A note on general involutional transformations

S. Madivanane
MATSCIENCE, The Institute of Mathematical Sciences, Madras-600 113, India

(Received 25 October 1983; accepted for publication 16 February 1984)

For two matrices A and B of order $n\times n$ which satisfy $x^n - 1 = 0$, it is shown that the transformation matrix T_{AB} which connects A and B via $AT_{AB} = T_{AB}B$ becomes general involutional when $AB = \omega BA$, ω being a primitive nth root of unity.

PACS numbers: 02.10.Sp, 03.65. – w

General involutional transformations are matrices satisfying the relation

$$T^n = kI, \quad k = \text{const.}$$

The properties of these matrices were studied by Santhanam *et al.* and the applications of them in physics were already known.[1] Recently Kim[2] developed a new method of matrix transformation and studied involutional transformations. For two matrices A and B of order $n\times n$ satisfying a monic quadratic equation with two roots λ_1 and λ_2, he observed that the transformation matrix T_{AB} (Ref. 2) which connects A and B via $AT_{AB} = T_{AB}B$ becomes involutional when the anticommutator of

$$\overline{A} = A - \tfrac{1}{2}(\lambda_1 + \lambda_2)I$$

and

$$\overline{B} = B - \tfrac{1}{2}(\lambda_1 + \lambda_2)I \tag{1}$$

is a c number. In this note we consider the case of the nth degree equation

$$x^n - 1 = 0, \tag{2}$$

and prove that T_{AB} becomes general involutional when A and B satisfy

$$AB = \omega BA, \tag{3}$$

where ω is a primitive nth root of unity.

Weyl[3] has shown that the abelian rotations in ray space defined by Eq. (3) is a finite group and has a single unique (apart from ordering) irreducible representation given by

$$A = (a_{rs}), \quad a_{rs} = \delta_{r,s-1}; \quad r<n,\ s\leqslant n, \qquad a_{ns} = \delta_{1,s}; \quad s\leqslant n, \tag{4}$$

$$B = (b_{rs}), \quad b_{rs} = \omega^s \delta_{rs}, \tag{5}$$

δ being the Kronecker symbol. In the limit of continuous spectrum and as n goes to ∞, the representation becomes that of Heisenberg and Schrödinger.

Obviously $A^n = B^n = I$ and the transformation matrix T_{AB} (Ref. 4) takes the form

$$T_{AB} = \sum_{k=0}^{n-1} A^{n-1-k} B^k, \tag{6}$$

which becomes in our case

$$T_{AB} \equiv T = (t_{k_1 k_2}),$$

where $t_{k_1 k_2} = \omega^{k_2(k_1 - k_2 - 1)}$; $k_1, k_2 = 0, 1, 2, \ldots, n-1$. It is easy to observe that

$$\operatorname{tr}(T^m) = \sum_{k_1, k_2, \ldots, k_m = 0}^{n-1} \omega^{k_2(k_1 - k_2 - 1) + k_3(k_2 - k_3 - 1) + \cdots + k_m(k_{m-1} - k_m - 1) + k_1(k_m - k_1 - 1)}, \tag{7}$$

where tr denotes the trace of a matrix.

For each fixed $(m-1)$-tuple $(l_2, \ldots, l_m)$ such that $0\leqslant l_2, \ldots, l_m \leqslant n-1$, put

$$\begin{aligned} k_m &= k_1 + l_m, \\ k_{m-1} &= k_1 + l_{m-1} \\ &\vdots \quad\ \vdots \\ k_2 &= k_1 + l_2. \end{aligned}$$

Then the terms in the above sum (7) reduce to

$$\omega^{-mk_1 + l_2(-l_2 - 1) + l_3(l_2 - l_3 - 1) + l_4(l_3 - l_4 - 1) + \cdots + l_m(l_{m-1} - l_m - 1)} \tag{8}$$

for each fixed $(m-1)$-tuple and for any k_1 such that $0\leqslant k_1 \leqslant n-1$.

On summing over k_1 in (8) we observe that the coefficient of each term

$$\omega^{l_2(-l_2 - 1) + l_3(l_2 - l_3 - 1) + \cdots + l_m(l_{m-1} - l_m - 1)}$$

is zero unless $m = n$. Hence, $\operatorname{tr}(T^m) = 0$ provided $m \neq n$. We now apply the Hamilton–Cayley theorem to conclude that

$$T^n = C_n I,$$

where

$$C_n = n^{-1} \operatorname{tr}(T^n).$$

ACKNOWLEDGMENT

I thank Professor T. S. Santhanam for suggesting this investigation and for his guidance.

[1] T. S. Santhanam, P. S. Chandrasekaran, and N. B. Menon, J. Math. Phys. **12**, 377 (1971), and reference therein.
[2] S. K. Kim, J. Math. Phys. **20**, 2153 (1979).
[3] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950) (English translation), pp. 272–280.
[4] S. K. Kim, J. Math. Phys. **20**, 2159 (1979). For special case of the matrices A and B that we consider, this transformation has already been given in A. Ramakrishnan, *L. Matrix Theory or the Grammar of Dirac Matrices* (Tata McGraw-Hill, India, 1972), pp. 7, 89, 91.

 0022-2488/84/102807-01$02.50

Simultaneity and reality of U(n) and SU(n) 3jm and 6j symbols

R. P. Bickerstaff
Institute for Nuclear Theory and Department of Physics, FM-15, University of Washington, Seattle, Washington 98195

(Received 24 January 1984; accepted for publication 18 May 1984)

A simple proof is given of the simultaneous existence of a real set of 3jm and 6j symbols for the unitary groups U(n) and SU(n); including the case of mixed tensor representations. We observe that simultaneity is incompatible with the conventional permutation symmetries for U(n) with $n > 3$. The relevance of these results to the Schur–Weyl duality of U(n) with the symmetric groups S_l is discussed.

PACS numbers: 02.20. + b, 11.30.Jw, 11.30.Kx, 03.65.Fd

I. INTRODUCTION

Since irreducible representations (irreps) of the unitary group U(n) remain irreducible on restriction to its unimodular subgroup SU(n), it is fairly simple to deduce that the 3jm and 6j symbols, arising in the Wigner–Racah (or coupling) algebra[1,2] may be the same for both groups. We shall refer to such an equality by the term "simultaneity." Many authors have naively concluded from this that the SU(n) algebra possesses all of the known properties of the U(n) algebra and sometimes carelessly use U(n) techniques for SU(n) groups. Simultaneity is however actually rather subtle because the entire series of "associated" irreps in U(n), which are related by Kronecker products with the one-dimensional determinantal irrep, become equivalent in SU(n) as a consequence of the determinantal (pseudoscalar) irrep subducing to the identity (scalar) irrep. It is most important to realize that only one of the equivalent irrep bases should be used in calculations. Thus, if the 3jm and 6j symbols are to be the same for U(n) and SU(n), it is necessary that for associated U(n) irreps the U(n) symbols must be identical.[3] This restriction could well be at variance with those phase choices necessary to realize certain desirable properties of the U(n) symbols, e.g., reality and simple permutation symmetries. Certainly, one is not entitled to assume that the existence of U(n) symbols with these properties implies that SU(n) symbols will also possess them.

An alternative way of seeing that simultaneity is nontrivial is to note than U(n) is isomorphic to $[\mathrm{U}(1)\times\mathrm{SU}(n)]/Z_n$, where $Z_n = C_n \times I_n$ with C_n being the n-fold cyclic group whose elements are the nth roots of unity and I_n the trivial group whose only element is the identity matrix of order n. For any direct product of groups, in particular $\mathrm{U}(n)\times C_n \simeq U(1)\times\mathrm{SU}(n)$, one can choose a factorizable (direct-product) basis such that the 3jm and 6j symbols for the product group are a product of 3jm and 6j symbols, respectively, for the two groups. In the present case we note that both C_n and U(1) are abelian, and therefore only have one-dimensional irreps, so that one may readily choose[2,4] all of their 3jm and 6j symbols to be + 1. Thus one might hope to equate U(n) and SU(n) 3jm and 6j symbols (as well as 9j symbols,[2] etc.) by using a basis which is simultaneously factorizable for $C_n \times \mathrm{U}(n)$ and $\mathrm{U}(1)\times\mathrm{SU}(n)$. Although this is possible it depends on certain choices being made for the U(n) symbols. In general, we only have that the U(n) symbols can be chosen the same as for $\mathrm{U}(1)\times\mathrm{SU}(n)$ and the former need not be factorizable, especially if various symmetries have been imposed. The possibility of 3jm and 6j symbols for U(n) which are simultaneously 3jm and 6j symbols for SU(n) is so attractive that it deserves very high priority in making choices for U(n), perhaps even if this means foregoing other symmetries. Nonetheless, nonfactorizable choices for U(n) do appear in the literature.[5]

Although it is well known from the existence of closed formulas[6] (and also from a more abstract argument[7]) that the 3jm and 6j symbols of SU(2) can be chosen real, a rigorous proof for general SU(n) groups is lacking. Real symbols can be of practical value, particularly in computer calculations, but despite their interest it is only recently that the general criteria for reality are becoming understood. We now know[8] that several finite groups, in particular the alternating groups A_l for all odd l greater than 8 and all even l greater than 15, do not admit real 3jm symbols. In addition, it is known that, even for groups which do admit real 3jm symbols, the 3jm factors[2] (which result when Racah's factorization lemma[9] is applied to the 3jm symbols in a subgroup-symmetry-adapted basis) cannot always be chosen real.[4,10] Thus the use of a symmetry-adapted basis may preclude real 3jm symbols.[8] On the other hand, since a 6j symbol is a (basis independent) contraction of products of four 3jm symbols together with what are essentially some "trivial" 3jm symbols involving the identity representation,[1,2] a real set of 6j symbols exists if the 3jm symbols can all be chosen real in any one basis. We shall use this to avoid explicit treatment of the 6j symbols.

It is known (at least in the case of fully covariant, or contravariant, tensor irreps) that real 3jm and 6j symbols for U(n) exist, as do real 3jm factors for the important subgroup symmetry schemes $\mathrm{U}(nm)\supset\mathrm{U}(n)\times U(m)$ and $U(n+m)\supset\mathrm{U}(n)\times\mathrm{U}(m)$. The existence of real sets can be related to the Schur–Weyl duality that exists between the unitary groups U(n) and the symmetric group S_l.[11,12] This duality can be used to calculate 3jm and 6j symbols, etc. for U(n) via Young symmetrizers in S_l.[13] Indeed the results for the more fundamental, but unsymmetrized, recoupling coefficients and isoscalar factors—respectively, related to 6j symbols and 3jm factors—can be expressed directly in terms of various coefficients for chains of symmetric groups.[14–20] Thus their values are independent of n and they display S_l symmetries,

 0022-2488/84/102808-07$02.50

evident[15] in U(n) as, for example, Regge[21] symmetries. In particular, since all irreps of S_l are of the first Frobenius–Schur kind,[22] all of the matrix irreps and various coefficients in S_l can be chosen real and reality of the U(n) coefficients follows. The n-independence is spoilt for the 3jm factors and 6j symbols by irrep dimensionality factors but nevertheless an n-independent calculation is still possible and the reality issue is unaltered. Neither the 3jm symbols nor the unsymmetrized coupling (or Clebsch–Gordan) coefficients are n-independent but by setting up a chain of unitary groups ending in U(1)$\times\cdots\times$U(1) (n times), with real 3jm factors for each stage, we see that real choices are possible. Such prescriptions alone though do not guarantee that the U(n) coefficients obtained have all the properties desirable of that group. Of concern is whether the U(n) 3jm and 6j symbols obtained via the duality relations are factorizable and in particular whether the real symbols thus obtained are factorizable. We may also ask whether they possess conventional symmetries under permutation or complex conjugation of irreps. To cast a little doubt on the reality issue we recall that the imposition of certain "maximal" symmetry requirements forces some SU(n) 6j symbols to be imaginary.[23] While the calculations of that paper do not explicitly use the symmetric groups they can—to the extent that they have proceeded—easily be cast in an n-independent form.[24–27]

A direct solution to this problem, by considering the possible phase choices subject to the symmetries implied by duality, has recently been proposed.[28] However, such an approach is necessarily rather intricate and because it proceeds by exhausting all possibilities it is difficult to be sure that the necessary conditions considered are also sufficient.[28] We also note that the argument omitted mixed tensor irreps of U(n) and furthermore the full permutational symmetry of the 3jm and 6j symbols was not explicitly considered. In any case it would be preferable to investigate the reality aspect in more general terms.

In this article we give, independently of symmetric group choices, a relatively straightforward proof of reality for both U(n) and SU(n), including the important (and oft neglected) case of mixed tensor irreps, and demonstrate that real coefficients for both groups may be identified with each other. We begin by recalling a general criterion for reality[29] and then show that the standard matrix irreps for U(n) and SU(n) obey that criterion simultaneously. The results are extended to 3jm factors and the implications for the permutational properties of the 3jm symbols, etc. are considered. Compatibility of the simultaneity property with duality, in conjunction with both reality and permutational symmetries, is then discussed. We conclude with some further remarks on our reality observations.

II. CRITERION FOR REALITY

Damhus[29] has previously noted that the 3jm symbols

$$\begin{pmatrix} \Gamma_1 & \Gamma_2 & \Gamma_3 \\ i_1 & i_2 & i_3 \end{pmatrix}^r$$

of a (compact) group G can be chosen real if and only if the matrix Γ_{123} with elements

$$\Gamma_{123}{}^{i_1 i_2 i_3}_{i_1' i_2' i_3'} = \int_G dR \left[\Gamma_1(R)^{*i_1'}{}_{i_1} \Gamma_2(R)^{*i_2'}{}_{i_2} \Gamma_3(R)^{*i_3'}{}_{i_3} \right] \tag{1a}$$

$$= \sum_r \begin{pmatrix} \Gamma_1 & \Gamma_2 & \Gamma_3 \\ i_1 & i_2 & i_3 \end{pmatrix}^{r*} \begin{pmatrix} \Gamma_1 & \Gamma_2 & \Gamma_3 \\ i_1' & i_2' & i_3' \end{pmatrix}^r \tag{1b}$$

can be chosen real. The real 3jm symbols are selected from the columns of Γ_{123} using the freedom in the product multiplicity separation r. (Observe that Γ_{123} can be real for complex 3jm symbols.) A sufficient condition for Γ_{123} to be real is[29] that G possess an involutory automorphism τ such that for every matrix irrep $\Gamma(R)$ we have

$$\Gamma(\tau(R)) = \Gamma(R)^*, \qquad \text{for all } R \in G. \tag{2}$$

We now show that both U(n) and SU(n) possess such an automorphism.

III. MATRIX IRREPS OF UNITARY GROUPS AND PROOF OF REALITY

Consider the mapping

$$\tau: A \to A^* \tag{3}$$

of $n\times n$ unitary matrices. It is trivial to show that it is an involutory automorphism for both U(n) and SU(n). Moreover, these defining matrices themselves form a matrix irrep which clearly satisfies(2).

Further representations of U(n) may be found by considering the tensor powers $A^{\times l}$ of A. They are just the invariant matrices of Schur.[11] The tensor powers may be transformed into a direct sum of irreducible invariant matrices $A^{\{\lambda\}}$ whose characteristics are the S-functions[30] $\{\lambda\}$ of the characteristic roots of A.[11] If A is a transformation of covariant vectors, then $A^{\{\lambda\}}$ may be considered as a transformation of their l th covariant tensor powers symmetrized so as to transform under S_l according to the irrep $[\lambda]$ corresponding to the partition λ of l. This is the Schur–Weyl duality[11,12] and stems from the appearance of $S_l \times$ U(n) as a subgroup of the wreath product group U(n) wr S_l where the U(n) factor in the subgroup is the diagonal subgroup of the base group U(n)$^{\times l} \subset$ U(n) wr S_l.[31] This group structure is quite general and not specific to U(n). In the case of U(n) however there is a one-to-one correspondence between the irrep $[\lambda]$ of S_l and the irrep $\{\lambda\}$ of U(n) which yields the simple branching rule under restriction of U(n)$^{\times l}$ to U(n)

$$\{1\}^{\times l} \to \sum |\lambda|_{S_l} \{\lambda\}, \tag{4}$$

where the sum is over all $[\lambda]$ in S_l consistent with symmetrization on n variables. Thus the number of copies of admissible $A^{\{\lambda\}}$ in the reduction of $A^{\times l}$ is given by the dimension $|\lambda|_{S_l}$ of $[\lambda]$. Contravariant and mixed tensor irreps,denoted by $\{\bar{\mu}\} \equiv \{\bar{\mu};0\}$ and $\{\bar{\mu};\nu\}$, respectively, in composite (or back-to-back) notation,[32,33] can also be constructed; either in a like manner or, more directly, by considering the associated series of irreps formed by the tensor products

$$A^{\{\bar{\mu};\nu\}} = (A^{\{1^n\}})^{\times p} \times A^{\{\lambda\}}. \tag{5}$$

Here, the S-function $\{1^n\}$ labels the (one-dimensional) determinantal irrep of U(n) and (for true irreps) p is a positive or

negative integer. Note that while any equivalent matrix irrep $\hat{A}^{\{\mu;\nu\}}$ can be resolved in the form of (5), in general the basis for $\hat{A}^{\{\lambda\}}$ would turn out to be different for differing members of the associated series. [Observe also that use of a dual basis, as suggested by the action of U(n) wr S_l on the tensor product basis for U(n)$^{\times l}$, neither guarantees nor unavoidably violates (5), since the action of the dual subgroups S_l and U(n) on that product space commute.]

Because irreps of U(n) remain irreducible on restriction to SU(n) these matrices are also irrep matrices of SU(n). However, the entire series of asociated irreps becomes equivalent to a single irrep in SU(n) and in general only one member of that series may be used at one time without leading to inconsistencies. It is conventional to choose that member of the associated series which becomes standard in SU(n) to be that covariant irrep corresponding to a partition into less than n parts. [In that case the power p in Eq. (5) is minus the first part of the partition μ for contravariant or mixed tensor irreps[33] and plus the last, i.e., nth, part of the partition ν for covariant irreps.]

The elements of an irreducible invariant matrix of A are homogeneous polynomials in the elements of A [11] and since in the canonical form the coefficients of these polynomials are real[34,35] we immediately have that

$$(\mathrm{A}^*)^{\{\bar{\mu};\nu\}} = (\mathrm{A}^{\{\bar{\mu};\nu\}})^*. \tag{6}$$

Thus the condition (2) is realized for all irreps of U(n) and SU(n). In particular it is realized when the irreps of U(n) are of the form (5).

IV. SIMULTANEITY

There is no *a priori* reason to expect that Γ_{123} is the same for U(n) and SU(n). However, we note that the $3jm$ symbols are defined,[1] up to a transformation in the product multiplicity, by

$$\sum_{i_1 i_2 i_1' i_2'} |\Gamma_3| \begin{pmatrix} \Gamma_1 & \Gamma_2 & \Gamma_3 \\ i_1' & i_2' & i_3' \end{pmatrix}^r \Gamma_1(R)^{i_1'}{}_{i_1} \Gamma_2(R)^{i_2'}{}_{i_2} \times \begin{pmatrix} \Gamma_1 & \Gamma_2 & \Gamma_3 \\ i_1 & i_2 & i_3 \end{pmatrix}^{r'} = \Gamma_3(R)^{*i_3'}{}_{i_3} \tag{7}$$

for all $R \in G$, where $|\Gamma_3|$ is the dimension of Γ_3. Since irreps of U(n) remain irreps on restriction to SU(n) it is clear from (7) that, for a given triad $(\Gamma_1\Gamma_2\Gamma_3)$, the $3jm$ symbols for U(n) are also $3jm$ symbols for SU(n). Indeed they may not differ by more than a unitary transformation in r and hence Eq. (1b) shows that Γ_{123} is the same for U(n) and SU(n). Furthermore, for those elements A of U(n) which are also elements of SU(n), $(A^{\{1^n\}})^{\times p} = +1$ and thus all matrix irreps of A constructed via (5) subduce to an identical matrix irrep in SU(n). (Note that this is only indirectly related to the factorization of groups discussed in the Introduction.) For such a basis then, Γ_{123} will also be the same for all triads $(\Gamma_1\Gamma_2\Gamma_3)$ of U(n) which become equivalent under SU(n). It now follows that by resolving the multiplicity for each triad in an analogous manner to the resolution made for real $3jm$ symbols in SU(n), all corresponding $3jm$ symbols in U(n) will be equal with each other and identical with the SU(n) $3jm$ symbols.

This result immediately follows on for $6j$ symbols of U(n) and SU(n), by the argument mentioned in the Introduction. However, the $6j$ symbols are basis independent and, therefore, neither their reality nor their simultaneity should depend on the use (as compared to existence) of irrep matrices constructed via (5). This prompts us to probe a little deeper. Indeed, given that a real set of $3jm$ symbols possessing a simultaneous resolution of multiplicity exists, we can clearly perform a basis transformation on the irrep matrices which may spoil both the simultaneity and the reality of the $3jm$ symbols but which, because it does not affect the multiplicity resolution, will affect neither the reality nor the simultaneity of the $6j$ symbols.

It is worthwhile noting that a U(n) triad will be related by equivalence in SU(n) to the U(n) triad $(\{\bar{\mu}_1;\nu_1\}, \{\bar{\mu}_2;\nu_2\}, \{\bar{\mu}_3;\nu_3\})$ if and only if it is of the form

$$(\{p_1^n\}\{\bar{\mu}_1;\nu_1\}, \{p_2^n\}\{\bar{\mu}_1;\nu_1\}, \{p_3^n\}\{\bar{\mu}_3;\nu_3\}),$$

where

$$p_1 + p_2 + p_3 = 0 \tag{8}$$

and $\{-p^n\}$ is interpreted as $\{\bar{p}^n\}$. The multiplicity of the product is clearly the same for all such triads. [In proving this result we have to show that the Kronecker product of two irreps in U(n) never contains more than one distinct member of the associated series. That this is the case easily follows by writing the product in terms of fully covariant tensor irreps, using powers of $\{1^n\}$, and then noting that all terms appearing in the Kronecker product of two covariant irreps involve partitions of the same number.] Notice that since all associated irreps become equivalent in SU(n) we would be quite entitled to relax the condition (8) within that group. However, to do so in the present context could be confusing as a triad violating (8) will not exist in U(n).

In connection with the last remark it deserves to be emphasized that simultaneity only refers to a numerical equality of the U(n) and SU(n) coefficients and merely allows one to retain the same coupling algebra on restriction from U(n) to SU(n). It does not mean that the tensor products of irrep matrices of elements outside of SU(n) may be reduced by SU(n) coupling coefficients. An example will illustrate this point. Let A be an element of SU(2) and let Z be an element of U(2) for which $\det Z \neq 1$. Then, while it is true that the same SU(2) coefficients which reduce $A^{\{1\}} \times A^{\{1\}}$ to $A^{\{0\}} \dot{+} A^{\{2\}}$ also reduce $Z^{\{1\}} \times Z^{\{1\}}$ to $Z^{\{1^2\}} \dot{+} Z^{\{2\}}$, the SU(2) coefficients which reduce $A^{\{0\}} \times A^{\{1\}}$ to $A^{\{1\}}$ are readily seen to reduce $Z^{\{1^2\}} \times Z^{\{1\}}$ to $Z^{\{21\}} = (\det Z) Z \neq Z^{\{1\}}$. It is clearly not possible to match the arguments of the entire set of SU(2) coefficients with all U(2) decompositions. This remark assumes considerable significance in any attempt to prove the necessity of (2) (cf. Ref. 8).

We should also note that simultaneity is not peculiar to the unitary groups. The cyclic groups provide many trivial examples of this, e.g., for both C_6 and its subgroup C_3 all coefficients may be chosen $+1$.

V. $3jm$ FACTORS

The above proofs apply to all n. However, in order to use Racah's factorization lemma to extend the results to $3jm$ factors for SU(n)$\supset$SU(n_1)$\times$SU(n_2), where $n = n_1 n_2$ or $n_1 + n_2$, we must show that the result for U(n) applies in the

appropriate symmetry-adapted basis. In fact, the mapping (3) in the defining irrep meets all the necessary requirements in any basis. We must show though that (6) still applies. This will be so if we can show that for the standard irrep matrices $A^{\{\lambda\}}$ used in (5), and which satisfy (6) for all A belonging to $U(n)$, $A^{\{\lambda\}}$ is diagonal in branching multiplicity and subgroup irrep labels for those A belonging to the subgroup, and further, that the subgroup irrep matrices appearing on the diagonal factorize into irrep matrices $A_{n_1}^{\{\mu_1\}} \times A_{n_2}^{\{\mu_2\}}$ of $U(n_1) \times U(n_2)$; for which we will also require that (6) holds.

Consider $n = n_1 n_2$. Let the n-dimensional vector space mapped by the defining irrep of $U(n)$ be a tensor product of n_1- and n_2-dimensional spaces so that for those elements A of $U(n)$ which also belong to $U(n_1) \times U(n_2)$ the defining irrep matrices are

$$A_n^{\{1\}} = A_{n_1}^{\{1\}} \times A_{n_2}^{\{1\}}. \tag{9}$$

It is well known that the $[\lambda]$-symmetrized parts of the l th tensor powers of this product space are just a direct sum of tensor products of the $[\mu_1]$-symmetrized parts of the l th tensor power of the first space with the $[\mu_2]$-symmetrized parts of the l th tensor power of the second space, where the multiplicity $g_{\mu_1\mu_2\lambda}$ in the direct sum is just the multiplicity of $[\lambda]$ in the inner product of $[\mu_1]$ and $[\mu_2]$ in S_l. (Such results are elegantly expressed using Littlewood's plethysm of S-functions.[34]) Clearly then, for those A_n belonging to $U(n_1) \times U(n_2)$

$$A_n^{\{\lambda\}} = \dot{+}_{\mu_1\mu_2} g_{\mu_1\mu_2\lambda} A_{n_1}^{\{\mu_1\}} \times A_{n_2}^{\{\mu_2\}}, \tag{10}$$

which certainly suffices if $A_{n_1}^{\{\mu_1\}}$ and $A_{n_2}^{\{\mu_2\}}$ are standard. For general $A_n^{\{\bar{\mu};\nu\}}$, we simply note that

$$A_n^{\{1^n\}} = \left(A_{n_1}^{\{1^{n_1}\}}\right)^{\times n_2} \times \left(A_{n_2}^{\{1^{n_2}\}}\right)^{\times n_1} \tag{11}$$

and the form of (10) is unaffected. If $A_{n_1}^{\{\mu_1\}}$ and $A_{n_2}^{\{\mu_2\}}$ appearing in (10) are nonstandard, as will frequently be the case, it will still be true with the present construction that the condition (6) is satisfied within both $U(n_1)$ and $U(n_2)$ and so reality of the $3jm$ symbols for both groups in these bases is assured. Since these bases can be obtained from those of the type (5) without affecting the simultaneity of the multiplicity resolutions in $U(n_1)$ and $U(n_2)$ and because the $3jm$ factors for $U(n) \supset U(n_1) \times U(n_2)$ do not depend on the bases of $U(n_1)$ or $U(n_2)$—but do depend on their product multiplicity resolutions—this is all that is required. Thus, applying the results of Secs. III and IV together with Racah's factorization lemma[2,9] we deduce that $3jm$ factors for $SU(n_1 n_2) \supset SU(n_1) \times SU(n_2)$ may be real and simultaneously $3jm$ factors for $U(n_1 n_2) \supset U(n_1) \times U(n_2)$.

The case $n = n_1 + n_2$ follows similarly by letting the vector space mapped by the defining irrep of $U(n)$ be a direct sum of n_1- and n_2-dimensional spaces. Hence, for those elements A of $U(n)$ which also belong to $U(n_1) \times U(n_2)$ the defining irrep matrices are

$$A_n^{\{1\}} = A_{n_1}^{\{1\}} \dot{+} A_{n_2}^{\{1\}}. \tag{12}$$

In analogy with Eqs. (10) and (11) we find

$$A_n^{\{\lambda\}} = \dot{+}_{\mu_1\mu_2} \Gamma_{\mu_1\mu_2\lambda} A_{n_1}^{\{\mu_1\}} \times A_{n_2}^{\{\mu_2\}} \tag{13}$$

(where $\Gamma_{\mu_1\mu_2\lambda}$, commonly referred to as the outer product multiplicity, is the multiplicity of $[\lambda]$ when the representation of S_l induced by the direct product representation $[\mu_1] \times [\mu_2]$ of $S_{l_1} \times S_{l_2}$, where $l_1 + l_2 = l$, is restricted to irreps of S_l) and

$$A_n^{\{1^n\}} = A_{n_1}^{\{1^{n_1}\}} \times A_{n_2}^{\{1^{n_2}\}}. \tag{14}$$

Thus we are also able to deduce that $3jm$ factors for $SU(n_1 + n_2) \supset SU(n_1) \times SU(n_2)$ may be real and simultaneously $3jm$ factors for $U(n_1 + n_2) \supset U(n)_1 \times U(n_2)$.

VI. PERMUTATION SYMMETRY

The matrix Γ_{123} is invariant under permutation of the irreps and thus permuted $3jm$ symbols may differ only by a (unitary) transformation in the product multiplicity,[1] i.e.,

$$\begin{pmatrix} \Gamma_a & \Gamma_b & \Gamma_c \\ i_a & i_b & i_c \end{pmatrix}^r = \sum_{r'} M_\pi \{\Gamma_1\Gamma_2\Gamma_3\}^r_{r'} \begin{pmatrix} \Gamma_1 & \Gamma_2 & \Gamma_3 \\ i_1 & i_2 & i_3 \end{pmatrix}^{r'}, \tag{15}$$

where abc is the permutation π of 123. Under normal circumstances a certain amount of choice exists for the permutation matrices M_π. The available choices depend on whether the triad $(\Gamma_1\Gamma_2\Gamma_3)$ is of one or another of the following types.[36]

1. $\Gamma_1 \neq \Gamma_2 \neq \Gamma_3 \neq \Gamma_1$. There are five fundamental transposition matrices in terms of which all M_π can be expressed. These five matrices are completely arbitrary, cf. Derome.[36] However, the conventional choice[2] is for them all to be the same diagonal matrix of real phases (and thus Hermitian). All eighteen transposition matrices are then the same and all eighteen cyclic permutation matrices are the identity.[36]

2. $\Gamma_1 = \Gamma_2 \neq \Gamma_3$. There are only three fundamental transposition matrices. The transposition matrix $M_{(12)}\{\Gamma_1\Gamma_2\Gamma_3\}$ must be Hermitian and will be a representation matrix of S_2. Its diagonal form is therefore fixed by whether the rth occurrence of Γ_3 is in the symmetric or antisymmetric part of $\Gamma_1 \times \Gamma_1$; because the diagonal element must be the character (which is $+1$ and -1, respectively) of either the identity irrep [2] or the alternating irrep $[1^2]$ in the class whose cycle structure[30,34] is (2^1). It is conventional[2,36] to choose this (fixed) diagonal form for all transposition matrices. The cyclic permutation matrices are thus all the identity once more.

3. $\Gamma_1 = \Gamma_2 = \Gamma_3$. There are only two fundamental transposition matrices and they generate the group S_3. It is conventional to choose the permutation matrices to be diagonal in the totally symmetric, $\Gamma_1 \otimes \{3\}$, and totally antisymmetric, $\Gamma_1 \otimes \{1^3\}$, parts of the Kronecker cube. An examination of the characters of the (one-dimensional) irreps [3] and $[1^3]$ in S_3 shows that this (fixed) diagonal form is the same as the choices made in case 2. If though the identity occurs in the mixed symmetry part, $\Gamma_1 \otimes \{21\}$, of the Kronecker cube then the permutation matrices cannot all be chosen completely diagonal because the irrep [21] is two-dimensional.[2,36] The group is said to be simple phase[37] if the mixed symmetry case never arises.

However, if the arguments of Secs. III–V were followed independently for each permuted triad then the permutation matrices would—except by fluke—be fixed in other than standard form. Obviously though, they would be real. One can therefore argue in a straightforward, but nontrivial, fashion that the conventional symmetries are consistent with

the reality condition.

More problematic is the simultaneity condition. The unitary groups are simple phase for $n = 1,2$, and 3[38,39] but for $n \geqslant 4$ they are not.[39,40] Consider the irrep $\{\overline{21};21\}$ which is equivalent to $\{432^{n-4}1\}$ in SU(n) and which exists for $n \geqslant 4$. The type 3 triad $(\{\overline{21};21\}\ \{\overline{21};21\}\ \{\overline{21};21\})$ is not simple phase.[2,39] Recalling (8) we observe that there are an infinite number of distinct triads in U(n) which become equivalent to this one in SU(n) and none of them are of type 3. Take U(4), for example. Letting $p_1 = 1, p_2 = -1$, and $p_3 = 0$ we obtain the type 1 triad $(\{\overline{1};32\}\ \{\overline{32};1\}\ \{\overline{21};21\})$ and letting $p_1 = p_2 = 1$ and $p_3 = -2$ we obtain the type 2 triad $(\{\overline{1};32\}\ \{\overline{1};32\}\ \{\overline{431};0\})$; both of which become equivalent in SU(4) to the non-simple-phase triad $(\{431\}\ \{431\}\ \{431\})$. Since the permutation matrices may be expressed as a contraction of 3jm symbols, by inverting (15) via the orthogonality relations,[2] we readily see that simultaneity requires the permutation matrices for each of these triads to be the same. Although demanding that the permutation matrices for a type 1 triad be identical to those for a type 2 triad would not be inconsistent with convention, demanding the permutation matrices for type 1 and 2 triads to have off-diagonal elements is definitely unconventional. Thus simultaneity of U(n) and SU(n) 3jm symbols is inconsistent with conventional U(n) permutation symmetries for $n > 3$.

Derome and Sharp[1] also discuss symmetries under complex conjugation of irreps in terms of Γ_{123}. However, it is known that the simplest choice, viz. the identity,[2] for the conjugation (i.e., 2jm, equivalently 1-jm, and A [1,2]) matrices, although consistent with simultaneity,[39] leads to a complex algebra.[23,27,28] This result is independent of the permutation matrices and we will therefore not pursue the matter other than to remark that the conjugation matrices in our case will be real.

VII. DUALITY

The precise relationships between general U(n) and S_l coefficients have recently been elegantly expressed in a simple unified manner for all coefficients (involving only covariant tensor irreps) which are of immediate interest.[41,42] For n-independence of the U(n) coefficients it is sufficient that the three duality factors[42] $D_n(\lambda_1\lambda_2,\lambda_3)$, $D_{n_1n_2}(\lambda,\mu_1\mu_2)$, and $D_{n_1+n_2}(\lambda,\mu_1\mu_2)$, which relate unitary group and symmetric group multiplicity separations, be chosen independently of n, n_1, and n_2. In fact for any set of symmetric group multiplicity separations it is trivial to show that all may be chosen to be the identity. This leads to the results for U(n) discussed in the Introduction. However, it is generally very difficult to show that the symmetric group multiplicity separations may be made so as to exhibit desirable U(n) properties, such as the simultaneity which would be, at least partially, necessary for the n-independent properties to carry over to SU(n).[28] [Of course one should be able to establish the existence of n-independent properties for SU(n) independently of U(n) but in the context of both the published literature and the present paper it is simpler to consider the issue via simultaneity.]

In a similar vein we observe that in the present scheme, and any other like it where U(n) symmetries have already been imposed, the duality factors are no longer completely arbitrary.[42] This is because there necessarily exist restrictions on the U(n) multiplicity separations which ensure that the imposed symmetries are maintained.[24] It is therefore not obvious to what extent the current results are compatible with n-independent duality factors. We may though make some observations.

One aspect which does emerge from the present investigation is that a duality consistent phase convention for SU(n) would necessarily be incompatible with the usual simplified permutation matrices for SU(n). For example, the triad $(\{\bar{0};431\}\ \{\bar{0};431\}\ \{\overline{6532};0\})$ which is legitimate in U(5), and equivalent to the type 2 triad $(\{431\}\ \{431\}\ \{6431\})$ in SU(5), is also legitimate in U(4) and because the product is one of fully covariant tensor irreps, coefficients involving it are related by duality.[42] However, in SU(4) that triad is equivalent to $(\{431\}\ \{431\}\ \{431\})$ which, as we have earlier remarked, is of type 3 and non-simple-phase.[2,39] The detailed argument is more involved than that in Sec. VI because the permutational symmetries of isoscalar factors and recoupling coefficients involve more than one permutation matrix as well as other transformations, but consideration of a few examples quickly reveals that the standard choice of permutation matrices for these triads is necessarily incompatible with n-independent properties. [It does not appear to have been noted before that the above possibility is in fact realized although it has often been remarked that duality requires the resolution of multiplicity for many type 1 triads to be made in the same manner as for type 2 triads with which they become equivalent in some SU(n). Indeed it is sometimes said that this provides "natural" phase choices.[27]]

Using the above as justification one could ignore all permutation symmetries, including for the time being the (12) transposition symmetry, and enquire whether n-independent duality factors might then be compatible with simultaneity. Consider $D_n(\lambda_1\lambda_2,\lambda_3)$ which is associated with the inner product $\{\lambda_1\}\times\{\lambda_2\}\supset\{\lambda_3\}$ in U(n) and the corresponding symmetric group outer product arising in the subduction of $[\lambda_3]$ to $[\lambda_1]\times[\lambda_2]$ under the restriction of S_{l_3} to $S_{l_1}\times S_{l_2}$.[42] The following argument shows that all such factors may be chosen unity and thus the recoupling coefficients for SU(n) may be chosen n-independently. We start with all $\{\lambda_1\}$, $\{\lambda_2\}$, $\{\lambda_3\}$ associated with partitions into one part. There are an infinite number of such products in any U(n) but in U(1) all are related by (8) and thus there is only one independent coupling in SU(1), which we could take as $\{0\}\times\{0\}\supset\{0\}$. Consider then the case $n = 1$. To ensure simultaneity the U(1) phases associated with these couplings should all be fixed to that for $\{0\}\times\{0\}\supset\{0\}$. We may though use the freedom in the symmetric group products, those being different in each case, to set $D_1(\lambda_1\lambda_2,\lambda_3) = +1$. For $n > 1$, the symmetric group phases must remain unaltered but the unitary group phases (which are no longer related by simultaneity) may now be used to set $D_n(\lambda_1\lambda_2,\lambda_3) = +1$ also. Thus these duality factors are independent of n. Next consider all $\{\lambda_1\}$, $\{\lambda_2\}$, $\{\lambda_3\}$, where $\{\lambda_3\}$ is associated with a partition into two parts and $\{\lambda_1\}$ and $\{\lambda_2\}$ are associated with partitions into two or less parts. No such products exist in U(1). Consider then $n = 2$. Some pro-

ducts will be associated in U(2) with those previously considered. For example $\{2\}\times\{11\}\supset\{31\}$ is related to $\{2\}\times\{0\}\supset\{2\}$ by $p_2=p_3=1$ in (8). Others will be associated with products which now arise for the first time. For example, $\{3\}\times\{21\}\supset\{42\}$ is related to $\{3\}\times\{1\}\supset\{31\}$. These couplings are not independent in SU(2) and thus to ensure simultaneity no free phase should be connected with them. However, the corresponding symmetric group products may be used to set $D_2(\lambda_1\lambda_2,\lambda_3)=+1$. For the remaining products, such as $\{1\}\times\{1\}\supset\{11\}$ and $\{3\}\times\{1\}\supset\{31\}$, it is possible to use the U(2) freedom to set $D_2(\lambda_1\lambda_2,\lambda_3)=+1$, regardless of the symmetric group phase choice. For $n>2$, simultaneity is no longer relevant and the unitary group phases alone may be used to set $D_n(\lambda_1\lambda_2,\lambda_3)=+1$. Proceeding in this manner for partitions into 3, 4, 5 parts etc., we may choose all of the duality factors $D_n(\lambda_1\lambda_2,\lambda_3)=+\mathrm{I}$, where I is the identity matrix whose order is that of the multiplicity. (Multiplicity, where it occurs, is readily seen to be n-independent and in no way complicates the line of argument.) Thus we have established the claim that all of these duality factors may still be chosen independently of n. Because all recoupling coefficients of SU(n) can be written in terms of products of covariant tensor irreps, which are valid in U(n), n-independence of those coefficients results. Notice that, in contrast to Sullivan's work,[28] the above argument is based on a sufficient condition.

We should now consider the compatibility of this result with both reality and the standard choice of a diagonal form for the (12) transposition matrix. In choosing the duality factors we have imposed a restriction which equates unitary group and symmetric group multiplicity choices, i.e., we may not now change one without an identical change in the other, or else the duality factors will be altered from their n-independent values. We have also had to restrict U(n) phases amongst themselves so as to preserve SU(n) equivalences. Thus restrictions also exist among the symmetric group phases. Consequently, one has to consider carefully whether sufficient freedom remains to choose all of the symmetric group coefficients to be real. Indeed it can be argued that the necessary freedom no longer exists. However, it is conceivable where a choice of reality may not be made that reality will nevertheless result as a consequence of previous choices. It is at least encouraging to know that U(n) choices displaying the property of simultaneity in conjunction with the property of reality do exist.

We can be more definite with regard to the (12) transposition phase. Consider the case $\lambda_1=\lambda_2$ and let n be the smallest value for which the U(n) product occurs. A simultaneous transformation of both the unitary group and symmetric group multiplicities will transform the transposition matrix to its (fixed) diagonal form[36] without affecting the duality factor. Invoking simultaneity in U(n) reveals that all transposition matrices related to this one by SU(n) equivalences will be identical. Furthermore, the duality results show that the transposition symmetry is independent of n. Next consider $\lambda_1\neq\lambda_2$. If two products $\{\lambda_1\}\times\{\lambda_2\}\supset\{\lambda_3\}$ and $\{\lambda_2\}\times\{\lambda_1\}\supset\{\lambda_3\}$ are related by equivalence in SU(n) then they are also related by equivalence to a product $\{\lambda'_1\}\times\{\lambda'_2\}\supset\{\lambda'_3\}$, where $\lambda'_1=\lambda'_2$. Hence such products have already been considered. Therefore, we are left with those cases for which the couplings are independent. (There is of course still only one transposition matrix: $M_{(12)}\{\lambda_1\lambda_2\lambda^*_3\}=M_{(12)}\{\lambda_2\lambda_1\lambda^*_3\}^{-1}$.) By making simultaneous transformations in both the symmetric and unitary groups $M_{(12)}$ may be arbitrarily chosen[36] without affecting the duality factors. Once more we may invoke simultaneity and duality to obtain those transposition matrices which may not be chosen. Thus duality for SU(n) is compatible with the conventional (12) transposition symmetry.

Finally we note that duality for SU(n) only required simultaneity for fully covariant tensor irreps. A similar duality exists for fully contravariant tensor irreps but the symmetric groups which act on the covariant and contravariant indices are different and consideration of mixed tensor irreps would involve a direct product of these groups and the outer product of their irreps. The question arises as to whether complete simultaneity, as espoused in the earlier sections, is compatible with n-independence. This would be extraordinarily demanding because it would vastly restrict the unitary group freedoms and thence also the symmetric group freedoms via the duality factors (which have only been defined for fully covariant or contravariant products). As an example, the product $\{1\}\times\{1\}\supset\{11\}$ is related via association in U(2) to the contravariant product $\{\bar{1}\}\times\{\bar{1}\}\supset\{\bar{1}\bar{1}\}$ so the corresponding duality factors could not be chosen without using symmetric group freedom. Furthermore, the degree of multiplicity where mixed tensor irreps are concerned is found to depend on n[27,39] because of the necessity to apply modification rules[33,43] for small n. Nevertheless, there are some cases where the recoupling coefficients for mixed tensor irreps are n-independent.[26,27] So far the details of this are not understood.

VIII. CONCLUDING REMARKS

The existence of real coefficients for U(n) and SU(n) has now been rigorously established, even for those coefficients involving mixed tensor irreps of U(n) which cannot be addressed via current duality arguments. The proof presented here utilizes an involutory automorphism (3) with the special property described by (2). Consequently,[29] that automorphism is also class inverting[44] and therefore U(n) and SU(n) are quasiambivalent.[44,45] This is in line with recent work[8,29,46] indicating that this property is relevant to the reality issue. Indeed one can derive, for all quasiambivalent groups, a formalism involving some "conjugating" matrices which factorize in a suitable subgroup basis and which otherwise play a role for both real and complex irreps analogous to that of the $2jm$ matrices for real irreps. Within this formalism, which will be reported elsewhere,[46] the reality issue may be more elegantly addressed; especially for $6j$ symbols, $3jm$ factors, and the various unsymmetrized coefficients. Indeed the criterion (2) for reality can be reduced to a character test. However, other issues such as simultaneity are not as transparent as in the present treatment.

Note added in proof: The Tannaka-Krein duality theorem for compact groups may be used, at least in the case of unitary matrix irreps, to show that the sufficient condition (2) for reality is also necessary.[47]

ACKNOWLEDGMENTS

This work stems from a collaboration with Ture Damhus on the reality of the Wigner–Racah algebra and I am deeply indebted to him for numerous helpful communications. I would also like to thank Phil Butler and Ric Haase for many communications on Schur–Weyl duality and Jack Sullivan for some useful comments on the original manuscript. This work was supported in part by the U. S. Department of Energy.

[1]J. -R. Derome and W. T. Sharp, J. Math. Phys. **6**, 1584 (1965).
[2]P. H. Butler, Philos. Trans. R. Soc. London, Ser. A **277**, 545 (1975).
[3]H. A. Jahn in an addendum to H. A. Jahn and H. van Wieringen, Ref. 13.
[4]P. H. Butler, *Point Group Symmetry Applications: Methods and Tables* (Plenum, New York, 1981).
[5]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **5**, 1723 (1964); S. I. So and D. Strottman, J. Math. Phys. **20**, 153 (1979).
[6]See for example M. Rotenberg, R. Bivins, N. Metropolis, and J. K. Wooten, *The 3j and 6j symbols* (MIT, Cambridge, MA, 1959); Y. A. Smorodinskii and L. A. Shelepin, Usp. Fiz. Nauk **106**, 3 (1972) [Sov. Phys. Usp. **15**, 1 (1972)].
[7]U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic, New York, 1959).
[8]R. P. Bickerstaff and T. Damhus, "A necessary and sufficient condition for the existence of real coupling coefficients for a finite group," Int. J. Quantum Chem. (in press).
[9]G. Racah, Phys. Rev. **76**, 1352 (1949).
[10]P. H. Butler, in *Symmetries in Science, Proceedings of Einstein Centennial Celebration Science Symposium on Symmetries in Science*, Carbondale, Illinois, 1979, edited by B. Gruber and R. S. Millman (Plenum, New York, 1980), p. 89.
[11]I. Schur, "Uber eine Klasse von Matrizen die sich einer gegebenen Matrix Zuordnen Lassen," inaugural dissertation, Berlin, 1901.
[12]H. Weyl, *The Classical Groups. Their Invariants and Representations* (Princeton U. P., Princeton, NJ, 1946).
[13]See for example: H. A. Jahn, Proc. R. Soc. London, Ser. A **205**, 192 (1951); H. A. Jahn and H. van Wieringen, Proc. R. Soc. London, Ser. A **209**, 502 (1951); J. P. Elliott, J. Hope, and H. A. Jahn, Philos. Trans. R. Soc. London, Ser. A **246**, 241 (1953); H. A. Jahn, Phys. Rev. **96**, 989 (1954); Philos. Trans. R. Soc. London, Ser. A. **253**, 27 (1960); I. G. Kaplan, Zh. Eksp. Teor. Fiz. **41**, 560 (1961) [Sov. Phys. JETP **14**, 401 (1962)]; Zh. Eksp. Teor. Fiz. **41**, 790 (1961) [Sov. Phys. JETP **14**, 568 (1962)]; H. Horie, J. Phys. Soc. Jpn. **19**, 1783 (1964); V. V. Vanagas, *Algebraic Methods in Nuclear Theory* (in Russian) (Mintus, Vilnius, 1971).
[14]P. Kramer, Z. Phys. **205**, 181 (1967); **216**, 68 (1968).
[15]P. Kramer and T. H. Seligman, Nucl. Phys. A **136**, 545 (1969); P. Kramer and T. H. Seligman, Z. Phys. **219**, 105 (1969).
[16]J. J. Sullivan, J. Math. Phys. **14**, 387 (1973); **16**, 756 (1975); **16**, 1707 (1975); in *Proceedings of the International Symposium on Mathematical Physics*, Mexico City, 5–8 Jan. 1976, Vol. I, p. 253; J. Math. Phys. **19**, 1674 (1978); **19**, 1681 (1978); **21**, 227 (1980).
[17]J. -Q. Chen, T. -R. Yu, and F. Wang, Phys. Energi. Fort. Phys. Nuclear. **3**, 216 (1979); J. -Q. Chen, J. Math. Phys. **22**, 1 (1981); J. -Q. Chen, Y. -J. Shi, D. H. Feng, and M. Vallieres, Nucl. Phys. A **393**, 122 (1983); J. -Q. Chen, M. Vallieres, and D. H. Feng, Phys. Energi. Fort. Phys. Nuclear. **7**, 443 (1983); J. -Q. Chen, M. -J. Gao, Y. -J. Shi, M. Vallieres, and D. H. Feng, Nucl. Phys. A **419** 77 (1984).
[18]I. T. Obukhovsky, Yu. F. Smirnov, and Yu. M. Tchuvil'sky, J. Phys. A **15**, 7 (1982).
[19]R. S. Nikam, G. G. Sahasrabudhe, and C. R. Sarma, J. Phys. A **16**, 1591 (1983).
[20]L. J. Somers, J. Math. Phys. **24**, 772 (1983); "Coupling coefficients for special unitary groups," Nijmegen University report THEF-NYM-82.13, 1982.
[21]T. Regge, Nuovo Cimento **10**, 544 (1958); **11**, 116 (1959).
[22]F. G. Frobenius and I. Schur, Sitzungsber. K. Preuss. Akad. Wiss. (Berlin), 186 (1906); T. Damhus, Linear Algebra Appl. **32**, 125 (1980) and references therein.
[23]R. P. Bickerstaff, P. H. Butler, M. B. Butts, R. W. Haase, and M. F. Reid, J. Phys. A **15**, 1087 (1982).
[24]R. P. Bickerstaff and B. G. Wybourne, J. Phys. G **7**, 272 (1981); **7**, 995 (E) (1981).
[25]R. P. Bickerstaff, G. C. Joshi, P. H. Butler, and R. W. Haase, Lett. Nuovo Cimento **34**, 348 (1982).
[26]R. W. Haase, "The Symmetric Group and the Unitary Group: An Application of Group-Subgroup Transformation Theory," Ph.D. thesis, University of Canterbury, 1982.
[27]R. W. Haase, "Algebraic formulas for SU_n 6*j* symbols," University of Canterbury report, 1982.
[28]J. J. Sullivan, J. Math. Phys. **24**, 424 (1983); **24**, 2542 (1983).
[29]T. Damhus, J. Math. Phys. **22**, 7 (1981).
[30]D. E. Littlewood and A. R. Richardson, Philos. Trans. R. Soc. London, Ser. A **233**, 99 (1934).
[31]G. D. James and A. Kerber, *The Representation Theory of the Symmetric Group, Encyclopedia of Mathematics and its Applications*, Vol. 16, (Addison-Wesley, Reading, MA, 1981).
[32]J. Abramsky and R. C. King, Nuovo Cimento A **67**, 153 (1970); R. C. King, J. Math. Phys. **11**, 280 (1970).
[33]G. R. E. Black, R. C. King, and B. G. Wybourne, J. Phys. A **16**, 1555 (1983).
[34]D. E. Littlewood, J. London Math. Soc. **11**, 49 (1936); *The Theory of Group Characters*, 2nd ed. (Oxford U. P., Oxford, 1958).
[35]D. E. Littlewood, Proc. London Math. Soc. (2) **40**, 370 (1936); (2)**43**, 226 (1937).
[36]J. -R. Derome, J. Math. Phys. **7**, 612 (1966); "Foundations of Racah algebra for a general group," Ph. D. thesis, University of Toronto, 1965.
[37]A. J. van Zanten and E. de Vries, J. Algebra **25**, 475 (1973); Can. J. Math. **27**, 528 (1975).
[38]J. -R. Derome, J. Math. Phys. **8**, 714 (1967).
[39]P. H. Butler and R. C. King, Can. J. Math. **26**, 328 (1974).
[40]J. -R. Derome and G. Jackimow, Can. J. Phys. **48**, 2169 (1970).
[41]R. W. Haase and P. H. Butler, J. Phys. A **17**, 47 (1984).
[42]R. W. Haase and P. H. Butler, J. Phys. A **17**, 61 (1984).
[43]R. C. King, J. Math Phys. **12**, 1588 (1971).
[44]W. T. Sharp, L. C. Biedenharn, E. de Vries, and A. J. van Zanten, Can. J. Math. **27**, 246 (1975).
[45]G. W. Mackey, Pac. J. Math. **8**, 503 (1958).
[46]R. P. Bickerstaff and T. Damhus, "On the reality of the Wigner–Racah algebra I. Quasiambivalent groups," University of Washington Report, 1984; "On the reality of the Wigner–Racah algebra II. Conjugating relations for 3*jm* and 6*j* symbols," University of Washington Report, 1984.
[47]R. P. Bickerstaff and T. Damhus, "Why and when are Clebsch–Gordon coefficients real?" in *Proceedings of the Thirteenth International Colloquium on Group Theoretical Methods in Physics*, Maryland, 21–25 May 1984, edited by W. W. Zachary (to be published).

Collectivity and geometry. III. The three-dimensional case in the $Sp(6) \supset Sp(2) \times \mathcal{O}(3)$ chain for closed shells

O. Castaños
Centro de Estudios Nucleares, UNAM, Apartado Postal 70543, México, D.F. 04510 Mexico

E. Chacón and M. Moshinsky[a)]
Instituto de Física, UNAM, Apartado Postal 20364, México, D.F. 01000 Mexico

(Received 27 July 1983; accepted for publication 16 December 1983)

As was indicated in previous papers, general Hamiltonians for systems of n particles in three-dimensional space can be formulated in the enveloping algebra of the symplectic group $Sp(6n)$. This group admits, among others, the subgroup $Sp(6) \times O(n)$ and, as has been noticed by many authors, collective Hamiltonians can be formulated in the enveloping algebra of Sp(6), so that their eigenstates can be characterized by a definite irreducible representation (irrep) of this group. The mathematical problem is then to determine the matrix elements of the generators of Sp(6) in a basis characterized by irreps of this group as well as of appropriate subgroups. In the present series of papers the subgroups chosen where $Sp(2) \times \mathcal{O}(3)$ as the Casimir operator of Sp(2) when $n \to \infty$ is formally related to the Bohr–Mottelson vibrational Hamiltonian (BMVH), while $\mathcal{O}(3)$ gives the angular momentum of the state. We give an algorithm for determining these states that closely parallels the procedure followed for the BMVH. Programs are being developed to convert our algorithm into a computational tool for determining collective excitations in nuclei.

PACS numbers: 02.20. + b, 03.65.Fd, 21.60.Ev, 21.60.Fw

I. INTRODUCTION

As indicated in the two previous papers of this series[1,2] (to be referred as I and II and whose equations will be quoted by their numbers followed by I or II) our purpose is to develop the mathematical background required for dealing with microscopic collective models. We noted[1] that an n-body system in three-dimensional space has as dynamical group the $6n$-dimensional symplectic group $Sp(6n)$ which admits, among others, the subgroup $Sp(6) \times O(n)$. Collective effects are introduced by the constraint that our states are restricted to a given irreducible representation (irrep) $(\omega_1\omega_2\omega_3)$ of the orthogonal group in n dimensions $O(n)$, which implies also[1] the irrep

$$[n/2 + \omega_3, n/2 + \omega_2, n/2 + \omega_1] \tag{1.1}$$

for Sp(6) as indicated in Eq. (60I). Furthermore while a general n-body Hamiltonian can be defined[1] in the enveloping algebra of $Sp(6n)$, it seemed reasonable[1] to restrict the collective Hamiltonians to the enveloping algebra of Sp(6). Thus our mathematical problem is to determine the matrix elements of the generators of Sp(6) in a basis characterized by definite irreps of Sp(6) and appropriate subgroups of this group.

The above program is not new as, from different points of view, it has been discussed by Rosensteel and Rowe,[3] Biedenharn *et al.*,[4] Filippov,[5] Vanagas,[6] and others. What is new though is that the subgroup of Sp(6) that we choose for our discussion is

$$Sp(6) \supset Sp(2) \times \mathcal{O}(3), \tag{1.2}$$

and the reason that we do this, as indicated in detail in I, is because of the relation of this chain,[1] when $n \to \infty$, with the Bohr–Mottelson vibrational Hamiltonian.

It is with respect to the states characterized by the irreps of the chain (1.2) that we want to determine the matrix elements of the generators of Sp(6). As we shall indicate in the next section, this implies calculating the matrix elements of the quadrupole tensor

$$q_{ij} = \sum_{s=1}^{n} x_{is}x_{js}\,, \tag{1.3}$$

where x_{is}; $i = 1,2,3$; $s = 1,2,\ldots, n$ are the $3n$ coordinates of the n-body system in three dimensions, with respect to eigenstates of a finite system of partial differential equations in six variables related to the q_{ij}. This set of differential equations was explicitly given in Ref. 7, but we managed to solve it only when $n \to \infty$. To obtain the eigenstates for arbitrary but finite n is a hard problem, and in this paper we shall tackle it only for the irrep $(\omega_1\omega_2\omega_3) = (000)$ of $O(n)$, i.e., the scalar representation of this group. In this case we have only one partial differential equation in six variables. The problem still remains difficult, but we will give an algorithm for finding the eigenstates by using group theoretical techniques similar to those employed in relation with the exact solution[8,9] of the Bohr–Mottelson vibrational Hamiltonian. As our eigenstates will be in configuration space, the matrix elements of q_{ij} with respect to them are feasible and thus our program can be implemented.

We note though that the program carried out in this paper is restricted to the irrep (000) of $O(n)$, which implies [as was indicated in the considerations following Eq. (60I)] that, in general, it corresponds to the irrep

$$[n/2 + \omega, n/2 + \omega, n/2 + \omega] \tag{1.4}$$

of Sp(6) in which ω is an arbitrary integer. The irrep corresponds then to a doubly closed shell nucleus,[7] in which the nucleons fill the levels of an harmonic oscillator potential up to and including the level of $\mathcal{N}$ quanta, if we choose

$$n = \tfrac{2}{3}(\mathcal{N} + 1)(\mathcal{N} + 2)(\mathcal{N} + 3), \tag{1.5a}$$

[a)] Member of the Instituto Nacional de Investigaciones Nucleares and El Colegio Nacional.

 0022-2488/84/102815-11$02.50

$$\omega = \tfrac{1}{6}\mathcal{N}(\mathcal{N}+1)(\mathcal{N}+2)(\mathcal{N}+3). \tag{1.5b}$$

In (1.5a) we denoted by n the total number of nucleons. Had we considered the translationally invariant formalism of I in which n was related to the number of Jacobi vectors, i.e., eliminating the center of mass coordinates, we should have written $n+1$ instead of n on the left-hand side of (1.5a).

The analysis in this paper is then only relevant for doubly closed shell nuclei though in subsequent papers of this series we will extend it to the case of open shells, i.e., when the irrep of Sp(6) has the general form (1.1).

II. THE DIFFERENTIAL EQUATION FOR EIGENSTATES

We know[1] that the basis for irreps of Sp($6n$), with its generators expressed as bilinear forms in the coordinates x_{is} and their canonically conjugate momenta p_{is}, as in (4I), are given by the eigenstates of the harmonic oscillator whose Hamiltonian is

$$H = \frac{1}{2}\sum_{i=1}^{3}\sum_{s=1}^{n}(p_{is}^2 + x_{is}^2), \quad p_{is} = -i\frac{\partial}{\partial x_{is}}, \tag{2.1}$$

where we took units in which $\hbar$, the mass of the particles, and the frequency of the oscillator are taken as 1. The oscillator states of even (odd) number of quanta correspond to the irreps $[\frac{1}{2}^{3n}]$ ($[\frac{1}{2}^{3n-1}\,\frac{3}{2}]$) of Sp($6n$).

Is it possible to extract from the eigenstates of the Hamiltonian (2.1) the basis for the chain of groups $\mathrm{Sp}(6)\supset \mathrm{Sp}(2)\times\mathcal{O}(3)$? The answer is yes if we note that the symmetry group of H is the $3n$-dimensional unitary group U($3n$) and we can further characterize the states by the chain of subgroups

$$\underset{N}{\mathrm{U}(3n)}\supset \underset{\nu}{\mathrm{O}(3n)}\supset \underset{L}{\mathcal{O}(3)}\times \underset{(\omega_1\omega_2\omega_3)}{\mathrm{O}(n)}, \tag{2.2}$$

where under each one of the groups we have given its irrep, noting that those of U($3n$), O($3n$), and $\mathcal{O}(3)$ are characterized by a single quantum number while the one of O(n) is given, as indicated in the previous section, by the partition $(\omega_1\omega_2\omega_3)$.

We note now that the group Sp($6n$) has, among other chains of subgroups, the following ones:

$$\mathrm{Sp}(6n)\supset \mathrm{Sp}(6)\times\mathrm{O}(n), \tag{2.3a}$$

$$\mathrm{Sp}(6n)\supset \mathrm{Sp}(2)\times\mathrm{O}(3n). \tag{2.3b}$$

It is well known[1,10] that if the irrep of Sp($6n$) is $[\frac{1}{2}^{3n}]$ or $[\frac{1}{2}^{3n-1}\,\frac{3}{2}]$, the irreps of Sp(6) and O(n) are complementary, i.e., if we fix the irrep of O(n), that of Sp(6) is given and vice versa. Furthermore this complementary relation also holds[1,11] for Sp(2) and O($3n$). Thus if the eigenstates of H in (2.1) are characterized by the irreps of the groups in the chain (2.2), we have that $(\omega_1\omega_2\omega_3)$ of O(n) gives the irrep (1.1) of Sp(6), while the irrep of Sp(2) can be denoted by Λ related to the ν of O($3n$) through Eq. (28I), i.e.,

$$\Lambda = \nu/2 + \tfrac{3}{4}n, \tag{2.4}$$

where we note that here we use the notation ν, Λ for what was l, λ in Eq. (28I). The irrep N of U($3n$) gives the total number of quanta which, as indicated in (52I), characterize also the irrep of the subgroup O(2) of Sp(2). Finally we denote by L, as indicated in (2.2), the irrep of $\mathcal{O}(3)$.

Can we find a coordinate system in the $3n$-dimensional configuration space of the x_{is} in which it is possible to write the H of (2.1) so that its eigenfunctions are characterized by irreps of the chain of groups (2.2)? Again the answer is yes and in fact the coordinate system is the one introduced by Zickendraht[11] and by Dzublik *et al.*[12] which was discussed in Eq. (15I). For particles in three-dimensional space we can write

$$x_{is} = \sum_{k=1}^{3}\rho_k D^{1}_{ki}(\vartheta_j) D^{1}_{n-3+k,s}(\phi), \tag{2.5}$$

where the 3×3 matrix $\mathbf{D}^1 = \|D^1_{ki}(\vartheta_j)\|$ is the defining representation (which is the reason for the 1 appearing as an upper index) of the $\mathcal{O}(3)$ group, in terms of its three Euler angles ϑ_j. The $n\times n$ matrix $\|D^1_{ts}(\phi)\|$ has the same meaning for the $O(n)$ group and it depends on $n(n-1)/2$ angular coordinates denoted by ϕ. As in (2.3) we deal only with the last three rows of the matrix, the actual number of angles involved[1] is

$$\tfrac{1}{2}n(n-1) - \tfrac{1}{2}(n-3)(n-4) = 3n-6, \tag{2.6}$$

to which if we add the 3 ϑ_j and 3 ρ_k we get $3n$, i.e., the number of x_{is}, as we should.

From the orthogonal properties of the matrix $\|D^1_{ts}(\phi)\|$ we have that

$$\mathbf{q} = \|q_{ij}\| = \tilde{\mathbf{D}}^1(\vartheta_j)\|\rho_i^2\delta_{ij}\|\mathbf{D}^1(\vartheta_j), \tag{2.7}$$

where $\tilde{}$ indicates the transposed matrix. Thus the ρ_k's appearing in (2.3) are related through $\|\rho_i^2\delta_{ij}\|$ with the values of the quadrupole tensor (1.3) in the frame of reference fixed in the body, while $\mathbf{D}^1(\vartheta_j)$ is the orthogonal matrix that takes us from the frame of reference fixed in space to the one fixed in the body.

Furthermore we replace in (2.5) the ρ_k's by[13]

$$\rho_k^2 = \tfrac{1}{3}\rho^2\{1 + 2b\cos[c - (2\pi/3)k]\}, \quad k = 1,2,3, \tag{2.8}$$

where $0\leqslant\rho\leqslant\infty$ while b and c are restricted to the values in the hatched triangle in the plane $x = b\cos c$, $y = b\sin c$ of Fig. 1.

With the help of (2.5) and (2.8) we can rewrite the Hamiltonian (2.1), as was shown in Eqs. (3.2) and (3.9)–(3.11) of Ref. 7, in terms of ρ, b, c and their derivatives, the components of the angular momentum L'_k, $k = 1,2,3$ in the frame of reference fixed in the body and the generators $\mathscr{L}'_{st} = -\mathscr{L}'_{ts}$ of the O(n) group also in a kind of "body fixed" frame. If we restrict ourselves to a definite irrep $(\omega_1\omega_2\omega_3)$ of O(n) we can then substitute these $\mathscr{L}'_{st}$ by their matrix representation in a basis characterized by the irreps of the chain[7] $\mathrm{O}'(n)\supset\mathrm{O}'(n-1)\supset\mathrm{O}'(n-2)$, again in a "body fixed" frame, as was done in Sec. IV of Ref. 7. Thus for a fixed irrep $(\omega_1\omega_2\omega_3)$ of O(n) the Hamiltonian H of (2.1) can be expressed as a finite matrix of partial differential equation

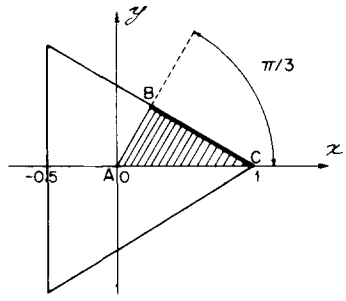


FIG. 1. The values of the variables b, c appearing in Eq. (2.8) are restricted to the lined triangle, where the coordinates are given by $x = b\cos c$, $y = b\sin c$. The points marked A, B, and C correspond to points with the same markings in Fig. 3 when we replace b, c by other variables.

functions of ρ, b, c and their derivatives as well as of L'_k, $k = 1,2,3$.

In the present paper we will be restricted to the scalar irrep of O(n), i.e., $(\omega_1\omega_2\omega_3) = (000)$ in which case we can replace all the matrices associated with the $\mathscr{L}'_{st}$ by zero, so that our matrix of partial differential equations becomes 1×1, i.e., a single equation. To obtain it we note that the oscillator Hamiltonian (2.1) is the same as H of (3.2) in Ref. 7, which is related to the H' of (3.9), also in Ref. 7, by

$$H = (\rho_1\rho_2\rho_3)^{-(n-3)/2}\rho^{-4}H'\rho^4(\rho_1\rho_2\rho_3)^{(n-3)/2}. \tag{2.9}$$

If we now put all $\mathscr{L}'_{st} = 0$ in H' we obtain from (3.10) and (3.11) of the same reference that

$$H = \frac{1}{2}\left(-\frac{1}{\rho^{3n-1}}\frac{\partial}{\partial\rho}\rho^{3n-1}\frac{\partial}{\partial\rho} + \frac{\mathscr{L}^2}{\rho^2} + \rho^2\right), \tag{2.10}$$

where

$$\begin{aligned}\frac{1}{2}\mathscr{L}^2 = &-(1 + b\cos 3c - 2b^2)\frac{\partial^2}{\partial b^2}\\ &-\frac{4}{b}\left(1 - \frac{1}{4}b\cos 3c - \frac{9}{4}b^2\right)\frac{\partial}{\partial b}\\ &-\frac{1}{b^2}(1 - b\cos 3c)\frac{\partial^2}{\partial c^2} - \frac{3}{b^2}\cot 3c\left(1\right.\\ &\left.+\frac{2}{3}b\sin 3c\tan 3c - b\sec 3c\right)\frac{\partial}{\partial c}\\ &+2b\sin 3c\,\frac{1}{b}\frac{\partial^2}{\partial b\,\partial c} + 3(n-3)b\frac{\partial}{\partial b}\\ &+\sum_{k=1}^{3}\frac{1 - bc_k}{4b^2 s_k^2}L'^2_k,\end{aligned} \tag{2.11}$$

where we used the shorthand notation

$$\begin{aligned}c_k &\equiv \cos[c - (2\pi/3)k],\\ s_k &\equiv \sin[c - (2\pi/3)k], \quad k = 1,2,3.\end{aligned} \tag{2.12}$$

As indicated in I, $\mathscr{L}^2$ is the Casimir operator of O($3n$), but now restricted to the scalar representation of its O(n) subgroup, and its eigenvalue, corresponding to the irrep ν of this group is[1]

$$\nu(\nu + 3n - 2). \tag{2.13}$$

The eigenstates associated with the irreps of the chain of groups (2.2), when $(\omega_1\omega_2\omega_3) = (000)$, can then be written as

$$\begin{aligned}|n; N\nu\sigma\tau LM\rangle = &\rho^{-(3n-1)/2}f^{\nu+(3n/2)-1}_{(N-\nu)/2}(\rho)\\ &\times\sum_K F^{N\nu\sigma\tau L}_K(b,c)D^L_{KM}(\vartheta_j).\end{aligned} \tag{2.14}$$

In (2.14) $f(\rho)$ can be obtained from (2.10) when we replace $\mathscr{L}^2$ by its eigenvalue (2.13), and its explicit form is given in terms of Laguerre polynomials in Eq. (20I). The $D^L_{KM}(\vartheta_j)$ are the standard Wigner functions[14] for the irreps of $\mathscr{O}(3)$. From (2.11) the $F_K(b,c)$, $K = L, L-1,\ldots,-L$ satisfy a set of $(2L+1)$ partial differential equations in b, c and their derivatives. An algorithm for determining the functions $F_K(b,c)$ is the main objective of this paper, and it will be implemented by a procedure entirely parallel to the one presented in Refs. 8 and 9 for the exact determination of the eigenstates of the Bohr–Mottelson vibrational Hamiltonian.

We note that the $F_K(b,c)$ have several upper indices which thus also appear in the ket (2.14). The number of particles is due to the fact that it appears as a parameter in the $\mathscr{L}^2$ of (2.11). The ν is due to the eigenvalue (2.13) of $\mathscr{L}^2$. The L, as well as the lower index K, is due to the matrix elements[8] of the L'^2_k, $k = 1,2,3$ in (2.11) with respect to the Wigner functions $D^L_{KM}(\vartheta_j)$. There are two more upper indices of $F_k(b,c)$, which we denoted by σ, τ, that are multiplicity indices required when we go from the chain O($3n$) to $\mathscr{O}(3)\times$O(n), keeping within the scalar representation of O(n). The explicit appearance of σ, τ will be seen in the following two sections.

Because of the complementarity relations discussed after (2.3), we note that the states (2.14) will also be the basis for irreducible representations of the chain of groups

$$\begin{array}{ccccc}\left[\frac{n}{2}\,\frac{n}{2}\,\frac{n}{2}\right] & & \Lambda & & N\\ & & \mathrm{Sp}(2) & \supset & O(2)\\ \mathrm{Sp}(6) & \supset & \times & & \\ & & \mathscr{O}(3) & \supset & \mathscr{O}(2)\\ & & L & & M\end{array}, \tag{2.15}$$

with the values of the irreps given above or below the groups and where Λ and ν are related by (2.4).

We note that the irrep of Sp(6) in (2.15) is not the one in (1.4) which, through (1.5), can be correlated with closed shell nuclei. We showed though in Sec. 9 of I, and particularly after (60I), that we can obtain the basis for the irrep (1.4) of Sp(6) by simply replacing

$$n \to n + 2\omega \tag{2.16}$$

in (2.14).

Once the kets (2.14) are explicitly available, in which we replace ν by Λ of (2.4) and n by $n + 2\omega$, our problem will be to find the matrix elements of q_{ij} of (1.3) with respect to them. As indicated in Eq. (30I) we can substitute the q_{ij} by

$$\begin{aligned}q_m = &(2/3)^{1/2}\rho^2\{(1/\sqrt{2})b\sin c[D^2_{2m}(\vartheta_j) + D^2_{-2m}(\vartheta_j)]\\ &+ b\cos cD^2_{0m}(\vartheta_j)\},\end{aligned} \tag{2.17a}$$

$$\bar{q} = \rho^2, \tag{2.17b}$$

where $m = 2,1,0,-1,-2$. Integrals involving ρ, ϑ_j, $j = 1,2,3$ are elementary so only the part involving the b, c need concern us and we shall indicate a possible analogy in their calculation with procedures used in Ref. 9 for determining the matrix elements of α_m with respect to the eigenstates of the Bohr–Mottelson vibrational Hamiltonian.

In the next two sections we outline the procedure for obtaining the kets (2.14).

III. POLYNOMIALS IN THE RAISING GENERATORS OF Sp(6) OF DEFINITE ANGULAR MOMENTUM

In paper I of this series we discussed the generators of Sp($6n$) and its subgroups in terms of the coordinates x_{is} and momenta p_{is}. For the analysis we are about to embark upon it is more convenient to discuss them using the creation and annihilation operators

$$\eta_{is} = (1/\sqrt{2})(x_{is} - ip_{is}), \tag{3.1a}$$

$$\xi_{is} = (1/\sqrt{2})(x_{is} + ip_{is}) \tag{3.1b}$$

in terms of which the generators of Sp(6) take the form[7]

$$B^\dagger_{ij} = \sum_{s=1}^{n} \eta_{is}\eta_{js}\,, \tag{3.2a}$$

$$C_{ij} = \frac{1}{2}\sum_{s=1}^{n} (\eta_{is}\xi_{js} + \xi_{js}\eta_{is}), \tag{3.2b}$$

$$B_{ij} = \sum_{s=1}^{n} \xi_{is}\xi_{js}\,, \tag{3.2c}$$

where $B^\dagger_{ij}$, C_{12}, C_{23}, C_{13} are the raising generators; C_{11}, C_{22}, C_{33} are the weight generators; and C_{21}, C_{32}, C_{31}, B_{ij} are the lowering generators of this group.

If we denote by

$$|0\rangle = \pi^{-(3n/4)} \exp(-\tfrac{1}{2}\rho^2) \tag{3.3}$$

the ground state of the oscillator Hamiltonian H of (2.1), we see that it is a state of lowest weight of Sp(6) as from (3.2b) and (3.2c) we observe that the lowering generators of this group, when applied to $|0\rangle$, give zero. Furthermore the weight generators acting on the ground state give the eigenvalue $n/2$, so $|0\rangle$ is the lowest weight state of an irrep of Sp(6) characterized by the partition given above this group in (2.15). It is clear then that the basis for the irrep $[n/2, n/2, n/2]$ of Sp(6) is given by

$$P_N(B^\dagger_{ij})|0\rangle, \tag{3.4}$$

where P_N are all possible homogeneous polynomials of degree $(N/2)$ in the $B^\dagger_{ij}$ where, from now on, N will be restricted to even values. Obviously the states (3.4) correspond to N quanta and thus to the irrep N of the O(2) subgroup of Sp(2) as was indicated in (2.15). Note that normally one would take P_N of (3.4) as a polynomial in all the raising generators which includes, besides the $B^\dagger_{ij}$, the C_{ij}, $i<j$. Using the commutation relations[7] of the generators (3.2) of Sp(6) we can though put all C_{ij}, $i<j$ to the right of the $B^\dagger_{ij}$, so that when acting on $|0\rangle$ they vanish leaving only the term (3.4).

In this section we wish also to characterize the polynomial in (3.4) by an angular momentum L and highest value $M = L$ for its projection, which implies

$$L_+ P_N(B^\dagger_{ij})|0\rangle = 0, \quad L_3 P_N(B^\dagger_{ij})|0\rangle = L P_N(B^\dagger_{ij})|0\rangle, \tag{3.5}$$

where L_i, $i = 1,2,3$ are the components of the total angular momentum and $L_\pm = L_1 \pm iL_2$.

To achieve our purpose we pass form the $B^\dagger_{ij'}$, $i, i' = 1,2,3$ in Cartesian components to $B^\dagger_{qq'}$ in spherical ones where $q, q' = 1,0,-1$, and then to the irreducible tensor form

$$B^\dagger_{lm} = \sum_{q,q'} \langle 1q,1q'|lm\rangle B^\dagger_{qq'}, \quad l = 0,2, \tag{3.6}$$

where, in what follows, we shall use the more compact notation

$$\overline{B}^\dagger = -\sqrt{3}B^\dagger_{00}, \quad B^\dagger_m = B^\dagger_{2m}, \quad m = 2,1,0,-1,-2. \tag{3.7}$$

The polynomial $P_N(B^\dagger_{ij})$ becomes then $P_N(\overline{B}^\dagger, B^\dagger_m)$ and as $\overline{B}^\dagger$ is a scalar with respect to the orthogonal group $\mathcal{O}(3)$ we can write the state satisfying (3.5) as

$$(\overline{B}^\dagger)^{(N-\nu)/2} P_{\nu L}(B^\dagger_m)|0\rangle, \tag{3.8}$$

where, from now on, ν will also be even and $P_{\nu L}$ will be a homogeneous polynomial of degree $(\nu/2)$ in the $B^\dagger_m$. It will only be in the next section, when we replace the boson creation operators η_{is} of (3.1a) by traceless bosons,[8] that ν will acquire the meaning it had in the previous section as the quantum number characterizing the irrep of the O($3n$) group.

We note that the $B^\dagger_m$, $m = 2,1,\ldots,-2$ functions only of the η_{is}, commute among themselves and are the components of a Racah tensor of order 2. The procedure of determining a state $P_{\nu L}(B^\dagger_m)$ satisfying (3.5), is then totally identical to that followed for the Bohr–Mottelson vibrational Hamiltonian in Sec. 3 of Ref. 8, where one wanted to obtain homogeneous polynomials $P(\eta_m)$, functions of the creation operators η_m, $m = 2,1,\ldots,-2$, that were also components of a Racah tensor of order 2. Following then the discussion of Ref. 8 we introduce some basic homogeneous polynomials denoted by (ν, L), that are of degree $(\nu/2)$ in $B^\dagger_m$ and satisfy (3.5), i.e.,

$$(2,2) = B^\dagger_2, \tag{3.9a}$$

$$(4,0) = \sum_{m=2}^{-2} (-)^m B^\dagger_m B^\dagger_{-m}\,, \tag{3.9b}$$

$$(4,2) = \sqrt{7}\sum_{mm'} \langle 2m,2m'|22\rangle B^\dagger_m B^\dagger_{m'}$$
$$= 2\sqrt{2}B^\dagger_0 B^\dagger_2 - \sqrt{3}(B^\dagger_1)^2, \tag{3.9c}$$

$$(6,0) = \sqrt{7}\sum_{mm'm''} (-)^{m''} \langle 2m,2m'|2,-m''\rangle B^\dagger_m B^\dagger_{m'} B^\dagger_{m''}\,, \tag{3.9d}$$

$$(6,3) = -\sqrt{\frac{14}{3}}\sum_{mm'm''m'''} \langle 2m,2m'|2m'''\rangle$$
$$\langle 2m''',2m''|33\rangle B^\dagger_m B^\dagger_{m'} B^\dagger_{m''}$$
$$= 2(B^\dagger_2)^2 B^\dagger_{-1} - \sqrt{6}B^\dagger_2 B^\dagger_1 B^\dagger_0 + (B^\dagger_1)^3. \tag{3.9e}$$

The polynomial $P_{\nu L}(B^\dagger_m)$ can be expressed as a product of powers of these basic homogeneous polynomials related to what are known as elementary permissible diagrams[15] (epd). The (6,3) appears in $P_{\nu L}(B^\dagger_m)$ only in the first power as its square is a function of the other epd as indicated in Eq. (3.21) of Ref. 8. Note incidentally that what is called here (ν,L) corresponds in Ref. 8 to $(\nu/2, L)$ as it is given there in terms of η_m rather than $B^\dagger_m$.

From Eq. (3.28) of Ref. 8 we see that the polynomials $P_{\nu L}(B^\dagger_m)$ require two extra indices, which in this paper we will denote by σ, τ, and thus for even L they take the form

$$P_{\nu L\sigma\tau}(B^\dagger_m)$$
$$= (2,2)^{L-(\nu/2)+2\sigma+3\tau}(4,2)^{(\nu-L)/2-2\sigma-3\tau}(4,0)^\sigma(6,0)^\tau, \tag{3.10a}$$

while for odd L we have

$$P_{\nu L\sigma\tau}(B^\dagger_m)$$
$$= (6,3)(2,2)^{L-(\nu/2)+2\sigma+3\tau}(4,2)^{(\nu-L-3)/2-2\sigma-3\tau}$$
$$\times(4,0)^\sigma(6,0)^\tau, \tag{3.10b}$$

We now write $\overline{B}^\dagger$ in an epd notation, i.e.,

$$(2,0) \equiv \overline{B}^\dagger = \sum_{i=1}^{3}\sum_{s=1}^{n} \eta_{is}\eta_{is}\,, \tag{3.11}$$

and see that a basis for the irrep $[n/2, n/2, n/2]$ of Sp(6) is given by the states

$$(2,0)^{(N-\nu)/2}P_{\nu L\sigma\tau}(B^{\dagger}_{m})|0\rangle, \tag{3.12}$$

where N, ν, L, σ, τ take all possible integer values (even only for N and ν) for which the exponents in (3.10) are nonnegative.

We note that the states (3.12) are characterized by the irreps $L, M = L$, and N of the subgroups $\mathcal{O}(3)$, $\mathcal{O}(2)$, and O(2) of Sp(6) appearing in the chain (2.15). They are *not* characterized by the irrep Λ of the subgroup Sp(2) or, equivalently, by the irrep of the O(3n) subgroup in the chain (2.2). To achieve this last objective we proceed, in analogy with what was done in Sec. 4 of Ref. 8, to replace the boson creation operators η_{is} by traceless boson operators $a^{\dagger}_{is}$.

IV. USE OF TRACELESS BOSON OPERATORS FOR DETERMINING STATES WITH GIVEN IRREP OF O(3n)

The eigenstates of the oscillator Hamiltonian (2.1) of N quanta will be homogeneous polynomials of degree N in the η_{is} acting on the ground state as they satisfy

$$\hat{N}P(\eta_{is})|0\rangle = NP(\eta_{is})|0\rangle, \tag{4.1}$$

where

$$\hat{N} = \sum_{s=1}^{n}\sum_{i=1}^{3}\eta_{is}\xi_{is} = \sum_{i=1}^{3}C_{ii} - \left(\frac{3n}{2}\right) \tag{4.2}$$

is the first-order Casimir operator of the U(3n) group.

If we want the $P(\eta_{is})$ to correspond to a definite irrep of the O(3n) subgroup of U(3n) they must satisfy the further condition[8] of being "harmonic" polynomials, i.e.,

$$\overline{B}P(\eta_{js})|0\rangle \equiv \sum_{i=1}^{3}\sum_{s=1}^{n}\xi_{is}\xi_{is}P(\eta_{js})|0\rangle = 0. \tag{4.3}$$

The states (3.12) satisfy (4.1) but not (4.3) and thus they do not correspond to a definite irrep of O(3n). There is, though, a method, originated by Vilenkin[16] and further developed by Lohe,[17] by which we can obtain polynomials based on the epd (3.9) that are "harmonic," and thus characterized by an irrep ν of O(3n), which implies that they also correspond to the irrep Λ of Sp(2) related to ν by (2.4).

This method is based on the introduction of traceless boson operators for the group O(3n) which, from Eq. (4.5) of Ref. 8, have in this case the form

$$a^{\dagger}_{is} = \eta_{is} - (2,0)(2\hat{N} + 3n)^{-1}\xi_{is}, \tag{4.4}$$

with (2,0) given by (3.11) and $\hat{N}$ by (4.2).

The traceless form of $B^{\dagger}_{ij}$ will now be denoted as $\beta^{\dagger}_{ij}$ which implies replacing in (3.2a) the η_{is} by $a^{\dagger}_{is}$ to get

$$\beta^{\dagger}_{ij} = B^{\dagger}_{ij} - (2,0)(2\hat{N} + 3n)^{-1}(C_{ij} + C_{ji}) + (2,0)^2(2\hat{N} + 3n + 4)^{-1}(2\hat{N} + 3n + 2)^{-1}B_{ij}, \tag{4.5}$$

where we made use of the definitions (3.2) and the commutation relations

$$(2\hat{N} + 3n)^{-1}\eta_{is} = \eta_{is}(2\hat{N} + 3n + 2)^{-1}, \tag{4.6a}$$

$$(2\hat{N} + 3n)^{-1}\xi_{is} = \xi_{is}(2\hat{N} + 3n - 2)^{-1}. \tag{4.6b}$$

As in the discussion preceeding (3.6), (3.7) we can replace $\beta^{\dagger}_{ij} = \beta^{\dagger}_{ji}$, $i, j = 1,2,3$ by their irreducible tensor form

$$\overline{\beta}^{\dagger} = \sum_{i=1}^{3}\beta^{\dagger}_{ii} = (2,0)^2(2\hat{N} + 3n + 4)^{-1}(2\hat{N} + 3n + 2)^{-1}\overline{B}, \tag{4.7a}$$

$$\beta^{\dagger}_{m} = B^{\dagger}_{m} - 2(2,0)(2\hat{N} + 3n)^{-1}Q_{m} + (2,0)^2(2\hat{N} + 3n + 4)^{-1}(2\hat{N} + 3n + 2)^{-1}B_{m}, \tag{4.7b}$$

where Q_m, B_m are related with $Q_{ij} \equiv \frac{1}{2}(C_{ij} + C_{ji})$, B_{ij} in the same form (3.6), (3.7) that $B^{\dagger}_{m}$ is related to $B^{\dagger}_{ij}$. It is then immediately clear that the states

$$P_{\nu L\sigma\tau}(\beta^{\dagger}_{m})|0\rangle \tag{4.8}$$

will not only satisfy equations (3.5), as well as (4.1) but now with eigenvalue ν instead of N, but also Eq. (4.3), i.e., they will be "harmonic" states and thus correspond to the irrep ν of O(3n). They can then be correlated with the kets (2.14) of the form

$$|n; N = \nu, \nu\sigma\tau L, M = L\rangle \equiv |n; \nu\sigma\tau L\rangle. \tag{4.9}$$

These kets are actually the most general we need to discuss as when calculating matrix elements of $\bar{q}$, q_m with respect to the states (2.14), the part involving the magnetic quantum numbers M, m, M' is given, because of the Wigner–Eckart theorem, by a standard Clebsch–Gordan coefficient. On the other hand, the dependence on N of the ket (2.14) is given through the function $f(\rho)$ appearing there, whose explicit form was derived in (30I).

In principle the state (4.8) can be determined as function of ρ, b, c and the Euler angles ϑ_j, $j = 1,2,3$ in the form (2.14) in which $M = L$ and $N = \nu$, where, from (30I), the dependence on ρ has the simple form $\rho^{\nu}\exp(-\rho^2/2)$. To obtain it explicitly we would need to use the commutation rules between $B^{\dagger}_{m}$, Q_m, B_m—that can be derived from those of $B^{\dagger}_{ij}$, C_{ij}, B_{ij} of (3.2) given in Eq. (A4) of Ref. 7—to order the polynomial (4.8) in the $\beta^{\dagger}_{m}$ of (4.7b) so that the $B^{\dagger}_{m}$ appear to the left of the Q_m and those in turn to the left of B_m. As B_m acting on $|0\rangle$ vanishes, while Q_m either vanishes or gives a constant eigenvalue, we see that the $P(\beta^{\dagger}_{m})|0\rangle$ becomes $\overline{P}(\overline{B}^{\dagger}, B^{\dagger}_{m})|0\rangle$, where the new polynomial is now a function of the $\overline{B}^{\dagger}$, $B^{\dagger}_{m}$ alone. From Dragt's theorem[9,18] we have now that

$$\overline{P}_{\nu L\sigma\tau}(\overline{B}^{\dagger}, B^{\dagger}_{m})|0\rangle = \pi^{-3n/4}2^{\nu/2}\overline{P}_{\nu L\sigma\tau}(\bar{q}, q_m)e^{-\rho^2/2}, \tag{4.10}$$

where $\bar{q}$, q_m are given by (2.17) from which we notice that $(\bar{q}/\rho^2) = 1$ and (q_m/ρ^2) is a function of b, c, ϑ_i only, so as the polynomial is homogeneous of degree $(\nu/2)$, the term ρ^{ν} mentioned above makes its appearance.

The procedure indicated in the previous paragraph, while feasible in principle, is cumbersome in practice. We shall proceed here to define our states in a slightly different fashion—parallel to the particle hole procedure developed in Ref. 9 for the Bohr–Mottelson vibrational Hamiltonian—and then discuss in Sec. V an algorithm for determining the states explicitly as functions of b, c when the angular momentum $L = 0$, while in Sec. VI we extend our analysis to arbitrary L.

For our alternative procedure, we first introduce an epd in traceless bosons which we shall denote by the square brackets

$$[\nu, L], \tag{4.11}$$

which are defined as in (3.9) but with $\beta^{\dagger}_{m}$ replacing $B^{\dagger}_{m}$. We also introduce an epd in the Cartesian annihilation operators ξ_{is}, $i = 1,2,3$; $s = 1,2,\ldots n$, by the definition

$$(\overline{2,2}) = B_2 = \frac{1}{2}\sum_{s=1}^{n}(\xi_{1s}^2 - \xi_{2s}^2) + i\sum_{s=1}^{n}\xi_{1s}\xi_{2s}\,, \quad (4.12)$$

which clearly corresponds to $L = M = 2$ and commutes with $\overline{B}$ appearing in (4.3).

We will now replace in (3.10) the ordinary boson epd's $(\nu\, L)$ by the traceless boson ones $[\nu, L]$ with the exception of [4,2] which we replace by $(\overline{2,2})$ of (4.12). The exponents have to change of course so that the polynomials acting on $|0\rangle$ still give a state of ν quanta and angular momentum L. We then get for L even that our state (4.9) is

$$|n; \nu L\sigma\tau) = [2,2]^{(\nu - 4\sigma - 6\tau + L)/4}(\overline{2,2})^{(L - \nu + 4\sigma + 6\tau)/4} \times [4,0]^{\sigma}[6,0]^{\tau}|0\rangle, \quad (4.13a)$$

while for L odd it becomes

$$|n; \nu L\sigma\tau) = [6,3][2,2]^{(\nu - 4\sigma - 6\tau + L - 9)/4} \times (\overline{2,2})^{(L - \nu + 4\sigma + 6\tau + 3)/4}[4,0]^{\sigma}[6,0]^{\tau}|0\rangle, \quad (4.13b)$$

where for fixed ν, L the σ, τ are restricted to values for which all exponents are nonnegative integers.

Note that the quantum numbers ν, L, and when they later appear N, M, are related with irreps of definite groups, so that states with different values of these quantum numbers are orthogonal. On the other hand σ, τ, restricted by inequalities coming from the fact that all exponents in (4.13) must be nonnegative integers, are multiplicity indices related to the chain $O(3n) \supset \mathcal{O}(3) \times O(n)$ when we have the irrep ν for $O(3n)$, L for $\mathcal{O}(3)$, and (000) for $O(n)$. Thus our set of states (4.13) while complete is not orthonormal with respect to the indices σ, τ though it can be made so in a variety of well-known ways.

We now proceed to consider the states (4.13) when the angular momentum $L = 0$.

V. STATES OF ZERO ANGULAR MOMENTUM

From Eq. (4.13a) we see that when $L = 0$ we have the kets

$$|n; \nu 0\sigma\tau) = [4,0]^{\sigma}[6,0]^{\tau}|0\rangle, \quad (5.1)$$

where

$$4\sigma + 6\tau = \nu, \quad (5.2)$$

with ν being the irrep of $O(3n)$ while σ, τ are nonnegative integers. Thus from (5.2) we see that for a fixed value of ν we have a finite number of states with $L = 0$.

We also observe that from (3.10a) and (3.12) the kets (5.1) can be written as

$$|n; \nu 0\sigma\tau) = \sum_{s,t}\overline{A}_{st}^{\sigma\tau}(2,0)^{(\nu/2) - 2s - 3t}(4,0)^{s}(6,0)^{t}\,|0\rangle, \quad (5.3)$$

as each monomial in the polynomial has $L = 0$ and the coefficients $\overline{A}_{st}^{\sigma\tau}$ can be chosen so that the states correspond to the irrep ν of $O(3n)$ which implies that

$$\overline{B}\,|n; \nu 0\sigma\tau) = 0. \quad (5.4)$$

Note that as all exponents in (5.3) must be nonnegative the s,t are restricted to nonnegative integers satisfying

$$4s + 6t \leqslant \nu. \quad (5.5)$$

From Dragt's theorem[9,18] we can, as indicated in (4.10), write the state (5.3) in configuration space as

$$|n; \nu 0\sigma\tau) = \pi^{-3n/4}2^{\nu/2} \times \sum_{s,t}\overline{A}_{st}^{\sigma\tau}\{2,0\}^{(\nu/2) - 2s - 3t}\{4,0\}^{s}\{6,0\}^{t}e^{-\rho^2/2}, \quad (5.6)$$

where $\{\nu, L\}$ are the epd (ν, L) in (3.9), (3.11) in which we replaced $B_m^{\dagger}$, $\overline{B}^{\dagger}$ by q_m, $\bar{q}$ of (2.17) so that

$$\{2,0\} = \rho^2, \quad (5.7a)$$

$$\{4,0\} = \tfrac{2}{3}\rho^4 b^2 = \tfrac{2}{3}\rho^4(1 - x), \quad (5.7b)$$

$$\{6,0\} = -(4/3\sqrt{3})\rho^6 b^3\cos 3c = -(2/3\sqrt{3})\rho^6(y - 3x + 2). \quad (5.7c)$$

In (5.7) we also wrote $\{\nu, L\}$ in terms of the variables x, y defined by

$$x = 1 - b^2, \quad (5.8a)$$

$$y = 1 - 3b^2 + 2b^3\cos 3c. \quad (5.8b)$$

We can now substitute in the ket (5.6) the σ, τ related by (5.2), by a single multiplicity index μ to denote the state as

$$|n; \nu\mu\} = \pi^{-3n/4}2^{\nu/2}\rho^{\nu}e^{-\rho^2/2}P_{\nu\mu}(x, y), \quad (5.9)$$

where the polynomial P has the form

$$P_{\nu\mu}(x, y) = \sum_{s,t}A_{st}^{\nu\mu}x^s y^t, \quad (5.10)$$

in which the $A_{st}^{\nu\mu}$ are of course different from $\overline{A}_{st}^{\sigma\tau}$ in (5.6), but the s,t are still restricted by the inequality (5.5).

We note that the ket(5.6) [or equivalently (5.9)] is an eigenstate of the operator $\mathscr{L}^2$ of (2.11) with eigenvalue (2.13). As this ket has $L = 0$ the matrix elements of the operators $L_k'^2$ with respect to it vanish so that we can replace $\mathscr{L}^2$ by $\mathscr{L}_0^2$ defined by (2.11) but now without the last term that depends on $L_k'^2$. As $\mathscr{L}_0^2$ is a function of b, c and their derivatives it acts only on the $P_{\nu\mu}(x, y)$ appearing in (5.9), so that the polynomial must satisfy

$$\mathscr{L}_0^2 P_{\nu\mu}(x, y) = \nu(\nu + 3n - 2)P_{\nu\mu}(x, y), \quad (5.11)$$

where from (2.11) and (5.8) we can write

$$\frac{1}{2}\mathscr{L}_0^2 = (8x^2 - 6x - 2y)\frac{\partial^2}{\partial x^2} + 18y(y - x)\frac{\partial^2}{\partial y^2} + 24y(x - 1)\frac{\partial^2}{\partial x\,\partial y} + [(6n + 4)x - (6n - 6)]\frac{\partial}{\partial x} + [(9n + 12)y - 9(n - 2)x]\frac{\partial}{\partial y}. \quad (5.12)$$

If we take the explicit polynomial expression (5.10) for $P_{\nu\mu}(x, y)$ we get, from (5.11) and (5.12), the following relation between the coefficients:

$$A_{s-1,t+1}^{\nu\mu} = \frac{(4s + 6t - \nu)(\nu + 3n - 2 + 4s + 6t)}{18(t + 1)(n - 2 + 2t)}A_{st}^{\nu\mu} - \frac{2(s + 1)(s + 4t + n - 1)}{3(t + 1)(n + 2t - 2)}A_{s+1,t}^{\nu\mu} - \frac{2(s + 1)(s + 2)}{9(t + 1)(n + 2t - 2)}A_{s+2,t-1}^{\nu\mu}. \quad (5.13)$$

If for a fixed ν we denote by d_{ν} the number of nonnegative integer values of s,t that satisfy the inequalities (5.5), we

see that this is the number of coefficients $A_{st}^{\nu\mu}$ we have to consider and for which from (5.13) we have a homogeneous system of d_ν linear equations. If the determinant of this system of equations is different from zero, i.e., if its rank r_ν is equal to d_ν, then there are no states of $L = 0$ for the given ν. If the determinant is zero and of rank r_ν, then $d_\nu - r_\nu$ will give the number of independent solutions for the coefficients $A_{st}^{\nu\mu}$, which we can distinguish by giving to μ the values

$$\mu = 1,2,..., d_\nu - r_\nu . \tag{5.14}$$

We arrive then at the conclusion that to determine the states (5.9) of angular momentum $L = 0$, all we need is to solve the system of linear equations (5.13) in which s,t are nonnegative integers restricted by (5.5). Furthermore we note that the states of $L = 0$ corresponding to a given ν are also given by (5.1) in which σ, τ satisfy (5.2). Thus $d_\nu - r_\nu$ is equal to the number of pairs of nonnegative integers (σ, τ) for which $4\sigma + 6\tau = \nu$ so that, for example, if $\nu = 18$, $(\sigma, \tau) = (0,3),(3,1)$ and in this case the number of independent solutions of the set of linear equation (5.13) is 2.

In Fig. 2 we indicate graphically how the coefficients $A_{st}^{\mu\nu}$ are related, which may provide ladder procedures for computational programs.

The states $|n; \nu\mu\}$ are not orthonormal in the index μ, so we would have to compute the scalar product $\{n; \nu\mu|n; \nu\mu'\}$ to find linear combinations of $|n; \nu\mu\}$ with that property. If our states with $L = 0$ had been given in terms of the variables ρ, b, c as in (2.14), the scalar products would require the volume element in those variables which has the form

$$\rho^{3n-1}\, d\rho (1 - 3b^2 + 2b^3 \cos 3c)^{(n-4)/2} b^4\, db \sin 3c\, dc , \tag{5.15}$$

with b, c restricted to the lined triangle in Fig 1. On the other hand if we use the variables ρ, x, y as in (5.9), then from (5.8) the volume element becomes

$$\rho^{3n-1}\, d\rho\, y^{(n-4)/2}\, dx\, dy . \tag{5.16}$$

We have to be careful though with the domain of integration which is drawn in Fig. 3. The upper line of the triangle in Fig. 1 corresponds to the line with $y = 0$ in Fig. 3, while the lines with $c = 0$, $\pi/3$ in Fig. 1 correspond to the upper and lower curved lines in Fig. 3 whose equations are given by

$$y = u_\pm , \quad u_\pm = 3x - 2 \pm 2(1 - x)^{3/2}. \tag{5.17}$$

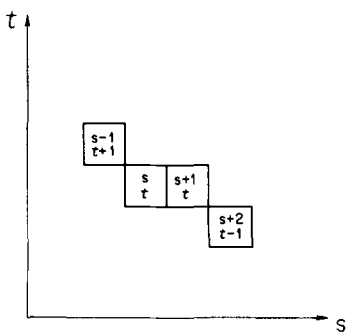


FIG. 2. The figure shows diagramatically how the coefficients A_{st} of (5.13) are connected. Each block has the values (s,t) of the coefficient with which it is related and only those associated with the blocks drawn are connected with coefficients appearing in (5.13).

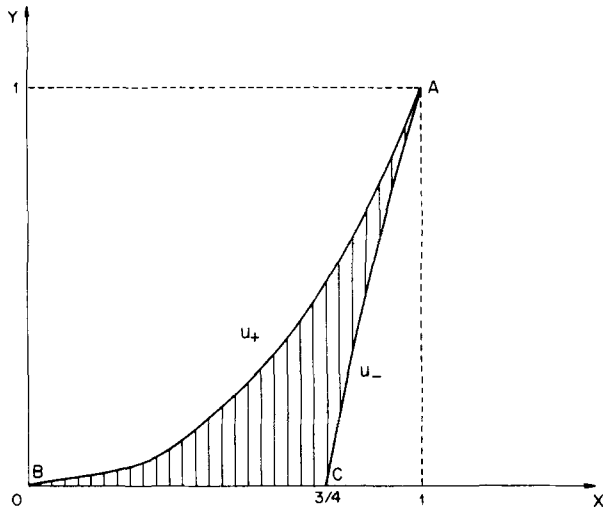


FIG. 3. The lined triangle of Fig. 1, to which allowed values of b, c are restricted, becomes the lined part of Fig. 3 in which the variables are $x = 1 - b^2, y = 1 - 3b^2 + 2b^3 \cos 3c$. The curves limiting the region are the $u_\pm$ of (5.17), while the points A, B, and C of Fig. 3 are the images of the points A, B, and C of Fig. 1.

Thus the relevant integrals for our scalar product take the form

$$\int_0^{3/4}\int_0^{u_+} P_{\nu\mu'}(x,y) P_{\nu\mu}(x,y)\, y^{(n-4)/2}\, dy\, dx + \int_{3/4}^{1}\int_{u_-}^{u_+} P_{\nu\mu'}(x,y) P_{\nu\mu}(x,y)\, y^{(n-4)/2}\, dy\, dx , \tag{5.18}$$

and they can be evaluated in an elementary fashion once the $A_{st}^{\nu\mu}$ of the polynomials $P_{\nu\mu}(x,y)$ of (5.10) have been determined.

Having obtained an algorithm for deriving the states characterized by irreps in the chain (2.15) when $L = 0$, we now turn our attention to the problem of arbitrary L.

VI. STATES OF ARBITRARY ANGULAR MOMENTUM AND MATRIX ELEMENTS OF THE GENERATORS OF Sp(6)

We start our discussion by considering the states with L even given by (4.13a). We replace in them the part of the type (5.1) of angular momentum zero, with the type (5.3) but now defined along the lines (5.7)–(5.10). We then have that we can write the state of L even as

$$|n; \nu L\lambda\, \mu) = [2,2]^{(\nu-\lambda+L)/4} (\overline{2,2})^{(-\nu+\lambda+L)/4} P_{\lambda\mu}|0\rangle, \tag{6.1}$$

where

$$P_{\lambda\mu} = \sum_{s,t} \overline{A}_{st}^{\lambda\mu} (2,0)^{(\lambda/2)-2s-3t} (4,0)^s (6,0)^t , \tag{6.2}$$

with the coefficients $\overline{A}_{st}^{\lambda\mu}$ obtained from the $A_{st}^{\lambda\mu}$ in (5.9) and (5.10) when we replace ν by λ and ρ^2, x, y by $\{2,0\}, \{4,0\}, \{6,0\}$ through the relations (5.7). We already indicated in Sec. V the system of linear algebraic equations (5.13) that the $A_{st}^{\lambda\mu}$ satisfy, from which we get $d_\lambda - r_\lambda$ independent sets of coefficients characterized by $\mu = 1,2,..., d_\lambda - r_\lambda$. Thus the $\overline{A}_{st}^{\lambda\mu}$ appearing in (6.2) are well defined, with the even λ restricted by the condition that for

fixed even ν, L the exponents in (6.1) should be nonnegative integers.

We now notice that $[2,2]=\beta_2^\dagger$ is given by (4.7b) when $m=2$. We see from (3.2) that $B_2^\dagger$, $(2,0)=\bar{B}^\dagger$ are quadratic expressions in the creation operators η_{is}, while B_2 is quadratic in the annihilation operators ξ_{is}, and $\hat{N}$, Q_2 are mixed, i.e., one annihilation and one creation operator. The traceless boson operator [2,2] appears in (6.1) to a certain power and so it would be convenient to expand it in normal form so that powers of $B_2^\dagger$, $(2,0)$, $\hat{N}$, Q_2, B_2 appear in this order where the actual dependence on $\hat{N}$ is more complicated than that of a power. This type of ordering is entirely similar to the one for powers of the traceless boson operator $a_m^\dagger$ that was given in Eq. (5.5) of Ref. 9 for the Bohr–Mottelson vibrational Hamiltonian. Using then the binomial theorem for powers of $\beta_2^\dagger$ and the commutation relations between the generators (3.2) of Sp(6) one can write

$$[2,2]^u=\sum_{v=0}^{u}\sum_{w=u-v}^{2u-2v}(2,2)^v(2,0)^w \times R^{\,u}_{vw}(\hat{N})Q_2^{\,2u-2v-w}(\overline{2,2})^{v+w-u}, \tag{6.3}$$

with v, w restricted by the fact that all exponents have to be nonnegative integers.

We note from (4.7b) that for $u=1$ the only terms $R^{\,u}_{vw}(\hat{N})$ that we have are

$$R^{\,1}_{10}(\hat{N})=1,\quad R^{\,1}_{01}(\hat{N})=-2(2\hat{N}+3n)^{-1},$$
$$R^{\,1}_{02}=[(2\hat{N}+3n+4)(2\hat{N}+3n+2)]^{-1}. \tag{6.4}$$

Furthermore in the Appendix we derive the recursion relation connecting

$$R^{\,u+1}_{v'w'}(\hat{N})\quad\text{with}\quad R^{\,u}_{vw}(\hat{N}), \tag{6.5}$$

so from (6.4) and (6.5) we have an algorithm to derive all the $R^{\,u}_{vw}(\hat{N})$ that we require.

We note now that in (6.2) we need to apply powers of [2,2] and $(\overline{2,2})$ to $P_{\lambda\mu}|0\rangle$. From (6.3) this implies that we have first to discuss the effect of $(\overline{2,2})$ and Q_2 on $P_{\lambda\mu}|0\rangle$ as powers of (2,2), (2,0) have a purely multiplicative action. To find this effect it is very convenient to write the epd (ν, L) as well as the operators $(\overline{2,2})$, Q_2 in terms of creation and annihilation opertors in spherical rather than Cartesian components, i.e.,

$$\eta_{\pm 1s}=\mp(1/\sqrt{2})(\eta_{1s}\pm i\eta_{2s}),\quad \eta_{0s}=\eta_{3s}, \tag{6.6a}$$

$$\xi_s^{\pm 1}=\mp(1/\sqrt{2})(\xi_{1s}\mp i\xi_{2s}),\quad \xi_s^{0}=\xi_{3s}, \tag{6.6b}$$

which have the commutation relations $[\xi_s^q,\eta_{q's'}]=\delta_{q'}^{q}\delta_{ss'}$ so that ξ_s^q acts as $\partial/\partial\eta_{qs}$. Furthermore, lowering the index q implies

$$\xi_{qs}=(-)^q\xi_s^{-q}, \tag{6.7}$$

and we shall use the scalar product notation

$$(\eta_q\cdot\eta_{q'})\equiv\sum_{s=1}^{n}\eta_{qs}\eta_{q's}. \tag{6.8}$$

We now introduce instead of the round bracket epd (ν, L) of (3.9), the angular ones $\langle\nu, L\rangle$ that will be functions of them and which will have a simpler form in terms of the scalar products (6.8) so as to get

$$\langle 2,2\rangle=(2,2)=(\eta_1\cdot\eta_1), \tag{6.9a}$$
$$\langle 2,0\rangle=(2,0)=-2(\eta_1\cdot\eta_{-1})+(\eta_0\cdot\eta_0), \tag{6.9b}$$
$$\langle 4,2\rangle=(1/2\sqrt{3})(4,2)+\tfrac{1}{3}(2,2)(2,0)$$
$$=(\eta_1\cdot\eta_1)(\eta_0\cdot\eta_0)-(\eta_1\cdot\eta_0)^2, \tag{6.9c}$$
$$\langle 4,0\rangle=\tfrac{1}{2}(4,0)-\tfrac{1}{3}(2,0)^2=2(\eta_1\cdot\eta_{-1})(\eta_0\cdot\eta_0)$$
$$+(\eta_1\cdot\eta_1)(\eta_{-1}\cdot\eta_{-1})-(\eta_1\cdot\eta_{-1})^2$$
$$-2(\eta_1\cdot\eta_0)(\eta_0\cdot\eta_{-1}), \tag{6.9d}$$
$$\langle 6,0\rangle=(1/6\sqrt{3})(6,0)+\tfrac{1}{6}(2,0)(4,0)-\tfrac{1}{27}(2,0)^3$$
$$=\sum_q\sum_{q'}\sum_{q''}\epsilon_{qq'q''}(\eta_1\cdot\eta_q)(\eta_0\cdot\eta_{q'})(\eta_{-1}\cdot\eta_{q''}), \tag{6.9e}$$
$$\langle 6,3\rangle=-(1/\sqrt{8})(6,3)$$
$$=(\eta_1\cdot\eta_0)[(\eta_1\cdot\eta_1)(\eta_0\cdot\eta_0)-(\eta_1\cdot\eta_0)^2]$$
$$-(\eta_1\cdot\eta_1)[(\eta_1\cdot\eta_1)(\eta_{-1}\cdot\eta_0)$$
$$-(\eta_1\cdot\eta_{-1})(\eta_1\cdot\eta_0)]. \tag{6.9f}$$

On the other hand the operators $(\overline{2,2})$ and Q_2 that act on the epd $\langle\nu, L\rangle$ have the form

$$(\overline{2,2})=(\xi_1\cdot\xi_1)=(\xi^{-1}\cdot\xi^{-1})\to\sum_{s=1}^{n}\frac{\partial^2}{\partial\eta_{-1s}^2}, \tag{6.10a}$$
$$Q_2=(\eta_1\cdot\xi_1)=-(\eta_1\cdot\xi^{-1})\to-\sum_{s=1}^{n}\eta_{1s}\frac{\partial}{\partial\eta_{-1s}}, \tag{6.10b}$$

with the arrows indicating that when these operators act on a state $P(\eta_{qs})|0\rangle$ we can apply the differential expressions to $P(\eta_{qs})$ and then the resulting polynomial to $|0\rangle$.

We note from (6.10) that the operators Q_2 and $(\overline{2,2})$ commute and furthermore from (6.9a) and (6.9c) that

$$[(\overline{2,2}),\langle 2,2\rangle]=0, \tag{6.11a}$$
$$[(\overline{2,2}),\langle 4,2\rangle]=0, \tag{6.11b}$$
$$[Q_2,\langle 2,2\rangle]=0, \tag{6.11c}$$
$$[Q_2,\langle 4,2\rangle]=0. \tag{6.11d}$$

To see then what is the effect of $(\overline{2,2})$ and Q_2 on $P_{\lambda\mu}|0\rangle$ we need to evaluate the action of these operators on products of powers of (2,2), (4,0), and (6,0) or, from (6.9), on those of $\langle 2,0\rangle$, $\langle 4,0\rangle$, and $\langle 6,0\rangle$. From the analysis given in the Appendix we see that

$$(\overline{2,2})\langle 2,0\rangle^r\langle 4,0\rangle^s\langle 6,0\rangle^t$$
$$=\langle 2,2\rangle\{4r(r-1)\langle 2,0\rangle^{r-2}\langle 4,0\rangle^s\langle 6,0\rangle^t$$
$$+2s(n+2s+4t-3)\langle 2,0\rangle^r\langle 4,0\rangle^{s-1}\langle 6,0\rangle^t\}$$
$$+\langle 4,2\rangle\{4s(s-1)\langle 2,0\rangle^{r+1}\langle 4,0\rangle^{s-2}\langle 6,0\rangle^t$$
$$+t(2n+4t-8)\langle 2,0\rangle^r\langle 4,0\rangle^s\langle 6,0\rangle^{t-1}$$
$$-8rs\langle 2,0\rangle^{r-1}\langle 4,0\rangle^{s-1}\langle 6,0\rangle^t\}, \tag{6.12a}$$
$$Q_2\langle 2,0\rangle^r\langle 4,0\rangle^s\langle 6,0\rangle^t$$
$$=\langle 2,2\rangle\{2r\langle 2,0\rangle^{r-1}\langle 4,0\rangle^s\langle 6,0\rangle^t\}$$
$$-\langle 4,2\rangle\{2s\langle 2,0\rangle^r\langle 4,0\rangle^{s-1}\langle 6,0\rangle^t\}. \tag{6.12b}$$

Note from (6.11) that the application of $(\overline{2,2})$, Q_2 to the right-hand side of (6.12) can again be calculated with the help

of (6.12). Thus we have an algorithm for applying the operator

$$[2,2]^{(\nu-\lambda+L)/4}(\overline{2,2})^{(-\nu+\lambda+L)/4} \tag{6.13}$$

to $P_{\lambda\mu}|0\rangle$ as required in (6.1) to get a polynomial $\overline{P}(\overline{B}^{\dagger}, B^{\dagger}_{m})$ function of the epd (6.9) acting on the ground state. By Dragt's theorem we can replace $\overline{B}$, $B^{\dagger}_{m}$ by $\bar{q}$, q_m, thus obtaining a polynomial in the epd $\{\nu, L\}$ that has to be multiplied by $\exp(-\rho^2/2)$ as indicated in (5.9). Thus we have a solution of the problem for even L in configuration space that can be reduced to the form (2.14) by the same steps followed in Sec. 3 of Ref. 9.

For odd L we have, as indicated in (4.13b), the extra traceless boson [6,3]. Its effect can be taken into account by a procedure entirely parallel to the one followed in Sec. 8 of Ref. 9 in the case of the Bohr–Mottelson vibrational Hamiltonian.

Thus we have an algorithm for deriving the states (2.14), i.e., getting the $F_K(b, c)$ that appear there. We intend to make available in the near future programs that allow us to obtain explicit analytic expressions for the $F_K(b, c)$ and with their help to calculate the matrix elements of the generators $\bar{q}$, q_m of Sp(6) given by (2.17). The other generators of Sp(6) can be obtained, as shown in (8I), by taking commutators of $\bar{q}$, q_m with the H of (2.1), so that their matrix elements are immediately available if those of $\bar{q}$, q_m are known.

We have thus a full algorithm for implementing the program outlined in paper I of this series when we consider a three-dimensional space but restrict ourselves to double closed shells for which, from the discussion at the end of Sec. I, we have to replace in our results of all the other sections n by $n + 2\omega$ and give to n and ω the values (1.5).

VII. CONCLUSION

We have discussed in detail the problem of determining the matrix elements of the generators of Sp(6) with respect to states characterized by irreps of the chain of subgroups (2.15) of this group, when the irrep of Sp(6) is (1.4), i.e., when it corresponds to the problem of doubly closed shell nuclei. While this problem is physically interesting, it is related in the two-dimensional case[2] just to the part discussed in Sec. 3 of II. The general case of two dimensions, analyzed in Secs. 4 and 5 of II is much more complex,[2] and we can expect that the difficulty of the problem will increase considerably when we go to three dimensions.

Several possibilities seems to exist for the extension of our analysis to the general irrep (1.1) of Sp(6), i.e., to the case of open shells. One could proceed, as in Sec. 4 of II, by finding the epd in the chain $U(3n) \supset \mathcal{O}(3) \times O(n)$ of (2.2) but now for a general irrep $(\omega_1\omega_2\omega_3)$ of $O(n)$ for states highest weight in this group. This procedure is feasible, particularly as it can be systematized by the powerful generating function techniques developed by Sharp[19] and his collaborators. The procedure may become very laborious though when we replace the epd by traceless ones and try to order powers of the latter in the normal form as was done in Secs. V and VI of this paper.

Another approach is to follow Filippov,[5] Rowe and Rosensteel,[7] and others in considering that collective states for open shells can be obtained by first determining the Slater determinant for the ground state and applying to it polynomials in the collective creation operators of the type (3.10) or (4.8). The Slater determinant can be obtained by filling compactly the levels of a harmonic oscillator in such a way that the intrinsic state corresponds[5,7] to one of highest weight for an irrep $[\omega_1\omega_2\omega_3]$ of U(3). From this state we can project out[5] ones of definite angular momentum L'' and projection M'' that can be denoted by the ket

$$\left|\begin{matrix}\omega_1\omega_2\omega_3 \\ \Omega L''M''\end{matrix}\right\rangle, \tag{7.1}$$

where Ω is the multiplicity index[20] that distinguishes between repeated irrep L'' of $\mathcal{O}(3)$ in the irrep $[\omega_1\omega_2\omega_3]$ of U(3).

From the way the ket (7.1) was constructed it is clear that when $\overline{B}$ of (4.3) is applied to this ket it vanishes, as otherwise we would have a state of two quanta less that satisfies the Pauli principle. Thus (7.1) is the lowest weight state of an irrep $\frac{1}{2}(\omega_1 + \omega_2 + \omega_3) + (3n/4)$ of Sp(2) as can be seen from the eigenvalue of the weight operator I_3 in (14I). From (2.4) the corresponding irrep of $O(3n)$ is $(\omega_1 + \omega_2 + \omega_3)$.

We can now define the ket

$$\left|\left[\frac{n}{2}+\omega_3, \frac{n}{2}+\omega_2, \frac{n}{2}+\omega_1\right]; N\nu\sigma\tau LM; \Omega L'L''\right\rangle$$
$$\equiv (2,0)^{(N-\nu)/2}\left[P_{\nu\sigma\tau L'}(\beta^{\dagger}_{m})\left|\begin{matrix}\omega_1\omega_2\omega_3 \\ \Omega L''\end{matrix}\right\rangle\right]_{LM}, \tag{7.2}$$

where the square bracket indicates the coupling of the angular momenta L', L'' to L and the polynomial $P(\beta^{\dagger}_{m})$ given by (4.8) is function of traceless boson operators. It is clear that $\overline{B}$ of (4.3) when applied to the square bracket part in the right-hand side of (7.2) vanishes as, from the discussion of Sec. 4 in Ref. 8, $\overline{B}$ commutes with P and, as indicated above, when acting on the state (7.1) it vanishes. Thus the square bracket in (7.2) is characterized by the irreps L of $\mathcal{O}(3)$, ν of $O(3n)$, and $(\omega_1\omega_2\omega_3)$ of $O(n)$. The latter comes from the fact that the state (7.1) corresponds to the irrep $[\omega_1\omega_2\omega_3]$ of U(3) and thus to the same one of U(n), and, as the levels are compactly filled,[7] to the irrep $(\omega_1\omega_2\omega_3)$ of $O(n)$.

These results imply that the ket (7.2) is characterized by the irrep (1.1) of Sp(6) as indicated by the first square bracket in it as well as by the L of $\mathcal{O}(3)$, M of $\mathcal{O}(2)$, and Λ of Sp(2), where Λ is related to ν by (2.4). The irrep N of the subgroup O(2) of Sp(2) is achieved with the factor $(2,0)^{(N-\nu)/2}$ in (7.2) as (2,0) is the raising operator of Sp(2), which is a scalar of $O(3n)$ and thus does not change the irrep ν of this group. The $\sigma, \tau, \Omega, L', L''$ appearing in the ket are the five multiplicity indices required to characterize the state as was indicated in the discussion following (29I).

While in principle we have in (7.2) the basis states for the general irrep (1.1) of Sp(6) characterized by the irreps of the subgroups in the chain (2.15), in practice it is difficult to translate this state from traceless bosons to ordinary bosons acting on definite kets of type (7.1) and even more so to get the states as definite wave functions in configuration space. Thus scalar products of kets of the type (7.2) or matrix elements of the generators of Sp(6) with respect to these kets, still present serious problems which we hope to solve in future publications.

