous inhomogeneous angular equation whose homogeneous solutions are the scalar wave equation angular function $Z_0$. It may be possible that a different technique to derive the solutions will obviate the necessity to use functions defined as solutions to inhomogeneous second order differential equations which are hard to work with. It is also possible that these new functions are expressible simply in terms of the functions already defined. This is not known.

The first pair of functions are the radial function which shall be called $j(r)$ which satisfies the inhomogeneous equation

$$(\mathcal{D}^*\Delta\mathcal{D} + 2i\omega r - \sigma_0^2)j(r) = r^2 h_0(r). \quad (2.41)$$

The analogous angular function will be denoted $A(\theta,\phi,t)$ and it satisfies the inhomogeneous equation

$$[(\eth^* - \cot\theta)\eth - 2a\omega\cos\theta + \sigma_0^2]A = \cos^2\theta Z_0. \quad (2.42)$$

The second pair of functions are the radial function which shall be called $w(r)$ which satisfies the inhomogeneous equation

$$(\mathcal{D}^*\Delta\mathcal{D} + 2i\omega r - \sigma_0^2)w(r) = h_0(r). \quad (2.43)$$

The analogous angular function, $T(\theta,\phi,t)$ is a solution of the equation

$$[(\eth^* - \cot\theta)\eth - 2a\omega\cos\theta + \sigma_0^2]T = Z_0. \quad (2.44)$$

Finally, it will be necessary to define the following pair of radial and angular functions for the case where $\omega = 0$. The radial function will be denoted $s(r)$ and the angular function $B(\theta,\phi)$. Their defining equations are, respectively

$$(\mathcal{D}^*\Delta\mathcal{D} - \sigma_0^2)S = r\Delta(h_0 + h_{-1})/\sigma_0 \quad (2.45)$$

and

$$[(\eth^* - \cot\theta)\eth + \sigma_0^2]B = \sin\theta\cos\theta(Y_{-1} - Y_1)/\sigma_0. \quad (2.46)$$

The three pair of functions just defined plus the functions defined previously allow the specification of our solutions to the almost-Killing equation in the Kerr spacetime, Eq. (1.1).

## III. SOLUTION OF THE EQUATIONS

Since the vacuum Kerr metric is Ricci-flat our equation (1.1) admits a commutation of covariant derivatives which makes it equivalent to

$$\nabla_\beta \nabla^\beta \xi_\alpha - \nabla^\beta \nabla_\alpha \xi_\beta = \epsilon \nabla_\alpha \nabla^\beta \xi_\beta, \quad (3.1)$$

where $\epsilon = c - 2$.

Defining

$$F_{\alpha\beta} = \nabla_\beta \xi_\alpha - \nabla_\alpha \xi_\beta, \quad J_\alpha = \epsilon \nabla_\alpha \nabla^\beta \xi_\beta, \quad (3.2)$$

we can write, suggestively, (3.1) as

$$\nabla^\beta F_{\alpha\beta} = J_\alpha. \quad (3.3)$$

The divergence of the left-hand side of this equation vanishes which implies that the "source" $J_\alpha$ is divergence free. Since the "source" is actually the gradient of the divergence of the almost-Killing vector itself, the divergence of the almost-Killing vector is a solution of the four-dimensional scalar wave equation, whose solutions are functions already defined [Eq. (2.13)]. Thus, although the "source" in the above "field equation" is explicitly a *functional* of the solution of the equation, it is also a known *function* of Boyer–Lindquist coordinates. It is this fact which allows us to find the solutions in terms of the Teukolsky functions. It is convenient in deriving what follows to express Eq. (3.3) in terms of the operators $\mathcal{D}$, $\mathcal{D}^*$, $\eth$, and $\eth^*$. Equation (3.3) is equivalent to the following set of four equations:

$$(r - ia\cos\theta)^{-2}\mathcal{D}((r - ia\cos\theta)^2\Phi_0) + (2)^{-1/2}(r - ia\cos\theta)^{-1}$$
$$\times[\eth^* - \cot\theta + ia\sin\theta(r - ia\cos\theta)^{-1}]\Phi_1$$
$$= -J_l/2, \quad (3.3a)$$

$$-(2)^{-1/2}[\Sigma(r - ia\cos\theta)]^{-1}[\eth(r - ia\cos\theta)^2\Phi_0]$$
$$+ (2\Sigma)^{-1}[\mathcal{D}^*\Delta - \Delta(r - ia\cos\theta)^{-1}]\Phi_1$$
$$= J_m/2, \quad (3.3b)$$

$$(r - ia\cos\theta)^{-1}\mathcal{D}((r - ia\cos\theta)\Phi_{-1})$$
$$+ (2)^{-1/2}(r - ia\cos\theta)^{-3}[\eth^*(r - ia\cos\theta)^2\phi_0]$$
$$= J_{\bar{m}}/2, \quad (3.3c)$$

$$\Delta(2\Sigma)^{-1}(r-ia\cos\theta)^{-2}\mathcal{D}^{+}((r-ia\cos\theta)^{2}\Phi_{0})-(2)^{-1/2}$$
$$\times(r+ia\cos\theta)^{-1}[\bar{\eth}-\cot\theta-ia\sin\theta(r-ia\cos\theta)^{-1}]\Phi_{-1}$$
$$=-J_{n}/2. \tag{3.3d}$$

Here we have defined the scalars $J_l, J_n, J_m$, and $J_{\bar{m}}$ by

$$J_{\alpha}=J_{n}l_{\alpha}+J_{l}n_{\alpha}+J_{\bar{m}}m_{\alpha}+J_{m}\bar{m}_{\alpha}. \tag{3.3e}$$

By analogy with the vector solutions of Maxwell's equation in flat-space (see, e.g., Morse and Feshbach[14]), we can distinguish the following classes of solutions for $\xi_\alpha$:

1. If $\xi_\alpha$ is divergenceless *and* if it is the gradient of a scalar, $\xi_\alpha=\nabla_\alpha\eta$, then $\eta$ satisfies the scalar wave equation $\nabla^\alpha\nabla_\alpha\eta=0$.

2. and 3. If $\xi_\alpha$ is divergenceless, but *not* the gradient of a scalar, then the almost-Killing vector is the vector-potential solution to Maxwell's equations in the Lorentz gauge. Two linearly independent classes of solutions of this type will be found below. The first one may be thought of as generating "electric"-type fields, the archtype being the timelike Killing vector $\partial/\partial t$. The second class may be thought of as generating "magnetic"-type fields, the archetype being the azimuthal Killing vector $\partial/\partial\phi$.

4. This class contains all solutions $\xi_\alpha$ with nonzero divergence.

### A. Solutions in class 1

It is most convenient to consider the projection of $\xi$, along the tetrad legs $\mathbf{l}$, $\mathbf{n}$, $\mathbf{m}$, and $\bar{\mathbf{m}}$, and write

$$\xi^{\alpha}=\xi_{n}l^{\alpha}+\xi_{l}n^{\alpha}+\xi_{\bar{m}}m^{\alpha}+\xi_{m}\bar{m}^{\alpha}. \tag{3.4}$$

Since the general solution to the scalar wave equation is given by $\eta=h_0(r)Z_0(\theta,\phi,t)$ [Eq. (2.13) above], it is a straightforward matter to compute the gradient and obtain the explicit components of $\xi$:

$$\begin{aligned}
\xi_{n}&=\Delta Z_{0}(\mathcal{D}^{+}h_{0})/2\Sigma=\Delta\sigma_{0}Z_{0}f_{-1}/2\Sigma,\\
\xi_{l}&=-Z_{0}\mathcal{D}h_{0}=-\sigma_{0}Z_{0}f_{1},\\
\xi_{m}&=-(2)^{-1/2}h_{0}(\eth Z_{0})/(r+ia\cos\theta)\\
&=-\sigma_{0}h_{0}Q_{1}/[(2)^{1/2}(r+ia\cos\theta)]\\
\xi_{\bar{m}}&=-(2)^{-1/2}h_{0}(\bar{\eth}Z_{0})/(r-ia\cos\theta)\\
&=\sigma_{0}h_{0}Q_{-1}/[(2)^{1/2}(r-ia\cos\theta)].
\end{aligned} \tag{3.5}$$

### B. Solutions in class 2

We are looking for vector-potential solutions to Maxwell's equations, subject to the additional restriction that they be divergenceless. Except for this last restriction, we can follow exactly the method of Chandrasekhar, which expresses the vector potential solution in terms of derivative operators acting on Teukolsky functions. Then, using gauge invariance, we can add the gradient of a scalar function to make the solution divergence free.

The solution to the vacuum Maxwell's equations for $\Phi_1$ and $\Phi_{-1}$ found by Teukolsky are given in Eq. (2.6). Using these expressions in Maxwell's equations, $\nabla^\beta F_{\alpha\beta}=0$, we find that they are satisfied identically if $\Phi_0$ is given by

$$\Phi_{0}=(2)^{-3/2}(r-ia\cos\theta)^{-2}[X_{1}(rg_{-1}-\Delta h_{-1}/\sigma_{1})$$
$$+X_{-1}(rg_{1}-\Delta h_{1}/\sigma_{1})-iag_{1}(\cos\theta X_{-1}-\sin\theta Z_{-1}/\sigma_{1})$$
$$-iag_{-1}(\cos\theta X_{1}+\sin\theta Z_{1}/\sigma_{1})]. \tag{3.6}$$

If both $\Phi_1$ and $\Phi_{-1}$ are zero, one can still have nonzero $\Phi_0$: This is the static Coulomb solution of Maxwell's equations, and represents the field about a Kerr black hole when an infinitesimal amount of charge is added.[15] Examination of the vacuum Maxwell's equations shows that this Coulomb field is given by

$$\Phi_{0}=M/(r-ia\cos\theta)^{2},\quad \Phi_{1}=\Phi_{-1}=0. \tag{3.7}$$

One can verify by direct substitution that the vector potential (and almost-Killing vector) corresponding to the solution (3.7) is just the Killing vector $\partial/\partial t$,

$$\left(\frac{\partial}{\partial t}\right)^{\alpha}=(\Delta/2\Sigma)l^{\alpha}+n^{\alpha}+ia\sin\theta[m^{\alpha}/(r-ia\cos\theta)$$
$$-\bar{m}^{\alpha}/(r+ia\cos\theta)](2)^{-1/2}. \tag{3.8}$$

For the case $\Phi_1\neq 0$, we turn to the equations which give the field quantities $\Phi_1$, $\Phi_{-1}$, and $\Phi_0$ in terms of the scalars $\xi_n$, $\xi_l$, $\xi_m$, and $\xi_{\bar{m}}$:

$$\Phi_{1}=[\eth\xi_{l}-(2)^{1/2}\mathcal{D}((r+ia\cos\theta)\xi_{m})]/[(2)^{1/2}(r+ia\cos\theta)], \tag{3.9}$$

$$\Phi_{-1}=-[\bar{\eth}^{+}(\Sigma\xi_{n})(2)^{1/2}+\Delta\mathcal{D}^{+}((r-ia\cos\theta)\xi_{\bar{m}}]/$$
$$[2\Sigma(r-ia\cos\theta)], \tag{3.10}$$

$$\Phi_{0}=(2\Sigma)^{-1}\{\mathcal{D}(\Sigma\xi_{n})+\mathcal{D}^{+}(\Delta\xi_{l})/2+(\bar{\eth}^{+}-\cot\theta)$$
$$\times[(r+ia\cos\theta)\xi_{m}]/(2)^{1/2}$$
$$-(\bar{\eth}-\cot\theta)[(r-ia\cos\theta)\xi_{\bar{m}}]/(2)^{1/2}$$
$$+(r+ia\cos\theta)[-2\xi_{n}-\Delta\xi_{l}/\Sigma+(2)^{1/2}ia\sin\theta\xi_{m}/$$
$$(r-ia\cos\theta)-(2)^{1/2}ia\sin\theta\xi_{\bar{m}}/(r+ia\cos\theta)]\}. \tag{3.11}$$

The equation for $\Phi_1$ involves only $\xi_l$ and $\xi_m$ while the equation for $\Phi_{-1}$ involves $\xi_n$ and $\xi_{\bar{m}}$. Following Chandrasekhar, we use the relations of Eqs. (2.23), (2.24), (2.28), and (2.29) to determine the solutions $\xi_n$, $\xi_l$, $\xi_m$, and $\xi_{\bar{m}}$ from the equations for $\Phi_1$ and $\Phi_{-1}$ and then use the equation for $\Phi_0$ to insure that the expressions for $\xi_n$, $\xi_l$, $\xi_m$, and $\xi_{\bar{m}}$ determined in this way are consistent. The solution of Eq. (3.9) with $\Phi_1$ given in Eq. (2.6) which will give our class 2 almost-Killing vectors is

$$\begin{aligned}
\xi'_{l}&=(2)^{1/2}rh_{1}X_{-1}/\sigma_{1},\\
\xi'_{m}&=-ia\cos\theta g_{-1}Z_{1}/[(r+ia\cos\theta)\sigma_{1}].
\end{aligned} \tag{3.12}$$

The class 2 solutions of Eq. (3.10) with $\Phi_{-1}$ given in Eq. (2.6) are

$$\begin{aligned}
\xi'_{n}&=(2)^{-1/2}r\Delta h_{-1}X_{1}/(\Sigma\sigma_{1}),\\
\xi'_{\bar{m}}&=-ia\cos\theta g_{1}Z_{-1}/[(r-ia\cos\theta)\sigma_{1}].
\end{aligned} \tag{3.13}$$

That these $\xi'$'s also satisfy Eq. (3.11) is verified by direct substitution. The primes in Eqs. (3.12) and (3.13) signify that the solutions are not yet divergence-free. In fact, their divergence can be calculated to be

$$\chi\equiv\nabla^{\alpha}\xi'_{\alpha}=(2)^{1/2}(g_{-1}X_{1}-g_{1}X_{-1})/(r+ia\cos\theta). \tag{3.14}$$

In the time independent case this divergence is zero.

For nonzero $\omega$, even in the flat space analog of this solution, this is not so. To the vector, $\xi'$ one must add the gradient of a scalar chosen to make the sum divergence-free. To wit, one writes

$$\xi^\alpha = \xi^{\alpha\,\prime} + \nabla^\alpha \eta. \tag{3.15}$$

Equations (3.14) and (3.15) imply an equation for $\eta$,

$$\nabla_\alpha \nabla^\alpha \eta = -(2)^{1/2}(g_{-1}X_1 - g_1X_{-1})/(r + ia\cos\theta). \tag{3.16}$$

Remarkably, it is possible to use the ladder relations to guess the solution of (3.16), viz.,

$$\eta = -(2)^{-1/2} i(g_1X_{-1} - g_{-1}X_1)/\omega. \tag{3.17}$$

That this is correct is most easily checked by using the scalar wave equation written in terms of the operators $D$, $D^+$, $\eth$, and $\eth^+$:

$$\nabla_\alpha \nabla^\alpha \eta = -\Sigma^{-1}(D^+\Delta D + 2i\omega r + (\eth^+ - \cot\theta)\eth - 2a\omega\cos\theta)\eta. \tag{3.18}$$

So, the complete class 2 solution is

$$\begin{aligned}
\xi_n &= \Delta(2\Sigma)^{-1}\{(2)^{1/2} rh_{-1}X_1/\sigma_1 - i(2)^{-1/2}\omega^{-1} \\
&\quad\times(\sigma_1 h_{-1}X_{-1} + \sigma_1 h_1 X_1 + 2i[(r^2+a^2)\omega - am]g_{-1}X_1/\Delta)\}, \\
\xi_l &= (2)^{1/2} rh_1X_{-1}/\sigma_1 + i(2)^{-1/2}\omega^{-1}\{\sigma_1 h_{-1}X_{-1} + \sigma_1 h_1 X_1 \\
&\quad - 2i[(r^2+a^2)\omega - am]g_1X_{-1}/\Delta\} \\
\xi_m &= \{-ia\cos\theta g_{-1}Z_1/\sigma_1 - (2i\omega)^{-1}[\sigma_1 g_1 Z_1 \\
&\quad - \sigma_1 g_{-1}Z_{-1} + 2(m/\sin\theta - a\omega\sin\theta)g_{-1}X_1]\}/(r + ia\cos\theta), \\
\xi_{\bar m} &= \{-ia\cos\theta g_1 Z_{-1}/\sigma_1 - (2i\omega)^{-1}[\sigma_1 g_1 Z_1 - \sigma_1 g_{-1}Z_{-1} \\
&\quad - 2(m/\sin\theta - a\omega\sin\theta)g_1X_{-1}]\}/(r - ia\cos\theta).
\end{aligned} \tag{3.19}$$

### C. Solutions of class 2 when $\omega = 0$

A note of explanation must be added concerning the explicit factor of $(\omega)^{-1}$ in the expression for the scalar $\eta$, Eq. (3.17). When $\omega$ is zero $g_1$ and $g_{-1}$ are equal to $h_0$ and $X_1$ and $X_{-1}$ are equal to $Z_0$. For nonzero $\omega$, $g_1$ and $g_{-1}$ will be equal to $h_0$ plus a term proportional to $\omega$; one can write as definitions

$$g_1 = h_0 + \omega k_1, \quad g_{-1} = h_0 + \omega k_{-1}. \tag{3.20}$$

By substituting these expressions for $g_1$ and $g_{-1}$ in Eq. (2.26) and its complex conjugate respectively and noting that

$$D\Delta D^+ = D^+\Delta D + 4i\omega r, \tag{3.21}$$

we derive the differential equations satisfied by $k_1$ and $k_{-1}$:

$$\begin{aligned}
(D^+\Delta D + 2i\omega r - \lambda_1 - 2)k_1 &= 2i\Delta h_1/\sigma_1 + (\lambda_1 + 2 - \sigma_0^2)h_0/\omega, \\
(D^+\Delta D + 2i\omega r - \lambda_1 - 2)k_{-1} &= -2i\Delta h_{-1}/\sigma_1 + (\lambda_1 + 2 - \sigma_0^2)h_0/\omega.
\end{aligned} \tag{3.22}$$

Adding these two equations gives

$$\begin{aligned}
&(D^+\Delta D + 2i\omega r - \lambda_1 - 2)(k_1 + k_{-1}) \\
&\quad = -2i\Delta(h_1 - h_{-1})/\sigma_1 + 2(\lambda_1 + 2 - \sigma_0^2)h_0/\omega.
\end{aligned} \tag{3.23}$$

Breuer[13] showed that the eigenvalues ${}_1\lambda + 2$ and $\sigma_0^2$ satisfy the relation

$${}_1\lambda + 2 - \sigma_0^2 = -2a\omega m/\sigma_0^2 + O(a^2\omega^2). \tag{3.24}$$

Then the right-hand side of Eq. (3.23) may be expanded as a power series in $\omega$. Using Eqs. (2.33) and (2.27) we get

$$-2i\Delta\left(\frac{\partial}{\partial r} + \frac{iam}{\Delta} - \frac{\partial}{\partial r} + \frac{iam}{\Delta}\right)\frac{h_0}{\sigma_0^2} - \frac{4amh_0}{\sigma_0^2} + O(\omega). \tag{3.25}$$

The terms in this expression zeroth order in $\omega$ cancel, allowing one to write Eq. (3.23) as

$$(D^+\Delta D + 2i\omega r - \lambda_1 - 2)(k_1 + k_{-1}) = O(\omega). \tag{3.26}$$

This means that we can redefine $k_1$ and $k_{-1}$ by the expression

$$k_1 + k_{-1} = \alpha_1 h_0 + \omega\bar{k}, \tag{3.27}$$

where $\alpha_1$ is a constant independent of $\omega$.

In exact analogy we can write as definitions

$$X_1 = Z_0 + \omega U_1, \quad X_{-1} = Z_0 + \omega U_{-1} \tag{3.28}$$

and show using Eqs. (2.28), (2.29), (2.31), (2.27), and (3.24), that we can redefine $U_1$ and $U_{-1}$ by

$$U_1 + U_{-1} = \alpha_2 Z_0 + \omega\bar{U}, \tag{3.29}$$

where $\alpha_2$ is also a constant independent of $\omega$. In terms of these new functions we may write Eq. (3.17) as

$$\begin{aligned}
\eta &= h_0 Z_0[-i(2)^{1/2}/\omega - i(\alpha_1 + \alpha_2)(2)^{-1/2}] \\
&\quad - \omega i(\bar{k}Z_0 + \bar{U}h_0 + k_1U_{-1} + k_{-1}U_1)/2.
\end{aligned} \tag{3.30}$$

When the gradient of $\eta$ is taken the first term in Eq. (3.30) generates the class 1 solution, Eq. (3.5) so it can be ignored. The relevant portion of the scalar $\eta$ does indeed become zero when $\omega$ is zero.

A final note about the solutions given by Eq. (3.19): If the flat-space analog of these solutions are considered with $\omega = 0$, then the functions $h_0^{nm}$ and $h_{\pm1}^{nm}$ are

$$h_0 = r^n, \quad h_{\pm1} = nr^{n-1}[n(n+1)]^{-1/2}. \tag{3.31}$$

The class 2 solutions are

$$\xi = \frac{(2)^{1/2} r^n Y_0^{nm}}{(n+1)}\frac{\partial}{\partial t} \tag{3.32}$$

which is the vector potential which generates the $n$-pole static electric fields.

### D. Solutions in class (3)

For a given solution to Eq. (3.3) with $J_\alpha = 0$ for the field tensor $F_{\alpha\beta}$, a second linearly independent solution can be generated by taking the dual of $F_{\alpha\beta}$. In terms of the Newman–Penrose scalars, $\Phi_1$, $\Phi_{-1}$, and $\Phi_0$, the dual solutions $\Phi_1^d$, $\Phi_{-1}^d$, $\Phi_0^d$ are gotten from the originals by multiplication by $-i$.[16] If we find a vector solution to Eqs. (3.9) and (3.10) with $\Phi_1, \Phi_{-1}$ given by Eq. (2.6) and $\Phi_0$ by Eq. (3.6), a second linearly independent solution must exist that generates the dual set. The tensor $F_{\alpha\beta}$ is determined by six functions: $\Phi_1, \Phi_{-1}, \Phi_0$, and $\bar\Phi_1, \bar\Phi_{-1}, \bar\Phi_0$. The barred scalars are defined by switching $m^\alpha$ and $\bar m^\alpha$ in Eq. (2.5), not by taking the complex conjugate of the unbarred scalars. The two procedures are the same only for a *real* $F_{\alpha\beta}$ and $\xi_\alpha$. The duals of the barred scalars are gotten from the unbarred ones by multiplication by $+i$. Suppose $\xi^\alpha$ is a solution to Eq. (1.1) from our class 2, then $\xi^\alpha$ generates a set of scalars, $\Phi_1, \Phi_{-1}, \Phi_0$, $\bar\Phi_1$, $\bar\Phi_{-1}$, and $\bar\Phi_0$, and $-i\xi^\alpha$ generates $-i\Phi_1$, $-i\Phi_{-1}$, $-i\Phi_0$, $-i\bar{\Phi}_1$, $-i\bar{\Phi}_{-1}$, and $-i\bar{\Phi}_0$. The linearly independent vector solution to Eqs. (3.9) and (3.10) which generates the correct dual solutions to $\Phi_1$ and $\Phi_{-1}$ given by Eq. (2.6) and $\Phi_0$ from Eq. (3.6) was first derived by Chandrasekhar[3] in the context of solving for the vector potential solutions to the vacuum Maxwell equations. In our notation, the class 3 solutions are

$$\begin{aligned}
\xi_n &= ah_{-1}\Delta(\cos\theta X_1 + \sin\theta Z_1/\sigma_1)/[(2)^{1/2}\Sigma\sigma_1],\\
\xi_l &= a(2)^{1/2}h_1(\cos\theta X_{-1} - \sin\theta Z_{-1}/\sigma_1)/\sigma_1,\\
\xi_m &= +iZ_1(rg_{-1} - \Delta h_{-1}/\sigma_1)/[(r+ia\cos\theta)\sigma_1],\\
\xi_{\bar{m}} &= +iZ_{-1}(rg_1 - \Delta h_1/\sigma_1)/[(r-ia\cos\theta)\sigma_1].
\end{aligned} \tag{3.33}$$

These also satisfy Eq. (3.11) for $\Phi_0$. Furthermore, the $\xi$'s of Eq. (3.33) are *already* divergenceless, so they are almost-Killing solutions.

In his paper,[3] Chandrasekhar derives the most general solution to our Eqs. (3.9) and (3.10) with $J_\alpha = 0$. His general solution is expressed as a sum of two parts, the first being our class 3 solution of Eq. (3.33) and the second is expressed in terms of two functions he calls $P_+$ and $P_-$ which are constrained by one equation. If two solutions to Eqs. (3.9) and (3.10) differ by the gradient of a scalar then Chandrasekhar's functions $P_+$ and $P_-$ will be equal. For the solutions of class 2 and class 3 this is not the case.

In the time independent case $X_1$ and $X_{-1}$ are equal to the $(n,m)$ spherical harmonic $Y_0^{nm}(\theta,\phi)$ and $Z_{+1}$ and $Z_{-1}$ are equal to the $(n,m)$ spin weight $+1$ and $-1$ spherical harmonics $Y_1^{nm}$ and $Y_{-1}^{nm}$, respectively. The class 3 solution containing only the $(1,0)$ angular functions is the Killing vector $(3/32\pi)^{1/2}(\partial/\partial\phi + 2a\partial/\partial t)$. The Minkowski space, time independent solutions of Eq. (3.33) containing the $(n,m)$ spherical harmonic generate the static $n$-pole magnetic fields.

### E. Solutions in class 4: Formulas for $\Phi_1$ and $\Phi_{-1}$

When the divergence of the almost-Killing vector is not required to be zero then this divergence must be a solution to the scalar wave equation,

$$\chi \equiv \nabla^\alpha \xi_\alpha = h_0(r)\, Z_0(\theta,\phi,t). \tag{3.34}$$

Then the components of $J_\alpha$ [the right-hand side of Eq. (3.3)] are:

$$\begin{aligned}
J_n &= \epsilon\Delta\sigma_0 Z_0 f_{-1}/2\Sigma,\\
J_l &= -\epsilon\sigma_0 Z_0 f_1,\\
J_m &= -\epsilon\sigma_0 h_0 Q_1[(2)^{1/2}(r+ia\cos\theta)],\\
J_{\bar{m}} &= \epsilon\sigma_0 h_0 Q_{-1}/[(2)^{1/2}(r-ia\cos\theta)].
\end{aligned} \tag{3.35}$$

Teukolsky derived decoupled equations for the field quantities $\Phi_1$ and $\Phi_{-1}$ even in the presence of sources.[1,2] These equations are second order differential equations for the field quantities $\Phi_1$ and $\Phi_{-1}$ with source terms denoted $P_1$ and $P_{-1}$. The equation for $\Phi_1$ when explicitly written with the operators $\mathcal{D}$, $\mathcal{D}^+$, $\eth$, and $\eth^+$ is

$$-(\Sigma)^{-1}[\mathcal{D}\mathcal{D}^+\Delta + 2i\omega r + \eth(\eth^+ - \cot\theta) - 2a\omega\cos\theta]\,\Phi_1 = P_1. \tag{3.36}$$

The source term in this equation is

$$\begin{aligned}
P_1 = &-[(r+ia\cos\theta)^{-1}\mathcal{D}(r+ia\cos\theta) + 2(r-ia\cos\theta)^{-1}]J_m\\
&-(2)^{-1/2}(r+ia\cos\theta)^{-1}[\partial - 2ia\sin\theta(r-ia\cos\theta)^{-1}]J_l.
\end{aligned} \tag{3.37}$$

When the expression for $J_m$ and $J_l$ from Eq. (3.35) are substituted into Eq. (3.37), this equation reduces to

$$P_1 = \epsilon(2)^{1/2}\sigma_0(h_0Q_1 + ia\sin\theta f_1 Z_0)/\Sigma. \tag{3.38}$$

The equation for $\Phi_{-1}$ is simplest when expressed as an equation for a function $\Omega$ defined by the relation

$$\Phi_{-1} = \Delta\Omega/[2(r-ia\cos\theta)^2]. \tag{3.39}$$

The equation for $\Omega$ is

$$-\Delta(2\Sigma)^{-1}[\mathcal{D}^+\mathcal{D}\,\Delta - 2i\omega r + \eth^+(\eth - \cot\theta) + 2a\omega\cos\theta]\,\Omega = P_{-1}, \tag{3.40}$$

where $P_{-1}$ is

$$\begin{aligned}
P_{-1} = &-(r-ia\cos\theta)^2\{-(2)^{-1/2}(r-ia\cos)^{-1}[\Sigma^{-1}\eth^+\Sigma\\
&-2ia\sin\theta(r-ia\cos\theta)^{-1}]J_n + [\Delta(2\Sigma)^{-1}\mathcal{D}^+\\
&+3\Delta(2\Sigma)^{-1}(r-ia\cos\theta)^{-1}]J_{\bar{m}}\}.
\end{aligned} \tag{3.41}$$

Using Eq. (3.35) this becomes

$$P_{-1} = \epsilon\sigma_0\Delta(2)^{-1/2}\Sigma^{-1}(-h_0Q_{-1} + ia\sin\theta f_{-1}Z_0). \tag{3.42}$$

The task now is to use the ladder relations, and the differential equations satisfied by the functions $f_1$, $f_{-1}$, $h_0$ and $Q_1$, $Q_{-1}$, and $Z_0$ [Eqs. (2.34)–(2.40)] to determine the solutions to Eqs. (3.36) and (3.40). It is quite surprising that the two equations turn out to have very simple solutions! The solution of Eq. (3.36) is

$$\Phi_1 = \sigma_0^2\epsilon f_1 Q_1/[(2)^{1/2}i\omega] \tag{3.43}$$

as can be verified by direct substitution. Similarly, the solution of Eq. (3.40) is

$$\Omega = \sigma_0^2\epsilon f_{-1}Q_{-1}/[(2)^{1/2}i\omega]. \tag{3.44}$$

One should notice that the solutions of the *homogeneous* differential equations [Eq. (3.36) and (3.40) with $T_1$ and $T_{-1}$ set equal to 0]

$$\Phi_1^{\text{hom}} = h_1Z_1, \tag{3.45}$$

$$\Omega^{\text{hom}} = h_{-1}Z_{-1}, \tag{3.46}$$

have a form remarkably similar to Eqs. (3.43) and (3.44). In fact the only difference is the substitutions

$$h_{\pm1} \to \epsilon\sigma_0^2 f_{\pm1}/[(2)^{1/2}i\omega],\quad Z_{\pm1} \to Q_{\pm1}. \tag{3.47}$$

### F. Solutions in class 4: The limit $\omega = 0$

When $\omega = 0$, $f_1$ is identically $h_1$, $f_{-1}$ is $h_{-1}$, $Q_1$ is $Z_1$, and $Q_{-1}$ is $Z_{-1}$. Therefore—except for the factor of $(\omega)^{-1}$—Eqs. (3.43) and (3.44) become identical to Eqs. (3.45) and (3.46). On the other hand, the source terms $J_n$, $J_l$, $J_m$, and $J_{\bar{m}}$ are nonzero even when $\omega = 0$. To resolve this apparent paradox we must consider the approach to zero frequency with somewhat greater care: Since the zero frequency limit of $f_1$ is $h_1$, purely as a formal definition, $f_1$ may be written as

$$f_1 = h_1 + \omega b_1. \tag{3.48}$$

Equation (3.20) above defined $k_{-1}$ by

$$g_{-1} = h_0 + \omega k_{-1}. \tag{3.49}$$

If both sides of Eq. (3.40) are acted on by the operator $\eth$, the resulting expression is

$$\sigma_1 h_1 = \sigma_0 f_1 + \omega \eth k_{-1}. \tag{3.50}$$

Rearranging this expression and using Eqs. (3.24) and (2.27) we find that to order $\omega$,

$$b_1 = -\eth k_{-1}/\sigma_0. \tag{3.51}$$

The function $b_1$, for all values of the frequency, satisfies an inhomogeneous differential equation derived by substituting Eq. (3.48) into Eq. (2.39), which is the equation satisfied by $f_1$, and making use of Eq. (2.12), viz.,

$$\eth\eth^+ \Delta b_1 + 2i\omega r b_1 - \sigma_0^2 b_1 = -2ih_0/\sigma_0 + (\sigma_0^2 - \lambda_1 - 2)\, h_1/\omega. \tag{3.52}$$

Similarly, by making the formal definitions

$$f_{-1} = h_{-1} + \omega b_{-1}, \quad Q_1 = Z_1 + \omega V_1, \quad Q_{-1} = Z_{-1} + \omega V_{-1} \tag{3.53}$$

differential equations for the new functions $b_{-1}$, $V_1$, and $V_{-1}$ can be derived in like manner:

$$\begin{aligned}
&(\eth^+\eth\,\Delta - 2i\omega r - \sigma_0^2)\, b_{-1} = 2ih_0/\sigma_0 + (\sigma_0^2 - \lambda_1 - 2)\, h_{-1}/\omega,\\
&[\eth(\eth^+ - \cot\theta) - 2a\omega\cos\theta + \sigma_0^2]\, V_1\\
&\quad = 2a\omega \sin\theta Z_0/\sigma_0 + (\lambda_1 + 2 - \sigma_0^2)\, Z_1/\omega,\\
&[\eth^+(\eth - \cot\theta) + 2a\omega\cos\theta + \sigma_0^2]\, V_{-1}\\
&\quad = 2a\omega\sin\theta Z_0/\sigma_0 + (\lambda_1 + 2 - \sigma_0^2)\, Z_{-1}/\omega.
\end{aligned} \tag{3.54}$$

In terms of these new functions the solutions to Eqs. (3.36) and (3.40) for the functions $\Phi_1$ and $\Omega$ are

$$\begin{aligned}
\Phi_1 &= \sigma_0^2\epsilon(2)^{-1/2}(i\omega)^{-1}[h_1 Z_1 + \omega(b_1 Z_1 + h_1 V_1) + \omega^2 b_1 V_1],\\
\Omega &= \sigma_0^2\epsilon(2)^{-1/2}(i\omega)^{-1}[h_1 Z_1 + \omega(b_{-1}Z_{-1} + h_{-1}V_{-1}) + \omega^2 b_{-1}V_{-1}].
\end{aligned} \tag{3.55}$$

Now it is apparent that the term proportional to $(\omega)^{-1}$ is, in both cases, the solution to the respective homogeneous differential equations and hence it may be dropped for small $\omega$ when considering the class 4 solutions. The functions $\Phi_1$ and $\Phi_{-1}$ may then be expressed in terms of the functions $b_1$, $b_{-1}$, $Z_1$, and $Z_{-1}$.

### G. Solutions in class 4: $\xi$'s obtained from the $\Phi$'s

In order to prove that the solutions given for $\Phi_1$ and $\Phi_{-1}$ are in fact consistent solutions of Eq. (3.3) with the specified source terms, the field function $\Phi_0$ should be independent of whether it is computed by quadratures from any one of Eqs. (3.3a), (3.3b), (3.3c), or (3.3d). To compute $\Phi_0$ from quadratures we make use of the relationships between the functions $f_1$, $f_{-1}$, $h_0$ and $Q_1$, $Q_{-1}$, $Z_0$ as expressed by Eqs. (2.34)–(2.36) and Eqs. (2.38)–(2.40), respectively. That the solutions presented for $\Phi_1$ and $\Phi_{-1}$ are consistent is verified. The function $\Phi_0$ has the form:

$$\begin{aligned}
\Phi_0 = \sigma_0\epsilon(2\omega i)^{-1}(r - ia\cos\theta)^{-2}\{&[(\sigma_0 r + i\omega r^2/\sigma_0)\, h_0 - \Delta f_{-1}]\, Z_0\\
&- iah_0[(\sigma_0\cos\theta - a\omega\cos^2\theta/\sigma_0)\, Z_0 + \sin\theta Q_1]\}.
\end{aligned} \tag{3.56}$$

Using Eqs. (3.9), (3.10), and (3.11) we solve by quadratures for the scalars $\xi'_n$, $\xi'_l$, $\xi'_m$, and $\xi'_{\bar m}$, these solutions being

$$\begin{aligned}
\xi'_n &= \Delta r f_{-1}\epsilon\sigma_0 Z_0/(2\Sigma i\omega),\\
\xi'_l &= r f_1\epsilon\sigma_0 Z_0/(i\omega),\\
\xi'_m &= -ia\cos\theta\sigma_0\epsilon h_0 Q_1/[(2)^{1/2}(i\omega)(r + ia\cos\theta)],\\
\xi'_{\bar m} &= -ia\cos\theta\sigma_0\epsilon h_0 Q_{-1}/[(2)^{1/2}(i\omega)(r - ia\cos\theta)].
\end{aligned} \tag{3.57}$$

The primes on the $\xi$'s signify that the solutions are not yet solutions to the almost-Killing equation [Eq. (1.1)] because we must go full circle and insure that the divergence of the $\xi$'s is given by Eq. (3.34). The divergence of the vector given by Eq. (3.57) is

$$\nabla^\alpha \xi'_\alpha = 3\epsilon h_0 Z_0. \tag{3.58}$$

In order to make a self-consistent solution to the almost-Killing equation, the gauge invariance of the "field tensor" defined in Eq. (3.2) must be used to add to the vector $\boldsymbol{\xi}'$ the gradient of a scalar function so that the divergence of the sum is equal to $\chi$ of Eq. (3.34). To this end, set

$$\xi^\alpha = \xi'^\alpha + \nabla^\alpha \eta. \tag{3.59}$$

The Laplacian operator when acting on the scalar $\eta$ must equal $(1-3\epsilon)\, h_0 Z_0$ in order for $\boldsymbol{\xi}$ to be a solution of Eq. (1.1). Thus the inhomogeneous equation for $\eta$ is

$$\begin{aligned}
&-(\eth^+\Delta\eth + 2i\omega r + (\eth^+ - \cot\theta)\,\eth - 2a\omega\cos\theta)\eta\\
&= (1-3\epsilon)\,\Sigma h_0 Z_0.
\end{aligned} \tag{3.60}$$

The solution to this equation is

$$\eta = -(1-3\epsilon)[j(r) Z_0(\theta,\phi,t) + a^2 h_0(r) A(\theta,\phi,t)] \tag{3.61}$$

as can be verified by direct substitution of Eq. (3.61) into Eq. (3.60) and making use of the relations of Eqs. (2.41) and (2.42). The complete class 4 solutions are given by

$$\begin{aligned}
\xi_n &= \Delta(2\Sigma)^{-1}[r f_{-1}\epsilon\sigma_0 Z_0/(i\omega)\\
&\quad - (1-3\omega)(\eth^+ j Z_0 + a^2\sigma_0 f_{-1} A)],\\
\xi_l &= [r f_1\epsilon\sigma_0 Z_0/(i\omega) + (1-3\epsilon)(\eth j Z_0 + a^2\sigma_0 f_1 A)],\\
\xi_m &= [-ia\cos\theta\sigma_0\epsilon h_0 Q_1/(i\omega)\\
&\quad + (1-3\epsilon)(j\sigma_0 Q_1 + a^2 h_0 \eth A)]/[(2)^{1/2}(r + ia\cos\theta)],\\
\xi_{\bar m} &= [-ia\cos\theta\sigma_0\epsilon h_0 Q_{-1}/(i\omega)\\
&\quad + (1-3\epsilon)(-j\sigma_0 Q_{-1} + a^2 h_0 \eth^+ A)]/[(2)^{1/2}(r - ia\cos\theta)].
\end{aligned} \tag{3.62}$$

In order to compute the $\omega = 0$ limit from the solution in this form it is necessary again to explore the composition of the term proportional to $(\omega)^{-1}$. This is done by using the functions defined in Eqs. (3.20), $k_{\pm1}$; (3.28), $U_{\pm1}$; (3.48), $f_1$; (3.53), $f_{-1}$ and $V_{\pm1}$. Substitution of these functions into Eqs. (3.57), which gives the terms proportional to $(\omega)^{-1}$, gives for the $\xi'$'s:

$$\begin{aligned}
\xi'_n &= \epsilon\sigma_0[\Delta(2\Sigma)^{-1} r h_{-1} X_1]/(i\omega)\\
&\quad + (i2\Sigma)^{-1}\epsilon\,\Delta r(b_{-1}X_1 - h_{-1}U_1 - \omega b_{-1}U_1),\\
\xi'_l &= \epsilon\sigma_0(r h_1 X_{-1})/(i\omega)\\
&\quad - i\epsilon\sigma_0 r(b_1 X_{-1} - h_1 U_{-1} - \omega b_1 U_{-1}),
\end{aligned}$$

$$\xi_m' = \epsilon\sigma_0(-ia\cos\theta g_{-1}Z_{+1}/[(2)^{1/2}(r+ia\cos\theta)i\omega]$$
$$-i\epsilon\sigma_0(-ia\cos\theta)(g_{-1}V_1 - k_{-1}Z_1 - \omega k_{-1}V_1)/$$
$$\times[(2)^{1/2}(r+ia\cos\theta)],$$
$$\xi_{\bar m}' = \epsilon\sigma_0(-ia\cos\theta g_1 Z_{-1})/[(2)^{1/2}(r-ia\cos\theta)i\omega]$$
$$-i\epsilon\sigma_0(-ia\cos\theta)(g_1V_{-1} - k_1Z_{-1} - \omega k_1V_{-1})/$$
$$\times[(2)^{1/2}(r-ia\cos\theta)]. \tag{3.63}$$

The terms in Eq. (3.63) proportional to $(\omega)^{-1}$ that remain are proportional to the scalars of Eqs. (3.12) and (3.13), the constant of proportionality being $-i\epsilon\sigma_0\sigma_1/[(2)^{1/2}\omega]$. The scalars of Eqs. (3.12) and (3.13) were defined in connection with the divergence-free class 2 solution. Thus if it is necessary, these terms in Eq. (3.63) proportional to $(\omega)^{-1}$ can be dropped from Eq. (3.62) if the scalar defined in Eq. (3.17) multiplied by the constant $-i\epsilon\sigma_0\sigma_1/[(2)^{1/2}\omega]$ is also dropped from Eq. (3.62) so that the divergence of what remains is still equal to $\chi$ of Eq. (3.34). For the sake of completeness, we present the time independent class 4 solutions which were derived after a lengthy calculation:

$$\xi_n = \Delta(2\Sigma)^{-1}[n(n+1)]^{-1/2}\{\epsilon[r^2h_{-1}Y_0^{nm} - 2iamr(h_{-1}T^{nm}$$
$$-[\eth^* w/(n(n+1))^{1/2} - h_{-1}/(n(n+1))]Y_0^{nm})]$$
$$+[n(n+1)]^{1/2}[(1-\epsilon)\eth^* j - \epsilon\eth^* S]Y_0^{nm}$$
$$+n(n+1)a^2h_{-1}[(1-\epsilon)A^{nm} - \epsilon B^{nm}]\},$$
$$\xi_l = [n(n+1)]^{-1/2}\{-\epsilon[r^2h_1Y_0^{nm} - 2iamr(h_1T^{nm}$$
$$-[\eth w/(n(n+1))^{1/2} - h_1/(n(n+1))]Y_0^{nm})]$$
$$-[n(n+1)]^{1/2}[(1-\epsilon)\eth j - \epsilon\eth s]Y_0^{nm}$$
$$-n(n+1)a^2h_1[(1-\epsilon)A^{nm} - \epsilon B^{nm}]\};$$
$$\xi_m = [n(n+1)]^{-1/2}(r+ia\cos\theta)^{-1}(2)^{-1/2}\{\epsilon\{a^2\cos^2\theta h_0 Y_1^{nm}$$
$$+2iam(ia\cos\theta)[h_0[\eth T^{nm}/(n(n+1))^{1/2} + Y_1^{nm}/(n(n+1))]$$
$$-wY_1^{nm}]\} - n(n+1)[(1-\epsilon)j - \epsilon s]Y_1^{nm}$$
$$-[n(n+1)]^{1/2}a^2h_0[(1-\epsilon)\eth A^{nm} - \epsilon\eth B^{nm}]\},$$
$$\xi_{\bar m} = [n(n+1)]^{-1/2}(r-ia\cos\theta)^{-1}(2)^{-1/2}\{-\epsilon\{a^2\cos^2\theta h_0 Y_{-1}^{nm}$$
$$-2iam(ia\cos\theta)[h_0[-\eth^* T^{nm}/(n(n+1))^{1/2} + Y_{-1}^{nm}/(n(n+1))]$$
$$-wY_{-1}^{nm}]\} + [(1-\epsilon)j - \epsilon s]Y_{-1}^{nm}$$
$$-[n(n+1)]^{1/2}a^2h_0[(1-\epsilon)\eth^* A^{nm} - \epsilon\eth^* B^{nm}]\}. \tag{3.64}$$

We can characterize the solutions of class 4 by the order of the angular function $Z_0$ with which we have expressed the divergence $\chi$ [of Eq. (3.34)] of the class 4 solutions. For the time independent solutions, the $(n=0,\ m=0)$ solution can, at least, be computed analytically. It is

$$\xi_{(0,0)} = +\left\{(4\pi)^{1/2}(3)^{-1}\left[\left(r + 2M + \frac{4M^2(r-M)}{\Delta} + \frac{2M(2M^2-a^2)}{(M^2-a^2)^{1/2}\Delta}\right)\frac{\Delta}{\Sigma}\right]\frac{\partial}{\partial r} + \left(\frac{a^2\sin\theta\cos\theta}{\Sigma}\right)\frac{\partial}{\partial\theta}\right\}. \tag{3.65}$$

## IV. DISCUSSION AND EXAMPLE

How do we know that we have found "all" the solutions of the original Eq. (1.1)? A strong indication (though not a rigorous proof) is obtained by solving Eq. (1.1) in flat spacetime. These flat space solutions (which are presented in another context for the time independent case in Morse and Feshbach[14]) can be proven to be a complete set of vector solutions. This follows from the properties of the spin-weight spherical harmonics. Our Kerr solutions are in one to one correspondence with these as can be seen by simply setting $a$ and $M$ equal to zero in our solutions.

Finally, it is interesting to work out one definite example of the way that the AKE can be used to extend a vector field in from infinity: We can compute those almost-Killing vectors which are also asymptotic Killing vectors in the sense of Eq. (1.3). (These asymptotic Killing vectors are by no means uniquely defined because to any asymptotic Killing vector one can add a solution to the almost-Killing equation which vanishes as $r$ gets very large.) In Minkowski space, there are ten independent Killing vectors, the generators of translations, of rotations and of Lorentz boosts. In this presentation of the asymptotic Killing vector solutions to the AKE, the axis of symmetry of the black hole will be taken to be the $z$ axis.

The Minkowski space translation Killing vectors and the almost-Killing vectors which approach them asymptotically are

$$\frac{\partial}{\partial t}:\ \frac{\partial}{\partial t},$$
$$\frac{\partial}{\partial z}:\ \left(\frac{\cos\theta\Delta}{\Sigma}\right)\frac{\partial}{\partial r} - \left(\frac{(r-M)\sin\theta}{\Sigma}\right)\frac{\partial}{\partial\theta},$$
$$\frac{\partial}{\partial x}:\ \left(\frac{d}{dr}(h_0^{1,1}\right)\left(\frac{\sin\theta\cos\phi\Delta}{\Sigma}\right)\frac{\partial}{\partial r} + \left(\frac{h_0^{1,1}\cos\theta\cos\phi}{\Sigma}\right)\frac{\partial}{\partial\theta}$$
$$-h_0^{1,1}\sin\theta\sin\phi\left\{\left[\left(1-\frac{2Mr}{\Sigma}\right)\Big/(\Delta\sin^2\theta)\right]\frac{\partial}{\partial\phi} - \left(\frac{2aMr}{\Delta\Sigma}\right)\frac{\partial}{\partial t}\right\},$$
$$\frac{\partial}{\partial y}:\ \frac{d}{dr}(h_0^{1,1})\left(\frac{\sin\theta\sin\phi\Delta}{\Sigma}\right)\frac{\partial}{\partial r} + \left(\frac{h_0^{1,1}\cos\theta\sin\phi}{\Sigma}\right)\frac{\partial}{\partial\theta}$$
$$+h_0^{1,1}\sin\theta\cos\phi\left\{\left[\left(\frac{1-2Mr}{\Sigma}\right)\Big/(\Delta\sin^2\theta)\right]\frac{\partial}{\partial\phi} - \left(\frac{2Mar}{\Delta\Sigma}\right)\frac{\partial}{\partial t}\right\}, \tag{4.1}$$

where the radial function $h_0^{1,1}$ is the solution to Eq. (2.15) with $(n,m)=(1,1)$,

$$h_0^{1,1} = (r^2 - m^2 + a^2)^{1/2}\cos\{(a/\delta)\ln[(r-r_-)/(r-r_+)]$$
$$-\tan^{-1}[a/(r-M)]\},$$
$$\delta = 2(M^2-a^2)^{1/2},\quad r_\pm = M \pm \delta/2. \tag{4.2}$$

The generators of the rotations in Minkowski space are the vectors

$$\mathbf{L}_Z = \frac{\partial}{\partial\phi},$$
$$\mathbf{L}_X = (\sin\phi)\frac{\partial}{\partial\theta} + (\cot\theta\cos\phi)\frac{\partial}{\partial\phi}, \tag{4.3}$$
$$\mathbf{L}_Y = (\cos\phi)\frac{\partial}{\partial\theta} - (\cot\theta\sin\phi)\frac{\partial}{\partial\phi}.$$

Alternately, these Killing vectors can be used to define the complex Killing vectors by

$$\mathbf{L}_+ = (2)^{-1/2}(L_Y + iL_X)$$
$$= \exp(i\phi)(r/2)[(1+\cos\theta)\,\mathbf{m} + (1-\cos\theta)\,\bar{\mathbf{m}}],$$
$$\mathbf{L}_- = (2)^{-1/2}(L_Y - iL_X)$$
$$= \exp(-i\phi)(r/2)[(1-\cos\theta)\mathbf{m} + (1+\cos\theta)\,\bar{\mathbf{m}}]. \quad (4.4)$$

The solutions to the almost-Killing equation asymptotically equal to $L_+$, $L_-$, and $L_Z$ are:

$$\mathbf{L}_Z : \frac{\partial}{\partial\phi},$$

$$\mathbf{L}_+ : -ia(2)^{-3/2}\left(\frac{d}{dr}(h_0^{1,1}) - \frac{iah_0^{1,1}}{\Delta}\right)$$
$$\times \Delta\Sigma^{-1}\exp(i\phi)\sin\theta(\cos\theta+1)\,\mathbf{l}$$
$$+ia(2)^{-1/2}\left(\frac{d}{dr}(h_0^{1,1}) + \frac{iah_0^{1,1}}{\Delta}\right)\exp(i\phi)\sin\theta(1-\cos)\,\mathbf{n}$$
$$+(\cos\theta+1)(r-ia\cos\theta)^{-1}\left[rh_0^{1,1} - \Delta\left(\frac{d}{dr}(h_0^{1,1}) - \frac{iah_0^{1,1}}{\Delta}\right)\Big/2\right]$$
$$\exp(i\phi)\mathbf{m} + (1-\cos\theta)(r+ia\cos\theta)^{-1}\left[rh_0^{1,1} - \Delta\left(\frac{d}{dr}(h_0^{1,1})\right.\right.$$
$$\left.\left.+\frac{iah_0^{1,1}}{\Delta}\right)\Big/2\right]\bar{\mathbf{m}}$$

$$\mathbf{L}_- : -ia(2)^{-3/2}\left(\frac{d}{dr}(h_0^{1,1}) + \frac{iah_0^{1,1}}{\Delta}\right)$$
$$\times\Delta\Sigma^{-1}\exp(-i\phi)\sin\theta(1-\cos\theta)\,\mathbf{l}$$
$$+ia(2)^{-1/2}\left(\frac{d}{dr}(h_0^{1,1}) - \frac{iah_0^{1,1}}{\Delta}\right)\exp(-i\phi)\sin\theta(1+\cos\theta)\mathbf{n}$$
$$+(1+\cos\theta)(r-ia\cos\theta)^{-1}\left[rh_0^{1,1} - \Delta\left(\frac{d}{dr}(h_0^{1,1})\right.\right.$$
$$\left.\left.+\frac{iah_0^{1,1}}{\Delta}\right)\Big/2\right]\exp(-i\phi)\,\mathbf{m}$$
$$+(1+\cos\theta)(r+ia\cos\theta)^{-1}\left[rh_0^{1,1} - \Delta\left(\frac{d}{dr}(h_0^{1,1})\right.\right.$$
$$\left.\left.-\frac{iah_0^{1,1}}{\Delta}\right)\Big/2\right]\exp(-i\phi)\,\bar{\mathbf{m}}. \quad (4.5)$$

The Killing vectors are the generators of Lorentz boosts in Minkowski space are the vectors

$$\mathbf{M}_Z = \frac{z\partial}{\partial t} + \frac{t\partial}{\partial z},$$
$$\mathbf{M}_X = \frac{x\partial}{\partial t} + \frac{t\partial}{\partial x}, \quad (4.6)$$
$$\mathbf{M}_Y = \frac{y\partial}{\partial t} + \frac{t\partial}{\partial y}.$$

The boost Killing vectors can be expressed in a more convenient form in order to compare with the solutions to the almost-Killing equation which approach them for large values of $r$:

$$\mathbf{M}_Z = (r+t)\cos\theta 2^{-1}\mathbf{l} + (r-t)\cos\theta\mathbf{n} - t\sin\theta(2)^{-1/2}(\mathbf{m}+\bar{\mathbf{m}}),$$

$$\mathbf{M}_+ = (2)^{-1/2}(\mathbf{M}_X + i\mathbf{M}_Y)$$
$$= (2)^{-3/2}(r+t)\sin\theta\exp(i\phi)\,\mathbf{l} + (2)^{-1/2}(r-t)\exp(i\phi)\sin\theta\mathbf{n}$$
$$+(2)^{-1}t(1+\cos\theta)\exp(i\phi)\,\mathbf{m} + (2)^{-1}t(\cos\theta-1)$$
$$\times\exp(i\phi)\,\bar{\mathbf{m}},$$
$$\mathbf{M}_- = (2)^{-1/2}(\mathbf{M}_X - i\mathbf{M}_Y)$$
$$= (2)^{-3/2}(r+t)\sin\theta\exp(-i\phi)\,\mathbf{l} + (2)^{-1/2}(r-t)$$
$$\times\exp(-i\phi)\sin\theta\mathbf{n} + (2)^{-1}t(\cos\theta-1)\exp(-i\phi)\,\mathbf{m}$$
$$+(2)^{-1}t(\cos\theta+1)\exp(-i\phi)\,\bar{\mathbf{m}}. \quad (4.7)$$

In Eqs. (4.7) and (4.4) the tetrad vectors are the Minkowski space analogs of the tetrad given in Eq. (2.3) obtained by setting $a=M=0$ in Eq. (2.3). [In all other equations in this paper the tetrad that is referred to is that of Eq. (2.3).] The solutions to the almost Killing equation which are asymptotically equal to $M_Z$, $M_+$, and $M_-$ are:

$$\mathbf{M}_Z : \Delta(2\Sigma)^{-1}\cos\theta[r+t-M(r^2-a^2)/\Delta]\,\mathbf{l}$$
$$+\cos\theta[r-t-M(r^2-a^2)/\Delta]\,\mathbf{n}$$
$$-(2)^{-1/2}(r-ia\cos\theta)^{-1}(r-M)\sin\theta(t-ia\cos\theta)\,\mathbf{m}$$
$$-(2)^{-1/2}(r+ia\cos\theta)^{-1}(r-m)\sin\theta(t+ia\cos\theta)\,\bar{\mathbf{m}},$$

$$M_+ : \Delta(2)^{-3/2}\Sigma^{-1}\left[(2r+t)\left(\frac{d}{dr}(h_0^{1,1}) - \frac{iah_0^{1,1}}{\Delta}\right) - \frac{(r^2+a^2)h_0^{1,1}}{\Delta}\right.$$
$$\left.+iaM\left(\frac{d}{dr}(b) - \frac{iab}{\Delta}\right)\right]\sin\theta\exp(i\phi)\,\mathbf{l}$$
$$+(2)^{-1/2}\left[(2r-t)\left(\frac{d}{dr}(h_0^{1,1}) + \frac{iah_0^{1,1}}{\Delta}\right) - \frac{(r^2+a^2)h_0^{1,1}}{\Delta}\right.$$
$$\left.-iaM\left(\frac{d}{dr}(b) + \frac{iab}{\Delta}\right)\right]\sin\theta\exp(i\phi)\,\mathbf{n}$$
$$+(2)^{-1}(r-ia\cos\theta)^{-1}[th_0^{1,1} - ia(1+\cos\theta)h_0^{1,1}$$
$$+iaMb](\cos\theta+1)\exp(i\phi)\,\mathbf{m}$$
$$+(2)^{-1}(r+ia\cos\theta)^{-1}[th_0^{1,1} + ia(\cos\theta-1)h_0^{1,1}$$
$$+iaMb](\cos\theta-1)\exp(i\phi)\,\bar{\mathbf{m}},$$

$$M_- : \Delta(2)^{-3/2}\Sigma^{-1}[(2r+t)\left(\frac{d}{dr}(h_0^{1,1}) + \frac{iah_0^{1,1}}{\Delta}\right) - \frac{(r^2+a^2)h_0^{1,1}}{\Delta}$$
$$\left.-iaM\left(\frac{d}{dr}(b) + \frac{iab}{\Delta}\right)\right]\sin\theta\exp(-i\phi)\,\mathbf{l}$$
$$+(2)^{1/2}\left[(2r-t)\left(\frac{d}{dr}(h_0^{1,1}) - \frac{iah_0^{1,1}}{\Delta}\right) - \frac{(r^2+a^2)h_0^{1,1}}{\Delta}\right.$$
$$\left.+iaM\left(\frac{d}{dr}(b) - \frac{iab}{\Delta}\right)\right]\sin\theta\exp(-i\phi)\,\mathbf{n}$$
$$+(2)^{-1}(r-ia\cos\theta)^{-1}[th_0^{1,1} - ia(\cos\theta-1)h_0^{1,1}$$
$$-iaMb](\cos\theta-1)\exp(-i\phi)\,\mathbf{m}$$
$$+(2)^{-1}(r+ia\cos\theta)^{-1}[th_0^{1,1} + ia(\cos\theta+1)h_0^{1,1}$$
$$-iaMb](\cos\theta+1)\exp(-i\phi)\,\mathbf{m}, \quad (4.8)$$

where the radial function $b(r)$ satisfies the inhomogeneous differential equation

$$\left(\frac{d}{dr}\Delta\frac{d}{dr} + \frac{a^2}{\Delta} - 2\right)b(r) = \frac{4rh_0^{1,1}}{\Delta}.$$

## ACKNOWLEDGMENTS

I wish to thank William Press and Larry Smarr for their encouragement and advice and especially for comments on the manuscript. I thank James York for helpful discussions.

[1]S.A. Teukolsky, Phys. Rev. Lett. **29**, 114 (1972).
[2]S.A. Teukolsky, Astrophys. J. **185**, 635 (1973).
[3]S. Chandrasekhar, Proc. Roy. Soc. A (GB) **349**, 1 1656 (1976).
[4]J.W. York, Ann. Inst. Henri Poincaré, XXI, 4, 319 (1974).
[5]R.H. Boyer and R.W. Lindquist, J. Math. Phys. **8**, 265 (1967).
[6]E.T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[7]W. Kinnersley, J. Math. Phys. **10**, 1195 (1969).
[8]B. Carter, Commun. Math. Phys. **10**, 280 (1968).
[9]S.A. Teukolsky and W.H. Press, Astrophys. J. **193**, 443 (1974).
[10]A.A. Starobinsky, Zh. Eksp. Teor. Fiz. **64**, 48 (1973) [Sov. Phys. JETP **37**, 28 (1974)].
[11]A.A. Starobinsky and S.M. Churilov, Zh. Eksp. Teor. Fiz. **65**, 3 (1973) [Sov. Phys. JETP **38**, 1 (1974)].
[12]W.H. Press, Astrophys. J. **175**, 243 (1972).
[13]R.A. Breuer, *Gravitational Perturbation Theory and Synchrotron Radiation* (Lecture Notes in Physics: 44, Springer-Verlag, Berlin, Heidelberg, 1975), p. 108.
[14]M.M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), part II, p. 1807.
[15]R.M. Wald, Phys. Rev. D **10**, 6, 1680 (1974).
[16]R. Penrose, Ann. Phys. (N.Y.) **10**, 171 (1960).

# Nonuniqueness in the inverse scattering problem

A. J. Devaney

*Eikonix Corporation, 103 Terrace Hall Avenue, Burlington, Massachusetts 01803*
(Received 21 June 1977)

The inverse scattering problem consists of determining the functional form of a scattering potential given the scattering matrix $A(k_0\mathbf{s}, k_0\mathbf{s}_0)$ for all scattering directions $\mathbf{s}$ and one or more values of the wave vector $k_0\mathbf{s}_0$. In this paper it is shown that within the framework of the first Born approximation the inverse scattering problem as defined above does not possess a unique solution. It is also shown that within the framework of exact (potential) scattering theory the problem does not admit a unique solution given only the scattering matrix for a single fixed value of the wave vector $k_0\mathbf{s}_0$ as data. The final section in the paper considers scattering experiments using incident fields other than plane waves and where knowledge of the scattered field at all points exterior to the scattering volume is available as data. It is found that, within the framework of exact scattering theory, the data generated by any *single* such experiment is not sufficient to uniquely specify the scattering potential while, within the framework of the first Born approximation, the data generated by any *finite* number of such experiments is not sufficient to uniquely specify the potential.

## 1. INTRODUCTION

A problem of considerable practical importance in optics, acoustics, and quantum mechanics is that of determining the structure of an unknown scattering potential from scattering data. The scattering data consists usually of either the intensity or the complex amplitude of the scattered field in the wave zone of the scatterer for cases when the field incident to the scatterer is a unit amplitude plane wave having a specified wavenumber $k_0$ and direction of propagation $\mathbf{s}_0$. The intensity of the scattered field in the wave zone is usually termed the differential cross section $d\sigma/d\Omega$ and is related to the complex amplitude of the scattered field according to the equation

$$\frac{d\sigma}{d\Omega} = |A(k_0\mathbf{s}, k_0\mathbf{s}_0)|^2. \tag{1}$$

The argument of the complex amplitude indicates its dependence on the wavevector $k_0\mathbf{s}_0$ of the incident plane wave and of the direction $\mathbf{s}$ at which it is measured.

In this paper we shall be concerned only with cases where the complex amplitude (or scattering matrix) $A(k_0\mathbf{s}, k_0\mathbf{s}_0)$ is available as data. This quantity can be measured directly in optical scattering experiments[1-3] or it can be analytically deduced from the differential cross section for certain classes of scattering potentials.[4] Depending on the number of scattering experiments performed $A(k_0\mathbf{s}, k_0\mathbf{s}_0)$ will be determined for one or more values of $k_0\mathbf{s}_0$ and for some set of scattering directions $\mathbf{s}$. The inverse scattering problem then consists of using this data to determine the functional form of the scattering potential $V(\mathbf{r})$. We shall assume throughout this paper that the scatterer is localized within a finite scattering volume $\tau$ and that the potential $V(\mathbf{r})$ is at least piecewise continuous in $\tau$.

It is known[5] that if the scattering matrix $A(k_0\mathbf{s}, k_0\mathbf{s}_0)$ is specified for *fixed* $k_0$ and *all* values of $\mathbf{s}_0$ and $\mathbf{s}$ then unique determination of the scattering potential is possible. Such a determination can be performed using an iterative algorithm presented in Ref. 5 or, in the case of weakly scattering potentials, by means of a procedure presented in Ref. 6 and further developed in Refs. 7–9. Another case known to yield a unique solution occurs when the scattering potential $V(\mathbf{r})$ is either independent of the wavenumber $k_0$ or depends on this quantity in a known way. In such situations a unique solution is obtained when $A(k_0\mathbf{s}, k_0\mathbf{s}_0)$ is specified for a fixed incident field direction of propagation $\mathbf{s}_0$ and for all values of the wavenumber $k_0$ and all scattering angles $\mathbf{s}$.[5]

It follows from symmetry that in cases where the scattering potential is spherically symmetric the scattering matrix is a function only of $k_0$ and the angle between the two unit vectors $\mathbf{s}$ and $\mathbf{s}_0$. Consequently, for such cases the scattering matrix need be specified only for a single arbitrary value of $k_0\mathbf{s}_0$ and all scattering directions $\mathbf{s}$ in order to uniquely specify the scattering potential.[10] Thus, by use of algorithms presented in the literature[5,10] a spherically symmetric scattering potential can be uniquely determined from scattering data obtained in a single experiment.

In this paper we present the results of an investigation into the problem of *uniquely* determining the structure of nonspherically symmetric scattering potentials from scattering data obtained in a single, or possibly finite number of experiments. This investigation was motivated primarily by statements appearing in the literature[11-13] to the effect that this should be possible at least within the framework of the Born approximation. Ineed, an algorithm has actually been devised[12,13] for determining the functional form of a weak scattering potential from a specification of the scattering matrix for a *single* fixed value of $k_0\mathbf{s}_0$ and all values of $\mathbf{s}$.

In Sec. 2 we address the uniqueness question for inverse scattering within the first Born approximation. The scattering data are assumed to consist of the scattering matrix given for all scattering directions $\mathbf{s}$ and for any prespecified finite set of values of $k_0\mathbf{s}_0$. It is shown that, with the exception of spherically symmetric scattering potentials, such data are *not* sufficient to uniquely specify the scattering potential. Indeed, in analogy to the *nonradiating distributions* (see Footnote 14) known to exist in radiation (source) problems, it is shown that an infinite number of scattering

 0022-2488/78/1907-1526$1.00  

potentials can be determined all of which are localized within any specified scattering volume and all of which produce scattering matrices which, within the Born approximation, vanish identically for all values of $\mathbf{s}$ and for any finite number of specified values of $k_0\mathbf{s}_0$.[14] Due to the linearity of the first Born approximation it follows that any one of these potentials can be added to any other potential without changing the scattering data observed in any finite number of experiments. Determination of the scattering potential from a finite number of experiments is thus nonunique within the first Born approximation unless auxillary information is available to rule out the presence of such "nonscattering" potentials within the scattering volume.

In Sec. 3 we treat the inverse scattering problem within the framework of exact scattering theory. A theorem is established which shows essentially that knowledge of the field everywhere *outside* a localized scattering potential is not sufficient to uniquely specify the field *within* the scattering volume. It follows that the scattering potential can *not* be uniquely specified from scattering data obtained in any *single* experiment and, in particular, from the scattering matrix given for a single fixed value of $k_0\mathbf{s}_0$ and all scattering directions $\mathbf{s}$.

The analysis presented in Secs. 2 and 3 assumes that only incident plane wave fields are used in a scattering experiment and that the data obtained in any such experiment is limited to the scattering matrix $A(k_0\mathbf{s}, k_0\mathbf{s}_0)$. In Sec. 4 we examine the impact on inverse scattering of using other types of incident waves and of allowing more extensive field measurements to be performed. It is shown that the nonuniqueness properties of inverse scattering within the Born approximation determined in Sec. 2 hold for arbitrary incident waves which are expressable as a finite sum of plane wave fields. In addition, the nonuniqueness result established in Sec. 3 for inverse scattering within the framework of exact scattering theory is shown to be applicable to cases where the incident field to the scatterer is entirely arbitrary.

It is also shown in Sec. 4 that knowledge of the scattered field in the wave zone (i.e., as $k_0 r \to \infty$) is completely *equivalent* to knowledge of the scattered field throughout all space exterior to the scattering volume. It follows that for the case of incident plane waves the scattering matrix is completely equivalent to knowledge of the scattered field throughout all space exterior to the scattering volume. Consequently, the nonuniqueness properties established in Secs. 2 and 3 remain valid even if one is allowed unlimited measurements of the field exterior to the scattering volume.

## 2. INVERSE SCATTERING WITHIN THE BORN APPROXIMATION

The scattering of a scalar wavefunction $\psi(\mathbf{r}, k_0\mathbf{s}_0)$ by a potential $V(\mathbf{r})$ is described by the reduced wave equation[15]

$$(\nabla^2 + k_0^2)\psi(\mathbf{r}, k_0\mathbf{s}_0) = V(\mathbf{r})\psi(\mathbf{r}, k_0\mathbf{s}_0). \tag{2}$$

The argument $k_0\mathbf{s}_0$ is included in the wavefunction $\psi$ to indicate its dependence on the wave vector of the incident plane wave $\exp(ik_0\mathbf{s}_0 \cdot \mathbf{r})$. For sufficiently weak scattering potentials the wavefunction $\psi$ is given approximately by the first Born approximation

$$\psi(\mathbf{r}, k_0\mathbf{s}_0) \approx \exp(ik_0\mathbf{s}_0 \cdot \mathbf{r}) - \frac{1}{4\pi}\int_\tau d^3r'\, V(\mathbf{r}')\exp(ik_0\mathbf{s}_0 \cdot \mathbf{r}') \times \frac{\exp(ik_0|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|}, \tag{3}$$

where $\tau$ denotes the scattering volume which is assumed to be finite. In the wave zone (i.e., as $k_0 r \to \infty$) Eq. (3) yields the following expression for $\psi$:

$$\psi(\mathbf{r}, k_0\mathbf{s}_0) \sim \exp(ik_0\mathbf{s}_0 \cdot \mathbf{r}) - \frac{1}{4\pi} A_B(k_0\mathbf{s}, k_0\mathbf{s}_0)\frac{\exp(ik_0 r)}{r}. \tag{4}$$

Here $A_B(k_0\mathbf{s}, k_0\mathbf{s}_0)$ denotes the scattering matrix in the Born approximation and is given by

$$A_B(k_0\mathbf{s}, k_0\mathbf{s}_0) = \int_\tau d^3r'\, V(\mathbf{r}')\exp(ik_0\mathbf{s}_0 \cdot \mathbf{r}')\exp(-ik_0\mathbf{s} \cdot \mathbf{r}'). \tag{5}$$

The Born approximation to the scattering matrix as given in Eq. (5) is intimately related to the threefold Fourier transform $\tilde{V}(\mathbf{k})$ of the scattering potential $V(\mathbf{r})$:

$$\tilde{V}(\mathbf{k}) = \int_\tau d^3r\, V(\mathbf{r})\exp(-i\mathbf{k} \cdot \mathbf{r}). \tag{6}$$

On comparing Eq. (5) with Eq. (6) we conclude that

$$A_B(k_0\mathbf{s}, k_0\mathbf{s}_0) = \tilde{V}[k_0(\mathbf{s} - \mathbf{s}_0)], \tag{7}$$

which implies that the scattering matrix determines $\tilde{V}(\mathbf{k})$ for all those frequency vectors $\mathbf{k}$ given by

$$\mathbf{k} = k_0(\mathbf{s} - \mathbf{s}_0). \tag{8}$$

For a given scattering experiment using an incident plane wave of fixed wave vector $k_0\mathbf{s}_0$ the values of $\mathbf{k}$ satisfying Eq. (8) lie on the surface defined by

$$\mathbf{k} \cdot \mathbf{k} = k^2 = 2k_0^2(1 - \mathbf{s}_0 \cdot \mathbf{s}). \tag{9}$$

It follows that if a (theoretically infinite) number of experiments were to be performed all using incident plane waves of fixed wavenumber $k_0$ but varying directions of propagation $\mathbf{s}_0$ the totality of scattering data so obtained allows $\tilde{V}(\mathbf{k})$ to be determined for all values of $\mathbf{k}$ lying within a sphere of radius $2k_0$.[6] A band-limited approximation $V_{\mathrm{bl}}(\mathbf{r})$ to the scattering potential is then immediately achievable by means of the relationship

$$V_{\mathrm{bl}}(\mathbf{r}) = \frac{1}{(2\pi)^3}\int_{k \le 2k_0} d^3k\, \tilde{V}(\mathbf{k})\exp(i\mathbf{k} \cdot \mathbf{r}), \tag{10}$$

where the Fourier amplitude $\tilde{V}(\mathbf{k})$ is that which is reconstructed from the scattering matrix.

Because the scattering potential $V(\mathbf{r})$ is assumed to be piecewise continuous and is localized within the finite volume $\tau$ its Fourier transform $\tilde{V}(\mathbf{k})$ is an entire analytic function of the three Cartesian components $k_x, k_y, k_z$ of the wave vector $\mathbf{k}$.[16] It follows from a theorem[17] in analytic function theory that $\tilde{V}(\mathbf{k})$ is uniquely determined for all values of $\mathbf{k}$ by its value within any finite *volume element* in $\mathbf{k}$ space. The extension of $\tilde{V}(\mathbf{k})$ from its value over a finite volume element to all of $\mathbf{k}$ space can, in principle, be performed by analytic continuation.[17] Because the set of scattering experiments discussed above yields $\tilde{V}(\mathbf{k})$ for all values of k lying within the sphere $k \leq 2k_0$ the theorem alluded to above implies that $\tilde{V}(\mathbf{k})$ is completely determined for all values of k by the scattering data. The complete reconstruction of $\tilde{V}(\mathbf{k})$ could, in principle, be performed using analytic continuation although a much more satisfactory (and realistic) approach would be to use the algorithm presented in Ref. 5.

The discussion presented above indicates that the key to unique determination of the scattering potential is to perform a sufficient number of scattering experiments to lead to a specification of $\tilde{V}(\mathbf{k})$ over any finite *volume* element in **k** space. Any single scattering experiment leads only to a specification of $\tilde{V}(\mathbf{k})$ over the *surface* defined in Eq. (9). To extend this surface to a volume requires an infinite number of experiments to be performed. One such series of experiments is that discussed above where $k_0$ is held fixed and $\mathbf{s}_0$ is varied. Alternatively, if $V(\mathbf{r})$ is independent of $k_0$ or depends on $k_0$ in a known way then $\mathbf{s}_0$ can be held fixed and $k_0$ varied as was indicated in the Introduction.

If the scattering potential is spherically symmetric then it is easily shown that $\tilde{V}$ is an analytic function of the single variable $k^2 = k_x^2 + k_y^2 + k_z^2$. In this case $\tilde{V}$ is completely determined by its value over any finite *interval* of $k^2$. According to Eq. (9) such an interval is generated by a single scattering experiment; e.g., by fixing $k_0\mathbf{s}_0$ and measuring the scattering matrix for all scattering directions **s**. As mentioned in the Introduction there are algorithms presented in the literature[5] for reconstructing spherically symmetric scattering potentials from such scattering data.

The discussion presented above should provide ample proof of the fact that with the exception of spherically symmetric potentials a finite number of scattering experiments simply do not generate sufficient information to lead to a unique reconstruction of the scattering potential. To reiterate: *the entire analytic function $\tilde{V}(k)$ of the three variables $(k_x, k_y, k_z)$ is uniquely specified for all values of $(k_x, k_y, k_z)$ if and only if it is specified over a finite volume element in* **k** *space*. A single scattering experiment leads to a specification of $\tilde{V}(\mathbf{k})$ only over the *surface* defined in Eq. (9) while a finite number of experiments leads to a specification of $\tilde{V}(\mathbf{k})$ only over a finite number of such surfaces.

To illustrate the nonuniqueness inherent to scattering potential reconstruction from a finite number of scattering experiments we shall now show that there exist an infinite number of (non spherically symmetric) scattering potentials all localized within any specified scattering volume $\tau$ and all of which produce a scattering matrix $A_B(k_0\mathbf{s}, k_0\mathbf{s}_0)$ which *vanishes identically for all values of* **s** *for any finite number of fixed values of* $k_0\mathbf{s}_0$.[14] Any one of these potentials can be added to any other potential $V(\mathbf{r})$ without changing the scattering matrix obtained from a finite number of scattering experiments performed on $V(\mathbf{r})$. Determination of $V(\mathbf{r})$ from the scattering data is thus nonunique unless auxillary information (such as *a priori* knowledge of spherical symmetry) is available to rule out the presence of such scattering potentials within the scattering volume.

To begin we shall construct a class of scattering potentials $F(\mathbf{r})$ which produce a scattering matrix which vanishes for all values of **s** and a *single* value of $k_0\mathbf{s}_0$. This class is then easily generalized to produce potentials having scattering matrices which vanish for any finite number of incident field wave vectors. The key to generating the required set of scattering potentials is to note from Eq. (7) that $A_B(k_0\mathbf{s}, k_0\mathbf{s}_0)$ is equal to the transform of the potential evaluated when $\mathbf{k} = k_0(\mathbf{s} - \mathbf{s}_0)$. Thus, by constructing a potential $F(\mathbf{r})$ whose transform $\tilde{F}(\mathbf{k})$ vanishes for these values of k we have constructed a potential which yields a scattering matrix which vanishes for a fixed value of $k_0\mathbf{s}_0$ and all values of **s**.

One such class of scattering potentials is generated by the relationship

$$\tilde{F}(\mathbf{k}) = [k^2 - 2k_0^2(1 - \mathbf{s}_0 \cdot \mathbf{s})]\tilde{\lambda}(\mathbf{k}), \tag{11}$$

where **s** is defined in Eq. (8). Here $\tilde{\lambda}(\mathbf{k})$ is an entire analytic function chosen to have a transform $\lambda(\mathbf{r})$ which is continuous with continuous first and second partial derivatives and to be localized within a prescribed scattering volume $\tau$ but is otherwise *arbitrary*. Substituting for **s** from Eq. (8) we find that

$$\tilde{F}(\mathbf{k}) = [k^2 + 2k_0\mathbf{s}_0 \cdot \mathbf{k}]\tilde{\lambda}(\mathbf{k}), \tag{12}$$

so that

$$\begin{aligned} F(\mathbf{r}) &= \frac{1}{(2\pi)^3}\int d^3k\, \tilde{F}(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}) \\ &= -[\nabla^2 + 2ik_0\mathbf{s}_0 \cdot \nabla]\lambda(\mathbf{r}). \end{aligned} \tag{13}$$

We can verify directly that the scattering potential $F(\mathbf{r})$ produces a scattering matrix $A_B(k_0\mathbf{s}, k_0\mathbf{s}_0)$ identically equal to zero. In particular, substituting $F(\mathbf{r})$ for $V(\mathbf{r})$ in Eq. (5) we obtain

$$\begin{aligned} A_B(k_0\mathbf{s}, k_0\mathbf{s}_0) = -\int_\tau d^3r\{[\nabla^2 + 2ik_0\mathbf{s}_0 \cdot \nabla]\lambda(\mathbf{r})\} \\ \times \exp(ik_0\mathbf{s}_0 \cdot \mathbf{r}) \exp(-ik_0\mathbf{s} \cdot \mathbf{r}). \end{aligned} \tag{14}$$

We now note that

$$\begin{aligned} (\nabla^2 + k_0^2)[\lambda(\mathbf{r}) \exp(ik_0\mathbf{s}_0 \cdot \mathbf{r})] \equiv \exp(ik_0\mathbf{s}_0 \cdot \mathbf{r})[\nabla^2 + 2ik_0\mathbf{s}_0 \cdot \nabla] \\ \times \lambda(\mathbf{r}). \end{aligned} \tag{15}$$

Substituting from Eq. (15) into Eq. (14) we find that

$$\begin{aligned} A_B(k_0\mathbf{s}, k_0\mathbf{s}_0) &= -\int_\tau d^3r\{(\nabla^2 + k_0^2)[\lambda(\mathbf{r}) \\ &\quad \times \exp(ik_0\mathbf{s}_0 \cdot \mathbf{r})]\} \exp(ik_0\mathbf{s} \cdot \mathbf{r}) \\ &= -\int_\tau d^3r[\lambda(\mathbf{r}) \exp(+ik_0\mathbf{s}_0 \cdot \mathbf{r})] \\ &\quad \times (\nabla^2 + k_0^2) \exp(-ik_0\mathbf{s} \cdot \mathbf{r}) = 0, \end{aligned} \tag{16}$$

where we have made use of the assumptions that $\lambda(\mathbf{r})$ is continuous with continuous first and second derivatives and is localized within $\tau$.

Finally, we note that the above construction technique can be repeated using a different value of $k_0\mathbf{s}_0$ (say $k'_0\mathbf{s}'_0$) and replacing $\lambda(\mathbf{r})$ by $F(\mathbf{r})$ as defined in Eq. (13). We then obtain the class of scattering potentials

$$F'(\mathbf{r}) = [\nabla^2 + 2ik'_0\mathbf{s}'_0 \cdot \nabla][\nabla^2 + 2ik_0\mathbf{s}_0 \cdot \nabla]\lambda(\mathbf{r}), \tag{17}$$

which yield scattering matrices which vanish for all values of $\mathbf{s}$ when the wave vector of the incident field is either $k_0\mathbf{s}_0$ or $k'_0\mathbf{s}'_0$. Continuing in this fashion it is evident that it is possible to construct a class of potentials whose scattering matrices vanish for all values of $\mathbf{s}$ when the wave vector of the incident field has any one of a finite number of prespecified values.

## 3. INVERSE SCATTERING WITHIN THE FRAMEWORK OF EXACT SCATTERING THEORY

The exact solution $\psi(\mathbf{r}, k_0\mathbf{s}_0)$ to Eq. (2) satisfies the integral equation

$$\psi(\mathbf{r}, k_0\mathbf{s}_0) = \exp(ik_0\mathbf{s}_0\cdot\mathbf{r}) - \frac{1}{4\pi} \times \int_\tau d^3r'\, V(\mathbf{r}')\psi(\mathbf{r}', k_0\mathbf{s}_0)\,\frac{\exp(ik_0|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|}, \quad (18)$$

which can be shown to possess a unique solution $\psi$ under rather weak conditions on the scattering potential $V(\mathbf{r})$. In the *inverse scattering problem* we are not concerned with solving Eq. (18) for $\psi$ in terms of the potential $V$ but, rather, are concerned with the problem of deducing the scattering potential $V(\mathbf{r})$ from knowledge of the scattering matrix

$$A(k_0\mathbf{s}, k_0\mathbf{s}_0) = \int_\tau d^3r'\, V(\mathbf{r}')\psi(\mathbf{r}', k_0\mathbf{s}_0)\,\exp(-ik_0\mathbf{s}\cdot\mathbf{r}'). \quad (19)$$

In this section we shall show that knowledge of the scattering matrix for a single fixed value of the incident field's wave vector $k_0\mathbf{s}_0$ and all scattering directions is *not* sufficient information to deduce uniquely the scattering potential $V(\mathbf{r})$. In order to establish this result we require the following Lemma.

*Lemma*: There exist an infinite number of continuous functions $\rho(\mathbf{r})$ localized within any specified volume $\tau$ and such that

$$\int_\tau d^3r'\rho(\mathbf{r}')\,\frac{\exp(ik_0|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|} = 0, \quad (20)$$

for all values of $\mathbf{r}$ lying outside the volume $\tau$.

The above Lemma follows immediately if we choose $\rho(\mathbf{r})$ to be given by

$$\rho(\mathbf{r}) = (\nabla^2 + k_0^2)\mu(\mathbf{r}), \quad (21)$$

where $\mu(\mathbf{r})$ is a continuous function with continuous first and second partial derivatives and is localized within $\tau$ but is otherwise arbitrary. Making use of Eq. (21) we find that

$$\begin{aligned}\int_\tau d^3r'\,\rho(\mathbf{r}')\,\frac{\exp(ik_0|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|} &= \int_\tau d^3r'[\nabla'^2 + k_0^2)\mu(\mathbf{r}')] \times \frac{\exp(ik_0|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|}\\ &= \int_\tau d^3r'\,\mu(\mathbf{r}')\left[(\nabla'^2 + k_0^2)\,\frac{\exp(ik_0|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|}\right] = -4\pi\mu(\mathbf{r}),\end{aligned} \quad (22)$$

where we made use of the assumed properties of $\mu(\mathbf{r}')$ and the fact that

$$(\nabla'^2 + k_0^2)\,\frac{\exp(ik_0|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|} = -4\pi\delta(\mathbf{r}-\mathbf{r}'), \quad (23)$$

with $\delta(\cdot)$ being the Dirac delta function.

The Lemma established above allows us to prove the following Theorem.

*Theorem*: If there exist a set of functions $(\psi, V)$ which satisfy the integral equation (18) for a fixed value of $k_0\mathbf{s}_0$ then there exists at least one other set of functions $(\psi', V')$ which also satisfy that equation for the given value of $k_0\mathbf{s}_0$ and which are such that:

(i) The regions of localization of both $V$ and $V'$ are identical and equal to the volume $\tau$.

(ii) Outside $\tau$, $\psi$, and $\psi'$ are everywhere equal.

*Theorem Proof*: It follows from the Lemma that we can find a set of functions $(\phi, \rho)$ such that

$$\phi(\mathbf{r}) = \int_\tau d^3r'\,\rho(\mathbf{r}')\,\frac{\exp(ik_0|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|} = 0, \quad (24)$$

for all values of $\mathbf{r}$ lying outside $\tau$. The function

$$\psi'(\mathbf{r}, k_0\mathbf{s}_0) = \psi(\mathbf{r}, k_0\mathbf{s}_0) + \phi(\mathbf{r}) \quad (25)$$

is thus everywhere equal to $\psi$ outside $\tau$. Moreover,

$$\psi'(\mathbf{r}, k_0\mathbf{s}_0) = \exp(ik_0\mathbf{s}_0\cdot\mathbf{r}) + \int_\tau d^3r'[V(\mathbf{r}')\psi(\mathbf{r}', k_0\mathbf{s}_0) + \rho(\mathbf{r}')]\,\frac{\exp(ik_0|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|}, \quad (26)$$

so that there exists a $V'(\mathbf{r})$ equal to

$$V'(\mathbf{r}) = \frac{V(\mathbf{r})\psi(\mathbf{r}, k_0\mathbf{s}_0) + \rho(\mathbf{r})}{\psi'(\mathbf{r}, k_0\mathbf{s}_0)} = \frac{V(\mathbf{r})\psi(\mathbf{r}, k_0\mathbf{s}_0) + \rho(\mathbf{r})}{\psi(\mathbf{r}, k_0\mathbf{s}_0) + \phi(\mathbf{r})}, \quad (27)$$

which has the same region of localization as $V$ (namely $\tau$) and which together with $\psi'$ satisfy Eq. (18). This completes the proof of the theorem.

The two scattered fields $\psi(\mathbf{r}, k_0\mathbf{s}_0)$ and $\psi'(\mathbf{r}, k_0\mathbf{s}_0)$ refered to in the above theorem are everywhere equal outside the scattering volume $\tau$ and consequently generate identical scattering data. It follows that the scattering data (i.e., the scattering matrix) generated from any single scattering experiment employing a unit amplitude incident plane wave $\exp(ik_0\mathbf{s}_0\cdot\mathbf{r})$ is not sufficient to uniquely specify the scattering potential. Indeed, we have actually shown that a unique reconstruction is not possible even if one is given the value of the scattered field at *all* points lying outside the scattering volume $\tau$. Although this later result appears to be stronger than the former they are actually equivalent since, as shown in the following section, the scattering matrix uniquely specifies the scattered field at all points lying outside the scattering volume $\tau$.

## 4. INVERSE SCATTERING USING DATA OTHER THAN THE SCATTERING MATRIX

Throughout this paper we have assumed that the scattering data from which the scattering potential is to be determined consists solely of the scattering matrix $A(k_0\mathbf{s}, k_0\mathbf{s}_0)$ or $A_B(k_0\mathbf{s}, k_0\mathbf{s}_0)$. It is reasonable to inquire of the possibility of uniquely specifying the scattering potential from a single or, possibly, finite number of scattering experiments if more *complete* scattering data were available. The scattering matrix is the coefficient of the spherical wave $-\exp(ik_0r)/(4\pi r)$ in the leading term of the asymptotic expansion of the scattered field when the field incident to the scatterer is the unit amplitude plane wave field $\exp(ik_0\mathbf{s}_0\cdot\mathbf{r})$. Thus, we might expect that complete knowledge of the scattered field (not just the leading term in its asymptotic expansion) throughout all space exterior to the scattering volume $\tau$ might provide more information than the scattering matrix and thus might lead to a unique specification of the scattering potential. Alternatively, the possibility exists that the use of incident fields other than the simple plane wave might possibly generate sufficient scattering data to yield a unique determination of the scattering potential in a single or finite number of scattering experiments.

Consider first the possiblity of generating additional information in a scattering experiment by (hypothetically) determining the scattered field at all points external to the scattering volume. In both the Born approximation and the exact theory the scattered field is given by an expression of the general form [cf., Eqs. (3) and (18)]

$$\psi(\mathbf{r},k_0\mathbf{s}_0)=\exp(ik_0\mathbf{s}_0\cdot\mathbf{r})-\frac{1}{4\pi}\int_\tau d^3r'\,S(\mathbf{r}')\,\frac{\exp(ik_0|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|},\tag{28}$$

where $S(\mathbf{r}')$ is a "source function" equal to the product of the scattering potential $V(\mathbf{r})$ with the field $\psi$ or, in the case of the Born approximation, the incident field $\exp(ik_0\mathbf{s}_0\cdot\mathbf{r})$. The second term on the right-hand side of Eq. (28) is easily shown to be the solution to the inhomogeneous Helmholtz equation

$$(\nabla^2+k_0^2)\phi(\mathbf{r})=S(\mathbf{r}),\tag{29}$$

which satisfies Sommerfeld's radiation condition (i.e., behaves as an outgoing wave at infinity). It follows from a theorem[18] pertaining to such wave fields that $\phi(\mathbf{r})$ *is specified uniquely at all points exterior to the volume in which* $S(\mathbf{r})$ *is localized (i.e., outside* $\tau$*) by the leading term in its asymptotic expansion.* Since the leading term in the asymptotic expansion of $\phi$ is simply the product of the scattering matrix with the spherical wave $-\exp(ik_0r)/(4\pi r)$ it follows that $\phi(\mathbf{r})$ and, hence, $\psi(\mathbf{r},k_0\mathbf{s}_0)$ is uniquely determined for all values of $\mathbf{r}$ lying outside $\tau$ by the scattering matrix.

We conclude from the discussion presented above that for the case of an incident plane wave $\exp(ik_0\mathbf{s}_0\cdot\mathbf{r})$ complete knowledge of the scattered field throughout all space exterior to the scattering volume $\tau$ is *completely equivalent* to knowledge of the scattering matrix $A(k_0\mathbf{s},k_0\mathbf{s}_0)$ evaluated at that particular value of $k_0\mathbf{s}_0$ equal to the wave vector of the incident plane wave. This result is true both within the framework of the Born approximation and in exact scattering theory. In conjunction with the results presented in Sec. 2 it leads to the assertion that: *within the framework of the Born approximation it is impossible to uniquely deduce the structure of a scattering potential from measurements of the field external to the scattering volume in any finite number of scattering experiments using incident plane waves.* The results presented in Sec. 3 show that the same assertion holds true within the framework of exact scattering theory if we allow only a single scattering experiment.

Consider now the use of incident fields other than plane waves in a scattering experiment. Denoting such an incident field by $\xi(\mathbf{r},k_0)$ we find that the scattered field $\psi(\mathbf{r},k_0)$ satisfies the integral equation

$$\psi(\mathbf{r},k_0)=\xi(\mathbf{r},k_0)-\frac{1}{4\pi}\int_\tau d^3r'\,V(\mathbf{r}')\psi(\mathbf{r}',k_0)\,\frac{\exp(ik_0|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|},\tag{30}$$

which, within the Born approximation, yields the following expression for $\psi$:

$$\psi(\mathbf{r},k_0)=\xi(\mathbf{r},k_0)-\frac{1}{4\pi}\int_\tau d^3r'\,V(\mathbf{r}')\xi(\mathbf{r}',k_0)\frac{\exp(ik_0|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|}.\tag{31}$$

A perusal of the theorem established in the previous section reveals that its proof does not depend on the nature of the incident wavefield to the scattering potential. Consequently, the theorem applies equally well to the integral equation (30) for arbitrary fixed incident fields $\xi(\mathbf{r},k_0)$. It follows that *within the framework of exact scattering theory it is impossible to uniquely deduce the structure of a potential from measurements of the field external to the scattering volume in any single scattering experiment.*

The situation is actually more complicated within the approximate scattering model provided by the Born approximation. Without loss of generality we can assume the scattering data consists of the leading term in the asymptotic expansion of $\psi$ which is easily shown to be

$$\psi(\mathbf{r},k_0)\sim\xi(\mathbf{r},k_0)-\frac{1}{4\pi}f(k_0\mathbf{s})\,\frac{\exp(ik_0r)}{r}.\tag{32}$$

Here

$$f(k_0\mathbf{s})=\int_\tau d^3r'\,V(\mathbf{r}')\xi(\mathbf{r}',k_0)\exp(-ik_0\mathbf{s}\cdot\mathbf{r}'),\tag{33}$$

is the generalization of the scattering matrix $A_B(k_0,k_0\mathbf{s}_0)$ to incident wavefields other than plane waves.

We shall consider only the case when the incident field is expressable as a sum of a finite number of plane waves $\exp(ik_0\mathbf{s}_0\cdot\mathbf{r})$ having identical wavenumbers $k_0$ but different directions of propagation $\mathbf{s}_0$[19]; i.e.,

$$\xi(\mathbf{r},k_0)=\sum_{\mathbf{s}_0}a(k_0\mathbf{s}_0)\exp(ik_0\mathbf{s}_0\cdot\mathbf{r}).\tag{34}$$

Substituting Eq. (34) into Eq. (33) yields

$$\begin{aligned}f(k_0\mathbf{s})&=\sum_{\mathbf{s}_0}a(k_0\mathbf{s}_0)\int_\tau d^3r'\,V(\mathbf{r}')\exp(ik_0\mathbf{s}_0\cdot\mathbf{r}')\exp(-ik_0\mathbf{s}\cdot\mathbf{r}')\\&=\sum_{\mathbf{s}_0}a(k_0\mathbf{s}_0)A_B(k_0\mathbf{s},k_0\mathbf{s}_0),\end{aligned}\tag{35}$$

where we have used the definition of the scattering matrix within the Born approximation given in Eq. (5).

The results presented in Sec. 2 showed that it is always possible to find a scattering potential $\overline{V}(\mathbf{r})$ which produces a scattering matrix $\overline{A}_B(k_0\mathbf{s},k_0\mathbf{s}_0)$ which vanishes identically for all values of $\mathbf{s}$ and any finite number of fixed values of $k_0\mathbf{s}_0$. Since the sum in Eq. (35) is over a finite number of terms we can replace $A_B(k_0\mathbf{s}, k_0\mathbf{s}_0)$ in this equation with the sum $A_B + \overline{A}_B$ without changing (the observable) $f(k_0\mathbf{s})$; i.e.,

$$\begin{aligned} f(k_0\mathbf{s}) &= \sum_{\mathbf{s}_0} a(k_0\mathbf{s}_0)[A_B(k_0\mathbf{s}, k_0\mathbf{s}_0) + \overline{A}_B(k_0\mathbf{s}, k_0\mathbf{s}_0)] \\ &= \sum_{\mathbf{s}_0} a(k_0\mathbf{s}_0) \int_\tau d^3r' [V(\mathbf{r}') + \overline{V}(\mathbf{r}')] \exp(ik_0\mathbf{s}_0 \cdot \mathbf{r}') \\ &\quad \times \exp(-ik_0\mathbf{s} \cdot \mathbf{r}') \\ &= \int_\tau d^3r' [V(\mathbf{r}') + \overline{V}(\mathbf{r}')] \xi(\mathbf{r}', k_0) \exp(-ik_0\mathbf{s} \cdot \mathbf{r}'). \end{aligned} \tag{36}$$

It follows immediately from Eq. (36) that it is not possible to uniquely specify the scattering potential from $f(k_0\mathbf{s})$ and, hence, from a finite number of scattering experiments using incident fields representable in the form given in Eq. (34).


[1]E. Wolf, J. Opt. Soc. Am. **60**, 18 (1970).
[2]H.G. Schmidt-Weinmar, J. Opt. Soc. Am. **63**, 547 (1973).
[3]H.G. Schmidt-Weinmar, J. Opt. Soc. Am. **65**, 999 (1975).
[4]R.G. Newton, J. Math. Phys. **9**, 2050 (1968).
[5]R.T. Prosser, J. Math. Phys. **10**, 1819 (1969).
[6]E. Wolf, Optics Commun. **1**, 153 (1969).
[7]W.H. Carter, J. Opt. Soc. Am. **60**, 306 (1970).
[8]W.H. Carter and P.-C. Ho, Appl. Opt. **13**, 162 (1974).
[9]P.-C. Ho and W.H. Carter, Appl. Opt. **15**, 313 (1976).
[10]R.G. Newton, SIAM Rev. **12**, 346 (1970).
[11]P.J. Wyatt, Appl. Opt. **7**, 1879 (1968).
[12]H.G. Schmidt-Weinmar, J. Opt. Soc. Am. **61**, 1578A (1971).
[13]H.G. Schmidt-Weinmar, J. Opt. Soc. Am. **65**, 1188A (1975).
[14]This result is the scattering problem counterpart of the well-known result in radiation (source) problems that there exist an infinite number of sources to the inhomogeneous Helmholtz equation all localized within any specified volume $\tau$ and all of which yield a radiation pattern which vanishes identically. The theory underlying such *nonradiating sources* is treated in: T. Erber, Fortschr. Phys. **9**, 343 (1961), G.H. Goedecke, Phys. Rev. **135**, B281 (1964); A.J. Devaney and E. Wolf, Phys. Rev. D **8**, 1044 (1973).
[15]A review of potential scattering and, in particular, a derivation of the results given in Eqs. (2)–(5) can be found in P.M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Sec. 9.3.
[16]This result is a three-dimensional analog of a well-known theorem that the Fourier transform of a continuous function which vanishes outside a finite interval is a boundary value of an entire analytic function. This theorem follows at once from a known result on analytic functions defined by definite integrals [cf., E.T. Copson, *An Introduction to the Theory of Functions of a Complex Variable* (Oxford University, London, 1962), Sec. 5.5]. The multidimensional form of the theorem is the Plancherel–Polya theorem (cf., Ref. 17, p. 352).
[17]B.A. Fuks, *Introduction to the Theory of Analytic Functions of Several Complex Variables* (American Mathematics Society, Providence, R.I., 1963).
[18]C. Müller, *Foundations of the Mathematical Theory of Electromagnetic Waves* (Springer, New York, 1969), p. 339. A procedure for actually determining a radiated field from its radiation pattern is given in A.J. Devaney and E. Wolf, J. Math. Phys. **15**, 234 (1974).
[19]The representation given in Eq. (34) is a discrete (sum) approximation to the well known plane wave expansion of solutions to the homogeneous Helmholtz equation [cf., A.J. Devaney and G.C. Sherman, SIAM Rev. **15**, 765 (1973)].

# Classical and relativistic vorticity in a semi-Riemannian manifold

Maurice J. Dupré and Steven I. Rosencrans

*Department of Mathematics, Tulane University, New Orleans, Louisiana 70118*
(Received 3 October 1977)

It is shown that a form of the Cauchy–Lagrange formula for the evolution of vorticity in a barotropic flow generalizes to the case of ideal fluid motion on higher-dimensional Riemannian or semi-Riemannian manifolds.

## 1. INTRODUCTION

An important theorem concerning the motion of an ideal fluid is the Cauchy—Lagrange formula, valid for barotropic flow (see Sec. 2), which expresses the vector $\omega_t/\rho_t$ as a linear function of $\omega_0/\rho_0$. Here $\omega_t$, the vorticity at time $t$, is defined as curl $\mathbf{v}_t$, where $\mathbf{v}_t$ is the velocity field at time $t$; $\rho_t$ is the density (mass per unit volume) at time $t$, not assumed constant. The precise formula is

$$\frac{\omega_t}{\rho_t}\bigg|_{F_t(x)} = F_{t*}\left(\frac{\omega_0}{\rho_0}\right)\bigg|_x, \tag{1}$$

where the matrix $F_{t*}$ is the Jacobian matrix of the diffeomorphism $F_t$ representing the fluid motion; $F_t(x)$ = position at time $t$ of the particle which was at $x$ at time 0.[1] Cauchy proved this formula in 1815. See Refs. 2—4. A consequence of (1) is that if the vorticity is zero at a point $x$ at time 0, it remains zero at all points $F_t(x)$ visited by the particle that was at $x$. (In particular, if the vorticity is zero everywhere at $t=0$, it remains zero everywhere, and hence locally $\mathbf{v}=\nabla\phi$ for some potential $\phi$. So potential flow persists for all time.)

Our object is to generalize the Cauchy—Lagrange formula to the case of ideal (compressible or incompressible) fluid motion in an $n$-dimensional semi-Riemannian manifold. Actually this relation cannot be directly generalized; one proves a covariant form, in which the vorticity is a 2-form rather than a vector field. For the incompressible case, a reference for this is Marsden.[5] Our method, however, is to view the problem in space—time, which both simplifies the calculations, and allows a simultaneous treatment of *relativistic* motion. We analyze in detail the isometric motion of a relativistic fluid. This has been discussed by Mason,[6] Ciubotariu,[7] and Trautman;[8] the present analysis is much simpler.

## 2. IDEAL FLUID FLOW IN $\mathbb{R}^3$

For the purposes of comparison, we briefly derive the classical results. There are two equations, expressing conservation of momentum and of mass. [In general a subscript $t$ indicates evaluation of a quantity at time $t$, but only when we wish to emphasize the evaluation. Thus, for example, $\mathbf{v}=\mathbf{v}(x,t)=\mathbf{v}_t$.]

$$\frac{\partial\mathbf{v}}{\partial t}+(\mathbf{v}\cdot\nabla)\mathbf{v}=-\frac{1}{\rho}\nabla p,\quad \frac{\partial\rho}{\partial t}+\nabla\cdot(\rho\mathbf{v})=0. \tag{2}$$

Here $p$ is the pressure field. A barotropic fluid motion is one in which $p$ and $\rho$ are functionally related. This may be either a physical property of the fluid or else a peculiarity of the particular motion under study. We define an operator

$$u=\frac{\partial}{\partial t}+\mathbf{v}\cdot\nabla. \tag{3}$$

By curling (2) and using some vector identities involving curl and divergence, one finds

$$\left[u,\frac{\omega}{\rho}\cdot\nabla\right]=0, \tag{4}$$

where the square bracket stands for the commutator (Lie bracket) of these differential operators. Thus in space—time $M=R^3\times\mathbb{R}$ the *flows* generated by these two vector-fields commute. The flows are, respectively,

$$\begin{aligned}&\text{flow of } u\text{: } G_t(x,s)=(F_{s+t}F_s^{-1}(x),s+t),\\ &\text{flow of } \frac{\omega}{\rho}\cdot\nabla\text{: } H_t(x,s)=(y(t),s),\end{aligned} \tag{5}$$

where $y(t)$ is the position $t$-units along the integral curve of $\omega/\rho$ *frozen at time* $s$:

$$\frac{dy}{dr}=\frac{\omega(y(r),s)}{\rho(y(r),s)},\quad y(0)=x.$$

The "quadrilateral" in Fig. 1 commutes because of (4), as does therefore the projection of this figure onto $\mathbb{R}^3$. We thus have the picture in Fig. 2 from which we can draw two conclusions:

*Proposition* 1: (1) Frozen $\omega/\rho$ lines are material lines, that is, they move with the fluid. (2) Frozen $\omega/\rho$-lines permute the particle paths. I.e., if each point of a given particle path is transported by the $\omega/\rho$-line frozen at the local time, the resulting curve is also a particle path.

*Remark*: Part (2) seems to be a new observation.

*Corollary*: If the vorticity is zero at a point $x$, it is zero at all points along the furture trajectory beginning at $x$.

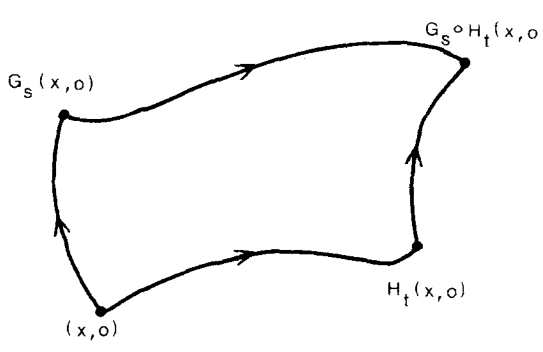



FIG. 1.

 0022-2488/78/1907-1532$1.00  

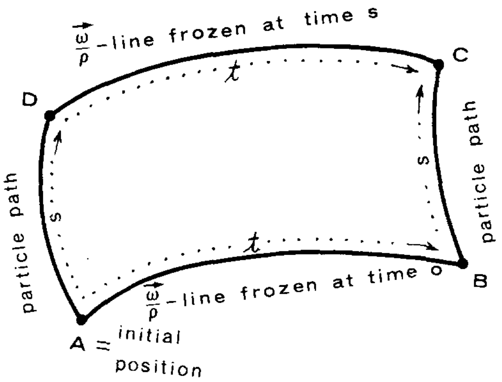



FIG. 2.

*Proof*: By hypothesis, $A=B$ in Fig. 1. But two distinct particle paths cannot originate from the same point. Thus $C=D$, which means $\omega=0$ at time $s$ on that particle path.

Equivalent consequences of (4) are the relations[8]

$$\frac{\omega}{\rho}\cdot\nabla\Big|_{G_t(x,s)} = G_{t*}\Big|_{(x,s)}\frac{\omega}{\rho}\cdot\nabla\Big|_{(x,s)}, \tag{6}$$

$$u\big|_{H_t(x,s)} = H_{t*}\big|_{(x,s)}u\big|_{(x,s)}. \tag{7}$$

Since $\omega/\rho\cdot\nabla$ has no time-term, (6) reduces to the Cauchy formula (1). The formula (7) does not seem to be of any particular use.

## 3. A LEMMA ABOUT LIE DERIVATIVES

Let $M$ be a semi-Riemannian manifold with (indefinite) metric tensor $g$. Thus $g$ is a second rank covariant tensor defining a nondegenerate bilinear pairing of each tangent space. We give $M$ the unique torsion-free metric connnection determined by $g$ and denote the associated covariant differentiation operator by $\nabla$. If $v$ is a vector field on $M$, then $L_v$ denotes the Lie differentiation with respect to $v$ and $\nabla_v$ is covariant differentiation with respect to $v$. Following Kobayashi and Nomizu,[9] we define the operator $A_v = L_v - \nabla_v$ for $v$ a vector field on $M$. All functions and tensor fields are assumed smooth unless otherwise specified. For economy of terminology we define a *tensor derivation* to be any derivation of the algebra of tensor fields which commutes with all contractions. The tensor derivations themselves form a Lie algebra.[9] As $L_v$ and $\nabla_v$ are tensor derivations and $L_v f = \nabla_v f$ for any function $f$ we conclude:

(a) for any vector field $v$, the operator $A_v$ is a tensor derivation vanishing on functions.

Moreover, as our connection is torsion-free ($[u,v] = \nabla_u v - \nabla_v u$), we have

(b) if $u$ and $v$ are vector fields, then $A_u v = -\nabla_v u$. Combining (a) and (b) gives

(c) if $u$ and $v$ are vector fields and $\lambda$ is a 1-form, then $\langle A_u\lambda, v\rangle = \langle\lambda, \nabla_v u\rangle$, where $\langle\ ,\ \rangle$ denotes the pairing of 1-forms and vector fields. If $v$ is any vector field, we denote by $v^*$ the dual 1-form given by $\langle v^*, w\rangle = g(v,w)$ for each vector field $w$.

We now observe that if $u, v, w$ are all vector fields, then by (c)

$$\langle A_u v^*, w\rangle = g(v, \nabla_w u),$$

and hence as $\nabla g = 0$,

$$\begin{aligned}\langle d[g(u,v)], w\rangle &= \nabla_w[g(u,v)] = g(\nabla_w u, v) + g(u, \nabla_w v)\\ &= \langle A_u v^*, w\rangle + \langle A_v u^*, w\rangle,\end{aligned}$$

and thus

$$d[g(u,v)] = A_u v^* + A_v u^*.$$

Taking $u=v$ and applying the definition $A_u = L_u - \nabla_u$, we find[10]:

*Lemma*: For any vector field $u$,

$$L_u(u^*) = \nabla_u u^* + \tfrac{1}{2}d[g(u,u)].$$

[Notice that $\nabla_u(v^*) = (\nabla_u v)^*$ as $\nabla g = 0$, so that the notation $\nabla_u v^*$ is unambiguous.]

*Remark*: Each mixed second rank tensor field $s$ can be thought of as an endomorphism of the tangent bundle of $M$ and defines a unique tensor derivation vanishing on functions which we can denote by $D[s]$. With this notation we can view (b) as stating that for each vector field $u$,

$$A_u = -D[\nabla u].$$

A similar computation shows that if $f$ is a function, then

$$L_{fu} - fL_u = -D[(df)\otimes u]. \tag{8}$$

The lemma, even though merely a fact about vector fields, can be given a physical interpretation. We think of $u$ as being the velocity field of some "generalized" fluid motion on $M$, so $a = \nabla_u u$ is the acceleration and $E = \frac{1}{2}g(u,u)$ is half the "square of speed." With this notation we can write the lemma as

$$L_u(u^*) = a^* + dE.$$

For any vector field $v$, we define $\Omega_v = dv^*$ as the generalized vorticity of $v$. As exterior and Lie differentiation operators commute, and $d^2=0$, and we have the following fact

*Corollary*: $L_u(\Omega_u) = \Omega_a$.

Thus the voricity change along flow lines is directly related to vorticity of acceleration.

## 4. NONRELATIVISTIC FLOW OF IDEAL MOTION IN A RIEMANNIAN MANIFOLD

Let $S$ be a Riemannian manifold. For a discussion of the correct formulation of the Euler equations, we refer the reader to Serrin.[2] (The problem is that the integral conservation laws leading to the Euler equations cannot be expressed intrinsically.) See also Dunic,[11] Ebin-Marsden,[12] Marsden,[13] and Szeptycki,[14] where existence and uniqueness questions are treated. We accept as a hypothesis that the correct generalization of the Euler equations is

$$\nabla_u v = -\frac{1}{\rho}\nabla p \tag{9}$$

on the space–time manifold $M = S\times\mathbb{R}$. Here $v$ is the (time-dependent) Eulerian velocity field on $S$, interpreted as a vector field on $M$; $u = \partial/\partial t + v$ [see (3)], another field on $M$; and $p$ and $\rho$ are functions on $M$.

The metric tensor of $M$ is $g$ = metric tensor of $S \otimes dt \otimes dt$. Actually, the dual form of (9) is more convenient,

$$\nabla_u v^* = -dp/\rho. \tag{10}$$

Note that $\nabla_u(v^*) = \nabla_u(u^*)$, since $\nabla_u(dt) = 0$. Thus if the flow is barotropic, which we assume, the Euler equation reduces to the statement

$\nabla_u(u^*)$ *is exact.*

The corollary then implies that $L_u(du^*) = 0$.

*Proposition* 2: (Covariant form of the Cauchy–Lagrange formula.) Let $v$ be a time-dependent vector field on a Riemannian manifold $S$, satisfying the Euler equation. Suppose pressure and density to be functionally related. Let $\{F_t : t \in \mathbb{R}\}$ be the family of diffeomorphisms of $S$ generated by $v$, and define the vorticity be to the 2-form $\omega_t = dv_t^*$. Then

$$F_t^*(\omega_t) = \omega_0.$$

*Proof*: Define $\mu_t : S \to M$ by $u_t(x) = (x, t)$, and note

$$\mu_t^*(u^*) = v_t^*.$$

Let the flow $\{G_t\}$ be defined in terms of the family $\{F_t\}$ exactly as in (5), and note that

$$u \text{ is the infinitesimal generator of } \{G_t\}. \tag{11}$$

Also it is clear that $\mu_s \circ F_s = G_s \circ \mu_0$. Finally,

$$\begin{aligned}
F_t^*(\omega_t) &= F_t^*(dv_t^*) \\
&= F_t^* d(\mu_t^* u^*) \\
&= F_t^* \mu_t^* du^* \\
&= (\mu_t \circ F_t)^* du^* \\
&= (G_t \circ \mu_0)^* du^* \\
&= \mu_0^* G_t^*(du^*) \\
&= \mu_0^* du^* \text{ by (10) and (11)} \\
&= dv_0^* \\
&= \omega_0.
\end{aligned}$$

*Remark*: For the incompressible case, see Ref. 5, p. 86.

## 5. ISOMETRIC MOTION OF A RELATIVISTIC FLUID

In this section we assume that $u$ is the generalization of the four-velocity field of a relativistic fluid, so we suppose that

$$g(u,u) = 1. \tag{12}$$

First let $\xi = fu$ where $f$ is a function. By (12), $g(u,u)$ is constant so that $g(u, \nabla_u u) = 0$, and we see velocity is perpendicular to acceleration. Thus on taking the inner product of $u$ with $\nabla_u \xi = (\nabla_u f)u + f\nabla_u u$, we find

$$g(\nabla_u \xi, u) = \nabla_u f. \tag{13}$$

On the other hand, applying the lemma of Sec. 3 to $\xi$ [and using $g(\xi,\xi) = f^2$] gives

$$L_\xi(\xi^*) = f(\nabla_u f)u^* + f^2 \nabla_u u^* + f df. \tag{14}$$

With these facts in mind, we can give an easy proof of the conservation of vorticity for isometric motion.

We first recall some basic facts. A vector field $k$ is called a Killing vector field provided $L_k(g) = 0$. As $\nabla g = 0$ we note that $L_k(g) = 0$ is equivalent to $A_k g = 0$. A simple calculation shows that for any vector field $v$ we have $L_v g = \mathrm{Sym}(\nabla v^*)$, where Sym denotes the symmetrizing operator. Consequently

$$\begin{aligned}
&k \text{ is a Killing vector field} \\
&\Longleftrightarrow \mathrm{Sym}(\nabla k^*) = 0 \\
&\Longleftrightarrow \text{for every vector field } w,\ g(\nabla_w k, w) = 0.
\end{aligned} \tag{15}$$

A relativistic fluid motion is called *isometric* if there is a nonvanishing Killing vector field parallel to 4-velocity. We thus assume that $\xi$ is a Killing vector field. Combining (13) and (15) gives

$$\nabla_u f = 0. \tag{16}$$

Now for any vector field $v$, $v^*$ is a contraction on $g \otimes v$, hence as $L_\xi$ is a tensor derivation, $L_\xi(v^*) = L_\xi(v)^*$ and therefore the left side of (14) vanishes. Thus in view of (16), (14) reduces to $f^2 \nabla_u u^* = -f df$ so that if $\xi$ is nowhere vanishing (as we can assume, then $f$ is everywhere positive without loss of generality) we can write

$$\nabla_u u^* = -d(\log f). \tag{17}$$

But this means the lemma of Sec. 3 (and corollary) applies to say

$$L_u(\Omega_u) = 0, \tag{18}$$

where as before, $\Omega_u = du^*$ is the generalized vorticity. Physically, (18) states that vorticity is conserved in isometric motion. Of course the lemma says (18) would hold whenever acceleration is potential, but this is really no more general, becuase in this latter case we arrive immediately at (17) and then work our way back[8] to $L_\xi g = 0$. *Thus the condition of isometric motion is equivalent to the condition of acceleration being a potential.* According to the corollary to the lemma of Sec. 3 and the Poincaré lemma for differential forms, vorticity can be conserved only when the motion is at least *locally* isometric. From (8) we have $L_\xi u = -(\nabla_u f)u$, which vanishes by (16), so that both $u$ and $g$ belong to $\mathrm{Ker}L_\xi$. As $L_\xi$ is a tensor derivation commuting with $d$, it follows that any tensor constructed out of $u$ and $g$ via operations of $\otimes$, $+$, scalar (constant) multiplication, $d$, contraction, Hodge star, will again belong to $\mathrm{Ker}L_\xi$. Such tensors include the physical tensors of interest in the case where $\dim M = 4$ and $g$ has signature $(+---)$. In particular, the comoving vorticity 2-form $\omega$, the expansion, and the comoving vorticity *vector* all belong to $\mathrm{Ker}L_\xi$ by our preceding remarks, and we therefore obtain as special case the results of Mason[6] and Ciubotariu.[7] However, as (8) makes quite clear, $\mathrm{Ker}L_\xi$ and $\mathrm{Ker}L_u$ are two different things: "shearing" due to $1/f$ can prevent some of these tensors from being conserved in an isometric motion.

A more general condition than the condition of isometric motion is the condition of Born rigidity. As (12) holds, the tensor

$$P_u = g - u^* \otimes u^*$$

defines (in mixed form) a field of projection operators onto the orthogonal complement of $u$ in $\tau M$, the tangent bundle of $M$. The motion is said to be *Born rigid* if $L_u(P_u)=0$. With $h=1/f$, a simple calculation shows

$$L_u(g)=hL_\xi(g)+2\,\mathrm{Sym}(\xi^*\otimes dh). \tag{19}$$

Hence if $\xi$ is a nonvanishing Killing vector field, then as $L_\xi g=0$,

$$\begin{aligned}L_u(g)&=2\,\mathrm{Sym}(\xi^*\otimes dh)\\&=2f\,\mathrm{Sym}\left[u^*\otimes\left(-\frac{df}{f^2}\right)\right]\\&=2\,\mathrm{Sym}[u^*\otimes\nabla_u u^*],\end{aligned}$$

where in the last step (17) has been applied. But by the lemma of Sec. 3 and (12), we have $\nabla_u u^*=L_u u^*$, hence $L_u(g)=L_u(u^*\otimes u^*)$ and therefore $L_u(P_u)=0$, which means the motion is Born rigid. As examples exist of non-isometric Born rigid motions, and in view of our preceding remarks (the corollary to the lemma) we conclude that, in general, vorticity is not conserved in Born rigid motion.

## ACKNOWLEDGMENT

The first author would like to acknowledge support under N.S.F. Grant MCS-76-05947A01.

[1] $\{F_t\}$ is *not* a flow: $F_{t+s}\neq F_t\circ F_s$, because the velocity field is time dependent.

[2] J. Serrin, *Handbuch der Physik*, edited by S. Flugge (Springer-Verlag, 1959), Vol. 8, Part 1.
[3] C. Truesdell, *The Kinematics of Vorticity* (Indiana U. P., Indiana, 1954).
[4] T.J.R. Hughes and J.E. Marsden, *A Short Course in Fluid Mechanics* (Publish or Perish, Boston, 1976).
[5] J.E. Marsden, *Applications of Global Analysis in Mathematical Physics* (Publish or Perish, Boston, 1974).
[6] D.P. Mason, J. Phys. A: Math. Gen. **8**, L58 (1975).
[7] C.D. Ciubotariu, Phys. Lett. A **40**, 369 (1972).
[8] A. Trautman, *Lectures on General Relativity*, edited by A. Trautman, F.A.E. Pirani, and H. Bondi (Prentice-Hall, Englewood Cliffs, New Jersey, 1965).
[9] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1963), p. 33, Corollary 3.7.
[10] This formula is stated but not proved on p. 85 of Ref. 4.
[11] M.D. Dunic, Publ. Inst. Math. Beograd., Ser. 6, **20**, 131 (1966).
[12] D.G. Ebin and J. Marsden, Ann. Math. (N.Y.) **92**, 102 (1970).
[13] J.E. Marsden, Comm. P.D.E. **1**, 215 (1976).
[14] P. Szeptycki, Bull. Acad. Polon, Sci. Sér. Math. Astron. Phys. **21**, 335 (1973).

# Spatial conformal flatness in homogeneous and inhomogeneous cosmologies

Alan Spero[a)]

*Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada and Department of Physics,* [b)] *Northern Michigan University, Marquette, Michigan 49855*

Duane A. Szafron[c)]

*Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*
(Received 25 January 1978)

In the past few years there has been a growing interest in cosmological models which are not spatially homogeneous. The assumption of spatial homogeneity simplifies the Einstein equations to ordinary differential equations. If the assumption of spatial homogeneity is relaxed, some other symmetries are needed to make the Einstein equations mathematically tractable. The recently discovered solutions of Szekeres have been found to possess an interesting type of symmetry: The three spaces orthogonal to the fluid flow are conformally flat. Herein, we prove a theorem restricting the possible inhomogeneous cosmologies with conformally flat 3-surfaces. We determine which spatially homogeneous models admit conformally flat 3-surfaces. This information, although interesting in its own right, will serve as a guide in determining those spatially homogeneous models that may be generalized by retaining spatial conformal flatness but relaxing the condition of spatial homogeneity.

## I. INTRODUCTION

Spatially homogeneous cosmological solutions to the Einstein equations have been extensively studied in the past fifteen years, the work being facilitated by Bianchi's original classification scheme for Lie groups and its various modifications.[1] This scheme has enabled researchers to study the dynamics of restricted classes of models individually. Lately, interest has turned to inhomogeneous models, where the same powerful group theoretic techniques are no longer applicable.

Szekeres[2] has recently found a family of exact solutions to the Einstein equations with a pressure-free matter source ("dust"), some of whose members may be interpreted as inhomogeneous cosmological models.[3] This family was later generalized by Szafron and Wainwright[4,5] to include perfect-fluid sources. We will henceforth refer to all these solutions, for $p=0$ or $p\neq 0$, as the "Szekeres solutions."

The Szekeres solutions have conformally flat, comoving, spacelike hypersurfaces.[6,7] These models suggest that the assumption of conformally flat space sections ("spatial conformal flatness") may provide the symmetries necessary to solve the Einstein equations when the condition of spatial homogeneity is relaxed.

Herein, we explore some of the restrictions imposed by spatial conformal flatness. In Sec. 2 we study that subclass of the Szekeres solutions which obey the barotropic equation of state $p=p(\mu)$, where $\mu$ is the relativistic energy density and $p$ is the pressure of the perfect fluid. In Sec. 3 we determine which spatially homogeneous models admit conformally flat slices. In this way all possible limiting cases of inhomogeneous cosmologies with conformally flat slices are exhibited. This should facilitate future attempts at generalizing the Szekeres solutions. Section 4 contains our conclusions.

Throughout, we use the orthonormal tetrad technique, as elucidated by MacCallum.[8] Our notation also follows that of MacCallum, and is, briefly:

Space–time metric signature: $+2$.

Orthonormal tetrad: $\{\mathbf{e}_a\}$, $a=0$ to 3.

Tetrad components are denoted by letters at the beginning of the Latin and Greek alphabets, coordinate components by letters near the end of the alphabets. Latin letters range from 0 to 3, while Greek letters range from 1 to 3.

Partial derivatives with respect to coordinates: $\partial B/\partial x\equiv B_x$.

Derivatives along tetrad vectors: $\mathbf{e}_a\cdot\boldsymbol{\partial}\equiv\partial_a$.

## II. THE SZEKERES SPACE-TIMES

The Szekeres space–times are solutions to the Einstein equations,

$$R_{ab}-\tfrac{1}{2}g_{ab}R=(\mu+p)u_au_b+pg_{ab}, \tag{2.1}$$

for the metrics of the form[1]

$$ds^2=-dt^2+e^{2A}dx^2+e^{2B}(dy^2+dz^2) \tag{2.2}$$

with the fluid flow vector

$$\mathbf{u}=\frac{\partial}{\partial t}. \tag{2.3}$$

(Note that the metric signature and notation used here differs from that of Szekeres[2] and Szafron.[5]) They divide naturally into two classes.[5] (In fact, this is a coordinate-independent division.[7])

Class I. $B_x\neq 0$ (2.4)

and $A$ and $B$ have the form

---

[a)]Work supported by an N. M. U. Faculty Research Grant.
[b)]Permanent address.
[c)]Work supported by a postgraduate scholarship from the National Research Council of Canada.

 0022-2488/78/1907-1536$1.00  

$$A = \ln[\eta(\phi/\eta)_x], \tag{2.5}$$

$$B = \ln[\phi/\eta], \tag{2.6}$$

where

$$\eta = \eta(x, y, z), \quad \phi = \phi(t, x). \tag{2.7}$$

Class II. $B_x = 0$ (2.8)

$$A = \ln[\lambda + \phi\zeta], \tag{2.9}$$

$$B = \ln[\phi/\eta], \tag{2.10}$$

where

$$\lambda = \lambda(t, x), \quad \phi = \phi(t) \tag{2.11}$$

$$\eta = \tfrac{1}{2}[1 + k(y^2 + z^2)], \quad k = \pm 1 \text{ or } 0, \tag{2.12}$$

$$\zeta = [\alpha(x)(y^2 + z^2) + \beta(x)y + \gamma(x)z + \delta(x)]\eta. \tag{2.13}$$

The functions $\alpha$, $\beta$, $\gamma$, and $\delta$ are arbitrary. Szafron[5] gives an algorithm for determining $\phi$ in Class I, and $\phi$ and $\lambda$ in Class II once the pressure $p$ is prescribed as a function of time; however, we shall not need more information than is given here.

The Szekeres space–times contain a perfect fluid whose flow is irrotational, geodesic, and normal to the hypersurfaces $\{t = \text{constant}\}$. Moreover, the flow's expansion tensor, $\theta_{\alpha\beta}$, has two equal eigenvalues:

$$\theta_2 = \theta_3 \tag{2.14}$$

as does the 3-Ricci tensor of the spacelike hypersurfaces. The co-moving, spacelike hypersurfaces $\{t = \text{constant}\}$ are conformally flat, as was shown by Berger *et al.*[6] for the Class I Szekeres solutions, and by Wainwright and Szafron (private communication) for all the Szekeres solutions. In fact, as Collins and Szafron[7] have shown, one may characterize the generalized Szekeres solutions by the above properties. We state this as a theorem, whose proof may be found in Collins and Szafron[7]:

*Theorem* 2.1: A space–time that contains a perfect fluid and satisfies:

(i) the fluid flow is geodesic and hypersurface-orthogonal ($\dot{\mathbf{u}} = \mathbf{0} = \boldsymbol{\omega}$);

(ii) the hypersurfaces to which the fluid flow is normal are conformally flat;

(iii) both the Ricci 3-tensor of the hypersurfaces and the fluid's expansion tensor have two equal eigenvalues,

is a solution to the Einstein equations if and only if it has the Szekeres line element.

The matter content of these space–times obeys an equation of state that is, in general, an unusual one; although the energy density may exhibit spatial variations, the pressure may not [$p = p(t)$ always]. In what follows we restrict our attention to that subclass of solutions satisfying a barotropic equation of state:

$$p = p(\mu). \tag{2.15}$$

Within this subclass, both $\mu$ and $p$ will be functions of time alone. The natural question which arises is: Do any spatially-inhomogeneous space–times belong to this subclass? [Wainright[9] has shown that all Szekeres solutions with $p = p(\mu)$ are locally rotationally symmetric (for definition see Ref. 10).] In fact, there are none, as we demonstrate below.

*Lemma* 2.2: In those Szekeres space–times with $\mu = \mu(t)$, both $\theta_{\alpha\beta}$ and $R^*_{\alpha\beta}$ are functions of time alone. [Note that in the Szekeres space–times $p = p(t)$. The assumption $p = p(\mu)$, $dp/d\mu \neq 0$ is a subcase of $\mu = \mu(t)$.]

*Proof*: We have

$$\partial_\alpha p = 0 = \partial_\alpha \mu. \tag{2.16}$$

Equation (A13) then implies

$$\partial_\alpha \theta_1 = -2\partial_\alpha \theta_2. \tag{2.17}$$

Taking $\partial_\alpha$ of (A7) and using (2.17) and the commutation relations (A2) gives

$$(\theta_1 - \theta_2)\,\partial_\alpha \theta_1 = 0; \tag{2.18}$$

whence we have

$$\partial_\alpha \theta_1 = \partial_\alpha \theta_2 = 0, \tag{2.19}$$

whether $\theta_1 = \theta_2$ or not. The equations obtained by taking $\partial_\alpha$ of (A11) together with (2.19) give

$$\partial_\alpha R^*_{11} = \partial_\alpha R^*_{22} = 0, \tag{2.20}$$

thus completing the proof.

*Theorem* 2.3: Any Szekeres solution with $\mu = \mu(t)$ is spatially homogeneous and is either: Robertson–Walker, of the Kantowski–Sachs type (see Sec. 3), or admits a group of Bianchi–Behr types I or $\mathrm{VI}_{-1}$.

*Proof*: By the lemma: $\partial_\alpha \theta_\beta = 0.$ (2.21)

We consider the two classes of Szekeres metrics separately:

$$\text{Class I:} \quad \theta_1(t) = B_t = \phi_t/\phi, \tag{2.22}$$

implying that $\phi$ has the form

$$\phi(t, x) = E(x) \exp\left[\int^t \theta_1\, dt\right]. \tag{2.23}$$

But then

$$\theta_3(t) = A_t = \phi_t/\phi = \theta_1, \tag{2.24}$$

so the expansion is isotropic, i.e., shear-free. Any perfect-fluid space–time in which the flow is geodesic, irrotational, and shear-free is necessarily Robertson–Walker,[11] (p. 135); thus the theorem is proven for Class I.

Class II: These are characterized by $B_x = 0$, so $a_1 = 0$. With (2.21), Eqs. (A9) and (A10) reduce to

$$(\theta_2 - \theta_1)\,(a_2 - n_{31}) = 0 \tag{2.25}$$

and

$$(\theta_2 - \theta_1)(a_3 + n_{12}) = 0. \tag{2.26}$$

If $\theta_1 = \theta_2$ these space–times must be Robertson–Walker by the previous argument; so assume $\theta_1 \neq \theta_2$. Then

$$0 = a_2 - n_{31} = -A_y e^{-B} \tag{2.27}$$

$$0 = a_3 + n_{12} = -A_z e^{-B}, \tag{2.28}$$

implying that $\zeta = 0$ in (2.13). Therefore,

$$\theta_3(t) = A_t = \lambda_t/\lambda; \tag{2.29}$$

so we must have

$$\lambda(t,x)=F(x)\exp[\int^t \theta_3\,dt], \tag{2.30}$$

where $F(x)$ is arbitrary. The metric now becomes

$$ds^2=-dt^2+\exp\left[2\int^t \theta_3\,dt\right]F^2(x)\,dx^2$$

$$+\frac{\phi^2(t)}{\eta^2(y,z)}\,(dy^2+dz^2). \tag{2.31}$$

By suitably redefining the $x$ coordinate we may set $F(x)=1$. This metric is spatially homogeneous and has the Killing vectors:

$$\boldsymbol{\xi}_1=[\tfrac{1}{2}k(z^2-y^2)+1]\frac{\partial}{\partial z}+kyz\frac{\partial}{\partial y}, \tag{2.32}$$

$$\boldsymbol{\xi}_2=kyz\frac{\partial}{\partial z}+\frac{1}{2}[k(y^2-z^2)+1]\frac{\partial}{\partial y}, \tag{2.33}$$

$$\boldsymbol{\xi}_3=y\frac{\partial}{\partial z}-z\frac{\partial}{\partial y}, \tag{2.34}$$

$$\boldsymbol{\xi}_4=\frac{\partial}{\partial x}. \tag{2.35}$$

We need only determine the possible Lie algebras. The commutators of the Killing vectors are:

$$[\boldsymbol{\xi}_1,\boldsymbol{\xi}_2]=k\boldsymbol{\xi}_3, \tag{2.36}$$

$$[\boldsymbol{\xi}_2,\boldsymbol{\xi}_3]=\boldsymbol{\xi}_1, \tag{2.37}$$

$$[\boldsymbol{\xi}_3,\boldsymbol{\xi}_1]=\boldsymbol{\xi}_2, \tag{2.38}$$

$$[\boldsymbol{\xi}_4,\boldsymbol{\xi}_i]=0\quad(i=1,2,3). \tag{2.39}$$

The following cases obtain:

(i) $k=0$, $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$, and $\boldsymbol{\xi}_4$ form a simply-transitive group of Bianchi–Behr type I;

(ii) $k=+1$, $\boldsymbol{\xi}_i$, and $\boldsymbol{\xi}_4$ form a multiply-transitive group with no simply-transitive subgroups. This is the Kantowski–Sachs case.[12,13]

(iii) $k=-1$, $\boldsymbol{\xi}_1+\boldsymbol{\xi}_3$, $\boldsymbol{\xi}_2$, and $\boldsymbol{\xi}_4$ form a simply-transitive group of Bianchi–Behr type $\mathrm{VI}_{-1}$. [*Note*: We find that the Kantowski–Sachs metrics can arise only when $k=+1$. Szekeres[2] has stated that the $k=\pm 1$ cases generalize those solutions given in Kantowski and Sachs.[13] In fact the $k=-1$ solutions are only the Bianchi–Behr type $\mathrm{VI}_{-1}$ solutions found by Kantowski and Sachs (see Ref. 12).] This completes the proof. In special cases other groups may be admitted as well as the ones indicated. For example, a space–time may admit a Bianchi–Behr type $\mathrm{VII}_0$ group in addition to its Bianchi–Behr type I group.

By combining the preceding two theorems, we obtain a theorem similar to that of Berger *et al.*[6] concerning the nature of possible inhomogeneous exact solutions to the Einstein equations with perfect fluid sources.

*Theorem* 2.4: Given Einstein's equations with a zero cosmological constant, a perfect fluid source, and:

(i) irrotational, geodesic flow whose expansion tensor has two equal eigenvalues;

(ii) conformally flat, comoving space slices, whose Ricci 3-tensor has two equal eigenvalues;

(iii) the equation of state $\mu=\mu(t)$;

the only solutions are spatially homogeneous ones that are either: Robertson–Walker, Kanotwski–Sachs, or admit a Bianchi–Behr type I or $\mathrm{VI}_{-1}$ group. Berger *et al.*[6] replace (i) with irrotational flow and spherical symmetry about a regular spatial origin; require flat comoving sections rather than (ii); and assume $p=p(\mu)$ with a nonzero speed of sound. With their assumptions, only the Robertson–Walker solutions arise.

Theorems of this type are special cases of the more general conjecture: hypersurface orthogonal space–times ($\boldsymbol{\omega}=0$) with geodesic flow ($\dot{\mathbf{u}}=0$) and $\mu=\mu(t)$, $p=p(t)$ are spatially homogeneous. We know of no counterexamples to this conjecture, however we know of no proof either.

## III. CONFORMALLY-FLAT SPATIALLY-HOMOGENEOUS SPACE-TIMES

We shall consider space–times satisfying the two conditions:

(A) There exists a group of isometries $G_r$ whose orbits in some open set of the space–time are spacelike hypersurfaces with either

(i) a subgroup $G_3$ of $G_r$, which acts simply transitively, or

(ii) no such subgroup $G_3$, and

(B) The surfaces of homogeneity are conformally flat.

If $r>3$, it can be shown[14] that $r=4$ or $r=6$. Those space–times with $r=6$ are the Robertson–Walker models (see Ref. 12), therefore they satisfy (Ai). Space–times satisfying (Aii) have $r=4$ and a subgroup $G_3$ whose orbits are two-dimensional and of constant positive curvature (see Refs. 12 or 15). We will refer to these space–times as Kantowski–Sachs since Kantowski and Sachs[13] have studied solutions to the Einstein field equations with dust matter content admitting such a group.

Let us first consider space–times satisfying (Ai). We choose an orthonormal tetrad $\{\mathbf{e}_a\}$ so that $\mathbf{e}_0$ is orthogonal to the hypersurfaces and

$$n_{\alpha\beta}=\mathrm{diag}(n_1,n_2,n_3),\quad a^\beta=(a,0,0),\quad an_1=0, \tag{3.1}$$

where $n^{\alpha\beta}$ and $a^\beta$ are given in terms of the rotation coefficients $\Gamma_{abc}=\mathbf{e}_a\cdot\nabla_b\mathbf{e}_c$ by

$$n^{\alpha\beta}=\tfrac{1}{2}\Gamma^{(\alpha}{}_{\nu\sigma}\epsilon^{\beta)\nu\sigma},\quad a_\beta=-\tfrac{1}{2}\Gamma^{\alpha}{}_{\alpha\beta}. \tag{3.2}$$

The proof that such a tetrad can be found is given by Ellis and MacCallum.[1] Although their paper assumes a perfect fluid energy momentum tensor, this assumption is not used in deriving the tetrad. Notice, however, that their fluid flow vector $\mathbf{u}$ must be replaced by our tetrad vector $\mathbf{e}_0$ which at this stage is independent of any fluid. Their fluid quantities, $\theta_{\alpha\beta}$, $\omega_{\alpha\beta}$, and $\Omega_\alpha$, here just relate to the timelike congruence defined by $\mathbf{e}_0$. The commutators $\{\mathbf{e}_\alpha\}$ are given by

$$\begin{aligned}&[\mathbf{e}_1,\mathbf{e}_2]=a\mathbf{e}_2+n_3\mathbf{e}_3,\\&[\mathbf{e}_2,\mathbf{e}_3]=n_1\mathbf{e}_1,\\&[\mathbf{e}_3,\mathbf{e}_1]=n_2\mathbf{e}_2-a\mathbf{e}_3.\end{aligned} \tag{3.3}$$

The group $G_3$ may be classified by placing $n_1$, $n_2$, and $n_3$ into one of the canonical forms of Table I.

TABLE I. Classification of space–times satisfying (Ai), given by Ellis and MacCallum,[1] due originally to Bianchi, and modified by Behr. Here, $h \equiv a^2/n_2n_3$.

| Group class | Bianchi–Behr group type | $a$ | $n_1$ | $n_2$ | $n_3$ | Bianchi type |
|---|---|---|---|---|---|---|
| A | I | 0 | 0 | 0 | 0 | I |
| | II | 0 | + | 0 | 0 | II |
| | $\mathrm{VII}_0$ | 0 | + | + | 0 | VII |
| | $\mathrm{VI}_0$ | 0 | + | − | 0 | VI |
| | IX | 0 | + | + | + | IX |
| | VIII | 0 | + | + | − | VIII |
| B | V | + | 0 | 0 | 0 | V |
| | IV | + | 0 | 0 | + | IV |
| | $\mathrm{VII}_h$ | + | 0 | + | + | VII |
| | $\mathrm{VI}_h$ | + | 0 | + | − | VI (III if $h=-1$) |

Eisenhart[16] has shown that a 3-space is conformly flat if and only if

$$0=R^*_{\alpha\beta\lambda}\equiv R^*_{\alpha\beta|\lambda}-R^*_{\alpha\lambda|\beta}+\tfrac{1}{4}(g^*_{\alpha\lambda}R^*{}_{|\beta}-g^*_{\alpha\beta}R^*{}_{|\lambda}), \qquad (3.4)$$

where * denotes a quantity in the 3-space and | is the covariant derivative in the 3-space. We expand (3.4) in the tetrad (3.1) and notethat $\partial_\alpha R^*{}_{\beta\lambda}=0$ since the 3-spaces are homogeneous,

$$0=-\Gamma^\mu_{\lambda\alpha}R^*_{\mu\beta}-\Gamma^\mu_{\lambda\beta}R^*_{\mu\alpha}+\Gamma^\mu_{\beta\alpha}R^*_{\mu\lambda}+\Gamma^\mu_{\beta\lambda}R^*_{\mu\alpha}. \qquad (3.5)$$

From Ellis and MacCallum[1] the 3-Ricci tensor of the homogeneous hypersurfaces is

$$R^*_{\alpha\beta}=-2\epsilon^{\sigma\lambda}{}_{(\alpha}n_{\beta)\sigma}a_\lambda+2n_{\alpha\delta}n^\delta{}_\beta+nn_{\alpha\beta}$$
$$+\delta_{\alpha\beta}(2a^2+n_{\sigma\lambda}n^{\sigma\lambda}-\tfrac{1}{2}n^2). \qquad (3.6)$$

In Appendix B we have written out: $\Gamma_{\alpha\beta\lambda}$, $R^*{}_{\alpha\beta}$, and Eqs. (3.5) in the tetrad (3.1).

*Lemma* 3.1: Space–times satisfy (Ai) and (B) if and only if they admit: a group of Bianchi–Behr type I, a type $\mathrm{VII}_0$ group with $n_1=n_2$, a type IX group with $n_1=n_2=n_3$, a group of type V, a type $\mathrm{VII}_h$ group with $n_2=n_3$ or a type $\mathrm{VI}_{-1}$ group with $a=n_2=-n_3$.

*Proof*: The conditions for conformally flat surfaces of homogeneity (3.5) in the tetrad (3.1) are equations (B3)–(B6). First we consider solutions admitting groups from class A ($a=0$). Equations (B5) and (B6) are identically satisfied. Equation (B3) minus (B4) yields

$$(n_1-n_2)(3n_1^2+2n_1n_2+3n_2^2-2n_3n_1-2n_3n_2-n_3^2)=0. \qquad (3.7)$$

Twice (B3) plus (B4) yields

$$(n_1-n_3)(3n_1{}^2+2n_1n_3+3n_3{}^2-2n_2n_1-2n_2n_3-n_2{}^2)=0. \qquad (3.8)$$

The only simultaneous solutions of (3.7) and (3.8) are

$$n_1=n_2=n_3;\quad n_1=n_2, n_3=0;\quad n_1=n_3, n_2=0;$$
$$\text{or } n_2=n_3, n_1=0. \qquad (3.9)$$

Comparing these possiblities with the canonical forms of the class A groups in Table I, we see that the Bianchi–Behr group types admitted are: type I, those type $\mathrm{VII}_0$ with $n_1=n_2$, and those type IX with $n_1=n_2=n_3$.

Now we consider solutions admitting groups from class B ($a\neq0$, $n_1=0$). Equation (B5) yields

$$(n_2-n_3)(n_2+n_3)=0. \qquad (3.10)$$

Equations (B3) and (B6) are then identically satisfied. Equation (B4) implies

$$(n_2-n_3)(2a^2-2n_2{}^2-n_3n_2-n_3{}^2)=0. \qquad (3.11)$$

The only simultaneous solutions of (3.10) and (3.11) are

$$n_2=n_3;\quad \text{or } n_2=-n_3,\quad a=\pm n_2. \qquad (3.12)$$

Comparing these possibilities with the canonical forms of the class B groups in Table I, we see that the Bianchi–Behr group types admitted are: type V, those type $\mathrm{VII}_h$ with $n_2=n_3$, and those type $\mathrm{VI}_{-1}$ with $a=n_2=-n_3$. Q.E.D.

*Lemma* 3.2: All space–times satisfying (Aii) satisfy (B).

*Proof*: Kantowski[12] (c.f. Ref. 15) has shown that we may write the metric in the form

$$ds^2=-dt^2+X^2(t)\,dr^2+Y^2(t)(d\theta^2+\sin^2\theta\,d\phi^2), \qquad (3.13)$$

where the hypersurfaces of homogeneity are given by $\{t=\text{constant}\}$. It is clear, after defining a new coordinate $\rho\equiv\exp(Xr/Y)$, that the hypersurfaces of homogeneity are conformally flat. We can combine Lemmas 3.1 and 3.2 into the following theorem:

*Theorem* 3.3: Space–times subject to (A) satisfy (B) if and only if they are Kantowski–Sachs or admit one of the following Bianchi–Behr groups: type I, type $\mathrm{VII}_0$ with $n_1=n_2$, type IX with $n_1=n_2=n_3$, type V, type $\mathrm{VII}_h$ with $n_2=n_3$ or type $\mathrm{VI}_{-1}$ with $a=n_2=-n_3$.

The results presented so far have been purely geometric. We shall now in addition demand that the space–times satisfy

(C) the Einstein field equations:

$$R_{ab}-\tfrac{1}{2}Rg_{ab}=T_{ab}, \qquad (3.14)$$

where the energy–momentum tensor is that of a perfect fluid

$$T_{ab}=\mu u_a u_b+p(g_{ab}+u_a u_b),$$
$$u_a u^a=-1,\quad \mu>0,\quad p\geq0. \qquad (3.15)$$

Later, we shall impose the additional restriction

(D) The fluid-flow vector $\mathbf{u}$ is orthogonal to the hypersurfaces of homogeneity, i.e., $\mathbf{u}=\mathbf{e}_0$.

Space–times not satisfying (D) are said to be "tilted.[17]"

*Theorem* 3.4: All space–times satisfying (A), (B), and (C) which do *not* satisfy (D) must admit a group belonging to one of the following types:

(i) $\mathrm{VII}_0$ with $n_1=n_2$ (these are precisely the solutions found by Demianski and Grischuk[18]),

(ii) V [if the space is locally rotationally symmetric (LRS), it will also admit groups of type $\mathrm{VII}_h$ in this case],

TABLE II. The 3-Ricci tensor for space—times subject to (Ai) and (B) in the tetrad (3.1).

| Bianchi—Behr group type | 3-Ricci tensor |
|---|---|
| I, $VII_0$ | $R^*_{\alpha\beta}=0$ |
| IX | $R^*_{11}=R^*_{22}=R^*_{33}=\frac{1}{2}n_1^2>0$ |
| V, $VII_h$ | $R^*_{11}=R^*_{22}=R^*_{33}=-2a^2<0$ |
| $VI_{-1}$ | $R^*_{11}=-4a^2,\ R^*_{22}=-2a^2,$ $R^*_{33}=-2a^2,\ R^*_{13}=2a^2$ |

(iii) $VI_{-1}$ with $a=n_2=-n_3$.

*Proof*: Theorem 3.3 implies that these solutions must be Kantowski—Sachs or admit one of the following Bianchi—Behr groups: I, $VII_0$, IX, V, $VII_h$, or $VI_{-1}$ with suitable restrictions. Kantowski[12] has shown that the Kantowski—Sachs solutions *must* satisfy (D). The Ricci 3-tensors of the allowed Bianchi—Behr type solutions appear in Table II. Notice that solutions admitting groups of Bianchi—Behr types I, $VII_0$, IX, $VII_h$, and V have hypersurfaces of homogeneity with isotropic $R^*_{\alpha\beta}$. In a 3-space, an isotropic Ricci tensor is equivalent to constant curvature.[16] From Theorem 4.2 of Ref. 17 we see that such solutions are "tilted" [i.e., do not satisfy (D)] only if they admit a group of type V or are Demianski—Grischuk solutions. The Demianski—Grischuk solutions admit a type $VII_0$ group.[17] The LRS type V solutions also admit a one-parameter family of type $VII_h$ groups.[1] Q.E.D.

*Lemma* 3.5: Any space—time that satisfies (A), (B), (C), and (D), and which admits a group of type IX is Robertson—Walker.

*Proof*: From Theorem 3.3 we see that $n_1=n_2=n_3\neq 0$ so the Jacobi identities of the tetrad [Eqs. (2.12) in Ref. 1] imply $\theta_1=\theta_2=\theta_3$. Q.E.D.

*Lemma* 3.6: Any space—time satisfying (A), (B), (C), and (D) which admits a type $VII_0$, $VII_h$, or $VI_{-1}$ group is LRS.

*Proof*: From Theorem 3.3 we see that in the type $VII_0$ case $n_1=n_2$, in the type $VII_h$ case $n_2=n_3$ and in the type $VI_{-1}$ case $n_2=-n_3$. Again Eqs. (2.12) of Ellis and MacCallum[1] imply $\theta_1=\theta_2$, $\theta_2=\theta_3$ and $\theta_2=\theta_3$, respectively. From Ref. 1 then, all cases are LRS. Q.E.D.

Note that the Kantowski—Sachs solutions are LRS since they have a nontrivial isotropy subgroup. Ellis and MacCallum[1] have shown that LRS solutions which satisfy (A), (C), and (D): admit a group of Type I if and only if they admit a group of type $VII_0$; are Robertson—Walker if they admit a group of type $VII_h$ or V. We summarize these results in the following theorem.

*Theorem* 3.7: Any space—time that satisfies (A), (B), (C), and (D), and which is *not* also LRS admits a group of type I or type V.

Table III summarizes the results of Sec. 3.

## IV. DISCUSSION

A study of the Szekeres solutions suggests that the assumption of conformally flat, spacelike slices may be a fruitful one when searching for inhomogeneous cosmological solutions to the Einstein equations. The results of Sec. 3, summarized in Table III, indicate that any family of inhomogeneous solutions with conformally flat slices can only contain spatially homogeneous solutions invariant under groups of type I, $VII_0$, IX, V, $VII_h$, or $VI_{-1}$, or be of the Kantowski—Sachs form. The Szekeres solutions extend a subset of the above listed spaces [containing the Robertson—Walker solutions, the Kantowski—Sachs solutions, and those LRS solutions admitting a group of type $VI_{-1}$ or both types I and $VII_0$ (Theorem 2.4)] to a family of inhomogeneous, perfect fluid solutions. The extended family has

TABLE III. A summary of results from Sec. 3.

| Space—times satisfying (A) | Those which satisfy (A) and (B) | Those which satisfy (A), (B), (C) but *not* (D) | Space—times that satisfy (A), (B), (C), (D) but are *not* LRS | Space—times that satisfy (A), (B), (C), and (D), and are LRS |
|---|---|---|---|---|
| I | all | none | some (those not admitting a $VII_0$ group) | some (those also admitting a $VII_0$ group) |
| $VII_0$ | $n_1=n_2$ | some (the Demianski—Grishchuk solutions, all of which have $n_1=n_2$) | none | all with $n_1=n_2$ (these also admit a I group) |
| IX | $n_1=n_2=n_3$ | none | none | all with $n_1=n_2=n_3$ (these are all R—W) |
| V | all | some | some (those that are not R—W) | some (only the R—W ones) |
| $VII_h$ | $n_2=n_3$ | some (only those that also admit a V group) | none | all with $n_2=n_3$ (these are all R—W) |
| $VI_{-1}$ (III) | $n_2=-n_3$ | some of those with $n_2=-n_3$ | none | all with $n_2=-n_3$ |
| K—S | all | none | none | all |
| II, $VI_0$, VII, IV, $VI_{h\neq -1}$ | none | none | none | none |

(a) irrotational, geodesic flow and an expansion tensor with two equal eigenvalues

(b) conformally flat, comoving hypersurfaces whose Ricci tensor has two equal eigenvalues

(c) a nonbarotropic equation of state.

Our results indicate that, if one wishes to generalize that subclass to perfect-fluid solutions satisfying (b) and having a barotropic equation of state, then some portion of property (a) must be discarded. If one discards either geodesic flow or two equal eigenvalues for the expansion tensor, then the extension must include other spatially-homogeneous models in addition to those listed in Theorem 2.4.

## ACKNOWLEDGMENTS

One of us (A. S.) wishes to thank the Department of Applied Mathematics at the University of Waterloo and, especially, Dr. C. B. Collins for their hospitality during his visit. We would also like to thank Dr. J. Wainwright for comments and suggestions.

## APPENDIX A: TETRAD FORM OF THE EINSTEIN EQUATIONS FOR THE SZEKERES METRIC

Given the Szekeres metric (2.2), choose the orthonormal tetrad:

$$\mathbf{e}_0=\frac{\partial}{\partial t},\quad \mathbf{e}_1=e^{-A}\frac{\partial}{\partial x},\quad \mathbf{e}_2=e^{-B}\frac{\partial}{\partial y},\quad \mathbf{e}_3=e^{-B}\frac{\partial}{\partial z}, \tag{A1}$$

whose commutators are

$$[\mathbf{e}_0,\mathbf{e}_\alpha]=-\theta_\alpha\mathbf{e}_\alpha \quad \text{(no sum)}, \tag{A2}$$

$$[\mathbf{e}_1,\mathbf{e}_2]=(n_{13}-a_2)\,\mathbf{e}_1+a_1\mathbf{e}_2, \tag{A3}$$

$$[\mathbf{e}_2,\mathbf{e}_3]=(n_{12}-a_3)\,\mathbf{e}_2+(n_{13}+a_2)\,\mathbf{e}_3, \tag{A4}$$

$$[\mathbf{e}_3,\mathbf{e}_1]=(n_{12}+a_3)\,\mathbf{e}_1-a_1\mathbf{e}_3, \tag{A5}$$

with

$$\theta_2=\theta_3. \tag{A6}$$

The Einstein field equations[8] are then

$$\partial_0(\theta_1+2\theta_2)+(\theta_1)^2+2(\theta_2)^2+\tfrac{1}{2}(\mu+3p)=0, \tag{A7}$$

$$\partial_1\theta_2+(\theta_1-\theta_2)\,a_1=0, \tag{A8}$$

$$\partial_2(\theta_1+\theta_2)+(\theta_2-\theta_1)(a_2-n_{13})=0, \tag{A9}$$

$$\partial_3(\theta_1+\theta_2)+(\theta_2-\theta_1)(a_3+n_{12})=0, \tag{A10}$$

$$\partial_0\theta_\alpha+\theta\theta_\alpha=-R_{\alpha\alpha}+\tfrac{1}{2}(\mu-p) \quad \text{(no sum)}, \tag{A11}$$

where

$$\theta\equiv\theta_1+\theta_2+\theta_3. \tag{A12}$$

The contracted Bianchi identities are

$$\partial_0\mu+(\mu+p)(\theta_1+2\theta_2)=0, \tag{A13}$$

$$\partial_\alpha p=0. \tag{A14}$$

## APPENDIX B

The nonvanishing rotation coefficients

$$\Gamma_{\alpha\beta\gamma}=\tfrac{1}{2}(\epsilon_{\beta\gamma\delta}n^\delta_\alpha+\epsilon_{\alpha\beta\delta}n_\gamma^\delta-\epsilon_{\gamma\alpha\delta}n_\beta^\delta+2\delta_{\gamma\beta}a_\alpha-2\delta_{\alpha\beta}a_\gamma)$$

in the tetrad (3.1) where

$$n_{\alpha\beta}=\mathrm{diag}(n_1,n_2,n_3),\quad a_\beta=(a,0,0)$$

are

$$\Gamma_{122}=-\Gamma_{221}=\Gamma_{133}=-\Gamma_{331}=a,\quad \Gamma_{123}=-\Gamma_{321}=\tfrac{1}{2}(n_1+n_3-n_2), \tag{B1a}$$

$$\Gamma_{132}=-\Gamma_{231}=\tfrac{1}{2}(-n_1-n_2+n_3),\quad \Gamma_{213}=-\Gamma_{312}=\tfrac{1}{2}(n_1-n_2-n_3). \tag{B1b}$$

The 3-Ricci tensor given by (3.6) in the tetrad (3.1) is

$$R^*_{11}=n_1(n_1-n_2-n_3)-N,\quad R^*_{22}=n_2(n_2-n_1-n_3)-N, \tag{B2a}$$

$$R^*_{33}=n_3(n_3-n_1-n_2)-N,\quad R^*_{23}=a(n_2-n_3), \tag{B2b}$$

$$R^*_{12}=R^*_{13}=0, \tag{B2c}$$

where

$$N=2a^2+\tfrac{1}{2}(n_1^2+n_2^2+n_3^2)-(n_1n_2+n_1n_3+n_2n_3). \tag{B2d}$$

The independent equations (3.5) in the tetrad (3.1) are

$$R^*_{123}=0\Rightarrow 2n_1^3-n_2^3-n_3^3-n_1^2n_2-n_1^2n_3+n_2n_3^2+n_2^2n_3=0, \tag{B3}$$

$$R^*_{231}=0\Rightarrow 2n_2^3-n_1^3-n_3^3-n_1n_2^2-n_2^2n_3+n_1n_3^2+n_1^2n_3+2a^2(n_3-n_2)=0, \tag{B4}$$

$$R^*_{331}=0\Rightarrow a(2n_1^2+3n_2^2-3n_3^2+n_1n_3-3n_1n_2)=0, \tag{B5}$$

$$R^*_{221}=0\Rightarrow a(2n_1^2+3n_3^2-3n_2^2+n_1n_2-3n_1n_3)=0. \tag{B6}$$


[1]G. F. R. Ellis and M. A. H. MacCallum, Commun. Math. Phys. **12**, 108 (1969).
[2]P. Szekeres, Commun. Math. Phys. **41**, 55 (1975).
[3]W. B. Bonnor and N. Tomimura, Mon. Not. R. Astron. Soc. **175**, 85 (1976).
[4]D. A. Szafron and J. Wainwright, J. Math. Phys. **18**, 1668 (1977).
[5]D. A. Szafron, J. Math. Phys. **18**, 1673 (1977).
[6]B. Berger, D. M. Eardley, and D. W. Olson, Phys. Rev. D **16**, 3086 (1977).
[7]C. B. Collins and D. A. Szafron, University of Waterloo Preprint (1977).
[8]M. A. H. MacCallum, *Cargese Lectures in Physics*, edited by E. Schatzman (Gordon and Breach, New York, 1973), Vol. 6.
[9]J. Wainwright, Gen. Rel. Grav. **8**, 797 (1977).
[10]G. F. R. Ellis, J. Math. Phys. **8**, 1171 (1967).
[11]G. F. R. Ellis, *General Relativity and Cosmology, Proceedings of The International School of Physics, "Enrico Fermi,"* edited by R. K. Sachs (Academic, New York, 1971).
[12]R. Kantowski, Ph.D. thesis, University of Texas (1966).
[13]R. Kantowski and R. K. Sachs, J. Math. Phys. **7**, 443 (1966).
[14]L. P. Eisenhart, *Continuous Groups of Transformations* (Dover, New York, 1961 and Princeton U. P., Princeton, N.J., 1933).
[15]C. B. Collins, J. Math. Phys. **18**, 2116 (1977).
[16]L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, N.J., 1926).
[17]A. R. King and G. F. R. Ellis, Commun. Math. Phys. **31**, 209 (1973).
[18]M. Demianski and L. P. Grischuk, Commun. Math. Phys. **25**, 233 (1972).

# A unified treatment of null and spatial infinity in general relativity. I. Universal structure, asymptotic symmetries, and conserved quantities at spatial infinity

Abhay Ashtekar[a)]

*Enrico Fermi Institute, Chicago, Illinois 60637*

R. O. Hansen[b)]

*Mathematics Department, University of California, Berkeley, California 94720*
(Received 12 September 1977)

A new definition of asymptotic flatness in both null and spacelike directions is introduced. Notions relevant to the null regime are borrowed directly from Penrose's definition of weak asymptotic simplicity. In the spatial regime, however, a new approach is adopted. The key feature of this approach is that it uses only those notions which refer to space–time as a whole, thereby avoiding the use of a initial value formulation, and, consequently, of a splitting of space–time into space and time. It is shown that the resulting description of asymptotic flatness not only encompasses the essential physical ideas behind the more familiar approaches based on the initial value formulation, but also succeeds in avoiding the global problems that usually arise. A certain 4-manifold—called Spi (spatial infinity)—is constructed using well-behaved, asymptotically geodesic, spacelike curves in the physical space–time. The structure of Spi is discussed in detail; in many ways, Spi turns out to be the spatial analog of I. The group of asymptotic symmetries at spatial infinity is examined. In its structure, this group turns out to be very similar to the BMS group. It is further shown that for the class of asymptotically flat space–times satisfying an additional condition on the (asymptotic behavior of the "magnetic" part of the) Weyl tensor, a Poincaré (sub-) group can be selected from the group of asymptotic symmetries in a canonical way. (This additional condition is rather weak: In essence, it requires only that the angular momentum contribution to the asymptotic curvature be of a higher order than the energy–momentum contribution.) Thus, for this (apparently large) class of space–times, the symmetry group at spatial infinity is just the Poincaré group. Scalar, electromagnetic and gravitational fields are then considered, and their limiting behavior at spatial infinity is examined. In each case, the asymptotic field satisfies a simple, linear differential equation. Finally, conserved quantities are constructed using these asymptotic fields. Total charge and 4-momentum are defined for arbitrary asymptotically flat space–times. These definitions agree with those in the literature, but have a further advantage of being both intrinsic and free of ambiguities which usually arise from global problems. A definition of angular momentum is then proposed for the class of space–times satisfying the additional condition on the (asymptotic behavior of the) Weyl tensor. This definition is intimately intertwined with the fact that, for this class of space–times, the group of asymptotic symmetries at spatial infinity is just the Poincaré group; in particular, the definition is free of super-translation ambiguities. It is shown that this angular momentum has the correct transformation properties. In the next paper, the formalism developed here will be seen to provide a platform for discussing in detail the relationship between the structure of the gravitational field at null infinity and that at spatial infinity.

## 1. INTRODUCTION

There are two distinct regimes in which the asymptotic behavior of the gravitational field has been investigated in detail[1]: at large separations from sources in null directions, and in spacelike directions. These investigations have yielded a great deal of information about properties of gravitating systems, information which has been crucial to the construction of mathematical models of isolated systems in general relativity.

In the null regime, the asymptotic structure was first examined in detail by Bondi, Van der Berg, and Metzner[2] and Sachs;[3] their results were reformulated and extended by Penrose.[4] Penrose's analysis provided a basis for later work[5] due to Newman, Newman and Penrose, Schmidt, Winicour, and others. These investigations have been crucial to the study of radiation—especially gravitational radiation—and also to the development of several ideas concerning global issues in general relativity. In particular, a great deal of the analysis of black holes, singularities, and, more recently, of $H$-spaces and of asymptotic quantization of zero rest-mass fields relies heavily on concepts introduced originally in the investigation of null infinity. In the spatial regime, major developments first came from the work of Arnowitt, Deser and Misner,[6] and Bergmann.[7] This work was later reformulated and extended by Geroch.[8] Independently, investigations have been made by O'Murchadha and York, Regge and Teitelboim, and others.[9] All these developments have also played an important role in the analysis of global problems. In particular, they have provided a basis for examining issues such as the positivity of (total) energy in general relativity, the

[a)]Supported in part by the grant S/GA/057.5 from the SRC to the University of Oxford and NSF contract PHY 76-81102 with the University of Chicago.
[b)]Supported in part by grant B/RG/513.4 from the SRC to the University of Oxford and NSF contract PHY 77-15191 with the University of California, Berkeley.

superspace formulation, and canonical quantization of gravity.

The key idea in both[4,8] sets of investigations is the following: One uses conformal transformations to bring "infinity" to a "finite distance"—more precisely, to attach suitable boundaries representing infinity—and then explores the asymptotic structure of the gravitational field by applying techniques from *local* differential geometry at points representing infinity. This procedure has several advantages: It avoids the heuristic considerations otherwise involved in taking limits, is manifestly coordinate independent, and, furthermore, simplifies computations. Vigorous work has been carried out using these techniques and a rich conceptual and mathematical structure has arisen in each regime.[1]

Unfortunately, however, very little is known about the connection between the two. Does asymptotic flatness in one regime, together with some simple and natural conditions, imply asymptotic flatness in the other?[10] In each regime, there arise groups of asymptotic symmetries. Is there any relation between these groups? Is there any relation between the conserved quantities which emerge from these groups? Not only are these issues unresolved, but in most cases, one does not know even how to *formulate* precise questions. Consider, for example, the notion of energy—momentum. In the spatial regime, one associates with isolated bodies a set of four numbers—the ADM 4-momentum—which represent the "total 4-momentum of the entire system including gravitation.[6,8]" In the null regime, one introduces another quantity—the Bondi 4-momentum—which represents the 4-momentum "left over" at a retarded instant of time.[2,4] It is natural to conjecture that the "difference"—in an appropriate sense—between the two quantities would represent the 4-momentum which has been radiated away until the retarded time under consideration. Unfortunately, the two existing descriptions are so disconnected from each other that it has not been possible to obtain even a precise formulation of this conjecture, let alone its proof or disproof! The essential difficulty is that the two vectors—as constructed—belong to entirely different vector spaces; hence one cannot even introduce the notion of their difference.

Why have the two formulations remained so disjoint from each other? It turns out that, in spite of close similarities, the two do differ from each other in a fundamental way: Whereas the standard framework for describing null infinity respects the four-dimensional character of space—time, that for spatial infinity requires a splitting of space—time into space and time. This difference permeates the two sets of analyses thoroughly. Thus, in the null regime, the asymptotic conditions refer to space—time as a whole, while in the spatial regime, they refer to spacelike 3-surfaces. As a consequence, null infinity turns out to be a boundary of the space—time manifold itself, and the physical fields which enter the discussion are all "four-dimensional" ones—the space—time metric, its curvature tensor, and various zero rest-mass fields. Spatial infinity, on the other hand, arises as a boundary of a spacelike three-surface and fields of interest in this case are all "three-dimensional" ones—Cauchy data for gravitation and matter fields. Thus, the usual formulation in the null case is "four dimensional" in spirit while the one in the spatial case is "*three*-dimensional." The essential reasons which have kept the two sets of analyses disjoint from each other can be traced back to this difference. Hence, a framework which attempts to unify the two must first overcome this difference: One of the two descriptions needs to be reformulated in the spirit of the other.

From aesthetic considerations, it is clearly the spatial description that needs reformulation. Technical considerations lead to the same conclusion: Quite apart from the issue of unification, the "three-dimensional" aspect of the usual description leads to some awkwardness in the spatial regime itself.

Recall[6,8] that, in this description, a space—time is said to be asymptotically flat at spatial infinity if it admits a surface on which the initial data approach the data on 3-planes in Minkowski space at an appropriate rate as one goes to infinity in spacelike directions. Unfortunately, however, the existence of *one* such Cauchy surface does not guarantee the existence of a "sufficient number" of them. Already for *linear* fields in *Minkowski space*, analogous results require a great deal of care in one's choice of asymptotic conditions: Even apparently minor modifications of the "correct" conditions have the effect that although the modified conditions are satisfied on one spacelike plane, after evolution, they need not be satisfied on a boosted one.[11] In general relativity, the situation is further complicated first by the absence of a "background" geometry and second by the nonlinearity of field equations. Indeed, the present state of affairs is again such that one does not even know how to formulate appropriate questions. Consider, for example, the notion of two Cauchy surfaces "boosted" relative to each other. In the usual descriptions of spatial infinity a precise formulation of this notion has always run into some global problem or another.[12] Hence, strictly speaking, one cannot even *ask* if asymptotic flatness of initial data sets will be preserved under boosts. The general procedure adopted so far essentially ignores all such global issues and just assumes—although often implicitly—that if there exists one asymptotically flat data set in the given space—time, then there exist "a sufficient number" of them. This assumption permeates the entire analysis and weakens many results substantially.[13]

All these considerations motivate the need of a new description of the asymptotic structure of the gravitational field at spatial infinity; a description which is "four-dimensional in spirit," which is free of global problems and which will serve as a platform for unification of the results obtained separately in the two regimes. Our purpose here is to obtain such a description. In the next paper, we shall discuss in detail the relation between the various notions introduced here in the spatial regime and their well-known analogs in the null regime.

The basic ideas underlying this work may be summarized as follows. In the null regime, we shall borrow the notion of asymptotic flatness directly from the traditional one—i.e., from Penrose's[14] definition of weak asymptotic simplicity. In the spatial regime, however, we adopt a new approach. To see this, consider, first Minkowski space–time. Recall that in the standard conformal completion, one obtains as its boundary not only null infinity $\mathcal{I}$, but also three additional points, $i^+$, $i^-$, and $i^0$. These three points represent, respectively, future timelike infinity, past timelike infinity, and spacelike infinity of Minkowski space.[15] The key idea in the new approach is to attach to space–times which are to be regarded as asymptotically flat, not only $\mathcal{I}$ but also a point "analogous to $i^0$." In the Minkowski case, $i^0$ can be characterized as the vertex of the "light cone at infinity," i.e., of $\mathcal{I}$. Therefore, given a space–time which is asymptotically flat in the null regime, one wishes to regard it as asymptotically flat also in the spatial regime provided one can attach to its null boundary $\mathcal{I}$ a single point $i^0$ such that, in the new completion, $\mathcal{I}$ is the null cone of $i^0$.

The essential difficulty arises of course in the specification of the details of this completion: One must introduce appropriate differential structure at $i^0$ and impose suitable conditions on the behavior of the conformally rescaled metric and of other physical fields. The selection of these conditions is a delicate issue. If the conditions are too weak, one might have too little structure available at $i^0$ to introduce physically interesting notions, or even worse, the completion might turn out to be so nonunique that the resulting analysis might inherit essential ambiguities. If, on the other hand, the conditions imposed are too strong, they might accommodate so few space–times that the resulting analysis might be totally uninteresting. These broad features are of course common to any analysis of asymptotics, and in particular to the analysis of null infinity. However, in the null regime, simplifications arise from the fact that $\mathcal{I}$ turns out to be a boundary of space–time. Thus, for example, as a direct consequence of this fact, smooth fields on the physical space–time satisfying conformally invariant equations automatically admit smooth extensions to $\mathcal{I}$ (after appropriate conformal rescalings) thereby simplifying the issue of differentiability conditions enormously. The situation is much more intricate in the spatial regime: One just does not expect physically interesting fields to acquire smooth limits at $i^0$. To see this, consider first the Maxwell field of a freely falling point charge in Minkowski space. After conformal completion, the appropriately rescaled field (i.e., the field which satisfies Maxwell's equations w.r.t. to the rescaled metric) is $C^\infty$ on $\mathcal{I}$. How does it behave at $i^0$? We claim that in fact it diverges, and, furthermore, cannot be made into a smooth, nonzero field by *any* rescaling using the conformal factor. One can see this rather easily geometrically. Fix any 3-plane in Minkowski space. Together with $i^0$ the plane becomes a compact submanifold—topologically a 3-sphere—of the completed space–time. Since the total charge on a compact spacelike 3-manifold must be zero, it follows that there must exist an effective "image" charge at $i^0$, and hence that the (rescaled) Maxwell field there must *diverge in a direction-dependent way*. The geometrical nature of the argument suggests that a similar situation might also exist in the gravitational case. This expectation is in fact confirmed by examples: The existence of a nonzero mass manifests itself in the singularity of the Weyl tensor at $i^0$![16] Thus, except in Minkowski space, one does not expect the (rescaled) metric to be even $C^2$ at $i^0$. As a result, the discussion of the differentiability conditions to be imposed on various fields becomes rather involved in the spatial regime.

This issue is discussed in Sec. 2, and a new definition of asymptotic flatness at null and spatial infinity is proposed. This definition is completely four dimensional in spirit: It is formulated using only the 4-manifold and "four-dimensional" fields theorem. Consequently, in the resulting analysis, global problems normally associated with the evolution of asymptotically flat initial data sets simply do not arise. However, given a space–time which is asymptotically flat in the sense of the new definition, one might introduce 3-surfaces and examine the issue of evolution of asymptotically flat initial data sets. It turns out that, due to the introduction of the point $i^0$, not only can one now formulate necessary notions to ask precise questions concerning this evolution, but also answer these questions in detail. This issue is examined in detail in Appendix B; it is shown that space–times which are asymptotically flat in the new sense do admit "a sufficient number" of asymptotically flat initial data sets. We emphasize that this result is *not* directly relevant to our analysis; it merely serves to connect the present framework with the ones in the literature. Note, however, that we have not *solved* the problems associated with the evolution of asymptotically flat data sets: The "four-dimensional" definition of asymptotic flatness simply circumlocutes these problems. In Appendix C we discuss the issue of existence of examples satisfying the new definition.

The "direction-dependence" of the limits of various physical fields at $i^0$ is almost inevitable: One can reach $i^0$ by moving away from sources in completely different spatial directions. Thus, the essential reason behind the intricate behavior of fields at $i^0$ is simply that $i^0$—the spatial boundary—is a *single point* and that the entire information about asymptotic behavior of fields at spatial infinity registers itself at this point. One is therefore led to look for a suitable "blowing up" of $i^0$ which can display all this information in terms of smooth fields on the blown-up structure. In Sec. 3, we present such a blowing-up procedure. The result is a certain 4-manifold which has the structure of a principal fibre bundle over the unit timelike hyperboloid in the tangent space at $i^0$, with the additive group of reals as the structure group. This 4-manifold is called Spi—spatial infinity.[17] The structure it inherits from its construction is examined. Apart from the fibre structure, it has a preferred (degenerate) horizontal metric (the pull back of the metric on the unit timelike hyperboloid at $i^0$) and a vertical vector field (the generator of the structure group). Intuitively, each point of Spi represents "an asymptotically distinct way of approaching infinity in spacelike directions," i.e., of approaching $i^0$. Consequently, even though physical fields admit only direction-dependent limits at $i^0$, they induce smooth fields on (and, in some cases, on cross sections of) Spi. Indeed in the final analysis, the situation at spatial infinity turns out to be rather similar to that at null infinity. In particular, as far as the universal structure at infinity is concerned, Spi plays essentially the same role in the spatial regime as $\mathscr{I}$ does in the null. This similarity is perhaps to be anticipated: A point of $\mathscr{I}$ can also be thought of as a distinct way of approaching infinity in null directions. Similarities—as well as differences—between the two regimes are also pointed out in Sec. 3.

In Sec. 4, we investigate the group $\mathcal{G}$ of asymptotic symmetries, i.e., the subgroup of the diffeomorphism group of Spi which leaves its universal structure invariant. In its structure, $\mathcal{G}$ turns out to be analogous to the BMS[2] group: It has an infinite dimensional, Abelian, normal subgroup—called the subgroup of Spi supertranslations—and a preferred four-dimensional Abelian normal subgroup—called the subgroup of Spi-translations—the quotient of the full group by the supertranslation subgroup being isomorphic with the Lorentz group. Furthermore, it turns out that there is a natural homomorphism from the group of allowable conformal rescalings of the unphysical metric onto the supertranslation subgroup. Hence, the full group $\mathcal{G}$ can also be realized as the semidirect product of the (quotient by the kernel of the above-mentioned homomorphism of the) group of conformal rescalings and the Lorentz group. Although this alternative description of the group is arrived at in a somewhat indirect fashion, it turns out to be the most useful one in the analysis of the asymptotic behavior of physical fields.

This analysis is carried out in Sec. 6. Specifically, we consider the zero rest-mass scalar field, the electromagnetic field, and the gravitational field. (Although the first of these is not of direct physical interest, it is included in the discussion to illustrate some mathematical techniques.) The (highest order) asymptotic behavior of these fields is described, respectively, by a scalar field $\phi$, a pair $(\mathbf{E}_a, \mathbf{B}_a)$ of vector fields, and a pair $(\mathbf{E}_{ab}, \mathbf{B}_{ab})$ of second rank symmetric trace-free tensor fields on the hyperboloid $\mathcal{K}$ of unit spacelike directions in the tangent space of $i^0$. (Thus, when regarded as fields on Spi, these fields are constant along fibres, reflecting the fact that they are invariant under Spi supertranslations.) $\mathbf{E}_a$ and $\mathbf{B}_a$ may be regarded as the asymptotic electric and magnetic fields *relative to the hyperboloid* $\mathcal{K}$;[18] and $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$, the electric and magnetic parts of the asymptotic Weyl curvature. Taking suitable limits of the field equations, we obtain the asymptotic equations for the asymptotic fields. It turns out that, although in the physical space–time one has nonlinear coupled differential equations, in the limit, asymptotic fields decouple and all the nonlinearities disappear.

Section 6 is devoted to conserved quantities. Since $i^0$ may be regarded as the "limit" of a sequence of 2-spheres with increasing radii (in the physical space–time), one expects to recover from asymptotic fields only those conserved quantities—such as total electric and magnetic charges, total 4-momentum and angular momentum—which are expressible as 2-surface integrals or limits thereof. (Thus, the situation is quite different from that at null infinity: Conserved quantities—such as the energy–momentum of test fields in, say, Minkowski space—expressible as 3-*surface integrals* in the physical space–time *can* be recovered from asymptotic fields on $\mathscr{I}$.) We first consider the electromagnetic case and obtain expressions for electric and magnetic charges in terms of integrals over 2-sphere cross sections of the hyperboloid $\mathcal{K}$ involving the fields $\mathbf{E}_a$ and $\mathbf{B}_a$, respectively. We then focus on the asymptotic gravitational fields $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$. It turns out that the *total* 4-momentum—including the contribution of the gravitational field itself—of the system can be expressed in terms of 2-sphere integrals involving $\mathbf{E}_{ab}$. (As one might expect, the analogous "conserved quantity" involving $\mathbf{B}_{ab}$, the "angular momentum monopole," vanishes identically.) The final definitions of all these quantities—the electric charge, the magnetic charge, and the 4-momentum—are essentially the same as those available in literature.[6,8,1] However, the present treatment has the advantage that, being "intrinsic," it is free of coordinate ambiguities, and, being "four-dimensional in spirit," it is free of global problems normally associated with the issue of preservation of asymptotic conditions under evolution. Finally, we introduce a new conserved quantity: the angular momentum. For this purpose, we have to make a restrictive assumption: It is only when the field $\mathbf{B}_{ab}$ on $\mathcal{K}$ vanishes that we can define angular momentum. Since $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$ together contain information about "$1/r^3$ part" of the Weyl curvature in the physical space–time, the condition on $\mathbf{B}_{ab}$ essentially requires that the "$1/r^3$ contribution" should arise *only* from the total energy–momentum of the isolated system. This condition serves two purposes. First, it enables one to introduce certain preferred cross sections on Spi, thereby reducing the infinite dimensional group $\mathcal{G}$ to the Poincaré group. Second, we can now introduce a new field which carries information about "the next order behavior" of—i.e., the "$1/r^4$ contribution" to—the magnetic part of the asymptotic curvature. This new field arises naturally as a tensor field on the preferred cross sections of Spi, rather than on on $\mathcal{K}$; it fails to be invariant under the action of translations on these cross sections. Angular momentum is defined using 2-sphere integrals of this field. This definition—being inseparably intertwined with the above mentioned Poincaré group—is free of the usual "supertranslation ambiguities."

The material covered in the main sections of the paper divides itself into two parts; the first part

(Secs. 3 and 4) deals with the universal structure at spatial infinity, the second (Secs. 5 and 6), with asymptotic fields and conserved quantities. In writing the second part, we have deliberately made as little reference to the first as possible. Thus, a reader interested mainly in conserved quantities may skip the discussion on universal structure—especially technicalities connected with Spi—without losing the main line of argument.

Before concluding this section, we wish to emphasize an important point which is often only implicit in discussions of infinity. General relativity, by itself, is a completely self-contained theory without the need of any detailed framework describing infinity: It is only because one is interested in capturing the intuitive notion of an isolated system in a mathematically precise fashion that one is interested in these frameworks. Therefore, the definitions one introduces, constructions one makes, and the notions one formulates in discussing the asymptotic structure of the gravitational field are arbitrary to some extent; their justification lies essentially in their utility. Thus, in principle, it is quite possible to have several distinct frameworks all of which are useful in different ways. In particular, the utility of one approach does not invalidate any other. The framework presented here is thus just one of the many possible ones and, in essence, its value lies only in its ability to introduce useful notions in the spatial regime and to relate them to their analog in the null regime.

## 2. ASYMPTOTIC CONDITIONS

In this section, we introduce a definition of asymptotic flatness at null and spatial infinity and discuss its relation with definitions used in the other formulations.

The basic idea in the present approach is to use the same notion of asymptotic flatness in the null regime as in Penrose's[14] definition of weak asymptotic simplicity and to simply supplement this definition by appropriate conditions to incorporate asymptotic flatness also in the spatial regime. How are these additional conditions to be chosen? One only has the following general set of criteria: (i) conditions should be formulated in a "four-dimensional" spirit, i.e., they should not require the introduction of any splitting of space—time into space and time; (ii) they should be strong enough to yield a sufficiently rich structure, a structure which can enable one to introduce physically interesting notions at spatial infinity and to relate them to those available at null infinity; and, (iii) they should be weak enough to allow a sufficient number of examples; space—times which are "obviously asymptotically flat" from physical considerations should, in particular, satisfy these conditions.

How do the conditions usually imposed at spatial infinity fare with respect to these criteria? To be specific, let us consider Geroch's[8] formulation of the Arnowitt—Deser—Misner[6] conditions (ADM—G conditions). Although these are "three dimensional" in spirit, they do indeed have the "correct" strength: They admit a wide class of examples and also enable one to introduce interesting physical notions. Thus, what we need is a set of conditions which are essentially as strong as these but which are more "four dimensional" in spirit; conditions which are manifestly free of global problems associated with the emphasis on "three dimensions." Let us therefore begin with a brief review of the ADM—G conditions. This review will also be useful later while examing the modifications called for by the new approach. The key notion in the ADM—G formulation is that of asymptotically flat initial data sets. One introduces this notion as follows. Fix a spacelike surface $\mathbf{T}$ with an initial data set $(q_{ab}, p_{ab})$—the intrinsic metric and the extrinsic curvature—satisfying the vacuum constraint equations outside some compact region representing sources. One adds to this $\mathbf{T}$ a single point $\Lambda$—the point at infinity—thereby obtaining a new 3-manifold $\widetilde{\mathbf{T}}$. (If $\mathbf{T}$ is topologically $\mathbb{R}^3$ this procedure is just the one point compactification of $\mathbf{T}$, so that $\widetilde{\mathbf{T}}$ is topologically $S^3$.) Various conditions on $q_{ab}$ and $p_{ab}$ are now formulated in terms of their behavior near $\Lambda$. More precisely, the data set is said to be asymptotically flat provided there exists a scalar field $\Omega$ on $\widetilde{\mathbf{T}}$ which is $C^2$ at $\Lambda$, $C^\infty$ (and positive) everywhere else, such that (i) $\Omega|_\Lambda = 0$, $\widetilde{D}_a\Omega|_\Lambda = 0$, and, $\mathbf{\Omega}_{ab} := \lim_{\to\Lambda}\Omega^{-1/2}(\widetilde{D}_a\widetilde{D}_b\Omega - 2\Omega^2 q_{ab})$ exists as a direction-dependent[19] tensor at $\Lambda$; (ii) there exists a metric $\widetilde{q}_{ab}$ on $\widetilde{\mathbf{T}}$ which is $C^0$ at $\Lambda$ and $C^\infty$ elsewhere with $\widetilde{q}_{ab} = \Omega^2 q_{ab}$ everywhere on $\mathbf{T}$; and, (iii) $\mathcal{R}_{ab} = \lim_{\to\Lambda}\Omega^{1/2}\widetilde{\mathcal{R}}_{ab}$ and $\mathbf{p}_{ab} = \lim_{\to\Lambda}\Omega p_{ab}$ exist as direction-dependent tensors at $\Lambda$. (Here, $\widetilde{D}_a$ is the derivative operator on $(\mathbf{T}, \widetilde{q}_{ab})$ and $\widetilde{\mathcal{R}}_{ab}$ is the Ricci tensor of $\widetilde{q}_{ab}$.) Conditions (i) and (ii) on the conformal factor $\Omega$ and the metric $q_{ab}$ are suggested immediately by the standard conformal completion of the Euclidean space. Condition (iii) [and also the precise differentiability requirement in (ii)], on the other hand, has a more subtle origin: It arises only after a careful investigation of examples.[8]

The idea now is to capture the essence of these conditions in a "four-dimensional" spirit. Consider, first, Cauchy surfaces in Minkowski space. To examine the asymptotic behavior of initial data sets on these surfaces à la ADM—G one must first complete each surface by adding a single point. Note, however, that the Penrose completion[15] of Minkowski space already provides a point, $i^0$, which can *simultaneously* serve as the point at infinity for *all* these surfaces! Indeed, in the completed space—time, together with $i^0$ each of these surfaces becomes, topologically, a 3-sphere. Moreover, the point $i^0$ itself does not refer to any preferred Cauchy surface; it can be introduced without having to make any splitting of space—time into space and time. The idea therefore is to introduce a point analogous to $i^0$ in more general contexts, a point which is at once the point at spatial infinity for all (well behaved) Cauchy surfaces. As remarked in the Introduction, this can be achieved by just attaching to $\mathcal{I}$—the null cone at infinity—its vertex. Next, we must specify the behavior of various fields at this newly attached point. It is here that we draw heavily on the ADM—G conditions: We wish to require that the various space—time fields—the conformal factor, the rescaled metric, its Weyl tensor, and possible matter fields—have exactly that behavior which can guarantee the existence of a "sufficient number" of initial data sets which are asymptotically flat in the sense of these conditions. More precisely, we wish to require that the behavior of fields be such that, given *any* three-dimensional, spacelike subspace of the tangent space at $i^0$, there exists at least one Cauchy surface (in the physical space—time) with an asymptotically flat initial data set, whose tangent space at $i^0$ in the completion coincides with the given subspace. If this requirement can be satisfied, not only will the resulting asymptotic conditions be of the "correct" strength, but they will also be free of the global problems discussed in the Introduction.

The ADM—G conditions on the conformal factor can be translated to four dimensions in a straightforward way; one has only to replace the 3-metric and the corresponding derivative operator by the 4-metric and its derivative operator. The continuity requirement on the 3-metrices can be satisfied only if the rescaled 4-metric is itself continuous at $i^0$. The requirement on extrinsic curvatures, however, is more severe: Since the extrinsic curvature involves the metric connection, the continuity of the rescaled metric, by itself, cannot guarantee the existence of even a single initial data set satisfying this requirement. One is therefore led to impose stronger conditions on the 4-metric. It is here that major complications arise. Consider, first, the obvious way to achieve the required strengthening. Demand that the rescaled 4-metric be $C^1$ at $i^0$. This condition would indeed do the job: It does guarantee the existence of a "sufficient number" of asymptotically flat initial data sets. Unfortunately, however, the condition is too strong: It turns out (as we shall see in Sec. 6) that discontinuities in the (rescaled) metric connection are, in essence, a measure of the total mass of the isolated system described by the given space—time, and hence, that the ADM 4-momentum associated with the space—time vanishes identically if the (rescaled) metric is $C^1$ at $i^0$![20] Thus, one is forced to make the awkward requirement that the 4-metric be better behaved than a $C^0$ field but not be $C^1$ at $i^0$. As the above remark on the ADM 4-momentum suggests, the appropriate condition turns out to be the following: Demand that not only should the metric be $C^0$ at $i^0$, but also the metric connection should admit a direction-dependent limit there. Finally, consider the ADM—G condition on the three-dimensional Ricci tensor. It is straightforward to show (using the source-free Einstein's equation) that one can translate this condition to a condition on the Weyl tensor of the 4-metric which is guaranteed to be satisfied by the differentiability requirements on the metric. Thus, we are led to the following definition.

*Definition*: A space—time $(\mathbf{M}, g_{ab})$ will be said to be *asymptotically empty and flat at null and spatial infinity* (AEFANSI) if:

(i) There exists a manifold $\bar{\mathbf{M}}$ with boundary $(\partial\bar{\mathbf{M}} =: \mathscr{I})$ equipped with a $(C^3)$ conformal structure, and, an imbedding of $\mathbf{M}$ into $\bar{\mathbf{M}}$ which displays $(\mathbf{M}, g_{ab})$ as a weakly asymptotically simple space—time,

(ii) There exists a manifold $\hat{\mathbf{M}}$ with a (Lorentz) metric $\hat{g}_{ab}$ and a conformal-structure-preserving imbedding $\psi$ of $\bar{\mathbf{M}}$ into $\hat{\mathbf{M}}$ (which is $C^4$ on $\bar{\mathbf{M}}$),

(iii) There exists a point $i^0$ in $\hat{\mathbf{M}}$ with the following properties:

(a) $\hat{\mathbf{M}}$ has a $C^{>1}$ differential structure at $i^0$, and $\hat{g}_{ab}$ is $C^{>0}$ at $i^0$,[21]

(b) In $\hat{\mathbf{M}}$, $\psi(\mathscr{I})$ is the null cone of $i^0$,

(c) The function $\Omega$ defined on $\psi(\mathbf{M})$ via $\psi_*(\hat{g}_{ab}) = \Omega^2 g_{ab}$ admits a $C^2$ extension at $i^0$, with $\Omega|_{i^0} = 0$, $(\hat{\nabla}_a \Omega)|_{i^0} = 0$, $(\hat{\nabla}_a \hat{\nabla}_b \Omega - 2\hat{g}_{ab})|_{i^0} = 0$; and finally,

(iv) The Ricci tensor $R_{ab}$ of $g_{ab}$ vanishes in the intersection in $\hat{\mathbf{M}}$ of the image of the physical space—time with some neighborhood of $\mathscr{I} \cup i^0$.

In essence, condition (i) guarantees that the space—time is asymptotically flat in null directions, while conditions (ii) and (iii) ensure that it is asymptotically flat in spacelike directions, and that the structures arising in the two regimes are as compatible as possible.[22] [The differentiability requirements (iii. a) on $(\hat{\mathbf{M}}, \hat{g}_{ab})$ just assures that the metric connection admits a regular direction dependent limit at $i^0$—i. e., has smooth "angular" behavior but possibly finite discontinuities in the "radial directions." For the definition of $C^{>n}$ differentiability, see Appendix A.] The manifold has been introduced in the definition just to include $i^0$ [23]—the point at spatial infinity—together with its differential and conformal structure in the completion; a simple attachment of $i^0$ to $\bar{\mathbf{M}}$ would have resulted in a "manifold with a corner," which, in turn, would have prevented us from making a straightforward use of local differential geometry at $i^0$. [The precise differentiability requirements in condition (ii) are motivated by examples. For details, see Appendix C.] Finally, we note that the various conditions in the definition can also be motivated without any reference to the ADM—G formalism: Conditions (i), (ii), and (iii. b) are geared to incorporate the intuitive idea that "$i^0$ be the vertex of the light cone of infinity;" (iii. c) ensures that the conformal factor $\Omega$ has the same asymptotic behavior (near $i^0$) as in Minkowski space, i. e., that it "falls off as $1/r^2$;" and, as shown in Appendix A, the differentiability requirements imposed via (iii. a) are precisely such as to ensure that Weyl curvature "falls off as $1/r^3$" in the physical space—time.

As explained in the Introduction, in this paper we are concerned more with the structure of spatial infinity than with the detailed relation between spatial and null infinity. *Consequently, for the purpose of this paper, of all conditions in the definition, only* *condition (iii.a) on differentiability of the metric $\hat{g}_{ab}$ at $i^0$, (iii.c) on the behavior of the conformal factor $\Omega$, and (iv) on the Ricci tensor $R_{ab}$ will be of direct relevance.* The other conditions will play an important role only in the next paper.

It follows directly from the above definition that every point in the physical space—time [more precisely, of its image $\psi(\mathrm{M})$ in $\hat{\mathrm{M}}$] is spacelike related to $i^0$. As a consequence, the point $i^0$ serves as the "spatial boundary" of the physical space—time very much as the 3-surface $\mathscr{I}$ serves as the null boundary. The fact that the spatial boundary consists of a single point is perhaps the most important aspect of the above completion; an aspect which adds to the completion complexity in some ways and richness in other. Thus, for example, the intricacy of differentiability requirements at $i^0$ can be easily traced back to this aspect: It is because $i^0$ is a single point that the limits attained by physical fields on space—time are forced to be direction-dependent. On the other hand, $i^0$ provides us with a *preferred* point in the completed space—time and this turns out to be useful in many ways. For example, isometries in physical space—times can be characterized and classified rather easily by examining their extensions to $i^0$.[20] Furthermore, the tangent space at $i^0$ will be seen to provide a natural home for various conserved quantities—ADM and BMS 4-momenta, multipole moments in the stationary case, etc.—making it easy to investigate the relation between them.

Note that, although the definition of AEFANSI space—times was arrived at by using the ADM—G framework, in the final version, *the definition itself makes no reference to initial data sets*; it refers only to "space—time" fields. Hence, in the present formulation, global problems associated with preservation of asymptotic conditions under evolutions simply to not arise. Nonetheless, having obtained the notion of AEFANSI space—times, one can, if one wishes, introduce spacelike 3-surfaces and ask for the status of these evolutions. *It turns out that a complete analysis of this issue can be made.* First, one can show that *AEFANSI space—times do admit "a sufficient number" of asymptotically flat initial data sets*: Given a spacelike, three-dimensional subspace of the tangent space at $i^0$, there exist as many asymptotically distinct spacelike 3-surfaces with asymptotically flat initial data sets in the physical space—time, which, in the completion, are tangential to the given subspace at $i^0$, as there are functions on a 2-sphere. Next, using the tangent space at $i^0$, one can now introduce, unambiguously, the notion of Cauchy surfaces which are boosted or time-translated with respect to each other. Therefore, one can meaningfully formulate questions about evolutions. Finally, using techniques from local differential geometry *at* $i^0$, one can also answer these questions in detail: Evolutions which are asymptotically regular do preserve the ADM—G conditions. Thus it appears that AEFANSI space—times offer an ideal home for discussions involving asymptotically flat initial data sets. Since in this paper we use only those notions which refer to space—time as a whole, a detailed discussion of the results quoted above has been relegated to Appendix B.

Finally, we remark that condition (iv) in the definition of AEFANSI space—times can be relaxed to admit zero rest-mass fields in a neighborhood of infinity.[24] Since, in a curved space—time, conformally invariant zero rest-mass equations are in general consistent only for spins less than $\frac{3}{2}$, we need to consider only these fields. Then (iv) may be replaced by:

(iv)′ Fields $\hat{\phi}$, $\hat{\psi}_A$, $\hat{F}_{ab}$ satisfying the zero rest-mass equations on $(\mathrm{M}, \hat{g}_{ab})$ for spins 0, $\frac{1}{2}$, and 1, respectively, are permissible sources at infinity provided $\hat{\phi}$, $\hat{\psi}_A$, and $\hat{F}_{ab}$ admit smooth extensions to $\mathscr{I}$, and, $\Omega^{1/2}\hat{\phi}$, $\Omega^{3/4}\hat{\psi}_A$, and $\Omega\hat{F}_{ab}$ admit regular direction-dependent limits[21] at $i^0$.

Here the conditions on the behavior at $\mathscr{I}$ are the usual[4] ones. Those at $i^0$ are motivated by the ones required in Minkowski space to guarantee finiteness of energy—momentum and, in the case of Maxwell field, also the finiteness of total charge.

## I. UNIVERSAL STRUCTURE AT SPATIAL INFINITY

### 3. *SPI*

In the previous section, we introduced the notion of AEFANSI space—times. In a sense, all that now remains is to examine the behavior of various physical fields in the neighborhood of the boundary $\mathscr{I} \cup i^0$, to introduce conserved quantities, and to obtain relations between them. However, the fact that $i^0$ is a single point introduces complications in such a program. In particular, various physical fields admit only *direction-dependent* limits at $i^0$ and hence it is rather awkward to examine their behavior using only the framework introduced so far. What is needed is some sort of "blowing up" of $i^0$: One might hope that in the limit, physical fields will register themselves as smooth fields on an appropriate blown-up structure. The purpose of this section is to obtain such a structure. Quite apart from simplifying the analysis of asymptotic fields, this blowing up turns out to be valuable in its own right. In particular, we shall see that it provides an arena for describing the universal structure in the spatial regime and plays a key role in the discussion of asymptotic symmetries.

How is this blowing up to be achieved? What is needed is an appropriate modification of the standard blowing-up procedures used in algebraic geometry; a modification which can incorporate the additional differentiable and metric structures available. Let us therefore begin by examining these structures. Consider first the differentiable structure. The completed manifold $\hat{\mathrm{M}}$ is only guaranteed to be $C^{>1}$ at $i^0$. Hence, using the differentiable structure one can only construct the first- and second-order tangent spaces there. Intuitively, this means that one cannot distinguish between two geometrical structures—e.g., submanifolds of $\hat{\mathbf{M}}$—which agree up to second order at $i^0$; indeed, one cannot even examine their higher-order behavior! This lack of distinguishability gives rise to severe constraints on the possible directions to proceed: One must obtain the required blowing up using only the first- and second-order behavior of geometrical objects at $i^0$.

Two classes of such objects at once present themselves: spacelike, three-dimensional submanifolds, and spacelike curves in $(\hat{\mathbf{M}}, \hat{g}_{ab})$. Intuitively, these represent two distinguished classes of paths to approach spatial infinity; in the physical space–time one can move away from sources along spacelike Cauchy surfaces or along inextendible spacelike curves. It turns out that the use of either one of these leads to essentially equivalent blown-up structures. However, the use of curves turns out to have two advantages in the analysis of the asymptotic behavior of physical fields. First, since these fields admit direction-dependent limits at $i^0$, they appear as *direction-independent*—in fact smooth—fields on the space of curves while they remain[8] direction-dependent as fields on the space of Cauchy surfaces. Second, it turns out that the use of Cauchy surfaces leads to the introduction of the initial value formulation and hence of "spatial" fields while the use of curves enables one to deal always with "space–time" fields, thereby simplifying the analysis. Therefore, in this paper we shall work with spacelike curves; equivalence of the resulting blown-up structure with the one obtained using Cauchy surfaces will be discussed elsewhere.[20]

To summarize, we wish to construct the required blown-up structure using various asymptotically distinct, inextendible spacelike curves in the physical space–time. The resulting structure will, in many ways, be the spatial analog of $\mathscr{I}$; $\mathscr{I}$ can also be constructed using certain inextendible curves—the null geodesics—in the physical space–time. Note that, in the null regime, one does not consider arbitrary null curves but only those which are geodesics; a point of $\mathscr{I}$ represents a "good" way of approaching infinity in null directions rather than an arbitrary one. Similarly, in the spatial case, one must specify some regularity conditions; we must choose only "good" ways of approaching $i^0$. Indeed, in the absence of such conditions, the blown-up structure will be infinite dimensional and hence not very useful.

How are these conditions to be selected? An obvious strategy presents itself: We should impose as many regularity conditions as the universal structure of AEFANSI space–times permits. First, we have $C^{>1}$ differentiability at $i^0$. So we shall require that the curves be $C^{>1}$ at $i^0$ and $C^3$ everywhere else. This will enable us to examine both velocity and acceleration—the first- and the second-order behavior—of curves at $i^0$. Next, we have the metric at $i^0$. We can use it to make demands on the parametrization of curves: Only those curves $p(\lambda)$ (with $p \in \hat{\mathbf{M}}$ and $\lambda \in \mathbb{R}$) are to be allowed for which $p(0)$ is $i^0$ and the tangent vector at $i^0$ is unit. The first of these two requirements is rather trivial. The second, however, is not: It can be imposed only because the metric *at* $i^0$ is universal. Note, in particular, that the metric at $i^0$ cannot even be rescaled; it follows[25] from conditions on the conformal factor in the definition of AEFANSI space–times that only such changes $\Omega \to \omega\Omega$ in the conformal factor are allowed for which $\omega$ is unity (and $C^{>0}$) at $i^0$, so that the metric at $i^0$ is conformally invariant. (Note also that we cannot demand the tangent vector to be unit at points other than $i^0$ since such a requirement would *not* be conformally invariant.) If the metric at $i^0$ were only $C^0$—rather than $C^{>0}$—these would be all the conditions we could impose. Then "good" ways of approaching $i^0$ would consist simply of equivalence classes of curves (satisfying the requirements stated above) where two curves are regarded as equivalent if they agree to first order at $i^0$, i.e., if they are tangential there. Since each of these equivalence classes can be characterized by its tangent vector at $i^0$, and since these vectors are required to be unit, the collection of all "good" ways would naturally acquire the structure of the unit timelike hyperboloid in the tangent space of $i^0$. Thus, using only the $C^0$ property of the metric, the blown-up structure would simply be this hyperboloid.[26]

However, the metric is $C^{>0}$ at $i^0$. Hence we can indeed distinguish between two curves which differ in the second order; not only does each curve carry a velocity vector at $i^0$ but, for each choice of a derivative operator, also an acceleration vector. What are the possible regularity conditions on the second order behavior of curves? An obvious choice is to demand that the curves be geodesics.[27] Unfortunately, for spacelike curves, the notion of geodesics is not conformally invariant. One must therefore select a particular metric in the conformal class available. The only distinguished metric in this class is the physical metric $g_{ab}$ itself. Unfortunately, this $g_{ab}$ is not even defined at $i^0$. Hence, we must first formulate the condition at points near $i^0$, re-express it in terms of an unphysical metric which *is* well-behaved at $i^0$ and then take the limit. Fix a spacelike curve $p(\lambda)$ in $\hat{\mathbf{M}}$ which is $C^{>1}$ at $i^0$ and $C^3$ elsewhere. Let $\eta^a$ be the tangent vector field to the curve. We wish to require that on $(\mathbf{M}, g_{ab})$, $\eta^a$ be geodesic; i.e., that $\eta^{[a}A^{b]} = 0$ where $A^b = \eta^a \nabla_a \eta^b$ is the acceleration of the curve relative to $g_{ab}$. In terms of the metric $\hat{g}_{ab}$ (whose restriction to $\mathbf{M}$ is $\Omega^2 g_{ab}$), this condition becomes $\eta^{[a}\hat{A}^{b]} + \Omega^{-1}\eta^{[a}\hat{\nabla}^{b]}\Omega = 0$, i.e., $\hat{h}_{ab}(\hat{A}^b + \Omega^{-1}\hat{\nabla}^b\Omega) = 0$ where $\hat{A}^b = \eta^a \hat{\nabla}_a \eta^b$ is the acceleration of the curve relative to $\hat{g}_{ab}$ and $\hat{h}_{ab} = \hat{g}_{ab} - (\hat{g}_{qp}\eta^p\eta^q)^{-1}\eta_a\eta_b$ is the projection operator in the 3-flat which is $\hat{g}_{ab}$—orthogonal to $\eta^a$. We can now take the limit. The resulting condition is $\lim_{\to i^0} \hat{h}_{ab}(\hat{A}^b + \Omega^{-1}\hat{\nabla}^b\Omega) = 0$. Note that, although $\hat{A}^b$ does depend on the particular choice of the metric from the conformal class, *the condition as a whole is conformally invariant*. Thus, the regularity condition on the second-order behavior of curves completely determines the components of the acceleration of the curve (relative to any $\hat{g}_{ab}$, in the conformal class, which is $C^{>0}$) at $i^0$ which are orthogonal to its tangent. ($\lim_{\to i^0} \Omega^{-1}\hat{h}_{ab}\hat{\nabla}^b\Omega$ depends only on the tangent vector to the curve at $i^0$.)[28] The component along the tangent vector, on the other hand, is completely unconstrained and carries all the interesting information about the second-order behavior of our curves. We summarize. A spacelike curve $p(\lambda)$ in $(\hat{\mathbf{M}}, \hat{g}_{ab})$, passing through $i^0$ will be said to be *regular* if and only if (i) it is $C^{>1}$ at $i^0$ and $C^3$ elsewhere; (ii) it is parametrized so that $p(0)$ is $i^0$ and the tangent vector to the curve, $\eta^a$, is unit *at* $i^0$; and (iii) $\eta^a$ satisfies $\lim_{\to i^0}\hat{h}_{ab}(\hat{A}^b + \Omega^{-1}\hat{\nabla}^b\Omega) = 0$. These are all the regularity conditions we can impose using only that structure which is universally present at $i^0$: The first of these conditions refers to the differentiable structure at $i^0$, the second to the metric at $i^0$, and the third, to the existence of (a family of conformally related) direction-dependent connections. Respectively, these conditions demand that regular curves be well-behaved submanifolds, that they be nicely parametrized, and that they be indistinguishable, asymptotically, from the geodesics in the physical space-time. The blowing up of $i^0$ will now be obtained from the collection of these regular curves.

Let $S$ denote the collection of equivalence classes of regular curves where two curves are regarded as equivalent if they have the same tangent and the same acceleration at $i^0$, i.e., if they agree to first and second order there. A point of $S$ can be characterized by the pair $(\eta^a, \hat{g}_{ab}\eta^a\hat{A}^b)$—the common tangent vector and the common tangential acceleration of the regular curves in the equivalence class, both evaluated at $i^0$—where the value of $\hat{g}_{ab}\eta^a\hat{A}^b$ is governed by the particular choice of the metric in the conformal class.[29] Intuitively, each of these points represents an asymptotically distinct, "good" way of approaching infinity in spacelike directions. Thus, $S$ is the blown up $i^0$.

What structure does this $S$ inherit from its construction? Note, first, that there is a natural projection mapping $\pi$ from $S$ onto the unit timelike hyperboloid $\mathcal{K}$ in the tangent space of $i^0$: $\pi$ sends each equivalence class of regular curves to the common tangent vector they have at $i^0$. Hence, one might expect that $S$ can be given the structure of a fibre bundle. This expectation is indeed correct. Fix a point on the hyperboloid $\mathcal{K}$ and consider the fibre $F$ over it. Points of this fibre $F$ represents various equivalence classes of regular curves which happen to have the same tangent vector at $i^0$. Hence, each point of $F$ can be labeled by the tangential component of the acceleration of the curves in the corresponding equivalence class. Fix a metric $\hat{g}_{ab}$. Then, for any curve $p(\lambda)$ with tangent vector field $\eta^a$, the tangential component $\hat{a}$ of acceleration is given by $\hat{a} = \hat{g}_{ab}\eta^a\hat{A}^b\,|_{i^0} = \hat{g}_{ab}\eta^a\eta^m\hat{\nabla}_m\eta^b\,|_{i^0}$. Since, by the definition of equivalence, the value of $\hat{a}$ is the same for all curves in any one equivalence class, the fibre $F$ can be coordinatized by $\hat{a}$. Recall that the tangential component of the acceleration is completely unconstrained. Hence, $\hat{a}$ can assume arbitrary real values. It therefore follows that $F$ is homeomorphic to the real line. Using these properties, one can easily endow $S$ with the structure of a fibre bundle.

How does the above coordinatization of fibres respond to conformal rescaling? Consider another $(C^{>0})$ metric $\tilde{g}_{ab}$ in the available conformal class. We must have $\tilde{g}_{ab} = \omega^2\hat{g}_{ab}$ for some function $\omega$ on $\hat{\mathbf{M}}$ which is $C^{>0}$ at $i^0$, $C^2$ elsewhere and which satisfies $\omega\,|_{i^0} = 1$. Consider $\tilde{a} = \tilde{g}_{ab}\eta^a\tilde{A}^b$ for any given regular curve with tangent vector $\eta^a$. It is easy to check that $\tilde{a} = \hat{a} + [\eta^a\hat{\nabla}_a\omega]_{i^0}$. Thus, the coordinatization is *not* conformally invariant. (Hence, there is *no* natural vector space structure on the fibres.) Note, however, that given any *two* points on the same fibre, say $p_1$ and $p_2$, we have, $\hat{a}_1 - \hat{a}_2 = \tilde{a}_1 - \tilde{a}_2$. Thus, given any fibre $F$, although there is no natural mapping from $F$ to the reals, *there does exist such a mapping from* $F\times F$ *to the reals*: Send $(p_1, p_2)$ to the real number $\hat{a}_1 - \hat{a}_2$. Furthermore, this mapping is onto and its kernel is just the diagonal subset of $F\times F$. Hence, it induces a natural free and transitive[30] action of the additive group of reals on each fibre $F$. Thus, $(S, \mathbb{R}, \mathcal{K})$ is in fact a principal fibre bundle[31] where the structure group $\mathbb{R}$ is just the additive group of reals.

To summarize, the result of the blowing up of $i^0$ is a 4-manifold which has the structure of a principal fibre bundle: The base space is the unit timelike hyperboloid in the tangent space of $i^0$, and the structure group is the additive group of reals. *This $S$ will be called Spi—spatial infinity.* From its very construction, $S$ inherits two tensor fields: a covariant, second rank, symmetric (degenerate) tensor field $h_{ab}$, the pullback to $S$ of the natural metric on the hyperboloid $\mathcal{K}$; and a vertical vector field $v^a$, the generator of the natural, one-parameter family of diffeomorphisms on $S$ induced by its structure group.[32]

Note that conformal rescalings of the (unphysical) metric induce, in a natural fashion, certain motions on Spi: Since under the rescaling $\hat{g}_{ab} \to \tilde{g}_{ab} = \omega^2\hat{g}_{ab}$, the labeling of fibres changes via $\hat{a} \to \tilde{a} = \hat{a} + (\eta^a\hat{\nabla}_a\omega)\,|_{i^0}$, one can obtain a natural action of these conformal transformations on Spi. (We shall see in the next section that the resulting transformations on Spi are precisely the supertranslations at spatial infinity.) Since this action leaves each fibre as a whole invariant—its projection on $\mathcal{K}$ vanishes—effectively, it reshuffles the "second-order" structure at $i^0$, leaving the "first-order" structure untouched. This interplay between conformal transformations in the completed space-time and "second-order" transformations at $i^0$ is a fundamental aspect of the universal structure at spatial infinity; we shall refer to it again in Secs. 5 and 6.

Finally, we remark that Spi could have been constructed via procedures which differ in some respects from the ones used above: There appears to exist a great deal of freedom in the precise choice of description of the "second-order structure" at $i^0$, i.e., of fibres of Spi. *In particular, we could have labeled these fibres using the induced connections on $C^{>1}$ curves, rather than acceleration.*[33] The particular procedure employed above is geared to bring out the similarity of the construction of Spi with that of $\mathscr{I}$. Thus, just as each point of $\mathscr{I}$ can be regarded, intuitively, as a "good" way of approaching infinity in null directions, a point of Spi can be so regarded in spacelike directions. Furthermore the resulting structures are also very similar in the two cases: $\mathscr{I}$ and Spi are both fibre bundles. As a result, the corresponding groups of asymptotic symmetries also turn out to be similar. There are, however, some important differences. For example, the structure of Spi is more rigid than that of $\mathscr{I}$: Spi is endowed with a *preferred* degenerate metric and a *preferred* vertical vector field, while $\mathscr{I}$ has available only a (conformal) class of such fields. Perhaps the most important difference is that whereas $\mathscr{I}$ serves as a boundary of space–time, Spi, being itself four dimensional, cannot.

## 4. ASYMPTOTIC SYMMETRIES AT SPATIAL INFINITY

Symmetry groups arise in physics as groups of transformations which preserve the structure of interest. What is the structure relevant to the analysis of the asymptotic behavior of the gravitational field at spatial infinity? It is just the universal structure of Spi: The fibre bundle character of $S$, the horizontal tensor field $h_{ab}$, and the vertical field $v^a$. The group of asymptotic symmetries at spatial infinity is therefore just that subgroup of the diffeomorphism group of $S$ which preserves this universal structure.[34] In this section, we shall first investigate this (sub-) group $\mathcal{G}$ in detail and then indicate how this group will be reduced to the Poincaré group in Sec. 6.