APPENDIX: PROOF OF EQS. (6.3) AND (6.12)

In the first part of this Appendix we shall determine the explicit form of $R^{\,u}_{vw}(\hat{N})$ appearing in Eq. (6.3), i.e.,

$$(\beta_2^\dagger)^u = \sum_v \sum_w (B_2^\dagger)^v (2,0)^w R^{\,u}_{vw}(\hat{N})(Q_2)^{2u-2v-w}(B_2)^{v+w-u}, \tag{A1}$$

where

$$\beta_2^\dagger = B_2^\dagger - 2(2,0)(2\hat{N}+3n)^{-1}Q_2 + (2,0)^2(2\hat{N}+3n+4)^{-1}(2\hat{N}+3n+2)^{-1}B_2 \tag{A2}$$

and the other variables entering in (A1) were defined in (3.7), (3.11), and after (4.7b). Multiplying (A1) by (A2) we obtain

$$(\beta_2^\dagger)^{u+1} = S_1 + S_2 + S_3, \tag{A3}$$

with

$$S_1 \equiv \sum_{vw} (B_2^\dagger)^{v+1}(2,0)^w R^{\,u}_{vw}(\hat{N})(Q_2)^{2u-2v-w}(B_2)^{v+w-u}, \tag{A4}$$

$$S_2 \equiv -2\sum_{vw}(2,0)(2\hat{N}+3n)^{-1}Q_2(B_2^\dagger)^v(2,0)^w \times R^{\,u}_{vw}(\hat{N})(Q_2)^{2u-2v-w}(B_2)^{v+w-u}, \tag{A5}$$

$$S_3 \equiv \sum_{vw}(2,0)^2(2\hat{N}+3n+4)^{-1}(2\hat{N}+3n+2)^{-1} \times B_2(B_2^\dagger)^v(2,0)^w R^{\,u}_{vw}(\hat{N})(Q_2)^{2u-2v-w}(B_2)^{v+w-u}. \tag{A6}$$

The sum S_1 has already all its factors in standard order as in the right side of (A1). On the other hand, in S_2 and S_3 we have to do some transpositions to put the factors in the standard order of Eq. (A1). For doing these exchanges of factors we need some commutators, given in the following list:

$$\begin{aligned}
&[Q_2, B_2^\dagger] = [Q_2, \hat{N}] = 0,\\
&[Q_2,(2,0)^w] = 2wB_2^\dagger(2,0)^{w-1},\\
&[B_2,B_2^\dagger] = [B_2, Q_2] = 0,\\
&[B_2,(2,0)^w] = 4w(2,0)^{w-1}Q_2 + 4w(w-1)B_2^\dagger(2,0)^{w-2}.
\end{aligned} \tag{A7}$$

We use also the relations given in Eq. (4.6). In this way we arrive at

$$\begin{aligned}
S_2 = &-2\sum_{vw}(B_2^\dagger)^v(2,0)^{w+1}(2\hat{N}+3n+4v+4w)^{-1} \times R^{\,u}_{vw}(\hat{N})(Q_2)^{2u-2v-w+1}(B_2)^{v+w-u}\\
&-\sum_{vw}4w(B_2^\dagger)^{v+1}(2,0)^w(2\hat{N}+3n+4v+4w)^{-1} \times R^{\,u}_{vw}(\hat{N})(Q_2)^{2u-2v-w}(B_2)^{v+w-u},
\end{aligned} \tag{A8}$$

and

$$\begin{aligned}
S_3 = &\sum_{vw}(2,0)^{w+2}(B_2^\dagger)^v(2\hat{N}+3n+4v+4w+4)^{-1}(2\hat{N}+3n+4v+4w+2)^{-1}R^{\,u}_{vw}(\hat{N}+2)(Q_2)^{2u-2v-w}(B_2)^{v+w-u+1}\\
&+\sum_{vw}4w(B_2^\dagger)^v(2,0)^{w+1}(2\hat{N}+3n+4v+4w)^{-1}(2\hat{N}+3n+4v+4w-2)^{-1}R^{\,u}_{vw}(\hat{N})(Q_2)^{2u-2v-w}(B_2)^{v+w-u+1}\\
&+\sum_{vw}4w(w-1)(B_2^\dagger)^{v+1}(2,0)^w(2\hat{N}+3n+4v+4w)^{-1}(2\hat{N}+3n+4v+4w-2)^{-1}\\
&\times R^{\,u}_{vw}(\hat{N})(Q_2)^{2u-2v-w}(B_2)^{v+w-u}.
\end{aligned} \tag{A9}$$

Substituting (A4), (A8), and (A9) on the right side of (A3) we obtain, after some changes of dummy indices,

$$(\beta_2^\dagger)^{u+1} = \sum_{vw}(B_2^\dagger)^v(2,0)^w R^{\,u+1}_{vw}(\hat{N})(Q_2)^{2u-2v-w+2}(B_2)^{v+w-u-1}, \tag{A10}$$

with

$$\begin{aligned}
R^{\,u+1}_{vw}(\hat{N}) = &(2\hat{N}+3n+4v+4w-4)^{-1}(2\hat{N}+3n+4v+4w-6)^{-1}\{R^{\,u}_{v,w-2}(\hat{N}+2) - 2(2\hat{N}+3n+4v+2w-4)R^{\,u}_{v,w-1}(\hat{N})\\
&+(2\hat{N}+3n+4v+2w-4)(2\hat{N}+3n+4v+2w-6)R^{\,u}_{v-1,w}(\hat{N})\}.
\end{aligned} \tag{A11}$$

The solution to this recurrence relation is found to be

$$R^{\,u}_{vw}(N) = \frac{2^{2u-2v-w}(-)^w u!(2N+3n+4v+2w-4)!!}{v!(v-u+w)!(2u-2v-w)!(2N+3n+4v+4w-4)!!}. \tag{A12}$$

We turn now to the proof of Eqs (6.12a) and (6.12b). Let us start with the second one. In this case all we need to know is that $-Q_2 \equiv \mathscr{C}_1^{-1}$ and

$$[\mathscr{C}_1^{-1},\langle 2,0\rangle^r] = -2r\langle 2,0\rangle^{r-1}\langle 2,2\rangle,$$
$$[\mathscr{C}_1^{-1},\langle 4,0\rangle^s] = 2s\langle 4,0\rangle^{s-1}\langle 4,2\rangle, \quad [\mathscr{C}_1^{-1},\langle 6,0\rangle] = 0. \tag{A13}$$

Then from the linear character of the operator $\mathscr{C}_1^{-1}$ we obtain (6.12b).

The operator $(\overline{2,2}) \to \Sigma(\partial^2/\partial\eta^2_{-1k})$ is not linear, so we have

$$\sum_{k=1}^n \frac{\partial^2}{\partial\eta^2_{-1k}}\langle 2,0\rangle^r\langle 4,0\rangle^s\langle 6,0\rangle^t = T_1 + T_2 + T_3 + T_4 + T_5 + T_6, \tag{A14}$$

where the definition of the T's, and their values obtained by direct computation, are

$$
\begin{aligned}
T_1 &\equiv \langle 4,0\rangle^s\langle 6,0\rangle^t(\overline{2,2})\langle 2,0\rangle^r \\
&= 4r(r-1)\langle 2,0\rangle^{r-2}\langle 4,0\rangle^s\langle 6,0\rangle^t\langle 2,2\rangle, \\
T_2 &\equiv \langle 2,0\rangle^r\langle 6,0\rangle^t(\overline{2,2})\langle 4,0\rangle^s \\
&= 2s(n-3+2s)\langle 2,0\rangle^r\langle 4,0\rangle^{s-1}\langle 6,0\rangle^t\langle 2,2\rangle \\
&\quad + 4s(s-1)\langle 2,0\rangle^{r+1}\langle 4,0\rangle^{s-2}\langle 6,0\rangle^t\langle 4,2\rangle, \\
T_3 &\equiv \langle 2,0\rangle^r\langle 4,0\rangle^s(\overline{2,2})\langle 6,0\rangle^t \\
&= 2t(2t+n-4)\langle 2,0\rangle^r\langle 4,0\rangle^s\langle 6,0\rangle^{t-1}\langle 4,2\rangle, \\
T_4 &\equiv 2\langle 6,0\rangle^t\sum_k\left(\frac{\partial\langle 2,0\rangle^r}{\partial\eta_{-1k}}\right)\left(\frac{\partial\langle 4,0\rangle^s}{\partial\eta_{-1k}}\right) \\
&= -8rs\langle 2,0\rangle^{r-1}\langle 4,0\rangle^{s-1}\langle 6,0\rangle^t\langle 4,2\rangle, \\
T_5 &\equiv 2\langle 4,0\rangle^s\sum_k\left(\frac{\partial\langle 2,0\rangle^r}{\partial\eta_{-1k}}\right)\left(\frac{\partial\langle 6,0\rangle^t}{\partial\eta_{-1k}}\right) = 0, \\
T_6 &\equiv 2\langle 2,0\rangle^r\sum_k\left(\frac{\partial\langle 4,0\rangle^s}{\partial\eta_{-1k}}\right)\left(\frac{\partial\langle 6,0\rangle^t}{\partial\eta_{-1k}}\right) \\
&= 8st\langle 2,0\rangle^r\langle 4,0\rangle^{s-1}\langle 6,0\rangle^t\langle 2,2\rangle.
\end{aligned}
\tag{A15}
$$

Adding these results we deduce (6.12a).

[1] M. Moshinsky, J. Math. Phys. **25**,1555 (1984).
[2] E. Chacón, P. Hess, and M. Moshinsky, J. Math. Phys. **25**, 1565 (1984).
[3] G. Rosensteel and D. J. Rowe, Ann. Phys. (N.Y.) **126**, 198 and 343 (1980).
[4] B. Buck, L. C. Biedenharn, and R. Y. Cusson, Nucl. Phys. **A317**, 205 (1979).
[5] G. F. Filippov and V. I. Ovcharenko, Yad. Fiz. **30**, 646 (1979); V. C. Vasilevsky, Ya. F. Smirnov, and G. F. Filippov, Yad. Fiz. **32**, 987 (1980).
[6] V. Vanagas, *Lecture Notes in Physics* (University of Toronto Press, Toronto, 1977); V. Vanagas, in *Group Theory and its Applications in Physics*, edited by T. H. Seligman (A.I.P., New York, 1980).
[7] O. Castaños, E. Chacón, A. Frank, P. Hess, and M. Moshinsky, J. Math. Phys. **23**, 2537 (1982).
[8] E. Chacón, M. Moshinsky, and R. T. Sharp, J. Math. Phys. **17**, 668 (1976).
[9] E. Chacón and M. Moshinsky, J. Math. Phys. **18**, 870 (1977).
[10] M. Moshinsky and C. Quesne, J. Math. Phys. **11**, 1631 (1970), and **12**, 1772 (1971); M. Kashiwara and M. Vergne, Invent. Math. **44**, 1 (1978).
[11] W. Zickendraht, J. Math. Phys. **10**, 30 (1969); and **12**, 1663 (1971).
[12] A. Ya. Dzublik, V. I. Ovcharenko, A. I. Steshenko, and G. F. Filippov, Yad. Fiz. **15**, 869 (1972) [Sov. J. Nucl. Phys. **15**, 487 (1972)].
[13] Yu. F. Smirnov and G. F. Filippov, Yad. Fiz. **27**, 73 (1978) [Sov. J. Nucl. Phys. **27**, 39 (1978)].
[14] M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York 1957), pp. 52–54.
[15] R. T. Sharp and C. S. Lam, J. Math. Phys. **10**, 2033 (1969).
[16] N. Y. Vilenkin, *Special Functions and Theory of Group Representations* (A.M.S. Translations, Providence, R. I., 1968).
[17] M. A. Lohe, thesis, University of Adelaide, 1974.
[18] A. J. Dragt, J. Math. Phys. **6**, 533 (1965).
[19] R. Gaskell, A. Peccia, and R. T. Sharp, J. Math. Phys. **19**, 2362 (1978).
[20] V. Bargmann and M. Moshinsky, Nucl. Phys. **23**, 177 (1961).

Application of double Gel'fand polynomials to the symmetric group and spin–isospin wave functions of cluster systems

Y. Fujiwara[a] and Y. C. Tang
School of Physics, University of Minnesota, Minneapolis, Minnesota 55455

H. Horiuchi
Department of Physics, Kyoto University, Kyoto 606, Japan

(Received 22 March 1983; accepted for publication 18 May 1984)

The theory of double Gel'fand polynomials is applied to irreducible representations of the symmetric and SU_4 groups with the aim to treat spin–isospin wave functions of nuclear cluster systems. Multiplicity-free recoupling coefficients of the symmetric group are connected with special types of Clebsch–Gordan coefficients of the unitary group. The standard phase conventions of the Yamanouchi basis and of the multiplicity-free recoupling coefficients are proved to be derivable from natural phase conventions of double Gel'fand polynomials and these special Clebsch–Gordan coefficients. By extending the concept of double Gel'fand polynomials, useful expansion formulas are derived with respect to the determinant associated with a matrix tensor product. A simple example of their application is given for normalization kernels of two-body systems composed of *s*-shell clusters and for SU_4 Clebsch–Gordan coefficients in the spin–isospin representation needed therein.

PACS numbers: 02.20. + b, 21.60.Gx

I. INTRODUCTION

The problem of the explicit construction of an irreducible representation (IR) of the unitary group U_n has a long history in which the boson operator technique, first applied to the angular-momentum algebra by Schwinger[1] and Bargmann[2] and further developed by Baird and Biedenharn,[3] Moshinsky,[4] and many other authors, was proved to be very successful in providing convenient basis states of boson polynomials with which many group-theoretical quantities have been calculated (see Louck[5] and references therein). These boson polynomials carry two state labels which may be chosen as two Gel'fand–Weyl patterns[6,7] sharing the same first row. In this paper, we refer to these "double tableaux boson polynomials" as double Gel'fand (DG) polynomials in conformity with the usage of other authors.[8] The structure of DG polynomials and their remarkable properties have been extensively investigated previously.[9–15] Thus we can now utilize these fruitful results on DG polynomials in the investigation of nuclear many-body problems. In fact, this technique was successfully applied to eigenvalue problems of normalization kernels (NK) of many-cluster systems in Ref. 16, where the SU_3 classification of the orbital wave functions was found to be very efficient under the assumption of equal width parameters for the harmonic oscillator wave functions.

On the other hand, because of the intimate relationship[6,17] between the symmetric group S_N and the unitary group U_n, many attempts have been made to utilize the techniques developed for the IR theory of the S_N group in order to discuss DG polynomials, especially, in the context of the double-coset theory.[13,18–22] We feel, however, that the opposite direction, involving applications of many nice properties of DG polynomials to the symmetric-group properties of nuclear wave functions, has not been fully developed. It was pointed out by Moshinsky[23] that Gel'fand states with the partition of N and the weights restricted to (11...1), referred to by him as special Gel'fand (SG) states, constitute the Yamanouchi basis[24] of the S_N group. By using this property, Moshinsky showed the way to construct explicit states with definite permutation symmetry in configuration and spin–isospin spaces. These states are essentially $U_n \times S_N$ (or $SU_n \times S_N$) DG polynomials embedded in $U_n \times U_N$ DG polynomials if we simply choose the canonical chain to specify the internal quantum numbers of U_n (or SU_n).

The purpose of this paper is to apply the theory of DG polynomials to spin–isospin wave functions of many-cluster systems. If spin–isospin wave functions of subunit clusters are described by simple DG polynomials with definite SU_4 IR labels, the total spin–isospin wave function is obtained by coupling these individual spin–isospin wave functions by SU_4 Clebsch–Gordan (CG) coefficients. This total wave function also has the nature of a DG polynomial, since it is the simultaneous basis of IR (BIR) of the SU_4 and S_N groups (N is the nucleon number of the total system). However, the internal quantum numbers of the S_N group are no longer specified in the standard manner. In this respect, it is indispensible to refine the theory of DG polynomials such that the internal quantum numbers of the S_N group are specified by direct-product representations of its subgroups. This problem of transforming a BIR of the S_N group from the standard representation to the representation with reduction type $S_{N_1} \times S_{N_2}$ ($N_1 + N_2 = N$) was solved by Elliott, Hope, and Jahn,[25] Kaplan,[26] Horie,[27] and Kramer[19] by introducing recoupling coefficients of the S_N group.

The discussion of DG polynomials in this paper is always done in the Bargmann space,[2] where we can embed entirely the boson operator technique and furthermore uti-

[a] Present address: Department of Physics, University of Michigan, Ann Arbor, Michigan 48109-1090.

 0022-2488/84/102826-12$02.50

lize many nice properties of integration with the Bargmann measure.[15] We follow the notation used in Ref. 15 for DG polynomials, since it is particularly convenient for the purpose of this paper. In Sec. II special DG (SDG) polynomials are defined and applied to the IR theory of the S_N group. Although all relationships mentioned there are already well known, we demonstrate these in order to fix the notation and to prepare necessary relationships which will be used in the following sections. In Sec. III it is shown by using SDG polynomials that multiplicity-free (MF) recoupling coefficients of the S_N group can be represented by simple types of MF U_n CG coefficents. This relationship gives us a new method to obtain explicit expressions of the MF recoupling coefficients. Also, one can use this relationship to discuss the phase conventions of the S_N and U_n groups in a unified manner. By using the results in Secs. II and III, two types of expansion formulas are presented and proved in Sec. IV with respect to the determinant associated with a matrix tensor product. Modified DG polynomials whose S_N internal quantum numbers are specified by direct-product representations of the subgroups are first introduced in Sec. III and further discussed in Sec. IV in a way tailored specifically to the spin–isospin wave functions of many-cluster systems in the SU_4 representation. In Sec. V a simple example for the application of the second expansion formula is given with respect to NK in the generator coordinate method (GCM) of two-cluster systems composed of s-shell clusters. The necessary SU_4 CG coefficients to obtain the kernels in spin–isospin representation are also calculated. Some concluding remarks are made in Sec. VI.

II. SPECIAL DOUBLE GEL'FAND POLYNOMIALS

Suppose $\varphi^{(nN)[f]}_{a,r}(R)$ with $n \leqslant N$ and $[f] = [f_1 \cdots f_n]$ are $n \times N$ DG polynomials with $n \times N$ argument matrix R and $D^{(N)[f]}_{r,s}(G)$ are representation matrices of IR of $GL(N;C)$ with $N \times N$ matrix G. As is well known, $D^{(N)[f]}_{r,s}(G)$ can also be represented by $N \times N$ DG polynomials by

$$D^{(N)[f]}_{r,s}(G) = \frac{1}{N_H[f]} \varphi^{(N,N)[f]}_{r,s}(G), \tag{2.1}$$

where $N_H[f]$ is the normalization constant of doubly highest-weight states of DG polynomials. In order to call $\varphi^{(nN)[f]}_{a,r}(R)$ [and also $\varphi^{(NN)[f]}_{r,s}(G)$] special DG (SDG) polynomials, we set the following restriction to the partition $[f]$ and to the internal quantum numbers r and s of the U_N group specified by Gel'fand patterns[23]:

$$\begin{aligned} &\text{(a)} \quad f_1 + \cdots + f_n = N, \\ &\text{(b)} \quad \text{each of the weights of } r \text{ or } s \text{ is 1.} \end{aligned} \tag{2.2}$$

Under these conditions, $\varphi^{(nN)[f]}_{a,r}(R)$ form a basis set of the IR of the S_N group with the partition $[f]$, and the representation matrices are given by $D^{(N)[f]}_{r,s}(\hat{P})$, where P is the matrix representation of the permutation $\hat{P} \in S_N$.

The content of the above theorem is clearly understood by using the square symbol[11,13,21,22] for the expansion coefficients of DG polynomials in a power series. We expand an $n \times m$ DG polynomial as

$$\varphi^{(nm)[f]}_{a,b}(R) = \sum_k \begin{bmatrix} f & b \\ a & k \end{bmatrix} \frac{R^k}{\sqrt{k!}}, \tag{2.3}$$

where the summation over $k \equiv (k_{ij})$ $(i = 1 \sim n, j = 1 \sim m)$ is only for non-negative integers satisfying

$$\begin{aligned} &\sum_{i=1}^{n} k_{ij} = w_j, \quad j = 1 \sim m, \\ &\sum_{j=1}^{m} k_{ij} = w'_i, \quad i = 1 \sim n, \end{aligned} \tag{2.4}$$

with $(w_1 \cdots w_m)$ and $(w'_1 \cdots w'_n)$ being weights of b and a, respectively. We use the shorthand notation

$$k! \equiv k_{11}! \cdots k_{nm}!, \quad R^k \equiv R^{k_{11}}_{11} \cdots R^{k_{nm}}_{nm}, \tag{2.5}$$

for the repeated products of matrix variables. For $[f]$ and r satisfying the condition (2.2), one can easily prove for s uniquely determined by P that

$$\begin{aligned} \varphi^{(NN)[f]}_{a,r}(P) &= \delta_{a,s} \varphi^{(NN)[f]}_{s,r}(P) \\ &= \delta_{a,s} \begin{bmatrix} f & r \\ s & P \end{bmatrix}. \end{aligned} \tag{2.6}$$

This means that the left side of Eq. (2.6) is nonzero only when each of the weights of a is also one. Thus, if we use the transformation formula of DG polynomials [Eq. (2.3.20) of Ref. 15], we get, for an arbitrary $\hat{P} \in S_N$ and a,

$$\begin{aligned} \hat{P} \varphi^{(nN)[f]}_{a,r}(R) &= \varphi^{(nN)[f]}_{a,r}(RP) \\ &= \sum_s D^{(N)[f]}_{s,r}(P) \varphi^{(nN)[f]}_{a,s}(R). \end{aligned} \tag{2.7}$$

The irreducibility of the representation of Eq. (2.7) is the consequence of the orthonormality[15] of SDG polynomials; namely, we have

$$\langle \varphi^{(NN)[f]}_{r,s}(R) | \varphi^{(NN)[f']}_{r',s'}(R) \rangle = \delta_{[f],[f']} \delta_{r,r'} \delta_{s,s'}, \tag{2.8}$$

where $[f]$, $[f']$, r, r', s, and s' all satisfy the condition (2.2). To show this, we only need to notice that for SDG polynomials the summation matrix $k = (k_{ij})$ in Eq. (2.3) has an exact one-to-one correspondence to the representation matrix P of S_N because of the weight conditions [Eq. (2.4)]. Therefore, we can write

$$\begin{aligned} \varphi^{(NN)[f]}_{r,s}(R) &= \sum_{\hat{P} \in S_N} \begin{bmatrix} f & s \\ r & P \end{bmatrix} R^P \\ &= N_H[f] \sum_{\hat{P} \in S_N} D^{(N)[f]}_{r,s}(P) R^P, \end{aligned} \tag{2.9}$$

where the last expression is derived from Eq. (2.1) and the second equality in Eq. (2.6). One substitutes Eq. (2.9) into Eq. (2.8) and performs the Bargmann integral[2] over $d\mu(R)$. By using $P^{-1} = {}^tP$ and the transposition property[15] of DG polynomials, Eq. (2.8) is rewritten into

$$\begin{aligned} &\sum_{\hat{P} \in S_N} D^{(N)[f]}_{r,s}(P) D^{(N)[f']}_{s',r'}(P^{-1}) \\ &\qquad = \delta_{[f],[f']} \delta_{r,r'} \delta_{s,s'} N_H[f]^{-2}. \end{aligned} \tag{2.10}$$

Equation (2.10) is nothing but the first orthogonality relation[28] of IR matrices of the S_N group. Thus one can see that the representation in Eq. (2.7) is actually irreducible and also obtain the following simple relationship[21,22,29] between the normalization constant $N_H[f_1 \cdots f_n]$ of the doubly highest-

weight state of DG polynomials and the dimension formula[28] $|f|_{S_N}$ of the IR of S_N with the partition $[f]=[f_1\cdots f_n]$:

$$(N_H[f_1\cdots f_n])^2=\frac{|f|_{S_N}}{N!},\quad \text{for } f_1+\cdots+f_n=N. \tag{2.11}$$

It is well known that the Gel'fand pattern r with each weight one has a one-to-one correspondence to the Yamanouchi symbol[24] $1\,r_2\cdots r_N$ for the standard scheme to specify BIR of the S_N group, which is based on the canonical chain

$$S_N\supset S_{N-1}\supset\cdots\supset S_2\supset S_1. \tag{2.12}$$

Therefore, we will write hereafter

$$r\equiv 1\,r_2\cdots r_{N-1}r_N \tag{2.13}$$

for the internal quantum numbers of DG polynomials with the condition (b) in Eq. (2.2). We follow Moshinsky's notation in Ref. 23 for the order of $1, r_2,\ldots,r_N$. Furthermore, the number r_N in Eq. (2.13) is, in fact, superfluous. However, we sometimes include it for convenience. If we write $r=r'\rho$ for Eq. (2.13) with $r'=1\,r_2\cdots r_{N-m}$, then $\rho=r_{N-m+1}r_{N-m+2}\cdots r_N$.[26] We also use the shorthand notation $[r_j]\,(j=1\sim N)$ to specify the Young partition constructed from $1\,r_2\cdots r_j$, so that $[r_N]=[f]$. The betweenness condition of Gel'fand patterns precisely corresponds to the condition that $[r_j]\,(j=1\sim N)$ should be all standard. Using these notations, BIR of the S_N group in Eq. (2.7) has the following vector coupling expression by U_n CG coefficients:

$$\begin{aligned}\varphi^{(nN)[f]}_{a,r}(R)&=[\cdots[[U_{[1]}(\mathbf{R}_1)U_{[1]}(\mathbf{R}_2)]_{[r_2]}U_{[1]}(\mathbf{R}_3)]_{[r_3]}\\&\quad\cdots U_{[1]}(\mathbf{R}_N)]_{[f]a},\end{aligned} \tag{2.14}$$

where $R=(\mathbf{R}_1\cdots\mathbf{R}_N)$ is an $n\times N$ matrix and $U_{[\Lambda]\Lambda}(\mathbf{R})$ is the monomial basis of an n-dimensional vector $\mathbf{R}$ with degree $\Lambda=\Lambda_1+\cdots+\Lambda_n$ [see Eqs. (2.2.4) and (2.2.1) of Ref. 15)].

For later discussion, we further introduce the following widely used notation.[28] Let $(r_i,f(r_i))$ be the row and the column of the box in a Young tableau r occupied by the number $i\,(i=1\sim N)$. To any two numbers i and j, one can define an axial distance

$$\tau_{ij}=f(r_i)-f(r_j)+r_j-r_i, \tag{2.15}$$

and a relative phase

$$\epsilon_{ij}=\begin{cases}+1, & r_i\leqslant r_j,\\ -1, & r_i>r_j.\end{cases} \tag{2.16}$$

The phase of Young tableau r is defined by

$$(-1)^r=\prod_{\substack{i<j\\ i,j\in r}}\epsilon_{ij}. \tag{2.17}$$

The associate representation (AR) $[\tilde f]$ of the S_N group and its internal quantum number $\tilde r$ are obtained from $[f]$ and r, respectively, by the interchange of rows and columns. For an arbitrary Young tableau r with the partition $[f]=[f_1\cdots f_n]$ of N, one can easily prove[19]

$$\rho_f\equiv(-1)^{r+\tilde r}\equiv\rho_{\tilde f}=\prod_{1=\mu<\nu}^{n}(-1)^{f_\mu f_\nu-f_\nu(f_\nu+1)/2}. \tag{2.18}$$

In the next section, we will show a method to derive the following well-known formula[24] for the matrix elements of the transposition $\hat P_{N-1,N}$ of the last two particles, using the MF recoupling coefficients of the S_N group:

$$\begin{aligned}\hat P_{N-1,N}\varphi^{(nN)[f]}_{a,r'r_{N-1},r_N}(R)&=\frac{1}{\tau_{N,N-1}}\varphi^{(nN)[f]}_{a,r'r_{N-1},r_N}(R)\\&\quad+\sqrt{1-\left(\frac{1}{\tau_{N,N-1}}\right)^2}\,\varphi^{(nN)[f]}_{a,r'r_Nr_{N-1}}(R),\end{aligned} \tag{2.19}$$

where $r'=1\,r_2\cdots r_{N-2}$. Using Eqs. (2.19), (2.7), and (2.15), one can directly prove

$$D^{(N)[\tilde f]}_{\tilde r\tilde s}(P_{N-1,N})=-(-1)^{r+s}D^{(N)[f]}_{r,s}(P_{N-1,N}), \tag{2.20}$$

or more generally[19]

$$D^{(N)[\tilde f]}_{\tilde r,\tilde s}(P)=\delta_P(-1)^{r+s}D^{(N)[f]}_{r,s}(P),\quad \forall\hat P\in S_N, \tag{2.21}$$

where δ_P is the sign of the permutation $\hat P$.

III. MULTIPLICITY-FREE RECOUPLING COEFFICIENTS OF THE SYMMETRIC GROUP

In order to define recoupling coefficients of the S_N group, we modify the SDG polynomials in Eq. (2.14) such that the internal quantum number $r=1\,r_2\cdots r_N$ is specified by direct-product representation of the subgroup

$$S_{N-m}\times S_m\subset S_N. \tag{3.1}$$

We adopt the canonical chain Eq. (2.12) for each of S_{N-m} and S_m, and take the Yamanouchi symbol r' of S_{N-m} so as to specify the partition of the first $N-m$ particles of S_N at the same time; namely, $r=r'\rho$ with $r'=1\,r_2\cdots r_{N-m}$ and $\rho=r_{N-m+1}\cdots r_N$. Then modified DG (MDG) polynomials with S_N internal quantum numbers specified in Eq. (3.1) are defined by the following vector coupling expression with U_n CG coefficients:

$$\begin{aligned}\tilde\varphi^{(nN)[f]}_{a;r',u(\mu)}(R)&\equiv[[\cdots[U_{[1]}(\mathbf{R}_1)U_{[1]}(\mathbf{R}_2)]_{[r_2]}\\&\quad\cdots U_{[1]}(\mathbf{R}_{N-m})]_{[r']}\\&\quad\times[\cdots[U_{[1]}(\mathbf{R}_{N-m+1})U_{[1]}(\mathbf{R}_{N-m+2})]_{[u_2]}\\&\quad\cdots U_{[1]}(\mathbf{R}_N)]_{[u]}]_{[f]a;\mu},\end{aligned} \tag{3.2}$$

where $u=1\,u_2\cdots u_m$ is the Yamanouchi symbol of S_m and $[r']=[r_{N-m}]$. Here μ is a multiplicity label to specify the U_n coupling $[r']\times[u]\to[f]$. We omit this label if we treat MF cases, i.e., $u=1\,1\cdots1$, $[u]=[m]$ or $u=1\,2\cdots m$, $[u]=[1^m]$. Furthermore, $[f]=[f_1\cdots f_n]$ with $f_1+\cdots+f_n=N$. With Eq. (2.14), Eq. (3.2) is compactly expressed as

$$\begin{aligned}\tilde\varphi^{(nN)[f]}_{a;r',u(\mu)}(R)&=\sum_{b,c}\langle[r']b\,[u]c|[f]a;\mu\rangle_n\\&\quad\times\varphi^{(n,N-m)[r']}_{b,r'}(\mathbf{R}_1\cdots\mathbf{R}_{N-m})\\&\quad\times\varphi^{(nm)[u]}_{c,u}(\mathbf{R}_{N-m+1}\cdots\mathbf{R}_N),\end{aligned} \tag{3.3}$$

where $\langle\cdots|\cdots\rangle_n$ denotes a U_n CG coefficient. The recoupling coefficients are defined to be transformation matrices for the unitary transformation between two different sets of BIR in Eqs. (3.3) and (2.14). Using Horie's notation,[27] we define them by

$$\tilde{\varphi}^{(nN)[f]}_{a;r',u(\mu)}(R) = \sum_{\rho} \langle r'\rho | r', u(\mu)\rangle \varphi^{(nN)[f]}_{a,r'\rho}(R), \tag{3.4}$$

and

$$\varphi^{(nN)[f]}_{a,r'\rho}(R) = \sum_{u(\mu)} \langle r', u(\mu) | r'\rho\rangle \tilde{\varphi}^{(nN)[f]}_{a;r',u(\mu)}(R). \tag{3.5}$$

We should note that the r'-dependence in the recoupling coefficients is in fact only in $[r']$. Furthermore, if we only treat MF recoupling coefficients, we can consider them as real, since the U_n CG coefficient in Eq. (3.3) becomes real. Therefore, we have

$$\langle r', u | r'\rho\rangle = \langle r'\rho | r', u\rangle, \quad \text{for } u = [m] \text{ or } [1^m]. \tag{3.6}$$

Using the properties of SDG polynomials, we can prove the following proposition with respect to the MF recoupling coefficients.[30]

Proposition:

(a) $\langle r'\rho | r', [1^m]\rangle$

$$= \frac{1}{\sqrt{m!}} \frac{\langle [r']H[1^m]\Lambda | [r]H\rangle_n}{\prod_{i=N-m+1}^{N} \langle [r_{i-1}]H[1]\mathbf{r}_i | [r_i]H\rangle_n}$$

$$\text{for } 1 \leqslant r_{N-m+1} < \cdots < r_N \leqslant n. \tag{3.7}$$

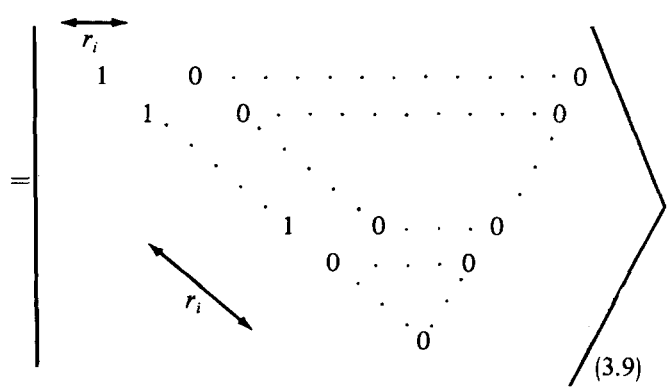
(b) $\langle r'\rho | r', [m]\rangle$

$$= \sqrt{\frac{\Lambda_1! \cdots \Lambda_n!}{m!}} \times \frac{\langle [r']H[m]\Lambda | [r]H\rangle_n}{\prod_{i=N-m+1}^{N} \langle [r_{i-1}]H[1]\mathbf{r}_i | [r_i]H\rangle_n}$$

$$\text{for } 1 \leqslant r_{N-m+1} \leqslant \cdots \leqslant r_N \leqslant n. \tag{3.8}$$

Here $[r_{N-m}] = [r'] \equiv [q] = [q_1 \cdots q_n]$ and $[r_N] = [r] \equiv [f] = [f_1 \cdots f_n]$ are both U_n partitions and $\Lambda = (\Lambda_1 \cdots \Lambda_n)$ $(\Lambda_1 + \cdots + \Lambda_n = m)$ with $\Lambda_\mu = f_\mu - q_\mu$ $(\mu = 1 \sim n)$. Also, H denotes the highest weight of U_n. Furthermore, $\mathbf{r}_i$ is defined by [cf. Eq. (2.2.2) of Ref. 15]

$\mathbf{r}_i \equiv (0 \cdots 010 \cdots 0)$



(3.9)

Proof: (a) We combine Eqs. (3.3) and (3.4) and set $[u] = [1^m]$ and $a = H$. We multiply $\varphi^{(n,N-m)[r']}_{H,r'}(\mathbf{R}_1 \cdots \mathbf{R}_{N-m})^*$ on both sides and integrate over the Bargmann measure $d\mu(\mathbf{R}_1) \cdots d\mu(\mathbf{R}_{N-m})$. Then, using Eq. (2.14) and changing ρ into $\rho' = \mathbf{r}'_{N-m+1} \cdots r'_N$, we obtain

$$\sum_{\rho'} \langle r'\rho' | r', [1^m]\rangle$$

$$\times \sum_{\alpha_1 \cdots \alpha_m} \cdots \sum_{\beta_1 \cdots \beta_{m-1}} \langle [r']H[1]\alpha_1 | [r'_{N-m+1}]\beta_1\rangle_n$$

$$\times \langle [r'_{N-m+1}]\beta_1[1]\alpha_2 | [r'_{N-m+2}]\beta_2\rangle_n$$

$$\cdots \langle [r'_{N-1}]\beta_{m-1}[1]\alpha_m | [r]H\rangle_n$$

$$\times R_{\alpha_1, N-m+1} R_{\alpha_2, N-m+2} \cdots R_{\alpha_m, N}$$

$$= \sum_{c} \langle [r']H[1^m]c | [r]H\rangle_n$$

$$\times \varphi^{(nm)[1^m]}_{c,1\,2\cdots m}(\mathbf{R}_{N-m+1} \cdots \mathbf{R}_N), \tag{3.10}$$

where the ρ' sum is over $r'_{N-m+1}, \cdots, r'_N$ under the condition $[r'_N] = [r]$. From the conservation of the weights, c in Eq. (3.10) should have the weights $(\Lambda_1 \cdots \Lambda_n)$ and also

$$\Lambda_\mu = \begin{cases} 1, & \text{for } \mu = r_{N-m+1}, \ldots, r_N, \\ 0, & \text{otherwise.} \end{cases} \tag{3.11}$$

This means that under the IR label $[1^m]$ the weight vector Λ completely specifies the U_n internal quantum number c. Therefore we can set uniquely $c = \Lambda$ in Eq. (3.10). It also means that the summation over $\alpha_1 \cdots \alpha_m$ on the left-hand side of Eq. (3.10) is just the permutation of $r_{N-m+1} \cdots r_N$ and, similarly, is the summation over $r'_{N-m+1} \cdots r'_N$ since $[r'_N] = [r]$. We change $\mathbf{R}_{N-m+1} \to \mathbf{R}_1, \ldots, \mathbf{R}_N \to \mathbf{R}_m$ in Eq. (3.10) and use

$$\varphi^{(nm)[1^m]}_{\Lambda,1\,2\cdots m}(R) = \varphi^{(nm)[1^m]}_{\Lambda,H}(R) = \frac{1}{\sqrt{m!}} \det\{R_{r_{N-m+1},1} \cdots R_{r_N,m}\}, \tag{3.12}$$

which is easily proved by the back-side vector coupling expression[15] of $\varphi^{(nm)[1^m]}_{\Lambda,H}(R)$. Then, multiplying by $R^*_{r_{N-m+1},1} \cdots R^*_{r_N,m}$ and integrating over $d\mu(R_{r_{N-m+1},1}) \cdots d\mu(R_{r_N,m})$, we obtain

$$\sum_{\rho'} \langle r'\rho' | r', [1^m]\rangle$$

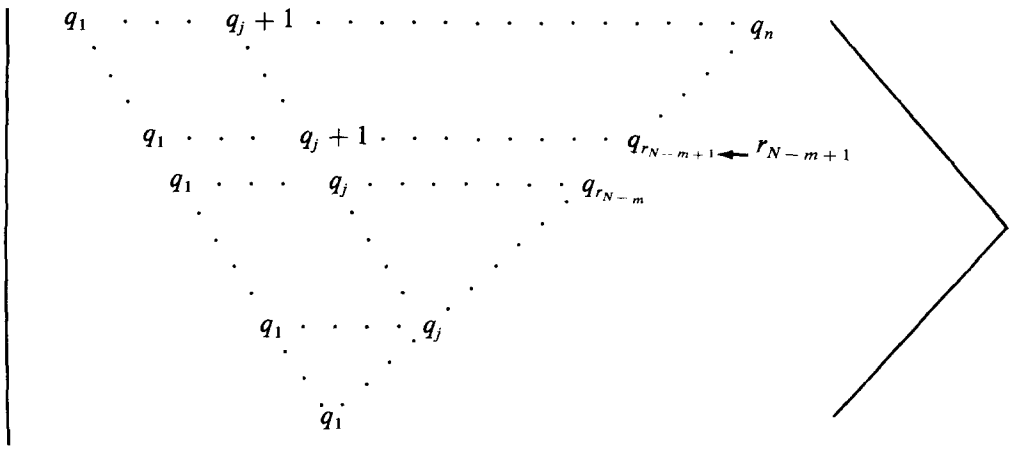
$$\times \sum_{\beta_1 \cdots \beta_{m-1}} \langle [r']H[1]\mathbf{r}_{N-m+1} | [r'_{N-m+1}]\beta_1\rangle_n$$

$$\times \langle [r'_{N-m+1}]\beta_1[1]\mathbf{r}_{N-m+2} | [r'_{N-m+2}]\beta_2\rangle_n$$

$$\cdots \langle [r'_{N-1}]\beta_{m-1}[1]\mathbf{r}_N | [r]H\rangle_n$$

$$= (1/\sqrt{m!}) \langle [r']H[1^m]\Lambda | [r]H\rangle_n. \tag{3.13}$$

Let us consider $\langle [r']H[1]\mathbf{r}_{N-m+1} | [r'_{N-m+1}]\beta_1\rangle_n$ on the left-hand side of Eq. (3.13). If we set $[r'] = [q]$, the possible form of the Gel'fand pattern β_1 is, for some $j \leqslant r_{N-m+1}$,



(3.14)

from the conservation of the weights and the betweenness condition for the coupling $[q]H\times[1]\mathbf{r}_{N-m+1}$. However, we should have $[q_1\cdots q_j+1\cdots]=[r'_{N-m+1}]$, so that $j=r'_{N-m+1}\leqslant r_{N-m+1}$. Noting that the $r'_{N-m+1}\cdots r'_N$ are just the permutation of $r_{N-m+1}\cdots r_N$ and that r_{N-m+1} is the smallest value among $r_{N-m+1}\cdots r_N$, we can conclude that the summation over r'_{N-m+1} in Eq. (3.13) is only for $r'_{N-m+1}=r_{N-m+1}$ and $\beta_1=H$. Repeating this process, we obtain $r'_{N-m+2}=r_{N-m+2}$, $\beta_2=H,\cdots,r'_{N-1}=r_{N-1}$, and $\beta_{m-1}=H$ in sequence. This completes the proof of Eq. (3.7).

The proof of (b) is also made in a similar manner as in (a). There we need

$$\langle R_{11}\cdots R_{1\Lambda_1}R_{2,\Lambda_1+1}\cdots R_{2,\Lambda_1+\Lambda_2}\cdots R_{n,\Lambda_1+\cdots+\Lambda_{n-1}+1}\cdots R_{nm}|\varphi^{(nm)[m]}_{\Lambda,11\cdots1}(R)\rangle=\sqrt{\frac{\Lambda_1!\cdots\Lambda_n!}{m!}}, \tag{3.15}$$

which is most easily proved as follows. We consider an expansion

$$R_{11}\cdots R_{1\Lambda_1}R_{2,\Lambda_1+1}\cdots R_{2,\Lambda_1+\Lambda_2}\cdots R_{n,\Lambda_1+\cdots+\Lambda_{n-1}+1}\cdots R_{nm}=\sum_{[g_2]\cdots[g_m]a}c([g_2]\cdots[g_m];a)[\cdots[\mathrm{U}_{[1]}(\mathbf{R}_1)\mathrm{U}_{[1]}(\mathbf{R}_2)]_{[g_2]}\cdots\mathrm{U}_{[1]}(\mathbf{R}_m)]_{[g_m]a}. \tag{3.16}$$

Then, Eq. (3.15) is calculated as $c([2]\cdots[m];\Lambda)$. Let us put $\mathbf{R}_1=\mathbf{R}_2=\cdots=\mathbf{R}_m=\mathbf{R}$ in Eq. (3.16) and use the formula

$$[\mathrm{U}_{[N_1]}(\mathbf{R})\mathrm{U}_{[N_2]}(\mathbf{R})]_{[g]a}=\delta_{[g],[N_1+N_2]}\sqrt{\binom{N_1+N_2}{N_1}}\,\mathrm{U}_{[N_1+N_2]a}(\mathbf{R}). \tag{3.17}$$

Then only $[g_2]=[2],\cdots,[g_m]=[m]$ survive for the sum in Eq.(3.16) and we obtain

$$R^{\Lambda_1}_1R^{\Lambda_2}_2\cdots R^{\Lambda_n}_n=\sum_a c([2]\cdots[m];a)\sqrt{m!}\,\mathrm{U}_{[m]a}(\mathbf{R}).$$

Comparing the weights, we know that $a=\Lambda$ only and the explicit expression[15] of $\mathrm{U}_{[m]\Lambda}(\mathbf{R})$ gives us the right-hand side of Eq. (3.15) for $c([2]\cdots[m];\Lambda)$. □

This proposition has very fruitful contents as will be seen in the following applications.

A. Application to $\langle r'\rho|r',[m]\rangle$

One can obtain an explicit expression of these MF recoupling coefficients by using the explicit expression of $\langle[q]H[\Lambda]\Lambda|[f]H\rangle_n$. This special U_n CG coefficient is given in Eq. (2.2.11) of Ref. 15 and is

$$\langle[q_1\cdots q_n]H[\Lambda]\Lambda|[f_1\cdots f_n]H\rangle_n=\left\{\prod_{\nu>\mu=1}^{n}\frac{(f_\mu-f_\nu+\nu-\mu)(f_\mu-f_\nu+\nu-\mu-1)\cdots(q_\mu-f_\nu+\nu-\mu)}{(f_\mu-q_\nu+\nu-\mu)(f_\mu-q_\nu+\nu-\mu-1)\cdots(q_\mu-q_\nu+\nu-\mu)}\right\}^{1/2}. \tag{3.18}$$

Using Eqs. (3.18) and (3.8), we first calculate $\langle r'\rho|r',[m]\rangle/\langle r'\bar{\rho}|r',[m-1]\rangle$ with $\rho=\bar{\rho}r_N$, which can be conveniently represented by using τ_{ij} in Eq. (2.15). From this recursion relation, we obtain the well-known result[27]

$$\langle r'\rho|r',[m]\rangle=\left\{\frac{1}{m!}\prod_{i>j=N-m+1}^{N}\left(1+\frac{1}{\tau_{ij}}\right)\right\}^{1/2}, \tag{3.19}$$

for $1\leqslant r_{N-m+1}\leqslant\cdots\leqslant r_N\leqslant n$. In order to remove this restriction on $r_{N-m+1}\cdots r_N$, we first prove Eq. (2.19) by using the $m=2$ case of Eq. (3.19). This can be done by using the relationships

$$\langle r'r_{N-1}r_N|\hat{P}_{N-1,N}|r'r_{N-1}r_N\rangle=2\langle r'r_{N-1}r_N|r',[2]\rangle^2-1, \tag{3.20}$$

and

$$\langle r'r_Nr_{N-1}|\hat{P}_{N-1,N}|r'r_{N-1}r_N\rangle=2\langle r'r_Nr_{N-1}|r',[2]\rangle\langle r'r_{N-1}r_N|r',[2]\rangle,\quad\text{for } r_{N-1}\neq r_N, \tag{3.21}$$

which are derived from the completeness of $|r',[2]\rangle$ and $|r',[11]\rangle$, and the orthonormality of the recoupling coefficients. If we further use the fact that the quantity on the left side of Eq.(3.21) is always positive (see Sec. III C about the phase convention), we can prove Eq. (2.19) and that Eq. (3.19) for the $m=2$ case is true even if $r_{N-1}>r_N$. The use of induction employing Eq. (2.19) makes it possible to prove that Eq. (3.19) is true for any arbitrary order of magnitudes of $r_{N-m+1}\cdots r_N$.

B. Application to $\langle r'\rho|r',[1^m]\rangle$

The next relationship between the two types of MF recoupling coefficients is easily proved by operating $\hat{P}=\hat{P}_1\hat{P}_2(\hat{P}_1\in S_{N-m},\hat{P}_2\in S_m)$ on Eq. (3.5) and using Eq. (2.21)[19];

$$\langle \tilde{r}',[1^m]|\tilde{r'\rho}\rangle = \kappa([r'][m];[r])(-1)^{r+r'}\langle r',[m]|r'\rho\rangle, \tag{3.22}$$

where $\kappa([r'][m];[r])$ is a real phase factor introduced in Ref. 19 and $r=r'\rho$. In order to obtain the explicit expression of $\langle r'\rho|r',[1^m]\rangle$ from Eqs. (3.22) and (3.19), we determine this phase factor by adopting a set of phase conventions which will be given in Eq. (3.29) of the next subsection. The discussion in Sec. III A and the combined use of Eq. (3.7) with the phase convention (iv) of Eq. (3.29) below lead us to the conclusions

$$\langle r'\rho|r',[m]\rangle\geqslant 0 \tag{3.23}$$

and

$$\langle r'\rho|r',[1^m]\rangle\geqslant 0, \quad \text{for } 1\leqslant r_{N-m+1}<\cdots<r_N\leqslant n. \tag{3.24}$$

In the Appendix, we shall prove that the phase rule[27] in Eqs. (3.23) and (3.24) gives us

$$\kappa([q][m];[f])=\prod_{1=\mu<\nu}^{n}(-1)^{(f_\mu-q_\mu)f_\nu} \tag{3.25}$$

with $[q]=[q_1\cdots q_n]$ and $[f]=[f_1\cdots f_n]$, so that Eq. (3.22) becomes

$$(-1)^{\rho}\langle r',[1^m]|r'\rho\rangle=\langle \tilde{r}',[m]|\tilde{r'\rho}\rangle, \tag{3.26}$$

where $(-1)^{\rho}$ is defined by Eq. (2.17) with r replaced by ρ. Using Eq. (3.19) with an arbitrary order for $r_{N-m+1}\cdots r_N$ and $\tilde{\tau}_{ij}=r_i-r_j+f(r_j)-f(r_i)=-\tau_{ij}$, one obtains from Eq. (3.26) the well-known result[27]

$$\langle r'\rho|r',[1^m]\rangle = \prod_{N-m+1=i<j}^{N}\epsilon_{ij}\left\{\frac{1}{m!}\prod_{i>j=N-m+1}^{N}\left(1-\frac{1}{\tau_{ij}}\right)\right\}^{1/2}. \tag{3.27}$$

Furthermore, if we use Eq. (3.7) again, we obtain the explicit expression of the following special U_n CG coefficient;

$$\langle [q_1\cdots q_n]H[1^m]r_{N-m+1}\cdots r_N|[f_1\cdots f_n]H\rangle_n = \left\{\prod_{r_i>\mu=1}^{n}\frac{f_\mu-f_{r_i}+r_i-\mu}{q_\mu-q_{r_i}+r_i-\mu}\right\}^{1/2}, \tag{3.28}$$

where $1\leqslant r_{N-m+1}<\cdots<r_N\leqslant n$ are determined from $[q]$ and $[f]$ uniquely and the product of r_i is over all $r_{N-m+1},\ldots,r_N$.

C. Discussion of the phase convention

It will be useful to summarize in this subsection the relationship between phase conventions used in the representation theory of the unitary group and in that of the symmetric group. In §2-2 of Ref. 15, it is shown that for the unitary group $U_l(l=1,2,\ldots)$ the phase conventions (i)~(iii) in

(i) $N_H[f_1\cdots f_l]\geqslant 0$,

(ii) $N\begin{bmatrix} f_1\cdots f_{l-1} \\ f_1\cdots f_l \\ q_1\cdots q_{l-1}\end{bmatrix}^{-1}\geqslant 0$,

(iii) $\langle [q_1\cdots q_{l-1}]H[\Lambda]\Lambda|[f_1\cdots f_l]H\rangle\geqslant 0$,

(iv) $\langle [q_1\cdots q_{l-1}]H[1^\Lambda]\Lambda|[f_1\cdots f_l]H\rangle\geqslant 0$,

$$\Lambda_\mu=f_\mu-q_\mu(\mu=1\sim l-1),\quad \Lambda_l=f_l,$$

$$\Lambda=|\Lambda|=\sum_{\mu=1}^{l}f_\mu-\sum_{\mu=1}^{l-1}q_\mu, \tag{3.29}$$

give us two kinds of vector coupling expressions for DG polynomials without any phase factor [see Eqs. (4.9) and (2.14)]. The conditions (i) and (ii) mean that the normalization constants of the doubly highest-weight states [Eq. (2.11)] and of the lowering operators of U_l, respectively, are non-negative and have already been adopted by Nagel and Moshinsky.[4,9] The condition (iii) is a natural extension of the phase convention in the angular momentum theory given by Condon and Shortley,[31] namely,

$$\langle j_1 j_1 j_2 j_3 - j_1|j_3 j_3\rangle\geqslant 0.$$

On the other hand, the positiveness of the off-diagonal matrix elements in Eq. (2.19), which was first adopted by Yamanouchi[24] quite arbitrarily, was connected by Moshinsky[23] with the condition (ii) by calculating directly the matrix elements in Eq. (2.19) by SG states. Moshinsky represented $\hat{P}_{N-1,N}$ by U_N generators $\{C_\alpha^\beta\}$ and expressed the off-diagonal matrix element of $\hat{P}_{N-1,N}$ as a product of two matrix elements of C_N^{N-1}. Then, if we use the fact proved by Nagel and Moshinsky[9] that the condition (ii) is equivalent to the non-negativeness of the matrix elements of C_N^{N-1}, we can immediately prove that the nonzero off-diagonal matrix elements of $\hat{P}_{N-1,N}$ are always positive. In fact, we have already used this result in Sec. III A to derive Eq. (2.19). However, as far as Eq. (2.19) is concerned, Moshinsky's proof[23] is much more direct than ours, since we assumed all the conditions (i)~(iii) implicitly.

The advantage of our method shows up if we discuss the phase convention of MF recoupling coefficients. The phase conventions Eqs. (3.23) and (3.24) for $\langle r'\rho|r',[m]\rangle$ and $\langle r'\rho|r',[1^m]\rangle$ were first adopted by Horie.[27] However, we showed in previous subsections that the non-negativeness of $\langle r'\rho|r',[m]\rangle$ can be derived from the phase conventions (i)~(iii) of Eq.(3.29), while Eq. (3.24) can be derived by just adding (iv) to (i)~(iii). Using these signs of phase factors, we can then obtain the explicit expression of the phase factor $\kappa([q][m];[f])$ in Eq. (3.25).

With the above discussion, we may say that the phase conventions of the two kinds of fundamental quantities in IR theory of the S_N group, namely, the matrix elements of $\hat{P}_{N-1,N}$ and MF recoupling coefficients $\langle r'\rho|r',[m]\rangle$ and $\langle r'\rho|r',[1^m]\rangle$ can be derived from the phase conventions (i)~(iv) in IR theory of the unitary group.

IV. EXPANSION FORMULAS

In this section, we derive two kinds of expansion formulas of the determinant associated with a matrix tensor pro-

duct, using the properties of SDG polynomials and MF recoupling coefficients derived in Sec. III.

A. The first expansion formula

Theorem 1:

$$\begin{vmatrix} A_{11}V_{11} & \cdots & A_{1N}V_{1N} \\ \vdots & & \vdots \\ A_{N1}V_{N1} & \cdots & A_{NN}V_{NN} \end{vmatrix} \equiv \det\{A \otimes V\}$$

$$= \sum_{[f]rs}{}' (-1)^{r+s} \varphi_{r,s}^{(NN)[f]}(A)\varphi_{\bar{r},\bar{s}}^{(NN)[\tilde{f}]}(V), \tag{4.1}$$

where the summation over $[f]$ is only for $f_1 + \cdots + f_N = N$ ($f_1 \geqslant \cdots \geqslant f_N \geqslant 0$) and each of the weights of r and s is 1.

Proof: Let us use Eqs. (2.9) (with $n = N$) and (2.21) for the right-hand side of Eq. (4.1). Then using $N_H[\tilde{f}] = N_H[f]$ from Eq. (2.11), we obtain

$I \equiv$right-hand side of Eq. (4.1)

$$= \sum_{P\in S_N} \sum_{\hat{P}'\in S_N} \delta_{P'} \sum_{[f]rs}{}' N_H[f]^2 D_{r,s}^{(N)[f]}(P) \times D_{r,s}^{(N)[f]}(P') A^P V^{P'}. \tag{4.2}$$

Here, by using the transposition property and the product formula of representation matrices, we find that

$$\sum_{r,s} D_{r,s}^{(N)[f]}(P) D_{r,s}^{(N)[f]}(P') = \sum_r D_{r,r}^{(N)[f]}(PP'^{-1}) = \chi^{[f]}(PP'^{-1}), \tag{4.3}$$

where $\chi^{[f]}$ is the character of the S_N group with the IR label $[f]$. We further notice from Eq. (2.11) that

$$N_H[f]^2 = \frac{|f|_{S_N}}{N!} = \frac{1}{N!}\sum_r D_{r,r}^{(N)[f]}(E) = \frac{1}{N!}\chi^{[f]}(E), \tag{4.4}$$

where E is the $N \times N$ unit matrix. Thus, if we use the second orthogonality relation[28] of irreducible characters, Eq. (4.2) becomes

$$I = \sum_{P\in S_N} \sum_{\hat{P}'\in S_N} \delta_{P'} \left\{ \frac{1}{N!} \sum_{[f]\in S_N} \chi^{[f]}(E)\chi^{[f]}(PP'^{-1}) \right\} \times A^P V^{P'}$$

$$= \sum_{P\in S_N} \sum_{\hat{P}'\in S_N} \delta_{P'} \cdot \delta_{E,PP'^{-1}} A^P V^{P'}$$

$$= \sum_{P\in S_N} \delta_P A^P V^P, \tag{4.5}$$

which is the determinant on the left-hand side of Eq. (4.1). □

B. The second expansion formula

Suppose the $N \times N$ matrix A in Eq. (4.1) takes a reduced form

$$\hat{A}^{(nm)} \equiv \begin{pmatrix} \hat{A}_{11} & \cdots & \hat{A}_{1m} \\ \vdots & & \vdots \\ \hat{A}_{n1} & \cdots & \hat{A}_{nm} \end{pmatrix}, \tag{4.6a}$$

with $P_\mu \times N_\nu$ matrices ($\mu = 1 \sim n$, $\nu = 1 \sim m$)

$$\hat{A}_{\mu\nu} \equiv \overbrace{\begin{pmatrix} A_{\mu\nu} & \cdots & A_{\mu\nu} \\ \vdots & & \vdots \\ A_{\mu\nu} & \cdots & A_{\mu\nu} \end{pmatrix}}^{N_\mu} \Big\} P_\mu, \tag{4.6b}$$

corresponding to the decomposition of the weights $N = N_1 + \cdots + N_m = P_1 + \cdots + P_n$. We can redefine an $n \times m$ matrix $A \equiv (A_{\mu\nu})$. Then, Eq. (4.1) is transformed into the second expansion formula using standard $n \times m$ DG polynomials $\varphi_{a,b}^{(nm)[f]}(A)$ and a new type of MDG polynomials $\tilde{\varphi}_{\bar{a},\bar{b}}^{(NN)[\tilde{f}]}(V)$ defined by the following vector-coupling expressions.

Definition: Let $\varphi_{a,b}^{(nm)[f]}(A)$ be an $n \times m$ DG polynomial with the partition $[f] = [f_1 \cdots f_l]$ ($l = \mathrm{Min}\{n,m\}$) of N ($f_1 + \cdots + f_l = N$) and the $n \times m$ argument matrix

$$A = (\mathbf{A}_1 \cdots \mathbf{A}_m) = \begin{pmatrix} \mathbf{A}_1^r \\ \vdots \\ \mathbf{A}_n^r \end{pmatrix}. \tag{4.7}$$

Corresponding to the Gel'fand patterns

$$\left| \begin{matrix} [f] \\ a \end{matrix} \right\rangle = \left| \begin{matrix} f_{1n} \cdots \cdots \cdots \cdots f_{nn} \\ f_{1,n-1} \cdots \cdots f_{n-1,n-1} \\ \cdots \cdots \\ f_{11} \end{matrix} \right\rangle = \left| \begin{matrix} [f_n] \\ [f_{n-1}] \\ \vdots \\ [f_1] \end{matrix} \right\rangle \tag{4.8a}$$

and

$$\left| \begin{matrix} [f] \\ b \end{matrix} \right\rangle = \left| \begin{matrix} g_{1m} \cdots \cdots \cdots \cdots g_{mm} \\ g_{1,m-1} \cdots \cdots g_{m-1,m-1} \\ \cdots \cdots \\ g_{11} \end{matrix} \right\rangle = \left| \begin{matrix} [g_m] \\ [g_{m-1}] \\ \vdots \\ [g_1] \end{matrix} \right\rangle \tag{4.8b}$$

with $[f_n] = [g_m] = [f]$, there are two types of vector coupling expressions by U_n and U_m CG coefficients[15]:

$$\varphi_{a,b}^{(nm)[f]}(A) = [\cdots[[\mathrm{U}_{[N_1]}(\mathbf{A}_1)\mathrm{U}_{[N_2]}(\mathbf{A}_2)]_{[g_2]}\mathrm{U}_{[N_3]}(\mathbf{A}_3)]_{[g_3]}\cdots\mathrm{U}_{[N_m]}(\mathbf{A}_m)]_{[f]a} \tag{4.9a}$$

$$= [\cdots[[\mathrm{U}_{[P_1]}(\mathbf{A}_1^r)\mathrm{U}_{[P_2]}(\mathbf{A}_2^r)]_{[f_2]}\mathrm{U}_{[P_3]}(\mathbf{A}_3^r)]_{[f_3]}\cdots\mathrm{U}_{[P_n]}(\mathbf{A}_n^r)]_{[f]b}, \tag{4.9b}$$

where

$$N_1 = g_{11}, \quad N_j = \sum_{i=1}^{j} g_{ij} - \sum_{i=1}^{j-1} g_{i,j-1} \quad (j = 2 \sim m) \tag{4.10a}$$

and

$$P_1 = f_{11}, \quad P_\beta = \sum_{\alpha=1}^{\beta} f_{\alpha\beta} - \sum_{\alpha=1}^{\beta-1} f_{\alpha,\beta-1} \quad (\beta = 2 \sim n) \tag{4.10b}$$

are the weights of vectors $\mathbf{A}_j$ and $\mathbf{A}_\beta^r$, respectively. Then, a MDG polynomial $\tilde{\varphi}_{\tilde{a},\tilde{b}}^{(NN)[\tilde{f}]}(V)$ with specific S_N internal quantum numbers $\tilde{a}$ and $\tilde{b}$ in the direct-product representation scheme of the subgroups $S_{P_1} \times \cdots \times S_{P_n} \subset S_N$ and $S_{N_1} \times \cdots \times S_{N_m} \subset S_N$ ($N = P_1 + \cdots + P_n = N_1 + \cdots + N_m$), respectively, can be defined from Eq. (4.9) by converting all the partitions of $[f]$, a, and b into the corresponding AR labels and by changing the $n \times m$ matrix A into an $N \times N$ matrix V; namely,

$$\tilde{\varphi}_{\tilde{a},\tilde{b}}^{(NN)[\tilde{f}]}(V) \equiv \left[\cdots \left[U_{[1^{N_1}]}(\mathbf{V}_1 \cdots \mathbf{V}_{N_1}) U_{[1^{N_2}]}(\mathbf{V}_{N_1+1} \cdots \mathbf{V}_{N_1+N_2}) \right]_{[\tilde{g}_2]} \cdots U_{[1^{N_m}]}(\mathbf{V}_{n-N_m+1} \cdots \mathbf{V}_N) \right]_{[\tilde{f}]\tilde{a}} \tag{4.11a}$$

$$= \left[\cdots \left[U_{[1^{P_1}]}(\mathbf{V}_1^r \cdots \mathbf{V}_{P_1}^r) U_{[1^{P_2}]}(\mathbf{V}_{P_1+1}^r \cdots \mathbf{V}_{P_1+P_2}^r) \right]_{[\tilde{f}_2]} \cdots U_{[1^{P_n}]}(\mathbf{V}_{N-P_n+1}^r \cdots \mathbf{V}_N^r) \right]_{[\tilde{f}]\tilde{b}}, \tag{4.11b}$$

where we have introduced a new notation

$$U_{[1^A]_c}(\mathbf{R}_1 \cdots \mathbf{R}_A) \equiv [\cdots [U_{[1]}(\mathbf{R}_1) U_{[1]}(\mathbf{R}_2)]_{[11]}] \cdots U_{[1]}(\mathbf{R}_A)]_{[1^A]_c} \tag{4.12}$$

for a SDG polynomial $\varphi_{c,11\cdots1}^{(NA)[1^A]}(\mathbf{R}_1 \cdots \mathbf{R}_A)$. The vector couplings in Eqs. (4.11) and (4.12) are all achieved by using U_N or S_N CG coefficients.

Theorem 2:

$$\det\{\hat{A}^{(nm)} \otimes V\} = \sqrt{\prod_{\mu=1}^{n} P_\mu! \prod_{\nu=1}^{m} N_\nu!} \sum_{[f]ab}{}' \kappa([f]a)\kappa([f]b) \times \varphi_{a,b}^{(nm)[f]}(A)\, \tilde{\varphi}_{\tilde{a},\tilde{b}}^{(NN)[\tilde{f}]}(V), \tag{4.13}$$

where the summation over $[f] = [f_1 \cdots f_l]$ ($l = \mathrm{Min}\{n,m\}$) is only for $f_1 + \cdots + f_l = N$, and a and b are only those in Eq. (4.8) satisfying the weight conditions of Eq. (4.10). The phase factor $\kappa([f]a)$, for example, is given by

$$\kappa([f]a) = \prod_{1=\alpha<\beta<\gamma}^{n} (-1)^{(f_{\alpha\gamma} - f_{\alpha,\gamma-1})f_{\beta\gamma}}. \tag{4.14}$$

Proof: We first derive a reduction formula of SDG polynomials by using Eq. (3.17). If we set $\mathbf{R}_{N-m+1} = \mathbf{R}_{N-m+2} = \cdots = \mathbf{R}_N = \mathbf{R}$ in Eqs. (3.5) and (3.3), we find

$$\varphi_{a,r'\rho}^{(nN)[f]}\left(\mathbf{R}_1 \cdots \mathbf{R}_{N-m} \overleftarrow{\mathbf{R} \cdots \mathbf{R}}^{\leftarrow m \rightarrow}\right) = \sqrt{m!}\langle r', [m] | r'\rho \rangle \times [\varphi_{*,r'}^{(n,N-m)[r']}(\mathbf{R}_1 \cdots \mathbf{R}_{N-m}) U_{[m]}(\mathbf{R})]_{[f]a}. \tag{4.15}$$

Repeating this and using Eq. (4.9a), we obtain

$$\varphi_{a,\rho_1\rho_2\cdots\rho_m}^{(nN)[f]}\left(\overleftarrow{\mathbf{R}_1 \cdots \mathbf{R}_1}^{\leftarrow N_1 \rightarrow} \overleftarrow{\mathbf{R}_2 \cdots \mathbf{R}_2}^{\leftarrow N_2 \rightarrow} \cdots \overleftarrow{\mathbf{R}_m \cdots \mathbf{R}_m}^{\leftarrow N_m \rightarrow}\right) = \prod_{\nu=1}^{m} \{\sqrt{N_\nu!}\langle \rho_1 \cdots \rho_{\nu-1}, [N_\nu] | \rho_1 \cdots \rho_\nu \rangle\} \times \varphi_{a,b}^{(nm)[f]}(R), \tag{4.16}$$

where $\rho_\nu = r_{N_1+\cdots+N_{\nu-1}+1} \cdots r_{N_1+\cdots+N_\nu}$ ($\nu = 1 \sim m$), $N_1 + \cdots + N_m = N$, and $f_1 + \cdots + f_n = N$. Furthermore, b in Eq. (4.16) is the Gel'fand pattern

$$|b\rangle \equiv \left| \begin{matrix} [\rho_1 \cdots \rho_{m-1}] \\ \vdots \\ [\rho_1] \end{matrix} \right\rangle \tag{4.17}$$

and N_ν are the weights of b. In particular, $\rho_1 = \overleftarrow{1\,1 \cdots 1}^{\leftarrow N_1 \rightarrow}$, $[\rho_1] = [N_1]$ and $\langle \cdot, [N_1] | \rho_1 \rangle = 1$. Let us now change $n \to N$, $a \to \sigma_1\sigma_2\cdots\sigma_n$ in Eq. (4.16) and adopt a similar discussion for the back-side vector coupling expression.[15] By this process, one obtains, for $\hat{A}^{(nm)}$ in Eq. (4.6),

$$\varphi_{\sigma_1\sigma_2\cdots\sigma_n,\rho_1\rho_2\cdots\rho_m}^{(NN)[f]}(\hat{A}^{(nm)}) = \prod_{\mu=1}^{n} \{\sqrt{P_\mu!}\langle \sigma_1 \cdots \sigma_{\mu-1}, [P_\mu] | \sigma_1 \cdots \sigma_\mu \rangle\} \times \prod_{\nu=1}^{m} \{\sqrt{N_\nu!}\langle \rho_1 \cdots \rho_{\nu-1}, [N_\nu] | \rho_1 \cdots \rho_\nu \rangle\} \times \varphi_{a,b}^{(nm)[f]}(A), \tag{4.18}$$

where $[f] = [f_1 \cdots f_l]$ ($l = \mathrm{Min}\{n,m\}$) with $f_1 + \cdots + f_l = N$ and a, σ_μ, P_μ are related in the similar manner as in b, ρ_ν, N_ν.

To prove Eq. (4.13), we substitute Eqs. (4.6) and (4.18) into Eq. (4.1), and transform the phase factors and the recoupling coefficients by the use of Eqs. (3.22) and (3.6). Thus we obtain

$$\det\{\hat{A}^{(nm)} \otimes V\} = \sum_{[f]}{}' \sum_{\sigma_1\cdots\sigma_n} \sum_{\rho_1\cdots\rho_m} \prod_{\mu=1}^{n} \{\kappa([\sigma_1 \cdots \sigma_{\mu-1}][P_\mu]; [\sigma_1 \cdots \sigma_\mu])\sqrt{P_\mu!}\langle \widetilde{\sigma_1 \cdots \sigma_\mu} | \widetilde{\sigma_1 \cdots \sigma_{\mu-1}}, [1^{P_\mu}] \rangle\} \times \prod_{\nu=1}^{m} \{\kappa([\rho_1 \cdots \rho_{\nu-1}][N_\nu]; [\rho_1 \cdots \rho_\nu])\sqrt{N_\nu!} \times \langle \widetilde{\rho_1 \cdots \rho_\nu} | \widetilde{\rho_1 \cdots \rho_{\nu-1}}, [1^{N_\nu}] \rangle\} \times \varphi_{a,b}^{(nm)[f]}(A)\, \varphi_{\widetilde{\sigma_1\cdots\sigma_n},\widetilde{\rho_1\cdots\rho_m}}^{(NN)[\tilde{f}]}(V). \tag{4.19}$$

The summation over $\sigma_1 \cdots \sigma_n$ and $\rho_1 \cdots \rho_m$ in Eq. (4.19) is taken in two steps; the first step is the summation over a and b with the constraints of the weights Eq. (4.10), and the second step is over $\sigma_1 \cdots \sigma_n$ and $\rho_1 \cdots \rho_m$ under the constraints

$$[\sigma_1 \cdots \sigma_\mu] = [f_\mu] \quad (\mu = 1 \sim n), \tag{4.20a}$$

and

$$[\rho_1 \cdots \rho_\nu] = [g_\nu] \quad (\nu = 1 \sim m). \tag{4.20b}$$

Thus, Eq. (4.19) becomes the right-hand side of Eq. (4.13) with, for example,

$$\kappa([f]a) = \prod_{\mu=1}^{n} \kappa([f_{\mu-1}][P_\mu]; [f_\mu]), \tag{4.21}$$

and

$$\tilde{\varphi}^{(NN)[\tilde{f}]}_{a,b}(V)$$
$$= \sum_{\sigma_1\cdots\sigma_n}{}' \sum_{\rho_1\cdots\rho_m}{}' \prod_{\mu=1}^{n} \langle \widetilde{\sigma_1\cdots\sigma_\mu} | \widetilde{\sigma_1\cdots\sigma_{\mu-1}}, [1^{P_\mu}] \rangle$$
$$\times \prod_{\nu=1}^{m} \langle \widetilde{\rho_1\cdots\rho_\nu} | \widetilde{\rho_1\cdots\rho_{\nu-1}}, [1^{N_\nu}] \rangle$$
$$\times \varphi^{(NN)[\tilde{f}]}_{\widetilde{\sigma_1\cdots\sigma_n},\widetilde{\rho_1\cdots\rho_m}}(V), \tag{4.22}$$

where Σ' denotes the summation under the condition expressed by Eq. (4.20). Equation (4.14) is derived by using Eq. (3.25). The equivalence between Eqs. (4.22) and (4.11) is apparent from the definition of the MF recoupling coefficients. □

Theorem 2 is the main result of this investigation. We will use it in the next section to calculate the matrix elements of double-coset generators between SU_4 spin–isospin wave functions of simple two-cluster systems.

V. APPLICATION OF THE SECOND EXPANSION FORMULA TO SIMPLE CLUSTER SYSTEMS

In this section, we give an example of the application of the second expansion formula with regard to GCM NK of simple two-cluster systems. Suppose C_1 and C_2 are s-shell clusters with mass numbers N_1 and N_2 ($N_1, N_2 \leqslant 4$), respectively, and $N = N_1 + N_2$ is the total mass number. Assuming (nonorthogonal) single-particle wave functions φ_1, φ_2 (or ψ_1, ψ_2) for the bra (or the ket) states, we calculate

$$G^{[f]}_N = \langle \varphi_1^{N_1} \varphi_2^{N_2} \left[U_{[1^{N_1}]}(\chi_1\cdots\chi_{N_1}) \right.$$
$$\times \left. U_{[1^{N_1}]}(\chi_{N_1+1}\cdots\chi_N) \right]_{[\tilde{f}]c} | \mathscr{A} | \psi_1^{N_1} \psi_2^{N_2}$$
$$\times \left[U_{[1^{N_1}]}(\chi_1\cdots\chi_{N_1}) U_{[1^{N_2}]}(\chi_{N_1+1}\cdots\chi_N) \right]_{[\tilde{f}]c} \rangle, \tag{5.1}$$

where $U_{[1^N]c}(\chi_1\cdots\chi_N)$ are $4\times N$ SDG polynomials, defined in Eq. (4.12), representing spin–isospin wave functions of s-shell clusters, and $\varphi_1^{N_1}\varphi_2^{N_2}$ is the shorthand notation for $\varphi_1(\mathbf{x}_1)\cdots\varphi_1(\mathbf{x}_{N_1})\varphi_2(\mathbf{x}_{N_1+1})\cdots\varphi_2(\mathbf{x}_N)$ and so on. Furthermore,

$$\mathscr{A} = \frac{1}{N_1!N_2!} \sum_{P\in S_N} \delta_P \hat{P}, \tag{5.2}$$

and the vector couplings in Eq. (5.1) are achieved by using SU_4 CG coefficients. The possible SU_4 IR labels $[f]$ are the AR of $[f] = [N-p,p]$ with $p = \mathrm{Max}\{0, N-4\} \sim \mathrm{Min}\{N_1, N_2\}$; namely, $[\tilde{f}] = [2^p 1^{N-2p}]$. Using the property of antisymmetry within C_1 and C_2, we can transform $\mathscr{A}$ in Eq. (5.1) as

$$\mathscr{A} \to \sum_{x=0}^{\mathrm{Min}\{N_1,N_2\}} (-1)^x \binom{N_1}{x}\binom{N_2}{x} \hat{P}_x, \tag{5.3}$$

where

$$\hat{P}_x = \hat{P}_{N_1-x+1,N_1+1}\hat{P}_{N_1-x+2,N_1+2}\cdots\hat{P}_{N_1,N_1+x} \tag{5.4}$$

is the so-called double-coset generator[18] corresponding to the number x of exchanged nucleons. Taking the overlaps in the spatial part, we obtain

$$G^{[f]}_N = \sum_{x=0}^{\mathrm{Min}\{N_1,N_2\}} C^{[\tilde{f}]}_x A_{11}^{N_1-x} (A_{12}A_{21})^x A_{22}^{N_2-x}, \tag{5.5}$$

where we have defined

$$A_{ij} = \langle \varphi_i | \psi_j \rangle \quad (i,j = 1 \text{ or } 2), \tag{5.6}$$

$$C^{[\tilde{f}]}_x = (-1)^x \binom{N_1}{x}\binom{N_2}{x} P_x(N_1N_2;[\tilde{f}]), \tag{5.7}$$

with

$$P_x(N_1N_2;[\tilde{f}])$$
$$= \langle \left[U_{[1^{N_1}]}(\chi_1\cdots\chi_{N_1}) U_{[1^{N_2}]}(\chi_{N_1+1}\cdots\chi_N) \right]_{[\tilde{f}]c} | \hat{P}_x |$$
$$\times \left[U_{[1^{N_1}]}(\chi_1\cdots\chi_{N_1}) U_{[1^{N_2}]}(\chi_{N_1+1}\cdots\chi_N) \right]_{[\tilde{f}]c} \rangle. \tag{5.8}$$

If we use a relationship such as

$$U_{[1^{N_1}]}(\chi_1\cdots\chi_{N_1})$$
$$= \left[U_{[1^{N_1-x}]}(\chi_1\cdots\chi_{N_1-x}) \right.$$
$$\times \left. U_{[1^x]}(\chi_{N_1-x+1}\cdots\chi_{N_1}) \right]_{[1^{N_1}]c}, \tag{5.9}$$

we find that $P_x(N_1N_2;[\tilde{f}])$ in Eq. (5.8) is nothing but a special type of MF 9-$[f]$ coefficient; namely,

$$P_x(N_1N_2;[\tilde{f}]) = \begin{bmatrix} [1^{N_1-x}] & [1^x] & [1^{N_1}] \\ [1^x] & [1^{N_2-x}] & [1^{N_2}] \\ [1^{N_1}] & [1^{N_2}] & [\tilde{f}] \end{bmatrix}. \tag{5.10}$$

There are many ways to calculate Eq. (5.8). One method is to use Eqs. (3.22) and (2.21), and transform Eq. (5.10) into the corresponding 9-$[f]$ coefficient [multiplied by $(-1)^x$] with all IR labels converted into AR. Since two-row 9-$[f]$ coefficients are equivalent to angular-momentum 9-j coefficients,[19] we know that Eq. (5.10) is simply a (stretched) 9-j coefficient.[18] Here we derive the explicit expression of Eq. (5.8) by using the second expansion formula of Sec. IV. To do this, we should note that

$$\left[U_{[1^{N_1}]}(\chi_1\cdots\chi_{N_1}) U_{[1^{N_2}]}(\chi_{N_1+1}\cdots\chi_N) \right]_{[\tilde{f}]c}$$
$$= \tilde{\varphi}^{(4N)[\tilde{f}]}_{c,\tilde{b}}(\chi) \tag{5.11}$$

is a one-side MDG polynomial introduced in Eq. (4.11a) with the $4\times N$ matrix variable $\chi = (\chi_1\cdots\chi_N)$ and

$$\left| \begin{matrix} [f] \\ b \end{matrix} \right\rangle = \left| \begin{matrix} N-p\,p \\ N_1 \end{matrix} \right\rangle. \tag{5.12}$$

Then, using the completeness and transformation formula of SDG polynomials, we can easily prove

$$P_x(N_1,N_2;[\tilde{f}]) = \frac{1}{N_H[\tilde{f}]} \tilde{\varphi}^{(NN)[\tilde{f}]}_{b,\tilde{b}}(P_x), \tag{5.13}$$

where P_x is the matrix representation of $\hat{P}_x$. Let us now set $n = m = 2, P_1 = N_1, P_2 = N_2, V = P_x$ in Eq. (4.13), multiply it by $\varphi^{(22)[f]}_{b,b}(A)$ and integrate over $d\mu(A)$. Then, using $\det\{\hat{A}^{(22)} \otimes P_x\} = A_{11}^{N_1-x} \times (-A_{12}A_{21})^x A_{22}^{N_2-x}$ and the explicit expression of 2×2 DG polynomials[15]

$$\varphi^{(22)[N-p,p]}_{[N_1],[N_1]}(A)$$
$$= N_H[N-p,p] \cdot \sum_{q=0}^{\mathrm{Min}\{N_1-p,N_2-p\}} \binom{N_1-p}{q}\binom{N_2-p}{q}$$
$$\times A_{11}^{N_1-p-q} (A_{12}A_{21})^q A_{22}^{N_2-p-q} (\det A)^p, \tag{5.14}$$

we obtain

$$C_x^{[2^p 1^{N-2p}]}$$

$$= (-1)^x \binom{N_1}{x}\binom{N_2}{x} \times \begin{bmatrix} [1^{N_1-x}] & [1^x] & [1^{N_1}] \\ [1^x] & [1^{N_2-x}] & [1^{N_2}] \\ [1^{N_1}] & [1^{N_2}] & [2^p 1^{N-2p}] \end{bmatrix}$$

$$= \sum_{q=\mathrm{Max}\{0,x-p\}}^{\mathrm{Min}\{x,N_1-p,N_2-p\}} (-1)^{x-q} \binom{p}{x-q} \times \binom{N_1-p}{q}\binom{N_2-p}{q}. \tag{5.15}$$

If we substitute Eq. (5.15) into Eq. (5.5) and change the summation order of x and q, we also find

$$G_N^{[N-p,p]} = \frac{1}{N_H[N-p,p]} \varphi_{[N_1],[N_1]}^{(22)[N-p,p]}(A). \tag{5.16}$$

Namely, GCM NK of two-s-shell-cluster systems are represented by the representation matrices of GL(2;C).[22,32]

For physical applications, we also need to calculate NK in spin–isospin representation; namely,

$$(G_N)^{TS}_{T_1S_1T_2S_2;T_1'S_1'T_2'S_2'} \equiv \langle \varphi_1^{N_1}\varphi_2^{N_2}\left[U_{[1^{N_1}]T_1S_1}(\chi_1\cdots\chi_{N_1}) \times U_{[1^{N_2}]T_2S_2}(\chi_{N_1+1}\cdots\chi_N)\right]_{TT_zSS_z} | \times \mathscr{A} | \psi_1^{N_1}\psi_2^{N_2}\left[U_{[1^{N_1}]T_1'S_1'}(\chi_1\cdots\chi_{N_1}) \times U_{[1^{N_2}]T_2'S_2'}(\chi_{N_1+1}\cdots\chi_N)\right]_{TT_zSS_z}\rangle, \tag{5.17}$$

in which SU_4 internal quantum numbers in $U_{[1^N]TT_zSS_z}(\chi_1\cdots\chi_N)$ are specified in spin–isospin representation and the vector couplings in Eq. (5.17) contain angular-momentum CG coefficients. By using reduced SU_4 CG coefficients in spin–isospin representation, Eqs. (5.17) and (5.1) are found to be related by the equation

$$(G_N)^{TS}_{T_1S_1T_2S_2;T_1'S_1'T_2'S_2'} = \sum_{[\tilde{f}]} B^{[\tilde{f}]TS}_{T_1S_1T_2S_2;T_1'S_1'T_2'S_2'} \cdot G_N^{[f]}, \tag{5.18}$$

where we have defined, for fixed N_1 and N_2,

$$B^{[\tilde{f}]TS}_{T_1S_1T_2S_2;T_1'S_1'T_2'S_2'} \equiv \langle [1^{N_1}]T_1S_1[1^{N_2}]T_2S_2 \| [\tilde{f}]TS\rangle \times \langle [1^{N_1}]T_1'S_1'[1^{N_2}]T_2'S_2' \| [\tilde{f}]TS\rangle. \tag{5.19}$$

These SU_4 CG coefficients can be calculated by extending Jahn's idea[33] in shell-model calculations. We first apply the double-coset expansion of $\mathscr{A}$ to Eq. (5.17). Then, following almost the same process as in the SU_4-representation case, we obtain

$$(G_N)^{TS}_{T_1S_1T_2S_2;T_1'S_1'T_2'S_2'} = \sum_{x=0}^{\mathrm{Min}\{N_1,N_2\}} C_x^{TS}(T_1S_1T_2S_2;T_1'S_1'T_2'S_2') \times A_{11}^{N_1-x}(A_{12}A_{21})^x A_{22}^{N_2-x}, \tag{5.20}$$

where

$$C_x^{TS}(T_1S_1T_2S_2;T_1'S_1'T_2'S_2') = (-1)^x \binom{N_1}{x}\binom{N_2}{x} \times \sum_{\substack{T_{11}S_{11}T_{12}S_{12} \\ T_{21}S_{21}T_{22}S_{22}}} \langle [1^{N_1-x}]T_{11}S_{11}[1^x]T_{12}S_{12} \| [1^{N_1}]T_1S_1\rangle \times \langle [1^x]T_{21}S_{21}[1^{N_2-x}]T_{22}S_{22} \| [1^{N_2}]T_2S_2\rangle \times \langle [1^{N_1-x}]T_{11}S_{11}[1^x]T_{21}S_{21} \| [1^{N_1}]T_1'S_1'\rangle \times \langle [1^x]T_{12}S_{12}[1^{N_2-x}]T_{22}S_{22} \| [1^{N_2}]T_2'S_2'\rangle \times \hat{T}_1\hat{T}_2\hat{T}_1'\hat{T}_2'\hat{S}_1\hat{S}_2\hat{S}_1'\hat{S}_2' \times \begin{Bmatrix} T_{11} & T_{12} & T_1 \\ T_{21} & T_{22} & T_2 \\ T_1' & T_2' & T \end{Bmatrix} \begin{Bmatrix} S_{11} & S_{12} & S_1 \\ S_{21} & S_{22} & S_2 \\ S_1' & S_2' & S \end{Bmatrix} \tag{5.21}$$

with $\hat{a} = \sqrt{2a+1}$. As can be easily seen from Eqs.(5.20) and (5.5), Eq. (5.18) is true for the coefficients of each x-nucleon exchange; i.e.,

$$C_x^{TS}(T_1S_1T_2S_2;T_1'S_1'T_2'S_2') = \sum_{[\tilde{f}]} B^{[\tilde{f}]TS}_{T_1S_1T_2S_2;T_1'S_1'T_2'S_2'} \cdot C_x^{[\tilde{f}]} \tag{5.22}$$

for $x = 0 \sim \mathrm{Min}\{N_1,N_2\}$.

Since $C_0^{[\tilde{f}]} = 1$ and

$$C_0^{TS}(T_1S_1T_2S_2;T_1'S_1'T_2'S_2') = \delta_{T_1T_1'}\delta_{S_1S_1'}\delta_{T_2T_2'}\delta_{S_2S_2'},$$

Eq. (5.22) for $x = 0$ gives us simply the orthogonality relationship of SU_4 CG coefficients. Using this and $x = 1$ case of Eqs. (5.22), (5.15), and (5.21) we can determine recursively the coefficients $B^{[\tilde{f}]TS}_{T_1S_1T_2S_2;T_1'S_1'T_2'S_2'}$ for all combinations of N_1 and N_2 (N_1, $N_2 \leqslant 4$).

We should note, however, that for many combinations of N_1 and N_2, the summation over $[\tilde{f}]$ in Eq. (5.22) is uniquely determined by T and S. In these cases, the coefficients in Eq. (5.22) are unity and NK in spin–isospin representation are automatically obtained from those in SU_4 representation. Furthermore, $B^{[\tilde{f}]TS}_{T_1S_1T_2S_2;T_1'S_1'T_2'S_2'}$ satisfies the following symmetry relations:

$$B^{[\tilde{f}]TS}_{T_1S_1T_2S_2;T_1'S_1'T_2'S_2'} = B^{[\tilde{f}]TS}_{T_1'S_1'T_2'S_2';T_1S_1T_2S_2} = B^{[\tilde{f}]ST}_{S_1T_1S_2T_2;S_1'T_1'S_2'T_2'}. \tag{5.23}$$

The coefficients $B^{[\tilde{f}]TS}_{T_1S_1T_2S_2;T_1'S_1'T_2'S_2'}$ for nontrivial cases of $N_2 \leqslant N_1 \leqslant 4$ are given in Table I under the phase convention

$$[\langle [1]_{\frac{1}{2}}{}_{\frac{1}{2}}[1]_{\frac{1}{2}}{}_{\frac{1}{2}} \| [11]10\rangle \times \langle [1]_{\frac{1}{2}}{}_{\frac{1}{2}}[1]_{\frac{1}{2}}{}_{\frac{1}{2}} \| [11]01\rangle] > 0, \tag{5.24}$$

which is related to the relative phase of spin-isospin wave functions of the deuteron and the quasideuteron. Using these coefficients and Eqs. (5.18), (5.16) and (5.14), we can derive easily the GCM NK of Eq. (5.17) in spin–isospin representation.

VI. CONCLUSION

Special double Gel'fand (SDG) polynomials,[23] with a partition of N and each of the weights equal to 1, are shown

TABLE I. Nontrivial values of $B^{[\tilde{f}]TS}_{T_1S_1T_2S_2;T_1'S_1'T_2'S_2'}$.

$N_1 N_2$	TS	$T_1S_1T_2S_2\,;\,T_1'S_1'T_2'S_2'$	$[\tilde{f}]/B$	
			[111]	[21]
		$10\frac{1}{2}\frac{1}{2};10\frac{1}{2}\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
2 1	$\frac{1}{2}\frac{1}{2}$			
		$10\frac{1}{2}\frac{1}{2};01\frac{1}{2}\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$
			[1111]	[22]
		1010;1010	$\frac{1}{2}$	$\frac{1}{2}$
	00			
		1010;0101	$-\frac{1}{2}$	$\frac{1}{2}$
2 2				
			[211]	[22]
		1001;1001	$\frac{1}{2}$	$\frac{1}{2}$
	11			
		1001;0110	$-\frac{1}{2}$	$\frac{1}{2}$
			[2111]	[221]
		$\frac{1}{2}\frac{1}{2}10;\frac{1}{2}\frac{1}{2}10$	$\frac{1}{2}$	$\frac{1}{2}$
3 2	$\frac{1}{2}\frac{1}{2}$			
		$\frac{1}{2}\frac{1}{2}10;\frac{1}{2}\frac{1}{2}01$	$\frac{1}{2}$	$-\frac{1}{2}$

to be very useful to discuss some of the fundamental quantities in the irreducible representation (IR) theory of the S_N group. Using extensively the vector-coupling expressions of $U_n \times S_N$ SDG polynomials, we have derived a simple relationship between multiplicity-free (MF) recoupling coefficients of the S_N group and special types of MF Clebsch-Gordan (CG) coefficients of the unitary group U_n. Since the explicit expressions of these MF U_n CG coefficients are already well known, we can derive explicit expressions[27] of the MF recoupling coefficients using this relationship. The Yamanouchi formula[24] for the transposition of the last two particles can also be derived, since it is represented by simple MF recoupling coefficients. The standard phase convention in this formula and that of the MF recoupling coefficients by Horie[27] are shown to result as a consequence of natural phase conventions of DG polynomials and the positiveness of these MF U_n CG coefficients.

For the purpose of applying SDG polynomials to spin–isospin wave functions of many-cluster systems, two kinds of expansion formulas of the determinant associated with a matrix tensor product have been proved with explicit expressions of phase factors. The first expansion formula is expressed in terms of SDG polynomials, while the second one is expressed in terms of standard DG polynomials and modified DG (MDG) polynomials with S_N internal quantum numbers specified by direct-product representations of its subgroups. The second expansion formula is then applied in Sec. V to calculate normalization kernels (NK) in the generator coordinate method (GCM) for simple two-body systems composed of s-shell clusters. Several nontrivial SU_4 CG coefficients are calculated in a simple derivation of these GCM NK in spin-isospin representation. The usefulness of the second expansion formula has also been demonstrated in Ref. 32 for NK problems of more complicated many-cluster systems. There, this formula was used to construct generating functions of spin–isospin wave functions in the SU_4 scheme and to generate GCM NK from the determinant associated with the overlap matrix of single particle wave functions.

The idea to apply the concept of DG polynomials to NK problems of cluster systems has also been proposed by Kramer and others.[22] However, it should be noted that there is a basic difference between our treatment and theirs. Namely, they started from the representation theory of the S_N group and tried to construct DG polynomials from representation matrices of double-coset generators by the subduction process. On the other hand, we started from the theory of DG polynomials and have derived some essential properties used in the representation theory of the S_N group. One advantage of our method is that we can discuss the phase-convention problem of the U_n and S_N groups in a unified manner. For example, the phase factor of Eq. (3.25) is explicitly given in this paper. The essential point of our approach to the NK problems is to use the powerful second expansion formula which has been proved entirely within the framework of the representation theory of the unitary group. By using this formula, we have shown that the treatment of the conjugate wave functions of spatial and spin–isospin parts for the permutation symmetry can be carried out in a very efficient manner.

ACKNOWLEDGMENTS

One of the authors (Y.F.) is indebted to Professor K. T. Hecht for helpful discussions during a visit to the University of Michigan, and also for a careful reading of the manuscript.

This research was supported in part by the U. S. Department of Energy under Contract No. DOE/DE-AC02-79 ER10364.

APPENDIX: PROOF OF EQS. (3.25) AND (3.26)

Noting that r' and $r = r'\rho$ in Eq. (3.22) are both standard Young tableau, we decompose $(-1)^r$ into

$$(-1)^r = (-1)^{r'\rho} = (-1)^{r'}(-1)^{[[r']\rho]}(-1)^{\rho}, \tag{A1}$$

where

$$(-1)^{[[r']\rho]} \equiv \prod_{\substack{i<j \\ i\in r',\, j\in\rho}} \epsilon_{ij} = \prod_{\substack{i\in[r'] \\ j\in[r'\rho]-[r']}} \epsilon_{ij} \tag{A2}$$

depends only on $[r']$ and $[r'\rho]$ since the permutations of $i = 1,2,\ldots,N-m$ or $j = N-m+1,\ldots,N$ in the Young tableau for each group change only the order of the product of $\Pi_{i\in r'}\Pi_{j\notin\rho}\epsilon_{ij}$. In Eq. (1), $(-1)^{\rho}$ is defined by Eq. (2.17) with r replaced by ρ, while $[r'\rho]-[r']$ in Eq. (2) denotes the aggregate of the boxes formed by ρ. Setting the numbers $N-m+1,\ldots,N$ in ρ so that $0\leqslant r_{N-m+1}\leqslant\cdots\leqslant r_N\leqslant n$, we easily obtain

$$(-1)^{[[r']\rho]} = \prod_{1=\mu<\nu}^{n} (-1)^{(f_\mu - q_\mu)q_\nu}, \tag{A3}$$

where $[r'] = [q] = [q_1\cdots q_n]$ and $[r] = [r'\rho] = [f] = [f_1\cdots f_n]$. Next, to calculate $(-1)^{\rho}$ and $(-1)^{\tilde{\rho}}$, we choose $0\leqslant\tilde{r}_{N-m+1}<\cdots<\tilde{r}_N\leqslant n$. For this specific order, we obtain

$$(-1)^{\rho} = \prod_{1=\mu<\nu}^{n} (-1)^{(f_\mu - q_\mu)(f_\nu - q_\nu)}, \tag{A4}$$

$$(-1)^{\tilde{\rho}} = 1.$$

Then, applying Eqs. (3.23) and (3.24) to Eq. (3.22), we obtain Eq. (3.25). Equation (3.26) is easily proved if we note that $(-1)^{\rho+\tilde{\rho}}$ is independent of the order of $r_{N-m+1}\cdots r_N$ from Eqs. (1) and (2.18).

[1]J. Schwinger, "On Angular Momentum," in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and C. Van Dam (Academic, New York, 1965), p. 229.
[2]V. Bargmann, Rev. Mod. Phys. **34**, 829 (1962).
[3]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[4]M. Moshinsky, J. Math. Phys. **4**, 1128 (1963).
[5]J. D. Louck, Am. J. Phys. **38**, 3 (1970).
[6]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1931).
[7]I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk SSSR **71**, 825 (1950); I. M. Gel'fand and M. I. Graev, Izv. Akad. Nauk SSSR Mekh. Mashinostr. **29**, 1329 (1965) [Am. Math. Soc. Transl. **65**, 116 (1967), Ser. 2].
[8]The terminology "double Gel'fand" has been used by many authors; for instance, see Refs. 13, 21, and 22 below. Furthermore, we suppress the word "boson" because we have in mind the polynomials of complex variables (Ref. 2).
[9]J. G. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965); Rev. Mex. Fis. **XIV**, 29 (1965).
[10]M. Moshinsky and E. Chacón, "Racah Coefficients and States with Permutation Symmetry," in *Spectroscopic and Group Theoretical Method in Physics*, Racah Memorial Volume, edited by F. Bloch, S. G. Cohen, A. deShalit, S. Sambursky, and I. Talmi (North-Holland, Amsterdam, 1968), p. 99.
[11]L. C. Biedenharn, A. Giovannini, and J. D. Louck, J. Math. Phys. **8**, 691 (1967).
[12]J. D. Louck and L. C. Biedenharn, J. Math. Phys. **14**, 1336 (1973).
[13]A. Antillon and T. H. Seligman, J. Math. Phys. **23**, 473 (1982).
[14]W. G. Harter and C. W. Patterson, "A Unitary Calculus for Electronic Orbitals," in *Lecture Notes in Physics 49* (Springer-Verlag, Berlin, 1976).
[15]Y. Fujiwara and H. Horiuchi, Mem. Fac. Sci. Kyoto Univ. Ser. A Phys. Astrophys. Geophys. Chem. **XXXVI**, 197 (1983).
[16]Y. Fujiwara and H. Horiuchi, Prog. Theor. Phys. **63**, 895 (1980); **65**, 1632 (1981); **65**, 1901 (1981).
[17]G. de B. Robinson, *Representation Theory of the Symmetric Group* (Univ. Toronto, Toronto, 1961).
[18]P. Kramer and T. H. Seligman, Nucl. Phys. A **136**, 545 (1969).
[19]P. Kramer, Z. Phys. **216**, 68 (1968).
[20]P. Kramer and M. Moshinsky, "Group Theory of Harmonic Oscillators and Nuclear Structure," in *Group Theory and its Applications*, edited by E. M. Loebl (Academic, New York, 1968), p. 339.
[21]P. Kramer, "Finite Representation of the Unitary Group and their Applications in Many-body Physics," in *Proceedings of the Fifth International Colloquium on Group Theoretical Methods in Physics*, Montréal, 1976, edited by R. T. Sharp and B. Kolman (Academic, New York, 1977).
[22]P. Kramer, G. John, and D. Schenzle, *Group Theory and the Interaction of Composite Nucleon Systems, Clustering Phenomena in Nuclei, Vol. 2* (Vieweg, Braunschweig, 1980).
[23]M. Moshinsky, J. Math. Phys. **7**, 691 (1966).
[24]T. Yamanouchi, Proc. Phys. Math. Soc. Jpn. **19**, 436 (1937).
[25]J. P. Elliott, J. Hope, and H. A. Jahn, Philos. Trans. R. Soc. A **246**, 241 (1953).
[26]I. G. Kaplan, J. Expl. Theor. Phys. USSR **41**, 560 (1961) [Soviet Phys. JETP **14**, 401 (1962)].
[27]H. Horie, J. Phys. Soc. Jpn. **19**, 1783 (1964).
[28]For example, see M. Hamermesh, *Group Theory and its Application to Physical Problems* (Addison-Wesley, Reading, MA, 1962).
[29]M. Ciftan and L. C. Biedenharn, J. Math. Phys. **10**, 221 (1969).
[30]The technique used in this section was suggested in Ref. 19 based on the idea discussed in Ref. 10. However, it should be noted that the two-side vector coupling expressions of DG polynomials and the concept of Bargmann integrals are used very efficiently in our discussion, as in Ref. 15.
[31]E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge U. P., Cambridge, 1935).
[32]Y. Fujiwara, Y. C. Tang, and H. Horiuchi, Prog. Theor. Phys. **70**, 809 (1983).
[33]H. A. Jahn and H. van Wieringen, Proc. R. Soc. London Ser. A **209**, 502 (1951).

General charge conjugation operators in simple Lie groups

R. V. Moody[a)]
Centre de Recherches de Mathématiques Appliquées, Université de Montréal, Montréal, Québec, Canada

J. Patera[b)]
California Institute of Technology, Pasadena, California 91125

(Received 22 February 1984; accepted for publication 18 May 1984)

A description of particular elements ("charge conjugation operators") found in any compact simple Lie group K is presented. Such elements R_i transform a physical state (weight vector of a basis of a representation space) into others with opposite "charge" (ith component of the weight), sometime changing also the sign of the state. It is demonstrated that exploitation of these elements and the finite subgroup N of K generated by them offer new powerful methods for computing with representations of the Lie group. Their application to construction of bases in representation spaces is considered in detail. It represents a completely new direction to the problem.

PACS numbers: 02.20. + b, 11.30.Er

I. INTRODUCTION

In this article we study certain elements R of order four, i.e., $R^4 = 1$, in connected compact simple Lie groups in order to demonstrate that they provide a new and powerful tool for applications. Although their importance has long been understood in the theory of Lie groups,[1] these elements have so far not been used in physics literature except for Refs. 2 and 3 (which are based entirely on this work) and they appear here for the first time in what might be called the theory of computation with Lie groups.

Intuitively these elements can be viewed in the following way: Given a simple Lie group K of rank l, then in a description of relevant physical states $|\lambda_1\lambda_2\cdots\lambda_l\rangle$, which are weight vectors in a representation space V of K, an important role is played by "quantum numbers" or "charges" λ_i, $i = 1,2...,l$, which are defined as eigenvalues of l suitably chosen linearly independent "diagonal" elements of the Lie algebra of K. The subject of our article is the elements R_i, $i = 1,2,...,l$, of K which permute the weight vectors of the same K-multiplet in such a way that $\lambda_i \to -\lambda_i$ if $\lambda_i \neq 0$. For lack of any better name we call R_i the charge conjugation operators (CCO) although it is only in special situations that one of them may coincide with the usual operator reversing electric charge.[4] It turns out that the action of R_i on $|\lambda_1\lambda_2\cdots\lambda_l\rangle$ is quite nontrivial. Besides reversing the charges (components of weights) they sometimes reverse the sign of the state or permute several states with the same "quantum numbers" (weights) when $\lambda_i = 0$. Let us underline the fact that there are no charge conjugating elements in K which would be of order 2 in all finite-dimensional representations of K.

The role which R_i may play in applications far exceeds the charge conjugation. In that respect Refs. 2 and 3, where they provide the main tool of the approach, are only modest illustrations of the possibilities. There all nonzero Clebsch–Gordon coefficients arising in a tensor product of two irreducible representation spaces of K are given by a small representative subset of them and any other coefficient is identified with one of the subset using CCO. Fortunately, the economy made this way rapidly increases with the rank l of K roughly being proportional to the order $|W|$ of the Weyl group W of K. Thus for instance, in the case of rank one group SU(2), the saving made by using CCO is the smallest because $|W| = 2$. It is equivalent to the well-known fact that from each pair of SU(2) Clebsch–Gordon coefficients $C(l_1,l_2l; m_2m_2m)$ and $C(l_1l_2l; -m_1 -m_2 -m)$ it suffices to calculate only one of them. However, for SU(n) the economy provided by CCO increases as $n!$.

In particle physics it is conceivable that the usual requirements of invariance of a (grand unified) model under the action of a reductive Lie subgroup K' of a semisimple group K is too strong and that all the conclusions drawn from the model would follow requiring only the invariance under the action of a finite subgroup F of K generated by R_i and possibly some other elements of finite order in K. In general, the finite subgroup N of K generated by R_i is of importance whenever K appears, even if its role so far has not been fully appreciated.

So far the possibilities of building N-invariant models which are not K'-invariant remain completely unexplored. They would closely resemble the K'-invariant ones in that the K-multiplets would be formed as direct sums of N-multiplets, but they would be simpler because N as a finite group has only finitely many irreducible representations.

Questions of this type motivate our undertaking although we do not address them directly in the article.

The first objective of the paper is to bring together what is known about CCO in a coherent way.

The second objective is related to the problem of construction of bases in representation spaces. Until now, in spite of the obvious importance of the problem, there is no satisfactory general method of construction. [*Note added in proof*: Daya-Nand Verma has produced an as yet unproved algorithm for constructing bases by the first of the methods below. It appears very promising.] Indeed, there are three well-known ways how to construct a basis. The first is a multiple application of generators to one basis vector. Although this is a general method in principle, practically it is so unruly that it is of use in spaces of low dimension only. Even sophisticated versions[5,6] developed for purely theoreti-

[a)] On leave from: Department of Mathematics, University of Saskatchewan, Saskatchewan, Canada.

[b)] On leave from: Centre de recherches de mathématiques appliquées, Université de Montréal, Montréal, Québec, Canada.

 0022-2488/84/102838-10$02.50

cal reasons did not make it any more useful in applications. The second way is to use chains of reductive subgroups. In special cases[7] this produces perhaps the most explicit and desirable form of representation theory. Unfortunately, in most cases it offers at best only a simplification of the construction. The third method based on subalgebras and their centralizers is practically restricted to low rank groups[8] or to very particular classes of representations.[9] Systematic exploitation of CCO and exploitation of the finite group $N \subset K$ generated by them leads to a new approach to the problem, where the group N plays the role of the Weyl group of K "lifted" into the representation space.

Let us emphasize that we are concerned here with methods which apply to any simple (and by an obvious extension to any reductive) Lie group K and therefore we ignore existing vast literature applicable only to groups of particular type(s).

The most obvious obstacle to construction of basis for representation spaces for anything beyond those of very low dimension is the sheer enormity of the number of vectors to be written down. The natural way out of this is to compute only the dominant weight spaces (which in general make only a tiny fraction of the entire space) and to use the group N to move outside them when necessary for some problem at hand. This approach leads to two fundamental problems

(i) Build a "good" basis for each of the dominant weight spaces.

(ii) Describe how to move about in the rest of the representation space.

Consequently, the second objective of our article is to describe an approach to basis construction, at least as far as it is possible at this time, and to illustrate various aspects of it by numerous examples, because in many cases of practical interest it offers a considerable help already in its present form.

The problem (i) of building orthonormal bases in dominant weight spaces is the truly difficult part of the construction. Our examples in Sec. VI illustrate two approaches to solving it. The first uses tensor products of simple spaces (practical in many situations), the second one involves the representation theory of subgroups of N (eigenspace decomposition of stabilizers of dominant weight vectors). A third approach would be exploitation of various subgroups of K.

Whenever a particular (reductive) subgroup $K' \subset K$ is of importance in an application, it should be reflected in the basis construction. That is, the bases in dominant weight subspaces [problem (i) above] have to be built using N' of K' rather than N of K. Naturally, in the simple situation when the corresponding branching rule for $K' \subset K$ contains each K'- irreducible component at most once, there are the usual shortcuts so that one faces the same problem but for smaller representations of the smaller group K'.

Moving between weight spaces of the representation [problem (ii)] is accomplished by two processes: (a) moving along N-orbits of the space; (b) crossing orbits. The first process is carried out by the group N whose action is completely described in Sec. III. The second process is carried out by transforming dominant weight vectors of one space to another dominant weight subspace. This involves computing the action of a few generators between a few subspaces once for all. The Figs. 3 and 4 illustrate the succinct way in which this information can be presented.

In Sec. II we work out in detail and with only simple means the CCO in the SU(2) case. This serves as an introduction to the general situation described in Secs. III and IV, followed by some examples (Secs. V and VI).

II. THE CHARGE CONJUGATION OPERATOR OF SU(2)

In order to specify an irreducible representation of SU(2) of dimension $L+1$ we use (L), where L is related to the highest weight $\Lambda = j\beta$ of the representation by

$$L = 2(\Lambda,\beta)/(\beta,\beta) = 2j\,. \tag{2.1}$$

Hence j is the familiar "angular momentum." A convenient orthonormal basis consists of the vectors (angular momentum states) denoted by $|{}^{L}_{M}\rangle$, such that

$$M = 2(\mu,\beta)/(\beta,\beta)\,,$$

where $\mu = M\beta/2$ is a weight of the weight system Ω_L of (L). Specifically,

$$M = 2(\Lambda - k\beta,\beta)/(\beta,\beta), \quad k \in \{0,1,..,L\}\,. \tag{2.2}$$

Thus

$$M \in \{L, L-2, \ldots, -L\}\,. \tag{2.3}$$

Assuming $(\beta,\beta) = 2$, the action of the Lie algebra spanned by generators $e_{\pm}$ and h satisfying the commutation relations

$$[e_+, e_-] = h, \quad [h, e_\pm] = \pm 2e_\pm \tag{2.4}$$

on the basis vectors $|{}^{L}_{M}\rangle$ of the space V^L is given by

$$h\,|{}^{L}_{M}\rangle = (\mu,\beta)|{}^{L}_{M}\rangle = M\,|{}^{L}_{M}\rangle\,,$$
$$e_\pm\,|{}^{L}_{M}\rangle = \tfrac{1}{2}\sqrt{(L \mp M)(L \pm M + 2)}\,|{}^{L}_{M\pm 2}\rangle\,. \tag{2.5}$$

There is up to an inversion only one CCO in the rank one group SU(2). It is defined by

$$R = \exp e_- \exp(-e_+)\exp e_- \tag{2.6}$$

from which it follows that

$$R\,|{}^{L}_{M}\rangle = (-1)^{(L-M)/2}|{}^{L}_{\bar{M}}\rangle\,. \tag{2.7}$$

Let us illustrate (2.7) by an example:

$$\begin{aligned} R\,|{}^{3}_{1}\rangle &= \exp e_- \exp(-e_+)(1 + e_- + \tfrac{1}{2}e_-^2 + \cdots)\,|{}^{3}_{1}\rangle \\ &= \exp e_-(1 - e_+ + \tfrac{1}{2}e_+^2 - \tfrac{1}{6}e_+^3 + \cdots) \\ &\quad \times(|{}^{3}_{1}\rangle + 2|{}^{3}_{1}\rangle + \sqrt{3}\,|{}^{3}_{3}\rangle) \\ &= (1 + e_- + \tfrac{1}{2}e_-^2 + \tfrac{1}{6}e_-^3 + \cdots) \\ &\quad \times(\sqrt{3}|{}^{3}_{3}\rangle - |{}^{3}_{1}\rangle) = -|{}^{3}_{1}\rangle\,. \end{aligned} \tag{2.8}$$

Here and through the rest of the article we write $-a$ as $\bar{a}$ in matrixlike symbols. Repeated application of (2.7) gives

$$\begin{aligned} R^2|{}^{L}_{M}\rangle &= (-1)^L\,|{}^{L}_{M}\rangle = (-1)^M|{}^{L}_{M}\rangle\,, \\ R^4|{}^{L}_{M}\rangle &= |{}^{L}_{M}\rangle\,, \end{aligned} \tag{2.9}$$

which demonstrates that R is an element of order 4.

Note that CCO of SU(2) could have been defined by

$$\tilde{R} = \exp e_+ \exp(-e_-)\exp e_+\,, \tag{2.10}$$

which would imply the interchange of M and $-M$ in (2.7). Indeed one can verify directly that

$$\tilde{R}\,|{}^{L}_{M}\rangle = (-1)^{(L+M)/2}|{}^{L}_{\bar{M}}\rangle\,. \tag{2.11}$$

Hence,

$$\tilde{R} = R^{-1}\,. \tag{2.12}$$

The matrix of R relative to the basis $\{|{}^{L}_{M}\rangle\}$ for $L = 0,1,2,3,...$ is

$$R = (1),\quad R = \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix},$$

$$R = \begin{pmatrix} 0 & 0 & 1 \\ 0 & \bar{1} & 0 \\ 1 & 0 & 0 \end{pmatrix},\quad R = \begin{pmatrix} 0 & 0 & 0 & \bar{1} \\ 0 & 0 & 1 & 0 \\ 0 & \bar{1} & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \tag{2.13}$$

and so on. The trace of any of the matrices R in (2.13) is the character of the element $R \in \mathrm{SU}(2)$ in the corresponding representation of SU(2). Thus tr R takes only three distinct values:

$$\mathrm{tr}\,R = \begin{cases} 1 & \text{for } L = 0 \quad (\mathrm{mod}\ 4) \\ 0 & \text{for } L = 1 \text{ or } 3 \quad (\mathrm{mod}\ 4)\,. \\ -1 & \text{for } L = 2 \quad (\mathrm{mod}\ 4) \end{cases} \tag{2.14}$$

It was shown in Ref. 10 that in SU(2) there is only one conjugacy class of elements of order 4 whose character values on irreducible representations are restricted to 0, ± 1,—Kostant's principal element—and, in notations of Kac (cf. Refs. 10, 11, and 12), its conjugacy class is given as $R \sim [11]$.

Subsequently we need the transformation of generators $e_{\pm}$ and h by R. For that consider the equalities

$$\begin{aligned} |{}^{1}_{1}\rangle &= h\,|{}^{1}_{1}\rangle = -R\,|{}^{1}_{1}\rangle \\ &= Rh\,|{}^{1}_{1}\rangle = RhR^{-1}R\,|{}^{1}_{1}\rangle = -RhR^{-1}|{}^{1}_{1}\rangle\,, \\ |{}^{1}_{1}\rangle &= e_{+}\,|{}^{1}_{1}\rangle = -R\,|{}^{1}_{1}\rangle = -Re_{-}|{}^{1}_{1}\rangle \\ &= -Re_{-}R^{-1}R\,|{}^{1}_{1}\rangle = -Re_{-}R^{-1}|{}^{1}_{1}\rangle\,, \\ |{}^{1}_{1}\rangle &= e_{-}|{}^{1}_{1}\rangle = R\,|{}^{1}_{1}\rangle = Re_{+}\,|{}^{1}_{1}\rangle \\ &= Re_{+}R^{-1}R\,|{}^{1}_{1}\rangle = -Re_{+}R^{-1}|{}^{1}_{1}\rangle\,, \end{aligned} \tag{2.15}$$

from which it follows immediately that

$$RhR^{-1} = -h,\quad Re_{\pm}R^{-1} = -e_{\mp}\,. \tag{2.16}$$

Finally, let us also point out that

$$R\,|{}^{L}_{0}\rangle = R^{-1}|{}^{L}_{0}\rangle = (-1)^{L/2}|{}^{L}_{0}\rangle\,. \tag{2.17}$$

III. CHARGE CONJUGATION OPERATORS OF ARBITRARY SIMPLE COMPACT LIE GROUP

All ideas of this section extend naturally to arbitrary simply connected compact Lie groups, but for simplicity we consider here only simple simply connected compact Lie groups. The purpose of this section is to bring together known facts relevant to CCO.

Let $\mathbf{k}$ be the Lie algebra of a simple simply connected compact Lie group K, $\mathbf{g}$ its complexification $\mathbf{k}_{\mathbf{C}}$, and G the simply connected complex group with Lie algebra $\mathbf{g}$ and with maximal compact subgroup K. We let T be a maximal torus of G and $\mathbf{h}$ be the corresponding Cartan subalgebra of $\mathbf{g}$. Thus $\mathbf{h} = \mathbf{t}_{\mathbf{C}}$, where $\mathbf{t}$ is the subalgebra of $\mathbf{g}$ corresponding to T, and $\mathbf{h}_{\mathbf{R}} := \sqrt{-1}\,\mathbf{t}$ is a real Euclidean space (under the Killing form) of dimension $l = \mathrm{rank}(G)$. Here the symbol $:=$ indicates that the left side is defined by the right one.

Relative to $\mathbf{h}$ we have the root space decomposition

$$\mathbf{g} = \mathbf{h} \oplus \bigoplus_{\alpha \in \Delta} \mathbf{g}^{\alpha} \tag{3.1}$$

of $\mathbf{g}$, where $\Delta \subset \mathbf{h}_{\mathbf{R}}^{*}$ (the dual space to $\mathbf{h}_{\mathbf{R}}$) is the root system of $\mathbf{g}$ relative to $\mathbf{h}$, and for each $\alpha \in \Delta$,

$$\mathbf{g}^{\alpha} = \{x \in g \mid [h,x] = \alpha(h)x \quad \text{for all } h \in \mathbf{h}\}. \tag{3.2}$$

Choosing an ordering of $\mathbf{h}_{\mathbf{R}}^{*}$ leads to an ordering on Δ. Let Δ^{+} denote the corresponding set of positive roots in Δ and let $\Pi = \{\alpha_1,...,\alpha_l\}$ be the corresponding set of simple roots.

For each $\beta \in \Delta$, $\mathbf{sl}^{\beta}(2) = \mathbf{g}^{\beta} + \mathbf{g}^{-\beta} + [\,\mathbf{g}^{\beta},\mathbf{g}^{-\beta}\,]$ is a subalgebra of $\mathbf{g}$ isomorphic to $\mathbf{sl}(2)$, and we chose $e_{\beta} \in g^{\beta}$, $e_{-\beta} \in \mathbf{g}^{-\beta}$, and $h_{\beta} \in \mathbf{h}$ such that

$$[h_{\beta},e_{\pm\beta}] = \pm 2\,e_{\pm\beta},\quad [e_{\beta},e_{-\beta}] = h_{\beta}\,. \tag{3.3}$$

These h_{β} are uniquely determined by (3.3) and satisfy

$$\lambda(h_{\beta}) = 2(\lambda,\beta)/(\beta,\beta) \tag{3.4}$$

for all $\lambda \in \bar{h}^{*}$. The choice of $e_{\beta} \in \mathbf{g}^{\beta}$, $e_{\beta} \neq 0$, is arbitrary, whereupon $e_{-\beta}$ is uniquely determined by $[e_{\beta},e_{-\beta}] = h_{\beta}$. At this time we leave this choice free.

Let $G^{\beta} \subset G$ be the connected subgroup whose Lie algebra is $\mathbf{sl}^{\beta}(2)$, and let $\mathrm{SU}^{\beta}(2) = G^{\beta} \cap K = \langle \exp(\mathbf{sl}^{\beta}(2) \cap \mathbf{k}) \rangle$ be the corresponding compact subgroup. Thus $G^{\beta} \cong \mathrm{SL}_{\mathbf{C}}(2)$ and $K^{\beta} \cong \mathrm{SU}(2)$.

A specific isomorphism of G^{β} and $\mathrm{SL}_{\mathbf{C}}(2)$ is established by identifying $\mathbf{sl}^{\beta}(2)$ and sl(2):

$$e_{\beta} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},\quad e_{-\beta} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},\quad h_{\beta} = \begin{pmatrix} 1 & 0 \\ 0 & \bar{1} \end{pmatrix}. \tag{3.5}$$

Consequently,

$$\exp(-e_{\beta}) = \begin{pmatrix} 1 & \bar{1} \\ 0 & 1 \end{pmatrix},\quad \exp e_{-\beta} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \tag{3.6}$$

and

$$R_{\beta} := \exp e_{-\beta} \exp(-e_{\beta}) \exp e_{-\beta} = \begin{pmatrix} 1 & \bar{1} \\ 1 & 0 \end{pmatrix} \in \mathrm{SU}(2) \tag{3.7}$$

as in (2.13).

Now let $\rho : K \to \mathrm{GL}(V)$ be a finite-dimensional (unitary) representation of K on a complex space V and let $d\rho : \mathbf{k} \to \mathrm{End}(V)$ be its differential, i.e., a representation of $\mathbf{k}$ on V. Both ρ and $d\rho$ have complexifications,

$$\rho_{\mathbf{C}} : G \to \mathrm{GL}(V),$$

$$d\rho_{\mathbf{C}} : \mathbf{g} \to \mathrm{End}(V).$$

Relative to T, V decomposes into weight spaces

$$V = \bigoplus_{\lambda \in \Omega} V(\lambda)\,, \tag{3.8}$$

where $\Omega \subset \mathbf{h}_{\mathbf{R}}^{*}$ is the weight system and for all $\lambda \in \Omega$

$$V(\lambda) = \{v \in V \mid d\rho_{\mathbf{C}}(h)\,v = \lambda(h)v \quad \text{for all } h \in \mathbf{h}\}\,. \tag{3.9}$$

For each $\beta \in \Delta$ we have, by restriction, representations

$$\rho^{\beta} : \mathrm{SU}^{\beta}(2) \to \mathrm{GL}(V),$$

$$\rho_{\mathbf{C}}^{\beta}: G^{\beta} \to \mathrm{GL}(V),$$

$$d\rho_{\mathbf{C}}^{\beta}: \mathbf{sl}_{\beta}(2) \to \mathrm{End}(V).$$

If we partition Ω into β-weight strings, $(\lambda + \mathbf{Z}\beta) \cap \Omega$, i.e.,

$$\lambda + q\beta, \lambda + (q-1)\beta, \ldots, \lambda, \ldots, \lambda - p\beta, \tag{3.10}$$

then the sums of the corresponding weight spaces

$$\bigoplus_{j=-p}^{q} V(\lambda + j\beta) \tag{3.11}$$

are $\mathrm{SU}^{\beta}(2)$-submodules. They can further be decomposed into $\mathrm{SU}^{\beta}(2)$-irreducible submodules. Each such submodule is a sum

$$\bigoplus_{k=0}^{s} \mathbf{C} v(\Lambda - k\beta), \tag{3.12}$$

where $\Lambda = \lambda + r\beta$ for some r and $v(\Lambda - k\beta) \in V(\Lambda - k\beta)$. The identification (3.5) of $\mathbf{sl}^{\beta}(2)$ and sl(2) leads to representations of $\mathbf{sl}(2)$ and $\mathrm{SL}_{\mathbf{C}}(2)$ on V such that

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \to d\rho_{\mathbf{C}}(e_{\beta}),$$

$$\begin{pmatrix} 1 & \bar{1} \\ 0 & 1 \end{pmatrix} \to \exp(d\rho_{\mathbf{C}}(-e_{\beta})) = \rho_{\mathbf{C}}(\exp(-e_{\beta})), \tag{3.13}$$

$$R_{\beta} \to \rho_{\mathbf{C}}(\exp e_{-\beta} \exp(-e_{\beta}) \exp e_{-\beta}), \text{ etc.}$$

An appropriate choice of $v(\Lambda - k\beta)$ allows us to identify them with $|{}^{L}_{M}\rangle$ of Sec. II, namely,

$$v(\Lambda - k\beta) \leftrightarrow |{}^{\Lambda(h_{\beta})}_{(\Lambda - k\beta)(h_{\beta})}\rangle = |{}^{L}_{M}\rangle . \tag{3.14}$$

Then also $e_{\pm\beta}$ and h_{β} act on (3.14) according to (2.5).

Although R_{β} is defined in terms of nonunitary operators, it lies in $\mathrm{SU}(2) \subset K$ and hence appears as a unitary operator on V. From now on we write R_{β} for $\rho_{\mathbf{C}}(R_{\beta})$. According to (2.7), one has

$$R_{\beta} |{}^{\Lambda(h_{\beta})}_{(\Lambda - k\beta)(h_{\beta})}\rangle = (-1)^{k} |{}^{\Lambda(h_{\beta})}_{(\Lambda - k\beta)(-h_{\beta})}\rangle, \tag{3.15}$$

which demonstrates the "charge conjugating role" of the operators R_{β}. Thus the general effect of R_{β} on V is the permutation of weight subspaces:

$$V(\lambda) \leftrightarrow V(\lambda - \lambda(h_{\beta})\beta) . \tag{3.16}$$

The Weyl reflection $r_{\beta}: \mathbf{h}_{\mathbf{R}}^{*} \to \mathbf{h}_{\mathbf{R}}^{*}$ is defined by

$$r_{\beta}\lambda = \lambda - \lambda(h_{\beta})\beta = \lambda - (2(\lambda, \beta)/(\beta, \beta))\beta . \tag{3.17}$$

Therefore,

$$R_{\beta} V(\lambda) = V(r_{\beta}\lambda) . \tag{3.18}$$

The Weyl group is by definition the group W generated by the r_{α}, $\alpha \in \Delta$. For all α, $\beta \in \Delta$, $r_{\alpha}^{2} = 1$, and $r_{\alpha} r_{\beta} r_{\alpha} = r_{r_{\alpha}\beta}$. It follows that W is generated by the $r_i := r_{\alpha_i}$, $i = 1,2,\ldots,l$.

Whereas, $r_{\beta}^{2} = 1$, it is obvious from (2.9) that we only have

$$R_{\beta}^{4} = 1. \tag{3.19}$$

From (2.9) and (3.15) one has

$$R_{\beta}^{2} | V(\lambda) = (-1)^{\lambda(h_{\beta})} . \tag{3.20}$$

Following Tits[13] we write

$$R_{\beta}^{2} = (-1)^{h_{\beta}} . \tag{3.21}$$

Corresponding to the generators $e_{\beta}, e_{-\beta}, h_{\beta}$ of $\mathbf{sl}^{\beta}(2)$, we have

$$e_{-\beta}, \quad e_{\beta}, \quad h_{-\beta} = -h_{\beta} \tag{3.22}$$

as a set of generators of $\mathbf{sl}^{\beta}(2)$. The operator $R_{-\beta}$ is then

$$R_{-\beta} = \exp e_{\beta} \exp(-e_{-\beta}) \exp e_{\beta} \leftrightarrow \begin{pmatrix} 0 & 1 \\ \bar{1} & 0 \end{pmatrix} \leftrightarrow R_{\beta}^{-1} .$$

Thus

$$R_{-\beta} = R_{\beta}^{-1} . \tag{3.23}$$

IV. THE FINITE GROUP GENERATED BY CHARGE CONJUGATION OPERATORS

Our primary interest here is the group N generated by $R_{\beta}, \beta \in \Delta$. As it stands N depends on the choice of e_{β} (hence, R_{β}), $\beta \in \Delta$. The most convenient form of N arises by the use of a Chevalley basis[5,6] of $\mathbf{g}$. According to Ref. 14, there is a choice of the $e_{\beta}, \beta \in \Delta$, such that the following occurs: For $\alpha, \beta \in \Delta$, where α and β are linearly independent with root string $\beta - p\alpha, \ldots, \beta, \ldots, \beta + q\alpha$,

$$[e_{\alpha}, e_{\beta}] = \pm(p+1) e_{\alpha+\beta} \text{ if } \alpha + \beta = \Delta . \tag{4.1}$$

The matter of sign is not essential to us here. With such a choice of basis,

$$\mathbf{g}_{\mathbf{Z}} := \sum \mathbf{Z} h_i + \sum_{\beta \in \Delta} \mathbf{Z} e_{\beta} \tag{4.2}$$

is a Lie ring. Most importantly, for all $\beta \in \Delta$ and for all $n \in \mathbf{N}$,

$$(1/n!)(ad e_{\beta})^{n}: \mathbf{g}_{\mathbf{Z}} \to \mathbf{g}_{\mathbf{Z}} . \tag{4.3}$$

More generally, Kostant has shown[6,15] that for every representation (ρ, V) of $\mathbf{g}$ there is a basis $v_1, \ldots, v_m$ of V consisting of weight vectors such that if we set

$$V_{\mathbf{Z}} = \bigoplus_{j} \mathbf{Z} v_j , \tag{4.4}$$

then for all $\beta \in \Delta$

$$\frac{1}{n!} (d\rho(e_{\beta}))^{n}: V_{\mathbf{Z}} \to V_{\mathbf{Z}} . \tag{4.5}$$

Here $\{e_{\beta}\}$ are assumed to be a Chevalley basis. Thus in particular,

$$R_{\beta}: V_{\mathbf{Z}} \to V_{\mathbf{Z}} . \tag{4.6}$$

Since $R_{-\beta} = R_{\beta}^{-1}$ we see that R_{β} is a bijective mapping of $V_{\mathbf{Z}}$.