Consider, then, diffeomorphisms of $S$ which preserve its fibre structure and leave invariant the fields $h_{ab}$ and $v^a$. Let $\xi^a$ denote the generator of such a diffeomorphism. Then, $\xi^a$ is a vector field on $S$ which satisfies, in particular, the following conditions: (i) $\pounds_\xi h_{ab}=0$ and (ii) $\pounds_\xi v^a=0$. Note, however, that the vector field $v^a$ is vertical and nowhere vanishing. Hence, each fibre of $S$ is just an integral curve of this $v^a$. Therefore, condition (ii) above already guarantees that the one-parameter family of diffeomorphisms generated by $\xi^a$ is fibre preserving. Thus, conditions (i) and (ii) are not just necessary but also sufficient to guarantee that (the one-parameter family of) diffeomorphisms generated by $\xi^a$ belong to $\mathcal{G}$. It is obvious that the collection of all such vector fields has the structure of a Lie algebra. We shall denote this Lie algebra by $\mathcal{L}_\mathcal{G}$.

Fix an element $\xi$ of $\mathcal{L}_\mathcal{G}$. Since $\pounds_v\xi^a=0$, we can project $\xi^a$ down to the base space $\mathcal{K}$ of $S$ unambiguously. Denote by $\boldsymbol{\xi}^a$ the projected vector field on $\mathcal{K}$. Then, it follows that $\pounds_\xi h_{ab}=0$ on $S$ if and only if $\pounds_{\boldsymbol{\xi}}\mathbf{h}_{ab}=0$ on $\mathcal{K}$, where $\mathbf{h}_{ab}$—the projection of $h_{ab}$ on $S$—is the natural metric on the hyperboloid $\mathcal{K}$. Thus, $\xi^a$ is in $\mathcal{L}_\mathcal{G}$ if and only if (i)′ $\pounds_{\boldsymbol{\xi}}\mathbf{h}_{ab}=0$ on $\mathcal{K}$, and (ii)′ $\pounds_\xi v^a=0$ on $S$.

Consider, first, the case when $\xi^a$ in $\mathcal{L}_\mathcal{G}$ is such that its projection $\boldsymbol{\xi}^a$ on $\mathcal{K}$ vanishes. Then, the condition (i)′ above is trivially satisfied. Furthermore, $\xi^a$ is itself a vertical vector field. Hence there exists a scalar field $f_\xi$ on $S$ such that $\xi^a=f_\xi v^a$. (Recall that $v^a$ is nowhere vanishing.) Condition (ii)′ is now satisfied if and only if $\pounds_v f_\xi=0$, i.e., if and only if $f_\xi$ is the pullback to $S$ of some scalar field $\mathbf{f}_\xi$ on $\mathcal{K}$. Thus, there is a one-to-one correspondence between (arbitrary) scalar fields on $\mathcal{K}$ and elements of $\mathcal{L}_\mathcal{G}$ whose projection on $\mathcal{K}$ vanishes. We shall call these elements of $\mathcal{L}_\mathcal{G}$ *infinitesimal Spi supertranslations*. What structure do we have on the collection $\mathcal{L}_S$ of these elements? First, there exists a natural vector space structure. (The correspondence between $\xi^a$ and $\mathbf{f}_\xi$ is vector space structure preserving.) Consider, next, the Lie bracket. Given any two elements $\xi^a$ and $\xi'^a$ of $\mathcal{L}_S$, we have $[\xi,\xi']^a=[f_\xi v, f_{\xi'}v]^a=0$. Thus, $\mathcal{L}_S$ is closed under the Lie bracket operation; in fact, $\mathcal{L}_S$ is an *Abelian* sub Lie-algebra of $\mathcal{L}_\mathcal{G}$. Consider, finally, the Lie bracket $[\mu,\xi]^a$ where $\mu^a$ is in $\mathcal{L}_S$ and $\xi^a$ in $\mathcal{L}_\mathcal{G}$. Then, we have $[\mu,\xi]^a=[\mu,f_\xi v]^a=(\pounds_\mu f_\xi)v^a$. Furthermore, $\pounds_v(\pounds_\mu f_\xi)=\pounds_\mu\pounds_v f_\xi=0$. Thus, the Lie bracket $[\mu,\xi]^a$ is itself an infinitesimal Spi supertranslation. Therefore, it follows that $\mathcal{L}_S$ is not only an Abelian Lie subalgebra of $\mathcal{L}_\mathcal{G}$, but an ideal!

Consider the quotient $\mathcal{L}_\mathcal{G}/\mathcal{L}_S$. By its construction, $\mathcal{L}_\mathcal{G}/\mathcal{L}_S$ has the structure of a Lie algebra. Each element of $\mathcal{L}_\mathcal{G}/\mathcal{L}_S$ is an equivalence class of infinitesimal Spi symmetries. Denote by $\{\mu^a\}$ the equivalence class to which $\mu^a$ in $\mathcal{L}_\mathcal{G}$ belongs. Then, $\{\mu^a\}=\{\mu'^a\}$ if and only if $\mu'^a-\mu^a$ is an infinitesimal Spi supertranslation, and therefore, in particular, a vertical vector field in $S$. Hence, it follows that all vector fields belonging to the same equivalence class $\{\mu^a\}$ give rise to the same vector field $\boldsymbol{\mu}^a$ on $\mathcal{K}$ when projected. That is, every element of $\mathcal{L}_\mathcal{G}/\mathcal{L}_S$ has an unambiguous projection. Furthermore, if $\{\mu^a\}$ is a nonzero element of $\mathcal{L}_\mathcal{G}/\mathcal{L}_S$, then the projected vector field $\boldsymbol{\xi}^a$ is also nonzero. What conditions does the projection $\boldsymbol{\xi}^a$ satisfy? It just satisfies the condition (ii)′ above: $\pounds_{\boldsymbol{\xi}}\mathbf{h}_{ab}=0$ on $\mathcal{K}$. Thus, the projection map provides us with a natural, onto, linear mapping from the space $\mathcal{L}_\mathcal{G}/\mathcal{L}_S$ to the space of Killing fields on $(\mathcal{K},\mathbf{h}_{ab})$ with trivial kernel. Hence, as vector spaces the two are isomorphic. Furthermore, using the fact that $\mathcal{L}_S$ is Abelian, it is easy to check that the isomorphism also preserves the Lie algebra structure. Recall, however, that the Lie algebra of Killing vector fields on a unit hyperboloid is isomorphic to the Lie algebra $\mathcal{L}_L$ of the Lorentz group $L$. Thus, $\mathcal{L}_\mathcal{G}/\mathcal{L}_S$ is the Lorentz Lie algebra.

We summarize. The Lie algebra $\mathcal{L}_\mathcal{G}$ has an infinite-dimensional Abelian (Lie) ideal $\mathcal{L}_S$ and the quotient $\mathcal{L}_\mathcal{G}/\mathcal{L}_S$ is just the Lorentz Lie algebra. This situation is quite analogous to that in the null regime: The BMS Lie algebra has exactly the same structure. The only difference in the two cases is the "size" of the supertranslation ideal: Whereas Spi supertranslations are in one-to-one correspondence with functions on the 3-manifold $\mathcal{K}$, BMS supertranslations correspond to functions on the 2-sphere of generators of $\mathscr{I}$.

The analogy goes even deeper. Consider the functions $\mathbf{f}(k)$ on $\mathcal{K}$ of the type $\mathbf{f}(k)=k_a\eta^a$ where $k_a$ is any covector at $i^0$ and $\eta^a$ is the position vector of points on the hyperboloid $\mathcal{K}$, in the tangent space of $i^0$. Next, consider infinitesimal supertranslations of the type $f(k)v^a$ on $S$. They form a four-dimensional Abelian Lie algebra. Denote it by $\mathcal{L}_T$. We claim that $\mathcal{L}_T$ is in fact a (Lie) ideal of $\mathcal{L}_{\mathcal{G}}$: Given any element $\mu^a$ of $\mathcal{L}_{\mathcal{G}}$, it is easy to verify that $[\mu, f(k)v]^a = (\mathcal{L}_\mu f(k))v^a$ is again in $\mathcal{L}_T$. Furthermore, it is easy to show that every Killing field in the physical space—time gives rise to a unique element of $\mathcal{L}_{\mathcal{G}}$, and in the case of Minkowski space—time, elements of $\mathcal{L}_{\mathcal{G}}$ which thus correspond to space—time translations are precisely the elements of $\mathcal{L}_T$.[20] Hence, we shall call elements of $\mathcal{L}_T$ *infinitesimal Spi-translations*. The existence of this four-dimensional ideal is yet another facet of the similarity between the Lie algebra $\mathcal{L}_{\mathcal{G}}$ and the BMS Lie algebra.

We now summarize the implications of the above analysis of infinitesimal transformations on *finite* ones, i.e., on elements of $\mathcal{G}$. The group $\mathcal{G}$ of asymptotic symmetries has an infinite-dimensional Abelian normal subgroup—the subgroup $\mathcal{S}$ of Spi supertranslations. (The vector space of generators of these supertranslations is *naturally* isomorphic to that of functions on the hyperboloid $\mathcal{K}$.) The quotient of $\mathcal{G}$ by this subgroup is just the Lorentz group: $\mathcal{G}$ is the semidirect product of the Spi supertranslation group and the Lorentz group. Finally, $\mathcal{G}$ admits a preferred four-dimensional *normal* subgroup—the subgroup $\mathcal{T}$ of Spi translations. Thus, in its structure, $\mathcal{G}$ is very analogous to the BMS group.[35]

We can now discuss the relation between conformal transformations on the completed space—time and Spi supertranslations, mentioned in Sec. 3. Fix a conformal transformation $\hat{g}_{ab}\to\tilde{g}_{ab}=\omega^2\hat{g}_{ab}$. Then $\omega|_{i^0}=1$ and $\omega$ must be $C^{>0}$ at $i^0$ and $C^2$ elsewhere. Recall that, given a metric $\hat{g}_{ab}$ which is $C^{>0}$ at $i^0$, each point of Spi can be labeled by the acceleration $\hat{a}$ of the corresponding (equivalence class of) curves at $i^0$. Since under $\hat{g}_{ab}\to\tilde{g}_{ab}=\omega^2\hat{g}_{ab}$ we have $\hat{a}\to\tilde{a} = \hat{a}+(\eta^a\hat{\nabla}_a\omega)|_{i^0}\equiv\hat{a}+\bar{f}$ (where $\bar{f}=(\eta^a\hat{\nabla}_a\omega)|_{i^0}$ is a smooth function on the hyperboloid $\mathcal{K}$), the action of the fixed conformal transformation on Spi is just the motion along each fibre $F$ a parameter distance given by the value of $\bar{f}$ at the point on $\mathcal{K}$ defined (via projection) by $F$. This motion is clearly the supertranslation defined by the function $\bar{f}$ on $\mathcal{K}$. Thus, there is a natural homomorphism from the group of conformal transformations on the unphysical space—time onto the supertranslation group. [The kernel of this homomorphism is, of course, the subgroup of conformal transformations $\hat{g}_{ab}\to\omega^2\hat{g}_{ab}$ for which $(\eta^a\hat{\nabla}_a\omega)|_{i^0}=0$.] What are the conformal rescalings corresponding to Spi translations? Since translations arise from functions (on $\mathcal{K}$) of the type $\bar{f}=\omega_a\eta^a$, it follows that the corresponding conformal rescalings $\hat{g}_{ab}\to\omega^2\hat{g}_{ab}$ are precisely those for which $\omega$ is $C^1$—rather than $C^{>0}$—at $i^0$. (Thus, if the rescaled metric at $i^0$ were $C^1$ rather than $C^{>0}$, one could have eliminated the "supertranslation freedom" entirely and obtained, as one's group of asymptotic symmetries at spatial infinity, the Poincaré group instead of the infinite-dimensional group $\mathcal{G}$. Note, however, that the presence of a $C^1$ metric would imply[20] vanishing of the ADM 4-momentum.) It is curious that the differentiability requirements at $i^0$ should govern so many aspects of the asymptotic behavior of the gravitational field.

Finally, consider the issue of the selection of a preferred Poincaré subgroup of $\mathcal{G}$. Fix in $S$ any 4-parameter family of cross sections which is left invariant by all Spi translations. Consider now the subgroup of $\mathcal{G}$ which leaves this family invariant. Since the translation subgroup of $\mathcal{G}$ is itself four-dimensional, it follows that the only supertranslations which will leave this family invariant are translations. As a result, the required subgroup of $\mathcal{G}$ will be a Poincaré group. In Sec. 6 we shall see that for the class of AEFANSI space—times satisfying an additional condition on the asymptotic behavior of the Weyl tensor, one can indeed select a translation invariant 4-parameter family of cross sections of $S$ in a canonical way. Thus, for this class of AEFANSI space—times, the group of asymptotic symmetries at spatial infinity is just the Poincaré group.

## II. PHYSICAL FIELDS AT SPATIAL INFINITY

## 5. ASYMPTOTIC FIELD EQUATIONS

In this section, we examine the asymptotic behavior of physical fields (at spatial infinity) and obtain the asymptotic field equations. Apart from its intrinsic interest, this discussion will prove to be crucial for definitions of conserved quantities in the next section.

The present section is divided into four parts. In the first, we consider the scalar field, in the second, the electromagnetic field, in the third, the gravitational field and, in the fourth, certain potentials for the gravitational field. The main ideas in the analysis are the same for all three fields. We begin by considering fields which, via conditions (iii.a) and (iv)′ in the definition of asymptotic flatness (Sec. 2) admit regular[21] direction-dependent limits at $i^0$. Since the limits depend only on "the direction of approach" to $i^0$, they induce (smooth) tensor fields on the hyperboloid $\mathcal{K}$ of unit spacelike vectors in the tangent space at $i^0$. Information about the (highest order) asymptotic behavior of fields is now coded in the corresponding tensor fields on $\mathcal{K}$. Finally, we consider the equations satisfied by the various physical fields on the completed space—time, and, by taking limits of these equations, obtain the asymptotic field equations for fields on $\mathcal{K}$.

### A. Scalar fields

The analysis of the asymptotic properties of zero rest-mass scalar fields is, by itself, not of direct physical interest. The main purpose of this subsection is rather to introduce certain mathematical techniques which will be used extensively for electromagnetic and gravitational fields later in this section.

Denote by $\phi$ a scalar field on the physical space–time $(\mathbf{M}, g_{ab})$ satisfying $\nabla^a\nabla_a\phi - \frac{1}{6}R\phi = 0$ where $\nabla$ and $R$ are respectively the derivative operator and the scalar curvature defined by $g_{ab}$. Then $\hat{\phi} = \Omega^{-1}\phi$ satisfies[4] $\hat{\nabla}^a\hat{\nabla}_a\hat{\phi} - \frac{1}{6}\hat{R}\hat{\phi} = 0$ on $(\mathbf{M}, \hat{g}_{ab})$. Condition (iv)′ in the definition of asymptotic flatness requires that we restrict ourselves to fields for which $\Omega^{1/2}\hat{\phi}$ admits a regular direction-dependent limit at $i^0$. Set $\boldsymbol{\phi}(\eta) = \lim_{\to i^0}\Omega^{1/2}\hat{\phi}$ where $\eta^a$ denotes the (unit) tangent to the curve of approach at $i^0$ along which the limit is taken. Then, $\boldsymbol{\phi}(\eta)$ induces a scalar field on $\mathcal{K}$ which we shall denote by $\boldsymbol{\phi}$. It is this $\boldsymbol{\phi}$ which represents the "asymptotic scalar field" corresponding to $\phi$. Regularity conditions on the limit require that, on $\mathcal{K}$, $\boldsymbol{\phi}$ be a smooth function and that the derivatives $\boldsymbol{\partial}_{a_1}\cdots\boldsymbol{\partial}_{a_n}\boldsymbol{\phi}(\eta)$ of $\boldsymbol{\phi}(\eta)$ with respect to the argument $\eta^a$ and the derivatives of $\hat{\phi}$ on $(\mathbf{M}, \hat{g}_{ab})$ be related via

$$\boldsymbol{\partial}_{a_1}\cdots\boldsymbol{\partial}_{a_n}\boldsymbol{\phi}(\eta) = \lim_{\to i^0}(\Omega^{1/2}\hat{\nabla}_{a_1})\cdots(\Omega^{1/2}\hat{\nabla}_{a_n})\Omega^{1/2}\hat{\phi}. \tag{1}$$

Furthermore, the $\boldsymbol{\partial}_a$ derivative turns out to be related to the derivative operator $\mathbf{D}_a$ defined on the hyperboloid $\mathcal{K}$ by its intrinsic metric $\mathbf{h}_{ab}$ as follows: $\mathbf{D}_{a_1}\cdots\mathbf{D}_{a_n}\boldsymbol{\phi}$ is the field induced on $\mathcal{K}$ by the direction dependent tensor $\mathbf{h}_{a_1}{}^{b_1}\cdots\mathbf{h}_{a_n}{}^{b_n}\boldsymbol{\partial}_{b_1}\cdots\boldsymbol{\partial}_{b_n}\boldsymbol{\phi}(\eta)$. (For details, see Appendix A.) In essence, the regularity conditions demand that, as one approaches $i^0$, $\Omega^{1/2}\hat{\phi}$ have (finite) discontinuities only in "radial directions," i.e., that it be smooth in its "angular behavior."

We can now obtain the asymptotic field equations satisfied by $\boldsymbol{\phi}$ on $\mathcal{K}$. Since

$$\hat{\nabla}^a\hat{\nabla}_a\hat{\phi} = \Omega^{-3/2}\{\Omega^{1/2}\hat{\nabla}^a\Omega^{1/2}\hat{\nabla}_a\Omega^{1/2}\hat{\phi} - \tfrac{1}{2}(\hat{\nabla}^a\hat{\nabla}_a\Omega)(\Omega^{1/2}\hat{\phi}) - \tfrac{3}{2}\Omega^{1/2}(\hat{\nabla}_a\Omega)(\hat{\nabla}^a\hat{\phi})\} \tag{2}$$

holds, since $\lim_{\to i^0}\hat{\nabla}_a\Omega^{1/2}$ is $\eta_a$, the unit tangent at $i^0$ to the curve of approach, and since[21] $\eta^a\boldsymbol{\partial}_a[\mathbf{T}^{m\cdots n}{}_{p\cdots q}(\eta)] = 0$ for any regular direction-dependent tensor $\mathbf{T}^{m\cdots n}{}_{p\cdots q}(\eta)$, we have the identity

$$\mathbf{D}^a\mathbf{D}_a\boldsymbol{\phi} - \boldsymbol{\phi} = \lim_{\to i^0}(\hat{\nabla}^a\hat{\nabla}_a\hat{\phi}). \tag{3}$$

We now use the field equation $\hat{\nabla}^a\hat{\nabla}_a\hat{\phi} - \frac{1}{6}\hat{R}\hat{\phi} = 0$ for $\hat{\phi}$. Because the metric $\hat{g}_{ab}$ is $C^{>0}$ at $i^0$, $\Omega^{1/2}\hat{R}_{abcd}$ admits a regular direction-dependent limit there[21] and hence $\lim_{\to i^0}\Omega R = 0$. Equation (3) therefore reduces to

$$\mathbf{D}^a\mathbf{D}_a\boldsymbol{\phi} - \boldsymbol{\phi} = 0. \tag{4}$$

Thus, the (leading order) asymptotic behavior of zero-rest-mass scalar fields in the physical space–time is described by scalar fields $\boldsymbol{\phi}$ on the hyperboloid $\mathcal{K}$ subject to Eq. (4).

### B. Electromagnetic fields

Fix, in the physical space–time, a Maxwell field $F_{ab}$ whose sources are confined to some world tube.[36] Outside this world tube, we have $\hat{\nabla}_a F^{ab} = 0$ and $\nabla_a {}^*F^{ab} = 0$. Set $\hat{F}_{ab} = F_{ab}$. Then, on $(\mathbf{M}, \hat{g}_{ab})$, $\hat{F}_{ab}$ satisfies $\hat{\nabla}_a\hat{F}^{ab} = 0$ and $\hat{\nabla}_a{}^*\hat{F}^{ab} = 0$. As before, condition (iv)′ in the definition of asymptotic flatness gives rise to a restriction on allowable fields: We shall consider only those fields for which $\mathbf{F}_{ab}(\eta) = \lim_{\to i^0}\Omega F_{ab}$ is a regular direction-dependent tensor at $i^0$.

As in the case of the scalar field, we first introduce the field on the hyperboloid $\mathcal{K}$ which is to represent the asymptotic electromagnetic field corresponding to $F_{ab}$. We can, of course, choose for this purpose the second-rank, skew tensor field induced at each point on the hyperboloid by $\mathbf{F}_{ab}(\eta)$. However, since, in general, $[\mathbf{F}_{ab}(\eta)]\eta^b$ fails to vanish, the induced tensor field will fail to have "its indices tangential to the hyperboloid," i.e., will fail to be a tensor field *on* $\mathcal{K}$. Hence, we proceed as follows. Set $\mathbf{E}_a(\eta) = (\mathbf{F}_{ab}(\eta))\eta^b$ and $\mathbf{B}_a(\eta) = ({}^*\mathbf{F}_{ab}(\eta))\eta^b = \frac{1}{2}\epsilon_{abcd}(\mathbf{F}^{cd}(\eta))\eta^b$ where $\epsilon_{abcd}$ is the alternating tensor at $i^0$ defined by the (universal) metric $\mathbf{g}_{ab}$ $(= \lim_{\to i^0}\hat{g}_{ab})$ there. These fields, $\mathbf{E}_a(\eta)$ and $\mathbf{B}_a(\eta)$ (being annihilated when contracted with $\eta^a$) do induce vector fields $\mathbf{E}_a$ and $\mathbf{B}_a$ on $\mathcal{K}$. Since $\mathbf{F}_{ab}(\eta)$ can be reconstructed from $\mathbf{E}_a(\eta)$ and $\mathbf{B}_a(\eta)$ $[\mathbf{F}_{ab}(\eta) = 2(\mathbf{E}_{[a}(\eta))\eta_{b]} + \epsilon_{abcd}\eta^c\mathbf{B}^d(\eta)]$, it is clear that the pair $(\mathbf{E}_a, \mathbf{B}_a)$ on $\mathcal{K}$ does contain all the information carried by $\mathbf{F}_{ab}(\eta)$. (It follows from regularity of the direction-dependent limit of $\Omega^{1/2}\hat{F}_{ab}$ that $\mathbf{E}_a$ and $\mathbf{B}_a$ are smooth vector fields on $\mathcal{K}$.) One might regard $\mathbf{E}_a$ as the "asymptotic electric field with respect to the hyperboloid $\mathcal{K}$" and $\mathbf{B}_a$ as the "asymptotic magnetic field with respect to $\mathcal{K}$." (Note, however, that $\mathcal{K}$ is *timelike*. Hence, this decomposition into "electric and magnetic parts"—although analogous to—is not quite the same as the usual decomposition relative to a given observer.)

We are now ready to obtain asymptotic field equations. Consider, first, the equation $\nabla_a F^{ab} = 0$. Since

$$\hat{\nabla}_a\hat{F}^{ab} = \Omega^{-3/2}\{\Omega^{1/2}\hat{\nabla}_a(\Omega\hat{F}^{ab}) - 2\Omega\hat{F}^{ab}\hat{\nabla}_a\Omega^{1/2}\} \tag{5}$$

holds, we have, after taking the limit (and using $\lim_{\to i^0}\hat{\nabla}_a\Omega^{1/2} = \eta_a$),

$$\boldsymbol{\partial}_a\mathbf{F}^{ab}(\eta) + 2\mathbf{E}^b(\eta) = 0. \tag{6}$$

Contracting this equation with $\eta^b$ and noting that the field induced by the direction-dependent tensor $\boldsymbol{\partial}_a\eta_b$ on $\mathcal{K}$ is precisely the intrinsic metric $\mathbf{h}_{ab}$ of $\mathcal{K}$, we obtain the first asymptotic equation

$$\mathbf{D}^a\mathbf{E}_a = 0. \tag{7a}$$

Projecting (6) into the hyperboloid, on the other hand, and using the expression for $\mathbf{F}_{ab}(\eta)$ in terms of $\mathbf{E}_a$ and $\mathbf{B}_a$, we obtain

$$\mathbf{D}_{[a}\mathbf{B}_{b]} = 0. \tag{8a}$$

Equations (7a) and (8a) are, together, completely equivalent to $\lim_{\to i^0}\Omega^{3/2}\hat{\nabla}^a\hat{F}_{ab} = 0$, and hence, can be regarded as the asymptotic field equation corresponding to $\nabla^a F_{ab} = 0$. The second field equation, $\nabla^a\, {}^*F_{ab} = 0$, yields, similarly,

$$\mathbf{D}_a\mathbf{B}^a = 0 \tag{7b}$$

and

$$\mathbf{D}_{[a}\mathbf{E}_{b]} = 0. \tag{8b}$$

Thus, the (first-order) asymptotic behavior of the electromagnetic field is described completely by two divergence-free and curl-free vector fields $\mathbf{E}_a$ and $\mathbf{B}_a$ on the hyperboloid $\mathcal{K}$.[37] We shall see in the next section that total electric and magnetic charge (in the physical space—time) can be evaluated using these asymptotic fields.

### C. Gravitational fields ds

For simplicity, we shall restrict our detailed discussion to the case in which the gravitational field is source-free near infinity and only comment briefly on the modifications required by the presence of sources. This procedure is appropriate especially since the presence of physically interesting sources at infinity—the zero rest-mass fields satisfying condition (iv)′ in the definition of asymptotic flatness[24]—affects only certain intermediate steps in the analysis, leaving final equations completely unaffected.

Because the Ricci tensor $R_{ab}$ of the physical metric $g_{ab}$ vanishes near infinity by assumption, the gravitational field there is completely described by the Weyl tensor $C_{abcd}$. As shown in Appendix, the $C^{>0}$ differentiability of $\hat{g}_{ab}$ guarantees that $C_{abc}{}^{d}$ $(=\hat{C}_{abc}{}^{d})$ is such that $\Omega^{1/2}C_{abc}{}^{d}$ admits a regular direction-dependent limit $\mathbf{C}_{abc}{}^{d}(\eta)$ at $i^0$. This limit describes the asymptotic gravitational field to first order.

Our task now is to obtain smooth tensor fields on the hyperboloid $\mathcal{K}$ induced by $\mathbf{C}_{abcd}(\eta)$. To achieve this goal, we essentially repeat the procedure used in the electromagnetic case. Set $\mathbf{E}_{ab}(\eta) = \lim_{\to i^0} \times \mathbf{C}_{ambn}(\eta)\eta^m\eta^n$ and $\mathbf{B}_{ab}(\eta) = \lim_{\to i^0} {}^*\mathbf{C}_{ambn}(\eta)\eta^m\eta^n = \epsilon_{ampq}\mathbf{C}^{pq}{}_{bn}(\eta)\eta^m\eta^n$, where $\epsilon_{abcd}$ is again the alternating tensor at $i^0$ comparable with the metric $\mathbf{g}_{ab}$ there and where indices are raised and lowered using $\mathbf{g}_{ab}$. Note that contractions of both $\mathbf{E}_{ab}(\eta)$ and $\mathbf{B}_{ab}(\eta)$ with $\eta^a$ vanish. Hence, $\mathbf{E}_{ab}(\eta)$ and $\mathbf{B}_{ab}(\eta)$ induce tensor fields on the hyperboloid which we denote by $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$, respectively. It follows directly from their definition that $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$ are symmetric and trace-free. Since $\mathbf{C}_{abcd}(\eta)$ can be expressed in terms of $\mathbf{E}_{ab}(\eta)$ and $\mathbf{B}_{ab}(\eta)$, the pair $(\mathbf{E}_{ab}, \mathbf{B}_{ab})$ on $\mathcal{K}$ may now be regarded as the asymptotic gravitational field. We shall refer to $\mathbf{E}_{ab}$ as the "electric part of the asymptotic curvature relative to $\mathcal{K}$" and $\mathbf{B}_{ab}$ as the "magnetic part."

Finally, we obtain the asymptotic field equations for the gravitational field. Since in the physical space—time, $R_{ab}=0$ near infinity, the only equation of interest there is the Bianchi identity $\nabla_{[a}C_{bc]de}=0$ on the Weyl tensor. In terms of the rescaled metric $\hat{g}_{ab}=\Omega^2 g_{ab}$, this equation becomes

$$\hat{\nabla}_{[m}\hat{C}_{ab]cd} = \Omega^{-1}(\hat{g}_{c[m}\hat{C}_{ab]pd}\hat{\nabla}^p\Omega + \hat{g}_{d[m}\hat{C}_{ab]cp}\hat{\nabla}^p\Omega). \tag{9}$$

Using the identity

$$\hat{\nabla}_{[m}\hat{C}_{ab]cd} = \Omega^{-1}[\Omega^{1/2}\hat{\nabla}_{[m}\Omega^{1/2}\hat{C}_{ab]cd} - \tfrac{1}{2}(\hat{\nabla}_{[m}\Omega^{1/2})\hat{C}_{ab]cd}] \tag{10}$$

in Eq. (9), multiplying by $\Omega$, and taking the limit of the resulting equation, one obtains

$$\boldsymbol{\partial}_{[m}\mathbf{C}_{ab]cd}(\eta) = \mathbf{g}_{c[m}\mathbf{C}_{ab]pd}(\eta)\eta^p + 2\mathbf{g}_{d[m}\mathbf{C}_{ab]cp}(\eta)\eta^p + \eta_{[m}\mathbf{C}_{ab]cd}(\eta). \tag{11}$$

Equation (11) is the asymptotic field equation for $\mathbf{C}_{abcd}(\eta)$. To obtain the required equations on $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$, we project Eq. (12) into the hyperboloid $\mathcal{K}$ and contract with $\eta$ in all possible ways. The result is

$$\mathbf{D}_{[a}\mathbf{E}_{b]c} = 0 \quad\text{and}\quad \mathbf{D}_{[a}\mathbf{B}_{b]c} = 0. \tag{12}$$

Equation (12) is completely equivalent to (11). [Note that equations on $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$ are analogous to the "curl equations" (8) in the electromagnetic case. Why are there no additional "divergence equations" analogous to (7)? It is simply because these are already contained in (12): Contracting over $a$ and $c$ in (12) and using the fact that $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$ are trace-free, we obtain $\mathbf{D}^a\mathbf{E}_{ab}=0$ and $\mathbf{D}^a\mathbf{B}_{ab}=0$.]

To summarize, the asymptotic behavior of the gravitational field is described completely (to first order) by two second rank, symmetric, trace-free tensor fields on $\mathcal{K}$. The asymptotic field equations—obtained by taking limits of the Bianchi identity—are a pair of *linear* differential equations on these fields. Although in arriving at this description we have assumed that the gravitational field is source-free near infinity, the final description itself continues to be valid in the presence of sources provided the stress-energy $T_{ab}$ in $(\mathrm{M}, g_{ab})$ remains finite—more precisely, admits a regular direction-dependent limit—as one approaches $i^0$ in $\hat{\mathbf{M}}$. [In this case, only Eq. (9) is modified; the rest of the equations remain unaltered.] The stress-energy of the zero rest-mass fields permitted by our definition of asymptotic flatness does satisfy this condition. Thus, the presence of these fields leaves no direct imprint on the first order behavior of the asymptotic gravitational field: In the asymptotic description, these fields simply decouple from the gravitational field.

### D. Gravitational potentials

We now wish to display certain natural potentials for asymptotic gravitational fields $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$. The existence of these potentials will play an important role in the discussion of conserved quantities. The key idea is to use the consequence

$$\hat{\nabla}_m\hat{C}_{abc}{}^m = \hat{\nabla}_{[b}\hat{S}_{a]c} \tag{13}$$

of the Bianchi identity on $\hat{R}_{abcd}$, where $\hat{S}_{ab}=\hat{R}_{ab} - \frac{1}{6}\hat{R}\hat{g}_{ab}$. The fields induced on the hyperboloid $\mathcal{K}$ by the limiting behavior of $\hat{S}_{ab}$ will provide the required potentials.

Since $\Omega^{1/2}\hat{R}_{abcd}$ admits a regular direction-dependent limit at $i^0$,[21] so does $\Omega^{1/2}\hat{S}_{ab}$. Set $\hat{\mathbf{S}}_{ab}(\eta) = \lim_{\to i^0}\Omega^{1/2}\hat{S}_{ab}$. (This limit will turn out not to be conformally invariant. Hence, we retain the "hat" even after the limit is taken. All other limits considered so far *are* conformally invariant; we therefore dropped the "hats" after the limit was taken.) How can we represent this $\hat{\mathbf{S}}_{ab}(\eta)$ in terms of smooth fields on the hyperboloid $\mathcal{K}$? Set $\hat{\mathbf{E}}(\eta) = (\hat{\mathbf{S}}_{ab}(\eta))\eta^a\eta^b$, $\hat{\mathbf{Q}}_a(\eta) = \mathbf{h}_a{}^m(\eta)\hat{\mathbf{S}}_{mn}(\eta)\eta^n$, and $\hat{\mathbf{U}}_{ab}(\eta) = \mathbf{h}_a{}^m(\eta)\mathbf{h}_b{}^n(\eta)\hat{\mathbf{S}}_{mn}(\eta)$. Clearly, $\hat{\mathbf{S}}_{ab}(\eta)$ can be reconstructed from the triplet $(\hat{\mathbf{E}}(\eta), \hat{\mathbf{Q}}_a(\eta), \hat{\mathbf{U}}_{ab}(\eta))$. Furthermore, contraction with $\eta^a$ annihilates both $\hat{\mathbf{Q}}_a(\eta)$ and $\hat{\mathbf{U}}_{ab}(\eta)$. Hence, each element of the triplet induces a smooth field on the hyperboloid $\mathcal{K}$. Denote these fields by $\hat{\mathbf{E}}$, $\hat{\mathbf{Q}}_a$, and $\hat{\mathbf{U}}_{ab}$, respectively.

Next, we wish to show that these fields serve as natural potentials for $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$. Note first that, using Eq. (9), Eq. (13) simplifies to

$$\hat{C}_{abcm}\hat{\nabla}^m\Omega$$

$$= \Omega(\hat{\nabla}_{[b}\hat{S}_{a]c}) = \Omega^{1/2}\hat{\nabla}_{[b}(\Omega^{1/2}\hat{S}_{a]c}) - \Omega^{1/2}\hat{S}_{c[a}\hat{\nabla}_{b]}\Omega^{1/2}. \tag{14}$$

Taking the limit of this equation, one obtains

$$2\mathbf{C}_{abcm}(\eta)\eta^m - \boldsymbol{\partial}_{[b}\hat{\mathbf{S}}_{a]c}(\eta) + \eta_{[b}\hat{\mathbf{S}}_{a]c}(\eta) = 0. \tag{15}$$

Next, contracting with $\eta^b$ and using the expression for the $\mathbf{D}$ derivative on $\mathcal{K}$ in terms of the $\boldsymbol{\partial}$ derivative, one has

$$\mathbf{E}_{ab} = -\tfrac{1}{4}(\mathbf{D}_a\hat{\mathbf{Q}}_b + \hat{\mathbf{E}}\mathbf{h}_{ab}). \tag{16}$$

Finally, using the asymptotic field Eq. (12) on $\mathbf{E}_{ab}$, it follows that $\hat{\mathbf{Q}}_a = \mathbf{D}_a\hat{\mathbf{E}}$, so that

$$\mathbf{E}_{ab} = -\tfrac{1}{4}(\mathbf{D}_a\mathbf{D}_b\hat{\mathbf{E}} + \hat{\mathbf{E}}\mathbf{h}_{ab}). \tag{17}$$

Thus, the scalar field $\hat{\mathbf{E}}$ on $\mathcal{K}$ serves as a potential for the electric part of the asymptotic curvature. Similarly, contracting (15) with $\epsilon^{abpq}\eta_q$ one obtains the potential $\mathbf{B}_{ab}$,

$$\mathbf{B}_{ab} = -\tfrac{1}{4}\boldsymbol{\epsilon}_{mnb}\mathbf{D}^m\hat{\mathbf{K}}^n{}_a, \tag{18}$$

where $\hat{\mathbf{K}}_{na} = \hat{\mathbf{U}}_{na} - \hat{\mathbf{E}}\mathbf{h}_{na}$ and where $\boldsymbol{\epsilon}_{abc}$ is the natural alternating tensor field on $(\mathcal{K}, \mathbf{h}_{ab})$.

While $\hat{\mathbf{E}}$ and $\hat{\mathbf{K}}_{ab}$ are "natural" potentials for $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$—they are obtained by taking the limit of a space—time field, which, in $(\mathbf{M}, \hat{g}_{ab})$ serves as a potential for $\hat{C}_{abcd}$ [Eq. (14)]—$\mathbf{B}_{ab}$ also admits a scalar potential which has no simple interpretation in terms of space—time fields. This new potential arises from the following fact about fields on hyperboloids: *Given any symmetric tensor field* $\mathbf{T}_{ab}$ *on* $\mathcal{K}$ *with* $\mathbf{D}_{[a}\mathbf{T}_{b]c} = 0$, *there exists a scalar field* $\mathbf{T}$ *on* $\mathcal{K}$ *satisfying* $\mathbf{T}_{ab} = \mathbf{D}_a\mathbf{D}_b\mathbf{T} + \mathbf{T}\mathbf{h}_{ab}$. Thus, because of the field Eq. (12) on $\mathbf{B}_{ab}$, we know that there also exist a scalar potential $\mathbf{B}$ for $\mathbf{B}_{ab}$ with

$$\mathbf{B}_{ab} = \mathbf{D}_a\mathbf{D}_b\mathbf{B} + \mathbf{B}\mathbf{h}_{ab} \tag{19}$$

Finally, let us analyze the behavior of $\hat{\mathbf{E}}$ and $\hat{\mathbf{K}}_{ab}$ under conformal rescalings of the unphysical metric. Let $\hat{g}'_{ab}$ be any other metric in the conformal class under consideration. Then $\hat{g}'_{ab} = \omega^2\hat{g}_{ab}$, where $\omega$ is $C^{>0}$ at $i^0$ ($C^2$ on $\mathcal{K}$) with $\omega|_{i^0} = 1$. Set $\boldsymbol{\alpha}(\eta) = (\lim_{\to i^0}\hat{\nabla}_a\omega)\eta^a$ and denote by $\boldsymbol{\alpha}$ the scalar field induced on $\mathcal{K}$ by $\boldsymbol{\alpha}(\eta)$. Then, it is easy to check that

$$\hat{\mathbf{E}}' = \hat{\mathbf{E}} \quad \text{and} \quad \hat{\mathbf{K}}'_{ab} = \hat{\mathbf{K}}_{ab} - (\mathbf{D}_a\mathbf{D}_b\boldsymbol{\alpha} + \boldsymbol{\alpha}\mathbf{h}_{ab}). \tag{20}$$

Thus, $\mathbf{E}$ is conformally invariant while $\hat{\mathbf{K}}_{ab}$ is not. (From now on we can drop the "hats" over $\mathbf{E}$.) Note, however, that $\hat{\mathbf{K}}_{ab}$ *is* invariant under certain conformal rescalings: If $\mathbf{D}_a\mathbf{D}_b\boldsymbol{\alpha} + \boldsymbol{\alpha}\mathbf{h}_{ab} = 0$, $\hat{\mathbf{K}}'_{ab} = \hat{\mathbf{K}}_{ab}$. In Sec. 5 we saw that there is a homomorphism from the group of conformal rescalings onto the group of supertranslations on Spi. It is easy to check that conformal rescalings which are in the kernel of this homomorphism are precisely those for which the function $\boldsymbol{\alpha}$ on $\mathcal{K}$ vanishes. Hence, the supertranslation group (on Spi) has a natural action on the potential $\hat{\mathbf{K}}_{ab}$. What are the supertranslations corresponding to $\boldsymbol{\alpha}$'s which satisfy $\mathbf{D}_a\mathbf{D}_b\boldsymbol{\alpha} + \boldsymbol{\alpha}\mathbf{h}_{ab} = 0$? These are precisely the translations! Thus, the potential $\hat{\mathbf{K}}_{ab}$ is invariant under translations but not under any other supertranslations. We shall use this fact in the next section to single out a preferred Poincaré subgroup of the group of asymptotic symmetries.

## 6. CONSERVED QUANTITIES

This section is divided into two parts. In the first, we consider general asymptotically flat space—times (in the sense of Sec. 2) and introduce definitions of total (electric and magnetic) charge and total 4-momentum in terms of asymptotic fields. In the second, we introduce an additional requirement on the asymptotic behavior of the Weyl tensor and, for the (somewhat) restricted class asymptotically flat space—times, obtain a definition of angular momentum. Since information about dynamics of the system cannot register itself at spatial infinity, all these quantities—unlike, e.g., the Bondi 4-momentum on $\mathscr{I}$—are "absolutely conserved;" they are associated with the space—time as a whole.

### A. Charge and 4-momentum

Recall that, asymptotically, the gravitational field completely decouples from the (zero rest-mass) sources: Each field satisfies a linear differential equation which makes no reference at all to other fields. This decoupling simplifies the analysis of conserved quantities considerably. In particular, in the definition of electric and magnetic charges, we need to consider only the asymptotic electromagnetic field, and, in the definitions of energy—momentum and angular momentum, only the gravitational field.

Fix an asymptotically flat space—time endowed with an electromagnetic field $F_{ab}$ and consider the asymptotic fields $\mathbf{E}_a$ and $\mathbf{B}_a$ induced by this $F_{ab}$ on $\mathcal{K}$. Since $\mathbf{D}^a\mathbf{E}_a = 0$ and $\mathbf{D}^a\mathbf{B}_a = 0$ on $\mathcal{K}$, it follows that the right-hand sides of

$$Q_E = \int_{S^2} \mathbf{E}^a\boldsymbol{\epsilon}_{abc}dS^{bc} \tag{21}$$

and

$$Q_B = \int_{S^2} \mathbf{B}^a\boldsymbol{\epsilon}_{abc}dS^{bc} \tag{22}$$

are independent of the particular choice of the 2-sphere cross section $S^2$ of $\mathcal{K}$ used in their evaluation.[38] $Q_E$ is to be interpreted as the total electric charge and $Q_B$, the total magnetic charge, of the isolated system under consideration. One can regard the 2-sphere cross sections of $\mathcal{K}$ as being "the limits of sequences of 2-spheres in the physical space–time as their radii tend to infinity." Since $Q_E$ and $Q_B$ are independent of the cross section,[39] they can be considered as "conserved quantities." (This definition of charge may seem a bit surprising since $\mathbf{E}_a$ is not the "limiting" electric field relative to any observer, but rather, relative to the (spacelike normals to the) hyperboloid $\mathcal{K}$. Recall, however, that in the physical space–time, the total charge is defined as the integral $\int {}^*F_{ab}dS^{ab}$ over any 2-sphere surrounding the charge, i.e., as the average of the "$t-r$ component" of $F_{ab}$ over all angles. Since $\mathbf{E}_a(\eta)=\mathbf{F}_{ab}(\eta)\eta^b$ is the "radial component" of the asymptotic Maxwell field, the integral in (21) is precisely the "average of the $t-r$ component" of the asymptotic Maxwell field $\mathbf{F}_{ab}(\eta)$. Similar remarks hold for the magnetic charge.)

Next, we wish to introduce the total 4-momentum (including the contribution of the gravitational field) of the given isolated system. For this, we consider the asymptotic gravitational field, i.e., the pair $(\mathbf{E}_{ab}, \mathbf{B}_{ab})$ on $\mathcal{K}$. Since, in special relativity, the 4-momentum of a system is intimately intertwined with the group of translations, one might expect the situation to be similar in the present case. This expectation is correct: The 4-momentum emerges as a linear mapping from the space of translations to the reals. Thus, the basic definition of 4-momentum *is* tied with asymptotic symmetries on Spi. However, we will also be able to give an alternate definition which refers only to the tangent space at $i^0$, thereby avoiding technicalities associated with Spi.

Recall, first, that there is a natural vector space preserving isomorphism between the space of functions on $\mathcal{K}$ and supertranslations on Spi, and that functions on $\mathcal{K}$ which thus correspond to translations are of the type $(\mathbf{f}(k))(\eta)=k_a\eta^a$ for some vector $k_a$ in the tangent space of $i^0$. Consider the linear mapping

$$\mathbf{f}(k)\rightarrow \tfrac{1}{2}\int_{S^2}\mathbf{E}^{ab}(\mathbf{D}_b\mathbf{f}(k))\epsilon_{amn}dS^{mn} \qquad (23a)$$

from the space of translations to the reals, where $\mathbf{S}^2$ is a 2-sphere cross section of the hyperboloid. Using the definition of $\mathbf{f}(k)$, it follows that $\mathbf{D}_a\mathbf{D}_b\mathbf{f}(k) = -\mathbf{f}(k)\mathbf{h}_{ab}$. Thus, $\mathbf{D}^a\mathbf{f}(k)$ is a conformal Killing field on $\mathcal{K}$. Since $\mathbf{E}_{ab}$ is both trace and divergence free, it follows that the integral in Eq. (22) is independent of the choice of the cross section. Thus, we have obtained a conserved quantity which takes values in the dual of the vector space of translations. This is the total 4-momentum. It is not difficult to show that this conserved quantity is essentially the same as the ADM 4-momentum.[6,8] (That is, the two agree when both are defined.) In Appendix B, it is shown that the present definition yields the expected answer for Kerr space–times. Finally, note that one cannot obtain a conserved quantity by replacing translations in Eq. (23) by arbitrary supertranslations: Unlike at null infinity,[40] supermomenta do not exist in the spatial regime.

Equation (23) suggests an alternate interpretation of 4-momentum; we may regard it as a covector at $i^0$. This interpretation arises because the vector field $\mathbf{D}^a\mathbf{f}(k)$ on $\mathcal{K}$ is precisely the same as the one obtained by projecting the constant vector field $k^a$ into the hyperboloid: $\mathbf{D}^b\mathbf{f}(k)=\mathbf{h}_a{}^b k^a = k^b-(\eta^n k_m)\eta^b$. Hence, the covector $P_a$ defined by

$$P_a k^a = \tfrac{1}{2}\int_{S^2}(\mathbf{E}_{ab}k^b)\epsilon^a{}_{mn}dS^{mn} \qquad (23b)$$

at $i^0$ may itself be regarded as 4-momentum. While in this interpretation 4-momentum is not directly linked with translations—as it normally is in physics—there is, nonetheless, the advantage that it is now more directly "attached" to the (completed) space–time manifold.

From a purely mathematical viewpoint, the key step in obtaining a conserved quantity is the construction of a curl-free 2-form (on $\mathcal{K}$) using asymptotic fields: The integral of such a form on any 2-sphere cross section of $\mathcal{K}$ is automatically independent of the cross section. Since $\mathbf{E}_{ab}$ is trace and divergence free, the 2-form $\mathbf{E}_{ab}\xi^b\epsilon^a{}_{mn}$ is clearly curl free if $\xi^b$ is a conformal Killing field on the hyperboloid. We have already used four conformal Killing fields to obtain the 4-momentum. There still remain the six Killing fields on $\mathcal{K}$.[41] What are the corresponding conserved quantities? It turns out that they vanish identically: If $\xi^a$ is a Killing field on $\mathcal{K}$, the 2-form $\mathbf{E}_{ab}\xi^b\epsilon^a{}_{mn}$ is exact; *using the expression for* $\mathbf{E}_{ab}$ *in terms of its potential* $\mathbf{E}$, one can easily express this 2-form as a curl of a 1-form. Thus, using Killing fields in place of $\mathbf{D}^a\mathbf{f}(k)$ in Eq. (23a) one does not obtain any nontrivial conserved quantities.

Recall that the magnetic part $\mathbf{B}_{ab}$ of the asymptotic gravitational field satisfies the same field equation as the electric part $\mathbf{E}_{ab}$. Hence, it would appear that, using $\mathbf{B}_{ab}$ in place of $\mathbf{E}_{ab}$ in Eqs. (23), one would obtain another conserved quantity, the "magnetic" analog of the 4-momentum, or, the "angular momentum monopole-moment." From physical considerations, one would hope that this quantity should vanish. This hope is indeed borne out: Since $\mathbf{B}_{ab}$ admits a (tensor) potential $\hat{\mathbf{K}}_{ab}$ with $\mathbf{B}_{ab}=-\frac{1}{4}\epsilon_{amn}\mathbf{D}^m\hat{\mathbf{K}}^n{}_b$, the 2-form $\mathbf{B}^{ab}\mathbf{D}_b\mathbf{f}(k)\epsilon_{amn}$ is exact and hence its integral on 2-sphere cross sections vanishes identically. (From a mathematical viewpoint this result may seem surprising since the algebraic symmetries and field equations for $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$ are identical. Note, however, that the "symmetry" between $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$ is broken via the introduction of potentials: Whereas $\mathbf{B}_{ab}$ admits a *tensor* potential $\hat{\mathbf{K}}_{ab}$, with $\epsilon^{mna}\mathbf{B}_{ab} = \frac{1}{2}\mathbf{D}^{[m}\hat{\mathbf{K}}^{n]}{}_b$, $\mathbf{E}_{ab}$ does not. This difference can be traced back to Eq. (13) which relates the divergence of the Weyl tensor with the derivative of the Ricci tensor; the dual of the Weyl tensor is not related to the Ricci tensor in an analogous manner.)

Finally, one might try to construct conserved quantities using $\mathbf{B}_{ab}$ and Killing fields on $\mathcal{K}$. However, they are all zero for the same reason that analogous quantities involving $\mathbf{E}_{ab}$ are zero: $\mathbf{B}_{ab}$ also admits a *scalar* potential $\mathbf{B}$ with $\mathbf{B}_{ab}=\mathbf{D}_a\mathbf{D}_b\mathbf{B}+\mathbf{B}\mathbf{h}_{ab}$ [Eq. (19)].

To summarize, a simple analysis of asymptotic fields yields only three nontrivial conserved quantities: The electric charge, the magnetic charge, and the total 4-momentum. The first two of these are scalars while the last one takes values either in the vector space dual to the space of translations or in the cotangent space at $i^0$.

### B. Angular momentum

To obtain a satisfactory definition of angular momentum, we must first overcome two apparently distinct obstacles.

The first of these is related to the general notion of angular momentum itself. Recall that, in special relativity, the notion of angular momentum is closely related with the presence of Lorentz subgroups of the Poincaré group: It arises as a linear mapping from the Lorentz Lie algebras to the reals. Since (the connected component of the identity of) the Poincaré group admits a four-parameter family of Lorentz subgroups and since none of these subgroups is preferred over any other, angular momentum is forced to be "origin dependent;" the structure of the Lie algebra of the Poincaré group then gives rise to the familiar transformation property under the action of translations. In the transition from the Minkowski space to asymptotically flat space–times, the Poincaré group has been replaced by the infinite dimensional group $\mathcal{G}$. Consequently, the symmetry group now admits "as many" Lorentz subgroups as there are supertranslations, rather than just a four-parameter family of them. If we were to consider a linear mapping from each of the corresponding Lorentz Lie algebras to the reals and obtain a conserved quantity, this quantity would have very little resemblance to one's intuitive notion of angular momentum: It would be defined relative to an "origin" lying in an *infinite dimensional space*! Thus, to obtain a definition which respects one's intuition about angular momentum, one must first suitably restrict the supertranslation freedom: We must introduce some additional structure at spatial infinity which can reduce the infinite dimensional group of asymptotic symmetries to the Poincaré group.

The second difficulty is that, as examples show,[42] none of the asymptotic fields (and potentials) introduced so far carries information about angular momentum. Intuitively, one might expect angular momentum to arise from 2-sphere integrals involving the "magnetic" part of the asymptotic curvature. The field $\mathbf{B}_{ab}$ is, however, quite unsuitable for this purpose: Both $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$ contain information only about the "$1/r^3$ part" of the asymptotic curvature and while (from examples) one expects the 4-momentum to appear at this order—as it did—one does not expect the angular momentum to do so. (Indeed, as we saw, all the conserved quantities that one can easily construct from $\mathbf{B}_{ab}$ vanish identically.) Therefore, before we can hope to define angular momentum, we need to introduce and analyze a new asymptotic field which can capture the "$1/r^4$ contribution" to the (magnetic part of the) asymptotic curvature.

It turns out that both these obstacles can be overcome at the same stroke. Impose, on the asymptotic behavior of the Weyl tensor, the following conditions,

$$\mathbf{B}_{ab}(\eta) \equiv \lim_{\to i^0} \Omega^{1/2} \, {}^*\hat{C}_{ambn}(\hat{\nabla}^m \Omega^{1/2})(\hat{\nabla}^n \Omega^{1/2}) = 0, \quad (24a)$$

and, that the "next order" contribution to the magnetic part be asymptotically well behaved, i.e., that

$$\hat{\beta}_{ab} := \lim_{\to i^0} {}^*\hat{C}_{ambn}(\hat{\nabla}^m \Omega^{1/2})(\hat{\nabla}^n \Omega^{1/2}) \quad (24b)$$

be a regular direction-dependent tensor at $i^0$. [Note that, $\hat{\beta}_{ab}$ is also the limit of the "radial" derivative, $(\hat{\nabla}^p \Omega^{1/2})\hat{\nabla}_p(\Omega^{1/2} \, {}^*\hat{C}_{ambn}(\hat{\nabla}^m\Omega^{1/2})(\hat{\nabla}^n\Omega^{1/2})$ of the field whose limit by Eq. (24a) must vanish. We retain the "hat" on $\hat{\beta}_{ab}$ because, as we shall see, it fails to be conformally invariant.] It turns out that the vanishing of $\mathbf{B}_{ab}$ introduces additional structure at infinity, thereby eliminating the "supertranslation freedom," while the existence of $\hat{\beta}_{ab}$ leads to the expression for angular momentum. In effect, the additional conditions just demand that, in the physical space–time, the "magnetic" part of the Weyl tensor should fall off one order faster than the "electric" part. These conditions are satisfied in Kerr space–times and, although no general result has been proven, there do exist heuristic arguments which suggests that they would be satisfied for a wide class of isolated systems. We now introduce the expression for *total* angular momentum (including the contribution of the gravitational field) of isolated systems whose asymptotic gravitational field satisfies this condition.

It is convenient to proceed in two steps. First, we shall introduce angular momentum as a set of skew tensors at $i^0$, with appropriate transformation properties, and then we shall present its more accurate description in terms of the structure of Spi and of the group $\mathcal{G}$ of asymptotic symmetries.

We begin by eliminating the supertranslation freedom using condition (24a). Recall, first, that the supertranslation freedom is essentially the same as the conformal freedom in the unphysical metric: A rescaling by a function which is $C^{>0}$ at $i^0$ corresponds to a supertranslation; while one by a function which is $C^1$ at $i^0$, to a translation. Hence, the elimination of supertranslations other than translations can be achieved simply by first singling out, from the conformal class of all metrics which are $C^{>0}$ at $i^0$, a preferred subclass of metrics whose *relative* conformal factor is $C^1$ at $i^0$, and then demanding that this subclass be left invariant by (restricted) asymptotic symmetries. The idea now is to use (24a) to select the required preferred subclass. Since $\mathbf{B}_{ab}$ vanishes, it follows that its potential $\hat{\mathbf{K}}_{ab}$ satisfies $\mathbf{D}_{[a}\hat{\mathbf{K}}_{b]c} = 0$. Furthermore, $\hat{\mathbf{K}}_{ab}$ is symmetric. Hence (by a result quoted in Sec. 5), there exists a scalar field $\hat{\mathbf{K}}$ on the hyperboloid $\mathcal{K}$ such that $\hat{\mathbf{K}}_{ab} = \mathbf{D}_a\mathbf{D}_b\hat{\mathbf{K}} + \hat{\mathbf{K}}\mathbf{h}_{ab}$. Hence, using the transformation property [Eq. (20)] of $\hat{\mathbf{K}}_{ab}$ under conformal rescalings it follows that one can always choose a conformal frame in which $\hat{\mathbf{K}}_{ab} = 0$, and, that the conformal factor relating any two such frames must be $C^1$ at $i^0$. Thus, when $\mathbf{B}_{ab} = 0$, the condition $\hat{\mathbf{K}}_{ab} = 0$ selects out a preferred subclass of conformally related metrics with the desired property. Finally, it is clear from the above discussion that the only supertranslations which will leave this subclass invariant are translations.

We are now ready to define angular momentum. Using the definition of $\hat{\beta}_{ab}(\eta)$, and the vanishing of $\mathbf{B}_{ab}$, one obtains,

$$\partial^a\hat{\beta}_{ab}(\eta)=\tfrac{1}{2}\ {}^*\mathbf{C}_{ambn}(\eta)\eta^m\hat{\mathbf{K}}^{an}(\eta), \tag{25}$$

where $\hat{\mathbf{K}}^{am}(\eta)=\hat{\mathbf{U}}^{am}(\eta)-\hat{\mathbf{E}}(\eta)\mathbf{h}^{am}$ is the direction-dependent field at $i^0$ which induces on the hyperboloid $\mathcal{K}$, the tensor potential $\hat{\mathbf{K}}^{am}$ for $\mathbf{B}^{am}$. Because we have restricted ourselves to conformal frames ($\hat{g}_{ab}$'s) in which $\hat{\mathbf{K}}_{ab}=0$, we have, on $\mathcal{K}$,

$$\mathbf{D}^a\hat{\beta}_{ab}=0, \tag{26}$$

where, as per our usual notation, $\hat{\boldsymbol{\beta}}_{ab}$ is the tensor field on $\mathcal{K}$, induced by $\hat{\beta}_{ab}(\eta)$ at $i^0$. Hence, given any skew tensor $F_{ab}$ in the tangent space of $i^0$, the right-hand side of

$$\hat{M}^{ab}F_{ab}:=\tfrac{1}{2}\int_{S^2}\hat{\beta}^{ab}\xi_b\epsilon_{amn}dS^{mn} \tag{27}$$

is independent of the 2-sphere cross section of $\mathcal{K}$, where $\xi^a$ is the (Killing) field induced on $\mathcal{K}$ by the direction-dependent vector $\epsilon^{abcd}F_{cd}\eta_b$ at $i^0$. Hence Eq. (27) defines a skew tensor $\hat{M}^{ab}$ at $i^0$. Under a conformal rescaling $\hat{g}_{ab}\to\hat{g}'_{ab}=\omega^2\hat{g}_{ab}$, where $\omega$ is $C^1$ at $i^0$ (i.e., under a rescaling which leaves our preferred subclass of conformal frames invariant), one has, on $\mathcal{K}$,

$$\hat{\boldsymbol{\beta}}_{ab}\to\hat{\boldsymbol{\beta}}'_{ab}=\hat{\boldsymbol{\beta}}_{ab}+2\epsilon_{mk(a}\mathbf{E}^k{}_{b)}\mathbf{D}^m\omega. \tag{28}$$

That is, under restricted conformal transformations (which correspond to *translations* rather than arbitrary supertranslations) the magnetic part $\hat{\boldsymbol{\beta}}_{ab}$ of asymptotic curvature picks up an electric part. Finally, using Eqs. (27) and (28), it follows that,

$$\hat{M}^{ab}\to\hat{M}'^{ab}=\hat{M}^{ab}+2P^{[a}\omega^{b]}, \tag{29}$$

where $\omega_a=(\hat{\nabla}_a\omega)\,|_{i^0}$ and where $P^a$ is the 4-momentum defined via Eq. (23). Since the natural isomorphism between the translation subgroup of the asymptotic symmetry group $\mathcal{G}$ and the tangent space at $i^0$ sends the translation corresponding to the conformal rescaling $\hat{g}_{ab}\to\hat{g}'_{ab}=\omega^2\hat{g}_{ab}$ to the vector $\omega^b$ at $i^0$, (29) is the usual transformation law for angular momentum under translations. The 4-parameter family of skew tensors $\hat{M}^{ab}$ at $i^0$, obtained via Eq. (27) in conformal frames selected by the condition $\hat{\mathbf{K}}_{ab}=0$ represents the angular momentum of the isolated system under consideration.

In terms of Spi, the situation may be summarized as follows. Given any metric $\hat{g}_{ab}$ in the conformal class available, one can coordinatize the fibres of Spi in a canonical way using the tangential component $\hat{a}$ of the acceleration of curves (representing points of Spi), at $i^0$. Fix a real number $r$ and denote by $C_r(\hat{g}_{ab})$ the cross section of Spi defined by $\hat{a}=r$. [Thus, $C_r$ is a mapping from the conformal class of metrics ($C^{>0}$ at $i^0$) on to the space of smooth cross sections of Spi.] Consider the subclass of metrics $\hat{g}_{ab}$ for which $\hat{\mathbf{K}}_{ab}=0$. It is easy to check that the image, under the mapping $C_r$ of this subclass is a 4-parameter family $f_r$ of cross sections which is left invariant by translations, but by no other supertranslations. In view of the remark made at the end of Sec. 4, it follows that, when this family is included in the universal structure on Spi, the asymptotic symmetry group reduces to the Poincaré group. Thus, the condition $\hat{\mathbf{K}}_{ab}=0$ enables us to select a preferred Poincaré subgroup of $\mathcal{G}$. Finally, note that this Poincaré subgroup is independent of the initial choice of the real number $r$: Since the structure group of Spi is in the center of $\mathcal{G}$, the action on Spi of the Poincaré subgroup selected above leaves invariant the family $f_r$ of preferred cross sections *for each real number r*.

How much "freedom" do we have in the selection of a preferred Poincaré subgroup of $\mathcal{G}$? The only additional structure that we have had to introduce is the condition $\hat{\mathbf{K}}_{ab}=0$. Note, however, that *some* such condition is essential to obtain a meaningful expression of angular momentum: It follows from Eq. (25) that, in a general conformal frame, $\hat{\beta}_{ab}$ would fail to be divergence free; hence, the resulting expression for angular momentum would fail to be independent of the 2-sphere cross section used in its definition. What is the most general condition that can be imposed on $\hat{\mathbf{K}}_{ab}$ to ensure that $\hat{\beta}_{ab}$ is divergence-free? Using the assumption $\mathbf{B}_{ab}=0$, it is easy to see that we must require $\hat{\mathbf{K}}_{ab}$ to be "pure trace," i.e., to be proportional to the metric $\mathbf{h}_{ab}$ on $\mathcal{K}$. Finally, the proportionality factor can be determined by demanding that the procedure should single out the "correct" Poincaré subgroup of $\mathcal{G}$ in the case of Minkowski space—time; the factor turns out to be zero. It is in this sense that the condition—and hence the resulting Poincaré subgroup of $\mathcal{G}$—is canonical.

How is the definition of angular momentum related to this Poincaré group? Note, first, that, given a metric $\hat{g}_{ab}$ in the preferred subclass, one obtains [using the 1-parameter family $C_r(\hat{g}_{ab})$ of cross sections], a natural lifting of the Lorentz group on the hyperboloid $\mathcal{K}$ to a Lorentz subgroup of the Poincaré group on Spi. Hence, the angular momentum—defined by Eq. (27)—may be regarded, more naturally, as a mapping from the vector spaces of Lorentz Lie algebras of this Poincaré group to the reals.

It is curious to note that, for each real number $r$, the set of cross sections contained in the family $f_r$ can be given, naturally, the structure of Minkowski space. Furthermore (the 1-parameter family of) these Minkowski spaces are naturally isomorphic to each other; the structure group of Spi provides these isomorphisms. Hence, we may identify them and call the resulting space the asymptotic Minkowski space. Since the action of the Poincaré group on Spi leaves each family $f_r$ invariant, it can be extended to the asymptotic Minkowski space; as might be expected, this action just coincides with that of the natural isometry group of the asymptotic Minkowski space. Finally, conserved quantities can be expressed as tensor fields on this Minkowski space: The total 4-momentum is represented by a constant vector field, and, the angular momentum, by a second rank, skew tensor field with the usual transformation property under the change of origin.

This concludes the discussion on angular momentum. Note, finally, that since $\hat{\beta}_{ab}$ is symmetric, trace-free, and divergence-free, one might imagine constructing another conserved quantity, $\int_{S^2} \hat{\beta}^{ab}(\hat{\mathbf{D}}_b \mathbf{f}(k)\epsilon_{amn} dS^{mn}$ from $\hat{\beta}_{ab}$ using the conformal Killing fields $\hat{\mathbf{D}}^a \mathbf{f}(k)$ on $\mathcal{K}$. This quantity vanishes identically; it is not hard to show that $\hat{\beta}_{ab}$ admits a tensor potential $\hat{k}_{ab}$ with $\hat{\beta}_{ab} = -\frac{1}{4}\epsilon_{amn}\mathbf{D}^m \hat{k}^n{}_b$, so that one can repeat the argument used to show that the "magnetic" analog of the 4-momentum vanishes.

## 7. DISCUSSION

In the preceding sections, we have presented a new description of the asymptotic structure of the gravitational field at spatial infinity. In the final picture, this description has turned out to be similar to that of null infinity in many ways. Thus, the universal structure of Spi is very analogous to that of $\mathcal{I}$: The asymptotic symmetry groups in the two regimes have the same broad features, the links between conserved quantities and asymptotic symmetries are parallel, and, in both cases, conserved quantities emerge as integrals of asymptotic fields over 2-sphere cross sections of the structure at infinity.

However, there do exist at least two important differences, both of which make the spatial description seem somewhat less "natural" than the null.

The first of these is that whereas in the null regime differentiability requirements on the completed manifold and the rescaled metric are simple,[43] in the spatial regime they have turned out to be quite awkward. Could we have, somehow, avoided these complications? Let us begin by analyzing, in intuitive terms, why these complications arose. An elementary calculation[4,16] shows that, in the case of the Schwarzschild space–time, the conformal curvature of the rescaled metric must diverge at $i^0$. Since any reasonable definition of asymptotic flatness must admit this space–time as an example, one is severely constrained in one's choice of differentiability requirements: The strongest condition that one can impose is that the metric be $C^1$ at $i^0$. If we had actually imposed this condition, our analysis would have been considerably simpler; in this case, the resulting description of spatial infinity would have inherited a much richer structure. For example, *without any extra assumptions on the behavior of the asymptotic curvature*, the Poincaré group would then have emerged as the asymptotic symmetry group. However, as remarked in Sec. 2, it turns out[20] that $C^1$-differentiability also implies the vanishing of the total (ADM) 4-momentum. (Intuitively, in this case the physical metric approaches the flat metric "as $1/r^2$" rather than "as $1/r$" at spatial infinity.) Hence the $C^1$-differentiability is simply too strong. What would have been the result if we had required the metric to be only $C^0$ at $i^0$? The situation would have been just the opposite: One would have obtained too little structure to make useful constructions. For example, since the analysis of the "second-order" structure at $i^0$ requires the use of a (possibly direction-dependent) connection, with a $C^0$ metric one could have examined only the "first-order" structure. Consequently, the "blown-up" structure would have been just the hyperboloid $\mathcal{K}$, and the asymptotic symmetry group, just the Lorentz group; one would have lost all supertranslations. The situation would have been even worse in the analysis of physical fields. For if $\hat{g}_{ab}$ were only $C^0$ at $i^0$, $\Omega^{1/2}\hat{C}_{abcd}$ need not have admitted a direction-dependent limits there. Consequently physically interesting asymptotic fields could not have even been introduced in the gravitational case. Thus, if one wishes to introduce $i^0$ at all, one is forced to accept the awkward differentiability condition on the rescaled metric $\hat{g}_{ab}$.

The second difference is that whereas in the null regime $\mathcal{I}$ provides, simultaneously, a boundary for the space–time manifold, an arena for asymptotic symmetries and a home for asymptotic fields, in the spatial regime, in a sense, the three roles have become disjoint, being played, respectively, by $i^0$, Spi, and the hyperboloid $\mathcal{K}$. Could we have introduced just one structure instead of all three? One can, in fact, imagine[44] an alternate approach in which one introduces neither $i^0$ nor Spi but only a timelike 3-manifold analogous to $\mathcal{K}$ which serves as the "spatial boundary" of the space–time in the same way as $\mathcal{I}$ serves as the "null boundary." Such an approach has several advantages: One can discuss spatial infinity by itself without any reference to $\mathcal{I}$, and one can deal, throughout, with smooth manifolds and smooth tensor fields. Note, however, that there are many advantages in having both $i^0$ and Spi at one's disposal. Thus, for example, as we shall see in the next paper, the presence of $i^0$ is crucial in relating the asymptotic structure at spatial infinity with that at null. Next, the tangent space at $i^0$ provides a natural common home for various conserved quantities making it easy to relate them. Isometries in the physical space–time can also be most conveniently analyzed and classified by examining their behavior near $i^0$.[20] Furthermore, the tangent space at $i^0$ also provides a natural arena for investigating the relation between these isometries and the corresponding conserved quantities. For example, one expects that, in stationary space–times, the "asymptotic rest frame" defined by the Killing field should coincide with that defined by the total (ADM) 4-momentum. While it seems rather difficult to obtain an unambiguous formulation of this conjecture in the absence of $i^0$, it is easy to obtain not only a formulation but also a proof using the tangent space at $i^0$. Similarly, the structure made available by Spi is useful in several ways. For example, it seems difficult to obtain a *"local"* characterization of the asymptotic symmetry group in the absence of Spi: If one has available only the hyperboloid $\mathcal{K}$—or, a suitable analog thereof—one must apparently introduce asymptotic symmetries as diffeomorphisms *in the space–time manifold* which preserve asymptotic conditions. While such a description of asymptotic symmetries may seem satisfactory mathematically, it is rather awkward from an aesthetic viewpoint; the main reason behind the introduction of boundaries is that one wishes to analyze the asymptotic structure of space—times using *local* differential geometry *at* their boundaries. Finally, it is the structure made available by Spi that enables one to introduce such notions as that of the "asymptotic Minkowski space at spatial infinity." The existence of these notions provides, in turn, mathematical tools to obtain precise formulations—as well as proofs—of intuitive conjectures concerning isolated systems. Thus, in absence of either $i^0$ or Spi, our analysis of the asymptotic structure of space—time would have been seriously hampered.

## ACKNOWLEDGMENTS

We are grateful to Robert Geroch, Anne Magnon-Ashtekar, Jerrold Marsden, Niall O'Murchadha, Bernd Schmidt, Paul Sommers, Martin Walker, and especially Roger Penrose for discussions. In addition, thanks are due to Anne Magnon-Ashtekar for pointing out an important error in a previous version of this paper, and to Paul Sommers for correspondence on his approach to spatial infinity. A part of this work was done while one of us (A. A.) held a visiting membership at the Max Planck Institut für Astrophysik, Munich; he wishes to thank the Munich relativity group for their hospitality.

## APPENDIX A: MATHEMATICAL PRELIMINARIES

In this Appendix, the differentiability conditions and conditions on tensor fields at $i^0$ are discussed.

We begin by defining direction-dependent tensors at $i^0$. Let $(\hat{\mathbf{M}}, g_{ab})$ be given as a $C^1$ manifold with $C^0$ metric $\hat{g}_{ab}$, where $\hat{g}_{ab}$ is $C^3$ on the image of $\mathbf{M}$ in $\hat{\mathbf{M}}$. Let $\Gamma$ be the family of $C^1$ spacelike curves in $\hat{\mathbf{M}}$, $C^3$ on the image of $\mathbf{M}$, passing through $i^0$; and let $\gamma \in \Gamma$ have unit tangent vector $\eta^a$ at $i^0$. A $C^3$ tensor field $T^{a\cdots b}{}_{c\cdots d}$ on the image of $M$ is said to have a *regular direction-dependent limit* at $i^0$ if:

(i) The limit of $T^{a\cdots b}{}_{c\cdots d}$ along $\gamma$ at $i^0$ exists for all $\gamma \in \Gamma$, and depends only on $\eta^a$. We write $\lim_{\to i^0} T^{a\cdots b}{}_{c\cdots d} = \mathbf{T}^{a\cdots b}{}_{c\cdots d}(\eta)$ for this limit.

(ii) The derivatives of all orders of $\mathbf{T}^{a\cdots b}{}_{c\cdots d}(\eta)$ with respect to $\eta^a$ exist; and if $\partial_e \mathbf{T}^{a\cdots b}{}_{c\cdots d}(\eta)$ denotes the derivative of $\mathbf{T}^{a\cdots b}{}_{c\cdots d}(\eta)$ with respect to $\eta^e$, then

$$\partial_e \mathbf{T}^{a\cdots b}{}_{c\cdots d}(\eta) = \lim_{\to i^0} \Omega^{1/2} \hat{\nabla}_e T^{a\cdots b}{}_{c\cdots d}$$

holds, where $\hat{\nabla}$ is the derivative operator associated with $\hat{g}_{ab}$.

The role of condition (i) is clear: A tensor field $T^{a\cdots b}{}_{c\cdots d}$ for which $\lim_{\to i^0} T^{a\cdots b}{}_{c\cdots d}$ exists defines a mapping from the unit hyperboloid $\mathcal{K}$ of spacelike vectors at $i^0$ to tensors at $i^0$ having the same index structure as $T^{a\cdots b}{}_{c\cdots d}$. Condition (ii) ensures that $\mathbf{T}^{a\cdots b}(\eta)$ is a smooth mapping, and that the derivative of $T^{a\cdots b}{}_{c\cdots d}(\eta)$ is given by the limiting formula which holds for direction-dependent tensors at $i^0$ in Minkowski space—time. (The derivative operator $\partial_a$ is defined by:

$$k^a \partial_a \mathbf{T}^{m\cdots n}{}_{p\cdots q}(\eta) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} [\mathbf{T}^{m\cdots n}{}_{p\cdots q}(\eta + \epsilon k) - \mathbf{T}^{m\cdots n}{}_{p\cdots q}(\eta)]$$

for any vector field $k^a$ on $\mathcal{K}$, and $\eta^a \partial_a \mathbf{T}^{m\cdots n}{}_{p\cdots q} = 0$.)

If $\mathbf{T}^{a\cdots b}{}_{c\cdots d}(\eta)$ is orthogonal to $\eta^a$ (on all its indices), it defines, naturally, a tensor field on $\mathcal{K}$ which we shall denote by $\mathbf{T}^{a\cdots b}{}_{c\cdots d}$. However, $\partial_e \mathbf{T}^{a\cdots b}{}_{c\cdots d}(\eta)$ will not, in general, be orthogonal to $\eta^a$ (on indices other than $e$) even if $\mathbf{T}^{a\cdots b}{}_{c\cdots d}(\eta)$ is. It is therefore useful to introduce a derivative operator $\mathbf{D}_a$ which acts on direction-dependent tensors at $i^0$ and which preserves orthogonality to $\eta^a$.

The metric $\hat{g}_{ab}$ has a limit at $i^0$, $\lim_{\to i^0} \hat{g}_{ab} = \mathbf{g}_{ab}$, which is *direction independent* (since $\hat{g}_{ab}$ is $C^{>0}$): $\partial_a \mathbf{g}_{bc}(\eta) = 0$. The direction-dependent tensor $\mathbf{h}_{ab}(\eta) = \mathbf{g}_{ab}(\eta) - \eta_a \eta_b$ is orthogonal to $\eta^a$, and so defines a tensor field $\mathbf{h}_{ab}$ on $\mathcal{K}$. This $\mathbf{h}_{ab}$ is the natural metric on $\mathcal{K}$. For $\mathbf{T}^{a\cdots b}{}_{c\cdots d}(\eta)$ orthogonal to $\eta^a$, $\mathbf{D}_e \mathbf{T}^{a\cdots b}{}_{c\cdots d}$ is defined by

$$\mathbf{D}_e \mathbf{T}^{a\cdots b}{}_{c\cdots d}(\eta) = \mathbf{h}_e{}^m(\eta) \mathbf{h}^a{}_n(\eta) \cdots \mathbf{h}^b{}_p(\eta) \mathbf{h}_c{}^q(\eta) \cdots \mathbf{h}_d{}^r(\eta) \times \partial_m \mathbf{T}^{n\cdots p}{}_{q\cdots r}(\eta).$$

Clearly, $\mathbf{D}_e \mathbf{T}^{a\cdots b}{}_{c\cdots d}(\eta)$ is orthogonal to $\eta$, and so corresponds to a tensor field $\mathbf{D}_e \mathbf{T}^{a\cdots b}{}_{c\cdots d}$ on $\mathcal{K}$. Furthermore, it is easily verified that $\mathbf{D}_a$ defines a derivative operator on $\mathcal{K}$. In fact, this derivative operator is the covariant derivative associated with $\mathbf{h}_{ab}$; for since $\partial_a \mathbf{g}_{bc}(\eta) = 0$, $\partial_a \eta_b = \mathbf{h}_{ab}(\eta)$ hold, we have $\mathbf{D}_a \mathbf{h}_{bc} = 0$.

We now define the $C^{>1}$ differentiable structure at $i^0$. Let $(U, \phi)$ and $(V, \psi)$ be charts containing $i^0$ in the $C^1$ atlas of $\hat{\mathbf{M}}$, such that the restrictions of $\phi$ and $\psi$ to the image of $\mathbf{M}$ are $C^4$ maps. The chart $(V, \psi)$ is said to be $C^{>1}$ compatible with $(U, \phi)$ at $i^0$ if $(\partial^2(\psi_i \circ \phi^{-1})/\partial x_j \partial x_k) \circ \phi$ and $(\partial^2(\phi_i \circ \psi^{-1})/\partial x_j \partial x_k) \circ \psi$ have direction-dependent limits at $i^0$ for all $j$, $k$, and $i$, where $\phi_i$ and $\psi_i$ are the $i$th component maps of $\phi$ and $\psi$. The family of all charts $C^{>1}$ compatible with $(U, \phi)$ at $i^0$ is the $C^{>1}$ differentiable structure on $\hat{\mathbf{M}}$ at $i^0$ compatible with $(U, \phi)$.

A $C^3$ tensor field $T^{a\cdots b}{}_{c\cdots d}$ on the image of $\mathbf{M}$ is said to be $C^{>0}$ at $i^0$ with respect to a $C^{>1}$ differentiable structure at $i^0$ if the first derivatives of its components in a chart (and thus in all charts) belonging to the $C^{>1}$ differentiable structure have regular direction-dependent limits at $i^0$. More generally, a tensor field whose components have vanishing derivatives up to order $k-1$ at $i^0$, whose $k$th derivatives are direction-independent at $i^0$, and whose $(k+1)$th derivatives have direction-dependent limits at $i^0$, is said to be of class $C^{>k}$ at $i^0$.

The condition that $\hat{g}_{ab}$ be $C^{>0}$ at $i^0$ with respect to a differentiable structure determines the differentiable structure uniquely; for if the first derivatives of the components of $\hat{g}_{ab}$ in two charts containing $i^0$ have direction-dependent limits at $i^0$, the two charts must be $C^{>1}$ compatible with each other.[20] The $C^{>1}$ differentiable structure with respect to which the metric $\hat{g}_{ab}$ is $C^{>0}$ is called the $C^{>1}$ differentiable structure on $\hat{\mathbf{M}}$ at $i^0$.

Finally, note that, from the requirement that $\hat{g}_{ab}$ be $C^{>0}$ at $i^0$, it follows that $\lim_{\to i^0}\Omega^{1/2}\hat{R}_{abcd} = \mathbf{R}_{abcd}(\eta)$ defines a direction-dependent tensor at $i^0$. Since the first derivatives of the components of $\hat{g}_{ab}$ have regular direction-dependent limits at $i^0$, so do the components of its connection. Condition (ii) of the definition of a regular direction-dependent tensor, and the definition of the curvature tensor in terms of the connection then imply the existence of $\mathbf{R}_{abcd}(\eta)$.

## APPENDIX B: ASYMPTOTICALLY FLAT INITIAL DATA SETS IN AEFANSI SPACE-TIMES

Fix an AEFANSI space-time $(\mathbf{M}, g_{ab})$ and denote by $(\hat{\mathbf{M}}, \hat{g}_{ab})$ any of its AEFANSI completions. For simplicity, in this appendix we shall not explicitly distinguish between $\mathbf{M}$ and its image $\psi(\mathbf{M})$ under the imbedding map $\psi$. Thus, we allow ourselves to write "$\hat{g}_{ab} = \Omega^2 g_{ab}$ on $\mathbf{M}$" rather than "$\psi_*(\hat{g}_{ab}) = \psi_*(\Omega^2) g_{ab}$ on $\mathbf{M}$."

We begin by showing that there do exist initial data sets in $(\mathbf{M}, g_{ab})$ which satisfy the ADM—G[8] conditions:

*Theorem*: Let $\hat{\mathbf{T}}$ be any three-dimensional, spacelike submanifold of $(\hat{\mathbf{M}}, \hat{g}_{ab})$ such that $i^0 \in \hat{\mathbf{T}}$, $\hat{\mathbf{T}}$ is $C^{>1}$ at $i^0$ and $C^3$ elsewhere. Then the initial data induced on $\mathbf{T} = \hat{\mathbf{T}} - i^0$ by $g_{ab}$ is asymptotically flat in the ADM—G sense.

*Sketch of the proof*: Identify $i^0$ with the point $\Lambda$ in the ADM—G framework (see Sec. 2), and use the pullback $\Omega_{\hat{\mathbf{T}}}$ of $\Omega$ to $\hat{\mathbf{T}}$ as the ADM—G conformal factor. Denote by $\hat{q}_{ab}$ the metric induced on $\hat{\mathbf{T}}$ by $\hat{g}_{ab}$ and by $q_{ab}$, the metric induced on $\mathbf{T}$ by $g_{ab}$. Then, on $\mathbf{T}$, $\hat{q}_{ab} = \Omega^2_{\hat{\mathbf{T}}} q_{ab}$. Since $\hat{g}_{ab}$ is $C^{>0}$ at $i^0$, and since $\hat{\mathbf{T}}$ is $C^{>1}$, it follows that $\hat{q}_{ab}$ must be $C^0$ at $i^0$. Next, consider the conformal factor $\Omega_{\hat{\mathbf{T}}}$. Because $\Omega$ is $C^2$ at $i^0$ with $\Omega\,|_{i^0} = 0$, $\hat{\nabla}_a\Omega\,|_{i^0} = 0$, and $(\hat{\nabla}_a\hat{\nabla}_b\Omega - 2\hat{g}_{ab})\,|_{i^0} = 0$, it follows that on $\hat{\mathbf{T}}$, $\Omega_{\hat{\mathbf{T}}}$ is also $\hat{C}^2$ at $i^0$ and satisfies $\Omega_{\hat{\mathbf{T}}}\,|_{i^0} = 0$, $\hat{D}_a\Omega_{\hat{\mathbf{T}}}\,|_{i^0} = 0$, and $(\hat{D}_a\hat{D}_b\Omega_{\hat{\mathbf{T}}} - 2\hat{q}_{ab})\,|_{i^0} = 0$, where $\hat{D}_a$ is the natural derivative operator on $(\mathbf{T}, \hat{q}_{ab})$. Next, we must show that $\Omega^{-1/2}(\hat{D}_a\hat{D}_b\Omega_{\hat{\mathbf{T}}} - 2\hat{q}_{ab})$ admits a direction-dependent limit at $i^0$. Recall first (from Appendix A) that, since $\hat{g}_{ab}$ is $C^{>0}$, its Riemann tensor $\hat{R}_{abcd}$ has the property that, in $\hat{\mathbf{M}}$, $\Omega^{1/2}\hat{R}_{abcd}$—and hence, $\Omega^{1/2}\hat{R}_{ab}$—admits a regular direction-dependent limit at $i^0$. Using the fact that the Ricci tensor $R_{ab}$ of $g_{ab}$ vanishes in a neighborhood of $i^0$ and the consequent expression for $\hat{R}_{ab}$ in terms of $\Omega$ and its derivatives, it now follows that $\Omega^{-1/2}(\hat{\nabla}_a\hat{\nabla}_b\Omega - 2\hat{g}_{ab})$ admits a regular direction-dependent limit at $i^0$. Finally, projecting the tensor field $\Omega^{-1/2}(\hat{\nabla}_a\hat{\nabla}_b\Omega - 2\hat{g}_{ab})$ into $(\mathbf{T}, \hat{q}_{ab})$ and using the fact that, since $\hat{\mathbf{T}}$ is $C^{>0}$ at $i^0$, the extrinsic curvature $\hat{p}_{ab}$ of $\hat{\mathbf{T}}$ in $(\hat{M}, \hat{g}_{ab})$ admits a regular direction-dependent limit, it follows that so does $\Omega_T^{-1/2}(\hat{D}_a\hat{D}_b\Omega_{\hat{\mathbf{T}}} - 2\hat{q}_{ab})$.

Consider, next, the conditions on the extrinsic curvature $p_{ab}$ of $\mathbf{T}$ in $(\mathbf{M}, g_{ab})$, and on the intrinsic curvature $\hat{\mathcal{R}}_{ab}$ of $(T, \hat{q}_{ab})$. It is easy to check that $p_{ab}$ is related to $\hat{p}_{ab}$ via $\hat{p}_{ab} = \Omega p_{ab} + \Omega^{-1}(\hat{n}^m\hat{\nabla}_m\Omega)\hat{q}_{ab}$, where $\hat{n}$ is the $\hat{g}$ unit normal to $\mathbf{T}$. Using various conditions on the conformal factor $\Omega$, it is straightforward to check that $\Omega^{-1}(\hat{n}^m\hat{\nabla}_m\Omega)$ admits a regular direction-dependent limit at $i^0$. Hence it follows that $p_{ab}$ satisfies the ADM—G condition. Consider, finally the Ricci tensor $\hat{\mathcal{R}}_{ab}$ of $\hat{q}_{ab}$. Since $\Omega^{1/2}\hat{R}_{abcd}$ admits a regular direction-dependent limit at $i^0$, so does $\Omega^{1/2}\hat{E}_{ab}$, where $\hat{E}_{ab} = \hat{C}_{ambn}\hat{n}^m\hat{n}^n$ is the electric part of the Weyl-curvature relative to $\mathbf{T}$. Furthermore, by the Gauss—Codazzi equations, one has $\hat{\mathcal{R}}_{ab} = \hat{E}_{ab} + \hat{p}_{am}\hat{p}^m{}_b - \hat{p}^m{}_m\hat{p}_{ab}$. Hence, on $\hat{\mathbf{T}}$, $\Omega_{\hat{\mathbf{T}}}^{1/2}\hat{\mathcal{R}}_{ab}$ admits a regular direction-dependent limit at $i^0$.

Thus, the initial data $(q_{ab}, p_{ab})$ induced on $\mathbf{T}$ by $g_{ab}$ does indeed satisfy all the ADM—G conditions.

Next, we wish to introduce the notion of initial data sets "boosted" and "time translated" relative to each other. Two asymptotically flat data sets $(\mathbf{T}, q_{ab}, p_{ab})$ and $(\mathbf{T}', q'_{ab}, p'_{ab})$ in $(\mathbf{M}, g_{ab})$ will be said to be boosted relative to each other, if, in the completion $\hat{\mathbf{M}}$, $\hat{\mathbf{T}} = \mathbf{T} \cup i^0$ and $\hat{\mathbf{T}}' = \mathbf{T}' \cup i^0$ fail to be tangential to each other. Thus, if two data sets are relatively boosted, they differ already "in the first order" at $i^0$. $(\mathbf{T}, q_{ab}, p_{ab})$ and $(\mathbf{T}', q'_{ab}, p'_{ab})$ will be said to be "time translated" w.r.t. each other if, in the completion, $\hat{\mathbf{T}}$ and $\hat{\mathbf{T}}'$ are tangential, and if the limiting extrinsic curvatures $\hat{p}_{ab}$ and $\hat{p}'_{ab}$ fail to agree at $i^0$; data sets which are relatively time-translated agree "to first order" at $i^0$ but not "to second-order." If $\hat{\mathbf{T}}$ and $\hat{\mathbf{T}}'$ are tangential, and if further limits of $\hat{p}_{ab}$ and $\hat{p}'_{ab}$ agree, then it is easy to show that limits of $\Omega_{\hat{\mathbf{T}}}^{1/2}\hat{\mathcal{R}}_{ab}$ and $\Omega_{\hat{\mathbf{T}}'}^{1/2}\hat{\mathcal{R}}'_{ab}$ also agree. Hence, in this case, the two data sets will be said to be asymptotically indistinguishable.

Fix an asymptotically flat initial data set $(\mathbf{T}', q'_{ab}, p'_{ab})$. "How many" asymptotically distinct, asymptotically flat data sets can one obtain via time translations from $(\mathbf{T}', q'_{ab}, p'_{ab})$? Let $(\mathbf{T}, q_{ab}, p_{ab})$ be one such data set. Then $\hat{\mathbf{T}}'$ and $\hat{\mathbf{T}}$ are tangential at $i^0$, $\hat{q}_{ab}$ and $\hat{q}'_{ab}$ agree at $i^0$ while $\hat{p}_{ab}$ and $\hat{p}'_{ab}$ have distinct limits. Therefore, the collection of data sets under consideration can be "labeled" by the limiting direction-dependent values of their extrinsic curvature. Note, however, that this limiting value is constrained: The Gauss—Codazzi equations imply that $\hat{D}_{[a}\hat{p}_{b]c} = \hat{q}_a{}^m\hat{q}_b{}^n\hat{q}_c{}^r\hat{R}_{mnrs}\hat{n}^s$ on $\mathbf{T}$, so that by multiplying by $\Omega_{\hat{\mathbf{T}}}^{1/2}$ and taking limits one obtains $\eth_{[a}\hat{\mathbf{p}}_{b]c}(\eta) = \hat{\mathbf{q}}_a{}^m\hat{\mathbf{q}}_b{}^n\hat{\mathbf{q}}_c{}^r\mathbf{R}_{mnrs}(\eta)\hat{\mathbf{n}}^s$, where $\hat{\mathbf{p}}_{ab}(\eta)$, $\hat{q}_{ab}$, $\mathbf{R}_{abcd}(\eta)$ and $\hat{n}^s$ are the (direction-dependent) limits of $i^0$ of $\hat{p}_{ab}$, $\hat{q}_{ab}$, $\hat{R}_{abcd}$, and $\hat{n}^s$ and where $\eth_a$ is the derivative w.r.t. the 2-sphere of directions "$\eta$" in the tangent space of $\hat{\mathbf{T}}$ at $i_0$. How many solutions $\hat{\mathbf{p}}_{ab}$ does the last equation admit? Note that, for any data set $(\mathbf{T}, q_{ab}, p_{ab})$ which is time translated relative to the given one, $(\mathbf{T}', q'_{ab}, p'_{ab})$, we have $\hat{\mathbf{q}}_a{}^m\hat{\mathbf{q}}_b{}^n\hat{\mathbf{q}}_c{}^r\hat{\mathbf{R}}_{mnrs}(\eta)\hat{\mathbf{n}}^s = \hat{\mathbf{q}}'_a{}^m\hat{\mathbf{q}}'_b{}^n\hat{\mathbf{q}}'_c{}^r\hat{\mathbf{R}}_{mnrs}(\eta)\hat{n}'^s$. Hence, $\Delta\hat{\mathbf{p}}_{ab}(\eta) \equiv \hat{\mathbf{p}}_{ab} - \hat{\mathbf{p}}'_{ab}$ satisfies $\eth_{[a}\Delta\hat{\mathbf{p}}_{b]c}(\eta) = 0$. Thus, there are as many solutions to the contraint equation on $\hat{\mathbf{p}}_{ab}$ as there are to $\eth_{[a}\Delta\hat{\mathbf{p}}_{b]c}(\eta) = 0$. Fortunately, one can write down the general solution to this last equation

$$\Delta\hat{\mathbf{p}}_{ab}(\eta) = \eth_a\eth_b\chi(\eta) + \tfrac{1}{2}\eta_a\eth_b\chi(\eta) + \eta_b\eth_a\chi(\eta) + \chi(\eta)\hat{\mathbf{q}}'_{ab}, \tag{B1}$$

where $\chi(\eta)$ is an arbitrary (smooth) function on the 2-sphere of directions $\eta$. Thus, there are as many solutions to the constraint equation as there are functions on a 2-sphere. Hence, we conclude that one can obtain as many asymptotically flat initial data sets via time translations of $(T', q'_{ab}, p'_{ab})$ as there are function on a 2-sphere. All these data sets agree with one another to the "first order" but not to the "second."

Finally, it is obvious from the above discussion that, restrictions to $\mathbf{M}$ of diffeomorphisms in $\hat{\mathbf{M}}$ which leave ($\mathcal{J}$ and) $i^0$ invariant and which are $C^{>1}$ at $i^0$ provide us with "asymptotically regular" evolusions in $(\mathbf{M}, g_{ab})$, evolutions which preserve the ADM—G asymptotic conditions.

## APPENDIX C: SOME AEFANSI SPACE-TIMES

Explicit constructions of the conformal completions for the Minkowski and Kerr geometries are given in this appendix, along with a discussion of the radiative perturbations of the Schwarzschild geometry. A general form for the space—time metric satisfying the "local conditions at $i^0$" required by the definition of AEFANSI space—times is also given.

The metric of Minkowski space in spherical coordinates is

$$ds^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2). \tag{C1}$$

Let $v$ and $w$ be given by

$$v = r + t, \quad w = r - t. \tag{C2}$$

In terms of these coordinates, the metric is

$$ds^2 = dv\, dw + \tfrac{1}{4}(v+w)^2(d\theta^2 + \sin^2\theta\, d\phi^2). \tag{C3}$$

Set $\hat{v} = 1/v$, $\hat{w} = 1/w$, and $\Omega = \hat{v}\hat{w}$; the conformally rescaled metric is

$$d\hat{s} = \Omega^2\, ds^2$$

$$= d\hat{v}d\hat{w} + \tfrac{1}{4}(\hat{v} + \hat{w})^2(d\theta^2 + \sin^2\theta\, d\phi^2). \tag{C4}$$

Comparison of (C4) with (C3) shows that the rescaled metric is flat (in fact, this rescaling of the Minkowski metric is induced by the action of the inversion about the origin in the flat-space conformal group). Thus, the introduction of coordinates $(\hat{t}, \hat{x}, \hat{y}, \hat{z})$ given by

$$\hat{t} = -\tfrac{1}{2}(\hat{v} - \hat{w}), \tag{C5a}$$

$$\hat{x} = \tfrac{1}{2}(\hat{v} + \hat{w}) \sin\theta \sin\phi, \tag{C5b}$$

$$\hat{y} = \tfrac{1}{2}(\hat{v} + \hat{w}) \sin\theta \cos\phi, \tag{C5c}$$

$$\hat{z} = \tfrac{1}{2}(\hat{v} + \hat{w}) \cos\theta, \tag{C5d}$$

leads to a metric which can be extended over all of $\mathbb{R}^4$. This extended manifold is a neighborhood of $i^0$ in $\hat{\mathbf{M}}$. The surface $\hat{t} = (\hat{x}^2 + \hat{y}^2 + \hat{z}^2)^{1/2}$ forms that portion of $\mathcal{J}^+$ to the past of $w = 0$; $\hat{t} = -(\hat{x}^2 + \hat{y}^2 + \hat{z}^2)^{1/2}$ is the part of $\mathcal{J}^-$ to the future of $v = 0$. The point $\hat{t} = \hat{x} = \hat{y} = \hat{z} = 0$ is, of course, $i^0$. The conditions on the metric and conformal factor are trivially satisfied; in fact $\hat{g}_{ab}$ is everywhere analytic and $\Omega$ satisfies $\hat{\nabla}_a\hat{\nabla}_b\Omega = 2\hat{g}_{ab}$ everywhere.

We now consider a more interesting example: The Kerr solution. In Boyer—Linquist coordinates, the metric[45] is given by

$$ds^2 = -\left(1 - \frac{2Mr}{r^2 + a^2\cos\theta}\right)dt^2$$

$$- \frac{4Mar\sin\theta}{r^2 + a^2\cos^2\theta}\, dt\, d\phi$$

$$+ \frac{r^2 + a^2\cos^2\theta}{r^2 - 2Mr + a^2}\, dr^2 + (r^2 + a^2\cos^2\theta)\, d\theta^2$$

$$+ \left(r^2 + a^2 + \frac{2Ma^2 r\sin^2\theta}{r^2 + a^2\cos^2\theta}\right)\sin^2\theta\, d\phi^2. \tag{C6}$$

Set

$$r^* = f(r)$$

$$= r + M/(M^2 - a^2)^{1/2}\Big[M + (M^2 - a^2)^{1/2}$$

$$\times \ln\left(\frac{r}{M + (M^2 - a^2)^{1/2}} - 1\right)$$

$$- [M - (M^2 - a^2)^{1/2}] \ln\left(\frac{r}{M - (M^2 - a^2)^{1/2}} - 1\right)\Big] \tag{C7}$$

and introduce coordinates $\hat{v}$ and $\hat{w}$ given[46] by

$$r^* = \frac{1}{2}\left[f\left(\frac{1}{\hat{v}}\right) + f\left(\frac{1}{\hat{w}}\right)\right], \tag{C8a}$$

$$t = \frac{1}{2}\left[f\left(\frac{1}{\hat{v}}\right) - f\left(\frac{1}{\hat{w}}\right)\right]. \tag{C8b}$$

Define $\bar{r} = 1/r$, and let $\Omega$ be given by

$$\Omega = \hat{v}\hat{w}\left[1 - 2M\frac{\hat{v}\hat{w}}{\bar{r}}\left(1 + \ln\frac{\bar{r}^2}{\hat{v}\hat{w}}\right)\right].$$

Then the conformally rescaled metric is

$$d\hat{S}^2 = \Omega^2\, ds^2$$

$$= \left[1 - 2M\frac{\hat{v}\hat{w}}{\bar{r}}\left(1 + \ln\frac{\bar{r}^2}{\hat{v}\hat{w}}\right)\right]^2$$

$$\times \left\{\left[\left(1 - \frac{2M\hat{v}}{1 + a^2\hat{v}^2}\right)^{-1}\left(1 - \frac{2M\hat{w}}{1 + a^2\hat{v}^2}\right)^{-1}\right.\right.$$

$$\times\left[\left(1 - \frac{2M\bar{r}}{1 + a^2\bar{r}^2\cos^2\theta} - \frac{1}{2}\frac{a^2\bar{r}^2\sin^2\theta}{1 + a^2\bar{r}^2}\right)\right.$$

$$\left.\left.\times\left(1 - \frac{2M\bar{r}}{1 + a^2\bar{r}^2} - \frac{2M\bar{r}}{1 + a^2\bar{r}^2\cos^2\theta}\right)\right]\, d\hat{v}d\hat{w}\right.$$

$$-\frac{1}{4}\Big[\frac{\hat{w}^2}{\hat{v}^2}\left(1-\frac{2M\hat{v}}{1+a^2\hat{v}^2}\right)^{-2}\left(\frac{a^2\bar{r}^2\sin^2\theta}{1+a^2\bar{r}^2}\right)$$

$$\times\left(1-\frac{2M\bar{r}}{1+a^2\bar{r}^2}-\frac{2M\bar{r}}{1+a^2\bar{r}^2\cos^2\theta}\right)\Big]\,d\hat{v}^2$$

$$-\Big[\frac{1}{4}\,\frac{\hat{v}^2}{\hat{w}^2}\left(1-\frac{2M\hat{w}}{1+a^2\hat{w}^2}\right)^{-2}\left(\frac{a^2\bar{r}^2\sin^2\theta}{1+a^2\bar{r}^2}\right)$$

$$\times\left(1-\frac{2M\bar{r}}{1+a^2\bar{r}^2}-\frac{2M\bar{r}}{1+a^2\bar{r}^2\cos^2\theta}\right)\Big]\,d\hat{w}^2$$

$$+\hat{w}^2\left(1-\frac{2M\hat{v}}{1+a^2\hat{v}^2}\right)^{-1}\left(\frac{2Ma\bar{r}\sin^2\theta}{1+a^2\bar{r}^2\cos^2\theta}\right)d\hat{v}\,d\phi$$

$$+\hat{v}^2\;1-\left(\frac{2M\hat{w}}{1+a^2\hat{w}^2}\right)^{-1}\left(\frac{2Ma\bar{r}\sin^2\theta}{1+a^2\bar{r}^2\cos^2\theta}\right)d\hat{w}\,d\phi$$

$$+\frac{\hat{v}^2\hat{w}^2}{\bar{r}^2}\,(1+a^2\bar{r}^2\cos^2\theta)\,d\theta^2$$

$$+\frac{\hat{v}^2\hat{w}^2}{\bar{r}^2}\left(1+a^2\bar{r}^2+\frac{2Ma^2\bar{r}^3\sin^2\theta}{1+a^2\bar{r}^2\cos^2\theta}\right)\sin^2\theta\,d\phi^2\Big\}. \qquad \text{(C9)}$$

Now let $(\hat{t},\hat{x},\hat{y},\hat{z})$ be given in terms of $(\hat{v},\hat{w},\theta,\phi)$ by (C5), as in Minkowski space, and extend the Kerr manifold to all values of these coordinates. The resulting manifold again forms a neighborhood of $i^0$; $\mathscr{I}^+$ and $\mathscr{I}^-$ are the same coordinate surfaces as in Minkowski space. However, the metric $\hat{g}_{ab}$ is only $C^0$ on $\mathscr{I}$ and $C^{>0}$ at $i^0$. The conformal factor $\Omega$ satisfies the required conditions on $\mathscr{I}$ and at $i^0$, and the direction-dependent tensor $\hat{\mathbf{K}}_{ab}(\eta)$—the tensor potential for $\mathbf{B}_{ab}(\eta)$—vanishes for this choice of conformal factor.

The remaining quantity of interest is the Weyl tensor. It is convenient to introduce a null tetrad $(l^a, m^a, \bar{m}^a, n^a)$, where in Boyer—Lindquist coordinates the tetrad vectors have components

$$l^a=\frac{1}{\sqrt{2}}\left(\frac{r^2-2Mr+a^2}{r^2+a^2\cos^2\theta}\right)^{1/2}\left(\frac{r^2+a^2}{r^2-2Mr+a^2},\,1,\,0,\,-\frac{a}{r^2-2Mr+a^2}\right), \qquad \text{(C10a)}$$

$$m^a=\frac{1}{\sqrt{2}}\,\frac{1}{(r^2+a^2\cos^2\theta)^{1/2}}\left(-ia\sin\theta,\,0,\,1,\,\frac{i}{\sin\theta}\right), \qquad \text{(C10b)}$$

$$n^a=\frac{1}{\sqrt{2}}\left(\frac{r^2-2Mr+a^2}{r^2+a^2\cos^2\theta}\right)^{1/2}\times\left(\frac{r^2+a^2}{r^2-2Mr+a^2},\,-1,\,0,\,-\frac{a}{r^2-2Mr+a^2}\right). \qquad \text{(C10c)}$$

In terms of this tetrad, the Weyl tensor for the Kerr solution can be written in the form

$$C_{abcd}+i\,{}^*C_{abcd}=\frac{M}{(r-ia\cos\theta)^3}\,[l_{[a}m_{b]}\bar{m}_{[c}n_{d]}+(l_{[a}n_{b]}-m_{[a}\bar{m}_{b]})(l_{[c}n_{d]}-m_{[c}\bar{m}_{d]})+\bar{m}_{[a}n_{b]}l_{[c}m_{d]}]. \qquad \text{(C11)}$$

To study the asymptotic behavior of the Weyl tensor at $i^0$, we introduce the tetrad $(\hat{\psi}_a, \hat{\rho}^a, \hat{m}^a, \hat{\bar{m}}^a)$, given by

$$\hat{\psi}_a=\frac{\Omega}{\sqrt{2}}\left(\frac{\hat{w}^{1/2}}{\hat{v}^{1/2}}l_a+\frac{\hat{v}^{1/2}}{\hat{w}^{1/2}}n_a\right), \qquad \text{(C12a)}$$

$$\hat{\rho}_a=\frac{\Omega}{\sqrt{2}}\left[-\left(\frac{\hat{w}}{\hat{v}}\right)^{1/2}l_a+\left(\frac{\hat{v}}{\hat{w}}\right)^{1/2}n_a\right], \qquad \text{(C12b)}$$

$$\hat{m}_a=\Omega m_a. \qquad \text{(C12c)}$$

The tetrad vectors $\hat{\psi}^a, \hat{\rho}^a, \hat{m}^a$, and $\hat{\bar{m}}^a$ have direction-dependent limits at $i^0$, and $\lim_{\to i^0}\hat{\rho}^a=\eta^a$. From the orthogonality relations for the tetrad, it is clear that the direction-dependent vectors defined at $i^0$ by $\hat{\psi}^a$, $\hat{m}^a$, and $\hat{\bar{m}}^a$ correspond to a triad $(\boldsymbol{\psi}^a, \mathbf{m}^a, \bar{\mathbf{m}}^a)$ on $\mathscr{K}$. In terms of the tetrad $(\hat{\psi}^a, \hat{\rho}^a, \hat{m}^a, \hat{\bar{m}}^a)$, the conformally rescaled Weyl tensor can be written

$$\hat{C}_{abcd}+i\,{}^*\hat{C}_{abcd}=-2\left(\frac{\bar{r}^3}{\hat{v}^2\hat{w}^2}\right)\left(\frac{M}{(1-ia\bar{r}\cos\theta)^3}\right)\times\left[1-\frac{2M\hat{v}\hat{w}}{\bar{r}}\left(1+\ln\frac{\bar{r}^2}{\hat{v}\hat{w}}\right)\right]^{-2}$$
$$\times[(\hat{\psi}_{[a}-\hat{\rho}_{[a})\hat{m}_{b]}(\hat{\psi}_{[c}+\hat{\rho}_{[c})\hat{\bar{m}}_{d]}-(2\hat{\psi}_{[a}\hat{\rho}_{b]}+\hat{m}_{[a}\hat{\bar{m}}_{b]})(2\hat{\psi}_{[c}\hat{\rho}_{d]}+\hat{m}_{[c}\hat{\bar{m}}_{d]})+(\hat{\psi}_{[a}+\hat{\rho}_{[a})\hat{\bar{m}}_{b]}(\hat{\psi}_{[c}-\hat{\rho}_{[c})\,\hat{m}_{d]}]. \qquad \text{(C13)}$$

To evaluate the mass and angular momentum, we need only obtain expressions for $\mathbf{E}_{ab}$ and $\boldsymbol{\beta}_{ab}$ on a fixed cross section of $\mathscr{K}$. For convenience, we choose the intersection $\Sigma$ of $\mathscr{K}$ with the tangent plane at $i^0$ to the hypersurface $\hat{v}=\hat{w}$. On this cross section, $\mathbf{E}_{ab}$ and $\boldsymbol{\beta}_{ab}$ take the form

$$\mathbf{E}_{ab}=M(\psi_a\psi_b+\mathbf{m}_{(a}\bar{m}_{b)}), \qquad \text{(C14a)}$$

$$\boldsymbol{\beta}_{ab}=3Ma\cos\theta\,(\psi_a\psi_b+\mathbf{m}_{(a}\bar{\mathbf{m}}_{b)}). \qquad \text{(C14b)}$$

[Note that, since $\mathbf{K}_{ab}$ vanishes (in this conformal frame), $\mathbf{B}_{ab}=0$.]