From the Chevalley basis the operators R_{α} and R_{α}^{-1} map $\mathbf{g}_{\mathbf{Z}}$ into itself (in the adjoint representation) and are automorphisms of $\mathbf{g}_{\mathbf{Z}}$ as a Lie ring. It is easy to check [cf. (2.16)] that $R_{\alpha} h_{\alpha} = -h_{\alpha}$ and R_{α} acts trivially on the orthogonal complement of h_{α} in $\mathbf{h}_{\mathbf{R}}$. Thus

$$R_{\alpha} h_{\beta} = h_{r_{\alpha}\beta} . \tag{4.7}$$

Also since for each $\gamma \in \Delta$, the generators $\pm e_{\gamma}$ are the only ones for $\mathbf{g}_{\mathbf{Z}}^{\alpha} = \mathbf{Z} e_{\gamma}$ (as $\mathbf{Z}$-modules),

$$R_{\alpha} e_{\beta} = \pm e_{r_{\alpha}\beta}, \quad \text{for all } \alpha, \beta \in \Delta. \tag{4.8}$$

From $R_{\alpha}(h_{\beta}) = R_{\alpha}[e_{\beta}, e_{-\beta}] = [R_{\alpha} e_{\beta}, R_{\alpha} e_{-\beta}]$ it follows that

$$R_{\alpha} e_{-\beta} = \pm e_{-r_{\alpha}\beta} , \tag{4.9}$$

where the sign is the same as in (4.8). From this we have

$$R_\alpha R_\beta R_\alpha^{-1} = R_\alpha \exp(-e_\beta)\exp e_{-\beta} \exp(-e_\beta) R_\alpha^{-1}$$
$$= \exp(\mp e_{r_\alpha\beta}) \exp(\pm e_{-r_\alpha\beta}) \exp(\mp e_{r_\alpha\beta})$$
$$= R_{r_\alpha\beta}^{\pm 1}. \tag{4.10}$$

It follows that the group N is generated by

$$R_i := R_{\alpha_i}, \quad i = 1,2,...,l. \tag{4.11}$$

A direct consequence of the well-known defining identities $(r_i r_j)^k = 1$ of the Weyl group are the identities

$$\underbrace{R_i R_j R_i R_j \cdots}_{k \text{ factors}} = \underbrace{R_j R_i R_j R_i \cdots}_{k \text{ factors}}. \tag{4.12}$$

Next consider the group

$$A := \langle R_1^2,...,R_l^2 \rangle \tag{4.13}$$

generated by R_i^2. According to (3.21), R_i^2 commute and the general element

$$R_1^{2\epsilon_1} \cdots R_l^{2\epsilon_l}, \quad \epsilon^i = 0 \text{ or } 1,\ i = 1,2,...,l,$$

of A acts as $(-1)^{\epsilon_1 h_1 + \cdots + \epsilon_l h_l}$.

The abelian group A can thus be identified as

$$\mathbf{h}_{\mathbf{Z}_2} := (\mathbf{Z}h_1 + \cdots + \mathbf{Z}h_l)/(2\mathbf{Z}h_1 + \cdots + 2\mathbf{Z}h_l)$$
$$\cong (\mathbf{Z}/2\mathbf{Z}) \times \cdots \times (\mathbf{Z}/2\mathbf{Z}) \quad (l \text{ factors}) \tag{4.14}$$

and with $\alpha \in A$ corresponding to $\tilde{\alpha} \in \mathbf{h}_{\mathbf{Z}_2}$ and acting as

$$(-1)^{\tilde{\alpha}} : (-1)^{\tilde{\alpha}}|_{V(\lambda)} = (-1)^{\lambda(\alpha)}. \tag{4.15}$$

It is convenient to write $\lambda(\tilde{\alpha})$ for $\lambda(\alpha)$ mod 2 so that $(-1)^{\lambda(\tilde{\alpha})} = (-1)^{\lambda(\alpha)}$. In view of (3.18) there is a natural mapping

$$\pi : N \to W \tag{4.16}$$

with $R_\alpha \mapsto r_\alpha$ and

$$RV(\lambda) = V(w\lambda) \tag{4.17}$$

for all weight spaces $V(\lambda)$ when $\pi(R) = w$. Clearly A is in the kernel of π. In fact A is the kernel of π and we have the exact sequence

$$1 \to A \to N \xrightarrow{\pi} W \to 1. \tag{4.18}$$

We can see as follows that (4.18) holds. Suppose that $R \in N$ and $\pi(R) = 1$. Then R stabilizes each weight space in every irreducible representation of K. In particular, $Ad(R)e_\alpha = \pm e_\alpha$ for each $\alpha \in \Delta$ so that $(AdR)^2 = 1$ and R^2 is in the center of K. Thus R^2 acts as a scalar on each irreducible representation (ρ, V). However, R stabilizes the top weight space (one dimensional) and using Kostant basis[6] we again see $R^2 = 1$. Now $R = \exp 2\pi i h$, for $h \in \mathbf{h}_{\mathbf{R}}$ and $R^2 = 1$, implies that $h = \frac{1}{2}(\epsilon_1 h_1 + \cdots + \epsilon_l h_l)$, $\epsilon_i = 0,1$. Thus $R = (-1)^{\overline{\alpha}} \in A$, where $\overline{\alpha}$ is $\epsilon_1 h_1 + \cdots + \epsilon_l h_l$ taken in $\mathbf{h}_{\mathbf{Z}_2}$.

From (4.18) we have for the order $|N|$ of N:

$$|N| = 2^l |W|. \tag{4.19}$$

Since A is a abelian and $A \lhd N$, the action $\alpha \mapsto nan^{-1}$ of N on A determines an action of W on $\mathbf{h}_{\mathbf{R}}$. This is easy to specify. Recall that N acts on $\mathbf{h}_{\mathbf{R}}$ by (4.7). Since $R_\beta^2 \mathbf{h}_{\mathbf{R}} = 1$ for all $\beta \in \Delta$, A acts trivially on $\mathbf{h}_{\mathbf{R}}$ and we have a representation of W on $\mathbf{h}_{\mathbf{R}}$. Precisely,

$$r_\alpha : h \to h - \alpha(h)h_\alpha. \tag{4.20}$$

This is no more than the action of W on $\mathbf{h}_{\mathbf{R}}$ induced by transposing the action of W on $\mathbf{h}_{\mathbf{R}^*}$: For all $w \in \mathbf{h}_{\mathbf{R}^*}$, $h \in \mathbf{h}_{\mathbf{R}}$,

$$w\varphi(h) = \varphi(w^{-1}h). \tag{4.21}$$

Since W stabilizes $\mathbf{h}_{\mathbf{Z}} = \Sigma\, \mathbf{Z}h_i$, it thus produces a modulo 2 action on $\mathbf{h}_{\mathbf{Z}_2}$. Let $\overline{\alpha} \in \mathbf{h}_{\mathbf{Z}_2}$ and $(-1)^{\overline{\alpha}}$ be the corresponding element of A. Then

$$w\cdot(-1)^{\overline{\alpha}} = (-1)^{w\overline{\alpha}}. \tag{4.22}$$

If (ρ, V) is any unitary representation of K and $V = \oplus_{\lambda \in \Omega} V(\lambda)$ is the weight space decomposition of V, then we have seen

$$RV(\lambda) = V(\pi(R)\lambda)$$

and $(-1)^{\overline{\alpha}}|_{V(\lambda)} = (-1)^{\lambda(\overline{\alpha})}$ [cf. (4.18) and (4.15)]. Important subgroups of N are those which stabilize a given weight space

$$N_\lambda = \{R \in N \mid RV(\lambda) = V(\lambda)\}. \tag{4.23}$$

As suggested by the notation, N_λ does in fact only depend on λ (not on the representation). An explicit description of N_λ is given by the following considerations. Each W-orbit, $W\lambda$ ($\lambda \in \Omega$), contains unique dominant element λ^+: defined by

$$\lambda^+(h_i) \geqslant 0 \quad \text{for all } i = 1,...,l. \tag{4.24}$$

For λ^+ dominant, let

$$J := \{i \in \{1,...,l\} \mid \lambda^+(h_i) = 0\}. \tag{4.25}$$

Then N_{λ^+} is the group generated by A and by the R_i, $i \in J$. Alternatively, if we define

$$W_J := \langle r_i \mid i \in J \rangle, \tag{4.26}$$

then N_{λ^+} is the full preimage $N_J := \pi^{-1}(W_J)$ of W_J in N,

$$1 \to A \to N_J \to W_J \to 1. \tag{4.27}$$

The cardinality of the set $W\lambda^+ = \{w\lambda^+ \mid w \in W\}$ is precisely the index

$$[W : W_J] = |W|/|W_J| \tag{4.28}$$

of W_J in W. This is trivial to compute since W_J is the Weyl group of the subroot system of Δ based on $\{\alpha_j \mid j \in J\}$. For λ not dominant, choose $w \in W$ such that $\lambda^+ = w^{-1}\lambda$ is dominant, and define N_{λ^+} as above. Then

$$N_\lambda = wN_{\lambda^+}w^{-1}. \tag{4.29}$$

Note that (4.29) makes sense since the choice of representation R of w in N for computing (4.29) is immaterial.

V. EXAMPLES

In order to illustrate the content of Sec. III and IV, we consider here some particular cases.

A. The group SU(3)

Consider the lowest faithful representation with the highest weight $\Lambda = (1,0)$ of the group SU(3). Its representation space V^Λ is spanned by the weight vectors

$$|{}^{10}_{10}\rangle, \quad |{}^{10}_{11}\rangle, \quad \text{and } |{}^{10}_{01}\rangle. \tag{5.1}$$

For simplicity we omit the highest weight in symbols like (5.1) whenever there can be no ambiguity. According to (3.15).

$$R_1|10\rangle = |\bar{1}1\rangle, \quad R_1|\bar{1}1\rangle = -|10\rangle, \quad R_1|0\bar{1}\rangle = |0\bar{1}\rangle,$$
$$R_2|10\rangle = |10\rangle, \quad R_2|\bar{1}1\rangle = |0\bar{1}\rangle, \quad R_2|0\bar{1}\rangle = -|\bar{1}1\rangle, \tag{5.2}$$

and therefore we have in the SU(3)-representation (1,0)

$$R_1 = \begin{pmatrix} 0 & \bar{1} & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad R_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & \bar{1} \\ 0 & 1 & 0 \end{pmatrix},$$
$$R_1R_2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \tag{5.3}$$
$$R_1^2 = \begin{pmatrix} \bar{1} & 0 & 0 \\ 0 & \bar{1} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad R_2^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \bar{1} & 0 \\ 0 & 0 & \bar{1} \end{pmatrix}, \text{ etc.}$$

By a direct computation one is led to the conclusion that the subgroup $N \subset \mathrm{SU}(3)$ generated by R_1 and R_2 is isomorphic to the octahedral group O of order 24 and that the above representation is the three-dimensional irreducible representation with determinant of all elements equal one. Adopting notations of Ref. 16, it is the representation Γ_4.

Since $N \subset \mathrm{SU}(3)$, every SU(3) representation (p,q) reduces with respect to N. That is

$$(p,q) \supset \bigoplus_{i=1}^{5} m_i \Gamma_i, \tag{5.4}$$

where m_i is the multiplicity of Γ_i in the reduction. The multiplicities are easily found for any (p,q) from the generating function (4.6)–(4.10) of Ref. 17.

The elements R_1 and R_2 of N lie in the same SU(3)-conjugacy class of regular elements of order 4 in SU(3), namely the one denoted by [2 1 1] in Table I of Ref. 10. Their character values on irreducible SU(3)-representations are restricted to $0, \pm 1$.

B. The group SU(n)

As in the previous case one finds a faithful matrix representation of $N \subset \mathrm{SU}(n)$ by considering the action of R_i, $i = 1,2,\ldots,n-1$, in the lowest faithful representation (1,0,...,0) of SU(n) according to (3.15):

$$R_i = I_{i-1} \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus I_{n-i-1}, \quad i = 1,2,\ldots,n-1, \tag{5.5}$$

where I_k is the $k \times k$ identity matrix. From (4.19) we find that the order of N is $|N| = 2^{n-1}n!$

It is obvious in (5.5) that all R_i belong to the same SU(n)-conjugacy class of rational elements of order 4, which is identified as [210...01] in Table 6 of Ref. 12. Except for SU(2) and SU(3), the R_i are not regular in SU(n) and consequently the set of their character values over all irreducible SU(n) representations is an unbounded set of integers. The elements R_i satisfy the following identities

$$R_i^4 = 1,$$
$$R_iR_j = R_jR_i, \quad \text{if } |i-j| > 1. \tag{5.6}$$
$$R_iR_jR_i = R_jR_iR_j \quad \text{if } |i-j| = 1.$$

The exact sequence (4.18) can be written as

$$1 \to \mathbf{Z}_2 \times \underbrace{\cdots}_{n-1 \text{ times}} \times \mathbf{Z}_2 \to N \to S_n \to 1. \tag{5.7}$$

Here the Weyl group SU(n) is isomorphic to the symmetric group S_n of n letters.

C. The groups USp(4) and O(5)

In the symplectic four-dimensional representation (1,0) of these groups, we have

$$R_1 = \begin{pmatrix} 0 & \bar{1} & & \\ 1 & 0 & & \\ & & 0 & \bar{1} \\ & & 1 & 0 \end{pmatrix} \text{ and } R_2 = \begin{pmatrix} 1 & & & \\ & 0 & \bar{1} & \\ & 1 & 0 & \\ & & & 1 \end{pmatrix} \tag{5.8}$$

relative to the basis of weight vectors. Therefore, also

$$R_1^4 = R_2^4 = 1,$$
$$R_1R_2R_1R_2 = R_2R_1R_2R_1. \tag{5.9}$$

Similarly in the five-dimensional orthogonal representation (0,1), one has

$$R_1 = \begin{pmatrix} 1 & & & & \\ & 0 & 0 & 1 & \\ & 0 & \bar{1} & 0 & \\ & 1 & 0 & 0 & \\ & & & & 1 \end{pmatrix} \text{ and } R_2 = \begin{pmatrix} 0 & \bar{1} & & & \\ 1 & 0 & & & \\ & & 1 & & \\ & & & 0 & \bar{1} \\ & & & 1 & 0 \end{pmatrix}, \tag{5.10}$$

which also satisfy the identities (5.9). The exact sequence (4.18) is in this case

$$1 \to \mathbf{Z}_2 \times \mathbf{Z}_2 \to N \to D_4 \to 1, \tag{5.11}$$

where D_4 is the dihedral group. The order of N is 32. The elements R_1 and R_2 are not conjugate to each other because they correspond to simple roots of different length. Their conjugacy classes are identified in Table 6 of Ref. 12 as [201] and [210], respectively, for R_1 and R_2. Both elements are rational, which implies that their characters take only integer values in any representation of the Lie group, but they are not regular which means that their character values are unlimited. For any given representation (a,b) the characters R_1 and R_2 are easily found from the generating function of Table V of Ref. 10.

D. The group G_2

In the lowest representation (0,1) of dim = 7 of G_2, one has R_1 and R_2 as

$$R_1 = 1 \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus 1 \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus 1,$$
$$R_2 = \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 & 1 \\ 0 & \bar{1} & 0 \\ 1 & 0 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix}, \tag{5.12}$$

relative to the weight vector basis. The group N generated by (5.12) is of order 48. One has the exact sequence

$$1 \to \mathbf{Z}_2 \times \mathbf{Z}_2 \to N \to D_6 \to 1, \tag{5.13}$$

where D_6 is the dihedral group, and the identities

$$R_1^4 = R_2^4 = 1,$$
$$R_1R_2R_1R_2R_1R_2 = R_2R_1R_2R_1R_2R_1 . \tag{5.14}$$

The elements R_1 and R_2 are rational and nonregular in G_2. Their G_2-conjugacy classes are, respectively [201] and [110]. Their characters in any representation of G_2 are found from the generating functions of Table VI of Ref. 10.

VI. CONSTRUCTING BASES IN REPRESENTATION SPACES

In this section we demonstrate and illustrate the reduction of the problem of constructing a basis and computing the matrix elements of generators to similar problems of much smaller size involving only the dominant weight vectors.

Given the general decomposition (3.8) of a space V in which a compact simple Lie group K acts irreducibly as a representation ρ, it is natural to consider the present problem as a construction of bases in every weight subspace $V(\lambda) \subset V$, where $\lambda \in \Omega$. Let us point out that in almost every application, and certainly in all of them in elementary particle physics, one chooses a basis of weight vectors relative to some Cartan subalgebra of $\mathbf{g}$ whenever an explicit use of a basis is made. Otherwise one could not associate quantum numbers with the basis vectors–physical states. The dimension of $V(\lambda)$ is the multiplicity of λ in Ω so that the basis construction in $V(\lambda)$ is a nontrivial problem only when $\dim V(\lambda) > 1$.

Using notations of Sec. III, let us recall[14] the following. For every weight $\lambda \in \Omega$ there exists a unique dominant weight $\lambda^+ \in \Omega$ such that $\lambda \in W\lambda^+$. If $\lambda = w\lambda^+$ then also $\lambda = w\,\mathrm{Stab}_W(\lambda^+)(\lambda^+)$. There exists a unique canonical w such that

$$\lambda = w\lambda^+, \quad w = r_i r_{i_2} \cdots r_{i_k} , \tag{6.1}$$

in which the number k of reflections r_i is minimal. We define

$$\tilde{w} = R_{i_1} R_{i_2} \cdots R_{i_k} . \tag{6.2}$$

Although w is unique, its expression as a word in the reflections $r_1,\ldots,r_l$ is not. Thus $\tilde{w}$ depends on the choice of the writing w in (6.1). If $r_{j_1} \cdots r_{j_m}$ is any other expression for w (minimal or not), then $\tilde{w}' = R_{j_1} \cdots R_{j_m}$ is some other preimage of w in N and $\tilde{s} := \tilde{w}^{-1}\tilde{w}'$ stabilizes the weight space $V(\lambda^+)$. Hence it is unavoidable to consider the effect of such elements $\tilde{s} \in N$ on the weight spaces they stabilize.

Let us assume that for each dominant weight $\lambda^+ \in \Omega$ we have an orthogonal basis [problem (i) of Introduction]

$$|\lambda^+\rangle_1,\ldots,|\lambda^+\rangle_m \quad \text{of} \quad V(\lambda^+), \dim V(\lambda^+) = m . \tag{6.3}$$

Then for λ as in (6.1) we define

$$\tilde{w}|\lambda^+\rangle_1, \tilde{w}|\lambda^+\rangle_2,\ldots,\tilde{w}|\lambda^+\rangle_m \tag{6.4}$$

as our basis for $V(\lambda)$, which solves part (a) of problem (ii) of the Introduction. In the following five examples we illustrate our approach to that problem.

Example 1 [SU(2)]: All $V(\lambda)$ are one dimensional. A basis of $V(\lambda^+)$ consists of $|{}^L_M\rangle$, $M \geqslant 0$. As a basis vector in $V(-\lambda^+)$ we take

$$|{}^L_{\bar{M}}\rangle = (-1)^{(L-M)/2} R\,|{}^L_M\rangle , \tag{6.5}$$

where we have used the phase factor in order to keep the convention in complete agreement with Sec. II.

Example 2 [SU(3) representation (1,0)]: There is only one dominant weight $\lambda^+ = (1,0)$ in Ω of multiplicity 1. The basis is given in (5.2).

Example 3 [the adjoint representation (2,0) of Sp(4)]: By assumption we know the basis vectors

$$|20\rangle, \quad |01\rangle, \quad |00\rangle_1, \quad |00\rangle_2.$$

Then the rest is given by

$$\begin{aligned} &R_1|20\rangle = |\bar{2}2\rangle, && R_2|01\rangle = |2\bar{1}\rangle, \\ &R_2R_1|20\rangle = |2\bar{2}\rangle, && R_1R_2|01\rangle = |\bar{2}1\rangle, \\ &R_1R_2R_1|20\rangle = |\bar{2}0\rangle, && R_2R_1R_2|01\rangle = |0\bar{1}\rangle . \end{aligned} \tag{6.6}$$

Example 4 [SU(3) representation (3,2) of dim=42]: Properties of dominant weights are shown on Fig. 1. Assuming that we have pairwise orthogonal dominant weight vectors

$$|21\rangle_i, \quad |02\rangle_i, \quad i = 1,2,$$
$$|10\rangle_j, \quad j = 1,2,3, \tag{6.7}$$

the basis is given by

$$\begin{aligned} &|32\rangle, && R_2|\bar{3}1\rangle = R_2R_1|32\rangle = |\bar{2}\bar{1}\rangle, \\ &R_1|32\rangle = |\bar{3}1\rangle, && R_1|5\bar{2}\rangle = R_1R_2|32\rangle = |\bar{5}3\rangle, \\ &R_2|32\rangle = |5\bar{2}\rangle, && R_2|\bar{5}3\rangle = R_2R_1R_2|32\rangle = |\bar{2}\bar{3}\rangle, \\ &|13\rangle, && R_2|\bar{1}4\rangle = R_2R_1|13\rangle = |3\bar{4}\rangle, \\ &R_2|13\rangle = |2\bar{3}\rangle, && R_1|3\bar{4}\rangle = R_1R_2R_1|13\rangle = |\bar{3}\bar{1}\rangle, \\ &|40\rangle, && \\ &R_1|40\rangle = |\bar{4}4\rangle, && R_2|\bar{4}4\rangle = R_2R_1|40\rangle = |0\bar{4}\rangle, \\ &|02\rangle_i, \; i = 1,2, && R_1|2\bar{2}\rangle_i = R_1R_2|02\rangle_i = |\bar{2}0\rangle_i, \\ &R_2|02\rangle_i = |2\bar{2}\rangle_i, && \end{aligned}$$

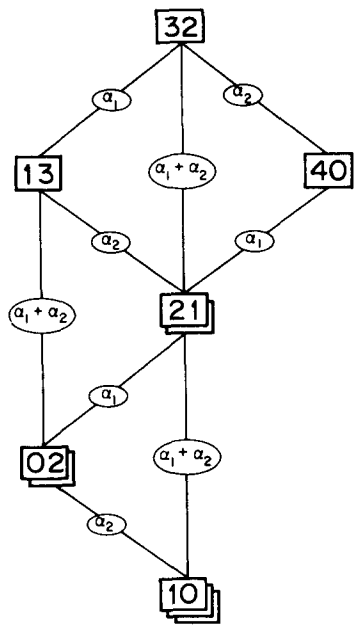


FIG. 1. Dominant weights of the weight system of the SU(3)-representation (3,2), their multiplicities, and the positive roots by which they differ.

$$
\begin{aligned}
&|10\rangle_j, j=1,2,3, \quad && R_2|\bar{1}1\rangle_j = R_2R_1|10\rangle_j = |0\bar{1}\rangle_j,\\
&R_1|10\rangle_j = |\bar{1}1\rangle_j, && \\
&|21\rangle_i,\ i=1,2, \quad && R_2|\bar{2}3\rangle_i = R_2R_1|21\rangle_i = |1\bar{3}\rangle_i,\\
&R_1|21\rangle_i = |\bar{2}3\rangle_i, && R_1|3\bar{1}\rangle_i = R_1R_2|21\rangle_i = |\bar{3}2\rangle_i,\\
&R_2|21\rangle_i = |3\bar{1}\rangle_i. && R_1|1\bar{3}\rangle_i = R_1R_2R_1|21\rangle_i = R_2|\bar{3}2\rangle_i\\
& && \quad = R_2R_1R_2|21\rangle_i = |\bar{1}\bar{2}\rangle_i.
\end{aligned}
\tag{6.8}
$$

The last example, although it still refers only to a group of rank 2, makes it obvious that it is impractical to explicitly write all basis vectors in larger spaces. Instead one should construct the basis for dominant weight subspaces and any other ones only when needed for a particular task at hand.

Example 5 [O(16)-representation (0000000^1_1) of dim=11 440]: The dominant weights and their multiplicities are found in Ref. 18; they are shown on Fig. 2. Assuming that an orthogonal basis in each of the three dominant weight subspaces of dim > 1 has been constructed, the rest of the construction is a mechanical application of CCO similar to (6.8). Thus from $|000000^1_1\rangle$ one gets $2^7\cdot 8!/7! = 1024$ other basis vectors, from each of the three $|000000^1_1\rangle_j$, $j=1,2,3$, one gets $2^7\cdot 8!/5!\cdot 4! = 1792$ others, from each of $|00100^0_0\rangle_k$, $k=1,\ldots,10$, one gets $2^6\cdot 8!/3!\cdot 2^4\cdot 5! = 448$ new ones, and from each $|100000^0_0\rangle_n$, $n=1,\ldots,35$, one gets $2^7\cdot 8!/2^6\cdot 7! = 16$ new basis vectors. Here the numbers are calculated using (4.28) and the orders of Weyl groups given for instance in Refs. 15 or 18.

Suppose now that we need the basis in a particular weight subspace, say $V(\lambda) = V(01\bar{1}0\bar{1}0^1_1)$. Applying R_i with subscripts corresponding to negative entries in the corresponding weight for as long as possible one finds

$$
V(000010^0_0) = R_5R_3R_7R_6R_4R_5V(01\bar{1}0\bar{1}0^1_1). \tag{6.9}
$$

Then applying R_i^{-1} in the inverse order to that in (6.9) to the basis vectors $|000010^0_0\rangle_i$, $i=1,2,3$, one gets the desired basis. Arrived at in this way the element $\tilde{w}$ of N is bound to be of minimal type.

Let us now exemplify the truly difficult part of our problem: construction of bases in dominant weight subspaces. Also here CCO operators provide a valuable tool.

Example 6 [the adjoint representation (1,1) of SU(3)]: The dominant weight subspace $V(00)$ is of dim = 2. Its basis can be constructed as follows: Consider the highest irreducible component in the tensor product (1,0) ⊗ (0,1) which is the adjoint representation. Its dominant weight vectors $|11\rangle$, $|00\rangle_A$, and $|00\rangle_B$ can be chosen as

$$
\begin{aligned}
&|11\rangle = |10\rangle|01\rangle,\\
&|00\rangle_A = (1/\sqrt{2})(|\bar{1}1\rangle|1\bar{1}\rangle + |10\rangle|\bar{1}0\rangle),\\
&|00_B = (1/\sqrt{2})(|\bar{1}1\rangle|1\bar{1}\rangle + |0\bar{1}\rangle|01\rangle).
\end{aligned}
\tag{6.10}
$$

Here the linearly independent but nonorthogonal vectors $|00\rangle_A$ and $|00\rangle_B$ span $V(00)$. Instead of (6.10) one could observe[19] that N acts irreducibly on $V(00)$. Its two-dimensional representation is generated by matrices

$$
m_1 = \begin{pmatrix} \bar{1} & 0 \\ 0 & 1 \end{pmatrix}, \quad m_2 = \frac{1}{2}\begin{pmatrix} \bar{1} & \sqrt{3} \\ \sqrt{3} & \bar{1} \end{pmatrix} \tag{6.11}
$$

(cf. Table IX, Ref. 16) and, by definition of N, also by R_1 and R_2. On $V(00)$ one can identify m_1 with R_1. Then the eigenvectors R_1

$$
R_1|00\rangle_{\pm} = \pm|00\rangle_{\pm} \tag{6.12}
$$

provide an orthogonal basis of $V(00)$. Using (6.10) one has explicitly

$$
\begin{aligned}
&|00\rangle_- = (1/\sqrt{2})(|\bar{1}1\rangle|1\bar{1}\rangle + |10\rangle|\bar{1}0\rangle),\\
&|00\rangle_+ = (1/\sqrt{6})(|\bar{1}1\rangle|1\bar{1}\rangle - |10\rangle|\bar{1}0\rangle + 2|0\bar{1}\rangle|01\rangle).
\end{aligned}
\tag{6.13}
$$

Example 7: As our next example let us construct the dominant weight basis vectors (6.7). For that consider the representation (3,2) as the highest irreducible component in (3,0) ⊗ (0,2). Since (3,0) and (0,2) have only one-dimensional weight spaces, their weight vectors $|^{30}_{\lambda}\rangle$ and $|^{02}_{\mu}\rangle$ provide a basis for our problem. For λ^+ of multiplicity one in Ω of (3,2), one has

$$
|^{32}_{32}\rangle = |^{30}_{30}\rangle|^{02}_{02}\rangle,\ |^{32}_{13}\rangle = |^{30}_{11}\rangle|^{02}_{02}\rangle,\ |^{32}_{40}\rangle = |^{30}_{30}\rangle|^{02}_{10}\rangle. \tag{6.14}
$$

The two-dimensional subspace $V(02)$ is stabilized by R_1,

$$
R_1V(02) = V(02). \tag{6.15}
$$

Hence, its basis can be taken as eigenvectors of R_1. In order to identify the eigenvalues, it suffices to notice that there are two α_1-strings passing through $V(02)$ of V corresponding to $SU^{\alpha_1}(2)$ representations of dimensions 5 and 3. Since according to (2.17), $R|^4_0\rangle = |^4_0\rangle$ but $R|^2_0\rangle = -|^2_0\rangle$, the R_1-eigenvalues are ± 1. Consequently, we can choose

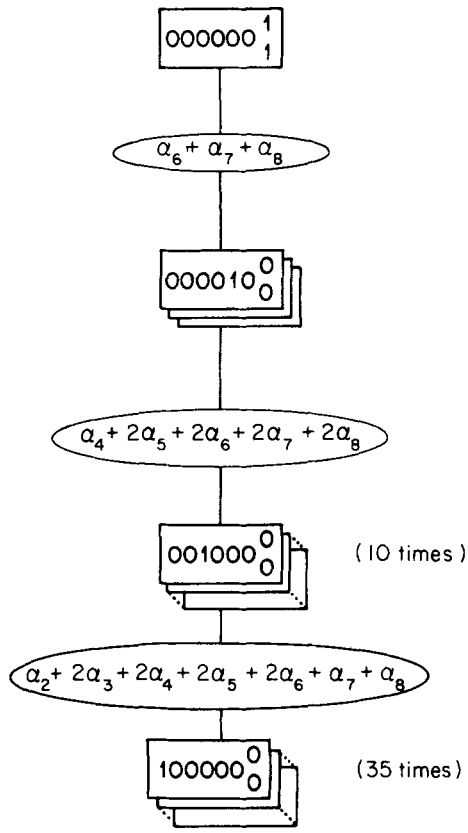


FIG. 2. Dominant weights of the representation (000000^1_1) of O(16) of dimension 11440, their multiplicities, and the positive roots by which they differ.

$$R_1|^{32}_{02}\rangle_1 = |^{32}_{02}\rangle_1 \quad \text{and } R_1|^{32}_{02}\rangle_2 = -|^{32}_{02}\rangle_2 \tag{6.16}$$

as the basis $V(02)$, or explicitly

$$|^{32}_{02}\rangle_1 = (1/\sqrt{2})(|^{32}_{12}\rangle|^{02}_{10}\rangle + |^{30}_{11}\rangle|^{02}_{11}\rangle),$$
$$|^{32}_{02}\rangle_2 = (1/\sqrt{6})(|^{30}_{12}\rangle|^{02}_{10}\rangle - |^{30}_{11}\rangle|^{02}_{11}\rangle + 2|^{30}_{00}\rangle|^{02}_{02}\rangle). \tag{6.17}$$

Then it is natural to also choose

$$|^{32}_{21}\rangle_1 = (1\sqrt{6})\, e_1|^{32}_{02}\rangle_1 = \tfrac{1}{2}(\sqrt{3}\,|^{30}_{11}\rangle|^{02}_{10}\rangle + |^{30}_{30}\rangle|^{02}_{11}\rangle), \tag{6.18}$$
$$|^{32}_{21}\rangle_2 = (1/\sqrt{2})\, e_1|^{32}_{02}\rangle_2 = (1\sqrt{12})(|^{30}_{11}\rangle|^{02}_{10}\rangle - \sqrt{3}|^{30}_{30}\rangle|^{02}_{11}\rangle + 2\sqrt{2}|^{30}_{21}\rangle|^{02}_{02}\rangle).$$

The coefficients above (other than the overall normalization) are a result of the application of $SU^{\beta}(2)$ generators along the corresponding β-string [cf. (3.14)] according to (2.5). The three-dimensional subspace $V(10)$ is stabilized by the subgroup $\langle R_2, R_1^2\rangle$ of N generated by R_2 and R_1^2. There are three α_2-strings passing through it of lengths 5, 2, and 1. Due to (2.17), one can thus require that

$$R_2|^{32}_{10}\rangle_1 = |^{32}_{10}\rangle_1 \text{ and } e_2^2|^{32}_{10}\rangle_1 \sim |^{32}_{14}\rangle = |^{30}_{12}\rangle|^{02}_{02}\rangle = R_1|^{32}_{13}\rangle,$$
$$R_2|^{32}_{10}\rangle_2 = -|^{32}_{10}\rangle_2 \text{ and } e_2|^{32}_{10}\rangle_2 \sim (1-R_2)\, e_{-2}|^{32}_{02}\rangle_2, \tag{6.19}$$
$$R_2|^{32}_{10}\rangle_3 = |^{32}_{10}\rangle_3 \text{ and } e_{\pm 2}|^{32}_{10}\rangle_3 = 0,$$

where $\sim$ indicates that both sides differ only by a constant nonzero factor.

Explicitly (6.19) is

$$|^{32}_{10}\rangle_1 = (1/\sqrt{6})(|^{30}_{1\bar{2}}\rangle|^{02}_{02}\rangle + |^{30}_{\bar{1}2}\rangle|^{02}_{2\bar{2}}\rangle + 2|^{30}_{00}\rangle|^{02}_{10}\rangle),$$
$$|^{32}_{10}\rangle_2 = (1/\sqrt{6})(|^{30}_{1\bar{2}}\rangle|^{02}_{02}\rangle - |^{30}_{\bar{1}2}\rangle|^{02}_{2\bar{2}}\rangle - \sqrt{2}|^{30}_{21}\rangle|^{02}_{1\bar{1}}\rangle - \sqrt{2}|^{30}_{11}\rangle|^{02}_{0\bar{1}}\rangle),$$
$$|^{32}_{10}\rangle_3 = (1/\sqrt{30})(3|^{30}_{11}\rangle|^{02}_{0\bar{1}}\rangle - 3|^{30}_{2\bar{1}}\rangle|^{02}_{1\bar{1}}\rangle + \sqrt{2}|^{30}_{1\bar{2}}\rangle|^{02}_{2\bar{2}}\rangle + \sqrt{2}|^{30}_{1\bar{2}}\rangle|^{02}_{02}\rangle - \sqrt{2}|^{30}_{00}\rangle|^{02}_{10}\rangle + \sqrt{6}|^{30}_{\bar{2}0}\rangle|^{02}_{20}\rangle). \tag{6.20}$$

Let us point out that using the product form (6.20) for each $|^{32}_{\lambda}\rangle$, the conditions (6.19) do not guarantee that $|^{32}_{10}\rangle_3$ lies in V because the product space is reducible and contains other SU(3) irreducible representations than (3,2). (In fact there are six linearly independent vectors $|^{\Lambda}_{10}\rangle$ in the product space corresponding to several highest weights Λ.) In order to assure that $|^{32}_{10}\rangle_3 \subset V$, one can proceed for instance as follows. A third vector of $V(10)$ which is linearly independent from $|^{32}_{10}\rangle_1$ and $|^{32}_{10}\rangle_2$, is $(1+R_2)\, e_{-2}^2|^{30}_{\bar{1}2}\rangle|^{02}_{02}\rangle$. Then $|^{32}_{10}\rangle_3$ is the linear combination of the three which is normalized and orthogonal to $|^{32}_{10}\rangle_1$ and $|^{32}_{10}\rangle_2$.

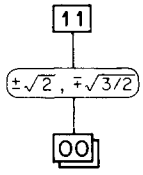


FIG. 3. Essentials of the basis for the adjoint representation of SU(3). The 1×2 matrix in the light rounded box gives the matrix elements of the generators $e_{\pm 1\pm 2}$ between the orthonormal dominant weight vectors; upper (lower) signs refer to matrix elements of $e_{-1-2}(e_{1+2})$.

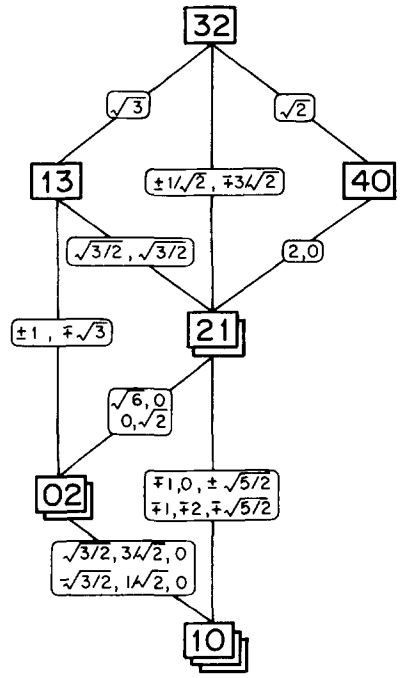


FIG. 4. Essentials of the basis for the SU(3) representation (3,2): dominant weights, their multiplicities and matrix elements of generators between orthonormal basis vectors of dominant weight subspaces. Upper (lower) signs correspond to lowering (raising) generators.

Finally, let us illustrate how all the relevant information concerning a basis and corresponding matrix elements can be presented and used [problem (ii)(b) of the Introduction].

Example 8: Let us continue Example 6. In Fig. 3 one finds the dominant weights of the adjoint representation of SU(3) and two matrices representing the action of the generator $e_{1+2} = e_1e_2 - e_2e_1$ on the chosen basis (6.13) of $V(00)$ and $e_{-1-2} = e_{-1}e_{-2} - e_{-2}e_{-1}$ on $V(11)$. Namely,

$$e_{-1-2}|11\rangle = \sqrt{1/2}\,|00\rangle_- - \sqrt{3/2}\,|00\rangle_+, \tag{6.21}$$
$$e_{1+2}|00\rangle_- = \sqrt{1/2}|11\rangle, \quad e_{1+2}|00\rangle_+ = \sqrt{3/2}|11\rangle. \tag{6.22}$$

Using Fig. 3 many other matrix elements can readily be found, for instance,

$$e_{-2}|\bar{1}2\rangle = e_{-2}R_1|11\rangle = R_1R_2^{-1}e_{-2}R_1|11\rangle = -R_1e_{-1-2}|11\rangle = -R_1(\sqrt{1/2}|00\rangle_- - \sqrt{3/2}|00\rangle_+). \tag{6.23}$$

Similarly, using (6.12) and (6.23), one has

$$e_{-2}|\bar{1}2\rangle = \sqrt{1/2}|00\rangle_- + \sqrt{3/2}|00\rangle_+. \tag{6.24}$$

Example 9: Consider again Example 7. On Fig. 4 we have summarized the relevant information, i.e., the basis vectors (6.14), (6.17), and (6.18) together with the matrix elements of generators relating them. Thus for instance, the nonzero matrix elements of $e_{-1-2} = e_{-1}e_{-2} - e_{-2}e_{-1}$ are read off Fig. 4 as

$$e_{-1-2}|32\rangle = (1/\sqrt{2})|21\rangle_1 - (3/\sqrt{2})|21\rangle_2,$$

and

$$e_{-1-2}|21\rangle_1 = -|10\rangle_1 + \sqrt{5/2}|10\rangle_3,$$

$$e_{-1-2}|21\rangle_2 = -|10\rangle_1 - 2|10\rangle_2 - \sqrt{5/2}|10\rangle_3,$$

and

$$e_{-1-2}|13\rangle = |02\rangle_1 = \sqrt{2}|02\rangle_2,$$

while for $e_{1+2} = e_1e_2 - e_2e_1$, one has

$$e_{1+2}|10\rangle_1 = |21\rangle_1 + |21\rangle_2,$$

$$e_{1+2}|10\rangle_2 = 2|21\rangle_2,$$

$$e_{1+2}|10\rangle_3 = -\sqrt{5/2}|21\rangle_1 + \sqrt{5/2}|21\rangle_2,$$

$$e_{1+2}|21\rangle_1 = -(1/\sqrt{2})|32\rangle,$$

$$e_{1+2}|21\rangle_1 = (3/\sqrt{2})|21\rangle_2,$$

$$e_{1+2}|02\rangle_1 = -|13\rangle, \quad e_{1+2}|02\rangle_2 = \sqrt{3}|13\rangle.$$

Proceeding as in the previous example one finds any other matrix elements of e_β, $\beta \in \Delta$, in terms of those of Fig. 4.

ACKNOWLEDGMENTS

Both authors are grateful for the hospitality of the Aspen Center for Physics where an early version of the article was put together.

Work was supported in part by the Natural Science and Engineering Research Council of Canada, by the Ministère de l'Education du Québec, and by the Fleischmann, Sloan, and Weingart Foundations.

[1] J. Tits, J. Algebra **4**, 96 (1966).
[2] I. G. Koh, J. Patera, and C. Rousseau, J. Math. Phys. **24**, 1955 (1983).
[3] I. G. Koh, J. Patera, and D. Rousseau, "Clebsch–Gordan coefficients for E(6) and SU(10) models of grand unification," J. Math. Phys. **25**, 2863(1984).
[4] R. Slansky, *Proceedings of the First Workshop on Grand Unification*, edited by P. Frampton, L. S. Glashow, and A. Yildiz (Math Science Press, New Hampshire, 1980); R. Slansky, Phys. Rep. **79**, 1 (1981); H. E. Haber, G. Segrè, and S. K. Soni, Phys. Rev. D **25**, 1400 (1982).
[5] C. Chevalley, Tôhoku Math. J. (2) **7**, 14 (1955).
[6] B. Kostant, *Proceedings of the Symposium on Pure Mathematics IX* (American Mathematical Society, Providence, RI, 1966).
[7] I. M. Gelfand and M. L. Tsetlin, Dok. Akad. Nauk SSSR **71**, 825 (1950).
[8] B. R. Judd, W. Miller, Jr., J. Patera, and P. Winternitz, J. Math. Phys. **15**, 1787 (1974).
[9] T. H. Seligman and R. T. Sharp, J Math. Phys. **24**, 769 (1983).
[10] R. V. Moody, R. T. Sharp, and J. Patera, J. Math. Phys. **24**, 2387 (1983).
[11] V. Kac, Functional Anal. Appl. **3**, 252 (1969).
[12] R. V. Moody and J. Patera, "Characters of Elements of Finite Order in Lie Groups," SIAM J. Algebraic Discrete Methods **5**, September (1984).
[13] J. Tits, Théorie des groupes, "Résumé des cours et travaux," Annuaire du Collège de France, (1980–1981).
[14] J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1972).
[15] N. Bourbaki, *Groupes et algèbres de Lie* (Hermann, Paris, 1968), Chaps. 4, 5e, and 6.
[16] J. Patera, R. T. Sharp, and P. Winternitz, J. Math. Phys. **19**, 2362 (1978).
[17] J Patera and R. T. Sharp, J. Phys. A: Math. Gen. **13**, 397 (1980).
[18] R. V. Moody and J.Patera, Bull. Am. Math. Soc. **7**, 337 (1982); M. Brenner, R. V. Moody, and J. Patera, *Tables of Dominant Weight Multiplicities for Simple Lie Algebras of Ranks<12* (Dekker, New York, 1984) (to be published).
[19] Corollary of Chap. VI. Sec. 1, 2 of Ref. 15.

Enveloping algebra annihilators and projection techniques for finite-dimensional cyclic modules of a semisimple Lie algebra

M. D. Gould[a] and S. A. Edwards[b]
Department of Mathematical Physics, University of Adelaide, Adelaide, South Australia 5001

(Received 1 September 1983; accepted for publication 24 February 1984)

Some results on the structure of finite-dimensional cyclic modules for a semisimple Lie algebra are presented. Cyclic modules arise naturally in constructing symmetry adapted states of a system using projection. Projecting out states with definite symmetry from an arbitrary state ψ is related to the properties of the cyclic module generated by ψ. An important example of a cyclic module is the tensor product of two irreducible modules $V(\lambda) \otimes V(\mu)$ which is cyclically generated by the vector $v^{\lambda}_{-} \otimes v^{\mu}_{+}$, where v^{λ}_{-} (resp., v^{μ}_{+}) is the minimal (resp., maximal) weight vector of $V(\lambda)$ [resp., $V(\mu)$]. For this particular case we determine the explicit form of the annihilator, in the universal enveloping algebra, of the cyclic vector $v^{\lambda}_{-} \otimes v^{\mu}_{+}$. It is hoped that this result may add new insight into the Clebsch–Gordan multiplicity problem. As an application of this result projection operators are constructed which project, from an arbitrary vector of weight λ, a maximal weight vector of weight λ.

PACS numbers: 02.20.Qs

I. INTRODUCTION

The theory of Lie groups has now been established as an invaluable tool in physical applications where they usually appear as the symmetry group of the system. The states of a physical system are then to comprise irreducible representations of the symmetry group. Lie groups afford not only convenient analytic methods but in practice are essential for determining selection rules and for the numerical solution of the equations of motion of the system by allowing the Hamiltonian to be broken into a convenient block form. Lie groups also provide suitable labels (i.e., quantum numbers) for physical states even though such Lie groups need not be symmetry groups.

An important problem in the analysis of quantum mechanical systems is the construction of wave functions with definite symmetry properties under a (semisimple) Lie group (which may be the symmetry group of the system or any group used to provide suitable quantum numbers for the physical states). Such wave functions can be obtained most naturally by means of projection operators. The method of projection has proved in the past to be a powerful tool for handling the various state labeling problems of physical interest. Projection operator techniques were successfully employed by Elliot[1,2] to the $U(3) \supset O(3)$ state labeling problem. Asherova and collaborators[3] have extensively developed the infinitesimal projection technique introduced by Löwdin[4] and Shapiro.[5] The methods of projection have also been applied to the Clebsch–Gordan multiplicity problem for a semisimple Lie group G (See Ref. 6) (i.e., the $G \times G \supset G$ state labeling problem). More recently these methods were applied[7] to give a solution to the symplectic group state labeling problem. A more detailed account of the various methods of projection can be found in Moshinsky *et al.*,[8] Asherova and Smirnov,[3] MacDonald,[9] and Edwards and Gould.[6]

In this paper we consider the problem of constructing projection operators for a semisimple Lie algebra which project out, from a weight vector of given weight $\lambda = (\lambda_1, \lambda_2, ..., \lambda_l)$, a maximal weight vector of weight λ. A solution to this problem, for the simple Lie algebras, has been proposed by Asherova *et al.*,[3] who represent their projection operators as a certain series in powers of the infinitesimal generators. We present here an alternative approach, applicable to any semisimple Lie algebra, which employs central projection using only the universal Casimir invariant of the associated Lie algebra. The emphasis in our work is on the structure of finite-dimensional cyclic modules for a semisimple Lie algebra. Cyclic modules arise naturally in constructing symmetry adapted states of a system using projection. Projecting out states with definite symmetry from an arbitrary state ψ, which is known to belong to a finite-dimensional Hilbert space, essentially reduces to analyzing the properties of the cyclic module generated by ψ.

To this end we consider in this paper some elementary properties of finite-dimensional cyclic modules for a semisimple Lie algebra. The structure of a finite-dimensional cyclic module is uniquely determined by the annihilator of the cyclic vector in the universal enveloping algebra of the Lie algebra. An important example of a cyclic module is the tensor product of two irreducible modules $V(\lambda) \otimes V(\mu)$ which is cyclically generated by the vector $v^{\lambda}_{-} \otimes v^{\mu}_{+}$, where v^{λ}_{-} (resp., v^{μ}_{+}) is the minimal (resp., maximal) weight vector of $V(\lambda)$ [resp., $V(\mu)$]. By adapting a result of Parthasarathy, Ranga Rao, and Varadarajan[10] we determine the explicit form of the annihilator of the cyclic vector $v^{\lambda}_{-} \otimes v^{\mu}_{+}$. This opens up the interesting possibility of treating the tensor product of two irreducible modules infinitesimally using an induced module construction analogous to that applied by Verma[11] and Gel'fand *et al.*[12] to the theory of Verma modules (see also Humphreys[13] and Dixmier[14]). It is hoped that

[a] Present address: School of Chemistry, The University of Western Australia, Nedlands, Western Australia 6009, Australia.
[b] School of Earth Sciences, Macquarie University, North Ryde, 2113, New South Wales, Australia.

 0022-2488/84/102848-08$02.50

this may add new insight into the Clebsch–Gordan multiplicity problem (cf. Ref. 6).

It shall be shown that all finite-dimensional cyclic modules, generated by a weight vector, may be imbedded, in a natural way, in the tensor product of two irreducible modules. This result yields, in particular, a method for constructing projection operators which project from a weight vector v^{λ} of weight λ a maximal weight vector of weight λ. Projection operators of this form have found considerable application in atomic and nuclear physics[3,9] and in particular are useful for the construction of raising and lowering operators for a semisimple Lie algebra (cf. Refs. 15 and 16). It is interesting to note that, since all finite-dimensional cyclic modules may be naturally imbedded in the tensor product of two irreducible modules, the approach of Ref. 6 to the Clebsch–Gordan multiplicity problem may be extended, in principle, to any finite-dimensional cyclic module. We remark, in this connection, that cyclic modules in fact appear naturally in several state labeling problems of physical interest (see Ref. 6).

The paper is organized as follows. In Sec. III we consider some elementary properties of finite-dimensional cyclic modules and some preliminary results on projection from a weight vector. In Sec. IV we determine the explicit form of the universal enveloping algebra annihilator of the cyclic vector $v^{\lambda}_{-} \otimes v^{\mu}_{+}$ of the tensor product of two irreducible modules $V(\lambda) \otimes V(\mu)$. In Sec. V we demonstrate that a finite-dimensional cyclic module may be naturally imbedded in the tensor product of two irreducible modules. This result is applied to the construction of projection operators which project, from a weight vector of weight λ, a maximal weight vector of weight λ. It is demonstrated that our projection operator construction, in a sense, is the simplest possible. We conclude with some remarks concerning cyclic modules and the application of our methods to the various state labeling problems of physical interest.

II. NOTATION AND FUNDAMENTALS

Our notation follows that of Humphreys.[13] We let L denote a complex semisimple Lie algebra of rank l and let U denote the universal enveloping algebra of L. Throughout we shall assume that a fixed choice of Cartan subalgebra $H \subseteq L$ has been made, as well as a choice of base $\Delta = \{\alpha_1,...,\alpha_l\} \subseteq H^*$. We denote the set of roots by Φ and the corresponding set of positive roots by Φ^+. The fundamental dominant weights are denoted $\{\Lambda_1,...,\Lambda_l\}$; they are defined from Δ via

$$\langle \Lambda_i, \alpha_j \rangle = 2(\Lambda_i, \alpha_j)/(\alpha_j, \alpha_j) = \delta_{ij},$$

where $(\ ,\ )$ denotes the inner product induced on H^* by the Killing form. The set of all integral linear combinations of the fundamental dominant weights is denoted $\Lambda \subseteq H^*$, and the sublattice of all dominant integral weights is denoted $\Lambda^+ \subset \Lambda$ (i.e., those weights $\lambda \in \Lambda$ such that $\langle \lambda, \alpha_i \rangle \geqslant 0$, $i = 1,...,l$).

It is convenient to fix a standard set of generators for L. For each positive root α, let h_α be the element of H satisfying $\kappa(h_\alpha, h) = \alpha(h)$, for all $h \in H$, where

$$\kappa(x,y) = \mathrm{tr}[adx\ ady], \quad x, y \in L$$

denotes the Killing form. Fix x_α and y_α by the requirements

$$[h, x_\alpha] = \alpha(h)x_\alpha,$$
$$[h, y_\alpha] = -\alpha(h)y_\alpha,$$
$$[x_\alpha, y_\alpha] = h_\alpha,$$

for all $h \in H$. For $\alpha = \alpha_i \in \Delta$ we write h_i, etc., for h_{α_i}. The standard generating set is then $\{x_i, y_i, h_i; i = 1,...,l\}$. We can decompose L as

$$L = B \oplus H \oplus N,$$

where B (resp., N) is the nilpotent subalgebra generated by the $\{x_i\}$ (resp., $\{y_i\}$). According to the P.B.W. theorem[13] the universal enveloping algebra U may be factorized as $U = U(B)U(H)U(N)$, where $U(B)$ [resp., $U(H), U(N)$] is the universal enveloping algebra of B (resp., H, N).

If V is a U-module we denote by V_λ the subspace spanned by vectors of weight λ; viz. $V_\lambda = \{v \in V \mid hv = \lambda(h)v, \quad \forall\, h \in H\}$. If V is finite dimensional it can be shown[13] that V is the direct sum of its weight spaces. We note that the elements of $U(B)$ and $U(N)$ may be classified according to their weights under the adjoint action of H in U. We have in fact the following easy result (cf. Ref. 13).

Lemma 1: Let $\Phi^+ = \{\beta_1,...,\beta_m\}$. Then we have the following.

(a) $U(B)$ [resp., $U(N)$] is spanned by basis monomials

$$x_{\beta_1}^{i_1} \cdots x_{\beta_m}^{i_m} (\text{resp, } y_{\beta_1}^{i_1} \cdots y_{\beta_m}^{i_m}), \quad i_r \in \mathbb{Z}^+.$$

In particular $U(B)$ [resp., $U(N)$] is the direct sum of its weight spaces.

(b) The weight spaces of $U(B)$ and $U(N)$ are finite dimensional.

Proof: Part (a) is an immediate consequence of the P.B.W. theorem and the fact that $x_1,...,x_m$ (resp., $y_1,...,y_m$) form a basis for B (resp., N).

As to part (b) we note that the basis monomial

$$x_{\beta_1}^{i_1} \cdots x_{\beta_m}^{i_m} \tag{1}$$

has weight $\Sigma_{r=1}^{m} i_r \beta_r$. Rewriting each β_r as a positive $\mathbb{Z}$-linear combination of simple roots we see that the basis monomial (1) has weight of the form

$$\sum_{i=1}^{l} k_i \alpha_i, \quad k_i \in \mathbb{Z}^+. \tag{2}$$

Clearly there is only a finite number of basis monomials (1) for which $\Sigma_r\, i_r \beta_r$ equals a prescribed weight (2). An analogous statement holds for the basis monomials of $U(N)$.

Q.E.D.

If V is a U-module and $v \in V$ then Uv is a submodule of V. We call V a cyclic U-module if there exists $v \in V$ such that $V = Uv$. We let $I(v) \subseteq U$ denote the left ideal in U which annihilates the vector v; i.e.,

$$I(v) = \{u \in U \mid uv = 0\}. \tag{3}$$

From the isomorphism theorem (cf. Ref. 17) we obtain the following isomorphism of U-modules:

$$Uv \cong U/I(v), \tag{4}$$

where the rhs denotes the set of left cosets of U modulo $I(v)$.

Let $V(\lambda)$ be a finite-dimensional irreducible U-module with highest weight $\lambda \in \Lambda^+$ and let v^λ_+ denote the highest weight vector of $V(\lambda)$. By definition of maximal weight vector and the P.B.W. theorem we have

$$V(\lambda) = Uv^\lambda_+ = U(N)U(H)U(B)v^\lambda_+ = U(N)v^\lambda_+ . \quad (5)$$

The left annihilator of v^λ_+, denoted $I_+(\lambda) = I(v^\lambda_+)$, is known[11,13] to be given explicitly by

$$I_+(\lambda) = \sum_{i=1}^{l} Ux_i + \sum_{i=1}^{l} Uy_i^{\langle \lambda+\delta,\alpha_i\rangle} + \sum_{i=1}^{l} U[h_i - \lambda(h_i)], \quad (6)$$

where δ denotes the half-sum of the positive roots (recall[13] that $\langle \delta,\alpha_i\rangle = 1$ for $i = 1,...,l$). As well as a highest weight vector the module $V(\lambda)$ possesses a unique (up to scalar multiples) lowest weight vector, denoted v^λ_-, of weight $-\lambda^*$, where λ^* denotes the highest weight of the irreducible module $V(\lambda)^*$ contragredient to $V(\lambda)$. In analogy with Eq. (5) we obtain

$$V(\lambda) = Uv^\lambda_- = U(B)U(H)U(N)v^\lambda_- = U(B)v^\lambda_- . \quad (7)$$

Similarly in analogy with Eq. (6) the left annihilator of v^λ_-, denoted $I_-(\lambda) = I(v^\lambda_-)$, is given explicitly by

$$I_-(\lambda) = \sum_{i=1}^{l} Uy_i + \sum_{i=1}^{l} Ux_i^{\langle \lambda^*+\delta,\alpha_i\rangle} + \sum_{i=1}^{l} U[h_i + \lambda^*(h_i)]. \quad (8)$$

We let $>$ denote the usual ordering induced on the weights by the positive roots; i.e., $\lambda > \mu$ if and only if $\lambda - \mu$ is a positive Z-linear combination of positive roots. If $V(\lambda)$ is finite dimensional and irreducible with highest (resp., lowest) weight λ (resp., $-\lambda^*$) then the weights $\mu \neq \lambda$ (resp., $\mu \neq -\lambda^*$) occurring in $V(\lambda)$ necessarily satisfy $\mu < \lambda$ (resp., $\mu > -\lambda^*$).

Finally we let C_L denote the universal Casimir invariant of L. If $V(\lambda)$ is irreducible with highest weight $\lambda \in \Lambda^+$, Schur's lemma implies C_L takes a constant value on $V(\lambda)$. This eigenvalue is given by the well-known formula[13]

$$\chi_\lambda(C_L) = (\lambda, \lambda + 2\delta).$$

III. FINITE-DIMENSIONAL CYCLIC MODULES

We consider here certain elementary properties of finite-dimensional cyclic modules. Unless otherwise stated $V = Uv^\mu$ denotes a finite-dimensional cyclic module generated by a weight vector v^μ of weight $\mu \in \Lambda$. Set $V^{(0)} = U(B)v^\mu$ and let $\Pi^{(0)}$ denote the set of distinct weights in $V^{(0)}$. For $\lambda \in \Pi^{(0)}$, let $m_0(\lambda)$ denote the multiplicity of the weight λ in $V^{(0)}$; i.e., $m_0(\lambda)$ is the dimension of the weight space $V^{(0)}_\lambda$.

Since V is finite dimensional, Weyl's theorem guarantees that V is a direct sum of irreducible submodules. Some of the properties of this decomposition are given in the following (notation as above).

Theorem 1: Let $V = Uv^\mu$ be a finite-dimensional cyclic module and let

$$V = \bigoplus_\lambda V(\lambda) \quad (9)$$

be the decomposition of V into irreducible submodules. Then the following hold.

(a) The highest weights occurring in the decomposition (9) are of the form $\lambda \in \Pi^{(0)} \cap \Lambda^+$. In particular $\lambda = \mu$ or $\lambda > \mu$.

(b) The irreducible module $V(\lambda)$ occurs in V with multiplicity $m_V(\lambda) \leqslant m_0(\lambda)$. In particular $m_V(\mu) \leqslant 1$.

(c) $m_V(\mu) = 1$ if and only if $v^\mu \notin \Sigma_{i=1}^{l} Ux_i v^\mu$.

Proof: Let $V(\lambda) \subseteq V$ be irreducible with highest weight vector v^λ_+. Since v^μ cyclically generates V it follows that v^μ cannot be orthogonal to $V(\lambda)$, i.e.,

$$\langle v^\mu \mid V(\lambda)\rangle \neq 0,$$

because

$$\begin{aligned} 0 \neq \langle v^\mu \mid V(\lambda)\rangle &\Leftrightarrow \langle v^\mu \mid U(N)v^\lambda_+\rangle \neq 0 \\ &\Leftrightarrow \langle U(B)v^\mu \mid v^\lambda_+\rangle \neq 0 \\ &\Leftrightarrow \langle V^{(0)} \mid P_0 v^\lambda_+\rangle \neq 0. \end{aligned}$$

Now let P_0 be the orthogonal projector onto $V^{(0)}$. Then we have [see Eq. (5)]

$$\begin{aligned} 0 \neq \langle v^\mu \mid V(\lambda)\rangle &\Leftrightarrow \langle v^\mu \mid U(N)v^\lambda_+\rangle \neq 0 \\ &\Leftrightarrow \langle U(B)v^\mu \mid v^\lambda_+\rangle \neq 0 \\ &\Leftrightarrow \langle V^{(0)} \mid P_0 v^\lambda_+\rangle \neq 0. \end{aligned}$$

Thus it follows that

$$P_0 v^\lambda_+ \neq 0. \quad (10)$$

If $W^{(+)} \subseteq V$ denotes the subspace spanned by maximal weight vectors, Eq. (10) implies that P_0 is 1–1 on $W^{(+)}$; i.e.,

$$P_0 W^{(+)} \subseteq V^{(0)}, \quad \text{with dim } W^{(+)} = \dim P_0 W^{(+)}.$$

If $W^{(+)}_\lambda \subseteq W^{(+)}$ denotes the subspace of maximal weight vectors of weight λ then $P_0 W^{(+)}_\lambda \subseteq V^{(0)}_\lambda$, and the irreducible module $V(\lambda)$ occurs in V with multiplicity

$$m(\lambda) = \dim W^{(+)}_\lambda = \dim P_0 W^{(+)}_\lambda \leqslant \dim V^{(0)}_\lambda = m_0(\lambda).$$

In particular, since $m_0(\mu) = \dim V^{(0)}_\mu = 1$, we see that the module $V(\mu)$ occurs with at most multiplicity 1. By definition, in order for $V(\lambda)$ to occur, we must have $\lambda \in \Lambda^+$. This proves parts (a) and (b).

As to part (c), set

$$Y = \sum_{i=1}^{l} Ux_i v^\mu .$$

By part (a) of the theorem we note that $Ux_i v^\mu$ is a cyclic module and the irreducible representations occurring in $Ux_i v^\mu$ have highest weights $\lambda > \mu$. Thus Y is a direct sum of irreducible submodules with highest weights $\lambda > \mu$ and it follows that $V(\mu)$ cannot occur in Y.

Thus if $V(\mu)$ occurs in V [i.e., $m_V(\mu) = 1$] then $Y \neq V$ whence $v^\mu \notin Y$. Conversely if $v^\mu \notin Y$ let V' denote the orthogonal complement of Y in V and let P' be the projection onto V'. Then $v^\mu_+ = P'v^\mu$ is a maximal weight vector of weight μ since $x_i v^\mu_+ = P' x_i v^\mu = 0$ (because $x_i v^\mu \in Y$, $i = 1,...,l$). Thus $V(\mu)$ must occur as a submodule of V. Note that because $v^\mu_+ = P'v^\mu$ cyclically generates V' it follows that $V' = V(\mu)$ is irreducible with highest weight μ and we have a decomposition $V = V(\mu) \oplus Y$. This completes the proof of the theorem.

Q.E.D.

The above result gives some information concerning the decomposition (9) of a cyclic module V into irreducible sub-

modules. However we have not yet determined the multiplicity with which the irreducible module $V(\lambda)$ occurs in the decomposition (9). As noted in Theorem 1 it suffices to determine the highest weight vectors which occur in V. Alternatively the decomposition (9) may be uniquely characterized by the ring of U-module homomorphisms (or intertwining operators) on the space V.

If V and W are U-modules, a U-module homomorphism (or intertwining operator) is defined as a vector space mapping T: $V \to W$ which intertwines the action of L and hence U; viz. $Tuv = uTv$ for all $u \in U$, $v \in V$. It is customary to denote the space of intertwining operators from V onto W by $\mathrm{Hom}_L(V, W)$. We set

$$(V{:}W) = \dim \mathrm{Hom}_L(\mathrm{V}, \mathrm{W}).$$

In view of Weyl's theorem of complete reducibility one may establish that

$$(V{:}W) = (W{:}V).$$

If V is a finite dimensional U-module then clearly the irreducible module $V(\lambda)$ occurs in V if and only if there exists a (nontrivial) intertwining operator from $V(\lambda)$ into V [or equivalently an intertwining operator from V onto $V(\lambda)$]. The irreducible module occurs in V with multiplicity $(V(\lambda){:}V)$. Moreover, if $V = Uv^{\mu}$ is cyclic then an intertwining operator $T{:}V \to V(\lambda)$ is uniquely determined by its action on the cyclic vector v^{μ}; $Tuv^{\mu} = uTv^{\mu}$ for all $u \in U$. We have the following result (notation as in Sec. II).

Lemma 2: Let $V = Uv^{\mu}$ be a finite-dimensional cyclic module and let $I(v^{\mu}) \subseteq U$ denote the left annihilator of v^{μ} in U. Suppose $V(\lambda)$ is finite-dimensional and irreducible and set

$$S(\lambda) = \{v \in V_{\mu}(\lambda) | I(v^{\mu})v = 0\}.$$

Then the irreducible module $V(\lambda)$ occurs in V with multiplicity $\dim S(\lambda)$.

Proof: Let $S(\lambda)$ be as in the statement of the lemma. In view of our previous remarks it suffices to show that $\dim S(\lambda) = \dim \mathrm{Hom}_L(V, V(\lambda))$. We set up a vector space homomorphism $\varphi{:}\mathrm{Hom}_L(V, V(\lambda)) \to S(\lambda)$, $T \to Tv^{\mu}$. Since T is an intertwining operator and $I(v^{\mu})v^{\mu} = 0$, we must have $I(v^{\mu})Tv^{\mu} = 0$ (and in particular Tv^{μ} has weight μ); i.e., $Tv^{\mu} \in S(\lambda)$. This establishes that φ is well defined. We show that φ is a vector space isomorphism. The cyclic nature of the vector v^{μ} guarantees that φ is 1–1 ($Tv^{\mu} = 0$ implies $uTv^{\mu} = Tuv^{\mu} = 0$ for all $u \in U$; i.e., $T = 0$). It remains to show φ is onto. Choose $v \in S(\lambda)$; we set up an intertwining operator $T{:}V \to V(\lambda)$ defined by $Tuv^{\mu} = uv^{\mu}$ for all $u \in U$. Since $I(v^{\mu})v^{\mu} = 0$, it is clear that T is a well-defined element of $\mathrm{Hom}_L(V, V(\lambda))$ and $\varphi(T) = v^{\mu}$. This shows that φ is onto whence φ is an isomorphism of vector spaces.

Q.E.D.

The above result on multiplicities looks formally promising but its applicability is determined by the ease with which the left annihilator of the cyclic vector v^{μ} may be obtained. The results of Theorem (1) nevertheless indicate that certain properties of the cyclic module V may be obtained simply from a knowledge of the weights in the space $U(B)v^{\mu}$. With regard to part (c) of Theorem 1, we note that in order for $V(\mu)$ to occur in the cyclic module V we must have $\mu \in \Lambda^{+}$ (this condition is necessary but not sufficient). In such a case we may construct a maximal weight state of weight μ by projection from the cyclic vector v^{μ}. We have in fact the following result.

Lemma 3: Suppose $\mu \in \Lambda^{+}$ and let $V = Uv^{\mu}$ be a finite-dimensional cyclic module. Set $V^{(0)} = U(B)v^{\mu}$ and let $\Pi^{(0)}$ denote the set of distinct weights in $V^{(0)}$. Put $\Gamma = \Lambda^{+} \cap \Pi^{(0)} \sim \{\mu\}$ and set

$$P = \prod_{\lambda \in \Gamma} \left(\frac{C_L - \chi_{\lambda}(C_L)}{\chi_{\mu}(C_L) - \chi_{\lambda}(C_L)} \right).$$

Then $v^{\mu}_{+} = Pv^{\mu}$ is a maximal weight vector of weight μ. The module $V(\mu)$ occurs in V if and only if $Pv^{\mu} \neq 0$.

Proof: In order to prove the result it clearly suffices to show that the universal Casimir element C_L separates the weight μ from the weights in Γ. We have

$$\begin{aligned} \chi_{\lambda}(C_L) - \chi_{\mu}(C_L) &= (\lambda, \lambda + 2\delta) - (\mu, \mu + 2\delta) \\ &= (\lambda - \mu, \lambda + \mu + 2\delta). \end{aligned} \tag{11}$$

For $\alpha \in \Phi^{+}$ we have $(\delta, \alpha) > 0$ and for $\lambda, \mu \in \Lambda^{+}$ we have $(\lambda + \mu, \alpha) \geqslant 0$ whence $(\lambda + \mu + 2\delta, \alpha) > 0$. If $\lambda \in \Gamma$ we have $\lambda > \mu$ whence $\lambda - \mu$ is a sum of positive roots and in such a case the rhs of Eq. (11) cannot vanish. This proves the result.

Q.E.D.

This result shows that one may construct projection operators which project from a weight vector v^{μ} of weight $\mu \in \Lambda^{+}$ a maximal weight vector of weight μ. Note that since only one copy of $V(\mu)$ can occur in the cyclic module $V = Uv^{\mu}$, $v^{\mu}_{+} = Pv^{\mu}$ is the unique (up to scalar multiples) maximal weight vector of weight $\mu \in \Lambda^{+}$ which can occur in V and hence any maximal weight state of weight $\mu \in \Lambda^{+}$ constructed from v^{μ} by projection necessarily yields the vector $v^{\mu}_{+} = Pv^{\mu}$.

The drawback in this method however is that one requires a knowledge of the weights occurring in the (finite-dimensional) space $U(B)v^{\mu}$. In Sec. V we shall consider an alternative construction of such projection operators which requires far less information. We note also that the projection operator P in Lemma 3 may be represented by the formal infinite product

$$P = \prod_{\substack{\lambda \in \Lambda^{+} \\ \lambda > \mu}} \left(\frac{C_L - \chi_{\lambda}(C_L)}{\chi_{\mu}(C_L) - \chi_{\lambda}(C_L)} \right).$$

An important example of a cyclic module is afforded by the tensor product of two irreducible modules $V(\lambda) \otimes V(\mu)$; $\lambda, \mu \in \Lambda^{+}$. We have the following result (notation as in Sec. II).

Lemma 4: $V(\lambda) \otimes V(\mu)$ is cyclically generated by the vector $v_0 = v^{\lambda}_{-} \otimes v^{\mu}_{+}$.

Proof: From the definition of maximal weight vector we have $x_i v^{\mu}_{+} = 0$, which implies $U(B)v^{\mu}_{+} = \mathbb{C}v^{\mu}_{+}$. Equation (7) then implies

$$U(B)v_0 = [U(B)v^{\lambda}_{-}] \otimes v^{\mu}_{+} = V(\lambda) \otimes v^{\mu}_{+}. \tag{12}$$

Now set $V_0 = Uv_0$ and consider the subspace of $V(\mu)$ defined by $V = \{v \in V(\mu) | V(\lambda) \otimes v \subseteq V_0\}$. Clearly V is a U-submodule of $V(\mu)$ since for $v \in V$, $x \in L$,

$$\omega \otimes xv = x(\omega \otimes v) - (x\omega) \otimes v \in V_0, \quad \text{for all } \omega \in V(\lambda),$$

which implies $xv \in V$ for all $x \in L$. However V is nontrivial

since, by Eq. (12), $v^{\mu}_{+} \in V$ whence irreducibility of $V(\mu)$ forces $V = V(\mu)$. Thus we must have $V_0 = Uv_0 = V(\lambda) \otimes V(\mu)$ and the result is proved

Q.E.D.

Note that the cyclic vector $v_0 = v^{\lambda}_{-} \otimes v^{\mu}_{+}$ has weight $\mu - \lambda^*$ (cf. Sec. II). The subspace $V^{(0)} = U(B)v_0$ in this case is given by [cf. Eq. (12)] $V^{(0)} = V(\lambda) \otimes v^{\mu}_{+}$. Hence, as a particular case of Theorem 1, we obtain the following well-known result.[13]

Lemma 5: The irreducible representations occurring in $V(\lambda) \otimes V(\mu)$ have highest weights of the form $\mu + \nu \in \Lambda^+$, where ν is a weight in $V(\lambda)$. Moreover the multiplicity $m(\mu + \nu{:}\lambda \otimes \mu)$ of $V(\mu + \nu)$ in $V(\lambda) \otimes V(\mu)$ is less than or equal to the multiplicity of the weight ν in $V(\lambda)$; i.e., $m(\mu + \nu{:}\ \lambda \otimes \mu) \leqslant \dim V_{\nu}(\lambda)$.

□

In the case of the tensor product of two irreducible modules the methods used in the proof of Theorem 1 may be extended to yield the following result originally due to Parthasarathy, Ranga Rao, and Varadarajan[10] (see also Ref. 6). We follow the argument presented in Ref. 6 [notation as in Lemma 5].

Theorem 2: Set $v_0 = v^{\lambda}_{-} \otimes v^{\mu}_{+}$, $V^{(0)} = V(\lambda) \otimes v^{\mu}_{+}$ and let $V_{\nu,\mu}(\lambda) \subseteq V_{\nu}(\lambda)$ be the subspace defined by

$$V_{\nu,\mu}(\lambda) = \{v \in V_{\nu}(\lambda) | y_i^{\langle \mu + \nu + \delta, \alpha_i \rangle} v = 0;\ i = 1,\ldots,l\}. \tag{13}$$

Put $(\lambda{:}\ \nu, \mu) = \dim V_{\nu,\mu}(\lambda)$. Then the multiplicity of the irreducible module $V(\mu + \nu)$ in $V(\lambda) \otimes V(\mu)$ is given by $m(\mu + \nu{:}\ \lambda \otimes \mu) = (\lambda{:}\ \nu, \mu)$.

Proof: Let P_0 be the orthogonal projection of $V(\lambda) \otimes V(\mu)$ onto $V^{(0)}$ and let $W^{(+)}_{\mu+\nu}$ denote the subspace of maximal weight vectors of weight $\mu + \nu \in \Lambda^+$. From the results of Theorem 1 P_0 is 1–1 on $W^{(+)}_{\mu+\nu}$ hence to prove the theorem it suffices to show that $P_0 W^{(+)}_{\mu+\nu} = V_{\nu,\mu}(\lambda) \otimes v^{\mu}_{+}$.

Now $P_0 W^{(+)}_{\mu+\nu} \subseteq V^{(0)}_{\mu+\nu} = V_{\nu}(\lambda) \otimes v^{\mu}_{+}$. We show that $P_0 W^{(+)}_{\mu+\nu} \subseteq V_{\nu,\mu}(\lambda) \otimes v^{\mu}_{+}$. Now since $V^{(0)}$ is stable under $U(B)$ it follows that the orthogonal complement of $V^{(0)}$ is stable under $U(N)$: viz.,

$$P_0 n(1 - P_0) = 0, \quad \text{for all } n \in U(N). \tag{14}$$

We note also that if $n \in U(N)$ then $P_0 n(v \otimes v^{\mu}_{+}) = (nv) \otimes v^{\mu}_{+}$, for all $v \in V(\lambda)$. Suppose now that $v^{\mu+\nu}_{+} \in W^{(+)}_{\mu+\nu}$ and set $v \otimes v^{\mu}_{+} = P_0 v^{\mu+\nu}_{+}$, $v \in V_{\nu}(\lambda)$. From Eq. (6) we have

$$y_i^{\langle \mu + \nu + \delta, \alpha_i \rangle} v^{\mu+\nu}_{+} = 0, \quad i = 1,\ldots,l$$

which implies in view of Eq. (14) above that

$$0 = P_0 y_i^{\langle \mu + \nu + \delta, \alpha_i \rangle} v^{\mu+\nu}_{+} = P_0 y_i^{\langle \mu + \nu + \delta, \alpha_i \rangle} P_0 v^{\mu+\nu}_{+}$$

$$= P_0 y_i^{\langle \mu + \nu + \delta, \alpha_i \rangle} v \otimes v^{\mu}_{+} = (y_i^{\langle \mu + \nu + \delta, \alpha_i \rangle} v) \otimes v^{\mu}_{+}.$$

Thus $y_i^{\langle \mu + \nu + \delta, \alpha_i \rangle} v = 0$; i.e., $v \in V_{\nu,\mu}(\lambda)$. This shows that

$$P_0 W^{(+)}_{\mu+\nu} \subseteq V_{\nu,\mu}(\lambda) \otimes v^{\mu}_{+}.$$

To prove the reverse inclusion we need to exhibit $v \otimes v^{\mu}_{+} \in V_{\nu,\mu}(\lambda) \otimes v^{\mu}_{+}$ as the image under P_0 of an element of $W^{(+)}_{\mu+\nu}$. To this end we construct an intertwining operator

$$T_v : V(\mu + \nu) \to V(\lambda) \otimes V(\mu)$$

satisfying

$$P_0 T_v v^{\mu+\nu}_{+} = v \otimes v^{\mu}_{+}.$$

We define T_v to be the operator with matrix elements[6]

$$\langle n_1(v' \otimes v^{\mu}_{+}) | T_v n_2 v^{\mu+\nu}_{+} \rangle = \langle n_1 v' | n_2 v \rangle,$$

for all $n_1, n_2 \in U(N)$. The consistency of this definition follows because $v \in V_{\nu,\mu}(\lambda)$ and because $n_1(v' \otimes v^{\mu}_{+}) = 0$ implies $n_1 v' = 0$; $n_1(v' \otimes v^{\mu}_{+}) = 0 \Rightarrow 0 = P_0 n_1(v' \otimes v^{\mu}_{+}) = (n_1 v') \otimes v^{\mu}_{+} \Rightarrow n_1 v' = 0$. The definition is complete since $V(\mu + \nu) = U(N) v^{\mu+\nu}_{+}$ and $V(\lambda) \otimes V(\mu) = U(N)U(B) v^{\lambda}_{-} \otimes v^{\mu}_{+} = U(N) V(\lambda) \otimes v^{\mu}_{+}$; i.e., any $\omega \in V(\lambda) \otimes V(\mu)$ can be written

$$\omega = n(v' \otimes v^{\mu}_{+}), \quad \text{for some } n \in U(N), \quad v' \in V(\lambda).$$

The intertwining property of T_v is easily checked.[6]

Hence $v \in V_{\nu,\mu}(\lambda)$ implies $T_v \in \mathrm{Hom}_L(V(\mu + \nu), V(\lambda) \otimes V(\mu))$ and $P_0 T_v v^{\mu+\nu}_{+} = v \otimes v^{\mu}_{+}$ (cf. Ref. 6). But $T_v v^{\mu+\nu}_{+} \in W^{(+)}_{\mu+\nu}$, whence $V_{\nu,\mu}(\lambda) \otimes v^{\mu}_{+} \subseteq P_0 W^{(+)}_{\mu+\nu}$.

We have thus proved $P_0 W^{(+)}_{\mu+\nu} = V_{\nu,\mu}(\lambda) \otimes v^{\mu}_{+}$ and since P_0 is 1–1 on $W^{(+)}_{\mu+\nu}$ we must have

$$m(\mu + \nu{:}\ \lambda \otimes \mu) = \dim W^{(+)}_{\mu+\nu} = \dim P_0 W^{(+)}_{\mu+\nu}$$

$$= \dim V_{\nu,\mu}(\lambda) = (\lambda{:}\ \nu, \mu).$$

This proves the theorem.

Q.E.D.

Corollary:

$$m(\mu + \nu{:}\ \lambda \otimes \mu) = m(\mu{:}\ \mu + \nu \otimes \lambda^*)$$

$$= m(\lambda{:}\ \mu + \nu \otimes \mu^*).$$

In particular,

$$m(\mu + \nu{:}\ \lambda \otimes \mu) = (\mu + \nu{:}\ \mu - \lambda^*, \lambda^*)$$

$$= \dim V_{\mu - \lambda^*, \lambda^*}(\mu + \nu).$$

Proof: To prove the above result we note that if V and W are finite-dimensional U-modules then we have a natural U-module isomorphism

$$V \otimes W \cong \mathrm{Hom}(V^*, W).$$

Now recall (see remarks preceding Lemma 2)

$$m(\mu + \nu{:}\ \lambda \otimes \mu) = \dim \mathrm{Hom}_L(V(\lambda) \otimes V(\mu), V(\mu + \nu)),$$

which equals the multiplicity of the trivial representation in the triple tensor product $V^*(\lambda) \otimes V^*(\mu) \otimes V(\mu + \nu)$. Rearranging the terms in this tensor product the required symmetry properties are seen to hold. In particular we have

$$m(\mu + \nu{:}\ \lambda \otimes \mu) = \dim \mathrm{Hom}_L(V(\mu), V^*(\lambda) \otimes V(\mu + \nu))$$

$$= \mathrm{m}(\mu{:}\ \mu + \nu \otimes \lambda^*).$$

In view of Theorem 2 we have

$$m(\mu{:}\ \mu + \nu \otimes \lambda^*) = (\mu + \nu{:}\ \mu - \lambda^*, \lambda^*)$$

$$= \dim V_{\mu - \lambda^*, \lambda^*}(\mu + \nu),$$

which establishes the result.

Q.E.D.

Since the tensor product module $V(\lambda) \otimes V(\mu)$ is cyclically generated by the vector $v^{\lambda}_{-} \otimes v^{\mu}_{+}$ one may, in principle, obtain the detailed structure of this space provided one can obtain the annihilator $I(v^{\lambda}_{-} \otimes v^{\mu}_{+})$ of the cyclic vector $v^{\lambda}_{-} \otimes v^{\mu}_{+}$ [cf. Eq. (4)]. It is our aim, in the following section to determine the explicit form of the annihilator $I(v^{\lambda}_{-} \otimes v^{\mu}_{+})$.

IV. ANNIHILATORS IN THE UNIVERSAL ENVELOPING ALGEBRA

We begin with some comments concerning the idea of degrees of vectors introduced in Ref. 6. Let $V(\lambda)$ be a finite-dimensional irreducible U-module and suppose $v \in V(\lambda)$. Following the notation of Ref. 6 we call the collection of integers $\mathbf{m} = (m_1, m_2, \ldots, m_l)$ (l = rank L = dim H) the degree of the vector v, where m_i is the smallest non-negative integer such that

$$x_i^{m_i+1} v = 0. \tag{15}$$

By analogy one may also consider the set of positive integers $\hat{\mathbf{m}} = (\hat{m}_1, \ldots, \hat{m}_l)$ (herein called the codegree of v), where $\hat{m}_i$ is the smallest positive integer such that

$$y_i^{\hat{m}_i+1} v = 0. \tag{16}$$

The degree $\mathbf{m}$ and codegree $\hat{\mathbf{m}}$ uniquely determine each other and are related via

$$\hat{m}_i = m_i + \langle \nu, \alpha_i \rangle. \tag{17}$$

This occurs due to the known result (see, e.g., Humphreys[13]) that the α_i-string through the weight ν is of length $\langle \nu, \alpha_i \rangle$. We note that if V is any finite-dimensional U-module and $v \in V_\nu$ then we may still define the degree (resp., codegree) $\mathbf{m}$ (resp., $\hat{\mathbf{m}}$) as in Eqs. (15) and (16) above. Due to Weyl's theorem of complete reducibility the degree and codegree of the vector v are still related via Eq. (17) (regardless of whether V is irreducible or not). It is interesting to note that if V is any finite-dimensional U-module and $v \in V_\nu$ then Eqs. (15)–(17) above imply the result

$$x_i^{m+1} v = 0 \text{ if and only if } y_i^{m+\langle \nu, \alpha_i \rangle + 1} v = 0. \tag{18}$$

It is our aim in this section to prove the following result (notation as in Sec. II and III).

Theorem 3:

$$I(v_-^\lambda \otimes v_+^\mu) = \sum_{i=1}^{l} U x_i^{\langle \lambda^* + \delta, \alpha_i \rangle} + \sum_{i=1}^{l} U y_i^{\langle \mu + \delta, \alpha_i \rangle} + \sum_{i=1}^{l} U[h_i - (\mu - \lambda^*)(h_i)]. \quad \square \tag{19}$$

We denote the left ideal on the rhs of Eq. (19) by $I(\lambda, \mu)$. In view of Eqs. (6) and (8) it is clear that $I(\lambda, \mu) \subseteq I(v_-^\lambda \otimes v_+^\mu)$. The proof of the opposite inclusion is not so obvious and requires some work. We set $V(\lambda, \mu) = U/I(\lambda, \mu)$. We begin by showing that $V(\lambda, \mu)$ is finite dimensional. We require first the following elementary result[10] (notation as in Sec. II).

Lemma 6:

(a) $I(\lambda, \mu) \cap U(N) = I(v_-^\lambda \otimes v_+^\mu) \cap U(N) = I_+(\mu) \cap U(N) = \sum_{i=1}^{l} U(N) y_i^{\langle \mu + \delta, \alpha_i \rangle}$.

(b) $I(\lambda, \mu) \cap U(B) = I(v_-^\lambda \otimes v_+^\mu) \cap U(B) = I_-(\lambda) \cap U(B) = \sum_{i=1}^{l} U(B) x_i^{\langle \lambda^* + \delta, \alpha_i \rangle}$.

Proof: It is clear that $I(v_-^\lambda \otimes v_+^\mu) \cap U(N) = I_+(\mu) \cap U(N)$ since, for $n \in U(N)$, we have

$$n \in I(v_-^\lambda \otimes v_+^\mu) \Leftrightarrow n(v_-^\lambda \otimes v_+^\mu) = 0 \Leftrightarrow v_-^\lambda \otimes (n v_+^\mu) = 0 \Leftrightarrow n \in I_+(\mu).$$

It also follows, by definition of $I_+(\mu)$ and $I(\lambda, \mu)$, that

$$I(\lambda, \mu) \cap U(N) = I_+(\mu) \cap U(N).$$

The result

$$I_+(\mu) \cap U(N) = \sum_{i=1}^{l} U(N) y_i^{\langle \mu + \delta, \alpha_i \rangle}$$

was proved as Lemma (2.1) of Ref. 10 and will not be reproduced here. An analogous argument holds for part (b). Q.E.D.

Lemma 7: (a) The space $V(\lambda, \mu)$ is finite dimensional. (b) $V(\lambda) \otimes V(\mu)$ is a submodule of $V(\lambda, \mu)$.

Proof: (a) Consider the left coset $\bar{1} = 1 + I(\lambda, \mu) \in V_{\mu - \lambda^*}(\lambda, \mu)$, and put $V_0(\lambda, \mu) = U(B)\bar{1}$.

We set up the following $U(B)$-intertwining operator:

$$T: V_0(\lambda, \mu) \to V(\lambda) \otimes v_+^\mu;$$

$$T\bar{1} = v_-^\lambda \otimes v_+^\mu, \quad Tb\bar{1} = b(v_-^\lambda \otimes v_+^\mu),$$

for all $b \in U(B)$. Since $I(\lambda, \mu) \cap U(B) = I(v_-^\lambda \otimes v_+^\mu) \cap U(B)$ it follows that T is a well-defined intertwining operator [in fact T determines an isomorphism of $U(B)$-modules]. We note also that T intertwines the action of $U(H)$. Thus, in particular, $V_0(\lambda, \mu)$ is finite dimensional and the weights occurring in $V_0(\lambda, \mu)$ are bounded above (with respect to partial ordering $>$ induced on the weights by the positive roots) by the weight $\mu + \lambda$. Moreover from the P.B.W. theorem we have

$$V(\lambda, \mu) = U(N)U(B)U(H)\bar{1} = U(N)U(B)\bar{1} = U(N)V_0(\lambda, \mu), \tag{20}$$

which shows that all weights in $V(\lambda, \mu)$ are bounded above by the weight $\mu + \lambda$. In an analogous way we have

$$V(\lambda, \mu) = U(B)U(N)\bar{1},$$

and we may establish that $U(N)\bar{1}$ is finite dimensional and that the weights in $V(\lambda, \mu)$ are bounded below by the weight $-(\mu^* + \lambda^*)$. Clearly there can be only a finite number of weights $\nu \in \Lambda$ satisfying the inequalities $\mu + \lambda > \nu > -(\mu^* + \lambda^*)$. This shows that only a finite number of distinct weights can occur in $V(\lambda, \mu)$. It remains to show that the weights occur with finite multiplicity.

Choose a weight basis $v_1, \ldots, v_d$ [d = dim $V(\lambda)$] for $V_0(\lambda, \mu) = U(B)\bar{1}$. In view of Eq. (20) we have $V(\lambda, \mu) = \Sigma_{i=1}^{d} U(N) v_i$. However Lemma (1) asserts that the weight spaces of $U(N)$ are finite dimensional which implies that the weight spaces of each $U(N)v_i$ are finite dimensional ($i = 1, \ldots, d$). Thus the weight spaces in $V(\lambda, \mu)$ are all finite dimensional whence the finite dimensionality of $V(\lambda, \mu)$ follows.

(b) We set up an intertwining operator

$$T: V(\lambda, \mu) \to V(\lambda) \otimes V(\mu), \tag{21}$$

defined by $T\bar{1} = v_-^\lambda \otimes v_+^\mu$, $T\bar{u} = u(v_-^\lambda \otimes v_+^\mu)$ for all $u \in U$. Since $I(\lambda, \mu) \subseteq I(v_-^\lambda \otimes v_+^\mu)$ it is easily checked that T is a well-defined intertwining operator. The cyclicity of the vector $v_-^\lambda \otimes v_+^\mu$ shows moreover that T is onto. This implies a U-module isomorphism

$$V(\lambda)\otimes V(\mu)\cong V(\lambda,\mu)/\ker T.$$

However part (a) shows that $V(\lambda,\mu)$ is finite dimensional and hence completely reducible which implies that $V(\lambda)\otimes V(\mu)$ must occur as a submodule of $V(\lambda,\mu)$.

Q.E.D.

We note that the proof of Lemma 7(a) implies that the weights in the spaces $V_0(\lambda,\mu)=U(B)\bar{1}$ and $V(\lambda)\otimes v_+^\mu = U(B)(v_-^\lambda\otimes v_+^\mu)$ are the same.

We are now in a position to set up the last step in the proof of Theorem 3. It clearly suffices to show, since $V(\lambda)\otimes V(\mu)\subseteq V(\lambda,\mu)$, that the irreducible modules in these spaces occur with the same multiplicities. We recall from Eq. (13) the definition of the spaces

$$V_{\nu,\mu}(\lambda)=\{v\in V_\nu(\lambda)\,|\,y_i^{\langle\mu+\nu+\delta,\alpha_i\rangle}v=0;\quad i=1,\ldots,l\}.$$

From Eq. (18) we see that $v\in V_{\nu,\mu}(\lambda)$ if and only if $v\in V_\nu(\lambda)$ satisfies

$$x_i^{\langle\mu+\delta,\alpha_i\rangle}v=0,\quad i=1,\ldots,l.$$

In view of the definition of the ideals $I(\lambda,\mu)$ [see Eq. (19)] we thus have the following equivalent definition:

$$V_{\nu,\mu}(\lambda)=\{v\in V(\lambda)\,|\,I(\mu^*,\mu+\nu)v=0\}. \tag{22}$$

Since $U(B)\bar{1}\cong V(\lambda)\otimes v_+^\mu$ [$U(B)$-isomorphism] Theorem (1) guarantees that the highest weights in $V(\lambda,\mu)$ and $V(\lambda)\otimes V(\mu)$ are of the form $\mu+\nu$, where ν is a weight in $V(\lambda)$ (cf. remark following Lemma 7). Since $V(\lambda,\mu)$ is cyclically generated by the vector $\bar{1}$ of weight $\mu-\lambda^*$, Lemma 2 implies that the irreducible module $V(\rho),\rho\in\Lambda^+$, occurs in $V(\lambda,\mu)$ with multiplicity

$$m(\rho: V(\lambda,\mu))=\dim\{v\in V_{\mu-\lambda^*}(\rho)\,|\,I(\lambda,\mu)v=0\}$$
$$=\dim V_{\mu-\lambda^*,\lambda^*}(\rho)=(\rho:\mu-\lambda^*,\lambda^*).$$

On the other hand the corollary to Theorem 2 implies that the irreducible module $V(\rho)$ occurs in $V(\lambda)\otimes V(\mu)$ with multiplicity $m(\rho:\lambda\otimes\mu)=(\rho:\mu-\lambda^*,\lambda^*)$. This establishes that the irreducible modules occurring in $V(\lambda)\otimes V(\mu)$ and $V(\lambda,\mu)$ occur with the same multiplicities whence we necessarily have $V(\lambda,\mu)\cong V(\lambda)\otimes V(\mu)$.

This shows that the intertwining operator of Eq. (21) determines a U-module isomorphism whence we must have $I(\lambda,\mu)=I(v_-^\lambda\otimes v_+^\mu)$ and Theorem 3 is proved.

We therefore have the U-module isomorphism

$$V(\lambda)\otimes V(\mu)\cong U/I(\lambda,\mu),$$

where $I(\lambda,\mu)=I(v_-^\lambda\otimes v_+^\mu)$ is the left ideal defined in Theorem 3. This result has important consequences with regard to the Clebsch–Gordan multiplicity problem (i.e., the $L\oplus L\supset L$ state labeling problem). However we shall not pursue this problem any further here. For a more detailed discussion of these points, we refer to Ref. 6.

V. CONSTRUCTION OF PROJECTION OPERATORS

We consider here an alternative construction of the projection operators P introduced in Lemma 3. Following our previous notation let Uv^μ be a finite-dimensional cyclic module where v^μ is a weight vector of weight $\mu\in\Lambda$. We shall show that the required projection operators may be constructed provided one knows the degree of the vector v^μ [see Eq. (15)]. To this end we suppose that the vector v^μ has degree $\mathbf{m}=(m_1,\ldots,m_l)$. From Eq. (16) we see that v^μ has codegree $\hat{\mathbf{m}}=(\hat{m}_1,\ldots,\hat{m}_l)$ where $\hat{m}_i=m_i+\langle\mu,\alpha_i\rangle$. We have the following result (notation as in Sec. II).

Theorem 4: Let $V=Uv^\mu$ be a finite-dimensional cyclic module and suppose v^μ has degree $\mathbf{m}$. Set

$$\nu=\sum_{i=1}^{l}m_i\Lambda_i\in\Lambda^+.$$

Then $\mu+\nu\in\Lambda^+$ and V is identifiable as a submodule of $V(\nu^*)\otimes V(\mu+\nu)$.

Proof: Let $\nu\in\Lambda^+$ be as in the statement of the theorem. Since v^μ has degree $m_i=\langle\nu,\alpha_i\rangle$ and codegree $\hat{m}_i=m_i+\langle\mu,\alpha_i\rangle$ we have

$$\langle\mu+\nu,\alpha_i\rangle=\langle\nu,\alpha_i\rangle+\langle\mu,\alpha_i\rangle=m_i+\langle\mu,\alpha_i\rangle=\hat{m}_i\geqslant 0;$$

i.e., $\mu+\nu\in\Lambda^+$.

We now note (see Lemma 4) that $V(\nu^*)\otimes V(\mu+\nu)$ is cyclically generated by the vector $v_-^{\nu^*}\otimes v_+^{\mu+\nu}$. Also, from Theorem 3, the vector $v_-^{\nu^*}\otimes v_+^{\mu+\nu}$ has annihilator $I(\nu^*,\mu+\nu)\subseteq U$ [where $I(\lambda,\mu)$ is the ideal defined in Eq. (19)]. In view of the definition of the weight ν we see that $v_-^{\nu^*}\otimes v_+^{\mu+\nu}$ is a weight vector of weight μ and degree $\mathbf{m}$. Since the cyclic vector v^μ also has degree $\mathbf{m}$ we necessarily have [see Eq. (3)]

$$I(\nu^*,\mu+\nu)\subseteq I(v^\mu);\quad\text{i.e., } I(\nu^*,\mu+\nu)v^\mu=0.$$

We now set up an intertwining operator

$$T: V(\nu^*)\otimes V(\mu+\nu)\to V,\quad T(v_-^{\nu^*}\otimes v_+^{\mu+\nu})=v^\mu$$
$$Tu(v_-^{\nu^*}\otimes v_+^{\mu+\nu})=uv^\mu,\quad\text{for all } u\in U.$$

Since $I(\nu^*,\mu+\nu)\subseteq I(v^\mu)$ it is easily verified that T is a well-defined U-module intertwining operator. Moreover the cyclicity of the vector $v^\mu\in V$ implies that T is onto. Thus we have a U-module isomorphism $V\cong V(\nu^*)\otimes V(\mu+\nu)/\ker T$. Since $V(\nu^*)\otimes V(\mu+\nu)$ is completely reducible V must occur as a submodule of $V(\nu^*)\otimes V(\mu+\nu)$. This proves the result.

Q.E.D.

The above result shows that we may naturally imbed any finite-dimensional cyclic module $V=Uv^\mu$ in the tensor product of two irreducible modules. In particular if we know the degree $\mathbf{m}$ of the cyclic vector v^μ one may gain useful information with regard to the possible irreducible modules which may occur in the space V. This then enables projection operators, which project from a weight vector v^μ of weight $\mu\in\Lambda^+$ a maximal weight vector of weight $\mu\in\Lambda^+$, to be constructed as follows (notation as in Theorem 4).

Lemma 8: Suppose $\mu\in\Lambda^+$ and v^μ has degree $\mathbf{m}=(m_1,\ldots,m_l)$. Set $\nu=\Sigma_{i=1}^l m_i\Lambda_i$ and put $\Lambda(\nu)=\{\mu+\nu-\lambda\,|\,\lambda\in\Pi(\nu)\}$, where $\Pi(\nu)$ denotes the set of distinct weights in $V(\nu)$. Finally set $\Gamma(\nu)=\Lambda(\nu)\cap\Lambda^+\sim\{\mu\}$ and

$$P=\prod_{\chi\in\Gamma(\nu)}\left(\frac{C_L-\chi_\lambda(C_L)}{\chi_\mu(C_L)-\chi_\lambda(C_L)}\right). \tag{23}$$

Then $v_+^\mu=Pv^\mu$ is a maximal weight vector of weight $\mu\in\Lambda^+$. The module $V(\mu)$ occurs in Uv^μ if and only if $Pv^\mu\neq 0$.

Proof: We note first that the weights occurring in the

irreducible module $V(\nu^*)$ are the negative of those occurring in the module $V(\nu)$ (see, e.g., Ref. 13). Thus the set of weights $\Lambda(\nu)$, as defined in the statement of the lemma, may be alternatively written

$$\Lambda(\nu) = \{\mu + \nu + \lambda \,|\lambda \in \Pi(\nu^*)\}, \tag{24}$$

where $\Pi(\nu^*)$ denotes the set of distinct weights in $V(\nu^*)$.

As shown in Theorem (4) the cyclic module $V = Uv^\mu$ occurs as a submodule of $V(\nu^*) \otimes V(\mu + \nu)$. In view of Eq. (24) and the argument used in Lemma 3 it follows that the projection operator of Eq. (23) projects the tensor product module $V(\nu^*) \otimes V(\mu + \nu)$ onto the irreducible module $V(\mu)$ (which, from Theorem 1, occurs with at most unit multiplicity). Thus P projects $V = Uv^\mu$ onto the irreducible submodule $V(\mu)$, whence $v^\mu_+ = Pv^\mu$ is necessarily a maximal weight state of weight $\mu \in \Lambda^+$. Clearly $V(\mu)$ occurs in V if and only if $Pv^\mu \neq 0$. This proves the result.

□

The above result illustrates that the required projection operators may be constructed solely from a knowledge of the weight μ and degree $\mathbf{m}$ of the vector v^μ and the weight spectrum of the irreducible module $V(\nu)$, where $\nu = \mathbf{m}\cdot\mathbf{\Lambda} = \Sigma_{i=1}^{l} m_i \Lambda_i$. For this latter problem we just need all $\lambda \in \Lambda^+$ such that $\lambda < \nu$ together with their Weyl-group conjugates (cf. Humphreys[13]). We need consider only those weights ρ in $V(\nu)$ such that $\rho \neq \nu$ and $\mu + \nu - \rho \in \Lambda^+$.

We remark that the integers m_i in Theorem 4 and Lemma 8 may be replaced by any set of integers $\{m_i'\}$ such that

$$x_i^{m_i'+1} v^\mu = 0.$$

However it is clear that the set of integers which yields the degree $\mathbf{m}$ of the vector v^μ will afford the most economical construction of the desired projection operators. This is because the greater the m_i' the greater the dimension of the space $V(\nu)$, $\nu = \Sigma_{i=1}^{l} m_i' \Lambda_i$, and hence the greater the number of factors occurring in the expression for the projection operator P.

It is interesting to note that in order for the projection operator P of Eq. (23) to project any given weight vector of weight $\mu \in \Lambda^+$ and degree $\mathbf{m}$ onto a maximal weight vector of weight μ it is necessary to keep all the factors occurring in the product of Eq. (23). This is because the cyclic vector $v^{\nu^*}_- \otimes v^{\mu+\nu}_+$ of the tensor product module $V(\nu^*) \otimes V(\mu + \nu)$ (notation as in Theorem 4) has degree $\mathbf{m}$ and weight μ and all factors in the expression for P are required to project $v^{\nu^*}_- \otimes v^{\mu+\nu}_+$ onto a maximal weight vector of weight μ. Thus, in this sense, the projection operators of Eq. (23) are the simplest possible choice. In any specific application however there may exist special properties of the vector v^μ which allows certain factors in the expression for P to be removed.

VI. CONCLUSIONS

The results of Sec. V (in particular Theorem 4) demonstrate that any finite-dimensional cyclic module may be naturally imbedded in the tensor product of two irreducible modules. This result therefore uniquely characterizes cyclic modules as direct summands of the tensor product of two irreducible modules. Since the multiplicity problem may be solved for the tensor product of two irreducible modules (see Ref. 6) this implies that the multiplicity problem may also be solved, in principle, for any finite-dimensional cyclic module.

The general state labeling problem may be stated as follows: let G be a semisimple Lie group and $H \subset G$ a semisimple subgroup. The finite-dimensional irreducible representations of the group G constitute completely reducible representations of the subgroup H and the irreducible representations of H usually occur with multiplicities. The problem of distinguishing the equivalent irreducible representations of H occurring in an irreducible representation of G is called the $H \subset G$ state labeling problem.

Our methods apply to irreducible representations of the group G which constitute cyclic representations of the subgroup H. This situation in fact occurs for several state labeling problems of physical interest, examples being given by the subgroup imbeddings $\mathrm{U}(n) \supset \mathrm{O}(n)$, $U(n) \supset \mathrm{Sp}(n)$, $G \times G \supset G$ (G semisimple), etc. (see Refs. 6 and 7). It is interesting to note, in connection with the subgroup imbeddings $\mathrm{U}(n) \supset \mathrm{O}(n)$ and $\mathrm{U}(n) \supset \mathrm{Sp}(n)$, that if the imbeddings are chosen correctly the cyclic vector under the subgroups $\mathrm{O}(n)$ or $\mathrm{Sp}(n)$ may be taken to be the $\mathrm{U}(n)$ maximal weight vector (cf. Ref. 6).

The results of Sec. IV have important consequences with regard to the Clebsch–Gordan multiplicity problem (i.e., the $G \times G \supset G$ state labeling problem). The explicit determination of the annihilator $I(\lambda, \mu) \subseteq U$ of the cyclic vector $v^\lambda_- \otimes v^\mu_+$ of $V(\lambda) \otimes V(\mu)$ yields a great deal of information concerning the tensor product module $V(\lambda) \otimes V(\mu)$. The results of Ref. 6 may in fact be attributed to the simple structure of the left ideal $I(\lambda, \mu)$. It is hoped that further information may now be obtained. As mentioned above, these methods apply not only to the Clebsch–Gordan problem but are likely to be applicable to other state labeling problems of physical interest.

Further work along these lines is now in progress.

[1] J. P. Elliott, Proc. R. Soc. London Ser. A **245**, 128, 562 (1958).
[2] J. P. Elliott and M. Harvey, Proc. R. Soc. London Ser. A **272**, 557 (1962).
[3] R. M. Asherova, Yu. F. Smirnov, and V. N. Tolstoy, Teor. Mat. Fiz. **8**, 255 (1971).
[4] P. O. Löwdin, Rev. Mod. Phys. **36**, 966 (1964).
[5] J. Shapiro, J. Math. Phys. **6**, 1680 (1965).
[6] S. A. Edwards and M. D. Gould, "Cyclic vectors, projection methods, and the multiplicity problem," J. Math. Phys. (to appear).
[7] M. D. Gould and E. G. Kalnins, "A projection based solution to the SP(2n) state labeling problem," J. Math. Phys. (to appear).
[8] M. Moshinsky, J. Patera, R. T. Sharp, and P. Winternitz, Ann. Phys. (N.Y.) **95**, 139 (1975).
[9] N. MacDonald, Adv. Phys. **19**, 371 (1970).
[10] K. R. Parthasarathy, R. Ranga Rao, and V. S. Varadarajan, Ann. Math. **85**, 383 (1967).
[11] D. N. Verma, Bull. Am. Math. Soc. **74**, 160 (1968); see also Yale dissertation (1966).
[12] I. N. Bernstein, I. M. Gel'fand, and S. I. Gel'fand, J. Funct. Anal. **5**, 1 (1971).
[13] J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1972).
[14] J. Dixmier, *Enveloping Algebras* (North-Holland, Amsterdam, 1977).
[15] R. M. Asherova, Yu. F. Smirnov, and V. N. Tolstoy, Teor. Mat. Fiz. **15**, 107 (1973).
[16] M. D. Gould, J. Math. Phys. **21**, 444 (1980); J. Math. Phys. **22**, 267 (1981).
[17] C. W. Curtis and I. Reiner, *Representation Theory of Finite Groups and Associative Algebras* (Wiley, New York, 1962).

Poisson bracket realizations of Lie algebras and subrepresentations of $(\mathrm{ad}^{\otimes k})_s$

Mircea Iosifescu and Horia Scutaru
Department of Theoretical Physics, Central Institute of Physics, Bucharest, Romania

(Received 10 March 1983; accepted for publication 16 February 1984)

A procedure which associates Poisson bracket realizations of a Lie algebra L to subrepresentations of the extension $(\mathrm{ad}^{\otimes k})_s$ of the adjoint action to the algebra of polynomials defined on the dual space L^* is pointed out. The procedure is applied, for $k = 2$, to the real forms of the semisimple Lie algebras of types D_3 and $B_2 \sim C_2$, in particular to the algebras so(4,2), so(4,1), and so(3,2)$\sim$sp(4,R). The results obtained for the algebra sp(4,R) have led to an algebraic foundation for the constraints satisfied by the dynamical variables for the classical limit of the generalized helium problem.

PACS numbers: 02.20.Qs, 02.20.Sv

I. INTRODUCTION

Let M be a symplectic manifold and let

$$\{f,g\} = \omega(X_f, X_g) = X_f g \tag{1.1}$$

be the Poisson bracket associated with the symplectic form ω on M and defined for two functions $f, g \in C^\infty(M)$; X_f is the Hamiltonian field associated to f and defined by $df = X_f \lrcorner \omega$. The product (1.1) organizes $C^\infty(M)$ as a Lie algebra.[1]

A *Poisson bracket* (PB) *realization*[2] of a Lie algebra L is, by definition, a Lie algebra homomorphism $\rho: L \to C^\infty(M)$.

PB realizations have received attention in the last fifteen years. The approaches to this subject were either "heuristical," i.e., constructions of specific PB realizations for various Lie algebras or "abstract"—e.g., proof of the existence on each coadjoint orbit of a Lie group of a PB realization generated by coordinate functions of the points belonging to the orbit.[3]

In several "heuristical" approaches,[4–7] the presence of a number of polynomial identities satisfied by the generators of the PB realization have been pointed out. References 4 and 5 consider PB realizations of the algebras so(4,1) and so(4,2) related to the Kepler problem; in Ref. 7 the presence of a set of polynomial relations between the generators of the so($n+2$) algebra is proved to be responsible for the possibility to define a set of n canonically conjugated pairs of coordinates and momenta as functions of the generators of this algebra.

In a previous work[8] concerned with PB realizations of the so(4,2) algebra, we have pointed out that the polynomial relations satisfied by the generators of the PB realization of so(4,2), obtained by Györgyi in Ref. 5, define a coordinate-independent PB realization of so(4,2), which contains, as particular cases, a number of "heuristic" PB realizations of this algebra. The polynomial relations satisfied by the generators characterize thus a whole class of PB realizations of so(4,2), which are related by a symplectomorphism.

We consider that the most interesting remark of Ref. 8 is that Györgyi's relations define at the same time a *linear representation* of so(4,2) on the space of polynomials (of degree 2) on the dual space so(4,2)* of so(4,2) and a *coadjoint orbit* of the corresponding Lie group SO(4,2).

The aim of the present paper is to generalize this remark and to give a procedure by which polynomial relations defining coadjoint orbits (and defining thus also the corresponding PB realizations of the Lie algebra) can be obtained from the symmetric part $(\mathrm{ad}^{\otimes k})_s = (\mathrm{ad} \otimes \mathrm{ad} \otimes \cdots \otimes \mathrm{ad})_s$ (k factors) of the direct power of the adjoint representation of the Lie algebra (Sec. II).[9–11] This procedure is applied subsequently (Sec. III) to all real forms of the semisimple Lie algebras of types $D_3 \sim A_3$ and $B_2 \sim C_2$ and to their orbits defined by the zeroes of second-degree homogeneous polynomials.

Particular emphasis is given to the algebra so(4,2). For this Lie algebra, the explicit equations of the coadjoint orbits associated to subrepresentations of $(\mathrm{ad} \otimes \mathrm{ad})_s$ have been deduced. Apart from the six-dimensional coadjoint orbit of SO(4,2), which has been studied in Ref. 8, two eight-dimensional families of orbits of SO(4,2) have been obtained (Sec. III). For one of these families, the explicit expression of the Kirillov–Kostant–Souriau (KKS) symplectic form has been deduced.

Special attention has also been paid to the algebras so(4,1) and so(3,2)$\sim$sp(4,R), in order to stress the fact that PB realizations of different real forms of the same complex semisimple Lie algebra are sometimes very different.

The polynomial identities satisfied by the six-dimensional PB realization of the sp(4,R) algebra are particularly interesting; indeed, these identities have recently been obtained as a classical limit for the quantum constraints in a quasiclassic treatment of the helium atom, in which the large quantum number is the dimension of the space.[12] The identities which appear in the classical limit, and which have been obtained in Ref. 12 using a heuristic procedure, are obtained in the present work from Lie algebraic considerations.

II. POISSON BRACKET REALIZATIONS OF A LIE ALGEBRA ASSOCIATED TO SUBREPRESENTATIONS OF THE PRODUCT $(\mathrm{ad}^{\otimes k})_s$

Let L be a Lie algebra and L^* be the dual space of L. Let

$$[x_i, x_j] = \sum_{k=1}^{n} c_{ij}^k x_k \quad (i,j = 1,\ldots,n = \dim L) \tag{2.1}$$

be the Lie product of the generators x_i, x_j of L. Let us introduce in L^* a basis $\{u_1,\ldots,u_n\}$, dual with respect to the basis

 0022-2488/84/102856-07$02.50

$\{x_1,...,x_n\}$ of L, i.e., defined by the condition

$$\langle u_i|x_j\rangle = \delta_{ij} \quad (i,j=1,...,n)\,. \tag{2.2}$$

An element $u\in L^*$ has the expression

$$u = \sum_{i=1}^{n} \xi_i(u)u_i\,, \tag{2.3}$$

where $\xi_i(u)$ are coordinate functions of the point u:

$$\xi_i(u) = \langle u|x_i\rangle\,. \tag{2.4}$$

The expression of the generator of the coadjoint action $\mathrm{Ad}^*(\exp tx_i)_u : L^* \to L^*$ is

$$X_{x_i}(u) = \sum c_{ij}^k \xi_k(u)\left(\frac{\partial}{\partial \xi_j}\right)_u\,. \tag{2.5}$$

It is interesting to observe that the realization $x_i \to X_{x_i}$ of L is an example of a quantum realization[13,14] or of a Jordan map[15,16] corresponding to the adjoint representation of L.

The generators X_{x_i}, $i=1,...,n$ are linearly dependent for any $u\in L^*$. Indeed, the rank of the $n\times n$ matrix

$$M_L(u) = \left|\left|\sum c_{ij}^k \xi_k(u)\right|\right| \quad (i,j=1,...,n) \tag{2.6}$$

is less than n for any $u\in L^*$. The generators X_{x_i} $(i=1,...,n)$ of the coadjoint action generate, in each point $u\in L^*$, a linear representation of the Lie algebra L, as

$$[X_{x_i},X_{x_j}] = X_{x_i}X_{x_j} - X_{x_j}X_{x_i} = \sum c_{ij}^k X_{x_k}\,. \tag{2.7}$$

The orbits of the coadjoint action of a Lie group have been organized as symplectic manifolds by Kirillov, Kostant, and Souriau,[17] the KKS symplectic form being defined, in each point $u\in\mathcal{O}_r$, by its value on pairs of vectors $X_{x_i},X_{x_j}\in T_u\mathcal{O}_r$,

$$\omega(X_{x_i},X_{x_j}) = -\langle u|[x_i,x_j]\rangle\,. \tag{2.8}$$

The symplectic form ω (2.8) associates to any function $f\in C^\infty(\mathcal{O}_r)$ a Hamiltonian field X_f, defined by $df = X_f \lrcorner \omega$. The Poisson bracket of two functions $f,g\in C^\infty(\mathcal{O}_r)$ is defined by (1.1); it has the following expression in terms of the coordinate functions $\xi_i(u)$ of a point $u\in L^*$:

$$\{f,g\} = \sum_{i,j,k} c_{ij}^k \xi_k \frac{\partial f}{\partial \xi_i}\frac{\partial g}{\partial \xi_j}\,. \tag{2.9}$$

In particular, if f and g are coordinate functions of a point $u\in\mathcal{O}_r$, we obtain

$$\{\xi_i,\xi_j\} = \sum_{k=1}^{n} c_{ij}^k \xi_k(u)\,, \tag{2.10}$$

i.e., the coordinate functions $\xi_i(u)$ of the point $u\in\mathcal{O}_r$ generate a PB realization of the Lie algebra L.

The extension $\widetilde{\mathrm{ad}}$ of the adjoint representation ad of L to a representation of L on the linear space $P(L^*)$ of all polynomials on L is given by

$$\widetilde{\mathrm{ad}}(x_i)p = X_{x_i}p = \sum_{j,k=1}^{n} c_{ij}^k \xi_k \frac{\partial p}{\partial \xi_j} = \{\xi_i,p\} \quad (p\in P(L^*))\,. \tag{2.11}$$

This representation is equivalent to the unique extension of the adjoint representation to the symmetric algebra $S(L)$ of L.[18] Indeed, it is well known that $S(L)$ and $P(L^*)$ are isomorphic as linear spaces and it is easy to see that this isomorphism intertwines the two representations of L defined above. From this remark or from the properties of the Jordan map $x_i \to X_{x_i}$ it follows that there exists a natural number k for each irreducible subrepresentation of ad such that it is equivalent with a subrepresentationof $(\mathrm{ad}^{\otimes k})_s$; in other words, boson spaces are symmetric.

Let π be a subrepresentation of $\widetilde{\mathrm{ad}}$. Let $\{p_1,...,p_m\}$ be a basis of a subspace invariant under π of $P(L^*)$, i.e., such that

$$(X_{x_i}p_j)(u) = (\pi(x_i)p_j)(u) = \sum_{k=1}^{m} \pi_{jk}(x_i)p_k(u)\,. \tag{2.12}$$

We define, for any such subrepresentation π, an algebraic submanifold of L^* by

$$\mathcal{O}_\pi \equiv \{u\in L^* | p_1(u) = \cdots = p_m(u) = 0\}\,. \tag{2.13}$$

From Whitney's theorem,[19] it follows that the regular part of $\mathcal{O}_\pi$, denoted $\mathcal{O}_\pi^\circ$, is a regular analytic submanifold of L^*, the dimension of which is equal to $\dim L^* - s$, where s is the maximum value attained on $\mathcal{O}_\pi$ by the dimension of the linear space generated by $(dp_1)_u,...,(dp_m)_u$, $u\in\mathcal{O}_\pi$. From the definition of $\mathcal{O}_\pi$ and from the property (2.12) it follows that the vectors $X_{x_1}(u),...,X_{x_n}(u)$ generate a linear subspace of the tangent space $T_u\mathcal{O}_\pi^\circ$ for any $u\in\mathcal{O}_\pi^\circ$. The dimension of this subspace is given by $r(u) = \mathrm{rank}||\Sigma\, c_{ij}^k\xi_k(u)||$. It is obvious that $\dim T_u\mathcal{O}_\pi^\circ = \dim L^* - s \geqslant r(u)$. Thus, in order to have $r(u) = \dim L^* - s$, it is sufficient to prove that $r(u)\geqslant \dim L^* - s$, for any $u\in\mathcal{O}_\pi^\circ$.

Hence, if it is possible to prove that on $\mathcal{O}_\pi^\circ$ we have

$$\mathrm{rank}\left|\left|\sum c_{ij}^k\xi_k\right|\right| \geqslant \dim \mathcal{O}_\pi^\circ\,, \tag{2.14}$$

then it results that $T_u\mathcal{O}_\pi^\circ$ is generated by the vectors $X_{x_1}(u),...,X_{x_n}(u)$ and that it is possible thus to define on $\mathcal{O}_\pi^\circ$ a KKS symplectic structure ω in a unique way by (2.8) and hence to define a Poisson bracket (2.9) associated to ω, for any $f,g\in C^\infty(\mathcal{O}_\pi^\circ)$. A preliminary analysis of the nondegeneracy of the manifold $\mathcal{O}_\pi$ defined by (2.13), i.e., a proof that

$$\dim \mathcal{O}_\pi > 0 \tag{2.15}$$

is required. The two conditions (2.14) and (2.15) are sufficient for the existence of a PB realization of the Lie algebra L associated to the linear representation π of this algebra.

Remark: If the representation (2.12) contains a one-dimensional subrepresentation of L, the basis of which is thus an invariant $I(u)$ of the algebra L, then the relation which corresponds to this invariant in the definition (2.13) of the submanifold $\mathcal{O}_\pi$ is $I(u) = k$, where k is an arbitrary, fixed real number.

III. INVARIANT SUBMANIFOLDS ASSOCIATED TO SUBREPRESENTATIONS OF $(\mathrm{ad}\otimes\mathrm{ad})_s$ FOR SEMISIMPLE LIE ALGEBRAS OF TYPES $D_3\sim A_3$ AND $B_2\sim C_2$

The construction described in Sec. II will be applied in this section to the semisimple Lie algebras of types $D_3\sim A_3$ and $B_2\sim C_2$. For these algebras, Ad-invariant submanifolds associated to subrepresentations of $(\mathrm{ad}\otimes\mathrm{ad})_s$ will be determined. The D_3-type algebras will be exemplified by so(4,2), and B_2 algebras will be exemplified by two real forms so(4,1) and so(3,2), in order to stress the fact that PB realizations of

TABLE I. Lie multiplication table of the algebra so(4, 2). The multiplication law is $_X{}^Y{}_{[X,Y]}$.