The mass is obtained by integrating $\mathbf{E}_{ab}$ over the cross section $\Sigma$, using $\psi^a$ as the integrating factor (on $\Sigma$, $\psi^a$ clearly represents the 4-velocity of the natural asymptotic rest frame of the system):

$$P_0=\int_\Sigma \psi_m M(\psi^m\psi^n+\tfrac{1}{2}\mathbf{m}^m\bar{\mathbf{m}}^n+\tfrac{1}{2}\bar{\mathbf{m}}^m\mathbf{m}^n)\,dS_n=-\frac{M}{4\pi}\int_0^{2\pi}\int_0^{\pi}\sin\theta\,d\theta\,d\phi=-M. \qquad \text{(C15)}$$

The only interesting component of the angular momentum is that about the symmetry axis. The corresponding tensor $F_{ab}$ has as its projection onto $\mathscr{K}$

$$\mathbf{f}_{ab}=i\cos\theta\,\mathbf{m}_{[a}\bar{\mathbf{m}}_{b]};$$

thus the component of the angular momentum about the symmetry axis is

$$\hat{M}^{mn}F_{mn} = \int_\Sigma \epsilon^{mnpq} f_{mn} \hat{\beta}_{pq}\, dS^q$$
$$= \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} 3Ma\cos^2\theta \sin\theta\, d\theta\, d\phi$$
$$= Ma. \tag{C16}$$

We next examine the effects at $i^0$ of the linearized radiation fields in the Schwarzschild background constructed by Bardeen and Press.[47] They begin with the Schwarzschild metric in coordinates $(u, r, \theta, \phi)$ such that the metric takes the form

$$ds^2 = -\left(1 - \frac{2M}{r}\right)du^2 - 2du\, dr$$
$$- r^2(d\theta^2 + \sin^2\theta\, \phi^2). \tag{C17}$$

The perturbations they consider are expressed in terms of an arbitrary complex-valued function $A(u, \theta, \phi)$ which is assumed to have a convergent expansion in spherical harmonics,

$$A(u,\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} A_{lm}(u) Y_{lm}(\theta,\phi). \tag{C18}$$

The perturbed Weyl tensor constructed from this function is shown to lead a regular geometry on $\mathscr{I}^+$ if the first $(l+2)$ derivatives of $A_{lm}(u)$ are everywhere bounded and if

$$\left(2^l \frac{(l-2)!}{(2l)!}\right)^2 \int_{-\infty}^{u} \left|A_{lm}^{(l+1)}(u_0)\right|^2 du_0 \ll M \tag{C19}$$

holds for all finite $u$, where the superscript parentheses denote differentiation with respect to $u_0$. However, the peeling behavior[4] of the Weyl tensor on $\mathscr{I}^-$ is guaranteed only if, in addition,

$$\lim_{u\to-\infty} A_{00}(u), \tag{C20a}$$

$$\lim_{u\to-\infty} A_{1m}(u), \tag{C20b}$$

$$\lim_{u\to-\infty} uA_{1m}^{(1)}(u), \tag{C20c}$$

$$\lim_{u\to-\infty} u^2 A_{1m}^{(2)}(u) \tag{C20d}$$

$$\lim_{u\to-\infty} u^{k-l+2} A_{lm}^{(k)}(u), \quad l \geq 2, \quad 0 \leq k \leq l+2 \tag{C20e}$$

all exist.

A similar analysis shows that conditions (C20) imply the existence of a direction-dependent limit for $\Omega^{1/2}\hat{C}_{abcd}$ at $i^0$ (in fact, somewhat weaker conditions suffice). The existence of the direction-dependent tensor $\hat{\beta}_{ab}$ also follows from (C20) if $\lim_{u\to-\infty} u[A_{00}(u) - \bar{A}_{00}(u)]$ exists, where the bar denotes complex conjugation. This last condition appears simply to preclude nonzero angular momentum monopoles at spatial infinity, and is to be expected if the perturbed metric coefficients are to be nonsingular[48] on $\mathscr{I}$. Thus, the appropriate behavior of the perturbed Weyl tensor components at $i^0$ follows from regularity requirements on $\mathscr{I}^+$ and $\mathscr{I}^-$. It is remarkable that conditions (C20) for regularity on $\mathscr{I}^-$ should imply the necessary conditions at $i^0$.

We conclude by giving a general form of the metric satisfying the "local" conditions at $i^0$ [i.e., condition (iii)] in the definition of AEFANSI space—times. The form is

$$ds^2 = \left(1 + \frac{v+w}{vw}\alpha\right) dv\, dw + \frac{w}{v^2}\beta\, dv^2 + \frac{v}{w^2}\gamma\, dw^2$$
$$+ \frac{v+w}{v}\delta_A\, dz^A\, dv + \frac{v+w}{w}\epsilon_A\, dz^A\, dw$$
$$+ \tfrac{1}{4}(v+w)^2\left(q_{AB} + \frac{v+w}{vw}\phi_{AB}\right) dz^A\, dz^B, \tag{C21}$$

where $q_{AB}\, dz^A dz^B$ is a 2-sphere metric (upper case Latin indices take the values 2 and 3). The metric coefficients $\alpha, \beta, \gamma, \delta_A, \epsilon_A$, and $\phi_{AB}$ are smooth functions of $v$, $w$, and $z^A$, continuous (and bounded) as $v\to\infty$ or $w\to\infty$, and smooth functions of $v/w$ and $z^A$ in the limit $v\to\infty$, $w\to\infty$ with $v/w$ finite and nonzero.

By the introduction of new coordinates $\hat{v} = 1/v$, $\hat{w} = 1/w$, and a conformal factor $\Omega = \hat{v}\hat{w}$, a metric is obtained which is $C^0$ as $\hat{v}\to 0$ or $\hat{w}\to 0$, and which, in coordinates $(\hat{t}, \hat{x}, \hat{y}, \hat{z})$ related to $(\hat{v}, \hat{w}, z^A)$ by the analog of Eqs. (C5), is $C^{>0}$ at $\hat{t} = \hat{x} = \hat{y} = \hat{z} = 0$. The conformal factor $\Omega$ vanishes where $\hat{v} = 0$ or $\hat{w} = 0$, and at $\hat{t} = \hat{x} = \hat{y} = \hat{z} = 0$ it is $C^{>2}$ and satisfies $\Omega = 0$, $\hat{\nabla}_a\Omega = 0$, $\hat{\nabla}_a\hat{\nabla}_b\Omega = 2\hat{g}_{ab}$. Thus, the coordinates $(\hat{x}, \hat{y}, \hat{z}, t)$ may be extended to form a neighborhood of $i^0$ in the obvious way.

Note, however, that there do exist additional conditions that the metric (C21) must satisfy before the space—time can qualify as AEFANSI: No field equations have yet been imposed on the metric. Thus, for the metric under consideration, there may exist no potentials (formed from the Ricci tensor $\hat{R}_{ab}$) for asymptotic fields $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$. Furthermore, $\mathbf{B}_{ab}$ need not vanish and hence $\hat{\beta}_{ab}$ need not exist. Finally, it is by no means clear that the metric (C21) is weakly asymptotic simple. The example of the Kerr solution shows that even if the rescaled metric can be made regular on $\mathscr{I}$ by a coordinate transformation, this transformation will in general be only $C^0$ at $i^0$.[49] However, it seems unlikely that the imposition of field equations, at least to the required asymptotic order, will seriously restrict the functional freedom in (C21); and regularity on $\mathscr{I}$ appears to play no essential role at $i^0$. In any case, it is known[4] that there exists a large class of weakly asymptotically simple space—times.

From the fact that the metric given in (C21) approaches a Minkowskian metric about as slowly as one would expect (it may be though of as Minkowski metric with correction terms of order "$1/r$") it seems reasonable to expect that any metric which is, intuitively, asymptotically flat at spatial infinity will satisfy at least the local conditions at $i^0$ to be AEFANSI.

[1]For a recent review, see the article by R. Geroch, in *Asymptotic Structure of Space—time*, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977).

[2]H. Bondi, A.W.K. Metzner, and M.J.G. Van der Berg, Proc. Roy. Soc. A (London) **289**, 21 (1962).

[3]R.K. Sachs, Proc. Roy. Soc. A (London) **270**, 103 (1962); Phys. Rev. **128**, 2851 (1962).

[4]R. Penrose, Phys. Rev. Lett. **10**, 66 (1963); Proc. Roy. Soc. A (London) **284**, 159 (1965); "Relativistic Symmetry Groups," in *Group Theory in Non-linear Problems*, Nato Advanced Study Institute, Series C, edited by A.O. Barut (Reidel; Dordrecht, Holland, 1974).

[5]A. Held, E.T. Newman, and R. Posadas, J. Math. Phys. **11**, 3145 (1970); R.W. Lind, J. Messmer, and E.T. Newman, J. Math. Phys. **13**, 1884 (1972); E.T. Newman and R. Penrose, J. Math. Phys. **7**, 863 (1966); Proc. Roy. Soc. A (London) **305**, 175 (1968); B.G. Schmidt, Commun. Math. Phys. **36**, 73 (1974); B.G. Schmidt, P. Sommers, and M. Walker, J. Gen. Rel. Grav. **6**, 489 (1975); J.H. Winicour, J. Math. Phys. **9**, 861 (1968); Phys. Rev. D **3**, 840 (1971); L.A. Tamburino and J.H. Winicour, Phys. Rev. **150**, 1039 (1966).

[6]R. Arnowitt, S. Deser, and C.W. Misner, Phys. Rev. **117**, 1595 (1960); **118**, 1100 (1960); **121**, 1566 (1961); **122**, 997 (1961); the article in *Gravitation, an Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).

[7]P.G. Bergmann, Phys. Rev. **124**, 274 (1961); *Asymptotic Properties of Gravitating Systems*, in the Proceedings of the Warszawa Meeting (Guthier-Villars, Paris, 1964).

[8]R. Geroch, J. Math. Phys. **13**, 956 (1972).

[9]N. O'Murchadha and J.W. York, Phys. Rev. D **16**, 428 (1974); T. Regge and C. Teitelboim, Ann. Phys. **88**, 286 (1974); Y. Choquet-Bruhat, A.E. Fischer, and J.E. Marsden, in the Proceedings of the Enrico Fermi International School of Physics (1976); R.O. Hansen, J. Math. Phys. **15**, 1 (1974).

[10]One does not expect asymptotic flatness in one regime, by itself, to imply asymptotic flatness in the other. Thus, for example, given a space—time which is asymptotically flat in both regimes, one can redistribute matter such that some neighborhood of $\mathscr{I}$ continues to remain empty but such that *every* neighborhood of spatial infinity contains some matter, thereby destroying the asymptotic flatness in the spatial regime but not in the null. The converse situation can be arranged by using zero rest-mass fields with appropriate strength and support.

[11]B.G. Schmidt and J. Stewart; private communication.

[12]The usual strategy to obtain a boosted Cauchy surface from the given one is to move along normal geodesics using a lapse function which resembles the boost lapse function in Minkowski space (see, e.g., Ref. 8). However, since this function diverges "like $r$," there is no *a priori* guarantee that geodesics will not start crossing before one can evolve by a sufficient amount. Other strategies run into similar problems. Although these can be avoided in special cases—e.g., in stationary space—times—the issue is unresolved in the general context.

[13]Consider, for example, the statement that the ADM 4-momentum is invariant under time evolution. Since the very definition of this quantity rests crucially on asymptotic flatness of initial data sets, the statement of invariance has nontrivial content only if one is guaranteed the preservation of asymptotic conditions under evolution.

[14]R. Penrose, in *Battelle Rencontres, Lectures in Mathematical Physics*, edited by C. DeWitt and J. A. Wheeler (Benjamin, New York, 1968).

[15]See Ref. 4, or, e.g., S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space—time* (Cambridge U.P., London, 1973).

[16]For example, in Schwarzschild space—time, the scalar $C^{abcd}C_{abcd} \sim 1/r^6$ at spatial infinity. Hence, in terms of the rescaled metric $\hat{g}_{ab} = \Omega^2 g_{ab}$ $(\Omega \sim 1/r^2)$, $\hat{C}_{abcd}\hat{C}^{abcd} \sim r^2$; $C_{abcd}$ diverges at $i^0$.

[17]We thank David Deutch for suggesting this terminology.

[18]One normally decomposes Maxwell field into electric and magnetic parts relative to *spacelike* 3-surfaces. $\mathscr{K}$, on the other hand, is *timelike*. Perhaps a better terminology would have been asymptotic pseudoelectric and asymptotic pseudomagnetic fields. The same remark holds for the decomposition of the Weyl tensor.

[19]That is, the limit is finite but depends on the direction of approach to $\Lambda$. For a precise definition, see Ref. 8 or Appendix A.

[20]A. Ashtekar, Lectures on Asymptotics at the University of Chicago. Lecture notes are being prepared for the Springer Series by B. Xanthopoulos.

[21]For definitions and detailed discussion, see Appendix A.

[22]A somewhat different but more natural definition will appear in Ref. 20; $\mathscr{I}$ will be introduced in terms of $i^0$ rather than $i^0$ in terms of $\mathscr{I}$.

[23]In the case of Minkowski space, the Einstein cylinder can serve as $\hat{\mathbf{M}}$; see, e.g., Ref. 15.

[24]More generally, arbitrary matter sources for which the stress-energy tensor field $T_{ab}$ (with this index structure) on $(\mathbf{M}, g_{ab})$ admits a regular direction-dependent limit at $i^0$ are permissible in the analysis of spatial infinity.

[25]Let $g_{ab}$ be the physical metric and $\hat{g}_{ab} = \Omega^2 g_{ab}$ and $\tilde{g}_{ab} = (\omega\Omega)^2 g_{ab}$ be any two rescaled metrics $C^{>0}$ at $i^0$. Then, $\Omega\,|_{i^0} = 0$, $(\hat{\nabla}_a\Omega)\,|_{i^0} = 0$, $(\hat{\nabla}_a\hat{\nabla}_b\Omega - 2\hat{g}_{ab})\,|\,i^0 = 0$, and $(\tilde{\nabla}_a\tilde{\nabla}_b\omega\Omega - 2\tilde{g}_{ab})\,|\,i^0 = 0$. Consistency of these conditions requires that $\omega$ be $C^{>0}$ at $i^0$ and $\omega\,|_{i^0} = 1$.

[26]In this case the resulting analysis would be essentially equivalent to that in Ref. 1 provided we require, *in addition*, that $\Omega^{1/2}\hat{C}_{abcd}$ admits a regular direction-dependent limit at $i^0$. (In Ref. 1 the same additional condition is required although it is formulated in terms of electric and magnetic parts of the Weyl tensor relative to a Cauchy surface.)

[27]Throughout, we use the term geodesics for curves whose tangent vector satisfies $\eta^m\,\eta^{[a}\nabla_m\eta^{b]} = 0$, rather than $\eta^m\nabla_m\eta^b = 0$. We do not use affine parametrization (w.r.t. the physical metric) because the parameter is ill behaved at $i^0$.

[28]Because the rescaled metric $\hat{g}_{ab}$ is $C^{>0}$ at $i^0$, it follows that $\Omega^{1/2}\hat{R}_{abcd}$—and hence $\Omega^{1/2}(\hat{\nabla}_a\hat{\nabla}_b\Omega - 2\hat{g}_{ab})$—admits a regular direction-dependent limit at $i^0$ (see Appendix A). Using this fact and the l'Hopital's rule, it follows that $\lim_{\to i^0}\Omega^{-1}\hat{h}_{ab}\hat{\nabla}^b\Omega$ exists as a direction-dependent tensor at $i^0$.

[29]Note that, given any two $C^{>1}$ curves with the same tangent vector $\eta^a$ at $i^0$, their tangential components of acceleration at $i^0$ agree w.r.t. one metric in the conformal class if and only if they agree w.r.t. any other metric. Hence, the equivalence relation is well defined.

[30]For example, no point of Spi is left invariant by the group action, and, the only group element whose action leaves Spi invariant is the identity.

[31]See, e.g., R.L. Bishop and R.J. Crittenden, *Geometry of Manifolds* (Academic, New York and London, 1964), page 41.

[32]Note that, since $h_{ab}$ is covariant and $v^a$ contravariant, these fields do not give rise to a (nondegenerate) metric on Spi.

[33]The resulting construction of Spi is the most natural one for discussing physical fields. It will be described in Ref. 20.

[34]Thus, the situation is the same as at null infinity; the BMS group can also be characterized as the subgroup of the diffeomorphism group of $\mathscr{I}$ which leaves all the universal structure invariant. See, e.g., the sixth paper in Ref. 5.

[35]Note, however, that there is an important difference between the BMS group and $\mathscr{G}$: While the center of the BMS group is the identity, the center of $\mathscr{G}$ is the one-parameter family of supertranslations corresponding to the structure group of Spi.

[36]This restriction may be weakened to allow charges whose current vector $J^a$ has the property that $\lim_{\to i^0}\Omega^{3/2}J^a = 0$, without affecting the final results.

[37]Note that, on $\mathscr{K}$, $\mathbf{D}_{[a}\mathbf{V}_{b]} = 0$, $\mathbf{D}^a\mathbf{V}_a = 0$ does *not* imply that $\mathbf{V}_a = 0$. Indeed, by the first of these equations, it follows that there exists a scalar $\mathbf{V}$ with $\mathbf{D}_a\mathbf{V} = \mathbf{V}_a$, and the second then reduces only to $\mathbf{D}^a\mathbf{D}_a\mathbf{V} = 0$. Since this last equation is hyperbolic, it admits lots of (regular) solutions.

[38]Note that, although the 2-forms $\mathbf{E}^a\epsilon_{abc}$ and $\mathbf{B}^a\epsilon_{abc}$ are both closed (i.e., curl-free), the integrals are not necessarily zero: Since $\mathscr{K}$ admits 2-sphere cross sections which cannot be contracted continuously to a point, closed 2-forms on $\mathscr{K}$ need not be exact.

[39]More precisely, they are invariant under the action of the

group $\mathcal{G}$ of asymptotic symmetries.

40 See last three papers in Ref. 5 or Ref. 1.

41 The hyperboloid $\mathcal{K}$ admits ten—the maximum possible number of—conformal Killing vectors. Furthermore, the Killing form on the Lie algebra is nondegenerate. Hence, one can divide the ten-dimensional vector space underlying the Lie algebra unambiguously into two complementary subspaces; a six-dimensional subspace of Killing vectors and a four-dimensional subspace of "pure" conformal Killing vectors. The former satisfy $D_{(a}\xi_{b)}=0$ while the latter, $D_{[a}\xi_{b]}=0$. (The "pure" conformal Killing vectors are gradients of functions $\mathbf{f}(K)$ on $\mathcal{K}$ which correspond to translations.)

42 See, e.g., the analysis of Kerr space—times in Appendix C.

43 Although it *has* been occasionally argued that the differentiability requirements in Ref. 4 are too strong. See, e.g., W.E. Couch and R.J. Torrence, J. Math. Phys. **13**, 69 (1973); W.H. Press and J.M. Bardeen, J. Math. Phys. **14**, 7 (1973). Stephaney Novak has investigated in detail the case when the rescaled metric fails to be $C^3$ on $\mathcal{I}$ (private communication).

44 See, e.g., P. Sommers, "Geometry of Space-like Infinity," preprint. It would not be surprising if Sommers' and the present approach turn out to be completely equivalent. Indeed, Sommers' hyperboloid "Psi" can, in essence, be constructed by associating, with each equivalence class of geodesics in the physical space—time which are tangential to each other at $i^0$, a boundary point to the given space—time.

45 R.P. Kerr, Phys. Rev. Lett. **11**, 237 (1963); R.H. Boyer and R.W. Lindquist, J. Math. Phys. **8**, 265 (1967).

46 The choice of these coordinates makes the conformally rescaled metric $C^{>0}$ (rather than $C^0$) at $i^0$. We thank Bernd Schmidt and Martin Walker for making available to us their calculations which led to this choice.

47 J.M. Bardeen and W.H. Press, the second paper in Ref. 43.

48 A.I. Janis and E.T. Newman, J. Math. Phys. **6**, 902 (1965).

49 That is, there do not appear to exist charts in which the rescaled metric connection admits direction-dependent limits as one approaches $i^0$ along $\mathcal{I}$.

# A technique for analyzing the structure of isometries

Abhay Ashtekar[a)] and Anne Magnon-Ashtekar[b)]

*Enrico Fermi Institute, Chicago, Illinois 60637*
(Received 16 January 1978)

A new technique is introduced to investigate the structure of isometry Lie algebras. Some general results are first proved by applying this technique to $n$-dimensional manifolds equipped with metrics of arbitrary signature. A restriction is then made to 3-manifolds representing the space of orbits of the timelike Killing field in stationary space–times. Under the assumption of asymptotic flatness at *spatial* infinity, a complete description of isometry Lie algebras of these 3-manifolds is obtained. As corollaries, several results about symmetries of stationary isolated systems in general relativity are proved.

## 1. INTRODUCTION

The purpose of this note is to introduce a new technique for analyzing the structure of Killing fields on metric manifolds, a technique which appears to be useful especially in the investigation of symmetries of isolated systems in general relativity.

In Sec. 2, we consider general $n$-dimensional manifolds equipped with (nondegenerate) metrics of arbitrary signature. With each point $p$ of the manifold, we associate two algebras, $\mathrm{I}_p$ and $\mathrm{C}_p$, of dimensions $n(n+1)/2$ and $(n+2)(n+1)/2$, respectively, with the property that there exists a natural imbedding of the isometry Lie algebra of the given metric manifold *into* (or, *onto*) $\mathrm{I}_p$, and of the conformal isometry Lie algebra *into* (or, *onto*) $\mathrm{C}_p$. These two algebras turn out to be powerful tools in the investigation of properties of Killing and conformal Killing fields. To illustrate their use, some general statements about isometries of $n$-manifolds are proved with their aid. While some of these results are well known, most, as far as we are aware, are new.

In Sec. 3, these tools are used to investigate symmetries of stationary isolated systems in general relativity. For this, a restriction is made to 3-manifolds representing the space of orbits of the timelike Killing field in stationary space–times which are asymptotically flat at spatial infinity. A complete description of isometry Lie algebras of these 3-spaces is obtained using the following technique: The algebra $\mathrm{C}_\Lambda$ associated with the point $\Lambda$ "at spatial infinity" is constructed and properties of Killing fields on the given 3-spaces are deduced by examining the imbedding mapping from the Lie algebra of Killing fields into $\mathrm{C}_\Lambda$. Using this description, several results concerning isometries of the given stationary space–times are established.

## 2. MATHEMATICAL PRELIMINARIES

### A. Killing data and the algebra $\mathrm{I}_p$

Fix an $n$-dimensional, connected, $C^\infty$ manifold (without boundary) M equipped with a $C^\infty$ metric $g_{ab}$. Let $\xi^a$ denote a Killing field on (M, $g_{ab}$). Then, since[1] $\nabla_a\nabla_b\xi_c = R_{cba}{}^d\xi_d$ where $\nabla$ and $R_{abc}{}^d$ denote respectively the derivative operator and the Rieman tensor on (M, $g_{ab}$), it follows that the pair $(\xi^a,\ \xi_{ab} := \nabla_a\xi_b)_p$ of tensors at any given point $p$ of M characterizes the Killing field $\xi^a$ completely: if $\xi^a|_p = 0$ and $\xi_{ab}|_p = 0$, then the Killing field $\xi^a$ must vanish *everywhere* on $M$. We shall refer to the pair $(\xi^a,\ \xi_{ab})_p$ as *Killing data* of $\xi^a$ at $p$.

Let us suppose, for a moment, that (M, $g_{ab}$) admits two Killing fields $\xi^a$ and $\eta^a$. Denote the corresponding data at $p$ by $(\xi^a,\ \xi_{ab})_p$ and $(\eta^a,\ \eta_{ab})_p$. Then, the commutator $[\xi,\ \eta]^a$ of the two Killing fields is again a Killing field and its Killing data at $p$ is given by the pair $(\xi^m\eta_m{}^a - \eta^m\xi_m{}^a,\ \xi_a{}^m\eta_{mb} - \eta_a{}^m\xi_{mb} - R_{mnab}\xi^m\eta^n)_p$, where indices are raised and lowered using the metric $g_{ab}|_p$.

This fact suggests the following construction. Fix any point $p$ of M and consider the $n(n+1)/2$-dimensional (real) vector space $V_p$ of pairs $(\xi^a,\ F_{ab})$ of vectors and skew-symmetric tensors at $p$. On this vector space, introduce the following bracket, $[\ ,\ ]_p$,

$$
\begin{aligned}
&[(\underset{1}{\xi}{}^a,\ \underset{1}{F}{}_{ab}),\ (\underset{2}{\xi}{}^a,\ \underset{2}{F}{}_{ab})]_p \\
&= (\underset{1}{\xi}{}^m \underset{2}{F}{}_m{}^a - \underset{2}{\xi}{}^m \underset{1}{F}{}_m{}^a,\ \underset{1}{F}{}_a{}^m \underset{2}{F}{}_{mb} - \underset{2}{F}{}_a{}^m \underset{1}{F}{}_{mb} \\
&\quad - R_{mnab}\underset{1}{\xi}{}^m \underset{2}{\xi}{}^n). \qquad (1)
\end{aligned}
$$

It is obvious that the bracket is linear in each element. Thus, we have acquired an $n(n+1)/2$-dimensional algebra associated with the point $p$. We shall denote it by $\mathrm{I}_p$. Note that the bracket is also skew symmetric in the two elements and that the "structure constants" of the algebra are completely determined by $g_{ab}|_p$ and $R_{abc}{}^d|_p$. (However, $\mathrm{I}_p$ is not an associative algebra, nor, in general, a Lie algebra.) We shall now prove several facts about isometry Lie algebras using the notion of Killing data and the existence of $\mathrm{I}_p$.

(i) It is obvious from the construction of $\mathrm{I}_p$ that there is a canonical mapping from the space of Killing fields on (M, $g_{ab}$) into [and, if (M, $g_{ab}$) admits the maximum possible number, $n(n+1)/2$, of Killing fields, onto] $V_p$: send each Killing field to the element of $V_p$ representing its Killing data at $p$. The mapping is clearly

[a)]Supported in part by the NSF contract PHY 76-81102 with the University of Chicago. Address from September, 78: Département de Physique, Université de clermont-Fd. 63170, Aubière, France.
[b)]Visiting Scholar. Permanent address: Département de Mathématiques Pures, Université de Clermont-Fd, 63370 Aubière, France.

 0022-2488/78/1907-1567$1.00  

linear, one to one and bracket preserving. *Thus, the isometry Lie algebra of* (M, $g_{ab}$) *is a sub (Lie)-algebra of* $I_p$ *for all* $p$ *in* M.

(ii) Under what conditions is $I_p$ a Lie algebra? Since the bracket is linear in each element and skew symmetric, one only needs to check the Jacobi identity. One obtains[1]

$$\oint_{1,2,3}[[(\underset{1}{\xi}^a, \underset{1}{F}_{ab}), (\underset{2}{\xi}^a, \underset{2}{F}_{ab})], (\underset{3}{\xi}^a, \underset{3}{F}_{ab})] = (0, M_{ab}),$$

with

$$M_{ab} = \oint_{1,2,3} \underset{1}{\xi}^m \underset{2}{\xi}^n (R_{mncb} \underset{3}{F}_a{}^c - R_{mna}{}^c \underset{3}{F}_{cb}) + 2\underset{1}{\xi}^{[m} \underset{2}{\xi}^{n]} R_{mcab} \underset{3}{F}_n{}^c, \tag{2}$$

where $\oint_{1,2,3}$ denotes the operation of adding terms obtained by cyclic permutations of 1, 2, and 3. If the data under consideration do actually arise from Killing fields, the Jacobi identity is, of course, automatically satisfied. Thus, *if* (M, $g_{ab}$) *admits three or more Killing fields, the Riemann tensor is algebraically constrained at each point of* M.

Next, using Eq. (2), it is straightforward to check that $M_{ab}=0$ for *all* triplets $(\underset{1}{\xi}^a, \underset{1}{F}_{ab})$, $(\underset{2}{\xi}^a, \underset{2}{F}_{ab})$, and $(\underset{3}{\xi}^a, \underset{3}{F}_{ab})$ in $V_p$ if and only if the Riemann tensor *at* $p$ is of constant curvature; $I_p$ *is a Lie-algebra if and only if* $R_{abcd} = [(2R/n(n-1)]g_{a[c}g_{d]b|p}$, where $R$ is the scalar curvature at $p$. Note that the condition on the Riemann tensor restricts its value only at the point $p$; the Riemann tensor is *not* required to be of constant curvature everywhere, not even in a neighborhood of $p$. This fact will play an important role throughout this section.

It is easy to check using Eq. (1) defining $[\ ,\ ]_p$ that if $R_{abcd} = [(2R/n(n-1)]\, g_{a[c}g_{d]b|p}$, the Lie algebra $I_p$ is isomorphic to the deSitter Lie algebra $D(n, \operatorname{sgn} g_{ab|p}, \operatorname{sgn} R_{|p})$, i.e., to the $n(n-1)/2$-dimensional Lie algebra of isometries of a $n$-manifold equipped with a metric of constant curvature, the signature of the metric being the same as that of $g_{ab|p}$ and the sign of the scalar curvature being the same as that of $R|_p$. Hence we are led to the following result: *given* (M, $g_{ab}$), *if there exists a point* $p$ *in* M *such that* $R_{abcd|p} = [(2R/n(n-1)]\, g_{a[c}g_{d]b|p}$, *then the isometry Lie algebra of* (M, $g_{ab}$) *is a sub-Lie algebra of* $D(n, \operatorname{sgn} g_{ab|p}, \operatorname{sgn} R|_p)$. In particular, one has the following result of interest to general relativity: *If the Riemann tensor of a space–time vanishes even at a single point, then the Lie algebra of its Killing fields is a sub-Lie algebra of the Poincaré Lie algebra.*

(iii) Let $n=2$. Then the Riemann tensor always satisfies $R_{abcd|p} = R g_{a[c}g_{d]b|p}$ for all $p$ in M. Thus, in this case, $I_p$ is always a (three-dimensional). Lie algebra. (This does not of course imply that every 2-manifold admits three Killing vectors: The scalar curvature $R$ need not be a constant function on M.) Hence, we recover the well-known result that the isometry Lie algebra of every 2-manifold is a sub-Lie algebra of the isometry Lie algebra of some 2-manifold with constant scalar curvature. This result is in turn useful in proving facts about metric manifolds which admit exactly two Killing fields: Using the definition of the bracket $[\ ,\ ]_p$, a complete list of possibilities relating the Abelian or non-Abelian character of the isometry Lie algebra with the sign of the scalar curvature and the signature of the metric on the integral manifolds of Killing fields can easily be obtained.

(iv) Let M be an $n$-manifold. Let there exist $n$ commuting Killing fields on (M, $g_{ab}$) which span the tangent space at some point $p$ of $M$. We shall show that $g_{ab}$ is necessarily flat and that (M, $g_{ab}$) cannot admit any additional Killing fields which commute with the given $n$. Note first that since the given $n$ Killing fields are closed under the Lie bracket, and since they span the tangent space at $p$, they span the tangent space at *every* point of $M$. [This follows from the fact that if, on an $n$-manifold, a Killing field vanishes on a $(n-1)$ surface, it must vanish everywhere, which, in turn, follows from the result that a Killing field is completely characterized by its Killing data at any one point.] Furthermore, since all Killing fields commute, one can choose suitable linear combinations to obtain $n$ orthonormal Killing fields whose derivatives vanish identically. Next, using the expression of the bracket $[\ ,\ ]_p$ and the fact that the isometry Lie algebra of (M, $g_{ab}$) is a sub-Lie algebra of $I_p$, it follows that the Riemann tensor $R_{abcd}$ of (M, $g_{ab}$) must vanish identically. Finally, using the expression of the bracket $[\ ,\ ]_p$ (and the fact that $R_{abcd}=0$), it follows that $I_p$ cannot admit any element which does not belong to the $n$-dimensional Abelian sub-Lie algebra of the elements of the type $(\xi^a, 0)$ and which commutes with every element of this sub-Lie algebra. Hence it follows that (M, $g_{ab}$) cannot admit any additional Killing fields which commute with all the $n$ given Killing fields.

The assumption that the Killing fields span the tangent space at some point is unecessary if either $n \leq 4$ or if $g_{ab}$ is positive definite: In these cases, the fact that the Killing fields commute itself implies that they must span the tangent space at each point of $M$.[2] Thus, one has the following result. *An n-dimensional metric manifold* (M, $g_{ab}$) *cannot admit more than n commuting Killing fields and* $g_{ab}$ *is necessarily flat if it does admit n, provided at least one of the following conditions is met*: (a) $n \leq 4$, (b) $g_{ab}$ *is positive definite.*

Results discussed above are only meant to illustrate the use of the notion of Killing data and the existence of the algebra $I_p$ to prove properties of Killing fields; the class of results that can be established using these techniques is by no means exhausted. Indeed, essentially every elementary fact about isometries can be proved in a rather simple manner via these techniques.

### B. Conformal Killing data and the algebra $C_p$

We shall now discuss, briefly, a generalization of the notion of Killing data and of the algebra $\mathbf{I}_p$, a generalization which will serve as a useful tool in the investigation of conformal isometry Lie algebras. The notions to be discussed in this subsection will also be of direct use in Sec. 3.

Consider again an $n$-dimensional manifold M, equipped with a metric $g_{ab}$ of arbitrary signature. Fix a conformal Killing field $\xi^a$ on (M, $g_{ab}$). Then, using the conformal Killing equation, one obtains

$$\nabla_a \xi_b = \nabla_{[a}\xi_{b]} + (1/n)(\nabla_m \xi^m) g_{ab},$$

$$\nabla_a \nabla_{[b}\xi_{c]} = R_{cba}{}^m \xi_m + (2/n) g_{a[c}\nabla_{b]}(\nabla_m \xi^m),$$

$$\begin{aligned}\nabla_a \nabla_b(\nabla_m \xi^m) &= [n/(n-2)]\{-\xi^m \nabla_m [R_{ab} - (R/2(n-1))g_{ab}] - (2/n)(\nabla_m \xi^m) \\ &\quad \times [R_{ab} - (R/2(n-1))g_{ab}] + R^m{}_a \nabla_{[m}\xi_{b]} + R^m{}_b \nabla_{[m}\xi_{a]}\}.\end{aligned}$$

Thus, the quadruplet $[\xi^a, \nabla_{[a}\xi_{b]}, \nabla_a \xi^a, \nabla_a(\nabla_m \xi^m)]_p$ at any given point $p$ in M suffices to determine the conformal Killing field $\xi^a$ *everywhere*. We shall therefore refer to this quadruplet as the *conformal Killing data* of $\xi^a$ at $p$ (relative to $g_{ab}$).

Consider now the vector space $V_p{}^c$ of quadruplets ($\xi^a$, $F_{ab}$, $\Phi$, $K_a$) at $p$, where $\xi^a$ is an arbitrary vector; $F_{ab}$, an arbitrary skew tensor; $\Phi$ an arbitrary number; and, $K_a$ an arbitrary covector. This space is clearly $(n+2)$ $(n+1)/2$-dimensional. Following the procedure used in the case of the Killing data, we now introduce a bracket on this $V_p{}^c$:

$$\{(\underset{1}{\xi}{}^a, \underset{1}{F}{}_{ab}, \underset{1}{\Phi}, \underset{1}{K}{}_a), (\underset{2}{\xi}{}^a, \underset{2}{F}{}_{ab}, \underset{2}{\Phi}, \underset{2}{K}{}_a)\}_p$$

$$= (\underset{3}{\xi}{}^a, \underset{3}{F}{}_{ab}, \underset{3}{\Phi}, \underset{3}{K}{}_a),$$

with

$$\begin{aligned}
\underset{3}{\xi}{}^a &= \underset{1}{\xi}{}^m \underset{2}{F}{}_m{}^a - \underset{2}{\xi}{}^m \underset{1}{F}{}_m{}^a - \underset{1}{\Phi}\underset{2}{\xi}{}^a + \underset{2}{\Phi}\underset{1}{\xi}{}^a,\\
\underset{3}{F}{}_{ab} &= \underset{1}{F}{}_a{}^m \underset{2}{F}{}_{mb} - \underset{2}{F}{}_a{}^m \underset{1}{F}{}_{mb} + 2\underset{2}{K}{}_{[a}\underset{1}{\xi}{}_{b]} - 2\underset{1}{K}{}_{[a}\underset{2}{\xi}{}_{b]}\\
&\quad - R_{abcd}\underset{1}{\xi}{}^c \underset{2}{\xi}{}^d,\\
\underset{3}{\Phi} &= \underset{1}{\xi}{}^a \underset{2}{K}{}_a - \underset{2}{\xi}{}^a \underset{1}{K}{}_a,\\
\underset{3}{K}{}_a &= \underset{1}{K}{}^m \underset{2}{F}{}_{ma} - \underset{2}{K}{}^m \underset{1}{F}{}_{ma} + \underset{1}{\Phi}\underset{2}{K}{}_a - \underset{2}{\Phi}\underset{1}{K}{}_a + [1/(n-2)]\, B_{abc}\underset{1}{\xi}{}^b \underset{2}{\xi}{}^c\\
&\quad + [2/(n-2)](\underset{1}{\Phi}\underset{2}{\xi}{}^b - \underset{2}{\Phi}\underset{1}{\xi}{}^b)S_{ab} + [2/(n-2)](\underset{1}{\xi}{}^b \underset{2}{F}{}^m{}_{(b}R_{a)m}\\
&\quad - \underset{2}{\xi}{}^b \underset{1}{F}{}^m{}_{(b}R_{a)m}).
\end{aligned} \tag{3}$$

Here, $S_{ab} = R_{ab} - (R/2(n-1))g_{ab}$, and, $B_{abc} = \nabla_{[b}S_{c]a}$ is the Bach tensor. (For $n > 3$, $B_{bcd} = \frac{1}{2}[(n-2)/(n-3)] \times \nabla^a C_{abcd}$ via the Bianchi identity.) The bracket is clearly linear in each element. We have obtained a $(n+2)(n+1)/2$-dimensional algebra. Denote it by $C_p$.

As in the case of the algebra $\mathbf{I}_p$, one can investigate the structure of the algebra $\mathbf{C}_p$. We shall mention only two aspects: (i) it follows from the definition of the bracket $\{\ ,\ \}_p$ that the conformal isometry Lie algebra of (M, $g_{ab}$) is a sub-Lie algebra of $\mathbf{C}_p$ for each $p$ in M; and, (ii) $\mathbf{C}_p$ is a Lie algebra (i.e., $\{\ ,\ \}_p$ satisfies the Jacobi identity) if and only if one of the following holds: (a) $n=2$; (b) $n=3$ and $B_{abc}|_p = 0$; or, (c) $n>3$ and, $C_{abcd}|_p = 0$, $\nabla^a C_{abcd}|_p = 0$. Again, one can use these facts to prove properties of conformal Killing vectors.

*Remarks*: (i) Note that, although the notion of conformal Killing fields is conformally invariant—it refers only to the conformal structure rather than to the Riemannian structure on M—the notion of conformal Killing data is not; to obtain the data from a conformal Killing field, we have used a specific metric in the conformal class. This is the reason behind the presence of nonconformally invariant terms (involving the Ricci tensor) in the structure constants of the bracket $\{\ ,\ \}_p$. The final results—e.g., the necessary and sufficient conditions for $\{\ ,\ \}_p$ to be a Lie bracket—are, however, conformally invariant, as they must be. It would, nonetheless, be desirable if the entire analysis could be carried out in a manifestly conformally invariant fashion. (ii) The condition for $\mathbf{I}_p$ to be a Lie algebra involves only the value of the Riemann tensor at $p$ while the analogous condition for $\mathbf{C}_p$ involves both, the value of the Weyl tensor *and* that of its derivative at $p$. This difference reflects the fact that whereas the metric structure is rigid of order one the conformal structure is rigid of order two.[3]

## 3. ASYMPTOTICALLY FLAT STATIONARY SPACE-TIMES

We shall now use the tools developed in the previous section to analyze symmetries of stationary isolated systems in the framework of general relativity.

Fix a stationary space–time (M, $g_{ab}$, $t^a$). Denote by S the manifold of orbits of $t^a$, and by $h_{ab}$, the natural metric on S. The space–time (M, $g_{ab}$, $t^a$) will be said to be *asymptotically flat at spatial infinity* provided there exists a $C^\infty$ 3-manifold $\hat{S}$ equipped with a $C^\infty$ metric $\hat{h}_{ab}$ satisfying the following conditions[4]: (i) As a point set, $\hat{S} = S \cup \Lambda$, where $\Lambda$ is a single point; (ii) $\hat{h}_{ab}$ is positive definite in a neighborhood of $\Lambda$ and $\hat{h}_{ab} = \Omega^2\, h_{ab}$ on S where $\Omega$ is a scalar field on $\hat{S}$ which is $C^2$ at $\Lambda$ and $C^\infty$ elsewhere; (iii) At $\Lambda$, $\Omega = 0$, $\hat{D}_a \Omega = 0$, and, $\hat{D}_a \hat{D}_b \Omega = 2\hat{h}_{ab}$, where $D$ is the derivative operator on (S, $\hat{h}_{ab}$); and (iv) there exists a neighborhood $N$ of $\Lambda$ in $\hat{S}$ such that in $\Pi^{-1}(S \cap N)$ Einstein's vacuum equation is satisfied, where $\pi$ is the natural projection mapping from M onto S.

Throughout this section we shall assume that the given stationary space–time (M, $g_{ab}$, $t^a$) is asymptotically flat in the sense of this definition; we wish to analyze the constraints imposed on the structure of additional Killing fields of (M, $g_{ab}$) by the assumption of asymptotic flatness at spatial infinity.

Recall, from Sec. 2, that if the curvature tensor of a metric manifold vanishes even at a single point, its isometry Lie algebra is a sub-Lie algebra of that of a flat manifold (of the same dimension and equipped with the metric of the same signature). Since $(S, h_{ab})$ is asymptotically Euclidean[4]—its curvature vanishes asymptotically—one might expect the isometry Lie algebra of $(S, h_{ab})$ to be a sub-Lie algebra of the isometry Lie algebra of the Euclidean space. In the first part of this section we shall show that this expectation is indeed correct. In the second part, we shall further assume that the total mass associated with the space–time is nonzero, and analyze the additional restrictions imposed by this assumption on the permissible isometries. In each case, the analysis of permissible Killing fields on $(S, h_{ab})$ will, in turn, yield information about the structure of isometries of $(M, g_{ab})$.

### A. General asymptotically flat space-times

Let us begin by analyzing the structure of the isometry Lie algebra of $(S, h_{ab})$. The key idea is to examine the behavior of Killing fields on $(S, h_{ab})$ *at* the point $\Lambda$ at infinity. Since the point $\Lambda$ belongs to the completion $\hat{S}$ and not to S itself, and since the metric $h_{ab}$ is not even defined at $\Lambda$, (recall that $\Omega|_\Lambda = 0$), we must first regard Killing fields on $(S, h_{ab})$ as conformal Killing fields on $(S, \hat{h}_{ab})$ and *then* look for their extensions to the point $\Lambda$. Let $\xi^a$ be a Killing field on $(S, h_{ab})$. Then, $\mathcal{L}_\xi \hat{h}_{ab} = 2\Omega^{-1}(\mathcal{L}_\xi \Omega)\hat{h}_{ab}$; $\xi^a$ is a *conformal* Killing field on $(S, \hat{h}_{ab})$. Since the metric $\hat{h}_{ab}$ is smooth at $\Lambda$, $\xi^a$ admits a smooth extension $\hat{\xi}^a$ to $\hat{S}$.[3] (On S, $\hat{\xi}^a = \xi^a$.) Hence, one can examine the conformal Killing data of $\hat{\xi}^a$ at $\Lambda$. Since $\hat{\xi}^a$ is not only a conformal Killing field on $(\hat{S}, \hat{h}_{ab})$ but also a *Killing* field on $(S, h_{ab})$, its data at $\Lambda$ are constrained. To see this, note first that given two asymptotically Euclidean spaces $(S, h_{ab})$ and $(S', h'_{ab})$ [satisfying conditions (i), (ii), and (iii)] with an isometry $i$ from one to another, the mapping $i$ extends uniquely to their completions $(\hat{S}, \hat{h}_{ab})$ and $(\hat{S}', \hat{h}'_{ab})$, the point $\Lambda$ in $\hat{S}$ being mapped to the point $\Lambda'$ in $\hat{S}'$, and the metric $\hat{h}_{ab}|_\Lambda$ to the metric $\hat{h}'_{ab}|_{\Lambda'}$.[5] Note that, due to condition (iii), one does not have the freedom of rescaling the metric at $\Lambda$: Only those conformal rescalings, $\Omega \to \omega\Omega$, are permissible for which $\omega$ is (smooth and unit) at $\Lambda$. (Thus, the metric at $\Lambda$ is "universal.") Consider now the one parameter family of diffeomorphisms generated by $\hat{\xi}^a$ on S. Each of these diffeomorphisms is an isometry on $(S, h_{ab})$. Hence, its natural extension to $\hat{S}$ must leave $\Lambda$ and the metric $\hat{h}_{ab}$ at $\Lambda$ invariant. Hence, $\hat{\xi}^a|_\Lambda = 0$, and $\hat{D}_a \hat{\xi}^a|_\Lambda = 0$.[6] Thus, two pieces of the conformal Killing data of $\hat{\xi}^a$ vanish identically at $\Lambda$ if $\hat{\xi}^a$ is a Killing field on $(S, h_{ab})$.

Consider now the six-dimensional subspace $V^k_\Lambda$ of the (ten-dimensional) vector space $V^c_\Lambda$ of conformal Killing data at $\Lambda$ of the form $(0, \hat{F}_{ab}, 0, \hat{K}_a)$. This subspace is closed under the conformal Killing bracket $\{\ ,\ \}_\Lambda$ [Eq. (3)]:

$$\{(0, \underset{1}{\hat{F}}_{ab}, 0, \underset{1}{\hat{K}}_a), (0, \underset{2}{\hat{F}}_{ab}, 0, \underset{2}{\hat{K}}_a)\}_\Lambda = (0, \underset{3}{\hat{F}}_{ab}, 0, \underset{3}{\hat{K}}_a)$$

where

$$\underset{3}{\hat{F}}_{ab} = \underset{1}{\hat{F}}_a{}^m \underset{2}{\hat{F}}_{mb} - \underset{2}{\hat{F}}_a{}^m \underset{1}{\hat{F}}_{mb} \quad \text{and} \quad \underset{3}{\hat{K}}_a = \underset{1}{\hat{K}}^m \underset{2}{\hat{F}}_{ma} - \underset{2}{\hat{K}}^m \underset{1}{\hat{F}}_{ma}. \tag{4}$$

Furthermore, the bracket $\{\ ,\ \}_\Lambda$ on this subspace is a Lie bracket: Commutation relations (4) are precisely the same as those satisfied by the generators of isometries in the Euclidean space. [See the expression of Killing bracket in Eq. (1).] Finally, it is obvious from the above discussion that there is a natural one to one mapping from the vector space of Killing fields on $(S, h_{ab})$ into the six-dimensional subspace $V^k_\Lambda$ of the space $V^c_\Lambda$ of conformal Killing data at $\Lambda$, and that this mapping is bracket preserving. Thus, *the isometry Lie algebra of $(S, h_{ab})$ is necessarily a sub-Lie algebra of the isometry algebra of the Euclidean 3-space.*

Let us now return to Killing fields on $(M, g_{ab})$. Let $\zeta^a$ be a Killing field on $(M, g_{ab})$ with the property that $\mathcal{L}_\zeta t^a = Kt^a$. (Since $\mathcal{L}_\zeta t^a$ must itself be a Killing field, it follows that $k$ must be a constant on M.) Set $\xi^a = h^a{}_b \zeta^b$, where $h_{ab} = g_{ab} + \lambda^{-1} t_a t_b$, $(\lambda = -t^a t_a)$ is a natural projection operator on the 3-flat orthogonal to $t^a$. ($h_{ab}$ on M is the pull back of the natural metric on S.) Then, $t_a \xi^a = 0$ and $\mathcal{L}_t \xi^a = 0$. Thus, $\xi^a$ induces, naturally, a vector field on S which we also denote by $\xi^a$.[7] It is easy to check that this $\xi^a$ is the generator of an isometry on $(S, h_{ab})$; $\mathcal{L}_\xi h_{ab} = 0$. Let L denote the Lie algebra of Killing fields on $(M, g_{ab})$ whose commutator with $t^a$ is a multiple of $t^a$. Clearly, the quotient Q of this **L** by $t^a$ is itself a Lie algebra. Furthermore, by above remarks, there is a natural imbedding of Q into the isometry Lie algebra of $(S, h_{ab})$.

The assumption of asymptotic flatness *at spatial infinitely* constrains only the Lie algebra **Q**; no essential restriction is imposed by this assumption on the structure of Killing fields on $(M, g_{ab})$ whose commutator with $t^a$ fails to be a multiple of $t^a$. Using the result obtained above concerning the isometry Lie algebra of $(S, h_{ab})$, we can conclude the following. *If $(M, g_{ab})$ is asymptotically flat at spatial infinity, Q is a sub-Lie algebra of the Lie algebra of Killing fields in the Euclidean space.* In particular, the dimension of Q can not exceed six, Q admits an Abelian Lie ideal of dimension less than or equal to three, and the quotient of Q by this ideal is a sub-Lie algebra of SO(3).

### B. Space-times with nonzero total mass: Further reduction of permissible isometries

Without additional restrictions on the class of space–times being considered, we cannot hope to obtain further constraints on Killing fields: If $(M, g_{ab})$ is flat, $(S, h_{ab})$ would be isometric with the Euclidean space and would therefore admit all six Killing fields. What we need therefore is a condition which rules out space–times whose asymptotic curvature "approaches zero too fast." A natural candidate is the following: Demand that the total mass associated with the space–time be nonzero. It turns out that this apparently weak condition imposes rather severe constraints on the permissible isometries; the Lie algebra Q can now be shown to be a sub-Lie algebra of SO(3).

Let us then assume that the total mass $m$ associated with the space–time is nonzero. Consider on S the scalar field $\hat{f}$ defined by[8] $\hat{f}=\lambda^{-2}\hat{h}^{ab}[(\hat{D}_a\lambda)(\hat{D}_b\lambda)+\omega_a\omega_b]$ where $\lambda=-t_at^a$ is the norm and $\omega_a=\epsilon_{abcd}t^b\nabla^ct^d$ is the twist of the stationary Killing field. Then, $\hat{f}|_\Lambda=4m^2$. Thus, in a neighborhood of $\Lambda$ in $\hat{S}$, $\hat{f}$ is positive. Furthermore, since Einstein's equation holds in $\Pi^{-1}(N\cap S)$ where $N$ is a neighborhood of $\Lambda$ in $\hat{S}$ and $\Pi$ is the natural projection from M to S, it follows[9] that $\hat{f}$ is $C^\infty$ everywhere in $N$ (including the point $\Lambda$). Let $\xi^a$ be a Killing field on (S, $h_{ab}$) representing an element of Q. [That is, let there exist a Killing field $\zeta^a$ on (M, $g_{ab}$) whose commutator with $t^a$ is a multiple of $t^a$ and let $\xi^b$ equal[7] $h_a{}^b\zeta^a$.] Then, it follows that $\mathcal{L}_\xi\lambda^{-2}h^{ab}[(\hat{D}_a\lambda)(\hat{D}_b\lambda)+\omega_a\omega_b]=0$ on S, and hence, $\mathcal{L}_{\hat{\xi}}=-2\hat{\Phi}\hat{f}$ on $\hat{S}$, where $\hat{\Phi}=\frac{1}{3}\hat{D}_a\hat{\xi}^a$. Taking the derivative of this equation and evaluating the result at $\Lambda$, one obtains $2\hat{f}\hat{K}_a|_\Lambda=-\hat{F}_a{}^m\hat{D}_m\hat{f}|_\Lambda$, where $\hat{K}_a=\hat{D}_a\hat{\Phi}$ and $\hat{F}_{ab}=\hat{D}_{[a}\hat{\xi}_{b]}$. Thus, *if the total mass associated with the space–time is nonzero, conformal Killing data of* $\hat{\xi}^a$ *at* $\Lambda$ *are further constrained*; the "$\hat{K}_a$ piece" of the data is completely determined by the "$\hat{F}_{ab}$ piece" and the value of $\hat{D}_a\log\hat{f}$ at $\Lambda$.[8] Consequently, the element of Q represented by $\xi^a$ can be now completely characterized by the value of $\hat{D}_{[a}\hat{\xi}_{b]}$ at $\Lambda$: If $\hat{D}_{[a}\hat{\xi}_{b]}|_\Lambda=0$, then $\xi^a=0$ everywhere on S. Since the vector space $V_\Lambda^s$ of second rank skew tensors at $\Lambda$ is only three-dimensional, we can now conclude that if the total mass $m$ associated with the space–time is nonzero, the dimension of Q can not exceed three. (Note that if $m=0$, the left side of the equation $2\hat{f}\hat{K}_a|_\Lambda=-\hat{F}_a{}^m\hat{D}_m\hat{f}|_\Lambda$ vanishes identically and the fourth piece, $\hat{K}_a$, of the conformal Killing data of $\hat{\xi}^a$ at $\Lambda$ remains unconstrained.)

How is the Lie algebra structure of Q constrained? Let us equip the three-dimensional vector space $V_\Lambda^s$ with the following bracket:

$$\{\underset{1}{\hat{F}}_{ab},\ \underset{2}{\hat{F}}_{ab}\}_\Lambda^s=\underset{1}{\hat{F}}_a{}^m\underset{2}{\hat{F}}_{mb}-\underset{2}{\hat{F}}_a{}^m\underset{1}{\hat{F}}_{mb}. \tag{5}$$

Then, it is obvious from Eq. (4) that the natural imbedding of the vector space underlying Q into $V_\Lambda^s$ [which sends the element of Q represented by the Killing field $\xi^a$ on (S, $h_{ab}$) to the skew tensor $\hat{D}_{[a}\hat{\xi}_{b]}|_\Lambda$] maps the Lie bracket between elements of Q to the bracket $\{\ ,\ \}_\Lambda^s$ of Eq. (5). Note, also that $\{\ ,\ \}_\Lambda^s$ is a Lie bracket and that ($V_\Lambda^s$, $\{\ ,\ \}_\Lambda^s$) is the Lie algebra of SO(3).

Thus, we have obtained the following result: *if* (M, $g_{ab}$) *is asymptotically flat and if the total mass associated with it is nonzero*, Q *is a sub-Lie algebra of* SO(3). In particular, the dimension of Q is either zero or one or three.

If Q is three-dimensional, (S, $h_{ab}$) is spherically symmetric at least in a neighborhood of $\Lambda$.[10] Denote the three Killing fields on (S, $h_{ab}$) by $\underset{i}{\xi}^a$, $i=1,2,3$, and the corresponding Killing fields on (M, $g_{ab}$) by $\underset{i}{\zeta}^a$($\underset{i}{\xi}^a=h_b{}^a\underset{i}{\zeta}^b$).[7] Using the fact that $\underset{i}{\xi}^a$ can be chosen so that their orbits are closed on (S, $h_{ab}$), it follows that $\underset{i}{\zeta}^a$ must commute with $t^a$ on M. Next, using this fact and the commutation relations of $\underset{i}{\xi}^a$, it follows that $\underset{i}{\zeta}^a$ themselves are Killing fields on (M, $g_{ab}$). Thus, (M, $g_{ab}$) is itself spherically symmetric and hence, by Birkhoff's[11] theorem, isometric with the Schwarzschild space–time in a neighborhood of infinity where Einstein's equation holds. Thus, we have the following result. *If a stationary space–time* (M, $g_{ab}$, $t^a$) *with nonzero total mass admits more than one Killing field whose commutator with* $t^a$ *is a multiple of* $t^a$, *then at least one of the following must hold*: (i) $g_{ab}$ *is isometric to the Schwarzschild metric outside a possible world tube, or,* (ii) (M, $g_{ab}$, $t^a$) *fails to be asymptotically flat at spatial infinity.*

The fact that Q cannot be of dimension two has an interesting consequence: *An asymptotically flat stationary space–time with nonzero mass can not be axisymmetric about two distinct axes unless it has additional isometries* (e.g., spherical symmetry). For, if it has no additional isometries, by Carter's[12] theorem, each axial Killing field must commute with the stationary Killing field and hence Q must be two-dimensional.

Note, finally, that we can classify Killing fields on (S, $h_{ab}$) by their behavior near the point $\Lambda$ at infinity. If the second piece, $\hat{F}_{ab}=\hat{D}_{[a}\hat{\xi}_{b]}$, of the conformal Killing data (w.r.t. a rescaled metric $\hat{h}_{ab}$) of a Killing field $\xi^a$ on (S, $h_{ab}$) vanishes at $\Lambda$, the one-parameter family of diffeomorphisms generated by $\hat{\xi}^a$ on $\hat{S}$ leaves not only the point $\Lambda$ but also the tangent space at $\Lambda$ invariant; its action is nontrivial only in the second jet over $\Lambda$. Such a Killing field may be called a *translation*. [Note also that, if $\hat{D}_{[a}\hat{\xi}_{b]}|_\Lambda=0$, the norm $h_{ab}\xi^a\xi^b$ of $\xi^a$ on (S, $h_{ab}$) remains bounded as one approaches $\Lambda$.] If $\hat{D}_{[a}\hat{\xi}_{b]}|_\Lambda\neq0$, then, although the action of $\hat{\xi}^a$ on $\hat{S}$ leaves the point $\Lambda$ invariant, it causes a rotation in the tangent space at $\Lambda$. Such a Killing field may be called a *rotation*. (As one might intuitively expect, its norm w.r.t. the metric $h_{ab}$ does grow unboundedly as the point $\Lambda$ is approached along any smooth curve.) Note that, in this terminology, although the notion of a "pure" translation is meaningful, that of a "pure" rotation is not: If $\hat{F}_{ab}|_\Lambda\neq0$, the value of the fourth piece $\hat{K}_a|_\Lambda$ of the conformal Killing data fails to be conformally invariant. This is precisely the situation one expects from the structure of the group of isometries in the Euclidean 3-space. Finally, it follows from our discussion above that if (S, $h_{ab}$) is a manifold of orbits of a stationary space–time with nonzero mass, which is asymptotically flat at spatial infinity, (S, $h_{ab}$) can not admit a translation Killing field.

## 4. DISCUSSION

There exist several methods of generating new solutions of Einstein's equation with one or more Killing fields. It is often the case that although one can generate solutions with relative ease, one can not associate any simple physical interpretation with these solutions. A major obstacle is that it is difficult to decide whether or not a given solution is asymptotically flat. Results obtained in Sec. 3 may turn out to be especially useful in making these decisions. Indeed, most of the new solutions obtained by these methods are relatively rich in isometries. Hence, by examining the isometry Lie algebras, one might be able to draw conclusions on the asymptotic behavior of these solutions. For example, if a given solution admits a timelike Killing field, and, in addition, more than one Killing field which commutes with the timelike one, one can conclude that the solution does not represent a new and interesting model for isolated systems: If its total mass is nonzero, either it is Schwarzschildean in a neighborhood of infinity or it must fail to be asymptotically flat. Thus, results obtained in Sec. 3 represent a curious interplay between local and global properties of space–times.

The discussion of Sec. 3 also yields some insight into the notion of asymptotic flatness at spatial infinity. On the other hand, the results obtained are essentially exhaustive: One has been able to prove most of the properties of Killing fields that one intuitively expects to hold in the case of stationary space–times which are asymptotically flat at spatial infinity. Since none of the conditions in the definition of asymptotic flatness was introduced for the express purpose of analyzing isometries, the fact that an exhaustive analysis is possible, and furthermore leads to intuitively expected results, provides a strong support in favor of this definition. On the other hand, every result in Sec. 3 is subject to the rather severe restriction of stationarity. Why was this restriction made? It is because, only in the case of stationary space–times does one have a completely unambiguous notion of asymptotic flatness at spatial infinity which is free of controversies and which does not refer to null infinity.[13] Thus, the major limitation of the present analysis stems directly from that of the notion of asymptotic flatness at spatial infinity itself. One can[14] similarly analyze the constraints on isometries imposed *only* by asymptotic flatness at null infinity. In this case, one does not need to restrict oneself to stationary space–times.

Finally, we wish to emphasize that the analysis made in Sec. 3 represents only an illustration of the use of techniques developed in Sec. 2 to a case of interest in general relativity; these techniques are in fact applicable to a wide variety of situations.

## ACKNOWLEDGMENTS

We wish to thank Robert Geroch, Charles Misner, and Bernard Schutz for discussions. One of us (A.M.A.) also thanks the Chicago relativity group for its hospitality.

[1]Our conventions are the following: $\nabla_{[a}\nabla_{b]}K_c = R_{abc}{}^{d}K_d$, $R_{ab} = R_{amb}{}^{m}$, and $R = R_a{}^{a}$.

[2]Let us assume that there exists a point $p$ at which the $n$ given Killing fields span a $m$ flat with $m<n$. Then, there must exist $k \equiv (n-m)$ independent Killing fields whose data at $p$ is of the type $(0, F_{ab})$. Using the fact that the $n$ Killing fields commute and the first piece of the Killing data in the expression of the bracket, $[\ ,\ ]_p$, between two data, it follows that $k>3$ and $n>4$. Let us now assume that $n>4$ and that the metric $g_{ab}$ is positive definite. Then, again using the expression of the bracket, $[\ ,\ ]_p$ it is easy to show that the Lie algebra of the $k$ Killing fields [whose data at $p$ is of the type $(0, F_{ab})$] is necessarily a sub-Lie algebra of SO($k$). This is however impossible since SO($k$) does not admit a $k$-dimensional *abelian* sub-Lie algebra. Hence the assumption that the Killing fields span a $m$ flat at $p$ with $m<n$ is inconsistent with the assumption that the $n$ Killing fields commute.

[3]For details, see, e.g., R. Geroch, Commun. Math. Phys. **13**, 180 (1969).

[4]For the motivation behind the conditions in the definition as well as for details, see R. Geroch, J. Math. Phys. **11**, 2580 (1970) and R.O. Hansen, J. Math. Phys. **15**, 1 (1974).

[5]See the first paper in Ref. 4.

[6]An alternative and more analytic proof is the following. Since $\hat{\xi}^a$ and $\hat{h}_{ab}$ are smooth everywhere on $(\hat{\mathbf{S}}, \hat{h}_{ab})$, it follows that $\mathcal{L}_{\hat{\xi}}\hat{h}_{ab}$ is also smooth. However, $\mathcal{L}_{\hat{\xi}}\hat{h}_{ab} = 2\Omega^{-1}(\hat{\xi}^m\hat{D}_m\Omega)\hat{h}_{ab}$. Next, $\lim_{\to\Lambda}\hat{D}_a\Omega^{1/2}$ exists by l'Hôpital's rule and is just the unit tangent vector to the curve of approach to $\Lambda$. Hence, $\lim_{\to\Lambda}\hat{\xi}^a = 0$. The result $\hat{\Phi} = 0$ follows from the fact that since $\hat{\xi}^a$ generates isometries on $(\mathbf{S}, h_{ab})$, $\mathcal{L}_{\hat{\xi}}\hat{h}_{ab}|_{\Lambda} = 0$.

[7]There is a natural isomorphism between tensor fields on **S** and tensor fields on **M** all of whose indices are orthogonal to $t^a$ and whose Lie derivative by $t^a$ vanishes. We shall not distinguish between tensor fields related by this isomorphism.

[8]$f$ is just one of the scalars constructed out of $g_{ab}$ and $t^a$ which could have been chosen for the present purpose. Any other scalar $\hat{f}'$ which is $C^2$ and nonzero at $\Lambda$ and which satisfies $\mathcal{L}_{\xi}\hat{f}' = \hat{\Phi}\hat{K}$ where $\hat{K}$ is $C^1$ and nonzero at $\Lambda$, will lead to a constraint on the values of $\hat{F}_{ab}|_{\Lambda}$ and $\hat{K}_a|_{\Lambda}$.

[9]See the second paper in Ref. 4.

[10]It suffices to show that there exists a neighborhood of $\Lambda$ at no point of which the tangent space is spanned by the three Killing fields. Suppose no such neighborhood exists. Then, using the fact that the Killing fields satisfy the commutation relations of SO(3), it follows that the norm $\lambda$ of $t^a$ must be constant in a neighborhood of $\Lambda$, and hence, that the total mass $m$ associated with $(\mathbf{M}, g_{ab}, t^a)$ must vanish.

[11]G. D. Birkhoff, *Relativity and Modern Physics* (Harvard U.P., Cambridge, Mass., 1923).

[12]B. Carter, Commun. Math. Phys. **17**, 233 (1970).

[13]In the nonstationary context, two possibilities present themselves: One might continue to use a "three-dimensional" notion of asymptotic flatness at spatial infinity, replacing the manifold of orbits **S** by a spacelike Cauchy surface, or, one might formulate the notion of asymptotic flatness in a completely new "four-dimensional" spirit. If **S** is replaced by a Cauchy surface, the rescaled metric $\hat{h}_{ab}$ cannot be $C^\infty$ at $\Lambda$ and there is some controversy about the precise degree of differentiability that one can demand. Also the question of uniqueness of the conformal completion is still open. In any case, the "four-dimensional" approaches appear to be more promising if one is interested in analyzing isometries of space–time as a whole. In particular, when technicalities concerning Sommers' ("four-dimensional") definition of asymptotic flatness are settled, making his boundary "Psi" completely unambiguous, one might be able to generalize the present analysis of isometries to nonstationary space–times. (See P.D. Sommers, J. Math. Phys. **19**, 549 (1978).

[14]A. Ashtekar and B.C. Xanthopoulos, "Isometries compatible with asymptotic flatness at null infinity: A complete description," (to appear in J. Math. Phys.).

# On the separability of the sine-Gordon equation and similar quasilinear partial differential equations

Anthony Osborne and Allan E. G. Stuart

*Department of Mathematics, The City University, St. John Street, London EC1V 4PB, England*
(Received 13 December 1977)

The separability of the sine-Gordon equation (SGE) is defined and studied in detail. We find a general class of dependent-variable transformations under which the SGE is separable. This class may be reduced to a two-parameter generalization of the usual transformation adopted, by requiring the transformations to reduce to the identity in the linear limit of the SGE (i.e., the Klein–Gordon equation). The method developed for studying the separability of the SGE is then applied to more general quasilinear equations and a discussion of the limitations of the method, and of separable solutions in general, is also given.

## 1. INTRODUCTION

Interest in nonlinear or, to be more precise, quasilinear wave equations has focused in recent years on the sine-Gordon equation (SGE)

$$\phi_{xx} - \phi_{tt} = \sin\phi, \tag{1.1}$$

where $\phi(x,t)$ is a scalar field in one space $(x)$ and one time $(t)$ dimension and the subscripts denote partial derivatives of $\phi$ with respect to $x$ and $t$. The SGE is Lorentz covariant, has a variational derivation and, as a mathematical model describing a variety of interesting wave and particle phenomena, has remarkable properties at both the classical and quantum levels. There is a vast literature on the subject and much of this can be traced from the pellucid review articles of Barone *et al.*,[1] Scott *et al.*[2] and Rajaraman.[3]

Now most of the classical studies of the SGE and its applications have concentrated, quite naturally, on the soliton solutions and some of the associated formal properties (infinite number of conservation laws, Bäcklund transformations, inverse method of solution, separable Hamiltonians etc.). However, the equation also has an underlying mathematical structure which needs to be investigated from a much wider point of view. For example, the initial value problem for the SGE and similar equations, such as the Korteweg—de Vries and its modified forms, can all be solved by linear methods using the inverse scattering formalism.[4] The question then arises as to whether this is merely a technique which happens to work for these equations, equivalent to, say, the reduction of an integral to standard form, or does it imply some deeper structure? If the latter is true, and this seems probable, and if, as seems likely, equations of this type are going to play an important part in the development of nonlinear physics, then it becomes desirable to develop a body of fundamental theory for these equations corresponding to that existing for, say, the linear partial differential equations of mathematical physics. As a possibly useful contribution to this development we present, in this paper, a report of an attempt to formulate and answer some of the questions that would arise in such a fundamental theory. In particular, we study the problem of the separability of the SGE and similar nonlinear equations.

Throughout the paper we shall assume, unless otherwise stated, that $\phi(x,t)$ is a continuous map from $(x,t)$ to the reals which is at least twice differentiable with respect to both $x$ and $t$. The domain of $\phi$ is thus the whole of $\mathbb{R}^2$, while its range is, in general, the 1-sphere (i.e., the space of real numbers modulo $2\pi$ with the usual metric).

## 2. SPECIFICATION OF THE PROBLEM

The question of the separability of the SGE first arose, in a somewhat implicit manner, in a 1971 paper by Lamb.[5] Lamb showed that (1.1) possesses a class of solutions of the form

$$\phi(x,t) = 4\tan^{-1}[X(x)T(t)], \tag{2.1}$$

where the single variable functions $X$ and $T$ are solutions of the uncoupled, ordinary differential equations

$$(X')^2 = pX^4 + mX^2 + q, \tag{2.2a}$$

$$(T')^2 = -qT^4 + (m-1)T^2 - p. \tag{2.2b}$$

The primes in (2.2) denoting ordinary derivatives while $p, q$ and $m$ are arbitrary constants which may be complex. For example, choosing $p=q=0$ gives single soliton solutions with speed $u = \sqrt{(1-1/m)}$.

More recently (1976), Zagrodzinski[6] extended Lamb's analysis by showing that both (1.1) and its "elliptic" variant

$$\phi_{xx} + \phi_{tt} = \sin\phi, \tag{2.3}$$

when combined into one equation

$$\phi_{xx} + \epsilon\phi_{tt} = \sin\phi \quad (\epsilon = \pm 1), \tag{2.4}$$

have solutions of the form

$$\phi(x,t) = \pm 4\tan^{-1}[X(x)T(t)] + (\pi/2)(1-\delta) \tag{2.5}$$

provided that

$$(X')^2 = pX^4 + \delta mX^2 + q, \tag{2.6a}$$

$$\epsilon(T')^2 = qT^4 + \delta(1-m)T^2 + p, \tag{2.6b}$$

where $\delta = \pm 1$ and $p, q, m \in \mathbb{C}$ the field of complex numbers. Note, however, that the two different values of $\delta$ do not give independent solutions of the SGE, and so one can use either value to generate all the solutions. In this paper we shall consistently use the value $\delta = +1$.

The existence of solutions of the type (2.5) show that the SGE and its elliptic variant are separable, not in terms of the original dependent variable $\phi$, but rather,

 0022-2488/78/1907-1573$1.00  

in terms of a new dependent variable

$$\psi(x,t)=\tan(\phi/4) \tag{2.7}$$

This leads us to pose the following questions:

(1) Is the dependent variable transformation (2.7) the only transformation of (2.4) which leads to a separable equation for $\psi(x,t)$ [i.e., one in which the functions $X(x)$ and $T(t)$ can be separated based on the assumption that $\psi(x,t)=X(x)T(t)$]?

(2) If there is more than one transformation $\phi\to g(\psi)$ which leads to a separable equation for $\psi$, then does this imply that the solutions (2.5) are only a subset of the set of all solutions of (2.4) of the form $\phi=g(XT)$?

Our answers to these and related questions, together with the relevant analysis, is given below. The plan of the rest of the paper and a summary of our main findings are as follows. In Sec. 3 we develop a method (for a limited class of equations) for deciding on whether a given nonlinear or quasilinear partial differential equation is separable or not. In Sec. 4 we use this method to investigate the separability of the SGE. We find that there is a general class of dependent variable transformations, expressed in terms of Jacobian elliptic functions,[7] under which the SGE is separable and of which (2.7) is a particular example. The occurrence of the elliptic functions is not surprising since the genealogy of separable solutions of (2.4) begins with the simple pendulum equation[8] which, of course, has elliptic function solutions in general. The existence of this general class of transformations leads to a wider set of separable solutions than that given by (2.5). However, we show that this set may be restricted to a two-parameter generalization of (2.7) by requiring that in the linear limit of the SGE, i.e., the Klein–Gordon equation, the transformations reduce to the identity. In Sec. 5 we apply our method to more general quasilinear equations and treat, in particular, the SGE in one time and two space dimensions. Finally, in Sec. 6, we discuss the limitations of our method and of separable solutions in general.

## 3. SEPARABILITY OF PARTIAL DIFFERENTIAL EQUATIONS

Any homogeneous, linear partial differential equation in two independent variables,

$$L[\phi(x,t)]=0, \tag{3.1}$$

is separable providing a function $f(x,t)$ exists such that

$$L[X(x)T(t)]=[g(x)+h(t)]f(x,t)X(x)T(t), \tag{3.2}$$

where $L$ is a linear partial differential operator and it is assumed that $\phi(x,t)=X(x)\,T(t)$.[9] If Eq. (3.1) has constant coefficients, then it is separable if $f(x,t)=1$. We are not concerned here with separability criteria depending on given initial and boundary conditions, but only with the separability of the operator $L$ itself.

The rule (3.2) is not readily extendible to the general nonlinear case. For example, putting $\psi=\tan(\phi/4)$ in (1.1) gives us the transformed equation

$$(1+\psi^2)(\psi_{xx}-\psi_{tt})-2\psi(\psi_x{}^2-\psi_t{}^2)=\psi(1-\psi^2) \tag{3.3}$$

which is separable, according to Lamb and Zagrodzinski, if $\psi(x,t)=X(x)T(t)$, but rule (3.2) is clearly not applicable. We thus have to look for a general rule which can be applied at least to certain types of nonlinear, partial differential equations.

Consider the equation

$$\sum_n P_n(\psi)Q_n(\psi_\alpha,\psi_{\alpha\alpha})=R(\psi), \tag{3.4}$$

where $\alpha=x$ and/or $t$, $\psi=\psi(x,t)$ and $P_n, Q_n$ and $R$ are polynomials of any given degree with each $Q_n$ being a sum of a polynomial in $\psi_\alpha$ and a polynomial in $\psi_{\alpha\alpha}$. By hypothesis, let $\psi(x,t)=X(x)T(t)$. Substitution in (3.4) then leads to the equation

$$\sum_n P_n(XT)Q_n(X'T,T'X,T'X',X''T,T''X)=R(XT). \tag{3.5}$$

This is, in general, an implicit relation between $X,T$ and their derivatives and so the linear approach breaks down. However, if the equation is separable, we can write $X=g(x)$ and $T=h(t)$, where $g$ and $h$ are 1–1 and differentiable over the domains and ranges considered. It then follows that

$$X'(x)=g'(x)=g'(g^{-1}(X))=(g'\circ g^{-1})(X)=g_1(X) \tag{3.6a}$$

and

$$T'(t)=h'(t)=h'(h^{-1}(T))=(h'\circ h^{-1})(T)=h_1(T). \tag{3.6b}$$

In general, $g(x)$ and $h(t)$ will contain more than one arbitrary constant and so $g_1(x)$ and $h_1(t)$ will also contain at least one arbitrary constant.

Now $X'$ may or may not be analytic in $X$ over the range considered. Similarly for $T'$. However, let us assume that each can be expanded in a power series as follows:

$$X'=\sum_{r=0}^{\infty}a_rX^{r+\lambda},\quad T'=\sum_{r=0}^{\infty}b_rT^{r+\rho}, \tag{3.7}$$

where $a_r$, $b_r\in\mathbb{R}$ and $\lambda$ and $\rho$ are possibly noninteger. If the $a_r$, $b_r$ exist and are finite, then Eq. (3.4) is separable. On the other hand, if $a_r$, $b_r$ do not exist, are infinite, or are all identically zero, then either our original hypothesis is incorrect and the equation is not separable in this form, or else the trivial solution $\psi(x,t)\equiv 0$ is the only separable solution of the equation. If the series (3.7) terminate, they represent first-order equations which can be solved for $X(x)$ and $T(t)$. However, if the series do not terminate then the question of uniform convergence must be looked into [i.e., the differential equations (3.7) may exist for certain values of $X,T$, but not others].

Using formal differentiation and multiplication of power series on (3.7) and substituting into (3.4) leads to an identity of the form

$$\sum_n P_n(XT)\times Q_n\left\{\sum_{r=0}^{\infty}a_rX^{r+\lambda}T,\ \sum_{r=0}^{\infty}b_rT^{r+\rho}X,\ \sum_{r=0}^{\infty}A_rX^{r+\lambda}T,\ \sum_{r=0}^{\infty}B_rT^{r+\rho}X\right\}=R(XT), \tag{3.8}$$

where the $A_r$ and $B_r$ are coefficients involving sums and products of the original expansion parameters. This identity then gives us a set of recurrence relations for the $a_r$, $b_r$, $\lambda$, and $\rho$ which, in principle, can be solved. There are two possible results. Either we get explicit and consistent expressions for the first-order differential equations (3.7), in which case (3.4) is separable, or else we are led to a contradiction and one of our original hypotheses is false. In the latter case the given partial differential equation may not be separable as it stands, or else the power series expansions for the derivatives of $X$ and $T$ may not be legitimate.

*Example*: Consider the Korteweg–de Vries (KdV) equation

$$\phi_t + \beta\phi\phi_x + \phi_{xxx} = 0 \quad (\beta \in \mathbb{R}), \tag{3.9}$$

where $\phi = \phi(x,t)$. This is not exactly of the form (3.4), but the method is readily extendible to the case where $Q_n$ contains $\psi_{\alpha\alpha\alpha}$, $\psi_{\alpha\alpha\alpha\alpha}$, etc. If we now assume that $\phi(x,t) = X(x)\,T(t)$ and carry through the procedure outlined above, we arrive at the ordinary differential equations

$$X' = a, \tag{3.10a}$$

$$T' = -\beta a T^2, \tag{3.10b}$$

where $a$ is an arbitrary constant. Solving these equations gives us the only separable solutions of the KdV equation as it stands, and these are

$$\phi(x,t) = (x+b)/(\beta t + c), \tag{3.11}$$

where $b$ and $c$ are arbitrary constants.

The method can also be applied to linear equations which are separable according to rule (3.2) and this is best illustrated by means of an example.

*Example*: Consider the partial differential equation

$$\phi_{xx} + \phi_y + \phi_{yy} = 0, \tag{3.12}$$

where $\phi = \phi(x,y)$. Let $\phi = X(x)Y(y)$ and, as there are no first derivatives of $X$ in (3.12), substitute the expansions

$$X'' = \sum_{r=0}^{\infty} a_r X^r, \quad Y' = \sum_{r=0}^{\infty} b_r Y^r. \tag{3.13}$$

This leads to the identity

$$\sum_{r=0}^{\infty} a_r X^r Y + \sum_{r=0}^{\infty} b_r Y^r X \left\{1 + \sum_{r=1}^{\infty} r b_r Y^{r-1}\right\} = 0 \tag{3.14}$$

and a consistent solution of the recurrence relations provided by this identity gives us the following pairs of ordinary differential equations:

$$Y' = aY, \quad X'' = -a(1+a)X, \tag{3.15a}$$

$$Y' = -(1+a)Y, \quad X'' = -a(1+a)X, \tag{3.15b}$$

where $a$ is the separation constant. Solving these equations then leads to the usual separable solutions of (3.12). Note that the nonlinear method, used above, gives easier equations for $X$ and $Y$ (one first-order and one second-order ordinary differential equation) than the elementary method based on rule (3.2) (two second-order equations), although the advantage in this case is minimal.

The method can also be used to separate linear equations which have separable solutions, but which are not separable by a single application of rule (3.2). For example, a fairly trivial equation which falls into this category is $\phi_{xx} + \phi_{xy} + \phi_{yy} = 0$.

We now consider the special case of (3.4) in which the only terms occurring in the $Q_n$ are $\psi_{xx}$, $\psi_{tt}$, $(\psi_x)^2$ and $(\psi_t)^2$, i.e., they are of the first degree in second derivatives and not exceeding the second degree in first derivatives. The equation then takes the form

$$(\psi_x)^2 P_1(\psi) + (\psi_t)^2 P_2(\psi) + \psi_{xx} P_3(\psi) + \psi_{tt} P_4(\psi) = R(\psi) \tag{3.16}$$

and assuming, in this instance, that $(X')^2$ and $(T')^2$ can be expanded in power series in $X$ and $T$, respectively, we arrive at the identity

$$\begin{aligned} &P_1(XT)\sum_{r=0}^{\infty} 2a_r X^{r+\lambda} T^2 + P_2(XT)\sum_{r=0}^{\infty} 2b_r T^{r+\rho} X^2 \\ &\quad + P_3(XT)\sum_{r=1}^{\infty}(r+\lambda)a_r X^{r+\lambda-1} T + P_4(XT)\sum_{r=1}^{\infty}(r+\rho)b_r T^{r+\rho-1} X \\ &= R(XT). \end{aligned} \tag{3.17}$$

In this case the problem is considerably simplified. Comparing coefficients in (3.17), it can be shown that $\lambda$ and $\rho$ cannot be fractional. Also, if $P_3(XT)$ contains a constant term the series for $(X')^2$ terminates, while if $P_4(XT)$ contains a constant term the series for $(T')^2$ terminates. Now the SGE falls into the special category represented by (3.16), but before we can show this we need the following lemma.

*Lemma* 3.1: An equation of the form

$$A\phi_{xx}(x,t) + B\phi_{tt}(x,t) = f(\phi), \tag{3.18}$$

where $A$ and $B$ are constants, is not separable as it stands unless $f(\phi) = \phi$. □

*Proof*: Equation (3.18) is a special case of (3.16), with slightly different notation, where $P_1 = P_2 \equiv 0$, $P_3(\phi) = A$ and $P_4(\phi) = B$. If we now assume separability and expand $X''$ and $T''$ in terms of $X$ and $T$, respectively, and expand $f(XT)$ as a power series in $XT$, then the recurrence relations obtained from the resulting identity have a consistent solution if, and only if, $f(XT) = XT$. ■

We can now look at the SGE and its elliptic variant as given by Eq. (2.4). If we exclude the trivial solutions, $\phi = 2n\pi$ $(n \in z)$, and the time-independent and space-independent solutions for which (2.4) automatically reduces to an ordinary differential equation, then, using Lemma 3.1, we see immediately that the equation is not separable as it stands. In order to put it into a separable form it is necessary, as we shall see later, to make a dependent-variable transformation. Using the usual transformation $\phi(x,t) = 4\tan^{-1}\psi(x,t)$ gives us an equation of the form (3.16) with $\epsilon P_1(\psi) = P_2(\psi) = -2\epsilon\psi$, $\epsilon P_3(\psi) = P_4(\psi) = \epsilon(1+\psi^2)$, and $R(\psi) = \psi(1-\psi^2)$. A consistent solution of the recurrence relations of the corresponding identity (3.17), then leads directly to the Zagrodzinski equations (2.6) for real values of the parameters. However, it is a simple matter to extend the specification of the power series expansions to include complex coefficients as well and hence to recover the whole set of separated equations given by Zagrodzinski.

## 4. TRANSFORMATIONS OF THE SGE WHICH LEAD TO SEPARABLE FORMS

In the last section we studied a method for separating (2.4) after it had been put into separable form via the dependent-variable transformation (2.7). We now turn our attention to the transformation itself and ask whether there are other transformations, either of the dependent variable or of the independent variables, which lead to separable forms of (2.4) or is (2.7) in some way unique? As an indication of the type of information that one can get from such an analysis consider, for example, the substitution of $\psi = \tan(\phi/2)$ into (2.4). In this case the equation reduces to

$$(1+\psi^2)(\psi_{xx}+\epsilon\psi_{tt}) - 2\psi(\psi_x{}^2+\epsilon\psi_t{}^2) = \psi(1+\psi^2), \qquad (4.1)$$

which, by the methods of Sec. 3, is not separable. Thus, in addition to the functional form of the transformation, it appears that the use of the quarter angle is also critical.

Now features such as the above are best studied once the transformations that lead to the separable forms of (2.4) are classified. Our first task, therefore, is to find and classify them. A useful Lemma, in this instance, is the following:

*Lemma* 4.1: Dependent-variable transformations are the only transformations of (2.4) which lead to separable forms. □

We imply here that there is no coordinatization of the domain of (2.4) which will lead to a separable equation. For example, the canonical form

$$4\frac{\partial^2\phi(\xi,\eta)}{\partial\xi\partial\eta} = \sin\phi \qquad (4.2)$$

is not separable.

*Proof*: A general transformation of the independent variables changes (2.4) into an equation with variable coefficients, but maintains the linearity of its derivatives. Thus, by a simple extension of Lemma 3.1, it is not separable. ■

The elimination of the independent-variable transformations allows us to concentrate on the dependent-variable transformations and we now state and prove a theorem about them.

*Theorem* 4.1: The only dependent-variable transformations which lead to separable versions of (2.4) are necessarily of the form

$$\phi = g(\psi(x,t)) = 2\cos^{-1}\mathrm{sn}\left\{\frac{\ln(\alpha\psi)}{k\beta}, k\right\}, \qquad (4.3)$$

where $\psi$ is the new dependent variable $\alpha$, $\beta$, $k \in \mathbb{R}-\{0\}$ with $k \leq 1$ and sn is a Jacobian elliptic function sine amplitude of modulus $k$. □

*Proof*: Applying the dependent-variable transformation $\phi = g(\psi)$ to (2.4) we obtain

$$(\psi_{xx}+\epsilon\psi_{tt})g'(\psi) + (\psi_x{}^2+\epsilon\psi_t{}^2)g''(\psi) = \sin g(\psi). \qquad (4.4)$$

We now assume that $\psi = X(x)T(t)$ and substitute the series expansions $(X')^2 = \sum_{r=0}^{\infty} 2a_rX^r$ and $(T')^2 = \sum_{r=0}^{\infty} 2b_rT^r$ into (4.4). This gives

$$\left\{\sum_{r=1}^{\infty} r(a_rX^{r-1}T+\epsilon b_rT^{r-1}X)\right\} g'(XT) + \left\{\sum_{r=0}^{\infty} 2(a_rX^rT^2+\epsilon b_rT^rX^2)\right\} g''(XT) = \sin g(XT). \qquad (4.5)$$

Since $g$ is a function of $XT$, $g'$, $g''$ and $\sin g$ must all be functions of $XT$. Thus, a necessary condition for (4.5) to be an identity is that the coefficients of $X^rT^r \forall r$ on the lhs are consistent with the coefficients of $(XT)^r$ on the rhs. Hence, on eliminating the zeros on the lhs, we get the reduced equation

$$\beta^2XTg'(XT) + \beta^2X^2T^2g''(XT) = \sin g(XT), \qquad (4.6)$$

where $\beta^2 = 2(a_2+\epsilon b_2)$. It follows, therefore, that a necessary condition for (4.4) to be separable is that $g(\psi)$ satisfies the ordinary differential equation

$$\beta^2\psi\{g'(\psi)+\psi g''(\psi)\} = \sin g(\psi). \qquad (4.7)$$

This is a quasilinear, variable coefficient equation for $g(\psi)$ and can be integrated in two stages. Firstly, multiplying the equation by $2g'$ turns the lhs into the derivative of $(\psi g')^2$ and hence leads to the integrated form

$$(\psi g')^2 = C - (2/\beta^2)\cos g, \qquad (4.8)$$

where $C$ is an arbitrary constant. Next, making the substitutions

$$h = \cos(g/2) \text{ and } z = \ln(\alpha\psi)/k\beta,$$

where $k^2(2+C\beta^2) = 4$, and $\alpha$ has been introduced to take care of the next integration constant, reduces (4.8) to

$$(h')^2 = (1-h^2)(1-k^2h^2), \qquad (4.9)$$

which is the defining equation for the Jacobian elliptic function $\mathrm{sn}(z,k)$. Working backwards through the subsituations then gives us Eq. (4.3) and hence completes the proof of the theorem. ■

To get the subclass of nonelliptic transformations we have to put $k=1$. Then, if we set $\beta = -n$ (not necessarily integral), (4.3) reduces to

$$\psi = (1/\alpha)\tan^n(\phi/4), \qquad (4.10)$$

which is a two-parameter generalization of (2.7). This leads to our next theorem.

*Theorem* 4.2: Separating (2.4), via the dependent-variable transformations (4.10), leads to solutions of the form $\phi = 4\tan^{-1}(\alpha^{1/n}XT)$, where the separating functions $X$ and $T$ are solutions of the ordinary differential equations

$$(X')^2 = pX^4 + mX^2 + q \qquad (4.11a)$$

and

$$\epsilon(T')^2 = \alpha^{2/n}qT^4 + (1-m)T^2 + \alpha^{-2/n}p, \qquad (4.11b)$$

respectively, and $p$, $q$, $m$ are arbitrary constants which may be complex.

*Proof*: This theorem looks fairly obvious and it is. However, the proof, although straightforward, is somewhat tedious and so we shall only mention the main steps. We start by substituting (4.10) into (4.4). This leads to the equation

$$n\psi\{1+(\alpha\psi)^{2/n}\}(\psi_{xx}+\epsilon\psi_{tt})$$
$$-\{(n+1)(\alpha\psi)^{2/n}+(n-1)\}(\psi_x{}^2+\epsilon\psi_t{}^2)$$
$$=n^2\psi^2\{1-(\alpha\psi)^{2/n}\}. \quad (4.12)$$

In order to eliminate terms such as $\psi^{2/n}$ we first assume that $\psi = X^nT^n$ and then use the series expansions $(X')^2 = \sum_{r=0}^{\infty} 2a_rX^r$ and $(T')^2 = \sum_{r=0}^{\infty} 2b_rT^r$. This gives us an identity in which the coefficients are independent of $n$. A consistent solution of the recurrence relations of this identity then leads directly to Eqs. (4.11) and completes the proof of the theorem. ∎

*Remark*: This theorem shows that the use of the general transformation (4.10) leads to separable solutions of (2.4) which are independent of $\alpha$ and $n$. Thus, there is no advantage to be gained by using (4.10) rather than its special case (2.7).

We now consider the general transformation (4.3) of Theorem 4.1. Substituting into (4.4) and assuming that $\psi = XT$ leads to the equation

$$-\beta XT(X''T+\epsilon T''X)$$
$$+\{\beta + k\,\mathrm{sn}(u,k)\mathrm{sn}(u+K,k)\}\{(X')^2T^2+\epsilon(T')^2X^2\}$$
$$=k\beta^2X^2T^2\,\mathrm{sn}(u,k)\mathrm{sn}(u+K,k), \quad (4.13)$$

where $u=\ln(\alpha\psi)/k\beta$ and $K$ is the complete elliptic integral of the first kind. We now run into a difficulty since the elliptic functions in (4.13) are only analytic about $\psi = 0$ if $k=1$. We have already dealt with this case, but it is interesting to see how the restriction comes about. Suppose we assume that it is legitimate to consider the SGE at small values of its amplitude $\phi$. In this case (2.4) reduces to the Klein–Gordon equation

$$\phi_{xx}+\epsilon\phi_{tt}=\phi, \quad (4.14)$$

which, of couse, is separable as it stands. However, other separable forms of (4.14) can be obtained by using the transformation $\phi = g(\psi)$, where $g$ is a solution of (4.7) with sin$g$ replaced by $g$. Thus, if we only consider those solutions of (4.7) which, in the limit of small $\phi$, reduce to a solution of

$$\beta^2\psi\{g'(\psi)+\psi g''(\psi)\}=g(\psi) \quad (4.15)$$

with $g(0)=0$, then this immediately restricts us to the set of nonelliptic transformations (4.10).

One can, of course, get elliptic transformations which are linear in the limit of small $\phi$, but only if we expand about finite values of $\psi$. For example, by considering the Taylor series expansion of $\mathrm{sn}(u,k)$ about $u=K$ it is easy to deduce that the function $\psi=\exp(a+bx+ct)$ leads to a transformation of this type. The resulting transformation is linear in the limit of small $\phi$ and, for appropriate choices of the constants involved, leads to traveling wave solutions of the SGE of the form

$$\phi=\pm 2\cos^{-1}\mathrm{sn}\{(\gamma/k)(x\pm vt),k\}, \quad (4.16)$$

where $v(<1)$ is the wave speed and $\gamma$ is the Lorentz factor $(1-v^2)^{-1/2}$. The waves have an amplitude of $2\pi$ and a wavelength of $4Kk/\gamma$, and in the limit $k=1$ reduce to a soliton or an antisoliton. Note, however, that although the soliton is a limiting case of (4.16), solutions with a finite wavelength (i.e., $0<k<1$) cannot be obtained via the Zagrodzinski prescription [Eqs. (2.5) and (2.6)]. Thus, the elliptic and nonelliptic transformations lead to different classes of solutions with only the solitons in common.

A feature worth mentioning, in the case of the elliptic transformation discussed above, is that the derivation of the transformation is constructive, i.e., it leads directly to the solution.

## 5. SEPARABLE SOLUTIONS OF A MORE GENERAL EQUATION

In this section we use the previous techniques to find solutions of the more general equation

$$\sum_{i=1}^{n}\epsilon_i\phi_{x_ix_i}=f(\phi), \quad (5.1)$$

whre $\epsilon_i=\pm 1$ and $\phi$ is a scalar field in $n$ dimensions. Since this is merely an extension to $n$ dimensions of the equations considered in the last two sections, we shall go through the details fairly rapidly.

Let $\phi=g(\psi)$ be a transformation of the dependent variable in (5.1). The equation can then be written as

$$\left\{\sum_{i=1}^{n}\epsilon_i\psi_{x_ix_i}\right\}g'(\psi)+\left\{\sum_{i=1}^{n}\epsilon_i\psi_{x_i}^2\right\}g''(\psi)=f(g). \quad (5.2)$$

Let $\psi(x_1,\ldots x_n)=\Pi_{i=1}^nX_i(x_i)$ and $(X_i')^2=\sum_{r=0}^{\infty}2a_{ir}X_i{}^r$. Then, following through the arguments which led to (4.7), we find that a necessary condition for (5.1) to be separable is that $g(\psi)$ satisfies the ordinary differential equation

$$\beta^2\psi\{g'(\psi)+\psi g''(\psi)\}=f(g), \quad (5.3)$$

where $\beta^2\in\mathbb{R}$. As before, this equation can be integrated once to give

$$(\psi g')^2=A+(2/\beta^2)\int f\,dg, \quad (5.4)$$

where $A$ is the constant of integration. Putting

$$\pm\sqrt{\{A+(2/\beta^2)\int f\,dg\}}=h(g), \quad (5.5)$$

separating the variables and integrating again, leads to the result

$$\ln(\alpha\psi)=\int dg/h(g), \quad (5.6)$$

where $\alpha$ is a second constant of integration.

In principle, (5.6) can be evaluated and the set of required transformations, $\phi=g(\psi)$, found for any particular $f(\phi)$. This is best illustrated by specific examples and we shall work through two of them after the following theorem.

*Theorem* 5.1: One set of solutions of (5.1) is always

$$\phi(x_1,\ldots,x_n)=g(\exp(a+\sum_{i=1}^{n}\pm m_i^{1/2}x_i)), \quad (5.7)$$

where

$$\sum_{i=1}^{n}\epsilon_im_i=\beta^2 \text{ and } a\in\mathbb{R}. \qquad \square$$

*Proof*: Using the transformation (5.6) on (5.1) leads to the only available separable equation (5.2). Now let $\psi(x_1,\ldots x_n)=\Pi_{i=1}^nX_i(x_i)$ and

$$(X_i')^2 = m_i X_i^2, \tag{5.8}$$

where $i = 1, \ldots n$ and $m_i \in \mathbb{R}$, and substitute into (5.2). This gives

$$\left\{\sum_{i=1}^{n} \epsilon_i m_i X_i \prod_{j\neq i} X_j\right\} g'(\psi)$$

$$+\left\{\sum_{i=1}^{n} \epsilon_i m_i X_i^2 \prod_{j\neq i} X_j^2\right\} g''(\psi) = f(g), \tag{5.9}$$

which can be rewritten as (5.3). But, by definition, $g(\psi)$ satisfies (5.3). Therefore (5.1) separates, via $g$, into the ordinary differential equations (5.8). Solving the set (5.8) gives us the functions

$$X_i = \exp(a_i \pm m_i^{1/2} x_i)\ \forall\, i, \tag{5.10}$$

where the $a_i$ are integration constants, and substitution into $\psi$ immediately leads to the separable solutions (5.7) with $a = \sum_{i=1}^{n} a_i$. This completes the proof of the theorem. ∎

*Example* 1: Consider the equation

$$\phi_{xx} - \phi_{tt} = e^{\phi} \tag{5.11}$$

which is a hyperbolic variant of Liouville's equation.[10] In this case the transformation (5.6) takes the form

$$\pm \ln(\alpha\psi) = \int dg/\{[A + (2/\beta^2)e^g]\}^{1/2}, \tag{5.12}$$

where $\alpha$, $A$ and $\beta$ are, in principle, arbitrary. However, if we want real solutions for $g(\psi)$, then we must choose either $A$ or $\beta^2$ to be negative. Performing the integral in (5.12) (which reduces to a standard form via the substitution $u = e^g$) then leads to the solutions

$$g(\psi) = \ln\left[\frac{|A|\beta^2}{2}\sec^2\left(\frac{|A|^{1/2}}{2}\ln(\alpha\psi)\right)\right],$$

$$\text{for } A < 0,\ \beta^2 > 0. \tag{5.13a}$$

$$g(\psi) = \ln\left[\frac{A|\beta^2|}{2}\operatorname{sech}^2\left(\frac{A^{1/2}}{2}\ln(\alpha\psi)\right)\right],$$

$$\text{for } A > 0,\ \beta^2 < 0. \tag{5.13b}$$

Now, from Theorem 5.1, we see that (5.11) has the following sets of traveling wave solutions:

$$\phi(x,t) = \ln\{\sec^2[a \pm mx \pm (m^2-1)^{1/2}t]\} + \ln 2, \tag{5.14a}$$

$$\phi(x,t) = \ln\{\operatorname{sech}^2[a \pm mx \pm (m^2+1)^{1/2}t]\} + \ln 2, \tag{5.14b}$$

where $a$, $m \in \mathbb{R}$ and where, for convenience, we have chosen $|A| = 4$ and $|\beta^2| = 1$. The waves in (5.14a) are periodic, while those of (5.14b) are solitary. Both solutions, however, are singular.

*Example* 2: Consider the SGE in two space dimensions,

$$\phi_{xx} + \phi_{yy} - \phi_{tt} = \sin\phi. \tag{5.15}$$

The transformations $\phi = g(\psi)$ which lead to separable forms of (5.15) are again those given by Eq. (4.3). Using Theorem 5.1 then gives us the traveling wave solutions

$$\phi(x,y,t) = \pm 2\cos^{-1}\operatorname{sn}[(\gamma/k)(x\cos\theta + y\sin\theta - vt), k], \tag{5.16}$$

where $\gamma$ is the Lorentz factor, $v$ is the wave speed and $\theta$ is the direction that the propagation vector makes with the $x$ axis. The cross sections of these waves, taken parallel to the direction of propagation, are identical to the one-dimensional waves of Eq. (4.16). However, their lengths in the direction perpendicular to motion are infinite. This means that two-dimensional solitons, of the type given by the $k = 1$ limit of (5.16), are not localized entities and thus do not possess some of the desirable properties of the corresponding one-dimensional solutions.[11]

Another interesting point about the two-dimensional SGE is that it does not possess the variety of separable solutions that exist for the one-dimensional equation. Thus, if we let $\psi(x,y,t) = X(x)Y(y)T(t)$ and go through the procedures of Sec. 3 in order to obtain the two-dimensional analogs of Eqs. (2.6), we find that we get only the degenerate cases

$$(X')^2 = mX^2,\ (Y')^2 = nY^2, \text{ and } (T')^2 = pT^2, \tag{5.17}$$

where the constants $m, n$ and $p$ satisfy the relation $m + n - p = \beta^2$. Hence, the solutions given by Theorem 5.1 [of which (5.16) is an example] are the only separable solutions of (5.15). This was not totally unexpected, of course, because of the symmetry between $x$ and $y$ in (5.15). However, it is unfortunate that this symmetry is not reflected in the soliton in the sense that the cross sections of the latter, along and perpendicular to its motion, are so different. On the other hand, the analysis leading to (5.17) does show that if there is a localized soliton solution of the two-dimensional SGE, then it is certainly not separable in terms of our definitions. As to whether there is any connection between dimensionality, localization and separability in the case of soliton solutions of the SGE is a question that might be worth investigating.

## 6. CONCLUDING REMARKS

We have defined and studied, in some detail, the separability and the existence of separable solutions of the SGE and similar quasilinear partial differential equations. The method we have used is to first make a dependent-variable transformation which reduces the original equation to a separable form, and then assume power series expansions for the derivatives of the separating functions, in terms of the functions themselves, in order to further reduce the separable form to an identity; the latter operation being somewhat reminiscent of the Frobenius method for ordinary differential equations. A consistent solution of the recurrence relations of this identity then gives us the separated ordinary differential equations corresponding to the original partial differential equation. In this manner we have demonstrated the existence of a general set of dependent-variable transformations which lead to separable forms of the SGE and have used them to discuss the existence and classification of separable solutions of the latter. Similar results have been obtained for some other quasilinear equations.

Our analysis, however, is still in its early stages. Thus, we have not discussed the "difficult" cases, nor have we considered the important questions of initial and boundary conditions. Now, it is well-known that boundary conditions are intimately connected with the separable solutions of linear partial differential equations.[9] Thus, it is almost certain that boundary conditions will play a prominent role in the separability problems of the nonlinear theory. On the other hand, in linear theory we have a principle of superposition, and the corresponding freedom to use Fourier series expansion enables us to fit a great variety of initial conditions either exactly or approximately. In the nonlinear theory we have no superposition principle and so the problem is that much harder.

As far as the separable solutions themeselves are concerned, they do, of course, form only a subset of the set of all solutions. Nevertheless, they do give some insight into the structure of the differential equation and into the structure and formation of soliton solutions. For example, in the case of the one-dimensional SGE, both the one-soliton and two-soliton solutions are separable, but the three-soliton solution is not. Whereas, for the two-dimensional SGE, only the single-soliton solution is separable. Furthermore, the dependent-variable transformation used used to obtain separable forms of the SGE are often the ones used to express more general solutions.[2]


[1]A. Barone, F. Esposito, C.J. Magee, and A.C. Scott, Riv. Nuovo Cimento **1**, 227 (1971).
[2]A.C. Scott, F. Chu, and D. McLaughlin, Proc. IEEE **61**, 1443 (1973).
[3]R. Rajaraman, Phys. Rep. C **21**, 227 (1975).
[4]M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Phys. Rev. Lett. **31**, 125 (1973).
[5]G.L. Lamb, Rev. Mod. Phys. **43**, 99 (1971).
[6]J. Zagrodzinski, Phys. Lett. A **57**, 213 (1976).
[7]H. Hancock, *Theory of Elliptic Functions* (Dover, New York, 1958).
[8]A.C. Bryan, C.R. Haines, and A.E.G. Stuart, "A classification of the separable solutions of the two-dimensional sine-Gordon equation and of its Laplacian variant" (preprint).
[9]H.F. Weinberger, *A First Course in Partial Differential Equations* (Xerox, Massachusetts, 1965), p. 68.
[10]H.T. Davis, *Introduction to Nonlinear Differential and Integral Equations* (Dover, New York, 1962), pp. 20 and 377.
[11]J.D. Gibbon and G. Zambotti, Nuovo Cimento B **28**, 1 (1975).