	L_1	L_2	L_3	A_1	A_2	A_3	B_1	B_2	B_3	B_4	C_1	C_2	C_3	C_4	M
L_1	0	L_3	$-L_2$	0	A_3	$-A_2$	0	B_3	$-B_2$	0	0	C_3	$-C_2$	0	0
L_2	$-L_3$	0	L_1	$-A_3$	0	A_1	$-B_3$	0	B_1	0	$-C_3$	0	C_1	0	0
L_3	L_2	$-L_1$	0	A_2	$-A_1$	0	B_2	$-B_1$	0	0	C_2	$-C_1$	0	0	0
A_1	0	A_3	$-A_2$	0	L_3	$-L_2$	B_4	0	0	$-B_1$	C_4	0	0	$-C_1$	0
A_2	$-A_3$	0	A_1	$-L_3$	0	L_1	0	B_4	0	$-B_2$	0	C_4	0	$-C_2$	0
A_3	A_2	$-A_1$	0	L_2	$-L_1$	0	0	0	B_4	$-B_3$	0	0	C_4	$-C_3$	0
B_1	0	B_3	$-B_2$	$-B_4$	0	0	0	$-L_3$	L_2	$-A_1$	$-M$	0	0	0	$-C_1$
B_2	$-B_3$	0	B_1	0	$-B_4$	0	L_3	0	$-L_1$	$-A_2$	0	$-M$	0	0	$-C_2$
B_3	B_2	$-B_1$	0	0	0	$-B_4$	$-L_2$	L_1	0	$-A_3$	0	0	$-M$	0	$-C_3$
B_4	0	0	0	B_1	B_2	B_3	A_1	A_2	A_3	0	0	0	0	$-M$	$-C_4$
C_1	0	C_3	$-C_2$	$-C_4$	0	0	M	0	0	0	0	$-L_3$	L_2	$-A_1$	B_1
C_2	$-C_3$	0	C_1	0	$-C_4$	0	0	M	0	0	L_3	0	$-L_1$	$-A_2$	B_2
C_3	C_2	$-C_1$	0	0	0	$-C_4$	0	0	M	0	$-L_2$	L_1	0	$-A_3$	B_3
C_4	0	0	0	C_1	C_2	C_3	0	0	0	M	A_1	A_2	A_3	0	B_4
M	0	0	0	0	0	0	C_1	C_2	C_3	C_4	$-B_1$	$-B_2$	$-B_3$	$-B_4$	0

different real forms of semisimple Lie algebras of the same type may be different. Denoting representations by Dynkin indices (upper lines) and by dimensions (lower lines) the Clebsch–Gordan series for $(\text{ad}\otimes\text{ad})_s$ are for D_3-type algebras,

$$\begin{array}{rrcccccccc} & ((1,0,1)\otimes & (1,0,1))_s & = & (2,0,2) & \oplus & (0,2,0) & \oplus & (1,0,1) \oplus & (0,0,0), \\ 125 = & (15\otimes & 15)_s & = & 84 & \oplus & 20 & \oplus & 15 \oplus & 1; \end{array} \tag{3.1}$$

and for B_2-type algebras,

$$\begin{array}{rrcccccccc} & ((0,2)\otimes & (0,2))_s & = & (0,4) & \oplus & (2,0) & \oplus & (0,1) \oplus & (0,0), \\ 55 = & (10\otimes & 10)_s & = & 35 & \oplus & 14 & \oplus & 5 \oplus & 1. \end{array} \tag{3.2}$$

Using the Clebsch–Gordan coefficients for so(6) (Ref. 20) and so(5),[21] the bases for the subrepresentations in (3.1) and (3.2) have been determined. The generators for all real forms of types D_3 and B_2 have been determined from these bases, using Cartan involutions and the Weyl unitary trick. In particular, denoting the generators of so(6) by M_{ij} ($M_{ij} = -M_{ji}$), with Lie products

$$[M_{ij},M_{kl}] = g_{ik}M_{jl} + g_{jl}M_{ik} - g_{il}M_{jk} - g_{jk}M_{il}\,, \tag{3.3}$$

with $g_{ij} = \delta_{ij}$, the transformations so(6)→so(4,2) are

$$\begin{aligned} &M_{jk}\to M_{jk} \quad (j,k = 1,2,3,4)\,, \quad M_{56}\to M_{56}\,, \\ &M_{j5}\to iM_{j5}\,, \quad M_{j6}\to iM_{j6} \quad (j = 1,2,3,4)\,, \end{aligned} \tag{3.4}$$

The transformations so(5)→so(4,1) are

$$\begin{aligned} &M_{jk}\to M_{jk} \quad (j,k = 1,2,3)\,, \\ &M_{j5}\to iM_{j5} \quad (j = 1,2,3)\,, \quad M_{45}\to iM_{45}\,. \end{aligned} \tag{3.5}$$

In order to obtain a basis for the so(4,2) algebra, such that the so(4,2) Lie product has the expression (3.3) with

$$\begin{aligned} &g_{11} = g_{22} = g_{33} = g_{44} = -g_{55} = -g_{66} = 1\,, \\ &g_{ij} = 0 \quad \text{if } i\neq j\,, \end{aligned} \tag{3.6}$$

we have to perform the supplementary transformation $M_{56}\to -M_{56}$. The following vectorial notations have been used for the generators of the so(4,2) and so(4,1) algebras:

$$\begin{aligned} &\mathbf{L} = (M_{23},M_{31},M_{12})\,, \quad \mathbf{A} = (M_{14},M_{24},M_{34})\,, \\ &\mathbf{B} = (M_{15},M_{25},M_{35})\,, \quad B_4 = M_{45}\,, \\ &\mathbf{C} = (M_{16},M_{26},M_{36})\,, \quad C_4 = M_{46}\,, \quad M = -M_{56}\,. \end{aligned} \tag{3.7}$$

In terms of these notations, the bases for the representations (0,2,0) and (1,0,1) of so(4,2) are those given in Tables II and III. For all semisimple Lie algebras we may identify the dual space L^* with the algebra L using the nondegeneracy of the Killing–Cartan bilinear form on L.

Let us apply now to the subrepresentations of $(\text{ad}\otimes\text{ad})_s$ in (3.1) and (3.2) the results of the previous section.

TABLE II. Basis vectors of the irreducible subrepresentation (1, 0, 1) of $(\text{ad}\times\text{ad})_s$ for the algebra so(4, 2). They are obtained by the application of Weyl's unitary trick to the basis vectors of the corresponding representation (1, 0, 1) of the algebra so(6). The generators of so(6) can be characterized as spherical tensors with respect to two commuting so(3) subalgebras; hence the labeling $(SM_S)(TM_T)$ used in Ref. 20 and adopted also in our tables. Table II uses notations (3.19).

S, M_S	T, M_T	Basis vector $f^{(1,0,1)}_{(S,M_S)(T,M_T)}$
1, −1	0, 0	$-(1/\sqrt{3})(\mathscr{L}_2 + i\mathscr{L}_1)$
1, 0	0, 0	$\sqrt{2/3}\,\mathscr{L}_3$
1, 1	0, 0	$(1/\sqrt{3})(\mathscr{L}_2 - i\mathscr{L}_1)$
0, 0	1, −1	$(1/\sqrt{3})(\mathscr{C}_4 + i\mathscr{B}_4)$
0, 0	1, 0	$\sqrt{(2/3)}\,\mathscr{M}$
0, 0	1, 1	$(1/\sqrt{3})(\mathscr{C}_4 - i\mathscr{B}_4)$
1, −1	1, −1	$-(1/\sqrt{6})[\mathscr{C}_1 + \mathscr{B}_2 - i(\mathscr{C}_2 - \mathscr{B}_1)]$
1, 0	1, −1	$(1/\sqrt{3})(\mathscr{B}_3 - i\mathscr{C}_3)$
1, 1	1, −1	$-(1/\sqrt{6})[\mathscr{C}_1 - \mathscr{B}_2 + i(\mathscr{C}_2 + \mathscr{B}_1)]$
1, −1	1, 0	$(1/\sqrt{3})(\mathscr{A}_2 + i\mathscr{A}_1)$
1, 0	1, 0	$-\sqrt{2/3}\,\mathscr{A}_3$
1, 1	1, 0	$-(1/\sqrt{3})(\mathscr{A}_2 - i\mathscr{A}_1)$
1, −1	1, 1	$(1/\sqrt{6})[\mathscr{C}_1 - \mathscr{B}_2 - i(\mathscr{C}_2 + \mathscr{B}_1)]$
1, 0	1, 1	$(1/\sqrt{3})(\mathscr{B}_3 + i\mathscr{C}_3)$
1, 1	1, 1	$(1/\sqrt{6})[\mathscr{C}_1 + \mathscr{B}_1 + i(\mathscr{C}_2 - \mathscr{B}_1)]$

A. The algebra so(4,2)

We first select the subrepresentations in (3.1) which satisfy condition (2.15). The following statements are true (for proofs cf. Ref. 10).

(i) Any representation of the algebra so(4,2), contained in the representation $(\mathrm{ad}\otimes\mathrm{ad})_s$ of this algebra and containing, as a subrepresentation, the representation (2,0,2) does not satisfy condition (2.15), i.e., leads to a submanifold (2.13) of dimension zero.

(ii) If the basis vectors of the representation $(1,0,1)\oplus(0,2,0)$ vanish, then the quadratic invariant of so(4,2)

$$I_2 = \mathbf{L}^2 + \mathbf{A}^2 + M^2 - \mathbf{B}^2 - B_4^2 - \mathbf{C}^2 - C_4^2 \tag{3.8}$$

vanishes.

These two statements lead to the following proposition.

Proposition 3.1: The only subrepresentations of the representation $(\mathrm{ad}\otimes\mathrm{ad})_s$ of the algebra so(4,2) to which it is possible to associate PB realizations of this algebra are

$$(1,0,1)\oplus(0,2,0)\,;\quad (1,0,1)\oplus(0,0,0)\,;$$
$$(0,2,0)\oplus(0,0,0)\,. \tag{3.9}$$

Let us now examine, one by one, the invariant submanifolds of so(4,2) which can be associated to these three representations of so(4,2).

1. Representation (1,0,1) ⊕ (0,2,0)

The properties of the submanifold (2.13) of so(4,2) associated to this representation have been analyzed in Ref. 8 (cf. also Ref. 10). This submanifold, denoted $\mathcal{O}_6$, is six dimensional. Condition (2.14) is satisfied on $\mathcal{O}_6$ (see Ref. 8). The KKS symplectic form ω has, on $\mathcal{O}_6$ and only on $\mathcal{O}_6$, the following expression[8]:

$$\omega_6 = \frac{1}{M}\sum_{i=1}^{4} dB_i\wedge dC_i - \sum_{i=1}^{4}\frac{B_i}{M^2}\,dM\wedge dC_i\,. \tag{3.10}$$

2. Representation (1,0,1) ⊕ (0,0,0)

Proposition 3.2: The manifold determined by (i) the equalities obtained equating to zero the basis vectors of representation (1,0,1) of so(4,2) and (ii) the equality [associated to representation (0,0,0) of so(4,2)]

$$I_2 = k^2 \tag{3.11}$$

is, for any $k\neq 0$, an eight-dimensional submanifold of the dual space so(4,2).

Proof: It has been proved in Ref. 8 that the six relations

$$\mathbf{C}\times\mathbf{B} - M\mathbf{L} = 0\,,\quad B_4\mathbf{C} - C_4\mathbf{B} - M\mathbf{A} = 0 \tag{3.12}$$

obtained by equating to zero six basis vectors of the representation (1,0,1) of so(4,2) (cf. Table II) lead to the vanishing of the other basis vectors of this representation. These six relations are functionally independent; they define, together with (3.11), for any fixed k, an eight-dimensional submanifold of the 15-dimensional linear space so(4,2).

Let us denote by $\mathcal{O}_8^k$ the manifold defined by Proposition 3.2. Denoting with $M_{\mathrm{so}(4,2)}(u)$ the matrix $M_L(u)$ (2.6) in which the c_{ij}^k are structure constants of the algebra so(4,2), we have the following proposition.

Proposition 3.3: rank $M_{\mathrm{so}(4,2)}(u)\geqslant 8$ for any $u\in\mathcal{O}_8^k$.

The proof, which is straightforward but tedious, will not be given here. (cf. Ref. 10). It consists in showing that, if a number of conveniently chosen eight-dimensional submatrices of $M_{\mathrm{so}(4,2)}(u)$ along its principal diagonal have vanishing determinants and if relations (3.12) are satisfied, then either $u=0$ or $u\in\mathcal{O}_6$.

Thus, properties (2.14) and (2.15) are verified for the subrepresentation $(1,0,1)\oplus(0,0,0)$ and hence, for any $k\neq 0$, the KKS symplectic form is nondegenerate on the manifold $\mathcal{O}_8^k$ associated to this representation.

Let us determine the explicit expression of this symplectic form. To do that, a result of Mukunda is of help. Mukunda's result,[7] adapted to noncompact algebras, in particular to the algebra so(4,2), is as follows: "If the generators M_{ij} of the algebra so(4,2) satisfy the set of identities

$$M_{ij}M_{kl} + M_{ik}M_{lj} + M_{il}M_{jk} = 0\,, \tag{3.13}$$

then there exist four pairs of canonically conjugated variables which are defined as functions of the generators M_{ij} by the expressions

$$q_r = (C^2 - M^2)^{1/2}B_r/M\,,\quad p_r = (C^2 - M^2)^{-1/2}C_r\,, \tag{3.14}$$

where $C^2 = \mathbf{C}^2 + C_4^2$."

The relations (3.13) reduce to the relations obtained by equating to zero the basis vectors of the representation (1,0,1). Hence, as the conditions (2.14) and (2.15) are satisfied for representation $(1,0,1)\oplus(0,0,0)$ the coordinate functions of a point of $\mathcal{O}_8^k$ generate a PB realization of the algebra so(4,2) with respect to the Poisson bracket associated to the KKS symplectic form on $\mathcal{O}_8^k$. It is possible thus to define on $\mathcal{O}_8^k$ a set of canonically conjugated coordinates q_i, p_i $(i=1,2,3,4)$; with respect to these coordinates, the KKS symplectic form on $\mathcal{O}_8^k$ takes the Darboux canonical expression

$$\omega_8 = \sum_{i=1}^{4} dq_i\wedge dp_i\,. \tag{3.15}$$

Replacing q_i and p_i by their expressions (3.14) we get the following proposition.

Proposition 3.4: (i) The expression of the Kirillov–Kostant–Souriau symplectic form on $\mathcal{O}_8^k$ is

$$\omega_8 = \frac{1}{M}\sum_{j=1}^{4} dB_j\wedge dC_j - \frac{1}{M^2}\sum_{j=1}^{4} B_j\,dM\wedge dC_j + \frac{1}{2}\frac{d(C^2-M^2)}{C^2-M^2}\wedge\frac{d(BC)}{M}\,, \tag{3.16}$$

with $(BC) = \mathbf{B}\cdot\mathbf{C} + B_4C_4$.

(ii) Denoting with ω the generic KKS symplectic form, defined by (2.8), we have

$$\omega_u = (\omega_8)_u \quad \text{if and only if} \quad u\in\mathcal{O}_8^k\,.$$

The proof of (ii) is straightforward. It consists in establishing that the expressions $(\omega-\omega_8)(X_{x_i},X_{x_j})$ vanish if and only if the polynomials $\mathcal{L}_i = B_4C_i - C_4B_i - MA_i$ and $\mathcal{A}_i = (\mathbf{C}\times\mathbf{B} - M\mathbf{L})_i$ $(i=1,2,3)$ vanish. The X_{x_i} are the generators (2.5) of the coadjoint action; their expressions for the so(4,2) algebra are given in Ref. 8. The functions $(\omega-\omega_8)(X_{x_i},X_{x_j})$ may be expressed as functions of $\mathcal{L}_i$ and $\mathcal{A}_i$, which are such that if $\mathcal{L}_i = \mathcal{A}_i = 0$ then $(\omega-\omega_8)(X_{x_i},X_{x_j}) = 0$. But, on $\mathcal{O}_8^k$, we have $\mathcal{L}_i = \mathcal{A}_i = 0$ $(i=1,2,3)$ and the direct

assertion follows. The converse statement results from

$$(\omega-\omega_8)(X_{L_i},X_{L_j})=(1/M)\mathscr{L}_k,$$
$$(\omega-\omega_8)(X_{L_i},X_{A_j})=(1/M)\mathscr{A}_k$$

[where (i,j,k) are cyclic permutations of (1,2,3)]. Thus $\mathscr{L}_i=\mathscr{A}_i=0$ $(i=1,2,3)$. To describe an SO(4,2)-invariant eight-dimensional submanifold of so(4,2)*, the points of which satisfy $\mathscr{L}=\mathscr{A}=0$, a supplementary equation is necessary which, in addition has to be invariant under the group action, i.e., an equation obtained by equating to a constant one of the polynomial invariants of the algebra so(4,2). As the cubic and quartic invariants, which have the expressions

$$I_3=\mathbf{L}\cdot\mathscr{L}+\mathbf{A}\cdot\mathscr{A}+M\mathscr{M}-\mathbf{B}\cdot\mathscr{B}-B_4\mathscr{B}_4-\mathbf{C}\cdot\mathscr{C}-C_4\mathscr{C}_4, \tag{3.17}$$

$$I_4=\mathscr{L}^2+\mathscr{A}^2+\mathscr{M}^2-\mathscr{B}^2-\mathscr{B}_4^2-\mathscr{C}^2-\mathscr{C}_4^2, \tag{3.18}$$

where

$$\mathscr{L}\equiv B_4\mathbf{C}-C_4\mathbf{B}-M\mathbf{A},\quad \mathscr{A}\equiv\mathbf{C}\times\mathbf{B}-M\mathbf{L},$$
$$\mathscr{M}\equiv-\mathbf{L}\cdot\mathbf{A},\quad \mathscr{B}\equiv\mathbf{C}\times\mathbf{A}+C_4\mathbf{L},$$
$$\mathscr{B}_4\equiv-\mathbf{L}\cdot\mathbf{C},\quad \mathscr{C}\equiv\mathbf{A}\times\mathbf{B}-B_4\mathbf{L},\quad \mathscr{C}_4=\mathbf{L}\cdot\mathbf{B}, \tag{3.19}$$

vanish if $\mathscr{L}=\mathscr{A}=0$ (cf. the statement beginning the proof of Proposition 3.2), the only equation providing supplementary information is $I_2=k$, which proves the converse statement.

3. Representation (0,2,0) ⊕ (0,0,0)

Proposition 3.5: The SO(4,2)-invariant submanifold of so(4,2)* associated to representation (0,2,0) ⊕ (0,0,0) and defined by the relations obtained equating to zero the basis vectors of representation (0,2,0) and by relation $I_2=k^2$ is, for any $k\neq0$, eight dimensional.

Proof: Equating to zero the basis vectors of representation (0,2,0), given in Table III, we obtain that the invariant I_2 has the expression

$$I_2=3(M^2-B^2)=k^2=3K, \tag{3.20}$$

and we get the following expressions for the generators **L**, **A**, and M:

$$\mathbf{L}=\pm\frac{\sqrt{K}}{B^2}(B_4\mathbf{C}-C_4\mathbf{B})-\frac{M}{B^2}(\mathbf{B}\times\mathbf{C}), \tag{3.21}$$

$$\mathbf{A}=\frac{M}{B^2}(B_4\mathbf{C}-C_4\mathbf{B})\mp\frac{\sqrt{K}}{B^2}(\mathbf{B}\times\mathbf{C}), \tag{3.22}$$

$$M^2=K+B^2 \tag{3.23}$$

(where $B^2=\mathbf{B}^2+B_4^2$), which prove the proposition.

Relations (3.21)–(3.23) can be given the following remarkable form:

$$B_4\mathbf{C}-C_4\mathbf{B}-M\mathbf{A}\pm\sqrt{I_2/3}\,\mathbf{L}=0, \tag{3.24}$$

$$\mathbf{C}\times\mathbf{B}-M\mathbf{L}\pm\sqrt{I_2/3}\,\mathbf{A}=0, \tag{3.25}$$

in which we have remembered that $I_2=3K$. We get similarly

$$\mathbf{A}\times\mathbf{C}-C_4\mathbf{L}\mp\sqrt{I_2/3}\,\mathbf{B}=0, \tag{3.26}$$

$$\mathbf{A}\times\mathbf{B}-B_4\mathbf{L}\pm\sqrt{I_2/3}\,\mathbf{C}=0, \tag{3.27}$$

$$\mathbf{L}\cdot\mathbf{A}\mp\sqrt{I_2/3}\,M=0, \tag{3.28}$$

$$\mathbf{L}\cdot\mathbf{B}\pm\sqrt{I_2/3}\,C_4=0, \tag{3.29}$$

$$\mathbf{L}\cdot\mathbf{C}\mp\sqrt{I_2/3}\,B_4=0. \tag{3.30}$$

Thus, if we abandon the condition to use only representations having as basis vectors polynomials, then the eight-dimensional manifold obtained in this section can be associated to a 15-dimensional representation, the basis vectors of which are the left-hand sides of (3.24)–(3.30).

This relation between the basic polynomials of the 20-

TABLE III. Basis vectors of the subrepresentation (0, 2, 0) of $(\mathrm{ad}\times\mathrm{ad})_s$, for the algebra so(4, 2).

S, M_S	T, M_T	Basis vector $f^{(0,2,0)}_{(S,M_S)(T,M_T)}$
0, 0	0, 0	$(1/\sqrt{6})(-\mathbf{L}^2-B_4^2-C_4^2+M^2)$
2, −2	0, 0	$\frac{1}{4}[(-L_1^2+L_2^2+A_1^2-A_2^2-B_1^2+B_2^2-C_1^2+C_2^2)+2i(L_1L_2-A_1A_2+B_1B_2+C_1C_2)]$
2, −1	0, 0	$\frac{1}{2}[(-L_2L_3+A_2A_3-B_2B_3-C_2C_3)+i(-L_1L_3+A_1A_3-B_1B_3-C_1C_3)]$
2, 0	0, 0	$(1/2\sqrt{6})[-(L_1^2+L_2^2-2L_3^2)+(A_1^2+A_2^2-2A_3^2)-(B_1^2+B_2^2-2B_3^2)-(C_1^2+C_2^2-2C_3^2)]$
2, 1	0, 0	$-\frac{1}{2}[(-L_2L_3+A_2A_3-B_2B_3-C_2C_3)-i(-L_1L_3+A_1A_3-B_1B_3-C_1C_3)]$
2, 2	0, 0	$\frac{1}{4}[(-L_1^2+L_2^2+A_1^2-A_2^2-B_1^2+B_2^2-C_1^2+C_2^2)-2i(L_1L_2-A_1A_2+B_1B_2+C_1C_2)]$
0, 0	2, −2	$-\frac{1}{4}[(\mathbf{B}^2+B_4^2-\mathbf{C}^2-C_4^2)-2i(\mathbf{B}\cdot\mathbf{C}+B_4C_4)]$
0, 0	2, −1	$\frac{1}{2}[(\mathbf{A}\cdot\mathbf{B}+C_4M)-i(\mathbf{A}\cdot\mathbf{C}-B_4M)]$
0, 0	2, 0	$-(1/2\sqrt{6})[\mathbf{B}^2-B_4^2+\mathbf{C}^2-C_4^2-2M^2+2\mathbf{A}^2]$
0, 0	2, 1	$\frac{1}{2}[(\mathbf{A}\cdot\mathbf{B}+C_4M)+i(\mathbf{A}\cdot\mathbf{C}-B_4M)]$
0, 0	2, 2	$-\frac{1}{4}[(\mathbf{B}^2+B_4^2-\mathbf{C}^2-C_4^2)+2i(\mathbf{B}\cdot\mathbf{C}+B_4C_4)]$
1, −1	1, −1	$(1/2\sqrt{2})\{[(\mathbf{B}\times\mathbf{L}+B_4\mathbf{A}-M\mathbf{C})_1+(\mathbf{C}\times\mathbf{L}+C_4\mathbf{A}+M\mathbf{B})_2]+i[(\mathbf{B}\times\mathbf{L}+B_4\mathbf{A}-M\mathbf{C})_2+(\mathbf{C}\times\mathbf{L}+C_4\mathbf{A}+M\mathbf{B})_1]\}$
1, 0	1, −1	$-\frac{1}{2}[(\mathbf{C}\times\mathbf{L}+C_4\mathbf{A}+M\mathbf{B})_3+i(\mathbf{B}\times\mathbf{L}+B_4\mathbf{A}-M\mathbf{C})_3]$
1, 1	1, −1	$-(1/2\sqrt{2})\{[(\mathbf{B}\times\mathbf{L}+B_4\mathbf{A}-M\mathbf{C})_1+(\mathbf{C}\times\mathbf{L}+C_4\mathbf{A}+M\mathbf{B})_2]+i[(\mathbf{B}\times\mathbf{L}+B_4\mathbf{A}-M\mathbf{C})_2-(\mathbf{C}\times\mathbf{L}+C_4\mathbf{A}+M\mathbf{B})_1]\}$
1, −1	1, 0	$\frac{1}{2}[(\mathbf{A}\times\mathbf{L}+B_4\mathbf{B}+C_4\mathbf{C})_1-i(\mathbf{A}\times\mathbf{L}+B_4\mathbf{B}+C_4\mathbf{C})_2]$
1, 0	1, 0	$(i/\sqrt{2})(\mathbf{A}\times\mathbf{L}+B_4\mathbf{B}+C_4\mathbf{C})_3$
1, 1	1, 0	$\frac{1}{2}[(\mathbf{A}\times\mathbf{L}+B_4\mathbf{B}+C_4\mathbf{C})_1+i(\mathbf{A}\times\mathbf{L}+B_4\mathbf{B}+C_4\mathbf{C})_2]$
1, −1	1, 1	$-(1/2\sqrt{2})\{[(\mathbf{B}\times\mathbf{L}+B_4\mathbf{A}-M\mathbf{C})_1+(\mathbf{C}\times\mathbf{L}+C_4\mathbf{A}+M\mathbf{B})_2]+i[-(\mathbf{B}\times\mathbf{L}+B_4\mathbf{A}-M\mathbf{C})_2+(\mathbf{C}\times\mathbf{L}+C_4\mathbf{A}+M\mathbf{B})_1]\}$
1, 0	1, 1	$\frac{1}{2}[(\mathbf{C}\times\mathbf{L}+C_4\mathbf{A}+M\mathbf{B})_3-i(\mathbf{B}\times\mathbf{L}+B_4\mathbf{A}-M\mathbf{C})_3]$
1, 1	1, 1	$-(1/2\sqrt{2})\{[(\mathbf{B}\times\mathbf{L}+B_4\mathbf{A}-M\mathbf{C})_1-(\mathbf{C}\times\mathbf{L}+C_4\mathbf{A}+M\mathbf{B})_2]+i[(\mathbf{B}\times\mathbf{L}+B_4\mathbf{A}-M\mathbf{C})_2+(\mathbf{C}\times\mathbf{L}+C_4\mathbf{A}+M\mathbf{B})_4]\}$

dimensional representation and the polynomials which are the bases of the 15-dimensional and one-dimensional representations is an intriguing phenomenon, which seems to be more general. For instance, it occurs also for the su(3) algebra[22] in the case of the ten-dimensional representation which appears in the subspace of polynomials of degree 3.

Using the relations (3.24)–(3.30) we can calculate the expressions of the cubic and quartic invariants (3.17) and (3.18) of the algebra so(4,2) on the family of eight-dimensional manifolds pointed out by Proposition 3.5. We obtain

$$I_3 = \mp (I_2/3)^{3/2}, \quad I_4 = I_2^2/3 . \tag{3.31}$$

The consideration of these invariants gives thus no additional information.

B. The algebras so(4,1) and so(3,2)

From the basic polynomials for the irreducible representation (3.2) of so(5) we obtain the basis vectors of the noncompact real forms so(4,1) and so(3,2) (Tables IV and V).

For the Lie algebra so(4,1) the following statements are true (cf. Ref. 11).

(i) The relations obtained equating to zero the basis vectors of the 14-dimensional representation (2,0) have as their only common solution the origin of so(4,1)*. A similar statement is true for the 35-dimensional representation (0,4).

(ii) The relations obtained equating to zero the basis vectors of the five-dimensional representation (0,1) define a seven-dimensional submanifold of so(4,1)*.

These statements lead to the following proposition.

Proposition 3.6: The only subrepresentation of the representation $(\mathrm{ad}\otimes\mathrm{ad})_s$ of the algebra so(4,1) to which correspond nondegenerate SO(4,1)-invariant submanifolds is $(0,1)\oplus(0,0)$. To this representation, a family of six-dimensional submanifolds denoted $\mathcal{O}_6^k$, defined by the equations

$$\mathbf{A}\times\mathbf{B} - B_4\mathbf{L} = 0, \quad \mathbf{L}^2 + \mathbf{A}^2 - \mathbf{B}^2 - B_4^2 = k^2 \tag{3.32}$$

(k = real constant) can be associated.

The quartic invariant $\mathscr{C}^2 + \mathscr{C}_4^2 - \mathscr{M}^2$ vanishes on any $\mathcal{O}_6^k$.

The following proposition, similar to 3.3, is true.

TABLE IV. Basis vectors of the subrepresentation (0, 1) of $(\mathrm{ad}\times\mathrm{ad})_s$ for the algebra so(4, 1). The labeling adopted in Ref. 21 for the so(6) basis functions has been conserved for the corresponding so(4, 1) basis functions obtained using Weyl's unitary trick.

S, M_S	T, M_T	Basis vector $f^{(0,1)}_{(S,M_S)(T,M_T)}$
0, 0	0, 0	$\frac{1}{6}\mathbf{L}\cdot\mathbf{A}$
$\frac{1}{2}, \frac{1}{2}$	$\frac{1}{2}, \frac{1}{2}$	$(1/2\sqrt{3})[-(\mathbf{A}\times\mathbf{B} - B_4\mathbf{L})_2 + i(\mathbf{A}\times\mathbf{B} - B_4\mathbf{L})_1]$
$\frac{1}{2}, -\frac{1}{2}$	$\frac{1}{2}, \frac{1}{2}$	$-(1/2\sqrt{3})[\mathbf{L}\cdot\mathbf{B} + i(\mathbf{A}\times\mathbf{B} - B_4\mathbf{L})_3]$
$\frac{1}{2}, \frac{1}{2}$	$\frac{1}{2}, -\frac{1}{2}$	$-(1/2\sqrt{3})[\mathbf{L}\cdot\mathbf{B} - i(\mathbf{A}\times\mathbf{B} - B_4\mathbf{L})_3]$
$\frac{1}{2}, -\frac{1}{2}$	$\frac{1}{2}, -\frac{1}{2}$	$(1/2\sqrt{3})[(\mathbf{A}\times\mathbf{B} - B_4\mathbf{L})_2 + i(\mathbf{A}\times\mathbf{B} - B_4\mathbf{L})_1]$

Proposition 3.7: For any $u\in\mathcal{O}_6^k$, rank $M_{\mathrm{so}(4,1)}(u)\geqslant 6$.

The complete proof of proposition 3.7, which is straightforward, is given in Ref. 11.

The manifolds $\mathcal{O}_6^k$ can be thus ascribed a nondegenerate KKS symplectic form; its explicit expression, obtainable from an adaptation of Mukunka's result[7] to the noncompact algebra so(4,1), is given by the following proposition.

Proposition 3.8: (1) The Kirillov–Kostant–Souriau symplectic form on $\mathcal{O}_6^k$ has the expression

$$\begin{aligned}\omega_6 = {} & \frac{1}{B_4}\sum_{i=1}^{3} dA_i\wedge dB_i - \sum_{i,j=1}^{3}\frac{B_iB_j}{B_4B^2}\,dA_i\wedge dB_j \\ & - \sum_{i=1}^{3}\frac{B_i}{B^2}\,dA_i\wedge dB_4 + \sum_{i,j=1}^{3}\frac{A_iB_j}{B_4B^2}\,dB_j\wedge dB_i \\ & - \frac{1}{B_4^2B^2}\sum_{i=1}^{3}(\mathbf{B}\times(\mathbf{A}\times\mathbf{B}))_i\,dB_4\wedge dB_i\end{aligned} \tag{3.33}$$

and can be brought to a Darboux canonical form $\omega = \Sigma dq_i\wedge dp_i$ by means of the transformation

$$q_i = (B^2)^{1/2}(B_4)^{-1}A_i, \quad p_i = (B^2)^{-1/2}B_i . \tag{3.34}$$

(2) Denoting with ω the generic KKS symplectic form defined by (2.8), we have $\omega_u = (\omega_6)_u$ if and only if $u\in\mathcal{O}_6^k$.

An analog of statement (i) is *no more true* for the algebra so(3,2), for which it becomes the following.

(i_1) The relations obtained by equating to zero the basis vectors of the 14-dimensional representation (2,0) define a

TABLE V. Basis vectors of the 14-dimensional subrepresentation (2, 0) of $(\mathrm{ad}\times\mathrm{ad})_s$, for the algebra so(4, 1). The corresponding so(3, 2) basis function are obtained from the so(4, 1) basis functions by the transformation $\mathbf{L}\to\mathbf{L}$, $\mathbf{A}\to i\mathbf{A}$, $\mathbf{B}\to\mathbf{B}$, $B_4\to iB_4$.

S, M_S	T, M_T	$f^{(2,0)}_{(S,M_S)(T,M_T)}$
0, 0	0, 0	$(1/4\sqrt{15})(L_1^2 + L_2^2 + L_3^2 + A_1^2 + A_2^2 + A_3^2)$
$\frac{1}{2}, \frac{1}{2}$	$\frac{1}{2}, \frac{1}{2}$	$(1/2\sqrt{3})[(\mathbf{B}\times\mathbf{L} + B_4\mathbf{A})_2 - i(\mathbf{B}\times\mathbf{L} + B_4\mathbf{A})_1]$
$\frac{1}{2}, -\frac{1}{2}$	$\frac{1}{2}, \frac{1}{2}$	$-(1/2\sqrt{3})[\mathbf{A}\cdot\mathbf{B} - i(\mathbf{B}\times\mathbf{L} + B_4\mathbf{A})_3]$
$\frac{1}{2}, \frac{1}{2}$	$\frac{1}{2}, -\frac{1}{2}$	$-(1/2\sqrt{3})[\mathbf{A}\cdot\mathbf{B} + i(\mathbf{B}\times\mathbf{L} + B_4\mathbf{A})_3]$
$\frac{1}{2}, -\frac{1}{2}$	$\frac{1}{2}, -\frac{1}{2}$	$-(1/2\sqrt{3})[(\mathbf{B}\times\mathbf{L} + B_4\mathbf{A})_2 + i(\mathbf{B}\times\mathbf{L} + B_4\mathbf{A})_1]$
1, 1	1, 1	$(1/4\sqrt{3})[L_1^2 - L_2^2 - A_1^2 + A_2^2 + B_1^2 - B_2^2 + 2i(L_1L_2 - A_1A_2 + B_1B_2)]$
1, 0	1, 1	$(1/2\sqrt{6})[-(\mathbf{A}\times\mathbf{L} + B_4\mathbf{B})_2 - L_1L_3 + A_1A_3 - B_1B_3 + i((\mathbf{A}\times\mathbf{L} + B_4\mathbf{B})_1 - L_2L_3 + A_2A_3 - B_2B_3)]$
1, −1	1, 1	$(1/4\sqrt{3})[(-L_1^2 - L_2^2 + A_1^2 + A_2^2 + B_3^2 - B_4^2) - 2i(\mathbf{A}\times\mathbf{L} + B_4\mathbf{B})_3]$
1, 1	1, 0	$(1/2\sqrt{6})[-(\mathbf{A}\times\mathbf{L} + B_4\mathbf{B})_2 + L_1L_3 - A_1A_3 + B_1B_3 + i((\mathbf{A}\times\mathbf{L} + B_4\mathbf{B})_1 + L_2L_3 - A_2A_3 + B_2B_3)]$
1, 0	1, 0	$(1/4\sqrt{3})(-2L_3^2 + 2A_3^2 + B_1^2 + B_2^2 - B_3^2 - B_4^2)$
1, −1	1, 0	$(1/2\sqrt{6})[(\mathbf{A}\times\mathbf{L} + B_4\mathbf{B})_2 - L_1L_3 + A_1A_3 - B_1B_3 + i((\mathbf{A}\times\mathbf{L} + B_4\mathbf{B})_1 + L_2L_3 - A_2A_3 + B_2B_3)]$
1, 1	1, −1	$(1/4\sqrt{3})[(-L_1^2 - L_2^2 + A_1^2 + A_2^2 + B_3^2 - B_4^2) + 2i(\mathbf{A}\times\mathbf{L} + B_4\mathbf{B})_3]$
1, 0	1, −1	$(1/2\sqrt{6})[(\mathbf{A}\times\mathbf{L} + B_4\mathbf{B})_2 + L_1L_3 - A_1A_3 + B_1B_3 + i((\mathbf{A}\times\mathbf{L} + B_4\mathbf{B})_1 - L_2L_3 + A_2A_3 - B_2B_3)]$
1, −1	1, −1	$(1/4\sqrt{3})[(L_1^2 - L_2^2 - A_1^2 + A_2^2 + B_1^2 - B_2^2) - 2i(L_1L_2 - A_1A_2 + B_1B_2)]$

four-dimensional submanifold of so(3,2)*. The 35-dimensional representation of so(3,2), contained in (3.2) leads to the trivial solution $u = 0$.

The four-dimensional manifold associated to the representation (2,0) is defined by the following six equations

$$\mathbf{A}^2 = \mathbf{B}^2 = B_4^2\,, \quad \mathbf{A}\cdot\mathbf{B} = 0\,, \quad \mathbf{A}\times\mathbf{B} - B_4\mathbf{L} = 0\,. \tag{3.35}$$

The quadratic and quartic invariants vanish on the manifold defined by (3.35). The analog of Proposition 3.6 for so(3,2) is the following.

Proposition 3.9: The representations of the algebra so(3,2), contained in $(\mathrm{ad}\otimes\mathrm{ad})_s$, to which correspond nondegenerate SO(3,2)-invariant submanifolds, are $(0,1)\oplus(0,0)$ and (2,0). The first representation leads to a family of six-dimensional submanifolds of so(3,2)*, defined by the equations $\mathbf{A}\times\mathbf{B} - \mathbf{L} = 0$ and $\mathbf{L}^2 - \mathbf{A}^2 - \mathbf{B}^2 + B_4^2 = k$; the second leads to a four-dimensional submanifold, defined by Eqs. (3.35).

C. The algebra sp(4,*R*)

Let us transpose the results obtained for the algebra so(3,2) to the algebra $\mathrm{sp}(4,R)\sim\mathrm{so}(3,2)$. Let us denote by $A_{11},A_{12},A_{21},A_{22},B_{11},B_{12},B_{22},C_{11},C_{12},C_{22}$ the basis of the algebra sp(4,*R*) for which the Lie products have the expressions

$$\begin{aligned}
&[A_{ij},A_{kl}] = \delta_{jk}A_{il} - \delta_{il}A_{kj}\,,\\
&[A_{ij},B_{kl}] = \delta_{jk}B_{il} - \delta_{jl}B_{ik}\,,\\
&[A_{ij},C_{kl}] = -(\delta_{ik}C_{jl} + \delta_{il}C_{jk})\,,\\
&[B_{ij},C_{kl}] = -(\delta_{ik}A_{jl} + \delta_{il}A_{jk} + \delta_{jk}A_{il} + \delta_{jl}A_{ik})\,.
\end{aligned} \tag{3.36}$$

The isomorphism between sp(4,*R*) and so(3,2) is given by the relations

$$\begin{aligned}
&L_1 = (i/2)(A_{12} + A_{21})\,, \quad L_2 = -\tfrac{1}{2}(A_{12} - A_{21})\,,\\
&L_3 = (i/2)(A_{11} - A_{22})\,, \quad A_1 = -\tfrac{1}{4}(B_{11} - B_{22} - C_{11} + C_{22})\,,\\
&A_2 = -(i/4)(B_{11} + B_{22} + C_{11} + C_{22})\,, \quad A_3 = \tfrac{1}{2}(B_{12} - C_{12})\,,\\
&B_1 = -(i/4)(B_{11} - B_{22} + C_{11} - C_{22})\,,\\
&B_2 = \tfrac{1}{4}(B_{11} + B_{22} - C_{11} - C_{22})\,,\\
&B_3 = (i/2)(B_{12} + C_{12})\,, \quad B_4 = -(i/2)(A_{11} + A_{22})\,.
\end{aligned} \tag{3.37}$$

Using this isomorphism, Proposition 3.7 becomes the following.

Proposition 3.10: The Sp(4,*R*)-invariant even-dimensional submanifolds of sp(4,*R*)* associated to subrepresentations of $(\mathrm{ad}\otimes\mathrm{ad})_s$ are the following.

(1) A family of six-dimensional manifolds associated to representation $(0,1)\oplus(0,0)$ and defined by the equations

$$(B_{11}C_{11} - A_{11}^2) - (B_{22}C_{22} + A_{22}^2) = 0\,, \tag{3.38}$$

$$B_{12}C_{22} + C_{12}B_{11} - A_{12}(A_{11} + A_{22}) = 0\,, \tag{3.39}$$

$$B_{12}C_{11} + C_{12}B_{22} - A_{21}(A_{11} + A_{22}) = 0\,, \tag{3.40}$$

$$(B_{11}C_{11} - A_{11}^2) + (B_{22}C_{22} - A_{22}^2) + 2(B_{12}C_{12} - A_{12}A_{21}) = k\,, \tag{3.41}$$

where k is an arbitrary real constant.

(2) A four-dimensional manifold associated to representation (2,0) and defined by the equations

$$(B_{11} + C_{22})(B_{22} + C_{11}) - (B_{12} - C_{12})^2 - (A_{11} + A_{22})^2 = 0\,, \tag{3.42}$$

$$(B_{11} - C_{22})(B_{22} - C_{11}) - (B_{12} + C_{12})^2 + (A_{11} + A_{22})^2 = 0\,, \tag{3.43}$$

$$(C_{11}C_{22} - C_{12}^2) - (B_{11}B_{22} - B_{12}^2) = 0\,. \tag{3.44}$$

The polynomials which define the six-dimensional submanifold of sp(4,*R*)* are particularly interesting: relations (3.38)–(3.41) have been recently obtained [see Ref. 12, relations (3.23)] as the classical limit of quantum constraints in the quasiclassic treatment of the helium atom, in which the large quantum number is the dimension of the space. To recover the identities [see Ref. 12, (3.23)], the generators A_{ij},B_{ij},C_{ij} (3.36) and (3.37) have to be replaced by F_{ij},P_{ij},X_{ij}, respectively.

The quantum constraints considered in Ref. 12 were implicitly incorporated in the Holstein–Primakoff representation. We proved that the identities (3.38)–(3.41), written in symmetrized form, appear as explicit constraints in the quantum problem considered in Ref. 12.

ACKNOWLEDGMENT

The authors are indebted to the referee for pointing out to them that the map $x_i \to X_{x_i}(u)$ defined by Eq. (2.5) is a Jordan map applied to the adjoint representation.

[1] V. I. Arnol'd, *Méthodes mathématiques de la mécanique classique* (Mir, Moscow, 1976).
[2] E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974). The term "Poisson bracket realization" replaces "canonical realization" used in our previous papers (see Refs. 8–11).
[3] J-M. Souriau, Nuovo Cimento A **49**, 197 (1967).
[4] H. Bacry, Nuovo Cimento A **41**, 222 (1966).
[5] G. Györgyi, Nuovo Cimento A **53**, 717 (1968).
[6] G. Györgyi, Nuovo Cimento A **62**, 449 (1969).
[7] N. Mukunda, J. Math. Phys. **8**, 1069 (1967).
[8] M. Iosifescu and H. Scutaru, J. Math. Phys. **21**, 2033 (1980).
[9] M. Iosifescu and H. Scutaru, "Canonical Realizations and Linear Representations," Preprint FT-205-81, Bucharest, 1981.
[10] M. Iosifescu and H. Scutaru, "Canonical Realizations of Lie Algebras and Subrepresentations of ad ⊗ ad," Preprint FT-217-82, Bucharest, 1982.
[11] M. Iosifescu and H. Scutaru, "Canonical Realizations of $B_2 \sim C_2$ Lie algebras," Preprint FT-223-82, Bucharest, 1982.
[12] L. D. Mlodinow and N. Papanicolaou, Ann. Phys. (NY) **131**, 1 (1981).
[13] A. Joseph, Commun. Math. Phys. **36**, 325 (1974).
[14] A. Joseph, "Realizations in Classical and Quantum Mechanics, Third International Colloquium on Group Theoretical Methods inPhysics," Marseille, 1974, Preprint I.H.E.S.
[15] P. Jordan, Z. Phys. **94**, 531 (1935).
[16] L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics, Theory and Application*, Vol. 8, *Encyclopedia of Mathematics and its Applications* (Addison–Wesley, Reading, MA, 1981).
[17] A. A. Kirillov, *Elementy Teorii Predstavleniia* (Nauka, Moscow, 1972).
[18] C. Chevalley, *Théorie des Groupes de Lie. II: Groupes Algébriques* (Hermann, Paris, 1951).
[19] V. S. Varadarajan, *Lie Groups, Lie Algebras and their Representations* (Prentice-Hall, Englewood Cliffs, NJ, 1974).
[20] K. T. Hecht and S. C. Pang, J. Math. Phys. **10**, 1571 (1969).
[21] K. T. Hecht, Nucl. Phys. **63**, 177 (1965).
[22] J. Dixmier, *Non-commutative Harmonic Analysis, Lecture Notes in Mathematics, No. 466* (Springer, Berlin, 1975).

Clebsch–Gordan coefficients for E_6 and SO(10) unification models[a)]

In-Guy Koh
Department of Physics, Sogang University, Seoul, Korea

J. Patera
Centre de Recherche de Mathématiques Appliquées, Université de Montréal, Montreal, Canada

C. Rousseau
Département de Mathématiques et de Statistique, Université de Montréal, Montreal, Canada

(Received 7 June 1983; accepted for publication 9 November 1983)

We illustrate here a new method for computing Clebsch–Gordan coefficients (CGC) for E_6 by computing CGC for the product $27 \otimes 27$ of the irreducible representation (100000) of E_6 with itself. These CGC are calculated thrice: once in a weight vector basis independent of any semisimple subgroup, then in a basis which refers to $SO(10) \subset E_6$, and finally in a basis referring to $SU(5) \subset SO(10) \subset E_6$.

PACS numbers: 02.20.Rt, 03.65.Fd

1. INTRODUCTION

The unification of strong and electromagnetic interactions with the weak ones based on a simple Lie group is regarded as one of the possible avenues of advancement for understanding of the fundamental interactions. Even if the ultimate unification should take place in the context of a supersymmetry, the Lie group part of it would be left to play a major role. It is generally accepted that the largest Lie group should contain $SU(3) \times SU(2) \times U(1)$ of the standard model via SU(5). Among the higher groups SO(10) and E_6 are considered most often.[1,2] Details of the approach as well as extensive bibliography can be found in recent reviews.[2]

Any exploitation of a particular model to the extent where experimental predictions or comparisons can be made inevitably involves some Clebsch–Gordan coefficients (CGC). The purpose of this paper is to provide CGC for models based on SO(10) and E_6. We proceed in a way which is complementary to our CGC computation for $SU(5) \supset SU(3) \times SU(2) \times U(1)$ in Ref. 3. Since $E_6 \supset SO(10)$, it is natural to set up the problem in terms of E_6.

The idea of our approach, similarly as Ref. 3, is to calculate and present only a list of representative CGC and to provide a prescription how any other CGC can be readily identified with one of the representatives. The main tool of such an identification are operators R_ξ which permute the basis vectors—physical states in representation spaces. These operators are new in physics literature (Refs. 3 and 4) although in mathematics they are known (cf. Ref. 5). Curiously, using our method it turns out that the larger the Lie group the easier is the CGC computation for relevant representations. The latter fact is somewhat obscured by the inevitably more cumbersome notations for larger groups.

We consider here the tensor product, $27 \otimes 27$, of the natural representations of E_6, identify the basis vectors of the representation space of 27 with the light fermions of the theory[3] and calculate the related CGC. Such CGC's are needed, for instance, for the study of the Higgs term of the Lagrangian. The physical states carry (additive) quantum numbers which are in a correspondence with the weights of the representation of the Lie group. If a symmetry is exact it is only a matter of convenience to consider with the group also a subgroup. In physics, however, the symmetries are broken. This forces one to prefer certain subgroups of the unifying group. In our case it is $E_6 \supset SO(10) \supset SU(5) \supset SU(3) \times SU(2) \times U(1)$. Thus the task in this paper is to consider the first two inclusions and provide a connection with the last one contained in Ref. 3.

The elementary explicit descriptions usually provided in physics literature for Lie algebras are not practical for a Lie algebra of the size of E_6. Although it would be quite simple to find the 27×27 matrices representing E_6 generators (using, for instance, the method of Ref. 4 where the G_2 and F_4 generators were found), it would be rather difficult to make much use of them. Fortunately, it is not necessary.

We assume that the reader is familiar with the root decomposition of the Lie algebra and with the algorithm for computing weights of a representation, starting from the highest one.[6] We make use only of the 12 basic generators of E_6

$$e_{\zeta_i}, e_{-\zeta_i}, \quad i = 1,2,3,4,5,6$$

corresponding to simple roots ζ_i. The simple roots are numbered as on Table I. Since our method was described earlier,[3,5] our explanations are brief.

Section 2 contains CGC's of the product $27 \otimes 27$ in E_6 in a basis independent of any subgroup; Sec. 3 (resp. 4) contains the CGC's for a basis relative to $SO(10) \subset E_6$ [resp. $SU(5) \subset SO(10) \subset E_6$]. Section 5 contains an example of CGC of $27 \otimes 27$ relative to the $SU(3) \times SU(2) \times U(1) \subset SU(5) \subset SO(10) \subset E_6$ basis.

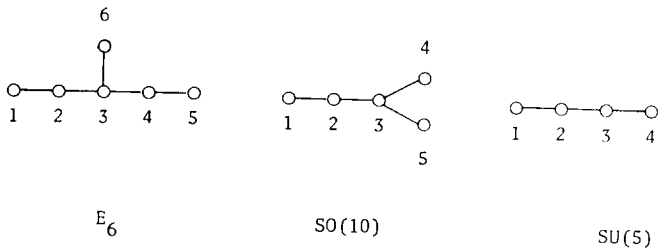
2. CLEBSCH–GORDAN COEFFICIENTS IN E_6-BASIS

The additive quantum numbers of a multiplet of elementary particles associated with an irreducible representation of E_6 are precisely the coordinates of weights of that representation (up to a normalization and a choice of basis in the weight space). For higher rank Lie algebras, only some of the quantum numbers may have a clear physical interpretation.

a) Work supported in part by the National Research Council of Canada, the Ministère de l'Education du Québec, the Korea Traders Foundation, and by the Korea Science and Engineering Foundation.

 0022-2488/84/102863-10$02.50

TABLE I. Numbering of simple roots of E_6, SO(10), and SU(5).



Consider the E_6 tensor product

$$27 \otimes 27 = 351' \oplus 351 \oplus 27 , \tag{2.1}$$

or equivalently

$$(100000) \otimes (100000) = (200000) \oplus (010000) \oplus (000010) , \tag{2.2}$$

where the E_6 representations are denoted by their highest weights (cf. Table II). Let us first find CGC's for the irreducible subspace (200000). Obviously,

$$(200000) = (100000)(100000) . \tag{2.3}$$

Other weight vectors of the (200000) subspace are obtained from (2.3) by successive application of the generators $e_{-\zeta_i}$ to (2.3) and then normalizing.

The action of $e_{\pm\zeta_i}$ on the basis weight vectors of 27 and $\overline{27}$ is given explicitly by

$$e_{\pm\zeta_1}(w) = \begin{cases} (w_1 \pm 2, w_2 \mp 1, w_3, w_4, w_5, w_6) & \text{if } w_1 \lesseqgtr 0 \\ 0 & \text{otherwise} \end{cases}$$

$$e_{\pm\zeta_2}(w) = \begin{cases} (w_1 \mp 1, w_2 \pm 2, w_3 \mp 1, w_4, w_5, w_6) & \text{if } w_2 \lesseqgtr 0 \\ 0 & \text{otherwise} \end{cases}$$

$$e_{\pm\zeta_3}(w) = \begin{cases} (w_1, w_2 \mp 1, w_3 \pm 2, w_4 \mp 1, w_5, w_6 \mp 1) & \text{if } w_3 \lesseqgtr 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.4}$$

$$e_{\pm\zeta_4}(w) = \begin{cases} (w_1, w_2, w_3 \mp 1, w_4 \pm 2, w_5 \mp 1, w_6) & \text{if } w_4 \lesseqgtr 0 \\ 0 & \text{otherwise} \end{cases}$$

$$e_{\pm\zeta_5}(w) = \begin{cases} (w_1, w_2, w_3, w_4 \mp 1, w_5 \pm 2, w_6) & \text{if } w_5 \lesseqgtr 0 \\ 0 & \text{otherwise} \end{cases}$$

$$e_{\pm\zeta_6}(w) = \begin{cases} (w_1, w_2, w_3 \mp 1, w_4, w_5, w_6 \pm 2) & \text{if } w_6 \lesseqgtr 0 \\ 0 & \text{otherwise} \end{cases} .$$

Using (2.4) one has

$$e_{-\zeta_1}(200000) \sim (010000) \sim (e_{-\zeta_1}(100000))(100000) + (100000)(e_{-\zeta_1}(100000))$$

so that

$$e_{-\zeta_1}\binom{200000}{200000} \sim \binom{200000}{010000} = \frac{1}{\sqrt{2}}(\bar{1}10000)(100000) + \frac{1}{\sqrt{2}}(100000)(\bar{1}10000) , \tag{2.5}$$

where we write $\bar{x}$ for $-x$. Here $\binom{200000}{010000}$ denotes a weight vector of the weight (010000) belonging to the irreducible subspace (200000). Continuing further one gets

$$e_{-\zeta_2}\binom{200000}{010000} = \binom{200000}{1\bar{1}1000} = \frac{1}{\sqrt{2}}(0\bar{1}1000)(100000) + \frac{1}{\sqrt{2}}(100000)(0\bar{1}1000) ,$$

$$e_{-\zeta_1}\binom{200000}{010000} \sim \binom{200000}{\bar{2}20000} = (\bar{1}10000)(\bar{1}10000) . \tag{2.6}$$

In this way one could, in principle, generate the whole subspace (200000). Next we go to the representation 351′. Its highest weight is (010000) and the corresponding weight vector (010000) must be orthogonal to (2.5). Choosing conveniently its arbitrary phase, we have

$$\binom{010000}{010000} = \frac{1}{\sqrt{2}}(\bar{1}10000)(100000) - \frac{1}{\sqrt{2}}(100000)(\bar{1}10000) . \tag{2.7}$$

Again starting from (2.7) one can generate the subspace 351′.

There are 10 different products of two 27-weight vectors with the dominant weight (000010). Four (resp. five) linear combinations of them lie in the subspace 351 (resp. 351′). The unique linear combination orthogonal to those nine ones is the highest weight vector $\binom{000010}{000010}$ of the subspace $\overline{27}$. It is listed in the last column of Table III. The subspace is generated directly from it.

TABLE II. Properties of relevant representations of E_6 and SO(10).

Lie algebra	Highest weight	Dimension	Other dominant weights	Reduction to the subalgebra D_5 in E_6 and A_4 in D_5
E_6	(100000)	27	—	(00001) ⊕ (10000) ⊕ (0)
	(000010)	$\overline{27}$	—	(00010) ⊕ (10000) ⊕ (0)
	(200000)	351′	(010000), 4(000010)	(00002) ⊕ (10001) ⊕ (20000) ⊕ (00001) ⊕ (10000) ⊕ (0)
	(010000)	351	5(000010)	(10001) ⊕ (00100) ⊕ (01000) ⊕ (00010) ⊕ (00001) ⊕ (10000)
D_5	(10000)	10	—	(1000) ⊕ (0001)
	(00010)	$\overline{16}$	—	(0010) ⊕ (1000) ⊕ (0)
	(00001)	16	—	(0100) ⊕ (0001) ⊕ (0)
	(01000)	45	5(00000)	(1001) ⊕ (0100) ⊕ (0010) ⊕ (0)
	(20000)	54	(01000), 4(00000)	(2000) ⊕ (1001) ⊕ (0002)
	(00100)	120	4(10000)	(0101) ⊕ (1010) ⊕ (0010) ⊕ (0100) ⊕ (1000) ⊕ (0001)
	(00002)	126	(00100), 3(10000)	(0020) ⊕ (0101) ⊕ (0002) ⊕ (0100) ⊕ (0001) ⊕ (0)
	(10001)	144	4(00010)	(0101) ⊕ (1100) ⊕ (0002) ⊕ (1001) ⊕ (0010) ⊕ (1000) ⊕ (0001)

TABLE III. CGC for (100000) ⊗ (100000) in E_6. Multiple weights are identified by subscripts. The subscripts i specifies the last $e_{-\zeta_i}$ used to derive the weight. This subscript characterizes the weight vector.

		(200000)											
		(200000)	(200000)	(010000)									
(100000)	(100000)	1	(010000)	(010000)									
	(100000)	(1̄10000)	$1/\sqrt{2}$	$1/\sqrt{2}$	(200000)				(010000)					(000010)
	(1̄10000)	(100000)	$1/\sqrt{2}$	$-1/\sqrt{2}$	$(000010)_4$	$(000010)_3$	$(000010)_2$	$(000010)_1$	$(000010)_4$	$(000010)_3$	$(000010)_2$	$(000010)_1$	$(000010)_6$	(000010)
		(000$\bar{1}$11)	(00010$\bar{1}$)	1/2				1/2				-1/2	$1/\sqrt{10}$
		(00010$\bar{1}$)	(000$\bar{1}$11)	1/2				-1/2				1/2	$1/\sqrt{10}$
		(00$\bar{1}$101)	(001$\bar{1}$1$\bar{1}$)	1/2	1/2			1/2	1/2			1/2	$-1/\sqrt{10}$
		(001$\bar{1}$1$\bar{1}$)	(00$\bar{1}$101)	1/2	1/2			-1/2	-1/2			-1/2	$-1/\sqrt{10}$
		(0$\bar{1}$1000)	(01$\bar{1}$010)		1/2	1/2			1/2	1/2			$1/\sqrt{10}$
		(01$\bar{1}$010)	(0$\bar{1}$1000)		1/2	1/2			-1/2	-1/2			$1/\sqrt{10}$
		(1̄10000)	(1$\bar{1}$0010)			1/2	1/2			1/2	1/2		$-1/\sqrt{10}$
		(1$\bar{1}$0010)	(1̄10000)			1/2	1/2			-1/2	-1/2		$-1/\sqrt{10}$
		(100000)	(1̄00010)				1/2				1/2		$1/\sqrt{10}$
		(1̄00010)	(100000)				1/2				-1/2		$1/\sqrt{10}$

In this way one can get all 729 ($= 27^2$) linear combinations of weight vectors of our problem and thus the CGC. However, it turns out that the CGC takes only a few distinct nonzero values. Namely,

$$1, 1/\sqrt{2} \text{ and } \tfrac{1}{2} \text{ in (200000) subspace},$$
$$\pm 1/\sqrt{2} \text{ and } \pm\tfrac{1}{2} \text{ in (010000)},$$
$$\pm 1/\sqrt{10} \text{ in (000010)}. \tag{2.8}$$

It is, therefore, natural to avoid listing all 729 CGC's in our problem in favor of a representative subset of them in each of the three irreducible subspaces (200000), (010000), and (000010). It was pointed out before[3,5] that it suffices to give CGC's which correspond to dominant weight vectors in each of the three irreducible subspaces. Table III contains the representative subset of CGC coefficients for $27 \otimes 27$. Each column of the table is denoted by two weights: the highest one in the first line of each column indicating one of the three subspaces, and the weight of an actual (dominant) weight vector.

Next we need a procedure to relate any other CGC with one of the entries of Table III. It is shown[5] that such a transformation is provided by the elements $R_{\zeta_i} \in E_6$, where ζ_i is a simple root of E_6, given by

$$R_{\zeta_i}(w) = \exp(e_{-\zeta_i})\exp(-e_{\zeta_i})\exp(e_{-\zeta_i})(w)$$
$$= \pm(r_i w) \quad \text{for } r_i w \neq w. \tag{2.9}$$

Here (w) denotes a weight vector of E_6 weight w; $r_i w$ is the reflection of the weight w

$$r_i w = w - (w,\zeta_i)\zeta_i. \tag{2.10}$$

The scalar product (w,ζ_i) of a weight $w = (w_1,\ldots,w_6)$ with the simple root ζ_i of E_6 is given by

$$(w,\zeta_i) = w_i. \tag{2.11}$$

We illustrate the action of the R_{ζ_i} by two examples.

Example 1: Let us consider (2.5). Then

$$R_{\zeta_1}(200000) = (R_{\zeta_1}(100000))(R_{\zeta_1}(100000)) \tag{2.12}$$

or

$$(\bar{2}20000) = (\bar{1}10000)(\bar{1}10000). \tag{2.13}$$

Similarly one finds that

$$(0\bar{2}2000) = R_{\zeta_2}R_{\zeta_1}(200000) = (0\bar{1}1000)(0\bar{1}1000), \tag{2.14}$$
$$(00\bar{2}202) = R_{\zeta_3}R_{\zeta_2}R_{\zeta_1}(200000) = (00\bar{1}101)(00\bar{1}101), \tag{2.15}$$

etc.

Example 2: Suppose we have given $(w_0) = (0000\bar{1}1)(001\bar{1}1\bar{1})$ and we want to find its contribution into the representation spaces 351, 351′, and 27. First

$$R_{-\zeta_5}R_{-\zeta_4}(w_0)$$
$$= R_{-\zeta_5}(0000\bar{1}1)(00010\bar{1}) = (000\bar{1}11)(00010\bar{1})$$
$$= \tfrac{1}{2}(000\bar{1}11)(00010\bar{1}) + (00010\bar{1})(000\bar{1}11)$$
$$+ \tfrac{1}{2}(000\bar{1}11)(00010\bar{1}) - (00010\bar{1})(000\bar{1}11) \equiv S + A. \tag{2.16}$$

The antisymmetric term A obviously entirely belongs to the space of 351. More precisely, one finds using Table III that

$$A = \tfrac{1}{2}\left(\binom{010000}{000010}_4 - \binom{010000}{000010}_6\right), \tag{2.17}$$

where we write the highest weight over the weight in order to avoid any ambiguity. The symmetric term S belongs to $351' \oplus \overline{27}$. One has

$$S = \binom{200000}{000010} = \tfrac{4}{5}\binom{200000}{000010}_4 - \tfrac{3}{5}\binom{200000}{000010}_3$$
$$+ \tfrac{2}{5}\binom{200000}{000010}_2 - \tfrac{1}{5}\binom{200000}{000010}_1 + (\sqrt{5}/\sqrt{2})\binom{000010}{000010}. \tag{2.18}$$

Remark: The elements $(000010)_i$ with different subscripts i correspond to different weight vectors of the same weight. The subscripts i indicate the last $e_{-\zeta_i}$ used to derive the vector. The subscript characterizes the vector.

3. CLEBSCH–GORDAN COEFFICIENTS IN SO(10) BASIS

Here we consider the same product (2.2) of E_6 representations. This time we want to calculate the corresponding CGC in the basis of SO(10)-weight vectors. Again it suffices to find only the linear combination of products which have SO(10)-dominant weights. Since $SO(10)\subset E_6$, there will be generally more SO(10) dominant weights than in the previous case for the same product space.

The following matrix[7] gives a convenient way to translate an E_6 weight $w=(w_1,...,w_6)$ into a weight $v=(v_1,...,v_5)$ of SO(10):

$$v=Pw^T=(w_2+w_3+w_4,w_6,w_3,w_4+w_5,w_1+w_2)\,, \tag{3.1}$$

i.e.,

$$P=\begin{pmatrix}0&1&1&1&0&0\\0&0&0&0&0&1\\0&0&1&0&0&0\\0&0&0&1&1&0\\1&1&0&0&0&0\end{pmatrix}. \tag{3.2}$$

Basic generators of SO(10) compatible with P are

$$\begin{aligned}e_{\xi_1}&=[e_{\xi_2},[e_{\xi_3},e_{\xi_4}]]\,,\\ e_{\xi_2}&=e_{\xi_6}\,,\\ e_{\xi_3}&=e_{\xi_3}\,,\\ e_{\xi_4}&=[e_{\xi_4},e_{\xi_5}]\,,\\ e_{\xi_5}&=[e_{\xi_1},e_{\xi_2}]\,.\end{aligned} \tag{3.3}$$

The reduction of the E_6 representations 351, 351′, 27, and $\overline{27}$ to SO(10) is found in Table II. Given that information, we proceed as before with the only difference that we must use only the generators (3.3) of SO(10) and all weight vectors must be interpreted as SO(10)-weight vectors using (3.1). Let us underline that (3.2) is one of many possible equivalent choices of the matrix P. A different choice would imply different linear combinations (3.3) for generators of SO(10). For extensive computations (3.2) is the rational choice to make. If, however, the simplicity of the relations (3.3) is of prime importance then one can set, for instance, $e'_{\xi_i}=e_{\xi_i}$, for $i=1,2,3,4$ and $e'_{\xi_5}=e_{\xi_6}$.

In order to specify an SO(10)-weight vector, one has to indicate

(i) the E_6 irreducible subspace, (200000), (010000), or (000010) to which it belongs;

(ii) the SO(10) irreducible subspace to which it belongs (the possibilities for each E_6 subspace are listed in Table II);

(iii) its SO(10) weight.

In the tables, the first three lines of each column of CGC contain just these three weights. In the text, for example

$$\begin{pmatrix}000010\\00010\\001\bar{1}0\end{pmatrix}$$

denotes the SO(10)-weight vector of the weight $(001\bar{1}0)$ in the SO(10) subspace (00010) of the irreducible E_6 space (000010). An E_6 weight cannot be mistaken for an SO(10) one because it has 6 components while the SO(10) one has only five of them.

First we translate the E_6 highest weight vectors of Table III into SO(10) higher weight vectors:

$$\begin{aligned}P\begin{pmatrix}200000\\200000\end{pmatrix}&=\begin{pmatrix}00002\\00002\end{pmatrix}=P\left(\begin{pmatrix}100000\\100000\end{pmatrix}\begin{pmatrix}100000\\100000\end{pmatrix}\right)\\&=\begin{pmatrix}00001\\00001\end{pmatrix}\begin{pmatrix}00001\\00001\end{pmatrix}.\end{aligned} \tag{3.4}$$

Similarly one gets

$$\begin{aligned}P\begin{pmatrix}010000\\010000\end{pmatrix}&=\begin{pmatrix}10001\\10001\end{pmatrix}=\frac{1}{\sqrt{2}}\begin{pmatrix}00001\\00001\end{pmatrix}\begin{pmatrix}00001\\10000\end{pmatrix}\\&\quad-\frac{1}{\sqrt{2}}\begin{pmatrix}00001\\10000\end{pmatrix}\begin{pmatrix}00001\\00001\end{pmatrix},\end{aligned} \tag{3.5}$$

and the last one of them

$$P\begin{pmatrix}000010\\000010\end{pmatrix}=\begin{pmatrix}00010\\00010\end{pmatrix} \tag{3.6}$$

is given in the last column of Table III as a linear combination of the products.

Application of SO(10) generators $e_{-\xi_i}$ to (3.4), (3.5), and (3.6) produces the dominant weight vectors in these subspaces. They are listed in Table IV to VI. However, the space $27\otimes 27$ contains other subspaces irreducible with respect to SO(10) than the three above. Their highest weight vectors are then constructed to be orthogonal to the vectors of the same SO(10) weight from subspaces (3.4)–(3.6). Then again successive application of $e_{-\xi_i}$'s produces the rest of our tables.

Any other CGC than those given in Table IV is obtained by the action of operators $R_{\xi i}$ of (2.9) with the only difference that ξ_i is one of the simple roots of the subgroup.

4. CGC IN SU(5) BASIS

In the previous section the CGC's of $27\otimes 27$ were computed in the basis of SO(10) weight vectors. Here we are interested in the same tensor product of E_6 representations but we want to compute the CGC's relative to a basis of weight vectors of $SU(5)\subset SO(10)\subset E_6$. An obvious procedure is an exact analog of Sec. 3. For that, one needs only the projection matrix[7] P' for SO(10)→SU(5):

$$P'=\begin{pmatrix}1&1&0&0&0\\0&0&1&0&1\\0&0&0&1&0\\0&1&1&0&0\end{pmatrix} \tag{4.1}$$

and the basic generators of SU(5) in terms of generators of SO(10) written in a way compatable with P'. These are

$$\begin{aligned}e_{\alpha_1}&=[e_{\xi_1},e_{\xi_2}]\,,\\ e_{\alpha_2}&=[e_{\xi_3},e_{\xi_5}]\,,\\ e_{\alpha_3}&=e_{\xi_4}\,,\\ e_{\alpha_4}&=[e_{\xi_2},e_{\xi_3}]\,.\end{aligned} \tag{4.2}$$

The representation 27 of E_6 decomposes as a representation of SO(10) as follows:

$$16\oplus\,+\,10\oplus 1\ \text{ or }\ (00001)\oplus(10000)\oplus(00000)\,.$$

Therefore, in order to calculate the CGC's of $27\otimes 27$ in the SO(10) or SU(5) basis it suffices to compute the CGC's for the

TABLE IV. CGC for the E_6 product (100000) (100000) in the SO(10) subgroup basis. On the left side of each line is the product identified in terms of SO(10) weights followed by the corresponding E_6 weights.

		(200000)
		(00002)
		(00002)
(00001) (00001)	(10000) (10000)	1

		(20000)
		(20000)
		(20000)
(10000) (10000)	(1̄10000) (1̄10000)	1

		(200000)	(010000)
		(10001)	(10001)
		(10001)	(10001)
(10000) (00001)	(1̄10000) (100000)	$1/\sqrt{2}$	$1/\sqrt{2}$
(00001) (10000)	(100000) (1̄10000)	$1/\sqrt{2}$	$-1/\sqrt{2}$

		(200000)	(010000)
		(00002)	(00100)
		(00100)	(00100)
(00001) (00101̄)	(100000) (01̄1000)	$1/\sqrt{2}$	$1/\sqrt{2}$
(00101̄) (00001)	(01̄1000) (100000)	$1/\sqrt{2}$	$-1/\sqrt{2}$

		(200000)	(010000)
		(00001)	(00001)
		(00001)	(00001)
(00001) (00000)	(100000) (11̄011̄0)	$1/\sqrt{2}$	$1/\sqrt{2}$
(00000) (00001)	(11̄011̄0) (100000)	$1/\sqrt{2}$	$-1/\sqrt{2}$

		(200000)	(010000)
		(20000)	(01000)
		(01000)	(01000)
(1̄1000) (10000)	(0001̄11) (1̄10000)	$1/\sqrt{2}$	$1/\sqrt{2}$
(10000) (1̄1000)	(1̄10000) (0001̄11)	$1/\sqrt{2}$	$-1/\sqrt{2}$

		(200000)				(010000)					(000010)
		(00002)			(10000)	(00100)				(10000)	(10000)
		(10000)$_2$	(10000)$_3$	(10000)$_5$	(10000)	(10000)$_2$	(10000)$_3$	(10000)$_5$	(10000)$_4$	(10000)	(10000)
(11̄11̄0)(011̄10)	(00101̄1̄)(001̄101)	1/2	1/2		$1/\sqrt{40}$	-1/2	-1/2		-1/2		$-1/\sqrt{10}$
(011̄10)(11̄11̄0)	(001̄101)(00101̄1̄)	1/2	1/2		$1/\sqrt{40}$	1/2	1/2		1/2		$-1/\sqrt{10}$
(11̄010)(0101̄0)	(00010 1̄)(00001̄1)	1/2			$-1/\sqrt{40}$	1/2			-1/2		$1/\sqrt{10}$
(0101̄0)(11̄010)	(00001̄1)(00010 1̄)	1/2			$-1/\sqrt{40}$	-1/2			1/2		$1/\sqrt{10}$
(101̄01)(00101̄)	(011̄11̄0)(01̄1000)		1/2	1/2	$-1/\sqrt{40}$		-1/2	-1/2			$1/\sqrt{10}$
(00101̄)(101̄01)	(01̄1000)(011̄11̄0)		1/2	1/2	$-1/\sqrt{40}$		1/2	1/2			$1/\sqrt{10}$
(10001̄)(00001)	(1̄0011̄0)(100000)			-1/2	$-1/\sqrt{40}$			1/2			$1/\sqrt{10}$
(00001)(10001̄)	(100000)(1̄0011̄0)			-1/2	$-1/\sqrt{40}$			-1/2			$1/\sqrt{10}$
(10000)(00000)	(1̄10000)(11̄011̄0)				$-4/\sqrt{40}$					$1/\sqrt{2}$	$-1/\sqrt{10}$
(00000)(10000)	(11̄011̄0)(1̄10000)				$-4/\sqrt{40}$					$-1/\sqrt{2}$	$-1/\sqrt{10}$

		(200000)				(010000)					(000010)
		(10001)				(10010)				(00010)	(00010)
		(00010)$_1$	(00010)$_2$	(00010)$_3$	(00010)$_5$	(00010)$_5$	(00010)$_3$	(00010)$_2$	(00010)$_1$	(00010)	(00010)
(11̄010)(1̄1000)	(00010 1̄)(0001̄11)	1/2	1/2					-1/2	-1/2	$-1/\sqrt{10}$	$1/\sqrt{10}$
(1̄1000)(11̄010)	(0001̄11)(00010 1̄)	1/2	1/2					1/2	1/2	$1/\sqrt{10}$	$1/\sqrt{10}$
(1̄0010)(10000)	(11̄0010)(1̄10000)	1/2							-1/2	$1/\sqrt{10}$	$-1/\sqrt{10}$
(10000)(1̄0010)	(1̄10000)(11̄0010)	1/2							1/2	$-1/\sqrt{10}$	$-1/\sqrt{10}$
(01̄100)(011̄10)	(0011̄11̄)(001̄101)		1/2	1/2			1/2	1/2		$-1/\sqrt{10}$	$-1/\sqrt{10}$
(011̄10)(01̄100)	(001̄101)(0011̄11̄)		1/2	1/2			-1/2	-1/2		$1/\sqrt{10}$	$-1/\sqrt{10}$
(001̄11)(00101̄)	(011̄010)(01̄1000)			1/2	1/2	1/2	1/2			$1/\sqrt{10}$	$1/\sqrt{10}$
(00101̄)(001̄11)	(01̄1000)(011̄010)			1/2	1/2	-1/2	-1/2			$-1/\sqrt{10}$	$1/\sqrt{10}$
(00011̄)(00001)	(1̄00010)(100000)				-1/2	-1/2				$1/\sqrt{10}$	$1/\sqrt{10}$
(00001)(00011̄)	(100000)(1̄00010)				-1/2	1/2				$-1/\sqrt{10}$	$1/\sqrt{10}$

		(200000)					(010000)					(000010)
		(20000)				(00000)	(01000)					(00000)
		(00000)$_5$	(00000)$_3$	(00000)$_2$	(00000)$_1$	(00000)	(00000)$_1$	(00000)$_2$	(00000)$_3$	(00000)$_4$	(00000)$_5$	(00000)
(00011̄)(0001̄1)	(1̄00010)(0101̄00)	1/2				$1/\sqrt{10}$				1/2	1/2	
(0001̄1)(00011̄)	(0101̄00)(1̄00010)	1/2				$1/\sqrt{10}$				-1/2	-1/2	
(0011̄1̄)(001̄11)	(1̄011̄00)(011̄010)	1/2	1/2			$-1/\sqrt{10}$			-1/2	+1/2	-1/2	
(001̄11)(0011̄1̄)	(011̄010)(1̄011̄00)	1/2	1/2			$-1/\sqrt{10}$			1/2	-1/2	+1/2	
(011̄00)(01̄100)	(1̄11̄001)(0011̄11̄)		1/2	1/2		$1/\sqrt{10}$		-1/2	-1/2			
(01̄100)(011̄00)	(0011̄11̄)(1̄11̄001)		1/2	1/2		$1/\sqrt{10}$		1/2	1/2			
(11̄000)(1̄1000)	(1̄10001̄)(0001̄11)			1/2	1/2	$-1/\sqrt{10}$	-1/2	-1/2				
(1̄1000)(11̄000)	(0001̄11)(1̄10001̄)			1/2	1/2	$-1/\sqrt{10}$	1/2	1/2				
(1̄0000)(10000)	(0001̄10)(1̄10000)				1/2	$1/\sqrt{10}$	-1/2					
(10000)(1̄0000)	(1̄10000)(0001̄10)				1/2	$1/\sqrt{10}$	1/2					
(00000)(00000)	(11̄011̄0)(11̄011̄0)											1

TABLE V. CGC for the SO(10) product (00001) ⊗ (00001).

		(00002)										
		(00002)	(00002)	(00100)								
(00001)	(00001)	1	(00100)	(00100)								
(00101)	(00001)	$1/\sqrt{2}$	$1/\sqrt{2}$	(00002)			(00100)				(10000)	
(00001)	(00101)	$1/\sqrt{2}$	$-1/\sqrt{2}$	$(10000)_2$	$(10000)_3$	$(10000)_5$	$(10000)_2$	$(10000)_3$	$(10000)_5$	$(10000)_4$	(10000)	
		(01010)	(11010)	1/2			1/2			-1/2	$1/\sqrt{8}$	
		(11010)	(01010)	1/2			-1/2			1/2	$1/\sqrt{8}$	
		(11110)	(01110)	1/2	1/2		1/2	1/2		1/2	$-1/\sqrt{8}$	
		(01110)	(11110)	1/2	1/2		-1/2	-1/2		-1/2	$-1/\sqrt{8}$	
		(10101)	(00101)		1/2	1/2		1/2	1/2		$1/\sqrt{8}$	
		(00101)	(10101)		1/2	1/2		-1/2	-1/2		$1/\sqrt{8}$	
		(10001)	(00001)			1/2			1/2		$-1/\sqrt{8}$	
		(00001)	(10001)			1/2			-1/2		$-1/\sqrt{8}$	

TABLE VI. CGC for the SO(10) product (10000) ⊗ (00001).

		(10001)					
		(10001)	(10001)				(00010)
(10000)	(00001)	1	$(00010)_1$	$(00010)_2$	$(00010)_3$	$(00010)_5$	(00010)
	(10000)	(10010)	$1/\sqrt{2}$				$1/\sqrt{5}$
	(11000)	(11010)	$1/\sqrt{2}$	$1/\sqrt{2}$			$-1/\sqrt{5}$
	(01100)	(01110)		$1/\sqrt{2}$	$1/\sqrt{2}$		$1/\sqrt{5}$
	(00111)	(00101)			$1/\sqrt{2}$	$1/\sqrt{2}$	$-1/\sqrt{5}$
	(00011)	(00001)				$1/\sqrt{2}$	$1/\sqrt{5}$

following SO(10) tensor products:

$$(10000)\otimes(00001),\ (10000)\otimes(10000),\ (00001)\otimes(00001).$$

This is done in two ways: first in the SO(10) basis (Tables V–VII), using the same technique as in Sec. 2, then in an SU(5) basis as in Sec. 3 (Tables VIII–X).

Remark: A E_6-weight vector of 27 is uniquely specified by its SO(10) weight because the representations 16, 10, and 1 of SO(10) have all weights distinct.

5. EXAMPLE

Finally, let us consider an example in which we combine the results of this article together with those of Ref. 3. Consider the product X of two E_6 weight vectors

$$X=\begin{pmatrix}100000\\00\bar{1}101\end{pmatrix}\begin{pmatrix}100000\\0\bar{1}1000\end{pmatrix} \tag{5.1}$$

each from the 27 space and let us express it in terms of the SU(3)×SU(2)×U(1) weight vectors. Using Table III of Ref. 3 such a weight vector is readily rewritten as a state of two light fermions of the SU(3)×SU(2)×U(1) theory.

First we rewrite X as a linear combination of contributions X_S and X_A, respectively, from E_6 subspaces 351′ and 351. Namely,

TABLE VII. CGC for the SO(10) product (10000) ⊗ (10000).

		(20000)												
		(20000)	(20000)	(01000)										
(10000)	(10000)	1	(01000)	(01000)										
(11000)	(10000)	$1/\sqrt{2}$	$1/\sqrt{2}$	(20000)				(01000)						(00000)
(10000)	(11000)	$1/\sqrt{2}$	$-1/\sqrt{2}$	$(00000)_1$	$(00000)_2$	$(00000)_3$	$(00000)_5$	$(00000)_1$	$(00000)_2$	$(00000)_3$	$(00000)_4$	$(00000)_5$	(00000)	
(10000)	(10000)			1/2				1/2					$1/\sqrt{10}$	
(10000)	(10000)			1/2				-1/2					$1/\sqrt{10}$	
(11000)	(11000)			1/2	1/2			1/2	1/2				$-1/\sqrt{10}$	
(11000)	(11000)			1/2	1/2			-1/2	-1/2				$-1/\sqrt{10}$	
(01100)	(01100)				1/2	1/2			1/2	1/2			$1/\sqrt{10}$	
(01100)	(01100)				1/2	1/2			-1/2	-1/2			$1/\sqrt{10}$	
(00111)	(00111)					1/2	1/2			1/2	1/2	1/2	$-1/\sqrt{10}$	
(00111)	(00111)					1/2	1/2			-1/2	-1/2	-1/2	$-1/\sqrt{10}$	
(00011)	(00011)						1/2				-1/2	1/2	$1/\sqrt{10}$	
(00011)	(00011)						1/2				1/2	-1/2	$1/\sqrt{10}$	

TABLE VIII. CGC for the SO(10) product (00001) ⊗ (00001) in the SU(5) subgroup basis.

		(00002)									
		(0200)	(00002)	(00100)							
		(0200)	(0200)	(1010)							
(00001)	(00001)	1	(1010)	(1010)	(00002)	(00100)					
(0010$\bar{1}$)	(00001)		$1/\sqrt{2}$	$1/\sqrt{2}$	(0101)	(0101)					
(00001)	(0010$\bar{1}$)		$1/\sqrt{2}$	$-1/\sqrt{2}$	(0101)	(0101)	(00002)				
			(01$\bar{1}$10)	(00001)	$1/\sqrt{2}$	$1/\sqrt{2}$	(0002)	(00002)	(00100)		
			(00001)	(01$\bar{1}$10)	$1/\sqrt{2}$	$-1/\sqrt{2}$	(0002)	(0100)	(0100)		
					(01$\bar{1}$10)	(01$\bar{1}$10)	1	(0100)	(0100)	(00002)	(00100)
						(1$\bar{1}$1$\bar{1}$0)	(00001)	$1/\sqrt{2}$	$1/\sqrt{2}$	(0002)	(0010)
						(00001)	(1$\bar{1}$1$\bar{1}$0)	$1/\sqrt{2}$	$-1/\sqrt{2}$	(0010)	(0010)
								($\bar{1}$0001)	(01$\bar{1}$10)	$1/\sqrt{2}$	$1/\sqrt{2}$
								(01$\bar{1}$10)	($\bar{1}$0010)	$1/\sqrt{2}$	$-1/\sqrt{2}$
										(1$\bar{1}$1$\bar{1}$0)	(1$\bar{1}$1$\bar{1}$0)

		(00002)			(00100)				(10000)
		(0101)			(0101)			(1000)	(1000)
		$(1000)_4$	$(1000)_3$	$(1000)_2$	$(1000)_4$	$(1000)_3$	$(1000)_2$	(1000)	(1000)
(01$\bar{1}$10)	($\bar{1}$01$\bar{1}$0)	1/2			1/2			$1/\sqrt{8}$	$-1/\sqrt{8}$
($\bar{1}$01$\bar{1}$0)	(01$\bar{1}$10)	1/2			-1/2			$-1/\sqrt{8}$	$-1/\sqrt{8}$
($\bar{1}$0010)	(010$\bar{1}$0)	1/2	-1/2		1/2	-1/2		$-1/\sqrt{8}$	$1/\sqrt{8}$
(010$\bar{1}$0)	($\bar{1}$0010)	1/2	-1/2		-1/2	1/2		$1/\sqrt{8}$	$1/\sqrt{8}$
($\bar{1}$1$\bar{1}$01)	(0010$\bar{1}$)		1/2	1/2		1/2	1/2	$-1/\sqrt{8}$	$1/\sqrt{8}$
(0010$\bar{1}$)	($\bar{1}$1$\bar{1}$01)		1/2	1/2		-1/2	-1/2	$1/\sqrt{8}$	$1/\sqrt{8}$
($\bar{1}$100$\bar{1}$)	(00001)			1/2			1/2	$1/\sqrt{8}$	$-1/\sqrt{8}$
(00001)	($\bar{1}$100$\bar{1}$)			1/2			-1/2	$-1/\sqrt{8}$	$-1/\sqrt{8}$

		(00002)			(00100)				(10000)
		(0200)		(0001)	(1010)			(0001)	(0001)
		$(0001)_3$	$(0001)_2$	(0001)	$(0001)_3$	$(0001)_2$	$(0001)_1$	(0001)	(0001)
(1$\bar{1}$010)	(010$\bar{1}$0)	-1/2		$1/\sqrt{24}$	-1/2		-1/2		$1/\sqrt{8}$
(010$\bar{1}$0)	(1$\bar{1}$010)	-1/2		$1/\sqrt{24}$	1/2		1/2		$1/\sqrt{8}$
(10$\bar{1}$01)	(0010$\bar{1}$)	1/2	1/2	$1/\sqrt{24}$	1/2	1/2	-1/2		$1/\sqrt{8}$
(0010$\bar{1}$)	(10$\bar{1}$01)	1/2	1/2	$1/\sqrt{24}$	-1/2	-1/2	1/2		$1/\sqrt{8}$
(1000$\bar{1}$)	(00001)		1/2	$-1/\sqrt{24}$		1/2			$-1/\sqrt{8}$
(00001)	(1000$\bar{1}$)		1/2	$-1/\sqrt{24}$		-1/2			$-1/\sqrt{8}$
(1$\bar{1}$1$\bar{1}$0)	(01$\bar{1}$10)			$\sqrt{3}/\sqrt{8}$				$1/\sqrt{2}$	$-1/\sqrt{8}$
(01$\bar{1}$10)	(1$\bar{1}$1$\bar{1}$0)			$\sqrt{3}/\sqrt{8}$				$-1/\sqrt{2}$	$-1/\sqrt{8}$

TABLE IX. CGC for the SO(10) product (10000) ⊗ (00001) in the SU(5) subgroup basis.

		(10001)
		(0101)
		(0101)
(10000)	(00001)	1

		(10001)
		(1100)
		(1100)
$(\bar{1}1000)$	(00001)	1

		(10001)
		(0002)
		(0002)
(10000)	$(01\bar{1}10)$	1

		(10001)
		(1001)
		(1001)
$(\bar{1}1000)$	$(01\bar{1}10)$	1

		(10001)				(00010)
		(0101)			(1000)	(1000)
		$(1000)_4$	$(1000)_3$	$(1000)_2$	(1000)	(1000)
(10000)	$(\bar{1}01\bar{1}0)$	$1/\sqrt{2}$			$1/2\sqrt{5}$	$1/\sqrt{5}$
$(0\bar{1}100)$	$(010\bar{1}0)$	$-1/\sqrt{2}$	$-1/\sqrt{2}$		$1/2\sqrt{5}$	$1/\sqrt{5}$
$(000\bar{1}1)$	$(0010\bar{1})$		$-1/\sqrt{2}$	$-1/\sqrt{2}$	$-1/2\sqrt{5}$	$-1/\sqrt{5}$
$(001\bar{1}\bar{1})$	(00001)			$1/\sqrt{2}$	$1/2\sqrt{5}$	$1/\sqrt{5}$
$(\bar{1}1000)$	$(\bar{1}11\bar{1}0)$				$2/\sqrt{5}$	$-1/\sqrt{5}$

		(10001)				(00010)
		(1100)		(0002)	(0010)	(0010)
		$(0010)_2$	$(0010)_1$	(0010)	(0010)	(0010)
$(\bar{1}1000)$	$(1\bar{1}010)$	$1/\sqrt{2}$				$-1/\sqrt{5}$
$(00\bar{1}11)$	$(0010\bar{1})$	$-1/\sqrt{2}$	$1/\sqrt{2}$			$-1/\sqrt{5}$
$(0001\bar{1})$	(00001)		$1/\sqrt{2}$			$1/\sqrt{5}$
(10000)	$(\bar{1}0010)$			$1/\sqrt{2}$	$1/\sqrt{2}$	$1/\sqrt{5}$
$(0\bar{1}100)$	$(01\bar{1}10)$			$1/\sqrt{2}$	$-1/\sqrt{2}$	$1/\sqrt{5}$

		(10001)				(00010)
		(1001)				(0000)
		$(0000)_4$	$(0000)_3$	$(0000)_2$	$(0000)_1$	(0000)
$(\bar{1}0000)$	$(01\bar{1}10)$	$1/\sqrt{2}$				$1/\sqrt{5}$
$(01\bar{1}00)$	$(\bar{1}0010)$	$1/\sqrt{2}$	$1/\sqrt{2}$			$-1/\sqrt{5}$
$(0001\bar{1})$	$(\bar{1}1\bar{1}01)$		$1/\sqrt{2}$	$1/\sqrt{2}$		$1/\sqrt{5}$
$(00\bar{1}11)$	$(\bar{1}100\bar{1})$			$1/\sqrt{2}$	$-1/\sqrt{2}$	$-1/\sqrt{5}$
$(\bar{1}1000)$	$(00\bar{1}10)$				$-1/\sqrt{2}$	$1/\sqrt{5}$

		(10010)
		(0001)
		(0001)
(10000)	$(1\bar{1}1\bar{1}0)$	1

TABLE X. CGC for the SO(10) product (10000) ⊗ (10000) in the SU(5) subgroup basis.

	(20000)
	(0002)
	(0002)
$(10000)(10000)$	1

	(20000)	(01000)
	(0002)	(0010)
	(0010)	(0010)
$(0\bar{1}100)(10000)$	$1/\sqrt{2}$	$1/\sqrt{2}$
$(10000)(0\bar{1}100)$	$1/\sqrt{2}$	$-1/\sqrt{2}$

	(20000)	(01000)
	(1001)	(1001)
	(1001)	(1001)
$(\bar{1}1000)(10000)$	$1/\sqrt{2}$	$1/\sqrt{2}$
$(10000)(\bar{1}1000)$	$1/\sqrt{2}$	$-1/\sqrt{2}$

	(20000)
	(2000)
	(2000)
$(\bar{1}1000)(\bar{1}1000)$	1

	(20000)	(01000)
	(2000)	(0100)
	(0100)	(0100)
$(00\bar{1}11)(\bar{1}1000)$	$1/\sqrt{2}$	$1/\sqrt{2}$
$(\bar{1}1000)(00\bar{1}11)$	$1/\sqrt{2}$	$-1/\sqrt{2}$

	(20000)				(01000)					(00000)
	(1001)				(1001)				(0000)	(0000)
	$(0000)_4$	$(0000)_3$	$(0000)_2$	$(0000)_1$	$(0000)_4$	$(0000)_3$	$(0000)_2$	$(0000)_1$	(0000)	(0000)
$(\bar{1}0000)(10000)$	$-1/2$				$-1/2$				$1/\sqrt{10}$	$1/\sqrt{10}$
$(10000)(\bar{1}0000)$	$-1/2$				$1/2$				$-1/\sqrt{10}$	$1/\sqrt{10}$
$(01\bar{1}00)(0\bar{1}100)$	$1/2$	$1/2$			$1/2$	$1/2$			$1/\sqrt{10}$	$1/\sqrt{10}$
$(0\bar{1}100)(01\bar{1}00)$	$1/2$	$1/2$			$-1/2$	$-1/2$			$-1/\sqrt{10}$	$1/\sqrt{10}$
$(0001\bar{1})(000\bar{1}1)$		$-1/2$	$-1/2$			$-1/2$	$-1/2$		$1/\sqrt{10}$	$1/\sqrt{10}$
$(000\bar{1}1)(0001\bar{1})$		$-1/2$	$-1/2$			$1/2$	$1/2$		$-1/\sqrt{10}$	$1/\sqrt{10}$
$(00\bar{1}11)(001\bar{1}\bar{1})$			$-1/2$	$-1/2$			$-1/2$	$-1/2$	$-1/\sqrt{10}$	$-1/\sqrt{10}$
$(001\bar{1}\bar{1})(00\bar{1}11)$			$-1/2$	$-1/2$			$1/2$	$1/2$	$1/\sqrt{10}$	$-1/\sqrt{10}$
$(\bar{1}1000)(1\bar{1}000)$				$1/2$				$1/2$	$-1/\sqrt{10}$	$-1/\sqrt{10}$
$(1\bar{1}000)(\bar{1}1000)$				$1/2$				$-1/2$	$1/\sqrt{10}$	$-1/\sqrt{10}$

$$X=\frac{1}{\sqrt{2}}X_S+\frac{1}{\sqrt{2}}X_A\,, \tag{5.2}$$

where

$$X_S=\frac{1}{\sqrt{2}}\begin{pmatrix}100000\\00\bar{1}101\end{pmatrix}\begin{pmatrix}100000\\0\bar{1}1000\end{pmatrix}+\frac{1}{\sqrt{2}}\begin{pmatrix}100000\\0\bar{1}1000\end{pmatrix}\begin{pmatrix}100000\\00\bar{1}101\end{pmatrix}, \tag{5.3}$$

$$X_A=\frac{1}{\sqrt{2}}\begin{pmatrix}100000\\00\bar{1}101\end{pmatrix}\begin{pmatrix}100000\\0\bar{1}1000\end{pmatrix}-\frac{1}{\sqrt{2}}\begin{pmatrix}100000\\0\bar{1}1000\end{pmatrix}\begin{pmatrix}100000\\00\bar{1}101\end{pmatrix}.$$

Since the weight $(0\bar{1}0101)$ is unique in (200000) and in (010000) we have

$$X_S=\begin{pmatrix}200000\\0\bar{1}0101\end{pmatrix}\quad\text{and}\quad X_A=\begin{pmatrix}010000\\0\bar{1}0101\end{pmatrix}. \tag{5.4}$$

The E_6 weight $(0\bar{1}0101)$ is reinterpreted as the SO(10) weight $(0101\bar{1})$ by means of (3.1). Then applying the SO(10) reflections $r_{\xi_4}r_{\xi_3}$ to $(0101\bar{1})$, we find that it belongs to the orbit (00100) or the SO(10) Weyl group. Hence, X_S belongs to the SO(10) irreducible subspace (00002) and X_A belongs to (00100), i.e.,

$$X_S=\begin{pmatrix}200000\\00002\\0101\bar{1}\end{pmatrix}\quad\text{and}\quad X_A=\begin{pmatrix}010000\\00100\\0101\bar{1}\end{pmatrix}. \tag{5.5}$$

Next we reduce the SO(10) weight of X_S and X_A to the SU(5) weight $(1\bar{1}11)$ using (4.1). This weight belongs to the SU(5) Weyl group orbit (0101). Indeed,

$$(0101)=r_{\alpha_2}(1\bar{1}11)\,.$$

Therefore, X_S and X_A can be written as the SU(5)-weight vectors

$$X_S=\begin{pmatrix}200000\\00002\\0101\\1\bar{1}11\end{pmatrix}\quad\text{and}\quad X_A=\begin{pmatrix}010000\\00100\\0101\\1\bar{1}11\end{pmatrix}.$$

Finally, we reduce $(1\bar{1}11)$ to $SU(3)\times SU(2)\times U(1)$ weight [Eq. (10) of Ref. 4] and get $(02)(0)(\bar{2})$ which is the highest weight vector of one of the irreducible $SU(3)\times SU(2)\times U(1)$ subspaces in (0101) of SU(5) [cf. Table 2 of Ref. 4]. Thus one has

$$X_S=\begin{pmatrix}200000\\00002\\0101\\(02)(0)(\bar{2})\\(02)(0)(\bar{2})\end{pmatrix},\quad X_A=\begin{pmatrix}010000\\00100\\0101\\(02)(0)(\bar{2})\\(02)(0)(\bar{2})\end{pmatrix}.$$

Similarly one can reinterpret the E_6-weight vectors $\binom{100000}{00\bar{1}101}$ and $\binom{100000}{0\bar{1}1000}$ in (5.1) as the SO(10)-weight vectors

$$\begin{pmatrix}100000\\00001\\01\bar{1}10\end{pmatrix}\quad\text{and}\quad\begin{pmatrix}100000\\00001\\0010\bar{1}\end{pmatrix}, \tag{5.6}$$

then as the SU(5)-weight vectors

$$\begin{pmatrix}100000\\00001\\0001\\0001\end{pmatrix}\quad\text{and}\quad\begin{pmatrix}100000\\00001\\0100\\0\bar{1}10\end{pmatrix}, \tag{5.7}$$

and finally as the dominant $SU(3)\times SU(2)\times U(1)$ weight vectors

$$K=\begin{pmatrix}100000\\00001\\0001\\(01)(0)(2)\\(01)(0)(2)\end{pmatrix},\quad L=\begin{pmatrix}100000\\00001\\0100\\(01)(0)(4)\\(01)(0)(4)\end{pmatrix}.$$

Let us recall that the top line in K and L gives the E_6 irreducible subspace, the second line gives the corresponding SO(10) subspace, the third line indicates the SU(5) subspace, and the fourth line gives the $SU(3)\times SU(2)\times U(1)$ subspace. The bottom line gives the actual $SU(3)\times SU(2)\times U(1)$ weight. It happens to also be the highest weight in our example.

ACKNOWLEDGMENTS

One of us (J.P.) is grateful for the hospitality of the Aspen Center for Physics, where the present work originated three years ago.

We are also grateful to Dr. J. Bagger for independent verification and some corrections of the tables.

[1]H. Georgi, *Particles and Fields* (1974) (APS/DPF Williamsburg), edited by C. E. Carlson (Amer. Inst. Phys., New York, 1975); H. Fritzsch and P. Minkowski, Ann. Phys. (N.Y.) **93**, 193 (1975); F. Gursey, P. Ramond, and P. Sikivie, Phys. Lett. B **60**, 177 (1975); R. C. King, Nucl. Phys. B **185**, 133 (1980); S. Okubo, Hadronic J. **5**, 7 (1981); Q. Feldman, T. Fulton, and P. T. Matthews, J. Phys. C: Nucl. Phys. **8**, 295 (1982); Y. Tosa and S. Okubo, Phys. Rev. D **23**, 3058 (1981); S. Okubo, preprint UR-820, University of Rochester, 1982; J. Harvey, D. Reiss, and P. Ramond, Nucl. Phys. B **199**, 223 (1982).

[2]M. Gell-Mann, P. Ramond, and R. Slansky, Rev. Mod. Phys. **50**, 721 (1978); R. Slansky, Phys. Rep. **79**, 1 (1981).

[3]In-Gyu Koh, J. Patera, and C. Rousseau, "Clebsch–Gordan coefficients for $SU(5)\supset SU(3)\times SU(2)\times U(1)$ theories," J. Math. Phys. **24**, 1955 (1983).

[4]J. Patera, J. Math. Phys. **11**, 3027 (1970); **12**, 384 (1971).

[5]R. V. Moody and J. Patera, J. Math. Phys. **25**, 2838 (1984).

[6]For instance, Supplement to E. B. Dynkin, Am. Math. Soc. Transl. Ser. 2, **2**, 245 (1957).

[7]W. McKay, J. Patera, and D. Sankoff, in *Computers in Nonassociative Rings and Algebras*, edited by R. Beck and B. Kolman (Academic, New York, 1977); cf. Table IV.

On the Fourier series representations of path integrals

Antoine Royer
Centre de Recherche de Mathématiques Appliquées, Université de Montréal, Montréal, Québec H3C 3J7, Canada

(Received 18 October 1983; accepted for publication 4 May 1984)

The direct transformation from the polygonal to the Fourier series representation of Feynman path integrals, via a change of integration variables, is effected explicitly for cases where the Lagrangian is of the form $L(x,\dot{x},t) = \frac{1}{2}\dot{x}^2 - V(x,t)$. This transformation involves a "functional Jacobian" stemming solely from the velocity term in the Lagrangian; this is because N-segment polygonal paths of the type contributing significantly to the integral, and their N-term Fourier series approximants, coalesce together as $N\to\infty$, but not their derivatives. We also consider integrals over paths with fixed means $T^{-1}\int_0^T dt\, x(t) = \text{const}$. The usefulness of the Fourier representation is illustrated with the harmonic oscillator case $V(x) = \frac{1}{2}\omega^2 x^2$, in both the free and fixed means situations; in particular, the Fourier evaluation of the path integrals trivially determines the large time phases (Maslov indices), and the ranges of ω^2 values for which the integrals are finite or infinite in the imaginary time cases.

PACS numbers: 02.30. + g, 03.65.Db, 03.70. + k, 04.60. + n

I. INTRODUCTION

Path integrals were initially defined in terms of polygonal paths,[1,2] i.e., as limits of discretized expressions. Soon after Feynman's introduction into physics of such path integrals,[2] Davison[3] proposed Fourier series (or more generally, complete sets of orthonormal functions) as convenient parametrizations of paths; the Fourier series path integral representation of the Feynman propagator (or Green's function) was written down in a rather obvious manner, and proven to be equivalent to the polygonal path integral representation when the Lagrangian is of the form

$$L(x,\dot{x},t) = \tfrac{1}{2}\dot{x}^2 - V(x,t), \tag{1.1}$$

where V is independent of $\dot{x}$. This was done by showing that the Fourier representation obeys the requisite composition law for propagators, and that it is identical to the polygonal representation in the small time limit.[3]

The Fourier series representation allows an efficient evaluation of the path integral in those cases where it can be evaluated in closed form,[4–6] i.e., quadratic potentials, and yields useful approximations and methods of numerical evaluation in the general case (1.1).[7]

The Fourier series representation has also been used in more general situations than (1.1)[4,8]; this however has drawn some criticism,[9,10] for it is suspected that when the Lagrangian contains terms like $x\dot{x}$, e.g., then the polygonal and Fourier series representations of a path integral may be different objects (this may be inferred, in particular, from the fact that the result of the polygonal path evaluation depends itself on the choice of "discretization procedure," or correspondence rule[11]). Unfortunately, Davison's proof[3] offers little insight as to why the two representations are equivalent in the case (1.1), but not necessarily otherwise.

In their book on path integrals, Feynman and Hibbs[12] use the Fourier representation to evaluate the propagator for the harmonic oscillator of frequency ω. They consider the passage from the polygonal to the Fourier representation as a linear transformation, and make mention of its "Jacobian" J; however, they bypass the direct evaluation of J, by fixing the final normalization by comparison with the known value in the case $\omega = 0$, on the "evident" presumption that J must be independent of ω.

A direct passage from the polygonal to the Fourier representation, by way of an explicit change of integration variables, was apparently first done by Chang,[13] for the case of the harmonic oscillator; specifically, the transformation is from N-segment polygonal paths to N-term Fourier series, with the limit $N\to\infty$ taken concurrently in both representations.

In this paper, we make the direct passage from the polygonal to the Fourier representation for cases where the Lagrangian is of the general form (1.1). This provides an alternative proof of the equivalence of the two representations, and some further insight. The transformation from one representation to the other involves what might be better called a "functional Jacobian." The value of this Jacobian is tied with the velocity term in the Lagrangian, which is perhaps not surprising when one recalls that that term defines the Wiener measure for paths.[1] The more specific reason is that, whereas N-segment polygonal paths and their N-term Fourier series approximants approach one another as $N\to\infty$, for those paths of the type contributing significantly to the integral, their time derivatives on the contrary diverge from one another (clearly implying that the polygonal and Fourier representations of a path integral may indeed be different objects when V is velocity dependent).

We also consider integrals over paths $x(t)$, $0\leqslant t\leqslant T$, constrained to having a fixed mean value $T^{-1}\int_0^T dt\, x(t)$.[14–16] Davison's proof[3] of the equivalence of the Fourier and polygonal representations does not apply to such *fixed means* path integrals, and it was the need to evaluate such integrals[16] which provided the initial motivation for the present study.

We illustrate the convenience of the Fourier representation by treating the harmonic oscillator case $V(x) = \frac{1}{2}\omega^2 x^2$, in both the free and fixed means situations. In particular, the Fourier series evaluation of the path integrals allows to very simply determine the large time phases (Maslov indices) and

 0022-2488/84/102873-12$02.50

the ranges of ω^2 values for which the propagators are finite or infinite in the imaginary time cases; these ranges do not seem to be so easily deducible otherwise.

In Sec. II, the basic expressions and definitions we shall use are written down, and the class of paths which contribute significantly to the integrals are identified. In Sec. III, we show that the N-segment polygonal paths belonging to the above class, and their N-term Fourier approximants, approach one another as $N\to\infty$. In Sec. IV, the change of integration path variables is effected. In Sec. V, we consider integrals over paths with constrained means. Section VI deals with the harmonic oscillator. We end with a short discussion in Sec. VII. Several appendixes contain the more technical details.

II. BASIC EXPRESSIONS

We consider path integrals of the form

$$K_\theta(x_a,t_a;x_b,t_b)=\int_{x_a,t_a}^{x_b,t_b}\mathscr{D}x(t)e^{-\theta S[x(t)]}, \tag{2.1}$$

where the functional

$$\begin{aligned}S[x(t)]&=\int_{t_a}^{t_b}dt\left[\frac{1}{2}\dot{x}(t)^2-V(x(t),t)\right]\\&\equiv S_{\rm vel}[x(t)]+S_{\rm pot}[x(t)].\end{aligned} \tag{2.2}$$

The symbol θ stands for either 1 or $i=(-1)^{1/2}$, and the integral is over all paths $x(t)$, $t_a\leqslant t\leqslant t_b$, with fixed endpoints $x(t_a)=x_a, x(t_b)=x_b$. Note that in cases where $V(x,t)=V(x)$ is independent of time,

$$(2.1)=\langle x_b|e^{-\theta(t_b-t_a)H_\theta}|x_a\rangle,$$

where $H_\theta=\frac{1}{2}d^2/dx^2-\theta^2V(x)$, i.e., $K_{\theta=i}$ is a real time propagator, while $K_{\theta=1}$ is a thermal, or imaginary time, propagator (with potential $-V$).

The explicit meaning of (2.1) is as follows[10,12]: Divide the time interval $[t_a,t_b]$ into N equal parts, and let

$$\begin{aligned}&t_0=t_a,\ x_0=x_a;\quad t_N=t_b,\quad x_N=x_b,\\&t_j=t_0+j\epsilon,\quad -\infty<x_j<\infty,\quad j=1,\ldots,N-1,\end{aligned} \tag{2.3}$$

where

$$\epsilon=(t_b-t_a)/N. \tag{2.4}$$

Define a N-segment *polygonal* path

$$x_P(t;x_1,x_2,\ldots,x_{N-1}) \tag{2.5}$$

as the succession of N straight segments $[(x_{j-1},t_{j-1}),(x_j,t_j)]$, $j=1,\ldots,N$, in the (x,t)-plane. The path integral (2.1) is then

$$\begin{aligned}&K_\theta(x_a,t_a;x_b,t_b)\\&\quad=\lim_{N\to\infty}(2\pi\epsilon/\theta)^{-N/2}\int_{-\infty}^{\infty}dx_1\,dx_2\cdots dx_{N-1}\\&\qquad\times\exp\{-\theta S[x_P(t;x_1\cdots x_{N-1})]\}.\end{aligned} \tag{2.6}$$

In $S[x_P(t;x_1\ldots x_{N-1})]$, the velocity term is evidently

$$S_{\rm vel}[x_P]=\frac{1}{2}\epsilon^{-1}\sum_{j=1}^{N}(x_j-x_{j-1})^2, \tag{2.7}$$

while the potential term is taken as the Riemann sum

$$S_{\rm pot}[x_P]=\epsilon\sum_{j=1}^{N}V(x_j,t_j). \tag{2.8}$$

The factor $(2\pi\epsilon/\theta)^{-N/2}$ in (2.6) is such as to guarantee a sensible value in the limit $N\to\infty$, the combination $(2\pi\epsilon/\theta)^{-1/2}\exp[-\frac{1}{2}\theta\epsilon^{-1}(x_j-x_{j-1})^2]$ tending to the δ-function $\delta(x_j-x_{j-1})$ as $N\to\infty$.

Because of the ϵ^{-1} in (2.7), it is the velocity term $S_{\rm vel}$ which determines the class of paths which contribute significantly to the integral (2.6) as $N\to\infty$. Given any real number α, let us call "α-path," and sometimes identify with a superscript α, a path $x_P(t;x_1\ldots x_{N-1})$ for which

$$|x_j-x_{j-1}|\leqslant\epsilon^\alpha\quad\text{for all } j=1,\ldots,N. \tag{2.9}$$

We now show that as $N\to\infty$, it is only α-paths with

$$\alpha<\tfrac{1}{2} \tag{2.10}$$

which contribute to the integral (2.6). Partition the integration volume $\{-\infty<x_j<\infty, j=1,\ldots,N-1\}$ into two parts: a part Ω_α within which (2.9) is obeyed for some real α, and the remainder Ω'_α. Within Ω'_α, (2.9) is violated for at least one value of j, so that

$$S_{\rm vel}[x_P]>\tfrac{1}{2}\epsilon^{-1}(\epsilon^\alpha)^2=\tfrac{1}{2}\epsilon^{2\alpha-1},\quad x_P\in\Omega'_\alpha. \tag{2.11}$$

Thus, the part of the integral (2.6) over Ω'_α is of order $\exp(-\text{const}\times N^{1-2\alpha})$ when $\theta=1$, and is negligible as $N\to\infty$ if α satisfies (2.10); it is also negligible when $\theta=i$, because of the rapid oscillations of the imaginary exponential inside Ω'_α [the oscillation wavelength $\epsilon|x_j-x_{j-1}|^{-1}<\epsilon^{1-\alpha}\to 0$ as $N\to\infty$ for at least one value of j]. We may thus replace (2.6) by

$$\begin{aligned}K&=\lim_{N\to\infty}\left(\frac{2\pi\epsilon}{\theta}\right)^{-N/2}\\&\quad\times\int_{\Omega_\alpha}dx_1\cdots dx_{N-1}\,e^{-\theta S[x_P]},\quad \alpha<\frac{1}{2}.\end{aligned} \tag{2.12}$$

In view of the above, it is usual to let the velocity term define a "measure" for paths,[17] i.e., to rewrite (2.1) as

$$\int d_\theta x\exp\left[\theta\int_{t_a}^{t_b}dt\,V(x(t),t)\right]. \tag{2.13}$$

As is well known,[12] one may express (2.1) as

$$K_\theta(x_a,t_a;x_b,t_b)=e^{-\theta S[x_{\rm cl}(t)]}\hat{K}_\theta(0,t_a;0,t_b), \tag{2.14}$$

where $x_{\rm cl}(t)$ is the classical trajectory between (x_a,t_a) and (x_b,t_b), at which the action is stationary, and $\hat{K}_\theta$ is given by (2.1), but with the potential V replaced by $(V'\equiv dV/dx)$[18]

$$\hat{V}(x,t)\equiv V(x_{\rm cl}(t)+x,t)-V(x_{\rm cl}(t),t)-xV'(x_{\rm cl}(t),t). \tag{2.15}$$

We may accordingly limit our considerations to the path integral

$$K_\theta\equiv K_\theta(0,0;0,1)=\int_{0,0}^{0,1}\mathscr{D}x(t)e^{-\theta S[x(t)]}. \tag{2.16}$$

Henceforth, it will be understood that

$$t_a=0,\ t_b=1;\ x_a=x(0)=x_b=x(1)=0, \tag{2.17}$$

whence, also,

$$\epsilon=1/N. \tag{2.17'}$$

An alternative, nondiscretized, representation of (2.1) is obtained by expressing the paths $x(t)$ as Fourier series, e.g.,[6] [for the case (2.17)]

$$x_F(t;b_1,b_2,\ldots)=\sum_{n=1}^{\infty}b_n\sin(\pi nt). \tag{2.18}$$

Then

$$K_\theta = \left(\frac{\theta}{2\pi}\right)^{1/2} \int_{-\infty}^{\infty} \frac{db_1}{\gamma_1}\frac{db_2}{\gamma_2} \cdots \exp\{-\theta S[x_F(t;b_1,b_2,\ldots)]\}, \tag{2.19}$$

with

$$\gamma_n = 2n^{-1}(\pi\theta)^{-1/2}.$$

Again, it is the velocity term

$$S_{\text{vel}}[x_F] = \left(\frac{1}{4}\right)\sum_{n=1}^{\infty} \pi^2 n^2 b_n^2 \tag{2.20}$$

that decides which frequency components in (2.18) contribute significantly to (2.19): indeed, the combination

$$\gamma_n^{-1} e^{-\theta(1/4)\pi^2 n^2 b_n^2} \to \delta(b_n) \text{ as } n\to\infty, \tag{2.21}$$

so the high-frequency components of (2.18) (on the scale of variation of $e^{-\theta\int V\,dt}$) only contribute factors 1 to (2.19); the requirement that this be so, for (2.19) to make sense, fixes the normalizers γ_n.

It is not really obvious that (2.19) is equivalent to the polygonal representation (2.6). Davison[3] demonstrated that equivalence, for the case of a slightly different Fourier representation, wherein it is $\dot{x}(t)$ which is expanded like (2.18), by showing that, like (2.6), the Fourier representation satisfies the semigroup law

$$K(x_a,t_a;x_b,t_b) = \int_{-\infty}^{\infty} dx\, K(x_a,t_a;x,t)K(x,t;x_b,t_b), \tag{2.22}$$

and that it is identical to (2.6) in the small time limit $t_b - t_a \to 0$. The equivalence of (2.19) with (2.6) will be shown more directly in Sec. IV, by explicitly transforming from (2.6) to (2.19) by way of a simple change of integration variables.

The Fourier parametrization (2.18) in terms of a pure sine series is the simplest, but it is not the only one possible. Mixed sine–cosine series may also be used, e.g. [for the case (2.17)],

$$x_F(t;\{a_n,b_n\}) = a_0 + \sum_{n=1}^{\infty}[a_n\cos(2\pi nt) + b_n\sin(2\pi nt)]. \tag{2.23}$$

Here, the endpoint constraints $x(0) = x(1) = 0$ become the condition

$$\sum_{n=0}^{\infty} a_n = 0, \tag{2.24}$$

which may be guaranteed by the insertion of an appropriate δ-function. We have

$$K_\theta = 2^{-1/2}\pi^{-1}\int_{-\infty}^{\infty} \frac{da_0}{\eta_0}\frac{da_1}{\eta_1}\cdots\frac{db_1}{\eta_1}\frac{db_2}{\eta_2} \cdots\delta\left(\sum_{n=0}^{\infty} a_n\right)\exp\{-\theta S[x_F(t;\{a_n,b_n\})]\}, \tag{2.25}$$

where the normalizers

$$\eta_0 = (\pi\theta)^{-1/2};\quad \eta_n = n^{-1}(\pi\theta)^{-1/2},\quad n\geqslant 1 \tag{2.26}$$

may again be guessed by inspection of the velocity term

$$S_{\text{vel}}[x_F(t;\{a_n,b_n\})] = \sum_{n=1}^{\infty}\pi^2 n^2(a_n^2 + b_n^2) \tag{2.27}$$

and the demand that

$$\eta_n^{-2}\exp[-\theta\pi^2 n^2(a_n^2 + b_n^2)] \to \delta(a_n)\delta(b_n) \text{ as } n\to\infty. \tag{2.28}$$

For integrals of the type (2.16), the pure sine series (2.18) is the more convenient; but when additional constraints are imposed on the paths, such as, e.g., $\int_0^1 dt\, x(t) = 0$, then other forms such as (2.23) may be more useful, as will be discussed in Sec. VI.

III. APPROXIMATING A N-SEGMENT POLYGONAL PATH BY A N-TERM FOURIER SERIES

A polygonal path $x_P(t;x_1\ldots x_{N-1})$ can always be expressed as an infinite Fourier series. However, in order to pass from the polygonal representation (2.6) to the Fourier representation (2.19) via a change of integration variables, we must pass from N-segment polygons to N-term Fourier series, such as to preserve the number of integration variables; the limit $N\to\infty$ is thus taken concurrently in both representations. The question then arises as to how closely is a N-segment polygonal path represented by a N-term Fourier series, and does the latter approach the former as $N\to\infty$ simultaneously in both quantities. We will see that the answer to the last question is no in general, but yes in the case of paths obeying the restriction (2.9) with $\alpha > 0$; however, even in the latter case when $\alpha < 1$, the derivatives of the polygonal and Fourier paths do not approach each other, but on the contrary diverge from each other as $N\to\infty$ (see Ref. 19).

Let us view the polygonal path $x_P(t;x_1\ldots x_{N-1})$, $0\leqslant t\leqslant 1$, with null endpoints $x_0 = x_N = 0$, as defined by a point or vector $(x_1,\ldots,x_{N-1})$ in some $(N-1)$-dimensional space. Other parametrizations of x_P may be obtained by effecting coordinate transformations in that space. If the vector $(a_1,\ldots,a_{N-1})$ is related to $(x_1,\ldots,x_{N-1})$ by some such transformation M,

$$a = M^{-1}x,\quad x = Ma, \tag{3.1}$$

we write

$$x_P(t;x_1,\ldots,x_{N-1}) = x_P(t;(Ma)_1,\ldots,(Ma)_{N-1}) \equiv x_P(t;a_1,\ldots,a_{N-1}). \tag{3.2}$$

Specifically, we shall require the following rotation:

$$x_i = \left(\frac{2}{N}\right)^{1/2}\sum_{n=1}^{N-1} a_n\sin\left(\frac{\pi in}{N}\right),\quad i = 1,\ldots,N-1, \tag{3.3}$$

$$a_n = \left(\frac{2}{N}\right)^{1/2}\sum_{i=1}^{N-1} x_i\sin\left(\frac{\pi ni}{N}\right),\quad n = 1,\ldots,N-1, \tag{3.3'}$$

where the matrix

$$M_{in} = (2/N)^{1/2}\sin(\pi in/N),\quad i = 1,\ldots,N-1,\quad n = 1,\ldots,N-1 \tag{3.3''}$$

is orthonormal [see Eq. (D3b) of Appendix D]:

$$(MM^T)_{in} = \left(\frac{2}{N}\right)\sum_{j=1}^{N-1}\sin\left(\frac{\pi ij}{N}\right)\sin\left(\frac{\pi jn}{N}\right) = \delta_{in}. \tag{3.3'''}$$

Thus, the Jacobian of the transformation (3.3) is 1.

Given now a point $(a_1,...,a_{N-1})$ in our $(N-1)$-dimensional space, we associate with it the following smooth curve in the form of a $(N-1)$ term Fourier series:

$$x_F(t;a_1,...,a_{N-1}) \equiv \left(\frac{2}{N}\right)^{1/2} \sum_{n=1}^{N-1} a_n \sin(\pi n t). \tag{3.4}$$

Again, relating the a_n's to x_i's by the transformation (3.3), we write

$$\begin{aligned} x_F(t;a_1,...,a_{N-1}) &= x_F(t;(M^T x)_1,...,(M^T x)_{N-1}) \equiv x_F(t;x_1,...,x_{N-1}) \\ &= (2N)^{-1} \sin(\pi\tau) \sum_{i=1}^{N-1} x_i (-)^{i-j} \frac{\sin(\pi i/N)}{\sin[\frac{1}{2}\pi(j+i+\tau)/N]\sin[\frac{1}{2}\pi(j-i+\tau)/N]}. \end{aligned} \tag{3.5}$$

In the last line, derived in Appendix A, j is any integer between 0 and N, and τ is defined by [note that $t_j = j/N$ in view of (2.17)]

$$t = t_j + \tau/N = (j+\tau)/N. \tag{3.6}$$

By construction, the curve $x_F(t;x_1,...,x_{N-1})$ passes through the points (x_j,t_j), $j=0,...,N$, $x_0 = x_N = 0$, in the (x,t)-plane [indeed, letting $\tau \to 0$ in (3.5), one obtains zero for all the terms of the sum over i, except the term $i=j$ which yields x_j]. However, in between these points, $x_F(t;x_1 \cdots x_{N-1})$ does not reproduce the straight segments of the polygonal path $x_P(t;x_1,...,x_{N-1})$, and may, in fact, differ from it by any arbitrary amount, whatever the value of N. [Let, e.g., all the x_i be zero, except x_j; then[20]

$$x_F(t;0,...,0,x_j,0,...,0) \cong x_j \frac{\sin(\pi\tau)}{\pi\tau} \quad \text{for } \frac{\tau}{N} \ll 1 \tag{3.7}$$

(see Fig. 1); by letting x_j be sufficiently large, one can have x_F differ from x_P by any amount at any point t outside the set $\{t_j, j=0,...,N\}$]. Thus, in general, and whatever the value of N,

$$S_{\text{vel}}[x_F(t;x_1 \cdots x_{N-1})] \neq S_{\text{vel}}[x_P(t;x_1 \cdots x_{N-1})], \tag{3.8a}$$

$$S_{\text{pot}}[x_F(t;x_1 \cdots x_{N-1})] \neq S_{\text{pot}}[x_P(t;x_1 \cdots x_{N-1})]. \tag{3.8b}$$

But now consider a path $x_F^\alpha(t;x_1 \cdots x_{N-1})$ for which (2.9) is satisfied for some $\alpha > 0$. We show in Appendix A that, for $|\tau| \leqslant \frac{1}{2}$,

$$\begin{aligned} &x_F^\alpha(t_j + \tau/N;x_1,...,x_{N-1}) \\ &\quad = x_j + \mathcal{O}(N^{-\alpha/4}), \quad \alpha > 0, \quad |\tau| \leqslant \tfrac{1}{2}. \end{aligned} \tag{3.9}$$

Since a similar relation obviously also holds for the corresponding polygonal path, i.e., $x_P^\alpha(t_j + \tau/N;x_1 \cdots x_{N-1}) = x_j + \mathcal{O}(N^{-\alpha})$, we have, for any $0 \leqslant t \leqslant 1$,

$$\begin{aligned} x_F^\alpha(t;x_1,...,x_{N-1}) &= x_P^\alpha(t;x_1,...,x_{N-1}) \\ &\quad + \mathcal{O}(N^{-\alpha/4}), \quad \alpha > 0 \end{aligned} \tag{3.10}$$

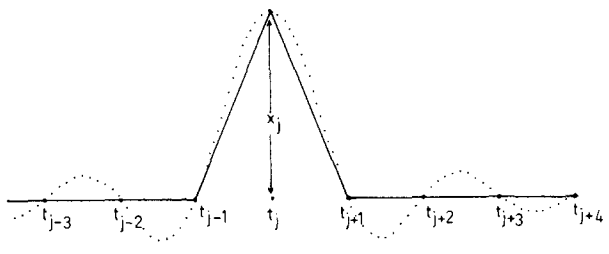


FIG. 1. Full line is the polygonal path $x_P(t;0,...,0,x_j,0,...,0)$; dotted curve is its Fourier approximant $x_F(t;0,...,0,x_j,0,...,0)$, Eq. (3.7).

It follows from (3.10) that

$$\begin{aligned} S_{\text{pot}}[x_F^\alpha(t;x_1 \cdots x_{N-1})] &= S_{\text{pot}}[x_P^\alpha(t;x_1 \cdots x_{N-1})] \\ &\quad + \mathcal{O}(N^{-\alpha/4}), \quad \alpha > 0 \end{aligned} \tag{3.11}$$

as is immediately inferred from considering (2.8) and (3.9). However, even though x_F^α and x_P^α approach one another as $N \to \infty$ when $\alpha > 0$, this is not the case with their derivatives for $0 < \alpha < 1$.[21] Indeed, $\dot{x}_P^\alpha$ and $\dot{x}_F^\alpha$ are both of order $\Delta x/\Delta t \cong N^{-\alpha}/N^{-1} = N^{1-\alpha}$ in magnitude, yet unequal in general, so that they actually diverge from each other as $N \to \infty$ when $\alpha < 1$. [Numerical examples are shown in Fig. 2. There, the paths obey (2.9) with $\alpha = \frac{1}{2}$. One can see that in the case $N = 10$, x_F and x_P are nearer to one another than in the case $N = 4$, but that the contrary holds for their derivatives.] We thus still have

$$S_{\text{vel}}[x_F^\alpha(t;x_1 \cdots x_{N-1})] \neq S_{\text{vel}}[x_P^\alpha(t;x_1 \cdots x_{N-1})], \quad \alpha < 1 \tag{3.12}$$

whatever the value of N.