# Existence of instantaneous Cauchy surfaces[a)]

Robert Budic, Mark J. Gotay, and Robert H. Gowdy[b)]

*Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742*

(Received 12 May 1977; revised manuscript received 8 February 1978)

Several properties of instantaneous Cauchy surfaces are obtained. It is shown that a strongly causal spacetime admits an instantaneous Cauchy surface through each of its points, that there is a close and reversible relationship between these surfaces and maximal open globally hyperbolic subsets, that every instantaneous Cauchy surface is contained in a maximal instantaneous Cauchy surface, and that the latter surface is a maximal achronal surface which separates spacetime into past, present, and future. Some other properties of instantaneous Cauchy surfaces are discussed along with a refinement of an earlier topology change property.

## I. INTRODUCTION

The preceding paper defined an instantaneous Cauchy surface to be an achronal set whose interior Cauchy development is maximal on the family of all such sets.[1] Several examples were considered, and it was argued that instantaneous Cauchy surfaces may have an important role to play in analyzing the structure of singular spacetimes and quantizing fields on such spacetimes.

Since one can construct spacetimes in which there are no nonempty achronal sets whatsoever, some restriction on the causal structure of spacetime is necessary if the spacetime is to admit the existence of an instantaneous Cauchy surface. The main result of this paper is an existence theorem which shows that instantaneous Cauchy surfaces can be found in all but the most pathological spacetimes.

In Sec. II we establish our notation and collect some well-known facts about globally hyperbolic sets. The main existence theorem is stated and proved in Sec. III. Section IV considers the properties of maximal and minimal instantaneous Cauchy surfaces. It is found, for example, that any instantaneous Cauchy surface is contained in a maximal instantaneous Cauchy surface which is also a maximal achronal set. Section V shows that a maximal instantaneous Cauchy surface is edgeless and that whenever two such surfaces have the same interior Cauchy development, they are homeomorphic.

## II. NOTATION AND USEFUL FACTS

The notation of this paper is chosen to be compatible with the monograph by Hawking and Ellis[2] and also with the monograph *Techniques of Differential Topology* by Penrose.[3] One slight departure from the Hawking and Ellis notation is that we only require Cauchy surfaces and partial Cauchy surfaces to be achronal instead of acausal. Domains of dependence, determined by timelike curves, are denoted by $\tilde{D}(S)$ as in Hawking and Ellis. The set that we almost always use is the interior Cauchy development $\mathrm{int}\tilde{D}(S)$ so that it is convenient to denote this set by $D^0(S)$.

Global hyperbolicity can be defined in several ways. We use the definition in Hawking and Ellis[2]:

*Definition*: A set $N$ is said to be *globally hyperbolic* if the strong causality assumption holds on $N$ and if, for any two points $p, q \in N$, $J^+(p) \cap J^-(q)$ is compact and contained in $N$.

With this definition, the global hyperbolicity of $N$ is tied to the causal structure of the whole spacetime $M$. In particular, we will use the following obvious consequence of this definition:

(GH0) If $N$ is a globally hyperbolic subset of a spacetime $M$ and $S$ is achronal relative to $N$, then $S$ is achronal relative to $M$.

We list below some additional properties of Cauchy developments and globally hyperbolic sets which will be used in what follows. We assume that $M$ is a time-orientable spacetime.

(GH1) For every achronal set $S$, $D^0(S)$ is globally hyperbolic.[4]

(GH2) Each open globally hyperbolic set $H$, considered as a spacetime, contains a Cauchy surface $S$.[5] In $H$, $H = D^0(S)$. In $M$, $H \subseteq D^0(S)$.

(GH3) Each open globally hyperbolic set can be foliated by the achronal sets of (GH2).[6]

(GH4) If $H$ is open and globally hyperbolic, and $S$ and $S'$ are as in (GH2), then $S$ and $S'$ are homeomorphic.[7]

## III. EXISTENCE

*Definition*: An achronal set $S$ is an *instantaneous Cauchy surface* if and only if, for any achronal set $S'$, $D^0(S) \subseteq D^0(S')$ implies $D^0(S) = D^0(S')$.

*Theorem* 1: If $M$ is a strongly causal spacetime and $p$ is a point of $M$, then there is an instantaneous Cauchy surface that includes $p$.

*Proof*: Let $G_p$ be the family of all open globally hyperbolic subsets of $M$ containing $p$. By strong causality, $p$ has an open globally hyperbolic neighborhood so that $G_p$ is not empty. Partially order $G_p$ by set inclusion and let $\{G_\alpha \mid \alpha \in A\}$ be a totally ordered subfamily of $G_p$. The set $\cup_\alpha \{G_\alpha\}$ is open and globally hyperbolic and is an upper bound for the subfamily $\{G_\alpha\}$. By Zorn's lemma there exists a maximal member $H$ of the family

[a)]Supported in part by the National Science Foundation Grant No. MPS 74-18386-A01.
[b)]Supported in part by the Alfred P. Sloan Foundation.

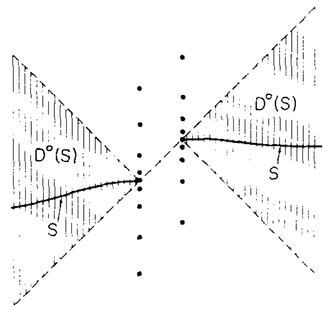



FIG. 1. A nonmaximal instantaneous Cauchy surface. The example consists of two-dimensional Minkowski spacetime with two sequences of points removed so that their limit points (also removed) are null related. The surface $S$ is an instantaneous Cauchy surface because the interior Cauchy development of any other achronal surface will necessarily encounter one of the removed points before it can include the development of $S$. Larger instantaneous Cauchy surfaces can be obtained by adjoining to $S$ points on the null line (dotted) that connects the limit points. These additional points do not contribute to the interior of the Cauchy development at all. Notice that the resulting larger instantaneous Cauchy surfaces may not be submanifolds because they can include isolated points.

$G_p$. Now $p \in H$, so by (GH3) and (GH0) we can find an achronal set $S$ through $p$ such that $H \subseteq D^0(S)$. Since $H$ is maximal $D^0(S) = H$ and $S$ is the desired instantaneous Cauchy surface through the point $p$.

*Remark*: A slight weakening of both the hypothesis and the conclusion of the existence theorem is possible. In a past(future)-distinguishing spacetime, one can show that an instantaneous Cauchy surface passes through each neighborhood of each point.[8] The method of proof is the same as above except for one point: It must be established that each neighborhood is intersected by an open globally hyperbolic set. This point can be established by a straightforward local construction.

In the process of proving the existence theorem, a connection between instantaneous Cauchy surfaces and maximal open globally hyperbolic sets has been intimated. In fact, for some purposes, one may be more interested in the maximal open globally hyperbolic subsets of a spacetime than in the instantaneous Cauchy surfaces. For this reason, it is useful to state the exact nature of this connection.

*Proposition* 1: An achronal set $S$ is an instantaneous Cauchy surface if and only if $D^0(S)$ is a maximal open globally hyperbolic set.

*Proof*: Suppose that $H$ is an open globally hyperbolic set such that $D^0(S) \subseteq H$. By (GH0) and (GH2) one can find some achronal set $S'$ with $D^0(S) \subseteq H \subseteq D^0(S')$. If $S$ is an instantaneous Cauchy surface, then $D^0(S) = H$ and $D^0(S)$ is a maximal open globally hyperbolic set.

From this result and (GH3), we see that the existence of instantaneous Cauchy surfaces is equivalent to the existence of maximal open globally hyperbolic subsets. We now restate the existence theorem in terms of these subsets.

*Proposition* 2: A strongly causal spacetime is covered by its maximal open globally hyperbolic subsets.

## IV. MAXIMAL AND MINIMAL INSTANTANEOUS CAUCHY SURFACES

If one is really interested in the "best possible achronal sets" in a spacetime, then the instantaneous Cauchy surfaces are not the last word. It is possible for one instantaneous Cauchy surface to be a proper subset of another. Figure 1 shows a simple example of this behavior. The interesting achronal sets, we find, are the maximal, $S_{\max}$, and the minimal, $S_{\min}$, instantaneous Cauchy surfaces. These sets are easily constructed from a given instantaneous Cauchy surface.

*Proposition* 3: If $M$ is a strongly causal spacetime and $S$ is an instantaneous Cauchy surface in $M$, then

(A) $S_{\min} = \cap_\alpha S_\alpha$, where $\{S_\alpha\}$ is the set of instantaneous Cauchy surfaces contained in $S$.

(B) $S_{\max} = {\sim} I(S)$, where $I(S) := I^+(S) \cup I^-(S)$.

*Proof*: (A) First we show that $\cap_\alpha S_\alpha$ is an instantaneous Cauchy surface. For each $\alpha$, $S_\alpha \subseteq S$ and so $D^0(S_\alpha) \subseteq D^0(S)$. Since each $S_\alpha$ and $S$ is an instantaneous Cauchy surface, $D^0(S_\alpha)$ and $D^0(S)$ are maximal globally hyperbolic sets by Proposition 1. Thus $D^0(S_\alpha) = D^0(S)$ for each $\alpha$. Consequently, if $p \in D^0(S)$, every timelike curve through $p$ must intersect $S_\alpha$ for each $\alpha$ and so every timelike curve through $p$ must intersect $\cap_\alpha S_\alpha$. It follows then that $D^0(S) \subseteq D^0(\cap_\alpha S_\alpha)$ and so from the maximality of $D^0(S)$, $D^0(S) = D^0(\cap_\alpha S_\alpha)$. Thus $D^0(\cap_\alpha S_\alpha)$ is maximal, and since $\cap_\alpha S_\alpha \subseteq S$, the set $\cap_\alpha S_\alpha$ is achronal. A final application of Proposition 1 completes the proof that $\cap_\alpha S_\alpha$ is an instantaneous Cauchy surface. Clearly it is the smallest contained in $S$.

(B) Since $S$ is achronal, $S \subseteq {\sim} I(S)$. In fact, one finds that ${\sim} I(S)$ is itself an achronal set. To show this, suppose that ${\sim} I(S)$ is not achronal and choose two points, $p \ll q$ in ${\sim} I(S)$. By strong causality, there exists some open globally hyperbolic subset $N \subseteq I^+(p) \cap I^-(q) \subseteq {\sim} I(S)$. By (GH0) and (GH2), there exists some achronal set $\Delta S \subseteq N$ such that $N \subseteq D^0(\Delta S)$. Since $\Delta S \subseteq {\sim} I(S)$, the set $S \cup \Delta S$ is achronal. Moreover, $D^0(S \cup \Delta S)$ contains $N$ and so is strictly larger than $D^0(S)$. This conclusion leads to a contradiction since, by Proposition 1, $D^0(S)$ is maximal. Thus ${\sim} I(S)$ is an achronal set containing $S$ and so is an instantaneous Cauchy surface. It is not difficult to see that ${\sim} I(S)$ is the largest achronal set containing $S$ (and hence the largest instantaneous Cauchy surface containing $S$); for any extension of ${\sim} I(S)$ would contain points in $I(S)$ and would not be achronal.

Some further properties of $S_{\max}$ and $S_{\min}$ are needed later. It is useful to state these properties explicitly.

*Proposition* 4: If $S$ is an instantaneous Cauchy surface in a strongly causal spacetime $M$ and $p \in S_{\min}$, then every timelike curve through $p$ intersects $D^0(S_{\min})$.

*Proof*: First note that for every $p \in S_{\min}$ either $I^+(p) \cap D^0(S) \neq \emptyset$ or $I^-(p) \cap D^0(S) \neq \emptyset$. Otherwise one has $D^0(S) = D^0(S_{\min}) = D^0(S_{\min} - \{p\})$ and $S_{\min} - \{p\}$ is an instantaneous Cauchy surface, contradicting the minimality of $S_{\min}$.

If $I^+(p) \cap D^0(S) \neq \emptyset$, then for any $q \in I^+(p) \cap D^0(S)$, $I^-(q) \cap I^+(p) \subseteq D^0(S)$. Any timelike curve through $p$ must intersect $I^-(q) \cap I^+(p)$ and so must intersect $D^0(S)$. Finally, if $I^-(p) \cap D^0(S) \neq \emptyset$, dual arguments complete the proof.

*Proposition* 5: $I(S_{\min}) \cup S_{\min} = I^+(D^0(S)) \cup I^-(D^0(S))$.

*Proof*: If $p \in I(S_{\min}) \cup S_{\min}$, then, for some timelike curve $\gamma$ through $p$, $\gamma \cap S_{\min} \neq \emptyset$. By the previous proposition, $\gamma \cap D^0(S) \neq \phi$. Thus, either $p \in I^+(D^0(S))$ or $p \in I^-(D^0(S))$, and consequently $I(S_{\min}) \cup S_{\min} \subseteq I^+(D^0(S)) \cup I^-(D^0(S))$.

To show the reverse inclusion, suppose $p \in I^+(D^0(S))$. Then for some $q \in D^0(S)$ and some timelike curve $\gamma$, $p$ and $q$ are respectively the future and past end points of $\gamma$. Let $\gamma'$ be any inextendible timelike curve containing $\gamma$. Since $\gamma'$ contains $q \in D^0(S_{\min}) = D^0(S)$, the set $\gamma' \cap S_{\min}$ must be nonempty. Since $p$ lies on a timelike curve that intersects $S_{\min}$, we have $p \in I(S_{\min}) \cup S_{\min}$. A similar argument can be made for $p \in I^-(D^0(S))$ so that $I^+(D^0(S)) \cup I^-(D^0(S)) \subseteq I(S_{\min}) \cup S_{\min}$.

An immediate consequence of Propositions 3 and 5 is

$$S_{\max} - S_{\min} = \sim [I(S_{\min}) \cup S_{\min}] = \sim [I^+(D^0(S)) \cup I^-(D^0(S))]$$

for any instantaneous Cauchy surface $S$. It is now easy to see that if $S$ and $S'$ are equivalent in the sense of having the same interior Cauchy development, then $S_{\max} - S_{\min} = S'_{\max} - S'_{\min}$. If $S$ and $S'$ are equivalent instantaneous Cauchy surfaces, then $S_{\max}$ and $S'_{\max}$ are homeomorphic if and only if $S_{\min}$ and $S'_{\min}$ are homeomorphic. We turn to this question in the next section.

## V. REFINEMENT OF THE TOPOLOGY CHANGE PROPERTY

Consider the topology change property (GH4) of Cauchy surfaces. In order to apply this result to instantaneous Cauchy surfaces in a straightforward way, the previous paper required one of the surfaces to be acausal so that it would lie in its interior Cauchy development. Here we show how this restriction can be removed.

*Theorem* 2: If $S$ and $S'$ are instantaneous Cauchy surfaces in a strongly causal spacetime and $D^0(S) = D^0(S')$, then $S_{\min}$ is homeomorphic to $S'_{\min}$ and $S_{\max}$ is homeomorphic to $S'_{\max}$.

*Proof*: From the comments following Proposition 5, it is sufficient to prove that $S_{\min}$ is homeomorphic to $S'_{\min}$. The property (GH4) cannot be applied directly because of the possibility that $S_{\min}$ and $S'_{\min}$ do not lie entirely within $D^0(S)$. As in the proofs[5-7] of properties (GH2)—(GH4), let $\gamma$ be a congruence of inextendible timelike curves and use this congruence to produce a map $f: S_{\min} \to S'_{\min}$.

First, show that the map $f$ is defined on $S$ and is one-to-one and onto. For convenience define $K := D^0(S)$. It follows from Proposition 4 that any timelike curve $\gamma_p \in \gamma$ through $p \in S_{\min}$ enters $K$ and so intersects $S'_{\min}$ at a point $f(p)$. Thus, $\gamma$ defines a one-to-one $f: S_{\min} \to S'_{\min}$. Similarly, $\gamma$ defines a one-to-one map $g: S'_{\min} \to S_{\min}$ such that $g = f^{-1}$ so that $f$ is onto.

Next, show that $f$ is continuous by expressing it locally as a composition of continuous maps. Choose any $p \in S_{\min}$ and choose neighborhoods $U_p$ of $p$ and $U_q$ of $q := f(p)$ with compact closures. Choose a point $\tilde{p} \in \gamma_p \cap U_p$ and a point $\tilde{q} \in \gamma_p \cap U_q$. From (GH3) we can pick achronal sets $\tilde{S}$ and $\tilde{S}'$ in $K$ such that $\tilde{p} \in \tilde{S}$, $\tilde{q} \in \tilde{S}'$, and $D^0(\tilde{S}) = D^0(\tilde{S}') = K$. The congruence $\gamma$ defines the one-to-one and onto maps $f_{\min}: S_{\min} \to \tilde{S}$, $\tilde{f}: \tilde{S} \to \tilde{S}'$, and $f'_{\min}: \tilde{S}' \to S'_{\min}$ so that $f = f'_{\min} \circ \tilde{f} \circ f_{\min}$. The function $f_{\min}$ will be continuous at $p$ if the sequence $\{\tilde{q}_n \mid \tilde{q}_n = f_{\min}(p_n)\}$ converges to $\tilde{q}$ whenever the sequence $\{p_n\}$ in $S_{\min}$ converges to $p$. If any subsequence of $\{\tilde{q}_n\}$ converges to a point $q' \in U_p$ but $q' \neq \tilde{q}$, then $q' \in \gamma$ and either $q' \in I^+(\tilde{q})$ or $q' \in I^-(\tilde{q})$. Suppose $q' \in I^+(\tilde{q})$. But then $I^+(\tilde{q})$ is a neighborhood of $q'$ and must contain $\tilde{q}_n$ for sufficiently large $n$. One then has $\tilde{q}_n \in I^+(\tilde{q})$ which contradicts the achronality of $\tilde{S}$. The same argument holds if $q' \in I^-(\tilde{q})$. If $\{\tilde{q}_n\}$ contains a subsequence $\{q_{n'}\} \subseteq M - U_p$, then the timelike curves of $\gamma$ which join $p_{n'}$ and $\tilde{q}_{n'}$ must intersect $\dot{U}$. As $\overline{U}$ is compact, the boundary $\dot{U}$ is also compact and this sequence of intersections will have a cluster point $q'$. One can then apply the previous argument to contradict the achronality of $\tilde{S}$. Thus $f_{\min}$ is continuous. This same argument can also be used to show that $f'_{\min}$ is continuous. Since (GH4) implies the continuity of $\tilde{f}$, we have established that $f$ is continuous.

A time reversal of the preceding argument shows that $f^{-1}$ is also continuous so that $f$ is continuous and open and therefore a homeomorphism.

## VI. DISCUSSION

Proposition 1 establishes a close and reversible connection between instantaneous Cauchy surfaces and maximal open globally hyperbolic sets. Theorem 1 shows that instantaneous Cauchy surfaces are plentiful in most of the spacetimes that one would wish to consider. Proposition 3 connects these instantaneous Cauchy surfaces to minimal and maximal instantaneous Cauchy surfaces.

For most purposes, it is the minimal and maximal instantaneous Cauchy surfaces that are of interest. Propositions 3–5 spell out a variety of properties of these surfaces. From part (B) of Proposition 3, a maximal instantaneous Cauchy surface can be thought of as a "global instant of time" because it divides spacetime into past, present, and future (compact partial Cauchy surfaces in a causally continuous spacetime share this same property[9]). Such a surface is an achronal boundary and is therefore a closed, edgeless, imbedded $C^{1-}$ three-dimensional submanifold of spacetime—a partial Cauchy surface. Proposition 4 shows that a minimal instantaneous Cauchy surface has an important global property in common with a spacelike hypersurface. However, it should be noticed that $S_{\min}$ is not, in general, an acausal or spacelike surface and can have null generators. Proposition 5 and the comments that follow it can be used to deduce the properties of $S_{\max} - S_{\min}$. The property that has been used in this paper is the fact that this set depends only on the maximal open globally hyperbolic set $D^0(S)$ and not on the particular instantaneous Cauchy surface $S$ that generates it. It is also quite easy to show that $S_{\max} - S_{\min}$ is generated by a congruence of null geodesics.

Theorem 2 requires considerable work to extend the property (GH4), but this work is necessary because one often prefers to give data on null hypersurfaces which need not be contained in their own Cauchy developments. As in the previous paper,[1] the most natural way to interpret Theorem 2 is to give various negative statements of it. Thus, we find that topology changes in maximal instantaneous Cauchy surfaces cannot occur through the regular evolution of hyperbolic field equations. If two such surfaces are not homeomorphic, then they must have distinct Cauchy developments and the spacetime cannot be globally hyperbolic.


[1] R.H. Gowdy, J. Math. Phys. **18**, 1798–1801 (1977).
[2] S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U.P., Cambridge, 1973).
[3] R. Penrose, *Technique of Differential Topology in Relativity* (SIAM, Philadelphia, 1972).
[4] Reference 3, p. 48, Propositions 5.22, 5.23.
[5] Reference 3, Theorem 5.25.
[6] Reference 3, Theorem 5.26.
[7] This result follows directly from the theorem cited above.
[8] We are using the definition of past-distinguishing which is given in Hawking and Ellis, Ref. 2, p. 192.
[9] R. Budic and R.K. Sachs, J. Math. Phys. **15**, 1302–9 (1974).

# Discrete finite nilpotent Lie analogs: New models for unified gauge field theory

Karl Kornacker

*Department of Biophysics, The Ohio State University, Columbus, Ohio 43210*
(Received 2 September 1977)

To each finite dimensional real Lie algebra with integer structure constants there corresponds a countable family of discrete finite nilpotent Lie analogs. Each finite Lie analog maps exponentially onto a finite unipotent group $G$, and is isomorphic to the Lie algebra of $G$. Reformulation of quantum field theory in discrete finite form, utilizing nilpotent Lie analogs, should elminate all divergence problems even though some non-Abelian gauge symmetry may not be spontaneously broken. Preliminary results in the new finite representation theory indicate that a natural hierarchy of spontaneously broken symmetries can arise from a single unbroken non-Abelian gauge symmetry, and suggest the possibility of a new unified group theoretic interpretation for hadron colors and flavors.

## I. INTRODUCTION

Clearly: (a) no experiment can determine whether physical space—time is actually discrete or continuous, finite or infinite; (b) if physical space—time is actually discrete, then the use of continuum models may unnecessarily complicate the theory of extremely microscopic phenomena such as quark confinement; (c) if physical space—time is actually finite, then the use of unbounded position operators may unnecessarily complicate the quantum dynamical theory of extremely macroscopic phenomena such as cosmic evolution, and may cause spurious infrared divergence problems; (d) systematic theoretical investigation of (b) and (c) requires a generalization of quantum field theory which is meaningful for some set of discrete finite space—time models. This paper presents a new mathematical approach to problem (d).

Three difficulties hinder any attempt to reformulate quantum field theory in discrete finite form: (1) quantitative measurements are commonly represented by elements of an infinite number field; (2) no finite number field is algebraically closed; (3) there is no known discrete finite analog of the exponential map from simple real Lie algebras onto analytic groups.[1] The first two difficulties are readily circumvented. The third, however, prevents the construction of discrete finite gauge field theories. For this reason prior studies of discrete finite space—time models[2] have contributed little toward the generalization of quantum field theory. In this paper we solve problem (3) and briefly consider some of the unique mathematical advantages which may be obtained by reformulating quantum field theories in discrete finite form. Specific models utilizing the new finite formalism will be discussed elsewhere.

## II. DISCRETE FINITE NILPOTENT LIE ANALOGS

Some authors have proposed that the finite Galois fields $GF(p^2)$ of prime characteristic $p\equiv 3 \pmod 4$ should be regarded as discrete finite analogs of the complex plane.[2] We cannot accept such a proposal for the following reason:

*Proposition I*: Let $K$ be a finite field of prime characteristic $p$. Let $\mathcal{E}: K\to K$ satisfy $\mathcal{E}(0)=1$. Then $\mathcal{E}(nx) = (\mathcal{E}(x))^n$ for all positive integers $n$ only if $\mathcal{E}(x)=1$.

*Corollary*: It is impossible to construct a nontrivial group homomorphism from an additive subgroup of $K$ onto a multiplicative subgroup of $K$.

*Proof*: It is known that $x^{p^r}\equiv x$ for all $x\in K$ and some positive integer $r$. Therefore the given conditions on $\mathcal{E}$ imply $1=\mathcal{E}(0)=\mathcal{E}(p^r x)=(\mathcal{E}(x))^{p^r}=\mathcal{E}(x)$.

Throughout the remainder of this paper we let $R$ denote a finite associative ring with unity $1_R$ and arbitrary fixed prime characteristic $p\geq 3$; let $C(R)$ denote the center of $R$; let $Z_p$ denote the subring (prime field) generated by $1_R$; let $R[x]$ denote the ring of polynomials in a single algebraically independent indeterminate over $R$; let $\mathcal{E}: R[x]\to R[x]$ denote the truncated exponential map defined by $\mathcal{E}(\varphi)=1_R+\sum_{s=1}^{p-1}\varphi^s/s!$; let $d/dx$: $R[x]\to R[x]$ denote the $R$-linear derivation defined by $d/dx(x)=1_R$; let $H: R[x]\to R[\xi]: x\to\xi$ denote a fixed unitary $R$-linear homomorphism; let $D: R[\xi]\to R[\xi]$ and exp: $R[\xi]\to R[\xi]$ denote maps defined by the commutative diagrams

$$\begin{array}{ccccccc} R[x] & \xrightarrow{d/dx} & R[x] & \quad & R[x] & \xrightarrow{\mathcal{E}} & R[x] \\ \downarrow{\scriptstyle H} & & \downarrow{\scriptstyle H} & & \downarrow{\scriptstyle H} & & \downarrow{\scriptstyle H} \\ R[\xi] & \xrightarrow{D} & R[\xi] & & R[\xi] & \xrightarrow{\exp} & R[\xi]. \end{array}$$

We say that $D$ is *faithful* if and only if the restricted map $D: \xi Z_p[\xi]\to Z_p[\xi]$ is injective. We say that the element $\eta\in C(R)$ is an *admissible value* of $\xi$ if and only if $(\xi-\eta)R[\xi]$ is a proper ideal in $R[\xi]$. If $\eta$ is an admissible value of $\xi$ in $C(R)$ then we let $\langle\ \rangle_\eta: R[\xi]\to R: \xi\to\eta$ denote the $R$-linear homomorphism with kernel $(\xi-\eta)R[\xi]$, and let $D_\eta$ denote the tangent vector at $\eta$ defined as the composite map $\langle\ \rangle_\eta\circ D$.

Proposition II uniquely characterizes the unitary $R$-linear homomorphism $H: R[x]\to R[\xi]$ which is most suitable for use in discrete finite quantum field theory.

*Proposition II*: $D$ is faithful if and only if the kernel of $H$ is $x^p R[x]$.

*Proof* (Sketch): If $D$ is faithful then it follows that


0022-2488/78/1907-1584$1.00

$\xi^{p-1} \notin Z_p$ and $\xi^p = 0$; therefore, the kernel of $\mathcal{H}$ is $x^p R[x]$. The converse follows by direct computation.

Proposition III and its corollary support the choice of $\mathcal{H}$ according to Proposition II.

*Proposition III*: If the kernel of $\mathcal{H}$ is $x^p R[x]$ then the following statements are equivalent for $\eta \in C(R)$:

1. $\eta$ is an admissible value of $\xi$;
2. $\mathcal{D}(\exp(\eta\xi)) = \eta \exp(\eta\xi)$;
3. $\eta^p = 0$.

*Corollary*: If the kernel of $\mathcal{H}$ is $x^p R[x]$ and $\eta \in C(R)$ is an admissible value of $\xi$ then the restricted map $\exp\colon \eta Z_p[\eta] \to Z_p[\eta]$ defines a group isomorphism *from* an additive group of admissible values of $\xi$ *onto* a multiplicative group of automorphisms on $\eta Z_p[\eta]$.

*Proof* (Sketch): $(1) \Rightarrow \eta^p = \langle \xi^p \rangle_\eta \Rightarrow (3)$; $(3) \Rightarrow (\xi - \eta)^p = 0 \Rightarrow (1)$; $(3) \Rightarrow (2)$ by direct computation; $(2) \Rightarrow \eta^p \xi^{p-1} = 0 \Rightarrow \mathcal{D}^{p-1}(\eta^p \xi^{p-1}) = 0 \Rightarrow (3)$. The conditions given in the corollary imply $\eta^p = 0$ (by Proposition III). If $2 \leq r \leq p$ is the lowest power of $\phi \in \eta Z_p[\eta]$ such that $\phi^r = 0$ then $\phi^{r-1} = (\exp(\phi) - 1_R)\phi^{r-2}$. It follows that $\exp(\phi) = 1_R$ if and only if $\phi = 0$. The rest of the corollary follows by direct computation.

Proposition IV uniquely characterizes the involutive automorphism $\mathcal{C}\colon Z_p[\eta] \to Z_p[\eta]$.

*Proposition IV*: There exists a unique involutive automorphism $\mathcal{C}\colon Z_p[\eta] \to Z_p[\eta]$ and $\mathcal{C}(\eta) = -\eta$.

*Proof* (Sketch): From $\eta^p = 0$ it follows that every automorphism on $Z_p[\eta]$ is defined by $\eta \to a\eta$ for some nonzero $a \in Z_p$. For an involutive automorphism $a^2 = 1_R \neq a$ and therefore $a = -1_R$.

Throughout the remainder of this paper we adopt the following conventions and notation: $\mathcal{H}$ is fixed as the unique $R$-linear homomorphism on $R[x]$ with kernel $x^p R[x]$; $\eta_r \in C(R)$ satisfies $\eta_r^r = 0$ and $\eta_r^{r-1} \neq 0$ for $3 \leq r \leq p$.

Now consider the countable dense ring of complex numbers $Z[ip^{1/p}]$ where $Z$ denotes the ring of rational integers. Clearly there is a natural homomorphism: $Z[ip^{1/p}] \to Z_p[\eta_r]$ defined by the commutative diagram

$$\begin{tikzcd} Z[ip^{1/p}] \arrow[r, "(\mathrm{mod}\ p)"] \arrow[dr] & Z_p[\eta_p] \arrow[d] \\ & Z_p[\eta_r] \end{tikzcd} \quad (\mathrm{mod}\,\eta_p^r).$$

Furthermore the involutive automorphism $\mathcal{C}_r$ acting on $Z_p[\eta_r]$ satisfies the commutative diagram

$$\begin{tikzcd} Z[ip^{1/p}] \arrow[r, "*"] \arrow[d] & Z[ip^{1/p}] \arrow[d] \\ Z_p[\eta_r] \arrow[r, "\mathcal{C}_r"] & Z_p[\eta_r], \end{tikzcd}$$

where $*$ denotes ordinary complex conjugation. Therefore, we call $\eta_r$ the $(p,r)$ *image* of $ip^{1/p}$, and identify $\mathcal{C}_r$ as the $(p,r)$ *analog of complex conjugation.*

In view of these observations we identify $\eta_r^2 Z_p[\eta_r^2]$, the set of all "real" (invariant under $\mathcal{C}_r$) admissible values of $\xi$ in $Z_p[\eta_r]$, as the $(p,r)$ *analog of the real axis*. The $(p,r)$ *analog of the complex plane* is then the $\eta_r^2 Z_p[\eta_r^2]$ module generated by $\{1, \eta_r\}$, or equivalently $\eta_r^2 Z_p[\eta_r]$. It follows that $\eta_r^3 Z_p[\eta_r^2]$ is the $(p,r)$ *analog of the imaginary axis*.

Every element of the multiplicative unitary group $\exp(\eta_r^3 Z_p[\eta_r^2])$ can be written as $\langle \exp(\eta_r \xi) \rangle_\phi$ for some $\phi \in \eta_r^2 Z_p[\eta_r^2]$. Also $\eta_r = \mathcal{D}_0(\exp(\eta_r \xi))$, so we may regard $\eta_r$ as the infinitesimal generator of $\exp(\eta_r^3 Z_p[\eta_r^2])$. It follows that the Lie algebra of $\exp(\eta_r^3 Z_p[\eta_r^2])$ is isomorphic to $\eta_r^3 Z_p[\eta_r^2]$. Therefore we regard $\exp(\eta_r^3 Z_p[\eta_r^2])$ as the $(p,r)$ *analog of* $U(1)$.

The following definitions together with proposition $V$ generalize the preceding constructions to include $(p,r)$ analogs of noncommutative real Lie algebras.

*Definition*: Let $\hat{L}$ be an abstract real $m$-dimensional Lie algebra with basis $\{\lambda_k\}_1^m$ chosen so that all structure constants of $\hat{L}$ are integers. Such a basis for $\hat{L}$ always exists if $\hat{L}$ is compact, but may also exist when $\hat{L}$ is noncompact. Let $L$ denote the Lie ring generated by $\{\lambda_k\}_1^m$. Let $L[\eta_p]$ denote the tensor product $Z_p[\eta_p] \otimes L$ with Lie product defined by $[\phi \otimes \lambda, \phi' \otimes \lambda'] = \phi\phi' \otimes [\lambda, \lambda']$. We call the finite nilpotent Lie algebra $\eta_p^2 L[\eta_p^2] = \eta_p^2 Z_p[\eta_p^2] \otimes L$ the abstract $p$ *analog of* $\hat{L}$.

*Definition*: Let $M_s[\eta_r]$ denote the semilocal ring of matrices of degree $s$ over the local[3] commutative ring $Z_p[\eta_r]$. Let $\rho_s\colon L[\eta_p] \to M_s[\eta_p]$ be a $Z_p[\eta_p]$-linear representation such that $\rho_s(1_R \otimes [\lambda, \lambda']) = \rho_s(1_R \otimes \lambda)\rho_s(1_R \otimes \lambda') - \rho_s(1_R \otimes \lambda')\rho_s(1_R \otimes \lambda)$ for all $\lambda, \lambda' \in L$. Let $\rho_s^{(r)}\colon L[\eta_p] \to M_s[\eta_r]$ denote the representation defined by the commutative diagram

$$\begin{tikzcd} L[\eta_p] \arrow[r, "\rho_s"] \arrow[dr, "\rho_s^{(r)}"'] & M_s[\eta_p] \arrow[d] \\ & M_s[\eta_r] \end{tikzcd} \quad (\mathrm{mod}\,\eta_p^r).$$

Then we call $\rho_s^{(r)}(\eta_p^2 L[\eta_p^2])$ a $(p,r)$ *analog of* $\hat{L}$. Note that the value of the *cutoff parameter* $r$ is determined by the choice of representation.

*Proposition V*: Let $R = M_s[\eta_r]$. Then the image of $\exp \circ \rho_s^{(r)}\colon \eta_p^2 L[\eta_p^2] \to R$ is a finite unipotent group whose Lie algebra is isomorphic to $\rho_s^{(r)}(\eta_p^2 L[\eta_p^2])$.

*Proof* (Sketch): The Baker–Campbell–Hausdorff theorem[4] assures that the specified image is a unipotent group. The rest of the proposition is proved by paraphrasing standard arguments.[5]

The remainder of this section deals with representation theory. We identify $M_s[\eta_r]$ with the ring of endomorphisms on a *free* $Z_p[\eta_r]$ module $F(r,s)$ whose basis contains $s$ elements. We say that a subset $T$ of $M_s[\eta_r]$ is reducible over $F(r,s)$ if and only if $F(r,s)$ contains a *free* proper submodule $F(r,s')$, $1 \leq s' < s$, which is invariant under every element of $T$; otherwise we say that $T$ is irreducible over $F(r,s)$. Clearly the reducibility of $T$ is independent of the choice of $F(r,s)$. For $s = m$ we choose $F(p,m)$ to be the free $Z_p[\eta_p]$ module with basis $\{1_R \otimes \lambda_k\}_1^m$ and let $\rho_m$ denote the regular (adjoint) representation defined by $\rho_m(1_R \otimes \lambda_k)\colon F(p,m) \to F(p,m)\colon 1_R \otimes \lambda_{k'} \to 1_R \otimes [\lambda_k, \lambda_{k'}]$.

*Proposition VI*: Let $Z_p \otimes L$ be a simple non-Abelian Lie algebra. Then

(1) $\rho_m$ is faithful;

(2) $\rho_m^{(r)}(\eta_p^2 L[\eta_p^2])$ is irreducible if and only if $r \geq 5$;

(3) $\exp \circ \rho_m^{(r)}(\eta_p^2 L[\eta_p^2])$ is isomorphic to a direct product of $m$ inequivalent $(p, r)$ analogs of U(1) if and only if $r = 4$.

*Proof* (Sketch): (1) and (2) are elementary consequences of the definitions. For $r=4$ there exists a Lie isomorphism $\rho_m^{(4)} \cong \sum_{k=1}^m \sigma_k$ (direct) where the skew-Hermitian singlet representation $\sigma_k: L[\eta_p] \to M_1[\eta_4]$ is defined by $\sigma_k(\phi \otimes \lambda_{k'}) = \delta_{kk'} \phi \eta_p (\mathrm{mod}\, \eta_p^4)$. If $r=3$ then $\eta_r^2 Z_p[\eta_r]$ contains no non-zero skew-Hermitian elements, while if $r \geq 5$ then VI.2 precludes a decomposition of $\rho_m^{(r)}$.

*Remark*: It is well known that if $\hat{L}$ is simple then $Z_p \otimes L$ is simple for all $p$ sufficiently large.

Finally, let $L'$ be a non-Abelian Lie subring of $L$. Let $F(p, m')$ denote the free submodule of $F(p, m)$ spanned by $1_R \otimes L'$. Let $G(p, m')$ denote the symmetry group of $F(p, m')$, that is, the subgroup of Lie automorphisms on $F(p, m)$ which leave $F(p, m')$ invariant. Clearly $G(p, m')$ contains a subgroup of inner automorphisms isomorphic to $\exp \circ \rho_m(\eta_p^2 L'[\eta_p^2])$. The group $G(p, m')$ may be regarded as the spontaneously broken symmetry group corresponding to the reduced skew-Hermitian Abelian multiplet representation $\rho_{m'}^{(4)}: L[\eta_p] \to M_{m'}[\eta_4]$ which is injective on $1_R \otimes L'$. No comparable model for spontaneously broken symmetries is possible if representation theory is limited to the representations of $\hat{L}$, because simple non-Abelian Lie algebras cannot have nontrivial Abelian representations.

## III. APPLICATIONS TO UNIFIED GAUGE FIELD THEORY

If the observed elementary hadrons are classified into multiplets by standard group theoretic methods, then it appears that only a small number of mathematically possible multiplets are realized in nature.[6] This confinement phenomenon, which includes quark confinement, is now generally thought to signify that all observed hadrons correspond to (colorless, flavored) $\mathrm{SU}(n)_{\mathrm{color}}$ singlets for some $n \geq 3$. The consensus is that color confinement will eventually be deduced from a unified gauge field theory carrying exact (not spontaneously broken) $\mathrm{SU}(n)_{\mathrm{color}}$ symmetry.[7] Little progress on this problem has been made, however, because there is no known method for constructing bounded solutions to quantum field equations which are invariant under an exact non-Abelian gauge group. Some suggestive results have been obtained with the aid of discrete (infinite) lattice models using finite-difference methods,[8] but such approaches leave infrared divergence problems unresolved.

Reformulation of exact color gauge field theories in discrete finite form, utilizing finite nilpotent Lie analogs, should produce several unique mathematical advantages: first, and perhaps most important, all divergence problems should be eliminated without the need for special renormalization procedures; second, all cutoffs should enter through representation theory without the need for special approximations; third, the singlet representations of the abstract non-Abelian color gauge group should be naturally distinguished, possibly suggesting a new unified group theoretic interpretation for color and flavor; and fourth, there should be a natural hierarchy of spontaneously broken gauge symmetries corresponding to the symmetry subgroups of the full automorphism group for the abstract color gauge group, possibly suggesting a new unified group theoretic interpretation for exact and spontaneously broken gauge symmetries (cf. Ref. 9). These observations indicate that further studies and applications of finite nilpotent Lie analogs may accelerate progress toward the solution of fundamental problems in quantum field theory.


## ACKNOWLEDGMENT

I am grateful to Hans Zassenhaus, Harry Allen, and Bernard Kripke for helpful discussions, and to Leo Lipetz for continued encouragement and support.



[1]D.J. Winter, *Abstract Lie Algebras* (M.I.T. Press, Cambridge, 1972), p. 34.
[2]E.G. Beltrametti and A.A. Blasi, J. Math. Phys. **9**, 1027 (1968), and references cited therein.
[3]B.R. McDonald, *Geometric Algebra over Local Rings* (Dekker, New York, 1976).
[4]V.S. Varadarajan, *Lie Groups, Lie Algebras, and Their Representations* (Prentice-Hall, Englewood Cliffs, N.J., 1974).
[5]C. Chevalley, *Theory of Lie Groups* (Princeton U.P., Princeton, N.J., 1946).
[6]S.L. Glashow, in *Gauge Theories and Modern Field Theory*, edited by R. Arnowitt and P. Nath (M.I.T. Press, Cambridge, 1976).
[7]R.A. Brandt *et al.*, Phys. Rev. D **16**, 1119 (1977).
[8]K. Wilson, in *Gauge Theories and Modern Field Theory*, edited by R. Arnowitt and P. Nath (M.I.T. Press, Cambridge, 1976).
[9]S. Weinberg, in Ref. 8.

# On a possible experiment to evaluate the validity of the one-speed or constant cross section model of the neutron-transport equation

C. E. Siewert[a]

*Boğaziçi Üniversitesi, Istanbul, Turkey*
(Received 16 January 1978)

The inverse problem for a half-space is solved (for isotropic scattering) to yield results that suggest an idealized experiment that could be used to evaluate in a new way the validity of the one-speed or constant cross section model of the neutron-transport equation.

## INTRODUCTION

Inverse problems in the theory of neutron diffusion have been discussed in recent years for finite[1,2] and infinite media.[3–5] Here we would like to investigate the half-space inverse problem for the one-speed or constant cross section model of the neutron-transport equation and to show how the established results suggest an experiment that could be used to evaluate the isotropic-scattering model of the neutron-transport equation.

## ANALYSIS

We consider the neutron-transport equation

$$\mu \frac{\partial}{\partial x} \psi(x,\mu) + \psi(x,\mu) = \frac{c}{2}\int_{-1}^{1} \psi(x,\mu')\,d\mu', \tag{1}$$

where $\psi(x,\mu)$ is the neutron angular flux, $x$ is the position variable measured in mean-free-paths, $\mu$ is the direction cosine, and

$$c = (\nu\Sigma_f + \Sigma_s)/\Sigma \tag{2}$$

is the mean number of secondary neutrons per collision. Traditionally for $c < 1$, we seek to solve Eq. (1) in a semi-infinite half-space such that

$$\psi(0,\mu) = F(\mu), \quad \mu > 0 \tag{3a}$$

and

$$\psi(\infty,\mu) = 0, \tag{3b}$$

where $F(\mu)$ is considered given. Here we consider that $F(\mu)$ is specified, that $\psi(0,-\mu)$, $\mu > 0$, can be measured experimentally, and that we wish to determine the mean number of secondaries $c$.

We know from the work of Chandrasekhar[6] that the exit flux can be computed from

$$\psi(0,-\mu) = \frac{c}{2} H(\mu) \int_0^1 H(x) F(x)\, x \frac{dx}{x+\mu}, \quad \mu > 0, \tag{4}$$

where $H(\mu)$ satisfies

$$H(\mu) = 1 + \frac{c}{2}\mu H(\mu) \int_0^1 H(x) \frac{dx}{x+\mu}. \tag{5}$$

It is clear that we cannot readily solve Eq. (4) for $c$ since $H(\mu)$ is a function of $c$. Moments of the exit distribution can be found by multiplying Eq. (4) by $\mu^\alpha$ and integrating over $\mu$. For example, after using Eq. (5), we can write

$$\psi_0 = \int_0^1 F(x)[H(x) - 1]\,dx, \tag{6a}$$

$$\psi_1 = \int_0^1 F(x)[-xH(x)\sqrt{1-c} + x]\,dx, \tag{6b}$$

and

$$\psi_2 = \int_0^1 F(x) \left[ xH(x) \left( \frac{c}{2} H_1 + x\sqrt{1-c} \right) - x^2 \right] dx, \tag{6c}$$

where

$$H_\alpha = \int_0^1 H(x)\, x^\alpha\, dx \tag{7a}$$

and

$$\psi_\alpha = \int_0^1 \psi(0,-\mu)\, \mu^\alpha\, d\mu. \tag{7b}$$

If we consider the special case of an isotropic incident flux, $F(\mu) = 1$, then the resulting version of Eq. (6a) yields

$$\psi_0^{(0)} = H_0 - 1 = (2/c)(1 - \sqrt{1-c}) - 1, \tag{8}$$

which can be solved for $c$ to yield

$$c = \frac{4\psi_0^{(0)}}{[\psi_0^{(0)} + 1]^2}. \tag{9}$$

Here we use

$$\psi_\alpha^{(\beta)} = \int_0^1 \psi^{(\beta)}(0,-\mu)\, \mu^\alpha\, d\mu, \tag{10}$$

where $\psi^{(\beta)}(x,\mu)$ denotes the solution of Eq. (1) corresponding to $F(\mu) = \mu^\beta$.

If we now consider $F(\mu) = \mu$, then Eqs. (6a) and (6b) can be used with the identity[6]

$$\sqrt{1-c}\, H_2 + (c/4) H_1^2 = \tfrac{1}{3} \tag{11}$$

to deduce

$$c = \frac{4\psi_1^{(1)}}{[\psi_0^{(1)} + \frac{1}{2}]^2}. \tag{12}$$

In a similar manner Eqs. (6a) and (6c) and the identity[7]

$$\sqrt{1-c}\, H_4 - (c/2)(\tfrac{1}{2}H_2^2 - H_3 H_1) = \tfrac{1}{5} \tag{13}$$

can be used to establish

$$c = \frac{4\psi_2^{(2)}}{[\psi_0^{(2)} + \frac{1}{3}]^2}. \tag{14}$$

[a] Permanent address: Nuclear Engineering Department, N. C. State University, Raleigh, N.C. 27607.

With the aid of Busbridge's identity[7] concerning moments of the $H$ function,

$$\sqrt{1-c}\,H_{2\alpha}+(c/4)(H_1H_{2\alpha-1}-H_2H_{2\alpha-2}+\cdots+H_{2\alpha-1}H_1)=\frac{1}{2\alpha+1}, \tag{15}$$

we can generalize Eqs. (9), (12), and (14) to obtain

$$c=\frac{4\psi_\beta^{(\beta)}}{[\psi_0^{(\beta)}+(\beta+1)^{-1}]^2},\quad \beta=0,1,2,3,\cdots. \tag{16}$$

Generally when we apply Eq. (1) to physical problems we consider $c$ to be a constant and thus clearly not a function of the boundary conditions. It thus seems feasible that the manner in which $c$, as computed from Eq. (16) and the experimentally measured $\psi^{(\beta)}(0,-\mu)$, varies with $\beta$ would be a reasonable measure of the accuracy with which Eq. (1) represents the given physical problem. It also seems feasible that the multigroup version of Eq. (16) would offer a definition of the transfer cross sections alternative to the traditional one. The finite-slab inverse problem solved in Ref. 2 for the multigroup model could serve a similar purpose.


## ACKNOWLEDGMENTS

The author is grateful to E. İnönü and Boğaziçi Üniversitesi for their kind hospitality and partial support of this work and to T. Yarman whose queries concerning the inverse problem were partly responsible for some of the work reported here. This work was also partially supported by the Centre d'Etudes Nucléaires de Saclay and the National Science Foundation through grant ENG. 7709405.



[1]S. Pahor, Phys. Rev. **175**, 218 (1968).
[2]C. E. Siewert, "The Inverse Problem for a Finite Slab," Nucl. Sci. Eng. (in press).
[3]K. M. Case, Phys. Fluids **16**, 1607 (1973).
[4]N. J. McCormick and I. Kuščer, J. Math. Phys. **15**, 926 (1974).
[5]C. E. Siewert, M. N. Özisik, and Y. Yener, Nucl. Sci. Eng. **63**, 95 (1977).
[6]S. Chandrasekhar, *Radiative Transfer* (Oxford U. P., London, 1950).
[7]I. W. Busbridge, *The Mathematics of Radiative Transfer* (Cambridge U. P., London, 1960).

# Lorentz transformations of observable and ghost particle states in quantum electrodynamics and in a massive gauge theory

K. Haller

*Department of Physics, University of Connecticut, Storrs, Connecticut 06268*

R. B. Sohn

*Department of Physics, University of Connecticut, Stamford, Connecticut 06903*
(Received 3 October 1977)

Derivations are given of the effect of Lorentz boosts on physical particle and ghost states in quantum electrodynamics. It is shown that the photon helicity is an invariant even though, in general, Lorentz boosts transform the transverse, longitudinal, and timelike components of the vector potential into each other. A similar calculation is made for an Abelian gauge theory in which the particles have dynamical mass.

## I. INTRODUCTION

Gauge theories in manifestly covariant formulations generate particle spectra that include ghosts as well as observable particles. Some ghost particles, such as the scalar ghost in quantum electrodynamics (qed), are excluded from the set of observable states by subsidiary conditions.[1] Others, like the zero helicity (longitudinal) ghost in qed, satisfy the appropriate subsidiary condition and are members of the set of allowed states, but have vanishing norms and therefore are still unobservable.

Under Lorentz transformations particle states are transformed into different particle states. This corresponds to the fact that an observation, carried out in an inertial reference frame $F$, which detects a particle with parameters $P$, would in general detect a particle with different parameters, $P'$, when made in a different inertial frame $F'$. Some aspects of Lorentz transformations, when carried out on particle states in gauge theories, become involved with the indefinite metric of the underlying Hilbert space and warrant special attention. For example, a Lorentz boost applied to a transversely polarized photon which is propagating in a direction not parallel to the boost, transforms the transverse photon partially into nontransverse ghost states. The fact that the photon mass is zero makes it necessary that the photon helicity be unaffected by the Lorentz boost, even though there is mixing between transverse and nontransverse states.[2] It is only in an indefinite metric space that these apparently inconsistent conditions can be reconciled. Consistency also requires that photon states that obey the subsidiary condition in one inertial frame do so in all inertial frames, and that this important condition for the covariance of the theory is not threatened by the indefinite metric and the mixing of transverse photons and ghosts in Lorentz transformations.[3]

For theories that can be formulated in a positive metric space, group theoretic methods can and have been used to study Lorentz transformations of particle states.[4] But within a positive metric space the transformations among the transverse and nontransverse photon states cannot be treated consistently. In order to accomodate the requirements of an indefinite metric space, we have constructed the generators that effect the Lorentz boost and explicitly calculated the transformed particle operators, including their ghost as well as their observable components. In this paper we will report on this calculation. We will also include a discussion of the mixing among the $+1$, $0$, and $-1$ helicity states of a massive Abelian gauge theory when a Lorentz transformation is carried out.

Although Lorentz transformations on observable and ghost states in gauge theories do not present any extraordinary mathematical difficulties, they are not explicitly carried out in texts, nor, to our knowledge, anywhere else in the literature. We address ourselves to this question in this paper, partly because the topic is of interest in its own right. The subject is also of special interest because it can be useful in identifying the particle spectra of gauge theories with spontaneously broken symmetries. Previously[5] we have discussed a model that incorporates spontaneous symmetry breaking and, in a later work, we hope to identify the particle spectrum of that model unambiguously by using the techniques developed here to examine how the relevant massive particles transform among themselves under a Lorentz transformation.

## II. LORENTZ TRANSFORMATION IN QUANTUM ELECTRODYNAMICS

The Lagrangian for the free Maxwell field in the Feynman gauge is[6]

$$\mathcal{L} = -\tfrac{1}{4}F_{\mu\nu}F_{\mu\nu} - G(x)\partial_\mu A_\mu + \tfrac{1}{2}G^2(x) \tag{1}$$

with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. $G(x)$ is a Lagrange multiplier field which keeps $\Pi_4$, the momentum conjugate to $A_4$, from vanishing identically. The Euler-Lagrange equations generated by $\mathcal{L}$ are

$$\Box A_\mu - \partial_\mu(\partial_\nu A_\nu) + \partial_\mu G(x) = 0 \tag{2}$$

and

$$G(x) = \partial_\mu A_\mu \tag{3}$$

 0022-2488/78/1907-1589$1.00  

and the canonical momenta adjoint to $A_\mu$ are

$$\Pi_j = \partial_0 A_j - i\partial_j A_4 \tag{4}$$

and

$$\Pi_4 = iG. \tag{5}$$

The Lorentz boost generator is given by

$$M_{4j} = i\int d\mathbf{x}\{x_0 \mathcal{P}_j - x_j \mathcal{H}\} - \int d\mathbf{x}\{[\partial_0 A_4]A_j - [\partial_0 A_j]A_4\}, \tag{6}$$

where $\mathcal{P}_j$ is the momentum density

$$\mathcal{P}_j = -\Pi_\mu \partial_j A_\mu \tag{7}$$

and $\mathcal{H}$ is the Hamiltonian density

$$\mathcal{H} = \tfrac{1}{2}\Pi_\mu\Pi_\mu + \tfrac{1}{4}F_{jk}F_{jk} + i[\Pi_j\partial_j A_4 - \Pi_4\partial_j A_j] - \partial_j[A_j\partial_i A_i]. \tag{8}$$

The last term on the right-hand side of Eq. (8) is a total divergence which makes no contribution to the Hamiltonian $H = \int d\mathbf{x}\mathcal{H}$ but does make an important contribution to $M_{4j}$ [Eq. (6)]. Unless this total divergence is included in $\mathcal{H}$, the $A_\mu$ and the $F_{\mu\nu}$ will not all transform properly among themselves like components of vectors and antisymmetric tensors respectively. The decomposition of $A_\mu$ and $\Pi_\mu$ into particle creation and annihilation operators proceeds as in Ref. 7, except that we now choose invariant integration over momentum space variables. We express $A_\mu$ as

$$A_\mu(x) = \frac{1}{(2\pi)^{3/2}}\int \frac{dk}{2k_0}\sum_{\lambda=1}^{4}\hat{\epsilon}^{\lambda}_{\mu}(\mathbf{k}) \times [A_\lambda(\mathbf{k})\exp(ik_\mu x_\mu) + A^\dagger_\lambda(\mathbf{k})\exp(-ik_\mu x_\mu)] \tag{9}$$

with $k_0 = |\mathbf{k}|$, where $A_\lambda(\mathbf{k})$ and $A^\dagger_\lambda(\mathbf{k})$ obey the commutation rule

$$[A_\lambda(\mathbf{k}),\ A^\dagger_{\lambda'}(\mathbf{k}')] = 2k_0\delta_{\lambda\lambda'}\delta(\mathbf{k}-\mathbf{k}').$$

Here $\hat{\epsilon}^4_\mu(\mathbf{k}) = \delta_{\mu,4}$ and $\hat{\epsilon}^j(\mathbf{k})$ indicates a set of unit 3-vectors of which $\hat{\epsilon}^{(3)}(\mathbf{k}) = \mathbf{k}/|\mathbf{k}|$; $\hat{\epsilon}^{(1)}(\mathbf{k})$ and $\hat{\epsilon}^{(2)}(\mathbf{k})$ are two unit vectors that, together with $\hat{\epsilon}^{(3)}(\mathbf{k})$ form an orthogonal triad. We express $\Pi_\mu$ as

$$\begin{aligned}\Pi_\mu(x) = \frac{1}{(2\pi)^{3/2}}\int \frac{d\mathbf{k}}{2}\Big\{&\sum_{\lambda=1}^{4} i\hat{\epsilon}^\lambda_\mu(\mathbf{k}) \times [-A_\lambda(\mathbf{k})\exp(ik_\mu x_\mu) + A^\dagger_\lambda(\mathbf{k})\exp(-ik_\mu x_\mu)] \\ &+ [i\hat{\epsilon}^3_\mu(\mathbf{k}) + \hat{\epsilon}^4_\mu(\mathbf{k})][-(A_3(\mathbf{k}) + iA_4(\mathbf{k}))\exp(ik_\mu x_\mu) \\ &+ (A^\dagger_3(\mathbf{k}) + iA^\dagger_4(\mathbf{k}))\exp(-ik_\mu x_\mu)]\Big\}.\end{aligned} \tag{10}$$

The space–time function $\exp(ik_\mu x_\mu)$ (with $k_\mu x_\mu = \mathbf{k}\cdot\mathbf{x} - k_0 x_0$) includes the explicit $c$-number time dependence characteristic of free fields. We exclude interactions between photons and charged particles because these interactions have little to do with the question of how single particle photon states appear in different inertial frames. Moreover, the definition of single particle states is very much complicated by the persistent effects of charged particle–photon interactions, and this too has motivated us to omit these effects in this work.

The operators that describe the scalar and the zero-helicity ghosts are

$$A_Q(\mathbf{k}) = (1/\sqrt{2})[A_3(\mathbf{k}) + iA_4(\mathbf{k})], \tag{11a}$$

$$A_R(\mathbf{k}) = (1/\sqrt{2})[A_3(\mathbf{k}) - iA_4(\mathbf{k})], \tag{11b}$$

and their adjoints are

$$A^\dagger_Q(\mathbf{k}) = (1/\sqrt{2})[A^\dagger_3(\mathbf{k}) - iA^\dagger_4(\mathbf{k})] \tag{12a}$$

and

$$A^\dagger_R(\mathbf{k}) = (1/\sqrt{2})[A^\dagger_3(\mathbf{k}) + iA^\dagger_4(\mathbf{k})]. \tag{12b}$$

The $^\dagger$ adjoint denotes the Hermitian adjoint, which, in an indefinite metric space, is not the representation-independent adjoint that relates self-adjointedness of an operator with the reality of its eigenvalues. The adjointing operation that satisfies this criterion is the $^\star$ adjoint, for which[8] $A^\star_Q = A^\dagger_R$ and $A^\star_R = A^\dagger_Q$. The Gupta–Bleuler subsidiary condition,

$$\partial_\mu A^{(+)}_\mu |n\rangle = 0, \tag{13}$$

translates into the momentum space equation

$$A_Q(\mathbf{k})|n\rangle = 0, \tag{14}$$

and the photon state $A^\dagger_Q|0\rangle$ represents the scalar photon forbidden by the subsidiary condition. The state $A^\dagger_R|0\rangle$ represents the zero-helicity photon, and it satisfies the subsidiary condition [Eq. (14)]. Since $\langle A_R|A^\star_R\rangle = 0$ and $\langle A_Q|A^\star_Q\rangle = 0$, both the forbidden scalar photon state and the zero-helicity photon state are ghosts. In a superposition of allowed states, i.e., those obeying Eq. (14), the zero-helicity components are not observable since they have zero norm and therefore zero probability of being detected.

Under a Lorentz transformation the vector potential transforms according to

$$A'_\mu(x_\alpha) = \Lambda_{\mu\nu}A_\nu[\Lambda^{-1}_{\alpha\beta}x_\beta]. \tag{15}$$

When the transformation is infinitessimal, $\Lambda_{\mu\nu}$ is given by

$$\Lambda_{\mu\nu} = \delta_{\mu\nu} + \omega_{\mu\nu} \tag{16}$$

and in the case of a pure Lorentz boost $\omega_{4j} = -i\delta\beta_j = -\omega_{j4}$, while all other components of $\omega_{\mu\nu}$ vanish; $\delta\beta_j = \delta v_j/c$, where $\delta\mathbf{v}$ is the infinitessimal velocity which the origin of the primed inertial frame has in the unprimed one. The infinitessimal change in $A_\mu(x)$, $\delta A_\mu(x) = A'_\mu(x) - A_\mu(x)$ is given by

$$\delta A_\mu(x) = \bar{\delta}A_\mu(x) + (\delta x_\nu)\partial_\nu A_\mu(x). \tag{17}$$

Here $\bar{\delta}A_\mu(x)$ is the part of the transformation that "scrambles" the vector components, and will be written as

$$\begin{aligned}\bar{\delta}A_\mu(x) = \frac{1}{(2\pi)^{3/2}}\int\frac{d\mathbf{k}}{2k_0}\Big\{&\sum_{\lambda=1}^{4}\delta\hat{\epsilon}^\lambda_\mu(\mathbf{k}) \times [A_\lambda(\mathbf{k})\exp(ik_\mu x_\mu) + A^\dagger_\lambda(\mathbf{k})\exp(-ik_\mu x_\mu)] + \sum_{\lambda=1}^{4}\hat{\epsilon}^\lambda_\mu(\mathbf{k}) \\ &\times[\delta A_\lambda(\mathbf{k})\exp(ik_\mu x_\mu) + \delta A^\dagger_\lambda(\mathbf{k})\exp(-ik_\mu x_\mu)]\Big\}.\end{aligned} \tag{18}$$

In $\bar{\delta}A_\mu(x)$ the "scrambling" of the components of $A_\mu$ is represented by the first order change in the momentum space photon operators, and in the unit vectors that mark the direction of propagation and the two transverse polarization directions. $(\delta x_\nu)\partial_\nu A_\mu(x)$ transforms the space–time point in the argument of $A_\mu(x)$.

In evaluating $\bar{\delta}A_\mu(x)$ we need to calculate the first order variations $\delta\hat{\epsilon}_\mu^\lambda(\mathbf{k})$. In order to permit a consistent interpretation of $A_3(\mathbf{k})$ and $A_3^\dagger(\mathbf{k})$ as the annihilation and creation operators, respectively, for photons polarized in the propagation direction, and $A_1(\mathbf{k})$, $A_2(\mathbf{k})$ and $A_1^\dagger(\mathbf{k})$ and $A_2^\dagger(\mathbf{k})$ as the appropriate operators for the transversely polarized photons, we must transform the three unit polarization vectors $\hat{\epsilon}^{(1)}(\mathbf{k})$, $\hat{\epsilon}^{(2)}(\mathbf{k})$, and $\hat{\epsilon}^{(3)}(\mathbf{k})$ so that they remain an orthonormal triad in all inertial frames. A representation that satisfied these requirements is $\hat{\epsilon}^{(3)}(\mathbf{k})=\mathbf{k}/|\mathbf{k}|$ and $\hat{\epsilon}^{(1)}(\mathbf{k})=\mathbf{e}^{(1)}(\mathbf{k})/|\mathbf{e}_1(\mathbf{k})|$, $\hat{\epsilon}^{(2)}(\mathbf{k})=\mathbf{e}^{(2)}(\mathbf{k})/|\mathbf{e}^{(2)}(\mathbf{k})|$, where $e_i^{(1)}$ and $e_i^{(2)}$ are the entries in a second rank antisymmetric tensor $t_{\mu\nu}$ with $t_{12}=e_3^{(2)}$, $t_{13}=-e_2^{(2)}$, $t_{23}=e_1^{(2)}$, $t_{14}=-ie_1^{(1)}$, $t_{24}=-ie_2^{(1)}$, and $t_{34}=-ie_3^{(1)}$. Lorentz transformation of the tensor $t_{\mu\nu}$ produces the following first order variations in the polarization unit vectors:

$$\delta\hat{\epsilon}^{(3)}(\mathbf{k})=-(1/v)[\mathbf{v}-(\mathbf{v}\cdot\hat{k})\hat{k}]\delta(\beta), \tag{19a}$$

$$\delta\hat{\epsilon}^{(1)}(\mathbf{k})=-(1/v)[\hat{\epsilon}^{(2)}\times\mathbf{v}-\hat{k}\cdot\mathbf{v}\,\hat{\epsilon}^{(1)}]\delta(\beta), \tag{19b}$$

and

$$\delta\hat{\epsilon}^{(2)}(\mathbf{k})=(1/v)[\hat{\epsilon}^{(1)}\times\mathbf{v}+\hat{k}\cdot\mathbf{v}\,\hat{\epsilon}^{(2)}]\delta(\beta). \tag{19c}$$

Use of Eqs. (19) in representing the variation $\bar{\delta}A_\mu$ leads to

$$\begin{aligned}\bar{\delta}\mathrm{A}(x)=&-\frac{\delta\beta}{(2\pi)^{3/2}}\int\frac{d\mathbf{k}}{2k_0}\{[\hat{\epsilon}^{(1)}(\hat{\mathbf{v}}\cdot\hat{\epsilon}^{(1)})+\hat{\epsilon}^{(2)}(\hat{\mathbf{v}}\cdot\hat{\epsilon}^{(2)})]\\&\times[A_3(\mathbf{k})\exp(ik_\mu x_\mu)+A_3^\dagger(\mathbf{k})\exp(-ik_\mu x_\mu)]\\&-\hat{\epsilon}^{(3)}[\hat{\epsilon}^{(1)}\cdot\hat{\mathbf{v}}(A_1(\mathbf{k})\exp(ik_\mu x_\mu)+A_1^\dagger(\mathbf{k})\exp(-ik_\mu x_\mu)\\&+\hat{\epsilon}^{(2)}\cdot\hat{\mathbf{v}}(A_2(\mathbf{k})\exp(ik_\mu x_\mu)+A_2^\dagger(\mathbf{k})\exp(-ik_\mu x_\mu))]\}\\&+\frac{1}{(2\pi)^{3/2}}\int\frac{d\mathbf{k}}{2k_0}\{\hat{\epsilon}^{(1)}[\delta A_1(\mathbf{k})\exp(ik_\mu x_\mu)+\delta A_1^\dagger(\mathbf{k})\\&\times\exp(-ik_\mu x_\mu)]+\hat{\epsilon}^{(2)}[\delta A_2(\mathbf{k})\exp(ik_\mu x_\mu)+\delta A_2^\dagger(\mathbf{k})\\&\times\exp(-ik_\mu x_\mu)]+\hat{\epsilon}^{(3)}[\delta A_3(\mathbf{k})\exp(ik_\mu x_\mu)+\delta A_3^\dagger(\mathbf{k})\\&\times\exp(-ik_\mu x_\mu)]\}\end{aligned} \tag{20a}$$

and

$$\begin{aligned}\bar{\delta}A_4(x)=&\frac{1}{(2\pi)^{3/2}}\int\frac{d\mathbf{k}}{2k_0}[\delta A_4(\mathbf{k})\exp(ik_\mu x_\mu)\\&+\delta A_4^\dagger(\mathbf{k})\exp(-ik_\mu x_\mu)].\end{aligned} \tag{20b}$$

In the remainder of this section we will use Eqs. (20), together with $\delta x_\nu\partial_\nu A_\mu(x)$ and the expression

$$\delta A_\mu(x)=-i\omega_{\alpha\beta}[M_{\alpha\beta},\,A_\mu(x)], \tag{21}$$

which we will substitute into Eq. (17) in order to find the explicit form of $\delta A_\lambda(\mathbf{k})$ [and $\delta A_\lambda^\dagger(\mathbf{k})$], the change in $A_\lambda(\mathbf{k})$ [and $A_\lambda^\dagger(\mathbf{k})$] that the Lorentz boost produces. $\delta x_\nu\partial_\nu A_\mu(x)$ can be written

$$\begin{aligned}\delta x_\nu\partial_\nu A_\mu(x)=&\frac{i\delta\beta_j}{(2\pi)^{3/2}}\int\frac{d\mathbf{k}}{2}\{[x_j-x_0\hat{k}_j]\sum_{\lambda=1}^{4}\hat{\epsilon}_\mu^\lambda(\mathbf{k})\\&\times[A_\lambda(\mathbf{k})\exp(ik_\mu x_\mu)-A_\lambda^\dagger(\mathbf{k})\exp(-ik_\mu x_\mu)]\}.\end{aligned} \tag{22}$$

We can replace $x_j\exp(ik_\mu x_\mu)$ in Eq. (22) by an integration by parts in momentum space. Modulo integration by parts we have

$$\begin{aligned}&\frac{\partial}{\partial k_j}\exp(ik_\mu x_\mu)\\&=ix_j\exp(ik_\mu x_\mu)-ix_0\frac{k_j}{k_0}\exp(ik_\mu x_\mu);\end{aligned} \tag{23}$$

the last term on the right-hand side of Eq. (23) stems from the mass shell constraint $k_jk_j=k_0^2$ in $\exp(ik_\mu x_\mu)$.

If this substitution is made in Eq. (22), we have

$$\begin{aligned}\partial_\nu x_\nu A_\nu(x\partial)=&-\frac{\delta\beta_j}{(2\pi)^{3/2}}\int\frac{d\mathbf{k}}{2}\sum_{\lambda=1}^{4}\left\{\left[\frac{\partial\hat{\epsilon}_\mu^\lambda(\mathbf{k})}{\partial k_j}A_\lambda(\mathbf{k})\right.\right.\\&\left.+\hat{\epsilon}_\mu^\lambda(\mathbf{k})\frac{\partial A_\lambda(\mathbf{k})}{\partial k_j}\right]\exp(ik_\mu x_\mu)\\&+\left[\frac{\partial\hat{\epsilon}_\mu^\lambda(\mathbf{k})}{\partial k_j}A_\lambda^\dagger(\mathbf{k})+\hat{\epsilon}_\mu^\lambda(\mathbf{k})\frac{\partial A_\lambda^\dagger(\mathbf{k})}{\partial k_j}\right]\\&\left.\times\exp(-ik_\mu x_\mu)\right\}.\end{aligned} \tag{24}$$

$\delta A_\mu(x)$ is evaluated by using the commutator in Eq. (21) and the expressions in Eqs. (6)–(8), and by making use of the identity given in Eq. (23). This leads to

$$\begin{aligned}\delta A_\mu(x)=&-\frac{\delta\beta_j}{(2\pi)^{3/2}}\int\frac{d\mathbf{k}}{2k_0}\sum_{\lambda=1}^{4}\hat{\epsilon}_\mu^\lambda\left\{\left[-k_0\frac{\partial A_\lambda}{\partial k_j}-k_0\right.\right.\\&\left.\times\sum_{\lambda',\mu'=1}^{4}\hat{\epsilon}_{\mu'}^\lambda\frac{\partial\hat{\epsilon}_{\mu'}^{\lambda'}}{\partial k_j}A_{\lambda'}+i\sum_{\lambda'=1}^{4}(\hat{\epsilon}_4^\lambda\hat{\epsilon}_j^{\lambda'}A_{\lambda'}-\hat{\epsilon}_j^\lambda\hat{\epsilon}_4^{\lambda'}A_{\lambda'})\right]\\&\times\exp(ik_\mu x_\mu)+\left[-k_0\frac{\partial A_\lambda^\dagger}{\partial k_j}-k_0\sum_{\lambda',\mu'=1}^{4}\hat{\epsilon}_{\mu'}^\lambda\frac{\partial\hat{\epsilon}_{\mu'}^{\lambda'}}{\partial k_j}A_{\lambda'}^\dagger\right.\\&\left.\left.+i\sum_{\lambda'=1}^{4}(\hat{\epsilon}_4^\lambda\hat{\epsilon}_j^{\lambda'}A_{\lambda'}^\dagger-\hat{\epsilon}_j^\lambda\hat{\epsilon}_4^{\lambda'}A_{\lambda'}^\dagger)\right]\exp(-ik_\mu x_\mu)\right\}.\end{aligned} \tag{25}$$

The resulting expressions for $\delta A_\lambda^\dagger(\mathbf{k})$ are

$$\delta A_1^\dagger(\mathbf{k})=\sqrt{2}\,\delta\beta\hat{\epsilon}^{(1)}(\mathbf{k})\cdot\hat{v}A_R^\dagger(\mathbf{k}) \tag{26a}$$

$$\delta A_2^\dagger(\mathbf{k})=\sqrt{2}\,\delta\beta\hat{\epsilon}^{(2)}(\mathbf{k})\cdot\hat{v}A_R^\dagger(\mathbf{k}) \tag{26b}$$

$$\delta A_R^\dagger(\mathbf{k})=\delta\beta\hat{\epsilon}^{(3)}(\mathbf{k})\cdot\hat{v}A_R^\dagger(\mathbf{k}) \tag{26c}$$

$$\begin{aligned}\delta A_Q^\dagger(\mathbf{k})=&-\delta\beta[\hat{\epsilon}^{(3)}\cdot\hat{v}A_Q^\dagger(\mathbf{k})+\sqrt{2}\hat{\epsilon}^{(1)}\cdot\hat{v}A_1^\dagger\\&+\sqrt{2}\hat{\epsilon}^{(2)}\cdot\hat{v}A_2^\dagger]\end{aligned} \tag{26d}$$

and the corresponding adjoint equations for the annihilation operators,

$$\delta A_1(\mathbf{k})=\sqrt{2}\,\delta\beta\hat{\epsilon}^{(1)}(\mathbf{k})\cdot\hat{v}A_Q(\mathbf{k}), \tag{26e}$$

$$\delta A_2(\mathbf{k})=\sqrt{2}\,\delta\beta\hat{\epsilon}^{(2)}(\mathbf{k})\cdot\hat{v}A_Q(\mathbf{k}), \tag{26f}$$

$$\delta A_Q(\mathbf{k})=\delta\beta\hat{\epsilon}^{(3)}(\mathbf{k})\cdot\hat{v}A_Q(\mathbf{k}), \tag{26g}$$

$$\begin{aligned}\delta A_R(\mathbf{k})=&-\delta\beta[\hat{\epsilon}^{(3)}\cdot\hat{v}A_R+\sqrt{2}\hat{\epsilon}^{(1)}\cdot\hat{v}A_1\\&+\sqrt{2}\hat{\epsilon}^{(2)}\cdot\hat{v}A_2].\end{aligned} \tag{26h}$$

To translate Eqs. (26) into Lorentz transformations on helicity states, we apply the boosts to creation operators acting on the vacuum, and define $A_{(+)}^\dagger=(A_1^\dagger+iA_2^\dagger)/\sqrt{2}$, $A_{(-)}^\dagger=(A_1^\dagger-iA_2^\dagger)/\sqrt{2}$, as well as $\hat{\epsilon}^{(+)}=(\hat{\epsilon}^{(1)}+i\hat{\epsilon}^{2})/\sqrt{2}$, and $\hat{\epsilon}^{(-)}=(\hat{\epsilon}^{(1)}-i\hat{\epsilon}^{(2)})/\sqrt{2}$. We define

$$|\mathbf{k},(+)\rangle=A_{(+)}^\dagger(\mathbf{k})|0\rangle, \tag{27a}$$

$$|\mathbf{k},(-)\rangle=A_{(-)}^\dagger(\mathbf{k})|0\rangle, \tag{27b}$$

$$|\mathbf{k},R\rangle=A_R^\dagger(\mathbf{k})|0\rangle, \tag{27c}$$

and

$$|\mathbf{k},Q\rangle=A_Q^\dagger(\mathbf{k})|0\rangle. \tag{27d}$$

The transformed states, to first order in $\delta\beta$, are

$$|\mathbf{k},(+)\rangle' = |\mathbf{k},(+)\rangle + \sqrt{2}\,\delta\beta\hat{\epsilon}^{(+)}(\mathbf{k})\cdot\hat{v}\,|\mathbf{k},R\rangle, \tag{28a}$$

$$|\mathbf{k},(-)\rangle' = |\mathbf{k},(-)\rangle + \sqrt{2}\,\delta\beta\hat{\epsilon}^{(-)}(\mathbf{k})\cdot\hat{v}\,|\mathbf{k},R\rangle, \tag{28b}$$

$$|\mathbf{k},R\rangle' = [1+\delta\beta\hat{\epsilon}^{(3)}(\mathbf{k})\cdot\hat{v}]\,|\mathbf{k},R\rangle, \tag{28c}$$

$$|\mathbf{k},Q\rangle' = [1-\delta\beta\hat{\epsilon}^{(3)}\cdot\hat{v}]\,|\mathbf{k},Q\rangle + \delta\beta[\hat{\epsilon}^{(+)}\cdot\hat{v}\,|\mathbf{k},-\rangle + \hat{\epsilon}^{(-)}\cdot\hat{v}\,|\mathbf{k},+\rangle]. \tag{28d}$$

Equations (28a) and (28b) indicate that under Lorentz transformation the photon's helicity is invariant. If the helicity is $+1$ or $-1$, then the generation of $|\mathbf{k},R\rangle$ components in the transformed wavefunction leaves the helicity unaffected because $|\mathbf{k},R\rangle$ is a zero norm ghost which is orthogonal to every other state vector in the physical subspace, and, in particular, is orthogonal to all $|\mathbf{k},(+)\rangle$ and $|\mathbf{k},(-)\rangle$ photon states. It is not orthogonal to $|\mathbf{k},Q\rangle$ states, but these latter are forbidden by the subsidiary condition, and inspection of Eqs. (28) demonstrates that they are never generated by Lorentz transformations of states allowed by the subsidiary condition. The helicity of $|\mathbf{k},(+)\rangle'$ [and $|\mathbf{k},(-)\rangle'$] is therefore trivially identical to that of $|\mathbf{k},(+)\rangle$ [and $|\mathbf{k},(-)\rangle$] respectively. It is the vanishing norm of the $|\mathbf{k},R\rangle$] state in the indefinite metric space that allows the photon's helicity to be invariant in a Lorentz transformation even though a Lorentz boost applied to a transverse state does generate nontransverse components. Similarly Eq. (28c) shows that the $|\mathbf{k},R\rangle$ ghost remains the pure $|\mathbf{k},R\rangle$ ghost state in a Lorentz transformation and neither develops transverse nor forbidden $|\mathbf{k},Q\rangle$ components. Following the usual nomenclature, we have referred to the $|\mathbf{k},R\rangle$ ghost as the zero-helicity photon, but properly speaking it has no helicity at all because its vanishing norm does not allow the definition of a helicity, nor any other expectation value. The Lorentz transformation of the forbidden ghost, $|\mathbf{k},Q\rangle$, is given for completeness, but has no physical significance since $|\mathbf{k},Q\rangle$ states are not admitted into the spectrum of observable states. Since the $|\mathbf{k},Q\rangle$ ghost is never generated by a Lorentz transformation of a transverse or zero-helicity ghost, and, equivalently, since the Lorentz transform of $A_Q(\mathbf{k})$ only involves further $A_Q(\mathbf{k})$ components, the subsidiary condition is easily explicitly shown to be invariant.

## III. GAUGE THEORIES WITH MASSIVE PARTICLES

In this section we will apply Lorentz boosts to the particle states of a massive gauge theory. This theory has previously been discussed by one of us[9] (KH). It results in a spin one massive boson, but it is a gauge theory and is formulated with a Gupta–Bleuler subsidiary condition in an indefinite metric space. It is therefore different from the Proca theory of massive vector bosons.[10]

The Lagrangian for this theory is

$$\mathcal{L} = -\tfrac{1}{4}F_{\mu\nu}F_{\mu\nu} - \tfrac{1}{2}M^2W_\mu W_\mu - G(x)\partial_\mu W_\mu + \tfrac{1}{2}G^2_{(x)}, \tag{29}$$

and the Euler Lagrange equations are

$$\partial_\mu F_{\mu\nu} - M^2 W_\nu + \partial_\nu G = 0. \tag{30}$$

The Hamiltonian density is

$$\mathcal{H} = \tfrac{1}{2}\Pi_\mu^2 + \tfrac{1}{2}M^2W_\mu^2 + \tfrac{1}{4}F_{jk}F_{jk} + i[\Pi_j\partial_jW_4 - \Pi_4\partial_jW_j] - \partial_j[W_j\partial_kW_k]. \tag{31}$$

Except for the case that in this section $k_0$ denotes $[|\mathbf{k}|^2+M^2]^{1/2}$, the momentum space decomposition of $W_\mu$ exactly parallels that of $A_\mu$ in Eq. (9), and the same unit polarization vectors are used in this case. The decomposition of $\Pi_\mu$ is given by

$$\begin{aligned}\Pi_\mu(x) = \frac{1}{(2\pi)^{3/2}}\int\frac{d\mathbf{k}}{2}\Big\{\sum_{\lambda=1}^{4} i\epsilon_\mu^\lambda(\mathbf{k}) &\times[-A_\lambda(\mathbf{k})\exp(ik_\mu x_\mu) + A_\lambda^\dagger(\mathbf{k})\exp(-ik_\mu x_\mu)]\\ &-\frac{|\mathbf{k}|}{k_0}[\epsilon_\mu^4(\mathbf{k})\,A_3(\mathbf{k}) - \epsilon_\mu^3(\mathbf{k})A_4(\mathbf{k})]\exp(ik_\mu x_\mu)\\ &+\frac{|\mathbf{k}|}{k_0}[\epsilon_\mu^4(\mathbf{k})A_3^\dagger(\mathbf{k}) - \epsilon_\mu^3(\mathbf{k})A_4^\dagger(\mathbf{k})]\exp(-ik_\mu x_\mu)\Big\}.\end{aligned} \tag{32}$$

The massive boson creation and annihilation operators are more complicated than the photon operators in qed. They are discussed extensively in Sec. II of Ref. 9, and the relevant ones will be defined here. $A_R^\dagger(\mathbf{k})$ is given by

$$A_R^\dagger(\mathbf{k}) = \frac{N(k)}{\sqrt{2}}[A_3^\dagger(\mathbf{k}) + i\frac{|\mathbf{k}|}{k_0}A_4^\dagger(\mathbf{k})], \tag{33}$$

where $N(k)=\sqrt{2}\,k_0[|\mathbf{k}|^2+k_0^2]^{-1/2}$ and the properly normalized zero-helicity state is given by

$$|\mathbf{k},R\rangle = \{[k_0^2+|\mathbf{k}|^2]^{1/2}/M\}A_R^\dagger(\mathbf{k})\,|0\rangle. \tag{34}$$

Equation (34) marks one of the most significant differences between this massive gauge theory and qed.

The state $|\mathbf{k},R\rangle$ is not a zero-norm ghost, but instead is a properly normalized state with $\langle\mathbf{k},R^\star|\mathbf{k}'R\rangle = \delta_{\mathbf{k}\mathbf{k}'}$. The helicities of the massive gauge particles therefore are not invariant under a Lorentz boost, but transform among themselves.

The massive boson forbidden by the subsidiary condition is $A_Q^\dagger(\mathbf{k})\,|0\rangle$,

where

$$A_Q^\dagger(\mathbf{k}) = \frac{N(\mathbf{k})}{\sqrt{2}}\Big[\frac{|\mathbf{k}|}{k_0}A_3^\dagger(\mathbf{k}) - iA_4^\dagger(\mathbf{k})\Big]. \tag{35}$$

The tensor composed of the components $\hat{\epsilon}^{(1)}$ and $\hat{\epsilon}^{(2)}$ in this case is $t_{12}=(|\mathbf{k}|/k_0)e_3^{(2)}$, $t_{13}=-(|\mathbf{k}|/k_0)e_2^{(2)}$, $t_{23}=(|\mathbf{k}|/k_0)e_1^{(2)}$, $t_{14}=-ie_1^{(1)}$, $t_{24}=-ie_2^{(1)}$, and $t_{34}=-ie_3^{(1)}$. The resulting expressions for the variations $\delta\hat{\epsilon}$ are

$$\delta\hat{\epsilon}^{(3)}(\mathbf{k}) = -\frac{k_0}{|\mathbf{k}|}\,\frac{1}{v}\,[\mathbf{v}-(\mathbf{v}\cdot\hat{k})\hat{k}]\,\delta\beta, \tag{36a}$$

$$\delta\hat{\epsilon}^{(1)}(\mathbf{k}) = -\frac{k_0}{|\mathbf{k}|}\,\frac{1}{v}\,[\hat{\epsilon}^{(2)}\times\mathbf{v} - (\hat{k}\cdot\mathbf{v})\hat{\epsilon}^{(1)}]\,\delta\beta, \tag{36b}$$

and

$$\delta\hat{\epsilon}^{(2)}(\mathbf{k}) = \frac{k_0}{|\mathbf{k}|}\,\frac{1}{v}\,[\hat{\epsilon}^{(1)}\times\mathbf{v} + (\hat{k}\cdot\mathbf{v})\hat{\epsilon}^{(2)}]\,\delta\beta. \tag{36c}$$

Lorentz boosts of the massive boson states are given by

$$\delta\,|\mathbf{k},(\pm)\rangle = \delta\beta\hat{\epsilon}^{(\pm)}\cdot\hat{v}\,\frac{M}{|\mathbf{k}|}\,|\mathbf{k},R\rangle \tag{37a}$$

and

$$\delta|\mathbf{k}, R\rangle = -\delta\beta\frac{M}{|\mathbf{k}|}[\hat{\epsilon}^{(+)}\cdot\hat{v}\,|\mathbf{k}, (-)\rangle + \hat{\epsilon}^{(-)}\cdot\hat{v}\,|\mathbf{k}, (+)\rangle], \tag{37b}$$

and in this case describe the transformations among the $+1$, $-1$, and 0 helicity states. As $M/|\mathbf{k}|$ becomes smaller, the amount of mixing of $+1$, 0, and $-1$ helicities in a Lorentz transformation decreases. This is consistent with the fact that there is no mixing at all in the case of massless photons. However, in the process of establishing the normalized state vectors for massive particle states, division by both $M$ and $|\mathbf{k}|$ may be equated to zero in Eq. (37) with safety.

[1]S. N. Gupta, Proc. Phys. Soc. (London) 63, 681 (1950); K. Bleuler, Helv. Phys. Acta 23, 567 (1950).
[2]See, for example, S. Gasiorowicz, *Elementary Particle Physics* (Wiley, New York, 1966), Chap. 4.
[3]A number of proofs have been given that the Gupta—Bleuler subsidiary condition is covariant. A very early one is: F. Belifante, Phys. Rev. 96, 780 (1954); A recent discussion appears in: F. Strocchi, "Locality and Covariance in QED and Gravitation. General Proof of Gupta-Bleuler Type Formulations," in *Lectures in Theoretical, Vol. XIVB*, edited by W. E. Brittin (Colorado Associated Univ. Press, Boulder, Colorado, 1973). The emphasis in the present work is not on proving covariance but on examining how the explicit Lorentz transformation of $A_Q$ remains consistent with covariance.
[4]H. E. Moses, J. Math. Phys. 9, 2039 (1968); 9, 16 (1968); 8, 1134 (1967).
[5]K. Haller and R. Sohn, Phys. Rev. D 14, 479 (1976).
[6]B. Lautrup, Kgl. Danske Videnskab. Selskab, Mat.—Fys. Medd. 35, No. 11, (1967); N. Nakanishi, Progr. Theor. Phys. 35, 1111 (1966); 38, 881 (1967); S. P. Tomczak and K. Haller, Nuovo Cimento B 8, 1 (1972).
[7]K. Haller, Acta Physica Austr. 42, 163 (1975); S. P. Tomczak and K. Haller, Nuovo Cimento B 8, 1 (1972).
[8]See, for example, N. Nakanishi, Progr. Theor. Phys. Suppl. 51, 1 (1972); K. L. Nagy, *State Vector Spaces with Indefinite Metric in Quantum Field Theory* (Noordhoff, Groningen, The Netherlands, 1960).
[9]K. Haller, Phys. Rev. D 8, 1796 (1973).
[10]A. Proca, J. Phys. Radium 7, 347 (1936); see also K. Nishijima, *Fields and Particles* (Benjamin, New York, 1969).

# Solution of a second-order integro–differential equation which occurs in laser modelocking[a]

P. L. Hagelstein[b]

*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*
*and Lawrence Livermore Laboratory, Livermore, California 94550*
(Received 5 December 1977)

We solve the following integro–differential equation for the eigenfunction $u(x)$:
$su(x)=\{d^2/dx^2-[1-6\operatorname{sech}^2(x)]\}u(x)-\epsilon\int_{-\infty}^{\infty}u(x')\operatorname{sech}(x')dx'\,(1+\nu^2d^2/dx^2)\operatorname{sech}(x)$,
where $s$ is the eigenvalue and $\epsilon$ and $\nu$ are arbitrary parameters which need not be small. This equation occurs in laser modelocking theory in the analysis of pulse stability, and $s$ is proportional to the rate of growth of perturbations. We expand the eigenfunction $u(x)$ in terms of a convenient basis set $\Lambda(x,k)$ satisfying
$-k^2\Lambda(x,k)=[d^2/dx^2+6\operatorname{sech}^2(x)]\Lambda(x,k)$.
We find two discrete eigenfunctions $u_0(x)$ and $u_1(x)$ and a continuum $u(x,s)$. We find that the lowest eigenvalue $s_0(\epsilon)$ is $-3$ at $\epsilon=0$ for finite or zero $\nu$, and that the point where $s_0=0$ the parameters $\epsilon$ and $\nu$ obey
$\epsilon=2/(1-\nu^2)$.
This is the zero growth point at which no eigenfunction $u(x)$ has a positive eigenvalue $s$.

## I. INTRODUCTION

In this paper we review the solution of a second order integro–differential eigenvalue equation which occurs in the theory of saturable absorber laser modelocking.[1] The modelocking of a laser occurs via temporal modulation of the electric field within the laser cavity to produce short pulses of light. In passive modelocking, the modulating element can be a fast saturable absorber which sharpens the pulse upon each passage through it by absorbing the wings of the pulse preferentially to the pulse maximum (the pulse maximum bleaches the absorber and sees less loss than the wings). The active gain medium has only a finite bandwidth which broadens the pulse upon each transit. A steady state operating condition is reached when the nonlinearity of the absorber balances the dispersion due to limited bandwidth in the gain medium.

The modelocking equations have been solved in closed form to yield pulse solutions and operating parameters for modelocked laser systems. The stability of the steady state pulses was a question of some interest because experimentally passively modelocked systems were often unstable, and before the recent theoretical work there were no guidelines by which one could design a stable system from the basic system parameters. The present paper is motivated by the stability analysis.

The linearized equation of motion for perturbations of the steady state pulse can be reduced in the case of well-separated pulses to the following integro–differential eigenvalue equation,

$$su(x)=\left(\frac{d^2}{dx^2}-[1-6\operatorname{sech}^2(x)]\right)u(x)-\epsilon\int_{-\infty}^{\infty}u(x')\operatorname{sech}(x')\,dx'\times\left(1+\nu^2\frac{d^2}{dx^2}\right)\operatorname{sech}(x),\tag{1}$$

where $s$ is the eigenvalue and $\epsilon$ and $\nu$ are arbitrary parameters which need not be small. The different terms in (1) can be assigned physical significance, the eigenvalue $s$ gives the growth rate of perturbations from pass to pass, and therefore conditions under which $s$ cannot be positive are of interest. The second derivative is a "diffusive" operator due to the limited bandwidth in frequency of the active gain medium and the "potential well" term is the normalized loss seen by the perturbation. The integral term is due to the additional saturation of the gain medium which is caused by the perturbation.

In the following solution of (1), no restrictions are placed on $\epsilon$ and the final results are valid even for complex $\epsilon$. The method of solution is to expand the eigenfunctions $u(x)$ in terms of an orthonormal set of functions satisfying the potential well eigenvalue equation (a degenerate Lamé equation)

$$-k^2\Lambda(x,k)=\left(\frac{d^2}{dx^2}+6\operatorname{sech}^2x\right)\Lambda(x,k).\tag{2}$$

The eigenfunctions $u(x)$ reduce to $\Lambda(x,k)$ when the parameter $\epsilon$ is zero in which case there are two discrete eigenfunctions and a continuum of eigenfunctions. As $\epsilon$ becomes nonzero, one finds that (1) retains two eigenfunctions and a continuum, and that of the spectra of eigenvalues, only the lowest one varies with $\epsilon$ and $\nu$.

In Sec. II we present the solutions to (2) and in the following two sections we solve for the eigenfunctions of (1) in the special case of $\nu$ equal to zero. This is done for the sake of clarity, since in this case the equations are simpler and hence more transparent. In Sec. V we treat the case of artitrary $\nu$.

[a]Work performed in part under the auspices of the U.S. Department of Energy, Contract No. W-7405-Eng-48.
[b]Work at MIT supported in part by the Fannie & John Hertz Foundation.

 0022-2488/78/1907-1594$1.00  

## II. SOLUTIONS OF THE DEGENERATE LAME EQUATION

The solutions of (2) are given in Refs. 2 and 3 and can be written:

$$\Lambda_0(x)=\sqrt{3/4}\ \mathrm{sech}^2x \quad (-k^2=4), \tag{3}$$

$$\Lambda_1(x)=\sqrt{3/2}\ \mathrm{sech}x\tanh x \quad (-k^2=1), \tag{4}$$

$$\Lambda(x,k)=\exp(ikx)\left[1-\frac{6}{1+ik}\left(\frac{e^x}{e^x+e^{-x}}\right)+\frac{12}{(1+ik)(2+ik)}\left(\frac{e^x}{e^x+e^{-x}}\right)^2\right] \quad (k^2\geq 0), \tag{5}$$

where the normalization is

$$\int_{-\infty}^{\infty}\Lambda_0^2(x)\,dx=\int_{-\infty}^{\infty}\Lambda_1^2(x)\,dx=1, \tag{6}$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Lambda(x,k)\Lambda^*(x,k')\,dx\,dk=1. \tag{7}$$

These functions constitute a complete orthonormal basis set. Completeness of the solutions of Sturm–Liouville equations is discussed in Refs. 3, 4, and 5 and the normalization follows from the evaluations of the integrals in (6) and (7). The main contribution of the integral in (7) comes from the wings in which the continuum eigenfunctions become complex exponentials, and so the normalization is similar to the case of the Fourier transform.

## III. SOLUTION FOR THE DISCRETE EIGENFUNCTIONS FOR $\nu=0$

There are two discrete eigenfunction solutions $u_0(x)$ and $u_1(x)$ to Eq. (1), the latter of which can be assigned immediately

$$u_1(x)=\Lambda_1(x)=\sqrt{3/2}\ \mathrm{sech}x\tanh x, \quad s_1=0, \tag{8}$$

since $u_1(x)$ is odd and its overlap integral with the hyperbolic secant is equal to zero. The remaining discrete eigenfunction can be found by assuming a solution of the form

$$u(x)=U_0\Lambda_0(x)+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}U(k)\,\Lambda(x,k)\,dk. \tag{9}$$

When the parameter $\nu$ is zero, Eq. (1) reduces to

$$su(x)=\left(\frac{d^2}{dx^2}-(1-6\,\mathrm{sech}^2x)\right)u(x)-\epsilon\,\mathrm{sech}x\int_{-\infty}^{\infty}u(x')\,\mathrm{sech}(x')\,dx', \tag{10}$$

which we shall consider in this and the following section in detail. Inclusion of the additional term when $\nu$ is nonzero presents no additional difficulties, as is shown in Sec. V.

Upon insertion of the solution (9) into (10), we obtain

$$(s_0-3)U_0\Lambda_0(x)+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(s_0+k^2+1)U(k)\Lambda(x,k)\,dk=-\epsilon\alpha_0\,\mathrm{sech}x, \tag{11}$$

where $\alpha_0$ is the overlap integral

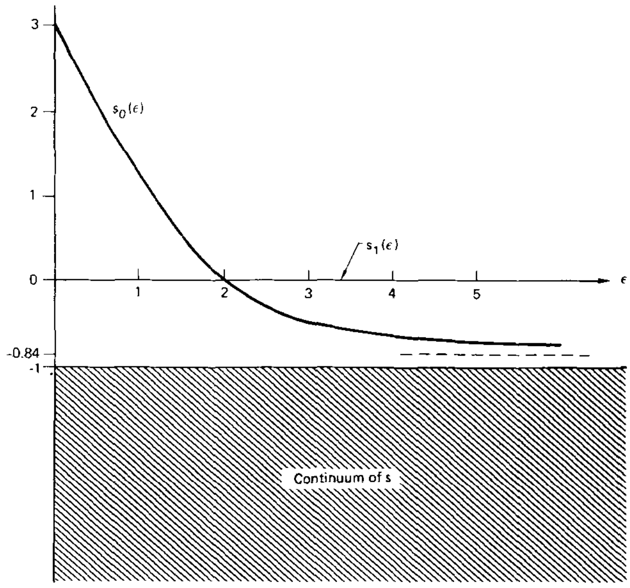



FIG. 1. Eigenvalues $s$ as a function of parameter $\epsilon$.

$$\alpha_0=\int_{-\infty}^{\infty}u_0(x)\,\mathrm{sech}x\,dx \tag{12}$$

which must be found self-consistently with $u_0(x)$ later on. The transform of the hyperbolic secant is

$$S_0=\int_{-\infty}^{\infty}\mathrm{sech}x\,\Lambda_0^*(x)\,dx=\frac{\sqrt{3}\,\pi}{4}, \tag{13}$$

$$S(k)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{sech}x\,\Lambda^*(x,k)\,dx=\sqrt{\pi/8}\,\mathrm{sech}(\pi k/2)\left(\frac{1+ik}{2-ik}\right), \tag{14}$$

where the integral in (14) is evaluated in Appendix A. Using the results (13) and (14), Eq. (11) becomes

$$(s_0-3)U_0\Lambda_0(x)+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(s_0+k^2+1)U(k)\Lambda(x,k)\,dk=-\epsilon\alpha_0\left[S_0\Lambda_0(x)+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}S(k)\Lambda(x,k)\,dk\right]. \tag{15}$$

Using orthogonality, it follows from (15) that

$$(s_0-3)U_0=-\epsilon\alpha_0S_0, \tag{16}$$

$$[s_0+k^2+1]U(k)=-\epsilon\alpha_0S(k). \tag{17}$$

One observes that the lowest eigenvalues $s_0$ is 3 only when the parameter $\epsilon$ is zero in which case the eigenfunction $u_0(x)$ is simply $\Lambda_0(x)$. The quantity $(s_0+k^2+1)$ is never zero for a discrete eigenfunction. However, it can be zero for continuum eigenfunctions, a case to be considered in the next section.

The lowest eigenfunction is

$$u_0(x)=-\frac{\epsilon\alpha_0S_0\Lambda_0(x)}{s_0-3}-\frac{\epsilon\alpha_0}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{S(k)\Lambda(x,k)}{s_0+k^2+1}\,dk, \tag{18}$$

where $\alpha_0$ must now be determined self-consistently with $u_0(x)$. Using (12), one obtains

$$1+\epsilon\left(\frac{S_0S_0^*}{s_0-3}+\int_{-\infty}^{\infty}\frac{S(k)S^*(k)}{s_0+k^2+1}\,dk\right)=0. \tag{19}$$

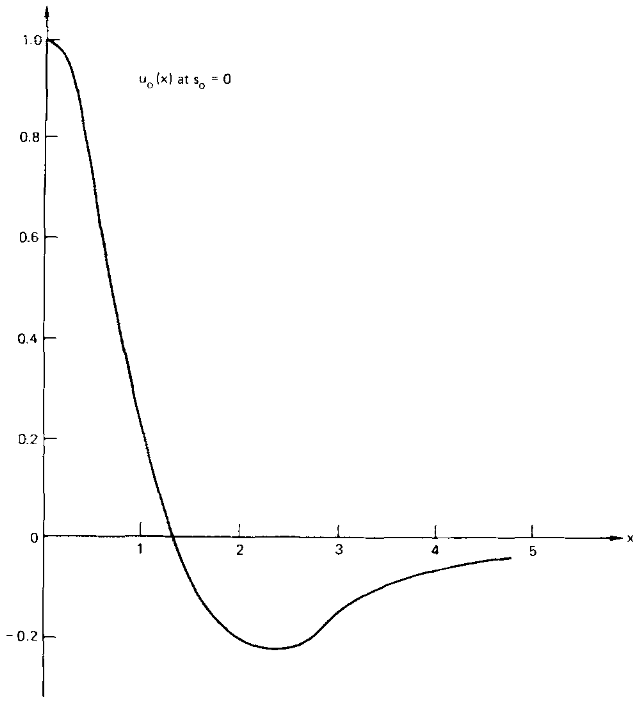



FIG. 2. Even eigenfunction $u_0(x)$ at $s_0 = 0$.

When the parameter $\epsilon$ is much less than unity, the eigenvalue $s_0$ is positive and nearly equal to 3. As $\epsilon$ increases, $s_0$ decreases and approaches $-0.84$ asymptotically as shown in Fig. 1. Of interest in the stability analysis is the value of $\epsilon$ at which $s_0$ becomes equal to zero, which can be found analytically by using the following results:

$$\frac{S_0 S_0^*}{s_0 - 3} \underset{s_0 \to 0}{\to} -\frac{\pi^2}{16} \tag{20}$$

$$\int_{-\infty}^{\infty} \frac{S(k)S^*(k)}{s_0 + k^2 + 1}\, dk \underset{s_0 \to 0}{\to} \frac{\pi}{8} \int_{-\infty}^{\infty} \frac{\mathrm{sech}^2(\pi k/2)}{4 + k^2}\, dk = \frac{\pi^2}{16} - \frac{1}{2}, \tag{21}$$

where the integral in (21) is evaluated in Appendix B. The required value of the parameter $\epsilon$ at which $s_0$ becomes equal to zero is found to be

$$\epsilon(s_0 = 0) = 2 \tag{22}$$

in the case of the parameter $\nu$ equal to zero. This is the boundary between stability and instability of the perturbation obeying (1) (for $\nu = 0$). When $\epsilon = 2$, the two discrete eigenfunctions $u_0(x)$ and $u_1(x)$ are degenerate and one can construct eigenfunctions which are a linear combination of $u_0(x)$ and $u_1(x)$.

In Fig. 2 we show $u_0(x)$ at $\epsilon = 2$ where $s_0 = 0$. The negative lobe of the function around $x = 2.5$ is due to the term proportional to $\epsilon\,\mathrm{sech}x$ in (1) (note that it is missing completely when $\epsilon = 0$) and grows larger the larger $\epsilon$ becomes.

## IV. SOLUTION FOR THE CONTINUUM EIGENFUNCTIONS

Assuming a solution of (1) of the form (9) fails for the continuum eigenfunctions because as defined $U(k)$ is singular on the real $k$ axis and the integral in (9) becomes undefined. Also, since (1) is being treated as an inhomogeneous equation (the integral is a constant found self-consistently with the eigenfunction solution), the solution in general is composed of a particular solution and a homogeneous solution which can be added in order to satisfy boundary conditions or constraints. In the previous section no mention was made of homogeneous solutions because in general they blew up asymptotically. These two problems are intimately related; the singularities on the real axis correspond to the homogeneous solutions.

One therefore assumes a solution to (1) of the form

$$\begin{aligned} u(x, s) &= U_0(s)\Lambda_0(x) + \frac{1}{\sqrt{2\pi}} \fint_{-\infty}^{\infty} U(s, k)\Lambda(x, k)\, dk \\ &\quad + \frac{U_+(s)}{\sqrt{2\pi}} \Lambda[x, \sqrt{-(s+1)}] + \frac{U_-(s)}{\sqrt{2\pi}} \Lambda[x, -\sqrt{-(s+1)}], \end{aligned} \tag{23}$$

where the functions $U_+(s)$ and $U_-(s)$ determine the amount of homogeneous solutions to be added to the particular solution and the principal value integral is used in anticipation of the singularities of $U(k)$ which occur at $s + k^2 + 1 = 0$. Upon substitution of (23) into (10), one obtains

$$\begin{aligned} (s - 3)U_0(s)\Lambda_0(x) &+ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (s + k^2 + 1)U(s, k)\Lambda(x, k)dk \\ &= -\epsilon\alpha(s)\left(S_0\Lambda_0(x) + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} S(k)\Lambda(x, k)\, dk\right), \end{aligned} \tag{24}$$

where the overlap integral $\alpha(s)$ is defined by

$$\alpha(s) = \int_{-\infty}^{\infty} \mathrm{sech}x\, u(x, s)\, dx \tag{25}$$

and the principal value integral is replaced by a Riemann integral since the singularities in the integrand have been removed by the factor $(s + k^2 + 1)$ under the integral. By orthonormality of the basis functions, one has from (24)

$$U_0(s) = -\frac{\epsilon\alpha(s)S_0}{s - 3}, \tag{26}$$

$$U(s, k) = -\frac{\epsilon\alpha(s)S(k)}{s + k^2 + 1}, \tag{27}$$

from which the continuum eigenfunctions are

$$\begin{aligned} u(x, s) &= -\frac{\epsilon\alpha(s)S_0\Lambda_0(x)}{s - 3} - \frac{\epsilon\alpha(s)}{\sqrt{2\pi}} \fint_{-\infty}^{\infty} \frac{S(k)\Lambda(x, k)}{s + k^2 + 1}\, dk \\ &\quad + \frac{U_+(s)}{\sqrt{2\pi}} \Lambda[x, \sqrt{-(s+1)}] + \frac{U_-(s)}{\sqrt{2\pi}} \Lambda[x, -\sqrt{-(s+1)}]. \end{aligned} \tag{28}$$

The determinantal equation for the eigenvalue $s$ is found by requiring $\alpha(s)$ to be determined self-consistently with $u(x, s)$,

$$\alpha(s)\left(1+\frac{\epsilon S_0 S_0^*}{s-3}+\epsilon \oint_{-\infty}^{\infty}\frac{S(k)S^*(k)}{s+k^2+1}dk\right)$$
$$=U_+(s)S^*[\sqrt{-(s+1)}]+U_-(s)S^*[-\sqrt{-(s+1)}], \quad (29)$$

which can be satisfied for all $s \leq -1$ since $U_+(s)$ and $U_-(s)$ are arbitrary.

## V. SOLUTION FOR THE LOWEST EIGENFUNCTION IN THE CASE OF FINITE $\nu$

When $\nu$ is finite, the procedure for constructing the lowest eigenfunction $u_0(x)$ is very similar to that used in Sec. III in the case of $\nu$ equal to zero. As before, we assume a solution of the form (9), which is

$$u(x)=U_0\Lambda_0(x)+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}U(k)\Lambda(x,k)\,dk, \quad (30)$$

which upon substitution into (1) yields

$$(s_0-3)U_0\Lambda_0(x)+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(s_0+k^2+1)U(k)\Lambda(x,k)\,dk$$
$$=-\epsilon\alpha_0\left(1+\nu^2\frac{d^2}{dx^2}\operatorname{sech}(x)\right)$$
$$=-\epsilon\alpha_0[(1+\nu^2)\operatorname{sech}(x)-2\nu^2\operatorname{sech}^3(x)], \quad (31)$$

where we have expanded the second derivative of the hyperbolic secant. The new result which is required is the transform of $\operatorname{sech}^3(x)$ which can be found to be

$$Q_0=\int_{-\infty}^{\infty}\operatorname{sech}^3(x)\Lambda_0^*(x)\,dx=\frac{\sqrt{3}\,3\pi}{16}, \quad (32)$$

$$Q(k)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\operatorname{sech}^3(x)\Lambda^*(x,k)\,dx$$
$$=-\sqrt{\pi/2}\,\frac{(1+k^2)}{8}\left(\frac{1+ik}{2-ik}\right)\operatorname{sech}\left(\frac{\pi k}{2}\right), \quad (33)$$

where the integral in (33) is evaluated through a procedure similar to that described in Appendix A.

From here, one can write down the lowest eigenfunction following the steps in Sec. III to be

$$u(x)=-\epsilon\alpha_0\left(\frac{(1+\nu^2)S_0\Lambda_0(x)}{s_0-3}+\frac{(1+\nu^2)}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{S(k)\Lambda(x,k)}{s_0+k^2+1}dk\right.$$
$$\left.-\frac{2\nu^2Q_0\Lambda_0(x)}{s_0-3}-\frac{2\nu^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{Q(k)\Lambda(x,k)}{s_0+k^2+1}dk\right). \quad (34)$$

Requiring $\alpha_0$ to be determined self-consistently yields

$$1+\epsilon\left[(1+\nu^2)\left(\frac{S_0S_0^*}{s_0-3}+\int_{-\infty}^{\infty}\frac{S(k)S^*(k)}{s_0+k^2+1}dk\right)\right.$$
$$\left.-2\nu^2\left(\frac{Q_0S_0^*}{s_0-3}+\int_{-\infty}^{\infty}\frac{Q(k)S^*(k)}{s_0+k^2+1}dk\right)\right]=0. \quad (35)$$

This result is similar in form to the result (19) found in the case of $\nu$ equal to zero.

We wish now to evaluate (35) when $s_0$ is zero, in which case we require the following results:

$$\frac{Q_0S_0^*}{-3}=-\frac{1}{3}\left(\frac{\sqrt{3}\,3\pi}{16}\right)\left(\frac{\sqrt{3}\pi}{4}\right)=-\frac{3\pi^2}{64}, \quad (36)$$

$$\int_{-\infty}^{\infty}\frac{Q(k)S^*(k)}{s_0+k^2+1}dk=-\frac{\pi}{32}\int_{-\infty}^{\infty}\left(\frac{1+k^2}{4+k^2}\right)\operatorname{sech}^2\left(\frac{\pi k}{2}\right)dk$$
$$=\frac{3\pi^2}{64}-\frac{1}{2}. \quad (37)$$

Using (36) and (37) we obtain from (35)

$$\epsilon=\frac{2}{1-\nu^2} \quad (38)$$

which is the required result.

In the stability analysis the parameter $\nu$ is proportional to the ratio of the steady state pulse bandwidth and the linewidth of the amplifier, and is much less than unity wherever the original model is valid. The analysis here is valid, however, for arbitrary $\nu$.

The construction of continuum eigenfunctions if they are desired proceeds analogously with the method used in Sec. IV, and we shall not concern ourselves with this generalization.

## SUMMARY

The integro–differential eigenvalue equation (1) has been solved and two discrete eigenfunctions $u_0(x)$ and $u_1(x)$ were found as well as a continuum of eigenfunctions. A determinantal equation was derived for the lowest eigenvalue $s_0$ and solved analytically at the point where $s_0$ equals zero which was the result of interest in the stability analysis from which (1) originated. The result is a relation between $\epsilon$ and $\nu^2$ which is valid at the stability boundary and can be written as

$$\epsilon=\frac{2}{1-\nu^2} \quad (\text{at } s_0=0). \quad (39)$$

In the physical problem where (1) originated the parameter $\nu^2$ is less than unity, and so whenever the gain saturation parameter $\epsilon$ takes on values larger than $2/(1-\nu^2)$ (for $\nu^2<1$) then the system is not unstable.

In the case of $\nu$ equal to zero, the continuum eigenfunctions were constructed and a determinantal equation was derived for the continuum eigenvalues $s$, which has solutions for all $s\leq -1$. This result holds also for $\nu$ not equal to zero.

## ACKNOWLEDGMENTS

The author acknowledges helpful discussions with Professor H.A. Haus as well as the careful reading of the manuscript by Professor F. Chu at MIT. The work was supported at the Research Laboratory of Electronics by the Joint Services Electronics Program, Contract DAAB07-75-C-1346. P. L. Hagelstein was a Hertz Foundation Fellow. Use of the MACSYMA Program (Mathlab Group, Project MAC, MIT) was greatly appreciated. The work of the Mathlab group is supported by ARPA Work Order 2095 under ONR Contract N00014-75-C-0661.

## APPENDIX A

We consider the evaluation of the following integral,

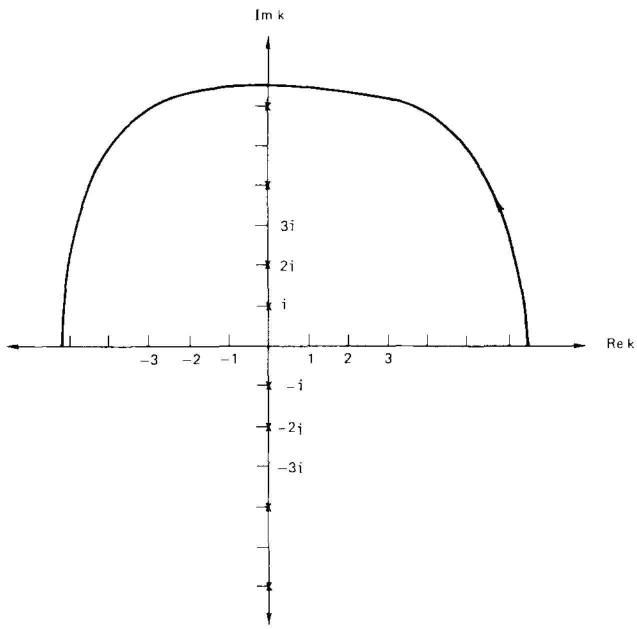



FIG. 3. Poles and contour integral for
$\frac{\pi}{8}\int_{-\infty}^{\infty} \frac{\mathrm{sech}^2(\pi k/2)}{k^2+4}\, dk.$

$$S(k)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx\, \mathrm{sech} x \exp(-ikx)\left[1-\frac{6}{1+ik}\left(\frac{e^x}{e^x+e^{-x}}\right)+\frac{12}{(1+ik)(2+ik)}\left(\frac{e^x}{e^x+e^{-x}}\right)^2\right]^*. \quad (A1)$$

The key step in the evaluation of (A1) is use of the identity[6]

$$\int_0^1 x^{\rho-1}(1-x)^{\sigma-1}{}_2F_1(\alpha,\beta\,|\,\gamma\,|\,x)\,dx = \frac{\Gamma(\rho)\Gamma(\sigma)}{\Gamma(\rho+\sigma)}\,{}_3F_2(\alpha,\beta,\rho\,|\,\gamma,\rho+\sigma\,|\,1), \quad (A2)$$

where

$$\mathrm{Re}\rho > 0, \quad \mathrm{Re}\sigma > 0, \quad \mathrm{Re}(\gamma+\sigma-\alpha-\beta) > 0. \quad (A3)$$

The hypergeometric function ${}_2F_1(\alpha,\beta\,|\,\gamma\,|\,x)$ can be written in terms of a series

$$\begin{aligned}{}_2F_1(\alpha,\beta\,|\,\gamma\,|\,x) &= 1+\frac{\alpha\beta}{\gamma}\frac{x}{1!}+\frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)}\frac{x^2}{2!}+\cdots \\ &+\frac{\alpha(\alpha+1)\cdots(\alpha+n-1)\beta(\beta+1)\cdots(\beta+n-1)}{\gamma(\gamma+1)\cdots(\gamma+n-1)} \\ &\times\frac{x^n}{n!}+\cdots. \end{aligned} \quad (A4)$$

The generalized hypergeometric function ${}_3F_2(\alpha,\beta,\gamma\,|\,\delta,\epsilon\,|\,x)$ can similarly be expressed as a series

$$\begin{aligned}{}_3F_2(\alpha,\beta,\gamma\,|\,\delta,\epsilon\,|\,x) &= 1+\frac{\alpha\beta\gamma}{\delta\epsilon}\frac{x}{1!}+\frac{\alpha(\alpha+1)\beta(\beta+1)\gamma(\gamma+1)}{\delta(\delta+1)\epsilon(\epsilon+1)}\frac{x^2}{2!}+\cdots \\ &+\frac{\alpha(\alpha+1)\cdots(\alpha+n-1)\beta(\beta+1)\cdots}{\delta(\delta+1)\cdots(\delta+n-1)} \\ &\times\frac{(\beta+n-1)\gamma(\gamma+1)\cdots(\gamma+n-1)}{\epsilon(\epsilon+1)\cdots(\epsilon+n-1)}\frac{x^n}{n!}+\cdots. \end{aligned} \quad (A5)$$

If one defines

$$(\alpha)_n=\alpha(\alpha+1)\cdots(\alpha+n-1), \quad (\alpha)_0=1, \quad (A6)$$

then these series can be written in the compact form

$${}_2F_1(\alpha,\beta\,|\,\gamma\,|\,x)=\sum_{n=0}^{\infty}\frac{(\alpha)_n(\beta)_n}{(\gamma)_n}\frac{x^n}{n!}, \quad (A7)$$

$${}_3F_2(\alpha,\beta,\gamma\,|\,\delta,\epsilon\,|\,x)=\sum_{n=0}^{\infty}\frac{(\alpha)_n(\beta)_n(\gamma)_n}{(\delta)_n(\epsilon)_n}\frac{x^n}{n!}. \quad (A8)$$

These functions are convenient here in that the notation becomes more compact and hopefully clearer. We note that the series terminates after $m$ steps if there is a negative integer, $-m$, in the first set of arguments of the function.

The summation in brackets in the integrand of (A1) can be rewritten as

$$\left[1-\frac{6}{1+ik}\left(\frac{e^x}{e^x+e^{-x}}\right)+\frac{12}{(1+k)(2+ik)}\left(\frac{e^x}{e^x+e^{-x}}\right)^2\right]^* = {}_2F_1\left(3,-2\,|\,1-ik\,|\,\frac{e^x}{e^x+e^{-x}}\right). \quad (A9)$$

The substitution

$$z=\frac{e^x}{e^x+e^{-x}} \quad (A10)$$

brings (A1) to the form of (A2), namely

$$S(k)=\frac{1}{\sqrt{2\pi}}\int_0^1 z^{(1-ik)/2}(1-z)^{(1+ik)/2} \times {}_2F_1(3,-2\,|\,1-ik\,|\,z)\frac{dz}{z(1-z)}. \quad (A11)$$

The following identifications are made:

$$\rho=\left(\frac{1-ik}{2}\right), \quad (A12)$$
$$\sigma=\left(\frac{1+ik}{2}\right),$$
$$\alpha=3,$$
$$\beta=-2,$$
$$\gamma=1-ik,$$

to yield the result

$$S(k)=\frac{1}{\sqrt{2\pi}}\left[\Gamma\left(\frac{1-ik}{2}\right)\Gamma\left(\frac{1+ik}{2}\right)\right]\Big/\Gamma(1) \times {}_3F_2\left(3,-2,\frac{1-ik}{2}\,|\,1-ik,1\,|\,1\right). \quad (A13)$$

The series terminates after two terms and with the identity for the gamma function,

$$\Gamma\left(\frac{1-ik}{2}\right)\Gamma\left(\frac{1+ik}{2}\right)=\pi\,\mathrm{sech}\left(\frac{\pi k}{2}\right), \quad (A14)$$

one obtains after some algebra

$$S(k)=\sqrt{\pi/8}\,\mathrm{sech}(\pi k/2)\left(\frac{1+ik}{2-ik}\right), \quad (A15)$$

which is the required result.

## APPENDIX B

In this appendix we consider the integral

$$I=\frac{\pi}{8}\int_{-\infty}^{\infty}\frac{\operatorname{sech}^2(\pi k/2)\,dk}{4+k^2} \tag{B1}$$

which can be evaluated simply by calculus of residues. The poles lie along the imaginary $k$ axis as shown in Fig. 3 and by closing the contour about either the upper or lower half plane we obtain

$$I=2\pi i\sum \operatorname{Res}\left(\frac{\pi}{8}\frac{\operatorname{sech}^2(\pi k/2)}{4+k^2}\right). \tag{B2}$$

The residue at $k=\pm 2i$ is due to a first order pole and is

$$\operatorname{Res}(\pm 2i)=\frac{\pi}{32i}. \tag{B3}$$

The residues at $k=\pm ni$ (for $n$ odd) are due to second order poles and are

$$\operatorname{Res}(\pm ni)=\frac{1}{2\pi}\left[\frac{2ni}{(4-n^2)^2}\right]. \tag{B4}$$

The integral can therefore be reduced to

$$I=\frac{\pi^2}{16}-\sum_{\substack{n_{\text{odd}}\\ n>0}}\frac{2n}{(4-n^2)^2}. \tag{B5}$$

This summation turns out to be trivial,

$$\begin{aligned}\sum_{n_{\text{odd}}}\frac{2n}{(4-n^2)^2}&=\frac{1}{4}\sum_{n_{\text{odd}}}\left[\frac{1}{(2-n)^2}-\frac{1}{(2+n)^2}\right]\\&=\frac{1}{4}\left[1+1+\frac{1}{3^2}+\frac{1}{5^2}+\cdots-\frac{1}{3^2}-\frac{1}{5^2}-\cdots\right]\\&=\frac{1}{2},\end{aligned} \tag{B6}$$

giving the required result

$$I=\frac{\pi^2}{16}-\frac{1}{2}. \tag{B7}$$


[1] H. A. Haus, "Theory of Modelocking with a Fast Saturable Absorber," J. Appl. Phys. **46**, 3049 (1975).
[2] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, New York, 1975).
[3] P. M. Morse and H. Feshback, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[4] F. B. Hildebrand, *Advanced Calculus for Applications* (Prentice-Hall, Englewood Cliffs, N.J., 1962).
[5] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953).
[6] I. S. Gradshteyn and I. W. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1975).

# A theory of classical limit for quantum theories which are defined by real Lie algebras

Kai Drühl

*Max-Planck-Institut zur Erforschung der Lebensbedingungen der wissenschaftlich-technischen Welt, D-813 Starnberg, Germany*

(Received 16 January 1978)

A theory of classical limit is developed for quantum theories, the basic observables of which correspond to elements in some real Lie algebra $L_0$. For both quantum and classical systems based on $L_0$ the basic observables are contained in a unique universal algebra. This is the universal enveloping algebra $\mathfrak{U}$ for the quantum case, and a universal commutative Poisson algebra $\mathfrak{B}$ for the classical case. $\mathfrak{U}$ and $\mathfrak{B}$ are connected by a system of contraction maps. For certain sequences of representations and of vector states defined by them renormalized expectation values of the quantum variables are shown to converge to values of the corresponding classical variables at some point in the classical phase space. The classical phase space is obtained as a limit of certain systems of coherent states. The general theory is illustrated by several examples and counterexamples.

## 1. INTRODUCTION

In this paper we develop a rigorous and general theory of classical limit for a large class of quantum theories. Heisenberg's discussion of this limit[1] is based on the canonical commutation relations for pairs of conjugate observables $P$ and $Q$:

$$[P,Q]=P\cdot Q-Q\cdot P=-i, \quad i\cdot i=-1, \tag{1.1}$$

in microscopic units where $\hbar=1$.

Heisenberg's uncertainty relations then give a nonvanishing lower bound for the dispersions of the basic observables in any arbitrary state,

$$\Delta^2(Q)\cdot\Delta^2(P)\geq\tfrac{1}{4}, \quad \Delta^2(A)=(f,A^2f)-(f,A,f)^2, \tag{1.2}$$

which is derived from (1.1). In the limit of large quantum numbers, i.e., for states where the expectation values of the basic observables are large compared to their dispersions, the expectations values of the physical quantities may then be replaced by their classical value. At the same time the commutator relation (1.1) is replaced by the Poisson bracket for the corresponding classical quantities $\tilde{P}$ and $\tilde{Q}$:

$$Q\to\tilde{Q}, \quad P\to\tilde{P}, \quad \{\tilde{P},\tilde{Q}\}=1 \tag{1.3}$$

(see Ref. 2 for a more recent analysis of classical limit based on canonical commutation relations).

We wish to emphasize here that the notion of classical limit and, in particular, the existence of a Poisson bracket for the classical quantities is in no way restricted to theories defined in terms of canonical commutation relations. A careful analysis of more general situations seems worthwhile, for example, with a view to relativistic particle theories, for which the analysis of classical limit continues to be of interest (see, e.g., Ref. 3 and 4).

For the class of theories we shall study here the basic observables are assumed to correspond to elements in a finite dimensional, real Lie algebra $L_0$.

By a quantum theory based on $L_0$ we mean a representation $\rho$ of $L_0$ by antisymmetric, linear operators on a dense linear subspace $D$ of some Hilbert space $H$, with essentially self adjoint operators $i\rho(X)$, $X$ in $L_0$, which are the basic quantum observables.

By a classical theory based on $L_0$ we mean a linear map $P$ of $L_0$ to the algebra of smooth functions on some canonical manifold $\Gamma$ such that

$$\{P(X),P(Y)\}=P([X,Y]), \tag{1.4}$$

where $\{\cdots,\cdots\}$ is the Poisson bracket on $\Gamma$.[5] The functions $P(X)$ then are the basic classical observables.

For both types of observables there exist universal algebras containing them. For the quantum case this is the universal enveloping algebra $\mathfrak{U}$ of $L_0$ (resp. of its complexification $L$), while for the classical case this is the associated Poisson algebra $\mathfrak{B}$. Our discussion of the classical limit is based on a certain algebraic relation between both algebras which may be considered a contraction analogous to notion of contraction for Lie algebras[6] [In Refs. 7 and 8 the inverse process of deformation of the Poisson algebra of classical observables is studied as a means of quantization. In fact the canonical linear map $\Psi:\mathfrak{B}\to\mathfrak{U}$ and its inverse discussed in Sec. 3 always lead to a deformation of $\mathfrak{B}$ in the sense of Refs. 7 and 8).]

In Sec. 2 we briefly state the generalization of Heisenberg's uncertainty relations appropriate for our case.

In Sec. 3 we define the algebra $\mathfrak{U}$ and a set of Poisson algebras isomorphic to $\mathfrak{B}$, discuss their universal properties and their algebraic relations which establish $\mathfrak{B}$ as a classical limit of $\mathfrak{U}$ on this purely algebraic level.

In Sec. 4 we consider sequences of representations of $\mathfrak{U}$ which can be contracted to a realization of $\mathfrak{B}$ by functions on a certain canonical manifold (phase space). This phase space appears as the limit of certain systems of coherent states in the sense of Ref. 9.

In Sec. 5 we give some examples and applications of our general theory and state a counterexample.

 0022-2488/78/1907-1600$1.00  

## 2. UNCERTAINTY RELATIONS

Let $\mathcal{H}$ be a complex Hilbert space with scalar product $(\cdots,\cdots)$ and $f$ a unit vector in $\mathcal{H}:(f,f)=1$. For any (unbounded) linear operator $A$ on $\mathcal{H}$ such that $f$ is in the domain $D(A)$ of definition of $A$ we write

$$\omega_f(A)=(f,Af) \tag{2.1}$$

for the expectation value of $A$ in the state $\omega_f$.

*Proposition* 1: Let $A$, $B$ be any two symmetric linear operators on $\mathcal{H}$, and $f$ be a unit vector contained in the domains of definition $D(A\circ A)$, $D(B\circ B)$, $D(A\cdot B)$, and $D(B\circ A)$. Then

$$\omega_f(AA)\cdot\omega_f(BB)\geq\tfrac{1}{4}\left|\omega_f(A\cdot B-B\cdot A)\right|^2, \tag{2.2}$$

and Eq. (2.2) is an equality if and only if

$$(A+irB)f=0 \text{ for some real number } r. \tag{2.3a}$$

In this case

$$\omega_f(A\cdot A)=-\tfrac{1}{2}i\omega_f(A\circ B-B\cdot A)\cdot r. \tag{2.3b}$$

*Proof*: For any real $t$ the vector $(A+itB)f$ is in $D(A)$ and $D(B)$. However $(A-itB)^*\supset A+itB$ implies that the real quadratic function,

$$\begin{aligned}p(t)&=\omega_f((A-itB)(A+itB))\\&=\omega_f(A\cdot A)+ti\omega_f(A\circ B-B\circ A)+t^2\omega_f(B^2)\geq 0\end{aligned}$$

is not negative for any $t$, which implies Eqs. (2.2) and (2.3). Q.E.D.

If we replace the operators $A$ and $B$ by $A_0=A-\omega_f(A)\cdot 1$ resp. $B_0$ the right hand side of Eq. 2.2 does not change and we obtain from Eq. (2.2) a lower bound for the product of the statistical mean-square deviations or dispersions:

$$\begin{aligned}&\Delta^2(A)=\omega_f(A\circ A)-\omega_f(A)^2,\\&\Delta^2(A)\cdot\Delta^2(B)\geq\tfrac{1}{4}\left|\omega_f(A\circ B-B\circ A)\right|^2.\end{aligned} \tag{2.3}$$

In particular these cannot vanish simultaneously in the state $\omega_f$ if $\omega_f(A\circ B-B\circ A)\neq 0$. For the case where $A,B$ satisfy "canonical commutation relations" inequality (2.3) is just Heisenberg's famous uncertainty relation. [1]

## 3. ENVELOPING ALGEBRAS AND POISSON ALGEBRAS

Let $L_0$ be a real Lie algebra and $L=L_0\oplus iL_0$ its complexification. If $\mathfrak{B}$ is a complex, associative algebra we call $\mathfrak{B}$ representation of $L$ a linear map $\rho:L\to\mathfrak{B}$ such that

$$\rho(X)\circ\rho(Y)-\rho(Y)\circ\rho(X)=\rho([X,Y]),\quad X,Y,[X,Y] \text{ in } L. \tag{3.1}$$

In particular, $\mathfrak{B}$ may be an algebra of linear operators on some complex vectorspace $V$. Denote by $\mathfrak{T}=\mathfrak{T}(L)$ the complex tensor algebra over $L$ with canonical map $\iota:L\to\mathfrak{T}(L)$, and consider the two sided ideal $\mathfrak{J}$ generated in $\mathfrak{T}$ by elements of the form

$$\iota(X)\circ\iota(Y)-\iota(Y)\circ\iota(X)-\iota([X,Y]),\quad X,Y \text{ in } L. \tag{3.2}$$

The quotient algebra $\mathfrak{U}=\mathfrak{U}(L)$ of $\mathfrak{T}$ by $\mathfrak{J}$ is called the universal enveloping algebra of $L$. There is a linear map $\hat{\iota}:L\to\mathfrak{U}$ defined by $\iota$. We summarize the most important properties of $\hat{\iota}:L\to\mathfrak{U}$ in the following proposition which is a statement of well known results (see, e.g., Ref. 10, Chap. 2, Ref. 11, Chap. 12.)

*Proposition* 2:

(I) The map $\hat{\iota}$ is an injective $\mathfrak{U}$ representation of $L$.

(II) For any $\mathfrak{B}$ representation $\rho$ of $L$ there is a unique homomorphism $\hat{\rho}:\mathfrak{U}\to\mathfrak{B}$ of associative algebras such that

$$\rho=\hat{\rho}\cdot\hat{\iota}.$$

(III) Let $n$ be a positive integer, and denote by $\mathfrak{U}^n$ the linear subspace of $\mathfrak{U}$ generated by all elements of the form

$$Z\circ 1,\hat{X}_1\circ\cdots\circ\hat{X}_m,\hat{X}_r=\hat{\iota}(X_r),X_r \text{ in } L,\quad Z \text{ in } \mathbb{C},\quad m\leq n.$$

Then

$$W^r\circ W^s \text{ is in } \mathfrak{U}^{r+s},$$

and

$$W^r\circ W^s-W^s\cdot W^r \text{ is in } \mathfrak{U}^{r+s-1} \text{ for } W^i \text{ in } \mathfrak{U}^i,\quad i=r,s.$$

(IV) The bilinear bracket operation: $V,W\to[V,W]=V\circ W-W\cdot V$ satisfies:

$$[V,W]=-[W,V], \tag{3.3a}$$

$$[V,W\cdot W']=[V,W]\circ W'+W\circ[V,W'], \tag{3.3b}$$

$$[V_1,[V_2,V_3]]+\text{cycl.}=0. \tag{3.3c}$$

(V) There exists a unique antilinear map $\sigma:\mathfrak{U}\to\mathfrak{U}$ such that:

(i) $\sigma\circ\sigma=\text{id},\quad\sigma[V+(a+ib)W]=\sigma(V)+(a-ib)\sigma(W),$

(ii) $\sigma(V\cdot W)=\sigma(W)\circ\sigma(V),\quad V,W \text{ in } U,\quad a,b \text{ in } \mathbb{R},$

(iii) $\sigma(\hat{X})=-\hat{X},\quad X \text{ in } L_0.$

We write $T^*$ for $\sigma(T)$.

An element $T$ in $\mathfrak{U}$ is called symmetric resp. antisymmetric if $T^*=T$ resp. $T^*=-T$. It follows from (V iii) alone that

$$[X^*,Y^*]^*=-[X,Y] \text{ for } X,Y \text{ in } L, \tag{3.4}$$

since $X,Y\to[X^*,Y^*]^*$ is bilinear and this equation holds on the real Lie subalgebra $L_0$.

The algebra $\mathfrak{U}$ is in fact uniquely characterized (up to ismorphism) by the existence of a $\mathfrak{U}$ representation satisfying Proposition 3(II). [11] This is the reason for calling it universal.

Now let $\mathfrak{B}$ be a complex, commutative associative algebra, $\mathfrak{B}$ is called a Poisson algebra if there exists a bilinear map $:A,B\to\{A,B\}$ of $\mathfrak{B}\times\mathfrak{B}$ to $\mathfrak{B}$ such that:

$$\{A,B\}=-\{B,A\}, \tag{3.5a}$$

$$\{A,B\circ B'\}=\{A,B\}B'+B\{A,B'\}, \tag{3.5b}$$

$$\{A_1,\{A_2,A_3\}\}+\text{cycl}=0, \tag{3.5c}$$

An example of a Poisson algebra is given by the algebra of smooth complex valued functions on a nondegenerate symplectic manifold with the Poisson bracket derived from the symplectic form. [5]

If $\mathfrak{B}$ is a Poisson algebra we call a $\mathfrak{B}$ realization of

$L$ a linear map $\pi : L \to \mathfrak{B}$ such that

$$\{\pi(X), \pi(Y)\} = i\pi([X, Y]). \tag{3.6}$$

An analog of Proposition 2 holds for $\mathfrak{B}$ realizations.

*Proposition* 3: Let $\mathfrak{A}$ be the complex symmetric tensor algebra over $L$, and $\tilde{\iota}$ the canonical map $\tilde{\iota} : L \to \mathfrak{A}$.

(I) There exists a bilinear map of $\mathfrak{A} \times \mathfrak{A}$ to $\mathfrak{A}$ satisfying (3.5) such that $\tilde{\iota}$ is an $\mathfrak{A}$ realization of $L$.

(II) For any $\mathfrak{B}$ realization $\pi$ of $L$ there exists a unique homomorphism $\tilde{\pi}$ of Poisson algebras, i.e., an algebraic homomorphism $\tilde{\pi} : \mathfrak{A} \to \mathfrak{B}$ such that $\{\tilde{\pi}(A), \tilde{\pi}(B)\} = \tilde{\pi}(\{A, B\})$, which satisfies $\pi = \tilde{\pi} \circ \tilde{\iota}$.

(III) Let $\mathfrak{A}^n$ be the space of symmetric tensors of degree $n$. Then

$T^r \circ T^s$ is in $\mathfrak{A}^{r+s}$,

$\{T^r, T^s\}$ is in $\mathfrak{A}^{r+s-1}$ for $T^i$ in $\mathfrak{A}^i$, $i = r, s$.

(IV) There exists a unique antilinear map $\tilde{\sigma} : \mathfrak{A} \to \mathfrak{A}$ such that

$$\tilde{\sigma}(\tilde{X}) = \tilde{X}^* = -\tilde{X}, \quad \tilde{\sigma}(A \cdot B) = \tilde{\sigma}(A) \circ \tilde{\sigma}(B),$$

$$A, B \text{ in } \mathfrak{A}, \quad \tilde{X} = \tilde{\iota}(X), \quad X \text{ in } L_0.$$

The proof of Proposition 3, which may be less well known than the corresponding results for enveloping algebras, is given in the Appendix.

It follows from the Def. (3.6) and (3.4) that

$$\{T^*, S^*\}^* = \{T, S\} \text{ for } T, S \text{ in } \mathfrak{A},$$

since this is true for $T, S$ in $\tilde{\iota}(L)$. Hence the symmetric elements in $\mathfrak{A}$ form a real Poisson subalgebra of $\mathfrak{A}$.

The Poisson algebra $\mathfrak{A}$ is in fact isomorphic to the algebra $\mathfrak{P}$ of complex polynomial functions on the real dual $L_0^*$ of $L_0$. This isomorphism is given by the linear map

$$\pi : X \to iP_X, \quad P_X(\gamma) = \gamma(X), \quad X \text{ in } L_0, \quad \gamma \text{ in } L_0^*.$$

The symmetric elements in $\mathfrak{A}$ then correspond to real valued functions of $L_0$. Let $(X^r)$ be a real basis in $L_0$. Then the functions

$$P^r = P_{X^r}$$

are a set of coordinate functions on all of $L_0^*$. (3.7a)

In terms of these coordinates the Poisson bracket on $\mathfrak{P}$ defined by the Poisson bracket on $\mathfrak{A}$ is given by the bidifferential operator

$$\{F, G\} = \sum_{r,s,t} \frac{\partial}{\partial p^r} F \cdot \frac{\partial}{\partial p^s} G \cdot C_t^{rs} \cdot P^t, \tag{3.7b}$$

where

$$[X^r, X^s] = \sum_t C_t^{rs} \cdot X^t.$$

We may as well take (3.7) as a definition of the Poisson bracket on $\mathfrak{P}$, show its independence of the basis chosen, and verify properties (3.5a)–(3.5c) by explicit calculation. However, this procedure offers little insight into how the Poisson bracket on $\mathfrak{A}$ resp. $\mathfrak{P}$ is related to the bracket operation on $\mathfrak{U}$ defined in Proposition 2.

As was the case for the universal enveloping algebra, the Poisson algebra $\mathfrak{A}$ is uniquely determined (up to isomorphism) by the existence of an $\mathfrak{A}$ realization of $L$ satisfying Proposition (2.II). Hence, we should call $(\mathfrak{A}, \{\cdots, \cdots\})$ the universal Poisson algebra over $L$, although this terminology does not seem to be used in the literature. [Note that we have not introduced any condition on the Poisson bracket which would correspond to the notion of nondegeneracy for the Poisson algebra of smooth functions on a symplectic manifold. Such a notion cannot be introduced within this purely algebraic framework, where all algebras are finitely generated. A typical example of a Poisson algebra (which is the only one we shall deal with in this paper) is the restriction of $\mathfrak{P}$ to some orbit $\Gamma$ of the coadjoint action of the group defined by $L_0$ in $L_0^*$. In general, it is not possible to give a complete algebraic characterization of $\Gamma$. In particular, the ideal of polynomials which vanish on $\Gamma$ may reduce to zero. In any case however the Poisson bracket as defined by (3.7) may be extended to arbitrary smooth functions on $\Gamma$, and then defines a nondegenerate bivector field on $\Gamma$.[13]]

Let us now describe an algebraic relation between the universal algebras $\mathfrak{U}$ and $\mathfrak{A}$ and their bracket operations, which is of central importance for our discussion of the classical limit.[11,12]

For any positive integer $n$ consider the quotient space and canonical projection:

$$\mathfrak{B}^n = \mathfrak{U}^n / \mathfrak{U}^{n-1}, \quad \pi^n : \mathfrak{U}^n \to \mathfrak{B}^n, \quad \mathfrak{B}^0 = \mathbb{C} \cdot 1. \tag{3.7}$$

The direct sum $\mathfrak{B} = \oplus_{n=0}^{\infty} \mathfrak{B}^n$ is a linear space.

*Proposition* 4:

(I) $\mathfrak{B}$ is a complex commutative associative algebra isomorphic to the symmetric tensor algebra $\mathfrak{A}$ if a multiplication law is defined by

$$\pi^r(T^r) \cdot \pi^s(T^s) = \pi^{r+s}(T^r \circ T^s), \quad T^i \text{ in } \mathfrak{U}^i.$$

(II) $\mathfrak{B}$ is a Poisson algebra isomorphic to the Poisson algebra $\mathfrak{A}$ if the Poisson bracket is defined by

$$\{\pi^r(T^r), \pi^s(T^s)\} = i\pi^{r+s-1}(T^r \cdot T^s - T^s \cdot T^r), \quad T^i \text{ in } \mathfrak{U}^i.$$

(III) There exists a unique linear isomorphism $\psi : \mathfrak{B} \to \mathfrak{U}$ such that

$$\psi[\pi^1(\hat{X}) \cdots \pi^1(\hat{X})] = \hat{X} \cdots \hat{X}, \quad X \text{ in } L.$$

$\psi$ has the properties:

(i) $\mathfrak{U}^n = \sum_{k=0}^{n} \psi(\mathfrak{B}^k), \quad \pi^n \cdot \psi(B^n) = B^n, \quad B^n \text{ in } \mathfrak{B}^n.$

(ii) $\psi(\{\pi^1(X), B\}) = i\{\hat{X}, \psi(B)\}, \quad X \text{ in } L, \quad B \text{ in } \mathfrak{B}.$

(iii) $\psi \circ \tilde{\sigma} = \sigma \circ \psi.$

The proofs of this proposition may be found in Ref. 12 and Ref. 10, Chap. 2. $\mathfrak{B}$ is called the associated algebra of $\mathfrak{U}$.

Let us just remark here that by Proposition 4 the important properties (3.5a)–(3.5c) of the Poisson bracket turn out to be a direct consequence of the properties

(3.3a)–(3.3b) which hold for the "commutator" bracket in any associative algebra.

The linear isomorphism $\psi$ is not multiplicative, i.e., in general

$$\psi(B\cdot B_1)=\psi(B_1\cdot B)\neq\psi(B)\cdot\psi(B_1).$$

On this purely algebraic level $\psi$ is a unique quantization map, while its inverse describes the classical limit. (The inverse map $\psi^{-1}$ may be used to define a generalized Moyal bracket on $\mathfrak{B}$, which is just the image under $\psi^{-1}$ of the commutator bracket on $\mathfrak{U}$. The Moyal bracket then is a deformation of the Poisson bracket on $\mathfrak{B}$. For an approach to quantization along these lines see Refs. 8 and 9.)

We shall henceforth identify the three isomorphic Poisson algebras $\mathfrak{A}$, $\mathfrak{B}$, and $\mathfrak{P}$. In particular, we denote for any element $T^n$ in $\mathfrak{U}^n$ by $\pi^n(T^n)$ the corresponding complex valued polynomial on $L_0^*$.

## 4. REPRESENTATIONS OF $\mathfrak{U}$ AND THEIR CLASSICAL LIMIT

Let $D$ be a dense linear subspace of some complex Hilbert space $\mathcal{H}$ with scalar product $(\cdots,\cdots)$, and $\rho:L\to L$ in $D$ be a representation of $L$ by linear operators on $D$ satisfying

$$(f,\rho(X)g)=(\rho(X^*)f,g),\quad f,g \text{ in } D,\quad X \text{ in } D. \tag{4.1}$$

In particular the elements in $L_0$ resp. $iL_0$ are represented by antisymmetric resp. symmetric operators.

By Proposition 2 we have a unique representation $\hat\rho$ of $\mathfrak{U}$ satisfying

$$(f,\hat\rho(T)g)=(\hat\rho(T^*)f,g),\quad f,g \text{ in } D,\quad T \text{ in } \mathfrak{U}. \tag{4.2}$$

Assume now that the symmetric operators $\rho(X)$, $X$ in $iL_0$, have unique selfadjoint extensions, which are the basic observables of some quantum theory. [Note that even in this case there may exist symmetric operators $\hat\rho(T)$, $T$ in $\mathfrak{U}$ which do not have any self-adjoint extension (Ref. 14, Chap. X.) These operators are not quantum observables in the strict sense of the term, although all their expectation values for vector states in $D$ are real.] In order to define the classical limit of this theory let us make precise the concept of "limit of large quantum numbers."

Let us introduce a real positive parameter $\lambda$ measuring the order of magnitude of the basic observables in a vector state $\omega^\lambda$ for a sequence $(\hat\rho^\lambda)$ of representations of $\mathfrak{U}$:

$$\lim_{\lambda\to\infty}\lambda^{-1}\omega^\lambda(X)=i\gamma(X)=\pi^1(X)(\gamma),\quad X \text{ in } L_0,\quad \gamma \text{ in } L_0^*.$$

$$\omega^\lambda(T)=(f^\lambda,\hat\rho^\lambda(T)f^\lambda),\quad f^\lambda \text{ in } D^\lambda,\quad T \text{ in } \mathfrak{U}. \tag{4.3}$$

The physical interpretation of this limit should be clearly understood. Assume we have chosen a basis $(X^r)_r$ in $L_0$ such that all structure constants are of order unity. This would correspond to a choice of "microscopic units" for the physical quantities defined by the symmetric elements $-iX^r$, $r$. For physical states where all these quantities assume large values of order $\lambda\gg 1$ we may choose "macroscopic units" by introducing a basis $(\lambda^{-1}X^r)_r=(Y^r)_r$ with structure constants of order $\lambda^{-1}$. The real numbers $\gamma(X^r)$ of order unity are then assumed to approximate the expectation values of the "rescaled" quantities $-iY^r$, $r$ in the sense of (4.3).

Now assume that the dispersions $\Delta^2$ for all basic observables are of the same order of magnitude, which is given by their lower bound in (2.2). Then $\Delta^2(-i\hat\rho(\hat X)\approx\lambda$, hence

$$\lim_{\lambda\to\infty}\lambda^{-2}\Delta^2[-i\hat\rho(\hat X)]$$
$$=\lim_{\lambda\to\infty}\{\lambda^{-2}\omega^\lambda[(-iX)(-iX)]-\pi^1(-iX)\circ\pi^1(-iX)(\gamma)\}=0 \tag{4.4}$$

for $X$ in $L_0$ by (4.3).

In this case a limit corresponding to (4.3) does exist for arbitrary elements in $\mathfrak{U}$.

*Proposition* 5: Let $(\hat\rho^\lambda,\omega^\lambda)_\lambda$ be a sequence of linear representations $\hat\rho^\lambda$ of $\mathfrak{U}$ satisfying (4.2), with vector state $\omega^\lambda$. Assume that:

(i) $\lim_{\lambda\to\infty}\lambda^{-1}\omega^\lambda(X)=\pi^1(X)(\gamma)$,

(ii) $\lim_{\lambda\to\infty}\lambda^{-2}\omega^\lambda(X\circ X)=[\pi^1(X)\cdot\pi^1(X)](\gamma)$, $X$ in $iL_0$, $\gamma$ in $L_0^*$,

(iii) $\lim_{\lambda\to\infty}|\lambda^{-m}\omega^\lambda(T^m)|<\infty$, $T^m$ in $\mathfrak{U}^m, m$.

Then

$$\lim_{\lambda\to\infty}\lambda^{-m}\omega^\lambda(T^m)=\pi^m(T)(\gamma),\quad T^m \text{ in } \mathfrak{U}^m. \tag{4.5}$$

*Proof*: By induction on $m$. Assume (4.5) holds for some $m$. Elements in $\mathfrak{U}^{m+1}$ are linear combinations of elements of the form

$$X\circ T^m,\quad X \text{ in } iL_0,\quad T^m \text{ in } \mathfrak{U}^m.$$

Hence,

$$\lambda^{-m-1}\omega^\lambda(X\circ T^m)$$
$$=\lambda^{-m}\pi^1(X)(\gamma)\omega^\lambda(T^m)+\lambda^{-m}\omega^\lambda\{[\lambda^{-1}X-\pi^1(X)(\gamma)]T^m\}$$
$$=\omega_1+\omega_2=\omega.$$

However,

$$|\omega_2|^2\le\omega^\lambda\{[\lambda^{-1}X-\pi^1(X)(\gamma)]^2\}\circ\lambda^{-2m}\circ\omega^\lambda(T^{m*}\circ T^m)\to 0$$

as $\lambda\to\infty$ by assumptions (ii) and (iii). Hence,

$$\lim_{\lambda\to\infty}\omega=\lim_{\lambda\to\infty}\omega_1=[\pi^1(X)\circ\pi^m(T^m)](\gamma)=\pi^{m+1}(X\circ T^m)(\gamma).$$

Q.E.D.

*Proposition* 6: Under the assumptions of Proposition 5 we have

$$\lim_{\lambda\to\infty}i\lambda\circ\lambda^{-r-s}\omega^\lambda(V^r\circ W^s-W^s\circ V^r)$$
$$=\{\pi^r(V^r),\pi^s(W^s)\}(\gamma). \tag{4.6}$$

*Proof*: Apply Proposition 5 and Proposition 4(II).

Q.E.D.

We may employ the quantization map $\psi$ to give an equivalent version of Proposition 5 resp. 6 applying to arbitrary elements in $\mathfrak{U}$. Consider the "scaling transformation" $\tilde\lambda:L\to L$,

$$\tilde\lambda:X\to\tilde\lambda X=\lambda^{-1}\circ X,\quad 0<\lambda \text{ in } \mathbb{R}, \tag{4.7}$$

and denote by the same symbol its unique extension to the symmetric tensor algebra $\mathfrak{A} \cong \mathfrak{B}$,

$$\tilde{\lambda} \circ B^n = \lambda^{-n} B^n, \quad B^n \text{ in } \mathfrak{B}^n.$$

Then we may replace (4.5) by

$$\lim_{\lambda \to \infty} \omega^\lambda[\psi(\tilde{\lambda} \cdot B)] = B(\gamma), \quad B \text{ in } \mathfrak{B} \cong \mathfrak{P}. \tag{4.5'}$$

*Proposition* 7: Under the assumptions of Proposition 5 assume furthermore that the representation $\rho^\lambda$ of $L$ restricted to $L_0$ can be integrated to a unitary representation of the group $\mathfrak{G}$ belonging to $L_0$. Let $g$ be an element of $\mathfrak{G}$, and $T_g^\lambda$ the unitary representative of $g$ in the representation obtained. Then $T_g^\lambda(D^\lambda)$ is in the domain of definition and invariant under the self-adjoint extension of $\rho^\lambda(X)$ for any $X$ in $iL_0$. Call $\hat{\rho}_g^\lambda$ the corresponding representation of $\mathfrak{U}$ on $T_g^\lambda(D^\lambda)$. Then the vector state $\omega_g^\lambda$

$$\omega_g^\lambda(V) = (T_g^{\lambda^{-1}} f^\lambda, \hat{\rho}_g^\lambda(V)\, T_g^{\lambda^{-1}} f^\lambda)$$

satisfies all assumptions of Proposition 5 with

$$\lim_{\lambda \to \infty} \omega_g^\lambda(\lambda^{-1} X) = \gamma(\pi^1(\alpha_g \cdot X)) = \pi^1(X)(\alpha_g^* \cdot \gamma), \tag{4.8}$$

where $\alpha$ is the adjoint representation of $\mathfrak{G}$ on $L_0$.

*Proof*: Let $\mathfrak{H} = (h_t)$ be a one-parameter subgroup of $\mathfrak{G}$, $\varphi_t$ the transformation of right multiplication by $h_t$ on $\mathfrak{G}$, and $X$ the left invariant vector field on $\mathfrak{G}$ belonging to $\mathfrak{H}$,

$$\lim_{t \to \infty} \frac{1}{t}[F \circ \varphi_t - F] = X(F), \quad F \text{ smooth on } \mathfrak{G}.$$

For some element $g$ in $\mathfrak{G}$ consider the subgroup $\mathfrak{H}' = (g h_t g^{-1})$ with corresponding group of transformations $(\varphi_t')$ satisfying

$$\alpha_g \circ X(F) = X'(F) = \lim_{t \to \infty} (1/t)(F \circ \varphi_t' - F), \text{ where } \alpha : g \to \alpha_g$$

is the adjoint representation. Now let $\rho$ be a representation of $L_0$ as above, and denote by $\hat{\rho}(iX)$ the self-adjoint extension of $\rho(iX)$, $X$ in $L_0$.

Then by Stone's theorem (Ref. 15, Sec. 5c) some vector $f$ is in the domain of definition of $\hat{\rho}(iX)$ if and only if

$$\lim_{t \to \infty} \frac{i}{t}(T_{h_t} \circ f - f) = f' \equiv \hat{\rho}(iX) \circ f$$

does exist in the norm topology on $\mathcal{H}$. In particular this holds for $f$ in $D$, and it follows that

$$\lim_{t \to 0} \frac{i}{t} T_g^{-1}(T_{h_t'} f - f)$$

$$= \lim_{t \to 0} \frac{i}{t}(T_{h_t} T_g^{-1} f - T_g^{-1} f)$$

$$= \hat{\rho}(iX)\, T_g^{-1} f, \quad h_t' = g h_t g^{-1},$$

does exist for any $f$ in $D$, $X$ in $L_0$, and

$$T_g^{-1} \hat{\rho}(i\alpha_g \circ X) f = \hat{\rho}(iX)\, T_g^{-1} f.$$

This shows the invariance of $T_g^{-1} D$ under all operators $\hat{\rho}(iX)$, $X$ in $L_0$. Furthermore,

$$(T_g^{-1} f, \ \hat{\rho}_g(V)\, T_g^{-1} f) = (f, \hat{\rho}_g(\hat{\alpha}_g \circ V) f),$$

where $\hat{\alpha}$ is the unique representations of $\mathfrak{G}$ on $\mathfrak{U}$ defined by $\alpha$. Q.E.D.

The dense set of state vectors

$$f_g^\lambda = T_g^{-1} f^\lambda, \quad f^\lambda \text{ in } D^\lambda, \ g \text{ in } \mathfrak{G}, \tag{4.9}$$

is a system of coherent state vectors in the sense of Ref. 9. By Proposition 7 this system corresponds in the limit $\lambda \to \infty$ to a certain orbit $\Gamma$ of the co-adjoint action of $\mathfrak{G}$ on $L_0^*$,

$$\lim_{\lambda \to \infty} \omega_g^\lambda = \alpha_g^* \cdot \gamma \ \text{ is in } \Gamma, \quad \gamma \text{ in } L_0, \quad g \text{ in } \mathfrak{G}. \tag{4.10}$$

In the same sense the universal enveloping algebra $\mathfrak{U}$ when considered on this system of states is "contracted" to a commutative Poisson algebra $\mathfrak{P}$ of smooth functions on this orbit,

$$\lim_{\lambda \to \infty} \omega_g^\lambda[\psi(\tilde{\lambda} \cdot B)] = B(\alpha_g^* \circ \gamma), \quad B \text{ in } \mathfrak{B} \cong \mathfrak{P}. \tag{4.11}$$

The orbit $\Gamma$ is a nondegenerate symplectic manifold,[13,5] and is just the phase space of the classical system obtained in the limit.

In the next section we study some examples for which

$$\omega_g^\lambda = \omega^\lambda \text{ for } \alpha_g^* \circ \gamma = \gamma, \quad \lambda \text{ finite}. \tag{4.12}$$

The corresponding systems of coherent states have been discussed in Ref. 9. We note that by (4.12) there is a direct correspondence between the system of states and $\Gamma$, by which the $\mathfrak{G}$ invariant symplectic measure on $\Gamma$ defines a corresponding measure on the system of coherent states.

## 5. EXAMPLES AND COUNTEREXAMPLES

### A. The special nilpotent algebra

In the simplest case $L_0$ is defined in terms of a suitable basis $(X_0, X_1, X_2)$ by the Lie brackets

$$[X_1, X_2] = X_0, \quad [X_0, X_1] = [X_0, X_2] = 0. \tag{5.1}$$

There does exist a unique representation $\rho$ of $L$ by antisymmetric operators which can be integrated to a representation of the group $\mathfrak{G}(L) = \mathfrak{G}$ and for which

$$R_0 = i\rho(X_0) = 1. \tag{5.2}$$

In this representation there does exist, for any real $r > 0$, a unique vector $f$ satisfying the conditions of Proposition 1 for the symmetric operators $R_r = i\rho(X_r)$, $r = 1, 2$ such that

$$(R_1 + ir \cdot R_2) f = 0.$$

It follows that:

$$\omega_f(R_1) = \omega_f(R_2) = 0, \quad \omega_f(R_0) = 1, \tag{5.3a}$$

$$\omega_f(R_1 \circ R_1) = r, \quad \omega_f(R_2 \circ R_2) = r^{-1}. \tag{5.3b}$$

The vector $f$ is contained in a unique minimal invariant domain $D$ of definition for the operators $R_r$, $r = 0, 1, 2$, which we use to define the representation $\hat{\rho}$ of $\mathfrak{U}$. Now consider for any real $\lambda > 0$ the isomorphism of $L$,

$$X_r \to {}^\lambda X_r, \quad {}^\lambda X_0 = \lambda X_0, \quad {}^\lambda X_{1,2} = \sqrt{\lambda}\, X_{1,2}. \tag{5.4}$$

Then the sequence $(\hat{\rho}^\lambda, \omega^\lambda)$ of representations defined by

$$\rho^\lambda(X) = \rho({}^\lambda X), \quad X \text{ in } L,$$
$$\omega^\lambda(T) = (f, \hat{\rho}^\lambda(T) f), \quad T \text{ in } \mathfrak{U}, \tag{5.5}$$

satisfies all conditions of Proposition 5 with $\gamma$ in $L_0$

defined by

$$\pi^1(iX_0)(\gamma)=1,$$
$$\pi^1(iX_1)(\gamma)=\pi^1(iX_2)(\gamma)=0. \tag{5.6}$$

For any $\lambda$, the set of vectors $(T_g^\lambda \circ f,\ g \text{ in } \mathfrak{G})$, defines a system of coherent states $\omega_g^\lambda$ in the sense of Ref. 9. One has

$$\omega_g^\lambda(iX_r)=(\overset{-1}{T}{}_g^\lambda f, \hat{\hat\rho}{}^\lambda(iX_r)\,\overset{-1}{T}{}_g^\lambda f), \tag{5.7}$$
$$=\lambda,\quad \text{for } r=0,$$
$$=\lambda\circ\sigma_r(g),\quad r=1,2,\quad \sigma_r(g) \text{ in } \mathbb{R}.$$

In terms of the original representation $\hat\rho$ we have

$$(\overset{-1}{T}{}_g^\lambda f,\ \hat{\hat\rho}(iX_r)\,\overset{-1}{T}{}_g^\lambda f), \tag{5.8}$$
$$=1, \qquad \text{for } r=0,$$
$$=\sqrt{\lambda}\,\sigma_r(g),\quad \text{for } r=1,2.$$

This shows how the usual discussion of the classical limit for the special nilpotent algebras[2] which uses the the representation $\rho$ only, together with the sequence of coherent state vectors $(T_g^\lambda f)$, is related to our general group theoretical formulation.

### B. The algebra of SU(2)

The algebra is given in terms of a basis $(X^1,X^2,X^3)$ by

$$[X^1,X^2]=X^3,\quad [X^2,X^3]=X^1,\quad [X^3,X^1]=X^2. \tag{5.9}$$

The irreducible unitary representation of the group are finite dimensional and classified by positive integers $m$. Denote by $\rho^m$ the corresponding representation of the algebra. There does exist a unique vector $f^m$ such that

$$[\rho^m(iX^1)+i\rho^m(iX^2)]f^m=0,$$
$$\rho^m(iX^3)f^m=\tfrac{1}{2}mf^m.$$

It follows from (2.3) that

$$\omega^m(iX^1)=\omega^m(iX^2)=0,\quad \omega^m(iX^3)=\tfrac{1}{2}m,$$
$$\omega^m(iX^1\circ iX^1)=\omega^m(iX^2\cdot iX^2)=\tfrac{1}{4}m.$$

Hence the sequence $(\hat\rho^m,\omega^m)$ satisfies the conditions (i), (ii) of Proposition 5, and it is not difficult to show that (iii) holds as well. Again, the vectors $\overset{-1}{T}{}_g^m\cdot f^m$, $g$ in $\mathfrak{G}$, define a system of coherent states in the sense of Ref. 9. The phase space $\Gamma$ is just the orbit defined by

$$\gamma^1\circ\gamma^1+\gamma^2\circ\gamma^2+\gamma^3\circ\gamma^3=\tfrac{1}{4},$$
$$\gamma^r=P_{X^r}(\gamma)=\pi(-iX^r)(\gamma),\quad \gamma \text{ in } L_0^*. \tag{5.10}$$

### C. The algebra of SL (2 IR)

The algebra $L_0$ is given in terms of a basis $(X^0,X^1,X^2)$ by

$$[X^0,X^1]=X^2,\quad [X^1,X^2]=-X^0,\quad [X^2,X^0=X^1]. \tag{5.11}$$

Representations of $L_0$ which can be integrated to unitary representations of the covering group $\mathfrak{G}=\overline{\mathrm{SL}(2,\mathbb{R})}$ have been first studied by Bargmann.[15] We consider here the so-called discrete classes only.

Denote by $H^r$ the self-adjoint operator corresponding to the symmetric element $iX^r$, $r=0,1,2$.

In the representation called $D_\lambda^+$ by Bargmann there does exist a unique vector $f^\lambda$ such that

$$Gf^\lambda=(H^1-iH^2)f^\lambda=0,$$
$$H^0f^\lambda=\lambda\circ f^\lambda,\quad \lambda>0, \tag{5.12}$$

hence

$$\omega_f(H^1)=\omega_f(H^2)=0,\quad \omega_f(H^0)=\lambda. \tag{5.13}$$

From Proposition 1 we conclude that

$$\omega_f(H^1\cdot H^1)=\omega_f(H^2\cdot H^2)=\tfrac{1}{2}\lambda. \tag{5.14}$$

There is a unique minimal invariant domain of definition $D^\lambda$ for all operators $H^r$, $r=0,1,2$ containing $f^\lambda$ which we use to define the representation $\hat\rho^\lambda$ of $\mathfrak{U}$. Again all conditions of Propositions 5 and 7 are fulfilled, and we obtain

$$\lim_{\lambda\to\infty}\omega^\lambda=\gamma,$$
$$\gamma^0=-\tfrac{1}{2},\quad \gamma^1=\gamma^2=0,\quad \gamma^r=P_{X^r}(\gamma)=-i\pi(X^r)(\gamma). \tag{5.15}$$

The phase space $\Gamma$ in $L_0^*$ is the submanifold defined by the equations

$$\gamma^0\cdot\gamma^0-\gamma^1\cdot\gamma^1-\gamma^2\circ\gamma^2=\tfrac{1}{4},\quad \gamma^0<0. \tag{5.16}$$

Using techniques described in Refs. 16 and 17 it can be shown that $\Gamma$ is isomorphic to the symplectic manifold of timelike geodesics on a relativistic spaceform of dimension two which has $\mathfrak{G}$ as a group of motions.

### D. The special nilpotent algebra: A counterexample

Consider the Hilbert space $\mathcal{H}=L^2(0,1)$ and the domain $D$ of infinitely often differentiable functions:

$$D=\{f \text{ in } C^\infty(0,1); f(0)=f(1)=0\}. \tag{5.17}$$

On $D$ we define the symmetric operators

$$R_0f=f, \tag{5.18a}$$

$$(R_1f)(x)=i\frac{d}{dx}f(x), \tag{5.18b}$$

$$(R_2f)(x)=xf(x). \tag{5.18c}$$

The map $\rho: X_r\to -iR_r$, $r=0,1,2$ defines a representation of the special nilpotent algebra of Sec. 5A and of its enveloping algebra $\mathfrak{U}$.

For vectors $f$ in $D$, Heisenberg's uncertainty relations hold,

$$\Delta^2(R_1)\circ\Delta^2(R_2)\geq\tfrac{1}{4}. \tag{5.19}$$

Since $R_2$ is bounded, $\rho$ cannot be integrated to a representation of the corresponding group. On the other hand there does exist for any real number $\alpha$ a self-adjoint extension $R_1^\alpha$ of $R_1$ with discrete spectrum, the eigenvectors of which are given by the functions

$$f_m^\alpha(x)=\exp(-i\alpha)\circ\exp(-2\pi imx),\quad R_1^\alpha f_m^\alpha=(\alpha+2\pi m)f_m^\alpha. \tag{5.20}$$

For any such vector the vector $R_2f_m^\alpha$ is not in the domain of $R_1^\alpha$, hence Proposition 1 does not apply to such vectors and, in fact, we have

$$\Delta_f^2(R_1)=0,\quad f=f_m^\alpha. \tag{5.21}$$

For the sequence $\rho^\lambda$ of representations constructed as in Sec. 5 A we have

$$\rho^\lambda(iX_2) = \sqrt{\lambda}\,\rho^\lambda(iX_2) = \sqrt{\lambda}\,R_2,$$

hence

$$\lim_{\lambda\to\infty} \lambda^{-1}\omega^\lambda(iX_2) = 0,$$

for any sequence of vector states, since $R_2$ is bounded. This shows that a classical limit leading to the usual phase space of the special nilpotent algebra does not exist for this type of quantum theory.

A successful discussion of the classical limit in this situation can be based on a different algebra defined by the equations

$$[Y_0, Y_1] = Y_2, \quad [Y_0, Y_2] = -Y_1, \quad [Y_1, Y_2] = 0, \tag{5.22}$$

with a representation $\rho^\lambda$ on $D$ given by

$$\begin{aligned} &\rho^\lambda(Y_0): f(x) \to \frac{\partial}{\partial x} f(x), \\ &\rho^\lambda(Y_1): f(x) \to i\lambda \sin x \cdot f(x), \\ &\rho^\lambda(Y_2): f(x) \to i\lambda \cos x \cdot f(x). \end{aligned} \tag{5.23}$$

We leave it as an exercise to the reader to demonstrate the existence of a classical limit in this case.

## APPENDIX: PROOF OF PROPOSITION 3

$\mathfrak{A}$ is generated by the elements in $\mathbb{C}\cdot 1$ and $\tilde{\iota}(L)$. For any linear map $\beta: L \to \mathfrak{A}$ the linear map $\tilde{\beta}: \mathfrak{A} \to \mathfrak{A}$ defined by

$$\tilde{\beta}(\tilde{X}_1 \circ\cdots\circ \tilde{X}_m) = \sum_{k=1}^{m} \tilde{X}_1 \circ\cdots\circ \beta(X_k) \circ\cdots\circ \tilde{X}_m, \quad \tilde{\beta}(1) = 0, \quad \tilde{X}_r = \tilde{\iota}(X_r) \tag{A1}$$

is a derivation on $\mathfrak{A}$, i.e., satisfies $\tilde{\beta}(S\cdot T) = \tilde{\beta}(S)\cdot T + S\circ\tilde{\beta}(T)$. Conversely any derivation on $\mathfrak{A}$ is uniquely determined by its restriction to $\tilde{\iota}(L)$ via (A1). For $\tilde{X}, \tilde{Y}$ in $\tilde{\iota}(L)$ define the bracket operation by

$$\{\tilde{X}, \tilde{Y}\} = i[X, Y] = \alpha_{\tilde{X}}(Y), \tag{A2}$$

where $\alpha_{\tilde{X}}$ is linear. Then $\tilde{\alpha}_{\tilde{X}}$ is a derivation, and we may for $A$ fixed in $\mathfrak{A}$ consider the map $\gamma_A: X \to \tilde{\alpha}_X(A)$. Putting

$$\{B, A\} = \tilde{\gamma}_A(B) \tag{A3}$$

we can easily see that the bracket thus defined satisfies (3.5a) and (3.5b). As for (3.5c) we observe that the trilinear map

$$\delta: A_1, A_2, A_3 \to \{A_1, \{A_2, A_3\}\} + \{A_2, \{A_3, A_1\}\} + \{A_3, \{A_1, A_2\}\}$$

is a derivation on any of its arguments if the remaining arguments are kept fixed. It follows that $\delta$ vanishes identically since it vanishes for $A_i$ in $\tilde{\iota}(L)$, $i = 1, 2, 3$, by Jacobi's identity. Finally let $\pi$ be a $\mathfrak{B}$ realization and $\tilde{\pi}: \mathfrak{A} \to \mathfrak{B}$ the unique homomorphism such that $\pi = \tilde{\pi}\circ\tilde{\iota}$. Then we have in $\mathfrak{B}$,

$$\{\tilde{\pi}(\tilde{X}), \tilde{\pi}(\tilde{Y})\} = \{\pi(X), \pi(Y)\} = i\pi([X, Y]) = \tilde{\pi}(\{\tilde{X}, \tilde{Y}\}).$$

It follows that the image of $\mathfrak{A}$ in $\mathfrak{B}$ under $\tilde{\pi}$ is a Poisson subalgebra generated by elements $z\circ 1$, $\tilde{\pi}(\tilde{X})$, $Z$ in $\mathbb{C}$, $X$ in $L$. Since any derivation on this subalgebra is uniquely determined by its action on a set of generators we conclude as above that

$$\{\tilde{\pi}(\tilde{X}), \tilde{\pi}(B)\} = \tilde{\pi}\{\tilde{X}, B\}, \text{ and finally}$$

$$\{\tilde{\pi}(A), \tilde{\pi}(B)\} = \pi(\{A, B\}), \text{ for } X \text{ in } L,\ A, B \text{ in } \mathfrak{A}.$$

The proofs of (III) and (IV) are straightforward and will not be spelled out here.[12] Q.E.D.


[1] W. Heisenberg, *Physikalische Prinzipien der Quantentheorie* (Leipzig, Hirzel, 1930).

[2] K. Hepp, "The classical limit of quantum mechanical correlation function," Commun. Math. Phys. **35**, 265–277 (1974).

[3] L. P. Staunton and S. Browne, "Classical limit of relativistic positive-energy theories with intrinsic spin," Phys. Rev. D **12**, 1026–37 (1975).

[4] H. W. Crater and J. Naft, "Classical limit and generalizations of the homogeneous quasipotential equation for scalar interactions," Phys. Rev. D **12**, 1165–77 (1975).

[5] R. Jost, "Poisson brackets (An unpedagogical lecture)," Rev. Mod. Phys. **36**, 572–9 (1964).

[6] E. Inönü and E. P. Wigner, "On the contraction of groups and their representations," Proc. Nat. Acad. Sci. **39**, 510–24 (1953).

[7] F. Bayen *et al.*, "Deformation theory and quantization. I. Deformation of symplectic structures," Preprint, Univ. California, UCLA-77-TEP-7, February 1977.

[8] F. Bayen *et al.*, "Deformation theory and quantization. II. Physical applications," Preprint, Univ. California, UCLA-77-TEP-8, March 1977.

[9] A. M. Perelomov, "Coherent states for arbitrary Lie group," Commun. Math. Phys. **26**, 222–36 (1972).

[10] J. Dixmier, *Algèbres enveloppantes* (Paris, Gauthier-Villar, 1974).

[11] L. Corwin *et al.*, "Graded Lie algebras in mathematics and physics (Bose-Fermi symmetry)," Rev. Mod. Phys. **47**, 573–603.

[12] K. Drühl, "A theory of classical systems based on graded Lie algebras. I. Graded Lie algebras, enveloping algebras and their Poisson bracket structure," Preprint (Max-Planck Institute, Starnberg, 1976).

[13] B. Kostant, Quantization and unitary representations. In: C. T. Taam (Ed.), *Lectures in Modern Analysis and Applications III.*, Lecture Notes in Mathematics 170 (Springer, Berlin, 1970).

[14] K. Maurin, *Methods of Hilbert spaces* (Polish Sci. Publ., Warszawa, 1972).

[15] V. Bargmann, "Irreducible unitary representations of the Lorentz group," Ann. Math. (N.Y.) **48**, 568–640 (1947).

[16] K. Drühl, "On the space–time interpretation of classical canonical systems. I. The general theory," Commun. Math. Phys. **49**, 277–88 (1976).

[17] K. Drühl, "On the space–time interpretation of classical canonical systems. II. Relativistic canonical systems," Commun. Math. Phys. **49**, 289–300 (1976).

# Exact vacuum solutions of Einstein's equation from linearized solutions[a), b)]

Basilis C. Xanthopoulos

*Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637*
(Received 16 January 1978)

It is proved that if $(M,g_{ab})$ is an exact vacuum solution of Einstein's equation, $l_a$ a null vector field and if $l_a l_b$ satisfies the linearized equation on background $(M,g_{ab})$, then $g_{ab}+l_a l_b$ is an exact vacuum solution. Applications to the search for asymptotically flat spacetimes are discussed.

## I. INTRODUCTION

It is of interest in general relativity to discover and interpret solutions of Einstein's equation. Additional solutions may provide more insight into the theory; their interpretation may permit the description of additional physical systems. We present here a result which might contribute to the discovery of more solutions as well as their physical interpretation. Specifically it selects a preferred subfamily of the linearized solutions—namely, certain linearized solutions which lead to exact solutions. This result might be useful because, while it is easier to obtain linearized solutions, exact solutions are, of course, the more interesting.

In Sec. II we prove the main result: if, for $l_a$ a null vector field, $l_a l_b$ is a linearized solution on the vacuum background $g_{ab}$, then $g_{ab}+l_a l_b$ is an exact vacuum solution. In Sec. III we point out that this result might be useful in the search for asymptotically flat spacetimes.

## II. THE THEOREM

Let $(M,g_{ab})$ be a one-parameter family of spacetimes, i.e., a $C^\infty$, four-dimensional manifold $M$ with nondegenerate metrics $g_{ab}(\lambda)$ of the Lorentz signature. Assume that for every value of the parameter $\lambda$ the metric $g_{ab}(\lambda)$ satisfies the vacuum Einstein equation $R_{ab}(g_{ab}(\lambda))=0$.[1] Equating $(d/d\lambda)(R_{ab}[g_{ab}(\lambda)])_{|\lambda=0}$ to zero, we obtain

$$\nabla^m\nabla_m h_{ab}-2\nabla^m\nabla_{(a}h_{b)m}+\nabla_a\nabla_b(g^{mn}h_{mn})=0, \tag{1}$$

where $g_{ab}=g_{ab}(0)$ is the background metric, $\nabla_a$ the derivative operator associated with $g_{ab}$, and $h_{ab}=(d/d\lambda)\times(g_{ab}(\lambda))_{|\lambda=0}$ is the first order change of the metric $g_{ab}$ along the family $g_{ab}(\lambda)$. Equation (1) is called the linearized equation; its solutions linearized fields. That is $(M, g_{ab}+\lambda h_{ab})$ is an approximate (for $\lambda\to 0$) solution of the Einstein equation.

Our result is the following.

*Theorem* 1: Let $(M, g_{ab})$ be an exact vacuum solution of Einstein's equation and let $l_a$ be a null vector field such that $h_{ab}=l_a l_b$ satisfies the linearized equation (1). Then $g_{ab}+l_a l_b$ is an exact vacuum solution.

The theorem says that in the class of linearized fields there is a preferred subclass, namely those of the form $l_a l_b$ for some null vector field $l_a$. Furthermore, the structure of the Einstein equation is such that $g_{ab}+l_a l_b$, which ought for "small" $l_a$ to represent an approximate solution, is, in fact, always an exact solution.

The proof consists of substituting $g_{ab}+l_a l_b$ into Einstein's equation, expanding and using the linearized equation (1).

We begin with Eq. (1) on the field $l_a l_b$, i.e.,

$$\nabla^m\nabla_m(l_a l_b)-2\nabla^m\nabla_{(a}[l_{b)}l_m]=0, \tag{2}$$

Set $x^a=l^m\nabla_m l^a$. Contracting (2) with $l^a l^b$ we obtain $x^a x_a=0$ while the nullness of $l^a$ yields $l_a x^a=0$. So, $l^a$ and $x^a$, as two real, null, and mutually orthogonal vector fields, must be parallel, i.e., $l^a$ must be geodesic. Now define

$$l^m\nabla_m l^a=\phi l^a,\quad \nabla^m l_m=\theta.$$

Then contraction of Eq. (2) with $l^a$ yields

$$(\nabla^m l^n)(\nabla_m l_n)=-(\dot\phi+\dot\theta)-(\phi+\theta)\phi \tag{3}$$

while the definition of $\theta$ yields

$$(\nabla^m l^n)(\nabla_n l_m)=\dot\phi-\dot\theta+\phi\theta, \tag{4}$$

where a dot denotes the directional derivative along $l^a$.

We next obtain the Ricci tensor of $g'_{ab}$. First note that, since $l_a$ is null, $g'_{ab}=g_{ab}+l_a l_b$ is again a nondegenerate metric, and in fact its inverse is $g'^{ab}=g^{ab}-l^a l^b$, where $l^a=g^{am}l_m$.

Let $\nabla'_a$ denote the derivative operator compatible with $g'_{ab}$. The connection tensor field[2] $C^m_{ab}$ which relates the two derivative operators, $\nabla_a$ and $\nabla'_a$, is easily found to be

$$C^m_{ab}=l^m\nabla_{(a}l_{b)}+l_{(a}\nabla_{b)}l^m-l_{(a}\nabla^m l_{b)}+\phi l^m l_a l_a \tag{5}$$

which, because of the nullness of $l^a$, satisfies $C^m_{mb}=0$, $l^b C^m_{ab}=\phi l^m l_a$, $l_m C^m_{ab}=-\phi l_a l_b$. The Ricci tensor of $g'_{ab}$, on the other hand, is given by

$$R'_{ab}=R_{ab}+2\nabla_{[m}C^m_{a]b}+2C^m_{a[b}C^n_{n]m}. \tag{6}$$

Substituting (5) into (6), using (2), (3), (4), and $R_{ab}=0$, we obtain finally $R'_{ab}=0$. QED

Note that the condition that a linearized field $h_{ab}$ be of the form $l_a l_b$ with $l_a$ null is not gauge invariant, i.e., it is not invariant under the addition to $h_{ab}$ of the symmetrized derivative of a vector field. Thus, it is conceivable that one could use the gauge freedom to make applicable the hypothesis of the theorem, i.e., to have, given $h_{ab}$, $h_{ab}+2\nabla_{(a}\xi_{b)}$ of the form $l_a l_b$, for a suitable choice of $\xi_b$.

---

a) Supported by the NSF contract PHY 76-81102 with the University of Chicago.

b) Presented as part of a thesis to the Department of Physics, University of Chicago, in partial fulfillment of the requirement for the Ph.D. degree.

 0022-2488/78/1907-1607$1.00  

We compare Theorem 1 with other known results. Kerr and Schild[3] found all vacuum solutions of Einstein's equation of the form $\eta_{ab}+l_a l_b$, where $\eta_{ab}$ is flat and $l_a$ is a null vector field whose divergence and twist do not both vanish at any point. What Theorem 1 says then is that the Kerr–Schild metrics could have been obtained by just solving the linearized equation for $l_a l_b$.[4] They, however, obtained these metrics using a tetrad approach, in which the linearity of the equations is not transparent. Theorem 1 is even stronger: It is applicable also in a curved background.

## III. ASYMPTOTICALLY FLAT SPACETIMES

An extensively studied class of solutions in general relativity are those which are asymptotically flat at null infinity,[5] i.e., which describe the spacetime of an isolated body. It turns out that it is rather difficult to find asymptotically flat solutions of Einstein's equation. Indeed, although this notion was developed in order to study gravitational radiation, no exact asymptotically flat radiative solution is yet available. In this section we remark that Theorem 1 may be useful in the search for asymptotically flat spacetimes.

Let $(M, g_{ab})$ be an asymptotically flat spacetime,[6] so in particular there exists a manifold $\tilde{M}=M\cup\mathscr{I}$ with smooth metric $\tilde{g}_{ab}$ and smooth scalar $\Omega$ such that on $M$ $\tilde{g}_{ab}=\Omega^2 g_{ab}$ and on $\mathscr{I}$ $\Omega=0$, $\tilde{n}_a=\tilde{\nabla}_a\Omega$ is nonzero and null and $\nabla_a\tilde{n}_b=0$. Let $h_{ab}$ be a linearized field on $(M, g_{ab})$. Then $h_{ab}$ is said to preserve asymptotic flatness to first order if $\Omega^2 h_{ab}$ admits a smooth extension to $\mathscr{I}$ such that $\Omega h_{ab}\tilde{n}^a\tilde{n}^b|_{\mathscr{I}}=0$, i.e., if the conditions in the definition are satisfied to first order.[7] Consider now the special case in which the linearized field is of the form $l_a l_b$. Obviously, it preserves asymptotic flatness to first order if and only if

(i) $\Omega l_a$ admits a smooth extension to $\mathscr{I}$ and (ii) there it is a multiple of $\tilde{n}^a$, i.e., $\Omega l_a\tilde{n}^a|_{\mathscr{I}}=0$. Actually, (ii) is a consequence of (i) and the linearized equation on $l_a l_b$. An analogous statement is also true in the full theory.

*Theorem* 2: Let $(M, g_{ab})$ be a vacuum asymptotically flat spacetime and let $l_a$ be a null vector field on $M$ such that $l_a l_b$ satisfies the linearized equation on $(M, g_{ab})$. If the vector field $\Omega l_a$ admits a smooth extension to $\mathscr{I}$, then the solution $(M, g_{ab}+l_a l_b)$ is also asymptotically flat.[8]

*Proof*: Since $(M, g_{ab})$ is asymptotically flat, there is a manifold with boundary $\tilde{M}=M\cup\mathscr{I}$ and a choice of a conformal factor $\Omega$ such that the conditions in the definition of Ref. 6 are satisfied. We choose the same manifold $\tilde{M}$ and the same conformal factor $\Omega$ to prove asymptotic flatness of $(M, g'_{ab}=g_{ab}+l_a l_b)$. Since smoothness of $\Omega l_a$ implies smoothness of $\tilde{g}'_{ab}=\Omega^2 g'_{ab}$, we only have to show that $\tilde{n}_a$ is null on $\mathscr{I}$ with respect to $\tilde{g}'_{ab}$ and that $\tilde{\nabla}'_a\tilde{n}_b=0$ on $\mathscr{I}$.

For the first set $\tilde{\zeta}_a=\Omega l_a$, so $\tilde{\zeta}_a$ is a smooth null vector field on $\tilde{M}$; moreover, $\tilde{\zeta}_a$ is geodesic in $(\tilde{M},\tilde{g}_{ab})$. Define $\Phi$, $\Theta$, $K$ by $\tilde{\zeta}^m\tilde{\nabla}_m\tilde{\zeta}_a=\Phi\tilde{\zeta}_a$, $\Theta=\tilde{\nabla}^m\tilde{\zeta}_m$ and $\tilde{\zeta}^m\tilde{n}_m=K$. Equation (3) expressed in terms of fields on $(\tilde{M},\tilde{g}_{ab})$ is

$$6K^2+\Omega^2(\tilde{\nabla}^m\tilde{\zeta}^n)(\tilde{\nabla}_m\tilde{\zeta}_n)+\Omega^2\tilde{\zeta}^m\tilde{\nabla}_m(\Phi+\Theta)-4\Omega\tilde{\zeta}^m\tilde{\nabla}_m K$$
$$+\Omega^2(\Phi+\Theta)\Phi-2\Omega K(\Phi+\Theta)=0.$$

Since the fields $K$, $\Phi$, $\Theta$, $\tilde{\zeta}^m$ are automically smooth on all of $\tilde{M}$, this equation implies that $K$ vanishes on $\mathscr{I}$. Thus, $\tilde{g}'^{ab}\tilde{n}_a\tilde{n}_b|_{\mathscr{I}}=[\tilde{g}^{ab}\tilde{n}_a\tilde{n}_b-K^2]_{\mathscr{I}}=0$, i.e., the nullness of $\tilde{n}_a$ is preserved.

For the second note that, since $g'_{ab}$ satisfies the vacuum Einstein's equation, we have that $\tilde{\nabla}'_a\tilde{n}_b|_{\mathscr{I}}=0$ if and only if $\lim_{\mathscr{I}}\Omega^{-1}\tilde{g}'^{ab}\tilde{n}_a\tilde{n}_b=0$.[9] But the latter follows from $K|_{\mathscr{I}}=0$ and $\lim_{\mathscr{I}}\Omega^{-1}\tilde{g}^{ab}\tilde{n}_a\tilde{n}_b=0$. ■

We consider now the relationship between the asymptotic gravitational fields of the two metrics $g_{ab}$ and $g'_{ab}$. The Einstein equation and asymptotic flatness imply that the Weyl tensor of $\tilde{g}_{ab}$ $\tilde{C}_{abcd}$ vanishes on $\mathscr{I}$, so $\Omega^{-1}\tilde{C}_{abcd}$ admits a smooth extension to $\mathscr{I}$. The asymptotic gravitational field is then described by

$$\tilde{K}_{ab}=\lim_{\mathscr{I}}\Omega^{-1}\tilde{C}_{ambn}\tilde{n}^m\tilde{n}^n.$$

[For example, in the Newman–Penrose notation, $\tilde{K}_{ab}=-2(\mathrm{Re}\Psi_2^0)\tilde{n}_a\tilde{n}_b+2\Psi_3^0\tilde{n}_{(a}\tilde{l}_{b)}+2\bar{\Psi}_3^0\tilde{n}_{(a}\bar{\tilde{l}}_{b)}-\Psi_4^0\tilde{l}_a\tilde{l}_b-\bar{\Psi}_4^0\bar{\tilde{l}}_a\bar{\tilde{l}}_b$. $\Psi_4^0$ is the radiation field. The Bondi mass is essentially the integral of $\Psi_2^0$ over a cross section of $\mathscr{I}$.] Thus, a spacetime is free of gravitational radiation if and only if $\tilde{K}_{ab}=\tilde{n}_{(a}\tilde{V}_{b)}$ for some vector field $\tilde{V}_b$ tangential to $\mathscr{I}$. We shall in fact show that

$$\tilde{K}'_{ab}-\tilde{K}_{ab}\sim\tilde{n}_a\tilde{n}_b \quad \text{on } \mathscr{I}. \tag{7}$$

So, the radiation fields of the two spacetimes $g_{ab}$ and $g'_{ab}$ are the same—in particular either both possess gravitational radiation or neither—but in general, given a cross section of $\mathscr{I}$, they may have different Bondi masses.

Finally, we sketch the proof of (7). Using the relation $C'_{abc}{}^d=C_{abc}{}^d-2\nabla_{[a}C_{b]c}^{\ d}+2C^m_{c[a}C_{b]m}^{\ d}$ between the respective Weyl tensors of the (vacuum) metrics $g'_{ab}$ and $g_{ab}$, using Eq. (5) and the nullness of $l^a$, it is easy to show that

$$(C'_{abcd}-C_{abcd})\,l^b l^d\sim l_a l_c. \tag{8}$$

Expressing (8) in terms of fields in the conformally completed spacetime and using that on $\mathscr{I}$ $\Omega l_a$ and $\tilde{n}_a$ are parallel (because they are both null and orthogonal on $\mathscr{I}$), we obtain (7).[10] Q.E.D.

## IV. DISCUSSION

Theorem 1 might be a useful tool in the search for exact solutions of the Einstein vacuum equation. Whenever the perturbations, i.e., the linearized solutions, of some spacetime have been obtained, one can ask for those perturbations which lead to exact solutions. If the original solution is asymptotically flat, one can possibly obtain additional asymptotically flat spacetimes. The main source of difficulty in the search of the perturbations of the type considered here is the fact that the condition $h_{ab}=l_a l_b$ with $l_a$ null is neither gauge invariant nor linear. Theorem 1 might also be a useful tool in the analysis and interpretation of certain known exact solutions. The idea would be to describe these solutions, via Theorem 1, in terms of linearized fields on a well understood background. In particular, the Kerr-Schild family and the plane wave solutions—which are precisely the vacuum solutions of the form $\eta_{ab}+l_a l_b$ with $\eta_{ab}$ flat and $l_a$ some null vector field—can be studied as perturbations of flat spacetime.

How large is the class of solutions that can be obtained via Theorem 1? It is apparently possible for a null $l^a$ satisfying the hypothesis of Theorem 1 to be neither a principal null direction of the Weyl tensor of $g_{ab}$ nor of $g'_{ab}$; the class may therefore be quite large. Unfortunately, we do not have any explicit examples. So, to get some feeling for how large is the class of solutions one might obtain from Theorem 1, we look at the Kerr–Schild class of solutions. These spacetimes are all algebraically special (in fact, of types [2,1,1] or [2,2]) and all admit at least one Killing field. The general solution in this class is determined by one arbitrary analytic function of one complex variable. For the present spacetimes, i.e., with an arbitrary curved background, one can prove the following. First, directly from Eq. (8), if $l^a$ is a repeated principal null direction of the Weyl tensor of $g_{ab}$, then $l^a$ must also be a repeated principal null direction of the Weyl tensor of $g'_{ab}$. Second, for $l^a$ a principal but *not* a repeated principal null direction of the Weyl tensor of $g_{ab}$, then $l^a$ need not even be a principal null direction of $g'_{ab}$. Furthermore, again for $l^a$ principal but not repeated, $l^a$ must have zero twist.[11] We do not know any further simple consequences on $g_{ab}$ and $l^a$ of the assumption that both $g_{ab}$ and $g_{ab}+l_a l_b$ are vacuum solutions. In fact, we know of no application of Theorem 1 to spacetimes other than the Kerr–Schild class and the plane waves.

Why does Einstein's equation have this curious feature?


## ACKNOWLEDGMENTS

I wish to thank Abhay Ashtekar for his encouragement and useful suggestions and Robert Geroch and Richard Hansen for many helpful discussions.


[1]Our conventions are: $\nabla_{[a}\nabla_{b]}K_c=\frac{1}{2}R_{abc}{}^{m}K_m$, $R_{ab}=R_{amb}{}^{m}$, $R=R_m{}^m$.

[2]We define $\nabla'_a K_b=\nabla_a K_b - C^m_{ab}K_m$, for all vector fields $K_a$.


[3]G. C. Debney, R. P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).


[4]Theorem 1 may be relevant to the claim of M. Gürses and F. Gürsey [J. Math. Phys. **16**, 2385 (1975)] that there exists a coordinate system in which the field equations are linear, "because the energy pseudotensor of the Kerr–Schild gravitational field vanishes."


[5]R. Geroch, in *Asymptotic Structure of Spacetime*, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977).


[6]For the definition, see R. Geroch and G. T. Horowitz, Phys. Rev. Lett. **40**, 203 (1978).

[7]In particular, that $\tilde{\nabla}_a\tilde{n}_b \mid \mathscr{I}=0$ to first order follows from $\Omega h_{ab}\tilde{n}^a\tilde{n}^b \mid \mathscr{I}=0$ and Einstein's equation.

[8]A discussion with R. O. Hansen on the possibility of the existence of such a theorem is acknowledged.

[9]See, e.g., Ref. 5, Eq. (6).

[10]If $\Omega l_a$ vanishes on $\mathscr{I}$ the proof is somewhat more difficult.

[11]So, all such solutions belong to the Newman–Tamburino family, E. T. Newman and L. Tamburino, J. Math. Phys. **3**, 902 (1962).

# $N$-body quantum scattering theory in two Hilbert spaces. II. Some asymptotic limits

Colston Chandler

*Physikalisches Institut der Universität Bonn, Endenicher Allee 11-13, D-5300 Bonn 1, FDR and Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131*

A. G. Gibson

*Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico 87131*
(Received 16 January 1978)

Within the framework of two-Hilbert space scattering theory the existence of the strong Abel limit of a certain operator is proved, leading to the following results. A generalized Lippmann identity is derived that is valid for all channels, rather than only two-body channels. On shell equivalence of the prior, post and AGS transition operators is rigorously proved, thus closing a gap in previous proofs. Results concerning the existence of the scattering operator as a strong, rather than weak, Abel limit are presented, and their implications with respect to the problem of unitarity are discussed. Finally, the possibility of exploiting operator limits of the Obermann-Wollenberg type is studied, with negative results.

## I. INTRODUCTION

In time-dependent nonrelativistic multichannel quantum scattering theory the question of in what sense the operators

$$W_{\beta\alpha}^{\pm}(t) \equiv P_\beta \exp(\pm iH_\beta t)\exp(\mp iH_\alpha t)P_\alpha - \delta_{\beta\alpha}P_\alpha \qquad (1.1)$$

approach zero as $t \to \infty$ has proved an interesting one. Here the operators $H_\beta$ and $H_\alpha$ are channel Hamiltonians, and the operators $P_\beta$ and $P_\alpha$ are the orthogonal projections of the $N$-particle Hilbert space $\mathcal{H}_N$ onto the respective channel subspaces. The symbol $\delta_{\beta\alpha}$ denotes the Kronecker delta. An early result was that[1-3]

$$\underset{t\to\infty}{\text{w-lim}}\, W_{\beta\alpha}^{\pm}(t) = 0 \quad (\beta, \alpha = \text{arbitrary channels}), \qquad (1.2)$$

which is important because it implies that the ranges of the channel wave operators $\Omega_\alpha^{\pm}$ are orthogonal subspaces of $\mathcal{H}_N$. That is,

$$\Omega_\beta^{+*}\Omega_\alpha^{+} = \delta_{\beta\alpha}P_\alpha \quad \text{and} \quad \Omega_\beta^{-*}\Omega_\beta^{-} = \delta_{\beta\alpha}P_\alpha. \qquad (1.3)$$

Later, in a paper on the problem of asymptotic completeness, Combes[4] proved that

$$\underset{t\to\infty}{\text{s-lim}}\, W_{\beta\alpha}^{\pm}(t) = 0 \quad (\beta, \alpha = \text{two-body channels}). \qquad (1.4)$$

The matter now rests at this point.

It is natural to ask if the restriction in Eq. (1.4) can be extended to include breakup channels. The detailed forms for $H_\alpha$ and $P_\alpha$ given, for example, by Hunziker[3] show that in Eq. (1.1) the operators $H_\beta$ and $H_\alpha$ may be replaced by commuting self-adjoint operators $T_\beta$ and $T_\alpha$. These commuting operators are, up to a constant, the kinetic energy operators of the various channels. The operator $T_\alpha$ has the further property that it commutes not only with $P_\alpha$ but also with the projections $P_\beta$ for all breakup channels of $\alpha$. Thus, if $\beta$ is a breakup channel of $\alpha$, $\|W_{\beta\alpha}^{\pm}(t)\psi\| = \|P_\beta P_\alpha\psi\|$ for all $\psi$ in $\mathcal{H}_N$. Since $P_\beta P_\alpha$ is not, in general, zero, it follows that the strong limit of $W_{\beta\alpha}^{\pm}(t)$ cannot be zero. Indeed, because of Eq. (1.2), the strong limit cannot even exist. Thus, in any multichannel theory in which breakup channels are included, the channels in Eq. (1.4) must be restricted so that $\beta$ is not a breakup channel of $\alpha$.[5]

Among the interesting implications of the result of the foregoing paragraph is that the method used by Combes to prove asymptotic completeness fails at energies $\lambda_0$ above the breakup threshold. An essential ingredient in his proof is that for all channels $\beta$ that are open at energy $\lambda_0$ the adjoint wave operators $\Omega_\beta^{\pm *}$ satisfy

$$\Omega_\beta^{\pm *}E_N(\lambda_0) = \underset{t\to\infty}{\text{s-lim}}\, P_\beta \exp(\pm iH_\beta t)\exp(\mp iH_N t)P_N E_N(\lambda_0). \qquad (1.5)$$

Here $H_N$ is the full $N$-particle Hamiltonian with spectral family $E_N(\cdot)$, and $P_N$ is the orthogonal projection of $\mathcal{H}_N$ onto the subspace of absolute continuity of $H_N$. Multiplying Eq. (1.5) from the right by $\Omega_\alpha^{\pm}$ and applying standard techniques of abstract time-dependent scattering theory, one obtains

$$\underset{t\to\infty}{\text{s-lim}}\, W_{\beta\alpha}^{\pm}(t)E_\alpha(\lambda_0) = 0 \quad (\beta, \alpha = \text{open channels}), \qquad (1.6)$$

as a necessary consequence of Eq. (1.5). Here $E_\alpha(\cdot)$ denotes the spectral family of $H_\alpha$. But the argument of the preceding paragraph is precisely that the strong limit does not exist if $\beta$ is a breakup channel of $\alpha$. One is forced to conclude that Eq. (1.5) is not true above the breakup threshold, a conclusion agreeing with previous less general results.[6]

Our curiosity being thus aroused, and motivated by other problems we encountered in a recent paper,[7] we decided to study strong Abel limits of the operators $W_{\beta\alpha}^{\pm}(t)$. This paper contains the results of our study.

In Sec. 2 we formulate the problem in a two Hilbert space framework and prove (Theorem 1) that the two Hilbert space analog of $W_{\beta\alpha}^{\pm}(t)$ has zero as a strong Abel limit. The proof requires that certain estimates be uniform in the channel indices, which forces us to introduce a technical assumption, Assumption ($\Sigma$). Concerning this unwanted intrusion, we are consoled by the fact that most physical systems appear to satisfy

 0022-2488/78/1907-1610$1.00  

the assumption. In particular, if only a finite number of channels are included, the assumption is satisfied.

In Sec. 3 we discuss some of the ramifications of Theorem 1.

We show (Theorem 2) in Sec. 3A that a generalized Lippmann identity[8,9] is an almost immediate consequence of Theorem 1. This identity has been used frequently in recent work on the $N$-particle problem.[10-14] The version used there is, however, valid only for two-body channels,[14] while the version given by Theorem 2 is valid for arbitrary channels.

In Sec. 3B we prove (Theorem 3) that a certain symmetric transition operator $T$ gives the same scattering operator as certain asymmetric operators $T^{(\pm)}$. The symmetric operator $T$ is the two Hilbert space version of the transition operators of Alt, Grassberger, and Sandhas,[15,16] while the asymmetric operators $T^{(\pm)}$ correspond to the prior and post operators used by Lovelace.[17]

Arguments for this equivalence have long been known,[1,15-17] and typically go as follows. One supposes that $T^{(1)}_{\beta\alpha}\ (E+i\epsilon)$ and $T^{(2)}_{\beta\alpha}\ (E+i\epsilon)$ are two candidates, so-called off shell extensions, for the transition operator from channel $\alpha$ to channel $\beta$. The difference of the two is then shown to have the form

$$T^{(1)}_{\beta\alpha}\ (E+i\epsilon)-T^{(2)}_{\beta\alpha}\ (E+i\epsilon)=\Delta_{\beta\alpha}(E+i\epsilon)[E+i\epsilon-H_\alpha], \tag{1.7}$$

where $\Delta_{\beta\alpha}(z)$ is an operator-valued function that is analytic in a neighborhood of the energy shell. Then, since $[E+i\epsilon-H_\alpha]$ vanishes on the energy shell, and since $\Delta_{\beta\alpha}(z)$ is well behaved on the energy shell, the difference in Eq. (1.7) is asserted to vanish on the energy shell. This is taken to mean that $T^{(1)}_{\beta\alpha}$ and $T^{(2)}_{\beta\alpha}$ correspond to the same scattering operator.

The proof that such an argument is flawed[18] is by counterexample. Let $\Delta_{\beta\alpha}(z)=\Omega_\beta^{+*}P_\alpha$. This is a particularly nice operator-valued analytic function, being bounded and independent of $z$. But with this choice of $\Delta_{\beta\alpha}$ the right side of Eq. (1.7) not only does not give zero contribution to the scattering operator $S_{\beta\alpha}$, but gives $S_{\beta\alpha}$ itself.[6]

The mathematically correct procedure is first to substitute the difference in Eq. (1.7) into the spectral integrals which relate transition operators $T_{\beta\alpha}(E+i\epsilon)$ to the scattering operator $S_{\beta\alpha}$, and then to prove that the limit as $\epsilon\to 0$ is zero. We use this rigorous procedure in Sec. 3B, thereby closing a gap in the previous proofs.

Section 3C is concerned with a further question concerning the spectral integrals relating the two Hilbert space transition operators $T(E+i\epsilon)$ and $T^{(\pm)}\ (E+i\epsilon)$ to the scattering operator $S$. It has been proved[6] that the spectral integrals are first to be evaluated and then the limit $\epsilon\to 0$ taken in the weak operator topology. If the formulas are true only in the weak topology, then problems are posed[6] for the standard time-independent proof of the unitarity of $S$. Hence the question arises whether the formulas are actually valid in the strong topology. Section 3C contains the results (Theorems 4, 5, and 6) that we can offer on the topic.

Finally, in Sec 3D we discuss a different limit process that was introduced by Obermann and Wollenberg.[19] Their limit is intermediate in "strength" between the strong limit as $t\to\infty$ and the strong Abel limit. We conclude (Theorem 7) that most of the limits sought in this paper do not exist in their sense.

## 2. THE MAIN THEOREM

The notation of Ref. 7 is adopted.

The dynamics of the full $N$-particle system is specified by the Hamiltonian $H_N$, which is a selfadjoint operator on the full $N$-particle Hilbert space $\mathcal{H}_N$. The spectral family of $H_N$ is denoted by $E_N(\lambda)$.

Asymptotically the particles are grouped into clusters. For a given clustering $A$ of particles there is a subspace $\mathcal{H}_A$ of $\mathcal{H}_N$ that contains the asymptotic states with that clustering. The projection operator $P_A$ that projects $\mathcal{H}_N$ onto $\mathcal{H}_A$ can be written

$$P_A=\sum_\alpha^{(A)}P_\alpha, \tag{2.1}$$

where the strong topology sum is over channels $\alpha$ with clustering $A$, and where $P_\alpha$ is the projection of $\mathcal{H}_N$ onto the subspace $\mathcal{H}_\alpha$ of asymptotic states appropriate to channel $\alpha$. For a given clustering $A$ the various $P_\alpha$ are mutually orthogonal.

The subspaces $\mathcal{H}_A$ are combined into a direct sum space,

$$\mathcal{H}\equiv\bigoplus_A\mathcal{H}_A, \tag{2.2}$$

which is one of the two Hilbert spaces of the theory, the other being $\mathcal{H}_N$. A (singular) mapping $J\colon\mathcal{H}\to\mathcal{H}_N$ is defined by

$$J\Phi\equiv\sum_A\phi_A \tag{2.3}$$

for all $\Phi=\oplus_A\phi_A$ in $\mathcal{H}$. The adjoint $J^*$ of $J$ is then given by

$$J^*\psi\equiv\bigoplus_A P_A\psi \tag{2.4}$$

for all $\psi$ in $\mathcal{H}_N$.

The asymptotic cluster Hamiltonians $H_A$ have the form

$$H_A=H_A^0+\hat{H}_A. \tag{2.5}$$

The operators $H_A^0$ are sums of Laplacians in appropriate variables and commute with $P_A$. The operators $H_A$ have on $\mathcal{H}_A$ the form given by

$$\hat{H}_AP_A=\sum_\alpha^{(A)}\lambda_\alpha P_\alpha. \tag{2.6}$$

The $\lambda_\alpha$ are sums of eigenvalues of appropriate subsystem Hamiltonians. The cluster Hamiltonians are combined into an operator $H$ that is defined on $\mathcal{H}$ by

$$H\Phi=\bigoplus_A H_A\phi_A. \tag{2.7}$$

The spectral family of the self-adjoint operator $H$ is denoted by $E(\lambda)$.

Wave operators $\Omega^\pm$ are now defined by

$$\Omega^\pm\equiv \underset{t\to\pm\infty}{\text{s-lim}}\,\Omega(t), \tag{2.8}$$

where $\Omega(t)$ is defined by

$$\Omega(t) \equiv \exp(iH_N t) J \exp(-iHt). \qquad (2.9)$$

These wave operators are partial isometries,

$$\Omega^{\pm *}\Omega^{\pm} = I \quad \text{and} \quad \Omega^{\pm}\Omega^{\pm *} = E_N^{\pm}, \qquad (2.10)$$

where $I$ is the identity on $\mathcal{H}$ and $E_N^{\pm}$ are the orthogonal projections of $\mathcal{H}_N$ onto the ranges of the wave operators $\Omega^{\pm}$.

The adjoint wave operators $\Omega^{\pm *}$ satisfy

$$\Omega^{\pm *} = \text{w-}\lim_{t\to\pm\infty} \Omega^*(t). \qquad (2.11)$$

where $\Omega^*(t)$ is the adjoint of $\Omega(t)$ defined in Eq. (2.9). As we have seen in Sec. 1 the weak limit in Eq. (2.11) may not, in general, be replaced by the strong limit.

The two Hilbert space analog of $W_{\beta\alpha}^{\pm}(t)$ defined in Eq. (1.1) is the operator

$$W(t) \equiv \Omega^*(t)\Omega(t) - I = \exp(iHt)(J^*J - I)\exp(-iHt). \qquad (2.12)$$

It is now necessary to assume something about the set of numbers $\lambda_\alpha$ in Eq. (2.6).

*Assumption* $(\Sigma)$: The set

$$\Sigma \equiv \text{closure of } \{x \mid x = \lambda_\alpha \text{ for some channel } \alpha\} \qquad (2.13)$$

has Lebesgue measure zero.

Assumption $(\Sigma)$ is automatically satisfied if there are only a finite number of possible bound states for each cluster of particles and, therefore, only a finite number of channels. This is thought to be commonly the case in nuclear physics where the interactions have short range. It is also true if the set of $\lambda_\alpha$'s has accumulation points which are either finite in number or themselves have a finite number of accumulation points. This is the case, for example, for interactions described by dilation analytic pair potentials such as the Coulomb potential. Although it seems reasonable to believe that all quantum mechanical problems of interest would satisfy Assumption $(\Sigma)$, we know of no general proof of that fact. What is known about the problem has been recently reviewed by Hunziker[20] and by Simon.[21]

The main mathematical result is the following.

*Theorem* 1: If Assumption $(\Sigma)$ is satisfied, then

$$\text{s-}\lim_{\epsilon\to 0^+} L^{\pm}(\epsilon) = 0, \qquad (2.14)$$

where

$$L^{\pm}(\epsilon) \equiv \epsilon\int_0^\infty dt \exp(-\epsilon t) W(\pm t). \qquad (2.15)$$

*Proof*: For all $\Phi = \oplus\, \phi_A$ in $\mathcal{H}$ the vector $L^{\pm}(\epsilon)\Phi$ can be written as

$$L^{\pm}(\epsilon)\Phi = \bigoplus_B L_{BA}^{\pm}(\epsilon)\,\phi_A, \qquad (2.16)$$

where the cluster matrix elements[7] $L_{BA}^{\pm}$ are given by

$$L_{BA}^{\pm}(\epsilon) \equiv \epsilon\int_0^\infty dt \exp(-\epsilon t) W_{BA}(\pm t), \qquad (2.17)$$

$$= \epsilon\int_0^\infty dt \exp(-\epsilon t)\exp(\pm iH_B t)(P_B P_A - \delta_{BA} P_A) \times \exp(\mp iH_A t). \qquad (2.18)$$

Since there are only a finite number of clusterings and hence only a finite number of operators $L_{BA}^{\pm}(\epsilon)$, it suffices to prove

$$\text{s-}\lim_{\epsilon\to 0^+} L_{BA}^{\pm}(\epsilon) = 0. \qquad (2.19)$$

Because $L_{BA}^{\pm}(\epsilon)$ is uniformly bounded, $\| L_{BA}(\epsilon)\| \leq 1$, it suffices to prove Eq. (2.19) on a dense subset of $\mathcal{H}_A$. Such a subset is provided by vectors $\phi_A$ of the form $\phi_A = \Sigma_\alpha^{(A)}\phi_\alpha$ with only a finite number of $\phi_\alpha$ different from zero. The linearity of $L_{BA}^{\pm}(\epsilon)$ then implies that it is sufficient to consider $\phi_A$ with only one nonzero $\phi_\alpha$. Since $L_{AA}^{\pm}(\epsilon) \equiv 0$, we are thus led to prove that

$$\text{s-}\lim_{\epsilon\to 0^+} L_{BA}^{\pm}(\epsilon)\phi_\alpha = 0 \quad (B \neq A), \qquad (2.20)$$

where

$$L_{BA}^{\pm}(\epsilon)\phi_\alpha = \epsilon\int_0^\infty dt \exp(-i\epsilon t)\exp(\pm iH_B t) P_B \times \exp(\mp iH_A t)\phi_\alpha \quad (B \neq A). \qquad (2.21)$$

Define now the operators [cf. Eq. (2.5)]

$$K_B \equiv \hat{H}_B - \lambda_\alpha I_N \qquad (2.22)$$

and

$$F_B \equiv H_B^0 - H_A^0, \qquad (2.23)$$

where $I_N$ is the identity on $\mathcal{H}_N$. By taking note of the commutation properties of these operators, one can write

$$\exp(\pm iH_B t)P_B \exp(\mp iH_A t)\phi_\alpha = \exp(\pm iK_B t)P_B \times \exp(\pm iF_B t)\phi_\alpha. \qquad (2.24)$$

The operator $F_B$ is self-adjoint with only absolutely continuous spectrum. Denote its spectral family by $F_B(\lambda)$. Then Eq. (2.21) may be written in the form

$$L_{BA}^{\pm}(\epsilon)\phi_\alpha = \epsilon\int_0^\infty dt \exp(-\epsilon t)\exp(\pm iK_B t)P_B \times \int_\lambda \exp(\pm i\lambda t)\, dF_B(\lambda)\phi_\alpha. \qquad (2.25)$$

Interchange of the order of integration is justified by Lemma 2 of Ref. 6, with $U(t,\lambda)$ and $B(t)$ of that lemma being identified with $\epsilon\exp(-\epsilon t \pm i\lambda t)$ and $\exp(\pm iK_B t)P_B$, respectively. The result after evaluating the Bochner integral is

$$L_{BA}^{\pm}(\epsilon)\phi_\alpha = \pm i\epsilon\int_\lambda [\lambda \pm i\epsilon + K_B]^{-1} P_B\, dF_B(\lambda)\phi_\alpha. \qquad (2.26)$$

We must prove that for every $\delta > 0$ there exists an $\epsilon_0 > 0$ such that $\|L_{BA}^{\pm}(\epsilon)\phi_\alpha\| < \delta$ for $\epsilon < \epsilon_0$.

First choose $a$ and $b$ such that

$$\|L_{BA}^{\pm}(\epsilon)[I_N - F_B(b) + F_B(a)]\phi_\alpha\| < (\delta/3) \qquad (2.27)$$

for all $\epsilon > 0$. This can be done since $L_{BA}^{\pm}(\epsilon)$ is uniformly bounded, and since

$$\text{s-}\lim_{\lambda\to\infty} F_B(\lambda) = I_N \quad \text{and} \quad \text{s-}\lim_{\lambda\to-\infty} F_B(\lambda) = 0. \qquad (2.28)$$

Next, consider the set $S[a,b] \equiv S \cap [a,b]$, where

$$S \equiv \text{closure of } \{x \mid x = \lambda_\beta - \lambda_\alpha \text{ for some } \beta \text{ with clustering } B\}. \qquad (2.29)$$

Since, by Assumption $(\Sigma)$, the set $S$ has Lebesgue measure zero, the set $S[a,b]$ is compact and has measure zero. It is therefore possible to cover $S[a,b]$ by a finite collection $N = \cup_{i=1}^{r}(\alpha_i, \beta_i)$ of disjoint open intervals such that

$$d = \inf\{|x-y| : x \in S[a,b] \text{ and } y \in [a,b]-N\} \tag{2.30}$$

is greater than zero and such that

$$\| F_B(N \cap [a,b])\phi_\alpha \| < (\delta/3). \tag{2.31}$$

This last requirement is possible because of the absolute continuity (with respect to Lebesgue measure) of $F_B$. Since $\| L^{\pm}_{BA}(\epsilon)\| \le 1$, one has

$$\| L^{\pm}_{BA}(\epsilon) F_B(N \cap [a,b])\, \phi_\alpha \| < (\delta/3). \tag{2.32}$$

It remains to show that

$$\| L^{\pm}_{BA}(\epsilon) F_B(N')\phi_\alpha \| < (\delta/3) \tag{2.33}$$

for all $\epsilon < \epsilon_0$, where $\epsilon_0$ is some positive number and $N' \equiv [a,b] - N \cap [a,b]$.