IV. TRANSFORMING FROM THE POLYGONAL TO THE FOURIER REPRESENTATION

In the polygonal representation (2.6), let us make the change of integration variables

$$\{x_1,x_2,...,x_{N-1}\} \to \{a_1,a_2,...,a_{N-1}\} \tag{4.1}$$

with the x_i's and a_n's related by the linear transformation (3.3) of unit Jacobian, i.e. [for the case (2.17)],

$$\begin{aligned} K_\theta = \lim_{N\to\infty} \left(\frac{N\theta}{2\pi}\right)^{N/2} &\int_{-\infty}^{\infty} da_1 \cdots da_{N-1} \\ &\times \exp\{-\theta S[x_P(t;a_1,...,a_{N-1})]\}. \end{aligned} \tag{4.2}$$

Because of (3.11) and in view of (2.12) (taking, e.g., $\alpha = \frac{1}{4}$), we can replace

$$S_{\text{pot}}[x_P(t;a_1,...,a_{N-1})] \to S_{\text{pot}}[x_F(t;a_1,...,a_{N-1})], \tag{4.3}$$

i.e., in the potential part of the action, the polygonal path $x_P(t;a_1 \ldots a_{N-1})$ can be replaced by its smooth Fourier approximant $x_F(t;a_1 \ldots a_{N-1})$. But this cannot be done with the velocity term S_{vel}, in view of (3.12). Indeed, substituting (3.3) into (2.7), we have (see Appendix C)

$$S_{\text{vel}}[x_P(t;a_1,...,a_{N-1})] = 2N \sum_{n=1}^{N-1} a_n^2 \sin^2\left(\frac{\frac{1}{2}\pi n}{N}\right), \tag{4.4}$$

whereas

$$S_{\text{vel}}[x_F(t;a_1,...,a_{N-1})] = 2N \sum_{n=1}^{N-1} a_n^2 \left(\frac{\frac{1}{2}\pi n}{N}\right)^2. \tag{4.5}$$

The terms $n \ll N$ in the above two sums are nearly equal, but

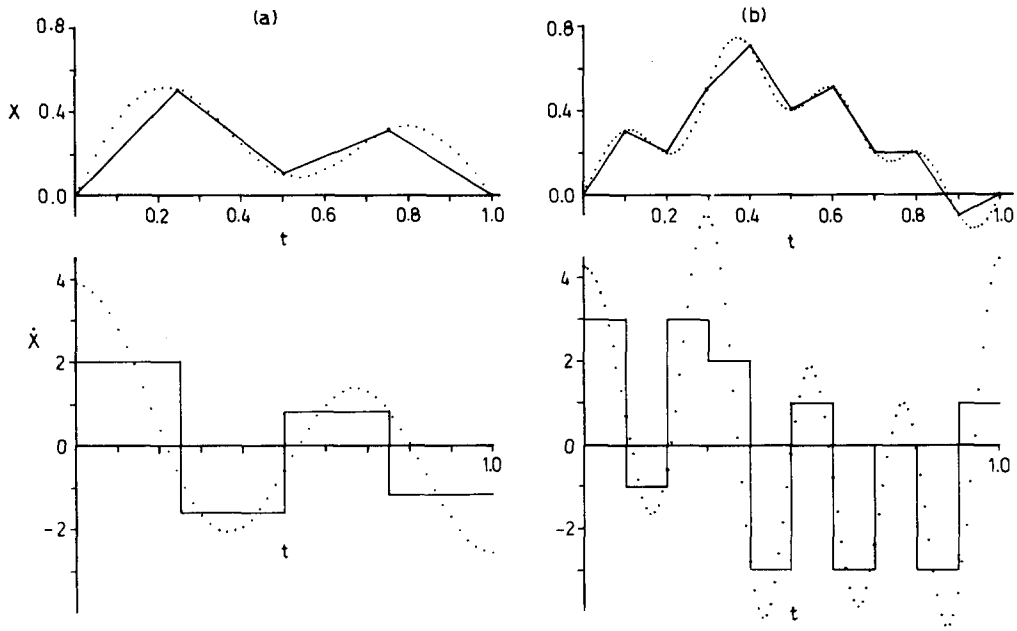


FIG. 2. Examples of N-segment polygonal paths $x_P(t;x_1,\ldots,x_{N-1})$, $0\leqslant t\leqslant 1$ (full lines), with their N-term Fourier series approximants $x_F(t;x_1,\ldots,x_{N-1})$ (dotted curves); $N=4$ in (a), and $N=10$ in (b). The upper parts of the figures show x_P and x_F, while the lower parts show their derivatives $\dot{x}_P$ and $\dot{x}_F$.

not the terms for which n/N is not negligible.

When N is very large, however, n/N not negligible implies $N\sin^2(\frac{1}{2}\pi n/N)$ and $N(n/N)^2$ both large, so that the exponentials

$$\exp\left[-\theta 2N\sin^2(\tfrac{1}{2}\pi n/N)a_n^2\right] \tag{4.6}$$

and

$$\exp\left[-\theta 2N(\tfrac{1}{2}\pi n/N)^2 a_n^2\right] \tag{4.7}$$

both act as unnormalized δ-functions of a_n [this is clear when $\theta=1$, (4.6)–(4.7) then being narrow Gaussians; when $\theta=i$, (4.6)–(4.7) are rapidly oscillating functions of a_n, of wavelengths proportional to $(Na_n)^{-1}$, so that inside an integral over a_n, they cause only a narrow region (of stationary phase) around $a_n=0$ to contribute]. Thus, inside an integral over a_n, (4.6) may be replaced by (4.7) multiplied by a factor matching the normalizations, specifically,

$$(4.6)\to\frac{\frac{1}{2}\pi n/N}{\sin(\frac{1}{2}\pi n/N)}\exp\left[-\theta 2N\left(\frac{\frac{1}{2}\pi n}{N}\right)^2 a_n^2\right]\quad(N\to\infty), \tag{4.8}$$

where the pre-exponential factor is the ratio of the areas of (4.6) and (4.7) as functions of a_n.

Doing (4.8) in (4.2)–(4.4), we get

$$K_\theta=\lim_{N\to\infty}\left(\frac{N\theta}{2\pi}\right)^{N/2}J_N\int_{-\infty}^{\infty}da_1\cdots da_{N-1}\times\exp\{-\theta S[x_F(t;a_1,\ldots,a_{N-1})]\}, \tag{4.9}$$

where the "functional Jacobian"

$$J_N=\prod_{n=1}^{N-1}\left(\frac{\frac{1}{2}\pi n/N}{\sin(\frac{1}{2}\pi n/N)}\right)=N^{-1/2}\prod_{n=1}^{N-1}\left(\frac{\pi n}{N}\right) \tag{4.10}$$

the second equality following from Eq. (D11) in Appendix D. Denoting

$$b_n=(2/N)^{1/2}a_n \tag{4.11}$$

we obtain (2.19)–(2.20).

The transformation from the polygonal to the sine–cosine representation (2.25) is done in a similar manner (see Appendix B).

V. INTEGRALS OVER PATHS WITH FIXED MEANS

In some problems[14–16] there arise path integrals of the form

$$G_\theta(x_a,t_a;x_b,t_b;X)=\int_{\substack{x_a,t_a\\ \bar{x}=X}}^{x_b,t_b}\mathscr{D}x(t)e^{-\theta S[x(t)]}, \tag{5.1}$$

wherein the integration is over paths $x(t)$ with mean

$$\bar{x}\equiv(t_b-t_a)^{-1}\int_{t_a}^{t_b}dt\,x(t) \tag{5.2}$$

equal to the fixed value X. The precise meaning of (5.1) is as in (2.6), but with a δ-function

$$\delta(\bar{x}_P(t;x_1,\ldots,x_{N-1})-X)=\delta\left(\epsilon\sum_{j=1}^{N}\frac{x_j+x_{j-1}}{2}-X\right) \tag{5.3}$$

inserted in the integrand.

In close analogy to the free $\bar{x}$ case, it is possible to express (5.1) as [compare Eq. (2.14)]

$$G_\theta(x_a,t_a;x_b,t_b;X)=e^{-\theta S[x_s(t)]}\tilde{G}_\theta(0,t_a;0,t_b;0), \tag{5.4}$$

where $\tilde{G}$ is given by (5.1), but with the potential $V(x,t)$ replaced by

$$\tilde{V}(x,t)=V(x_s(t)+x,t)-V(x_s(t),t)-xV'(x_s(t),t)$$

and $x_s(t)$ is that path, subject to the constraints

$$x_s(t_a)=x_a,\quad x_s(t_b)=x_b,\text{ and }\bar{x}_s=X,$$

which makes the action S stationary. We shall accordingly limit our considerations to the integral

$$G_\theta\equiv G_\theta(0,0;0,1;0)=\int_{\substack{0,0\\ \bar{x}=0}}^{0,1}\mathscr{D}x(t)e^{-\theta S[x(t)]}. \tag{5.5}$$

When transforming to the Fourier representation, the mean value $\bar{x}_P$ in the δ-function (5.3) may, in view of (3.10), be replaced by $\bar{x}_F$. Thus, G_θ in the pure sine series representation is given by (2.19) with

$$\delta(\bar{x}_F(t;b_1,\ldots,b_{N-1})) = \delta\left(2\pi^{-1}\sum_{n=1,3,5,\ldots}\frac{b_n}{n}\right) \tag{5.6}$$

inserted in the integrand, and the mixed sine–cosine representation of G_θ is given by (2.25) with[22]

$$\delta(\bar{x}_F(t;\{a_n,b_n\})) = \delta(a_0) \tag{5.7}$$

inserted.

VI. HARMONIC OSCILLATOR

We now illustrate the considerations of the preceding sections with the harmonic oscillator case

$$V(x) = \tfrac{1}{2}\omega^2 x^2 \tag{6.1}$$

(for which $\hat{V} = \tilde{V} = V$). Here,

$$S_{\text{pot}}[x_F(t;b_1,b_2,\ldots)] = -\left(\frac{\omega^2}{4}\right)\sum_{n=1}^{\infty} b_n^2 \tag{6.2}$$

so that (2.19) becomes

$$K_\theta = \left(\frac{\theta}{2\pi}\right)^{1/2}\int_{-\infty}^{\infty}\prod_{n=1}^{\infty}\left[\frac{1}{2}n(\pi\theta)^{1/2}db_n\right] \times\exp\left[-\left(\frac{\theta}{4}\right)\sum_{n=1}^{\infty} b_n^2(\pi^2n^2-\omega^2)\right] \tag{6.3}$$

$$= \left(\frac{\theta}{2\pi}\right)^{1/2}\prod_{n=1}^{\infty}\left[1-\left(\frac{\omega}{n\pi}\right)^2\right]^{-1/2}, \tag{6.4}$$

i.e., in view of formula (D7),

$$K_\theta = (\theta/2\pi)^{1/2}(\omega/\sin\omega)^{1/2}. \tag{6.5}$$

In the case $\theta = 1$, (6.4)–(6.5) is true only if $\pi^2n^2 - \omega^2 \geqslant 0$ for each $n = 1,2,\ldots$,i.e., if

$$\omega^2 \leqslant \pi^2. \tag{6.6}$$

Otherwise, $K_{\theta=1}$ is infinite (the action being unbounded below). Note that when $\omega^2 < 0$, i.e., ω is imaginary, the sine function becomes a hyperbolic sine.

Throughout the range (6.6), $(\omega/\sin\omega)$ conserves the same sign, so that there is no problem of interpreting the square root (6.5). In the case $\theta = i$ however, (6.4)–(6.5) holds for all values of ω^2, and the meaning of (6.5) must be made more specific, since $\sin\omega$ undergoes periodic sign changes as ω ranges from 0 to ∞. But clearly, the intermediate form (6.4) forces the interpretation

$$(\sin\omega)^{1/2} = \exp\left[\tfrac{1}{2}i\pi\,\mathrm{Int}(\omega/\pi)\right]|\sin\omega|^{1/2}, \tag{6.7}$$

where $\mathrm{Int}(\omega/\pi)$, the integral part of (ω/π), is the number of factors $(1-\omega^2/n^2\pi^2)$ which are negative in (6.4). The phase in (6.7) is the so-called Maslov index for the harmonic oscillator.[23] Note that $\mathrm{Int}(\omega/\pi)$ is the number of sign changes suffered by $\sin(x)$ between $x = 0$ and $x = \omega$, so that (6.7) really corresponds to the most natural interpretation of the square root of a function $f(x)$ undergoing sign changes: Each time $f(x)$ changes sign, an additional factor $(-1)^{1/2} = e^{i\pi/2}$ is produced, whence

$$f(x)^{1/2} = e^{(1/2)ni\pi}|f(x)|^{1/2}, \tag{6.8}$$

where n is the number of times $f(x)$ changes sign between 0 and x.

To calculate the fixed means integral (5.1), the sine-cosine parametrization (2.25) with (5.7) is the more convenient: We then have

$$G_\theta = \left(\frac{\theta}{2\pi}\right)^{1/2}\int_{-\infty}^{\infty}\prod_{n=1}^{\infty}(n^2\theta\pi\,da_n\,db_n)\delta\left(\sum_{n=1}^{\infty}a_n\right) \times\exp\left[-\theta\sum_{n=1}^{\infty}\left(\pi^2n^2-\frac{1}{4}\omega^2\right)(a_n^2+b_n^2)\right]. \tag{6.9}$$

Introducing the representation

$$\delta(x) = (2\pi)^{-1}\int_{-\infty}^{\infty}dk\,e^{ikx} \tag{6.10}$$

for $\delta(\Sigma_{n=1}^{\infty}a_n)$, we obtain, on interchanging the order of the k and $\{a_n\}$ integrations,

$$G_\theta = (2\pi)^{-3/2}\theta^{1/2}\int_{-\infty}^{\infty}dk\prod_{n=1}^{\infty}\left\{\left[1-\left(\frac{\frac{1}{2}\omega}{n\pi}\right)^2\right]^{-1} \times\exp\left[-\frac{1}{4}\theta^{-1}k^2\left(\pi^2n^2-\frac{1}{4}\omega^2\right)^{-1}\right]\right\} \tag{6.11}$$

$$= \theta\pi^{-1}2^{-1/2}\prod_{n=1}^{\infty}\left[1-\left(\frac{\frac{1}{2}\omega}{\pi n}\right)^2\right]^{-1} \times\left[\sum_{m=1}^{\infty}\left(\pi^2n^2-\frac{1}{4}\omega^2\right)^{-1}\right]^{-1/2}, \tag{6.12}$$

that is, in view of formulas (D7) and (D8),

$$G_\theta = \theta\pi^{-1}(\tfrac{1}{2}\omega/\sin\tfrac{1}{2}\omega)\zeta(\tfrac{1}{2}\omega)^{-1/2} \tag{6.13a}$$

$$= \theta\pi^{-1}(\omega/\sin\omega)^{1/2}\xi(\tfrac{1}{2}\omega)^{-1/2}, \tag{6.13b}$$

where we defined

$$\zeta(z) = (z^{-1}-\cot z)/z = \tfrac{1}{3}+\tfrac{1}{45}z^2+\tfrac{2}{945}z^4+\cdots, \tag{6.14}$$

$$\xi(z) = (\tan z - z)/z^3 = \tfrac{1}{3}+\tfrac{2}{15}z^2+\tfrac{17}{315}z^4+\cdots. \tag{6.15}$$

In the case $\theta = 1$, (6.11)–(6.13) is true only if $\pi^2n^2 - \frac{1}{4}\omega^2 \geqslant 0$ for every $n = 1,2,\ldots$, i.e., if

$$\omega^2 \leqslant 4\pi^2. \tag{6.16}$$

Otherwise, $G_{\theta=1}$ is infinite (the action $S[x(t)]_{\bar{x}=0}$ being unbounded below). In the case $\theta = i$, the square roots in (6.13) must be interpreted in the natural manner (6.8).

On the choice of Fourier parametrization

One could also elect to evaluate K_θ by using the sine–cosine representation (2.25), or to evaluate G_θ via the pure sine representation (2.19) with (5.6). These alternatives, however, do not allow us to determine correctly the ranges of ω^2 values for which these integrals are finite or infinite in the case $\theta = 1$. As a general rule, it seems that the Fourier parametrization to be chosen preferably is that which allows to see unambiguously for which values of ω^2 the action $S[x(t)]$ or $S[x(t)]_{\bar{x}=0}$ is bounded or unbounded below. Consider, e.g., the sine–cosine expansion

$$S[x_F(t;\{a_n,b_n\})]$$

$$= -a_0^2\omega^2 + \sum_{n=1}^{\infty}(a_n^2 + b_n^2)\left(\pi^2 n^2 - \frac{1}{4}\omega^2\right). \quad (6.17)$$

Without the constraint (2.24), one would conclude that (6.17) is unbounded below when $\omega^2 > 0$, because of the term $-a_0^2\omega^2$. But with (2.24), we do not know how to deduce when (6.17) is bounded or unbounded below. The sine–cosine representation of K_θ is identical to (6.9), except for an extra integral over a_0, resulting in an additional factor ω^{-1}, and a term ω^{-2} added to the sum $\Sigma_{n=1}^{\infty}(\pi^2 n^2 - \frac{1}{4}\omega^2)^{-1}$ in (6.12). The final answer is again (6.5). But in the case $\theta = 1$, the integral over a_0 diverges for $\omega^2 \geqslant 0$, whereas we know, from the calculation (6.3)–(6.5), that $K_{\theta=1}$ is finite for (6.6). This apparent contradiction has its origin in the change of the order of the k and $\{a_n\}$ integrations done in passing from (6.9)–(6.10) to (6.11), which is evidently not permissible when $0 \leqslant \omega^2 \leqslant \pi^2$.

The same kinds of undesirable features are met when we try to evaluate G_θ in the pure sine representation (2.19) with (5.6). Note that in the evaluation (6.9)–(6.16) of G_θ, the constraint (2.24) and the concomitant change of the order of the k and $\{a_n\}$ integrations, (6.9)–(6.10) to (6.11), cause no problem, because the value of ω^2 separating the regions where $S[x(t)]_{\bar{x}=0}$ is bounded or unbounded below is determined by the term containing b_1, and the b_n's are not subject to any constraint. Another manner of calculating G_θ is given in Refs. 15a,b; but this again does not allow us to correctly deduce the range (6.16).

VII. CONCLUSION

We transformed, by way of a change of integration variables, from the polygonal to the Fourier representations of path integrals when the action is of the form (1.1). The possibility of effecting such a transformation is due to three factors. The first is that it is only a restricted class of polygonal paths which contribute to the integral. The second factor is that, within that restricted class of paths, N-segment polygonal paths and their N-term Fourier approximants approach one another arbitrarily closely as $N \to \infty$, although their derivatives diverge from each other. The third factor is that the velocity terms in both representations induce effective δ-functions, though of different normalizations; the key step is thus the matching of these normalizations, whence emerges the "functional Jacobian" of the transformation.

The above ascertainments provide valuable insight, and allow us to understand how the polygonal and Fourier representations of a path integral may be different objects when additional velocity-dependent terms are present in the action.

In the harmonic oscillator case, the Fourier series representation trivially determines the large time phase in the real time case, and the ranges of ω^2 values for which the propagator is finite or infinite in the imaginary time case, in both the free and fixed means situations. The calculation of the large time phase by other means is much more elaborate,[23] and it does not seem so easy to deduce the above ω^2 ranges outside the Fourier series representation.

APPENDIX A: DEMONSTRATION OF EQS. (3.5) AND (3.9)

Proof of Eq. (3.5): Substituting (3.3') into (3.4), we have

$$x_F(t;x_1,\ldots,x_{N-1}) = \left(\frac{2}{N}\right)\sum_{i=1}^{N-1}\sum_{n=1}^{N-1} x_i \sin\left(\frac{\pi n i}{N}\right)\sin(\pi n t). \quad (A1)$$

Since $\sin(\pi n i/N) = 0$ at $n = N$, we can let the sum over n run up to N; using then (D1d), and then (D2b), we find

$$(A1) = \left(\frac{2}{N}\right)\sum_{i=1}^{N-1} x_i\left(\frac{1}{4}\right) \times\left\{\frac{\sin[\pi(N+\frac{1}{2})(t-i/N)]}{\sin[\frac{1}{2}\pi(t-i/N)]} - \frac{\sin[\pi(N+\frac{1}{2})(t+i/N)]}{\sin[\frac{1}{2}\pi(t+i/N)]}\right\}. \quad (A2)$$

Introducing the notation (3.6), we have

$$\sin[\pi(N+\tfrac{1}{2})(t\pm i/N)] = \sin\{\pi[(j\pm i)+\tfrac{1}{2}(j\pm i+\tau)/N+\tau]\} = (-)^{j\pm i}\sin[\tfrac{1}{2}\pi(j\pm i+\tau)/N+\pi\tau]. \quad (A3)$$

Substituting into (A2) and using (D1g) with $a = \frac{1}{2}\pi(j+\tau+i)/N$, $b = \pi\tau$, $c = \frac{1}{2}\pi(j+\tau-i)/N$, we obtain (3.5) [note that $(-)^{j\pm i} = (-)^{i-j}$].

Proof of Eq. (3.9): Now let the x_i's obey (2.9) for some $\alpha > 0$. Setting

$$i - j = l, \quad 0 < |\tau| \leqslant \tfrac{1}{2}, \quad (A4a)$$

$$x_{j+l} = x_j + N^{-\alpha} f(l), \quad (A4b)$$

where we have, because of (2.9) [with (2.17)],

$$|f(l+1) - f(l)| \leqslant 1, \quad |f(l)| \leqslant l, \quad (A5)$$

we rewrite (3.5) as

$$x_F(t;x_1,\ldots,x_{N-1}) = (2N)^{-1}\sum_{l=-j}^{N-j}(-)^l[x_j + N^{-\alpha}f(l)]G(l), \quad (A6)$$

where

$$G(l) = H(l)L(l),$$

$$H(l) = \frac{\sin[\pi(j+l)/N]}{\sin[\pi(j+\frac{1}{2}l+\frac{1}{2}\tau)/N]}, \quad (A7)$$

$$L(l) = \frac{\sin(\pi\tau)}{\sin[\frac{1}{2}(\tau-l)/N]}.$$

In the last part of this Appendix, we show that (*cte* denote constants independent of N)

$$|G(l)| \leqslant \begin{cases} (cte)N, & l = 0, \\ (cte)N/|l|, & l \neq 0, \end{cases} \quad (A8)$$

and

$$|\Delta G(l)| \leqslant \begin{cases} (cte)N, & l = 0, \\ (cte)N/l^2, & l \neq 0, \end{cases} \quad (A9)$$

where we denote, for any function $F(l)$ of the integer l,

$$\Delta F(l) \equiv F(l+1) - F(l). \quad (A10)$$

We will use the relations

$$\Delta(fG) = (\Delta f)G + f(\Delta G) + (\Delta f)(\Delta G) \quad (A11)$$

and

$$\sum_{l=l_1}^{l_2} (-)^l F(l) = \sum_{l\,\text{even}} \Delta F(l) + E, \tag{A12}$$

where E consists of possible unpaired initial or final terms [i.e., $E = \epsilon_{l_1} F(l_1) + (1-\epsilon_{l_2} F(l_2)$, where $\epsilon_m = 1$ if m is even, $\epsilon_m = 0$ if m is odd].

With the above results, we have, for the second sum in (A6),

$$\begin{aligned} &|N^{-1-\alpha}\sum (-)^l f(l)G(l)| \\ &\quad \leqslant (cte)N^{-\alpha}\left(cte' + \sum_{l\,\text{even},\neq 0} l^{-1}\right) \\ &\quad \leqslant (cte)N^{-\alpha}\ln(N) \end{aligned} \tag{A13}$$

(since $|l| \leqslant N$) [the end-point correction E is of order $N^{-\alpha}$ since $|f(l)G(l)| \leqslant (cte)N$; it has been incorporated in cte', together with the term $l=0$].

We now estimate the first sum in (A6). Let us break up

$$\frac{1}{2} x_j N^{-1} \sum_l (-)^l G(l) \equiv A_1 + A_2, \tag{A14}$$

where A_1 is the part $|l| \leqslant N^{\alpha/4}$, and A_2 the part $|l| \geqslant N^{\alpha/4}$. Using (A8)–(A12), we have

$$|A_2| \leqslant (cte) \sum_{l\,\text{even},> N^{\alpha/4}} l^{-2} \leqslant (cte) N^{-\alpha/4}. \tag{A15}$$

To estimate A_1, three cases are considered:

Case (i): $N^{\alpha/2} \leqslant j \leqslant N - N^{\alpha/2}$. Since $|l| \leqslant N^{\alpha/4}$, we have $|l/j| \leqslant (cte)N^{-\alpha/4}$ and $|l/N| \leqslant N^{(\alpha/4)-1}$, whence [cf. Eqs. (A7)]

$$H(l) = 1 + \mathcal{O}(N^{-\alpha/4}), \tag{A16}$$

$$\sin[\tfrac{1}{2}\pi(\tau - l)/N] = \tfrac{1}{2}\pi(\tau - l)/N + \mathcal{O}(N^{-3(1-\alpha/4)}). \tag{A17}$$

Substituting into A_1, Eq. (A14), and using the formula[24]

$$\csc(\pi\tau) = \pi^{-1} \sum_{l=-\infty}^{\infty} \frac{(-)^l}{\tau - l} \tag{A18}$$

we deduce

$$A_1 = x_j + \mathcal{O}(N^{-\alpha/4}). \tag{A19}$$

Case (ii): $0 \leqslant j \leqslant N^{\alpha/2}$. We have, from (A8)–(A12),

$$|N^{-1}\sum (-)^l G(l)| \leqslant (cte), \tag{A20}$$

whence $|A_1| \leqslant (cte)x_j$. But since $|x_j| \leqslant (cte)N^{-\alpha/2}$ in view of (2.17) and (A4b) (with $j=0$, $l \leqslant N^{\alpha/2}$), we can again write Eq. (A19).

Case (iii): $N - N^{\alpha/2} \leqslant j \leqslant N$. This case is symmetric with case (ii).

The above results together yield (3.9).

Proof of (A8)–(A9): Denote

$$H(t,t') = \sin(\pi t')/\sin[\tfrac{1}{2}\pi(t+t')], \quad 0 \leqslant t,t' \leqslant 1. \tag{A21}$$

We have

$$|H(t,t')| \leqslant cte_1, \tag{A22}$$

$$\left|\left(\frac{d}{dt'}\right)H(t,t')\right| \leqslant cte_2 \begin{cases} (t+t')^{-1} & \text{for } 0 \leqslant t \leqslant \tfrac{1}{2} \\ [(1-t)+(1-t')]^{-1} & \text{for } \tfrac{1}{2} \leqslant t \leqslant 1. \end{cases} \tag{A23}$$

This is shown, for the case $0 \leqslant t \leqslant \frac{1}{2}$ [so that $0 \leqslant \frac{1}{2}(t+t') \leqslant \frac{3}{4}$] by using $|\sin(\pi t')| \leqslant \pi t'$ if $0 \leqslant t' \leqslant 1$, $|\sin(\frac{1}{2}\pi x)| \leqslant \frac{1}{2}Kx$ if $0 \leqslant x \leqslant \frac{3}{4}$, where $K = \sin(3\pi/4)/(3\pi/4)$, and $|\cos(\pi x)| \leqslant 1$, $t'/(t+t') \leqslant 1$. The case $\frac{1}{2} \leqslant t \leqslant 1$ follows from $H(t,t') = H(1-t, 1-t')$; specifically, $cte_1 = 2\pi/K$, $cte_2 = (2\pi/K) + 2\pi^2/K$.

With $H(l)$ defined as in (A7), i.e.,

$$H(l) = H(t = (j+\tau)/N,\ t' = (j+l)/N) \tag{A24}$$

with

$$0 \leqslant j \leqslant N, \quad -j \leqslant l \leqslant N - j, \quad |\tau| \leqslant \tfrac{1}{2} \tag{A25}$$

we have

$$|H(l)| \leqslant cte, \tag{A26}$$

$$|\Delta H(l)| \leqslant \begin{cases} cte, & l = 0 \\ cte/|l|, & l \neq 0 \end{cases}. \tag{A27}$$

The case $l \neq 0$ in (A27) follows from

$$|\Delta F(l)| = |F(l+1) - F(l)| \leqslant \operatorname*{Max}_{l \leqslant x \leqslant l+1} \left|\frac{dF(x)}{dx}\right| \tag{A28}$$

(A23) and (A25); the case $l = 0$ follows from

$$|\Delta F(l)| \leqslant |F(l+1)| + |F(l)| \tag{A29}$$

and (A22).

Denoting now

$$\hat{L}(t) = \sin(\pi\tau)/\sin(\tfrac{1}{2}\pi t), \quad |\tau| \leqslant \tfrac{1}{2}, \quad |t| \leqslant 1 \tag{A30}$$

we have

$$|\hat{L}(t)| \leqslant \left|\frac{\pi\tau}{t}\right|, \quad \left|\left(\frac{d}{dt}\right)\hat{L}(t)\right| \leqslant \left|\frac{\pi\tau}{t^2}\right| \tag{A31}$$

following from

$$|x| \leqslant |\sin(\tfrac{1}{2}\pi x)| \leqslant \tfrac{1}{2}\pi|x| \text{ if } |x| \leqslant 1 \tag{A32}$$

and $|\cos(\pi x)| \leqslant 1$. With

$$L(l) = \hat{L}((\tau - l)/N), \quad |\tau| \leqslant \tfrac{1}{2}, \quad |\tau - l| \leqslant N \tag{A33}$$

as defined in (A7), we have

$$|L(l)| \leqslant \begin{cases} cte, & l = 0, \\ cte/|l|, & l \neq 0, \end{cases} \tag{A34}$$

$$|\Delta L(l)| \leqslant \begin{cases} cte\,N, & l = 0, \\ cte\,N/l^2, & l \neq 0, \end{cases} \tag{A35}$$

following from (A31), (A28), and (A29). Equation (A8)–(A9) follow from the above and (A11).

APPENDIX B: TRANSFORMING TO THE SINE-COSINE REPRESENTATION

In this Appendix, we transform from the polygonal representation (2.6) to the sine–cosine Fourier representation (2.25). Let us first rewrite (2.6) as [for the case (2.17)]

$$\begin{aligned} K_\theta = \lim_{N\to\infty} \left(\frac{N\theta}{2\pi}\right)^{N/2} \int_{-\infty}^{\infty} dx_0\, dx_1 \cdots dx_{N-1} \\ \times \delta(x_0) \exp\{-\theta S[x_P(t; x_0, x_1, \ldots, x_{N-1})]\}, \end{aligned} \tag{B1}$$

wherein the integration is over all polygonal paths $x_P(t; x_0, x_1, \ldots, x_{N-1})$ which have equal but free endpoints

$$x_0 = x_N. \tag{B2}$$

Letting N be even, we make the change of integration variables

$$\{x_0,x_1,\ldots,x_{N-1}\}\to\{a_0',a_1',\ldots,a_{(1/2)N}',b_1',\ldots,b_{(1/2)N-1}'\} \tag{B3}$$

via the transformation

$$x_j=\left(\frac{2}{N}\right)^{1/2}\left[2^{-1/2}a_0'+\sum_{n=1}^{(1/2)N}a_n'\cos\left(\frac{2\pi nj}{N}\right)+\sum_{n=1}^{(1/2)N-1}b_n'\sin\left(\frac{2\pi nj}{N}\right)\right],\quad j=0,1,\ldots,N-1 \tag{B4}$$

of unit Jacobian [see (C6)]. Note that if extended to the case $j=N$, (B4) satisfies (B2), i.e.,

$$x_0=x_N=\left(\frac{2}{N}\right)^{1/2}\left[2^{-1/2}a_0'+\sum_{n=1}^{(1/2)N}a_n'\right]. \tag{B5}$$

The polygonal paths may be parametrized by either the x_j's or the $\{a_n',b_n'\}$, and we denote

$$x_P(t;x_0,x_1,\ldots,x_{N-1})\equiv x_P(t;\{a_n',b_n'\}_N). \tag{B6}$$

Substituting (B4) into the discretized expression (2.7) of $S_{\rm vel}$, we obtain (see Appendix C)

$$S_{\rm vel}[x_P(t;\{a_n',b_n'\}_N)]=N\left[\sum_{n=1}^{(1/2)N}a_n'^2\sin^2\left(\frac{\pi n}{N}\right)+\sum_{n=1}^{(1/2)N-1}b_n'^2\sin^2\left(\frac{\pi n}{N}\right)\right]. \tag{B7}$$

Introducing the smooth Fourier approximant

$$x_F(t;\{a_n',b_n'\}_N)=\left(\frac{2}{N}\right)^{1/2}\left[2^{-1/2}a_0'+\sum_{n=1}^{(1/2)N}a_n'\cos(2\pi nt)+\sum_{n=1}^{(1/2)N-1}b_n'\sin(2\pi nt)\right] \tag{B8}$$

to the polygonal path (B6), we have

$$S_{\rm vel}[x_F(t;\{a_n',b_n'\}_N)]=N\left[\sum_{n=1}^{(1/2)N}a_n'^2\left(\frac{\pi n}{N}\right)^2+\sum_{n=1}^{(1/2)N-1}b_n'^2\left(\frac{\pi n}{N}\right)^2\right]. \tag{B9}$$

For the same reasons which allow the substitution (4.8), we may replace in the integral (B1)

$$\exp\left[-\theta Na_n'^2\sin^2\left(\frac{\pi n}{N}\right)\right]\to\frac{(\pi n/N)}{\sin(\pi n/N)}\exp\left[-\theta Na_n'^2\left(\frac{\pi n}{N}\right)^2\right] \tag{B10}$$

and likewise with the b_n' terms. We thereby get, using (B5),

$$K_\theta=\lim_{N\to\infty}\left(\frac{N\theta}{2\pi}\right)^{N/2}J_N\times\int_{-\infty}^{\infty}da_0'\,da_1'\cdots da_{(1/2)N}'\,db_1'\cdots db_{(1/2)N-1}'\times\delta\left(N^{-1/2}a_0'+2^{1/2}N^{-1/2}\sum_{n=1}^{(1/2)N}a_n'\right)e^{-\theta S[x_F]} \tag{B11}$$

where the functional Jacobian

$$J_N=\prod_{n=1}^{(1/2)N}\left[\frac{(\pi n/N)}{\sin(\pi n/N)}\right]\prod_{n=1}^{(1/2)N-1}\left[\frac{(\pi n/N)}{\sin(\pi n/N)}\right]=N^{-N}(2\pi)^{N-1}(\tfrac{1}{2}N)!(\tfrac{1}{2}N-1)! \tag{B12}$$

in view of formula (D12). Denoting

$$a_0=N^{-1/2}a_0',\quad a_n=(2/N)^{1/2}a_n',\quad b_n=(2/N)^{1/2}b_n',\quad n\geqslant 1, \tag{B13}$$

we recover (2.25)–(2.27).

APPENDIX C: EVALUATION OF $\Sigma_{j=0}^{N-1}(x_{j+1}-x_j)^2$ IN FOURIER COORDINATES

Let us denote

$$\Delta\equiv\frac{1}{2}\sum_{j=0}^{N-1}(x_{j+1}-x_j)^2=\sum_{j=0}^{N-1}x_j^2-\sum_{j=0}^{N-1}x_jx_{j+1}, \tag{C1}$$

where we assume $x_N=x_0$ [so that $\Sigma_{j=0}^{N-1}x_{j+1}^2=\Sigma_{j=0}^{N-1}x_j^2$].

Pure sine case: Here, $x_0=x_N=0$, so that the sums in (C1) run from 1 to $N-1$. Introducing the transformation (3.3), and defining a matrix L by

$$L_{ij}=\delta^{Kr}_{i+1,j} \tag{C2}$$

so that, e.g., $(Lx)_j=x_{j+1}$, we rewrite (C1) as

$$\Delta=x^Tx-x^TLx=a^TM^TMa-a^TM^TLMa. \tag{C3}$$

Using

$$M^TM=1,\quad (M^TLM)_{mn}=\delta_{mn}\cos(\pi n/N)+f_{mn} \tag{C4}$$

following from (D3) and (D4), we find

$$\Delta=\sum_{n=1}^{N-1}a_n^2\left[1-\cos\left(\frac{\pi n}{N}\right)\right]=2\sum_{n=1}^{N-1}a_n^2\sin^2\left(\frac{\frac{1}{2}\pi n}{N}\right) \tag{C5}$$

(note that $a^Tfa=0$ since $f_{mn}=-f_{nm}$).

Sine–cosine case: Introducing the transformation (B4), we find, using (D5)

$$\sum_{j=0}^{N-1}x_j^2=\sum_{n=0}^{(1/2)N}a_n'^2+\sum_{n=1}^{(1/2)N-1}b_n'^2 \tag{C6}$$

implying that the transformation (B4) is a rotation, of unit Jacobian. We also have, on using (D6) with $k=1$,

$$\begin{aligned}\sum_{j=0}^{N-1}x_jx_{j+1}&=\left(\frac{2}{N}\right)\sum_{j=0}^{N-1}\sum_{n,m}a_n'a_m'\cos\left(\frac{2\pi mj}{N}\right)\cos\left[\frac{2\pi n(j+1)}{N}\right]\\&+\sum_{n,m}b_n'b_m'\sin\left(\frac{2\pi nj}{N}\right)\sin\left[\frac{2\pi n(j+1)}{N}\right]\\&+\sum_{n,m}a_n'b_m'\cos\left(\frac{2\pi nj}{N}\right)\sin\left[\frac{2\pi m(j+1)}{N}\right]\\&+\sin\left(\frac{2\pi mj}{N}\right)\cos\left[\frac{2\pi n(j+1)}{N}\right]\end{aligned} \tag{C7}$$

$$=\sum_{n=0}^{(1/2)N}a_n'^2\cos\left(\frac{2\pi n}{N}\right)+\sum_{m=1}^{(1/2)N-1}b_n'^2\cos\left(\frac{2\pi n}{N}\right) \tag{C7'}$$

[the last two sums in (C7) cancel one another]. Thus

$$\Delta=\sum_{n=0}^{(1/2)N}a_n'^2\left[1-\cos\left(\frac{2\pi n}{N}\right)\right]+\sum_{n=1}^{(1/2)N-1}b_n'^2\left[1-\cos\left(\frac{2\pi n}{N}\right)\right]$$

$$= 2\sum_{n=0}^{(1/2)N} a_n'^2 \sin^2\left(\frac{\pi n}{N}\right) + 2\sum_{n=1}^{(1/2)N-1} b_n'^2 \sin^2\left(\frac{\pi n}{N}\right). \quad \text{(C8)}$$

APPENDIX D: TRIGONOMETRIC IDENTITIES

In this Appendix, we derive some useful trigonometric identities. Let us first list a few standard formulas to which we will need to refer:

$$\sin(a+b) = \sin(a)\cos(b) + \cos(a)\sin(b), \quad \text{(D1a)}$$

$$\cos(a+b) = -\sin(a)\sin(b) + \cos(a)\cos(b), \quad \text{(D1b)}$$

$$\cot(a) \pm \cot(b) = \pm \sin(a \pm b)/[\sin(a)\sin(b)], \quad \text{(D1c)}$$

$$\sin(a)\sin(b) = \tfrac{1}{2}[\cos(a+b) - \cos(a-b)], \quad \text{(D1d)}$$

$$\cos(a)\cos(b) = \tfrac{1}{2}[\cos(a+b) + \cos(a-b)], \quad \text{(D1e)}$$

$$\sin(a)\cos(b) = \tfrac{1}{2}[\sin(a+b) + \sin(a-b)], \quad \text{(D1f)}$$

$$\frac{\sin(a+b)}{\sin(a)} - \frac{\sin(c+b)}{\sin(c)} = \frac{\sin(b)\sin(c-a)}{\sin(a)\sin(c)}. \quad \text{(D1g)}$$

From evaluating the geometric sum $\Sigma_{k=0}^{N}(e^{ix})^k$, one deduces

$$\sum_{k=0}^{N} \cos(kx) = \frac{\cos(\frac{1}{2}Nx)\sin[\frac{1}{2}(N+1)x]}{\sin(\frac{1}{2}x)} \quad \text{(D2a)}$$

$$= \frac{\frac{1}{2}\sin[(N+\frac{1}{2})x]}{\sin(x)} + \frac{1}{2} \quad \text{(D2b)}$$

$$= \tfrac{1}{2}\sin(Nx)\cot(\tfrac{1}{2}x) + \cos^2(\tfrac{1}{2}Nx), \quad \text{(D2c)}$$

$$\sum_{k=0}^{N} \sin(kx) = \frac{\sin(\frac{1}{2}Nx)\sin[\frac{1}{2}(N+1)x]}{\sin(\frac{1}{2}x)} \quad \text{(D2d)}$$

$$= \tfrac{1}{2}\sin(Nx) + \sin^2(\tfrac{1}{2}Nx)\cot(\tfrac{1}{2}x) \quad \text{(D2e)}$$

the cases c and e following from (D1a) with $a = \frac{1}{2}Nx$, $b = \frac{1}{2}x$.

Lemma: Given integers m,n satisfying

$$1 \leqslant m,n \leqslant N-1 \quad \text{(D3a)}$$

we have

$$\sum_{j=1}^{N-1} \sin\left(\frac{\pi jm}{N}\right)\sin\left(\frac{\pi jn}{N}\right) = \tfrac{1}{2}N\delta_{mn}, \quad \text{(D3b)}$$

$$\sum_{j=1}^{N-1} \sin\left(\frac{\pi jm}{N}\right)\cos\left(\frac{\pi jn}{N}\right) = \begin{cases} 0, & \text{for } m \pm n \text{ even}, \\ \frac{1}{2}\dfrac{\sin(\pi m/N)}{\sin[\frac{1}{2}\pi(m+n)/N]\sin[\frac{1}{2}\pi(m-n)/N]}, & \\ & m \pm n \text{ odd}. \end{cases} \quad \text{(D3c)}$$

Proof: We first note that we can extend the sums on the left-hand side of (D3b) and (D3c) to $\Sigma_{j=0}^{N}$, the terms $j=0$ and N being zero; this will allow us to use Eqs. (D2). Using (D1d) and then (D2c), we have that the left-hand side of (D3b) equals

$$\tfrac{1}{2}\{\tfrac{1}{2}\sin[\pi(m-n)]\cot[\tfrac{1}{2}\pi(m-n)/N] + \cos^2[\tfrac{1}{2}\pi(m-n)] - \tfrac{1}{2}\sin[\pi(m-n)]\cot[\tfrac{1}{2}\pi(m+n)/N] - \cos^2[\tfrac{1}{2}\pi(m+n)]\}.$$

The first term equals $\frac{1}{2}N\delta_{mn}$; the third term is zero; the second and fourth terms are zero when $m \pm n$ is odd, and cancel one another when $m \pm n$ is even. Whence (D3b). Using (D1f) and (D2e), we have that the left-hand side of (D3c) equals

$$\tfrac{1}{2}\{\tfrac{1}{2}\sin[\pi(m+n)] + \sin^2[\tfrac{1}{2}\pi(m+n)]\cot[\tfrac{1}{2}\pi(m+n)/N] + \tfrac{1}{2}\sin[\pi(m+n)] + \sin^2[\tfrac{1}{2}\pi(m-n)]\cot[\tfrac{1}{2}\pi(m-n)/N]\}.$$

The first and third terms are zero. The second and fourth terms are zero when $m \pm n$ is even; when $m \pm n$ is odd, we have $\sin^2[\frac{1}{2}\pi(m \pm n)] = 1$, whence (D3c) on using (D1c). QED

Corollary: Given (D3a), we have

$$\sum_{j=1}^{N-1} \sin\left(\frac{\pi mj}{N}\right)\sin\left[\frac{\pi n(j+1)}{N}\right] = \frac{1}{2}N\delta_{mn}\cos\left(\frac{\pi n}{N}\right) + f_{mn}, \quad \text{(D4a)}$$

where

$$f_{mn} = g_{mn}\sin(\pi n/N) = -f_{nm} \quad \text{(D4b)}$$

and g_{mn} is the right-hand side of (D3c).

Proof: Follows from using (D1a) with $a = \pi nj/N$, $b = \pi n/N$, and then applying (D3).

Lemma: Given integers m and n satisfying

$$0 \leqslant m \leqslant \tfrac{1}{2}N, \quad 0 \leqslant n \leqslant \tfrac{1}{2}N, \quad N \text{ even} \quad \text{(D5a)}$$

we have

$$\sum_{j=0}^{N-1} \cos\left(\frac{2\pi nj}{N}\right)\cos\left(\frac{2\pi mj}{N}\right) = \begin{cases} \frac{1}{2}N\delta_{mn}, & m,n \neq 0,\tfrac{1}{2}N, \\ N\delta_{mn}, & m \text{ or } n = 0 \text{ or } \tfrac{1}{2}N, \end{cases} \quad \text{(D5b)}$$

$$\sum_{j=0}^{N-1} \sin\left(\frac{2\pi nj}{N}\right)\sin\left(\frac{2\pi mj}{N}\right) = \begin{cases} \frac{1}{2}N\delta_{mn}, \\ 0, \end{cases} \quad \text{(D5c)}$$

$$\sum_{j=0}^{N-1} \cos\left(\frac{2\pi nj}{N}\right)\sin\left(\frac{2\pi mj}{N}\right) = 0. \quad \text{(D5d)}$$

In (D5c), and in Eqs. (D6) below, the upper and lower values apply, as in (D5b) and (D6a), to the cases $\{m \neq 0 \text{ or } \frac{1}{2}N \text{ and } n \neq 0 \text{ or } \frac{1}{2}N\}$ and $\{m = 0 \text{ or } \frac{1}{2}N \text{ or } n = 0 \text{ or } \frac{1}{2}N\}$, respectively.

Proof: The cases m or $n = 0$ or $\frac{1}{2}N$ are obvious. Cases $m,n \neq 0, \frac{1}{2}N$:

Using (D1e) and then (D2c), we have

$$\sum_{j=0}^{N} \cos\left(\frac{2\pi nj}{N}\right)\cos\left(\frac{2\pi mj}{N}\right) = \frac{1}{2}\left\{\frac{1}{2}\sin[2\pi(n-m)]\cot\left[\frac{\pi(n-m)}{N}\right] + \cos^2[\pi(n-m)] + \frac{1}{2}\sin[2\pi j(n+m)] \times \cot\left[\frac{\pi(n+m)}{N}\right] + \cos^2[\pi(n+m)]\right\}$$

The first term equals $\frac{1}{2}N\delta_{mn}$, the second and fourth terms equal 1, and the third term is zero, whence (D5b) since the term $j = N$ equals 1. (D5c) is shown similarly. Using (D1f) and then (D2e), we have

$$\sum_{j=0}^{N} \cos\left(\frac{2\pi nj}{N}\right)\sin\left(\frac{2\pi mj}{N}\right) = \frac{1}{2}\left\{\frac{1}{2}\sin[2\pi(m+n)] + \sin^2[\pi(m+n)]\cot\left[\frac{\pi(m+n)}{N}\right]\right.$$

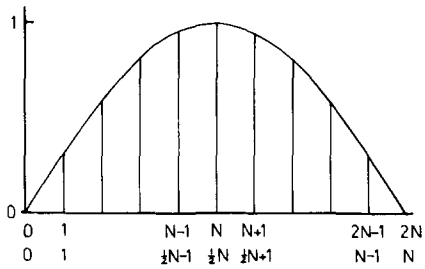


FIG. 3. Illustrating the derivations of Eqs. (D11) and (D12). For (D11), the curve represents $\sin(\frac{1}{2}\pi n/N)$ vs n, the values of n being the upper ones. For (D12), the curve is $\sin(\pi n/N)$ vs n, the values of n being the lower ones.

$$+\frac{1}{2}\sin[2\pi(m-n)]$$
$$+\sin^2[\pi(m-n)]\cot\left[\frac{\pi(m-n)}{N}\right]\Bigg\}$$
$$=0,$$

whence (D5d), since the term $j=N$ is zero. QED

Corollary: Given (D5a), we have

$$\sum_{j=0}^{N-1}\cos\left(\frac{2\pi mj}{N}\right)\cos\left[\frac{2\pi n(j+k)}{N}\right]$$
$$=\begin{cases}\frac{1}{2}N\delta_{mn}\cos(2\pi nk/N), & m,n\neq 0,\frac{1}{2}N,\\ N\delta_{mn}, & m \text{ or } n=0 \text{ or } \frac{1}{2}N,\end{cases} \tag{D6a}$$

$$\sum_{j=0}^{N-1}\cos\left(\frac{2\pi mj}{N}\right)\sin\left[\frac{2\pi n(j+k)}{N}\right]$$
$$=\begin{cases}\frac{1}{2}N\delta_{mn}\sin(2\pi nk/N),\\ 0,\end{cases} \tag{D6b}$$

$$\sum_{j=0}^{N-1}\sin\left(\frac{2\pi mj}{N}\right)\sin\left[\frac{2\pi n(j+k)}{N}\right]$$
$$=\begin{cases}\frac{1}{2}N\delta_{mn}\cos(2\pi nk/N),\\ 0,\end{cases} \tag{D6c}$$

$$\sum_{j=0}^{N-1}\sin\left(\frac{2\pi mj}{N}\right)\cos\left[\frac{2\pi n(j+k)}{N}\right]$$
$$=\begin{cases}-\frac{1}{2}N\delta_{mn}\sin(2\pi nk/N),\\ 0.\end{cases} \tag{D6d}$$

Proof: Follows from using (D1a) – (D1b) with $a=2\pi nj/N$, $b=2\pi nk/N$, and then applying (D5).

We also need the following standard formulas[25]:

$$\sin(z)=z\prod_{n=1}^{\infty}\left[1-\left(\frac{z}{n\pi}\right)^2\right], \tag{D7}$$

$$\cot(z)=z^{-1}+2z\prod_{n=1}^{\infty}(z^2-n^2\pi^2)^{-1}, \tag{D8}$$

$$\prod_{n=0}^{N-1}\sin\left(x+\frac{n\pi}{N}\right)=2^{-N+1}\sin(Nx). \tag{D9}$$

From (D9), we deduce

$$\prod_{n=1}^{N-1}\sin\left(\frac{\pi n}{N}\right)=2^{-N+1}\lim_{x\to 0}\frac{\sin(Nx)}{\sin(x)}=N2^{-N+1}, \tag{D10}$$

whence (see Fig. 3)

$$\prod_{n=1}^{N-1}\sin\left(\frac{\frac{1}{2}\pi n}{N}\right)$$
$$=\left[\prod_{n=1}^{2N-1}\sin\left(\frac{\frac{1}{2}\pi n}{N}\right)\right]^{1/2}=(2N2^{-2N+1})^{1/2}$$
$$=N^{1/2}2^{-N+1} \tag{D11}$$

and

$$\prod_{n=1}^{(1/2)N-1}\sin\left(\frac{\pi n}{N}\right)\prod_{n=1}^{(1/2)N}\sin\left(\frac{\pi n}{N}\right)=\prod_{n=1}^{N-1}\sin\left(\frac{\pi n}{N}\right)$$
$$=N2^{-N+1}. \tag{D12}$$

[1] N. Wiener, J. Math. Phys. (Cambridge, Mass.) **2**, 131 (1923); Proc. London Math. Soc. **22**, 454 (1924); "Generalized Harmonic Analysis," Acta. Math. **55**, 117 (1930).

[2] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).

[3] B. Davison, Proc. R. Soc. London Ser. A **225**, 252 (1954).

[4] W. K. Burton and A. H. deBorde, Nuovo Cimento **2**, 197 (1955).

[5] H. Davies, Proc. Cambridge Philos. Soc. **53**, 199 (1957).

[6] The Fourier series evaluation of path integrals is reviewed in detail in S. G. Brush, Rev. Mod. Phys. **33**, 79 (1961).

[7] J. D. Doll and L. E. Myers, J. Chem. Phys. **71**, 2880 (1979).

[8] H. Dekker, Phys. Lett. A **65**, 388 (1978); **69**, 241 (1978); Phys. Rev. A **19**, 2102 (1979); Physica A **103**, 586 (1980). J. D. Bekenstein and L. Parker, Phys. Rev. D **23**, 2850 (1981).

[9] S. S. Chang, Phys. Rev. D **17**, 2595, 2611 (1978).

[10] F. Langouche, D. Roekaerts, and E. Tirapegui, (a) Phys. Rev. A **21**, 1344 (1980); (b) *Functional Integration and Semiclassical expansions* (Reidel, Dordrecht, 1982), pp. 132–135.

[11] See Ref. 10(b), Chap. II and references therein.

[12] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965), Sec. 3–11.

[13] S. S. Chang, "Inconsistency in Feynman and Hibbs' evaluation of path integrals by Fourier expansions," preprint 4210-91, Syracuse University, New York, 1977 (unpublished), cited and partly reproduced in Ref. 10(b).

[14] Reference 12, Sec. 10-3.

[15] (a) A. Siegel and T. Burke, J. Math. Phys. **13**, 1681 (1972); **14**, 2018 (1973); (b) M. M. Mizrahi, J. Math. Phys. **17**, 566 (1976); (c) L. W. Bruch and H. E. Revercomb, J. Chem. Phys. **58**, 751 (1973); (d) E. P. Gross, J. Stat. Phys. **21**, 215 (1979); **31**, 115 (1983).

[16] A. Royer, "Simple approximations for quantum partition functions and integrals over paths with fixed means" (unpublished).

[17] See Ref. 1 for the case $\theta=1$; for the case $\theta=i$, see J. R. Klauder and I. Daubechies, Phys. Rev. Lett. **48**, 117 (1982); J. Math. Phys. **23**, 1806 (1982).

[18] This transformation is achieved by setting $x(t)=x_{\rm cl}(t)+y(t)$ and replacing the integral over paths $x(t)$ by an integral over paths $y(t)$. Strictly speaking, since $x(t)$ and $y(t)$ are polygonal paths, so should $x_{\rm cl}(t)$; but in the limit $N\to\infty$, the exact $x_{\rm cl}(t)$ and its derivatives can be approached arbitrarily closely by N-segment polygonal paths.

[19] As is well known, two functions may approach one another in some limit, without their derivatives doing so. Let for instance $f(t)=N^{-1/2}\sin(Nt)$ and $g(t)=0$; we have $f(t)\to g(t)$, whereas $\dot f(t)=N^{1/2}\cos(Nt)$ and $\dot g(t)\equiv 0$ diverge from one another, as $N\to\infty$.

[20] More exactly: from (3.5),

$$x_F(t;0\cdots 0,x_j,0\cdots 0)=x_j(2N)^{-1}\sin(\pi\tau)\sin(\pi t_j)/$$
$$\times[\sin(\pi t_j+\tfrac{1}{2}\pi\tau/N)\sin(\tfrac{1}{2}\pi\tau/N)]$$
$$=x_j(2N)^{-1}\frac{\sin[\pi(t-\tau/N)]}{\sin[\pi(t-\frac{1}{2}\tau/N)]}$$
$$\times\frac{\sin(\pi\tau)}{\sin(\frac{1}{2}\pi\tau/N)}$$

$[\to(3.7)$ as $N\to\infty]$.

[21] To demonstrate these statements, one might be tempted to argue that since the coefficients a_n of the (infinite) Fourier expansion $\Sigma_{n=1}^{\infty}a_n\sin(\pi nt)$ of a polygon behave like n^{-2} as $n\to\infty$, then by neglecting the terms $n>N$, one makes an error of order $\Sigma_{n=N}^{\infty}n^{-2}=\mathcal{O}(N^{-1})\to 0$, whereas for the derivatives, the corresponding error is of order $\Sigma_{n=N}^{\infty}n^{-1}=\mathcal{O}(\log N)\to\infty$. This argument, however, does not apply to our case, because the number of sides of the polygonal path $x_P(t;x_1\cdots x_{N-1})$ is also N, so that the a_n's depend on N.

[22] In fact, we here have exactly $\bar x_P=N^{-1}\Sigma_{j=1}^{N}\frac{1}{2}(x_{j-1}+x_j)=N^{-1}\Sigma_{j=0}^{N-1}x_j=a_0=\bar x_F$; the second equality is because $x_0=x_N$, and

the third one follows from [see (D5)]

$$\sum_{j=0}^{N-1} \cos\left(\frac{2\pi nj}{N}\right) = \sum_{j=0}^{N-1} \sin\left(\frac{2\pi nj}{N}\right) = 0 \text{ for } n \neq 0.$$

[23]For a more elaborate derivation of that phase, see e.g.: Ph. Combe, R. Rodriguez, G. Rideau, and M. Sirugue-Collin, Rep. Math. Phys. **13**, 279 (1978).

[24](a) I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965). Formula (A18) follows from entry 1.422.3 [the entry 1.422.6 is incorrect, and should be replaced by (A18)]; (b) M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), entry 4.3.93.

[25]Reference 24(a): entries 1.431,1, 1.392.1; Ref. 24(b), entries 4.3.91, 4.3.89.

The expression for the triple vector product solid harmonic

T. A. Antone[a)]
Department of Mathematics, Ahmadu Bello University, Zaria, Nigeria

(Received 21 December 1983; accepted for publication 4 May 1984)

The solid harmonic $y_{LM}[(\mathbf{r}_1\wedge\mathbf{r}_2)\wedge\mathbf{r}_3]$ was expressed in terms of the spherical harmonics $Y_{L_1M_1}(\hat{\mathbf{r}}_1)$, $Y_{L_2M_2}(\hat{\mathbf{r}}_2)$, and $Y_{L_3M_3}(\hat{\mathbf{r}}_3)$, where the coefficients of the expansion were expressed in terms of $9j$ symbols. Here we present a simpler form of those coefficients expressed in terms of $6j$ symbols.

PACS numbers: 02.30.Gp

I. INTRODUCTION

The author[1] recently presented the expression

$$Y_{LM}[(\mathbf{r}_1\wedge\mathbf{r}_2)\wedge\mathbf{r}_3]=\sum_{L_1L_2L_{12}L_3M_1M_2}A(L_1,L_2,L_{12},L_3,L)\begin{pmatrix}L_1 & L_2 & L_{12}\\ M_1 & M_2 & -M_{12}\end{pmatrix}\begin{pmatrix}L_{12} & L_3 & L\\ M_{12} & M_3 & -M\end{pmatrix}\times Y_{L_1M_1}(\hat{\mathbf{r}}_1)Y_{L_2M_2}(\hat{\mathbf{r}}_2)Y_{L_3M_3}(\hat{\mathbf{r}}_3),\tag{1}$$

where the coefficients $A(L_1,L_2,L_{12},L_3,L)$ were expressed in terms of $9j$ symbols in the form

$$A(L_1,L_2,L_{12},L_3,L)=(-1)^{L+M_3}(2L_{12}+1)\sum_{rst}(-1)^rK(r,s,t,L_1,L_2,L_3,L)\begin{Bmatrix}L_1 & L_2 & L_{12}\\ s & t & L_3\\ L-r & r & L\end{Bmatrix},\tag{2}$$

where r, s, and t take the values

$r=L, L-1, L-2,...0,$

$s=r, r-2, r-4,...,1$ or 0,

$t=L-r, L-r-2, L-r-4,...,1$ or 0,

and

$$K(r,s,t,L_1,L_2,L_3,L)=4\pi(r_1r_2r_3)^L\begin{pmatrix}s & L-r & L_1\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}t & r & L_2\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}t & s & L_3\\ 0 & 0 & 0\end{pmatrix}\times\frac{2^{1/2(s+t-L)}(2L+1)(2s+1)(2t+1)r!(L-r)!}{[\frac{1}{2}(r-s)]![\frac{1}{2}(L-r-t)]!(r+s+1)!!(L-r+t+1)!!}\left[\frac{(2L_1+1)(2L_2+1)(2L_3+1)(2L)!}{(2r)!(2L-2r)!}\right]^{1/2}.\tag{3}$$

It is the purpose of this article to express those coefficients in terms of $6j$ symbols in the form

$$A(L_1,L_2,L_{12},L_3,L)=(-1)^{M_3-(1/2)(L+L_{12})}(r_1r_2r_3)^LC(L_{12},L_3,L)\sum_{pj_1j_2}C(j_1,j_2,L_{12})B(p,j_1,j_2)\begin{Bmatrix}L_1 & j_1 & 2p\\ j_2 & L_2 & L_{12}\end{Bmatrix},\tag{4}$$

where

$$C(x,y,z)=2^{x+y-z}\left[\frac{4\pi(2z+1)(2x+1)(2y+1)(x-y+z)!(y-x+z)!(x+y+z+1)!}{(x+y-z)!}\right]^{1/2}\times\frac{[\frac{1}{2}(z+x)]![\frac{1}{2}(z+y)]!}{[\frac{1}{2}(z-x)]![\frac{1}{2}(z-y)]!(z+x+1)!(z+y+1)!},\tag{5}$$

with the factorials implying integer arguments and

$$B(p,j_1,j_2)=\frac{n!n!(4p+1)2^{2n}(2p)!(n+p)!\sqrt{(2L_1+1)(2L_2+1)(2j_1+1)(2j_2+1)}}{p!p!(n-p)!(2n+2p+1)!}\begin{pmatrix}j_1 & 2p & L_1\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}j_2 & 2p & L_2\\ 0 & 0 & 0\end{pmatrix},\tag{6}$$

with $n=\frac{1}{2}(L-L_{12})$.

II. CALCULATION OF THE COEFFICIENTS

Hage Hassan *et al.*[2] gave the expansion

$$Y_{jm}(\mathbf{r}_1\wedge\mathbf{r}_2)=(-1)^{m-(1/2)j}(r_1r_2)^j\sum_{j_1j_2m_1}C(j_1,j_2,j)\begin{pmatrix}j_1 & j_2 & j\\ m_1 & m_2 & -m\end{pmatrix}Y_{j_1m_1}(\hat{\mathbf{r}}_1)Y_{j_2m_2}(\hat{\mathbf{r}}_2),\tag{7}$$

[a)] Current address: Department of Applied Mathematical Studies, University of Leeds, England.

 0022-2488/84/102885-03$02.50

which we use to write

$$Y_{LM}[(\mathbf{r}_1\wedge\mathbf{r}_2)\wedge\mathbf{r}_3]=(-1)^{M-(1/2)L}(r_1r_2r_3\sin\theta_{12})^L\sum_{j_{12}L_3u_{12}}C(j_{12},L_3,L)\begin{pmatrix}j_{12}&L_3&L\\u_{12}&M_3&-M\end{pmatrix}Y_{j_{12}u_{12}}(\widehat{\mathbf{r}_1\wedge\mathbf{r}_2})Y_{L_3M_3}(\hat{\mathbf{r}}_3),\tag{8}$$

but

$$(r_1r_2\sin\theta_{12})^{j_{12}}Y_{j_{12}u_{12}}(\widehat{\mathbf{r}_1\wedge\mathbf{r}_2})=Y_{j_{12}u_{12}}(\mathbf{r}_1\wedge\mathbf{r}_2).$$

Therefore on using Eq. (7) once more we write Eq. (8) as

$$\begin{aligned}Y_{LM}[(\mathbf{r}_1\wedge\mathbf{r}_2)\wedge\mathbf{r}_3]=(-1)^{M-(1/2)L}(r_1r_2r_2)^L&\sum_{j_1j_2j_{12}L_3u_{12}}(-1)^{u_{12}-(1/2)L_{12}}C(j_1j_2,j_{12})C(j_{12},L_3,L)\\&\times\begin{pmatrix}j_1&j_2&j_{12}\\u_1&u_2&-u_{12}\end{pmatrix}\begin{pmatrix}j_{12}&L_3&L\\u_{12}&M_3&-M\end{pmatrix}\sin\theta_{12}^{L-j_{12}}Y_{j_1u_1}(\hat{\mathbf{r}}_1)Y_{j_2u_2}(\hat{\mathbf{r}}_2)Y_{L_3M_3}(\hat{\mathbf{r}}_3).\end{aligned}\tag{9}$$

Since $L-j_{12}=2n$ is an even integer we write

$$\sin\theta_{12}^{2n}=(1-\cos^2\theta_{12})^n=\sum_q\frac{(-1)^q n!}{q!(n-q)!}\cos^{2q}\theta_{12},$$

and use the relationship

$$cos^{2q}\theta_{12}=4\pi\sum_{pu}\frac{2^{(p-q)}(2q)!}{(q-p)!(2q+2p+1)!!}Y^*_{2p,u}(\hat{\mathbf{r}}_1)Y_{2p,u}(\hat{\mathbf{r}}_2),\tag{10}$$

together with Eq. (4.6.5) in Edmonds[3] in the form

$$Y_{j_1m_1}(\hat{\mathbf{r}})Y_{j_2m_2}(\hat{\mathbf{r}})=\sum_{j_3m_3}(-1)^{m_3}\sqrt{\frac{(2j_1+1)(2j_2+1)(2j_3+1)}{4\pi}}\begin{pmatrix}j_1&j_2&j_3\\0&0&0\end{pmatrix}\begin{pmatrix}j_1&j_2&j_3\\m_1&m_2&-m_3\end{pmatrix}Y_{j_3m_3}(\hat{\mathbf{r}}),\tag{11}$$

to write

$$\begin{aligned}\sin\theta_{12}^{2n}Y_{j_1u_1}(\hat{\mathbf{r}}_1)Y_{j_2u_2}(\hat{\mathbf{r}}_2)=&\sum_{pqL_1L_2M_1M_2u}\frac{(-1)^{q+u+M_1+M_2}n!(2q)!2^{p-q}(4p+1)}{q!(n-q)!(q-p)!(2p+2q+1)!!}\sqrt{(2j_1+1)(2j_2+1)(2L_1+1)(2L_2+1)}\\&\times\begin{pmatrix}j_1&2p&L_1\\0&0&0\end{pmatrix}\begin{pmatrix}j_2&2p&L_2\\0&0&0\end{pmatrix}\begin{pmatrix}j_1&2p&L_1\\u_1&-u&-M_1\end{pmatrix}\begin{pmatrix}j_2&2p&L_2\\u_2&u&-M_2\end{pmatrix}Y_{L_1M_1}(\hat{\mathbf{r}}_1)Y_{L_2M_2}(\hat{\mathbf{r}}_2),\end{aligned}\tag{12}$$

which requires that $u=u_1-M_1=M_2-u_2$ or $u_1+u_2=M_1+M_2$.

We now use Eq. (6.2.6) in Edmonds in the form

$$\begin{aligned}&\begin{pmatrix}L_1&j_1&2p\\-M_1&u_1&-u\end{pmatrix}\begin{pmatrix}2p&j_2&L_2\\u&u_2&-M_2\end{pmatrix}\\&\quad=(-1)^{u_2+M_1}\sum_{L_{12}}(2L_{12}+1)\begin{pmatrix}j_1&j_2&L_{12}\\u_1&u_2&-M_{12}\end{pmatrix}\begin{pmatrix}L_1&L_2&L_{12}\\M_1&M_2&-M_{12}\end{pmatrix}\begin{Bmatrix}L_1&j_1&2p\\j_2&L_2&L_{12}\end{Bmatrix}\end{aligned}\tag{13}$$

to write Eq. (12) as

$$\begin{aligned}\sin\theta_{12}^{2n}Y_{j_1u_1}(\hat{\mathbf{r}}_1)Y_{j_2u_2}(\hat{\mathbf{r}}_2)=&\sum_{pL_1L_2L_{12}M_1M_2}B(p,j_1,j_2)(2L_{12}+1)\begin{pmatrix}j_1&j_2&L_{12}\\u_1&u_2&-M_{12}\end{pmatrix}\begin{pmatrix}L_1&L_2&L_{12}\\M_1&M_2&-M_{12}\end{pmatrix}\\&\times\begin{Bmatrix}L_1&j_1&2p\\j_2&L_2&L_{12}\end{Bmatrix}Y_{L_1M_1}(\hat{\mathbf{r}}_1)Y_{L_2M_2}(\hat{\mathbf{r}}_2).\end{aligned}\tag{14}$$

Here we have used Eq. (A.1.2) in Edmonds and the duplication formula for the Gamma function to evaluate the summation over q as follows:

$$\sum_q\frac{(-1)^q(2q)!(p+q)!}{q!(n-q)!(q-p)!(2p+2q+1)!}=\frac{2^{n-p}n!(2p)!(n+p)!}{p!p!(n-p)!(2n+2p+1)!}.\tag{15}$$

Substituting Eq. (14) in Eq. (9) and using the orthonormality of the $3j$ symbol given by Eq. (3.7.8) in Edmonds in the form

$$\sum_{u_1u_2}\begin{pmatrix}j_1&j_2&j_{12}\\u_1&u_2&-u_{12}\end{pmatrix}\begin{pmatrix}j_1&j_2&L_{12}\\u_1&u_2&-M_{12}\end{pmatrix}=\frac{1}{2L_{12}+1}\delta_{j_{12}L_{12}}\delta_{u_{12}M_{12}},\tag{16}$$

we finally arrive at Eq. (1) where the coefficients $A(L_1,L_2,L_{12},L_3,L)$ are given by Eq. (4).

[1]T. A. Antone, J. Math. Phys. **24**, 1060 (1983).
[2]H. Hage Hassan, G. Grent, J. P. Gazeau, and K. Kibler, J. Phys. A: Math. Gen. **13**, 2623 (1980).
[3]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton, U.P., Princeton, NJ, 1960).

Eigenvalues and degeneracies for n-dimensional tensor spherical harmonics

Mark A. Rubin
Theory Group, Department of Physics, The University of Texas, Austin, Texas 78712
Carlos R. Ordóñez
Center for Theoretical Physics, Department of Physics, The University of Texas, Austin, Texas 78712

(Received 1 November 1983; accepted for publication 25 April 1984)

Symmetric transverse traceless tensor harmonics of arbitrary rank are constructed on spheres S^n of dimensionality $n \geqslant 3$, and the associated eigenvalues of the Laplacian are computed. It is shown that these tensor harmonics span the space of symmetric transverse traceless tensors on S^n and are eigenfunctions of the quadratic Casimir operator of the group $O(n+1)$. The dimensionalities of the eigenspaces of the Laplacian are computed for harmonics of rank 1 and rank 2.

PACS numbers: 02.30.Gp, 02.20. + b, 02.40. + m, 04.60. + n

I. INTRODUCTION

The task of computing determinants of second-order differential operators is a common one in quantum field theory. In the classical theory of a tensor field $t_{c_1\cdots c_m}(x)$ the equations of motion are obtained by varying the action functional $S[t_{c_1\cdots c_m}]$; in the corresponding quantum field theory, the equations which determine the expectation value $\bar{t}_{c_1\cdots c_m}(x)$ of the field in the vacuum state are obtained by varying the effective action functional $\Gamma[\bar{t}_{c_1\cdots c_m}]$. To first order in $\hbar$, Γ is given by the expression[1]

$$\Gamma[\bar{t}_{c_1\cdots c_m}] = S^E[\bar{t}_{c_1\cdots c_m}] + \tfrac{1}{2}\ln\det S_2^E[\bar{t}_{c_1\cdots c_m}], \quad (1.1)$$

where S^E is the Euclideanized action functional, and S_2^E is the second functional derivative of S^E with respect to $\bar{t}_{c_1\cdots c_m}(x)$.

The formal expression $\ln\det S_2^E$ must be calculated by a suitable regularization procedure which assigns to this quantity a finite value. A convenient regularization procedure is the zeta-function method[2]:

$$\ln\det S_2^E = -\frac{d}{ds}(\zeta_{S_2^E}(s))|_{s=0}, \quad (1.2)$$

where the generalized zeta function $\zeta_{S_2^E}(s)$ is defined as

$$\zeta_{S_2^E}(s) \equiv \sum_l d_l(\lambda_l)^{-s}. \quad (1.3)$$

The sum in (1.3) is taken over all distinct nonzero eigenvalues λ_l of S_2^E; d_l is the degeneracy of λ_l, i.e., the number of linearly independent eigenfunctions associated with λ_l. [The actual summation in Eq. (1.3) is performed for values of the parameter s large enough that the sum converges; the value of $\zeta_{S_2^E}(s)$ for other values of s is determined by analytic continuation in s.]

S_2^E is usually a second-order differential operator on tensor fields $t_{c_1\cdots c_m}(x)$, and it often occurs that λ_l and d_l may be simply determined from a knowledge of the eigenvalues and degeneracies of the covariant Laplacian $\nabla^a\nabla_a$ acting on tensor fields of the type in question. In the case that the manifold on which the tensor fields are defined is an n-dimensional sphere S^n, or contains S^n as a subspace, the computation of the effective action Γ requires a knowledge of the eigenvalues Λ_l and degeneracies D_l of the tensor spherical harmonics, i.e., tensor fields $T_{c_1\cdots c_m}(x)$ on S^n satisfying

$$\tilde{\nabla}^a\tilde{\nabla}_a T^{(l)}_{c_1\cdots c_m(x)} = \Lambda_l T^{(l)}_{c_1\cdots c_m(x)}. \quad (1.4)$$

(Throughout this paper $\tilde{\nabla}_a$ denotes the covariant derivative operator on S^n.)

If the quantum field theory under study involves the gravitational field, it is necessary to know the values of Λ_l and D_l for scalar fields, vector fields, and symmetric second-rank tensor fields,[3,4] i.e., $m = 0,1$, and 2. For $m = 0$ the Λ_l's and D_l's are associated with the usual (scalar) n-spherical harmonics and may be found in standard references.[5] For $m = 1$ they have been computed for arbitrary n in Ref. 6. However, to the best of our knowledge, the eigenvalues and degeneracies for second-rank symmetric tensor harmonics have only been computed[7,8] for spheres of dimension $\leqslant 4$. The calculation of quantum-gravitational effects in Kaluza–Klein theories in which the internal dimensions form an n-sphere[9,10] requires the knowledge of Λ_l and D_l for $m = 2$ and arbitrary n; this is the motivation behind the present work.

This paper is organized as follows: In Sec. II we review the properties and construction of scalar harmonics on S^n. In Sec. III we employ the methods of Refs. 3 and 8 to construct "canonical" sets of symmetric transverse traceless (ST^2) tensor harmonics of arbitrary rank and compute the associated eigenvalues Λ_l [Eq. (3.9b)]. We demonstrate that these canonical tensors span the space of ST^2 tensors on S^n. (Rank 1 and rank 2 tensors on compact spaces can always be decomposed in terms of scalars and ST^2 tensors[4,11,12]; there is no loss of generality in considering only ST^2 tensor harmonics, since from these the eigenvalues and degeneracies associated with non-ST^2 harmonics can be obtained.[4,13]) In Sec. IV we demonstrate that the canonical tensors are eigenfunctions of the quadratic Casimir operator of $O(n+1)$; using this information, we compute the degeneracies D_l for ST^2 tensor harmonics of rank 1 and rank 2 [Eqs. (4.37a) and (4.37b)].

II. SCALAR HARMONICS

The eigenvalues $\Lambda_l(n,0)$ and degeneracies $D_l(n,0)$ of $\tilde{\nabla}^a\tilde{\nabla}_a$ acting on scalar functions on S^n of radius r are[5]

$$\Lambda_l(n,0) = -\frac{l(l+n-1)}{r^2}, \quad (2.1a)$$

$$D_l(n,0) = \frac{(l+n-2)!}{l!(n-1)!}(2l+n-1), \quad l = 0,1,\ldots. \quad (2.1b)$$

 0022-2488/84/102888-07$02.50

The associated eigenfunctions can be most easily constructed if we imbed S^n in $(n+1)$-dimensional Euclidean space R^{n+1}. Denote the Cartesian coordinates on R^{n+1} by $x^{\bar{\alpha}}, \bar{\alpha} = 0,1,\ldots,n$. (Greek indices will always take on the $n+1$ values $0,\ldots,n$; bars over indices will always denote Cartesian components.) Let $C_{\bar{\alpha}_1\cdots\bar{\alpha}_l}$ be a symmetric traceless rank-l tensor on R^{n+1} whose components in the Cartesian coordinate system are constant. Then, for any such tensor $C_{\bar{\alpha}_1\cdots\bar{\alpha}_l}$, the function

$$T^{(l)}(x) = r^{-l} C_{\bar{\alpha}_1\cdots\bar{\alpha}_l} x^{\bar{\alpha}_1}\cdots x^{\bar{\alpha}_l}, \qquad r = (x^{\bar{\alpha}} x_{\bar{\alpha}})^{1/2}, \tag{2.2}$$

upon restriction to S^n of radius r, is an eigenfunction of $\tilde{\nabla}^a \tilde{\nabla}_a$ corresponding to $\Lambda_l(n,0)$.

We can see that the functions (2.2) include all eigenfunctions of the S^n Laplacian, since the number of linearly independent symmetric traceless rank-l tensors in R^{n+1} is precisely $D_l(n,0)$.[14] To demonstrate that the $T^{(l)}$'s are indeed eigenfunctions of $\tilde{\nabla}^a \tilde{\nabla}_a$, we shall derive a formula, Eq. (2.9), which will also prove useful in the tensor case.

We introduce spherical polar coordinates in R^{n+1}:

$$x^\alpha = (x^0, x^a), \quad a = 1,\ldots,n, \qquad x^0 = r = (x^{\bar{\alpha}} x_{\bar{\alpha}})^{1/2}. \tag{2.3a}$$

(Latin indices always take on the n values $1,\ldots,n$.) The metric in these coordinates takes the form

$$g_{\alpha\beta} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \tilde{g}_{ab} \end{pmatrix}, \tag{2.3b}$$

where $\tilde{g}_{ab}$ is the metric on the unit S^n and is a function only of the n coordinates x^a. The only nonzero Christoffel symbols[15] are

$$\Gamma^0_{ab} = -r\tilde{g}_{ab}, \qquad \Gamma^a_{0b} = \Gamma^a_{b0} = (1/r)\delta^a_b, \qquad \Gamma^a_{bc} = \tilde{\Gamma}^a_{bc}, \tag{2.4}$$

where $\tilde{\Gamma}^a_{bc}$ is the Christoffel symbol constructed out of $\tilde{g}_{ab}$.

Let $t_{\gamma_1\cdots\gamma_m}$ be a symmetric mth-rank tensor on R^{n+1} which is also in the tangent space to S^n. That is, the Cartesian components $t_{\bar{\gamma}_1\cdots\bar{\gamma}_m}$ satisfy

$$x^{\bar{\gamma}_1} t_{\bar{\gamma}_1\cdots\bar{\gamma}_m} = 0, \tag{2.5a}$$

and the only nonzero components of $t_{\gamma_1\cdots\gamma_m}$ in spherical polar coordinates are those of the form

$$t_{\gamma_1\cdots\gamma_m} = t_{c_1\cdots c_m} \quad \text{(i.e., none of the } \gamma\text{'s} = 0). \tag{2.5b}$$

Working in spherical polar coordinates, and using (2.4), (2.5), and the symmetry of $t_{\gamma_1\cdots\gamma_m}$,

$$\nabla^\alpha \nabla_\alpha t_{c_1\cdots c_m} = (\nabla^0\nabla_0 + \nabla^a\nabla_a) t_{c_1\cdots c_m}, \tag{2.6}$$

$$\nabla^0\nabla_0 t_{c_1\cdots c_m} = \left(\frac{\partial}{\partial r} - \frac{m}{r}\right)\left(\frac{\partial}{\partial r} - \frac{m}{r}\right) t_{c_1\cdots c_m}, \tag{2.7}$$

$$\nabla^a\nabla_a t_{c_1\cdots c_m} = \left(\tilde{\nabla}^a\tilde{\nabla}_a + \frac{n}{r}\frac{\partial}{\partial r} - \frac{m(n+1)}{r^2}\right) t_{c_1\cdots c_m}, \tag{2.8}$$

where ∇^α is the covariant derivative operator on R^{n+1}. We therefore obtain the following relation between the Laplacians on R^{n+1} and S^n acting on symmetric tensors satisfying (2.5):

$$\tilde{\nabla}^a\tilde{\nabla}_a t_{c_1\cdots c_m} = \left[\nabla^\alpha\nabla_\alpha - \left(\frac{\partial}{\partial r} + \frac{n-m}{r}\right) \times \left(\frac{\partial}{\partial r} - \frac{m}{r}\right) + \frac{m}{r^2}\right] t_{c_1\cdots c_m}. \tag{2.9}$$

(This equation has been given in Ref. 3.)

We now apply (2.9) to the rank-zero tensor $T^{(l)}$ given by (2.2). Since $\nabla^\alpha\nabla_\alpha T^{(l)} = \nabla^{\bar{\alpha}}\nabla_{\bar{\alpha}} T^{(l)}$, we find that $T^{(l)}$ does indeed satisfy

$$\tilde{\nabla}^a\tilde{\nabla}_a T^{(l)} = \Lambda_l(n,0) T^{(l)}, \tag{2.10}$$

with $\Lambda_l(n,0)$ as given in (2.1a).

III. TENSOR HARMONICS: EIGENVALUES

In analogy with the scalar harmonics, ST^2 tensor harmonics of rank m can be constructed by restricting to S^n tensors on R^{n+1}, provided these tensors satisfy (2.5). The Cartesian components of these tensors are

$$T^{(l)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m} = r^{-l} C_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m\bar{\alpha}_1\cdots\bar{\alpha}_{l-m}} x^{\bar{\beta}_1}\cdots x^{\bar{\beta}_m}\cdots x^{\bar{\alpha}_1}\cdots x^{\bar{\alpha}_{l-m}}, \quad m = 0,1,\ldots, \quad l = m, m+1,\ldots. \tag{3.1}$$

The quantity $C_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\alpha}_{l-m}}$ is an $(l+m)$th-rank tensor on R^{n+1} whose Cartesian components are constant, and which has the following properties: it is symmetric in all the $\bar{\alpha}_i$ indices; it is antisymmetric under interchange of $\bar{\gamma}_i$ with the $\bar{\beta}_i$ to the right; it is traceless under contraction of any pair of indices; it is symmetric under interchange of any pair $\bar{\gamma}_i\,\bar{\beta}_i$ with any other pair $\bar{\gamma}_j\,\bar{\beta}_j$.

(The above conditions insure that $T^{(l)}_{\gamma_1\cdots\gamma_m}$ will be an eigenfunction of $\tilde{\nabla}^a\tilde{\nabla}_a$. Other authors[3,8] impose additional constraints on $C_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m\bar{\alpha}_1\cdots\bar{\alpha}_{l-m}}$ which insure that there will be precisely one linearly independent Cartesian tensor per linearly independent tensor harmonic. In principle, one can then, as in the scalar case, determine the degeneracy of each eigenvalue by counting the number of linearly independent $C_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m\bar{\alpha}_1\cdots\bar{\alpha}_{l-m}}$ tensors satisfying all the conditions. In practice, this counting is quite difficult and is rendered unnecessary by the method of Sec. IV.)

We first make use of Eq. (2.9) to demonstrate that the tensors (3.1) are eigenfunctions of $\tilde{\nabla}^a\tilde{\nabla}_a$. In employing (2.9) with $m \neq 0$ we must keep in mind that this equation is only valid in a coordinate system of the spherical polar sort (2.3). The term $\nabla^\alpha\nabla_\alpha t_{c_1\cdots c_m}$ is, of course, covariant under arbitrary coordinate transformations in R^{n+1}, and may be evaluated in any coordinate system (the natural choice being the Cartesian system). However, the remaining terms are only covariant under coordinate transformations on S^n; that is, coordinate transformations which do not mix up the coordinate $x^0 = r$ with the angular coordinates x^a. So to make use of Eq. (2.9) we must first obtain the components of $T^{(l)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m}$ in a spherical polar coordinate system.

Pick a point $\mathscr{P}$ on the sphere on radius r, and orient the $(n+1)$-dimensional Cartesian frame so that at $\mathscr{P}$ the $x^{\bar{0}}$ direction is normal to the sphere. The $x^{\bar{a}}$ directions will then be

tangent to the sphere at $\mathscr{P}$. Choose coordinates x^a in S^n such that, at $\mathscr{P}$, the coordinate axes are parallel to the $x^{\bar{a}}$ axes and the metric is proportional to the Kronecker delta

$$r^2\tilde{g}_{ab} = r^2\tilde{g}^{1/n}\delta_{ab}, \tag{3.2}$$

where $\tilde{g} = \det \tilde{g}_{ab}$. The square norm of a vector tangent to S^n at $\mathscr{P}$ and pointing in the a' direction is

$$V^{\bar{a}'}V_{\bar{a}'} = V^{a'}V_{a'} \quad \text{(no summation)} \tag{3.3}$$

or

$$(V^{\bar{a}'})^2 = r^2 g^{1/n}(V^{a'})^2 \tag{3.4}$$

using (2.3b) and (3.2);

$$V^{\bar{a}'} = r\tilde{g}^{1/2n}V^{a'}. \tag{3.5}$$

But the components of a vector in two different coordinate systems are related by[15]

$$V^{\bar{a}'} = \frac{\partial x^{\bar{a}'}}{\partial x^a}V^a. \tag{3.6}$$

Hence

$$\frac{\partial x^{\bar{a}}}{\partial x^a} = r\tilde{g}^{1/2n}\delta^{\bar{a}}_a, \quad \frac{\partial x^{\bar{0}}}{\partial x^a} = 0. \tag{3.7}$$

This gives us the components of the tensor harmonic (3.1) in a coordinate system of the form (2.3):

$$T^{(l)}_{c_1\cdots c_m} = r^m\tilde{g}^{m/2n}T^{(l)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m}. \tag{3.8}$$

Using (2.9), (3.1), and (3.8), we find that

$$\tilde{\nabla}^a\tilde{\nabla}_a T^{(l)}_{c_1\cdots c_m} = \Lambda_l(n,m)T^{(l)}_{c_1\cdots c_m}, \tag{3.9a}$$

$$\boxed{\Lambda_l(n,m) = -[l(l+n-1)-m]/r^2.} \tag{3.9b}$$

We now demonstrate that the set of all tensors of the form (3.1) spans the space of ST^2 tensors on S^n.

Let $t_{\bar{\gamma}_1\cdots\bar{\gamma}_m}(x)$ be an arbitrary R^{n+1} tensor tangent to S^n [i.e., satisfying (2.5)]. Then if $V^a_{\bar{\gamma}}$ is the $\bar{\gamma}$th Cartesian component of the ath linearly independent (covariant) vector tangent to the sphere ($a = 1,\ldots,n$),

$$t_{\bar{\gamma}_1\cdots\bar{\gamma}_m}(x) = t_{a_1\cdots a_m}(x)V^{a_1}_{\bar{\gamma}_1}(x)\cdots V^{a_m}_{\bar{\gamma}_m}(x). \tag{3.10}$$

But each vector $V^a_{\bar{\gamma}}$ can be expressed as

$$V^a_{\bar{\gamma}}(x) = \hat{c}^a_{\bar{\gamma}\bar{\beta}}(x)(x^{\bar{\beta}}/r), \tag{3.11}$$

where $\hat{c}^a_{\bar{\gamma}\bar{\beta}}(x)$ is antisymmetric in $\bar{\gamma}$ and $\bar{\beta}$. This can be easily seen by considering the position vector

$$y^{\bar{\beta}} = \begin{pmatrix}1\\0\\0\\\vdots\end{pmatrix}.$$

Any vector normal to this can be obtained by contracting $y^{\bar{\beta}}$ with an $n+1$ by $n+1$ tensor of the form

$$\hat{c}^a_{\bar{\gamma}\bar{\beta}}(y) = \begin{pmatrix}0 & -c_1 & \cdots & -c_n\\ c_1 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots\\ c_n & 0 & \cdots & 0\end{pmatrix}. \tag{3.12}$$

The $\hat{c}^a_{\bar{\gamma}\bar{\beta}}(x)$'s at other points $x^{\bar{\beta}}$ on S^n can be obtained from (3.12) by coordinate transformations which will, of course, preserve the antisymmetry in $\bar{\gamma}$ and $\bar{\beta}$. Group theoretically, the $\hat{c}^a_{\bar{\gamma}\bar{\beta}}(x)$'s are those generators of SO($n+1$) which are not contained in that SO(n) subgroup of SO($n+1$) which leaves the position vector $x^{\bar{\beta}}$ unchanged.

Using (3.11), (3.10) becomes

$$t_{\bar{\gamma}_1\cdots\bar{\gamma}_m}(x) = r^{-l}\hat{t}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m}(x)x^{\bar{\beta}_1}\cdots x^{\bar{\beta}_m}, \tag{3.13}$$

where

$$\hat{t}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m}(x) \equiv t_{a_1\cdots a_m}(x)\hat{c}^{a_1}_{\bar{\gamma}_1\bar{\beta}_1}(x)\cdots\hat{c}^{a_m}_{\bar{\gamma}_m\bar{\beta}_m}(x). \tag{3.14}$$

Expand each component of $\hat{t}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m}(x)$ in terms of scalar spherical harmonics (the index q distinguishes between different harmonics corresponding to the same eigenvalue):

$$\hat{t}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m}(x) = \sum_{l=0}^{\infty}\sum_q \hat{c}^{(l,q)}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m}T^{(l,q)}(x), \tag{3.15a}$$

where the $\hat{c}^{(l,q)}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m}$'s are constants. Using (2.2),

$$\begin{aligned}&\hat{t}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m}(x)\\ &= \sum_{l=0}^{\infty}\sum_q \hat{c}^{(l,q)}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m}r^{-l}C^{(l,q)}_{\bar{\alpha}_1\cdots\bar{\alpha}_l}x^{\bar{\alpha}_1}\cdots x^{\bar{\alpha}_l}\\ &= \sum_{l=m}^{\infty}\sum_q \hat{c}^{(l-m,q)}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m}r^{-(l-m)}C^{(l-m,q)}_{\bar{\alpha}_1\cdots\bar{\alpha}_{l-m}}x^{\bar{\alpha}_1}\cdots x^{\bar{\alpha}_{l-m}}.\end{aligned} \tag{3.15b}$$

Using (3.15) and (3.14) in (3.13), and defining

$$c^{(l,q)}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m\bar{\alpha}_1\cdots\bar{\alpha}_m} \equiv \hat{c}^{(l-m,q)}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m}C^{(l-m,q)}_{\bar{\alpha}_1\cdots\bar{\alpha}_{l-m}}, \tag{3.16}$$

we obtain

$$\begin{aligned}&t_{\bar{\gamma}_1\cdots\bar{\gamma}_m}(x)\\ &= \sum_{l=m}^{\infty}\sum_q r^{-l}c^{(l,q)}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m\bar{\alpha}_1\cdots\bar{\alpha}_{l-m}}x^{\bar{\beta}_1}\cdots x^{\bar{\beta}_m}x^{\bar{\alpha}_1}\cdots x^{\bar{\alpha}_{l-m}}.\end{aligned} \tag{3.17}$$

Thus, an arbitrary mth-rank tensor tangent S^n can be written as a linear combination of tensors of the form

$$\begin{aligned}&t^{(l)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m}(x)\\ &= r^{-l}c^{(l)}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m\bar{\alpha}_1\cdots\bar{\alpha}_{l-m}}x^{\bar{\beta}_1}\cdots x^{\bar{\beta}_m}x^{\bar{\alpha}_1}\cdots x^{\bar{\alpha}_{l-m}},\end{aligned} \tag{3.18}$$

where the R^{n+1}-tensor with constant Cartesian components $c^{(l)}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m\bar{\alpha}_1\cdots\bar{\alpha}_{l-m}}$ is antisymmetric in each $\bar{\gamma}_i\,\bar{\beta}_i$ pair, and symmetric and traceless in the α indices. In general, there will be more than one way to express a general tensor in terms of the tensors (3.18) if we add no further constraints on the $c^{(l)}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m\bar{\alpha}_1\cdots\bar{\alpha}_{l-m}}$'s.

A necessary and sufficient condition that $t^{(l)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m}$ be traceless is that $c^{(l)}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m\bar{\alpha}_1\cdots\bar{\alpha}_{l-m}}$ be traceless in the $\bar{\gamma}$ indices (hence traceless in all the $\bar{\gamma}$ and $\bar{\beta}$ indices, due to the antisymmetry under the $\bar{\gamma}_i \leftrightarrow \bar{\beta}_i$ interchange). Imposing this constraint, and taking the R^{n+1}-divergence of (3.18), we find that if $t^{(l)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m}$ is traceless, a necessary and sufficient condition for $\partial^{\bar{\gamma}_1}t^{(l)}_{\bar{\gamma}_1\bar{\gamma}_2\cdots\bar{\gamma}_m}$ to vanish is that $c^l_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m\bar{\alpha}_1\cdots\bar{\alpha}_{l-m}}$ be totally traceless. However, using (2.4) it can be shown that $\partial^{\bar{\gamma}_1}t^{(l)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m} = 0$ implies $\tilde{\nabla}^{c_1}t^{(l)}_{c_1c_2\cdots c_m} = 0$, and vice versa, for any $t^{(l)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m}$ tangent to S^n.

Thus, any transverse traceless tensor on S^n can be expressed as a linear combination of tensors of the form (3.18), where $c^{(l)}_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m\bar{\alpha}_1\cdots\bar{\alpha}_{l-m}}$ satisfies all the constraints placed on the $C_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\gamma}_m\bar{\beta}_m\bar{\alpha}_1\cdots\bar{\alpha}_{l-m}}$'s of (3.1) except for the symmetry

under interchange of $\bar{\gamma}_i\,\tilde{\beta}_i$ pairs, which constraint is clearly necessary and sufficient for $t^{(l)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m}$ to be symmetric. We conclude that any ST^2 tensor can in fact be constructed out of the $T^{(l)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m}$'s defined in (3.1), though not necessarily in a unique manner.

IV. TENSOR HARMONICS: DEGENERACIES

Under the action of O($n+1$), the group of rotations and inversions about the origin in R^{n+1}, each of the canonical tensor harmonics (3.1) transforms into another tensor of the same form; i.e., a tensor harmonic corresponding to the same values of l and m, but (in general) with a different $C_{\bar{\gamma}_1\bar{\beta}_1\cdots\bar{\alpha}_{l-m}}$. We shall show below that, for all $n\geqslant 3$ and for $m=0$, 1, and 2, the set of canonical tensors corresponding to a given l and m transforms *irreducibly* under the action of O($n+1$). We shall determine the irreducible representation (irrep) corresponding to a given l, m, and n; the degeneracy $D_l(m,n)$ of the Laplacian eigenvalue Λ_l(m,n) is then given by the well-known formulas for the dimensionalities of irreps of O($n+1$).[17]

The simplest case is, naturally, $m=0$. We have already established, in Sec. II of this paper, an isomorphism between the scalar harmonics of degree l on S^n and constant symmetric traceless Cartesian tensors of rank l on R^{n+1}. Such tensors are known to transform according to the irrep of O($n+1$) labeled by the dominant weight vector (l,0,...,0).[18] (The number of components of the weight vector is $[(n+1)/2]\equiv$the integer part of $(n+1)/2$). The scalar harmonics, therefore, also transform according to this irrep. The dimensionality of this irrep is, in fact, equal to the scalar degeneracy (2.1b).

We now turn to the transformation properties of the $m\neq 0$ harmonics. Each Cartesian component of a given tensor harmonic can be uniquely decomposed in terms of scalar harmonics:

$$T^{(l)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m(x)} = \sum_q \hat{T}^{(l,q)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m} T^{(l,q)}(x), \tag{4.1}$$

where $\hat{T}^{(l,q)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m}$ is constant, and q labels distinct scalar harmonics corresponding to the same value of l. The discussion above shows that, if we keep the coordinates $x^{\bar{\alpha}}$ fixed and choose a new Cartesian basis related to the old one by an O($n+1$) transformation, the quantities $T^{(l,q)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m}$ will transform according to the (m,0,...) irrep of O($n+1$). Conversely, if we keep the basis vectors fixed and perform an O($n+1$) transformation on the $x^{\bar{\alpha}}$ the $T^{(l,q)}(x)$ will transform according to the (l,0...) irrep. So, $T^{(l)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m(x)}$ transforms according to the direct product (l,0,...)$\otimes$(m,0,...). For $l\geqslant m$ and $n\geqslant 3$, this direct product decomposes into the following irreps,[19,20] each with multiplicity 1:

$$\begin{aligned}
&(l,0,\ldots,)\otimes(m,0,\ldots)\\
&\quad=(l+m,0,\ldots)\oplus(l+m-2,0,\ldots)\oplus\cdots\oplus(l-m,0,\ldots)\\
&\qquad\oplus(l+m-1,1,\ldots)\oplus(l+m-3,1,\ldots)\cdots\\
&\qquad\oplus(l-m+1,1,\ldots)\\
&\qquad\oplus\cdots\\
&\qquad\vdots\\
&\qquad\oplus(l,m,0,\ldots).
\end{aligned}\tag{4.2}$$

In other words, $T^{(l)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m(x)}$ can be written as a sum of tensors each of which transforms according to one of the irreps on the rhs of (4.2). Our goal is to show that, in fact, only one of the tensors in this sum is nonzero, and to identify the irrep to which it belongs.

To accomplish this, we shall first show that each $T^{(l)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m(x)}$ is an eigenfunction of the quadratic Casimir operator C_2 of O($n+1$), with an eigenvalue equal to the C_2 eigenvalue of the (l,m,0,...) irrep of O($n+1$). We shall then show that, for $m=1$ and $m=2$ (the relevant cases in quantum gravity), none of the other irreps on the rhs of (4.2) has the same C_2 eigenvalue, no matter what the value of l or $n\geqslant 3$. We shall thus have established that the $T^{(l)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m(x)}$'s transform according to (l,m,0,...).

The quadratic Casimir operator C_2 is defined as[21]

$$C_2 = g^{ij}X_iX_j, \tag{4.3}$$

X_i is the ith generator of the group, and g^{ij} is the group metric. The indices i and j in (4.3) are summed over all the generators of the group. To determine the action of C_2 on $T^{(l)}_{\bar{\gamma}_1\cdots\bar{\gamma}_m}(x)$, we require a representation of X_i as a differential operator on tensors.

If a continuous group acts on a manifold to produce coordinate transformations, the action of the generator X_i on tensors on the manifold is that of $\mathcal{L}_{\xi_i}$, the Lie derivative operator in the direction of the Killing vector field $\xi_i^{\ a}(x)$ corresponding to X_i (see Refs. 22–24). The coordinate representation of C_2 is therefore

$$C_2(x) = (1/N)\,g^{ij}\mathcal{L}_{\xi_i}\mathcal{L}_{\xi_j}. \tag{4.4}$$

We have introduced the "textbook-dependent normalization constant" $(1/N)$ so that we may choose our normalization for g^{ij} and $\xi_i^{\ a}(x)$ in a manner which is convenient for our calculation[25,26]:

$$g^{ij}\xi_i^{\ a}(x)\xi_j^{\ b}(x) = g^{ab}(x), \tag{4.5}$$

where $g^{ab}(x)$ is the (contravariant) metric tensor on S^n (g^{ij} is, or course, independent of x^a). The result of $C_2(x)$ acting on scalars, which we already know to belong to (l,0,...), can then be used to fix $(1/N)$ so as to agree with the conventions of any given author.

When acting on a scalar, $\mathcal{L}_{\xi_i}=\xi_i^{\ a}\partial_a$. Furthermore, Lie differentiation leaves tensor character unchanged. Therefore,

$$\begin{aligned}
g^{ij}\mathcal{L}_{\xi_i}\mathcal{L}_{\xi_j}T^{(l)} &= g^{ij}\xi_i^{\ a}\partial_a\xi_j^{\ b}\partial_b T^{(l)}\\
&= [\partial^a\partial_a + g^{ij}\xi_i^{\ a}(\partial_a\xi_j^{\ b})\partial_b]\,T^{(l)}
\end{aligned}\tag{4.6}$$

using (4.5) ($\partial^a\equiv g^{ab}\partial_b$). The covariant constancy of the metric yields

$$0=\tilde{\nabla}_a\,g^{bc} = g^{ij}[\xi_i^{\ b}\tilde{\nabla}_a\xi_j^{\ c} + \xi_j^{\ c}\tilde{\nabla}_a\xi_i^{\ b}]. \tag{4.7}$$

Equations (4.5) and (4.7), combined with the equation satisfied by a Killing vector,

$$\tilde{\nabla}^a\xi_i^{\ b} + \tilde{\nabla}^b\xi_i^{\ a} = 0, \tag{4.8}$$

show that

$$g^{ij}\xi_i^{\ a}(\partial_a\xi_j^{\ b}) = -\Gamma^b_{ac}\,g^{ac}. \tag{4.9}$$

Therefore,

$$g^{ij}\mathcal{L}_{\xi_i}\mathcal{L}_{\xi_j}T^{(l)} = \tilde{\nabla}^a\tilde{\nabla}_a T^{(l)} = -[l(l+n-1)/r^2]T^{(l)}. \tag{4.10}$$

Making use of Ref. 27, we find that the C_2 eigenvalue of the $(p,q,0,...)$ irrep of $O(n+1)$ is

$$C_2(p,q) = 2[p(p+n-1)+q(q+n-3)], \tag{4.11}$$

so the C_2 eigenvalue of $(l,0,...)$ is

$$C_2(l,0) = 2l(l+n-1). \tag{4.12}$$

Comparing (4.10) and (4.12), we conclude

$$C_2(x) = -2r^2 g^{ij}\mathcal{L}_{\xi_i}\mathcal{L}_{\xi_j}. \tag{4.13}$$

The Lie derivative of a covariant tensor of rank m is

$$\mathcal{L}_{\xi_i}t_{c_1\cdots c_m} = \xi_i^a\partial_a t_{c_1\cdots c_m} + \sum_{R=1}^{m} t_{c_1\cdots c_{R-1}ac_{R+1}\cdots c_m}\partial_{c_R}\xi_i^a. \tag{4.14}$$

Using[28] (4.5), (4.14), and the symmetry of $T^{(l)}_{c_1\cdots c_m}(x)$,

$$\begin{aligned} g^{ij}\mathcal{L}_{\xi_i}\mathcal{L}_{\xi_j}T^{(l)}_{c_1\cdots c_m} &= \partial^a\partial_a T^{(l)}_{c_1\cdots c_m} + \partial_a T^{(l)}_{c_1\cdots c_m} g^{ij}\xi_i^b\partial_b\xi_j^a \\ &\quad + 2m\partial_a T^{(l)}_{b(c_1\cdots c_{m-1}|}g^{ij}\xi_i^a\partial_{|c_m)}\xi_j^b \\ &\quad + mT^{(l)}_{a(c_1\cdots c_{m-1}|}g^{ij}\xi_i^b\partial_b\partial_{|c_m)}\xi_j^a \\ &\quad + mT^{(l)}_{a(c_1\cdots c_{m-1}}g^{ij}\partial_{c_m)}\xi_i^b\partial_b\xi_j^a \\ &\quad + m(m-1)T^{(l)}_{ab(c_1\cdots c_{m-2}}g^{ij}\partial_{c_{m-1}|}\xi_i^{(a}\partial_{|c_m)}\xi_j^{b)}. \end{aligned} \tag{4.15}$$

We now proceed to evaluate the various combinations of Killing vectors and their derivatives which appear on the rhs of (4.15). In addition to (4.5) and (4.8), the following formula,[16] derivable from (4.8), will prove useful:

$$\nabla_a\nabla_b\xi_{ic} = -R^d_{\ abc}\xi_{id}. \tag{4.16}$$

Using these we obtain

$$g^{ij}\xi_i^{(a}\partial_b\xi_j^{c)} = -\Gamma^{(a}_{bd}g^{c)d}, \tag{4.17a}$$

and

$$\begin{aligned} g^{ij}\xi_i^a\partial_b\partial_c\xi_j^d &= -2\Gamma^d_{e(c|}g^{ij}\xi_i^a\partial_{|b)}\xi_j^e + \Gamma^e_{bc}g^{ij}\xi_i^a\partial_e\xi_j^d \\ &\quad -(\partial_b\Gamma^d_{ec})g^{ea} + \Gamma^e_{bc}\Gamma^d_{ef}g^{fa} \\ &\quad -\Gamma^d_{eb}\Gamma^e_{fc}g^{fa} - g^{de}R^a_{\ bce}. \end{aligned} \tag{4.17b}$$

It will prove convenient to reexpress the quantities in the last two terms of (4.15) using

$$g^{ij}\partial_c\xi_i^a\partial_d\xi_j^b = g^{ij}\partial_c(\xi_i^a\partial_d\xi_j^b) - g^{ij}\xi_i^a\partial_c\partial_d\xi_j^b. \tag{4.18}$$

We require the antisymmetric part of $g^{ij}\xi_i^a\partial_b\xi_j^c$, in addition to the symmetric part (4.17a). It will simplify not only the calculation of this quantity but the rest of the calculation as well to employ "equatorial plane projection" (EP^2) coordinates on the sphere. In this coordinate system each point on S^n is simply labeled by n of its $n+1$ Cartesian coordinates in R^{n+1}, say $x^{\bar{a}}$, $\bar{a} = 1,...,n$. Such a system, of course, assigns *one* set of coordinates to *two* different points, one in the "northern hemisphere" and the other in the "southern hemisphere." [The "equator" in these coordinates is the $(n-1)$-sphere which is the intersection of S^n with the plane in R^{n+1} $x^{\bar{0}} = 0$.] Since, however, we shall only be dealing with quantities defined at a single point on S^n, rather than integrated quantities, this will not be a drawback. We shall denote the EP^2 coordinate system by hats over the indices: $x^{\hat{a}}$, $\hat{a} = 1,...n$.

The components of the Killing vectors on S^n in the EP^2 coordinates are given explicitly in Chap. 13 of Ref. 16. They are

$$\xi_I^{\hat{a}} = \Omega_I^{\hat{a}}{}_{\hat{b}}x^{\hat{b}}, \tag{4.19}$$

$$\xi_p^{\hat{a}} = \alpha_p^{\hat{a}}\left[1 - \frac{1}{r^2}\sum_{\hat{a}=1}^{n}(x^{\hat{a}})^2\right]^{1/2}. \tag{4.20}$$

The quantities $\Omega_I^{\hat{a}}{}_{\hat{b}}$ and $\alpha_p^{\hat{a}}$ are both independent of $x^{\hat{a}}$. the $\Omega_I^{\hat{a}}{}_{\hat{b}}$'s are matrix representations of the generators of $SO(n)$; the Killing vectors (4.19) generate rotations of the sphere about the axis $x^{\bar{a}} = 0$, $a = 1,...n$. The quantities $\alpha_p^{\hat{a}}$ are n linearly independent tangent vectors to S^n to $x^{\hat{a}} = 0$. The Killing vectors (4.20) correspond to the generators of $SO(n+1)$ which do not leave the axis $x^{\bar{a}} = 0$ fixed. From (4.19) and (4.20) we obtain the formula

$$g^{ij}\xi_i^{[\hat{a}}\partial_{\hat{b}}\xi_j^{\hat{c}]} = g^{IJ}\Omega_I^{[\hat{a}}{}_{\hat{d}}\Omega_J^{\hat{c}]}{}_{\hat{b}}x^{\hat{d}}. \tag{4.21}$$

The Christoffel symbols in the EP coordinate system are

$$\Gamma^{\hat{a}}_{\hat{b}\hat{c}} = (1/r^2)x^{\hat{a}}g_{\hat{b}\hat{c}}. \tag{4.22}$$

Since r^2 is constant on S^n,

$$\partial_{\hat{d}}\Gamma^{\hat{a}}_{\hat{b}\hat{c}} = (1/r^2)[g^{\hat{a}}_{\hat{d}}g_{\hat{b}\hat{c}} + x^{\hat{a}}\partial_{\hat{d}}g_{\hat{b}\hat{c}}]. \tag{4.23}$$

Let us evaluate all our quantites at the "north pole" of the coordinate system; denoting this point by "0,"

$$x^{\hat{a}}|_0 = 0. \tag{4.24}$$

Equations (4.22) and (4.23) then give

$$\Gamma^{\hat{a}}_{\hat{b}\hat{c}}|_0 = 0, \tag{4.25}$$

and

$$\partial_{\hat{d}}\Gamma^{\hat{a}}_{\hat{b}\hat{c}}|_0 = (1/r^2)g^{\hat{a}}_{\hat{d}}\,g_{\hat{b}\hat{c}}, \tag{4.26}$$

since the metric and its derivatives are nonsingular. Using (4.9), (4.17), (4.18), (4.21), and (4.24)–(4.26), the $x^{\hat{a}}$ independence of the quantities $\Omega_I^{\hat{a}}{}_{\hat{b}}$, and the relation (valid at all points in all coordinate systems) between the Riemann tensor and the metric on S^n,

$$R_{abcd} = (1/r^2)[g_{bc}\,g_{ad} - g_{ac}\,g_{bd}], \tag{4.27}$$

we obtain

$$g^{ij}\xi_i^{\hat{a}}\partial_{\hat{b}}\xi_j^{\hat{c}}|_0 = 0, \tag{4.28a}$$

$$g^{ij}\xi_i^{\hat{a}}\partial_{\hat{a}}\partial_{\hat{b}}\xi_j^{\hat{c}}|_0 = -(1/r^2)g^{\hat{c}}_{\hat{b}}, \tag{4.28b}$$

$$g^{ij}\partial_{\hat{a}}\xi_i^{\hat{b}}\partial_{\hat{b}}\xi_j^{\hat{c}}|_0 = -[(n-1)/r^2]g^{\hat{c}}_{\hat{a}}, \tag{4.28c}$$

$$g^{ij}\partial_{\hat{a}}\xi_i^{(\hat{b}}\partial_{\hat{c}}\xi_j^{\hat{d})}|_0 = \frac{1}{r^2}g^{\hat{e}(\hat{b}}g^{\hat{d})\hat{f}}[g_{\hat{a}\hat{c}}\,g_{\hat{e}\hat{f}} - g_{\hat{f}\hat{c}}\,g_{\hat{e}\hat{a}}]. \tag{4.28d}$$

From (4.15) and (4.28), we conclude that

$$g^{ij}\mathcal{L}_{\xi_i}\mathcal{L}_{\xi_j}T^{(l)}_{\hat{c}_1\cdots\hat{c}_m}|_0 = [\partial^{\hat{a}}\partial_{\hat{a}} - m(m+n-1)/r^2]T^{(l)}_{\hat{c}_1\cdots\hat{c}_m}|_0. \tag{4.29}$$

Employing the usual rules of covariant differentiation and the values of the Christoffel symbols in the EP^2 coordinates we find that

$$\tilde{\nabla}^{\hat{a}}\tilde{\nabla}_{\hat{a}} T^{(l)}_{\hat{c}_1\cdots\hat{c}_m|_0} = [\partial^{\hat{a}}\partial_{\hat{a}} - m/r^2]\, T^{(l)}_{\hat{c}_1\cdots\hat{c}_m}|_0, \tag{4.30}$$

so

$$g^{ij}\mathscr{L}_{\xi_i}\mathscr{L}_{\xi_j} T^{(l)}_{\hat{c}_1\cdots\hat{c}_m}|_0 = [\tilde{\nabla}^{\hat{a}}\tilde{\nabla}_{\hat{a}} - m(m+n-2)/r^2]\, T^{(l)}_{\hat{c}_1\cdots\hat{c}_m}|_0. \tag{4.31}$$

But this is a tensor equation with respect to coordinate transformations on S^n, and therefore is valid in all coordinate systems on S^n. Furthermore, EP^2 coordinates could have been chosen with the "north pole" at any point; therefore, at every point of S^n,

$$g^{ij}\mathscr{L}_{\xi_i}\mathscr{L}_{\xi_j} T^{(l)}_{c_1\cdots c_m} = [\tilde{\nabla}^{a}\tilde{\nabla}_{a} - m(m+n-2)/r^2]\, T^{(l)}_{c_1\cdots c_m}. \tag{4.32}$$

Using (3.9), (4.13), and (4.32), we find that $T^{(l)}_{c_1\cdots c_m}$ is an eigenfunction of $C_2(x)$:

$$C_2(x) T^{(l)}_{c_1\cdots c_m} = 2[l(l+n-1) + m(m+n-3)]\, T^{(l)}_{c_1\cdots c_m}. \tag{4.33}$$

The C_2 eigenvalue corresponding to $T^{(l)}_{c_1\cdots c_m}$ is the same as that for the $(l,m,0,\ldots)$ irrep of $O(n+1)$. [See Eq. (4.11).] This leaves two possibilities for the transformation properties of $T^{(l)}_{c_1\cdots c_m}$.

(A) If, for a given set of values of l, m, and n ($l \geqslant m$, of course), the C_2 eigenvalue of the $(l,m,0\ldots)$ irrep of $O(n+1)$ does not equal the C_2 eigenvalue of any of the other irreps on the rhs of (4.2), then the tensors $T^{(l)}_{c_1\cdots c_m}$ transform irreducibly under $(l,m,0,\ldots)$, and the number of linearly independent such $T^{(l)}_{c_1\cdots c_m}$'s is equal to the dimensionality of this irrep.

(B) Some subset of the other irreps in (4.2) have the same C_2 eigenvalue as $(l,m,0,\ldots)$. Then, without investigating the action of the higher-rank Casimir operators on $T^{(l)}_{c_1\cdots c_m}$, we can only conclude that $T^{(l)}_{c_1\cdots c_m}$ can be expressed as a sum of tensors each of which transforms irreducibly according to one of the irreps which have this C_2 eigenvalue in common.

For the cases of greatest interest, $m = 1$ and $m = 2$, possibility (B) can be ruled out by explicit computation using the decomposition formula (4.2) and Eq. (4.11) for the C_2 eigenvalues. For $m = 1$ and $m = 2$ it turns out that, for all $n \geqslant 3$ and all $l \geqslant m$, the C_2 eigenvalue of $(l,m,0,\ldots)$ is distinct from the C_2 eigenvalues of the other irreps in (4.11).

Making use of Ref. 17, we find for all *even* $n \geqslant 4$,

$$D_l(n,1) = \frac{l(l+n-1)(2l+n-1)}{(n-2)(n-3)} \times \prod_{j=3}^{n/2}\left[\frac{(j+l-1)(n+l-j)}{(n-1-j)(j-2)}\right], \tag{4.34a}$$

$$D_l(n,2) = \frac{(n+1)(l-1)(l+n)(2l+n-1)}{(n-1)(n-2)(n-3)} \times \prod_{j=3}^{n/2}\left[\frac{(l+n-j)(j+l-1)j(n+1-j)}{(n-j)(n-1-j)(j-1)(j-2)}\right], \tag{4.34b}$$

where the products are set equal to 1 when $n = 4$. For all *odd* $n \geqslant 5$,

$$D_l(n,1) = \frac{l(l+n-1)}{(n-2)} \prod_{j=3}^{(n+1)/2}\left[\frac{(j+l-1)(n+l-j)}{(n-1-j)(j-2)}\right], \tag{4.35a}$$

$$D_l(n,2) = \frac{(l-1)(l+n)}{(n-2)} \times \prod_{j=3}^{(n+1)/2}\left(\frac{(l+n-j)(l-1+j)j(n+1-j)}{(n-j)(n-1-j)(j-1)(j-2)}\right). \tag{4.35b}$$

For $n = 3$,

$$D_l(n,1) = 2l(l+2), \tag{4.36a}$$

$$D_l(n,2) = 2(l-1)(l+3). \tag{4.36b}$$

Since $\Pi_{i=a}^{b} i = b!/(a-1)!$, (4.34)–(4.36) may all be written as

$$D_l(n,1) = \frac{l(l+n-1)(2l+n-1)(l+n-3)!}{(n-2)!(l+1)!}, \quad l \geqslant 1 \tag{4.37a}$$

$$D_l(n,2) = \frac{(n+1)(n-2)(l+n)(l-1)(2l+n-1)(l+n-3)!}{2(n-1)!(l+1)!}, \quad l \geqslant 2. \tag{4.37b}$$

ACKNOWLEDGMENTS

We wish to thank Arno Bohm, Steven Christensen, Marek Demianski, Piotr Kielanowski, Richard Matzner, Pat Moylan, Malcolm Perry, Larry Shepley, Marco Tarlini, Xerxes Tata, Peter Tomas, Robert Wald, and Steven Weinberg for helpful discussions. Subsequent to the completion of this work, we learned that Alan Chodos and Eric Myers[10] have concurrently computed the eigenvalues and degeneracies for transverse vetor and ST^2 tensor harmonics on S^n, with the same motivation and conclusions as our own. We are grateful to them for comparing results prior to publication. We are particularly indebted to John Gilbert for extremely valuable group-theoretic guidance during the course of this work.

This work was supported in part by the Robert A. Welch Foundation, The University of Texas Center for Theoretical Physics, and the National Science Foundation under Grants PHY-8304629 and PHY-8205717.

[1]See, e.g., E. S. Abers and B. W. Lee, Phys. Rep. C **9**, 1 (1973); P. Ramond, *Field Theory: A Modern Primer* (Benjamin/Cummings, Reading, MA, 1981).
[2]P. Candelas and D. J. Raine, Phys. Rev. D **15**, 1494 (1977); S. W. Hawking, Commun. Math. Phys. **55**, 133 (1977).
[3]G. W. Gibbons and M. J. Perry, Nucl. Phys. B **146**, 90 (1978).
[4]S. M. Christensen and M. J. Duff, Nucl. Phys. B **170** [FS1], 480 (1980).
[5]Bateman Manuscript Project, *Higher Transcendental Functions* (McGraw-Hill, New York, 1955), Vol. II.
[6]I. T. Drummond and G. M.Shore, Ann. Phys. (NY) **117**, 89 (1979); G. M. Shore, Ann. Phys. (NY) **117**, 121 (1979).
[7]E. g., S. L. Adler, Phys. Rev. D **6**, 3445 (1972) and **8**, 2400 (1973); B. L. Hu, J.Math. Phys. **15**, 1748 (1874); R. T. Jantzen, J. Math. Phys. **19**, 1163 (1978).

[8]E. M. Lifshitz, J. Phys. (Moscow) **10**, 116 (1946); E. M. Lifshitz and I. M. Khalatnikov, Adv. Phys. **12**, 185 (1963).
[9]C. R. Ordóñez and M. A. Rubin, University of Texas Theory Group preprint (in preparation).
[10]A. Chodos and E. Myers, Yale University preprint YTP-83-18.
[11]J. W. York, Jr., J. Math. Phys. **14**, 456 (1973); W. V.D. Hodge, *Theory and Application of Harmonic Integrals* (Cambridge U.P., Cambridge, 1952).
[12]S. M. Christensen, M. J. Duff, G. W. Gibbons, and M. Roček, "One-Loop Effects in Supergravity with a Cosmological Constant" (unpublished).
[13]C. R. Ordóñez and M. A. Rubin, University of Texas Theory Group preprint UTTG-5-84 (to appear in J. Math. Phys).
[14]A. P. Lightman, W. H. Press, R. H. Price, and S. A.Teukolsky, *Problem Book in Relativity and Gravitation* (Princeton U.P., Princeton, 1975), Prob. 3.16. Remember to subtract the tracelessness conditions to obtain (2.1b).
[15]Except as noted, our differential-geometric conventions are those of Ref. 16.
[16]S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, New York, 1972).
[17]W. Miller, Jr., *Symmetry Groups and Their Applications* (Academic, New York, 1972), Theorem 9.8, p. 361, and Theorem 9.9, p. 362. These formulas are for SO($n+1$). They may be applied directly to O($n+1$) except for $n=3$, in which case the dimensionality of the irrep of O(4) labeled by the dominant weight vector $(l,m,0,...)$ is equal to the sum of the dimensionalities of the $(l,m,0,...)$ and $(l,-m,0,...)$ irreps of SO(4). See Ref. 19, comments after Eq. (2.5), p. 499.
[18]M. Gourdin, *Basics of Lie Groups* (Editions Frontieres, Gif sur Yvette, 1982).
[19]D. A. Levine, Trans. Am. Math. Soc. **144**, 493 (1969). Note comments after Eq. (2.5), p. 499.
[20]S. G.Mihlin, *Higher-dimensional Integral Equations* (Fitzmatgiz, Moscow, 1962) [English translation (Pergamon, New York, 1965)].
[21]M. Hamermesh, *Group Theory* (Addison–Wesley, Reading, MA 1962).
[22]B. F. Schutz, *Geometrical Methods of Mathematical Physics* (Cambridge U.P., Cambridge, 1980).
[23]Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds, and Physics* (North-Holland, Amsterdam, 1982).
[24]R. Geroch, "Differential Geometry" (unpublished).
[25]A. Zee, "Grand Unification and Gravity—Selected Topics," in *Grand Unified Theories and Related Topics: Proceedings of the 4th Kyoto Summer Institute* (World Scientific, Singapore, 1981).
[26]P. Candelas and S. Weinberg, Nucl. Phys. B **237**, 397 (1984), Appendix A.
[27]A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (PWN-Polish Scientific Publishers, Warsaw, 1977).
[28]We employ the following conventions:

$$t_{(c_1\cdots c_m)} \equiv (1/m!)\times(\text{sum over all index permutations of } t_{c_1\cdots c_m}),$$

$$t_{[c_1\cdots c_m]} \equiv (1/m!)\times\big[(\text{sum over all even index permutations of } t_{c_1\cdots c_m} - (\text{sum over all odd index permutations of } t_{c_1\cdots c_m})\big].$$

Differential equations for the cuspoid canonical integrals

J. N. L. Connor
A. A. Noyes Laboratory of Chemical Physics, California Institute of Technology, Pasadena, California 91125 and Department of Chemistry, University of Manchester, Manchester M13 9PL, England [a]

P. R. Curtis
Department of Chemistry, University of Manchester, Manchester M13 9PL, England [b]

(Received 23 November 1983; accepted for publication 2 March 1984)

Differential equations satisfied by the cuspoid canonical integrals $I_n(\mathbf{a})$ are obtained for arbitrary values of $n \geqslant 2$, where $n-1$ is the codimension of the singularity and $\mathbf{a} = (a_1, a_2, \ldots, a_{n-1})$. A set of linear coupled ordinary differential equations is derived for each step in the sequence $I_n(0,0,\ldots,0,0) \rightarrow I_n(0,0,\ldots,0,a_{n-1}) \rightarrow I_n(0,0,\ldots,a_{n-2},a_{n-1}) \rightarrow \ldots \rightarrow I_n(0,a_2,\ldots,a_{n-2},a_{n-1}) \rightarrow I_n(a_1,a_2,\ldots,a_{n-2},a_{n-1})$. The initial conditions for a given step are obtained from the solutions of the previous step. As examples of the formalism, the differential equations for $n = 2$ (fold), $n = 3$ (cusp), $n = 4$ (swallowtail), and $n = 5$ (butterfly) are given explicitly. In addition, iterative and algebraic methods are described for determining the parameters $\mathbf{a}$ that are required in the uniform asymptotic cuspoid approximation for oscillating integrals with many coalescing saddle points. The results in this paper unify and generalize previous researches on the properties of the cuspoid canonical integrals and their partial derivatives.

PACS numbers: 02.30.Hq, 02.30.Bi, 03.80. + r

I. INTRODUCTION

There has been considerable interest recently in the properties of the cuspoid canonical integrals $I_n(\mathbf{a})$.[1-9] The integral $I_n(\mathbf{a})$ is defined by[2,10]

$$I_n(a_1,a_2,\ldots,a_{n-1}) = \int_{-\infty}^{\infty} \exp[i(a_1 u + a_2 u^2 + \cdots + a_{n-1}u^{n-1} + u^{n+1})]\,du\,, \quad (1.1)$$

where the coefficients a_j are real and n is a positive integer with $n \geqslant 2$. The name cuspoid canonical integral comes from catastrophe theory, where the polynomial in the exponent of Eq. (1.1) is the canonical form for the cuspoid catastrophes, with $n-1$ the codimension of the singularity.[10-13]

The integral $I_n(\mathbf{a})$, together with its partial derivatives $\partial I_n/\partial a_1$, $\partial I_n/\partial a_2,\ldots,\partial I_n/\partial a_{n-1}$ occur in the asymptotic analysis of many short wavelength scattering theories involving atoms, molecules, and nuclear heavy ions as well as in acoustic, electromagnetic, and water-wave propagation, and in other problems. References to a large number of relevant papers can be found in Refs. 4, 6, and 9.

We have recently investigated[4-6,8,9] the differential equations satisfied by $I_n(\mathbf{a})$ and its partial derivatives for the special cases of $n = 2$ (fold), $n = 3$ (cusp), and $n = 4$ (swallowtail). For $n = 2$, Eq. (1.1) is proportional to Airy's integral[14] $\mathrm{Ai}(x)$ and the differential equation for $\mathrm{Ai}(x)$ is well known.[15] When $n = 3$, Eq. (1.1) is identical with Pearcey's integral $P(x,y)$.[16-18] The differential equations for $P(x,y)$ were first derived by Pearcey[17,18] and used by him to numerically evaluate $P(x,y)$. For the swallowtail canonical integral ($n = 4$), which is also written $S(x,y,z)$, the relevant differential equations have been obtained by ourselves and Farrelly[8] and used in numerical computations.[8,9]

The purpose of the present paper is to derive the differential equations satisfied by the cuspoid canonical integrals for any value of n. Our treatment generalizes and unifies the previous work[4-6,8,9,15,17,18] in which each value of $n = 2,3,4$ was treated as a separate case.

We derive the differential equations in Sec. II. The basic idea is to obtain a set of linear coupled ordinary differential equations for each step in the sequence $I_n(0,0,\ldots,0,0) \rightarrow I_n(0,0,\ldots,0,a_{n-1}) \rightarrow I_n(0,0,\ldots,a_{n-2},a_{n-1}) \rightarrow, \ldots, \rightarrow I_n(0,a_2,\ldots,a_{n-2},a_{n-1}) \rightarrow I_n(a_1,a_2,\ldots,a_{n-2},a_{n-1})$. The solutions of the differential equations for a given step provide the initial conditions for the next step, with the initial conditions at the origin $(0,0,\ldots,0,0)$ being obtained directly from Eq. (1.1). An advantage of this scheme is that it allows the numerical evaluation of the derivatives $\partial I_n/\partial a_1$, $\partial I_n/\partial a_2,\ldots,\partial I_n/\partial a_{n-1}$ as well as I_n at a given value of $\mathbf{a}$.

Section III provides concrete examples of our formalism. We show for the special cases of $n = 2,3,4$ that our equations are equivalent to those obtained earlier. We also give the differential equations for the butterfly canonical integral ($n = 5$), as they have not been reported in the literature before.

An important use of I_n and $\partial I_n/\partial a_1$, $\partial I_n/\partial a_2,\ldots,\partial I_n/\partial a_{n-1}$ is in uniform asymptotic approximations for one-dimensional oscillating integrals with many nearly coincident stationary phase or saddle points.[1,2,10,19-23] In Section IV we show how the iterative and algebraic methods of Ref. 9 for calculating the parameters $\mathbf{a}$ that are required in I_n and its partial derivatives can be generalized from the case of $n = 4$ to the case of arbitrary n. Thus the techniques reported in this paper allow in principle the application of uniform asymptotic approximations for integrals with an arbitrary number of nearly coincident saddle points. Our conclusions are in Sec. V.

II. DIFFERENTIAL EQUATIONS

The integral representation (1.1) for $I_n(\mathbf{a})$ is conditionally convergent. This has the disadvantage that repeated differentiation under the integral sign eventually produces a

[a] Permanent address.

[b] Present address: Marconi Space and Defense Systems Ltd., The Grove, Warren Lane, Stanmore, Middlesex HA7 4LY, England.

 0022-2488/84/102895-08$02.50

divergent integral.[8,24] To avoid this problem, we deform the contour of integration C_n so that it starts and ends in sectors of the complex u plane where the integrand is exponentially small. We can write

$$I_n(\mathbf{a}) = \int_{C_n} \exp[iP_n(u)]\,du\,, \tag{2.1}$$

with

$$P_n(u) = u^{n+1} + \sum_{j=1}^{n-1} a_j u^j\,. \tag{2.2}$$

When n is an even integer, we choose C_n to start on the ray $\arg u = \pi - \frac{1}{2}\pi/(n+1)$ and finish on the ray $\arg u = \frac{1}{2}\pi/(n+1)$. For n odd, the contour $\arg u = \pi + \frac{1}{2}\pi/(n+1)$ to $\arg u = \frac{1}{2}\pi/(n+1)$ is an appropriate choice for C_n.

We now wish to derive a set of differential equations for the integration $I_n(0,0,...,0,0) \to I_n(a_1,a_2,...,a_{n-2},a_{n-1})$. To accomplish this we use a sequence of steps $I_n(0,0,...,0,0) \to I_n(0,0,...,0,a_{n-1}) \to I_n(0,0,...,a_{n-2},a_{n-1}) \to,...,\to I_n(0,a_2,...,a_{n-2},a_{n-1}) \to I_n(a_1,a_2,...,a_{n-2},a_{n-1})$, such that in each step k all the a_j are constant except for a_k.

We derive the differential equation for the variable a_1 first. From Eq. (2.2) we have

$$P_n'(u) = (n+1)u^n + \sum_{j=1}^{n-1} j a_j u^{j-1}\,. \tag{2.3}$$

Now partial differentiation of Eq. (2.1) with respect to a_1 is equivalent to multiplication of the integrand by iu. It then follows from Bleistein[20] that $I_n(\mathbf{a})$ satisfies the differential equation

$$P_n'\left(-i\frac{\partial}{\partial a_1}\right) I_n(a_1,a_2,...,a_{n-1}) = 0\,, \tag{2.4}$$

because

$$\int_{C_n} P_n'(u)\exp[iP_n(u)]\,du = 0\,.$$

The operator in Eq. (2.4) is given explicitly by

$$P_n'\left(-i\frac{\partial}{\partial a_1}\right) = (-i)^n(n+1)\frac{\partial^n}{\partial a_1^n} + \sum_{j=1}^{n-1}(-i)^{j-1} j a_j \frac{\partial^{j-1}}{\partial a_1^{j-1}}\,. \tag{2.5}$$

We then obtain from Eq. (2.4) the differential equation for $I_n(\mathbf{a})$ in the variable a_1:

$$\frac{\partial^n I_n}{\partial a_1^n} + \sum_{j=1}^{n-1}(-i)^{j-n-1}\frac{j}{n+1} a_j \frac{\partial^{j-1} I_n}{\partial a_1^{j-1}} = 0\,. \tag{2.6}$$

To proceed further we introduce the functions $I_n^{(p)}(\mathbf{a})$ defined by

$$I_n^{(p)}(\mathbf{a}) = \frac{\partial^p I_n(\mathbf{a})}{\partial a_1^p}\,, \quad p = 0,1,... \tag{2.7}$$

with the convention that

$$I_n^{(0)}(\mathbf{a}) \equiv I_n(\mathbf{a})$$

and

$$I_n^{(p)}(\mathbf{a}) \equiv 0 \quad \text{when } p < 0\,.$$

For $n = 4$ (swallowtail canonical integral), $I_4^{(1)}$ and $I_4^{(2)}$ are the same as the functions H and G, respectively, that were introduced in Ref. 8. Our strategy now is to derive sets of coupled linear ordinary differential equations for each of the variables $a_2,a_3,...,a_{n-1}$ in which the $I_n^{(p)}$ are treated as independent functions.

Differentiating Eq. (2.6) p times with respect to a_1 with $p = 0,1,...,n-2$ gives

$$\frac{\partial^n I_n^{(p)}}{\partial a_1^n} + \sum_{j=1}^{n-1}(-i)^{j-n-1}\frac{j}{n+1} a_j \frac{\partial^{j-1} I_n^{(p)}}{\partial a_1^{j-1}} + (-i)^{-n}\frac{p}{n+1} I_n^{(p-1)} = 0\,, \quad p = 0,1,...,n-2\,. \tag{2.8}$$

Next we use differential identities satisfied by $I_n^{(p)}$ to change the integration variable from a_1 to the variable a_j, $j = 2,3,...,n-1$.

From the integral representation (2.1) for $I_n(\mathbf{a})$ we deduce that

$$I_n^{(p)}(\mathbf{a}) = i^p \int_{C_n} u^p \exp[iP_n(u)]\,du\,, \quad p = 0,1,... \tag{2.9}$$

and differentiation of this equation q times ($q = 0,1,2,...$) with respect to a_j gives

$$\frac{\partial^q I_n^{(p)}}{\partial a_j^q} = i^{p+q}\int_{C_n} u^{p+jq}\exp[iP_n(u)]\,du\,, \quad p,q = 0,1,...,\ j = 1,2,...,n-1\,. \tag{2.10}$$

If we change labels $p \to r$, $q \to s$, $j \to l$ in Eq. (2.10) we have

$$\frac{\partial^s I_n^{(r)}}{\partial a_l^s} = i^{r+s}\int_{C_n} u^{r+ls}\exp[iP_n(u)]\,du\,, \quad r,s = 0,1,...,\ l = 1,2,...,n-1\,. \tag{2.11}$$

It is now clear that if r is chosen to be

$$r = p + jq - ls\,,$$

then we obtain from Eqs. (2.10) and (2.11) the identity

$$\frac{\partial^q I_n^{(p)}}{\partial a_j^q} = i^{q(1-j)-s(1-l)}\frac{\partial^s I_n^{(p+jq-ls)}}{\partial a_l^s}\,, \quad \begin{matrix} p,q,s = 0,1,..., \\ j,l = 1,2,...,n-1. \end{matrix} \tag{2.12}$$

This relation lets us change from the variable a_j on the left-hand side to the variable a_l on the right-hand side. For example for the case $j = 1$ required below we have

$$\frac{\partial^q I_n^{(p)}}{\partial a_1^q} = i^{s(l-1)}\frac{\partial^s I_n^{(p+q-ls)}}{\partial a_l^s}\,, \quad \begin{matrix} p,q,s = 0,1,..., \\ l = 1,2,...,n-1\,. \end{matrix} \tag{2.13}$$

The relation (2.12) encompasses a large number of individual identities. For the swallowtail case $n = 4$ some of them are reported in Ref. 8.

We next return to the differential equation (2.8) for $I_n^{(p)}$ and set

$$a_j = 0 \quad \text{for } j < l \quad \text{with } l = 1,2,...,n-1\,. \tag{2.14}$$

In addition for each l we only keep terms which have $p = 0,1,...,l-1$. We obtain

$$\frac{\partial^n I_n^{(p)}}{\partial a_1^n} + \sum_{j=l}^{n-1}(-i)^{j-n-1}\frac{j}{n+1} a_j \frac{\partial^{j-1} I_n^{(p)}}{\partial a_1^{j-1}} + (-i)^{-n}\frac{p}{n+1} I_n^{(p-1)} = 0\,,$$
$$a_j = 0 \quad \text{for } j < l\,,\ p = 0,1,...,l-1\,,\ l = 1,2,...,n-1\,. \tag{2.15}$$

We next wish to convert Eq. (2.15) into a set of differential equations with respect to a_l using the identity (2.13). However, inspection of Eq. (2.13) shows that for fixed p,q,l there is still a choice to be made for s. We choose s so that the superscript $k = p + q - ls$ on the right-hand side of Eq. (2.13) is zero or is the smallest possible integer for a given p,q,l. We choose s in this way because it introduces the minimum number of independent functions $I_n^{(k)}$ into Eq. (2.15) when the identity (2.13) is used. Our choice for s is automatically satisfied if for a given p,q,l we calculate it from the equation

$$s_q = \text{int}[(p+q)/l\,], \tag{2.16}$$

where int = integer part of. Note that s_q is also a function of p and l, but this has not been indicated explicitly in Eq. (2.16) in order to simplify the notation.

With this choice for s, Eq. (2.13) becomes for $q = n$

$$\frac{\partial^n I_n^{(p)}}{\partial a_1^n} = i^{s_n(l-1)} \frac{\partial^{s_n} I_n^{(p+n-ls_n)}}{\partial a_l^{s_n}}, \quad \begin{matrix} p = 0,1,\ldots, \\ l = 1,2,\ldots,n-1, \end{matrix} \tag{2.17}$$

where

$$s_n = \text{int}\left(\frac{p+n}{l}\right). \tag{2.18}$$

We now use the identity (2.17) to change variable from a_1 to a_l in Eq. (2.15) obtaining

$$\frac{\partial^{s_n} I_n^{(n+p-ls_n)}}{\partial a_l^{s_n}} + \sum_{j=l}^{n-1} i^{(s_{j-1}-s_n)(l-1)+n+1-j} \frac{j}{n+1} \times a_j \frac{\partial^{s_{j-1}} I_n^{(j-1+p-ls_{j-1})}}{\partial a_l^{s_{j-1}}} + i^{n-s_n(l-1)} \frac{p}{n+1} I_n^{(p-1)} = 0 \tag{2.19}$$

$a_j = 0$ for $j<l$, $p = 0,1,\ldots,l-1$, $l = 1,2,\ldots,n-1$,

and

$$s_{j-1} = \text{int}[(p+j-1)/l\,]. \tag{2.20}$$

Equation (2.19) is the main result of this paper. It represents $n-1$ sets of coupled ordinary linear differential equations. Each set involves only a single variable a_l with the condition $a_j = 0$ for $j<l$. For each value of l, there are l equations ($p = 0,1,\ldots,l-1$) which contain l independent functions $I_n^{(p)}$.

Equation (2.19) represents sufficient equations to evaluate I_n and $\partial I_n/\partial a_1, \partial I_n/\partial a_2, \ldots, \partial I_n/\partial a_{n-1}$ at any point $(a_1,a_2,\ldots,a_{n-1})$. The procedure to accomplish this is as follows

(a) Set $l = n-1$ in Eq. (2.19). This gives a set of $n-1$ differential equations in the variable a_{n-1} with $a_1 = a_2 = \cdots = a_{n-2} = 0$.

(b) Evaluate $I_n^{(p)}$ at the origin (0,0,...,0) for all required values of $p = 0,1,\ldots,n-1$ (see below). This supplies all the necessary initial conditions at the origin for Eq. (2.19) with $l = n-1$.

(c) Integrate numerically from (0,0,...,0,0) to $(0,0,\ldots,0,a_{n-1})$ using a standard integration routine for coupled ordinary linear differential equations.

(d) Set $l = n-2$ in Eq. (2.19). This gives a set of $n-2$ differential equations in the variable a_{n-2} with $a_j = 0$ for $j<n-2$.

(e) Use the identity (2.12) to transform the solutions at $(0,0,\ldots,0,a_{n-1})$ obtained in step (c) into initial conditions for Eq. (2.19) with $l = n-2$ [step (d)].

(f) Integrate from $(0,0,\ldots,0,0,a_{n-1})$ to $(0,0,\ldots,0,a_{n-2},a_{n-1})$.

(g) Repeat steps (d), (e), (f) for $l = n-3, n-4,\ldots,3,2$. We have now reached the point $(0,a_2,a_3,\ldots,a_{n-2},a_{n-1})$.

(h) Finally set $l = 1$ and integrate Eq. (2.19) [which is now equivalent to Eq. (2.6)] from $(0,a_2,\ldots,a_{n-1})$ to $(a_1,a_2,\ldots,a_{n-1})$ using as initial conditions the solutions from the previous integration step together with Eq. (2.12). At $(a_1,a_2,\ldots,a_{n-1})$ we have now evaluated I_n. Its derivatives $\partial I_n/\partial a_1, \partial I_n/\partial a_2, \ldots, \partial I_n/\partial a_{n-1}$ are obtained from the solutions of Eq. (2.19) together with the identity (2.17).

In step (b), it is necessary to calculate the initial conditions at the origin (0,0,...,0) for Eq. (2.19) with $l = n-1$. The allowed values of p in this case are $p = 0,1,\ldots,n-2$. If we evaluate the superscripts s_n, $n+p-(n-1)s_n$, s_{n-2}, and $n-2+p-(n-1)s_{n-2}$ in Eq. (2.19) then it is found that the set of differential equations for $l = n-1$ consists of terms of the form

$$\frac{\partial I_n^{(p)}}{\partial a_{n-1}}, \quad p = 0,1,\ldots,n-2 \quad \text{and} \quad \frac{\partial^2 I_n^{(0)}}{\partial a_{n-1}^2}, \tag{2.21}$$

where the term of second order comes from $p = n-2$. The initial conditions at the origin that we must calculate are therefore, of the form

$$I_n^{(p)}, \quad p = 0,1,\ldots,n-2 \quad \text{and} \quad \frac{\partial I_n^{(0)}}{\partial a_{n-1}}. \tag{2.22}$$

However Eqs. (2.1) and (2.9) show that

$$\frac{\partial I_n^{(0)}}{\partial a_{n-1}} = -i^{-n} I_n^{(n-1)}. \tag{2.23}$$

Hence, we can fix the initial conditions at the origin for Eq. (2.19) with $l = n-1$ provided we can calculate

$$I_n^{(p)}(0,0,\ldots,0), \quad p = 0,1,\ldots,n-1. \tag{2.24}$$

Now Eq. (2.9) shows that

$$I_n^{(p)}(0,0,\ldots,0) = i^p \int_{-\infty}^{\infty} u^p \exp(iu^{n+1}) du, \quad p = 0,1,\ldots,n-1, \tag{2.25}$$

which can be written in the alternative form

$$I_n^{(p)}(0,0,\ldots,0) = i^p \int_0^{\infty} u^p \exp(iu^{n+1}) du + (-1)^p i^p \int_0^{\infty} u^p \exp(i(-1)^{n+1} u^{n+1}) du, \quad p = 0,1,\ldots,n-1. \tag{2.26}$$

The integrals in Eq. (2.26) can now be evaluated with the help of the results[25,26]

$$\int_0^{\infty} u^p \exp(iu^{n+1}) du = \frac{1}{n+1} \Gamma\left(\frac{p+1}{n+1}\right) \exp\left[i \frac{\pi}{2} \left[\frac{p+1}{n+1}\right]\right], \quad p = 0,1,\ldots,n-1, \tag{2.27}$$

and

$$\int_0^\infty u^p \exp(-iu^{n+1})du$$

$$= \frac{1}{n+1}\Gamma\left[\frac{p+1}{n+1}\right]\exp\left[-i\frac{\pi}{2}\left[\frac{p+1}{n+1}\right]\right]. \quad (2.28)$$

$p = 0,1,...,n-1$.

Consider first the case $n = 2m$ = even positive integer. Applying Eqs. (2.27) and (2.28) to Eq. (2.26) shows that

$$I^{(p)}_{2m}(0,0,...,0)$$

$$= \frac{2}{2m+1}\Gamma\left[\frac{p+1}{2m+1}\right]\cos\left[\frac{\pi}{2}\left[\frac{p+1}{2m+1}+p\right]\right],$$

$m = 1,2,...,\ p = 0,1,...,2m-1$. (2.29)

Next consider the case $n = 2m+1$ = odd positive integer. If $p = 2q+1$ = odd positive integer, then the integrand of Eq. (2.25) is an odd function of u and

$$I^{(2q+1)}_{2m+1}(0,0,...,0) = 0, \quad m = 1,2,...,\ q = 0,1,...,m-1. \quad (2.30)$$

If however, $p = 2q$ = even integer, then Eqs. (2.26) and (2.27) give

$$I^{(2q)}_{2m+1}(0,0,...,0) = \frac{1}{m+1}\Gamma\left[\frac{2q+1}{2m+2}\right] \times\exp\left[i\frac{\pi}{2}\left[\frac{2q+1}{2m+2}+2q\right]\right],$$

$m = 1,2,...,\ q = 0,1,..,m$. (2.31)

An alternative way of deriving Eqs. (2.29)–(2.31) is from the exact series representation of $I_n(a_1,a_2,...,a_{n-1})$ obtained in Ref. 2.

III. EXAMPLES

In this section we consider in more detail the cases $n = 2$ (fold), $n = 3$ (cusp), $n = 4$ (swallowtail), and $n = 5$ (butterfly). We verify that Eq. (2.19) is equivalent to known results for $n = 2,3,4$. The differential equations we derive for the butterfly canonical integral ($n = 5$) are new and have not been reported in the literature before. In order to write down explicitly the differential equations contained in Eq. (2.19) for a given value of n, we have written a simple symbolic algebraic computer program based on the FORMAT statements of FORTRAN.

A. The fold canonical integral $I_2(a_1)$

We have $n = 2$ so that $l = 1$, $p = 0$ and Eq. (2.19) becomes

$$\frac{d^2 I^{(0)}_2}{da_1^2} - \frac{1}{3}a_1 I^{(0)}_2 = 0. \quad (3.1)$$

The fold integral $I_2(a_1)$ is related to the Airy integral $\mathrm{Ai}(a_1)$ by

$$I_2(a_1) = (2\pi/3^{1/3})\mathrm{Ai}(a_1/3^{1/3}),$$

and if we set $a_1 = x$, Eq. (3.1) becomes

$$\frac{d^2\,\mathrm{Ai}(x)}{dx^2} - x\,\mathrm{Ai}(x) = 0,$$

which is a well-known result.[15]

B. The cusp canonical integral $I_3(a_1,a_2)$

In this case $n = 3$ and either $l = 2$, $p = 0,1$ with $a_1 = 0$ or $l = 1$, $p = 0$. The differential equations obtained from Eq. (2.19) are for $l = 2$, $p = 0$

$$\frac{\partial I^{(1)}_3}{\partial a_2} + i\frac{1}{2}a_2 I^{(1)}_3 = 0, \quad a_1 = 0 \quad (3.2)$$

for $l = 2$, $p = 1$,

$$\frac{\partial^2 I^{(0)}_3}{\partial a_2^2} + i\frac{1}{2}a_2\frac{\partial I^{(0)}_3}{\partial a_2} + i\frac{1}{4}I^{(0)}_3 = 0, \quad a_1 = 0 \quad (3.3)$$

and for $l = 1$, $p = 0$

$$\frac{\partial^3 I^{(0)}_3}{\partial a_1^3} - i\frac{1}{4}a_1 I^{(0)}_3 - \frac{1}{2}a_2\frac{\partial I^{(0)}_3}{\partial a_1} = 0. \quad (3.4)$$

The cusp canonical integral $I_3(a_1,a_2)$ is the same as Pearcey's integral $P(x,y)$ provided we set $x = a_2$, $y = a_1$ and Eqs. (3.2)–(3.4) are identical with Eqs. (2.12), (2.13), and (2.3), respectively, of Ref. 8.

Note that Eqs. (3.2) and (3.3) for $I^{(1)}_3$ and $I^{(0)}_3$ are not coupled together. Because of this we can solve Eq. (3.2) for the boundary condition $\partial P(0,y)/\partial y = 0$ at $y = 0$ [see Eq. (2.30)] obtaining

$$\left.\frac{\partial P(x,y)}{\partial y}\right|_{y=0} = 0. \quad (3.5)$$

This result also follows immediately from Eq. (1.1). Equation (3.5) is one of the initial conditions required for Eq. (3.4). The two remaining initial conditions can also be obtained explicitly from Eq. (3.3) in terms of Bessel functions of orders $\pm\frac{1}{4}$ and $\pm\frac{3}{4}$ and argument $x^2/8$ [see Eqs. (5) and (8) of Ref. 5].

C. The swallowtail canonical integral $I_4(a_1,a_2,a_3)$

When $n = 4$, we have the following possibilities

$l = 3$, $p = 0,1,2$ with $a_1 = a_2 = 0$,

$l = 2$, $p = 0,1$ with $a_1 = 0$,

$l = 1$, $p = 0$.

Equation (2.19) then yields the following differential equations: for $l = 3$

$$\left.\begin{aligned}
&(p=0)\quad \frac{\partial I^{(1)}_4}{\partial a_3} + \frac{3}{5}a_3 I^{(2)}_4 = 0,\\
&(p=1)\quad \frac{\partial I^{(2)}_4}{\partial a_3} - \frac{3}{5}a_3\frac{\partial I^{(0)}_4}{\partial a_3} - \frac{1}{5}I^{(0)}_4 = 0,\\
&(p=2)\quad \frac{\partial^2 I^{(0)}_4}{\partial a_3^2} + \frac{3}{5}a_3\frac{\partial I^{(1)}_4}{\partial a_3} + \frac{2}{5}I^{(1)}_4 = 0,
\end{aligned}\right\} a_1 = a_2 = 0, \quad (3.6)$$

for $l=2$

$$\left.\begin{aligned}
&(p=0)\quad \frac{\partial^2 I_4^{(0)}}{\partial a_2^{\,2}}+i\frac{2}{5}a_2 I_4^{(1)}+i\frac{3}{5}a_3\frac{\partial I_4^{(0)}}{\partial a_2}=0\,,\\
&(p=1)\quad \frac{\partial^2 I_4^{(1)}}{\partial a_2^{\,2}}-\frac{2}{5}a_2\frac{\partial I_4^{(0)}}{\partial a_2}+i\frac{3}{5}a_3\frac{\partial I_4^{(1)}}{\partial a_2}-\frac{1}{5}I_4^{(0)}=0\,,
\end{aligned}\right\}\ a_1=0\,, \tag{3.7}$$

and for $l=1, p=0$

$$\frac{\partial^4 I_4^{(0)}}{\partial a_1^{\,4}}+\frac{1}{5}a_1 I_4^{(0)}-i\frac{2}{5}a_2\frac{\partial I_4^{(0)}}{\partial a_1}-\frac{3}{5}a_3\frac{\partial^2 I_4^{(0)}}{\partial a_1^{\,2}}=0\,. \tag{3.8}$$

In previous work[8,9,27] the swallowtail canonical integral has been denoted $S(x,y,z)$ so we must make the correspondences $a_1=z$, $a_2=y$, $a_3=x$. Equations (3.6)–(3.8) are equivalent to those reported in Secs. 3.2–3.5 of Ref. 8 if it is also noted that $H=I_4^{(1)}$ and $G=I_4^{(2)}$.

D. The butterfly canonical integral $I_5(a_1,a_2,a_3,a_4)$

We now have $n=5$ and the possible values for l and p are

$$\begin{aligned}
&l=4\,,\quad p=0,1,2,3\quad \text{with } a_1=a_2=a_3=0\,,\\
&l=3\,,\quad p=0,1,2\quad \text{with } a_1=a_2=0\,,\\
&l=2\,,\quad p=0,1\quad \text{with } a_1=0\,,\\
&l=1\,,\quad p=0\,.
\end{aligned}$$

From Eq. (2.19) we can obtain the following sets of differential equations: for $l=4$

$$\left.\begin{aligned}
&(p=0)\quad \frac{\partial I_5^{(1)}}{\partial a_4}-i\frac{2}{3}a_4 I_5^{(3)}=0\,,\\
&(p=1)\quad \frac{\partial I_5^{(2)}}{\partial a_4}-\frac{2}{3}a_4\frac{\partial I_5^{(0)}}{\partial a_4}-\frac{1}{6}I_5^{(0)}=0\,,\\
&(p=2)\quad \frac{\partial I_5^{(3)}}{\partial a_4}-\frac{2}{3}a_4\frac{\partial I_5^{(1)}}{\partial a_4}-\frac{1}{3}I_5^{(1)}=0\,,\\
&(p=3)\quad \frac{\partial^2 I_5^{(0)}}{\partial a_4^{\,2}}-i\frac{2}{3}a_4\frac{\partial I_5^{(2)}}{\partial a_4}-i\frac{1}{2}I_5^{(2)}=0\,,
\end{aligned}\right\}\ a_1=a_2=a_3=0\,, \tag{3.9}$$

for $l=3$

$$\left.\begin{aligned}
&(p=0)\quad \frac{\partial I_5^{(2)}}{\partial a_3}+i\frac{1}{2}a_3 I_5^{(2)}-\frac{2}{3}a_4\frac{\partial I_5^{(0)}}{\partial a_3}=0\,,\\
&(p=1)\quad \frac{\partial^2 I_5^{(0)}}{\partial a_3^{\,2}}+i\frac{1}{2}a_3\frac{\partial I_5^{(0)}}{\partial a_3}+\frac{2}{3}a_4\frac{\partial I_5^{(1)}}{\partial a_3}+i\frac{1}{6}I_5^{(0)}=0\,,\\
&(p=2)\quad \frac{\partial^2 I_5^{(1)}}{\partial a_3^{\,2}}+i\frac{1}{2}a_3\frac{\partial I_5^{(1)}}{\partial a_3}+\frac{2}{3}a_4\frac{\partial I_5^{(2)}}{\partial a_3}+i\frac{1}{3}I_5^{(1)}=0\,,
\end{aligned}\right\}\ a_1=a_2=0\,, \tag{3.10}$$

for $l=2$

$$\left.\begin{aligned}
&(p=0)\quad \frac{\partial^2 I_5^{(1)}}{\partial a_2^{\,2}}-\frac{1}{3}a_2 I_5^{(1)}-\frac{1}{2}a_3\frac{\partial I_5^{(0)}}{\partial a_2}+i\frac{2}{3}a_4\frac{\partial I_5^{(1)}}{\partial a_2}=0\,,\\
&(p=1)\quad \frac{\partial^3 I_5^{(0)}}{\partial a_2^{\,3}}-\frac{1}{3}a_2\frac{\partial I_5^{(0)}}{\partial a_2}+i\frac{1}{2}a_3\frac{\partial I_5^{(1)}}{\partial a_2}+i\frac{2}{3}a_4\frac{\partial^2 I_5^{(0)}}{\partial a_2^{\,2}}-\frac{1}{6}I_5^{(0)}=0\,,
\end{aligned}\right\}\ a_1=0\,, \tag{3.11}$$

and for $l=1, p=0$

$$\frac{\partial^5 I_5^{(0)}}{\partial a_1^5}+i\frac{1}{6}a_1 I_5^{(0)}+\frac{1}{3}a_2\frac{\partial I_5^{(0)}}{\partial a_1} -i\frac{1}{2}a_3\frac{\partial^2 I_5^{(0)}}{\partial a_1^2}-\frac{2}{3}a_4\frac{\partial^3 I_5^{(0)}}{\partial a_1^3}=0\,. \tag{3.12}$$

Equations (3.9)–(3.12) are new results for the butterfly canonical integral.

IV. UNIFORM ASYMPTOTIC CUSPOID APPROXIMATION FOR OSCILLATING INTEGRALS

An important use of I_n and its derivatives $\partial I_n/\partial a_1$, $\partial I_n/\partial a_2,\ldots,\ \partial I_n/\partial a_{n-1}$ is in the derivation of uniform asymptotic expansions[1,2,10,19–23] for integrals of the form

$$\int_{-\infty}^{\infty} g(t)\exp[if(\alpha;t)/\hbar]dt\,,\quad \hbar\to 0\,, \tag{4.1}$$

where $f(\alpha;t)$ is assumed to be real for real values of t and to have n coalescing real or complex saddle points defined by

$$\frac{\partial f(\alpha;t)}{\partial t}=0\quad\text{for } t=t_j(\alpha)\,,\quad j=1,2,\ldots,n\,. \tag{4.2}$$

Note that the positions of the saddle points depend on a set of parameters α.

A key step in the analysis is a local one-to-one uniformly analytic change of variables $t\to u(\alpha;t)$ defined by[1,2,10,19–23]

$$f(\alpha;t)=u^{n+1}+\sum_{i=1}^{n-1}a_{n-i}u^{n-i}+A\,, \tag{4.3}$$

where the new parameters $\{a_{n-i}\}$ and A depend on α but not on t.

In practical applications of uniform approximations, it is necessary to devise methods for calculating $\{a_{n-i}\}$ and A. The equations determining these quantities are[1,2,10,19–23]

$$f_j=u_j^{n+1}+\sum_{i=1}^{n-1}a_{n-i}u_j^{n-i}+A\,,\quad j=1,2,\ldots,n\,, \tag{4.4}$$

where the u_j satisfy

$$(n+1)u_j^n+\sum_{i=1}^{n-1}(n-i)a_{n-i}u_j^{n-i-1}=0\,,\quad j=1,2,\ldots,n \tag{4.5}$$

and for notational convenience we have written $f_j=f(\alpha;t_j)$.

For $n=2$ (fold) and $n=3$ (cusp), explicit formulas exist for the new parameters[4,28–30] in terms of the $\{f_j\}$. For $n\geqslant 4$, however, this is no longer true. To overcome this problem for the swallowtail case $n=4$, we have developed iterative and algebraic techniques for the calculation of the parameters.[9] In this section we outline how these two techniques can be generalized from the case $n=4$ to arbitrary values of n, thereby allowing the application of uniform asymptotic techniques to oscillating integrals with an arbitrary number of coalescing saddle points.

A. Iterative method

The iterative method[9] is applicable when the $\{a_{n-i}\}$ are not close to the caustic associated with the transformation (4.3). On the caustic surface, two or more of the t_j (and hence also the f_j and u_j) have coalesced.

First we simplify Eq. (4.4). Multiplying Eq. (4.5) by u_j shows that

$$u_j^{n+1}=-(u+1)^{-1}\sum_{i=1}^{n-1}(n-i)a_{n-i}u_j^{n-i}\,,\quad j=1,2,\ldots,n\,, \tag{4.6}$$

and substituting Eq. (4.6) into Eq. (4.4) gives

$$f_j=\sum_{i=1}^{n-1}\frac{i+1}{n+1}a_{n-i}u_j^{n-i}+A\,,\quad j=1,2,\ldots,n\,. \tag{4.7}$$

The t_j and u_j are either purely real or occur as complex conjugate pairs and it is important to consider how they are to be ordered. We order the t_j and u_j so that all complex conjugate pairs are next to each other and are before any purely real t_j and u_j. Next we construct the quantities

$$\sum_{i=1}^{n-1}\frac{i+1}{n+1}a_{n-i}(u_{2k-1}^{n-i}-u_{2k}^{n-i})=f_{2k-1}-f_{2k}\,,\quad k=1,2,\ldots,\text{int}(n/2) \tag{4.8}$$

and

$$\sum_{i=1}^{n-1}\frac{i+1}{n+1}a_{n-i}(u_{2k-1}^{n-i}+u_{2k}^{n-i}-u_{2k+1}^{n-i}-u_{2k+2}^{n-i}) =f_{2k-1}+f_{2k}-f_{2k+1}-f_{2k+2}\,,\quad k=1,2,\ldots,\text{int}(n/2)-1\,, \tag{4.9}$$

where as before int = integer part of. When n is an odd integer, we also need the extra equation

$$\sum_{i=1}^{n-1}\frac{i+1}{n+1}a_{n-i}\left[\frac{1}{2}(u_{n-2}^{n-i}+u_{n-1}^{n-i})-u_n^{n-i}\right] =\frac{1}{2}(f_{n-2}+f_{n-1})-f_n\,. \tag{4.10}$$

Note there is always one real root of Eq. (4.5) when n is odd and in Eq. (4.10) we label this root to be u_n.

The left- and right-hand sides of Eqs. (4.8)–(4.10) are both either real or purely imaginary and so can always be written in a purely real form. For given values of $\{u_j\}$ and $\{f_i\}$, Eqs. (4.8)–(4.10) represent $n-1$ *linear* equations in the $n-1$ variables a_{n-i} with $i=1,2,\ldots,n-1$. The iterative procedure to solve them is as follows

(a) Make an initial guess for the vector $\mathbf{a}^{(0)}=(a_1^{(0)},a_2^{(0)},\ldots,a_{n-1}^{(0)})$.

(b) Solve the polynomial Eq. (4.5) for its roots $u_j(\mathbf{a}^0)$ $j=1,2,\ldots,n$.

(c) Order the u_j in the same way as the saddle points t_j of the function $f(\alpha;t)$, i.e., we have $u=u_j\leftrightarrow t=t_j$ for $j=1,2,\ldots,n$.

(d) Solve the linear set of Eqs. (4.8)–(4.10) to obtain a new value for the vector of coefficients $\mathbf{a}^{(1)}$.

(e) Return to (b) and iterate until convergence is achieved.

(f) The remaining coefficient A can then be obtained from Eq. (4.7).

Notice this iterative technique fails if two or more saddle points have coalesced, i.e., if they are on the caustic. In practice it is expected that the method will also become ill conditioned as the $\{a_{n-i}\}$ approach the caustic.

B. Algebraic method

Unlike the iterative scheme, the algebraic method[4,9,30] is expected to be most useful for $\{a_{n-i}\}$ actually on the caustic. The basic idea is to obtain a set of simultaneous polynomial equations for the $\{a_{n-i}\}$.

First, we define the elementary symmetric polynomials[31]

$$\begin{aligned}
\tau_1\{r_i\} &= \sum_{i=1}^{n} r_i \,, \\
\tau_2\{r_i\} &= \sum_{i>j}^{n}\sum_{j=1}^{n} r_i r_j \,, \\
\tau_3\{r_i\} &= \sum_{i>j}^{n}\sum_{j>k}^{n}\sum_{k=1}^{n} r_i r_j r_k \,, \\
&\vdots \qquad\qquad \vdots \\
\tau_n\{r_i\} &= r_1 r_2 r_3 \cdots r_n \,,
\end{aligned} \tag{4.11}$$

together with

$$s_m\{r_i\} = \sum_{i=1}^{n} r_i^m \quad m = 1,2,\ldots \,. \tag{4.12}$$

Second, we introduce the following symmetric functions of the $\{f_i\}$

$$\begin{aligned}
\bar{f}_1 &= \binom{n}{1}^{-1} \tau_1\{f_i\} \,, \\
\bar{f}_2 &= \binom{n}{2}^{-1} \tau_2\{f_i - \bar{f}_1\} \,, \\
\bar{f}_3 &= \binom{n}{3}^{-1} \tau_3\{f_i - \bar{f}_1\} \,, \\
&\vdots \qquad\qquad \vdots \\
\bar{f}_j &= \binom{n}{j}^{-1} \tau_j\{f_i - \bar{f}_1\} \,, \\
&\vdots \qquad\qquad \vdots \\
\bar{f}_n &= \binom{n}{n}^{-1} \tau_n\{f_i - \bar{f}_1\} \,,
\end{aligned} \tag{4.13}$$

where the $\binom{n}{j}$ are binomial coefficients. Notice that the $\bar{f}_j$ are all real even if some of the f_i are complex. Since everything on the right-hand sides of Eq. (4.13) is assumed to be known, the $\{\bar{f}_j\}$ are also known quantities. We now write each $\tau_j\{f_i - \bar{f}\}$ in terms of $\{\tau_j\{f_i\}\}$ and then in terms of $\{s_j\{f_i\}\}$. We now use Eq. (4.7) to express the $s_j\{f_i\}$ in terms of $\{s_m\{u_i\}\}$. At this point we have obtained $\{\bar{f}_j\}$ in terms of $\{s_m\{u_i\}\}$. The final step is to use Newton's formulas[32] for the sums of powers of the roots of the polynomial Eq. (4.5) to write $\{s_m\{u_i\}\}$ and hence, $\{\bar{f}_j\}$ in terms of $\{a_{n-i}\}$. It should be noted that the algebraic steps just described involve a huge amount of manipulation for $n \geqslant 4$ and it is only feasible to carry them out with the help of a symbolic algebraic computer program such as REDUCE or MACSYMA.

The result of the operations outlined above is a set of $n-1$ polynomial equations in the variables a_{n-i}, $i = 1,2,\ldots,n-1$ and one equation in the $\{a_{n-i}\}$ and A. The solution of these equations for given input data $\{f_j\}$ then yields the desired values of $\{a_{n-i}\}$ and A.

The main advantage of the algebraic method is that it remains valid on the caustic. Indeed in this circumstance the set of polynomial equations for the $\{a_{n-i}\}$ and A are expected to be easier to solve numerically than when the $\{a_{n-i}\}$ are off the caustic. In addition, if $f(\alpha;t)$ has a special symmetry which results in some of the a_{n-i} being identically zero, the set of polynomial equations will simplify and also be easier to solve. Thus in practice the iterative and algebraic methods for determining $\{a_{n-i}\}$ and A are expected to be complementary.

V. SUMMARY AND CONCLUSIONS

In this paper we have derived the differential equations satisfied by the cuspoid canonical integral $I_n(\mathbf{a})$ and its partial derivatives for any value of n. Our method consists in finding a set of coupled linear ordinary differential equations for each step in the sequence $I_n(0,0,\ldots,0,0) \rightarrow I_n(0,0,\ldots,0,a_{n-1}) \rightarrow I_n(0,0,\ldots,a_{n-2},a_{n-1}) \rightarrow \cdots \rightarrow I_n(0,a_2,\ldots,a_{n-2},a_{n-1}) \rightarrow I_n(a_1,a_2,\ldots,a_{n-2},a_{n-1})$. The results obtained in Sec. II provide everything that is required for the method (differential equations, transformations, and initial conditions).

In practical implementations of this technique for the numerical evaluation of $I_n(\mathbf{a})$ and its partial derivatives, it is also necessary to consider the stability of the differential equations when they are integrated. Previous calculations for $n = 3$ (cusp)[4–6,18] and $n = 4$ (swallowtail)[8] indicate that the method will be numerically stable when all the saddle points of $P_n(u)$ [see Eq. (2.2)] are real, because this condition implies that the independent solutions of the differential equations are then all of oscillatory type. It is important to note that the region of $(a_1,a_2,\ldots,a_{n-1})$ space which corresponds to all real saddle points for $P_n(u)$ is also the region where $I_n(\mathbf{a})$ has the richest structure. Outside of this region the independent solutions of the differential equations will contain exponentially increasing solutions which will eventually limit the accuracy to which $I_n(\mathbf{a})$ can be computed.

We have also considered the problem of determining the parameters $\mathbf{a}$ for use in the uniform asymptotic cuspoid approximation. We have shown how the iterative and algebraic methods developed earlier[9] for the special case of $n = 4$ can be generalized to arbitrary values of n.

Finally, we wish to emphasize that the treatment presented in this paper has unified and generalized previous researches on the cuspoid canonical integrals, in which each value of n had been treated as a special case.

ACKNOWLEDGMENTS

J.N.L.C. thanks Professor R. A. Marcus for his hospitality at the A.A. Noyes Laboratory of Chemical Physics, California Institute of Technology, Pasadena, California and NATO for a Senior Scientist award. P.R.C. thanks the States of Jersey Education Committee for a research studentship. The algebraic calculations were carried out on the CDC 7600 computer of the University of Manchester Regional Computer Centre.

[1]J. N. L. Connor, Mol. Phys. **26**, 1217 (1973).
[2]J. N. L. Connor, Mol. Phys. **27**, 853 (1974). On p. 861, lines 25 and 26, the words "even" and "odd" should be interchanged.
[3]F. J. Wright, J. Phys. A: Math. Gen. **13**, 2913 (1980).
[4]J. N. L. Connor and D. Farrelly, J. Chem. Phys. **75**, 2831 (1981).
[5]J. N. L. Connor and D. Farrelly, Chem. Phys. Lett. **81**, 306 (1981).
[6]J. N. L. Connor and P. R. Curtis, J. Phys. A: Math. Gen. **15**, 1179 (1982).
[7]J. J. Stamnes and B. Spjelkavik, Opt. Acta **30**, 1331 (1983).
[8]J. N. L. Connor, P. R. Curtis, and D. Farrelly, Mol. Phys. **48**, 1305 (1983). In Eq. (3.46) for $\frac{1}{3}\Gamma(\frac{1}{3})$ read $\frac{2}{3}\Gamma(\frac{1}{3})$.
[9]J. N. L. Connor, P. R. Curtis, and D. Farrelly, J. Phys. A: Math. Gen. **17**, 283 (1984). In Table **4**, for $c = -20.957$ read $c = -20.597$.
[10]J. N. L. Connor, Mol. Phys. **31**, 33 (1976).
[11]R. Thom, *Structural Stability and Morphogenesis. An Outline of a General Theory of Models*, translated from the French edition, as updated by the author, by D. H. Fowler, with a Foreword by C. H. Waddington (Benjamin, Reading, MA, 1975).
[12]T. Poston and I. Stewart, *Catastrophe Theory and its Applications* (Pitman, London, 1978).
[13]R. Gilmore, *Catastrophe Theory for Scientists and Engineers* (Wiley, New York, 1981).
[14]G. B. Airy, Trans. Cambridge Philos. Soc. **6**, 379 (1838).
[15]J. C. P. Miller, *The Airy Integral, Giving Tables of Solutions of the Differential Equation* $y'' = xy$ (Cambridge U.P., Cambridge, England, 1971), Part-Vol. B.
[16]L. Brillouin, Ann. Sci. École Norm. Sup., 3rd Ser. **33**, 17 (1916).
[17]T. Pearcey, Philos. Mag. **37**, 311 (1946).
[18]T. Pearcey and G. W. Hill, editors, *The Effect of Aberrations Upon the Optical Focus* (Commonwealth, Scientific and Industrial Research Organization, Melbourne, Australia, 1963).
[19]D. Ludwig, Commun. Pure Appl. Math. **19**, 215 (1966).
[20]N. Bleistein, J. Math. Mech. **17**, 533 (1967).
[21]S. O. Rice, Bell Syst. Tech. J. **47**, 1970 (1968).
[22]F. Ursell, Proc. Cambridge Philos. Soc. **72**, 49 (1972).
[23]J. Martin, Proc. Cambridge Philos. Soc. **76**, 211 (1974).
[24]G. H. Hardy, Quart. J. **41**, 226 (1910), reprinted in *Collected Papers of G. H. Hardy*, Vol. IV, edited by a Committee appointed by The London Mathematical Society (Clarendon, Oxford, 1969), p. 460.
[25]C. A. Stewart, *Advanced Calculus* (Methuen, London, 1951), 3rd ed., Sec. 12.7.
[26]W. Gröbner and N. Hofreiter, *Integraltafel*, Zweiter Teil *Bestimmte Integrale*, Funfte, verbesserte Auflage (Springer, Vienna, Austria, 1973), Sec. 333, p. 118, items (11a) and (11b).
[27]J. N. L. Connor, P. R. Curtis, C. J. Edge, and A. Laganà, J. Chem. Phys. **80**, 1362 (1984).
[28]C. Chester, B. Friedman, and F. Ursell, Proc. Cambridge Philos. Soc. **53**, 599 (1957).
[29]J. N. L. Connor and R. A. Marcus, J. Chem. Phys. **55**, 5636 (1971).
[30]D. H. Peregrine and R. Smith, Philos. Trans. R. Soc. London, Ser. A **292**, 341 (1979).
[31]A. Mostowski and M. Stark, *Introduction to Higher Algebra*, translated from the Polish by J. Musielak (Pergamon, Oxford, 1964), p. 346.
[32]J. V. Uspensky, *Theory of Equations* (McGraw-Hill, New York, 1948), pp. 260–262.

Discriminant, transmission coefficient, and stability bands of Hill's equation

Joseph B. Keller
Departments of Mathematics and Mechanical Engineering, Stanford University, Stanford, California 94305

(Received 3 February 1984; accepted for publication 25 May 1984)

The discriminant $\Delta(k^2)$ of Hill's equation is shown to be related to the transmission coefficient $T(k)e^{i\theta(k)}$ of one period of the potential by $\Delta(k^2) = [2/T(k)]\cos[k\pi + \theta(k)]$. This result is used to find the boundaries of the stability bands.

PACS numbers: 02.30.Hq

I. INTRODUCTION

Hill's equation is the following second-order ordinary differential equation, in which the potential $Q(x) = Q(x+\pi)$ is periodic with period π:

$$y''(x,k^2) + [k^2 - Q(x)]y(x,k^2) = 0. \tag{1}$$

The Floquet exponents, which govern the behavior of solutions of this equation, are determined by the discriminant $\Delta(k^2)$ (see Ref. 1). We shall prove the following theorem, which relates $\Delta(k^2)$ to the transmission coefficient $T(k)e^{i\theta(k)}$ associated with one period of the potential.

Theorem 1: For $k^2 > 0$,

$$\Delta(k^2) = [2/T(k)]\cos[k\pi + \theta(k)]. \tag{2}$$

The significance of this theorem is that information about either the discriminant or the transmission coefficient can shed light on the other. This will be illustrated by using it to find the boundaries of the stability bands.

Proof of the theorem: The discriminant is defined by $\Delta(k^2) = y_1(\pi,k^2) + y_2'(\pi,k^2)$ where y_1 and y_2 are two solutions of (1) satisfying the initial conditions

$$y_1(0,k^2) = 1, \quad y_1'(0,k^2) = 0; \quad y_2(0,k^2) = 0, \quad y_2'(0,k^2) = 1. \tag{3}$$

To define the transmission coefficient we modify (1) by setting $Q(x) = 0$ for $x < 0$ and $x > \pi$, keeping $Q(x)$ unchanged for $0 \leqslant x \leqslant \pi$. Then we introduce a complex solution $y_L(x,k)$ of the modified equation corresponding to a wave of unit amplitude incident from the left. It is partially reflected with the complex reflection coefficient $R(k)e^{i\varphi(k)}$ and partially transmitted with the complex transmission coefficient $T(k)e^{i\theta(k)}$. Thus $y_L(x,k)$ must have the form

$$y_L(x,k) = \begin{cases} e^{ikx} + R(k)e^{i\varphi(k)}e^{-ikx}, & x \leqslant 0, \\ T(k)e^{i\theta(k)}e^{ikx}, & x \geqslant \pi. \end{cases} \tag{4}$$

In addition $y_L(x,k)$ must satisfy (1) in the interval $0 < x < \pi$ and be continuous with a continuous derivative. These conditions uniquely determine the solution. They also uniquely determine R, φ, T, and θ provided we require that both R and T be real and non-negative, and that φ and θ lie in the interval $-\pi < \varphi, \theta \leqslant \pi$.

Next we introduce another complex solution $y_R(x,k)$ of the modified equation, corresponding to a wave of unit amplitude incident from the right. By using the constancy of the Wronskian of y_L and y_R, it can be shown that both solutions have the same transmission coefficient. Furthermore, from the constancy of the Wronskian of y_L and $\bar{y}_R$, the complex conjugate of y_R, it follows that the reflection coefficient in y_R is $-Re^{i(2\theta-\varphi)}$. Thus y_R must have the form

$$y_R(x,k) = \begin{cases} T(k)e^{i\theta(k)}e^{-ikx}, & x \leqslant 0, \\ e^{-ikx} - R(k)e^{i[2\theta(k)-\varphi(k)]}e^{ikx}, & x \geqslant \pi. \end{cases} \tag{5}$$

The continuity of y_R and y_R', together with (1) in the interval $0 < x < \pi$, uniquely determine y_R.

Since y_L and y_R are linearly independent solutions of (1) in the interval $0 < x < \pi$, we can express y_1 and y_2 as linear combinations of them in the closure of that interval. By using the initial conditions (3), we find the linear combinations to be

$$y_1(x,k^2) = \frac{1}{2}y_L(x,k) + \frac{1}{2Te^{i\theta}}(1 - Re^{i\varphi})y_R(x,k), \quad 0 \leqslant x \leqslant \pi, \tag{6}$$

$$y_2(x,k^2) = \frac{1}{2ik}y_L(x,k) - \frac{1}{2ikTe^{i\theta}}(1 + Re^{i\varphi})y_R(x,k), \quad 0 \leqslant x \leqslant \pi. \tag{7}$$

Finally we use (6) and (7) in the definition of $\Delta(k^2)$, and use the fact that $T^2 + R^2 = 1$, which follows from the constancy of the Wronskian of y_L and $\bar{y}_L$. The result is just (2), which proves the theorem.

The left side of (2) is even in k, so the right side must also be even. This implies that $T(k) = T(-k)$ and $\theta(k) = -\theta(-k)$. This same conclusion follows from the observation that $\bar{y}_L(x,-k) = y_L(x,k)$ since both of these functions are solutions of the same equation with the same incident wave.

II. APPLICATION TO STABILITY BOUNDARIES

As an application of Theorem 1, we shall determine k_n^+ and k_n^-, the boundaries of the nth stability band. They are roots of the equations $\Delta(k^2) = 2$ and $\Delta(k^2) = -2$, respectively. Upon using (2) for $\Delta(k^2)$ in these two equations, we obtain $\cos[k\pi + \theta(k)] = \pm T(k)$. Now taking the inverse cosine yields

$$k\pi + \theta(k) = \cos^{-1}[\pm T(k)] = (n + \tfrac{1}{2})\pi - (-1)^n \times \sin^{-1}[\pm T(k)] = (n + \tfrac{1}{2})\pi \mp (-1)^n \sin^{-1}T(k). \tag{8}$$

In order to rewrite (8) in a more convenient form, we introduce the function F defined by

$$F^{-1}(k) \equiv k\pi + \theta(k). \tag{9}$$

Then the left side of (8) can be inverted, i.e., solved for k, with the result

 0022-2488/84/102903-02$02.50

$$k_n^{\pm} = F\left[(n+\tfrac{1}{2})\pi \mp (-1)^n \sin^{-1} T(k_n^{\pm})\right]. \tag{10}$$

We have written $k_n^{\pm}$ to emphasize the dependence of the solution on n and on the choice of sign. Of course $\sin^{-1} T$ must be assigned its principal value.

We can proceed further if $T(k_n^{\pm})$ is small, i.e., if $|T(k_n^{\pm})| \ll 1$. Then $\sin^{-1} T = T + O(T^3)$. Upon expanding F for T small, we get from (10)

$$k_n^{\pm} = F[(n+\tfrac{1}{2})\pi] \mp (-1)^n F'[(n+\tfrac{1}{2})\pi] \times T\{F[(n+\tfrac{1}{2})\pi]\} + O(T^2). \tag{11}$$

This result yields $k_n^{\pm}$ in terms of two functions F and T, which are determined by the phase and amplitude of the transmission coefficient. Then the width $|B_n|$ of the nth stability band is given by

$$|B_n| = |k_n^{+} - k_n^{-}| = 2F'[(n+\tfrac{1}{2})\pi] T\{F[(n+\tfrac{1}{2})\pi]\} + O(T^2). \tag{12}$$

The results (11) and (12) can be made more explicit by asymptotic methods, such as the WKB method, to calculate T and θ. Certain asymptotic results which can be obtained in this way are given by Weinstein and Keller.[2]

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research, the Air Force Office of Scientific Research, the Army Research Office, and the National Science Foundation.

[1]W. Magnus and S. Winkler, *Hill's Equation* (Interscience, New York, 1966).
[2]M. Weinstein and J. B. Keller, "Hill's equation with a large potential," SIAM J. Appl. Math. (to be published).

Lax-pairs, spectral problems, and recursion operators

Walter Strampp
Fachbereich 17, Mathematik, Gesamthochschule Kassel, Wilhelmshöher Allee 73, 3500 Kassel, West Germany

(Received 1 November 1983; accepted for publication 24 February 1984)

We present various examples for the connection between Lax-pairs and recursion operators. From this connection a new method for constructing recursion operators is derived. As an application we find recursion operators to some integrable equations newly found by Wadati *et al.* [J. Phys. Soc. Jpn. Lett. **46**, 1965 (1979); J. Phys. Soc. Jpn. **47**, 1698 (1979)].

PACS numbers: 02.30.Jr

I. STATEMENT OF THE PROBLEM

Let $u=(u^1,...,u^n)^T$ denote a vector valued function depending upon the space variable x and the time variable t. By $u_t=(u_t^1,...,u_t^n)^T$, $u_k=(u_k^1,...,u_k^n)^T$, where $u_k^1=D^k u^1=(\partial^k/\partial x^k)(u^1)$, we denote the partial derivatives of u with respect to t and x. Finally the symbol $f[u]$ is used for functions depending upon u and higher partial derivatives u_k up to arbitrary but finite order.

The object of this paper is the evolution equation

$$u_t=K[u],\quad K=(K^1,...,K^n)^T, \tag{1}$$

and its linearized version

$$v_t=A(K[u])v,\quad v=(v^1,...,v^n)^T, \tag{2}$$

where the operator $A(K[u])$ is given by the matrix

$$A_{jl}(K[u])=\frac{\partial K^j[u]}{\partial u_k^l}D^k$$

($D^k=\partial^k/\partial x^k$). An evolution equation

$$u_t=M[u]$$

is called a symmetry for (1) if

$$A(K[u])M[u]-A(M[u])K[u]=0, \tag{3}$$

for all solutions u of Eq. (1), holds. (See Ref. 1 for a motivation of this notation.) We note that the equation

$$u_t=u_1$$

gives a trivial symmetry for Eq. (1).

Next, we shall consider the problem of generating symmetries starting from the trivial one. This problem has been studied first within the algebra of polynomial functions $f[u]$.[1] The following theorem has been obtained by Olver.[1]

Theorem 1: Let

$$R[u]=a^p[u]D^p+a^{p-1}[u]D^{p-1}+\cdots+a^0[u]$$

be an operator with polynomial coefficients.

Assume that for all solutions of (1) the identity

$$R[u]_t=[A(K[u]),R[u]] \tag{4}$$

holds. Then, for all $n>0$ the evolution equations

$$u_t=(R[u])^n u_1$$

will form symmetries of Eq. (1).

Operators satisfying condition (4) are called recursion operators.

In order to obtain interesting applications of the above theorem it is necessary to allow recursion operators that involve the inverse derivative D^{-1}. This requires the construction of a suitable function space on which D^{-1} is well defined. For a nice exposition of these constructions the reader may consult Ref. 2, where the concept of extended symmetries is introduced.

Now let $M[u]$ be a function depending upon $u_0, u_1,...,u_p$ as well as upon $u_{-1}, u_{-2},...,u_{-q}$, where u^l_{-h} means $D^{-h}u^2$. The operator of linearization is given by

$$A_{jl}(M[u])=\frac{\partial M^j[u]}{\partial u_k^l}D^k,$$

where k now ranges from $-q$ to p. Extended symmetries for Eq. (1) are then given by evolution equations

$$u_t=M[u];$$

where condition (3) is satisfied for all solutions of Eq. (1).

Extended symmetries thus are extensions of Lie–Bäcklund symmetries. Now, the following theorem can be found in Ref. 2.

Theorem 2: Let

$$R[u]=a^k[u]D^k+a^{k-1}[u]D^{k-1}+\cdots+a_0[u]+a_{-1}[u]D^{-1}+\cdots+a^{-m}[u]D^{-m}$$

be an operator defined on a suitable function space. Assume that for all solutions of Eq. (1) identity (4) holds. Then, for all $n>0$ the evolution equations

$$u_t=(R[u])^n u_1$$

will form symmetries of Eq. (1).

(We emphasize that the assumption of polynomial dependence of the functions f, M, a^h upon $u_0, u_1,...$and $u_{-1}, u_{-2},...$is dropped in Ref. 2.)

The theory of recursion operators (also called strong symmetries) has been developed further by various authors.[3–6] In particular, the hereditary property of strong symmetries has been studied. The hereditary property is a sufficient condition of the fact that all flows generated by a strong symmetry commute.

Equation (4) possesses the shape of a Lax equation. Now let us assume that there exists a further Lax-pair for Eq. (1). Let us denote this by the following spectral problem:

$$L(u,\varphi,\lambda)=0, \tag{5a}$$

$$P(u,\varphi,\lambda)=\varphi_t, \tag{5b}$$

 0022-2488/84/102905-05$02.50

where L and P are linear in $\varphi = (\varphi^1,...,\varphi^n)^T$ and λ is constant.

The interrelations between problems (4) and (5) have been revealed in Ref. 6 where, furthermore, a constructive approach for obtaining strong symmetries has been given. This approach proceeds by using solely the isospectral eigenvalue problem (5a). The next step in this algorithm is the calculation of the gradient of the eigenvalue λ with respect to the solution u. If one can then find the eigenvalue equation that the gradient satisfies, one obtains the recursion operator as the adjoint of the gradient.

In this paper we shall attack the problem by a more direct approach. We shall demonstrate that for various equations there exists a transformation $T(\varphi,\lambda)$ mapping the eigenfunctions of (5) into eigenfunctions $\psi = (\psi^1,...,\psi^n)^T$ satisfying the equations

$$R[u]\psi = f(\lambda)\psi, \quad A(K[u])\psi = \psi_t.$$

This motivates the following method for obtaining recursion operators.

Let us be given an evolution equation (1) together with the Lax-pair (5). From Eq. (1) we immediately obtain its linearization. Suppose we can find a transformation $T(\varphi,\lambda)$ mapping eigenfunctions φ of Eqs. (5) into solutions of the linearized equation

$$A(K[u])\psi = \psi_t. \tag{6}$$

Then we have to bring the eigenvalue equation

$$L(u, T^{-1}\psi, \lambda) = 0$$

into the form

$$R[u]\psi = f(\lambda)\psi. \tag{7}$$

Since λ was a constant eigenvalue, it follows from Eqs. (6) and (7) that Eq. (4) holds for all solutions u of Eq. (1). Hence, $R[u]$ is a recursion operator for Eq. (1).

II. INTRODUCTORY EXAMPLES

In this section we shall consider first Burgers' equation for demonstrating the connection between Lax-pairs and recursion operators. Next, we shall consider KdV and Schrödinger's equation, which both arise from the AKNS inverse scattering scheme.

Burgers' equation is given by

$$u_t = K[u] = u_2 + uu_1. \tag{8}$$

A Lax-pair for Eq. (8) is formed by

$$D\varphi + \tfrac{1}{2}u\varphi = \lambda\varphi, \tag{9a}$$

$$\varphi_t = D^2\varphi + uD\varphi. \tag{9b}$$

The linearization of Eq. (8) becomes

$$\psi_t = A(K[u])\psi = D^2\psi + uD\psi + u_1\psi. \tag{10}$$

Now it is easy to see that Eq. (9b) is carried over to Eq. (10) by the transformation

$$\psi = D\varphi.$$

If we insert $\varphi = D^{-1}\psi$, Eq. (9a) takes the form

$$\psi + \tfrac{1}{2}uD^{-1}\psi = \lambda D^{-1}\psi.$$

By differentiation, the equation

$$D\psi + \tfrac{1}{2}u\psi + \tfrac{1}{2}u_1D^{-1}\psi = \lambda\psi$$

follows showing that

$$R[u] = D + \tfrac{1}{2}u + (1/2)u_1D^{-1}$$

is a recursion operator for Eq. (8).

Next, we consider the KdV equation

$$u_t = K[u] = u_3 + uu_1. \tag{11}$$

The Lax-pair is given by

$$D^2\varphi + \tfrac{1}{6}u\varphi = \lambda\varphi, \tag{12a}$$

$$\varphi_t = 4D^3\varphi + uD\varphi + \tfrac{1}{2}u_1\varphi. \tag{12b}$$

Equation (11) is linearized by

$$\psi_t = A(K[u])\psi = D^3\psi + uD\psi + u_1\psi. \tag{13}$$

The transformation

$$\psi = D(\varphi^2)$$

maps solutions of Eqs. (12a) and (12b) into solutions of Eq. (13). Inserting $\varphi = (D^{-1}\psi)^{1/2}$ into Eq. (12a) yields

$$D^2\psi + \tfrac{2}{3}u\psi + \tfrac{1}{3}u_1D^{-1}\psi = 4\lambda\psi.$$

Thus, we obtain the well-known recursion operator

$$R[u] = D^2 + \tfrac{2}{3}u + \tfrac{1}{3}u_1D^{-1}.$$

The final example in this section is the Schrödinger equation

$$iq_t = -q_2 - 2|q|^2q,$$

which must be treated as a two-component system

$$u_t = K[u], \tag{14}$$

where $u = (u^1,u^2)^T$ and

$$K[u] = \begin{pmatrix} iu_2^1 - 2i(u^1)^2u^2 \\ -iu_2^2 + 2iu^1(u^2)^2 \end{pmatrix}.$$

Obviously, Eq. (14) yields the scalar Schrödinger equation upon setting $u^2 = \bar{u}^1$.

The AKNS spectral problem

$$D\varphi^1 - u^1\varphi^2 = \lambda\varphi^1, \quad -D\varphi^2 + u^2\varphi^1 = \lambda\varphi^2, \tag{15}$$

and

$$\begin{aligned} \varphi_t^1 &= (2i\lambda^2 - iu^1u^2)\varphi^1 + (iu_1^1 + 2i\lambda u^1)\varphi^2, \\ \varphi_t^2 &= (-iu_1^2 + 2i\lambda u^2)\varphi^1 - (2i\lambda^2 - iu^1u^2)\varphi^2 \end{aligned} \tag{16}$$

give a Lax-pair for Eq. (14) if we eliminate the eigenvalue λ from Eq. (16). The transformation

$$\psi^1 = (\varphi^1)^2, \quad \psi^2 = (\varphi^2)^2$$

now maps solutions φ of Eqs. (15) and (16) into solutions ψ of the linearization of Eq. (14) given by

$$\psi_t = A(K[u])\psi, \tag{17}$$

where $A(K[u])$ is the matrix

$$\begin{pmatrix} iD^2 - 4iu^1u^2 & -2i(u^1)^2 \\ +2i(u^2)^2 & -iD^2 + 4iu^1u^2 \end{pmatrix}.$$

The role of the squared eigenfunctions for the determination of all equations integrable by the AKNS scheme has been stressed already in Ref. 7. There it is shown that the squared eigenfunctions satisfy the eigenvalue problem

$$R[u]\psi = 2\lambda\psi,$$

where

$$R[u] = \begin{pmatrix} D - 2u^1 D^{-1} u^2 & -2u^1 D^{-1} u^1 \\ 2u^2 D^{-1} u^2 & -D + 2u^2 D^{-1} u^1 \end{pmatrix}.$$

Thus $R[u]$ forms a recursion operator.

III. JAULENT–MIODEK EQUATIONS

In Ref. 8 Jaulent and Miodek introduced a new class of equations being integrable by the inverse scattering method. The underlying isospectral eigenvalue problem differs from the AKNS scheme by the fact that the parameter λ^2 appears in the eigenvalue equation. For convenience we restrict ourselves here to the following equation $u_t = K[u]$:

$$u_t^1 = \tfrac{1}{4}u_3^2 - u^1 u_1^2 - \tfrac{1}{2}u_1^1 u^2, \quad u_t^2 = -u_1^1 - \tfrac{3}{2}u^2 u_1^2. \tag{18}$$

Equations (18) appear as integrability conditions of the following spectral problem:

$$D^2\varphi + (\lambda^2 - \lambda u^2 - u^1)\varphi = 0, \tag{19a}$$

$$\varphi_t = \tfrac{1}{4}u_1^2\varphi - \tfrac{1}{2}(u^1 + 2\lambda)\varphi. \tag{19b}$$

Obviously, the scalar problem (19) can be cast into the form (5) by introducing the eigenvector $(\varphi, D\varphi)$. It is not necessary for us to use this form here.

The equations of linearization now become

$$\psi_t = A(K[u])\psi, \tag{20}$$

where

$$A(K[u]) = \begin{pmatrix} -\tfrac{1}{2}u^2 D - u_1^2 & \tfrac{1}{4}D^3 - u^1 D - (1/2)u^1 \\ -D & \tfrac{3}{2}u^2 D - (3/2)u_1^2 \end{pmatrix}.$$

The transformation

$$\psi^1 = -(u^2 - \lambda)D(\varphi)^2 + \tfrac{1}{2}u_1^2(\varphi)^2, \quad \psi^2 = D(\varphi)^2$$

maps solutions of the spectral problem (19) into solutions of the equation of linearization (20). By setting

$$\varphi = (D^{-1}\psi^2)^{1/2},$$

the eigenvalue equation (19a) becomes, upon integration,

$$(-\tfrac{1}{4}D^2 + u^1 + \tfrac{1}{2}u_1^1 D^{-1})\psi^2 + (u^2 + \tfrac{1}{2}u_1^2 D^{-1})\psi^2 = \lambda^2\psi^2.$$

This is equivalent to the system of equations

$$\begin{pmatrix} 0 & -\tfrac{1}{4}D^2 + u^1 + \tfrac{1}{2}u_1^1 D^{-1} \\ 1 & u^2 + \tfrac{1}{2}u_1^2 D^{-1} \end{pmatrix}\begin{pmatrix}\psi^1 \\ \psi^2\end{pmatrix} = \lambda\begin{pmatrix}\psi^1 \\ \psi^2\end{pmatrix},$$

since the relation

$$\psi^1 = -(u^2 - \lambda + \tfrac{1}{2}u_1^2 D^{-1})\psi^2$$

holds. Therefore, we obtain the operator

$$R[u] = \begin{pmatrix} 0 & -\tfrac{1}{4}D^2 + u^1 + \tfrac{1}{2}u_1^1 D^{-1} \\ 1 & u^2 + \tfrac{1}{2}u_1^2 D^{-1} \end{pmatrix}$$

as recursion operator for Eq. (18). This result has been derived by using different methods in Refs. 8 and 9.

IV. WADATI–KONNO–ICHIKAWA SPECTRAL PROBLEMS

We now turn to generalized versions of the AKNS scheme which are known in the literature as Wadati–Konno–Ichikawa spectral problems. The idea is to consider the following eigenvalue problem:

$$D\varphi^1 + F(\lambda)\varphi^1 = G(\lambda)u^1\varphi^2,$$

$$D\varphi^2 - F(\lambda)\varphi^2 = G(\lambda)u^2\varphi^1.$$

(See Ref. 10.) In a more special version this idea already appears in Ref. 11 where Kaup and Newell solve the inverse scattering problem for a derivative Schrödinger equation.

Let us consider first Dym's equation

$$u_t + K[u] = 2D^3\{(1+u)^{-1/2}\}. \tag{21}$$

The reader may consult Refs. 12 and 13 for an explanation as to how this equation arises from the Wadati–Konno–Ichikawa scheme. In Ref. 12 the following scalar spectral problem is stated:

$$D^2\varphi + (1+u)\lambda^2\varphi = 0, \tag{22a}$$

$$\varphi_t = 4(1+u)^{-1/2}\lambda^2 D\varphi - 2D\{(1+u)^{-1/2}\}\lambda^2\varphi. \tag{22b}$$

By simple manipulations Eqs. (22) can be written as

$$-(1+u)^{-1}D^2\varphi = \lambda^2\varphi, \tag{23a}$$

$$\varphi_t = -4(1+u)^{-3/2}D^3\varphi - 2D\{(1+u)^{-3/2}\}D^2\varphi, \tag{23b}$$

so that we have a classical Lax-pair for Eq. (21).

The linearized equation $\psi_t = A(K[u])\psi$ can be conveniently written in the form

$$\psi_t = -D^3\{(1+u)^{-3/2}\psi\}. \tag{24}$$

Solutions of the spectral problem (23) are mapped into solutions of the linearized equation (24) by the transformation

$$\psi = D^3(\varphi)^2.$$

The inverse transformation $\varphi = (D^{-3}\psi)^{(1/2)}$ yields

$$\{2(1+u)D^{-2} + u_1 D^{-3}\}\psi = -(1/2\lambda^2)\psi,$$

by inserting into Eq. (23a). Therefore, we obtain the recursion operator

$$R[u] = 2(1+u)D^{-2} + u_1 D^{-3},$$

which can also be found in Ref. 2.

Next, we shall use the algorithm developed in Sec. I for finding a recursion operator of the following equation $u_t = K[u]$:

$$u_t^1 = iD^2(u^1 v^{-1}),$$

$$u_t^2 = -iD^2(u^2 v^{-1}), \quad v = (1 - u^1 u^2)^{1/2}. \tag{25}$$

If we set $u^2 = -\bar{u}^1$, Eq. (25) becomes the Schrödinger-type equation

$$u_t^1 = iD^2\{u^1(1 + |u^1|^2)^{-1/2}\}.$$

Equation (25) is the integrability condition for the following Wadati–Konno–Ichikawa spectral problem (see Ref. 13):

$$D\varphi^1 + i\lambda\varphi^1 = \lambda u^1\varphi^2, \quad D\varphi^2 - i\lambda\varphi^2 = \lambda u^2\varphi^1, \tag{26a}$$

and

$$\varphi_t^1 = -2iv^{-1}\lambda^2\varphi^1 + \{2u^1 v^{-1}\lambda^2 + iD(u^1 v^{-1})\lambda\}\varphi^2,$$

$$\varphi_t^2 = \{2u^2 v^{-1}\lambda^2 - iD(u^2 v^{-1})\lambda\}\varphi^1 + 2iv^{-1}\lambda^2\varphi^2. \tag{26b}$$

Let us briefly demonstrate the steps for obtaining a recursion operator. First of all, we write Eqs. (26) as

$$v^{-2}(D\varphi^1 + iu^1 D\varphi^2) = -i\lambda\varphi^1,$$
$$v^{-2}(D\varphi^2 - iu^2 D\varphi^2) = i\lambda\varphi^2, \tag{27a}$$

and

$$\varphi_t^1 = 2iv^{-1}D\{v^{-2}(D\varphi^1 + iu^1 D\varphi^2)\} + D(u^1 v^{-1})v^{-2}(D\varphi^2 - iu^2 D\varphi^1),$$
$$\varphi_t^2 = -2iv^{-1}D\{v^{-2}(D\varphi^2 - iu^2 D\varphi^1)\} + D(u^2 v^{-1})v^{-2}(D\varphi^1 + iu^1 D\varphi^2), \tag{27b}$$

yielding a Lax-pair for Eq. (25).

The linearized equation $\psi_t = A(K[u])\psi$ can be written as

$$\psi_t^1 = iD^2\{(v^{-1} + \tfrac{1}{2}u^1u^2v^{-3})\psi^1\} + iD^2\{\tfrac{1}{2}(u^1)^2v^{-3}\psi^2\},$$
$$\psi_t^2 = -iD^2\{\tfrac{1}{2}(u^2)^2v^{-3}\psi^1\} - iD^2\{(v^{-1} + \tfrac{1}{2}u^1u^2v^{-3})\psi^2\}. \tag{28}$$

The transformation

$$\psi^1 = D(\varphi^1)^2, \quad \psi^2 = D(\varphi^2)^2$$

maps solutions of the spectral problem (27) into solutions of the linearized equation (28). Now, the recursion operator is found by inserting $\varphi^1 = (D^{-1}\psi^1)^{1/2}$, $\varphi^2 = (D^{-1}\psi^2)^{1/2}$ into the eigenequation (27a). This yields

$$R[u]\psi = (1/\lambda)\psi,$$

where $R[u]$ is given by the matrix

$$\begin{pmatrix} -2iD^{-1} + u^1D^{-1}(iu^2v^{-2}/(1+u^1u^2v^{-2})) & -u^1D^{-1}(iu^1v^{-2}/(1+u^1u^2v^{-2})) \\ u^2D^{-1}(iu^2v^{-2}/(1+u^1u^2v^{-2})) & -2iD^{-1} - u^2D^{-1}(iu^1v^{-2}/(1+u^1u^2v^{-2})) \end{pmatrix}.$$

Thus $R[u]$ constitutes a recursion operator for Eq. (25).

Our final example is the derivative Schrödinger equation

$$iu_t = -D^2u + iD(|u|^2u)$$

(see Ref. 11), which can be written as a system $u_t = K[u]$:

$$u_t^1 = iD^2u^1 + D((u^1)^2u^2),$$
$$u_t^2 = -iD^2u^2 + D(u^1(u^2)^2), \tag{29}$$

where $u^2 = \bar{u}^1$. In Ref. 11 the following spectral problem of the Wadati–Konno–Ichikawa type is stated for Eq. (29):

$$D\varphi^1 + i\lambda^2\varphi^1 = u^1\lambda\varphi^2, \quad D\varphi^2 - i\lambda^2\varphi^2 = u^2\lambda\varphi \tag{30a}$$

and

$$i\varphi_t^1 = A\varphi^1 + B\varphi^2, \quad i\varphi_t^2 = C\varphi^1 - A\varphi^2, \tag{30b}$$

where

$$A = 2\lambda^4 + \lambda^2u^1u^2,$$
$$B = 2i\lambda^3u^1 - \lambda u_1^1 + i\lambda(u^1)^2u^2,$$
$$C = 2i\lambda^3u^2 + \lambda u_1^2 + i\lambda u^1(u^2)^2.$$

Performing the algorithm for obtaining a recursion operator for Eq. (29) we obtain the following results. The solutions φ of the spectral problem (30) are transformed into solutions ψ to the linearized equation $\psi_t = A(K[u])\psi$ under the transformation

$$\psi^1 = D(\varphi^1)^2, \quad \psi^2 = -D(\varphi^2)^2.$$

Furthermore, from Eq. (30a) it follows that ψ satisfies the equation

$$R[u]\psi = (1/\lambda^2)\psi,$$

where

$$R[u] = \begin{pmatrix} -2iD^{-1} + 2u^1D^{-1}(u^2D^{-1}) & -2u^1D^{-1}(u^1D^{-1}) \\ -2u^2D^{-1}(u^2D^{-1}) & -2iD^{-1} + 2u^2D^{-1}(u^1D^{-1}) \end{pmatrix}.$$

Therefore, $R[u]$ is a recursion operator for Eq. (29).

V. CONCLUDING REMARKS

The equations $u_t = K[u]$, which we have considered thus far, are completely integrable systems. A common feature of those equations is the integrability by the inverse scattering method as well as the existence of infinitely many commuting symmetries. The latter property is related to the Hamiltonian structure of the equations. Commuting symmetries are connected to involutive conserved densities by the Hamiltonian structure.

Given an isospectral eigenvalue problem

$$L(u,\varphi,\lambda) = 0,$$

there are in general different hierarchies of time evolution equations for the eigenfunctions

$$P(u,\varphi,\lambda) = \varphi_t$$

leading to hierarchies of integrable equations. Within one hierarchy all equations represent symmetries for each other, i.e., they generate commuting flows.

The equations of Secs. II and III are basic equations of their respective hierarchies, i.e., they are of lowest order in the derivatives. If we denote such a hierarchy by $u_t = K_n[u]$, where $K_n[u]$ depends only upon u and derivatives of u, then the hierarchy is obtained recursively as $K_n[u] = R[u]^n u_1$. This fact is remarkable since $R[u]$ involves the inverse operator D^{-1}.

The equations with underlying eigenvalue problem of the Wadati–Konno–Ichikawa type do not share this property. They possess infinitely many symmetries $u_t = K_n[u]$, where $K_n[u]$ depends only upon u and derivatives of u (see Ref. 13). However, these symmetries cannot be constructed by using recursion operators. The recursion operators that we have obtained are leading to extended symmetries for the equations in the respective hierarchy. For instance, consider Dym's equation $u_t = 2D^3(1+u)^{-1/2}$. The Hamiltonian formulation is $u_t = D^3(\delta/\delta u)H$ with the Hamiltonian

$H = 4(1 + u)^{1/2}$. In Ref. 13 infinitely many conserved densities $\rho[u]$ depending upon u and its derivatives are constructed. By Noether's theorem the densities yield symmetries $u_t = D^3(\delta/\delta u)\rho$. The recursion operator $R[u] = 2(1 + u)D^{-2} + u_1 D^{-3}$ now generates extended symmetries for the Dym equation, for example, $R[u]u_1 = 2(1 + u)u_{-1} + u_1 u_{-2}$.

[1]P. J. Olver, J. Math. Phys. **18**, 1212 (1977); M. Wadati, Stud. Appl. Math. **59**, 153 (1978).
[2]O. V. Kaptsov, Sov. Math. Dokl. **25**, 173 (1982).
[3]F. Magri, J. Math. Phys. **18**, 1405 (1977) and J. Math. Phys. **19**, 1156 (1978).
[4]B. Fuchssteiner, Nonlinear Anal. TMA, 849 (1979); A. S. Fokas and B. Fuchssteiner, Lett. Nuovo Cimento **28**, 299 (1980); B. Fuchssteiner and A. S. Fokas, Physica D **4**, 47 (1981).
[5]B. G. Konopelchenko and V. G. Dubrovsky, "General nth order differential spectral problem: general structure of the integrable equations, nonuniqueness of recursion operators and gauge invariance," Institute of Nuclear Physics, Novosibirsk, Preprint 83-26 (to appear in Ann. Phys.); N. H. Ibragimov and A. B. Shabat, Dokl. Akad. Nauk **244**, 57 (1979); B. G. Konopelchenko and V. G. Mokhnachnev, J. Phys. A: Math. Gen. **13**, 3113 (1980).
[6]A. S. Fokas and R. L. Anderson, J. Math. Phys. **23**, 1066 (1982).
[7]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[8]M. Jaulent and I. Miodek, Lett. Math. Phys. **1**, 243 (1976).
[9]M. Boiti, F. Pempinelli, and G. Z. Tu, Prog. Theor. Phys. **69**, 48 (1983).
[10]M. Wadati, K. Konno, and Y-H. Ichikawa, J. Phys. Soc. Jpn. Lett. **46**, 1965 (1979).
[11]D. J. Kaup and A. C. Newell, J. Math. Phys. **19**, 798 (1978).
[12]M. Wadati, Y-H. Ichikawa, and T. Shimizu, Prog. Theor. Phys. **64**, 1959 (1980).
[13]M. Wadati, K. Konno, and Y-H. Ichikawa, J. Phys. Soc. Jpn. **47**, 1698 (1979).

On the Schrödinger operator with periodic point interactions in the three-dimensional case

S. Fassari
Istituto Matematico "G. Castelnuovo," Università degli Studi, Roma, Italy

(Received 18 August 1983; accepted for publication 16 February 1984)

We prove that it is possible to define the self-adjoint operator which gives sense to the merely formal expression $-\Delta - \Sigma_{y\in L}\lambda\delta(\cdot - y)$ (where L is a certain lattice of $\mathbb{R}^3$) as the limit when $\epsilon\to 0_+$ in the resolvent sense of the net $H_\epsilon = -\Delta + \Sigma_{y\in L}\,\lambda(\epsilon)\epsilon^{-2}V(\cdot - y/\epsilon)\,\lambda(\epsilon)$ being a real-valued, $C^\infty[0,1]$ function with $\lambda(0) = 1$ and $V\in L^\infty$ is such that supp V is contained in the Wigner–Seitz cell. By using the direct integral decomposition, we reduce the problem to the convergence of the reduced Hamiltonian $H_\epsilon(\theta) = -\Delta_\theta + \lambda(\epsilon)\epsilon^{-2}V(\cdot/\epsilon)$. In order to find the limit when $\epsilon\to 0_+$ of $[H_\epsilon(\theta) - E]^{-1}$, we also study the properties of its integral kernel.

PACS numbers: 02.30.Tb, 03.65.Nk

I. INTRODUCTION

In this paper we give an alternative proof of a theorem, due to Grossman, Høegh-Khron, and Mebkout,[1] about Schrödinger Hamiltonians with periodic point interactions in the three-dimensional case.

In Ref. 1 the authors define the Hamiltonian "$-\Delta - \Sigma_{y\in L}\lambda\delta(\cdot - y)$" (where L is a certain lattice of $\mathbb{R}^3$) as the strong limit (in the resolvent sense) of the net of Hamiltonians $H_y^\omega = p^2 - \Sigma_{y\in\overline{L}}|\phi_y^\omega\rangle\langle\phi_y^\omega|$ when $\omega\to\infty$ and $|\overline{L}|\to\infty$, $\overline{L}$ being a finite subset of L and $\phi_y^\omega = (2\pi)^{-3/2}[\lambda(\omega)]^{1/2}\chi_\omega(p)e^{iyp}$ (χ_ω is the characteristic function of the ball of radius ω in the momentum space).

They show that if $\lambda(\omega)$ is chosen suitably infinitesimal, such a limit exists. Subsequently they exploit the direct integral decomposition, that is the typical decomposition for Hamiltonians with periodic potentials in order to get the spectral analysis of the limit operator.

In our way of proceeding the direct integral decomposition plays a more important role; we take as approximating net $H_\epsilon = -\Delta + \Sigma_{y\in L}\lambda(\epsilon)V_\epsilon(\cdot - y)(\epsilon\in(0,1])$ where $V_\epsilon = \epsilon^{-2}V(x/\epsilon)$, $V\in\mathbb{L}^\infty$ being such that supp $V\subset C$ (C is the Wigner–Seitz cell), and $\lambda(\epsilon)\in C^\infty[0,1]$ is a real-valued function with $\lambda(0) = 1$.

We exploit the direct integral decomposition in order to reduce the convergence of the net H_ϵ to the convergence of the reduced Hamiltonians $H_\epsilon(\theta) = -\Delta_\theta + \lambda(\epsilon)V_\epsilon$ as operators in the space $\mathbb{L}^2(C;d^3x)$.

The proof of the convergence of the resolvents $[H_\epsilon(\theta) - E]^{-1}$ is not very different from proving the convergence of the resolvents $[-\Delta + \lambda(\epsilon)V_\epsilon - E]^{-1}$ in the space $\mathbb{L}^2(\mathbb{R}^3;d^2x)$; this statement is based on the fact that the integral kernel of the resolvent of the free Hamiltonian with θ-boundary conditions $(-\Delta_\theta)$ may be written as the sum of $G_{\sqrt{E}}(x-y)$ (the free Green function) and a term $\tilde{G}_\theta(x,y;E)$, which is a C^∞ function. Once we have computed the limit of the net $[H_\epsilon(\theta) - E]^{-1}$ and proved that this limit is again a resolvent of a self-adjoint operator $H_\alpha(\theta)$, we use the direct integral decomposition in order to define the self-adjoint operator $\tilde{H}_\alpha$ which has $H_\alpha(\theta)$ as its fibers. Here $\tilde{H}_\alpha$ is an unbounded self-adjoint operator in the space

$$\mathbb{L}^2([0,2\pi)^3, d^3\theta/(2\pi)^3; \mathbb{L}^2(C;d^3x)) = \mathrm{H} = \int_{[0,2\pi)^3}^{\oplus}\mathbb{L}^2(C,d^3x)\frac{d^3\theta}{(2\pi)^3},$$

and it defines, by the unitary equivalence between H and $\mathbb{L}^2(\mathbb{R}^3, d^3x)$, a self-adjoint operator H_α in the space $\mathbb{L}^2(\mathbb{R}^3, d^3x)$, which is just the limit of the net H_ϵ.

The spectral analysis of the operator H_α follows directly from the expression of $[H_\alpha(\theta) - E]^{-1}$.

II. THE INTEGRAL KERNEL OF $(-\Delta_\theta + \lambda V - E)^{-1}$

We begin by recalling that in the three-dimensional case it is not possible to define a self-adjoint operator, which gives sense to the purely formal expression "$-\Delta + \Sigma_{y\in L}\lambda\delta(\cdot - y)$", by the KLMN theorem.

Therefore it is quite natural, in order to solve our problem, to make use of the techniques of convergence for unbounded operators. Let H_ϵ be the net as described in the introduction. By the direct integral decomposition we get

$$UH_\epsilon U^{-1} = \int_{[0,2\pi)^3}^{\oplus} H_\epsilon(\theta)\frac{d^3\theta}{(2\pi)^3}, \tag{2.1}$$

where U is the unitary equivalence between $\mathbb{L}^2(\mathbb{R}^3,d^3x)$ and H defined in Ref. 2, XIII.16 (Theorem XIII.97). Here $H_\epsilon(\theta) = -\Delta_\theta + \lambda(\epsilon)V_\epsilon(x)$, the reduced Hamiltonian, is self-adjoint on $D(-\Delta_\theta)$, by the Kato–Rellich theorem.

Therefore, by (2.1), it is sufficient to study the convergence of the net $H_\epsilon(\theta)$. We first note that the Green function of $-\Delta_\theta$, i.e., the integral kernel of $(-\Delta_\theta - E)^{-1}$, may be written in the following way:

$$G_\theta(x,y;\sqrt{E}) = G_{\sqrt{E}}(x,y) + \tilde{G}_\theta(x,y;\sqrt{E}) \tag{2.2}$$

$$(\forall E\in\mathbb{C}|[0,+\infty) \text{ with } \mathrm{Im}\sqrt{E} > 0, x,y\in C, \text{ and } \theta\in[0,2\pi)^3),$$

where $G_{\sqrt{E}}(x,y) = (4\pi)^{-1}e^{i\sqrt{E}|x-y|}/|x-y|$ is the free Green function and $\tilde{G}_\theta(x,y\sqrt{E})$ is a function such that $\forall f\in C_0^\infty(C)$ the function

$$(\tilde{G}_\theta f)_{\sqrt{E}}(x) = \int_C \tilde{G}_\theta(x,y;\sqrt{E})f(y)d^3y$$

 0022-2488/84/102910-08$02.50

satisfies the equation $(-\Delta - E)\psi = 0$ on C [obviously the θ dependence is defined by $(-\Delta_\theta - E)^{-1} f \in D(-\Delta_\theta)$, $\forall f \in \mathbb{L}^2(C; d^3x)$].

By making use of the "image charges" method, $\tilde{G}_\theta(x,y;\sqrt{E})$ may be expressed as a uniformly convergent series in the region $\mathrm{Im}\sqrt{E} > 0$:

$$\tilde{G}_\theta(x,y;\sqrt{E}) = \sum_{\substack{\mathbf{m}\in\mathbf{Z}^3 \\ \mathbf{m}\neq 0}} \frac{\exp(i\sqrt{E}\,|x - y + \Sigma_{i=1}^3 m_i \mathbf{a}_i|)}{4\pi|x - y + \Sigma_{i=1}^3 m_i \mathbf{a}_i|} e^{-i\mathbf{m}\cdot\theta}, \tag{2.3}$$

where $\{\mathbf{a}_i\}_{i=1}^3$ is a basis for L [obviously $\tilde{G}_\theta(x,y;\sqrt{E})$ is independent of the basis chosen].

So we see that the operators $V^{1/2}(-\Delta_\theta - E)^{-1}V_\parallel^{1/2}$, $(-\Delta_\theta - E)^{-1}V_\parallel^{1/2}$, and $V^{1/2}(-\Delta_\theta - E)^{-1}$ (here we use Simon's notation in Ref. 3) belong to the J_2 class. In fact we have $\forall E \in \mathbb{C}\setminus[0,\infty)$ with $\mathrm{Im}\sqrt{E} > 0$ and $\forall\theta \in [0,2\pi)^3$:

$$\begin{aligned}
&\|V^{1/2}(-\Delta_\theta - E)^{-1}V_\parallel^{1/2}\|_{J_2}^2 \\
&\quad = \int_C\int_C |V(x)|\,|G_\theta(x,y;\sqrt{E})|^2 |V(y)| d^3x\, d^3y \\
&\quad \leqslant 2\int_C\int_C |V(x)|(|G_{\sqrt{E}}(x,y)|^2 \\
&\qquad + |\tilde{G}_\theta(x,y;\sqrt{E})|^2)|V(y)| d^3x\, d^3y,
\end{aligned} \tag{2.4}$$

and since V is obviously a Röllnik function, (2.4) is bounded by

$$2\left(\frac{\|V\|_R^2}{(4\pi)^2} + \int_C\int_C |V(x)|\,|\tilde{G}_\theta(x,y;\sqrt{E})|^2 |V(y)| d^3x\, d^3y\right).$$

Furthermore, one easily verifies

$$\begin{aligned}
\|\tilde{G}_\theta(\sqrt{E})\|_\infty &\leqslant \sum_{\substack{m\in L \\ m\neq 0}} \frac{\exp(-\mathrm{Im}\sqrt{E}\,|m| - \mathrm{diam}\, C|)}{4\pi|\,|m| - \mathrm{diam}\, C|} \\
&\equiv F(\sqrt{E}) < \infty \quad (\mathrm{Im}\sqrt{E} > 0),
\end{aligned}$$

so that

$$\begin{aligned}
&\|V^{1/2}(-\Delta_\theta - E)^{-1}V_\parallel^{1/2}\|_{J_2}^2 \\
&\quad \leqslant 2\{\|V\|_R^2/(4\pi)^2 + [F(\sqrt{E})]^2\|V\|_1^2\},
\end{aligned} \tag{2.5}$$

for $\mathrm{Im}\, E^{1/2} > 0$.

In the same way we can evaluate the J_2 norm of $(-\Delta_\theta - E)^{-1}V_\parallel^{1/2}$ and $V^{1/2}(-\Delta_\theta - E)^{-1}$:

$$\begin{aligned}
&\|(-\Delta_\theta - E)^{-1}V_\parallel^{1/2}\|_{J_2}^2 \\
&\quad < 2\left(\int_C\int_C |G_{\sqrt{E}}(x,y)|^2 |V(y)| d^3x\, d^3y\right. \\
&\qquad \left. + \int_C\int_C |\tilde{G}_\theta(x,y;\sqrt{E})|^2 |V(y)| d^3x\, d^3y\right) \\
&\quad \leqslant 2(\||\,|G_{\sqrt{E}}|^2 * |V|\,\|_1 \\
&\qquad + \|V\|_1(\mathrm{vol}\, C)[F(\sqrt{E})]^2 \\
&\quad \leqslant 2\|V\|_1\{\|G_{\sqrt{E}}\|_2^2 + [F(\sqrt{E})]^2(\mathrm{vol}\, C)\},
\end{aligned} \tag{2.6}$$

where we have exploited Young's inequality in order to estimate the $\mathbb{L}^1$ norm of the convolution (see Ref. 4, IX.4).

Furthermore, we get by the dominated convergence theorem

$$\begin{aligned}
&\|V^{1/2}(-\Delta_\theta - E)^{-1}V_\parallel^{1/2}\|_{J_2}^2 \\
&\quad \leqslant 2\{\|V^{1/2}G_{\sqrt{E}}V_\parallel^{1/2}\|_{J_2}^2 \\
&\qquad + [F(\sqrt{E})]^2\|V\|_1^2\} \underset{\mathrm{Im}\sqrt{E}\to\infty}{\to} 0,
\end{aligned} \tag{2.7}$$

$\forall\theta \in [0,2\pi)^3$, $\mathrm{Im}\sqrt{E} > 0$.

The last property guarantees that one can find $K > 0$ so that the Born series is uniformly convergent for $\mathrm{Im}\sqrt{E} > K$; we can then write

$$\begin{aligned}
&(-\Delta_\theta + \lambda V - E)^{-1} \\
&= (-\Delta_\theta - E)^{-1} - \lambda(-\Delta_\theta - E)^{-1}V_\parallel^{1/2} \\
&\times\left(\sum_{l=0}^\infty [-\lambda V^{1/2}(-\Delta_\theta - E)^{-1}V_\parallel^{1/2}]^l\right)V^{1/2}(-\Delta_\theta - E)^{-1} \\
&= (-\Delta_\theta - E)^{-1} - \lambda(-\Delta_\theta - E)^{-1}V_\parallel^{1/2} \\
&\times[1 + \lambda V^{1/2}(-\Delta_\theta - E)^{-1}V_\parallel^{1/2}]^{-1}V^{1/2}(-\Delta_\theta - E)^{-1}.
\end{aligned} \tag{2.8}$$

We see also, setting $\sqrt{E} = \omega$, that the map $\omega \to V^{1/2}(-\Delta_\theta - \omega^2)^{-1}V_\parallel^{1/2}$ from the half-plane $\mathrm{Im}\,\omega > 0$ into the $J_2[\mathbb{L}^2(C,d^3x)]$-space is analytic in the J_2 norm. Then, by the analytic Fredholm theory (see Ref. 5, VI.5 and Ref. 2, XIII.4), we obtain that $[1 + \lambda V^{1/2}(-\Delta_\theta - \omega^2)^{-1}V_\parallel^{1/2}]^{-1}$ is meromorphic in $D \equiv \{\omega \in \mathbb{C} | \mathrm{Im}\,\omega > 0\}$, analytic in $D\setminus S_\theta$, where S_θ is a discrete subset of D (S_θ is contained in the axis $\mathrm{Re}\,\omega = 0$, $-\Delta_\theta + \lambda V$ being self-adjoint); the residues at the poles are finite rank operators and if $\omega \in S_\theta$ it follows that $\exists\psi(\theta) \in \mathbb{L}^2(C;d^3x)$ such that

$$\lambda V^{1/2}(-\Delta_\theta - \omega^2)^{-1}V_\parallel^{1/2}\psi(\theta) = -\psi(\theta). \tag{2.9}$$

By the map $\omega \to \omega^2$, we get a 1–1 correspondence between the poles of $[1 + \lambda V^{1/2}(-\Delta_\theta - \omega^2)^{-1}V_\parallel^{1/2}]^{-1}$ and the negative eigenvalues of $-\Delta_\theta + \lambda V$ and the correspondence preserves the multiplicity. In fact, if $\phi(\theta) = (-\Delta_\theta - \omega^2)^{-1}V_\parallel^{1/2}\psi(\theta)$, we get

$$(-\Delta_\theta + \lambda V)\phi(\theta) = \omega^2\phi(\theta).$$

[Note that $\forall\psi(\theta) \in \mathbb{L}^2(C,d^3x), V_\parallel^{1/2}\psi(\theta) \in \mathbb{L}^2(C,d^3x)$, V being a bounded function with supp $V \subset C$; so $\phi(\theta) \in D(-\Delta_\theta) \equiv D(-\Delta_\theta + \lambda V)$, $-\Delta_\theta + \lambda V$ being defined by Rellich's theorem.]

III. THE LIMIT OF $[H_\epsilon(\theta) - E]^{-1}$

After having summarized the main properties of the Hamiltonian $-\Delta_\theta + \lambda V$ and of its resolvent, we pass to study the convergence of the net $H_\epsilon(\theta)$.

From the second resolvent equation we have, for $\mathrm{Im}\,\omega > 0$ and $\mathrm{Re}\,\omega \neq 0$,

$$\begin{aligned}
[H_\epsilon(\theta) - \omega^2]^{-1} &= (-\Delta_\theta - \omega^2)^{-1} - \lambda(\epsilon)(-\Delta_\theta - \omega^2)^{-1}V_\epsilon(-\Delta_\theta - \omega^2)^{-1} \\
&\quad + \lambda^2(\epsilon)(-\Delta_\theta - \omega^2)^{-1}V_\epsilon(-\Delta_\theta - \omega^2)^{-1}V_\epsilon(-\Delta_\theta - \omega^2)^{-1} - \lambda^3(\epsilon)(-\Delta_\theta - \omega^2)^{-1}V_\epsilon(-\Delta_\theta - \omega^2)^{-1} \\
&\quad \times V_{\epsilon\parallel}^{1/2}[1 + \lambda(\epsilon)V_\epsilon^{1/2}(-\Delta_\theta - \omega^2)^{-1}V_{\epsilon\parallel}^{1/2}]^{-1}V_\epsilon^{1/2}(-\Delta_\theta - \omega^2)^{-1}V_\epsilon(-\Delta_\theta - \omega^2)^{-1}
\end{aligned} \tag{3.1}$$

(we have used the expression of $[H_\epsilon(\theta)-\omega^2]^{-1}$ given by (2.8) in writing the third term and the fourth one).

The second term in the rhs of (3.1) converges to zero in the J_1 norm uniformly in θ. Indeed,

$$\|(-\Delta_\theta-\omega^2)^{-1}V_\epsilon(-\Delta_\theta-\omega^2)^{-1}\|_{J_1}\leqslant\|(-\Delta_\theta-\omega^2)^{-1}V_{\epsilon\|}^{1/2}\|_{J_2}^2$$

$$\leqslant 2\|V_\epsilon\|_1\{\|G_\omega\|_2^2+[F(\omega)]^2(\mathrm{vol}\,C)\}=2\epsilon\|V\|_1\{\|G_\omega\|_2^2+[F(\omega)]^2(\mathrm{vol}\,C)\}\underset{\epsilon\to 0_+}{\to}0, \tag{3.2}$$

where we have exploited the estimate (2.6).

Also the third term converges to zero in the J_1 norm uniformly in θ:

$$\|[(-\Delta_\theta-\omega^2)^{-1}V_{\epsilon\|}^{1/2}][V_\epsilon^{1/2}(-\Delta_\theta-\omega^2)^{-1}V_{\epsilon\|}^{1/2}][V_\epsilon^{1/2}(-\Delta_\theta-\omega^2)^{-1}]\|_{J_1}$$

$$\leqslant\|(-\Delta_\theta-\omega^2)^{-1}V_{\epsilon\|}^{1/2}\|_{J_2}^2\|V_\epsilon^{1/2}(-\Delta_\theta-\omega^2)^{-1}V_{\epsilon\|}^{1/2}\|_{J_2}$$

$$\leqslant 4\epsilon\|V\|_1\{\|G_\omega\|_2^2+[F(\omega)]^2(\mathrm{vol}\,C)\}\{\|V\|_R^2/(4\pi)^2+\epsilon^2[F(\omega)]^2\|V\|_1^2\}^{1/2}\underset{\epsilon\to 0_+}{\to}0, \tag{3.3}$$

where we have used the estimates (2.5) and (2.6) and the fact that the series on the rhs of (2.3) converges uniformly in $x,y\in C$ and $\theta\in[0,2\pi)^3$.

Now we turn to study the fourth term.

By (2.7) we see that $\forall\epsilon\in(0,1]\exists\omega(\epsilon)>0$ such that

$$\|V_\epsilon^{1/2}(-\Delta_\theta-\omega^2)^{-1}V_{\epsilon\|}^{1/2}\|_{J_2}^2\leqslant 2\{\|V^{1/2}G_{\epsilon\omega}V_\|^{1/2}\|_{J_2}^2+\epsilon^2[F(\omega)]^2\|V\|_1^2\}<1,$$

for all ω with $\mathrm{Im}\,\omega>\omega(\epsilon)$. So in the region $\mathrm{Im}\,\omega>\omega(\epsilon)$, the fourth term is represented by the operator having as its integral kernel

$$\sum_{l=3}^{\infty}[-\lambda(\epsilon)]^l\int_C\cdots\int_C G_\theta(x,x_1;\omega)V_\epsilon(x_1)G_\theta(x_1,x_2;\omega)\cdots V_\epsilon(x_l)\,\overline{G_\theta(x_l,y;\omega)}\,d^3x_1\cdots d^3x_l$$

$$=\sum_{l=3}^{\infty}\epsilon^l[-\lambda(\epsilon)]^l\int_C\cdots\int_C G_\theta(x,\epsilon x_1;\omega)V(x_1)G_\theta(\epsilon x_1,\epsilon x_2;\omega)\cdots V(x_l)G_\theta(\epsilon x_l,y;\omega)d^3x_1\cdots d^3x_l$$

$$=\sum_{l=3}^{\infty}[-\epsilon\lambda(\epsilon)]^l\int_C\cdots\int_C\frac{1}{\epsilon^{l-1}}G_\theta(x,\epsilon x_1;\omega)V(x_1)[G_{\epsilon\omega}(x_1,x_2)+\epsilon\tilde{G}_\theta(\epsilon x_1,\epsilon x_2;\omega)]\cdots$$

$$\times V(x_{l-1})[G_{\epsilon\omega}(x_{l-1},x_l)+\epsilon\tilde{G}_\theta(\epsilon x_{l-1},\epsilon x_l;\omega)]V(x_l)\,\overline{G_\theta(\epsilon x_l,y;\omega)}\,d^3x_1\cdots d^3x_l$$

$$=\epsilon[-\lambda(\epsilon)]^3[G_\theta(\cdot,\epsilon\cdot;\omega)V_\|^{1/2}]\{V^{1/2}[G_{\epsilon\omega}+\epsilon\tilde{G}_\theta(\epsilon\cdot,\epsilon\cdot;\omega)]V_\|^{1/2}\}$$

$$\times\Big(\sum_{l=0}^{\infty}[-\lambda(\epsilon)]^l\{V^{1/2}[G_{\epsilon\omega}+\epsilon G_\theta(\epsilon\cdot,\epsilon\cdot;\omega)]V_\|^{1/2}\}^l\Big)\{V^{1/2}[G_{\epsilon\omega}+\epsilon\tilde{G}_\theta(\epsilon\cdot,\epsilon\cdot;\omega)]V_\|^{1/2}\}[V^{1/2}\,\overline{G_\theta(\epsilon\cdot,\cdot;\omega)}], \tag{3.4}$$

where we have represented the integral operators with their kernels. Let $A_\epsilon^\theta(\omega)$, $B_\epsilon^\theta(\omega)$, $C_\epsilon^\theta(\omega)$ be the J_2 $[\mathbb{L}^2(C,d^3x)]$ operators with integral kernels

$$A_\epsilon^\theta(x,y;\omega)=G_\theta(x,\epsilon y;\omega)V_\|^{1/2}(y),$$

$$B_\epsilon^\theta(x,y;\omega)=V^{1/2}(x)[G_{\epsilon\omega}(x,y)+\epsilon\tilde{G}_\theta(\epsilon x,\epsilon y;\omega)]V_\|^{1/2}(y),$$

$$C_\epsilon^\theta(x,y;\omega)=V^{1/2}(x)\,\overline{G_\theta(\epsilon x,y;\omega)}.$$

We begin by studying the limit of the net $B_\epsilon^\theta(\omega)$. For $\forall\omega\in\mathbb{C}$ with $\mathrm{Im}\,\omega>0$ we have

$$\|B_\epsilon^\theta(\omega)\|_{J_2}\leqslant\|V^{1/2}G_0V_\|^{1/2}\|_{J_2}+[F(\omega)]\|V\|_1. \tag{3.5}$$

So $B_\epsilon^\theta(\omega)$ is uniformly bounded and by dominated convergence we get that $B_\epsilon^\theta(\omega)\underset{\epsilon\to 0_+}{\overset{J_2}{\to}}V^{1/2}G_0V_\|^{1/2}$ uniformly in θ.

Furthermore for each fixed ϵ, the map $\omega\to B_\epsilon^\theta(\omega)$ from $D\equiv\{\omega\in\mathbb{C}|\ \mathrm{Im}\,\omega>0\}$ into the space $J_2[\mathbb{L}^2(C,d^3x)]$ is analytic with respect to the J_2 norm; hence by the analytic Fredholm theory $[1+\lambda(\epsilon)B_\epsilon^\theta(\omega)]^{-1}$ is meromorphic in D, analytic in $D\setminus S_\epsilon(\theta)$, where $S_\epsilon(\theta)$ is a discrete subset of D and the residues at the poles are finite rank operators.

In this way we can define the operator on the right-hand side of (2.4) in all the half-plane $\mathrm{Im}\,\omega>0$ except $S_\epsilon(\theta)$ and we write it as

$$-\epsilon\lambda^3(\epsilon)A_\epsilon^\theta(\omega)B_\epsilon^\theta(\omega)[1+\lambda(\epsilon)B_\epsilon^\theta(\omega)]^{-1}B_\epsilon^\theta(\omega)C_\epsilon^\theta(\omega).$$

We have already seen that $B_\epsilon^\theta(\omega)\underset{\epsilon\to 0_+}{\overset{J_2}{\to}}V^{1/2}G_0V_\|^{1/2}$; we now turn to find the limit of $C_\epsilon^\theta(\omega)$.

We first note that $C_\epsilon^\theta(\omega)$ is $(\|\ \|_{J_2})$-bounded uniformly in ϵ and θ, since we have

$$\|C_\epsilon^\theta(\omega)\|_{J_2}^2=\int_C\int_C|V(x)|\,|G_\theta(\epsilon x,y;\omega)|^2d^3x\,d^3y$$

$$\leqslant 2\|V\|_1\{\|G_\omega\|_2^2+[F(\omega)]^2(\mathrm{vol}\,C)\}. \tag{3.6}$$

We also note that $\forall f\in\mathbb{L}^2(C,d^3x)$, $[\,\overline{G_\theta(\omega)}*f\,]\in D(-\Delta_\theta)\subset C^0(C)$. Hence by dominated convergence we have

$$\|V^{1/2}[\,\overline{G_\theta(\omega)}*f\,](\epsilon\cdot)-V^{1/2}[\,\overline{G_\theta(\omega)}*f\,](0)\|_2^2$$

$$\leqslant\int_C|V(x)|\|([\,\overline{G_\theta(\omega)}*f\,](\epsilon x)-[\,\overline{G_\theta(\omega)}*f\,](0)\|_\infty^2\,d^3x\underset{\epsilon\to 0_+}{\to}0, \tag{3.7}$$

$\forall f \in \mathbb{L}^2(C,d^3x)$, $\forall \theta \in [0,2\pi)^3$, and $\mathrm{Im}\,\omega > 0$.

Hence we have shown that $C^\theta_\epsilon(\omega) \underset{\epsilon\to 0_+}{\overset{s}{\to}} V^{1/2}(G_\theta(\omega),\cdot)$ and with the same arguments one proves that $[A^\theta_\epsilon(\omega)]^* \underset{\epsilon\to 0_+}{\overset{s}{\to}} V_\parallel^{1/2}(G_\theta(\omega),\cdot)$, which implies that $A^\theta_\epsilon(\omega) \underset{\epsilon\to 0_+}{\overset{w}{\to}} G_\theta(\omega)(V_\parallel^{1/2},\cdot)$.

We now prove that all these convergences are actually in the J_2 norm.

We first note that $V^{1/2}(G_\theta(\omega),\cdot) = C^\theta_0(\omega) = s-\lim_{\epsilon\to 0_+} C^\theta_\epsilon(\omega)$ belongs to the J_2 class, since $C^\theta_0(\omega)$ is a finite rank operator.

Furthermore, by (3.6) we have

$$\|C^\theta_\epsilon(\omega)\|^2_{J_2} \le \|V\|_1 \|G_\theta(\omega)\|^2_2 = \|C^\theta_0(\omega)\|^2_{J_2}, \tag{3.8}$$

and in the same way

$$\|A^\theta_\epsilon \omega\|^2_{J_2} \le \|A^\theta_0(\omega)\|^2_{J_2}. \tag{3.9}$$

Hence, by dominated convergence, we get

$$\begin{aligned}\|C^\theta_\epsilon(\omega)\|^2_{J_2} &= \int_C\int_C |G_\theta(\epsilon x,y;\omega)|^2 \\ &\times |V(x)|d^3x\,d^3y \underset{\epsilon\to 0_+}{\to} \int_C\int_C |G_\theta(0,y;\omega)|^2 \\ &\times |V(x)|d^3x\,d^3y = \|C^\theta_0(\omega)\|^2_{J_2},\end{aligned} \tag{3.10}$$

and similarly $\|A^\theta_\epsilon(\omega)\|^2_{J_2} \underset{\epsilon\to 0_+}{\to} \|A^\theta_0(\omega)\|^2_{J_2}$.

Let $(\cdot,\cdot)_{J_2}$ be the inner product of $J_2[\mathbb{L}^2(C,d^2x)]$ regarded as a Hilbert space. For each fixed $B \in J_2$ we have

$$\begin{aligned}|(B,A^\theta_\epsilon(\omega))_{J_2}| &\le \sum_{l=1}^\infty |(\phi_l, B^* A^\theta_\epsilon(\omega)\phi_l)| \\ &\le \Big(\sum_{l=1}^\infty \|B\phi_l\|_2^2\Big)^{1/2}\Big(\sum_{l=1}^\infty \|A^\theta_\epsilon(\omega)\phi_l\|_2^2\Big)^{1/2} \\ &= \|B\|_{J_2}\|A^\theta_\epsilon(\omega)\|_{J_2} \le \|B\|_{J_2}\|A^\theta_0(\omega)\|_{J_2},\end{aligned} \tag{3.11}$$

for any orthonormal bases $\{\phi_l\}_{l=1}^\infty$ of $\mathbb{L}^2(C,d^3x)$. Since $A^\theta_\epsilon(\omega) \underset{\epsilon\to 0_+}{\overset{w}{\to}} A^\theta_0(\omega) \in J_2$, we have that $(\phi_l, B^* A^\theta_\epsilon(\omega)\phi_l) \underset{\epsilon\to 0_+}{\to} (\phi_l, B^* A^\theta_0(\omega)\phi_l)$ for each fixed $B \in J_2$, i.e., the net of successions

$$\{f^\theta_\epsilon(l)\}_{l=1}^\infty = \{(\phi_l, B^* A^\theta_\epsilon(\omega)\phi_l)\}_{l=1}^\infty \in l_1(\mathbb{N}),\ \forall \epsilon \in (0,1]$$

converges pointwisely to $\{f^\theta_0(l)\}_{l=1}^\infty = \{(\phi_l, B^* A^\theta_0(\omega)\phi_l)\}_{l=1}^\infty \in l_1(\mathbb{N})$.

Hence, by dominated convergence, we get, for each fixed $B \in J_2$,

$$\begin{aligned}|(B,A^\theta_\epsilon(\omega))_{J_2} - (B,A^\theta_0(\omega))_{J_2}| &\le \sum_{l=1}^\infty |(\phi_l, B^* A^\theta_\epsilon(\omega)\phi_l) - (\phi_l, B^* A^\theta_0(\omega)\phi_l)| \underset{\epsilon\to 0_+}{\to} 0,\end{aligned}$$

for $\mathrm{Im}\,\omega > 0$, $\forall \theta \in [0,2\pi)^3$.

By taking $B = A^\theta_0(\omega)$, we get $(A^\theta_0(\omega), A^\theta_\epsilon(\omega))_{J_2} \underset{\epsilon\to 0_+}{\to} \|A^\theta_0(\omega)\|^2_{J_2}$, which implies

$$\begin{aligned}\|A^\theta_0(\omega) - A^\theta_\epsilon(\omega)\|^2_{J_2} &= \|A^\theta_0(\omega)\|^2_{J_2} + \|A^\theta_\epsilon(\omega)\|^2_{J_2} - (A^\theta_0(\omega), A^\theta_\epsilon(\omega))_{J_2} \\ &\quad - (A^\theta_\epsilon(\omega), A^\theta_0(\omega))_{J_2} \underset{\epsilon\to 0_+}{\to} 0,\end{aligned} \tag{3.12}$$

exploiting the fact that $\|A^\theta_\epsilon(\omega)\|^2_{J_2} \underset{\epsilon\to 0_+}{\to} \|A^\theta_0(\omega)\|^2_{J_2}$.

Hence we have shown that

$$A^\theta_\epsilon(\omega) \underset{\epsilon\to 0_+}{\overset{\|\ \|_{J_2}}{\to}} A^\theta_0(\omega),$$

and in the same way

$$C^\theta_\epsilon(\omega) \underset{\epsilon\to 0_+}{\overset{\|\ \|_{J_2}}{\to}} C^\theta_0(\omega),$$

for $\mathrm{Im}\,\omega > 0$ and $\forall \theta \in [0,2\pi)^3$.

It remains to determine the limit of the net $\epsilon[1 + \lambda(\epsilon)B^\theta_\epsilon(\omega)]^{-1}$. Following the steps in the proof of Lemma 3.1 in Ref. 6 adapted to our space $\mathbb{L}^2(C,d^3x)$, one obtains the following norm convergent Laurent expansion around $\epsilon = 0$:

$$(V^{1/2}G_0V_\parallel^{1/2} + 1 + \epsilon)^{-1} = \epsilon^{-1}E_{(-1)} + \sum_{m=0}^\infty (-\epsilon)^m T^{m+1}, \tag{3.13}$$

where $E_{(-1)}$ is the projector onto the N-dimensional eigenspace of $V^{1/2}G_0V_\parallel^{1/2}$ to the eigenvalue -1,

$$\begin{aligned}E_{(-1)} &= -(2\pi i)^{-1}\oint_{\Gamma(-1)} dz(V^{1/2}G_0V_\parallel^{1/2} - z)^{-1} \\ &= \sum_{j=1}^N \frac{(\tilde\phi_j,\cdot)}{(\tilde\phi_j,\phi_j)}\phi_j,\end{aligned} \tag{3.14}$$

where $N = \dim[E_{(-1)}\mathbb{L}^2(C,d^3x)]$ and $\Gamma_{(-1)}$ surrounds in the usual way only the isolated eigenvalue -1 of $V^{1/2}GV_\parallel^{1/2}$. The $\{\phi_j\}_{j=1}^N$ are such that

$$V^{1/2}G_0V_\parallel^{1/2}\phi_j = -\phi_j,$$
$$\tilde\phi_j = (\mathrm{sgn}\,V)\phi_j,\ (\tilde\phi_i,\phi_j) = 0,\ \forall i \ne j.$$

Here, T is a bounded operator given by

$$\begin{aligned}T &= (2\pi i)^{-1}\oint_{\Gamma(-1)} dz(z+1)^{-1}(V^{1/2}G_0V_\parallel^{1/2} - z)^{-1} \\ &= \|\ \| - \lim_{\epsilon\to 0_+}(V^{1/2}G_0V_\parallel^{1/2} + 1 + \epsilon)^{-1}(1 - E_{(-1)}).\end{aligned} \tag{3.15}$$

(We recall that these operators are in $L[\mathbb{L}^2(C,d^3x)]$, while in Ref. 6 they were in $L[\mathbb{L}^2(\mathbb{R}^3,d^3x)]$.)

By the $\|\ \|_{J_2}$-analyticity of $\lambda(\epsilon)B^\theta_\epsilon(\omega)$ for $\mathrm{Im}\,\omega > 0$, we obtain the following expansion for any $\theta \in [0,2\pi)^3$:

$$\begin{aligned}&\lambda(\epsilon)\{V^{1/2}[G_{\epsilon\omega} + \epsilon\tilde G_\theta(\epsilon\cdot,\epsilon\cdot;\omega)]V_\parallel^{1/2}\} \\ &\quad = V^{1/2}G_0V_\parallel^{1/2} + \epsilon\{\lambda'(0_+)V^{1/2}G_0V_\parallel^{1/2} + [i\omega/4\pi \\ &\qquad + \tilde G_\theta(0;\omega)]|V^{1/2}\rangle\langle V_\parallel^{1/2}|\} + O(\epsilon^2).\end{aligned} \tag{3.16}$$

In the case that -1 is not an eigenvalue of $V^{1/2}G_0V_\parallel^{1/2}$ (for example, when $V \ge 0$) we have $E_{(-1)} = 0$, which implies that the $\|\ \|$ limit of $\epsilon[1 + \lambda(\epsilon)B^\theta_\epsilon(\omega)]^{-1}$ is zero.

If -1 is a simple eigenvalue of $V^{1/2}G_0V_\parallel^{1/2}$, which implies that $E_{(-1)} = \phi(\tilde\phi,\cdot)/(\tilde\phi,\phi)$, and if $(V_\parallel^{1/2},\phi) \ne 0$, by repeating the proof of the Theorem 3.2. of Ref. 6 adapted to

our space $L\,[\mathbb{L}^2(C,d^3x)]$, one proves

$$\| \ \| - \lim_{\epsilon\to 0_+}\epsilon[1+\lambda(\epsilon)B^\theta_\epsilon(\omega)]^{-1} = \{[i\omega/4\pi+\tilde G_\theta(0;\omega)]|(V^{1/2}_\parallel,\phi)|^2 - \lambda'(0_+)(\tilde\phi,\phi)\}^{-1}(\tilde\phi,\cdot)\phi, \tag{3.17}$$

$\forall\theta\in[0,2\pi)^3$ and for $\operatorname{Im}\omega>0$.

Now let -1 be an eigenvalue of multiplicity N (obviously $N<\infty$, since $V^{1/2}G_0V^{1/2}_\parallel$ is Hilbert–Schmidt).

As in Ref. 6 we have the following cases: (i) $(V^{1/2}_\parallel,\phi_j)=0$, $\forall j=1,\dots N\,(1\leqslant N<\infty)\Leftrightarrow V^{1/2}_\parallel\in\{E_{(-1)}\times\mathbb{L}^2(C,d^3x)\}^\perp$; and (ii) $(V^{1/2}_\parallel,\phi_{j_0})\neq 0$ for some j_0. Both the cases can in turn be divided into (a) $\lambda'(0_+)\neq 0$, and (b) $\lambda'(0_+)=0$.

The simplest one is the case (ii) (a). Here we have nothing else to do but to repeat the steps for finding (3.17) with the matrical notation due to the fact that $\{E_{(-1)}\mathbb{L}^2(C,d^3x)\}$ is N-dimensional. One obtains

$$\| \ \| - \lim_{\epsilon\to 0_+}\epsilon[1+\lambda(\epsilon)B^\theta_\epsilon(\omega)]^{-1} = \sum_{j,l=1}^{N}\left(\left[\frac{i\omega}{4\pi}+\tilde G_\theta(0;\omega)\right](\tilde\phi_j,V^{1/2})(V^{1/2}_\parallel,\phi_l) - \lambda'(0_+)(\tilde\phi_j,\phi_l)\right)^{-1}(\tilde\phi_l,\cdot)\phi_j, \tag{3.18}$$

where $[\ ]^{-1}$ denotes the inverse of the $N\times N$ matrix $([i\omega/4\pi+\tilde G_\theta(0;\omega)](\tilde\phi_j,V^{1/2})(V^{1/2}_\parallel,\phi_l)-\lambda'(0_+)(\tilde\phi_j,\phi_l))$. [We see that in the case $\lambda'(0_+)=0$, the matrix would not be invertible, having the jth line and the jth column equal to zero if $(V^{1/2}_\parallel,\phi_j)=0$.] By (3.18), we obtain in the case (i)(a),

$$\| \ \| - \lim_{\epsilon\to 0_+}\epsilon[1+\lambda(\epsilon)B^\theta_\epsilon(\omega)]^{-1} = \sum_{j,l=1}^{N}[-\lambda'(0_+)(\tilde\phi_j,\phi_l)]^{-1}(\tilde\phi_l,\cdot)\phi_j = -[\lambda'(0_+)]^{-1}\sum_{j=1}^{N}\frac{(\tilde\phi_j,\cdot)}{(\tilde\phi_j,\phi_j)}\phi_j. \tag{3.19}$$

In the other two cases (i) (b) and (ii) (b), the $L\,[\mathbb{L}^2(C,d^3x)]$-valued function $\epsilon[1+\lambda(\epsilon)B^\theta_\epsilon(\omega)]^{-1}$ has norm convergent Laurent expansions around $\epsilon=0$, which are quite similar to the correspondent ones obtained in Ref. 6 [(3.43) and (3.59), respectively]. In fact we have to transfer all the expressions in our space $L\,[\mathbb{L}^2(C,d^3x)]$ and to replace the coefficients of the Taylor series of $\lambda(\epsilon)V^{1/2}G_{\epsilon\omega}V^{1/2}_\parallel$ with those ones of $\lambda(\epsilon)[V^{1/2}G_{\epsilon\omega}V^{1/2}_\parallel+\epsilon V^{1/2}G_\theta(\epsilon\cdot,\epsilon\cdot;\omega)V^{1/2}_\parallel]=\lambda(\epsilon)B^\theta_\epsilon(\omega)$, i.e., we have to add the k th coefficient of the expansion of $\epsilon\lambda(\epsilon)V^{1/2}\tilde G_\theta(\epsilon\cdot,\epsilon\cdot;\omega)V^{1/2}_\parallel$ to the k th coefficient of the expansion of $\lambda(\epsilon)V^{1/2}G_{\epsilon\omega}V^{1/2}_\parallel$. (We have already written the first-order coefficient $[\lambda(\epsilon)B^\theta_\epsilon(\omega)]_1=\{\lambda'(0_+)V^{1/2}G_0V^{1/2}_\parallel+[i\omega/4\pi+\tilde G_\theta(0;\omega)]|V^{1/2}\rangle\langle V^{1/2}_\parallel|\}$; the second-order one is, for example,

$$[\lambda(\epsilon)B^\theta_\epsilon(\omega)]_2 = \tfrac12\{\lambda''(0_+)V^{1/2}G_0V^{1/2}_\parallel+2[i\omega/4\pi+\tilde G_\theta(0;\omega)]\times\lambda'(0_+)|V^{1/2}\rangle\langle V^{1/2}_\parallel|+2[\omega^2\hat C+\hat C_\theta(\omega)]\},$$

where $\hat C$ and $\hat C_\theta(\omega)$ are the J_2-$[\mathbb{L}^2(C,d^3x)]$ operators, with kernels given by

$$\hat C(x,y) = -(8\pi)^{-1}V^{1/2}(x)|x-y|V^{1/2}_\parallel(y),$$

$$\hat C_\theta(x,y;\omega) = V^{1/2}(x)\left(\frac{d}{d\epsilon}\tilde G_\theta[\epsilon(x-y);\omega]\right)_{\epsilon=0}\times V^{1/2}_\parallel(y).)$$

At this point we can compute the limit of the net

$$-\epsilon\lambda^3(\epsilon)A^\theta_\epsilon(\omega)B^\theta_\epsilon(\omega)[1+\lambda(\epsilon)B^\theta_\epsilon(\omega)]^{-1}B^\theta_\epsilon(\omega)C^\theta_\epsilon(\omega),$$

which, in its turn, will provide us with the limit of $[H_\epsilon(\theta)-\omega^2]^{-1}$. When -1 is not an eigenvalue of $V^{1/2}G_0V^{1/2}_\parallel$, we obtain that

$$\| \ \| - \lim_{\epsilon\to 0_+} -\epsilon\lambda^3(\epsilon)A^\theta_\epsilon(\omega)B^\theta_\epsilon(\omega)\times[1+\lambda(\epsilon)B^\theta_\epsilon(\omega)]^{-1}B^\theta_\epsilon(\omega)C^\theta_\epsilon(\omega)=0, \tag{3.20}$$

$\forall\omega\in\mathbb{C}$ with $\operatorname{Re}\omega\neq 0$ and $\operatorname{Im}\omega>0$, $\forall\theta\in[0,2\pi)^3$. Then using (3.1)–(3.3) and (3.20), we get

$$\| \ \| - \lim_{\epsilon\to 0_+}[H_\epsilon(\theta)-\omega^2]^{-1} = (-\Delta_\theta-\omega^2)^{-1}, \tag{3.21}$$

$\forall\omega\in\mathbb{C}$ with $\operatorname{Re}\omega\neq 0$, $\operatorname{Im}\omega>0$, and $\forall\theta\in[0,2\pi)^3$. By setting $E=\omega^2$ ($\operatorname{Im}E^{1/2}>0$), this is equivalent to saying

$$\| \ \| - \lim_{\epsilon\to 0_+}[H_\epsilon(\theta)-E]^{-1} = (-\Delta_\theta-E)^{-1}, \tag{3.22}$$

$\forall E\in\mathbb{C}\setminus\mathbb{R}$ and $\forall\theta\in[0,2\pi)^3$, i.e., $H_\epsilon(\theta)$ converges in the norm resolvent sense to $-\Delta_\theta$ for any $\theta\in[0,2\pi)^3$.