The set $N'$ consists of a finite number of disjoint closed intervals $[x_i, y_i]$. Thus

$$L^{\pm}_{BA}(\epsilon)F_B(N')\phi_\alpha = \sum_i L^{\pm}_{BA}(\epsilon) F_B([x_i, y_i])\phi_\alpha \tag{2.34}$$

$$= \sum_i (\pm i\epsilon) \int_{x_i}^{y_i} [\lambda \pm i\epsilon + K_B]^{-1} P_B \, dF_B(\lambda)\phi_\alpha. \tag{2.35}$$

The spectral integrals in Eq. (2.35) can be converted to Bochner integrals by an integration by parts.[22,23] The result is

$$\begin{aligned}\int_{x_i}^{y_i} [\lambda \pm i\epsilon + K_B]^{-1} P_B \, dF_B(\lambda)\phi_\alpha &= [y_i \pm i\epsilon + K_B]^{-1} P_B F_B(y_i)\phi_\alpha \\ &\quad - [x_i \pm i\epsilon + K_B]^{-1} P_B F_B(x_i)\phi_\alpha \\ &\quad + \int_{x_i}^{y_i} d\lambda [\lambda \pm i\epsilon + K_B]^{-2} P_B F_B(\lambda)\phi_\alpha.\end{aligned} \tag{2.36}$$

On the set $N'$ the operators $[\lambda \pm i\epsilon + K_B]^{-k} P_B$ have bounds given by

$$\| [\lambda \pm i\epsilon + K_B]^{-k} P_B \| \le d^{-k} \quad (k = 1, 2). \tag{2.37}$$

The right side of Eq. (2.36) is therefore uniformly bounded in $\epsilon$, and

$$\left\| \int_{x_i}^{y_i} [\lambda \pm i\epsilon + K_B]^{-1} P_B \, dF_B(\lambda)\phi_\alpha \right\| \le d^{-2}(2d + y_i - x_i) \| \phi_\alpha \|. \tag{2.38}$$

Combining this result with Eq. (2.35) yields

$$\| L^{\pm}_{BA}(\epsilon) F_B(N')\phi_\alpha \| \le \epsilon \sum_i d^{-2}(2d + y_i - x_i) \| \phi_\alpha \|. \tag{2.39}$$

It is now clear that $\epsilon_0$ can be chosen so that Eq. (2.33) is true for all $\epsilon < \epsilon_0$.

Combining Eqs. (2.27), (2.31), and (2.33) finishes the proof of the theorem. QED

To gain a certain perspective on Theorem 1, it is useful to rewrite it once more in terms of the channel and cluster matrix elements.[7]

*Corollary*: For all channels $\beta$ and $\alpha$

$$\operatorname*{s-lim}_{\epsilon \to 0^+} \epsilon \int_0^\infty dt \exp(-\epsilon t) W^{\pm}_{\beta\alpha}(t) = 0, \tag{2.40}$$

where $W^{\pm}_{\beta\alpha}(t)$ is defined by Eq. (1.1). If Assumption $(\Sigma)$ is satisfied, then

$$\operatorname*{s-lim}_{\epsilon \to 0^+} L^{\pm}_{BA}(\epsilon) = 0, \tag{2.41}$$

where $L^{\pm}_{BA}(\epsilon)$ is defined by Eq. (2.18).

The proof is immediate from Eq. (2.14) upon taking channel or cluster matrix elements. Assumption $(\Sigma)$ is not needed for Eq. (2.40) since its sole purpose was to enable us to handle an infinite number of channels simultaneously.

## 3. RAMIFICATIONS

### A. The Lippmann identity

An alternative form of Theorem 1 is interesting in that it provides a generalization of the so-called Lippmann identity.[8,9]

*Theorem* 2: If Assumption $(\Sigma)$ is satisfied, then

$$\operatorname*{s-lim}_{\epsilon \to 0^+} (\mp i\epsilon) \int_\lambda (\lambda - H \mp i\epsilon)^{-1} (J^* J - I)\, dE(\lambda) = 0, \tag{3.1}$$

and

$$\operatorname*{s-lim}_{\epsilon \to 0^+} (\pm i\epsilon) \int_\lambda dE(\lambda)(J^* J - I)(\lambda - H \pm i\epsilon)^{-1} = 0. \tag{3.2}$$

*Proof*: Substitute the spectral resolutions

$$\exp(\pm iHt) = \int_\lambda \exp(\pm i\lambda t)\, dE(\lambda), \tag{3.3}$$

for the left- or right-hand exponentials of Eq. (2.15). Apply Lemma 2 of Ref. 6 to justify the interchange of Bochner and spectral integrations. Equations (3.1) and (3.2) result immediately from evaluation of the Bochner integrals. QED

*Corollary*: Let $E_A(\lambda)$ be the spectral family associated with the cluster Hamiltonian $H_A$. If Assumption $(\Sigma)$ is satisfied, then

$$\operatorname*{s-lim}_{\epsilon \to 0^+} (\mp i\epsilon) \int_\lambda (\lambda - H_B \mp i\epsilon)^{-1} P_B P_A \, dE_A(\lambda) = \delta_{BA} P_A, \tag{3.4}$$

and

$$\operatorname*{s-lim}_{\epsilon \to 0^+} (\pm i\epsilon) \int_\lambda dE_B(\lambda) P_B P_A (\lambda - H_A \pm i\epsilon)^{-1} = \delta_{BA} P_A. \tag{3.5}$$

*Proof*: The proof is immediate upon taking cluster matrix elements of Eqs. (3.1) and (3.2). QED

Note that just as with Eq. (2.41) the cluster labels $A$ and $B$ in Eqs. (3.4) and (3.5) can be replaced with channel labels $\alpha$ and $\beta$. If this is done, Assumption $(\Sigma)$ can be dropped since its sole function was to allow us to handle a possibly infinite number of channels simultaneously.

We also remark that the Lippmann identity is usually expressed[8,9] in the form

$$\operatorname*{s-lim}_{\epsilon \to 0^+} (\mp i\epsilon)(\lambda - H_B \mp i\epsilon)^{-1} | \phi_\alpha(\lambda)\rangle = \delta_{BA} | \phi_A(\lambda)\rangle, \tag{3.6}$$

where $|\phi_\alpha(\lambda)\rangle$ is an improper eigenfunction of $H_A$ with improper eigenvalue $\lambda$. The spectral integrals in Eqs. (3.4) and (3.5) reflect, we believe, the proper mathematical way to deal with these improper eigenfunctions. Another difference between Eq. (3.6) and Eq. (3.4) is the lack of the projection operator $P_B$ in Eq. (3.6). This is the reason why Eq. (3.6) is valid only when $A$ has two fragments.[14] On the other hand, Eqs. (3.4) and (3.5) are valid for all clusterings $A$ and $B$.

### B. On-shell equivalence of $T^{(\pm)}$ and $T$

For scattering with short-range interactions we have previously discussed[7] the question of whether the symmetric transition operator of the type of Alt, Grassberger, and Sandhas and the nonsymmetric operators of the type used by Lovelace correspond to the same scattering amplitude. This question has been answered previously in the affirmative,[15-17] but with less rigor than is presented here.

The symmetric operator $T(z)$ is defined on the domain of $H$ by

$$T(z) \equiv (z-H)\{J^*(z-H_N)^{-1}J-(z-H)^{-1}\}(z-H), \qquad (3.7)$$

where $z$ is understood to be in the resolvent set of $H_N$, the full $N$-body Hamiltonian. The connection between $T(z)$ and the scattering operator $S$ is given by the formulas[6]

$$S-I = \operatorname*{w-lim}_{\epsilon\to 0^+}(-2\pi i)\int_\lambda\int_\mu dE(\lambda)\delta_\epsilon(\lambda-\mu) \times T([\lambda+\mu+i\epsilon]/2)\,dE(\mu), \qquad (3.8)$$

$$= \operatorname*{w-lim}_{\epsilon_1\to 0^+}\ \operatorname*{s-lim}_{\epsilon_2\to 0^+}(-2\pi i)\int_\lambda\int_\mu dE(\mu)\delta_{\epsilon_1}(\lambda-\mu) \times T(\lambda+i\epsilon_2)dE(\lambda), \qquad (3.9)$$

$$= \operatorname*{s-lim}_{\epsilon_1\to 0^+}\ \operatorname*{w-lim}_{\epsilon_2\to 0^+}(-2\pi i)\int_\lambda\int_\mu dE(\lambda)T(\lambda+i\epsilon_2)\delta_{\epsilon_1}(\lambda-\mu)dE(\mu). \qquad (3.10)$$

In Eqs. (3.8)–(3.10) the function $\delta_\epsilon(x)$ is defined by $\delta_\epsilon(x) \equiv (\epsilon/\pi)(\epsilon^2+x^2)^{-1}$. The spectral integrals are repeated (not double) integrals. The ones in Eq. (3.8) may be done in either order, while in the other equations the $\mu$ integration must be done first.

The nonsymmetric operators are defined by

$$T^{(+)}(z) \equiv (z-H)\{J^*(z-H_N)^{-1}J-(z-H)^{-1}J^*J\}(z-H), \quad (3.11)$$

$$T^{(-)}(z) \equiv (z-H)\{J^*(z-H_N)^{-1}J-J^*J(z-H)^{-1}\}(z-H). \quad (3.12)$$

The question is, then, whether $T^{(\pm)}(z)$ can be substituted for $T(z)$ in Eqs. (3.8)–(3.10).

To establish that $T^{(\pm)}$ can replace $T$ in Eqs. (3.8)–(3.10), it is sufficient to prove that if the differences

$$\Delta^+(z) \equiv T(z)-T^{(+)}(z) = (J^*J-I)(z-H) \qquad (3.13)$$

and

$$\Delta^-(z) \equiv T(z)-T^{(-)}(z) = (z-H)(J^*J-I) \qquad (3.14)$$

are substituted into the spectral integrals of those equations and the various $\epsilon$ limits taken, the resulting limits should be zero. This is straightforward[7] when Eq. (3.13) is used in Eq. (3.9) and when Eq. (3.14) is used in Eq. (3.10).

To go farther, one must substitute the differences in Eqs. (3.13) and (3.14) into Eqs. (3.8)–(3.10), and express the resulting integrals in terms of integrals of the resolvent operators. These resolvent operators are expressed, in turn, as Bochner integrals involving the exponentials $\exp(\pm iHt)$. The order of Bochner and spectral integration is reversed, resulting in the following equations:

$$(-2\pi i)\int_\lambda\int_\mu dE(\lambda)\delta_\epsilon(\lambda-\mu)\Delta^\pm([\lambda+\mu+i\epsilon]/2)dE(\mu) = L^\mp(\epsilon); \qquad (3.15)$$

$$(-2\pi i)\int_\lambda\int_\mu dE(\mu)\delta_{\epsilon_1}(\lambda-\mu)\Delta^-(\lambda+i\epsilon_2)dE(\lambda) = (\epsilon_1^{-1}\epsilon_2+1)L^+(\epsilon_1)+(\epsilon_1^{-1}\epsilon_2-1)L^-(\epsilon_1); \qquad (3.16)$$

$$(-2\pi i)\int_\lambda\int_\mu dE(\lambda)\delta_{\epsilon_1}(\lambda-\mu)\Delta^+(\lambda+i\epsilon_2)dE(\mu) = (\epsilon_1^{-1}\epsilon_2+1)L^-(\epsilon_1)+(\epsilon_1^{-1}\epsilon_2-1)L^+(\epsilon_1). \qquad (3.17)$$

The operators $L^\pm(\epsilon)$ are defined in Eq. (2.15).

It is apparent from Assumption (W4) of Ref. 6, which for multichannel systems of the type discussed here was verified in the appendix of that paper, that

$$\operatorname*{w-lim}_{\epsilon\to 0^+} L^\pm(\epsilon) = 0. \qquad (3.18)$$

The assertion of Theorem 1 is that the weak limit can be replaced by a strong limit if Assumption ($\Sigma$) is true.

When combined with the preceding analysis, the remark of the previous paragraph yields a rigorous proof of the following theorem.

*Theorem* 3: Equations (3.8) and (3.9) are true with the operator $T$ replaced by $T^{(\pm)}$. Equation (3.10) is true with the operator $T$ replaced by $T^{(-)}$. In addition, if Assumption ($\Sigma$) is true, then Eq. (3.10) is true with $T$ replaced by $T^{(+)}$.

This theorem establishes, for all practical purposes, the rigorous equivalence of the operators $T(z)$ and $T^{(\pm)}(z)$ to the extent that they yield the same scattering operator for particles with short-range interactions.

### C. On replacing weak limits by strong limits

The reason for the presence of weak limits in Eqs. (3.8)–(3.10) is that the weak limit in Eq. (2.11) cannot be replaced by the strong limit. Theorem 1 raises a new possibility, however, which is based on the following theorem.

*Theorem* 4: If Assumption ($\Sigma$) is true, then

$$\Omega^{\pm *} = \operatorname*{s-lim}_{\epsilon\to 0^+}\epsilon\int_0^\infty dt\,\exp(-\epsilon t)\Omega^*(\pm t)E_N^\pm, \qquad (3.19)$$

where $\Omega^*(t)$ is the adjoint of $\Omega(t)$ defined in Eq. (2.9) and where $E_N^\pm$ is defined in Eq. (2.10).

*Proof*: Starting with Eq. (2.8), it is an easy matter to prove that

$$0 = \operatorname*{s-lim}_{\epsilon\to 0^+}\epsilon\int_0^\infty dt\,\exp(-\epsilon t)\Omega^*(\pm t)\{\Omega(\pm t)-\Omega^\pm\}. \qquad (3.20)$$

Combining Eq. (3.20) with Theorem 1 and Eq. (2.10) yields Eq. (3.19) QED

Since the time independent formulas for the scattering operator $S$ involve only Abel limits and not the direct time limits, the appearance of $\Omega^{\pm *}$ as a strong Abel limit reopens the question of whether the weak limits in Eqs. (3.8)–(3.10) can be replaced by strong limits.

In this direction we can prove the following.

*Theorem* 5: If Assumption ($\Sigma$) is true, and if $S$ is unitary, then the weak limits in Eqs. (3.8) and (3.9) can be replaced by strong limits.

*Proof*: From Eq. (2.8) it is apparent that

$$\operatorname*{s-lim}_{\epsilon\to 0^+}\epsilon\int_0^\infty dt\exp(-\epsilon t)\{\Omega^*(t)\Omega(-t)-I\}$$
$$=\operatorname*{s-lim}_{\epsilon\to 0^+}\int_0^\infty dt\exp(-\epsilon t)\{\Omega^*(t)\Omega^- - I\}. \tag{3.21}$$

Using Eq. (3.19) and the well-known fact that $E_N^+=E_N^-$ is necessary and sufficient for $S$ to be unitary, one can rewrite the right side of Eq. (3.21). The result is

$$S-I=\Omega^{+*}\Omega^- - I$$
$$=\operatorname*{s-lim}_{\epsilon\to 0^+}\epsilon\int_0^\infty dt\exp(-\epsilon t)\{\Omega^*(t)\Omega(-t)-I\}. \tag{3.22}$$

The procedure used in the proof of Theorem 4 of Ref. 6 now proves the theorem for Eq. (3.8). To prove the result for Eq. (3.9), one follows the proof of Theorem 5 of Ref. 6. Equation (3.53) of Ref. 6, can, however, now be replaced by

$$S-I=\operatorname*{s-lim}_{\epsilon\to 0^+}\epsilon\int_0^\infty dt\exp(-\epsilon t)$$
$$\times[\exp(+iHt)J^*\Omega^-\exp(-iHt)$$
$$-\exp(-iHt)J^*\Omega^-\exp(iHt)]. \tag{3.23}$$

Equation (3.23) follows from the fact that $E_N^+=E_N^-$, Theorem 4, and from the intertwining property. The procedure used to prove Theorem 5 of Ref. 6 now yields the desired result for Eq. (3.9). QED

The assumptions of Theorem 5 are not sufficient to allow one to prove that the weak limit in Eq. (3.10) can be replaced by a strong limit. To that end, we can offer the following result.

*Theorem* 6: Suppose that

$$0=\operatorname*{s-lim}_{\epsilon\to 0^+}\epsilon\int_0^\infty dt\exp(-\epsilon t)\Omega^*(t)(I_N-E_N^+), \tag{3.24}$$

where $I_N$ is the identity on $\mathcal{H}_N$, and that Assumption ($\Sigma$) is true. Then, the weak limits can be replaced by strong limits in Eqs. (3.8)–(3.10).

*Proof*: If Eq. (3.24) is true, then Eq. (3.19) is true with $E_N^+$ replaced by $I_N$. The proof of Theorem 5 of Ref. 6 then is true with all limits being strong limits. QED

Equation (3.24) is especially interesting in that it is a necessary, but doubtless not sufficient, condition for $\Omega^+$ to be asymptotically complete in the sense of Kato.[24] Assuming that $H_N$ has no singularly continuous spectrum, and assuming that $\Omega^+$ is asymptotically complete so that $P_N=E_N^+$, then $I_N-E_N^+$ is the orthogonal projection of $\mathcal{H}_N$ onto the subspace spanned by the eigenvectors of $H_N$. In this case Eq. (3.24) then follows from the fact that $H$ has only absolutely continuous spectrum, and hence $\epsilon(\epsilon+i\lambda-H)^{-1}$ has strong limit zero as $\epsilon\to 0$, uniformly in $\lambda$.

A proof of Eq. (3.24) from first principles is therefore of great intrinsic interest as a possible first step in a proof of asymptotic completeness.

Note also that although in Theorem 6 no assumptions about $\Omega^-$ are made, the weak limits in Eqs. (3.8)–(3.10) can be replaced by strong limits. In particular one can imagine that $E_N^+\neq E_N^-$, and, hence, that $S$ is not unitary. It follows that the replacement of weak limits in Eqs. (3.8), (3.9) with strong limits is not as directly tied to the unitarity of $S$, as Theorem 5 might suggest.

In summary, Theorem 4 does not imply that the weak limits in Eqs. (3.8)–(3.10) can be replaced by strong limits. There seems to be some, albeit indirect, connection between replacing the weak by strong limits and the unitarity of the scattering operator $S$. The objections raised in Ref. 6 to unitarity proofs based on differences $T(\lambda+i0)-T^*(\lambda+i0)$, evaluated as strong limits, therefore still stand unanswered.

### D. The Obermann-Wollenberg theory

The fact that the weak limit in Eq. (2.11) cannot be replaced by a strong limit means that Kato's two Hilbert space theorem (Theorem 6.3 of Ref. 24) on asymptotic completeness is not valid in the multichannel case in general (see, however, Ref. 4 for a related result in a more limited context). Obermann and Wollenberg,[19] within the context of two-particle scattering, have, however, developed a similar theory that involves only a restricted sort of Abel limit. In their theory one might hope to replace Eq. (2.11) by the stronger condition that

$$\lim_{\epsilon\to 0^+}\epsilon\int_0^\infty dt\exp(-\epsilon t)\|\Omega^*(\pm t)E_N^\pm\psi-\Omega^{\pm *}\psi\|^2=0. \tag{3.25}$$

The assumptions of Theorem 6.3 of Ref. 24 are sufficient to guarantee Eq. (3.25) with $E_N^\pm$ replaced by $P_N$. In the two-particle theory of Obermann and Wollenberg the validity of Eq. (3.25), with $E_N^\pm$ replaced by $P_N$, is necessary and sufficient for the wave operators to be complete (Theorem 3 of Ref. 19).

Alas, that aspect of the approach of Obermann and Wollenberg does not generalize to multichannel scattering, as the following theorem shows.

*Theorem* 7: Let $E_A^\pm$ be the orthogonal projections of $\mathcal{H}_N$ onto the ranges of the cluster wave operators

$$\Omega_A^\pm\equiv\operatorname*{s-lim}_{t\to\pm\infty}\exp(iH_Nt)\exp(-iH_At)P_A. \tag{3.26}$$

Then, Eq. (3.25) is not true for $\psi\in E_A^\pm\mathcal{H}_N\subset P_N\mathcal{H}_N$, $A\neq 0$.

*Proof*: A necessary and sufficient condition for Eq. (3.25) to be true is that

$$0=\lim_{\epsilon\to 0^+}\epsilon\int_0^\infty dt\exp(-\epsilon t)(\psi,[E_N^\pm\Omega(\pm t)\Omega^*(\pm t)E_N^\pm-E_N^\pm]\psi). \tag{3.27}$$

Substitute $E_N^\pm E_A^\pm=E_A^\pm$ and $JJ^*=I_N+\sum' P_B$, where $\sum'$ means sum over $B\neq 0$, into Eq. (3.27). The result, for $\psi=E_A^\pm\psi$, is

$$\epsilon\int_0^\infty dt\exp(-\epsilon t)(\psi,[E_N^\pm\Omega(\pm t)\Omega^*(\pm t)E_N^\pm-E_N^\pm]\psi)$$
$$=\sum_{B\neq 0}\epsilon\int_0^\infty dt\exp(-\epsilon t)\|P_B\exp(\mp iH_Nt)E_A^\pm\psi\|^2$$
$$\geqslant\epsilon\int_0^\infty dt\exp(-\epsilon t)\|P_A\exp(\mp iH_Nt)E_A^\pm\psi\|^2. \tag{3.28}$$

Using the method of Kato (Theorem X.3.3 of Ref. 25) we prove that

$$\operatorname*{s-lim}_{t\to\pm\infty}\exp(iH_At)\exp(-iH_Nt)E_A^\pm=\Omega_A^{\pm *}. \tag{3.29}$$

From this it is simple to see that, as $\epsilon \to 0^+$, the right side of Eq. (3.28) approaches

$$\|\Omega_A^{\pm *}\psi\|^2 = \|E_A^{\pm}\psi\|^2 = \|\psi\|^2. \tag{3.30}$$

Thus, when the limit is taken in Eq. (3.25), the limit, if it exists, is greater than $\|\psi\|^2$, and, hence, cannot be zero. QED

*Corollary*: If the limits exist,

$$\text{s-}\lim_{\epsilon\to 0^+}\epsilon\int_0^\infty dt\,\exp(-\epsilon t)P_N[\Omega(\pm t)\Omega^*(\pm t) - I_N]P_N \neq 0. \tag{3.31}$$

This corollary is stated because of the parallel structure of the integral in Eq. (3.31) and that of Eq. (2.15).

## ACKNOWLEDGMENTS

The authors are indebted to T. Trucano and to Professor R.C. Metzler for discussions concerning Theorem 1. C. Chandler is grateful to the Physikalisches Institut, and particularly to Professor W. Sandhas, for their hospitality during the preparation of the final manuscript at the University of Bonn. He is also indebted to the Minna-James-Heinemann-Stiftung, in collaboration with the NATO Senior Scientists Program, and to the University of New Mexico for financial support during that time.

[1]H. Ekstein, Phys. Rev. **101**, 880 (1956).
[2]J.M. Jauch, Helv. Phys. Acta **31**, 661 (1958).
[3]W. Hunziker, in *Lectures in Theoretical Physics*, edited by A.O. Barut and W.E. Britten (Gordon and Breach, New York 1968), Vol. X-A.
[4]J.M. Combes, Nuovo Cimento A 64, 111 (1969). See also Ref. 21.
[5]In all other cases the strong limit of $W^{\pm}_{\beta\alpha}(t)$ is indeed zero. See Theorem 5 of M. Wollenberg, *On the Inverse Problem in the Abstract Theory of Scattering*, ZIMM der AdW der DDR preprint (1977).
[6]C. Chandler and A.G. Gibson, J. Math. Phys. **14**, 1328 (1973).
[7]C. Chandler and A.G. Gibson, "$N$-body quantum scattering theory in two Hilbert spaces. I. The basic equations," J. Math. Phys. **18**, 2336 (1977).
[8]B.A. Lippmann, Phys. Rev. **102**, 264 (1956).
[9]M.L. Goldberger and K.M. Watson, *Collision Theory* (Wiley, New York, 1964), p. 198.
[10]Gy. Bencze, Nucl. Phys. A **210**, 568 (1973).
[11]E.F. Redish, Nucl. Phys. A **225**, 16 (1974).
[12]D.J. Kouri and F.S. Levin, Nucl. Phys. A **253**, 395 (1975).
[13]D.J. Kouri, H. Krüger and F.S. Levin, Phys. Rev. D **15**, 1156 (1977).
[14]P. Benoist-Guental, Phys. Lett. B **56**, 413 (1975).
[15]E.O. Alt, P. Grassberger, and W. Sandhas, Nucl. Phys. B **2**, 167 (1967).
[16]P. Grassberger and W. Sandhas, Nucl. Phys. B **2**, 181 (1967).
[17]C. Lovelace, Phys. Rev. B **135**, 1225 (1964). See also Ref. 1 and papers cited there.
[18]Calculations[8–14] involving the Lippmann identity often use the same type of argument, and may thus be similarly flawed.
[19]P. Obermann and M. Wollenberg, "Abel Wave Operators, Part I: General Theory," Math. Nachr., (to be published).
[20]W. Hunziker, Acta Phys. Austr. Suppl. **XVII**, 43 (1977).
[21]B. Simon, Commun. Math. Phys. **55**, 259 (1977).
[22]C. Chandler and A.G. Gibson, Indiana Univ. Math. J. **25**, 443 (1976).
[23]C. Chandler and A.G. Gibson, *Long-Range Invariance Principle Revisited*, Indiana Univ. Math. J. (to be published).
[24]T. Kato, J. Func. Anal. **1**, 342 (1967).
[25]T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, Berlin, 1976), 2nd ed.

# An investigation of some of the kinematical aspects of plane symmetric space-times[a)]

George T. Carlson, Jr. and John L. Safko

*University of South Carolina, Columbia, South Carolina 29208*
(Received 20 October 1977)

A short review of the literature on plane symmetric space-times (PSSTS) is given in the Introduction. The rest of the paper concerns itself with an investigation of some of the kinematical aspects of PSSTS, i.e., properties of PSSTS which do not depend on the field equations. In particular, the existence of four special coordinate systems is considered. It is shown that the existence of these coordinate systems is not guaranteed for a general $C^k$ ($k \leqslant 1$) plane symmetric metric (PSM). For $k = 2$, two of the coordinate systems exist in a weak sense whereas the existence of the other two is not guaranteed in any sense. A local intrinsic type classification is introduced in Sec. 3, and it is shown that the existence of an extra Killing vector is correlated to the classification. Finally, the local equivalence of two given PSSTS is considered in Sec. 4. It is shown that some algebraic equations arise from the analysis. These algebraic equations may lead directly to the solution of the problem of the local equivalence of two given PSSTS.

## INTRODUCTION

PSSTS were defined by Taub.[1] Taub found a static solution ($g_s$) to $R_{\mu\nu}=0$. Davis and Ray[2] have shown that there is a homogeneous solution ($g_h$) to $R_{\mu\nu}=0$ and pointed out that there is apparently no natural way to join $g_s$ and $g_h$, thus obtaining a solution in an extended manifold. Novotný[3] has pointed out that $g_s$ and $g_h$ are the only known solutions to $R_{\mu\nu}=0$ with plane symmetry and have given as generalization of $g_s$ and $g_h$, solutions to $R_{\mu\nu}-\frac{1}{2}g_{\mu\nu}R+\Lambda g_{\mu\nu}=0$ with plane symmetry. Bonnor[4] has shown that a Robinson–Trautman metric, which contains a singular hypersurface $\rho=0$, may be transformed to $g_s$ in the region $\rho<0$ and may be transformed to $g_h$ in the region $\rho>0$.

Horský[5] has given a physical interpretation of $g_s$ as the field of a plane shell and has solved the dynamical problem of collapsing plane shells of dust. Horský and Novotný[6] matched $g_s$ to the interior of a homogeneous thick plane disk and later, Horský and Horská[7] matched $g_s$ to the interior of an inhomogeneous thick plate. Davis and Ray[8] showed that $g_s$ could be interpreted physically as a field of ghost neutrinos and later that $g_h$ could also be interpreted as a field of ghost neutrinos.[2] Davis and Ray[9] found the general form of the metric for plane symmetric neutrino fields when $T_{\mu\nu}\neq 0$. Plane symmetric self-gravitating fluids with pressure equal to the energy density were studied by Tabensky and Taub.[10] Static plane symmetric zero-rest mass scalar field were analyzed by Singh,[11] and Sistero[12] analyzed some non static plane symmetric zero-rest mass scalar fields. A large class of solutions of the exterior Einstein–Maxwell equations with plane symmetry was found by Letelier and Tabensky.[13] Banerjee and Chakrabarty[14] corrected an error of sign in the Letelier and Tabensky article and studied some plane symmetric charged dust distributions. Humi and LeBritton[15] found some interior solutions to the plane symmetric Einstein–Maxwell equations. Tiwari and Nayak[16] found plane symmetric vacuum solutions of the Brans–Dicke field equations and later,[17] plane symmetric interior solutions of the Brans–Dicke field equations. Pandy and Sharma[18] have studied PSSTS from the point of view of imbedding class.

No claim of completeness is intended for the above summary of the literature on PSSTS, but clearly, quite extensive work has been done on plane symmetry. However, we feel that this symmetry is still not understood as well as the more popular spherical and cylindrical symmetries. The following work deals with some aspects of PSSTS which are independent of the field equations and hence is a study of the kinematics of PSSTS.

Taub's[1] definition of PSSTS is given in Sec. 1. It is pointed out that the plane symmetry of a space–time may not obtain globally.

Section 2 is a discussion of admissible transformations. Four different coordinate systems are analyzed. The idea is to start with a general PSM and look for admissible transformations which simplify the form of the metric.

In Sec. 3 we introduce an intrinsic local type classification for PSSTS. Under this classification a plane symmetric space–time (PSST) is locally one of three possible types. It is shown that under certain conditions there is a one-to-one correlation between an extra Killing vector and the local type.

Section 4 concludes this discussion with an analysis of the problem of the local equivalence of two given PSSTS. It is shown that the transformation equations can be reduced to algebraic equations. If the algebraic equations are not satisfied identically, they may yield the desired transformation, or imply that the transformation does not exist.

## 1. PLANE SYMMETRIC SPACE-TIMES

Taub[1] defined PSSTS as space–times which admit the three-parameter group of transformations

$$\bar{x}^1=\cos(\theta)x^1+\sin(\theta)x^2+a, \tag{1.1a}$$

$$\bar{x}^2=-\sin(\theta)x^1+\cos(\theta)x^2+b, \tag{1.1b}$$

$$\bar{x}^3=x^3,\quad \bar{x}^4=x^4, \tag{1.1c}$$

[a)]An abstract of this paper (called Review of Plane Symmetric Space–Times) was contributed (by first named author) to the Eighth International Conference on General Relativity and Gravitation, University of Waterloo, 1977.

 0022-2488/78/1907-1617$1.00  

as a three-parameter group of isometries. The infinitesimal generators of this group of transformations are

$$k^{\mu}_{(a)} = \delta^{\mu}_1, \ \ k^{\mu}_{(b)} = \delta^{\mu}_2, \tag{1.2a}$$

$$k^{\mu}_{(\theta)} = x^2\delta^{\mu}_1 - x^1\delta^{\mu}_2. \tag{1.2b}$$

Killing's equations imply that with respect to the above coordinate system the metric must have the form

$$ds^2 = A(dx^{1^2} + dx^{2^2}) + Bdx^{3^2} + 2Cdx^3dx^4 + Ddx^{4^2}, \tag{1.3}$$

where $A$, $B$, $C$, and $D$ are suitably smooth [say $c^k$ $(k>0)$] functions of $x^3$ and $x^4$ only. In order that (1.3) be a space–time metric, $A$, $B$, $C$, and $D$ must satisfy the Lorentz signature requirements

$$A > 0, \tag{1.4a}$$

$$B + D + [(B-D)^2 + 4C^2]^{1/2} > 0, \tag{1.4b}$$

$$B + D - [(B-D)^2 + 4C^2]^{1/2} < 0. \tag{1.4c}$$

Note that the inequalities (1.4) imply that $\det(g_{\mu\nu}) = A^2(BD - C^2) < 0$.

Equivalently, a PSST is a space–time which admits maximally symmetric two-dimensional subspaces whose metric has positive eigenvalues and zero curvature.[19]

Unless one specifies that the transformations given by Eqs. (1.1) are isometries for all $x^3$ and $x^4$, the plane symmetry of a space–time may not obtain globally. For example, consider the metric

$$ds^2 = A(dx^{1^2} + dx^{2^2}) + Bdx^{3^2} + 2Cdx^3dx^4 + Ddx^{4^2} + h(x^3)h_{\mu\nu}(x)dx^\mu dx^\nu, \tag{1.5}$$

where $A$, $B$, $C$, and $D$ are as before, $h(x^3)$ is the "test function" defined by

$$h(x^3) = \exp[-1/(1-x^{3^2})], \ \ |x^3| \le 1; \ h(x^3) = 0, \ \ |x^3| \ge 1, \tag{1.6}$$

and where the $h_{\mu\nu}(x)$ are suitably smooth functions of the $x^\mu$ chosen so as to give (1.5) the correct signature in the region $|x^3| < 1$. This metric is plane symmetric for $|x^3| \ge 1$ but not necessarily so for $|x^3| < 1$. We shall consider space–times which are globally plane symmetric. In this case the transformations given by Eqs. (1.1) are isometries for all $x^3$ and $x^4$.

## 2. ADMISSIBLE TRANSFORMATIONS

A PSM $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$ can by means of a coordinate transformation be reduced to the form given by Eq. (1.3). A coordinate system such as this is said to be natural for plane symmetry. An admissible transformation[20] of a PSST is a transformation which leaves the form of the generators given by Eqs. (1.2) invariant. In order to deduce the general form for the admissible transformations of a PSST, consider the general transformation

$$x'^\mu = F^\mu(x^1, x^2, x^3, x^4). \tag{2.1}$$

The transformation given by Eqs. (2.1) is an admissible transformation for PSSTS if

$$\delta^\nu_1 = \delta^\mu_1 \frac{\partial F^\nu}{\partial x^\mu}, \ \ \delta^\nu_2 = \delta^\mu_2 \frac{\partial F^\nu}{\partial x^\mu}, \tag{2.2a}$$

$$F^2\delta^\nu_1 - F^1\delta^\nu_2 = (x^2\delta^\mu_1 - x^1\delta^\mu_2)\frac{\partial F^\nu}{\partial x^\nu}. \tag{2.2b}$$

From these equations it follows that

$$x'^1 = x^1, \ \ x'^2 = x^2, \tag{2.3a}$$

$$x'^3 = F(x^3, x^4), \ \ x'^4 = G(x^3, x^4), \tag{2.3b}$$

is the general form for the admissible transformations of a PSST.

Suppose that the functions $F$ and $G$ are of class $C^{k+1}$. In general, Eqs. (2.3) need not have an inverse which is of class $C^{k+1}$ The inverse function theorem tells us that in regions where the Jacobian

$$J = \begin{vmatrix} \dfrac{\partial F}{\partial x^3} & \dfrac{\partial F}{\partial x^4} \\ \dfrac{\partial G}{\partial x^3} & \dfrac{\partial G}{\partial x^4} \end{vmatrix} = \frac{\partial F}{\partial x^3}\frac{\partial G}{\partial x^4} - \frac{\partial F}{\partial x^4}\frac{\partial G}{\partial x^3} \ne 0, \tag{2.4}$$

Eqs. (2.3) are locally invertible and the inverse functions

$$x^3 = f(x'^3, x'^4), \ \ x^4 = g(x'^3, x'^4) \tag{2.5}$$

will also be of class $C^{k+1}$. When this is the case, Eqs. (2.3) define a local admissible coordinate transformation for PSSTS. Admissible coordinate transformations transform natural coordinates into natural coordinates.

One can use invertible admissible transformations to express the metric in many different forms. This be becomes a useful exercise when a simplification for the form given by Eq. (1.3) is achieved. We shall consider the following four forms of a PSM: (A) harmonic, (B) orthogonal, (C) the Taub[1] form, and (D) Petrov's[21] form for conformal reducible metrics of type II.

### A. Plane symmetric space-times in harmonic coordinates

We shall say that a PSST is naturally harmonic if a natural coordinate system exists such that

$$ds^2 = A'(dx'^{1^2} + dx'^{2^2}) + B'dx'^{3^2} + 2C'dx'^3dx'^4 + D'dx'^{4^2}, \tag{2.6}$$

where $A'$, $B'$, $C'$, and $D'$ are $C^k$ $(k>0)$ functions of $x'^3$ and $x'^4$ only, satisfy the Lorentz signature requirements and the harmonic coordinate conditions,[22] $0 = (\sqrt{-g'}\,g'^{\mu\nu})_{,\nu}$. These conditions may be written out explicitly as follows:

$$(A'D'(C'^2 - B'D')^{-1/2})_{,3} = (A'C'(C'^2 - B'D')^{-1/2})_{,4}, \tag{2.7a}$$

$$(A'C'(C'^2 - B'D')^{-1/2})_{,3} = (A'B'(C'^2 - B'D')^{-1/2})_{,4}, \tag{2.7b}$$

where the subscripts 3 and 4 refer to $x'^3$ and $x'^4$, respectively.

Now we proceed to show that every $C^3$ PSST is locally equivalent to a $C^1$ naturally harmonic PSST.

Given a PSST with metric given by Eq. (1.3) subject to the inequalities (1.4), the four independent solutions (when they exist) of the equation

$$g^{\mu\nu}\psi_{,\mu\nu} = \psi_{,\nu}\Gamma^\nu, \tag{2.8}$$

where

$$\Gamma^{\nu}=\tfrac{1}{2}g^{\mu\rho}g^{\nu k}\{g_{k\mu,\rho}+g_{k\rho,\mu}-g_{\mu\rho,k}\}. \tag{2.9}$$

and

$$(g^{\mu\nu})=\begin{bmatrix}1/A & 0 & 0 & 0\\ 0 & 1/A & 0 & 0\\ 0 & 0 & D/(BD-C^2) & -C/(BD-C^2)\\ 0 & 0 & -C/(BD-C^2) & B/(BD-C^2)\end{bmatrix} \tag{2.10}$$

define a local system of harmonic coordinates.[22] We want to look for solutions of Eq. (2.8) in the form of Eqs. (2.3). Note first of all that $\psi=x^1$ (or $x^2$) is a solution of Eq. (2.8) if $\Gamma^1$ (or $\Gamma^2$) vanishes identically. Inspection of Eqs. (2.9) and (2.10) show that $\Gamma^1$ and $\Gamma^2$ vanish identically, and therefore $x'^1=x^1$ and $x'^2=x^2$ are two independent solutions of Eq. (2.8).

We now look for solutions of the form $\psi(x^3, x^4)$. Equation (2.8) becomes

$$D\psi_{,33}-2C\psi_{,34}+B\psi_{,44}=(BD-C^2)[\psi_{,3}\Gamma^3+\psi_{,4}\Gamma^4]. \tag{2.11}$$

If $B$ (or $D$) is not equal to zero, Eq. (2.11) may be divided by $B$ (or $D$) yielding a form of Eq. (2.11) suitable for the Cauchy—Kowalewski theorem. However, the conditions of this theorem (analyticity) are too restrictive to be generally applicable to general relativity. Hawking and Ellis[23] point out that a $C^{2-}$ metric guarantees the existence of unique geodesics. They give examples where a $C^{1-}$ metric is used to describe gravitational shock waves (Choquet-Bruhat and Penrose), thin mass shells (Israel), and solutions containing pressure free matter where the geodesic flow lines have two- or three-dimensional caustics (Papepetrou and Hamoui). Synge[24] claims that the metric should be $C^1$ across 3-surfaces of discontinuity and $C^3$ everywhere else. If the $g_{\mu\nu}$ are $C^k$, $\det(g_{\mu\nu})$ is $C^k$. Since $\det(g_{\mu\nu})\neq 0$ the $g^{\mu\nu}$ will also be $C^k$.

Now rewrite Eq. (2.11) as follows:

$$(D\psi_{,3}-C\psi_{,4})_{,3}+(-C\psi_{,3}+B\psi_{,4})_{,4}+P\psi_{,3}+Q\psi_{,4}=0, \tag{2.12a}$$

$$P=C_{,4}-D_{,3}-(BD-C^2)\Gamma^3, Q=C_{,3}-B_{,4}-(BD-C^2)\Gamma^4. \tag{2.12b}$$

If $A$, $B$, $C$, and $D$ are $C^2$, $P$ and $Q$ will be $C^1$ and we are guaranteed the existence of a unique (corresponding to suitable initial data) local weak ($C^1$) solution of the hyperbolic (i.e., $BD-C^2<0$) second order linear partial differential equation (2.12a) for the function $\psi$ of the two independent variables $x^3$ and $x^4$. If $A$, $B$, $C$, and $D$ are $C^3$, $P$ and $Q$ will be $C^2$ and this guarantees the existence of a unique (corresponding to suitable initial data) local $C^2$ solution in the usual sense.[25]

Suppose $A$, $B$, $C$ and $D$ are $C^3$. It is possible to find two solutions, say $F(x^3, x^4)$ and $G(x^3, x^4)$ of Eq. (2.12a) in a neighborhood of an initial noncharacteristic curve $a(s)$, $s\in[0,s_0]$, if the initial data are chosen as follows [Haak and Wendland (25)]:

$$F/_{a(s)}=\tilde{F}(s),\ \ [(DF_{,3}-CF_{,4})\dot{X}^4+(CF_{,3}-BF_{,4})\dot{X}^3]/_{a(s)}=V(s), \tag{2.13a}$$

$$G/_{a(s)}=\tilde{G}_{(s)},\ \ [(DG_{,3}-CG_{,4})\dot{X}^4+(CG_{,3}-BG_{,4})\dot{X}^3]/_{a(s)}=W(s), \tag{2.13b}$$

$$a(s),\ \tilde{F}(s),\ \tilde{G}(s)\in C^2([0,s_0]),\ V(s),\ W(s)\in C^1([0,s_0]), \tag{2.13c}$$

where a "dot" means differentiation with respect to the parameter $s$. From Eqs. (2.13) it follows that

$$J/_{a(s)}=(F_{,3}G_{,4}-F_{,4}G_{,3})/_{a(s)}\equiv\tilde{J}(s)=(V\tilde{G}-W\tilde{F})/(B\dot{X}^{3^2}+2C\dot{X}^3\dot{X}^4+D\dot{X}^{4^2}). \tag{2.14}$$

The quantity $B\dot{X}^{3^2}+2C\dot{X}^3\dot{X}^4+D\dot{X}^{4^2}\neq 0$ for a noncharacteristic curve and the initial data may be chosen so that $V\tilde{G}-W\tilde{F}\neq 0$. This shows that the initial data may be chosen so that $\tilde{J}(s)\neq 0$. It then follows [in the same way as the solution to Eq. (2.12a) was constructed] that $J\neq 0$ in a neighborhood of $a(s)$.

### B. Plane symmetric space-times in orthogonal cordinates

We shall say that a PSST is naturally orthogonal if a natural coordinate system exists such that

$$ds^2=A(dx^{1^2}+dx^{2^2})+Bdx^{3^2}+Ddx^{4^2}, \tag{2.15}$$

where $A$, $B$, and $D$ are $C^k$ ($k>0$) functions of $x^3$ and $x^4$ only and satisfy the signature requirements (1.4) with $C=0$.

Now we proceed to show that a $C^k$ naturally harmonic PSST is locally equivalent to a $C^k$ naturally orthogonal PSST.

Consider the naturally harmonic PSST defined by Eqs. (2.6) and (2.7). Consider also the change of basis

$$w^1=dx'^1,\ \ w^2=dx'^2,\ \ w^3=dx'^3, \tag{2.16a}$$

$$w^4=\eta(C'dx'^3+D'dx'^4), \tag{2.16b}$$

where $\eta$ is a function of $x'^3$ and $x'^4$. With respect to this new basis, the metric takes the form

$$g=A'(w^{1^2}+w^{2^2})+(B'-C'^2/D')w^{3^2}+(1/\eta^2D')w^{4^2}, \tag{2.17}$$

in the regions where $D'\neq 0$. (We use the letter $g$ to denote the metric with respect to a noncoordinate basis.) Now we require that the new basis be a coordinate basis, i.e., that

$$x^1=x'^1,\ \ x^2=x'^2,\ \ x^3=x'^3, \tag{2.18a}$$

$$x^4=G(x'^3, x'^4),\ \ w^4=dx^4. \tag{2.18b}$$

Equations (2.16b) and (2.18b) lead to

$$\eta C'=\frac{\partial G}{\partial x'^3},\ \ \eta D'=\frac{\partial G}{\partial x'^4}, \tag{2.19a}$$

$$\frac{\partial}{\partial x'^4}(\eta C')=\frac{\partial}{\partial x'^3}(\eta D'). \tag{2.19b}$$

Equation (2.19b) is the integrability condition for Eqs. (2.19a) and ensures that $w^4=dx^4$ will be an exact differential. The $C^k$ solution of Eq. (2.19b) [via Eq. (2.7a)] is $\eta=A'/(C'^2-B'D')^{1/2}$. $G$ is obtainable from Eqs. (2.19a) by integration and therefore $G$ is also $C^k$. The Jacobian of the transformation given by Eqs. (2.18) is $J=\eta D'\neq 0$ in the regions where $D'\neq 0$. Apart from the possibility $D'=0$, the new basis is a coordinate basis and the metric is given by

$$g=ds^2_{R_1}=A(dx^{1^2}+dx^{2^2})+\left(B-\frac{C^2}{D}\right)\left[dx^{3^2}-\left(\frac{1}{A}\right)^2 dx^{4^2}\right]. \tag{2.20}$$

The primes have been removed, indicating that $A'$, $B'$, $C'$, and $D'$ have been evaluated in terms of $x^3$ and $x^4$. The subscript $R_1$ indicates that the metric is defined in the regions $R_1$ for which $D'\neq 0$.

We may also transform (locally) the naturally harmonic PSM to a naturally orthogonal PSM via the change of basis

$$w^1=dx'^1,\ \ w^2=dx'^2,\ \ w^4=dx'^4, \tag{2.21a}$$

$$w^3=\eta(B'dx'^3+C'dx'^4). \tag{2.21b}$$

With respect to the new basis the metric takes the form

$$g=A'(w^{1^2}+w^{2^2})+\left(\frac{1}{\eta^2B'}\right)w^{3^2}+\left(D'-\frac{C'^2}{B'}\right)w^{4^2}, \tag{2.22}$$

in the regions where $B'\neq 0$. The requirement that the new basis be a coordinate basis is satisfied with the choice $\eta=A'/(C'^2-B'D')^{1/2}$.

The Jocobian of the transformation given by Eqs. (2.21) is $J=\eta B'\neq 0$ in the regions where $B'\neq 0$. Apart from the possibility $B'=0$, the new basis is a coordinate basis and the metric is given by

$$g=ds^2_{R_2}=\bar{A}(d\bar{x}^{1^2}+d\bar{x}^{2^2})+(\bar{C}^2\bar{B}^{-1}-\bar{D})[\bar{A}^{-2}d\bar{x}^{3^2}-d\bar{x}^{4^2}]. \tag{2.23}$$

The "bars" indicate that $A'$, $B'$, $C'$, and $D'$ have been evaluated in terms of $\bar{x}^3$ and $\bar{x}^4$ and the subscript $R_2$ indicates that the metric is defined in the regions $R_2$ for which $B'\neq 0$.

Since we have an explicit form for the local transformation of a naturally harmonic PSST to a naturally orthogonal PSST we may discuss some of the global properties of this transformation. If $D'$ (or $B'$) is definite (strictly greater than or less than zero), the transformation, Eqs. (2.18) [or Eqs. (2.21)] from a naturally harmonic PSST to a naturally orthogonal PSST is global. On the other hand, we have to admit the possibility that the hypersurfaces $D'=0$ and $B'=0$ exist for some naturally harmonic PSSTS. If this is the case, then either $D'=0$ and $B'=0$ have an empty intersection or they do not.

Let us first of all consider the case where $(D'=0)\cap(B'=0)=\emptyset$. In this case consider a covering of the naturally harmonic PSST with four coordinate patches: $\bar{R}_1$ for which $D'>0$, $\underline{R}_1$ for which $D'<0$, $\bar{R}_2$ for which $B'>0$, and $\underline{R}_2$ for which $B'<0$. In order to visualize this situation, lets consider the $x'^3$, $x'^4$ plane in which $D'=0$ and $B'=0$ appear as curves. For simplicity, we shall represent these curves by parallel straight lines. [See Fig. 1.]

The region $\underline{R}_1$ contains $B'=0$, and the region $\bar{R}_2$ contains $D'=0$. The naturally harmonic PSST may be mapped to the naturally orthogonal PSST which is covered by the four coordinate patches $\bar{R}_1$, $\underline{R}_1$, $\bar{R}_2$, and $\underline{R}_2$ and for which the metric is given by

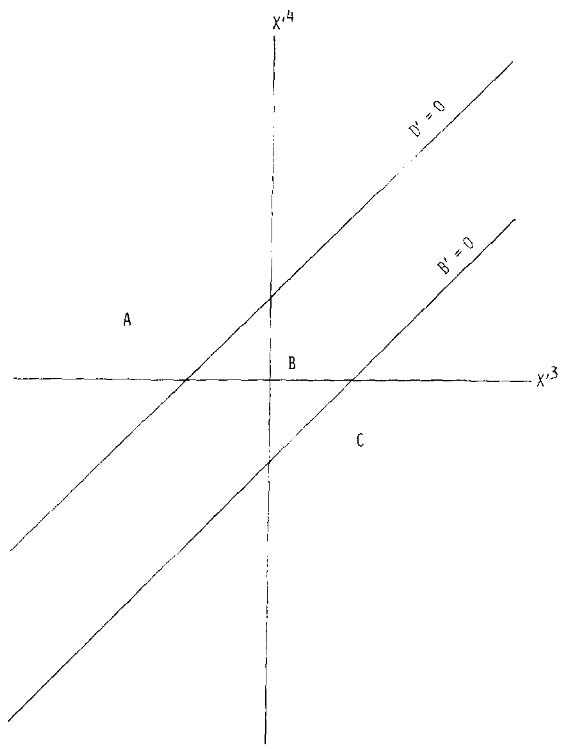



FIG. 1. The region $B$ is $\bar{R}_2\cap\underline{R}_1$, $A$ is the region $\bar{R}_1\cap\bar{R}_2$ and the region C is $\underline{R}_2\cap\underline{R}_1$. By their definition, the regions $A$, $B$, $C$ do not contain the curves $D'=0$ and $B'=0$. These are the overlap regions for the covering $\bar{R}_1\cup\underline{R}_1\cup\bar{R}_2\cup\underline{R}_2$.

$$ds^2_{\bar{R}_1}=A(dx^{1^2}+dx^{2^2})+(C^2D^{-1}-B)\,[A^{-2}dx^{4^2}-dx^{3^2}], \tag{2.24a}$$

$$ds^2_{\underline{R}_1}=A(dx^{1^2}+dx^{2^2})+(B-C^2D^{-1})[dx^{3^2}-A^{-2}dx^{4^2}], \tag{2.24b}$$

$$ds^2_{\bar{R}_2}=\bar{A}(d\bar{x}^{1^2}+d\bar{x}^{2^2})+(\bar{C}^2\bar{B}^{-1}-\bar{D})[\bar{A}^{-2}d\bar{x}^{3^2}-d\bar{x}^{4^2}], \tag{2.24c}$$

$$ds^2_{\underline{R}_2}=\bar{A}(d\bar{x}^{3}+d\bar{x}^{2^2})+(\bar{D}-\bar{C}^2\bar{B}^{-1})[d\bar{x}^{4^2}-\bar{A}^{-2}d\bar{x}^{3^2}] \tag{2.24d}$$

in the respective patches. In the overlap region $A$, $ds^2_{\bar{R}}$ is related to $ds^2_{\bar{R}_2}$ via the relations $x^1=\bar{x}^1$, $x^2=\bar{x}^2$, $x^3=x'^3$, $x^4=G(x'^3,x'^4)$, $\bar{x}^3=F(x'^3,x'^4)$, $\bar{x}^4=x'^4$. One can express $x'^3$ and $x'^4$ in terms of $x^3$ and $x^4$ and then $\bar{x}^3$ and $\bar{x}^4$ may be expressed in terms of $x^3$ and $x^4$. The metrics $ds^2_{\bar{R}_2}$ and $ds^2_{\underline{R}_1}$ are similarly related in the overlap regions $B$ as are $ds^2_{\underline{R}_1}$ and $ds^2_{\underline{R}}$ in the overlap region $C$. Therefore if $(D'=0)\cap(B'=0)=\emptyset$, the naturally harmonic PSST is globally equivalent to the rather complex naturally orthogonal PSST (2.24).

In the case where $(D'=0)\cap(B'=0)\neq\emptyset$, the situation is more complicated. Consider, once again, the $x'^3$, $x'^4$ plane and a neighborhood $U$ of one of the intersection points. For simplicity, we represent the curves $D'=0$ and $B'=0$ by intersecting line segments in $U$ as shown in Fig. 2. Note that the point $P$ where $D'=0$ and $B'=0$ intersect is not in the orthogonal PSST constructed above. However, the point $P$ is well defined in the naturally harmonic PSST.

We end this section with the following remark. A

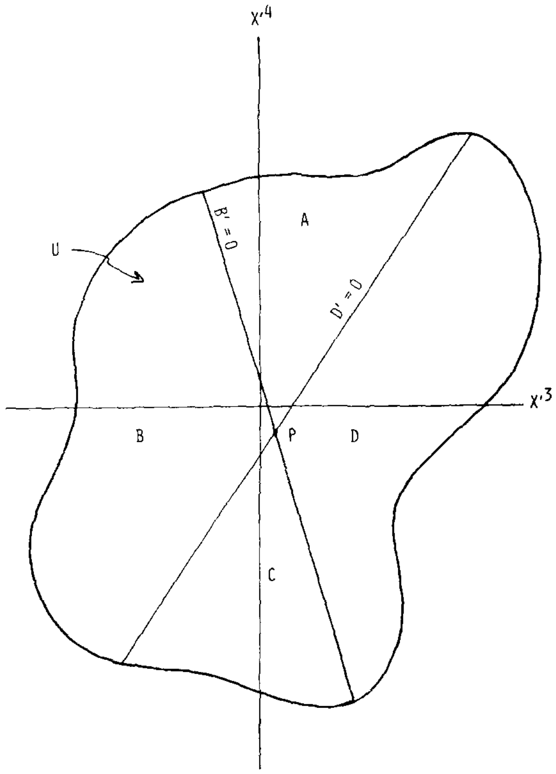



FIG. 2. The overlap regions for the covering $\bar{R}_1 \cup \underline{R}_1 \cup \bar{R}_2 \cup \underline{R}^2 \cup P$ of the naturally harmonic plane symmetric space—time are $A = \bar{R}_2 \cap \bar{R}_1$, $B = \underline{R}_2 \cap \bar{R}_1$, $C = \underline{R}_2 \cap \underline{R}_1$, and $D = \bar{R}_2 \cap \underline{R}_1$.

general $C^3$ PSST is locally equivalent to a $C^1$ naturally orthogonal PSST via its local equivalence to a $C^1$ naturally harmonic PSST.

### C. The Taub form of plane symmetric space-times

The Taub[1] form for a PSM is

$$ds^2 = A(dx^{1^2} + dx^{2^2}) + B(dx^{3^2} - dx^{4^2}), \tag{2.25}$$

where $A$ and $B$ are $C^k$ (say $k \geq 0$) functions of $x^3$ and $x^4$ only.

A $C^1$ naturally orthogonal PSST is locally equivalent to a $C^0$ Taub PSST. To see this consider the $C^1$ naturally orthogonal PSST given by Eq. (2.18) and make the change of basis

$$w^1 = dx^1, \quad w^2 = dx^2, \tag{2.26a}$$

$$w^3 = a dx^3 + \left(\frac{-D}{B}\right)^{1/2} b\, dx^4, \tag{2.26b}$$

$$w^4 = b dx^3 + \left(\frac{-D}{B}\right)^{1/2} a\, dx^4. \tag{2.26c}$$

With respect to this new basis, the metric takes the form

$$g = A(w^{1^2} + w^{2^2}) + [B/(a^2 - b^2)](w^{3^2} - w^{4^2}). \tag{2.27}$$

The new basis will be a coordinate basis provided

$$\frac{\partial a}{\partial x^4} = \frac{\partial}{\partial x^3}\left[\left(\frac{-D}{B}\right)^{1/2} b\right], \tag{2.28a}$$

$$\frac{\partial b}{\partial x^4} = \frac{\partial}{\partial x^3}\left[\left(\frac{-D}{B}\right)^{1/2} a\right]. \tag{2.28b}$$

Equations (2.28) are in a form suitable for the Cauchy–Kowalewsky theorem, however we shall manage with less stringent conditions than the conditions of this theorem.

Write Eqs. (2.28) as follows:

$$a_{,4} - \left(\frac{-D}{B}\right)^{1/2} b_{,3} - \left[\left(\frac{-D}{B}\right)^{1/2}\right]_{,3} b = 0, \tag{2.29a}$$

$$b_{,4} - \left(\frac{-D}{B}\right)^{1/2} a_{,3} - \left[\left(\frac{-D}{B}\right)^{1/2}\right]_{,3} a = 0, \tag{2.29b}$$

Since $B$ and $D$ are definite $C^1$ functions of $x^3$ and $x^4$, $(-D/B)^{1/2}$ is $C^1$ and therefore $[(-D/B)^{1/2}]_{,3}$ is $C^0$. Hence, Eqs. (2.29) are a special case of the linear system of two first order partial differential equations discussed in Chap. 7 of Ref. 25. It turns out that the system (2.29) is hyperbolic. Furthermore, the conditions of the theorem which guarantees the existence of a local, weak, unique (corresponding to suitable initial data) solution are satisfied [see Chap. 8 of Ref. (25)]. Therefore, we are guaranteed the existence of local $C^0$ solutions $a(x^3, x^4)$ and $b(x^3, x^4)$ of Eqs. (2.29). Since $F$ and $G$ are obtained by integration of $a$ and $b$, $F$ and $G$ will also be $C^0$ functions of $x^3$ and $x^4$. The Jacobian of the transformation given by Eqs. (2.26) is $= (-D/B)^{1/2}(a^2 - b^2)$. The initial data may be chosen so that $J$ does not equal zero. The new basis is a coordinate basis and the metric takes the form

$$g = ds^2 = A'(dx'^{1^2} + dx'^{2^2}) + (B'/(a'^2 - b'^2)) \times [dx'^{3^2} - dx'^{4^2}], \tag{2.30}$$

where the primes indicate that $A$, $B$, $a$, and $b$ have been evaluated in terms of $x'^3$ and $x'^4$.

### D. Plane symmetric space-times in Petrov's form for conformal reducible metrics of type II.

The metric of a PSST is conformal reducible type II. Petrov[21] has found all the solutions to $R_{\mu\nu} = kg_{\mu\nu}$ for metrics of this type. The coordinate system used there is

$$ds^2 = H^2\{e_1\beta^2 dx^{1^2} + e_2 dx^{2^2} + e_3\gamma^2 dx^{3^2} + e_4 dx^{4^2}\}, \tag{2.31}$$

where $H = H(x^1, x^2, x^3, x^4)$, $\beta = \beta(x^1, x^2)$, and $\gamma = \gamma(x^3, x^4)$. The $e_\mu$ are chosen to give (2.37) the correct signature. The Petrov form for conformal reducible metrics of type II with plane symmetry is

$$ds^2 = A\{dx^{1^2} + dx^{2^2} + B dx^{3^2} - dx^{4^2}\}, \tag{2.32}$$

where $A$ and $B$ are positive definite $C^k$ (say $k \geq 0$) functions of $x^3$ and $x^4$ only. Now we set up the equations which must be solvable if a naturally orthogonal $C^1$ PSST is locally equivalent to a PSST with metric given by (2.32).

Write the naturally orthogonal $C^1$ PSM as follows,

$$ds^2 = A\{dx^{1^2} + dx^{2^2} + B dx^{3^2} + D dx^{4^2}\}. \tag{2.33}$$

The change of basis

$$w^1 = dx^1, \quad w^2 = dx^2,$$

$$w^3 = a dx^3 + \left(\frac{-D}{B}\right)^{1/2} (ab/\sqrt{B + b^2})\, dx^4, \tag{2.34}$$

$$w^4 = b dx^3 + \left(\frac{-D}{B}\right)^{1/2} \sqrt{B + b^2}\, dx^4,$$

brings the metric to the desired form

$$g = A\left[w^{1^2} + w^{2^2} + \left(\frac{(B+b^2)}{a^2}\right) w^{3^2} - w^{4^2}\right]. \tag{2.35}$$

The new basis is a coordinate basis, provided

$$\frac{\partial a}{\partial x^4} = \frac{\partial}{\partial x^3}\left\{\left(\frac{-D}{B}\right)^{1/2}(ab/\sqrt{B+b^2}\right\}, \tag{2.36a}$$

$$\frac{\partial b}{\partial x^4} = \frac{\partial}{\partial x^3}\left\{\left(\frac{-D}{B}\right)^{1/2}\sqrt{B+b^2}\right\}, \tag{2.36b}$$

Although Eqs. (2.36) are in a form suitable for the Cauchy–Kowalewski theorem, the quantities involved are not analytic. We instead put Eqs. (2.36) into a form used by Courant,[26]

$$\frac{\partial U}{\partial x^4} + M\frac{\partial U}{\partial x^3} + N = 0, \tag{2.37}$$

where

$$U = \begin{pmatrix} a \\ b \end{pmatrix}, \quad -N = \begin{pmatrix} (ab/\sqrt{B+b^2})\left[\left(\frac{-D}{B}\right)^{1/2}\right]_{,3} \\ \sqrt{B+b^2}\left[\left(\frac{-D}{B}\right)\right]^{1/2}_{,3} \end{pmatrix} \tag{2.38}$$

$$-M = \begin{pmatrix} \left(\frac{-D}{B}\right)^{1/2} b/\sqrt{B+b^2} & \left(\frac{-D}{B}\right)^{1/2} aB/\sqrt{B+b^2} \\ 0 & \left(\frac{-D}{B}\right)^{1/2} b\left(\sqrt{B+b^2}\right)^{-1/2} \end{pmatrix}. \tag{2.39}$$

The matrix $M$ has a double real eigenvalue $\lambda = -(-D/B)^{1/2}b/\sqrt{B+b^2}$ for real $b$. $M$ does not have two linearly independent left eigenvectors. Courant[26] has shown that $C^1$ solutions of Eq. (2.44) exist when $M$ and $N$ are $C^1$ functions of $U, x^3$, and $x^4$ and when the matrix $M$ has two real eigenvalues and two linearly independent left eigenvectors. However, the problem we are considering violates these conditions.

The problem of determining the conditions under which a naturally orthogonal PSST is locally equivalent to a PSST with metric given by (2.38) is essentially the same as determining the conditions under which there exist a solution of Eq. (2.37) where $U$, $M$, and $N$ are given by Eqs. (2.38) and (2.39) such that $J = a(-D/B)^{1/2}B/\sqrt{B+b^2} \neq 0$.

## 3. LOCAL TYPE CLASSIFICATION FOR PLANE SYMMETRIC SPACE-TIMES

In Sec. 4 we will consider the problem of the local equivalence of two given PSSTS. The following local type classification allows one to distinguish three local types of PSSTS.

Consider the PSM given by Eq. (1.3) along with the inequalities (1.4). For admissible transformations [Eqs. (2.3)] we have

$$A'(x'^3, x'^4) = A(x^3, x^4), \tag{3.1a}$$

$$g'_{ij} = \frac{\partial x^l}{\partial x'^j}\frac{\partial x^m}{\partial x'^j} g_{lm}, \quad i, j, l, m = 3, 4. \tag{3.1b}$$

Note that $A$ is a scalar function with respect to admissible transformations. Restricting ourselves to admissible transformations, the quantities

$$V_\mu = \frac{\partial A}{\partial x^\mu} \tag{3.2}$$

are the components of a vector and

$$V^\mu V_\mu = \left\{D\left(\frac{\partial A}{\partial x^3}\right)^2 - 2C\left(\frac{\partial A}{\partial x^3}\right)\left(\frac{\partial A}{\partial x^4}\right) + B\left(\frac{\partial A}{\partial x^4}\right)^2\right\}\Delta^{-1}, \tag{3.3a}$$

$$\Delta \equiv BD - C^2, \tag{3.3b}$$

is a scalar, which depending on $A$, $B$, $C$, and $D$, may be greater than, less than, or equal to zero. We shall say that a PSST is respectively type I, II, III if $V^\mu V_\mu$ is greater than, less than, or equal to zero. This is a local classification as a PSST may be of mixed type over a large domain.

Most of the PSSTS discussed in the references given in the Introduction admit an extra Killing vector $\xi^\mu$ for which $\xi^1 = 0 = \xi^2$. When this is the case, Killing's equations imply that $\xi^3$ and $\xi^4$ are independent of $x^1$ and $x^2$ that

$$A_{,3}\xi^3 + A_{,4}\xi^4 = 0. \tag{3.4}$$

Suppose that $A_{,4} \neq 0$. In this case

$$\xi^\mu \xi_\mu = (\xi^3)^2\{DA_{,3}^2 - 2CA_{,3}A_{,4} + BA_{,4}^2\}/(A_{,4}^2). \tag{3.5}$$

Comparing Eqs. (3.3) and (3.5) we see that

$$V^\mu V_\mu = [(A_{,4})^2/(\Delta(\xi^3)^2]\xi^\mu\xi_\mu. \tag{3.6}$$

From Eq. (3.6) if a PSST admits an extra Killing vector $\xi^\mu$ for which $\xi^1 = 0 = \xi^2$, the type is I, II, III if $\xi^\mu$ is respectively timelike, spacelike, null.

## 4. LOCAL EQUIVALENCE OF TWO PLANE SYMMETRIC SPACE-TIMES

Suppose that one has the general solution to Einstein's field equations with plane symmetry. This solution generally consists of a class of solutions depending on some arbitrary functions and constants of integration. At this point one is faced with the problem of determining the subclass of solutions which cannot be obtained from one another by a coordinate transformation. The general problem of the equivalence of two given symmetric quadratic forms is discussed in Eisenhart.[27] We shall follow the outline of Eisenhart's discussion.

Let $S_T(g)$ be the class of all solutions to Einstein's field equations for a PSM $g$ and energy–momentum tensor $T$. Consider $S_{1T}(g_1)$ and $S_{2T}(g_2) \in S_T(g)$. $g_1$ and $g_2$ will have the general form (1.3). We have shown that every $C^3$ PSM is locally equivalent to a $C^1$ naturally orthogonal PSM. In the following analysis we will assume that $g_1$ and $g_2$ are at least $C^3$ so that we may use their local naturally orthogonal form. This involves a slight loss of generality since the results will not apply to $C^2$ and $C^1$ PSMS.

Suppose that we have determined that $g_1$ and $g_2$ are locally of the same type. Then

$$ds_1^2 = A(dx^{1^2} + dx^{2^2}) + Bdx^{3^2} - Ddx^{4^2}, \tag{4.1}$$

$$ds_2^2 = A'(dx'^{1^2} + dx'^{2^2}) + B'dx'^{3^2} - D'dx'^{4^2}, \tag{4.2}$$

If there is an admissible transformation of the form

$$x^1 = x'^1, \ x^2 = x'^2, \tag{4.3a}$$

$$x^3 = F(x'^3, x'^4), \tag{4.3b}$$

$$x^4 = G(x'^3, x'^4), \tag{4.3c}$$

such that $ds_1^2 = ds_2^2$, then the space—times are locally equivalent. That is, if the nonlinear first order equations:

$$A(F,G) = A'(x'^3, x'^4), \tag{4.4a}$$

$$B(F,G)\left(\frac{\partial F}{\partial x'^3}\right)^2 - D(F,G)\left(\frac{\partial G}{\partial x'^3}\right)^2 = B'(x'^3, x'^4), \tag{4.4b}$$

$$B(F,G)\left(\frac{\partial F}{\partial x'^4}\right)^2 - D(F,G)\left(\frac{\partial G}{\partial x'^4}\right)^2 = -D'(x'^3, x'^4), \tag{4.4c}$$

$$B(F,G)\left(\frac{\partial F}{\partial x'^3}\right)\left(\frac{\partial F}{\partial x'^4}\right) - D(F,G)\left(\frac{\partial G}{\partial x'^3}\right)\left(\frac{\partial G}{\partial x'^4}\right) = 0 \tag{4.4d}$$

have solutions, the two space—times are locally equivalent.

If we use the notation

$$\left.\begin{aligned} a &= \frac{\partial F}{\partial x'^3} \\ b &= \frac{\partial F}{\partial x'^4} \end{aligned}\right\}, \quad \left.\begin{aligned} e &= \frac{\partial G}{\partial x'^3} \\ f &= \frac{\partial G}{\partial x'^4} \end{aligned}\right\}, \quad \left.\begin{aligned} g &= \frac{\partial A}{\partial F} \\ h &= \frac{\partial A}{\partial G} \end{aligned}\right\}, \quad \left.\begin{aligned} i &= \frac{\partial A'}{\partial x'^3} \\ j &= \frac{\partial A'}{\partial x'^4} \end{aligned}\right\}, \tag{4.5}$$

and provided $g \neq 0$, we can differentiate Eq. (4.4a) and reduce Eqs. (4.4b)—(4.4d) to

$$(Bh^2 - g^2D)e^2 - 2iBhe + (Bi^2 - g^2B') = 0, \tag{4.6a}$$

$$(Bh^2 - g^2D)f^2 - 2jBhf + (Bj^2 + g^2D') = 0, \tag{4.6b}$$

$$(Bh^2 - g^2D)ef - iBhf - jBhe + Bij = 0. \tag{4.6c}$$

Once $e$ and $f$ have been determined as solutions of Eqs. (4.6), $a$ and $b$ are given by

$$a = (i - he)/g, \tag{4.7a}$$

$$b = (j - hf)/g. \tag{4.7b}$$

Equations (4.6a,b) are quadratic equations for $e$ and $f$ respectively. Their solution may be analyzed in terms of two cases: (1) $Bh^2 - g^2D \neq 0$ ($g_1$ and $g_2$ are type I or II), (2) $Bh^2 - g^2D = 0$ ($g_1$ and $g_2$ are type III).

*Case* (1): $Bh^2 - g^2D \neq 0$. Equations (4.6a,b) yield

$$\begin{aligned} e &= [iBh \pm \{(iBh)^2 - (Bh^2 - g^2D)(Bi^2 - g^2B')\}^{1/2}]/(Bh^2 - g^2D), \\ f &= [jBh \pm \{(jBh)^2 - (Bh^2 - g^2D)(Bj^2 + g^2D')\}^{1/2}]/(Bh^2 - g^2D). \end{aligned} \tag{4.8}$$

in order that $e$ and $f$ be real, the quantities $(i, B')$ and $(j, D')$ must satisfy

$$\begin{aligned} (i,B') &\equiv (iBh)^2 - (Bh^2 - g^2D)(Bi^2 - g^2B') \geq 0, \\ (j,D') &\equiv (jBh)^2 - (Bh^2 - g^2D)(Bj^2 + g^2D') \geq 0. \end{aligned} \tag{4.9}$$

Finally Eq. (4.6c) gives us

$$-ijg^2BD + [\pm\sqrt{(i, B')}][\pm\sqrt{j, D'}] = 0. \tag{4.10}$$

Equation (4.10) was obtained by differential iteration of Eq. (4.4a) and therefore Eq. (4.10) must be satisfied simultaneously with Eq. (4.4a). If these two equations may be solved for $F$ and $G$ as real functions of $x'^3$ and $x'^4$ then the space—times are locally equivalent. On the other hand, if these two equations are not compatible the space—times are not locally equivalent. It might occur that Eq. (4.10) is satisfied identically as a result of Eq. (4.4a). If this is the case we must integrate Eqs. (4.8). These equations are of the form

$$\frac{\partial G}{\partial x'^3} = H_1(F, G, x'^3, x'^4), \quad \frac{\partial G}{\partial x'^4} = H_2(F, G, x'^3, x'^4), \tag{4.11}$$

where $H_1$ and $H_2$ stand for the right-hand sides of Eqs. (4.8). The integrability conditions are

$$\frac{\partial H_1}{\partial x'^4} = \frac{\partial H_2}{\partial x'^3}, \quad \frac{\partial H_3}{\partial x'^4} = \frac{\partial H_4}{\partial x'^3}, \tag{4.12}$$

where $H_3$ and $H_4$ are the right-hand sides of Eqs. (4.7).

Note that the $H$'s may vary implicitly with respect to $x'^3$ and $x'^4$ through $F$ and $G$. $a, b, e$ and $f$ will appear in Eqs. (4.12) and may be eliminated from these equations via Eqs. (4.7)—(4.8). Therefore, Eqs. (4.12) are of the form

$$H_5(F, G, x'^3, x'^4) = 0, \ H_6(F, G, x'^3, x'^4) = 0. \tag{4.13}$$

Equations (4.13) must be satisfied simultaneously with Eq. (4.4a). If these equations can be solved for $F$ and $G$ as functions of $x'^3$ and $x'^4$, the space—times are locally equivalent. If these equations are not compatible, the space—times are not locally equivalent. However, once again Eqs. (4.13) may be satisfied identically as a result of Eq. (4.4a). If this is the case, we are left with the problem of determining if solutions exist for the over determined system of four quasilinear first order partial differential equations (4.7) and (4.8).

*Case* (2): $Bh^2 - g^2D = 0$. Then

$$-i^2D' + j^2B' = 0$$

which is just the condition that $g_2$ be type III and therefore tells us nothing new. However, the integrability conditions will once again lead to equations of the form of Eqs. (4.13).

The above analysis is not valid if $g = 0$. If this is the case, the analysis may be repeated for $h \neq 0$. If $g = h = 0$, $A$ is a constant and none of the special results obtained above are valid. However, this does not involve too great a loss of generality since none of the PSSTS discussed in the references given in the Introduction have $A =$ constant.

## SUMMARY

The major thrust of the above analysis was to examine some of the kinematical aspects of PSSTS. The major result was the discovery that the special forms of PSMS, harmonic and orthogonal and some of the special orthogonal forms, are not really general forms for a PSM. Given a $C^k$ PSM in the form of Eq. (1.3), there is no guarantee that the special forms discussed in part (2) exist if $k \leq 1$. If $k = 2$, the naturally harmonic and naturally orthogonal forms exist only in a weak sense and

and there is no guarantee that the Taub and Petrov forms exist. If $k=3$, there is no guarantee that the Petrov form exists. Of course, if the given PSST is analytic, the Cauchy–Kowalewski theorem guarantees the existence of each of the forms discussed in Sec. 2. As an example of where this result may be significant, consider a $C^{1-}$ PSM in the form of Eq. (1.3). As pointed out by Hawking and Ellis,[23] the examples of Choquet-Bruhat and Penrose, Israel, Papepetrou, and Hamoui show that this metric may describe a physically interpretable discontinuity. There is no guarantee that this discontinuity will appear in one of the special forms for a PSM discussed in Sec. 2.

One can distinguish three local types of PSSTS via the intrinsic local type classification given in Sec. 3. Under certain conditions the existence of an extra Killing vector is correlated with the type classification. Finally, if one has a class of solutions of the field equations with plane symmetry, one must face the problem of finding the subclass of solutions which cannot be related to each other by coordinate transformations. The analysis of Sec. 4 may greatly simplify the solution of this problem.

## ACKNOWLEDGMENTS

We would like to thank J.R. Ray (Clemson University) for helpful discussions over the period of time 1976–1977. It became evident during these discussions that a more detailed treatment of the kinematical aspects of plane symmetric space–times was needed.

[1]A.H. Taub, Ann. Math. (N.Y.) **53**, 472–90 (1951).
[2]T.M. Davis and J.R. Ray, J. Math. Phys. **17**, 1049–50 (1976).
[3]J. Novotný, Scripta Fac. Sci. Nat. Ujep Brumensis, Physica **5**, 211–2 (1975).
[4]W.B. Bonnor, Phys. Lett. A **31**, 269 (1970).
[5]J. Horský, Czech. J. Phys. B **18**, 569–83 (1968).
[6]J. Horský and J. Novotný, J. Phys. A (Gen. Phys.) **2**, 251–6 (1969).
[7]J. Horský and M. Horská, Acta Phys. Polon. B **6**, 171–2 (1975).
[8]T.M. Davis and J.R. Ray, Phys. Rev. D **9**, 331–3 (1974).
[9]T.M. Davis and J.R. Ray, J. Math. Phys. **16**, 75–9 (1975).
[10]R. Tabensky and A.H. Taub, Commun. Math. Phys. **29**, 61–77 (1973).
[11]T. Singh, Gen. Rel. Grav. **5**, 657–62 (1974).
[12]R.F. Sistero, Gen. Rel. Grav. **7**, 487–9 (1976).
[13]P.S. Letelier and R.R. Tabensky, J. Math. Phys. **15**, 594–5 (1974).
[14]A. Banerjee and N. Chakrabarty, J. Math. Phys. **18**, 265–6 (1977).
[15]M. Humi and J.A. LeBritton, Phys. Rev. D **11**, 2689–91 (1975).
[16]R.N. Tiwari and B.K. Nayak, J. Phys. A (Math. Gen.) **9**, (1976).
[17]R.N. Tiwari and B.K. Nayak, Phys. Rev. D **14**, 395–8 (1976).
[18]S.N. Pandy and S.P. Sharma, Gen. Rel. Grav. **8**, 147–53 (1977).
[19]S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, New York, 1972), Chap. 13.
[20]A.Z. Petrov, *Einstein Spaces* (Pergamon, New York, 1969), Chap. 4, Sec. 25, p. 156.
[21]A.Z. Petrov, *Einstein Spaces* (Pergamon, New York, 1969), Chap. 9, Sec. 51.
[22]V. Fock, *The Theory–Time and Gravitation* (second revised edition) (Pergamon, New York, 1964), see sections on harmonic coordinates.
[23]S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space–Time* (Cambridge U.P., Cambridge, 1973), see Chap. 8, Sec. 4, p. 286.
[24]J.L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1971), 4th printing, see Chap. 1, Sec. 1.
[25]W. Haak and W. Wendland, *Lectures on Partial and Pfaffian Differential Equations* (Pergamon, New York, 1972), see Chap. 3, Sec. 2, Theorem 5.
[26]R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Vol. II (Interscience, New York, 1962), see Chap. 5, Sec. 7.
[27]L.P. Eisenhart, *Riemannian Geometry* (Princeton U.P., Princeton, New Jersey), 5th printing, p. 23.

# Characterization of certain stationary solutions of Einstein's equations[a)]

D. M. Chitre

*Department of Physics, Montana State University, Bozeman, Montana 59715*

(Received 23 January 1978)

A scheme is proposed to characterize Tomimatsu–Sato solutions of Einstein's equations. It categorizes an infinite series of solutions, the lowest (nonflat) members of which are the Schwarzschild solution for the static case and the Kerr solution for the stationary case.

Einstein's equations for stationary, axisymmetric systems give rise to an infinite sequence of potentials, which are then used to write down infinitesimal transformations to generate new solutions.[1] So far, it has not been clear how this formalism relates to the known stationary, axisymmetric solutions, like the Schwarzschild or the Kerr solution. Furthermore, Tomimatsu and Sato[2] found more such solutions, characterized by a certain distortion parameter $\delta$, $\delta$ being the parameter classifying a series of the Weyl metrics. Since, $\delta$ can take any positive real value for the Weyl metrics, it has been a puzzle as to why the TS solutions should exist only for the integral values of $\delta$.

We will show here that there is an intimate connection between the hierarchy of potentials and the TS solutions. We will characterize the solutions by algebraic relationships among the potentials, which, then necessarily give only the integral valued Weyl solutions for the static case and should give the TS solutions for the stationary case.

We start with Einstein's field equations as written down by Kinnersley.[3] We take the metric of the form

$$ds^2 = f_{AB}dx^A dx^B - \exp(2\Gamma)\,\delta_{MN}\,dx^M\,dx^N, \quad A,B=1,2, \quad M,N=3,4, \tag{1}$$

where $f_{AB}$, $\Gamma$ are functions of $x^3$, $x^4$.

We have

$$f^{AX}f_{XB} = -\rho^2\delta^A{}_B, \tag{2}$$

where indices are raised and lowered using $\epsilon_{AB}=\pm 1$.

The Einstein field equations imply the existence of potentials $\psi_{AB}$ such that

$$\nabla\psi_{AB} = -\rho^{-1}f_A{}^X\tilde{\nabla}f_{XB}, \tag{3}$$

where $\nabla$ and $\tilde{\nabla}$ are the two-dimensional gradient operators

$$\nabla = (\partial_3, \partial_4), \quad \tilde{\nabla} = (\partial_4, -\partial_3). \tag{4}$$

Analogous to the Ernst[4] formulation, the complex combination

$$H_{AB} = f_{AB} + i\psi_{AB} \tag{5}$$

satisfies

$$\nabla H_{AB} = -i\rho^{-1}f_A{}^X\tilde{\nabla}H_{XB}. \tag{6}$$

It was shown[5] that these field equations imply the existence of an infinite heierarchy of potentials $\overset{n+1}{H}_{AB}$, which obey the following field equations:

$$\nabla\overset{n+1}{H}_{AB} = -i\rho^{-1}f_A{}^X\tilde{\nabla}\overset{n+1}{\tilde{H}}_{XB}, \tag{7}$$

where $\overset{n+1}{\tilde{H}}_{AB}$ is constructed from lower order potentials as

$$\overset{n+1}{\tilde{H}}_{AB} = i(\overset{1n}{N}_{AB} + H_{AX}\overset{n}{H}{}^X{}_B), \tag{8}$$

where

$$\nabla\overset{1n}{N}_{AB} = H^*_{XA}\nabla\overset{n}{H}{}^X{}_B. \tag{9}$$

The hierarchy is constructed starting with $n=1$, with

$$\overset{1}{H}_{AB} \equiv H_{AB}.$$

(Notice that the hierarchy of potentials is defined only up to a constant. We would use this gauge freedom, whenever necessary.)

It was also observed[5] that these higher potentials when computed for the flat spacetime were related to each other by certain algebraic equations. The same was noticed[6] to be true for the Schwarzschild case. We will see here that one can write down a general scheme classifying the solutions according to one algebraic relationship among the potentials.

Appropriate field equations given in (7) can be combined to give

$$\nabla(i\overset{n+1}{H}_{11} - \overset{n}{H}_{12}) = -i\rho^{-1}[f_{12}\tilde{\nabla}(i\overset{n+1}{H}_{11} - \overset{n}{H}_{12}) - f_{11}\tilde{\nabla}(i\overset{n+1}{H}_{21} - \overset{n}{H}_{22})], \tag{10}$$

and

$$\nabla(i\overset{n+1}{H}_{21} - \overset{n}{H}_{22}) = -i\rho^{-1}[f_{22}\tilde{\nabla}(i\overset{n+1}{H}_{11} - \overset{n}{H}_{12}) - f_{21}\tilde{\nabla}(i\overset{n+1}{H}_{21} - \overset{n}{H}_{22})], \tag{11}$$

which imply that

$$i\overset{n+1}{H}_{11} = \overset{n}{H}_{12} \tag{12}$$

if and only if

$$i\overset{n+1}{H}_{21} = \overset{n}{H}_{22}. \tag{13}$$

Taking the gradient of Eq. (8) and using Eq. (9), one obtains

$$\nabla(i\overset{n+2}{H}_{11} - \overset{n+1}{H}_{12}) = i[(H_{11}+H^*_{11})\nabla(i\overset{n+1}{H}_{21} - \overset{n}{H}_{22}) - (H^*_{21}+H_{12})\nabla(i\overset{n+1}{H}_{11} - \overset{n}{H}_{12}) + (i\overset{n+1}{H}_{21} - \overset{n}{H}_{22})\nabla H_{11} - (i\overset{n+1}{H}_{11} - \overset{n}{H}_{12})\nabla H_{12}] \tag{14}$$

and

a)Work supported in part by the National Science Foundation under Grant PHY76-12246.

$$\nabla(i\overset{n+2}{H}_{21} - \overset{n+1}{H}_{22})$$
$$= i[(H^*_{12} + H_{21})\nabla(i\overset{n+1}{H}_{21} - \overset{n}{H}_{22}) - (H^*_{22} + H_{22})\nabla(i\overset{n+1}{H}_{11} - \overset{n}{H}_{12}) + (i\overset{n+1}{H}_{21} - \overset{n}{H}_{22})\nabla H_{21} - (i\overset{n+1}{H}_{11} - \overset{n}{H}_{12})\nabla H_{22}]. \tag{15}$$

It follows from Eqs. (12), (13), and (14) that if

$$i\overset{n+1}{H}_{11} = \overset{n}{H}_{12} \tag{16}$$

for some value of $n$, then it would hold good for all higher values, also.

We classify our solutions by the above single (complex) algebraic condition. Thus, the $l$th order solutions are characterized by the fact that $l$ is the lowest integer for which Eq. (16) is satisfied.

We find that, for the static case, the Weyl solution

$$f_{11} = \left(\frac{x-1}{x+1}\right)^l \tag{17}$$

is the $l$th order solution in our classification, where $x$ and $y$ are prolate spheroidal coordinates. For example, for $l=1$, the metric is the Schwarzschild solution, where the unit of length is chosen to be mass, $m$. A simple computation gives

$$i\overset{2}{H}_{11} = 2i(x-1)y = \overset{1}{H}_{12}. \tag{18}$$

For $l=2$, the relevant higher potentials are

$$\begin{aligned} &\overset{1}{H}_{12} = 2i(xy - 2y), \quad \overset{2}{H}_{11} = \frac{2(x-1)^2(x+2)y}{(x+1)^2} \\ &\overset{2}{H}_{12} = i(x-1)^2(6y^2-2), \quad \overset{3}{H}_{11} = (x-1)^2(6y^2-2), \end{aligned} \tag{19}$$

showing that the condition (16) is satisfied for $l=2$.

For the stationary case, one finds that the Kerr solution satisfies equation (16) for $n=1$, i.e., one can choose a suitable gauge in which the appropriate potentials are

$$\begin{aligned} f_{11} &= \frac{p^2x^2 + q^2y^2 - 1}{(px+1)^2 + q^2y^2}, \quad \psi_{11} = \frac{2qy}{(px+1)^2 + q^2y^2}, \\ f_{12} &= \frac{2q[x(1-y^2) + p(x^2-y^2)]}{(px+1)^2 + q^2y^2}, \\ \psi_{21} &= -\frac{2y[(px+1)^2 - q^2]}{p[(px+1)^2 + q^2y^2]}, \end{aligned} \tag{20}$$

and

$$i\overset{2}{H}_{11} = \overset{1}{H}_{12} (= f_{12} + i\psi_{21} + 2ixy)$$

Since the Weyl solutions given in Eq. (17) are the static cases of TS solutions, it is evident that the higher order stationary solutions in our classification scheme should contain the TS solutions; a straightforward (but rather tedious) calculation to show that it is indeed the case will be done elsewhere.[7]

The question arises as to what, if any, are the solutions other than the TS solutions, which fall in our classification scheme. Let us restrict ourselves to the simplest possible case, i.e., the static $n=1$, case. Using Eq. (3), one can show that the condition

$$i\overset{2}{H}_{11} = \overset{1}{H}_{12}$$

reduces to the following two equation:

$$(\xi^2 - 1)\left(x\frac{\partial \xi}{\partial x} - \xi\right) = \xi^2\left(\frac{1}{y}\frac{\partial \xi}{\partial y} - y\frac{\partial \xi}{\partial y}\right) \tag{21}$$

and

$$(\xi^2 - 1)\left(y\frac{\partial \xi}{\partial y} - \xi\right) = \xi^2\left(\frac{1}{x}\frac{\partial \xi}{\partial x} - x\frac{\partial \xi}{\partial x}\right), \tag{22}$$

where

$$\xi = \frac{f_{11} + 1}{1 - f_{11}} \tag{23}$$

It is quite clear that $\xi = x$ or $\xi = y$ satisfies Eqs. (21) and (22). The most general solutions of these equations are given by

$$F\left[\frac{\xi}{x}, \frac{\xi^2 - 1}{y^2 - 1}\right] = 0 \tag{24}$$

and

$$G\left[\frac{\xi}{y}, \frac{\xi^2 - 1}{x^2 - 1}\right] = 0, \tag{25}$$

where $F$ and $G$ are arbitrary functions. It is not clear whether one can find suitable forms for $F$ and $G$ such that one can produce a solution for $\xi$ which is neither $x$ nor $y$.

## ACKNOWLEDGMENTS

I would like to thank Dr. W. Kinnersley and Mr. T.C. Jones for useful conversations.

[1]R. Geroch, J. Math. Phys. 13, 394 (1972).
[2]A. Tomimatsu and H. Sato, Prog. Theor. Phys. 50, 95 (1973).
[3]W. Kinnersley, J. Math. Phys. 18, 1529 (1977).
[4]F.J. Ernst, Phys. Rev. 167, 1175 (1968).
[5]W. Kinnersley and D.M. Chitre, J. Math. Phys. 18, 1538 (1977).
[6]T.C. Jones, private communication.
[7]D.M. Chitre and T.C. Jones, to be published.

# Noncompletely reducible representations of the Poincaré group associated with the generalized Lorentz gauge

G. Rideau

*Université Paris VII and ER 77, Paris, Cedex 05, France*

(Received 20 June 1977)

From complete resolution of a cohomological equation determining the 1-cocycle of an extension of one mass-null scalar representation by a vectorial mass-null representation of the Poincaré group, we build a one-parameter family of inequivalent noncompletely reducible representations of this group. Each of them leads to a quantum field theory by the Fock quantization process to a description of the electromagnetic field which turns out to be identical with field theories built by others in the general framework of the generalized Lorentz gauge.

## INTRODUCTION

In the Gupta—Bleuler formulation of Maxwell's theory a noncompletely reducible representation $U_0(a,\Lambda)$ of the Poincaré group $\mathcal{P}$ is basically used.

This representation is realized in a space of 4-components function $\varphi_\mu(k)$, $\mu=0,1,2,3$, defined on the future cone $C_+$. If we introduce the "variable"

$$\omega(k)=k^\mu\varphi_\mu(k),\quad k\in C_+, \tag{1}$$

and eliminate, for instance, $\varphi_0(k)$, we can write at least formally

$$U_0(a,\Lambda)=\begin{vmatrix} V(a,\Lambda) & T_0(a,\Lambda)\,W(a,\Lambda)\\ 0 & W(a,\Lambda)\end{vmatrix}, \tag{2}$$

where:

(1) $V(a,\Lambda)$ is a representation of $\mathcal{P}$ realized in some space $E$ of three-component functions of $C_+$, according to

$$(V(a,\Lambda)\varphi)_i(k)=\exp(ia\cdot k)\left(\Lambda_i{}^j-\Lambda_i{}^0\,\frac{(\Lambda^{-1}k)^j}{|\Lambda^{-1}\mathbf{k}|}\right)\varphi_j(\Lambda^{-1}k),$$
$$k\in C_+, \tag{3}$$

(2) $W(a,\Lambda)$ is a representation of $\mathcal{P}$ realized in the space $F$ of the $\omega(k)$ according to

$$W(a,\Lambda)\,\omega(k)=\exp(ia\cdot k)\,\omega(\Lambda^{-1}k). \tag{4}$$

(3) $T_0(a,\Lambda)$ is a column operator mapping $F$ into $E$ according to

$$(T_0(a,\Lambda)\,\omega)_i(k)=\Lambda_i{}^0\,\frac{\omega(\mathbf{k})}{|\Lambda^{-1}\mathbf{k}|},\quad k\in C_+. \tag{5}$$

Here we understand that Latin indices go from 1 to 3 and Greek from 0 to 3.

Obviously, $T_0(a,\Lambda)$ is a solution of the cohomological equation

$$T((a_1,\Lambda_1)(a_2,\Lambda_2))=T(a,\Lambda_1)+V(a_1\Lambda_1)\,T(a_2,\Lambda_2)\,W(a_1\Lambda_1)^{-1}, \tag{6}$$

which ensures that (2) is actually a representation. It is the equation determining the so-called one-cocycle of extension of representation $W$ by representation $V$,

$$T(a,\Lambda)\in Z^1(W,V).$$

Up to now, nothing has been said about the topological structure of $E$ and $F$. To avoid difficulties associated with the vertex of the cone, we shall take for $F$ the space of $C^\infty$ functions with compact support on $C_+$ and for $E$ the space of functions with components in $F$. We assume that $T(a,\Lambda)$ is a continuous mapping from $F$ to $E$ depending continuously on $(a,\Lambda)$. Then each solution of Eq. (6) gives rise to a continuous representation of $\mathcal{P}$ in $E+F$. Some of them are proper to a field quantization process leading to eventually distinct descriptions of the quantum electromagnetic field. Our paper is devoted to a comprehensive study of the solutions of (6) in this framework, and of their applications in field theory, especially with respect to the resulting gauge condition and the form of the field equations. It is somewhat noteworthy that the description of the electromagnetic field in Laudau gauge (cf. Refs. 1 and 2) is obtained as a special case of our general construction.

## 1. RESOLUTION OF (6)

It will be convenient to identify $F$ with the space $\mathcal{D}_0(\mathbb{R}^3)$ of $C^\infty$ functions on $\mathbb{R}^3$ with compact support, turning to zero in a neighborhood of the origin. $\mathcal{D}_0(\mathbb{R}^3)$ provided with the topology induced by the usual topology of $\mathcal{D}(\mathbb{R}^3)$ is a nuclear space. Therefore, to each component of $T(a,\Lambda)$ we can associate a distribution kernel $T_i(a,\Lambda;\mathbf{k},\mathbf{k}')\in(\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^3))'$ with the proviso that (cf. Ref. 3)

$$\int d\mathbf{k}'T_i(a,\Lambda;\mathbf{k},\mathbf{k}')\,\varphi(\mathbf{k}') \tag{7}$$

is in $\mathcal{D}_0(\mathbb{R}^3)$ when $\varphi(\mathbf{k})$ is in $\mathcal{D}_0(\mathbb{R}^3)$. (We choose the symbolic notation common among the physicists and we use the measure $d\mathbf{k}$ instead of the invariant measure on $C^+$ in order to enjoy the usual distribution properties.)

Concerning the elements in $\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^3)$, we further need the following technical lemma:

*Lemma* 1: Let $\varphi(\mathbf{k},\mathbf{k}')\in\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^3)$ with support in the domain

$$\epsilon<|\mathbf{k}|<A,\quad \epsilon<|\mathbf{k}'|<A,\quad 0<\epsilon<A<\infty. \tag{8}$$

We can write

$$\varphi(\mathbf{k},\mathbf{k}')=\sum_i\left[1-\hat{f}_i\left(\frac{\mathbf{k}-\mathbf{k}'}{2A}\right)\right]\psi_i(\mathbf{k},\mathbf{k}')+\varphi_0(\mathbf{k},\mathbf{k}'), \tag{9}$$

where $\varphi_0(\mathbf{k},\mathbf{k}')$ and $\psi_i(\mathbf{k},\mathbf{k}')$, $i=1,2,3$ are in $\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^3)$, and the $\hat{f}_i(\mathbf{r})$, $i=1,2,3$, are Fourier transforms of functions in $\mathcal{D}(\mathbb{R}^3)$ such that $(1-\hat{f}_i(\mathbf{k}))$ is dif-

 0022-2488/78/1907-1627$1.00  

ferent from zero when $0<|\mathbf{k}|\leq 1$ and proportional to $k_i$ in a sufficiently small neighborhood of the origin.

*Proof*: Let $\alpha(\mathbf{k})$ be a $C^\infty$ function such that

$$\alpha(\mathbf{k})=1,\quad |\mathbf{k}|\leq \tfrac{1}{2},$$

$$\alpha(\mathbf{k})=0,\quad |\mathbf{k}|>1.$$

Then

$$\varphi_0(\mathbf{k},\mathbf{k}')=\varphi(\mathbf{k},\mathbf{k})\left[1-\alpha\left(\frac{\mathbf{k}'}{\epsilon}\right)\right]\alpha\left(\frac{\mathbf{k}'}{2A}\right) \tag{10}$$

is in $\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^3)$. The difference $\varphi(\mathbf{k},\mathbf{k}')-\varphi_0(\mathbf{k},\mathbf{k}')$ is zero when $\mathbf{k}=\mathbf{k}'$. Therefore, we can find $\psi_i'(\mathbf{k},\mathbf{k}')$ in $\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^3)$ [and support in (8)] such that

$$\varphi(\mathbf{k},\mathbf{k}')=\sum_i (k_i-k_i')\,\psi_i'(\mathbf{k},\mathbf{k}')+\varphi_0(\mathbf{k},\mathbf{k}'). \tag{11}$$

Now let $f_i(\mathbf{x})$, $i=1,2,3$, be three functions in $\mathcal{D}(\mathbb{R}^3)$ with Fourier transforms such that $1-\hat{f}_i(\mathbf{k})$ is different from zero when $0<|\mathbf{k}|\leq 1$ and proportional to $k_i$ in a neighborhood of the origin. Then we can write

$$\varphi(\mathbf{k},\mathbf{k}')=\sum\left[1-\hat{f}_i\left(\frac{\mathbf{k}-\mathbf{k}'}{2A}\right)\right]\psi_i(\mathbf{k},\mathbf{k}')+\varphi_0(\mathbf{k},\mathbf{k}'),$$

where the functions $\psi_i(\mathbf{k},\mathbf{k}')$ are defined by

$$\psi_i(\mathbf{k},\mathbf{k}')=k_i-k_i'\Big/\left[1-f_i\left(\frac{\mathbf{k}-\mathbf{k}'}{2A}\right)\right]\psi_i'(\mathbf{k},\mathbf{k}')\quad i=1,2,3$$

are obviously in $\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^3)$ [and support in (8)].

We now solve (6) step by step.

*Proposition* 1: $T_i(a,\mathrm{I\!I};\mathbf{k},\mathbf{k}')$ has the following form,

$$T_i(a,\mathrm{I\!I};\mathbf{k},\mathbf{k}')=\beta_{i\rho}(\mathbf{k})\,a^\rho\delta(\mathbf{k},\mathbf{k}')+[1-\exp(ia_{\bullet}(k-k'))]\,t_i(\mathbf{k},\mathbf{k}'), \tag{12}$$

where $\beta_{i\rho}(\mathbf{k})$ are $C^\infty$ functions outside the origin and $t_i(\mathbf{k},\mathbf{k}')$ is an element of $(\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^3))'$ and applies $\mathcal{D}_0(\mathbb{R}^3)$ into itself.

*Proof*: From (6) results

$$[1-\exp(ia'(k-k'))]\,T_i(a,\mathrm{I\!I};\mathbf{k},\mathbf{k}')=[1-\exp(ia_{\bullet}(k-k'))]\,T_i(a',\mathrm{I\!I};\mathbf{k},\mathbf{k}'). \tag{13}$$

If $f(\mathbf{x})$ is in $\mathcal{D}(\mathbb{R}^3)$ with Fourier transform equal to one at the origin, we get from (13) and the continuity in $\mathbf{a}$ of $T_i\,\mathbf{a},\mathrm{I\!I};\mathbf{k},\mathbf{k}')$

$$(1-\hat{f}(\mathbf{k}-\mathbf{k}'))\,T_i(a,\mathrm{I\!I};\mathbf{k},\mathbf{k}')=[1-\exp(ia_{\bullet}(k-k'))]\int d\mathbf{a}' f(\mathbf{a}')\,T_i(\mathbf{a}',\mathrm{I\!I};\mathbf{k},\mathbf{k}'). \tag{14}$$

Let us denote by $\mathcal{D}_{\epsilon,A}(\mathbb{R}^6)$ the space of $C^\infty$ functions on $\mathbb{R}^6$ with compact support in the domain (8). If $\varphi(\mathbf{k},\mathbf{k}')$ is in $\mathcal{D}_{\epsilon,A}(\mathbb{R}^6)$, using (14) and (9), we get

$$\begin{aligned}\int T_i(\mathbf{a},\mathrm{I\!I};\mathbf{k},\mathbf{k}')\,\varphi(\mathbf{k},\mathbf{k}')\,d\mathbf{k}\,d\mathbf{k}'&=(2A)^3[1-\exp(ia_{\bullet}(k-k'))]\sum_j d\mathbf{k}\,d\mathbf{k}'\,\psi_j(a,\mathrm{I\!I};\mathbf{k},\mathbf{k}')\\&\times\int d\mathbf{a}' f_j(2A\mathbf{a}')\,T_i(\mathbf{a}'\,\mathrm{I\!I};\mathbf{k},\mathbf{k}')+\int\varphi(\mathbf{k},\mathbf{k})\,d\mathbf{k}\\&\times\int d\mathbf{k}'\,T_i(a,\mathrm{I\!I};\mathbf{k},\mathbf{k}')\left[1-\alpha\left(\frac{\mathbf{k}'}{\epsilon}\right)\right]\alpha\left(\frac{\mathbf{k}'}{2A}\right).\end{aligned}$$

By assumption

$$\int d\mathbf{k}'\,T_i(a,\mathrm{I\!I};\mathbf{k},\mathbf{k}')\left[1-\alpha\left(\frac{\mathbf{k}'}{\epsilon}\right)\right]\alpha\left(\frac{\mathbf{k}'}{2A}\right)$$

is in $\mathcal{D}_0(\mathbb{R}^3)$. Furthermore, the mapping

$$\varphi(\mathbf{k},\mathbf{k}')\to\psi_i(\mathbf{k},\mathbf{k}')$$

is obviously linear and continuous with respect to the topology induced on $\mathcal{D}_{\epsilon,A}(\mathbb{R}^6)$ by the topology of $\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^3)$. Therefore, restricted to $\mathcal{D}_{\epsilon,A}(\mathbb{R}^6)$, $T_i(a,\mathrm{I\!I};\mathbf{k},\mathbf{k}')$ is in $\mathcal{D}'_{\epsilon A}(\mathbb{R}^6)$ with the following form,

$$T_i(a,\mathrm{I\!I};\mathbf{k},\mathbf{k}')=[1-\exp(ia_{\bullet}(k-k'))]\,t_i(\mathbf{k},\mathbf{k}')+\beta_i(\mathbf{k},a)\,\delta(\mathbf{k}-\mathbf{k}'), \tag{14'}$$

where $t_i(\mathbf{k},\mathbf{k}')\in\mathcal{D}'_{\epsilon A}(\mathbb{R}^6)$ and $\beta_i(\mathbf{k},a)$ is a $C^\infty$ function when $\epsilon<|\mathbf{k}|<A$. But, $\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^3)$ is the inductive limit of the spaces $\mathcal{D}_{\epsilon A}(\mathbb{R}^6)$ when $\epsilon\to 0$ and $A\to\infty$, and therefore $(\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^3))$ is the projective limit of the spaces $\mathcal{D}'_{\epsilon A}(\mathbb{R}^6)$. This easily implies the validity of (14′) for any $\varphi(\mathbf{k},\mathbf{k}')\in\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^3)$ but, this time, $t_i(\mathbf{k},\mathbf{k}')\in(\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^3))'$ and $\beta_i(\mathbf{k},a)$ is a $C^\infty$ function outside the origin. Finally,

$$T_i(a+a',\mathrm{I\!I};\mathbf{k},\mathbf{k}')=T_i(a,\mathrm{I\!I};\mathbf{k},\mathbf{k}')+\exp(ia(k-k'))\,T_i(a',\mathrm{I\!I};\mathbf{k},\mathbf{k}')$$

and the continuity in $a$ imply $\beta_i(\mathbf{k},a)$ is linear in $a$.