We get the same result in the case (i) (a). In fact

$$\begin{aligned}\| \ \| - \lim_{\epsilon\to 0_+} &-\epsilon\lambda^3(\epsilon)A^\theta_\epsilon(\omega)B^\theta_\epsilon(\omega)\times[1+\lambda(\epsilon)B^\theta_\epsilon(\omega)]^{-1}B^\theta_\epsilon(\omega)C^\theta_\epsilon(\omega)\\ &= +[\lambda'(0_+)]^{-1}|G_\theta(\omega)\rangle\langle V^{1/2}_\parallel|V^{1/2}G_0\times V^{1/2}_\parallel(E_{(-1)})V^{1/2}G_0V^{1/2}_\parallel|V^{1/2}\rangle\langle G_\theta(\omega)|\\ &= +[\lambda'(0_+)]^{-1}\left(\sum_{j=1}^{N}|(V^{1/2}_\parallel,\phi_j)|^2\right)(\tilde\phi_j,\phi_j)^{-1}\times(G_\theta(\omega),\cdot)G_\theta(\omega)=0,\end{aligned} \tag{3.23}$$

$(V^{1/2}_\parallel,\phi_j)=0$ being $\forall j=1,\dots,N$.

Hence also in case (i) (a) $H_\epsilon(\theta)$ converges in the norm resolvent sense to $-\Delta_\theta$ for any $\theta\in[0,2\pi)^3$.

We now consider the case when -1 is a simple eigenvalue of $V^{1/2}G_0V^{1/2}_\parallel$ and $(V^{1/2}_\parallel,\phi)=(\tilde\phi,V^{1/2})\neq 0$. Here we have obtained

$$\begin{aligned}\| \ \| - \lim_{\epsilon\to 0_+} &-\epsilon\lambda^3(\epsilon)A^\theta_\epsilon(\omega)B^\theta_\epsilon(\omega)\times[1+\lambda(\epsilon)B^\theta_\epsilon(\omega)]^{-1}B^\theta_\epsilon(\omega)C^\theta_\epsilon(\omega)\\ &= -\left(\frac{i\omega}{4\pi}+\tilde G_\theta(0;\omega)-\frac{\lambda'(0_+)(\tilde\phi,\phi)}{|(V^{1/2}_\parallel,\phi)|^2}\right)^{-1}\times(G_\theta(\omega),\cdot)G_\theta(\omega),\end{aligned} \tag{3.24}$$

$\forall\omega\in\mathbb{C}$ with $\operatorname{Re}\omega\neq 0$, $\operatorname{Im}\omega>0$, and $\forall\theta\in[0,2\pi)^3$. Hence,

$$\| \ \| - \lim_{\epsilon\to 0_+}[H_\epsilon(\theta)-\omega^2]^{-1} = (-\Delta_\theta-\omega^2)^{-1} - [i\omega/4\pi+\tilde G_\theta(0;\omega)-\alpha]^{-1}\times(G_\theta(\omega),\cdot)G_\theta(\omega), \tag{3.25}$$

$\forall\omega\in\mathbb{C}$ with $\operatorname{Re}\omega\neq 0$, $\operatorname{Im}\omega>0$, $\forall\theta\in[0,2\pi)^3$, and

$\alpha = \lambda'(0_+)(\tilde{\phi},\phi)/|(V_\parallel^{1/2},\phi)|^2$. By setting $E=\omega^2(\mathrm{Im}\, E^{1/2}>0)$, this is equivalent to saying

$$\begin{aligned}\| \ \| &- \lim_{\epsilon\to 0_+} [H_\epsilon(\theta)-E]^{-1} \\ &= (-\Delta_\theta - E)^{-1} - [i\sqrt{E}/4\pi + \tilde{G}_\theta(0;\sqrt{E}) - \alpha]^{-1} \\ &\quad \times (G_\theta(\sqrt{E}),\cdot)G_\theta(\sqrt{E}), \end{aligned} \tag{3.26}$$

$\forall E\in\mathbb{C}\setminus\mathbb{R}$ and $\forall\theta\in[0,2\pi)^3$.

We now note that

$$\begin{aligned}&\mathrm{Ker}\{(-\Delta_\theta - E)^{-1} - [i\sqrt{E}/4\pi + \tilde{G}_\theta(0;\sqrt{E}) - \alpha]^{-1} \\ &\quad \times (G_\theta(\sqrt{E}),\cdot)G_\theta(\sqrt{E})\} = \{0\},\end{aligned}$$

since there is no $\mathbb{L}^2(C,d^3x)$ function satisfying $(-\Delta_\theta - E)^{-1}f = G_\theta(E^{1/2})*f = \text{const}\, G_\theta(\sqrt{E})$. Hence we are in the hypotheses of the Trotter–Kato theorem (see Ref. 5, Theorem VIII.22), and by exploiting it, we get that the rhs of (3.26) is the resolvent of a self-adjoint operator $H_\alpha(\theta)$, which is the limit in the norm resolvent (n.r.) convergence of the net $\{H_\epsilon(\theta)\}_{\epsilon\in(0,1}$, $\forall\theta\in[0,2\pi)^3$.

Case (ii) (a) is quite similar, since we have

$$\begin{aligned}\| \ \| &- \lim_{\epsilon\to 0_+} -\epsilon\lambda^3(\epsilon)A_\epsilon^\theta(\omega)B_\epsilon^\theta(\omega) \\ &\quad \times [1+\lambda(\epsilon)B_\epsilon^\theta(\omega)]^{-1}B_\epsilon^\theta(\omega)C_\epsilon^\theta(\omega) \\ &= -[i\omega/4\pi + \tilde{G}_\theta(0;\omega) - \alpha]^{-1}(G_\theta(\omega),\cdot)G_\theta(\omega), \end{aligned} \tag{3.27}$$

where α is given by $\lambda'(0_+)\,[\Sigma_{j=1}^N |(V_\parallel^{1/2},\phi_j)|^2/(\tilde{\phi}_j,\phi_j)]^{-1}$.

Hence also in this case we obtain that $H_\epsilon(\theta)$ converges in the n.r. sense to a self-adjoint operator $H_\alpha(\theta)$, with resolvent given by the rhs of (3.26) with $\alpha = \lambda'(0_+)\,[\Sigma_{j=1}^N |(V_\parallel^{1/2},\phi_j)|^2/(\tilde{\phi}_j,\phi_j)]^{-1}$, $\forall\theta\in[0,2\pi)^3$.

We now have to deal with the most delicate cases, i.e., (i) (b) and (ii) (b), where $\lim_{\epsilon\to 0_+}\epsilon[1+\lambda(\epsilon)B_\epsilon^\theta(\omega)]^{-1}$ does not exist, because of the nonanalyticity of the $L\,[\mathbb{L}^2(C,d^3x)]$ function $\epsilon[1+\lambda(\epsilon)B_\epsilon^\theta(\omega)]^{-1}$ at $\epsilon=0$.

We only discuss case (ii) (b), case (i) (b) being quite similar. We have seen that in this case $\epsilon[1+\lambda(\epsilon)B_\epsilon^\theta(\omega)]^{-1}$ has a norm convergent Laurent expansion around $\epsilon=0$ given by

$$\epsilon[1+\lambda(\epsilon)B_\epsilon^\theta(\omega)]^{-1} = (1/\epsilon)D_{-1}(\theta) + D_0(\theta) + O(\epsilon), \tag{3.28}$$

$\forall\theta\in[0,2\pi)^3$, where D_{-1} is of the type $F(\theta)[E_{(-1)} - (V_\parallel^{1/2},E_{(-1)}V^{1/2})^{-1}(E^*_{(-1)}V_\parallel^{1/2},\cdot)E_{(-1)}V^{1/2}]$, so that we get

$$\begin{aligned}D_{-1}(\theta)V^{1/2} &= F(\theta)[E_{(-1)}V^{1/2} - (V_\parallel^{1/2},E_{(-1)}V^{1/2})^{-1} \\ &\quad \times (E^*_{(-1)}V_\parallel^{1/2},V^{1/2})E_{(-1)}V^{1/2}] = 0 \end{aligned} \tag{3.29}$$

[which implies $D^*_{-1}(\theta)V_\parallel^{1/2}=0$.] Furthermore we have

$$(V_\parallel^{1/2},D_0(\theta)V^{1/2}) = [i\omega/4\pi + \tilde{G}_\theta(0;\omega)]^{-1},\ \forall\theta\in[0,2\pi)^3, \tag{3.30}$$

and

$$\begin{aligned}&(V^{1/2}G_0V_\parallel^{1/2})D_{-1}(\theta)(V^{1/2}G_0V_\parallel^{1/2}) = D_{-1}(\theta), \\ &(V^{1/2}G_0V_\parallel^{1/2})D_0(\theta)(V^{1/2}G_0V_\parallel^{1/2}) = D_0(\theta), \\ &\quad \forall\theta\in[0,2\pi)^3.\end{aligned} \tag{3.31}$$

For any f and g in $\mathbb{L}^2(C,d^3x)$, we have

$$\begin{aligned}&-\lambda^3(\epsilon)(f,A_\epsilon^\theta(\omega)B_\epsilon^\theta(\omega)\{\epsilon[1+\lambda(\epsilon)B_\epsilon^\theta(\omega)]^{-1}\} \\ &\quad \times B_\epsilon^\theta(\omega)C_\epsilon^\theta(\omega)g) \\ &= -\lambda^3(\epsilon)(f,A_\epsilon^\theta(\omega)B_\epsilon^\theta(\omega)[(1/\epsilon)D_{-1}(\theta)] \\ &\quad \times B_\epsilon^\theta(\omega)C_\epsilon^\theta(\omega)g) - \lambda^3(\epsilon)(f,A_\epsilon^\theta(\omega)B_\epsilon^\theta(\omega) \\ &\quad \times D_0(\theta)B_\epsilon^\theta(\omega)C_\epsilon^\theta(\omega)g) + O(\epsilon), \end{aligned} \tag{3.32}$$

$\forall\omega\in\mathbb{C}$ with $\mathrm{Re}\,\omega\neq 0$, $\mathrm{Im}\,\omega>0$, and $\forall\theta\in[0,2\pi)^3$.

We now study the limit of the second term on the rhs of (3.32):

$$\begin{aligned}&-\lambda^3(\epsilon)(f,A_\epsilon^\theta(\omega)B_\epsilon^\theta(\omega)D_0(\theta)B_\epsilon^\theta(\omega)C_\epsilon^\theta(\omega)g) \\ &\quad \underset{\epsilon\to 0_+}{\to} -(f,A_0^\theta(\omega)(V^{1/2}G_0V_\parallel^{1/2})D_0(\theta) \\ &\quad \times (V^{1/2}G_0V_\parallel^{1/2})C_0^\theta(\omega)g) \\ &= -([A_0^\theta(\omega)]^*f,D_0(\theta)C_0^\theta(\omega)g) \\ &= -(f,G_\theta(\omega))(V_\parallel^{1/2},D_0(\theta)V^{1/2})(G_\theta(\omega),g) \\ &= -[i\omega/4\pi + \tilde{G}_\theta(0;\omega)]^{-1}(f,G_\theta(\omega))(G_\theta(\omega),g). \end{aligned} \tag{3.33}$$

Hence we have to compute the limit of the first term on the rhs of (3.32).

From the expansion,

$$\begin{aligned}B_\epsilon^\theta(\omega) &= V^{1/2}G_{\epsilon\omega}V_\parallel^{1/2} + \epsilon V^{1/2}G_\theta(\epsilon\cdot,\epsilon\cdot;\omega)V_\parallel^{1/2} \\ &= V^{1/2}G_0V_\parallel^{1/2} + \epsilon[i\omega/4\pi + \tilde{G}_\theta(0;\omega)] \\ &\quad \times |V^{1/2}\rangle\langle V_\parallel^{1/2}| + O(\epsilon^2),\end{aligned}$$

we get

$$\begin{aligned}&-\lambda^3(\epsilon)[(1/\epsilon)(f,A_\epsilon^\theta(\omega)B_\epsilon^\theta(\omega)D_{-1}(\theta)B_\epsilon^\theta(\omega)C_\epsilon^\theta(\omega)g)] \\ &\quad = -(1/\epsilon)(f,A_\epsilon^\theta(\omega)D_{-1}(\theta)C_\epsilon^\theta(\omega)g) + O(\epsilon), \end{aligned} \tag{3.34}$$

where we have exploited the fact that

$$(V^{1/2}G_0V_\parallel^{1/2})D_{-1}(\theta)(V_\parallel^{1/2},\cdot)V^{1/2} = 0, \tag{3.35}$$

by (3.29), and the first equality of (3.31).

Now we note that $\forall f\in\mathbb{L}^2(C,d^3x)$, $[G_\theta(\omega)*f](\epsilon x)$ is a C^1 function, belonging to $D(-\Delta_\theta)$. So we may write

$$\begin{aligned}[G_\theta(\omega)*f](\epsilon x) &= [G_\theta(\omega)*f](0) \\ &\quad + \epsilon\left(\frac{d}{d\epsilon}[G_\theta(\omega)*f](\epsilon x)\right)_{\epsilon=0} + O(\epsilon^2), \end{aligned} \tag{3.36}$$

and this implies

$$\begin{aligned}&V^{1/2}(x)[G_\theta(\omega)*f](\epsilon x) \\ &\quad = V^{1/2}\{[G_\theta(\omega)*f](0)\} + \epsilon\{x\cdot\nabla[G_\theta(\omega)*f]|_0\} \\ &\quad \times V^{1/2}(x) + O(\epsilon^2), \end{aligned} \tag{3.37}$$

i.e.,

$$\begin{aligned}C_\epsilon^\theta(\omega)f &= |V^{1/2}\rangle\langle G_\theta(\omega)|f\rangle \\ &\quad + \epsilon[x\cdot\nabla[G_\theta(\omega)*f]|_0]V^{1/2} + O(\epsilon^2). \end{aligned} \tag{3.38}$$

In the same way,

$$\begin{aligned}[A_\epsilon^\theta(\omega)]^*f &= |V_\parallel^{1/2}\rangle\langle G_\theta(\omega)|f\rangle \\ &\quad + \epsilon[x\cdot\nabla[G_\theta(\omega)*f]|_0]V_\parallel^{1/2} + O(\epsilon^2). \end{aligned} \tag{3.39}$$

By expanding the matrix element in (3.34) we get

$$
\begin{aligned}
&-(1/\epsilon)(f,A^{\theta}_{\epsilon}(\omega)B^{(\theta)}_{\epsilon}(\omega)D_{-1}(\theta)B^{\theta}_{\epsilon}(\omega)C^{\theta}_{\epsilon}(\omega)g)\\
&\quad=-(1/\epsilon)(f,G_{\theta}(\omega))(V_{\parallel}^{1/2},D_{-1}(\theta)V^{1/2})(G_{\theta}(\omega),g)\\
&\qquad-(\{x\cdot\nabla[G_{\theta}(\omega)*f]|_0\}V_{\parallel}^{1/2},D_{-1}(\theta)V^{1/2})\\
&\qquad\times(G_{\theta}(\omega),g)-(f,G_{\theta}(\omega))(V_{\parallel}^{1/2},D_{-1}(\theta)V^{1/2}\\
&\qquad\times\{x\cdot\nabla[G_{\theta}(\omega)*g]|_0\})+O(\epsilon)=O(\epsilon),
\end{aligned}
\tag{3.40}
$$

since $D_{-1}(\theta)V^{1/2}=0$. Hence we have shown that $\forall f,g\in\mathbb{L}^2(C,d^3x)$,

$$
\begin{aligned}
&-\lambda^3(\epsilon)(f,A^{\theta}_{\epsilon}(\omega)B^{\theta}_{\epsilon}(\omega)\{\epsilon[1+\lambda(\epsilon)B^{\theta}_{\epsilon}(\omega)]^{-1}\}\\
&\quad\times B^{\theta}_{\epsilon}(\omega)C^{\theta}_{\epsilon}(\omega)g)\underset{\epsilon\to 0_+}{\to}-[i\omega/4\pi+\tilde{G}_{\theta}(0;\omega)]^{-1}\\
&\qquad\times(f,G_{\theta}(\omega))(G_{\theta}(\omega),g),
\end{aligned}
\tag{3.41}
$$

$\forall\omega\in\mathbb{C}$ with $\mathrm{Im}\,\omega>0$, $\mathrm{Re}\,\omega\neq 0$, and $\forall\theta\in[0,2\pi)^3$.

By setting $E=\omega^2(\mathrm{Im}\sqrt{E}>0)$, (3.41) implies

$$
\begin{aligned}
&[H_{\epsilon}(\theta)-E]^{-1}\\
&\quad\overset{w}{\underset{\epsilon\to 0_+}{\to}}(-\Delta_{\theta}-E)^{-1}-[i\sqrt{E}/4\pi+\tilde{G}_{\theta}(0;\sqrt{E})]^{-1}\\
&\quad\times G_{\theta}(\sqrt{E})(G_{\theta}(\sqrt{E}),\cdot),
\end{aligned}
\tag{3.42}
$$

$\forall E\in\mathbb{C}\setminus\mathbb{R}$ and $\forall\theta\in[0,2\pi)^3$.

By noting that the w limit is the resolvent of the self-adjoint operator $H_{\alpha}(\theta)$ with $\alpha=0$, defined by (3.26), we have

$$
\begin{aligned}
&[H_{\epsilon}(\theta)-E]^{-1}\\
&\quad\overset{s}{\underset{\epsilon\to 0_+}{\to}}(-\Delta_{\theta}-E)^{-1}-[i\sqrt{E}/4\pi+\tilde{G}_{\theta}(0;\sqrt{E})]^{-1}\\
&\quad\times G_{\theta}(\sqrt{E})(G_{\theta}(\sqrt{E}),\cdot).
\end{aligned}
\tag{3.43}
$$

When case (i) (b) holds, we find

$$
[H_{\epsilon}(\theta)-E]^{-1}\overset{s}{\underset{\epsilon\to 0_+}{\to}}(-\Delta_{\theta}-E)^{-1}.
$$

Hence we have proved that in all the cases the net of the reduced Hamiltonians converges in the resolvent sense to a self-adjoint operator. We summarize what we have proved in a theorem.

Theorem 3.1: Let $V\in\mathbb{L}^{\infty}(C,d^3x)$ have compact support such that supp $V\subset C$, C being the Wigner–Seitz cell. Let $\lambda(\epsilon)$ be a $C^{\infty}[0,1]$-real valued function with $\lambda(0)=1$.

We distinguish four cases.

(a) -1 is not an eigenvalue for the $J_2[\mathbb{L}^2(C,d^3x)]$ operator $V^{1/2}G_0V_{\parallel}^{1/2}$, with integral kernel given by $V^{1/2}(x)V_{\parallel}^{1/2}(y)/4\pi|x-y|$.

(b) -1 is a simple eigenvalue of $V^{1/2}G_0V_{\parallel}^{1/2}$ and $(V_{\parallel}^{1/2},\phi)=\overline{(\phi,V^{1/2})}\neq 0$, ϕ being such that $V^{1/2}GV_{\parallel}^{1/2}\phi=-\phi$ $[\tilde{\phi}=(\mathrm{sgn}\,V)\phi]$.

(c) -1 is an eigenvalue of multiplicity $N\geqslant 1$ (obviously $N<\infty$) for $V^{1/2}G_0V_{\parallel}^{1/2}$ and $V_{\parallel}^{1/2}\in\{E_{(-1)}\mathbb{L}^2(C,d^3x)\}^{\perp}$.

(d) -1 is an eigenvalue of multiplicity $N\geqslant 2$ for $V^{1/2}G_0V_{\parallel}^{1/2}$ and $V_{\parallel}^{1/2}\notin\{E_{(-1)}\mathbb{L}^2(C,d^3x)\}^{\perp}$.

If $\lambda'(0_+)\neq 0$, the net of the reduced Hamiltonians $H_{\epsilon}(\theta)=-\Delta_{\theta}+\lambda(\epsilon)\epsilon^{-2}V(x/\epsilon)$ converges in the norm resolvent sense when $\epsilon\to 0_+$ to a self-adjoint operator, which is $\forall\theta\in[0,2\pi)^3$: (1) $-\Delta_{\theta}$ in cases (a) and (c); and (2) $H_{\alpha}(\theta)$, i.e., the operator with resolvent given by

$$
\begin{aligned}
&[H_{\alpha}(\theta)-E]^{-1}\\
&\quad=(-\Delta_{\theta}-E)^{-1}-[i\sqrt{E}/4\pi+\tilde{G}_{\theta}(0;\sqrt{E})-\alpha]^{-1}\\
&\qquad\times G_{\theta}(\sqrt{E})(G_{\theta}(\sqrt{E}),\cdot),
\end{aligned}
$$

with $\alpha=\lambda'(0_+)(\tilde{\phi},\phi)/|(V_{\parallel}^{1/2},\phi)|^2$ in case (b) and $\alpha=\lambda'(0_+)(\Sigma_{i=1}^{N}|(V_{\parallel}^{1/2},\phi_i)|^2/(\tilde{\phi}_i,\phi_i))^{-1}$ in case (d), $\{\phi_i\}_{i=1}^{N}$ being an orthogonal basis for the spectral subspace $E_{(-1)}\mathbb{L}^2(C,d^3x)$.

When $\lambda'(0_+)=0$, we find the same limits, but in cases (c) and (d), norm resolvent convergence has to be replaced by a strong one. [We note that $-\Delta_{\theta}=H_{\infty}(\theta)$.]

IV. THE LIMIT OF $[H_{\epsilon}-E]^{-1}$

After having proved the convergence in the resolvent sense of the reduced Hamiltonians $-\Delta_{\theta}+\lambda(\epsilon)\epsilon^{-2}V(x/\epsilon)$, $\forall\theta\in[0,2\pi)^3$, it is immediate to define the limit of the net $H_{\epsilon}=-\Delta+\Sigma_{y\in L}\lambda(\epsilon)\epsilon^{-2}V(\cdot-y/\epsilon)$, L being the lattice in $\mathbb{R}^3$ having C as its Wigner–Seitz cell.

Exploiting the direct integral decomposition theory, we define a self-adjoint operator H_{α} ($H_{\infty}=-\Delta$) in $\mathbb{L}^2(\mathbb{R}^3,d^3x)$ by setting

$$
UH_{\alpha}U^{-1}=\int_{[0,2\pi)^3}^{\oplus}H_{\alpha}(\theta)\frac{d^3\theta}{(2\pi)^3}.
\tag{4.1}
$$

Obviously we have, $\forall E\in\mathbb{C}\setminus\mathbb{R}$,

$$
U[H_{\alpha}-E]^{-1}U^{-1}=\int_{[0,2\pi)^3}^{\oplus}[H_{\alpha}(\theta)-E]^{-1}\frac{d^3\theta}{(2\pi)^3}.
\tag{4.2}
$$

We show that H_{α} is the limit, in the strong resolvent convergence, of the net H_{ϵ}. In fact, $\forall f\in\mathbb{L}^2(\mathbb{R}^3,d^3x)$,

$$
\begin{aligned}
&\|(H_{\alpha}-E)^{-1}f-(H_{\epsilon}-E)^{-1}f\|^2_{\mathbb{L}^2(\mathbb{R}^3,d^3x)}\\
&\quad=\|U[H_{\alpha}-E]^{-1}U^{-1}(Uf)\\
&\qquad-U[H_{\epsilon}-E]^{-1}U^{-1}(Uf)\|^2_{\mathrm{H}}\\
&\quad=\int_{[0,2\pi)^3}\|[H_{\alpha}(\theta)-E]^{-1}(Uf)_{\theta}\\
&\qquad-[H_{\epsilon}(\theta)-E]^{-1}(Uf)_{\theta}\|^2_{\mathbb{L}^2(C,d^3x)}d^3\theta/(2\pi)^3,
\end{aligned}
\tag{4.3}
$$

$\forall E\in\mathbb{C}\setminus\mathbb{R}$.

Now, $[H_{\epsilon}(\cdot)-E]^{-1}(Uf)_{(\cdot)}$ converges pointwisely to $[H_{\alpha}(\cdot)-E]^{-1}(Uf)_{(\cdot)}$ when $\epsilon\to 0_+$ (by Theorem 3.1); the limit belongs to $\mathbb{L}^2([0,2\pi)^3,d^3\theta/(2\pi)^3;\mathbb{L}^2(C,d^3x))$, since

$$
\begin{aligned}
&\int_{[0,2\pi)^3}\|[H_{\alpha}(\theta)-E]^{-1}(Uf)_{\theta}\|^2_{\mathbb{L}^2(C,d^3x)}\frac{d^3\theta}{(2\pi)^3}\leqslant|\mathrm{Im}\,E|^{-2},\\
&\int_{[0,2\pi)^3}\|(Uf)_{\theta}\|^2_{\mathbb{L}^2(C,d^3x)}\frac{d^3\theta}{(2\pi)^3}\\
&\quad=|\mathrm{Im}\,E|^{-2}\|f\|^2_{\mathbb{L}^2(\mathbb{R}^3,d^3x)}.
\end{aligned}
\tag{4.4}
$$

Therefore, by the dominated convergence theorem,

$$
\begin{aligned}
&\int_{[0,2\pi)^3}\|[H_{\epsilon}(\theta)-E]^{-1}(Uf)_{\theta}\\
&\qquad-[H_{\alpha}(\theta)-E]^{-1}(Uf)_{\theta}\|^2_{\mathbb{L}^2(C,d^3x)}d^3\theta/(2\pi)^3\underset{\epsilon\to 0_+}{\to}0.
\end{aligned}
\tag{4.5}
$$

Let us recall that, in order to prove norm resolvent convergence of H_{ϵ} to H_{α} from norm resolvent convergence of $H_{\epsilon}(\theta)$ to $H_{\alpha}(\theta)$, it is necessary to show that

$[H_\epsilon(\cdot)-E]^{-1}\underset{\epsilon\to 0_+}{\to}[H_\alpha(\cdot)-E]^{-1}$ in $\mathbb{L}^\infty([0,2\pi)^3,d^3\theta/(2\pi)^3;\mathbb{L}^2(C,d^3x))$, i.e., we have to prove

$$\sup_{\theta\in[0,2\pi)^3}\|[H_\alpha(\theta)-E]^{-1}[H_\epsilon(\theta)-E]^{-1}\|\underset{\epsilon\to 0_+}{\to}0, \tag{4.6}$$

which, by (3.2) and (3.3), is equivalent to

$$\sup_{\theta\in[0,2\pi)^3}\|\{[i\sqrt{E}/4\pi+\tilde{G}_\theta(0;\sqrt{E})-\alpha]^{-1}G_\theta(\sqrt{E})\times(G_\theta(\sqrt{E}),\cdot)-\lambda^3(\epsilon)A^\theta_\epsilon(\sqrt{E})B^\theta_\epsilon(\sqrt{E})\times\{\epsilon[1+\lambda(\epsilon)B^\theta_\epsilon(\sqrt{E})]^{-1}\}B^\theta_\epsilon(\sqrt{E})C^\theta_\epsilon(\sqrt{E})\|\underset{\epsilon\to 0_+}{\to}0. \tag{4.7}$$

We have seen that $B^\theta_\epsilon(\sqrt{E})\underset{\epsilon\to 0_+}{\overset{\|\ \|_{J_2}}{\to}}V^{1/2}G_0V^{1/2}$ uniformly in θ; furthermore, also $A^\theta_\epsilon(\sqrt{E})$ and $C^\theta_\epsilon(\sqrt{E})$ converge in the J_2 norm uniformly in θ to their limits $A^\theta_0(\sqrt{E})$ and $C^\theta_0(\sqrt{E})$, as it follows by using the uniform convergence of the series on the rhs of (2.3) and dominated convergence.

Hence, in order to prove (4.7), one needs to show that the convergence of $\epsilon[1+\lambda(\epsilon)B^\theta_\epsilon(\sqrt{E})]^{-1}$ is uniform with respect to θ. For example, in the case that -1 is a simple eigenvalue of $V^{1/2}G_0V^{1/2}_\parallel$ and $V^{1/2}_\parallel\in\{E_{(-1)}\mathbb{L}^2(C,d^3x)\}^\perp$, we have to show

$$\sup_{\theta\in[0,2\pi)^3}\|\{[i\sqrt{E}/4\pi+\tilde{G}_\theta(0;\sqrt{E})]|(V^{1/2}_\parallel,\phi)|^2-\lambda'(0_+)(\tilde{\phi},\phi)\}^{-1}(\tilde{\phi},\cdot)\phi-\epsilon[1+\lambda(\epsilon)\times B^\theta_\epsilon(\sqrt{E})]^{-1}\|\underset{\epsilon\to 0_+}{\to}0, \tag{4.8}$$

which is equivalent to proving that the first-order coefficient of the Taylor expansion of $\epsilon[1+\lambda(\epsilon)B^\theta_\epsilon(\sqrt{E})]^{-1}$ is uniformly bounded in θ; one can easily verify this by writing the expression corresponding to (3.37) in Ref. 6. Hence in the cases where we had norm resolvent convergence of $H_\epsilon(\theta)$ to $H_\alpha(\theta)$ for $\epsilon\to 0_+$, we have, $\forall E\in\mathbb{C}\setminus\mathbb{R}$,

$$\|(H_\alpha-E)^{-1}-(H_\epsilon-E)^{-1}\|_{L(\mathbb{L}^2(\mathbb{R}^3,d^3x))}=\|[H_\alpha(\cdot)-E]^{-1}-[H_\epsilon(\cdot)-E]^{-1}\|_{\mathbb{L}^\infty([0,2\pi)^3,d^3\theta/(2\pi)^3;\mathbb{L}^2(C))}\underset{\epsilon\to 0_+}{\to}0. \tag{4.9}$$

So we have completely proved the following theorem.

Theorem 4.1: Let L be a lattice of $\mathbb{R}^3$ and C its Wigner–Seitz cell. Let $V(x)$ be a $\mathbb{L}^\infty(\mathbb{R}^3;d^3x)$ function such that supp $V\subset C$. Furthermore let $\lambda(\epsilon)$ be a $C^\infty[0,1]$-real valued function with $\lambda(0_+)=1$. Then the net of Hamiltonians $H_\epsilon=-\Delta+\Sigma_{y\in L}\lambda(\epsilon)\epsilon^{-2}V(\cdot-y/\epsilon)$ converges in the resolvent sense to the self-adjoint unbounded operator given by

$$UH_\alpha U^{-1}=\int^{\oplus}_{[0,2\pi)^3}H_\alpha(\theta)\frac{d^3\theta}{(2\pi)^3}.$$

The convergence is in the norm resolvent sense in all the cases where by Theorem 3.1, the net of the reduced Hamiltonians $H_\epsilon(\theta)$ converges in the norm resolvent sense to $H_\alpha(\theta)$ when $\epsilon\to 0$, and in the strong one otherwise.

Remark: The operator we have obtained is actually the same one of Theorem 5.2 (Ref. 1) since

$$\begin{aligned}\tilde{G}_\theta(0;\sqrt{E})&=\sum_{\substack{\mathbf{m}\in\mathbb{Z}^3\\ \mathbf{m}\neq 0}}\frac{\exp(i\sqrt{E}\,|\Sigma_{i=1}^3m_i\mathbf{a}_i|)}{4\pi|\Sigma_{i=1}^3m_i\mathbf{a}_i|}e^{-i\mathbf{m}\cdot\theta}\\&=\lim_{\omega\to\infty}(2\pi)^3\left(\sum_{\substack{\mathbf{l}\in L^*\\ |\mathbf{l}+\mathbf{k}|<\omega}}\frac{1}{|\mathbf{l}+\mathbf{k}|^2-E}-\int_{|p|<\omega}\frac{d^3p}{p^2-E}\right)\\&=\lim_{\omega\to\infty}(2\pi)^{-3}\left(\sum_{\substack{\mathbf{l}\in L^*\\ |\mathbf{l}+\mathbf{k}|<\omega}}\frac{1}{|\mathbf{l}+\mathbf{k}|^2-E}-4\pi\omega\right)-\frac{i\sqrt{E}}{4\pi},\end{aligned}$$

where L^* is the dual lattice. By substituting into the expression of $[H_\alpha(\theta)-E]^{-1}$, we have

$$[H_\alpha(\theta)-E]^{-1}=(-\Delta_\theta-E)^{-1}-[g_E(0;\mathbf{k})-\alpha]^{-1}(G_\theta(\sqrt{E}),\cdot)G_\theta(\sqrt{E}),$$

where

$$g_E(0;\mathbf{k})=\lim_{\omega\to\infty}(2\pi)^{-3}\left(\sum_{\substack{\mathbf{l}\in L^*\\ |\mathbf{l}+\mathbf{k}|<\omega}}\frac{1}{|\mathbf{l}+\mathbf{k}|^2-E}-4\pi\omega\right).$$

ACKNOWLEDGMENTS

This research is part of my graduation thesis. I would thank my supervisor Professor G. dell'Antonio for having suggested this problem to me and for having taken a great interest in this work.

[1] A. Grossmann, R. Høegh-Krohn, and M. Mebkhout, Commun. Math. Phys. 77, 87 (1980).

[2] M. Reed and B. Simon, *Analysis of Operators. Method of Modern Mathematical Physics IV* (Academic, New York, 1978).

[3] B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, NJ, 1971).

[4] M. Reed and B. Simon, *Fourier Analysis, Self-Adjointness-Methods of Modern Mathematical Physics II* (Academic, New York, 1975).

[5] M. Reed and B. Simon, *Functional Analysis. Methods of Modern Mathematical Physics I* (Academic, New York, 1972).

[6] S. Albeverio, F. Gesztesy, and R. Høegh-Khron, Ann. Inst. H. Poincaré A 37, 1 (1982).

Global reduction of a dynamical system on a foliated manifold

Ida Cattaneo Gasparini
Dipartimento di Metodi e Modelli Matematici per le Scienze Applicate, Università degli Studi di Roma "La Sapienza," Italy

(Received 30 November 1983; accepted for publication 20 April 1984)

Conditions for the global reductions of a dynamical system defined on foliated manifold M are given. They are expressed by a local condition on the topology of one single leaf and a global condition on the transverse bundle to the foliation. The link of this condition is shown with the existence of a normalizer of the Lie algebra of the vector fields tangent to the foliation in the Lie algebra of all the vector fields of M. This normalizer contains all the derivations of the functions constant on the leaves.

PACS numbers: 02.40. + m

I. INTRODUCTION

In this paper we consider the following global problem: If Y is a dynamical system defined on a foliated manifold $(M,\mathcal{F})$, under what conditions on the vector field Y, on the topology of the leaves and of the manifold M, is $M/\mathcal{F}$ a differentiable Hausdorff manifold so that the dynamical system Y can be globally reduced to a system on the manifold $M/\mathcal{F}$?

The space of the leaves is studied through the properties of the transverse bundle of the foliation and by a systematic use of a connection introduced in a previous paper of 1969,[1,2] which allows a transport of the transverse bundle of the foliation along the foliation itself. This connection is a global operator in an "almost parallelizable manifold and, in the particular case of foliated manifold, it gives the connection introduced by Bott.[3] As this operator can be useful in other contexts where there are vector fields intrinsically linked to the problem, we shall collect several facts that have been obtained using it. (Sec. II). In Sec. IV we define "almost regular foliations" in a connected smooth manifold and we prove that under a global condition on the transverse bundle of the foliation and a local condition on the topology of one single leaf, $M/\mathcal{F}$ is a Hausdorff, smooth manifold (Theorem 4.3). The separability property of $M/\mathcal{F}$ is necessary to ensure the uniqueness of the integral curves of the projected vector fields. If, in addition, the parallel frames are complete, the manifold M is a fiber bundle with the leaves as fibers. In Sec. V we prove that if Y is parallel along the leaves and if the foliation and the transverse bundle satisfy the conditions of Theorem 4.3, then the dynamical system Y can be globally reduced on $M/\mathcal{F}$ (Theorem 5.2). Finally, in Sec. VI we give an interpretation of this theorem in terms of the existence of a transitive normalizer of the Lie algebra of the vector fields tangent to the foliation in the infinite-dimensional algebra of all the vector fields tangent to M and we point out that this normalizer contains all the derivations of the functions constant on the leaves.

It appears to us that we are using the most natural and simple building blocks for the study of $M/\mathcal{F}$. We remark that independent of the problem of the reduction, Theorem 4.3 may have an interest in many physical situations in which foliations arise naturally in state manifold and we want to "quotient out" the leaves: in general relativity for instance or in problems of observability in control theory.

II. TOTAL LIE DIFFERENTIAL (TLD) OR LIE CONNECTION

Let M be a smooth manifold of dimension n, $F(M)$ the ring of smooth functions on M, and $\mathcal{H}(M)$ the F-module of smooth vector fields on M. A linear connection associates to each piecewise smooth path $\sigma: I = (0,1) \to M$ a linear map of tangent spaces in $\sigma(0)$ and $\sigma(1)$ depending on σ. From an algebraic point of view a linear connection can equivalently be defined as a rule which assigns to each $X \in \mathcal{H}(M)$ a map ∇ of $\mathcal{H}(M)$ into itself called covariant differentiation satisfying the following axioms:

(i) $\nabla_X(Y' + Y'') = \nabla_X Y' + \nabla_X Y''$,

(ii) $\nabla_X fY = Xf \cdot Y + f\nabla_X Y$,

(iii) $\nabla_{fX+gZ} Y = f\nabla_X Y + g\nabla_Z Y$,

with X, Y, $Z \in \mathcal{H}(M)$ and $f, g \in F(M)$.

If X is a vector field of M, we shall indicate with $L(X)$ the Lie derivative associated to X. It is classically known that the Lie derivative is not a connection operator as it does not satisfy (iii). In fact we have

$$L(fX)Y = fL(X)Y - (Yf)\cdot X, \tag{1}$$

namely the map $X \to L(X)$ is not linear over the functions.

Now consider a field of frames $\{X_i\}$ $(i = 1,...,n)$ on an open set U of M. For any vector field $Z = Z^r X_r (r = 1,...n)$ we defined the operator

$$D_Z \overset{\text{def}}{=} Z^r L(X_r), \tag{2}$$

which satisfies all the axioms of a connection operator. We have called it the total Lie differential (TLD) or Lie connection.

The Lie connection associated to a frame field $\{X_i\}$ does not change if the frame field changes with a transformation with constant coefficients. The operator D_Z is then a global operator on an almost parallelizable manifold, namely a manifold such that the structure group of the tangent bundle can be reduced to a discrete group. The advantages of the TLD are the following.

(1) If on M there are some vector fields with particular geometrical or physical meanings, for instance, symplectic automorphism on a symplectic manifold or isometric vector fields on Riemannian manifold or vector fields tangent to a foliation on a foliated manifold, then this operator can be intrinsically linked to them.

 0022-2488/84/102918-04$02.50

(2) It is possible to define a covariant derivative of a connection with respect to a TLD, a notion which is not defined with respect to any other connection.

We shall recall here some results[4] that have been obtained using the TLD: (a) a sufficient condition is given for the Euler characteristic of a reductive homogeneous space (in particular a compact homogeneous space) to be null; and (b) for the p-sectional curvature and Pontrjagin classes to be null; (c) a proof that Darboux's theorem is valid globally in a symplectic manifold which admits a symplectic connection with curvature and torsion zero; and finally (d) the characterization of simple connected manifolds with connection ω which can have a structure of homogeneous reductive spaces with respect to a Lie group of ω-affine transformations.

III. PARALLEL TRANSVERSE BUNDLE

Let $(M,\mathcal{F})$ be a smooth connected manifold, foliated with a smooth foliation $\mathcal{F}$. Let n be the dimension of M and q the codimension of the foliation.[5,6] M is the union of disjoint connected submanifolds $\{L_i\}$ of dimension $n-q$ and every point of M has a neighborhood U with a system of local coordinates $(x^A)=(x^1,\ldots,x^n):U\to\mathbb{R}^n$ $(A=1,\ldots,n)$ such that the connected components of $L\cap U$ which are called local leaves are described by equations

$$x^{n-q+1}=\text{const},\ldots,x^n=\text{const}. \tag{3}$$

If $x\in U_i\cap U_j$ and if $(x^{1'},\ldots,x^{n'})$ are the local coordinates in U_j, the changes of coordinates are given by

$$\begin{aligned} &x^{a'}=h_1(x^a,x^\alpha), \qquad a,b=1,2,\ldots,n-q,\\ &x^{\alpha'}=h_2(x^\alpha), \quad \alpha,\beta=n-q+1,\ldots,n. \end{aligned} \tag{4}$$

Coordinates for which $x^\alpha=\text{const}$ on the leaves are said to be distinguished coordinates or adapted to the foliation $\mathcal{F}$.

Definition 3.1: A differentiable map of the manifold $(M,\mathcal{F})$ to a manifold N is said to be distinguished if it is constant on the leaves. The stalk of germs of distinguished functions is the structural stalk of the foliated structure.

Consider now the fiber bundles $T(M),E,Q,R(Q)$ having the same base manifold M. $T(M)$ is the tangent bundle of M, E is the $(n-q)$-dimensional integrable sub-bundle of $T(M)$ consisting of the vectors tangent to the leaves of the foliation $\mathcal{F}$. Q is the quotient bundle $Q=T(M)/E$ and is said to be a transverse bundle to $\mathcal{F}$. Finally $R(Q)$ is the principal $\mathrm{GL}(q,R)$ bundle associated to Q. We shall indicate with $\Gamma(E)$ and $\Gamma(Q)$ the spaces of smooth sections of the vector bundles E,Q.

Choose now a selected splitting $T(M)=E\oplus Q$ and let $\{X_a\}$ and $\{X_\alpha\}$ be local trivializations of the fiber bundles E and Q over an open set $U\subset M$. Then if D is the total Lie differential we have for any $Z\in E/U$ and $Y\in Q/U$

$$\begin{aligned} D_Z Y &= Z^a L(X_a)(Y^\alpha X_\alpha)\\ &= Z^a(\{L(X_a(Y^\alpha X_\alpha)\}_E+\{L(X_a(Y^\alpha X_\alpha)\}_Q). \end{aligned} \tag{5}$$

Let $[Y]$ be the equivalence class of Y in the quotient bundle Q. We have then

$$[D_Z Y]\overset{\text{def}}{=}Z^a(L(X_a)Y)_Q. \tag{6}$$

Remark 3.2: As we project over Q and we transport along the leaves, we see from formula (1) that the Lie derivative satisfies, in that case, the property of linearity over functions, so that we have

$$[D_Z Y]=(L(Z)Y)_Q. \tag{7}$$

For the transport of the transverse Q-projections of the bundle along the leaves, the TLD coincides then with the Lie derivatives. It follows that the TLD is globally defined on the leaves without any further assumption on the leaves (to be almost parallelizable). We obtain then the Bott's connection as a particular case of our TLD.

Definition 3.3: The transverse bundle Q is parallel along the leaves[7] if there is a frame field $\{X_\alpha\}$ of Q such that

$$[D_Z X_\alpha]=0, \quad \forall Z\in\Gamma E. \tag{8}$$

Definition 3.4: We say that the structural group of Q is reducible to a discrete distinguished group if the transition functions of Q

$$g_{ij}:U_i\cap U_j\to\mathrm{GL}(q,\mathbb{R}) \tag{9}$$

are functions constant along the connected components of the intersections of $U_i\cap U_j$ with each leaf.

Proposition 3.5: Let $(M,\mathcal{F})$ be a foliated manifold. If Q is the transverse bundle, then Q is parallel along $\mathcal{F}$ iff the structural group of Q can be reduced to a discrete distinguished group.

Proof: Let $\{X_\alpha\}$ be a trivialization of Q with parallel frames. $Y\in\Gamma(Q)$ is parallel along $\mathcal{F}$ if

$$\begin{aligned} &[L(Z)Y]=0, \text{ i.e., } [(ZY^\alpha)X_\alpha+Y^\alpha L(Z)X_\alpha]=0,\\ &\quad \forall Z\in\Gamma(E). \end{aligned} \tag{10}$$

As by assumption $[L(Z)X_\alpha]=0$, it follows that

$$ZY^\alpha=0. \tag{11}$$

Thus, if we change the frame $\{X_\alpha\}\in R(Q)$ in a frame $\{Y_a\}$ parallel along the leaves, the components of the vectors of the new frame are functions constant along the leaves. The structural group of Q, can then be reduced to a distinguished discrete group.

Conversely, choose locally a parallel frame $\{X_\alpha\}$. As by hypothesis the structural group of the transverse fiber bundle can be reduced to a discrete distinguished group, then the parallel frame can be globally defined on the leaves and Q is parallel along $\mathcal{F}$.

Definition 3.6: We say that the foliation satisfies the property of K-transitivity (K,T) if there exists a $K\in\mathbb{R}^+$ such that for any two points $x\neq y$ belonging to the same leaf L if $\varphi_t x$ and $\varphi_t y$ are the integral curves of any transverse vector field passing through the points x and y for $t=0$, the points $\varphi_{t_1}x$ and $\varphi_{t_1}y$ belong to the same leaf for any $t_1<K$.

IV. MANIFOLD OF THE LEAVES[8]

Definition 4.1: We say that a foliation $\mathcal{F}$ on the differentiable manifold $(M,\mathcal{F})$ is almost regular (a.r.) if: (i) the local submersions which describe the local leaves are proper; (ii) for any neighborhood U of an arbitrary point $p\in M$, $\forall L\in\mathcal{F}$, if $U\cap L\neq 0$ then $U\cap L$ consists of a finite number of local leaves.

Theorem 4.2: If M is a connected, smooth manifold and $\mathcal{F}$ is a foliation on M of codimension q, the foliation $\mathcal{F}$ is almost regular iff: (1) $\mathcal{F}$ contains a locally closed leaf; (2) the

structure group of the transverse bundle Q can be reduced to a finite distinguished group.

Proof: Let U be an open set with adapted coordinates such that the local leaves are given by Eq. (3). These local leaves are level sets of the local submersions $f: U \to \mathbb{R}^q$ with $f(p) = (x^{n-q+1}(p), \ldots, x^n(p))$ $(p \in U)$. As it is assumed that M is almost regular these local leaves are closed in M.

Consider now the universal cover $\tilde{L}$ of a leaf L and a homomorphism φ of the fundamental group $\pi_1(L)$ of L in $\mathrm{Diff}(R^q)$. The trivial foliation $\tilde{L} \times R^q$ goes in the local leaves of M if we consider the action of $\pi_1(L)$ on $\tilde{L}$ and of φ on $\mathbb{R}^q$, namely if we consider equivalent $(x,y) \sim (x',y')$ with $x \in \tilde{L}$ and $y \in \mathbb{R}^q$ when there exists a $g \in \pi_1(L)$ such that $x' = gx, y' = \varphi(g)y$. For instance, in the Mobius band $\tilde{S}^1 = \mathbb{R}$ and $\varphi: \pi_1(S^1) \to \mathrm{Diff}\, R$ gives the homomorphism $Z \to Z_2$. Remark now that this group is the holonomy group[9] of the connection D which transports the transverse bundle Q or $R(Q)$ parallel along the leaves. As it is supposed that the number of local leaves, for any L, is finite, the group $\varphi(L)$ is finite. The structure group of Q can therefore be reduced to a finite distinguished group.

Conversely, let $(X_\alpha) \in \Gamma R(Q)$ be a transverse frame field parallel along the leaves, and let $V = V^\alpha X_\alpha$ be any transverse vector such that the integral curve $\psi(t)$ of V passes for $t = 0$ through the point x. The map $\mathbb{R}^q \to M: (V^1, \ldots, V^q) \to \exp(V) \equiv \psi(1)$ in a sufficiently small neighborhood of the point $0 \in \mathbb{R}^q$ defines a diffeomorphism between the q-disk D^q_x transverse to the foliation and centered in x and a neighborhood U of x in M. As the vectors of D^q_x are vectors of Q, D^q_x can be transported by parallelism along the leaves, and in particular along the leaf L which is supposed locally close. A diffeomorphism can then be defined between the leaves in a neighborhood of L. As L is locally closed, we deduce that every leaf is locally closed, so that condition (1) is satisfied. Condition (2) is a consequence of the fact that, as the structural group of $R(Q)$ is distinguished finite, the holonomy of D along the leaves is finite. The number of local leaves is then finite.

Theorem 4.3: Let $(M, \mathcal{F})$ be a connected, smooth manifold foliated with an a.r. foliation satisfying the property of K-transitivity. Then the space of the leaves $M/\mathcal{F}$ is Hausdorff and has the structure of a differentiable V-manifold.

Proof: As $\mathcal{F}$ is a.r., for any $x \in M$, any $L \in \mathcal{F}$, and any adapted neighborhood of x, the intersection of $U \cap L$ is given by a finite number of local leaves, and as the leaves are locally closed (Theorem 4.2) and $\mathcal{F}$ satisfies the property (K,T), we can find a q-disk D^q_x (with $x \in L$) so small that $D^q_x \cap L = x$ and $D^q_x \cap L' = \emptyset$ for any $L' \neq L$. Consequently, for any two points x and x', we can define two neighborhoods on $M/\mathcal{F}$ such that $U_x \cap V_{x'} = \emptyset$.

Moreover $M/\mathcal{F}$ can be endowed with a structure of differentiable V-manifold; in fact the disk D^q_x with $x \in L$ can be parallel transported along L to any other point y of L. The disk D^q_x in y [denoted as $D^q_x(y)$] intersects the same leaves as the disk D^q_x. In a sufficiently small neighbborhood of the leaf L, we obtain thus a diffeomorphism between the open sets $U_x = \exp D^q_x$ and $U_y = \exp D^q_x(y)$. The canonical map $\pi: M \to M/\mathcal{F}$ is then an open map.

By the action of the finite group $\varphi \in \mathrm{Diff}\, \mathbb{R}^q$, we identify the local open sets of $\mathbb{R}^q$ which correspond to an open set U of $M/\mathcal{F}$. As φ is acting properly discontinuously, a local uniformizing system of open sets can be introduced in the Hausdorff space $M/\mathcal{F}$ and a structure of a V-manifold can be defined on $M/\mathcal{F}$. If φ is the identity we have the usual differentiable structure. V-manifold has been introduced by Satake.[10] Satake proved that differential forms, integration, and de Rham theorems can be generalized to these manifolds.

In the case where M is compact and φ is the identity we have a well-known result that M is a fiber bundle.

Theorem 4.4: If M is compact and the structural group of the transverse bundle Q can be reduced to the identity then all the leaves are diffeomorphic, M is a fiber bundle, $M/\mathcal{F}$ is the base space, and the leaves are the fibers.

Proof: If M is compact the transverse fields $\{X_\alpha\} \in \Gamma R(Q)$ are complete and if $\varphi_\beta(t)$ is the flow generated by the field X_β then $((\varphi_\beta)_t)_* E_x = E_{(\varphi_\beta)_t x}$. The distribution E_x is then invariant by the flow $(\varphi_\beta)_t$ generated by the vector field X_β. If x and y belong to the same leaf L_0 and $\varphi_\beta(0)x = x$, $\varphi_\beta(0)y = y$, then for any $t \in \mathbb{R}$, $\varphi_\beta(t)x$ and $\varphi_\beta(t)y$ belong to the same leaf L_1. From the differentiability of the solutions $\varphi_\beta(t, x_0)$ with respect to the initial values x_0, t_0 we see that the flows $\{\varphi_\alpha\}$ generated by the $\{X_\alpha\}$ define a diffeomorphism between the leaves of the foliation.

Corollary 4.3: $M/\mathcal{F}$ is an almost parallelizable manifold.

Proof: The frames of $R(Q)$ transported by parallelism along the leaves give, through the canonical projection, global frame fields over $M/\mathcal{F}$.

V. REDUCTION OF A DYNAMICAL SYSTEM

Consider now on $M/\mathcal{F}$ a dynamical system given in a local system of coordinates by the differential equation $\dot{x} = Y(x)$.

The problem of reducing locally a dynamical system with respect to a foliation has been treated recently in this Journal in a paper by Marmo *et al.*[11]

Definition 5.1: A vector field in a foliated manifold $(M,\mathcal{F})$ is projectable with respect to the foliation, if it transforms distinguished functions in distinguished functions, namely if f is a function constant on the leaves, Yf must be a function constant on the leaves. From the classical relation $L(X)L(Y)f - L(Y)L(X)f = L[X,Y]f$, we have if f is constant on the leaves and $X \in \Gamma(E)$

$$L(X)L(Y)f = 0 \Leftrightarrow L[X,Y]f = 0, \tag{12}$$

it follows that Y is projectable iff $[X,Y] \in \Gamma(E) \forall X \in \Gamma(E)$.

We now extend the action of the operator D_X and Definition 3.3 of parallelism to any $Y \in \mathcal{H}(M)$.

Definition 5.2: We say that a vector field Y on $(M,\mathcal{F})$ is parallel along the leaves if

$$[D_x Y]_Q = 0, \text{ i.e., } [L(X)Y]_Q = 0, \ \forall X \in \Gamma(E). \tag{13}$$

Proposition 5.3: A vector field Y on a foliated manifold $(M,\mathcal{F})$ is projectable with respect to a foliation iff it is parallel along the leaves. We obtain then the following.

Theorem 5.4: Let Y be a dynamical system on a smooth connected manifold $(M,\mathcal{F})$ foliated with a (K,T) foliation of codimension q. If the foliation has one single leaf locally

closed and if the structural group of the transverse bundle Q, can be reduced to a finite distinguished group, and Y is parallel along the leaves, then the given system can be globally reduced to a system on $M/\mathcal{F}$.

Proof: From Theorems 4.2 and 4.3, $M/\mathcal{F}$ is a smooth Hausdorff manifold. From Proposition 5.3 Y is projectable on $M/\mathcal{F}$. The system can thus be globally reduced to a system on the V-manifold $M/\mathcal{F}$. Remark that if $Y \in \Gamma(E)$, the projected vector field is the null vector of the reduced system on $M/\mathcal{F}$.

We can thus conclude that under the given assumption the system can be always globally reduced and that by integrating the reduced dynamical system one determines the given flow modulo the movement along the leaves or one can reduce the given system to a system on the leaves.

VI. ASSOCIATED TRANSITIVE NORMALIZER

We shall now give an interpretation to the previous results in terms of Lie algebras. We recall the following.

Definition 6.1: The normalizer $N_g(L)$ of a subalgebra L of a Lie algebra g is defined by

$$N_g(L) = \{X \in g, [X,L] \subset L\}. \tag{14}$$

The normalizer of L contains L as an ideal, and, from Jacobi's identity it is easily proved that it is a Lie algebra.

Let $(M,\mathcal{F})$ be a foliated manifold with a foliation of codimension q. Let $\mathcal{H}(M)$ be the Lie algebra of the vector fields of M, L the Lie algebra of the vector fields tangent to E, and L' the normalizer of L in $\mathcal{H}(M)$.

Proposition 6.2: The normalizer of L in $\mathcal{H}(M)$ consists of the vectors parallel along the leaves with respect to the TLD $D_X \forall X \in \Gamma(E)$.

Proposition 6.3: The normalizer represents the set of the derivations of the functions constant along the leaves. It is then of fundamental importance in the study of the manifold of the leaves and in the problems of reduction of a dynamical system.

Naturally this set of derivations must be "large" enough, namely its dimension over $\mathbb{R}$ must be n. In an equivalent way we say that the normalizer must be transitive on M.

ACKNOWLEDGMENT

This work was partially supported by Grant 291 and by a grant from M.P.I.

[1] I. Cattaneo Gasparini, "Operatori intrinseci di derivazione su una varietà parallelizabile," Rend. Accad. Naz. Lincei, Ser. VIII **6**, 682 (1969).

[2] I. Cattaneo Gasparini, "Group action and differential operators," in *Lecture Notes in Control and Information Sciences*, Vol. 39, in *Feedback Control of Linear and Nonlinear Systems* (Springer-Verlag, Berlin, 1982), pp. 30–40.

[3] R. Bott, "On a topological obstruction to integrability," in *Proceedings of the International Congress of Mathematics, Nice*, 1970 (Gauthiers-Villars, Paris, 1971) Vol. 1, p. 27–36.

[4] I. Cattaneo Gasparini, "Algebra di Lie e caratteristica di uno spazio omogeneo compatto," Rend. Accad. Naz. Lincei, Ser. VIII **54**(3) (1973); I. Cattaneo Gasparini and G. De Cecco, "Algebra di Lie e curvature p-sezionali di uno spazio omogeneo compatto," Rend. Accad. Naz. Lincei, Ser. VIII 56 (1974); I. Cattaneo Gasparini and L. Gualandri, "Connessioni di Lie e deformazioni su una varietà simplettica piatta," Rend. Mat. Ser. III **1**, 7 (1981); A. Carfagna D'Andrea and R. Di Febo Marinelli, "Sulla rigidità di una connessione secondo Konstant," Rend. Mat. Ser. II **11**, 231 (1978).

[5] G. Reeb, "Sur certaines propriètès topologiques des varièteès feuilletès," Actualités Sci. Indust. (Hermann, Paris, 1952); Ch. Ehresmann, "Structures feuilletées," Proc. 5 Can. Math. 1961.

[6] A. Haefliger, "Varietès feuilletées," Ann. Scuola Norm. Sup. Pisa **16** (1964); H. B. Lawson, Jr. "Foliations," Bull. Am. Math. Soc. **80** (1974).

[7] Transverse bundle parallel along the leaves have been introduced by Conlon and deep results obtained for codimension two foliations: L. Conlon, "Transversally parallelizable foliations of codimension two," Trans. Am. Math. Soc. **194** (1974); R. A. Blumenthal, "Foliated manifolds with flat basic connection," J. Differential Geometry **16** (1981). For transversally complete foliations see P. Molino, "Feuilletages transversalment complets," Ann. Sci. Ec. Norm. Super. **4** (1977). See also Kamber and P. Tondeur, "Foliated bundles and characteristic classes," *Lecture Notes in Mathematics* (Springer, Berlin, 1975), p. 493.

[8] For a very general setting of this problem see H. J. Sussman, "A generalization of the closed subgroup theorem," J. Differential Geometry **10** (1975). Our approach is more geometrical.

[9] A. Lichnerowicz, *Théorie globale des connexions et des groupes d'holonomie* (Cremonese, Rome, 1955); S. Kobayashi and Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963).

[10] I. Satake, "On the generalization of the notion of manifold," Proc. Math. Acad. Sci. USA (1956); J. Girbau and L. Nicolau, "Pseudo differential operators on V-manifolds," Collectanea Math. Vol. XXX (1979).

[11] G. Marmo, E. J. Saletan, and A. Simoni, "A general setting for reduction of dynamical systems," J. Math. Phys. **20**, 856 (1979); R. Hermann and A. J. Krener, "Nonlinear controllability and observability," IEEE Trans. Autom. Control **XXII**, 728 (1977).

The Haar integral for Lie supergroups

D. Williams and J. F. Cornwell
Department of Physics, University of St. Andrews, St. Andrews, Fife, Scotland

(Received 21 December 1983; accepted for publication 16 February 1984)

Each supermanifold can be considered as equivalent to a certain family of real manifolds. The Haar integral of a Lie supergroup is defined using this equivalence. A useful general formula is derived together with explicit construction methods of the simple Lie supergroups of types SPL$(n|m;E_L)$, OSP$(n|2r;E_L)$, and $B(n;E_L)$, and the extended Poincaré group.

PACS numbers: 02.40.Vh, 11.30.Pb, 02.20.Qs

I. INTRODUCTION

The mathematical theory of supermanifolds and supergroups has been developed by many authors following basically two approaches.

(1) The sheaf of algebras of C^∞ functions over a differentiable manifold locally homeomorphic to $\mathbb{R}^n$ is extended to a $\mathbb{Z}_2$-graded sheaf of algebras such that the algebra contains both commuting and anticommuting functions. This line of argument has been pursued by Berezin and Kac,[1] Berezin and Leites,[2] and Kostant.[3] Berezin has in fact considered this and (2).

(2) Commuting and anticommuting coordinates are introduced in the base manifold by making the coordinates copies of either the even or odd parts of a Grassmann algebra. There are many variations in this development, mainly in respect of the topology endowed on the coordinates.

We shall follow the latest development in this second approach that was initiated by Rogers.[4,5] (See these articles for references to earlier work and the relationship between the various approaches. See also the review by Berezin.[6])

Integration on supermanifolds has, up to now, followed the heuristic[6] scheme proposed by Berezin[7] in the context of path integral quantization. This has been taken to be the correct method of integration on supermanifolds constructed as in scheme (2) and is widely used in the physics literature. (See the review by Nieuwenhuizen.[8])

The topology endowed on supermanifolds by Rogers[4] gives us the opportunity to construct a theory of integration based on the normal measure theoretic approach. That is, as for complex and quaternionic manifolds, we view the manifold as a differentiable manifold locally homeomorphic to $\mathbb{R}^n$ with additional Grassmann analytic structure.

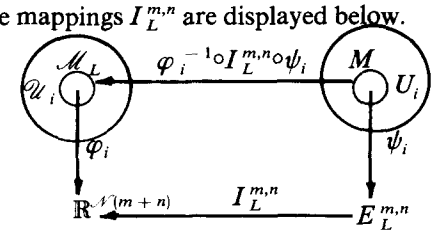
The plan of this article is as follows. In Sec. II we consider measure and integration on supermanifolds. In Sec. III we extend this discussion to cover Lie supergroups. In Secs. IV, V, and VI expressions are obtained for the invariant integrals on the supergroups SPL$(n|m;E_L)$, OSP$(n|2r;E_L)$, and $B(n;E_L)$, respectively. Section VII contains a derivation of the invariant integral for the extended Poincaré groups and for superspace defined as a coset space. In Sec. VIII we give a short discussion of the Berezin integration theory. Our notation is collected, for convenience, in an appendix.

II. MEASURE AND INTEGRATION ON SUPERSPACE AND SUPERMANIFOLDS

Using the topology endowed on superspace $E_L^{m,n}$ and consequently on supermanifolds, Rogers[5] was able to establish the fact that to each real supermanifold M and given value for L there is a topologically equivalent real differentiable manifold $\mathcal{M}_L$. It is convenient therefore to define the topological homeomorphisms $I_L^{m,n}$ as follows.

Definition 2.1: (a) For each L let I_L be the topological homeomorphism $I_L: E_L \to \mathbb{R}^{2^{\mathcal{N}}}$. (b) For each L let $I_L^{m,n}$ be the topological homeomorphism $I_L^{m,n}: E_L^{m,n} \to \mathbb{R}^{2^{\mathcal{N}}(m+n)}$. Clearly we can define the inverses $(I_L^{m,n})^{-1}$.

The mappings $I_L^{m,n}$ are displayed below.



In this figure M is any real supermanifold, $\mathcal{M}_L$ the equivalent real manifold (for fixed L), $\mathcal{U}_i$ is a coordinate neighborhood in $\mathcal{M}_L$ corresponding to the Grassmann coordinate neighborhood U_i with ψ_i, φ_i, respectively, their coordinate charts.

To avoid cumbersome notation we will write simply I_L instead of $I_L^{m,n}$, where it is clear to do so, and also I_L instead of $\varphi_i^{-1} \circ I_L^{m,n} \circ \psi_i$. We can then regard I_L as the map $I_L: M \to \mathcal{M}_L$, for which the inverse I_L^{-1} is defined (since $\mathcal{M}_L$ was constructed from M).

Now to each open subset $A \subset E_L^{m,n}$ there corresponds an open subset $\mathcal{A} = I_L(A) \subset \mathbb{R}^{2^{\mathcal{N}}(m+n)}$. In terms of measure theory (cf. Cohn[9]) the open subsets of $\mathbb{R}^{2^{\mathcal{N}}(m+n)}$ generate the Borel sets of $\mathbb{R}^{2^{\mathcal{N}}(m+n)}$, so that we can construct an equivalence relation between the Borel sets of $E_L^{m,n}$ and the Borel sets of $\mathbb{R}^{2^{\mathcal{N}}(m+n)}$. It is then natural to define the measure of any subset of $E_L^{m,n}$ as the Lebesgue measure of the corresponding subset of $\mathbb{R}^{2^{\mathcal{N}}(m+n)}$. For every function $f: E_L^{m,n} \to E_L$ we have the functions $f \circ I_L^{-1}: \mathbb{R}^{2^{\mathcal{N}}(m+n)} \to E_L$, $I_L \circ f: E_L^{m,n} \to \mathbb{R}^{2^{\mathcal{N}}}$ and $I_L \circ f \circ I_L^{-1}: \mathbb{R}^{2^{\mathcal{N}}(m+n)} \to \mathbb{R}^{2^{\mathcal{N}}}$, which we can regard as alternative descriptions of the original function f. It is

 0022-2488/84/102922-11$02.50

natural, therefore, to define the integral of f to be the integral of the components of $I_L \circ f \circ I_L^{-1}$ rewritten as an element of E_L. Formally the definitions are as follows.

Definition 2.2: (a) The measure μ of a set of $A \subset E_L^{m,n}$ is defined to be the Lebesgue measure μ_L of the set $I_L(A) \subset \mathbb{R}^{\mathcal{N}(m+n)}$, i.e.,

$$\mu(A) = \mu_L(I_L(A)) \in \mathbb{R}.$$

(b) The integral of a function $f: E_L^{m,n} \to \mathbb{R}$ is defined to be the real number obtained in evaluating the integral of $f \circ I_L^{-1}: \mathbb{R}^{\mathcal{N}(m+n)} \to \mathbb{R}$. That is

$$\int f\, d\mu = \int f \circ I_L^{-1}\, d\mu_L \in \mathbb{R}.$$

(c) Let p_i, q_j be the projection functions,

$$p_i: E_L \to \mathbb{R}, \quad q_j: E_L \to \mathbb{R}, \quad i = 0,1,\ldots,\mathcal{N}-1, \quad j = 1,2,\ldots,\mathcal{N}$$

such that p_i (resp., q_j) projects the component in the direction $\mathbf{e}_i$ (resp., $\mathbf{f}_j$), so that $p_k(x+\theta) = p_k(x_i\mathbf{e}_i + \theta_j\mathbf{f}_j) = x_k$ and $q_l(x+\theta) = q_l(x_i\mathbf{e}_i + \theta_j\mathbf{f}_j) = \theta_l$. Then the integral of a function $g: E_L^{m,n} \to E_L$ is defined to be

$$\int g\, d\mu = \sum_{i=0}^{\mathcal{N}-1} \mathbf{e}_i \int p_i \circ g\, d\mu + \sum_{j=1}^{\mathcal{N}} \mathbf{f}_j \int q_j \circ g\, d\mu,$$

so that

$$\int g\, d\mu = \sum_{i=0}^{\mathcal{N}-1} \mathbf{e}_i \int p_i \circ g \circ I_L^{-1}\, d\mu_L + \sum_{j=1}^{\mathcal{N}} \mathbf{f}_j \int q_j \circ g \circ I_L^{-1}\, d\mu_L.$$

The definition of measure and integration for functions $h: E_L^{m,n} \to E_L^{p,q}$ is a straightforward extension of the definition above and need not be given separately. In keeping with the standard coordinate representation on $\mathbb{R}^n$ in which for $x^i \in \mathbb{R}$, $i = 1,2,\ldots,n$, we have $d\mu = \Pi_i\, dx^i$ we will write $\hat{d}x = d\mu$ or $\hat{d}\theta = d\mu$ for the line element in E_{L0} and E_{L1}, respectively, the volume element in $E_L^{m,n}$ can then be written

$$\hat{d}^m x\, \hat{d}^n\theta = \prod_\mu \hat{d}x^\mu \prod_\nu \hat{d}\theta^\nu = \prod_\mu \prod_i dx_i^\mu \prod_\nu \prod_j d\theta_j^\nu,$$

where $i = 0,1,\ldots,\mathcal{N}-1, j = 1,2,\ldots,\mathcal{N}, \mu = 1,2,\ldots,m$, and $\nu = 1,2,\ldots,n$.

We see that the integral, as defined, provides a positive integral on $E_L^{m,n}$ (i.e., $f \geqslant 0$ implies that $\int f \geqslant 0$) and that the set of integrable functions on $E_L^{m,n}$ can be constructed from the set of Lebesgue integrable functions on $\mathbb{R}^{\mathcal{N}(m+n)}$. We will denote this set by $\mathcal{L}^1(E_L^{m,n}, E_L)$. It is clear that we can construct the spaces $\mathcal{L}^p(E_L^{m,n}, E_L)$ for $p \geqslant 1$ modeled on the L^p spaces of normal integration theory. We now give the definition of measure and integration on supermanifolds.

Definition 2.3: Let M be a G^∞ supermanifold over $E_L^{m,n}$, let $\mathcal{M}_L = I_L(M)$ be the related manifold over $\mathbb{R}^{\mathcal{N}(m+n)}$ and let $d\mu_L(\mathcal{M}_L)$ be a volume form on $\mathcal{M}_L$. Then (a) the integral of a function $f: M \to \mathbb{R}$ is given by

$$\int f\, d\mu(M) = \int f \circ I_L^{-1}\, d\mu_L(\mathcal{M}_L)$$

for all $f \circ I_L^{-1} \in \mathcal{L}^1(\mathcal{M}_L, \mathbb{R})$. If $f \circ I_L^{-1} \in \mathcal{L}^1(\mathcal{M}_L, \mathbb{R})$ then $f \in \mathcal{L}^1(M, \mathbb{R})$, where $\mathcal{L}^1(\mathcal{M}_L, \mathbb{R})$ is the set of integrable functions defined on $\mathcal{M}_L$ etc.

(b) The integral of a function $g: M \to E_L$ is given by

$$\int g\, d\mu(M) = \sum_i \mathbf{e}_i \int p_i \circ g \circ I_L^{-1}\, d\mu_L(\mathcal{M}_L) + \sum_j \mathbf{f}_j \int q_j \circ g \circ I_L^{-1}\, d\mu_L(\mathcal{M}_L).$$

Now if $\{x^\mu, \theta^\nu\}$ are coordinate functions on an open set $U_k \subset M$, then $\{x_i^\mu, \theta_j^\nu\}$ are coordinate functions of $I_L(U_k) = \mathcal{U}_k \subset \mathcal{M}_L$ and the volume form can be written

$$d\mu_L(\mathcal{M}_L) = h(x_i, \theta_j) \prod_{\mu,i} dx_i^\mu \prod_{\nu,j} d\theta_j^\nu, \tag{2.1}$$

where $h(x_i, \theta_j): I_L \circ U \to \mathbb{R}$. This is conveniently abbreviated to give

$$d\mu(M) = h(x,\theta)\hat{d}^m x\, \hat{d}^n \theta, \tag{2.2}$$

with $h(x,\theta): U \to \mathbb{R}$, and such that

$$h(x,\theta) = h(x_i, \theta_j). \tag{2.3}$$

Note that h must be a real valued function since we have defined a real valued measure.

Given a set of 1-forms

$$\psi_s^\alpha = \sum_{i=0}^{\mathcal{N}-1} \sum_{\mu=1}^{m} f_{s\mu}^{\alpha i}(x_k, \theta_l) dx_i^\mu + \sum_{i=\mathcal{N}}^{2\mathcal{N}-1} \sum_{\mu=m+1}^{m+n} f_{s\mu}^{\alpha i}(x_k, \theta_l) d\theta_{i-(\mathcal{N}-1)}^{\mu-(m+1)}$$

for $\alpha = 1,2,\ldots,m+n; s = 0,1,\ldots,2\mathcal{N}-1$ and $f_{s\mu}^{\alpha i}: \mathbb{R}_L^{\mathcal{N}(m+n)} \to \mathbb{R}$ on an open subset $\mathcal{U}_\alpha$ of $\mathcal{M}_L$ that are linearly independent, a real-valued volume form on $\mathcal{U}_\alpha$ can be constructed as

$$|\det[f_{s\mu}^{\alpha i}(x_k,\theta_l)]| \prod_{\mu,i} dx_i^\mu \prod_{\nu,j} d\theta_j^\nu = |\det[f_{s\mu}^{\alpha i}(x_k,\theta_l)]|\, \hat{d}^m x\, \hat{d}^n \theta, \tag{2.4}$$

where $f_{s\mu}^{\alpha i}$ is in the (α, s) row and (μ, i) column of the $(m+n)\mathcal{N} \times (m+n)\mathcal{N}$ matrix of coefficients of the one forms. The function $|\det[f_{s\mu}^{\alpha i}(x_k,\theta_l)]|$ will be called the weight function on $\mathcal{U}_\alpha$ (or U_α).

We now have to consider what happens when we need several coordinate charts to cover M. As in normal manifold theory we need simply to insist that the Jacobian of the coordinate change on each $\mathcal{U}_i \cap \mathcal{U}_j$ is equal to 1 so that there is no "change of scale" over the manifold. The Jacobian here, being the one for the charts of $\mathcal{M}_L$. It is not the "Super Jacobian" for M. We assume that we are dealing with orientable manifolds. This means that we define a supermanifold M to be orientable if $\mathcal{M}_L$ is orientable. All Lie supergroups are therefore orientable together with super coset spaces, which are defined in an analogous way to normal coset spaces.

III. INVARIANT INTEGRALS FOR LIE SUPERGROUPS

The theory of invariant (Haar) integrals for topological groups is well known. (See Nachbin[10] or Hewitt and Ross.[11]) Since every Lie supergroup G is equivalent to a Lie group $\mathcal{G}_L$ (for given L) we are thus able, following Sec. II, to define the invariant integral over a Lie supergroup G as equal to that over its related Lie group $\mathcal{G}_L$ (for given L). In this way

the existence of an invariant integral for each supergroup is guaranteed.

Our problem, then, is to learn how to evaluate invariant integrals for this class of Lie group. We will see that for the simple Lie supergroups SPL$(n|m;E_L)$, OSP$(n|2r;E_L)$, $B(n;E_L)$ and certain of their real forms this can be expressed in a very convenient way (see Secs. V, VI, and VII) in terms of the (known) Haar integrals over the simple Lie groups. In this section we examine techniques for evaluating Haar integrals over Lie supergroups in general.

Definition 3.1: If G is a Lie supergroup, $f \in \mathscr{L}^1(G,E_L)$ and $U,V \in G$ the left translation $Uf \in \mathscr{L}^1(G,E_L)$ is defined by $Uf(V) = f(U^{-1}V)$. Similarly the right translation is defined by $f(V)U = f(VU)$. Clearly if e is the identity of G then $ef = f = fe$. Also $U(Vf) = UV(f)$, $(fU)V = (f)UV$ and $(Uf)V = U(fV)$ for each $U,V \in G$.

Definition 3.2: (a) A positive integral on a Lie supergroup G is said to be left invariant if for every $f \in \mathscr{L}^1(G,E_L)$ and $U,V \in G$ we have

$$\int f(U^{-1}V)d\mu(V) = \int f(V)d\mu(V). \tag{3.1}$$

(b) Similarly a positive integral is right invariant if

$$\int f(VU)d\mu(V) = \int f(V)d\mu(V). \tag{3.2}$$

(c) A supergroup for which the left and right invariant integrals are equal is said to be unimodular.

In particular we note that if a subset $A \subset G$ is measurable and $\mu(A)$ is left and right invariant then,

$$\mu(UA) = \mu(A) = \mu(AV) \in \mathbb{R}_+ \tag{3.3}$$

for all $U,V \in G$.

Definition 3.3: A semisimple Lie supergroup is a Lie supergroup with a semisimple Lie superalgebra (see Kac[12]).

Proposition 3.4: The left and right invariant integrals over a semisimple Lie supergroup are identical for each L.

Proof: If a Lie supergroup G is semisimple it is equal to its closed commutator subgroup. It follows that this is true for each related Lie group $\mathscr{G}_L$. Then by Proposition 15 of Ref. 10 (p. 83), $\mathscr{G}_L$ is unimodular. Hence G is unimodular.

We can also see that any connected Lie supergroup considered as a Lie group is the semidirect product of two components, one of which corresponds to the Lie subalgebra of the Lie superalgebra the other being a connected nilpotent group. By Proposition 1.4 of Helgason (Ref. 13, p. 366), the nilpotent part is always unimodular. Now a semidirect product of two groups is unimodular if both groups are separately unimodular. [See Hewitt and Ross[11] p. 210, Sec. (15.29). The fact that the functional $\delta(h)$ is equal to 1 is most easily seen from the statement in Sec. (15.23) of the same book.] Thus a Lie supergroup is unimodular if the Lie group corresponding to the Lie subalgebra ($\mathscr{L}_0$) of its Lie superalgebra is unimodular. In particular, since the Poincaré group is unimodular, the extended Poincaré group is also unimodular. (It also coincides with its closed commutator subgroup.)

We note that since the Lebesgue integral is a Haar integral for the additive group $\mathbb{R}^n$ (see Nachbin[10]) we already have one example of an invariant integral for a supergroup, i.e., the additive group $E_L^{m,n}$.

We now narrow our attention to linear Lie supergroups which are defined as follows.

Definition 3.5: An $(m|n)$-dimensional linear Lie supergroup G is defined to be a Lie supergroup with a faithful representation by matrices from $M(p|q;E_L)$ depending on m even and n odd parameters. The group operation is given by matrix multiplication.

Since $x \in E_{L0}(\theta \in E_{L1})$ can be written $x = x_i\mathbf{e}_i(\theta = \theta_j\mathbf{f}_j)$ and the $\mathbf{e}_i$'s($\mathbf{f}_j$'s) can be manipulated as matrices (matrix representations obeying normal matrix multiplication can in fact be constructed for each L), we can deduce that, if G is a linear Lie supergroup then $\mathscr{G}_L$ is a linear Lie group for each L. One can therefore consider the matrix representation of G as a "coded" matrix representation of $\mathscr{G}_L$.

As with linear Lie groups we do not insist that a linear Lie supergroup requires only one chart. We require simply that if several charts are required then each chart can be expressed as a set of matrices.

Now we further restrict our attention to left invariant integrals. It is possible to carry through the following procedure for right invariant integrals but all groups of interest are unimodular so that this is unnecessary.

In Sec. II, we stated how the weight function is obtained from a linearly independent set of 1-forms on a manifold. There is a well-known prescription for finding left invariant 1-forms (sometimes called Maurer–Cartan 1-forms) on a linear Lie group (cf. Chevalley[14]), which is conveniently expressed in the following lemma.

Lemma 3.6: Let $\mathscr{U}$ be a matrix representation of a chart containing the identity e of a linear Lie group $\mathscr{G}$ parametrized by $\{x^\mu\}$, $x^\mu \in \mathbb{R}$. Let $a^\sigma = (\partial\mathscr{U}/\partial x^\sigma)_e$. Evaluate the matrix elements A_σ^μ from the equations

$$\mathscr{U}^{-1}\frac{\partial\mathscr{U}}{\partial x^\mu} = A_\sigma^\mu a^\sigma, \quad \mu = 1,2,... \tag{3.4}$$

The left invariant 1-forms w^μ are then given by $w^\mu = A_\sigma^\mu dx^\sigma$ in the chart $\mathscr{U}$, and a left invariant weight function is given by $|\det[A_\sigma^\mu]|$.

Theorem 3.7: Let G be an $(m|n)$-dimensional linear Lie group and let U be a matrix representation of a chart containing the identity $e \in G$ expressed in terms of the parameters $\{x^\mu,\theta^\nu\}$. Let $\alpha_\mu = (\partial U/\partial x^\mu)_e$ and $\beta_\nu = (\partial U/\partial\theta^\nu)_e$. Suppose further that if U_α, $\alpha = 1,2,...$, is an atlas for G the Jacobian of coordinate change on each $U_\alpha \cap U_\beta$ is equal to one. The left invariant weight function $\mathscr{W}_G$ can then be obtained from the equations

$$U^{-1}\frac{\partial U}{\partial x^\mu} = a_\mu^\sigma\alpha_\sigma + b_\mu^\rho\beta_\rho$$

and

$$U^{-1}\frac{\partial U}{\partial\theta^\nu} = c_\nu^\sigma\alpha_\sigma + d_\nu^\rho\beta_\rho$$

as

$$\mathscr{W}_G = |(\det[\,p_0(a_\mu^\sigma(x_0^\mu,0))]\det[p_0(d_\nu^\rho(x_0^\mu,0))])^{\mathscr{N}}|.$$

Proof: Consider first the equation

$$U^{-1}\frac{\partial U}{\partial x^\mu} = a_\mu^\sigma\alpha_\sigma + b_\mu^\rho\beta_\rho.$$

Multiply by $\mathbf{e}_i$ on the left, then recall from the Appendix that $aX = (-1)^{|a|\,|X|}Xa$ for $a \in E_L$, $X \in \tilde{M}(p|q;E_L)$ to obtain for each i

$$U^{-1}\mathbf{e}_i \frac{\partial U}{\partial x^\mu} = a^\sigma_\mu \mathbf{e}_i \alpha_\sigma + b^\rho_\mu \mathbf{e}_i \beta^\rho.$$

Now expand a^σ_μ and b^ρ_μ in terms of the projection functions p_k and q_l [see Definition 2.2(c)] noting that by construction $|a| = 0$ and $|b| = 1$, and[15] that $\mathbf{e}_i(\partial U/\partial x^\mu) = (\partial U/\partial x^\mu_i)$ to obtain

$$U^{-1}\frac{\partial U}{\partial x^\mu_i} = p_0(a^\sigma_\mu)\mathbf{e}_i\alpha_\sigma + p_k(a^\sigma_\mu)\mathbf{e}_k\mathbf{e}_i\alpha_\sigma + q_j(b^\rho_\mu)\mathbf{f}_j\mathbf{e}_i\beta_\rho,$$

with $k = 1,2,\ldots,\mathcal{N}-1$ and $j = 1,2,\ldots,\mathcal{N}$.

If we now consider the Taylor expansion of a^σ_μ in all the real variables x^η_i, θ^ξ_j, $i \neq 0$ it is clear that $p_0(a^\sigma_\mu(x,\theta)) = p_0(a^\sigma_\mu(x_0,0))$ so that

$$U^{-1}\frac{\partial U}{\partial x^\mu_i} = p_0(a^\sigma_\mu(x_0,0))\mathbf{e}_i\alpha_\sigma + p_k(a^\sigma_\mu)\mathbf{e}_k\mathbf{e}_i\alpha_\sigma + q_j(b^\rho_\mu)\mathbf{f}_j\mathbf{e}_i\beta_\rho. \tag{3.5}$$

By a similar calculation we obtain

$$U^{-1}\frac{\partial U}{\partial \theta^\nu_j} = -q_l(c^\sigma_\nu)\mathbf{f}_l\mathbf{f}_j\alpha_\sigma + p_0(d^{\,l}_\nu(x_0,0))\mathbf{f}_j\beta_\rho + p_k(d^\rho_\nu)\mathbf{e}_k\mathbf{f}_j\beta_\rho. \tag{3.6}$$

Suppose the left invariant 1-forms on $\mathcal{G}_L$ written as a matrix equation take the form

$$\begin{bmatrix} w^\mu_i \\ \omega^\nu_j \end{bmatrix} = \begin{bmatrix} P & Q \\ R & S \end{bmatrix} \begin{bmatrix} dx^\sigma_k \\ d\theta^\rho_l \end{bmatrix}$$

with the indices ordered such that

$$[dx^\sigma_k, d\theta^\rho_l]^t = [dx^1_0,\ldots,dx^m_0,dx^1_1,\ldots,dx^m_1,\ldots,dx^m_{\mathcal{N}-1},d\theta^1_1,\ldots,d\theta^n_1,d\theta^1_2,\ldots,d\theta^n_{\mathcal{N}}]^t$$

and the indices in the matrix $[w^\mu_i,\omega^\nu_j]^t$ having the same ordering. A comparison of Lemma 3.5 with Eqs. (3.5) and (3.6) (assuming the same index ordering) then shows that the left invariant weight function for G can be written as

$$\mathscr{W}_G = |\det \left[\begin{array}{ccccccc|ccccccc} p_{11} & p_{12} & & & & & & q_{11} & q_{12} & & & & & \\ 0 & p_{22} & p_{23} & & p_{ij} & & & 0 & q_{22} & & & q_{ij} & & \\ & 0 & p_{33} & & & & & & 0 & & & & & \\ & & 0 & & & & & & & 0 & & & & \\ & & & 0 & & & & & & & 0 & & & \\ & 0 & & & 0 & & & & & & & 0 & & \\ & & & & & 0 & & & 0 & & & & 0 & \\ & & & & & & 0\ p_{\mathcal{N}\mathcal{N}} & & & & & & & 0\ q_{\mathcal{N}\mathcal{N}} \\ \hline 0 & r_{12} & r_{13} & & & & & s_{11} & s_{12} & & & & & \\ & 0 & r_{23} & & r_{ij} & & & 0 & s_{22} & & & s_{ij} & & \\ & & 0 & & & & & & 0 & & & & & \\ & & & 0 & & & & & & 0 & & & & \\ & & & & 0 & & & & & & 0 & & & \\ & 0 & & & 0 & & & & & & & 0 & & \\ & & & & & 0 & & & 0 & & & & 0 & \\ & & & & & & 0\ r_{\mathcal{N}-1,\mathcal{N}} & & & & & & & 0\ s_{\mathcal{N}\mathcal{N}} \end{array}\right]|,$$

where p_{ij}, q_{ij}, r_{ij}, and s_{ij} are $m\times n$, $n\times n$, $m\times n$, and $n\times m$ submatrices, such that

$$(p_{ii})^\sigma_\mu = p_0(a^\sigma_\mu(x_0,0)), \quad i = 0,\ldots,\mathcal{N}-1;\ \sigma,\mu = 1,\ldots,m,$$

$$(s_{ii})^\rho_\nu = p_0(d^\rho_\mu(x_0,0)), \quad i = 1,\ldots,\mathcal{N};\ \rho,\mu = 1,\ldots,n,$$

and where p_{ij}, q_{ij}, s_{ij} are zero for $i > j$ and r_{ij} is zero for $i \geqslant j$. The values of the elements of the other submatrices are irrelevant, since with this structure $\mathscr{W}_G = |(\Pi_{i=0}^{\mathcal{N}-1}\det p_{ii})(\Pi_{j=1}^{\mathcal{N}}\det s_{ii})|$.

Thus,

$$\mathscr{W}_G = |(\det[\,p_0(a^\sigma_\mu(x_0,0))\,]\det[\,p_0(d^\rho_\nu(x_0,0))\,])^{\mathcal{N}}|.$$

To extend this to a Lie supergroup G containing several charts there are two cases to consider: (1) G is one connected component, (2) G has several components.

In the first case, the answer for real manifolds is to choose the Jacobian determinant on the overlapping regions to have value 1. We can achieve this in the case of a Lie supergroup by insisting that $\det[\,p_0(\partial(x^\mu,\theta^\nu)/\partial(y^\sigma,\phi^\rho))\,] = 1$ on each $U_\alpha \cap U_\beta$, where $(\partial(x^\mu,\theta^\nu)/\partial(y^\sigma,\phi^\rho))$ denotes the super Jacobian matrix. The weight function is then the same in each chart, that is, for each U_α, $\mathscr{W}_{U_\alpha} = \mathscr{W}_G((x^\mu,\theta^\nu)_\alpha)$. Note that requiring that the superdeterminant of the coordinate change is equal to 1 does not achieve this result.

In the second case, it is clear that each component of a many component group is a topological copy of the component connected to the identity. So that we just need to use this fact to obtain an integral over the whole group.

IV. LEFT INVARIANT INTEGRAL FOR THE LIE SUPERGROUP SPL($n|m;E_L$)

It is clear from the work of Rogers[5] that Lie supergroups can be constructed by starting with a real Lie superalgebra $L = L_0 \oplus L_1$ and forming the "graded tensor product" $E_{L0} \otimes L_0 \oplus E_{L1} \otimes L_1$, where L_0, L_1 are the even, respectively, odd parts of the superalgebra. This then forms a Lie module and thus defines a Lie supergroup (or possibly several Lie supergroups). This procedure was proposed by Rittenberg and Scheunert,[16] as a formal algebraic construction. That is they defined the "group" as the exponential of the algebra and did not impose any topology on the resulting "group." (Note that the supergroups in this paper that use the conjugation operations are not Lie supergroups according to the definition of Rogers.)

To carry out this construction we need to use the exponential function for G. This is straightforward to obtain from the exponential function defined on $\mathcal{G}_L$, and it does, in fact, satisfy all the properties endowed on it in the literature so that we will not detail it here.

The simple Lie superalgebras have been classified (see Kac[12] and Scheunert[17]) and their real forms determined by Kac,[12] and by Parker[18] who gives construction methods. It is thus straightforward to define real simple Lie supergroups, and since we can always construct faithful matrix representations for the superalgebras which obey the commutator $[X,Y] = XY - (-1)^{|X||Y|} YX$ (noting that in some cases it is necessary to use the adjoint representation for this, e.g., spl$(n|n)/I_{2n}$), it is clear that these will be linear Lie supergroups.

For terminology for the superalgebras, we follow Scheunert so that we can consistently use small latin letters to denote the algebras and Lie modules, and large latin letters for the corresponding groups. The Lie module obtained from a Lie superalgebra will be denoted as above, i.e., spl$(n|m;\mathbb{R})$ is a real superalgebra for which spl$(n|m;E_L)$ is the corresponding Lie module.

The supergroup SPL$(n|m;E_L)$ is defined by

$$\mathrm{SPL}(n|m;E_L) = \left\{ u = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \quad \mathrm{sdet}\, U = 1 \right\} \tag{4.1}$$

and its Lie module is

$$\mathrm{spl}(n|m;E_L) = \left\{ X = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \ \mathrm{str}\, X = 0 \right\}. \tag{4.2}$$

Here A,a are $n \times n$ matrices with entries from E_{L0}; B,b are $n \times m$ matrices with entries from E_{L1}; C,c are $m \times n$ matrices with entries from E_{L1}; and D,d are $m \times m$ matrices with entries from E_{L0}. sdet and str have their usual definitions if we make the identification $\epsilon_\phi = e_1 \simeq 1 \in \mathbb{R}$.

It is well known that spl$(n|m;\mathbb{R})$ and hence, spl$(n|m;E_L)$ also has a $\mathbb{Z}$ grading such that we can write

$$\begin{aligned} X &= X_{-1} \oplus X_0 \oplus X_1 \\ &= \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} \oplus \begin{bmatrix} 0 & 0 \\ c & 0 \end{bmatrix}. \end{aligned}$$

Clearly each X_i, $i = -1,0,1$ is a submodule of spl$(n|m;E_L)$.

The supergroup SPL$(n|m;E_L)$ can be decomposed in a similar way to give

$$\begin{aligned} U = U_{-1}U_0U_1 &= \begin{bmatrix} A'' & B' \\ C' & D' \end{bmatrix} \\ &= \begin{bmatrix} I & B'D'^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A'' - B'D'^{-1}C' & 0 \\ 0 & D' \end{bmatrix} \begin{bmatrix} I & 0 \\ D'^{-1}C & I \end{bmatrix} \\ &= \begin{bmatrix} I & B \\ 0 & I \end{bmatrix} \begin{bmatrix} A' & 0 \\ 0 & D' \end{bmatrix} \begin{bmatrix} I & 0 \\ C & I \end{bmatrix}, \end{aligned} \tag{4.3}$$

where the last line is simply a reparametrization. We can now identify the subgroups U_{-1}, U_0, U_1, with the submodules X_{-1}, X_0, X_1, respectively, and so construct the supergroup in a very straightforward way. Note that X_{-1}, X_1 are abelian and $\exp(X_i) = I + X_i = U_i$, $i = -1,1$.

Now recall that the subgroup U_0 has a one-dimensional invariant subgroup which for our purposes is best factored out to give

$$U = \begin{bmatrix} I & B \\ 0 & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} e^{y/n} I_n & 0 \\ 0 & e^{y/m} I_m \end{bmatrix} \begin{bmatrix} I & 0 \\ C & I \end{bmatrix} \tag{4.4}$$

with $\det A = \det D = 1$, so that

$$\begin{aligned} U^{-1} = &\begin{bmatrix} I & 0 \\ -C & I \end{bmatrix} \begin{bmatrix} e^{-y/n} I_n & 0 \\ 0 & e^{-y/m} I_m \end{bmatrix} \\ &\times \begin{bmatrix} A^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} I & -B \\ 0 & I \end{bmatrix}. \end{aligned} \tag{4.5}$$

In this form the block matrices A,B,C,D depend on disjoint sets of parameters and $y \in E_{L0}$. If $n = m$ then the supergroup is no longer simple, the invariant subgroup of U_0 is a multiple of the identity and hence an invariant subgroup of the whole supergroup. It is normal practice in this case to factor it out to obtain a simple supergroup. But this leads to a group composition rule also involving the factorization unless the adjoint representation is constructed. To find a Haar integral in this case it is better to leave the ideal in, then construct the required supergroup as a coset space. This decomposition can be carried out for the real forms of spl$(n|m;\mathbb{C})$ with $L_0 = \mathrm{su}^*(n) \oplus \mathrm{su}^*(m) \oplus \mathbb{R}$ but not for those with real form $L_0 = \mathrm{su}(p,q) \oplus \mathrm{su}(r,s) \oplus i\mathbb{R}$.

Denote the parameters of the matrix blocks A,B,C,D by $x^\alpha, \theta^\beta, \theta^\gamma, x^\delta$, respectively. Consider first a parameter from the set $\{x^\alpha\}$. We then have

$$U^{-1} \frac{\partial U}{\partial x^\alpha} = \begin{bmatrix} A^{-1} \dfrac{\partial A}{\partial x^\alpha} & 0 \\ -CA^{-1} \dfrac{\partial A}{\partial x^\alpha} & 0 \end{bmatrix},$$

similarly,

$$U^{-1} \frac{\partial U}{\partial \theta^\beta} = \begin{bmatrix} A^{-1} \dfrac{\partial B}{\partial \theta^\beta} DC & A^{-1} \dfrac{\partial B}{\partial \theta^\beta} D \\ -CA^{-1} \dfrac{\partial B}{\partial \theta^\beta} & -CA^{-1} \dfrac{\partial B}{\partial \theta^\beta} D \end{bmatrix},$$

$$U^{-1} \frac{\partial U}{\partial \theta^\gamma} = \begin{bmatrix} 0 & 0 \\ \dfrac{\partial C}{\partial \theta^\gamma} & 0 \end{bmatrix},$$

$$U^{-1} \frac{\partial U}{\partial x^\delta} = \begin{bmatrix} 0 & 0 \\ D^{-1} \dfrac{\partial D}{\partial x^\delta} C & D^{-1} \dfrac{\partial D}{\partial x^\delta} \end{bmatrix},$$

and

$$U^{-1}\frac{\partial U}{\partial y}=\begin{bmatrix} I_n & 0 \\ 0 & I_m \end{bmatrix}.$$

Now applying Theorem 3.7, the weight function of $G=\mathrm{SPL}(n|m;E_L)$ is given by

$$\begin{aligned}\mathscr{W}_G &= \left(\left|\det\left[\left\{A^{-1}\frac{\partial A}{\partial x^\alpha}\right\}(x_0^\alpha,0)\right]\det\left[\left\{D^{-1}\frac{\partial D}{\partial x^\delta}\right\}(x_0^\delta,0)\right]\right.\right.\\ &\quad\times\det\left[\left\{\frac{\partial C}{\partial\theta^\gamma}\right\}(0,0)\right]\\ &\quad\left.\left.\times\det\left[\left\{A^{-1}\frac{\partial B}{\partial\theta^\beta}D\right\}(x_0^\alpha,x_0^\delta,0)\right]\right|\right)^{\mathscr{N}}\\ &=\mathscr{W}_A\mathscr{W}_D\mathscr{W}_C\left(\left|\det\left[\left\{A^{-1}\frac{\partial B}{\partial\theta^\beta}D\right\}(x_0^\alpha,x_0^\delta,0)\right]\right|\right)^{\mathscr{N}},\end{aligned}$$

where we have written $\mathscr{W}_A=\mathscr{W}_{\begin{bmatrix}A&0\\0&I\end{bmatrix}}$ etc. and the expression $\{A^{-1}(\partial A/\partial x^\alpha)\}$ means the matrix of coefficients obtained when $A^{-1}(\partial A/\partial x^\alpha)$ is expanded in terms of the superalgebra of A, etc.

Clearly we can choose a parametrization such that $\mathscr{W}_C=1$ and $\mathscr{W}_A,\mathscr{W}_B$ are the weight functions for $\mathrm{SL}(n;E_L)$ and $\mathrm{SL}(m;E_L)$, respectively, i.e., the Lie supergroups obtained by extending the domain of the parameters of $\mathrm{SL}(n;\mathbb{R})$ from $\mathbb{R}$ to E_{L0} and $\mathscr{W}_{\mathrm{SL}(n;E_L)}=(\mathscr{W}_{\mathrm{SL}(n,\mathbb{R})})^{\mathscr{N}}$. Now B is an arbitrary matrix with entries from E_{L1} so that we can choose its parametrization to be such that $[B]^{ij}=\theta^{ij}$, for $i=1,..,n, j=1,..,m$, and $\theta^{ij}\in E_{L1}$. Then $\partial B/\partial\theta^{ij}$ is independent of the position it is evaluated at and it is a routine calculation to show that $|\det[\{A^{-1}(\partial B/\partial\theta^\beta)D\}]|=1$. The weight function for $\mathrm{SPL}(n|m;E_L)$ can then be written

$$\mathscr{W}_{\mathrm{SPL}(n|m;E_L)}=(\mathscr{W}_{\mathrm{SL}(n;\mathbb{R})}\mathscr{W}_{\mathrm{SL}(m;\mathbb{R})})^{\mathscr{N}}.$$

V. LEFT INVARIANT INTEGRAL FOR THE LIESUPERGROUP OSP($n|2r;E_L$).

This section closely follows the previous one. The Lie superalgebra has a $\mathbb{Z}$ grading into three subsets only in the special case of $\mathrm{osp}(2|2r;\mathbb{R})$. We treat the general case in which there is a decomposition into five subsets (cf. Kac[12]) as $X=X_{-2}\oplus X_{-1}\oplus X_0\oplus X_1\oplus X_2$ such that $\{X_{-2}\oplus X_{-1}\}$, $\{X_0\}$, and $\{X_1\oplus X_2\}$ are subalgebras.

Our definition of $\mathrm{osp}(n|2r;\mathbb{R})$ is given by

$$\mathrm{osp}(n|2r;\mathbb{R})=\{X\in\mathrm{spl}(n|2r;\mathbb{R}),X^{st}h+(-1)^{|X|}hX=0\}\quad(5.1)$$

The Lie module is then defined by

$$\mathrm{osp}(n|2r;E_L)=\{X\in\mathrm{spl}(n|2r;E_L),X^{st}h+(-1)^{|X|}hX=0\},\quad(5.2)$$

where

$$h=\begin{bmatrix} J & 0 & 0\\ 0 & 0 & I_r\\ 0 & -I_r & 0\end{bmatrix}$$

and J is any symmetric, non singular matrix.

Our definition of supertranspose is given by

$$X^{st}=\begin{bmatrix}A & B\\ C & D\end{bmatrix}^{st}=\begin{bmatrix}A^t & (-1)^{|X|}C^t\\ (-1)^{|X|+1}B^t & D^t\end{bmatrix}.\quad(5.3)$$

These definitions differ from those often given.[16,19,20] We formulate them in this way to allow consistency throughout the supermatrix algebra, in which we follow Leites,[21] Eq. (5.2) above is also immediately applicable to the whole of the Lie supermodule.

If we put $J=I_n$, a general element of $\mathrm{osp}(n|2r;E_L)$ can be written

$$X=\begin{bmatrix} a & y & z\\ -z^t & b & c\\ y^t & d & -b^t\end{bmatrix}\quad(5.4)$$

with $a\in\mathrm{so}(n;E_L)$, $b\in\mathrm{gl}(r;E_L)$, c and d symmetric with entries from E_{L0} and y and z having entries from E_{L1}. We then have the decomposition

$$\begin{aligned}X&=X_{-2}\oplus X_{-1}\oplus X_0\oplus X_1\oplus X_2\\ &=\begin{bmatrix}0&0&0\\0&0&0\\0&d&0\end{bmatrix}\oplus\begin{bmatrix}0&y&0\\0&0&0\\y^t&0&0\end{bmatrix}\oplus\begin{bmatrix}a&0&0\\0&b&0\\0&0&-b^t\end{bmatrix}\oplus\begin{bmatrix}0&0&z\\-z^t&0&0\\0&0&0\end{bmatrix}\oplus\begin{bmatrix}0&0&0\\0&0&c\\0&0&0\end{bmatrix}\end{aligned}\quad(5.5)$$

with the sets $\{X_{-2}\oplus X_{-1}\}$, $\{X_0\}$, and $\{X_1\oplus X_2\}$ as subalgebras. We can exponentiate each of these separately and construct a group element

$$\begin{aligned}U&=\exp[X_{-2}\oplus X_{-1}]\exp[X_0]\exp[X_1\oplus X_2]\\ &=\begin{bmatrix}I_n & Y & 0\\ 0 & I_r & 0\\ Y^t & D+Y^tY/2 & I_r\end{bmatrix}\begin{bmatrix}A\in\mathrm{SO}(n;E_{L0}) & 0 & 0\\ 0 & B\in\mathrm{GL}_e(r;E_{L0}) & 0\\ 0 & 0 & (B^t)^{-1}\end{bmatrix}\begin{bmatrix}I_n & 0 & Z\\ -Z^t & I_r & C-Z^tZ/2\\ 0 & 0 & I_r\end{bmatrix}.\end{aligned}\quad(5.6)$$

We note that

$$U^{-1}=\begin{bmatrix}I_n & 0 & -Z\\ Z^t & I_r & -C-Z^tZ/2\\ 0 & 0 & I_r\end{bmatrix}\begin{bmatrix}A^{-1} & 0 & 0\\ 0 & B^{-1} & 0\\ 0 & 0 & B^t\end{bmatrix}\begin{bmatrix}I_n & -Y & 0\\ 0 & I_r & 0\\ Y^t-D+Y^tY/2 & I_r\end{bmatrix}.\quad(5.7)$$

Here Z and Y are conveniently parametrized by a different odd parameter (θ^ν) in each position and C and D are parametrized by a different even parameter (x^μ) at each position. $\mathrm{GL}_e(r;E_L)$ is the component of $\mathrm{GL}(r;E_L)$ connected to the identity.

This construction certainly meets the condition normally given for the orthosymplectic group[16] [see Eq. (5.9) below], and in fact meets the more restrictive definition

$$\mathrm{SOSP}(n|2r;E_L) = \{\, U \in \mathrm{SPL}(n|2r;E_L), U^{st}hU = h \,\}. \tag{5.8}$$

It is clear that the only element in common between the three subgroups is the identity so that the representation is faithful and that the product of any two matrices of this form can again be decomposed into this form; so we can deduce that this construction does give us the whole group, provided we choose a suitable parametrization for the even parts.

The definition often given[16,19,20] for the orthosymplectic groups

$$\mathrm{OSP}(n|2r;E_L) = \{\, U \in PL\,(n|2r;E_L), U^{st}hU = h \,\} \tag{5.9}$$

is clearly obtained from the above by replacing $\mathrm{SO}(n;E_L)$ by $\mathrm{O}(n;E_L)$ and $\mathrm{GL}_e(r;E_L)$ by $\mathrm{GL}(r;E_L)$.

Now denote the parameters of the blocks A, B, C, D, Y, and Z by $x^\alpha, x^\beta, x^\gamma, x^\delta, \theta^\xi$, and θ^ξ, respectively, and using the procedures of the previous section we obtain

$$U^{-1}\frac{\partial U}{\partial x^\alpha} = \begin{bmatrix} A^{-1}\dfrac{\partial A}{\partial x^\alpha} & 0 & A^{-1}\dfrac{\partial A}{\partial x^\alpha}Z \\ -Z^t A^{-1}\dfrac{\partial A}{\partial x^\alpha} & 0 & -Z^t A^{-1}\dfrac{\partial A}{\partial x^\alpha}Z \\ 0 & 0 & 0 \end{bmatrix}$$

$$U^{-1}\frac{\partial U}{\partial x^\beta} = \begin{bmatrix} 0 & 0 & -ZB^t\dfrac{\partial (B^t)^{-1}}{\partial x^\beta} \\ B^{-1}\dfrac{\partial B}{\partial x^\beta}Z^t & B^{-1}\dfrac{\partial B}{\partial x^\beta} & B^{-1}\dfrac{\partial B}{\partial x^\beta}\left(C+\dfrac{Z^tZ}{2}\right) \\ & & +\left(-C+\dfrac{Z^tZ}{2}\right)B^t\dfrac{\partial (B^t)^{-1}}{\partial x^\beta} \\ 0 & 0 & B^t\dfrac{\partial (B^t)^{-1}}{\partial x^\beta} \end{bmatrix},$$

$$U^{-1}\frac{\partial U}{\partial x^\gamma} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \dfrac{\partial C}{\partial x^\gamma} \\ 0 & 0 & 0 \end{bmatrix},$$

$$U^{-1}\frac{\partial U}{\partial x^\delta} = \begin{bmatrix} -ZB^t\dfrac{\partial D}{\partial x^\delta}BZ^t & -ZB^t\dfrac{\partial D}{\partial x^\delta}B & -ZB^t\dfrac{\partial D}{\partial x^\delta}B\left(C+\dfrac{Z^tZ}{2}\right) \\ \left(-C+\dfrac{Z^tZ}{2}\right)B^t\dfrac{\partial D}{\partial x^\delta}BZ^t & \left(-C+\dfrac{Z^tZ}{2}\right)B^t\dfrac{\partial D}{\partial x^\delta}B & \left(-C+\dfrac{Z^tZ}{2}\right)B^t\dfrac{\partial D}{\partial x^\delta} \\ & & \times B\left(C+\dfrac{Z^tZ}{2}\right) \\ B^t\dfrac{\partial D}{\partial x^\delta}BZ^t & B^t\dfrac{\partial D}{\partial x^\delta}B & B^t\dfrac{\partial D}{\partial x^\delta}(B^t)^{-1} \end{bmatrix},$$

$$U^{-1}\frac{\partial U}{\partial \theta^\xi} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

with

$$a_{11} = -ZB^t\frac{\partial Y^t}{\partial \theta^\xi}A + A^{-1}\frac{\partial Y}{\partial \theta^\xi}AZ + ZB^t\left(\frac{Y^t}{2}\frac{\partial Y}{\partial \theta^\xi} - \frac{\partial Y^t}{\partial \theta^\xi}\frac{Y}{2}\right)BZ^t,$$

$$a_{12} = A^{-1}\frac{\partial Y}{\partial \theta^\xi}B - ZB^t\left(\frac{Y^t}{2}\frac{\partial Y}{\partial \theta^\xi} - \frac{\partial Y^t}{\partial \theta^\xi}\frac{Y}{2}\right)B,$$

$$a_{13} = -ZB^t\frac{\partial Y^t}{\partial \theta^\xi}AZ + A^{-1}\frac{\partial Y}{\partial \theta^\xi}B\left(C+\frac{Z^tZ}{2}\right) - ZB^t\left(\frac{Y^t}{2}\frac{\partial Y}{\partial \theta^\xi} - \frac{\partial Y^t}{\partial \theta^\xi}\frac{Y}{2}\right)B\left(C+\frac{Z^tZ}{2}\right),$$

$$a_{21} = -\left(-C+\frac{Z^tZ}{2}\right)B^t\frac{\partial Y^t}{\partial\theta^\xi}A - Z^tA^{-1}\frac{\partial Y}{\partial\theta^\xi}BZ^t$$
$$+\left(-C+\frac{Z^tZ}{2}\right)B^t\left(\frac{Y^t}{2}\frac{\partial Y}{\partial\theta^\xi}-\frac{\partial Y^t}{\partial\theta^\xi}\frac{Y}{2}\right)BZ^t,$$
$$a_{22} = -Z^tA^{-1}\frac{\partial Y}{\partial\theta^\xi}B+\left(-C+\frac{Z^tZ}{2}\right)B^t\left(\frac{Y^t}{2}\frac{\partial Y}{\partial\theta^\xi}-\frac{\partial Y^t}{\partial\theta^\xi}\frac{Y}{2}\right)B,$$
$$a_{23} = -\left(-C+\frac{Z^tZ}{2}\right)B^t\frac{\partial Y^t}{\partial\theta^\xi}AZ - Z^tA^{-1}\frac{\partial Y}{\partial\theta^\xi}B\left(C+\frac{Z^tZ}{2}\right)$$
$$+\left(-C+\frac{Z^tZ}{2}\right)B^t\left(\frac{Y^t}{2}\frac{\partial Y}{\partial\theta^\xi}-\frac{\partial Y^t}{\partial\theta^\xi}\frac{Y}{2}\right)B\left(C+\frac{Z^tZ}{2}\right),$$
$$a_{31} = -B^t\frac{\partial Y^t}{\partial\theta^\xi}A+B^t\left(\frac{Y^t}{2}\frac{\partial Y}{\partial\theta^\xi}-\frac{\partial Y^t}{\partial\theta^\xi}\frac{Y}{2}\right)BZ^t,$$
$$a_{32} = B^t\left(\frac{Y^t}{2}\frac{\partial Y}{\partial\theta^\xi}-\frac{\partial Y^t}{\partial\theta^\xi}\frac{Y}{2}\right)B, \quad \text{and}$$
$$a_{33} = -B^t\frac{\partial Y^t}{\partial\theta^\xi}AZ+B^t\left(\frac{Y^t}{2}\frac{\partial Y}{\partial\theta^\xi}-\frac{\partial Y}{\partial\theta^\xi}\frac{Y}{2}\right)B\left(C+\frac{Z^tZ}{2}\right),$$

$$\text{and}\quad U^{-1}\frac{\partial U}{\partial\theta^\zeta} = \begin{bmatrix} 0 & 0 & \dfrac{\partial Z}{\partial\theta^\zeta} \\ \dfrac{\partial Z^t}{\partial\theta^\zeta} & 0 & \dfrac{Z^t}{2}\dfrac{\partial Z}{\partial\theta^\zeta}+\dfrac{\partial Z^t}{\partial\theta^\zeta}\dfrac{Z}{2} \\ 0 & 0 & 0 \end{bmatrix}.$$

The fact that each of these can be expressed in terms of the Lie superalgebra is readily checked.

Now as for $\mathrm{SPL}(n|m;E_L)$ we can use Theorem 3.7 to obtain for $G=\mathrm{SOSP}(n|m;E_L)$

$$\mathscr{W}_G = \left(\left|\det\left[\left\{A^{-1}\frac{\partial A}{\partial x^\alpha}\right\}(x_0^\alpha,0)\right]\det\left[\left\{B^{-1}\frac{\partial B}{\partial x^\beta}\right\}(x_0^\beta,0)\right]\right.$$
$$\left.\times\det\left[\left\{B^t\frac{\partial D}{\partial x^\delta}B\right\}(x_0^\beta,x_0^\delta,0)\right]\det\left[\left\{A^{-1}\frac{\partial Y}{\partial\theta^\xi}B\right\}(x_0^\alpha,x_0^\beta,0)\right]\right|\right)^{\mathscr{N}}$$
$$= \mathscr{W}_A\mathscr{W}_B\left(\left|\det\left[\left\{B^t\frac{\partial D}{\partial x^\delta}B\right\}(x_0^\beta,x_0^\delta,0)\right]\det\left[\left\{A^{-1}\frac{\partial Y}{\partial\theta^\xi}B\right\}(x_0^\alpha,x_0^\beta,0)\right]\right|\right)^{\mathscr{N}}, \tag{5.10}$$

where $\mathscr{W}_A$ is the weight function for $\mathrm{SO}(n;E_L)$ and $\mathscr{W}_B$ is the weight function for $\mathrm{GL}_e(r;E_L)$. Now let us insist that the parametrization of the matrices D and Y is such that $\partial D/\partial x^\delta$ and $\partial Y/\partial\theta^\xi$ are constant on G. It is then a matter of routine manipulation to obtain

$$\mathscr{W}_G = \mathscr{W}_A\mathscr{W}_B(|\det[B(x_0^\beta,0)]|)^{\mathscr{N}(2r+1)}.$$

A parametrization that is often used for $\mathrm{GL}(n,\mathbb{R})$ is

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{1r} \\ x_{21} & & \\ & & \\ x_{r1} & & x_{rr} \end{bmatrix}$$

$x_{ij}\in\mathbb{R}$, $\det[X]\neq 0$. If we insist that $\det[X]>0$ then we are able to use this parametrization here for $\mathrm{GL}_e(n;E_L)$ (with the obvious extension of the domain of the parameters). The weight function for $\mathrm{GL}(n,\mathbb{R})$ parametrized in this way is $(1/[\det X])r$. We now have

$$\mathscr{W}_G = \mathscr{W}_A(|\det[B(x_0^\beta,o)]|)^{\mathscr{N}(r+1)}. \tag{5.11}$$

VI. LEFT INVARIANT INTEGRAL FOR THE LIE SUPERGROUP B($n;E_L$)

The Lie superalgebra $b(n;\mathbb{C})$ is defined by[17]

$$b(n;\mathbb{C}) = \{X\in\mathrm{spl}(n|n;\mathbb{C}),\ X^p = -(-1)^{|X|}X\}.$$

Here if

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

then

$$X^p = \begin{bmatrix} d^t & -(-1)^{|X|}b^t \\ (-1)^{|X|}c^t & a^t \end{bmatrix}.$$

The defining condition restricts the matrix X to the form

$$\begin{bmatrix} a & b \\ c & -a^t \end{bmatrix}$$

with $b^t=b$ and $c^t=-c$ so that there is a Z grading of the algebra such that

$$X = X_{-1} \oplus X_0 \oplus X_1$$
$$= \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} a & 0 \\ 0 & -a^t \end{bmatrix} \oplus \begin{bmatrix} 0 & 0 \\ c & 0 \end{bmatrix}.$$

The real form $b(n;\mathbb{R})$ gives us the Lie module

$$b(n;E_L) = \{X \in \mathrm{spl}(n|n;E_L),\ X^\rho = -X\} \tag{6.1}$$

and since $\exp(X^\rho) = (\exp X)^\rho$ for $|X| = 0$ we can define the supergroup $\mathrm{B}(n;E_L)$ as

$$\mathrm{B}(n;E_L) = \{U \in \mathrm{SPL}(n|n;E_L), UU^\rho = 1\}. \tag{6.2}$$

This supergroup has been previously defined by Ne'eman[19] and Rittenberg[20] [who call it $\mathrm{P}(n-1)$]. Now since every $U \in \mathrm{B}(n;E_L)$ is invertible we can write it in the form

$$U = \begin{bmatrix} I & B \\ 0 & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ C & I \end{bmatrix}$$

and use the definition (6.2) to determine that $C = -C^t$, $B = B^t$, and $\mathrm{D} = (A^t)^{-1}$, $A \in \mathrm{SL}(n;E_L)$. We can thus identify the group decomposition with that of the algebra. We note that

$$U^{-1} = \begin{bmatrix} I & 0 \\ -C & I \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & A^t \end{bmatrix} \begin{bmatrix} I & -B \\ 0 & I \end{bmatrix}.$$

It is now easy to evaluate the weight function $\mathscr{W}_{\mathrm{B}(n;E_L)}$ simply by examining the result for $\mathrm{SPL}(n|m;E_L)$. We find

$$\mathscr{W}_{\mathrm{B}(n;E_L)} = \mathscr{W}_{\mathrm{SL}(n;E_L)}. \tag{6.3}$$

There is one other real form of $\mathrm{B}(n;\mathbb{C})$ which exists only for even n with $L_0 = \mathrm{su}^*(n)$. This can be treated by exactly the same procedure.

VII. INVARIANT INTEGRAL FOR THE SUPER POINCARÉ GROUPS AND SUPERSPACE

We consider a slight generalization of the super Poincaré group with four odd generators (extension to any number is straightforward) as given for instance by Salam and Strathdee.[22]

The super Poincaré group can be considered as the semidirect produce of the Lorentz group (over E_{L0}) with the supertranslation group (defined below). In terms of the Lorentz generators ($M_{\mu\nu}$) translations (P_λ) and supersymmetry generators (Q^α) which we take to be the generators of a real superalgebra, we have

$$\{M_{\mu\nu}\} \circledS \{P_\lambda, Q^\alpha\}.$$

Now the $M_{\mu\nu}$ can be taken to generate the Lie supergroup $\mathrm{SL}(2,\mathbb{C}E_L)$ considered as a six parameter real Lie supergroup, and the translations and supertranslation generators give a supertranslation group defined by

$$(x^\mu,\theta^\nu)\circ(y^\mu,\phi^\nu) = (x^\mu + y^\mu + a^\mu_{\nu\sigma}\theta^\nu\phi^\sigma, \theta^\nu + \phi^\nu),$$

with $a^\mu_{\nu\sigma} \in E_{L0}$. This is more general than the group normally given[22] since the weight function, as we will see, is independent of the coefficients.

For this supergroup we use a different technique to that of the previous three sections. We obtain the weight function for the supersymmetry group and then combine the result with that for $\mathrm{SL}(2,\mathbb{C}E_L)$.

We thus refer to example (e) [Sec. (15.17) p. 199] of Hewitt and Ross[11] using their terminology. The associated Lie group meets the laid down conditions and the left invariant integral is thus given by

$$\int \frac{f}{|J(\sigma_{(x,y)})|}\, \hat{d}^4x\, \hat{d}^4\theta,$$

where $J(\sigma_{(x,y)})$ is the Jacobian determinant of left translations in the group. It is easy to see that this is 1 since the Jacobian matrix is upper triangular with 1's along the diagonal. We note that since the group is connected and nilpotent it is unimodular.

Now if we choose a parametrization for $\mathrm{SL}(2,\mathbb{C}E_L)$ as

$$\begin{bmatrix} e^{z^5} & z^6 e^{z^5} \\ z^7 e^{z^5} & z^6 z^7 e^{z^5} + e^{-z^5} \end{bmatrix}$$

with $z^\mu = x^\mu + iy^\mu$, $\mu = 5,\ldots,7$; $x^\mu, y^\mu \in E_{L0}$, we can use Theorem 3.6 to obtain

$$\mathscr{W}_{\mathrm{SL}(2,\mathbb{C}E_L)} = \exp\{4\mathscr{N}x_0^5\}.$$

Now by Proposition 29 (p. 100) of Nachbin,[10] the left invariant integral for a semidirect product is given by

$$\int f(x)dx = \int \frac{f(y,x)}{\delta^H(a_z)}\, dy\, dz,$$

with $G = H \circledS K$, dy the left invariant integral on H, dz that on K and dx that on G. By Proposition 28 (p. 99) of Nachbin, $\delta^H(a_z) = 1$ since the super Poincaré, super Lorentz and supertranslation groups are each unimodular. We can thus write the invariant integral for the super Poincaré group as

$$\int f \exp\{4\mathscr{N}x_0^5\}\, \hat{d}^7x\, \hat{d}^3y\, \hat{d}^4\theta.$$

We note that the extension to any number of odd generators will not affect this result and neither will the inclusion of "central charges" since in this case the group considered is (super Poincaré group)$\circledS$ (an abelian group generated by the central charges in the superalgebra).

To finish this section we note that the superspace used in many theories is constructed as a coset space G/H = (super Poincaré)/(super Lorentz). But this is just the supertranslation group given above. Now since G/H is again a group, the G invariant integral on G/H is precisely the invariant integral on G/H. (See Hewitt and Ross[11] p. 206.)

VIII. BEREZIN INTEGRATION

The theory we have developed has been based, in a natural way, on the topology of $E_L^{m,n}$. This is closely analogous to integration on complex manifolds. This gives results that are very different from the theory normally used in the physics literature, i.e., (1) integration is not the "inverse" of differentiation (or even the same as differentiation!). Of course, neither is the Lebesgue integral.

(2) The Jacobian used for transformations is that for the components (x_i^μ, θ_j^ν) not the "super Jacobian" or Berezin function.

(3) The result of integration is an element of E_L, i.e., $\int: f \to E_L$ for $f \in \mathscr{L}^1(M,E_L)$, not a real number.