*Proposition* 2: $T_i(0,\Lambda;\mathbf{k},\mathbf{k}')$ has the following form,

$$\begin{aligned}T_i(0,\Lambda;\mathbf{k},\mathbf{k}')&=t_i(\mathbf{k},\mathbf{k}')-\left(\Lambda_i{}^j-\Lambda_i\,\frac{{}^0(\Lambda^{-1}k)^j}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right)t_j(\mathbf{\Lambda}^{-1}\mathbf{k},\mathbf{\Lambda}^{-1}\mathbf{k}')\,\frac{|\mathbf{\Lambda}^{-1}\mathbf{k}'|}{|\mathbf{k}'|}\\&\quad+\theta_i(\Lambda,\mathbf{k})\,\delta(\mathbf{k}-\mathbf{k}')-\theta_{ij}(\Lambda,\mathbf{k})\,\frac{\partial}{\partial k_j'}\,\delta(\mathbf{k}-\mathbf{k}'),\end{aligned} \tag{15}$$

where $\theta_i(\Lambda,\mathbf{k})$, $\theta_{ij}(\Lambda,\mathbf{k})$ are continuous mappings from $\mathrm{SL}(2,\mathbb{C})$ into the set of $C^\infty$ functions on $\mathbb{R}^3$ outside the origin. [By definition, if $T(\mathbf{k},\mathbf{k}')\in(\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^3))'$, we denote by $T(\mathbf{\Lambda}^{-1}\mathbf{k},\mathbf{\Lambda}^{-1}\mathbf{k}')\,|\mathbf{\Lambda}^{-1}\mathbf{k}'|/|\mathbf{k}'|$ the distribution defined by

$$\int T(\mathbf{\Lambda}^{-1}\mathbf{k},\mathbf{\Lambda}^{-1}\mathbf{k}')\,\frac{|\mathbf{\Lambda}^{-1}\mathbf{k}'|}{|\mathbf{k}'|}\,\varphi(\mathbf{k},\mathbf{k}')\,dk\,dk'=\int T(\mathbf{k},\mathbf{k}')\,\frac{|\mathbf{\Lambda}\mathbf{k}'|}{|\mathbf{k}|}\,\varphi(\mathbf{\Lambda}\mathbf{k},\mathbf{\Lambda}\mathbf{k}')\,dk\,dk'.$$

This definition is understood in all the following.]

*Proof*: From (6) and

$$(a,\Lambda)=(a,\mathrm{I\!I})(0,\Lambda)=(0,\Lambda)(\Lambda^{-1}a,\mathrm{I\!I})$$

we get

$$\begin{aligned}[1-\exp(ia(k-k'))]\,T_i(0,\Lambda;\mathbf{k},\mathbf{k}')&=T_i(a,\mathrm{I\!I};\mathbf{k},\mathbf{k}')-\left(\Lambda_i{}^j-\Lambda_i{}^0\,\frac{(\Lambda^{-1}k)^j}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right)\\&\quad\times T_j(\Lambda^{-1}a,\mathrm{I\!I};\mathbf{\Lambda}^{-1}\mathbf{k},\mathbf{\Lambda}^{-1}\mathbf{k}')\,\frac{|\mathbf{\Lambda}^{-1}\mathbf{k}'|}{|\mathbf{k}'|}.\end{aligned}$$

If $f(\mathbf{a})$ is again a function in $\mathcal{D}(\mathbb{R}^3)$ with Fourier transform equal to one at the origin, we can write

$$(1-\hat{f}(\mathbf{k}-\mathbf{k}'))\,T_i(0,\Lambda,\mathbf{k},\mathbf{k}')=\int f(\mathbf{a})\,T_i(\mathbf{a},\mathrm{I\!I};\mathbf{k},\mathbf{k}')\,d\mathbf{a}-\left(\Lambda_i{}^j-\Lambda_i{}^0\,\frac{(\Lambda^{-1}k)^j}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right)$$

$$\times \int f(\mathbf{a})\, T_j(\Lambda^{-1}\mathbf{a}, I; \mathbf{\Lambda}^{-1}\mathbf{k}, \mathbf{\Lambda}^{-1}\mathbf{k}') \frac{|\mathbf{\Lambda}^{-1}\mathbf{k}'|}{|\mathbf{k}'|}\, d\mathbf{a}$$

But, using Proposition 1 and the relation

$$\delta(\mathbf{k}-\mathbf{k}') = \delta(\mathbf{\Lambda}^{-1}\mathbf{k} - \mathbf{\Lambda}^{-1}\mathbf{k}') \frac{|\mathbf{\Lambda}^{-1}\mathbf{k}'|}{|\mathbf{k}'|}$$

this is equivalent to

$$\begin{aligned}(1-\hat{f}(\mathbf{k}-\mathbf{k}'))\, T_i(0,\Lambda;\mathbf{k},\mathbf{k}') &= \frac{1}{i}\frac{\partial \hat{f}}{\partial kj}(0)\left[\beta_{ij}(\mathbf{k}) - \left(\Lambda_i{}^l - \Lambda_i{}^0 \frac{(\Lambda^{-1}k)^l}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right)\Lambda_j{}^\rho \beta_{l\rho}(\mathbf{\Lambda}^{-1}\mathbf{k})\right] \\ &\times \delta(\mathbf{k}-\mathbf{k}') + (1-\hat{f}(\mathbf{k}-\mathbf{k}'))\left[t_i(\mathbf{k},\mathbf{k}') - \left(\Lambda_i{}^j - \Lambda_i{}^0\frac{(\Lambda^{-1}k)^j}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right)\right. \\ &\left.\times t_j(\mathbf{\Lambda}^{-1}\mathbf{k}, \mathbf{\Lambda}^{-1}\mathbf{k}')\frac{|\mathbf{\Lambda}^{-1}\mathbf{k}'|}{|\mathbf{k}'|}\right].\end{aligned}$$

We now apply this formula to the evaluation of

$$\int T_i(0,\Lambda;\mathbf{k},\mathbf{k}')\,\varphi(\mathbf{k},\mathbf{k}')\,d\mathbf{k}\,d\mathbf{k}'$$

when $\varphi(\mathbf{k},\mathbf{k}')$ is in $\mathcal{D}_{\epsilon,A}(\mathbb{R}^6)$ and, therefore, can be put in the form (9). If we notice further that

$$\frac{\partial\varphi}{\partial k_j'}(\mathbf{k},\mathbf{k}) = -\frac{1}{2A}\psi_j(\mathbf{k},\mathbf{k})\frac{\partial \hat{f}_i}{\partial k_j}(0)$$

we get (15) after considerations similar to those made at the end of the proof of Proposition 1.

Incidentally, we obtain the relation

$$i\theta_{ij}(\Lambda,\mathbf{k}) = \beta_{ij}(\mathbf{k}) - \Lambda_j{}^\sigma\left(\Lambda_i{}^l - \Lambda_i{}^0\frac{(\Lambda^{-1}k)^l}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right)\beta_{l\sigma}(\mathbf{\Lambda}^{-1}\mathbf{k}). \tag{16}$$

*Lemma* 2: $\theta_i(\Lambda,\mathbf{k})$ and $\theta_{ij}(\Lambda,\mathbf{k})$ verify the following cohomological equations:

$$\theta_i(\Lambda_1\Lambda_2,\mathbf{k}) = \theta_i(\Lambda_1,\mathbf{k}) + \left((\Lambda_1)_i{}^l - (\Lambda_1)_i{}^0\frac{(\Lambda^{-1}k)^l}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right)\theta_l(\Lambda_2,\mathbf{\Lambda}_1^{-1}\mathbf{k}) \tag{17}$$

$$\begin{aligned}\theta_{ij}(\Lambda_1\Lambda_2,\mathbf{k}) &= \theta_{ij}(\Lambda_1,\mathbf{k}) \\ &+ \left((\Lambda_1)_i{}^l - (\Lambda_1)_i{}^0\frac{(\Lambda_1^{-1}k)^l}{|\mathbf{\Lambda}_1^{-1}\mathbf{k}|}\right) \\ &\times\left((\Lambda_1)_j{}^m - (\Lambda_1)_j{}^0\frac{(\Lambda_1^{-1}k)^m}{|\mathbf{\Lambda}_1^{-1}\mathbf{k}|}\right)\theta_{lm}(\Lambda_2,\mathbf{\Lambda}_1^{-1}\mathbf{k}).\end{aligned} \tag{18}$$

*Proof*: It is almost obvious from applying (6) to $T_i(0,\Lambda,\mathbf{k},\mathbf{k}')$ and Proposition 2, if we note that $\varphi(\mathbf{k},\mathbf{k}')$ and $(\partial\varphi/\partial k_j)(\mathbf{k},\mathbf{k}')$ are "independent variables."

*Lemma* 3: We have

$$\beta_{i0}(\mathbf{k}) = \delta\frac{k_i}{|\mathbf{k}|} - \beta_{ij}(\mathbf{k})\frac{k_j}{|\mathbf{k}|}, \tag{19}$$

where $\delta$ is some constant.

*Proof*: Let us introduce

$$\delta_l(\mathbf{k}) = \beta_{l\rho}(\mathbf{k})\,k^\rho, \quad k^0 = |\mathbf{k}| \tag{20}$$

and let us rewrite (16) in terms of $\beta_{lm}(\mathbf{k})$ and $\delta(\mathbf{k})$. If we substitute the resulting expression of $\theta_{ij}(\Lambda,\mathbf{k})$ in (18) with $\Lambda_1=\Lambda$, and $\Lambda_2=\Lambda^{-1}$, we get the equation

$$\left(\Lambda_i{}^l - \Lambda_i{}^0\frac{(\Lambda^{-1}k)^l}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right)\delta_l(\mathbf{\Lambda}^{-1}\mathbf{k}) = \delta_i(\mathbf{k}). \tag{21}$$

The lemma will be true if we show that the general solution of (21) has the form

$$\delta_i(\mathbf{k}) = \delta k_i, \quad \delta \text{ being some constant.} \tag{22}$$

This is the object of:

*Lemma* 4: A vectorial function infinitely differentiable outside the origin and verifying (21) is a multiple of **k**.

*Proof*: Let $\gamma$ be the point on $C^+$ with coordinates

$$\gamma_0 = \gamma_3 = \tfrac{1}{2}, \quad \gamma_1 = \gamma_2 = 0$$

and let $\Gamma$ be the subgroup of $SL(2,\mathbb{C})$ which leaves $\gamma$ invariant.

Writing (21) at $\gamma$ for $\Lambda\in\Gamma$, we find

$$\delta_1(\gamma) = \delta_2(\gamma) = 0, \quad \delta_3(\gamma) = \delta/2,$$

where $\delta$ is some constant. Now

$$\delta_i(\mathbf{k}) = \left((\Lambda_k^{-1})_i{}^l - (\Lambda_k^{-1})_i{}^0\frac{\gamma^l}{|\boldsymbol{\gamma}|}\right)\delta_l(\boldsymbol{\gamma}), \tag{23}$$

where $\Lambda_k\in SL(2,\mathbb{C})$ is such that

$$\Lambda_k^{-1}\gamma = k, \quad k = (|\mathbf{k}|, \mathbf{k}). \tag{24}$$

But then, the right member of (23) is precisely $\delta k_i$. Taking (16) into account, the lemma implies the following expression of $\theta_{ij}(\Lambda,\mathbf{k})$,

$$\begin{aligned}i\theta_{ij}(\Lambda,\mathbf{k}) &= \beta_{ij}(\mathbf{k}) - \left(\Lambda_i{}^l - \Lambda_i{}^0\frac{(\Lambda^{-1}k)^l}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right) \\ &\times\left(\Lambda_j^m - \Lambda_j{}^0\frac{(\Lambda^{-1}k)^m}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right)\beta_{lm}(\mathbf{\Lambda}^{-1}\mathbf{k}) - \delta\frac{k_i}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\Lambda_j{}^0.\end{aligned} \tag{25}$$

*Proposition* 3: $\theta_i(\Lambda,\mathbf{k})$ has the following general form,

$$\begin{aligned}\theta_i(\Lambda,\mathbf{k}) &= \lambda\frac{\Lambda_i{}^0}{|\mathbf{\Lambda}^{-1}\mathbf{k}|} \\ &+ \left(\Lambda_i{}^l - \Lambda_i{}^0\frac{(\Lambda^{-1}k)^l}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right)t_l'(\mathbf{\Lambda}^{-1}\mathbf{k}) - t_i'(\mathbf{k}),\end{aligned} \tag{26}$$

where $\lambda$ is a constant and $t_i'(\mathbf{k})$ are $C^\infty$ functions outside the origin.

*Proof*: As above, we take $\mathbf{k}=\boldsymbol{\gamma}$, and $\Lambda_1,\Lambda_2\in SL(2,\mathbb{C})$ in (17). We find

$$\theta_i(\Lambda,\boldsymbol{\gamma}) = 2\lambda\Lambda_i{}^0 + \left(\Lambda_i{}^l - \Lambda_i{}^0\frac{\gamma^l}{|\boldsymbol{\gamma}|} - \delta_i{}^l\right)\mu_l, \quad \Lambda\in\Gamma,$$

where $\lambda$ is a constant and $\mu_l$ the components of a 3-vector.

Now, with $\Lambda_k$ as in (24), we get

$$\theta_i(\Lambda_k\Lambda,\boldsymbol{\gamma}) = \theta_i(\Lambda_k,\boldsymbol{\gamma}) + \left((\Lambda_k)_i{}^l - (\Lambda_k)_i{}^0\frac{k^l}{|\mathbf{k}|}\right)\theta_l(\Lambda,\mathbf{k}).$$

Using

$$\Lambda_k\Lambda = (\Lambda_k\Lambda\Lambda^{-1}_{\Lambda^{-1}k})\,\Lambda_{\Lambda^{-1}k}, \quad \Lambda_k\Lambda\Lambda^{-1}_{\Lambda^{-1}k}\in\Gamma,$$

we have

$$\begin{aligned}\theta_i(\Lambda_k\Lambda,\boldsymbol{\gamma}) &= \theta_i(\Lambda_k\Lambda\Lambda^{-1}_{\Lambda^{-1}k},\boldsymbol{\gamma}) \\ &+ \left((\Lambda_k\Lambda\Lambda^{-1}_{\Lambda^{-1}k})_i{}^l - (\Lambda_k\Lambda\Lambda^{-1}_{\Lambda^{-1}k})_i{}^0\frac{\gamma^l}{|\boldsymbol{\gamma}|}\right)\theta_l(\Lambda_{\Lambda^{-1}k},\boldsymbol{\gamma}).\end{aligned}$$

Finally, from the cohomological properties of $(\Lambda_i{}^j - \Lambda_i{}^0(\Lambda^{-1}k)^j/|\Lambda^{-1}\mathbf{k}|)$ and the relation

$$(\Lambda\Lambda^{-1}_{\Lambda^{-1}k})_i{}^0 = \left(\Lambda_i{}^l - \Lambda_i{}^0 \frac{(\Lambda^{-1}k)^l}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right)(\Lambda^{-1}_{\Lambda^{-1}k})_l{}^0 + \frac{|\gamma|}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\,\Lambda_i{}^0$$

we get (26) with

$$t_i'(\mathbf{k}) = (\Lambda_k^{-1})_i{}^0 + \left((\Lambda_k^{-1})_i{}^0 \frac{\gamma^l}{|\boldsymbol{\gamma}|}\right)(\mu_l + \theta_l(\Lambda_k, \gamma)).$$

Putting together all these results, we can state the following theorem.

*Theorem* 1: $T_i(a, \Lambda; \mathbf{k}, \mathbf{k}')$ has the following general form,

$$\begin{aligned}T_i(a, \Lambda; \mathbf{k}, \mathbf{k}') = \lambda\, &\frac{\Lambda_i^0}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\,\delta(\mathbf{k}-\mathbf{k}')\\ &- i\mu\,\frac{k_i}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\left(\Lambda_j{}^0\frac{\partial}{\partial k_j'}\,\delta(\mathbf{k}-\mathbf{k}')\right.\\ &\left.+\, i(\Lambda^{-1}a)^0\,\delta(\mathbf{k}-\mathbf{k}')\right)\\ &+ \exp(ia(k-k'))\left(\Lambda_i{}^j - \Lambda_i{}^0\frac{(\Lambda^{-1}k)^j}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right)\\ &T_j(\mathbf{\Lambda}^{-1}\mathbf{k}, \Lambda^{-1}\mathbf{k}') - T_j(\mathbf{k}, \mathbf{k}'),\end{aligned} \tag{27}$$

where $\lambda$ and $\mu$ are some constants and $T_i(\mathbf{k}, \mathbf{k}')$ are some distributions in $(\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^3))'$ applying $\mathcal{D}_0(\mathbb{R}^3)$ into $\mathcal{D}_0(\mathbb{R}^3)$.

*Proof*: From (6), we have

$$\begin{aligned}T_i(a, \Lambda; \mathbf{k}, \mathbf{k}') = T_i(0, \Lambda; \mathbf{k}, \mathbf{k}') + &\left(\Lambda_i{}^j - \Lambda_i{}^0\frac{(\Lambda^{-1}k)^j}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right)\\ &\times T_j(\Lambda^{-1}a, I; \mathbf{\Lambda}^{-1}\mathbf{k}, \mathbf{\Lambda}^{-1}\mathbf{k}')\,\frac{|\mathbf{\Lambda}^{-1}\mathbf{k}'|}{|\mathbf{k}'|}.\end{aligned}$$

We here insert (15) and (12) and take in account (26), (25), and (19). If we note the relations:

$$\exp(ia(k-k'))\,\delta(\mathbf{\Lambda}^{-1}\mathbf{k} - \mathbf{\Lambda}^{-1}\mathbf{k}')\,\frac{|\mathbf{\Lambda}^{-1}\mathbf{k}'|}{|\mathbf{k}'|} = \delta(\mathbf{k}-\mathbf{k}'),$$

$$\begin{aligned}\exp(ia(k-k'))\,&\frac{\partial}{\partial k_j'}\,\delta(\mathbf{\Lambda}^{-1}\mathbf{k} - \mathbf{\Lambda}^{-1}\mathbf{k}')\,\frac{|\mathbf{\Lambda}^{-1}\mathbf{k}'|}{|\mathbf{k}'|}\\ &= \left(\Lambda_l{}^j - \Lambda_l{}^0\frac{(\Lambda^{-1}k)^j}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right)\frac{\partial}{\partial k_l}\,\delta(\mathbf{k}-\mathbf{k}')\\ &\quad - i\left((\Lambda^{-1}a)^j - (\Lambda^{-1}a)^0\frac{(\Lambda^{-1}k)^j}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right)\delta(\mathbf{k}-\mathbf{k}')\end{aligned}$$

we finally get with (27)

$$\begin{aligned}T_i(\mathbf{k}, \mathbf{k}') = &- t_i'(\mathbf{k}, \mathbf{k}') + t_i'(\mathbf{k})\,\delta(\mathbf{k}-\mathbf{k}')\\ &- i\beta_{ij}\,\frac{\partial}{\partial k_j'}\,\delta(\mathbf{k}-\mathbf{k}').\end{aligned}$$

*Remark*: The last two terms in the right member of (27) turn out to be the trivial 1-cocycle generated by the $T_i(\mathbf{k}, \mathbf{k}')$'s. From the point of view of representation theory, these terms are unmaterial and are easily eliminated if we perform the transformation

$$(\varphi, \omega) \to (\varphi + T\omega, \omega),$$

where $T$ is the operator from $E$ to $F$ corresponding to the distribution kernels $T_i(\mathbf{k}, \mathbf{k}')$. Therefore, in the following, we are concerned only with the first two terms in the right member of (27), so that we are finally faced with a two (complex) parameter family of representations of the Poincaré group, written

$$\begin{aligned}\varphi_i(\mathbf{k}) \xrightarrow{(a,\Lambda)} \exp(ia_\cdot k)&\left((\Lambda_i{}^j - \Lambda_i{}^0\frac{(\Lambda^{-1}k)^j}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\right)\varphi_j(\mathbf{\Lambda}^{-1}\mathbf{k})\\ &+ \lambda\frac{\Lambda_i^0}{|\mathbf{\Lambda}^{-1}\mathbf{k}|}\,\omega(\mathbf{\Lambda}^{-1}\mathbf{k}) - i\mu\,\frac{k_i}{|\mathbf{k}|}\left(\Lambda^0{}_l\,\frac{\partial\omega}{\partial k_l}(\mathbf{\Lambda}^{-1}\mathbf{k})\right.\\ &\left.+\, ia_0\omega(\mathbf{\Lambda}^{-1}\mathbf{k})\right)\end{aligned} \tag{28}$$

$$\omega(\mathbf{k}) \xrightarrow{(a,\Lambda)} \exp(ia_\cdot k)\,\omega(\mathbf{\Lambda}^{-1}\mathbf{k}), \tag{28}$$

where $\varphi_i(\mathbf{k})$ and $\omega(\mathbf{k})$ are in $\mathcal{D}_0(\mathbb{R}^3)$.

## 2. CONTINUOUS INVARIANT SESQUILINEAR FORM AND EQUIVALENCE

Let us assume $\lambda \neq 0$. Then, $\omega(\mathbf{k})$ in (28) can be replaced by a "variable" $\varphi_0(\mathbf{k})$ such that

$$\omega(\mathbf{k}) = \frac{1}{\lambda}\,k^\mu\varphi_\mu(\mathbf{k}), \quad k^0 = |\mathbf{k}|$$

and with

$$\Omega(\mathbf{k}) = k^\mu\varphi_\mu(\mathbf{k}),$$

we get the following form of (28),

$$\begin{aligned}\varphi_\mu(\mathbf{k}) \xrightarrow{(a,\Lambda)} \exp(ia_\cdot k)&\left\{\Lambda_\mu{}^\nu\varphi_\nu(\mathbf{\Lambda}^{-1}\mathbf{k}) - i\,\frac{\mu}{\lambda}\,\frac{k_\mu}{|\mathbf{k}|}\right.\\ &\left.\times\left(\Lambda^0{}_l\,\frac{\partial\Omega}{\partial k_l}(\mathbf{\Lambda}^{-1}\mathbf{k}) + ia_0\Omega(\mathbf{\Lambda}^{-1}\mathbf{k})\right)\right\},\end{aligned} \tag{29}$$

i.e., a representation realized in the direct sum $L$ of four spaces $\mathcal{D}_0(\mathbb{R}^3)$. It is clear that the representations with the same value of $\mu/\lambda$ are equivalent. We put in the following $d = i\mu/\lambda$, the corresponding representation being denoted by $U_d(a, \Lambda)$.

If $B_{d,d'}(\varphi, \varphi')$ is a nondegenerate separately continuous sesquilinear form invariant when $U_d(a, \Lambda)$ acts on $\varphi$ and $U_{d'}(a, \Lambda)$ acts on $\varphi'$, we can state the following proposition.

*Proposition* 4: $B_{d,d'}(\varphi, \varphi')$ exists only if $d' = \bar{d}$, and then it is written

$$\begin{aligned}B_{d,\bar{d}}(\varphi, \varphi') = -a\int\frac{d\mathbf{k}}{|\mathbf{k}|}&\left(\overline{\varphi_\mu'}(\mathbf{k})\,\varphi^\mu(\mathbf{k}) + d\,\frac{\overline{\Omega'}(\mathbf{k})\,\Omega(\mathbf{k})}{|\mathbf{k}|^2}\right)\\ &+ b\int\frac{d\mathbf{k}}{|\mathbf{k}|}\,\overline{\Omega'}(\mathbf{k})\,\Omega(\mathbf{k}),\end{aligned} \tag{30}$$

where $a$ and $b$ are some constants and $a \neq 0$.

*Proof*: From our assumption, we can write

$$B_{d,d'}(\varphi, \varphi') = \int \overline{\varphi_\mu'}(\mathbf{k})\,B^{\mu\nu}(\mathbf{k}, \mathbf{k}')\,\varphi_\nu(\mathbf{k}')\,d\mathbf{k}\,d\mathbf{k}',$$

where $B^{\mu\nu}(\mathbf{k}, \mathbf{k}') \in (\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^3))'$.

The invariance with respect to spatial translations implies

$$B^{\mu\nu}(\mathbf{k}, \mathbf{k}') = b^{\mu\nu}(\mathbf{k})\,\delta(\mathbf{k}-\mathbf{k}'), \quad b^{\mu\nu}(\mathbf{k}) \in \mathcal{D}_0'(\mathbb{R}^3).$$

From the invariance with respect to time translation, we get

$$k_\rho b^{\rho\mu}(\mathbf{k}) = \frac{1}{d'}\,k^\mu\beta(\mathbf{k}), \quad \beta^{\mu\rho}(\mathbf{k})k_\rho = \frac{1}{d}\,k^\mu\beta(\mathbf{k}), \quad \beta(\mathbf{k}) \in \mathcal{D}_0'(\mathbb{R}^3). \tag{31}$$

The Lorentz invariance is expressed by

$$b^{\mu\nu}(\mathbf{k}) = \Lambda_\rho{}^\mu \Lambda_\sigma{}^\nu b^{\rho\sigma}(\Lambda\mathbf{k}) \frac{|\Lambda\mathbf{k}|}{|\mathbf{k}|} + k^\mu k^\nu \Lambda^0{}_l \frac{\partial}{\partial k_l} \frac{\beta(\Lambda\mathbf{k})}{|\mathbf{k}|} \tag{32}$$

which implies

$$|\mathbf{k}|\beta(\mathbf{k}) = |\Lambda\mathbf{k}|\beta(\Lambda\mathbf{k}).$$

Therefore, since $\beta(\mathbf{k}) \in \mathcal{D}_0'(\mathbb{R}^3)$, we get

$$\beta(\mathbf{k}) = \beta/|\mathbf{k}|.$$

From this, we deduce the following general solution of (32):

$$b^{\mu\nu}(\mathbf{k}) = \frac{1}{|\mathbf{k}|}\left(-ag^{\mu\nu} + bk^\mu k^\nu + \beta\frac{k^\mu k^\nu}{|\mathbf{k}|^2}\right).$$

Now, (31) gives the relation

$$\beta = -a\bar{d}' = -ad.$$

Therefore, $d = \bar{d}'$ and we get (30) with $\beta = -da$; $a \neq 0$ expresses the nondegeneracy.

*Corollary* 1: There exists a nondegenerate sesquilinear form invariant with respect to $U_d(a,\Lambda)$ if and only if $d$ is real. Its general form is given by (30).

*Corollary* 2: The representations $U_d(a,\Lambda)$ and $U_{d'}(a,\Lambda)$ are equivalent if, and only if, $d = d'$.

*Proof*: If $U_d(a,\Lambda)$ and $U_{d'}(a,\Lambda)$ are equivalent, we have

$$U_d(a,\Lambda)A = AU_{d'}(a,\Lambda),$$

where $A$ is an invertible continuous operator on $L$. Therefore, $B_{\bar{d},d}(\varphi', A\varphi)$ is a nondegenerate, sesquilinear form invariant with respect to $U_{\bar{d}}(a,\Lambda)$ and $U_{d'}(a,\Lambda)$, and from Proposition 4, $d' = d$.

## 3. INTERTWINING OPERATORS

As we explain later, we need for the quantization process an intertwining operator between $U_d(a,\Lambda)$ and the following representation of $\mathcal{P}$, denoted $\mathcal{U}(a,\Lambda)$:

$$f_\mu(k) \xrightarrow{(a,\Lambda)} \exp(ia.k)\Lambda_\mu{}^\nu f_\nu(\Lambda^{-1}k), \quad k \in \mathbb{R}^4$$

where by assumption each $f_\mu(k)$ is in $\mathcal{D}_0(\mathbb{R}^4)$. [We postpone to the next section a discussion about this choice of the space of test functions, instead of $\mathcal{S}(\mathbb{R}^4)$.]

We begin by looking for a separately continuous sesquilinear form $B_d(\varphi, f)$ invariant when $U_d(a,\Lambda)$ is acting on $\varphi$ and $\mathcal{U}(a,\Lambda)$ is acting on $f$. We can state

*Proposition* 5: For any $d$, there exists $B_d(\varphi, f)$ given by

$$B_d(\varphi,f) = a\int \frac{d\mathbf{k}}{|\mathbf{k}|}\overline{\varphi_\mu}(\mathbf{k})\left[f^\mu(|\mathbf{k}|,\mathbf{k}) - \bar{d}\frac{k^\mu}{|\mathbf{k}|}\frac{\partial\Omega}{\partial k_0}(|\mathbf{k}|,\mathbf{k})\right]$$
$$+ a\bar{d}\int\frac{d\mathbf{k}}{|\mathbf{k}|}\frac{\bar{\omega}(\mathbf{k})\Omega(|\mathbf{k}|,\mathbf{k})}{|\mathbf{k}|^2}$$
$$+ b\int\frac{d\mathbf{k}}{|\mathbf{k}|}\bar{\omega}(\mathbf{k})\Omega(|\mathbf{k}|,\mathbf{k}), \tag{33}$$

where $a$, $b$ are some constants and

$$\omega(\mathbf{k}) = k^\mu\varphi_\mu(\mathbf{k}), \quad k_0 = |\mathbf{k}|, \quad \Omega(\mathbf{k}) = k^\rho f_\rho(\mathbf{k}), \quad k \in \mathbb{R}^4.$$

*Proof*: As before, to $B_d(\varphi, f)$ correspond distribution kernels $B^{\mu\nu}(\mathbf{k}, k') \in (\mathcal{D}_0(\mathbb{R}^3)\otimes\mathcal{D}_0(\mathbb{R}^4))'$. From the invariance with respect to space–time translations and proceeding as in the proof of Proposition 1, we first get:

$$B_d(\varphi,f) = \int d\mathbf{k}\,\varphi_\mu(\mathbf{k})\,\alpha^{\mu\nu}(\mathbf{k})f_\nu(|\mathbf{k}|,\mathbf{k})$$
$$-\int d\mathbf{k}\,\varphi_\mu(\mathbf{k})\,\beta^{\mu\nu}(\mathbf{k})\frac{\partial f_\nu}{\partial k_0}(|\mathbf{k}|,\mathbf{k}), \tag{34}$$

where

$$\alpha^{\mu\nu}(\mathbf{k}) \in (\mathcal{D}_0(\mathbb{R}^3))' \text{ and } \beta^{\mu\nu}(\mathbf{k}) = \bar{d}\frac{k^\mu}{|\mathbf{k}|}k_\rho\alpha^{\rho\mu}(\mathbf{k}). \tag{35}$$

The Lorentz invariance is expressed by

$$\Lambda_\rho{}^\mu\Lambda_\sigma{}^\nu\alpha^{\rho\sigma}(\Lambda\mathbf{k})\frac{|\Lambda\mathbf{k}|}{|\mathbf{k}|} + \bar{d}k^\mu\Lambda^0{}_l\frac{\partial}{\partial k_l}\frac{(\Lambda k)_\rho\alpha^{\rho\sigma}(\Lambda\mathbf{k})}{|\mathbf{k}|}\Lambda_\sigma{}^\nu$$
$$= \alpha^{\mu\nu}(\mathbf{k}) \tag{36}$$

which implies

$$(\Lambda k)_\rho\alpha^{\rho\sigma}(\Lambda\mathbf{k})\Lambda_\sigma{}^\nu\frac{|\Lambda\mathbf{k}|}{|\mathbf{k}|} = k_\rho\alpha^{\rho\nu}(\mathbf{k}). \tag{37}$$

Since $k_\rho\alpha^{\rho\mu}(\mathbf{k})$ is in $\mathcal{D}_0'(\mathbb{R}^3)$, the general solution of (37) is

$$k_\rho\alpha^{\rho\mu}(\mathbf{k}) = \alpha\frac{k^\mu}{|\mathbf{k}|}, \quad \alpha \text{ const.} \tag{37'}$$

Substituting in (36) and observing that when some distribution $T^{\mu\nu}(\mathbf{k})$ in $\mathcal{D}_0'(\mathbb{R}^3)$ verifies

$$\Lambda_\rho{}^\mu\Lambda_\sigma{}^\nu T^{\rho\sigma}(\Lambda\mathbf{k}) = T^{\mu\nu}(\mathbf{k}),$$

then it has the general form

$$T^{\mu\nu}(\mathbf{k}) = ag^{\mu\nu} + bk^\mu k^\nu, \quad a, b, \text{ constants.}$$

We finally get [$\alpha = a$ from (37′)]

$$\alpha^{\mu\nu}(\mathbf{k}) = a\frac{g^{\mu\nu}}{|\mathbf{k}|} + a\bar{d}\frac{k^\mu k^\nu}{|\mathbf{k}|^3} - a\bar{d}\frac{k^\mu g^{\nu 0}}{|\mathbf{k}|^2} + b\frac{k^\mu k^\nu}{|\mathbf{k}|}.$$

We get (33), after substitution into (34).

*Corollary* 3: For any real $d < \infty$, there exist intertwinning operators $\Pi$ between $U_d(a,\Lambda)$ and $\mathcal{U}(a,\Lambda)$. They have the general form

$$(\Pi f)_\mu(\mathbf{k}) = \alpha\left(f_\mu(|\mathbf{k}|,\mathbf{k}) - d\frac{k\mu}{|\mathbf{k}|}\frac{\partial\Omega}{\partial k_0}(|\mathbf{k}|,\mathbf{k})\right)$$
$$+ \beta k_\mu\Omega(|\mathbf{k}|,\mathbf{k}), \tag{38}$$

where $\alpha$, $\beta$ are some constants and $\Omega(k) = k^\rho f_\rho(k)$.

*Proof*: Let $\Pi$ be an intertwining operator between $U_d(a,\Lambda)$ and $\mathcal{U}(a,\Lambda)$. If $B_{d,\bar{d}}(\varphi,\varphi')$ is a nondegenerate sesquilinear form invariant with respect to $U_d(a,\Lambda)$ and $U_{\bar{d}}(a,\Lambda)$, then $B_{d,\bar{d}}(\Pi f,\varphi')$ is a sesquilinear form invariant with respect to $\mathcal{U}(a,\Lambda)$ and $U_d(a,\Lambda)$. Using (33) and (30) we get (38) after a trivial identification.

## 4. QUANTIZATION PROCESS

In this section, we show that, for any $d$ real and finite, we are able to build a field theory of the electromagnetism where the representation of the Poincaré group is generated by $U_d(a,\Lambda)$ and where Maxwell's equations are valid only in a restricted sense as in Gupta–Bleuler theory (see also Ref. 1).

We first build the space $\mathcal{J}$, the elements of which are vectors $\Psi$ with components $\psi^{(n)}_{\mu_1,\ldots,\mu_n}(\mathbf{k}_1,\ldots,\mathbf{k}_n)$ in the $n$th tensorial power of $\mathcal{D}_0(\mathbb{R}^3)$, depending symmetrically on $\mathbf{k}_i$, $\mu_i$ and vanishing for $n > N(\Psi)$. We extend to $\mathcal{J}$ the representation $U_d(a,\Lambda)$ in the usual way and define on $\mathcal{J}$ a Hermitian nondegenerate invariant form $\beta_d(\Phi,\Psi)$,

$$\beta_d(\Phi,\Psi)=\sum_{n=0}^{\infty}(-1)^n\int\prod_{i=1}^{n}\frac{d\mathbf{k}_i}{|\mathbf{k}_i|}\,\Psi^{(n)}_{\mu_1\ldots\mu_n}(\mathbf{k}_1,\ldots,\mathbf{k}_n)\times\prod_{j=1}^{n}\left(g^{\mu_j\nu_j}+d\,\frac{k_j^{\mu_j}k_j^{\nu_j}}{|\mathbf{k}_j|^2}\right)\Phi^{(n)}_{\nu_1,\ldots,\nu_n}(\mathbf{k}_1,\ldots,\mathbf{k}_n).$$

For each $\varphi$ in $L$, we define $a^+(\varphi)$ on $\mathcal{J}$ by

$$(a^+(\varphi)\Psi)^{(n)}_{\mu_1\ldots\mu_n}(\mathbf{k}_1,\ldots,\mathbf{k}_n)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varphi_{\mu_j}(\mathbf{k}_j)\Psi^{(n-1)}_{\mu_1,\ldots,\hat{\mu}_j,\ldots,\mu_n}(\mathbf{k}_1,\ldots,\hat{\mathbf{k}}_j,\ldots,\mathbf{k}_n)\quad(40)$$

and $a(\varphi)$ by conjugation with respect to $\beta_d(\Phi,\Psi)$,

$$(a(\varphi)\Psi)^{(n)}_{\mu_1,\ldots,\mu_n}(\mathbf{k}_1,\ldots,\mathbf{k}_n)=-\sqrt{n+1}\int\frac{d\mathbf{k}}{|\mathbf{k}|}\,\varphi_\mu(\mathbf{k})\left(g^{\mu\nu}+d\,\frac{k^\mu k^\nu}{|\mathbf{k}|^2}\right)\times\Psi^{(n+1)}_{\nu\mu_1\ldots\mu_n}(\mathbf{k}\,\mathbf{k}_1,\ldots,\mathbf{k}_n).\quad(41)$$

In the following, we introduce, as usual, $a^\mu(\mathbf{k})$ and $a^{+\mu}(\mathbf{k})$ such that, symbolically

$$a(\varphi)=\int a^\mu(\mathbf{k})\,\varphi_\mu(\mathbf{k})\,d\mathbf{k},\quad a^+(\varphi)=\int a^{+\mu}(\mathbf{k})\,\varphi_\mu(\mathbf{k})\,d\mathbf{k}.\quad(42)$$

If now $f_\mu(x)$, $x\in\mathbb{R}^4$, $\mu=0,1,2,3$ are real functions with Fourier transforms $\hat{f}_\mu(\mathbf{k})$ in $\mathcal{D}_0(\mathbb{R}^4)$, we define the vector potential field $A^\mu(x)$ by

$$A(f)=\int A^\mu(x)f_\mu(x)\,d^4x=a^+(\Pi\hat{f})+a(\overline{\Pi\hat{f}}),\quad(43)$$

where

$$(\Pi\hat{f})_\mu(\mathbf{k})=f_\mu(|\mathbf{k}|,\mathbf{k})-d\,\frac{k_\mu}{|\mathbf{k}|}\,\frac{\partial\Omega}{\partial k_0}(|\mathbf{k}|,\mathbf{k})+bk_\mu\Omega(|\mathbf{k}|,\mathbf{k})\quad(44)$$

with $\Omega(k)=k^\rho\hat{f}_\rho(k)$.

Since $\Pi$ is an intertwining operator between $U(a,\Lambda)$ and $U_d(a,\Lambda)$, it is clear that under the action of Poincaré group, we shall have

$$A_\mu(x)\overset{(a,\Lambda)}{\to}(\Lambda^{-1})_\mu{}^\nu A_\nu(\Lambda x+a),$$

i.e., the vector potential field is a true vector with respect to the Poincaré group.

*Proposition* 6: Let $B(x)$ be the scalar field defined by

$$B(f)=i\int_{k\in C_+}d\mathbf{k}\big(k_\mu a^{+\mu}(\mathbf{k})\hat{f}(k)-k_\mu a^\mu(\mathbf{k})\overline{\hat{f}(k)}\big),\quad(45)$$

where $f$ is a real function on $\mathbb{R}^4$ with Fourier transform in $\mathcal{D}_0(\mathbb{R}^4)$. Then we have the following field equations:

$$\Box A_\mu(x)=-2d\partial_\mu B(x),\quad(46)$$

$$\partial^\mu A_\mu(x)=(1-2d)\,B(x).\quad(47)$$

*Proof*: By definition of $A_\mu(x)$ as operator distribution,

$$\int(\Box A^\mu(x))f_\mu(x)\,d^4x=\int A^\mu(x)f'_\mu(x)\,d^4x,\quad f'_\mu(x)=\Box f_\mu(x).$$

But

$$(\Pi\hat{f}')_\mu(\mathbf{k})=2dk_\mu\Omega(|\mathbf{k}|,\mathbf{k}),\quad k_0=|\mathbf{k}|.$$

From (44) and (45), we get (46). Similarly, we have

$$\partial^\mu A_\mu(f)=-A(f'),\quad f'_\mu(x)=\partial_\mu f.$$

But, with $\hat{f}(k)$ in $\mathcal{D}_0(\mathbb{R}^4)$,

$$(\Pi f')_\mu(\mathbf{k})=-i(1-2d)\,k_\mu\hat{f}(|\mathbf{k}|,\quad k_0=|\mathbf{k}|.$$

(47) now results from (44) and (45).

*Corollary* 4: The electromagnetic field $F_{\mu\nu}(x)$ verifies the following field equation,

$$\partial^\mu F_{\mu\nu}(x)=-\partial_\nu B(x).\quad(48)$$

*Proof*: It is obvious from the preceding proposition and

$$F_{\mu\nu}(x)=\partial_\mu A_\nu(x)-\partial_\nu A_\mu(x).$$

Let us define the subspace $\mathcal{J}'$ of the physical vectors as the set of $\Psi$ in $\mathcal{J}$ such that

$$k^\mu a_\mu(\mathbf{k})\Psi=0.$$

It is readily seen that $\beta_d(\Phi,\Psi)$ is nonnegative on $\mathcal{J}'$. Furthermore, from the commutation relations between $a^\mu(\mathbf{k})$ and $a^{+\mu}(\mathbf{k})$, we easily deduce that

$$\beta_d(B(f)\Phi,\Psi)=0$$

for any $\Phi,\Psi\in\mathcal{J}'$.

Thus Maxwell's equations are valid in mean on $\mathcal{J}'$. The same holds for the Lorentz condition, except when $d=\frac{1}{2}$ where we get a true operator equation corresponding to the Landau gauge. Obviously, the true physical space is the quotient $\mathcal{J}'/\mathcal{J}''$ where $\mathcal{J}''$ is the kernel of $\beta_d(\Phi,\Psi)$ restricted to $\mathcal{J}'$ [see also (1)]. It turns out that this quotient does not depend on $d$ and must be identified with the Fock space built only on the transverse components of the vector potential field.

But if the field theories we have obtained substantially have the same physical content as Gupta–Bleuler theory, nevertheless they display some unusual features. In particular, the Hamiltonian cannot be diagonalized and the vector potential field increases linearly with the time. Indeed, let us introduce

$$b^\mu(\mathbf{k})=\left(g^{\mu\nu}-\frac{d}{2}\,\frac{k^\mu k^\nu}{|\mathbf{k}|^2}\right)a_\nu(\mathbf{k}),\quad(49)$$

$$b^{+\mu}(\mathbf{k})=\left(g^{\mu\nu}-\frac{d}{2}\,\frac{k^\mu k^\nu}{|\mathbf{k}|^2}\right)a^+_\nu(\mathbf{k})\quad(50)$$

with commutation relations independent of $d$,

$$[b^\mu(\mathbf{k}),b^{+\nu}(\mathbf{l})]=-\frac{1}{|\mathbf{k}|}\,g^{\mu\nu}\delta(\mathbf{k}-\mathbf{l})\,I.\quad(51)$$

Then we have the following proposition.

*Proposition* 7: The spatial impulsion operators and the Hamiltonian respectively have the following forms:

$$P_i = -\int |\mathbf{k}|\, d\mathbf{k}\, k_i b^{+\mu}(\mathbf{k})\, b_\mu(\mathbf{k}), \quad i=1,2,3, \tag{52}$$

$$P_0 = -\int_{k\in C_+} d\mathbf{k}\, |\mathbf{k}|^2 \left( b^{+\mu}(\mathbf{k})\, b_\mu(\mathbf{k}) + \frac{d}{|\mathbf{k}|}\, (k_\rho b^{+\rho}(\mathbf{k})) \times (k_\sigma b^\sigma(\mathbf{k})) \right). \tag{53}$$

*Proof*: By definition

$$(P_i\Psi)^{(n)}_{\mu_1,\ldots,\mu_n}(\mathbf{k}_1,\ldots,\mathbf{k}_n) = \left(\sum_j (k_j)_i\right) \Psi^{(n)}_{\mu_1,\ldots,\mu_n}(\mathbf{k}_1,\ldots,\mathbf{k}_n), \quad i=1,2,3,$$

$$(P_0\Psi)^{(n)}_{\mu_1,\ldots,\mu_n}(\mathbf{k}_1,\ldots,\mathbf{k}_n) = \sum_j |\mathbf{k}_j| \left( g_{\mu_j}^{\ \nu_j} - \frac{d (k_j)_{\mu_j} (k_j)^{\nu_j}}{|\mathbf{k}_j|^2} \right) \Psi^{(n)}_{\mu_1,\ldots,\mu_j,\ldots,\mu_n}(\mathbf{k}_1,\ldots,\mathbf{k}_n).$$

This is equivalent to:

$$P_i = -\int d\mathbf{k}\, |\mathbf{k}|\, k_i a^{+\mu}(\mathbf{k}) \left( g_\mu^{\ \nu} - d\, \frac{k_\mu k^\nu}{|\mathbf{k}|^2} \right) a_\nu(\mathbf{k}),$$

$$P_0 = -\int d\mathbf{k}\, |\mathbf{k}|^2 a^{+\mu}(\mathbf{k}) \left( g_\mu^{\ \nu} - 2d\, \frac{k_\mu k^\nu}{|\mathbf{k}|^2} \right) a_\nu(\mathbf{k}).$$

Equations (52) and (53) now result directly from (49) and (50).

*Corollary* 5: $P_0$ cannot be diagonalized when $d \neq 0$.

*Proof*: We can write

$$P_0 = -\int_{k\in C_+} |\mathbf{k}|^2 d\mathbf{k} \left( g_\mu^{\ \nu} + d\, \frac{k_\mu k^\nu}{|\mathbf{k}|^2} \right) b^{+\mu}(\mathbf{k})\, b_\nu(\mathbf{k}).$$

Since the commutation rules between $b^{+\mu}(\mathbf{k})$ and $b_\nu(\mathbf{k})$ are diagonal, $P_0$ can be put in a diagonal form if this can be done for the matrix $\{g_\mu^{\ \nu} + d(k_\mu k^\nu)/|\mathbf{k}|^2\}$. But it is readily seen this is impossible.

*Proposition* 8: The vector potential operator has the following form,

$$\begin{aligned} A^\mu(x) = &\int d\mathbf{k} [b^{+\mu}(\mathbf{k}) \exp(ik_\bullet x) + b^\mu(\mathbf{k}) \exp(-ik_\bullet x)] \\ &+ d\, \frac{\partial}{\partial x_\mu} \int \frac{d\mathbf{k}}{|\mathbf{k}|^2} \left\{ \left( \frac{1}{2i} - x_0 |\mathbf{k}| \right) k_\rho b^{+\rho}(\mathbf{k}) \exp(ik_\bullet x) \right. \\ &\left. - \left( \frac{1}{2i} + x_0 |\mathbf{k}| \right) k_\rho b^\rho(\mathbf{k}) \exp(-ik_\bullet x) \right\} \\ &- ib\, \frac{\partial}{\partial x_\mu} \int_{k\in C_+} d\mathbf{k} [k_\rho b^{+\rho}(\mathbf{k}) \exp(ik_\bullet x) - k_\rho b^\rho(\mathbf{k}) \\ &\times \exp(-ik_\bullet x)]. \end{aligned} \tag{54}$$

*Proof*: In terms of $a^\mu(\mathbf{k})$ and $a^{+\mu}(\mathbf{k})$, (54) is written

$$\begin{aligned} A^\mu(x) = &\int_{k\in C_+} d\mathbf{k} [a^{+\mu}(\mathbf{k}) \exp(ik_\bullet x) + a^\mu(\mathbf{k}) \exp(-ik_\bullet x)] \\ &- d\, \frac{\partial}{\partial x_\mu} \left[ x_0 \int \frac{d\mathbf{k}}{|\mathbf{k}|} \Big[ (k_\rho a^{+\rho}(\mathbf{k}) \exp(ik_\bullet x) \right. \\ &\left. + k_\rho a^\rho(\mathbf{k}) \exp(-ikx) \Big] \right] \\ &- ib\, \frac{\partial}{\partial x_\mu} \int_{k\in C_+} d\mathbf{k} [k_\rho a^{+\rho}(\mathbf{k}) \exp(ikx) \\ &- k_\rho a^\rho(\mathbf{k}) \exp(-ik_\bullet x)]. \end{aligned} \tag{55}$$

Let $f_\mu(x)$ be real functions on $\mathbb{R}^4$ with Fourier transform in $\mathcal{D}_0(\mathbb{R}^4)$. Let us form $\int A^\mu(x) f_\mu(x)\, d^4x$. If we notice

$$\int \exp(ik_\bullet x)\, x_0\, \frac{\partial f_\mu}{\partial x_\mu}\, d^4x = -\frac{\partial}{\partial k_0}\, k^\mu \hat{f}(k),$$

(43) results readily from (55).

To conclude this section, we make some remarks and comments, mainly concerning the choice of the test function space.

It is usual to take these test functions in $\mathcal{S}(\mathbb{R}^4)$ or $\mathcal{D}(\mathbb{R}^4)$. However, the restrictions to $C_+$ of their Fourier transform is not a basic space for distribution theory since their derivatives of any order are not defined at the vertex of the cone. If we want to define the $a^\mu(\mathbf{k})$ [and $a^{+\mu}(\mathbf{k})$] as operator distribution, we must impose regularity conditions at this point. These conditions may be more or less drastic, but we must have it, when $d \neq 0$. Our choice of functions with Fourier transforms in $\mathcal{D}_0(\mathbb{R}^4)$ is merely a matter of mathematical convenience and simplicity. Nonetheless, we must notice that, in the subspace of physical vectors, the cumbersome term in (29) does not contribute and we can use $\mathcal{S}(\mathbb{R}^4)$ [or $\mathcal{D}(\mathbb{R}^4)$] as test functions space.

This conclusion is implicit in formula (54), which shows that our vector potential differs from the vector potential of the Gupta–Bleuler theory only by a gradient. But one cannot think that a convenient gauge transformation brings back all the previous construction to the Gupta–Bleuler case because such a gauge transformation restricted to the one-particle subspace would transform $U_d(a,\Lambda)$ into $U_0(a,\Lambda)$; and, from Corollary 2, this is impossible.

Finally, we add a few words about the discarded case $\lambda = 0$ in (28). All corresponding representations are equivalent. Furthermore, there exist only degenerate sesquilinear forms, so that we shall be unable to uniquely define annihilation operators as adjoints of creation operators. Therefore, these representations are not suitable for a quantization process. It can be applied only when starting with the representation defined on the quotient of $E+F$ by the kernel of the degenerate forms. But this representation turns out to be the direct sum of the unitary representations with zero mass and helicities $=1$, $-1$, and 0, for which the quantization is straightforward.

## 5. CONCLUSION

Let us assume $d \neq \frac{1}{2}$. Then $B(x)$ can be eliminated between (46) and (47) and we get the unique field equation

$$\Box A_\mu(x) = \lambda \partial_\mu (\partial^\rho A_\rho(x)), \quad \lambda = \frac{2d}{2d-1}. \tag{56}$$

It can be derived from the Lagrangian density,

$$\mathcal{L}(x) = -\frac{1}{2}\frac{\partial A^\mu(x)}{\partial x^\nu}\frac{\partial A_\mu(x)}{\partial x_\nu} + \frac{\lambda}{2}(\partial^\rho A_\rho(x))^2. \qquad (56')$$

For $\lambda \neq 1$, $\mathcal{L}(x)$ is nonsingular, so that we can perform, *a priori*, canonical quantization. The corresponding commutation relations are

$$[A^\mu(x), A^\nu(y)]_{x_0=y_0=0} = [\Pi^\mu(x), \Pi^\nu(y)]_{x_0=y_0=0}, \qquad (57a)$$

$$[\Pi^\mu(x), A^\nu(y)]_{x_0=y_0=0} = \frac{1}{i} g^{\mu\nu}\delta(\mathbf{x}-\mathbf{y}), \qquad (57b)$$

with

$$\Pi^\mu(x) = -\frac{\partial A^\mu(x)}{\partial x_0} + \lambda g^{\mu 0}(\partial_\rho A^\rho(x)). \qquad (58)$$

Now, it can be verified that the expression (54) of $A^\mu(x)$, when the constant $b$ is equal to zero, provides a complete solution of (56) and (57), which has been derived from purely group theoretical considerations. Thus, we know that (56) and (57) can be solved, at least if we assume that the distribution operators are defined on suitable test function spaces.

Conversely, if we start from the field equations (56) and the canonical quantization (57), what solutions can we get and what test function spaces are needed for their definition? We first note that the commutation relations (57) imply that $A_\mu(x)$ is a distribution in $\boldsymbol{\lambda}$, depending parametrically on $x_0$. Therefore, it is convenient to write the solutions of (56) in a form well suited to the Cauchy problem. Among various possibilities, we take the following

$$A_\mu(x) = a_\mu(x) + \frac{\lambda}{2}\partial_\mu x_0 T, \qquad (61)$$

where the operator distributions $a_\mu(x)$ and $T(x)$ must satisfy

$$\Box a_\mu(x) = 0, \quad \Box T(x) = 0,$$
$$\frac{\partial T}{\partial x_0} = \frac{1}{1-\lambda}\partial^\mu a_\mu(x). \qquad (62)$$

The first two equations imply that the Fourier transforms, with respect to $\mathbf{x}$, of $a_\mu(x)$ and $T(x)$ are distributions in the dual variable $\mathbf{k}$, multiplied by $\exp(\pm i x_0 |\mathbf{k}|)$. But such a factor is not a $C^\infty$ function in $\mathbf{k}$ at the origin. Therefore, either the Fourier transforms of $a_\mu(x)$ and $T(x)$ are zero order distributions, or we must impose restrictions on the behavior at the origin of the admissible test functions. It is just what we have done previously in defining $A_\mu(x)$ as a distribution on test functions with Fourier transforms in $\mathcal{D}_0(\mathbb{R}^4)$. If we adopt such a point of view, we can write, since any contribution from the origin is vanishing,

$$a_\mu(x) = \int_{k\in C_+} d\mathbf{k}[a^{+\mu}(\mathbf{k})\exp(ik_\cdot x) + a^\mu(\mathbf{k})\exp(-ik_\cdot x)]$$

and, using the canonical commutation relations, we get the same commutation relations as for the $a^\mu(\mathbf{k})$ and $a^{+\mu}(\mathbf{k})$ in Sec. 4.

Thus, it can be thought that there is a full equivalence between our group theoretical construction completed by the Fock process and the Lagrangian formalism completed by the canonical commutation relations. Nevertheless, this equivalence takes place only on a formal level insofar as, from the Lagrangian point of view, the mathematical problem is not well defined from the beginning. It is only at the end of the calculation process that this problem can be precisely settled. On the contrary, in the group theoretical approach, there are no ambiguities at all and the quantization process is founded on a firm ground. Furthermore, whereas in the Lagrangian version, we cannot treat directly the case $d=\frac{1}{2}$ $(\lambda=\infty)$, it does not matter from the group theoretical point of view. To tell the truth, one can raise the objection that (46) and (47) are derivable from a Lagrangian formalism for any finite $d$. But then, the ghost field $B(x)$ occurs explicitly in the Lagrangian, which is singular, (see Ref. 4) the canonical quantization does not make sense, and the mathematical problem is even less determined than when starting with (56′). Once more, it is the group theoretical treatment which brings about consistency and closeness.

It must be pointed out that group theoretical treatment leads to a kind of representations called "undecomposable representations" in Ref. 5, where their appearance in theoretical physics has been strongly related to the existence of zero mass particles. From this point of view, our work shows that the set of undecomposable representations is not restricted to the family of representations induced by undecomposable representations of the little group. It seems to us that this opens a new and very large domain of research.

Finally, it is noteworthy that the field theories we built above are such that $\Box\partial^\mu A_\mu(x)=0$. It has been shown in Ref. 6 that this gauge condition is the simplest linear gauge condition which is conformal invariant. It may be asked whether all electromagnetic field theories verifying this gauge condition have been obtained in the present paper. We hope to soon give an answer to this question.


## ACKNOWLEDGMENTS

We are indebted to Dr. J.C. Houard and Dr. M. Irac for many fruitful discussions, especially with respect to the Lagrangian point of view.



[1] F. Strocchi and A.S. Wightman, J. Math. Phys. 15, 2198 (1974).

[2] G. Rideau, Lett. Math. Phys. 1, 17 (1975).

[3] F. Trèves, *Topological Vector Spaces, Distributions and Kernels* (Academic, New York, 1967).

[4] S.P. Tomczak and K. Haller, Nuovo Cimento B 8, 1 (1972).

[5] Dj. Parvisi, "Sur les représentations de certains groupes relativistes," Thèse de Doctorat d'Etat, Dijon, 1975.

[6] F. Bayen and M. Flato, J. Math. Phys. 17, 1112–4 (1976).

# Multiplicity-free 6-j symbols and Weyl coefficients of U($n$): Explicit evaluation

M. K. F. Wong

*Fairfield University, Fairfield, Connecticut 06430*
(Received 14 June 1977)

An explicit expression has been obtained for all known multiplicity-free 6-$j$ symbols of U($n$), i.e., 6-$j$ symbols of the following three types, where anyone of its columns consists of (1) two totally symmetric representations, (2) one totally symmetric and one conjugate to the totally symmetric, and (3) two conjugate to the totally symmetric. The symmetry properties of the multiplicity-free 6-$j$ symbols of U($n$) under permutation of columns, inversion of any two columns, and conjugation are given. Some general theorems concerning the multiplicity-free 6-$j$ symbols of U($n$) or more precisely, the multiplicity-free 6-$j$ symbols of the "SU($n$) type" have been obtained. Since the Weyl coefficients of U($n$) are basically 6-$j$ symbols of U($n-1$), we also conclude that the Weyl coefficients of U($n$) have been explicitly obtained. This result implies that the $d$ function of U($n$) can be completely and explicitly written down in terms of the Weyl coefficients.

## 1. INTRODUCTION

The problem of classifying multiplicity-free 6-$j$ symbols of U($n$) is closely connected with the classification of the multiplicity-free 3-$j$ symbols of U($n$). In the latter case it is known that the 3-$j$ symbol is multiplicity-free if one of the three irreducible representations is totally symmetric, i.e., of the type ($p$,0), or conjugate to the totally symmetric representation, i.e., of the type ($\dot{p}$,0). Extending this result from the 3-$j$ symbol to the 6-$j$ symbol, we can say that a sufficient condition for a 6-$j$ symbol of U($n$) to be multiplicity-free is the following: (1) Any one of its columns consists of two totally symmetric representations, i.e., of the type ($p,\dot{0}$), ($q,\dot{0}$), (2) one is of the type ($p,\dot{0}$), one is of the type ($\dot{q}$,0), (3) two are of the type ($\dot{p}$,0), ($\dot{q}$,0). This is a sufficient condition, but need not be a necessary condition because we do not know whether a 3-$j$ symbol can be accidentally multiplicity free even if none of the three representations are of the type ($p,\dot{0}$) or ($\dot{p}$,0). In other words, it may be possible that the complement to the null space of the tensor operator is one dimensional. We do not know of any such cases, but we do not have any proof that they do not exist. We suspect that the condition is both necessary and sufficient, but we shall leave this point to be clarified in the future. At any rate, for all practical purposes, the three cases mentioned above are the only known cases for the multiplicity-free 6-$j$ symbols of U($n$). In this paper we shall give an explicit expression for all the three cases mentioned above. In fact, we find that they are all equivalent to each other, differing from each other by at most a phase factor.

Besides its intrinsic value, the 6-$j$ symbol of U($n$) is also of interest in connection with the finite transformation $d$ matrix of U($n$), since it is connected with the Weyl coefficient of U($n+1$), as Holman[1] and Wong[2] have pointed out. However, Holman has only shown that the Weyl coefficient, as a 6-$j$ symbol, can be written as a sum over the product of four 3-$j$ symbols, but did not succeed in explicity evaluating the 6-$j$ symbol, whereas we wish to give in this paper an explicit evaluation of the 6-$j$ symbol, and not merely as a product of four 3-$j$ symbols.

There are two ways to attempt an explicit evaluation of the 6-$j$ symbol. One is to simplify the expression where the 6-$j$ symbol is written as a sum over the product of four 3-$j$ symbols. This is the method used by Racah[3] to obtain the Racah coefficient of U(2). However, this method is extremely laborious, and without the genius of Racah, we can hardly expect to make any progress on this line. The other way is to obtain the 6-$j$ symbol from a 9-$j$ symbol, by putting one of the terms in the 9-$j$ symbol equal to zero. It is along this line that we shall proceed to obtain an explicit evaluation of the 6-$j$ symbol of U($n$).

There have been two formulas given connecting a 9-$j$ symbol in U($n$) with an isoscalar factor in U($n+1$). One was given by Ališauskas, Jucys, and Jucys,[4,5] where five terms in the 9-$j$ symbol are totally symmetric, i.e., ($p,\dot{0}$), ($q,\dot{0}$), ($p-q,\dot{0}$), ($r,\dot{0}$), and ($r',\dot{0}$). The other was given, independently, by Wong[2] (henceforth referred to as I), where two terms are totally symmetric, i.e., ($W_{n},\dot{0}$) and ($W'_{n},\dot{0}$) while three terms are conjugate to the totally symmetric representation, i.e., ($\dot{p}$,0), ($\dot{q}$,0), and ($\dot{p-q}$,0). By putting $q=p$, these two expressions give us immediately an explicit expression for the 6-$j$ symbol of U($n$) for cases (1) and (2) above. The third case can be easily related to the first case by taking conjugation on all six representations, as we shall show in Secs. 2 and 3.

Thus the statement we are making is actually very simple. Essentially we are saying that the multiplicity-free 6-$j$ symbols of U($n$) can be explicitly evaluated from the known multiplicity-free isoscalar factor of U($n+1$), where, in particular, one can put $p=q$ and $m_{nn}=m'_{nn}=0$. We shall call this particular isoscalar factor "mfppssif" for "multiplicity-free $p$, $p$, semistretched isoscalar factor."

The multiplicity-free 3-$j$ symbols of U($n$) have been evaluated by many authors.[4,6,7] In particular, Chacón *et al.*[6] have obtained the 3-$j$ symbol of U($n$) as a sum over $n-1$ variables, while Ališauskas *et al.*[4] have obtained, in addition to the result above, another expession, which, in our present case, reduces to a sum over $n-2$ variables.

Thus in this paper we report the following four results.

1. We have obtained explicit expressions for all known multiplicity-free 6-$j$ symbols of U($n$), i.e., 6-$j$ symbols of the

 0022-2488/78/1907-1635$1.00  

three types mentioned above. 2. We have obtained symmetries of the multiplicity-free 6-$j$ symbol of $\mathrm{U}(n)$ under permutation of columns and inversion of any two columns, and also under conjugation. 3. We have obtained some general theorems concerning the multiplicity-free 6-$j$ symbols of the "$\mathrm{SU}(n)$ type", where there is a zero at the end of each irreducible representation. 4. We have obtained explicit expressions for all the Weyl coefficients of $\mathrm{U}(n)$, and therefore explicit expressions for all the finite transformation $d$ matrices of $\mathrm{U}(n)$.

This paper contains six sections. In Sec. 2 we give the definition of the multiplicity-free 6-$j$ symbols of $\mathrm{U}(n)$ and their symmetry properties under permutation of columns and inversion of any two columns, and also under conjugation. In Sec. 3 we give an explicit expression for the three types of multiplicity-free 6-$j$ symbols of $\mathrm{U}(n)$. In Sec. 4 we discuss the 6-$j$ symbol of $\mathrm{U}(2)$. Since all 6-$j$ symbols in $\mathrm{U}(2)$ are multiplicity-free, our method completely solves the problem of the 6-$j$ symbol of $\mathrm{U}(2)$. We show how our result agrees with Racah.[3] In Sec. 5 we state and prove some general theorems concerning the multiplicity-free 6-$j$ symbols of $\mathrm{U}(n)$. All these theorems deal with 6-$j$ symbols when there is a zero at the end of each irreducible representation. We call these 6-$j$ symbols of the "$\mathrm{SU}(n)$ type". In Sec. 6 we discuss the Weyl coefficients of $\mathrm{U}(n)$ and offer some suggestions as to future topics of research.

## 2. DEFINITION OF THE MULTIPLICITY-FREE 6-*j* SYMBOL OF U(*n*) AND ITS SYMMETRY PROPERTIES

We shall use the following notation for the different coupling coefficients. A Clebsch–Gordan coefficient is denoted by

$$C\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}.$$

A 3-$j$ symbol is denoted by a paranthesis only; thus,

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$$

is a 3-$j$ symbol. A 6-$j$ symbol is denoted by a brace:

$$\begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & J & j_{23} \end{Bmatrix}.$$

We define the multiplicity-free 6-$j$ symbol of $\mathrm{U}(n)$ by first recalling the definition of the 6-$j$ symbol of $\mathrm{U}(2)$. A 6-$j$ symbol is the transformation coefficient through the following two different ways of coupling:

$$\begin{aligned} &1.\ j_1 \text{ and } j_2 \to j_{12},\ j_{12} \text{ and } j_3 \to J, \\ &2.\ j_2 \text{ and } j_3 \to j_{23},\ j_{23} \text{ and } j_1 \to J. \end{aligned} \tag{2.1}$$

As is well known,[8] the 6-$j$ symbol can be written as a product of four 3-$j$ symbols, i.e.,

$$\begin{aligned} \begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & J & j_{23} \end{Bmatrix} &= [(\dim j_{12} \dim j_{23})]^{-1/2}(-1)^{j_1+j_2+j_3+J} \\ &\times \sum_{m_1, m_2} C\begin{pmatrix} j_1 & j_2 & j_{12} \\ m_1 & m_2 & m_1+m_2 \end{pmatrix} \\ &\times C\begin{pmatrix} j_{12} & j_3 & J \\ m_1+m_2 & M-m_1-m_2 & M \end{pmatrix} \\ &\times C\begin{pmatrix} j_2 & j_3 & j_{23} \\ m_2 & M-m_1-m_2 & M-m_1 \end{pmatrix} \\ &\times C\begin{pmatrix} j_1 & j_{23} & J \\ m_1 & M-m_1 & M \end{pmatrix} \\ &= \sum_{\text{all } 6m} \begin{pmatrix} j_1 & j_2 & j_{12}^* \\ m_1 & m_2 & -m_{12} \end{pmatrix}\begin{pmatrix} j_1^* & J & j_{23}^* \\ -m_1 & M & -m_{23} \end{pmatrix} \\ &\times \begin{pmatrix} j_3^* & j_2^* & j_{23} \\ -m_3 & -m_2 & m_{23} \end{pmatrix} \\ &\times \begin{pmatrix} j_3 & J^* & j_{12} \\ m_3 & -M & m_{12} \end{pmatrix} \\ &\times (-1)^{j_1+j_2+j_3+J+j_{12}+j_{23}+m_3+m_{23}+M}. \end{aligned} \tag{2.2}$$

Now we generalize the definition of the 6-$j$ symbol from $\mathrm{U}(2)$ to $\mathrm{U}(n)$, using the phase convention given in I. Thus, for example,

$$(-1)^{j_1} \to (-1)^{\epsilon_n(m_{1n}^{(1)} - m_{nn}^{(1)})},$$

$$(-1)^{m_1} \to (-1)^{\sum_{i=1}^{n-1} \frac{1}{2} z_i^{(1)}},$$

where

$$z_i = (i+1)\sum_{j=1}^{i} m_{ji} - i\sum_{j=1}^{i+1} m_{j,i+1} \tag{2.3}$$

and $\epsilon_n$ is defined in I. More explicitly,

$$\begin{aligned} \epsilon_n &= 1/2 && \text{for } n = 2+4k \\ &= 1 && \text{for } n = 3+4k, \\ &= 3/2 && \text{for } n = 4+4k, \\ &= 0 && \text{for } n = 5+4k, \quad k = 0,1,2,\cdots. \end{aligned} \tag{2.4}$$

In Eq. (2.2) an asterisk denotes the conjugation operation.[9] Thus

$$m^*_{ij} = m_{1n} - m_{j-i+1,j}. \tag{2.5}$$

The symmetry properties of the 6-$j$ symbol of an arbitrary group have been discussed by Derome and Sharp.[10–12] In connection with Derome's work, we would like to make two observations. The first observation concerns the very existence of the 3-$j$ symbol itself such that its absolute value is invariant under every permutation of the $j$'s and of the corresponding $m$'s. It has been shown by Derome[11] that in general the 3-$j$ symbol cannot be chosen in symmetric form if $j_1, j_2, j^*_3$ are equivalent representations, except in the case of $\mathrm{SU}(3)$.[12] However, in the situation we are dealing with, i.e., one of the three representations being of the form $(p,\dot{0})$ or $(\dot{p},0)$, this case cannot arise, since if two are of the form $(p,\dot{0})$, $(p,\dot{0})$, the third must be $(p,p,\dot{0})$, which is conjugate to

$(p,0)$ only for U(3), for which the symmetric 3-$j$ symbol can be defined. Similarly, for two $(\dot{p},0)$, $(\dot{p},0)$, the third must be $(\dot{p},0,0)$, which is conjugate to $(p,p,0)$ only for U(3). Thus the existence of the multiplicity-free 3-$j$ symbol of U($n$) is assured. The second observation concerns the possibility of extending the phase conventions in U(2) to U($n$). It has been shown by Wigner[13] that if one is dealing with a simply reducible (SR) group, then the symmetry properties of the 3-$j$ symbols of that group are essentially the same as R(3), and the phase conventions of U(2) and R(3) can be extended to that group. The definition of a simply reducible group is (1) it is ambivalent, i.e., every element of the group is in the same class as its inverse and (2) it is multiplicity free. Now in the case of U($n$), $n\neq 2$, we are dealing with the multiplicity-free cases only, so the second condition is automatically satisfied. The first condition can be stated alternatively as: The conjugate representation is equivalent to the original representation. Now in the phase convention we have chosen in I, the conjugate representation has the same phase as the original representation. Therefore, in the present situation, the symmetry properties of the multiplicity-free 3-$j$ symbols of U($n$) with regard to the phase convention are essentially the same as a simply reducible group, such as U(2). We can therefore extend the phase convention from U(2) to U($n$), for the multiplicity-free case.

We now state the symmetry properties of the multiplicity-free 6-$j$ symbol of U($n$).

1. Interchange columns 1 and 2.

$j_1 \to j_2, j_3 \to J^*, J \to j_3^*, j_{23} \longleftrightarrow j_{23}^*$

We have

$$\begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & J & j_{23} \end{Bmatrix} = \begin{Bmatrix} j_2 & j_1 & j_{12} \\ J^* & j_3^* & j_{23}^* \end{Bmatrix}. \tag{2.6}$$

2. Interchange columns 1 and 3.

$j_1 \to j_{12}^*, j_{12} \to j_1^*, j_3 \to j_{23}^*,$

$j_{23} \to j_3^*, J \longleftrightarrow J^*.$

We have

$$\begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & J & j_{23} \end{Bmatrix} = \begin{Bmatrix} j_{12}^* & j_2 & j_1^* \\ j_{23}^* & J & j_3^* \end{Bmatrix}. \tag{2.7}$$

3. Inversion of columns 1 and 2.

$j_1 \longleftrightarrow j_3, \quad j_2 \to J^*, \quad J \to j_2^*, \quad j_{12} \longleftrightarrow j_{12}^*, \quad j_{23} \longleftrightarrow j_{23}^*.$

We have

$$\begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & J & j_{23} \end{Bmatrix} = \begin{Bmatrix} j_3 & J^* & j_{12}^* \\ j_1 & j_2^* & j_{23}^* \end{Bmatrix}. \tag{2.8}$$

4. Inversion of columns 1 and 3.

$j_1 \to j_3^*, \quad j_3 \to j_1^*, \quad j_{12} \to j_{23}^*,$

$j_{23} \to j_{12}^*, j_2 \longleftrightarrow j_2^*, J \longleftrightarrow J^*.$

We have

$$\begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & J & j_{23} \end{Bmatrix} = \begin{Bmatrix} j_3^* & j_2^* & j_{23}^* \\ j_1^* & J^* & j_{12}^* \end{Bmatrix}. \tag{2.9}$$

Note that these symmetry properties are not entirely the same as Derome and Sharp,[10] or Resnikoff[14] for SU(3). This is because, basically, there is a difference in the definition of the 6-$j$ symbol between various authors. Our definition is taken from Edmonds,[8] which agrees with the conventional definition of the 6-$j$ symbol for U(2) or R(3).

5. Symmetry of $mf$ 6-$j$ symbol of U($n$) under conjugation.

Under conjugation, we have

$$\begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & J & j_{23} \end{Bmatrix} = \begin{Bmatrix} j_1^* & j_2^* & j_{12}^* \\ j_3^* & J^* & j_{23}^* \end{Bmatrix}. \tag{2.10}$$

Combining (2.7) and (2.9) with (2.10), we can also write

$$\begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & J & j_{23} \end{Bmatrix} = \begin{Bmatrix} j_3 & j_2 & j_{23} \\ j_1 & J & j_{12} \end{Bmatrix} \tag{2.11}$$

$$= \begin{Bmatrix} j_{12} & j_2^* & j_1 \\ j_{23} & J & j_3 \end{Bmatrix} \tag{2.12}$$

etc.

Equation (2.10) is particularly useful in obtaining type 3 6-$j$ symbols of U($n$) from type 1, as we shall show in the next section.

## 3. EXPLICIT EVALUATION OF THE THREE TYPES OF MULTIPLICITY-FREE 6-j SYMBOLS OF U(n)

Let us start with the explicit evaluation of the multiplicity-free 6-$j$ symbol of U($n$) of the first type. Without loss of generality, we can choose $j_1$ and $j_3$ to be totally symmetric, and $j_2$ to be of the form $[m_{1,n-1},\dots,m_{n-1,n-1},0]$. Thus the 6-$j$ symbol of type 1 is of the form

$$\begin{Bmatrix} [W_{n+1},\dot{0}] & [m_{1,n-1},\dots,m_{n-1,n-1},0] & [m'_{1n},\dots,m'_{nn}] \\ [p,\dot{0}] & [m_{1,n+1},\dots,m_{n,n+1}] & [m_{1n},\dots,m_{nn}] \end{Bmatrix}$$

with the constraints

$$p=m_{1n}+\dots+m_{nn}-m_{1,n-1}-\dots-m_{n-1,n-1}=m_{1,n+1}+\dots+m_{n,n+1}-m'_{1n}-\dots-m'_{nn}, \tag{3.1}$$

$$W_{n+1}=m_{1,n+1}+\dots+m_{n,n+1}-m_{1n}-\dots-m_{nn}=m'_{1n}+\dots+m'_{nn}-m_{1,n-1}-\dots-m_{n-1,n-1}. \tag{3.2}$$

Equation (3.1) and (3.2) follow necessarily from the formula for the Wigner coefficients of $U(n)$ for the totally symmetric representation, since otherwise the Wigner coefficient is equal to zero.

Now Ališauskas, Jucys, and Jucys[4] have given a formula relating a doubly stretched 9-$j$ symbol in $U(n-1)$ with an isoscalar factor of a totally symmetric tensor operator in $U(n)$. We can obtain their result by using the technique developed in I for the derivation of Eq. (1.17) in I, except that the states $(\dot p,0)$, $(\dot q,0)$, $(p\dot - q,0)$ are changed to the totally symmetric ones $(p,\dot 0)$, $(q,\dot 0)$, $(p-q,\dot 0)$ respectively. The result is

$$X\begin{vmatrix} [q,\dot 0]_{n-1} & [m'_{1,n-1},\dots,m'_{n-1,n-1}] & [m_{1,n-1},\dots,m_{n-1,n-1}] \\ [p-q,\dot 0]_{n-1} & [W'_n,\dot 0]_{n-1} & [W_n,\dot 0]_{n-1} \\ [p,\dot 0]_{n-1} & [m'_{1n},\dots,m'_{n-1,n}] & [m_{1n},\dots,m_{n-1,n}] \end{vmatrix}$$
$$=\{\dim[W_n,\dot 0]_{n-1}\dim[p,\dot 0]_{n-1}\dim(m'_{1,n-1},\dots,m'_{n-1,n-1})\dim(m_{1,n-1},\dots,m_{n-1,n-1})\}^{-1/2}$$
$$\times\left\{\frac{\mathcal{M}(m_{1n},\dots,m_{n-1,n})\mathcal{M}(m'_{1,n-1},\dots,m'_{n-1,n-1})}{\mathcal{M}(m'_{1n},\dots,m'_{n-1,n})\mathcal{M}(m_{1,n-1},\dots,m_{n-1,n-1})}\,\frac{W'_n!(p-q)!q!}{W_n!\,p!}\right\}^{1/2}$$
$$\times\left\langle\begin{matrix} m_{1n},\dots,m_{n-1,n},0 \\ m_{1,n-1},\dots,m_{n-1,n-1}\end{matrix}\left\|\begin{matrix} p,0 \\ q,0\end{matrix}\right\|\begin{matrix} m',\dots,m'_{n-1,n},0 \\ m'_{1,n-1},\dots,m_{n-1,n-1}\end{matrix}\right\rangle. \quad (3.3)$$

Equation (3.3) differs from (B1) of Ališauskas, Jucys, and Jucys by only a dimensional factor, which is due to difference of definition of the 9-$j$ symbol of $U(n-1)$.

By setting $q=p$ in Eq. (3.3), we obtain

$$\begin{Bmatrix} [W_{n+1},\dot 0] & [m_{1,n-1},\dots,m_{n-1,n-1},0] & [m'_{1n},\dots,m'_{nn}] \\ [p,\dot 0] & [m_{1,n+1},\dots,m_{n,n+1}] & [m_{1n},\dots,m_{nn}]\end{Bmatrix}$$
$$=(-1)^{-\epsilon_n(m_{1,n+1}-m_{n,n+1})+p+W_{n+1}+m_{1,n-1}}\{\dim(m_{1,n-1},\dots,m_{n-1,n-1},0)\dim(m_{1n},\dots,m_{nn})\}^{-1/2}$$
$$\times\left\{\frac{\mathcal{M}(m_{1,n+1},\dots,m_{n,n+1})\mathcal{M}(m_{1,n-1},\dots,m_{n-1,n-1},0)}{\mathcal{M}(m'_{1n},\dots,m'_{nn}\mathcal{M}(m_{1n},\dots,m_{nn})}\right\}^{1/2}$$
$$\times\left\langle\begin{matrix} m_{1,n+1},\dots,m_{n,n+1},0 \\ m_{1n},\dots,m_{nn}\end{matrix}\left\|\begin{matrix} p,\dot 0 \\ p,\dot 0\end{matrix}\right\|\begin{matrix} m'_{1n},\dots,m'_{nn},0 \\ m_{1,n-1},\dots,m_{n-1,n-1},0\end{matrix}\right\rangle. \quad (3.4)$$

Now the isoscalar factor "mfppssif" in (3.4) has been evaluated by many authors, and we shall quote only two results, one by Chacón, Ciftan, and Biedenharn[6] (also obtained by Ališauskas, Jucys, and Jucys[4]), and the other by Ališauskas *et al.*[4] Both results can be expressed in terms of $S_{nm}$ defined by Chacón *et al.*[6] The first result is expressed as a sum over $n-1$ variables. The second over $n-2$ variables. The first one is

$$\left\langle\begin{matrix}[m]_n \\ [m]_{n-1}\end{matrix}\left\|\begin{matrix}p,\dot 0 \\ p,\dot 0\end{matrix}\right\|\begin{matrix}[m']_n \\ [m']_{n-1}\end{matrix}\right\rangle=\delta_{\Sigma_i m_{in},p+\Sigma_i m'_{in}}\delta_{\Sigma_i m_{i,n-1},p+\Sigma_i m'_{i,n-1}}$$
$$\times\frac{S_{nn}([m]_n,[m]_n)S_{n,n-1}([m']_n,[m']_{n-1})S_{n-1,n-1}([m]_{n-1},[m']_{n-1})}{S_{nn}([m]_n,[m']_n)S_{n,n-1}([m]_n,[m]_{n-1})}$$
$$\times S_{n-1,n-1}([m']_{n-1},[m']_{n-1})\sum_{[r]_{n-1}}(-1)^{r_{1,n-1}+\cdots+r_{n-1,n-1}}$$
$$\times\left\{\frac{S_{n,n-1}([m]_n,[m']_{n-1}+[r]_{n-1})S_{n-1,n-1}([m']_{n-1}+[r]_{n-1},[m']_{n-1}+[r]_{n-1})}{S_{n,n-1}([m']_n,[m']_{n-1}+[r]_{n-1})S_{n-1,n-1}([m]_{n-1},[m']_{n-1}+[r]_{n-1})}\right.$$
$$\left.\times\frac{1}{S_{n-1,n-1}([m']_{n-1}+[r]_{n-1},[m']_{n-1})}\right\}^2, \quad (3.5)$$

where

$$S_{nm}(h_1,\dots,h_n;q_1,\dots,q_m)=\left(\frac{\Pi_{k=1}^m\Pi_{s=1}^k(h_s-q_k+k-s)!}{\Pi_{k=1}^{n-1}\Pi_{s=k+1}^n(q_k-h_s+s-k-1)!}\right)^{1/2}.$$

The second formula is

$$\left\langle \begin{matrix}[m]_n \\ [m]_{n-1}\end{matrix} \left\| \begin{matrix} p,\dot{0} \\ p,\dot{0}\end{matrix} \right\| \begin{matrix}[m']_n \\ [m']_{n-1}\end{matrix}\right\rangle = \delta_{\Sigma_i m_{in},p+\Sigma_i m'_{in}} \delta_{\Sigma_i m_{i,n-1},p+\Sigma_i m'_{i,n-1}}$$

$$\times \frac{S_{nn}([m]_n,[m]_n) S_{n-1,n-1}([m']_{n-1},[m']_{n-1}) S_{n,n-1}([m]_n,[m]_{n-1})}{S_{n,n-1}([m']_n,[m']_{n-1}) S_{n-1,n-1}([m]_{n-1},[m']_{n-1})}$$

$$\times S_{nn}([m]_n,[m']_n) \sum_{\substack{r_{jn} \\ j=2}}^{n} (-1)^{\phi_i}$$

$$\times \left( \frac{S_{nn}([r]_n,[r]_n) S_{n,n-1}([r]_n,[m']_{n-1})}{S_{nn}([m]_n,[r]_n) S_{nn}([r]_n,[m']_n) S_{n,n-1}([r]_n,[m]_{n-1})} \right)^2, \tag{3.6}$$

where

$$\phi_i = \sum_{j=2}^{n} m_{jn} - \sum_{j=2}^{n} r_{jn}.$$

When applying the formula for $S_{nm}$ on $[r]_n$, it is understood that, since $r_{1n}$ does not exist, all factors containing $r_{1n}$ should be automatically removed.

At first sight, Eq. (3.6) seems to contain a sum over $n-1$ variables. However, for the semistretched case, $m_{nn}=m'_{nn}$, it can be seen that $r_{nn}=m_{nn}=m'_{nn}$. Thus the sum over $r_{nn}$ can be trivially done. In our present case, since $m_{nn}=m'_{nn}=0$, $r_{nn}=0$. Thus for the "mfppssif" of U($n$), (3.6) is a sum over $n-2$ variables only.

The multiplicity-free 6-$j$ symbol of U($n$) of type 2 is obtained from Eq. (1.17) of I, by putting $q=p$. The result can be expressed in terms of the 6-$j$ symbol of type 1, since they differ only by a phase factor:

$$\begin{Bmatrix} [W_{n+1},\dot{0}] & [m_{1,n-1},\dots,m_{n-1,n-1},0] & [m'_{1n},\dots,m'_{nn}] \\ [\dot{p},0] & [m_{1,n+1},\dots,m_{n,n+1}] & [m_{1n},\dots,m_{nn}] \end{Bmatrix}$$
$$=(-1)^{-y(m_{1n},\dots,m_{nn})-y(m_{1,n+1},\dots,m_{n,n+1})} \begin{Bmatrix} [W_{n+1},\dot{0}] & [m_{1,n-1},\dots,m_{n-1,n-1},0] & [m'_{1n},\dots,m'_{nn}] \\ [p,\dot{0}] & [m_{1,n+1},\dots,m_{n,n+1}] & [m_{1n},\dots,m_{nn}] \end{Bmatrix}, \tag{3.7}$$

where

$$y([m]_n) = \sum_{i=2}^{n-1} m_{in} \quad \text{for } n=4,6,8,\dots,2k$$
$$=0 \quad \text{for } n=2,3,5,\dots,2k+1.$$

The multiplicity-free 6-$j$ symbol of type 3 can be obtained from that of type 1 by using Eq. (2.10). We have

$$\begin{Bmatrix} [\dot{W}_{n+1},0] & [m_{1,n-1},\dots,m_{n-1,n-1},0]^* & [m'_{1n},\dots,m'_{nn}]^* \\ [\dot{p},0] & [m_{1,n+1},\dots,m_{n,n+1}]^* & [m_{1n},\dots,m_{nn}]^* \end{Bmatrix}$$
$$=\begin{Bmatrix} [W_{n+1},\dot{0}] & [m_{1,n-1},\dots,m_{n-1,n-1},0] & [m'_{1n},\dots,m'_{nn}] \\ [p,\dot{0}] & [m_{1,n+1},\dots,m_{n,n+1}] & [m_{1n},\dots,m_{nn}] \end{Bmatrix}. \tag{3.8}$$

Thus we have obtained the multiplicity-free 6-$j$ symbol of U($n$) for all three types where the totally symmetric or conjugate to the totally symmetric representations occur in the first column. If, however, they occur in the second or third column, then we can use Eqs. (2.6) and (2.7) to change them back to the first column. Finally, if in type 2, the conjugationi is inverted, i.e., first row first column is of the form $(\dot{p},0)$ and second row first column is of the form $(W_{n+1},\dot{0})$, then we can use either (2.8), (2.11), or (2.10) to relate it to (3.7).

## 4. 6-$j$ symbols of U(2)

Since in U(2), all 3-$j$ symbols are multiplicity free, we conclude that all the 6-$j$ symbols of U(2) can be obtained from Eq. (3.4). The result, of course, must agree with Racah's result, and we proceed to demonstrate this fact.

If we write the 6-$j$ symbol of U(2) in terms of the angular momentum label $j$, we have the following relation:

$$\begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & J & j_{23} \end{Bmatrix} = \begin{Bmatrix} [W_3,O] & [m_{11},O] & [m'_{12},m'_{22}] \\ [p,O] & [m_{13},m_{23}] & [m_{12},m_{22}] \end{Bmatrix}; \tag{4.1}$$

then

$$j_1=\tfrac{1}{2}W_3, j_{12}=\tfrac{1}{2}(m'_{12}-m'_{22}),$$
$$j_2=\tfrac{1}{2}m_{11}, j_{23}=\tfrac{1}{2}(m_{12}-m_{22}), \tag{4.2}$$
$$j_3=\tfrac{1}{2}p, J=\tfrac{1}{2}(m_{13}-m_{23}).$$

Now if we evaluate the 6-$j$ symbol of U(2) according to (3.4) and (3.6), we find that it reduces to a sum over a single variable. The resulting formula is identical to the "doubly stretched" 9-$j$ symbol of U(2), found, e.g., in Eq. (4) of Sharp,[15] with one term equal to zero (i.e., $d=0$ in Sharp's formula). Since a 9-$j$ symbol with one term equal to zero is clearly a 6-$j$ symbol, this immediately gives us the required result. More explicitly, after putting $d=0$ in Sharp's Eq. (4), we can write it as a hypergeometric function ${}_4F_3(W,1)$ equivalent to Minton's[16] Eq. (7). Then ${}_4F_3(W,1)$ is connected with Racah's expression through Minton's Eq. (3a). Thus our method gives the complete result for the 6-$j$ symbol of U(2).

The multiplicity-free 6-$j$ symbols of other U($n$) groups can be evaluated explicitly using (3.4) and (3.5) or (3.6). However, instead of writing down those expressions explicitly for the other groups, which one can obviously do, we shall show that there are some general theorems concerning the multiplicity-free 6-$j$ symbols of U($n$). This we do in the next section.

## 5. SOME GENERAL THEOREMS CONCERNING MULTIPLICITY-FREE 6-j SYMBOLS OF U(n)

In this section we would like to state and prove some general theorems concerning the multiplicity-free 6-$j$ symbols of U($n$). These theorems are all connected with irreducible representations with one or more zeros at the end. Therefore they can all be called 6-$j$ symbols of the "SU($n$) type," since for an irreducible representation of SU($n$), $m_{nn}=0$.

*Theorem* 1: A multiplicity-free 6-$j$ symbol of the "SU($n$) type" is equivalent to a multiplicity-free 6-$j$ symbol of U($n-1$). Mathematically, this theorem says:

$$\begin{aligned}
&\begin{Bmatrix} [W_n,\dot{0}] & [m'_{1,n-1},\ldots,m'_{n-1,n-1},0] & [m'_{1n},\ldots,m'_{n-1,n},0] \\ [p,\dot{0}] & [m_{1n},\ldots,m_{n-1,n},0] & [m_{1,n-1},\ldots,m_{n-1,n-1},0] \end{Bmatrix}_{\mathrm{U}(n)} \\
&=(-1)^{\epsilon_n(W_n+p+m_{1n}+m'_{1,n-1})-\epsilon_{n-1}(W_n+p+m_{1n}-m_{n-1,n}+m'_{1,n-1}-m'_{n-1,n-1})} \\
&\times\left(\frac{\dim[m'_{1n},\ldots,m'_{n-1,n}]_{\mathrm{U}(n)}\dim[m_{1,n-1},\ldots,m_{n-1,n-1}]_{\mathrm{U}(n-1)}}{\dim[m'_{1n},\ldots,m'_{n-1,n},0]_{\mathrm{U}(n)}\dim[m_{1,n-1},\ldots,m_{n-1,n-1},0]_{\mathrm{U}(n)}}\right)^{1/2} \\
&\times\begin{Bmatrix} [W_n,\dot{0}] & [m'_{1,n-1},\ldots,m'_{n-1,n-1}] & [m'_{1n},\ldots,m'_{n-1,n}] \\ [p,\dot{0}] & [m_{1n},\ldots,m_{n-1,n}] & [m_{1,n-1},\ldots,m_{n-1,n-1}] \end{Bmatrix}_{\mathrm{U}(n-1)}
\end{aligned} \tag{5.1}$$

where $\epsilon_n$ has been defined in Sec. 2.

Equation (5.1) is quite simple and easy to remember. Apart from the phase factor and dimensional factor, which are connected with the definition of the 6-$j$ symbol, Eq. (5.1) says that in order to change from a multiplicity-free 6-$j$ symbol in U($n$), with a zero at the end of each irreducible representation, to a multiplicity-free 6-$j$ symbol in U($n-1$), all one has to do is "drop off the zero at the end of each irreducible representation." We shall give two proofs for Theorem 1.

*Proof* 1: By direct calculation using (3.6)! This was actually the way we arrived at the theorem. Note that the actual calculation using (3.6) is not trivial, since the $S_{nm}$ functions with or without a zero at the end are not the same. However, it turns out that after multiplying all the eleven terms of $S_{nm}$ together in (3.6), the extra factors all cancel out, and the theorem is proved.

*Proof* 2: After completing the proof of Theorem 1 through the (laborious!) method above, we realized that there is a more direct, and, we think, elegant, proof. This is through the isoscalar factor (mfppssif) between U($n+1$) and U($n$). Using (3.4) again, we find that (5.1) is equivalent to the following relation between the mfppssif of U($n+1$) and U($n$).

$$\begin{aligned}
&\left\langle \begin{matrix} [m_{1n},\ldots,m_{n-1,n},0,0] \\ [m_{1,n-1},\ldots,m_{n-1,n-1},0] \end{matrix} \middle\| \begin{matrix} p,\dot{0} \\ p,\dot{0} \end{matrix} \middle\| \begin{matrix} [m'_{1n},\ldots,m'_{n-1,n},0,0] \\ [m'_{1,n-1},\ldots,m'_{n-1,n-1},0] \end{matrix} \right\rangle_{\mathrm{U}(n+1)} \\
&=\left\langle \begin{matrix} [m_{1n},\ldots,m_{n-1,n},0] \\ [m_{1,n-1},\ldots,m_{n-1,n-1}] \end{matrix} \middle\| \begin{matrix} p,\dot{0} \\ p,\dot{0} \end{matrix} \middle\| \begin{matrix} [m'_{1n},\ldots,m'_{n-1,n},0] \\ [m'_{1,n-1},\ldots,m'_{n-1,n-1}] \end{matrix} \right\rangle_{\mathrm{U}(n)}
\end{aligned} \tag{5.2}$$

Equation (5.2) is obviously true, since all one does is "drop off a zero" at the end of each irreducible representation on the left hand side of the equation.

It is interesting to apply Theorem 1 to SU(3), and compare with the work of Resnikoff[14] (1965) and Ališauskas[17] (1969).

Theorem 1, when applied to SU(3), gives, in the most general case, i.e., two irreducible representations totally symmetric, the same result as ($\pi$.5b) of Ališauskas,[17] apart from a dimensional factor, which is due to difference of definition. In addition, when there are three irreducible representations totally symmetric, one gets the result of ($\pi$.2) of Ališauskas, again, apart from the dimensional factor. [Note that $m=0$ in Ališauskas' Eq. ($\pi$.2).]

With regard to Resnikoff's work, it is interesting to note that Resnikoff obtained the correct number of terms when calculating the 6-$j$ symbol of SU(3) with three irredicuble representations totally symmetric. However, some of his terms are unfortunately incorrect. As a result, Resnikoff did not realize that his Eq. (4.6) is none other than the 6-$j$ symbol of U(2). We would like to express our admiration for his courageous effort in calculating the 6-$j$ symbol of SU(3). We hope he will accept our correction to his formula.

The following four terms on the neumerator of Resnikoff Eq. (4.7a),

$$\rho_{23}!(\lambda_1-\lambda_{13})!(\lambda_3-\lambda_{13})!\rho_{35}!,$$

should read

$$(\lambda_{12}+\mu_{12})!(\lambda_{13}+\mu_{13}-\lambda_1)!(\lambda+\mu-\mu_{13})!(\lambda_1-\mu_{13})!.$$

The following two terms on the denominator of (4.7b),

$$(\rho_{32}-s)!(k_{10}-\rho_{24}-\rho_{43}-s)!,$$

should read

$$(\rho_{32}-\rho_{12}+s)!(\rho_{43}+\rho_{21}-\rho_{23}-s)!$$

*Corollary* 1: Theorem 1 is still true, with the appropriate change of phase factors, when the irreducible representations in (5.1) are changed from type 1 to types 2 and 3.

The corollary can be easily proved by using Eqs. (3.7) and (3.8). When applied to SU(3), this corollary gives the results of Ališauskas ($\pi$.3) and ($\pi$.4).

The next series of theorems deal with 6-$j$ symbols of U($n$) with three irreducible representations totally symmetric.

*Theorem 2a*: A 6-$j$ symbol in U($n$) with three irreducible representations totally symmetric is reducible to a sum over one variable only, and can therefore be expressed as a generalized hypergeometric function with unit argument.

*Proof*: Let us say in the 6-$j$ symbol in U($n-1$)

$$\begin{Bmatrix} [W_n,\dot{0}] & [m']_{n-1} & [m']_n \\ [p,\dot{0}] & [m]_n & [m]_{n-1} \end{Bmatrix}$$

$[m']_{n-1}$ is totally symmetric. Then from (3.6) we have the term

$$S^2_{n,n-1}([r]_n,[m']_{n-1})=S^2_{n,n-1}([\cdot r_{2n},\dots,r_{nn}],[m'_{1,n-1},O,\dots,O])=\sum_{[r]_n}\frac{\cdots r_{3n}!\cdots r_{nn}!}{\cdots(-r_{3n})!\cdots(-r)_{nn})!}.$$

The result is that $r_{3n}=r_{4n}=\cdots=r_{nn}=0$. Thus there is only one sum left, i.e., a sum over $r_{2n}$.

*Theorem 2b*: When a 6-$j$ symbol in U($n-1$) has three irreducible representations totally symmetric, then $[m']_n$ can have at most two rows nonzero in the Young tableaux, $[m]_{n-1}$ can have at most two rows nonzero in the Young tableaux, and have at most three rows nonzero in the Young tableaux.

*Proof*: From Eq. (3.6) and the definition of $S_{nm}$, we find that Theorem 2b is true, for otherwise, the 6-$j$ symbol will have factorials of negative integers.

*Theorem 2c*: All 6-$j$ symbols of U($n$), $n>3$, with three irreducible representations totally symmetric are equivalent to the 6-$j$ symbol of U(3).

*Proof*: Combining Theorems 1 and 2b, we can drop off all the zeroes at the end of each irreducible representation. Since the largest number of nonzro rows in the Young tableaux of the irreducible representations in the 6-$j$ symbol is three, we conclude that U(3) is the smallest group to which all other groups for $n>3$ can be reduced.

*Theorem 2d*: All 6-$j$ symbols of U($n$), $n>3$, with three irreducible representations totally symmetric can be written as a ${}_4F_3$ function with unit argument.

*Proof*: Using Theorem 2c, we only need to calculate the 6-$j$ symbol in U(3) of the form

$$\begin{Bmatrix} [W_3,\dot{0}] & [m'_{12},0,0] & [m'_{13},m'_{23},0] \\ [p,\dot{0}] & [m_{13},m_{23},m_{33}] & [m_{12},m_{22},0] \end{Bmatrix}. \tag{5.4}$$

We find, however, that this 6-$j$ symbol in U(3) can be written as a ${}_4F_3$ function with unit argument.

*Theorem* 2*e*: All 6-$j$ symbols of $\mathrm{U}(n)$ with three irreducible representations totally symmetric can be written as a ${}_4F_3$ function with unit argument.

*Proof*: We know that for U(2), the 6-$j$ symbol can also be written as a ${}_4F_3$ function with unit argument. Therefore Theorem 2e is true for all $n$.

*Theorem* 2*f*: All 6-$j$ symbols of $\mathrm{U}(n)$ with three irreducible representations totally symmetric are reducible to a 6-$j$ symbol in U(2).

*Proof*: We finally compute the expression in (5.4) and find that it is reducible to a 6-$j$ symbol in U(2). More precisely,

$$\begin{Bmatrix} [W_3,\dot{0}] & [m'_{12},0,0] & [m'_{13},m'_{23},0] \\ [p,\dot{0}] & [m_{13},m_{23},m_{33}] & [m'_{12},m_{22},0] \end{Bmatrix}$$
$$= \left[ \frac{(m_{12}-m_{33}+1)!(m_{22}-m_{33})!m_{33}!(m_{13}+2)!(m_{23}+1)!(m'_{12}+1)}{(m_{13}-m_{33}+2)!(m_{23}-m_{33}+1)!(m'_{12}-m_{33}+1)!(m_{12}+1)!m_{22}!} \frac{(m'_{12}+1)!(m'_{13}-m_{33}+1)!(m'_{23}-m_{33})!}{(m'_{13}+1)!m'_{23}!(m'_{12}-m_{33}+1)!} \right]^{1/2}$$
$$\times(-1)^{m_{13}+m_{23}} \begin{Bmatrix} [W_3-m_{33},0] & [m'_{12}-m_{33},0] & [m'_{13}-m_{33},m'_{23}-m_{33}] \\ [p-m_{33},0] & [m_{13}-m_{33},m_{23}-m_{33}] & [m_{12}-m_{33},m_{22}-m_{33}] \end{Bmatrix}_{\mathrm{U}(2)} \tag{5.5}$$

$\times$ measure factor $\times$ dimensional factor, where

$$\text{measure factor} = \left[ \frac{(m_{13}+2)!(m_{23}+1)!m_{33}!m'_{12}!(m'_{13}-m_{33}+1)!(m'_{23}-m_{33})!}{(m_{13}-m_{33}+2)!(m_{23}-m_{33}+1)!(m'_{12}-m_{33})!(m'_{13}+1)!m'_{23}!} \frac{(m_{12}-m_{33}+1)!(m_{22}-m_{33})!}{(m_{12}+1)!m_{22}!} \right]^{1/2}, \tag{5.6}$$

$$\text{dimensional factor} = \left[ \frac{\dim[m'_{13},m'_{23},0]_{\mathrm{U}(3)} \dim[m_{12},m_{22},0]_{U(3)}}{(m'_{13}-m'_{23}+1)(m_{12}-m_{22}+1)} \right]^{-1/2}. \tag{5.7}$$

*Corollary* 2: All Theorems 2a–f are true, with the appropriate change of phase factors, when one or more of the irreducible representations in the 6-$j$ symbol of $\mathrm{U}(n)$ are changed to their conjugate representations.

The proof follows from Eqs. (3.7) and (3.8).

## 6. WEYL COEFFICIENTS OF U(n)

What we have said about the multiplicity-free 6-$j$ symbols of $\mathrm{U}(n)$ can be equally applied to the Weyl coefficients of $\mathrm{U}(n)$. In particular, we wish to stress that the Weyl coefficients of $\mathrm{U}(n)$ are now explicitly known. If one uses Eq. (3.6), then the Weyl coefficients of $\mathrm{U}(n)$, being equivalent to a multiplicity-free 6-$j$ symbol in $\mathrm{U}(n-1)$, can be written as a sum over $n-2$ variables. We thus assert that the representation function of $\mathrm{U}(n)$ is also explicitly known, in that we can write down the $d$-function of $\mathrm{U}(n)$ completely and explicitly if we so wish. The $d$-function of $\mathrm{U}(n)$ can be compared with the double boson polynomial of $\mathrm{U}(n)$. In I, we have given the relation for U(3). Similar relations can obviously be obtained for higher order groups. The double boson polynomial can be obtained through Moshinsky's work[18] for the highest weight, and the lowering operators of Nagel and Moshinsky.[19] However, Holman[1] has already given an explicit expression for the $d$ functions of $\mathrm{U}(n)$ in Eqs. (2.16) and (2.19) of Ref. 1. Also an explicit expression has been given by Ciftan[10] for U(4). We can therefore regard this part of the problem as essentially solved.

Another question, however, which still remains somewhat mysterious, is the Regge[21] symmetry of the 6-$j$ symbols of $\mathrm{U}(n)$. We have shown in I that the 144 Regge symmetries of U(2) can be regarded as the symmetry of the Weyl coefficients of U(3)*U(3) for the following state:

$$\left| \begin{array}{cccc} & & J_2+J_3-K_1 & \\ & J_2 & & K_2-J_0 \\ K_2 & & J_2+J_3-K_3 & 0 \\ & J_3 & & K_2-J_1 \\ & & J_2+J_3-K_1 & \end{array} \right\rangle. \tag{6.1}$$

The 144 symmetries of Regge are then made up of separate permutations of $J_0$, $J_1$, $J_2$, and $J_3$, with separate permutations of $K_1$, $K_2$, and $K_3$, subject to the constraint

$$J_0+J_1+J_2+J_3-K_1-K_2-K_3=0. \tag{6.2}$$

The mysterious part is that neither Eq. (3.5) nor Eq. (3.6) gives the full Regge symmetry. If one writes Eq. (3.5) out in terms of $J_i$ and $K_i$ above, one finds that it is symmetric between the interchange of $J_0 \longleftrightarrow J_1$, and $J_2 \longleftrightarrow J_3$, but does not give the full 144 symmetries. Similarly, Eq. (3.6), as we have shown in Sec. 4, is essentially Racah's expression, and therefore does not give the full Regge symmetry explicitly.

This leads us to suspect that there may be yet other expressions for the 6-$j$ symbol or Weyl coefficient of $\mathrm{U}(n)$ which will display the full Regge symmetry at a glance. If these expressions can be found, then new Regge symmetries for U(3), U(4), etc., can also be found.

[1] W.J. Holman III, Nuovo Cimento A **4**, 904 (1971).
[2] M.K.F. Wong, J. Math. Phys. **17**, 1558 (1976).
[3] G. Racah, Phys. Rev. **62**, 438 (1942).
[4] S.J. Ališauskas, A.A.A. Jucys, and A.P. Jucys, J. Math. Phys. **13**, 1329 (1972).
[5] S.J. Ališauskas, V.V. Vanagas, and A.P. Jucys, Dokl. Akad. Nauk SSSR **197**, 804 (1971).
[6] E. Chacón, M. Cliftan, and L.C. Biedenharn, J. Math. Phys. **13**, 577 (1972).
[7] L. Banyai, N. Marinescu, I. Raszillier, and V. Rittenberg, Phys. Lett. **14**, 156 (1965); A.A. Jucys, Liet. Mat. Rink. **8**, 597 (1968); Liet. Fiz. Rink. **10**, 5 (1970).
[8] A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1960).
[9] G.E. Baird and L.C. Biedenharn, J.Math. Phys. **5**, 1723 (1964).
[10] J.-R. Derome and W.T. Sharp, J. Math. Phys. **6**, 1584 (1965).
[11] J.-R. Derome, J. Math. Phys. **7**, 612 (1966).
[12] J.-R. Derome, J. Math. Phys. **8**, 714 (1967).
[13] E.P. Wigner, "On the Matrices Which Reduce the Kronecker Product of Representations of S.R. Groups" in *Selected Papers on the Quantum Theory of Angular Momentum,* edited by L.C. Biedenharn and H. Van Dam (Academic, New York, 1965). See also E.P. Wigner, *Group Theory* (Academic, New York, 1959).
[14] M. Resnikoff, J. Math. Phys. **8**, 79 (1967).
[15] R.T. Sharp, Nucl. Phys. A **95**, 222 (1967).
[16] B.M. Minton, J. Math. Phys. **11**, 3061 (1970).
[17] S.J. Ališauskas, Liet. Fiz. Rink. **9**, 641 (1969).
[18] M. Moshinsky, Nucl. Phys. **31**, 384 (1962).
[19] J.G. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965).
[20] M. Ciftan, J. Math. Phys. **10**, 1635 (1969).
[21] T. Regge, Nuovo Cimento **11**, 116 (1959).

# CUMULATIVE AUTHOR INDEX

*All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering each volume is published annually. An* (E) *after the page number indicates an Erratum.*