The Berezin theory for supermanifolds is based on the "sheaf" definition mentioned in the Introduction (see Leites[21] for a full exposition and Rogers[4] for the relationship

between the alternatives). It is sufficient for our purposes to consider a superdomain. That is, consider a region R of $\mathbb{R}^n$ parametrized by $\{u^\mu, \mu = 1,2,...,n\}$, a set of elements of a Grassmann algebra $\{\xi_i, i = 1,2,...,L\}$ and the set of functions

$$H = \Big\{ f \,|\, f(u^\mu,\xi_i) = f_0(u^\mu) + \sum_{i=1}^{L} f_i(u^\mu)\xi_i + \sum_{\substack{i,j=1 \\ i<j}}^{L} f_{ij}(u^\mu)\xi_{i\wedge j} + \cdots + f_{12\ldots L}(u^\mu)\xi_{1\wedge 2\wedge\cdots\wedge L}, \quad \text{with each } f_{ij\cdots l}(u^\mu) \in C^\infty(R) \Big\}. \tag{8.1}$$

A superdomain is then the region R together with the set of functions (sheaf) H, it is denoted by $S^{n,L}$ and is said to have coordinate system (u,ξ).

The Berezin integral of a function $f \in H$ is then defined by[21]

$$\int^B f = (-1)^{(-L(L-1)/2 + Ln)} \int_R f_{12\ldots L}(u^\mu) d^n u. \tag{8.2}$$

Note that the sign on the rhs of (8.2) does not appear in earlier articles (cf. Ref. 6). Apart from the sign, this is as used in the physics literature. It is easy to see that $\int^B : H \to \mathbb{R}$. Leites defines the transformation of coordinates to be given by the "super Jacobian" of the coordinate change.

It was shown by Rogers[4] that a G^∞ -supermanifold can be constructed from a sheaf supermanifold, but unfortunately the Berezin integral cannot be transferred to give a notion of measure on a G^∞ -supermanifold.

It is clear that Berezin integration is a linear functional defined on H, that is little more than a projection map. It is stated in Ref. 6 that invariant integrals can be constructed with $\int^B$ and that the result for SPL$(p|q;\mathbb{R})$ (U(p,q)) is zero.

APPENDIX: NOTATIONS AND CONVENTIONS

Our conventions follow Rogers[4,5] very closely, with some influence from Leites[21] and some from personal preference.

A. Grassman algebra and superspace

Let L be a positive integer and denote the basis elements of $\mathbb{R}^L$ by $\epsilon_i (i = 1,2,...,L)$. Let E_L denote the Grassmann algebra over $\mathbb{R}^L$ with antisymmetric product given by $\wedge$. Then E_L has basis $\epsilon_0, \epsilon_i, \epsilon_{i\wedge j}, \ldots, \epsilon_{i\wedge j\wedge k}, \ldots, \epsilon_{i\wedge j\wedge k\wedge\cdots\wedge l}$, where ϵ_0 is the unit element of E_L. For notational convenience we have defined $\epsilon_{i\wedge j} = \epsilon_i \wedge \epsilon_j$ etc. so that

$$\epsilon_{i\wedge j\wedge\cdots\wedge k\wedge l\wedge\cdots\wedge m} = -\epsilon_{i\wedge j\wedge\cdots\wedge l\wedge k\wedge\cdots\wedge m}$$

(of course $\epsilon_0 \wedge \epsilon_i = \epsilon_i$ etc.).

E_L has a natural $\mathbb{Z}$ grading in which (for $l = 1,2,...,L$) the homogeneous part $E_L^{(l)}$ consists of all real linear combinations of basis elements $\epsilon_{i\wedge j\wedge\cdots\wedge k}$ involving l indices, with $E_L^0 = \{\mathbb{R}\epsilon_0\}$. Elements of $E_L^{(l)}$ will be said to be of level l, denoted by $l(x) = l$ if $x \in E_L^{(l)}$. This induces a $\mathbb{Z}_2$ grading $E_L = E_{L0} \oplus E_{L1}$, with $E_{L0} = \{E_L^{(l)}, l \text{ even}\}$ and $E_{L1} = \{E_L^{(l)}, l \text{ odd}\}$. Then E_{L0}, E_{L1}, E_L are vector spaces over $\mathbb{R}$ with dim E_{L0} = dim $E_{L1} = 2^{L-1} = \mathcal{N}$ and dim $E_L = 2\mathcal{N}$.

The degree or parity of a homogeneous $\mathbb{Z}_2$ graded element $x \in E_{La}$ is defined by $|x| = a$, $a \in \{0,1\}$. It will be said that x is "even" if $|x| = 0$, and is "odd" if $|x| = 1$.

It is convenient to denote the basis elements of E_{L0} by $\mathbf{e}_i$, $i = 0,1,...,\mathcal{N} - 1$ and correspondingly the basis elements of E_{L1} by $\mathbf{f}_j, j = 1,2,...,\mathcal{N}$ with the assignment $\mathbf{e}_0 = \epsilon_0$, and the restrictions that if $i < k$ then $l(\mathbf{e}_i) \leqslant l(\mathbf{e}_k)$ and $l(\mathbf{f}_i) \leqslant l(\mathbf{f}_k)$. (Where a specific basis is needed it is better to revert to the ϵ basis rather than have complicated index assignments.)

A typical element of E_{L0} will be denoted by $x = x_i \mathbf{e}_i$ (summation implied) $i = 0,1,...,\mathcal{N} - 1$, $x_i \in \mathbb{R}$, and a typical element of E_{L1} will be denoted by $\theta = \theta_j \mathbf{f}_j$ (summation implied) $j = 1,2,...,\mathcal{N}$, $\theta_j \in \mathbb{R}$. We will write an element of E_L as $x + \theta$.

We define a norm on E_L by

$$\|x + \theta\| = \left(\sum_{i=1}^{\mathcal{N}} (x_i)^2 + \sum_{j=1}^{\mathcal{N}} (\theta_j)^2 \right)^{1/2}$$

and the corresponding distance function as $d(x + \theta, x' + \theta') = \|x - x' + \theta - \theta'\|$. E_L is then a topological vector space with a Hausdorff topology. The basis vectors in E_L ($\mathbf{e}_i, \mathbf{f}_j$, or equivalently $\epsilon_{i\wedge j\wedge\cdots\wedge k}$) are defined to be orthonormal. The complex Grassmann space is defined by $\mathbb{C}E_L = E_L \oplus iE_L$.

Superspace is defined as $E_L^{m,n} = E_{L0}^m \times E_{L1}^n$ $= E_{L0} \times \cdots \times E_{L0} \times E_{L1} \times \cdots \times E_{L1}$, i.e., the Cartesian product of m copies of E_{L0} and n copies of E_{L1}. A typical element of $E_L^{m,n}$ is then of the form $(x^1, x^2, ..., x^m, \theta^1, \theta^2, ...\theta^n)$, where $x^\mu = \sum_{i=0}^{\mathcal{N}-1} x_i^\mu \mathbf{e}_i$, $x_i^\mu \in \mathbb{R}$ and $\theta^\nu = \sum_{j=1}^{\mathcal{N}} \theta_j^\nu \mathbf{f}_j$, $\theta_j^\nu \in \mathbb{R}$. This will be written (x,θ) unless we wish to consider explicitly the real variables x_i^μ and θ_j^ν in which case we will write (x_i,θ_j). That is, the Greek indices which appear as superscripts distinguish the various copies of E_{L0} or E_{L1}, while the Latin indices which appear as subscripts indicate the component "within" E_{L0} or E_{L1}. Summation is assumed over all repeated indices including those that are suppressed, e.g.,

$$\alpha_\mu x^\mu = \sum_{\mu=1}^{m} a_\mu x_i^\mu e_i = \sum_{\mu=1}^{m} \sum_{i=0}^{\mathcal{N}} a_\mu x_i^\mu \mathbf{e}_i.$$

B. Analysis

Differentiation is defined as by Rogers[4] (see also Boyer and Gitter[15] for a full discussion of this). We prefer to denote it in the normal way, i.e., $\partial/\partial x^\mu$, or $\partial/\partial\theta^\nu$. Differentiation with respect to the components x_i^μ and θ_j^ν is also defined in the normal way and is denoted by $\partial/\partial x_i^\mu$ and $\partial/\partial\theta_j^\nu$. Functions $f : E_L^{m,n} \to E_L$ will be said to be G^1 if they are once differentiable with respect to x^μ, θ^ν for all μ,ν and G^∞ if they are infinitely differentiable.

C. Differential geometry

We denote manifolds over $E_L^{m,n}$ by large Latin letters, e.g., M, and the corresponding manifolds over $\mathbb{R}^{\mathcal{N}(m+n)}$ by script letters $\mathcal{M}_L$, the subscript being a reminder that we are in fact considering a whole family of manifolds (one for each value of L). The dimension of M is denoted by $(m|n)$. The space of vector fields on M is denoted by $D^1(M)$ and admits a grading $D^1(M) = D^1(M)_0 \oplus D^1(M)$ (see Rogers[4]). The "tangent" space at a point p of M is denoted by $ST_p(M)$, and again

is graded, the even part, which "corresponds" to the tangent space of $\mathscr{M}_L$, being denoted by $T_p(M)$.

Lie supergroups are denoted by large latin letters, e.g., G, and their corresponding Lie groups by large script letters, e.g., $\mathscr{G}_L$. The "tangent" space at the identity is denoted by $\mathrm{ST}_e(G)$ and forms a "Lie super module" (or "Lie left B_L-module" in the terminology of Rogers[4]). The even part of $ST_e(G)$ is denoted by $T_e(G)$ and called a "Lie module." The basis of the Lie supermodule of a Lie supergroup G of dimension $(m|n)$ is denoted by $\{\alpha_\mu, \beta_\nu\}$ with $|\alpha_\mu| = 0$, $|\beta_\nu| = 1$, $\mu = 1,2,...,m$ and $\nu = 1,2,...,n$. A basis of the Lie algebra of $\mathscr{G}_L$ is then given by $\{\mathbf{e}_i\alpha_\mu, \mathbf{f}_j\beta_\nu\}$.

D. Super matrix algebra

For an extensive review of this subject see Leites.[21]

Super matrices are partitioned in block form $M = \left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]$. The set $M(p|q;E_L)$ is defined to be such that A is $p\times p$ and D is $q\times q$ with entries from E_{LO}, while B and C have dimensions $p\times q$ and $q\times p$ and have entries from E_{L1}. The set $\bar{M}(p|q;E_L)$ is defined to be such that A is $p\times p$ and D is $q\times q$ with entries from E_{L1} and B and C are $p\times q$ and $q\times p$ and have entries from E_{L1}. We define $\tilde{M}(p|q;E_L) = M(p|q;E_L)\oplus\bar{M}(p|q;E_L)$. Parity of matrices is defined by $|M| = 0$ if $M\in M(p|q;E_L)$ and $|M| = 1$ if $M\in\bar{M}(p|q;E_L)$. Multiplication by parameters, from E_L is defined by

$$aX = \begin{bmatrix} aI_p & 0 \\ 0 & (-1)^{|a|}aI_q \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

for $a\in E_{L\alpha}, \alpha\in\{0,1\}$ this then satisfies $aX = (-1)^{|a||X|}Xa$.

Matrix supergroups are subsets of $M(p|q;E_L)$ so that every group element is *even*. The "tangent" space $ST_e(G)$ consists of both odd and even matrices with $T_e(G)$ even.

[1]F. A. Berezin and G. I. Kac, Mat. Sb. [Math. USSR. Sb.] **11**, 311–325 (1970).
[2]F. A. Berezin and D. A. Leites, Dokl. Akad. Nauk SSSR [Sov. Math. Dokl.] **16**, 1218–1222 (1975).
[3]B. Kostant, *Lecture Notes in Mathematics*, Vol. 570 (Springer-Verlag, Berlin, 1977).
[4]A. Rogers, J. Math. Phys. **21**, 1352–1365 (1980).
[5]A. Rogers. J. Math. Phys. **22**, 939–945 (1981).
[6]F. A. Berezin, Yad. Fiz. **29**, 1670 (1979) [Sov. J. Nucl. Phys. **29**, 857–866 (1979)].
[7]F. A. Berezin, *The Method of 2nd Quantization* (Academic, New York, 1966).
[8]P. van Nieuwenhuizen, Phys. Repo. **68**, (4) (1981).
[9]D. L. Cohn, *Measure Theory* (Birkhauser, Cambridge, MA, 1980).
[10]L. Nachbin, *The Haar Integral* (Van Nostrand, New York, 1965).
[11]E. Hewitt and K. A. Ross, *Abstract Harmonic Analysis* (Springer-Verlag, Berlin, 1963).
[12]V. G. Kac, Adv. Maths. **26**, 8–96 (1979).
[13]S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[14]C. Chevalley, *Theory of Lie Groups I* (Princeton University, Princeton, NJ, 1946).
[15]C. P. Boyer and S. Gitter, IIMAS Report No. 303, Universidad Nacional Autonoma de Mexico.
[16]V. Rittenberg and M. Scheunert, J. Math. Phys. **19**, 709–713 (1978).
[17]M. Scheunert, *The Theory of Lie Superalgebras*, Vol. 716 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1979).
[18]M. Parker, J. Math. Phys. **21**, 689–697 (1981).
[19]Y. Ne'eman, *Superalgebras, Supergroups and Geometric Gauging* in *Cosmology and Gravitation*, edited by G. Bergmann and V. de Sabbata (NATO ASI Series, Erice, 1980), Vol. 5, p. B58.
[20]V. Rittenberg, *A Guide to Lie Superalgebras*, Vol. 79 in *Lecture Notes in Physics* (Springer-Verlag, Berlin, 1978).
[21]D. A. Leites, Usp. Mat. Nauk (Russian Math. Surveys) **35**, 1–64 (1980).
[22]A. Salam and J. Strathdee, Phys. Rev. D **11**, 1521–1535 (1975).

A novel class of Bessel function integrals

M. L. Glasser
Department of Mathematics and Computer Science, Clarkson College of Technology, Potsdam, New York 13676

(Received 12 October 1983; accepted for publication 16 February 1984)

A number of definite integrals over the unit interval involving Bessel functions with argument $[\alpha^2 x^{-1} + \beta^2(1-x)^{-1}]^{1/2}$ are evaluated in closed form.

PACS numbers: 02.90. + p, 02.30. + g

I. INTRODUCTION

The integrals to be discussed in this paper have arisen in a calculation of the screening properties of a charged impurity located inside and near the surface of a metal subjected to a magnetic field. Their unusual nature and the surprise that they could be evaluated at all, thus avoiding a tedious numerical analysis, suggested that there might be sufficient interest to justify publishing this short paper.

The integrals are

$$\phi_\nu(\alpha,\beta) = \int_0^1 (uu')^{-3/2} t_0^{-\nu} J_\nu(t_0)\,du, \tag{1}$$

$$\psi_0(\alpha,\beta) = \int_0^1 u^{1/2}(u')^{-3/2} J_0(t_0)\,du, \tag{2}$$

and

$$C_0(\alpha,\beta) = \int_0^1 (uu')^{-1/2} J_0(t_0)\,du. \tag{3}$$

For completeness, although they are of a simpler nature, I include

$$A_\nu = \int_0^1 (uu')^{(3-\nu)/2} J_\nu(\lambda\,(uu')^{-1/2})\,du, \quad \nu = 0,1,2 \tag{4}$$

and

$$B_0 = \int_0^1 u^{1/2}(u')^{-3/2} J_0(\lambda\,(uu')^{-1/2})\,du. \tag{5}$$

The notation is

$$u' = 1-u, \quad t_0 = \sqrt{\frac{\alpha^2}{u'} + \frac{\beta^2}{u}}, \tag{6}$$

where α and β are positive constants and $\mathrm{Re}\,\nu > -1$, unless otherwise specified.

II. EVALUATION

Let $\mathscr{L}_p$ and $\mathscr{L}_t^{-1}$ denote the Laplace transform and its inverse.[1] Then we have the integral representation[2]

$$t^{-\nu} J_\nu(t) = 2^\nu \mathscr{L}_{1/4}^{-1}\{p^{-\nu-1} e^{-t^2/p}\} \quad (\mathrm{Re}\,\nu > -1) \tag{7}$$

so, since the interchange of the order of integration causes no problems,

$$\phi_\nu(\alpha,\beta) = 2^\nu \mathscr{L}_{1/4}^{-1}\left\{p^{-\nu-1}\int_0^1 (uu')^{-3/2} \times \exp\left[-\frac{1}{p}\left(\frac{\alpha^2}{u'} + \frac{\beta^2}{u}\right)\right] du\right\}. \tag{8}$$

The substitution $x = (u')^{-1} - 1$ reduces the inner integral to

$$\exp[-p^{-1}(\alpha^2+\beta^2)] \int_0^\infty (x^{-1/2} + x^{-3/2}) \times \exp\{-p^{-1}(\alpha^2 x + \beta^2 x^{-1})\} dx \tag{9}$$

which is a tabulated Laplace transform.[3] Thus,

$$\phi_\nu(\alpha,\beta) = [2^\nu \pi^{1/2}(\alpha+\beta)/\alpha\beta\,]\,\mathscr{L}_{1/4}^{-1}\{p^{-(\nu-1/2)-1} \times \exp[-p^{-1}(\alpha+\beta)^2]\}. \tag{10}$$

The latter is a tabulated inverse Laplace transform,[4] whence

$$\phi_\nu(\alpha,\beta) = (2\pi)^{1/2}(\alpha\beta)^{-1}(\alpha+\beta)^{3/2-\nu} J_{\nu-1/2}(\alpha+\beta). \tag{11}$$

The case of the second integral $\psi_0(\alpha,\beta)$ is somewhat more complicated. The procedure above leads to

$$\psi_0(\alpha,\beta) = \mathscr{L}_{1/4}^{-1}\left\{p^{-1}\exp[-p^{-1}(\alpha^2+\beta^2)]\int_0^\infty (x+1)^{-1}x^{-3/2} \times \exp[p^{-1}(\alpha^2 x + \beta^2 x^{-1})]dx\right\}. \tag{12}$$

To evaluate the x-integral in (12) we use the Parseval formula for the Laplace transform, noting that[5]

$$\mathscr{L}_z^{-1}\{(x+1)^{-1}\} = e^{-z}, \tag{13}$$

$$\mathscr{L}_z\{x^{-3/2}\exp[-p^{-1}(\beta^2 x + \alpha^2 x^{-1})]\} = (\pi p)^{1/2}\alpha^{-1}\exp[-2\alpha\,p^{-1/2}(z+\beta^2 p^{-1})^{1/2}] \tag{14}$$

so that the x-integral becomes

$$(\pi p)^{1/2}\alpha^{-1}\int_0^\infty \exp[-z - 2\alpha\,p^{-1/2}(z+\beta^2 p^{-1})^{1/2}]dz = 2\alpha^{-1}(\pi p)^{1/2}\exp(\beta^2/p) \times \int_{\beta p^{-1/2}}^\infty \exp[-2\alpha\,p^{-1/2}t - t^2]\,t\,dt. \tag{15}$$

From Ref. 6 we derive the formula

$$\int_b^\infty \exp[-xt - t^2]\,t\,dt = \frac{1}{2}\left\{\exp[-bx - b^2] - 2^{-1}\pi^{1/2} \times x e^{x^2/4}\,\mathrm{erfc}\left(\frac{1}{2}x + b\right)\right\}, \tag{16}$$

so that

$$\psi_0(\alpha,\beta) = \pi^{1/2}\alpha^{-1}\mathscr{L}_{1/4}^{-1}\{p^{-1/2}(\exp[-p^{-1}(\alpha+\beta)^2] - \alpha^{-1}(\pi/p)^{1/2}\,\mathrm{erfc}[p^{-1/2}(\alpha+\beta)])\}. \tag{17}$$

The first inverse Laplace transform in (17) is tabulated[7]; the second is not, but it can be shown that

 0022-2488/84/102933-02$02.50

$$\mathcal{L}_t\{p^{-1}\operatorname{erfc}(ap^{-1/2})\} = -(2/\pi)\operatorname{si}(at^{1/2}). \tag{18}$$

Thus,

$$\psi_0(\alpha,\beta) = 2\{\alpha^{-1}\cos(\alpha+\beta) + \operatorname{si}[2^{-1}(\alpha+\beta)]\}. \tag{19}$$

The same procedure applied to $C_0(\alpha,\beta)$ yields the rhs of (12), where $x^{-3/2}$ is replaced by $x^{-1/2}$. A similar application of Parseval's formula gives

$$C_0(\alpha,\beta) = \pi\mathcal{L}^{-1}_{1/4}\{p^{-1}\operatorname{erfc}[p^{-1/2}(\alpha,\beta)]\}, \tag{20}$$

and by (18) this is reduced to

$$C_0(\alpha,\beta) = -2\operatorname{si}(\alpha+\beta). \tag{21}$$

The integrals in (4) are much simpler. By the substitution $x = 2^{-1}(uu')^{-1/2}$ they reduce to tabulated Hankel transforms[8] (our notation for the Hankel transform and the Struve functions are taken from this reference).

$$\begin{aligned} A_0 &= 8(2\lambda)^{-1/2}\mathscr{H}_0\{x^{1/2}(x^2-1)^{-1/2}\theta(x-1);2\lambda\} \\ &= 4\lambda^{-1}\cos(2\lambda), \\ A_1 &= 4(2\lambda)^{-1/2}\mathscr{H}_1\{x^{1/2}(x^2-1)^{-1/2}\theta(x-1);2\lambda\} \\ &= 2\lambda^{-1}\sin(2\lambda), \\ A_2 &= 2(2\lambda)^{-1/2}\mathscr{H}_2\{x^{-3/2}(x^2-1)^{-1/2}\theta(x-1);2\lambda\} \\ &= (\pi/\lambda)^{1/2}J_{3/2}(2\lambda), \end{aligned} \tag{22}$$

where $\theta(x)$ denotes the Heaviside unit step function. By this procedure we have

$$\begin{aligned} B_0 &= \lambda^{-1}\cos(2\lambda) - 2\lambda^{-1} \\ &\quad + 4J_0(2\lambda) - 2\pi[J_0(2\lambda)\mathbf{H}_1(2\lambda) - J_1(2\lambda)\mathbf{H}_0(2\lambda)] \\ &\quad + 2(2\lambda)^{-1/2}\mathscr{H}_0\{x^{-3/2}(x^2-1)^{1/2}\theta(x-1);2\lambda\}. \end{aligned} \tag{23}$$

The latter Hankel transform is not tabulated, but it can be shown to be

$$(2\lambda)^{-1/2}\cos(2\lambda) + (2\lambda)^{1/2}\operatorname{si}(2\lambda) \tag{24}$$

which completes the evaluation.

The example (21) suggests the conjecture that for any function f for which the integral exists,

$$\int_0^1\left(\frac{1}{u}+\frac{1}{u'}\right)^{1/2} f(t_0)du = F(\alpha+\beta). \tag{25}$$

ACKNOWLEDGMENT

This work has been supported by the National Science Foundation.

[1] A. Erdelyi *et al.*, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. I, Chap. IV.
[2] Reference 1, Eq. (4.14(30)).
[3] Reference 1, Eq. (4.5(27), (28)).
[4] Reference 1, Eq. (5.5 (40)).
[5] Reference 1, Eq. (4.1 (5)).
[6] Reference 1, Eq. (4.5 (21)).
[7] Reference 1, Eq. (5.5 (37)).
[8] A. Erdelyi *et al.*, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. II, Chap. VIII.

Post-Newtonian extensions of the Runge–Lenz vector

F. Argüeso and J. L. Sanz
Departamento de Física Teórica, Universidad de Santander, Santander, Spain

(Received 18 November 1983; accepted for publication 15 March 1984)

We obtain the most general post-Newtonian extension of the Runge–Lenz vector corresponding to a very large class of two-body relativistic systems whose equations of the relative motion can be derived from a post-Newtonian Hamiltonian depending on four parameters which include the electromagnetic and gravitational cases. Assuming a couple of conditions, we fix the arbitrariness and obtain a unique post-Newtonian Runge–Lenz vector, whose associated symmetries generate the same algebra as in the Keplerian case.

PACS numbers: 03.30. + p, 03.20. + i

I. INTRODUCTION

In Newtonian mechanics, a two-body system interacting through a Coulombian potential $V(r) = k/r$ ($k = \text{const}$) admits, in the center-of-mass frame, a vector first integral $\mathbf{A}_N$ orthogonal to the relative angular momentum, which is called the Runge–Lenz vector:

$$\mathbf{A}_N \equiv \mathbf{v} \wedge \mathbf{J}_N - kr^{-1}\mathbf{r}, \tag{1}$$

$$\mathbf{J}_N \equiv \mu \mathbf{r} \wedge \mathbf{v}, \quad \mu \equiv m_1 m_2 (m_1 + m_2)^{-1},$$

$$\mathbf{r} \equiv \mathbf{r}_1 - \mathbf{r}_2, \quad \mathbf{v} \equiv \mathbf{v}_1 - \mathbf{v}_2. \tag{2}$$

From a practical point of view, $\mathbf{A}_N$ can be used to obtain directly the relative orbit of the system through a trivial calculation involving the scalar product $(\mathbf{r}\cdot\mathbf{A}_N)$. Moreover, the dynamical symmetries associated to $(2\mu^{-1}|E_N|)^{-1/2}\,\mathbf{A}_N$ generate—together with the generators of the spatial rotations associated to the vector first integral $\mathbf{J}_N$—the Lie algebra corresponding to $T(3)\times SO(3)$ ($E_N = 0$), SO(3,1) ($E_N > 0$), or SO(4) ($E_N < 0$), where E_N is the relative Newtonian energy.

Basic questions addressed this paper are the following: Given post-Newtonian equations of motion that correspond to a two-body system, does there exists any post-Newtonian extension of the Runge–Lenz vector? If such an extension exists, is there uniqueness? Are the symmetries mentioned above broken?

In Sec. II, we prove that such a post-Newtonian extension is possible for a very large class of systems which include the relevant physical interactions (electromagnetism and gravitation). Two arbitrary functions of the variables (E_N, J_N) appear in the general solution corresponding to the extension. One of them can be fixed by assuming a particular coordinate system, in the plane of the relative motion, in order to obtain the most simple equation for the orbit. This calculation is presented in Sec. III, whereas we study the Lie algebra generated by the symmetries associated—via Noether's inverse theorem—to the previous extension of $\mathbf{A}_N$ in Sec. IV.

II. EXTENSION OF THE RUNGE–LENZ VECTOR

Let us consider a relativistic two-body system, constituted by two structureless pointlike particles, whose equations of motion at the post-Newtonian level, in the center-of-mass frame, are obtained from the Hamiltonian

$$H(r,(\mathbf{r}\cdot\mathbf{p}),p) = \frac{p^2}{2\mu} - \frac{k}{r} + \frac{1}{c^2}\left\{a\frac{p^4}{\mu^3} - \frac{k}{r} \times \left[b\frac{p^2}{\mu^2} - d\frac{k}{\mu r} + e\frac{(\mathbf{r}\cdot\mathbf{p})^2}{r^2\mu^2}\right]\right\}, \tag{3}$$

where k is a constant, c the velocity of light in vacuum, $\mathbf{p}$ is the relative momentum [i.e., $\mathbf{p} = \mu\mathbf{v} + O(c^{-2})$] and a,b,d,e are arbitrary functions of the dimensionless variable μm^{-1}, $m \equiv m_1 + m_2$. The structure (3) corresponds to a nonsingular (except at $r = 0$) relativistic extension, invariant under time translations and spatial rotations, of a classical Hamiltonian with a Coulombian potential.

The expression (3) coincides, in the center-of-mass frame, with Darwin's Hamiltonian[1,2] (DH) and Einstein–Infeld–Hoffmann's Hamiltonian[3,4] (EIHH) for the particular values

$$\text{DH:} \quad a \equiv \tfrac{1}{8}(-1 + 3\mu/m), \quad b \equiv \tfrac{1}{2}\mu/m, \quad d \equiv 0,$$

$$e \equiv \tfrac{1}{2}\mu/m, \quad k \equiv -e_1 e_2, \tag{4}$$

$$\text{EIHH:} \quad a \equiv \tfrac{1}{8}(-1 + 3\mu/m), \quad b \equiv \tfrac{1}{2}(3 + \mu/m),$$

$$d = \tfrac{1}{2}, \quad e \equiv \tfrac{1}{2}\mu/m, \quad k \equiv Gm_1 m_2. \tag{5}$$

The DH (EIHH) describes the electromagnetic (gravitational) interaction of two pointlike particles up to the post-Newtonian level in the framework of classical electrodynamics (general relativity).

We shall look for a post-Newtonian extension of the Runge–Lenz vector $\mathbf{A}_N$, orthogonal to the relative post-Newtonian angular momentum $\mathbf{J} = \mathbf{r} \wedge \mathbf{p}$, being a vector first integral of the equations of motion derived from (1).

From the expression $\mathbf{v} = \partial H/\partial \mathbf{p}$, one easily obtains

$$\mathbf{p} = \mu\mathbf{v} + \frac{1}{c^2}\left\{-4a\mu v^2\mathbf{v} + \frac{2k}{r}\left[b\mathbf{v} + e\frac{(\mathbf{r}\cdot\mathbf{v})}{r^2}\mathbf{r}\right]\right\}, \tag{6}$$

$$\mathbf{J} = \left\{1 + \frac{1}{c^2}\left[-4av^2 + 2b\frac{k}{\mu r}\right]\right\}\mathbf{J}_N. \tag{7}$$

Then, we shall consider the decomposition

$$\mathbf{A} = \mathbf{A}_N + (1/c^2)[f\mathbf{A}_N + g\,\mathbf{J}_N \wedge \mathbf{A}_N]. \tag{8}$$

Our objective is to calculate the functions f and g, which we shall assume invariant under spatial rotations. It is clear that

 0022-2488/84/102935-04$02.50

f, g must be arbitrary functions of three independent scalars, for instance (r, E_N, J_N), where $J_N \equiv |\mathbf{J}_N|$.

Here $\mathbf{A}$ is a vector first integral if its Poisson bracket with the Hamiltonian (1) vanishes:

$$[\mathbf{A}, H] = 0. \tag{9}$$

In order to calculate this bracket in a simple way, we can rewrite the extension (8) in the form

$$\mathbf{A} = \mathbf{A}_0 + (1/c^2)[\alpha \mathbf{A}_0 + \beta \mathbf{J} \wedge \mathbf{A}_0], \quad \alpha, \beta (r, J, E), \tag{10}$$

with

$$\mathbf{A}_0 = \left(\frac{p^2}{\mu} - \frac{k}{r}\right)\mathbf{r} - \frac{(\mathbf{r}\cdot\mathbf{p})}{\mu}\mathbf{p}, \quad J = |\mathbf{J}|, \quad E \equiv H. \tag{11}$$

The introduction of this decomposition into Eq. (9) leads to a couple of partial differential equations (with respect to the variable r) for α and β, whose integrations give

$$\alpha = \frac{J^2}{A_0^2}\frac{k}{\mu r}\left[2(4a - b - e)\frac{E}{\mu} + (4a - 2b + d - 2e) \times \frac{k}{\mu r} + e\left(\frac{J}{\mu r}\right)^2\right] + \tilde{\alpha}(J, E), \tag{12}$$

$$\beta = \frac{1}{A_0^2}\left\{(4a - 2b + d - e)\frac{k^2}{J}\left(2\frac{E}{\mu} + \frac{k^2}{J^2}\right) \times \arctan\left(\frac{\mu k r - J^2}{J(\mathbf{r}\cdot\mathbf{p})}\right) + \frac{k}{\mu r}(\mathbf{r}\cdot\mathbf{p})\left[2(4a - b - e)\frac{E}{\mu} + (4a - 2b + d - e) \times \left(\frac{k^2}{J^2} + \frac{k}{\mu r}\right) + e\left(\frac{J}{\mu r}\right)^2\right]\right\} + \tilde{\beta}(J, E), \tag{13}$$

where $\tilde{\alpha}, \tilde{\beta}(J, E)$ are arbitrary functions of their arguments and $A_0 \equiv |\mathbf{A}_0|$.

Reintroducing the value of $\mathbf{p}$ given by Eq. (6), we can rewrite the formulas (10) in the form (8) with

$$f \equiv \frac{J_N^2}{A_N^2}\frac{k}{\mu r}\left[2(-20a + 3b + e)\frac{E_N}{\mu} + (-12a + 2b + 2e + d)\frac{k}{\mu r} - e\left(\frac{J_N}{\mu r}\right)^2\right] + \alpha^*(J_N, E_N), \tag{14}$$

$$g \equiv \frac{1}{A_N^2}\left\{(4a - 2b + d - e)\frac{k^2}{J_N}\left(2\frac{E_N}{\mu} + \frac{k^2}{J_N^2}\right) \times \arctan\left(\frac{\mu k r - J_N^2}{\mu J_N(\mathbf{r}\cdot\mathbf{v})}\right) + \frac{k}{r}(\mathbf{r}\cdot\mathbf{v})\left[(4a - 2b + d - e)\frac{k^2}{J_N^2} + (-12a + 2b + d + e)\frac{k}{\mu r} - e\left(\frac{J_N}{\mu r}\right)^2\right]\right\} + \beta^*(J_N, E_N), \tag{15}$$

where $\alpha^*, \beta^*(J_N, E_N)$ are arbitrary functions of their arguments and $A_N \equiv |\mathbf{A}_N|$. In order to fix this arbitrariness, we shall use the equation of the orbit and fix the modulus of $\mathbf{A}$.

III. EQUATION OF THE ORBIT

As in Newtonian mechanics, the orbit can be obtained making the scalar product $(\mathbf{r}\cdot\mathbf{A})$, where $\mathbf{A}$ is given by the expressions (14) and (15). If we denote by θ the angle between $\mathbf{r}$ and $\mathbf{A}$,

$$rA\cos\theta = -kr + \frac{J_N^2}{\mu} + \frac{1}{c^2}\{frA\cos\theta + gJAr\sin\theta\}. \tag{16}$$

Substituting r by $J^2[\mu(A\cos\theta + k)]^{-1}$ in the post-Newtonian term and J_N^2 by the expression obtained from Eq. (7),

$$J_N^2 = J^2\left\{1 + \frac{4}{c^2}\left[4a\frac{E}{\mu} + (4a - b)\frac{k}{\mu r}\right]\right\}, \tag{17}$$

Eq. (16) can be rewritten, after a tedious calculation, as follows:

$$1/r = B + F\cos(\theta - \delta\theta + \theta_0) + G\cos 2(\theta + \theta_0), \tag{18}$$

where

$$B \equiv \frac{\mu k}{J^2}\left\{1 + \frac{1}{c^2}\left[(-8a + 2b + e)\frac{E}{\mu} + \left(-8a + 4b + \frac{1}{2}e - 2d\right)\frac{k^2}{J^2}\right]\right\}, \tag{19}$$

$$F \equiv \frac{\mu A}{J^2}\left\{1 + \frac{1}{c^2}\left[-4a\frac{E^2J^2}{\mu^2A^2} - (16a - 6b + 2d)\frac{Ek^2}{\mu A^2} - 2(4a - 2b + d)\frac{k^4}{A^2J^2}\right]\right\}, \tag{20}$$

$$G \equiv -(1/c^2)e(\mu k A^2/2J^4), \tag{21}$$

$$\theta_0 \equiv \frac{1}{c^2}\left[J\beta^*(J, E) - \frac{\pi}{2}\frac{k^2}{A^2}\left(2\frac{E}{\mu} + \frac{k^2}{J^2}\right) \times (-4a + 2b - d + e)\right], \tag{22}$$

$$\delta \equiv (1/c^2)(-4a + 2b - d + e)(k^2/J^2). \tag{23}$$

It is obvious that the equation of the orbit (18) is written in the most simple way if one elects $\theta_0 \equiv 0$, which is equivalent to the following election for β^*:

$$\beta^* = \frac{\pi k^2}{2JA^2}\left(2\frac{E}{\mu} + \frac{k^2}{J^2}\right)(-4a + 2b + e - d). \tag{24}$$

Then, Eq. (18) represents a perturbed conic: the term $F\cos(\theta - \delta\theta)$ produces a perihelion advance (at any period the perihelion advances the angle $2\pi\delta$ rad), whereas the term $G\cos 2\theta$ produces a deformation of the conic of period π.

Substituting the parameters which characterize the Darwin and EIH Hamiltonians, Eqs. (4) and (5), we obtain

$$\delta_D = \tfrac{1}{2}(e_1e_2/cJ)^2, \quad \delta_{EIH} = 3(Gm_1m_2/cJ)^2. \tag{25}$$

The last result, δ_{EIH}, constitutes one of the classical tests of general relativity.

Here $F\cos(\theta - \delta\theta)$ is a "secular" term, because it produces observable effects for sufficiently large time intervals (e.g., a perihelion advance of 43 arcsec/century for Mercury in the gravitational case). Trivially, this effect does not exist for particular Hamiltonians such that $4a + d = 2b + e$, whereas there is no "periodic" perturbation for $e \equiv 0$.

In order to determine a unique extension of the Runge–Lenz vector, we shall assume that the modulus of **A** depends on E and J in the same form as in Newtonian mechanics, i.e.,

$$A^2 = k^2 + 2EJ^2/\mu. \tag{26}$$

Taking into account Eqs. (8), (14), and (15), we obtain the following expression:

$$A^2 = k^2 + 2\frac{EJ^2}{\mu} + \frac{2}{c^2}\left[A^2\alpha^* + 26a\left(\frac{EJ}{\mu}\right)^2\right]. \tag{27}$$

Therefore, if we assume the modulus given by Eq. (26), α^* must be given by

$$\alpha^* = -26a(E/\mu)(2+\mu k^2/EJ^2)^{-1}, \tag{28}$$

and β^* is reduced to, see Eq. (24),

$$\beta^* = \pi k^2/2J^3. \tag{29}$$

Summing up, under the assumption of a modulus for **A** given by Eq. (26) and an equation of the orbit referred to **A** expressed in the most simple form, we have obtained a unique post-Newtonian extension of the Runge–Lenz vector which is given by Eqs. (8), (14), (15), (28), and (29).

IV. ASSOCIATED SYMMETRIES

Let J_i, A_j $(i,j,k = 1,2,3)$ be the components of the post-Newtonian angular momentum and extension of the Runge–Lenz vector, respectively. A direct calculation leads to the following Poisson brackets between them:

$$[J_i, J_j] = \epsilon_{ij}{}^k J_k, \quad [J_i, A_j] = \epsilon_{ij}{}^k A_k,$$
$$[A_i, A_j] = -(2E/\mu)\epsilon_{ij}{}^k J_k. \tag{30}$$

Therefore the Poisson algebra corresponding to the first integrals (J_i, A_j) is the same as in Newtonian mechanics, but now E appears instead of the Newtonian energy.

On the other hand, one can associate to the constants of the motion $\zeta_a(x^i, p_j)$ $(a,b,c = 1,\ldots,r)$ the infinitesimal generators

$$\zeta_a \equiv -\frac{\partial \zeta_a}{\partial p_i}\frac{\partial}{\partial x^i} + \frac{\partial \zeta_a}{\partial x^i}\frac{\partial}{\partial p_i}, \tag{31}$$

under which the Hamiltonian H is invariant[5]:

$$\zeta_a H = \frac{d\zeta_a}{dt} = 0. \tag{32}$$

It is very easy to prove, taking into account the structure (31), the following relations:

$$\frac{\partial}{\partial x^i}\{[\zeta_a,\zeta_b] - C^c_{ab}\zeta_c\} = \{[\zeta_a,\zeta_b] - C^c_{ab}\zeta_c - \mathbf{S}_{ab}\}p_i, \tag{33}$$

$$\frac{\partial}{\partial p_i}\{[\zeta_a,\zeta_b] - C^c_{ab}\zeta_c\} = -\{[\zeta_a,\zeta_b] - C^c_{ab}\zeta_c - \mathbf{S}_{ab}\}x^i, \tag{34}$$

where $C^c_{ab}(x^i, p_j)$ and $\mathbf{S}_{ab}$ is defined by

$$\mathbf{S}_{ab} = \zeta_c\left(\frac{\partial C^c_{ab}}{\partial x^i}\frac{\partial}{\partial p_i} - \frac{\partial C^c_{ab}}{\partial p_i}\frac{\partial}{\partial x^i}\right). \tag{35}$$

The Einstein convention on summation is considered and $[\ ,\ ]$ denotes the Poisson (Lie) bracket applied to a couple of functions (generators). Therefore, if the first integrals ζ_a generate the Poisson algebra

$$[\zeta_a,\zeta_b] = C^c_{ab}\zeta_c, \tag{36}$$

from Eqs. (33) and (34) one obtains the following Lie brackets, corresponding to the symmetries of the Hamiltonian,

$$[\zeta_a,\zeta_b] = C^c_{ab}\zeta_c + \mathbf{S}_{ab}. \tag{37}$$

By applying all of this to the first integrals J_i, A_j [see Eqs. (30)], we obtain generators J_i, A_j—via Eqs. (31)—which satisfy the following relations, according to Eq. (37),

$$[\mathbf{J}_i,\mathbf{J}_j] = \epsilon_{ij}{}^k\mathbf{J}_k, \quad [\mathbf{J}_i,\mathbf{A}_j] = \epsilon_{ij}{}^k\mathbf{A}_k,$$
$$[\mathbf{A}_i,\mathbf{A}_j] = -(2/\mu)\epsilon_{ij}{}^k(E\mathbf{J}_k + J_k\mathbf{H}), \tag{38}$$

where **H** is the generator associated to the Hamiltonian H. The algebra (38) is the same that appears in the Keplerian case, but now E and J^i represent post-Newtonian quantities.

Obviously, if we apply the previous technique to the first integrals J_i, N_j, where $N_j \equiv A_j$ for $E \equiv 0$ and $N_j \equiv ((2/\mu)|E|)^{-1/2}A_j$ for $E \neq 0$, we obtain generators J_i, N_j, which satisfy

$$[\mathbf{J}_i,\mathbf{J}_j] = \epsilon_{ij}{}^k\mathbf{J}_k, \quad [\mathbf{J}_i,\mathbf{N}_j] = \epsilon_{ij}{}^k\mathbf{N}_k,$$
$$[\mathbf{N}_i,\mathbf{N}_j] = K\epsilon_{ij}{}^k\mathbf{J}_k, \tag{39}$$

with $K \gtrless 0$ for $E \lessgtr 0$, respectively. Trivially, the same relations appear at the Newtonian level and represent the invariance of H by the group $T(3)\times SO(3)$ $(E \equiv 0)$, $SO(3,1)(E > 0)$, and $SO(4)$ $(E < 0)$, respectively.

V. CONCLUSIONS

We have considered a two-body relativistic system whose equation of the relative motion can be derived from a post-Newtonian Hamiltonian depending on four parameters. For particular values of these parameters we recoup the Darwin and EIH Hamiltonians, in the center-of-mass frame, which apply for the electromagnetic and gravitational interactions, respectively.

We have proven, in Sec. II, that it is possible to extend the Runge–Lenz vector to the post-Newtonian order and given an analytical expression **A**, see Eqs. (8), (14), and (15), in which appear two arbitrary functions. We fix such an arbitrariness [see Eqs. (28) and (29)] by impossing a couple of conditions: one is related to the modulus of the vector **A** and the other takes into account the simplest writing of the equation of the relative orbit as referred to **A**.

In Sec. III we have given the explicit form of the relative orbit, which, in general, is a perturbed conic characterized by a perihelion advance given by Eq. (24).

With respect to a possible breaking of the Keplerian symmetries we have answered in the negative [see Sec. IV] because the associated symmetries (of the Hamiltonian) to **A** generate the same algebra that appears at the Newtonian level [Eq. (38)].

It remains as an unsolved problem the question of a possible extension at a higher order than the post-Newtonian level. We conjecture that this will not be possible up to the c^{-3} order, due to the manifestation of radiative effects for certain types of interactions (e.g., electromagnetism). For certain interactions where these radiative effects appear at higher order (e.g., order c^{-5} for gravitation), perhaps one can find a global first integral vector (smoothly defined over all the phase space) being a $(n-1)/2$ post-Newtonian extension of the Runge–Lenz vector if radiation phenomena appear at order c^{-n}.

[1]C. G. Darwin, Philos. Mag. **39**, 537 (1920).
[2]L. Landau and E. Lifchitz, *Theorie des champs* (Mir, Moscow, 1970), p. 225.
[3]A. Einstein, L. Infeld, and D. Hoffmann, Ann. Math. **39**, 65 (1938).
[4]Reference 2, p. 437.
[5]E. J. Saletan and A. H. Cromer, *Theoretical Mechanics* (Wiley, New York, 1971), pp. 221–224.

Geometrical models for quantum logics with conditioning

Platon C. Deliyannis
Department of Mathematics, Illinois Institute of Technology, Chicago, Illinois 60616

(Received 31 January 1984; accepted for publication 2 March 1984)

After obtaining a representation of a quantum logic by means of projection operators on the state space, geometrical conditions are imposed on a cone in an abstract Banach space which allow us to show that certain projections leaving this cone invariant will form a quantum logic with conditioning. Several examples are also presented.

PACS numbers: 03.65.Bz

I. INTRODUCTION

Logics in which a state conditioning process has been defined are important because it is only on this background that the study of general stochastic processes can be based. In abstract terms such logics have been studied by Pool in Refs. 1 and 2, and the present writer in Ref. 3 and 4. The results in this last paper will be used in Sec. II to obtain a geometrical model for such logics, i.e., a representation of the events in a conditioned logic by projection operators on a Banach space leaving the state cone invariant. This geometric background becomes the motivation for the rest of the paper, whose main aim is to show how logics with conditioning can be constructed geometrically.

This construction can, of course, go through only in certain spaces; the exact hypotheses we use are stated in Sec. III, where some elementary consequences are also drawn. In Sec. IV the projections which will form the logic $\mathscr{L}$ are defined and $\mathscr{L}$ is shown to be a complete lattice. Complements are obtained in Sec. V and in Sec. VI the action of $\mathscr{L}$ on the dual space, a useful technical tool, is studied. Orthomodularity of $\mathscr{L}$ is the main result in Sec. VII and the states of $\mathscr{L}$ are investigated in Sec. VIII. The Appendix is devoted to various examples, among which the classical Hilbert space model and a commutative case are included.

Logics consisting of projection operators on a normed space leaving a given cone invariant, analogous to those we construct here, have been studied by Alfsen and Shultz in Ref. 5, as part of what they call noncommutative spectral theory. The connection between their approach and the present exposition is not yet clear. Their basic assumption is a "projection axiom" which, roughly speaking, postulates the existence of enough projections of the desired type. On the contrary, in our approach we impose conditions of straightforward geometrical nature on the cone which is to be left invariant and deduce the existence of the projections. We have been unable as yet to show that the projection axiom follows from our assumptions, but on the other hand we know of no example to the contrary.

Having represented the events by means of idempotent operators on the space $\mathscr{V}$, one is tempted to proceed and extend this representation to observables as well. This is quite straightforward to do, at least for bounded observables. There are several difficulties to overcome, however, before a complete theory can be developed. The last result in Sec. II, for example, shows that the spectral calculus will not work the usual way; on the other hand it is not hard to show that the nonzero elements of the point spectrum of an observable are the nonzero eigenvalues of the corresponding operator. Another possibility that would preserve the spectral calculus is to represent events and observables as functionals on $\mathscr{V}$. This stumbles across the unsolved question as to whether or not an observable is completely determined by its expected values in all states. We shall therefore refrain from presenting in detail any of the above ideas.

II. THE GEOMETRICAL MODEL

First we list the axioms we need to build the geometrical basis.

Consider given an ortholattice $\mathscr{L}$, write $\mathscr{M}$ for the convex set of all states and $\mathscr{P}$ for the set of pure states (extreme points of $\mathscr{M}$). The support of a state m is defined to be the event $L_m = \inf\{A \mid mA = 1\}$ (whenever it exists). We impose the following structure on the set $\mathscr{M}$.

(M1): (i) $\mathscr{M}$ is quite full; that is, if $mA = 1$ for all $m \in \mathscr{M}$ for which $mB = 1$, then $B \leqslant A$.

(ii) Every state m has a support L_m and $m(L_m) = 1$.

(iii) Every $m \in \mathscr{M}$ is a countable mixture of pure states: $m = \Sigma_{i=1}^{\infty} a_i p_i$, where $a_i > 0$, $\Sigma_{i=1}^{\infty} a_i = 1$, $p_i \in \mathscr{P}$ and convergence is pointwise on $\mathscr{L}$.

(M2): The map $(p,q) \to \langle p|q\rangle = p(L_q)$ defines a transition probability functional on $\mathscr{P} \times \mathscr{P}$, i.e., has range in the interval [0,1] and satisfies

(i) $\langle p|q\rangle = \langle q|p\rangle$;

(ii) $\langle p|q\rangle = 1$, if and only if $p = q$.

As we shall see presently, the map $\langle\ |\ \rangle$ extends by convexity of $\mathscr{M} \times \mathscr{M}$. Keeping this in mind we state the following.

(M3): A conditioning process is defined on $\mathscr{M}$, i.e., to each event $A \in \mathscr{L}$ there corresponds a map $m \to m_{:A}$ from $\mathscr{M} \cup \{\theta\}$ to itself (where θ is a formal "vacuous state" $\notin \mathscr{M}$) with the following properties.

(i) $m_{:A} = \theta$, if and only if $\theta A = 0$; also $\theta_{:A} = \theta$.

(ii) $(m_{:A})_{:A} = m_{:A}$.

(iii) $m_{:A} = m$, if and only if $mA = 1$.

(iv) If $nA = 1$, then $\langle m|n\rangle = m(A)\,\langle m_{:A}|n\rangle$.

We shall call a conditioning "pure" if $m_{:A} \in \mathscr{P} \cup \{\theta\}$ for all events A and all $m \in \mathscr{P}$.

Before proceeding with our construction we should verify that the functional $\langle\ |\ \rangle$ extends from $\mathscr{P} \times \mathscr{P}$ to $\mathscr{M} \times \mathscr{M}$. To this end first note that if $\Sigma a_i p_i$ and $\Sigma b_j q_j$ are two ways of writing a state as a mixture, then $\Sigma a_i \langle p_i|r\rangle = \Sigma b_j \langle q_j|r\rangle$ for

 0022-2488/84/102939-08$02.50

all $r \in \mathscr{P}$. Now let $m = \Sigma a_i p_i = \Sigma b_j q_j$ and $n = \Sigma c_k r_k = \Sigma d_l s_l$; then we have $\Sigma_{i,k} a_i c_k \langle p_i | r_k \rangle = \Sigma_{j,k} b_j c_k \langle q_j | r_k \rangle = \Sigma_j b_j \Sigma_k c_k \langle r_k | q_j \rangle = \Sigma_j b_j \Sigma_l d_l \langle s_l | q_j \rangle = \Sigma_{j,l} b_j d_l \langle q_j | s_l \rangle$, which shows that $\langle m|n \rangle$ can be defined unambiguously as $\Sigma_{i,k} a_i c_k \langle q_i | r_k \rangle$. Trivially this functional on $\mathscr{M} \times \mathscr{M}$ is symmetric and convex in each variable.

It is shown in Ref. 4 that the results therein follow from the above axioms; so we shall not hesitate to use those that we shall need.

The vector space we shall use is the set $\mathscr{V}$ of all maps $\mathscr{L} \to \mathbb{R}$ of the form $am - bn$, where $a,b \geqslant 0$ and $m,n \in \mathscr{M}$. Thus every vector in $\mathscr{V}$ is a countably additive finite signed measure on $\mathscr{L}$. We shall identify the vacuous state θ of Axiom M3 with the zero vector in $\mathscr{V}$. We write $\mathscr{C}$ for the convex cone with base $\mathscr{M}$.

Note that finite convex combinations of vectors in $\mathscr{M}$ correspond to mixtures. More can be said, as we shall presently see.

We equip $\mathscr{V}$ with the base norm $\| \ \|_1$ generated by $\mathscr{M}$: $\|u\|_1 = \inf\{a + b \mid u = am - bn;\ a,b \geqslant 0;\ m,n \in \mathscr{M}\}$, and observe that $\|m\|_1 = 1$ for all $m \in \mathscr{M}$.

Proposition 2.1: If $m = \Sigma_i a_i m_i$ is a mixture, then the series converges to m in the $\| \ \|_1$.

Proof: We have $m - \Sigma_{i=1}^{N} a_i m_i = \Sigma_{i=N+1}^{\infty} a_i m_i = (\Sigma_{i=N+1}^{\infty} a_i) n_N$, where n_N is the state $\Sigma_{i=N+1}^{\infty} a_i / (\Sigma_{j=N+1}^{\infty} a_j)\, m_i$; thus $\|m - \Sigma_{i=1}^{N} a_i m_i\|_1 = \Sigma_{i=N+1}^{\infty} a_i$ which approaches 0 as $N \to \infty$.

Theorem 2.2: $\mathscr{V}$ is complete under $\| \ \|_1$ and $\mathscr{C}$ is closed in $\mathscr{V}$.

Proof: Let $(u_k)_{k \in \mathbb{N}}$ be a Cauchy sequence, select $k_1 < k_2 < \cdots$ so that $\|u_{k_{N+1}} - u_{k_N}\|_1 < 1/2^N$ and write v_N for u_{k_N}; it suffices to show that the sequence $(v_N)_{N \in \mathbb{N}}$ converges. We next select $a_N, b_N \geqslant 0$ and $m_N, n_N \in \mathscr{M}$ such that $v_{N+1} - v_N = a_N m_N - b_N n_N$ and $a_N + b_N < 1/2^{N-1}$, which is possible by the definition of the norm. Then we write $a = \Sigma_N a_N$, $b = \Sigma_N b_N$ and $m = \Sigma_N (a_N/a) m_N$, $n = \Sigma_N (b_N/b) n_N$ and note that $v = am - bn \in \mathscr{V}$; on the other hand, $\|v - v_N\|_1 \leqslant \Sigma_{k=N-1}^{\infty} (a_k + b_k)$ which approaches zero. To see that $\mathscr{C}$ is closed, let $u_k \in \mathscr{C}$ and $\lim \|u - u_k\|_1 = 0$; this implies that we can select $a_k, b_k \geqslant 0$ and $m_k, n_k \in \mathscr{M}$ so that $u - u_k = a_k m_k - b_k n_k$ and $a_k + b_k$ goes to zero. Then $u = u_k + a_k m_k - b_k n_k$ and $u(A) = u_k(A) + a_k m_k(A) - b_k n_k(A) \geqslant u_k(A) - a_k - b_k$, which implies that $u(A) \geqslant \liminf_{k \to \infty} u_k(A) \geqslant 0$, since $u_k \in \mathscr{C}$. Thus u is a positive countably additive measure on $\mathscr{L}$ and thus in $\mathscr{C}$.

Theorem 2.3: The functional $\langle \ | \ \rangle$ extends to a bilinear nondegenerate functional on $\mathscr{V} \times \mathscr{V}$ whose restriction to $\mathscr{C} \times \mathscr{C}$ is nonnegative, and such that $|\langle u|v \rangle| \leqslant \|u\|_1 \|v\|_1$. This functional is an inner product if and only if for all m, $n \in \mathscr{M}$ we have $\langle m|n \rangle^2 \leqslant \langle m|m \rangle \langle n|n \rangle$.

Proof: Let $u_i = a_i m_i - b_i n_i$ with $a_i, b_i \geqslant 0$ and $m_i, n_i \in \mathscr{M}$ $(i = 1,2)$. If $u_1 = u_2$ then, evaluating at I we get $a_1 - b_1 = a_2 - b_2$ and so we have $[a_1/(a_1 + b_2)] m_1 + [b_2/(a_1 + b_2)] n_2 = [a_2/(a_2 + b_1)] m_2 + [b_1/(a_2 + b_1)] n_1 \in \mathscr{M}$; thus, for any $m \in \mathscr{M}$ we obtain $a_1 \langle m_1|m \rangle + b_2 \langle n_2|m \rangle = a_2 \langle m_2|m \rangle + b_1 \langle n_1|m \rangle$, or $a_1 \langle m_1|m \rangle - b_1 \langle n_1|m \rangle = a_2 \langle m_2|m \rangle - b_2 \langle n_2|m \rangle$. Using this it follows at once that for arbitrary u_1 and u_2 we can define $\langle u_1|u_2 \rangle$ unambiguously as $a_1 a_2 \langle m_1|m_2 \rangle - a_1 b_2 \langle m_1|n_2 \rangle - a_1 b_2 \langle m_1|n_2 \rangle + b_1 b_2 \langle n_1|n_2 \rangle$ and that the resulting functional is bilinear. Now take u, $v \in \mathscr{C}$; we have $u = am$, $v = bn$ with a, $b \geqslant 0$ and $m,n \in \mathscr{M}$ so $\langle u|v \rangle = ab \langle m|n \rangle \geqslant 0$. If $\langle u|v \rangle = 0$ for all $v \in \mathscr{V}$, then $\langle u|p \rangle = 0$ for all $p \in \mathscr{P}$ and so $u(L_p) = 0$; but then, using propositions 2 and 4 of Ref. 4 we obtain $u(A) = 0$ for all $A \in \mathscr{L}$, i.e., $u = 0$; thus $\langle \ | \ \rangle$ is nondegenerate. Finally, since $\langle am - bn|am - bn \rangle = a^2 \langle m|m \rangle - 2ab \langle m|n \rangle + b^2 \langle n|n \rangle$, we see that $\langle u|u \rangle \geqslant 0$ for all $u = am - bn$ if and only if the discriminant $\langle m|n \rangle^2 - \langle m|m \rangle \langle n|n \rangle$ is nonpositive.

Axiom (M3) will now be used to obtain a representation of the events in $\mathscr{L}$ as operators on $\mathscr{V}$.

Theorem 2.4: To each event $A \in \mathscr{L}$ there corresponds an operator $P_A : \mathscr{V} \to \mathscr{V}$ such that (i) $P_A^2 = P_A$; (ii) $\langle P_A u|v \rangle = \langle u|P_A v \rangle$ (we call this symmetry); and (iii) $P_A \mathscr{C} \subsetneq \mathscr{C}$.

Proof: We first define P_A on $\mathscr{C}$ by $P_A(am) = am(A)\, m_{:A}$, and note that trivially $P_A(bu) = bP_A(u)$, and that (iii) holds; by (M3) (ii) we also have $P_A^2 = P_A$. Proposition 5 of Ref. 4 now gives property (ii) above for u, $v \in \mathscr{M}$, and hence for $u,v \in \mathscr{C}$. Now observe that $\langle P_A(u + v)|m \rangle = \langle u + v|P_A m \rangle = \langle u|P_A m \rangle + \langle v|P_A m \rangle = \langle P_A u|m \rangle + \langle P_A v|m \rangle = \langle P_A u + P_A v|m \rangle$ for $u,v,m \in \mathscr{C}$, which shows that P_A is also additive on the cone; but $\mathscr{C}$ generates $\mathscr{V}$ and thus P_A extends by linearity to $\mathscr{V}$, retaining properties (i) and (ii).

The algebraic structure of $\mathscr{L}$ can be completely described by means of these operators. Using the results in Sec. 4 of Ref. 4 we obtain easily the following.

(1) $A \leqslant B$ if and only if $P_A \leqslant P_B$ in the sense that $P_A = P_A P_B$.

(2) $P_{A'}$ is the largest [in the sense of the partial order defined in (1)] element P_B whose product with P_A is zero.

(3) The events A and B are disjoint if and only if $P_A P_B = 0$.

(4) The events A and B are compatible if and only if P_A and P_B commute. In such a case $P_{A \wedge B} = P_A P_B$.

It is also straightforward to see that in general we have $P_{A \wedge B} = P_A \wedge P_B$, where the infimum on the operators is taken with respect to the partial ordering defined in (1). In case $\langle \ | \ \rangle$ in an inner product, a more explicit formula for $P_{A \wedge B}$ can be obtained by working in the completion of $\mathscr{V}$ with respect to the norm $\|u\|_2 = \sqrt{\langle u|u \rangle}$.

(5) If the functional $\langle \ | \ \rangle$ is an inner product, then $P_{A \wedge B}$ is the limit of $(P_A P_B)^k$, as $k \to \infty$, in the strong topology. The probability $m(A)$ of occurrence of the event A in the state m is given, geometrically, by the length $\|P_A m\|_1$. In case of a pure conditioning, a useful alternative exists.

(6) If the conditioning is pure, then for $p \in \mathscr{P}$ we have $p(A) = \sqrt{\langle P_A p|p \rangle}$.

To see this we apply (M2) (iii) and (M3) (iv): since $p_{:A}(A) = 1$, we have $\langle p|p_{:A} \rangle = p(A) \langle p_{:A}|p_{:A} \rangle = p(A)$, and so $p(A)^2 = \langle P_A p|p \rangle$.

Formulas for the supremum are difficult to obtain, even in the simplest case of disjoint events.

(7) Let the conditioning be pure and the functional $\langle \ | \ \rangle$ be an inner product. Then, for A, $B \in \mathscr{L}$ disjoint, we have $P_{A \vee B} = P_A + P_B + (I - P_{A'})(I - P_B)$.

Write Q for the right-hand side and note that the two factors of the third term are commuting idempotents, and so the three terms of Q are orthogonal idempotents; thus Q is an idempotent, and as each term is symmetric, Q is also. By direction computation, using the obvious facts that $P_{A \vee B}P_A = P_A P_{A \vee B} P_{A'} = P_B$, etc., we obtain that $P_{A \vee B}Q = P_{A \vee B}$. This implies in particular that $Q - P_{A \vee B}$ is a symmetric idempotent. Noting that according to (4) above we have $P_{A'}P_{B'} = P_{A' \wedge B'} = P_{(A \vee B)'}$, take a pure state p and compute $\langle Qp|p\rangle = \langle P_A p|p\rangle + \langle P_B p|p\rangle + \langle p|p\rangle - \langle P_{A'}p|p\rangle - \langle P_{B'}p|p\rangle + \langle P_{(A \vee B)'}p|p\rangle = p(A)^2 + p(B)^2 + 1 - p(A')^2 - p(B')^2 + p((A \vee B)')^2 = p(A)^2 + p(B)^2 + 1 - [1 - p(A)]^2 - [1 - p(B)]^2 + [1 - p(A \vee B)]^2$; using the fact that $p(A \vee B) = p(A) + p(B)$ (since A and B are disjoint), we see that the last expression simplifies to $p(A \vee B)^2 = \langle P_{A \vee B}p|p\rangle$. Thus $\langle Qp|p\rangle = \langle P_{A \vee B}p|p\rangle$ or $\langle (Q - P_{A \vee B})p|p\rangle = 0$, hence $\|Qp - P_{A \vee B}p\|_2 = 0$. So, for any pure state p we have $Qp = P_{A \vee B}p$, and since Q and $P_{A \vee B}$ are continuous in the topology of $\| \ \|_1$, Axiom M1 (iii) and Proposition 2.1 imply the same relation for $p \in \mathscr{M}$, hence for $p \in \mathscr{C}$, and thus for all vectors in $\mathscr{V}$.

III. THE GEOMETRICAL HYPOTHESES

From this point on we shall be concerned with the reverse question: what conditions need we impose on a normed space with a cone and a bilinear functional so that it is possible to construct a conditioned logic consisting of symmetric idempotents leaving the cone invariant.

In this section we shall list the conditions we need and draw some basic preliminary conclusions. Our hypotheses fall in two groups. Those that we call necessary, because they are valid in the spaces we constructed in the previous section, and the others which may or may not be valid in these spaces, but are apparently essential—these we call sufficient.

We write $\mathscr{V}$ for a given vector space over the reals $\mathbb{R}$, $\mathscr{C}$ for a cone in $\mathscr{V}$; $\mathscr{M}$ is a base for $\mathscr{C}$, $\mathscr{P}$ the set of extreme points of $\mathscr{M}$, and $\mathscr{C}$ the set $\{ap|a \geqslant 0; p \in \mathscr{P}\}$.

Necessary hypotheses

(H1) The cone $\mathscr{C}$ generates the space $\mathscr{V}$ and has a convex base $\mathscr{M}$ which defines the base norm $\| \ \|_1$.

(H2) $\mathscr{V}$ is complete with respect to $\| \ \|_1$ and $\mathscr{C}$ is closed in $\mathscr{V}$.

(H3) The bilinear nondegenerate functional $\langle \ | \ \rangle$ restricted to $\mathscr{C} \times \mathscr{C}$ takes non-negative values.

(H4) For all $u, v \in \mathscr{V}$ we have $|\langle u|v\rangle| \leqslant \|u\|_1 \|v\|_1$.

(H5) For each $p \in \mathscr{P}$ we have $\langle p|p\rangle = 1$.

(H6) For each $p \in \mathscr{P}$ write $p^\perp$ for the set $\{r \in \mathscr{P} | \langle r|p\rangle = 0\}$; then the only vector in $\mathscr{P}$ orthogonal to all vectors in $p^\perp$ is the vector p itself.

(H7) Each $m \in \mathscr{M}$ has the form $m = \Sigma_{i=1}^{\infty} a_i p_i$ with $p_i \in \mathscr{P}$, $a_i > 0$, $\Sigma_{i=1}^{\infty} a_i = 1$ and where convergence is in the $\| \ \|_1$.

(H8) For each $q \in \mathscr{P}$ there is a unique $r_0 \in p^\perp$ which maximizes $\langle r|q\rangle$ over $r \in p^\perp$; this r_0 is such that $q - \langle q|r_0\rangle r_0$ is orthogonal to all $r \in p^\perp$.

Sufficient hypotheses

(H9) The functional $\langle \ | \ \rangle$ is an inner product such the metric it defines on $\mathscr{P}$ is equivalent to that defined by $\| \ \|_1$.

(H10) For each $p \in \mathscr{P}$ let $\mathscr{K}_p = \{v \in \mathscr{V} | \langle v|q\rangle = 0$ for all $q \in p^\perp\}$. Then every $\| \ \|_1$-closed hyperplane supporting $\mathscr{C}$ at p contains the space $\mathscr{K}_p$.

Several examples are presented in the Appendix.

As in Sec. II we shall write $\|u\|_2$ for $\sqrt{\langle u|u\rangle}$.

Lemma 3.1: The set $\mathscr{P}$ is complete in the $\| \ \|_2$.

Proof: We shall show that $\mathscr{P}$ is complete in the $\| \ \|_1$ and then apply (H9). All we need then is that $\mathscr{P}$ is closed in the $\| \ \|_1$, since $\mathscr{V}$ is complete with respect to this norm. So let $\|u - p_k\|_1 \to 0$ with $p_k \in \mathscr{P}$; since $\mathscr{M}$ is the intersection of $\mathscr{C}$ and the unit $\| \ \|_1$-sphere, we see that $\mathscr{M}$ is $\| \ \|_1$ closed, and so $u \in \mathscr{M}$. By (H4) we have $\langle u|u\rangle = \lim_k \langle p_k|p_k\rangle = 1$; but since $u \in \mathscr{M}$, (H7) gives $u = \Sigma_{j=1}^{\infty} b_j q_j$ with $b_j > 0$, $\Sigma_{j=1}^{\infty} b_j = 1$, and q_j distinct $\in \mathscr{P}$, which implies $\Sigma_{j,k} b_j b_k \langle q_j|q_k\rangle = 1 = (\Sigma_j b_j)(\Sigma_k b_k)$, i.e., that $\Sigma_{j,k} b_j b_k [1 - \langle q_j|q_k\rangle] = 0$, and since all terms are $\geqslant 0$, that $b_j b_k [1 - \langle q_j|q_k\rangle] = 0$ for all j, k. Thus, if $b_{j_0} \neq 0$ we have $b_j = 0$ for all $j \neq j_0$ since the q_j are all distinct and therefore $\langle q_j|q_k\rangle \neq 1$ for $j \neq k$. Since $u \neq 0$ ($\|u\|_2 = 1$!) we have $u = q_{j_0} \in \mathscr{P}$.

Lemma 3.2: If $u_k \in \mathscr{C}$ and $\|u - u_k\|_2 \to 0$, then $\|u - u_k\|_1 \to 0$ also.

Proof: We have $u_k = a_k p_k$ with $a_k \geqslant 0$ and $p_k \in \mathscr{P}$. Since $\|p_k\|_2 = 1$, $\lim a_k$ exists and we call it a; if $a = 0$ then $u = 0$ and $\|u - u_k\|_1 = a_k \to 0$. So let $a \neq 0$ which implies $\|p_k - u/a\|_2 \to 0$ whence, by (H9), $\|p_k - u/a\|_1 \to 0$, i.e., $\|u - u_k\|_1 \to 0$.

In the previous section we had no opportunity to verify that (H6) and (H8) hold. We do this now.

First (H6). Take any $p \in \mathscr{P}$ and note that $p^\perp = \{r \in \mathscr{P} | L_r \leqslant L'_p\}$. Using Proposition 2 of Ref. 4 we see that $m(L'_p) = \Sigma_i m(L_{r_i})$ for suitable $r_i \in p^\perp$; thus, if q is orthogonal to all $r \in p^\perp$, we have $q(L_{r_i}) = 0$ and so $q(L'_p) = 0$, or $q(L_p) = 1$, i.e., $q = p$.

For (H8) we shall need to assume a pure conditioning. First note that for any $p \in \mathscr{P}$ we have $p(A) = \langle p|p_{:A}\rangle$, because by (M3) (iv) we have $\langle p|p_{:A}\rangle = p(A)\langle p_{:A}|p_{:A}\rangle$ and we have assumed $p_{:A} \in \mathscr{P}$. But then, again by (M3) (iv), we obtain $\langle q|r\rangle = \langle q|q_{:A}\rangle\langle q_{:A}|r\rangle$ for any $r \in \mathscr{P}$ for which $r(A) = 1$; thus for all such r we have $\langle q|r\rangle \leqslant \langle q|q_{:A}\rangle$. Specializing to the case where $A = L'_p$, we see that $\langle q|r\rangle$ ($r \in p^\perp$) is maximized for $r = q_{:L'_p}$. The desired orthogonality relation is just (M3) (iv). As for uniqueness, let $\langle q|q_{:A}\rangle = \langle q|r\rangle$ for some $r \in p^\perp (A = L'_p)$ and note that (M3) (iv) implies either $\langle q|r\rangle = 0$, or $\langle q_{:A}|r\rangle = 1$; the first implies $\langle q|q_{:A}\rangle = 0$, i.e., $q(A) = 0$, or $q = p$, and the second means $r = q_{:A}$.

A few more preliminary remarks.

Note that for any linear T: $\mathscr{V} \to \mathscr{V}$ we have $\|T\|_1 = \sup\{\|Tm\|_1 | m \in \mathscr{M}\}$. Let A be this supremum and observe that $A \leqslant \|T\|_1$, since $\|m\|_1 = 1$; on the other hand, $\|Tu\|_1 = \|a\ Tm - b\ Tn\|_1 \leqslant a\|Tm\|_1 + b\|Tn\|_1 \leqslant (a + b)A$ which implies $\|Tu\|_1 \leqslant \|u\|_1 A$ or $\|T\|_1 \leqslant A$.

We shall write $\overline{\mathscr{V}}$ for the completion of $\mathscr{V}$ under the norm $\| \ \|_2$.

Proposition 3.3: Let $\overline{T}$: $\overline{\mathscr{V}} \to \overline{\mathscr{V}}$ be a linear, $\| \ \|_2$ continuous map such that $\overline{T}\mathscr{C} \subseteq \mathscr{C} \cup \{0\}$. Then $\overline{T}\mathscr{C} \subseteq \mathscr{C}$ (hence $\overline{T}\mathscr{V} \subseteq \mathscr{V}$) and $\|\overline{T}|\mathscr{V}\|_1 \leqslant \|\overline{T}\|_2$.

Proof: Let $m = \Sigma_i a_i p_i$ with $a_i > 0$ and $\Sigma_i a_i = 1, p_i \in \mathscr{P}$; since the series also converges to m in the norm $\| \ \|_2$ we have $\overline{T}m = \Sigma_i a_i \overline{T}p_i = \Sigma_i a_i b_i q_i$ with $b_i \geqslant 0$ and $q_i \in \mathscr{P}$. Since $b_i \leqslant \|\overline{T}\|_2$, we have $\|\Sigma_{i>N} a_i b_i q_i\|_1 \leqslant \Sigma_{i>N} a_i b_i \leqslant (\Sigma_{i>N} a_i)\|\overline{T}\|_2$ which shows that $\Sigma_i a_i b_i q_i$ converges to $\overline{T}m$ in the norm $\| \ \|_1$ also. Since $\mathscr{C}$ is closed for this norm, we see that $\overline{T}m \in \mathscr{C}$, and since the cone generates $\mathscr{V}$, $\overline{T}\mathscr{V} \subseteq \mathscr{V}$. Now look at the above estimates for $N = 1$ to obtain $\|Tm\|_1 \leqslant \|\overline{T}\|_2$, which, combined with the previous remark, shows that $\|\overline{T}\,|\,\mathscr{V}\|_1 \leqslant \|\overline{T}\|_2$.

It will be useful to keep in mind that since $\| \ \|_1$ is the base norm generated by $\mathscr{M}$ it is countably additive on $\mathscr{M}$ and on $\mathscr{C}$.

IV. CONSTRUCTION OF THE LOGIC $\mathscr{L}$

We shall write $\mathscr{L}$ for the set of all operators $P: \mathscr{V} \to \mathscr{V}$ for which

L(i) $P^2 = P$;

L(ii) $\langle Pu|v\rangle = \langle u|Pv\rangle$, for all $u, v \in \mathscr{V}$.

L(iii) $P\mathscr{C} \subseteq \mathscr{C} \cup \{0\}$;

L(iv) $(\mathscr{R}_P \cap \mathscr{P})^{\perp\perp} = \mathscr{R}_P \cap \mathscr{P}$,

where $\mathscr{R}_P$ is the range of the operator P and, for any $S \subseteq \mathscr{P}$ we write $S^\perp$ for the set $\{q \in \mathscr{P} \,|\, \langle q|r\rangle = 0 \text{ for all } r \in S\}$.

Using (H3) and (H6) we see at once that for each $p \in \mathscr{P}$ the operator $P_p: u \to \langle u|p\rangle p$ is in $\mathscr{L}$. Clearly O and $I \in \mathscr{L}$ also.

Proposition 4.1: Any $P \in \mathscr{L}$ extends uniquely by continuity to an orthoprojection $\overline{P}$ on $\overline{\mathscr{V}}$ which leaves $\mathscr{C} \cup \{0\}$ invariant.

Corollary: For any nonzero $P \in \mathscr{L}$ we have $\|P\|_1 = 1$, and $P\mathscr{C} \subseteq \mathscr{C}$.

Proof: Since $P = \overline{P}\,|\,\mathscr{V}$ Proposition 3.3 implies $\|P\|_1 \leqslant \|P\|_2 = 1$; but $P^2 = P$ implies $\|P\|_1^2 \geqslant \|P\|_1$ and since $P \neq 0$ we have $\|P\|_1 \geqslant 1$.

Lemma 4.2: For $0 \neq p \in \mathscr{L}$ and $m \in \mathscr{M}$, we have $Pm = m$ if and only if $\|Pm\|_1 = 1$.

Proof: One way it is obvious, so let $\|Pm\|_1 = 1$ and $m = \Sigma_i a_i p_i$ with $a_i > 0$, $\Sigma_i a_i = 1$ and $p_i \in \mathscr{P}$. Then $1 = \Sigma_i a_i \|Pp_i\|_1$ and so $\|Pp_i\|_1 = 1$; but $Pp_i \in \mathscr{C}$ hence $\|Pp_i\|_2 = 1$ also, which by Schwartz implies $Pp_i = p_i$; but then $Pm = m$ also.

Proposition 4.3: For any $P \in \mathscr{L}$, $\mathscr{R}_P \cap \mathscr{C}$ is a face of $\mathscr{C}$.

Proof: Let $u = a_1 u_1 + a_2 u_2$ with $u_1, u_2 \in \mathscr{C}$, $a_1, a_2 > 0$, $a_1 + a_2 = 1$, and $Pu = u$; without loss we may assume that $\|u\|_1 = 1 = \|u_2\|_1$ since all vectors are in $\mathscr{C}$ and $\| \ \|_1$ is additive. Then we use $u = a_1 Pu_1 + a_2 Pu_2$ to obtain $1 = a_1\|Pu_1\|_1 + a_2\|Pu_2\|_1$ which implies $\|Pu_i\|_1 = 1$, and since $u_i \in \mathscr{M}$, $Pu_i = u_i$.

On the set $\mathscr{L}$ we impose now the partial ordering given by $P \leqslant Q$ if and only if $PQ = P$ (i.e. $\mathscr{R}_P \subseteq \mathscr{R}_Q$). Note that by L(ii) this is equivalent to $P = QP$ and so P and Q commute.

Lemma 4.4: We have $P \subseteq Q$ if and only if for all $m \in \mathscr{M}$, $Pm = m$ implies $Qm = m$.

Proof: One way this is obvious, so we assume that, for all $m \in \mathscr{M}$, $Pm = m$ implies $Qm = m$. Take any $u \in \mathscr{R}_P$, and note that it has the form am-bn with $m, n \in \mathscr{R}_P \cap \mathscr{M}$ [because if $u = a_1 m_1 + b_1 n_1$, then $u = Pu = a_1\|Pm_1\|_1(Pm_1/\|Pm_1\|_1) + a_2\|Pm_2\|_1(Pm_2/\|Pm_2\|_1)$]. But then $m, n \in \mathscr{R}_Q$ and so $Qu = u$. Thus $P \leqslant Q$ follows.

Theorem 4.5: Under the partial order defined above, $\mathscr{L}$ is a complete lattice.

Proof: It suffices to show that any family of operators in $\mathscr{L}$ has an infimum in $\mathscr{L}$. First we do it for two operators, P and Q. We consider their extensions $\overline{P}$ and $\overline{Q}$ to $\overline{\mathscr{V}}$ which exist by Proposition 4.1, and write $\overline{R}$ for their infimum $\overline{P} \wedge \overline{Q}$; we know that $\overline{R}$ is an orthoprojection on $\overline{\mathscr{V}}$ and we shall show that it preserves $\mathscr{C} \cup \{0\}$. We know that $\overline{R}$ is the limit in the weak topology of the Hilbert space $\overline{\mathscr{V}}$ of $(\overline{PQ})^k$; take $p \in \mathscr{P}$ and note that $(\overline{PQ})^k p = a_k p_k$ with $a_k \geqslant 0$ and $p_k \in \mathscr{P}$, since both $\overline{P}$ and $\overline{Q}$ preserve $\mathscr{C} \cup \{0\}$. Since $\|p_k\|_2 = 1$, the limit of (a_k) exists; call it a. If $a = 0$, we are done; if $a \neq 0$, then (p_k) converges to $\overline{R}p/a$ in the $\| \ \|_2$, and since $\mathscr{P}$ is complete in the corresponding distance, we have $\overline{R}p/a \in \mathscr{P}$, hence $\overline{R}p \in \mathscr{C} \cup \{0\}$. By Proposition 3.3 $\overline{R}$ leaves $\mathscr{V}$ invariant; setting $R = \overline{R}\,|\,\mathscr{V}$ we see that R satisfies L(i), L(ii), and L(iii) and trivially $\mathscr{R}_R = \mathscr{R}_P \cap \mathscr{R}_Q$. Thus the proof that R is the infimum of P and Q will be complete as soon as we compute $(\mathscr{R}_R \cap \mathscr{P})^{\perp\perp}$. We observe that $\mathscr{R}_R \cap \mathscr{P} \subseteq (\mathscr{R}_R \cap \mathscr{P})^{\perp\perp}$ and since $\mathscr{R}_R \subseteq \mathscr{R}_P$ we have $(\mathscr{R}_R \cap \mathscr{P})^{\perp\perp} \subseteq (\mathscr{R}_P \cap \mathscr{P})^{\perp\perp} = \mathscr{R}_P \cap \mathscr{P}$ and similarly $(\mathscr{R}_R \cap \mathscr{P})^{\perp\perp} \subseteq \mathscr{R}_Q \cap \mathscr{P}$; thus $(\mathscr{R}_R \cap \mathscr{P})^{\perp\perp} \subseteq \mathscr{R}_P \cap \mathscr{R}_Q \cap \mathscr{P} = \mathscr{R}_R \cap \mathscr{P}$. To finish the proof of the theorem we need to show that if (P_j) is a decreasing net in $\mathscr{L}$, then its infimum exists. The proof of this follows roughly the same lines as the previous argument. We consider the extensions $\overline{P}_j$ to be Hilbert space $\overline{\mathscr{V}}$ and let $\overline{P}$ be the infimum of the $\overline{P}_j$ in $\overline{\mathscr{V}}$; since $\overline{P}$ is the weak limit of the $\overline{P}_j$, we can argue as above to get that $\overline{P}$ leaves $\mathscr{C} \cap \{0\}$ invariant. It is straightforward now to adapt the previous arguments to show that $P = \overline{P}\,|\,\mathscr{V} \in \mathscr{L}$ and is the infimum of the family (P_j).

Theorem 4.6: For any $P \in \mathscr{L}$ we have $P = \sup\{P_p \,|\, p \in \mathscr{R}_P \cap \mathscr{P}\}$.

Proof: For any $p \in \mathscr{R}_P \cap \mathscr{P}$ we have $Pp = p$, hence $PP_p = P_p$ or $P_p \leqslant P$, and P is an upper bound. Now let $P_p \leqslant Q$ for all $p \in \mathscr{R}_P \cap \mathscr{P}$ and suppose that $Pm = m$; write m as a mixture of pure p_i, $m = \Sigma_i a_i p_i$ and apply Proposition 4.3 to get $Pp_i = p_i$, i.e., $p_i \in \mathscr{R}_P \cap \mathscr{P}$. Then $P_{p_i} \leqslant Q$, which means $Qp_i = p_i$, or $Qm = m$.

V. COMPLEMENTARY OPERATORS

The *complement* of an operator $P \in \mathscr{L}$ is an operator $Q \in \mathscr{L}$ such that, for all $m \in \mathscr{M}$, we have $Pm = m$ if and only if $Qm = 0$ and $Pm = 0$ if and only if $Qm = m$.

Using Lemma 4.4 we see that complements, if they exist, are unique. We shall write P' for the complement of P and note that $(P')' = P$ while $P \geqslant Q$ implies $P' \leqslant Q'$ (assuming of course that all the complements exist). Clearly it suffices to test only for $m \in \mathscr{P}$.

Theorem 5.1: For each $p \in \mathscr{P}$, P_p' exists.

Proof: Consider any $q \in \mathscr{P}$, $q \neq p$, and write R_q for the vector $\langle q|r_0\rangle r_0$, where r_0 is as in (H8); define Rp to be the zero vector. By (H8) we have $\langle q|r\rangle = \langle Rq|r\rangle$ for all $r \in p^\perp$, and considering any two $q_1, q_2 \in \mathscr{P}$ we observe that $\langle q_1|Rq_2\rangle = \langle Rq_1|Rq_2\rangle$ as well as $\langle q_2|Rq_1\rangle = \langle Rq_2|Rq_1\rangle$. Thus $\langle q_1|Rq_2\rangle = \langle Rq_1|q_2\rangle$ for any $q_1, q_2 \in \mathscr{P}$. Arguing as in

Sec. II we can extend R to an orthoprojection on $\mathscr{V}$ which evidently preserves $\mathscr{E} \cup \{0\}$; since $(\mathscr{R}_R \cap \mathscr{P})^\perp = p^\perp$ we have L(iv) by using (H6). So we have $R \in \mathscr{L}$. To show that $R = P'_p$, note that $P_p m = m$ implies $m = p$ and so $Rm = 0$; conversely, if $Rm = 0$ we have either $m = p$ or $\langle m|r_0\rangle = 0$, where r_0 maximizes $\langle m|r\rangle$ over $r \in p^\perp$; but then $\langle m|r\rangle = 0$ for all $r \in p^\perp$ and by (H6) we have $m = p$. Finally, $P_p m = 0$ means $\langle m|p\rangle = 0$, or $m \in p^\perp$, i.e., $Rm = m$.

Theorem 5.2: Every element $P \in \mathscr{L}$ has a complement.

Proof: We shall show that $P' = \inf\{P'_p | p \in \mathscr{R}_P \cap \mathscr{P}\}$. Let R be this infimum and suppose that $Rq = q$; then $P'_p q = q$ or $P_p q = 0$ for all $p \in \mathscr{R}_P \cap \mathscr{P}$, and so $\langle q|p\rangle = 0$ for all such p. Since $Pq \in \mathscr{R}_P \cap \mathscr{E}$ we have $\langle q|Pq\rangle = 0$ and so $Pq = 0$. Thus $Rq = q$ implies $Pq = 0$. Conversely, let $Pq = 0$; then $P_p q = 0$ for all $p \in \mathscr{R}_P \cap \mathscr{P}$, hence $P'_p q = q$ for all such p, which implies $Rq = q$. Now let $Pq = q$; then by definition $P'_q \geqslant R$ and so $RP_q = 0$ which means that $Rq = 0$. For the converse note that $r \in (\mathscr{R}_P \cap \mathscr{P})^\perp$ implies $P'_p r = r$ for all $p \in \mathscr{R}_P \cap \mathscr{P}$ hence $Rr = r$, i.e., note that we have $(\mathscr{R}_P \cap \mathscr{P})^\perp \subseteq \mathscr{R}_R \cap \mathscr{P}$. Now, if $Rq = 0$, then $q \in (\mathscr{R}_R \cap \mathscr{P})^\perp$, hence $q \in (\mathscr{R}_P \cap \mathscr{P})^{\perp\perp} = \mathscr{R}_P \cap \mathscr{P}$, or $Pq = q$.

VI. ACTION ON THE DUAL SPACE

A linear functional f on $\mathscr{V}$ is *positive* if $f(u) \geqslant 0$ for all $u \in \mathscr{C}$. We shall write $\mathscr{C}^*$ for the set of all positive functionals which are continuous with respect to the norm $\| \ \|_1$, and $\mathscr{V}^*$ for the vector space $\mathscr{C}^* - \mathscr{C}^*$. Clearly $\mathscr{V}^*$ is a subspace of the dual space of $\mathscr{V}$.

We now define a special useful functional in $\mathscr{V}^*$. If a, $b \geqslant 0$ and $m, n \in \mathscr{M}$ we see that $am - bn = 0$ implies $a - b = 0$; thus the map $t: am - bn \to a - b$ is single valued and, by the same token, linear. Clearly, t is positive, $\|t\|_1 = 1$ and it extends $\| \ \|_1 | \mathscr{C}$.

The cone $\mathscr{C}^*$ defines a natural partial order in $\mathscr{V}^*$: $f \leqslant g$ if and only if $g - f \in \mathscr{C}^*$, i.e., if and only if $f(m) \leqslant g(m)$ for all $m \in \mathscr{M}$.

Lemma 6.1: The functional t is an order unit in $\mathscr{V}^*$.

Proof: Since we have $|f(u)| \leqslant \|f\|_1 \|u\|_1$ for all u, we obtain $|f(u)| \leqslant \|f\|t(u)$ for $u \in \mathscr{C}$, and so $-\|f\|_1 t \leqslant f \leqslant \|f\|_1 t$.

Lemma 6.2: For all $f \in \mathscr{C}^*$ and all $\| \ \|_1$ continuous operators T on $\mathscr{V}$ leaving $\mathscr{C}$ invariant we have $T^* f (\equiv f \circ T)$ in $\mathscr{V}^*$.

Proof: We have $f = g - h$ with $g, h \in \mathscr{C}^*$; since $T\mathscr{C} \subseteq \mathscr{C}$ we have $T^* g$ and $T^* h$ in $\mathscr{C}^*$, and so $T^* f \in \mathscr{C}^* - \mathscr{C}^*$.

Lemma 6.3: If $P, Q \in \mathscr{L}$ and $P^* t = Q^* t$, then $P = Q$.

Proof: Let $m \in \mathscr{M}$ and $Pm = m$; then $1 = \|Pm\|_1 = t(Pm) = (P^* t)(m)$ $= (Q^* t)(m) = t(Qm) = \|Qm\|_1$, hence by Lemma 4.2 we have $Qm = m$, i.e., $P \leqslant Q$; similarly $Q \leqslant P$, and so $P = Q$.

Lemma 6.4: The following two properties of an operator $P \in \mathscr{L}$ are equivalent.

(i) For all $f \in \mathscr{C}^*$, $f | \mathscr{N}_{P'} \cap \mathscr{C} = 0$ implies $f | \mathscr{N}_{P'} = 0$.

(ii) For all $f \in \mathscr{C}^*$, $P^* f = 0$ implies $P'^* f = f$.

Here $\mathscr{N}_T$ denotes the null space of the map T.

Proof: We assume (i) and $P^* f = 0$ with $f \in \mathscr{C}^*$. Then $f(Pu) = 0$ for all $u \in \mathscr{V}$ and so $f(v) = 0$ for all $v \in \mathscr{R}_P$. In particular, for any $m \in \mathscr{M}$, $P'm = 0$ implies $f(m) = 0$ since $Pm = m$. But this means that $f | \mathscr{N}_{P'} \cap \mathscr{C} = 0$, and thus $f | \mathscr{N}_{P'} = 0$ holds; since for any $u \in \mathscr{V}$ we have $u - P'u \in \mathscr{N}_{P'}$, we obtain $f(u) - f(P'u) = 0$, i.e., $f = P'^* f$. Now we assume (ii) and $f | \mathscr{N}_{P'} \cap \mathscr{C} = 0$, which means that $f(m) = 0$ for all $m \in \mathscr{M}$ for which $P'm = 0$. Then $f(Pm) = 0$ for any $m \in \mathscr{M}$, and so $P^* f = 0$. But this implies $P'^* f = f$, or $f(P'u) = f(u)$; in particular, $P'u = 0$ implies $f(u) = 0$, i.e., $f | \mathscr{N}_{P'} = 0$.

Lemma 6.5: Properties (i) and (ii) stated in Lemma 6.4 are stable under suprema.

Proof: First for two operators. Suppose that (ii) holds for P and Q, and suppose that $(P \vee Q)^* f = 0$; since $(P \vee Q)P = P$, we get $P^* f = 0$ and similarly $Q^* f = 0$. Thus $P'^* f = f$ and $Q'^* f = f$, from which we obtain $f = [(P'Q')^k]^* f$; so, for any $p \in \mathscr{P}$ we have $f(p) = f((P'Q')^k p)$. But $(P'Q')^k p \in \mathscr{E}$ and converges to $(P' \wedge Q')p$ in the norm $\| \ \|_2$ and hence by Lemma 3.2 in the norm $\| \ \|_1$ also. Since f is continuous for this norm, we have $f(p) = f((P' \wedge Q')p)$ which implies $(P' \wedge Q')^* f = f$; since $P' \wedge Q' = (P \wedge Q)'$ property (ii) holds for $P \vee Q$. In a similar way, we assume that (ii) holds for each P_j and show that it holds for $P = \sup(P_j)$, where (P_j) is increasing. To this end, $P^* f = f$ and note that it implies $P_j^* f = 0$, whence $P'_j{}^* f = f$. For $p \in \mathscr{P}$ we have again that $P'_j p$ converges in the norm $\| \ \|_2$ to $P'p$ (since P'_j form a decreasing net with infimum P') and thus also in the norm $\| \ \|_1$ and so we have $P'^* f = f$, just as before.

Theorem 6.6: For any $P \in \mathscr{L}$ and any $f \in \mathscr{C}^*$ we have that $P^* f = 0$ implies $P'^* f = f$.

Proof: By Lemma 6.5 and Theorem 4.6 all we need is to show that (ii) holds for each $P_p (p \in \mathscr{P})$; using Lemma 6.4 we verify (i) instead. To do this we must compute the null space $\mathscr{N}_{P_{p'}}$, which of course consists of all vectors orthogonal to $\mathscr{R}_{P_{p'}}$. As we have noted elsewhere, $\mathscr{R}_P = \mathscr{R}_P \cap \mathscr{C} - \mathscr{R}_P \cap \mathscr{C}$ for any $P \in \mathscr{L}$, and so, using the definition of the complement we see that $\mathscr{R}_{P_{p'}} = \mathscr{N}_{P_p} \cap \mathscr{C} - \mathscr{N}_{P_p} \cap \mathscr{C}$. Now $\mathscr{N}_{P_p} \cap \mathscr{C}$ consists of all am with $a \geqslant 0$ and $m \in \mathscr{M}$, m orthogonal to p, and so $\mathscr{R}_{P_{p'}} = \{am - bn | a, b \geqslant 0$ and $m, n \in \mathscr{M}$ orthogonal to $p\}$. Thus $\mathscr{N}_{P_{p'}}$ consists of all vectors u which are orthogonal to all $am - bn$ as above. But obviously this boils down to saying that u is orthogonal to all $q \in p^\perp$, i.e., that $u \in \mathscr{K}_p$. Thus we want to show that for any $f \in \mathscr{C}^*$ for which $f(p) = 0$ (which is what $f | \mathscr{N}_{P_{p'}} \cap \mathscr{C} = 0$ means) we also have $f | \mathscr{K}_p = 0$. Axiom (H10) gives us this.

VII. THE STRUCTURE OF $\mathscr{L}$

From this point on all operators are in $\mathscr{L}$.

Lemma 7.1: For $f \in \mathscr{C}^*$, the following are equivalent: (i) $f \leqslant P^* t$; and (ii) $P^* f = f$ and $\|f\|_1 \leqslant 1$.

Proof: Assume (i), i.e., that $f(m) \leqslant \|Pm\|_1$ for all $m \in \mathscr{M}$. Then $f(P'm) = 0$ and so $P'^* f = 0$ which by Theorem 6.6 gives $P^* f = f$; since $\|Pm\|_1 \leqslant 1$, $\|f\|_1 \leqslant 1$ follows. Now assume (ii); then $f(m) = f(Pm) \leqslant \|f\|_1 \|Pm\|_1 \leqslant \|Pm\|_1$ $= t(Pm) = (P^* t)(m)$ for all $m \in \mathscr{M}$.

Lemma 7.2: Let $f \in \mathscr{C}^*$; if $f \leqslant P^* t$ and $f \leqslant Q^* t$, then $f \leqslant (P \wedge Q)^* t$.

Proof: Using Lemma 7.1 we obtain $f(m) = f(Pm) = f(Qm)$ for all $m \in \mathscr{M}$, and thus $f(p) = f((PQ)^k p)$ for all $p \in \mathscr{P}$ and $k > 0$. Since $(PQ)^k p$

$\in \mathscr{C} \cup \{0\}$ and converges to $(P \wedge Q)p$ in the $\| \ \|_2$ we use Lemma 3.2 once more to obtain convergence in the norm $\| \ \|_1$ and thus $f(p) = f((P \wedge Q)p)$. But this means that $f = (P \wedge Q)^* f$, and since $\|f\|_1 \leqslant 1$ we have $f \leqslant (P \wedge Q)^* t$.

Lemma 7.3: If $P^*Q^*t \leqslant Q^*t$, then P and Q commute.

Proof: Since $Q^*t \leqslant t$ we also have $P^*Q^*t \leqslant P^*t$ and Lemma 7.2 gives us $P^*Qt \leqslant (P \wedge Q)^*t$. On the other hand, if $f \in \mathscr{C}^*$ and $f \leqslant t$, we have $P^*Q^*f \leqslant P^*Q^*t$ and so $P^*Q^*f \leqslant (P \wedge Q)^*t$. Lemma 7.1 now implies that $P^*Q^*f = (P \wedge Q)^*P^*Q^*f = (P \wedge Q)^*f$ since $(P \wedge Q)P = (P \wedge Q)Q = P \wedge Q$. Since t is an order unit for $\mathscr{V}^*$ this last relation holds for any $f \in \mathscr{C}^*$ and so for any $f \in \mathscr{V}^*$; but then we have $P^*Q^* = (P \wedge Q)^*$, or $QP = P \wedge Q$, so QP is a symmetric operator and therefore P and Q commute.

Proposition 7.4: If $P \leqslant Q$, then P' and Q commute.

Proof: All we need to do is show $P'^*Q^*t \leqslant Q^*t$ and then apply Lemma 7.3. Since we have $PQ = QP = P$ we have at once $P^*Q^*t = P^*t \leqslant Q^*t$, which implies that $Q^*t - P^*Q^*t \in \mathscr{C}^*$; but this functional also vanishes under the action of P^*, and so by Theorem 6.6 we obtain $Q^*t - P^*Q^*t = P'^*(Q^*t - P^*Q^*t)$, i.e., $P'^*Q^*t + P^*Q^*t = Q^*t$.
Since $P^*Q^*t \in \mathscr{C}^*$, we have $P'^*Q^*t \leqslant Q^*t$ as desired.

Proposition 7.5: If $PQ = 0$, then $P^*t + Q^*t = (P \vee Q)^*t$.

Proof: First the special case where $Q = P'$: clearly $t - P^*t \in \mathscr{C}^*$ and vanishes under the action of P^*, which implies that $t - P^*t = P'^*(t - P^*T) = P'^*t$, or $t = P^*t + P'^*t$. Now in general we note that the four operators P, Q, P', Q' commute pairwise by Proposition 7.4. In particular $P' \wedge Q' = P'Q'$ and also P' commutes with $P \vee Q$ since $P \leqslant P \vee Q$. Now act on both sides of $P^*t + P'^*t = t$ by $(P \vee Q)^*$ to obtain $(P \vee Q)^*t = P^*t + R^*t$, where $R = P'(P \vee Q) = P' \wedge (P \vee Q)$. We shall show that $R = Q$; since both $P \vee Q$ and P' are $\geqslant Q$ (because $PQ = 0$) we have $R \geqslant Q$ and all we need is $R \leqslant Q$. So let $Rm = m$; then $P'm = m$ and $(P \vee Q)m = m$, whence $(P \vee Q)'m = 0$, or $Q'P'm = 0$, which implies $Q'm = 0$, i.e., $Qm = m$.

Theorem 7.6: The partially ordered complemented set $\mathscr{L}$ is a complete atomic ortholattice.

Proof: We have already seen that $\mathscr{L}$ is a complete lattice; clearly, the projections $P_p (p \in \mathscr{P})$ are the atoms of $\mathscr{L}$, and the zero and the identity projections are the extreme elements. It is also trivial that $0' = I$ and $I' = 0$, while for any $P \in \mathscr{L}$ we have $P \wedge P' = 0$ and $P \vee P' = I$ (because $PP' = 0$). So all that is left is the orthomodular law. Let $P \leqslant Q$, and observe that since $P(P' \wedge Q) = 0$ we have by Proposition 6.5 that $P^*t + (P' \wedge Q)^*t = [P \vee (P' \wedge Q)]^*t$. Since we want to show that $P \vee (P' \wedge Q) = Q$, we must show that $P^*t + (P' \wedge Q)^*t = Q^*t$. By hypothesis we have $P \leqslant Q$, or $PQ = Q$, and also that P' and Q commute, so that $P' \wedge Q = P'Q = QP'$; thus $P^*t + (P' \wedge Q)^*t = Q^*P^*t + Q^*P'^*t = Q^*(P^*t + P'^*t) = Q^*t$.

VIII. THE STATES OF THE LOGIC $\mathscr{L}$

For each $m \in \mathscr{M}$ write $\hat{m}$ for the map $P \to \|Pm\|_1 = t(Pm) = (P^*t)(m)$.

Theorem 8.1: For each $m \in \mathscr{M}$ the map $\hat{m}$ is a completely additive state of $\mathscr{L}$.

Proof: Since $m = \Sigma_i a_i p_i$ implies $\hat{m} = \Sigma_i a_i \hat{p}_i$ it suffices to verify the above statement for $m \in \mathscr{P}$. So take $p \in \mathscr{P}$ and note that $0 \leqslant \hat{p}(P) \leqslant 1$ for all P and $p(I) = 1$. By Proposition 7.5, $\hat{p}$ is finitely additive, since $\hat{p}(P \vee Q) = \hat{p}(P) + \hat{p}(Q)$ for disjoint P, Q. So it suffices to show that if (P_j) is a decreasing net with infimum 0, then $\inf_j \hat{p}(P_j) = 0$ also. But $\inf_j(P_j) = 0$ means that $\lim_j \|P_j p\|_2 = 0$ hence again by Lemma 3.2 $\lim_j \|P_j p\|_1 = 0$, i.e., $\lim_j \hat{p}(P_j) = 0$.

The state functionals can be evaluated by using only the inner product. To this end we first establish a refinement of Theorem 4.6.

Theorem 8.2: Let (p_i) be a family of pairwise orthogonal vectors in $\mathscr{R}_P \cap \mathscr{P}$ maximal with respect to inclusion. Then $P = \sup(P_{p_i})$.

Proof: Let $Q = \sup(P_{p_i}) \leqslant P$. If $Q \neq P$, then $R = P \wedge Q$ by orthomodularity, hence there is some $r \in \mathscr{P}$ such that $Rr = r$. Then $Pr = r$ and $Q'r = r$, so $r \in \mathscr{R}_P \cap \mathscr{P}$ and $Qr = 0$, which implies that $P_{p_i} r = 0$ or $\langle r | p_i \rangle = 0$. But this contradicts maximality of the family (p_i).

Now we can compute $\hat{m}(P)$: take any maximal orthogonal (p_i) in $\mathscr{R}_P \cap \mathscr{P}$ and note that $\hat{m}(P) = \Sigma_i \hat{m}(P_{p_i}) = \Sigma_i \|P_{p_i} m\|_1 = \Sigma_i \langle m | p_i \rangle \|p_i\|_1 = \Sigma_i \langle m | p_i \rangle$.

As can be seen by means of examples, not all states have necessarily the form $\hat{m}$ with $m \in \mathscr{M}$. A specific case is presented in the Appendix. We shall now characterize the spaces for which all states have the form $\hat{m}$.

A *frame function* is a map φ: $\mathscr{P} \to [0,1]$ such that for any maximal orthogonal set $\{p_i\}$ in $\mathscr{P}$ we have $\Sigma_i \varphi(p_i) = 1$.

Theorem 8.2 shows that for each $m \in \mathscr{M}$ the map $\varphi_m : p \to \langle m | p \rangle$ is a frame function. We shall say that $\mathscr{M}$ is a *Gleason* set if the converse holds, i.e., if every frame function of $\mathscr{P}$ is of the form φ_m for some $m \in \mathscr{M}$.

Theorem 8.3: Suppose that every orthogonal set in $\mathscr{P}$ is at most countable. Then $\{\hat{m} | m \in \mathscr{M}\}$ exhausts the set of all states of $\mathscr{L}$ if and only if $\mathscr{M}$ is a Gleason set.

Proof: Suppose $\mathscr{M}$ is a Gleason set and let s be a state of $\mathscr{L}$. Again, Theorem 8.2 shows that the map $p \to s(P_p)$ is a frame function and so $s(P_p) = \langle m | p \rangle$ for some $m \in \mathscr{M}$, hence $s = \hat{m}$. For the converse, let φ be a frame function of $\mathscr{P}$. We note that, given $P \in \mathscr{L}$, the number $\Sigma_i \varphi(p_i)$, where $\{p_i\}$ is a maximal orthogonal set in $\mathscr{R}_P \cap \mathscr{P}$, is independent of $\{p_i\}$ because if we select and fix a maximal orthogonal set $\{q_j\}$ in $\mathscr{R}_{P'} \cap \mathscr{P}$ the set $\{p_i, q_j\}$ is maximal orthogonal in $\mathscr{P}$ and so $\Sigma_i \varphi(p_i) + \Sigma_j \varphi(q_j) = 1$, or $\Sigma_i \varphi(p_i) = 1 - \Sigma_j \varphi(q_j)$ which depends only on P. Write $\hat{\varphi}(P)$ for this number and note that $\hat{\varphi}$ is a state of $\mathscr{L}$. By hypothesis all states have the form $\hat{m}$ and so $\varphi(p) = \hat{\varphi}(P_p) = \hat{m}(P) = \langle m | p \rangle$.

It is not hard to see that countability of the orthogonal sets in $\mathscr{P}$ holds if, for example, the space $\mathscr{V}$ is separable with respect to the $\| \ \|_1$.

Now suppose that indeed $\mathscr{M}$ is the set of all states of $\mathscr{L}$. We shall show that a conditioning exists, i.e., that the three Axioms (M1), (M2), and (M3) of Sec. II hold, where the transition probability functional is just our inner product. First observe that for any $m \in \mathscr{M}$, the support L_m exists since $\mathscr{L}$ is complete. If we observe also that $\mathscr{R}_{\inf(P_i)} = \cap \mathscr{R}_{P_i}$, we

have at once that Axiom M1 (ii) holds; the other two parts of Axiom M1 are obvious. Clearly for $q \in \mathscr{P}$ the support is P_q, and so $p(P_q)$ is indeed the inner product $\langle p|q\rangle$, and (M2) holds. Now for (M3). We define the state θ to be just the zero vector, and we let for $m \in \mathscr{M}$ and $P \in \mathscr{L}$ the state $m_{:P}$ to be $Pm/\|Pm\|_1$, if $Pm \neq 0$ and let $m_{:P} = 0$ if $Pm = 0$. Then (i), (ii), and (iii) trivially follow, while, if $n(P) = 1$ we have $Pn = n$ and so $m(P)\langle m_{:P}|n\rangle = \|Pm\|_1\langle Pm/\|Pm\|_1, n\rangle = \langle Pm|n\rangle = \langle m|Pn\rangle = \langle m|n\rangle$, i.e., (iv) holds. We also note that the conditioning is pure, and therefore, according to Ref. 4, it is the only conditioning possible on $\mathscr{L}$.

APPENDIX: EXAMPLES

First some general remarks.

In case dim $\mathscr{V}$ is finite and we select $\mathscr{M}$ to be compact we do not have to worry about (H2), (H7), and (H9). In such cases we identify $\mathscr{V}$ with $\mathbb{R}^n$ and take $\langle x|y\rangle$ to be $\Sigma_{i=1}^n x_i y_i$. Also note that (H4) will follow from (H5), since the inner product is continuous in the $\|\ \|_1$: $\langle m|n\rangle = \Sigma_{i,j} a_i b_j \langle p_i|q_j\rangle \leqslant (\Sigma_i a_i)(\Sigma_j b_j) = 1$, whence $|\langle a_1 m_1 - b_1 n_1 | a_2 m_2 - b_2 n_2\rangle| \leqslant (a_1 + b_1)(a_2 + b_2)$ and so $|\langle u|v\rangle| \leqslant \|u\|_1\|v\|_1$. So all we have to pay attention to are (H1), (H3), (H5), (H6), (H8), and (H10).

Example 1. We select in $\mathbb{R}^n$ the convex base $\mathscr{M}$ to be the intersection of the unit ball with the hyperplane $x_n = \sqrt{2}/2$, i.e., the set

$$\{(x_1, x_2, \ldots x_{n-1}, \sqrt{2}/2) | x_1^2 + x_2^2 + \cdots + x_{n-1}^2 \leqslant \tfrac{1}{2}\}.$$

The extreme points form the set

$$\mathscr{P} = \{(x_1, x_2, \ldots, x_{n-1}, \sqrt{2}/2) | x_1^2 + x_2^2 + \cdots + x_{n-1}^2 = \tfrac{1}{2}\}$$

and so (H5) holds. Since

$$\langle x|y\rangle = \sum_{i=1}^{n-1} x_i y_i + \tfrac{1}{2} \geqslant -\sqrt{x_1^2 + \cdots + x_{n-1}^2} \times \sqrt{y_1^2 + \cdots + y_{n-1}^2} + \tfrac{1}{2} \geqslant -\tfrac{1}{2} + \tfrac{1}{2} = 0$$

we get (H5) at once. By the Cauchy inequality we see that the only vector orthogonal to $p = (x_1, x_2, \ldots, x_{n-1}, \sqrt{2}/2)$ is $(-x_1, -x_2, \ldots, -x_{n-1}, \sqrt{2}/2)$ and so (H6) and (H8) hold. This same remark also shows that the subspace $\mathscr{K}_p$ consists of all $v = (v_1, v_2, \ldots, v_n)$ for which $\Sigma_{i=1}^{n-1} x_i v_i \pm v_n(\sqrt{2}/2) = 0$, i.e., those $v = (v_1, v_2, \ldots, v_{n-1}, 0)$ for which $\Sigma_{i=1}^{n-1} x_i v_i = 0$. Now the hyperplane through p supporting $\mathscr{C}$ is

$$\left\{(z_1, z_2, \ldots, z_n) | \sum_{i=1}^{n-1} x_i z_i = z_n \sqrt{2}/2\right\}$$

and it obviously contains $\mathscr{K}_p$; thus (H10) holds.

Example 2. In $\mathbb{R}^3$, we select $\mathscr{M}$ to be a regular polygon of n sides in the plane $z = \delta$, inscribed in a circle of radius r, centered on the z axis; clearly we want $r = \sqrt{1 - \delta^2}$ so that (H5) holds. As before, (H1) holds by construction. We rotate $\mathscr{M}$ so that a vertex, say p_0 is in the xz plane, in which case $\mathscr{P}$ consists of the points $p_k = (\sqrt{1 - \delta^2}\cos(2k/n), \sqrt{1 - \delta^2}\sin(2k/n), \delta)$ $(k = 0, 1, \ldots, n-1)$. Because of the rotational symmetry it suffices to consider $\langle p_0|p_k\rangle$ which is equal to $(1 - \delta^2)\cos(2k/n) + \delta^2$. To satisfy (H3) we must choose $\delta^2 \geqslant -\cos(2k\pi/n)/[1 - \cos(2k\pi/n)]$ for $k = 0, 1, \ldots, n$. The maximum of the right-hand side occurs for $k = N$ when $n = 2N$ and for $k = N$ and $N + 1$ when $n = 2N + 1$; this implies that we must select $\delta^2 \geqslant \frac{1}{2}$ and $\delta^2 \geqslant \cos(\pi/n)/[1 + \cos(\pi/n)]$, respectively. We claim that selecting equality will yield (H6), because it is equivalent to orthogonality of p_0 and p_N for $n = 2N$ and p_0 and $p_k (k = N, N + 1)$ for $n = 2N + 1$. Indeed, for n even there is only one vector orthogonal to p_0 and so symmetry gives (H6); for n odd we have $p_0^\perp = \{p_N, p_{N+1}\}$, $p_N^\perp = \{p_0, p_{2N}\}$, and $p_{N+1}^\perp = \{p_0, p_1\}$ which means that $p_0^{\perp\perp} = p_0$.

This leaves us with (H8) and (H10); as we shall see, in case n is even (H8) holds but (H10) fails, while in case n is odd it is the other way around. This serves to show independence of the axioms. So take the case $n = 2N$. Since $p_0^\perp = \{p_N\}$ the maximizing vector is, for any p_k, the vector p_N; also $p_k - \langle p_k|p_N\rangle p_N$ is orthogonal to p_N which implies (H8). On the other hand $\mathscr{K}_{p_0}$ is spanned by p_0 and p_N, and is therefore, the y axis; since there are infinitely many planes supporting $\mathscr{C}$ through p_0, (H10) fails. Finally take the case $n = 2N + 1$. We have $p_0^\perp = \{p_N, p_{N+1}\}$ and so the vector maximizing $\langle p_k|r\rangle$ is p_N if $0 < k < N$ and p_{N+1} if $N + 1 < k < 2N$; in the first case the desired orthogonality fails for p_{N+1} and in the second for p_N. However, (H10) holds, because now $\mathscr{K}_{p_0}$ is $\{0\}$.

Example 3—the classical case. Let $\mathscr{H}$ be a Hilbert space whose inner product we write as $(\ |\ )$ and norm as $\|\ \|$. All statements made about $\mathscr{H}$ and its operators which are not proved here can be found in Refs. 6 and 7.

Let $\mathscr{V}$ be the space of all self-adjoint trace class operators on $\mathscr{H}$; the inner product $\langle T|S\rangle = \mathrm{tr}(TS)$ is defined. We take $\mathscr{M}$ to be the convex set of all positive operators T with $\mathrm{tr}(T) = 1$, so $\mathscr{C}$ is just the set of all positive operators; $\mathscr{C}$ spans $\mathscr{V}$ and so (H1) holds.

Now we compute the base norm. Since $T = T^+ - T^-$ we have $\|T\|_1 \leqslant \mathrm{tr}(T^+) + \mathrm{tr}(T^-) = \mathrm{tr}(|T|)$. On the other hand, if $T = aA - bB$ with $a, b \geqslant 0$ and $A, B \in \mathscr{M}$ we have $T \leqslant aA$, $-T \geqslant bB$; using the basis of eigenvectors of T to compute the trace, we obtain $\mathrm{tr}(T^+) \leqslant a$ and $\mathrm{tr}(T^-) \leqslant b$, so $\mathrm{tr}(|T|) \leqslant a + b$, which implies $\mathrm{tr}(|T|) \leqslant \|T\|_1$, whence equality. Thus (H2) follows. It is also well known that $\mathscr{P}$ consists of all projections of rank 1; for any $\varphi \in \mathscr{H}$ with $\|\varphi\| = 1$ we shall write $\varphi \otimes \varphi$ for the projection $\psi \to (\psi|\varphi)\varphi$. It is clear that (H3) and (H5) are trivial to verify; (H4) follows at once from $\mathrm{tr}(T^2) \leqslant [\mathrm{tr}|T|]^2$. Because $\langle \varphi \otimes \varphi | \psi \otimes \psi\rangle = |(\varphi|\psi)|^2$, (H6) is immediate, while (H7) is part of the spectral theorem for trace class operators. For (H8) we note that if $p = \varphi \otimes \varphi$ then $p^\perp = \{\psi \otimes \psi | \psi$ orthogonal to $\varphi\}$; given $q = \omega \otimes \omega \in \mathscr{P}$ we see that $\langle q|r\rangle$ is maximized for $r = \psi_0 \otimes \psi_0 = r_0$, where $\psi_0 = \pi\omega/\|\pi\omega\|$ and π is the orthoprojection in $\mathscr{H}$ onto the orthocomplement of φ. The inner product of $q - \langle q|r_0\rangle r_0 = \omega \otimes \omega - (\pi\omega) \otimes (\pi\omega)$ and the vectors $\psi \otimes \psi \in p^\perp$ is $|(\omega|\psi)|^2 - |(\pi\omega|\psi)|^2 = |(\omega|\psi)|^2 - |(\omega|\pi\psi)|^2 = 0$ since $\pi\psi = \psi$. To verify (H9) we must compute the norms of $\varphi \otimes \varphi - \psi \otimes \psi$. We easily see that $\|\varphi \otimes \varphi - \psi \otimes \psi\|_2 = \sqrt{2[1 - |(\varphi|\psi)|^2]}$; for the other, we compute the eigenvalues of the operator $\varphi \otimes \varphi - \psi \otimes \psi$ which turn out to be $\pm\sqrt{1 - |(\varphi|\psi)|^2}$ and so the two distances differ by a factor of $\sqrt{2}$ and (H9) holds.

More work is required for (H10). We want to show that for any $f \in \mathscr{V}^*$ which is $\geqslant 0$ on $\mathscr{C}$ and such that $f(p) = 0$ we

also have $f(v) = 0$ for all $v \in \mathscr{K}_p$. It is now immediate that any $f \in \mathscr{V}^*$ has the form $T \to \mathrm{tr}(TA)$, where A is a bounded operator on $\mathscr{H}$; using T of the form $\varphi \otimes \varphi$ we see that f is $\geqslant 0$ on $\mathscr{C}$ if and only if A is a positive operator. Now fix $p = \varphi \otimes \varphi \in \mathscr{P}$ and observe that $\mathscr{K}_p$ consists of all A of trace class such that $\mathrm{tr}(A(\varphi \otimes \varphi)) = 0$ and $\mathrm{tr}(A(\psi \otimes \psi)) = 0$ for all ψ orthogonal to φ; this boils down to $(A\varphi|\varphi) = 0$ and $(A\psi|\psi) = 0$ for all ψ orthogonal to φ. Now suppose that the given functional is determined by the (positive) operator B, and write $B = C^2$ with C positive; since the functional vanishes at $p = \varphi \otimes \varphi$, we have $(B\varphi|\varphi) = 0$, hence $C\varphi = 0$, and so $C\psi$ is orthogonal to φ for every ψ orthogonal to φ. We want to show that the functional vanishes to each operator A as described above, i.e., to show $\mathrm{tr}(C^2A) = 0$, or $\mathrm{tr}(CAC) = 0$. Take a basis $\varphi, \psi_1, \psi_2, \ldots$ and compute $\mathrm{tr}(CAC) = (CAC\varphi|\varphi) + \Sigma_i(CAC\psi_i|\psi_i) = \Sigma_i(AC\psi_i|C\psi_i)$. Since $C\psi_i$ is orthogonal to φ, we have each term in this sum equal to zero, and so $\mathrm{tr}(BA) = 0$, and (H10) holds.

Example 4—a commutative case. We let $\mathscr{V}$ be the space l^1 of real summable sequences, $\mathscr{C}$ the cone of sequences with non-negative elements and $\mathscr{M}$ the convex set of all $x = (x_i)_{i=1}^{\infty} \in \mathscr{C}$ with $\Sigma_i x_i = 1$. Clearly the base norm is the l^1 norm and $\mathscr{P}$ is the set of all $x \in \mathscr{M}$ with all but one element zero. We define $\langle x|y\rangle$ to be $\Sigma_i x_i y_i$ and observe that (H1)–(H7) and (H9) are trivial to verify. For (H8) we need only consider $p = (1,0,0,\ldots)$, as all other cases are similar; but note that any $q \in \mathscr{P}$, $q \neq p$ is necessarily in $p^{\perp}$ and so (H8) holds. It will take a little longer to verify (H10). Obviously $\mathscr{V}^*$ consists of all $x \to \Sigma_i a_i x_i$ with $(a_i)_{i=1}^{\infty}$ bounded and $\mathscr{C}^*$ of all those with $a_i \geqslant 0$. Consider (without loss of generality) the point $p \in \mathscr{P}$, $p = (1,0,0,\ldots)$ and note that $p^{\perp}$ consists of all $q_k = (0,0,\ldots,0,1,0,\ldots)$ where 1 occurs at the k th place ($k > 1$). We must show that if $f \in \mathscr{C}^*$ and $f(p) = 0$, then $f(v) = 0$ for all v orthogonal to all q_k. Now $v = (v_1, v_2, \ldots)$ is orthogonal to q_k means $v_k = 0$, so we must show that $f(v) = 0$ for all $v = (v_1,0,\ldots)$, provided $f(p) = 0$. But $f(p) = 0$ means $a_1 = 0$, and so $f(v) = a_1 v_1 = 0$.

We shall now show that the system of example 1 admits of states that do not have the form $\hat{m}$. This can be done directly, or by showing that $\mathscr{M}$ is not a Gleason set. The second way is easier. Consider the set $\mathscr{A}$ of all antipodal pairs in $\mathscr{P}$, i.e., all pairs (p,q) with p orthogonal to q; recall that $\{q\}$ is just $p^{\perp}$. Now select from each pair in $\mathscr{A}$ a point p, define $\varphi(p)$ arbitrarily between 0 and 1 and define $\varphi(q)$ as $1 - \varphi(p)$. Since antipodal pairs form maximal orthogonal sets, φ is a frame function. However, every frame function of the form $\hat{m}$ is continuous, and we can obviously select our φ to be discontinuous.

On the other hand, in example 3, $\mathscr{M}$ is a Gleason set. This is, of course, the content of the famous Gleason theorem (Ref. 8) and the reason we called such convex sets Gleason sets.

It is also quite straightforward to see that $\mathscr{M}$ in example 4 is a Gleason set.

[1] J. C. T. Pool, Commun. Math. Phys. **9**, 118 (1968).
[2] J. C. T. Pool, Commun. Math. Phys. **9**, 212 (1968).
[3] P. C. Deliyannis, J. Math. Phys. **17**, 653 (1976).
[4] P. C. Deliyannis, J. Math. Phys. **19**, 2341 (1978).
[5] E. M. Alfsen and F. W. Shultz, Mem. Am. Math. Soc. **172** (1976).
[6] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U. P., Princeton, NJ, 1955).
[7] T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966).
[8] A. M. Gleason, J. Math. Mech. **6**, 885 (1957).

Anharmonic oscillator with general polynomial potential

G. S. Sharma and L. K. Sharma
Department of Applied Physics, Government Engineering College, Jabalpur-482011, India

(Received 18 May 1983; accepted for publication 2 March 1984)

This paper derives simple analytical formulas for the energy eigenvalues $E_n(\lambda)$ of one-dimensional anharmonic oscillators characterized by the potentials $\omega^2x^2 + \Sigma_{\alpha=2}^{m}\lambda_\alpha x^{2\alpha}$. For doubly anharmonic oscillators, over a wide range of n and λ, these energy values agree well with the numerical values calculated by earlier workers.

PACS numbers: 03.65.Db, 11.10.Jj

I. INTRODUCTION

The study of one-dimensional anharmonic oscillators (AHO's) has evoked much interest because of its varied application in field theory[1] and molecular physics.[2] A general survey of the various applications may be found in the literature.[3,4]

The perturbation expansion of $E_n(\lambda)$ in powers of λ is not convergent but asymptotic.[5,6] These investigations have been confined to the calculation of the lowest few eigenvalues for different λ. Variational techniques using a harmonic oscillator (HO) basis have yielded better results.[2] For higher eigenvalues, however, the variational calculations become quite cumbersome. The Hill determinant method has also been used[7] for obtaining eight eigenvalues of the AHO for different λ in the range $0 < \lambda \leqslant 100$. Once again for higher eigenvalues as well as for large values of λ the numerical errors in the technique become quite appreciable. A variety of other methods have also been formulated to obtain approximate analytic relations which yield the values of $E_n(\lambda)$ to high accuracy. Hioe and Montroll[3] and Hioe *et al.*[4,8] have obtained a number of such relations which are accurate in different regimes of values of the quantum number n and the anharmonicity constant λ. In one regime the energy eigenvalues differ slightly from the HO levels (the near harmonic regime), in the other they differ slightly from the pure quartic oscillator eigenvalues (the near quartic regime). In a transition region between two such regimes, this work offers no specific way to decide which relation is more reliable. However numerical computations made by these workers have yielded results for λ varying from very small to very large values, but only for low-lying levels. The most comprehensive work on this problem is by Banerjee *et al.*[9] and Banerjee.[10] They have obtained $E_n(\lambda)$ for x^4, x^6, and x^8 oscillators for both n and λ varying from very small to very high values. However, this work has the shortcoming that it provides no single analytical formula, for a general power of x, that yields $E_n(\lambda)$ accurately for all n and λ.

Some AHO's with more complicated potential have also been investigated recently.[11,12] Doubly anharmonic systems of the type $\alpha x^2 + \beta x^4 + \gamma x^6$ have been studied analytically.[13–18] In the study of this problem,[16,19] it is found that a suitable three-term difference equation involving contiguous terms exists which permits the analytic study of the system in a manner which was not possible for $\lambda_m\, x^{2m}$ oscillators. The eigenvalues of the rotating harmonic oscillator have been similarly investigated.

The main object of this paper is to approximate analytically the energy levels of oscillators having a Hamiltonian of the general form

$$H = p^2 + \omega^2x^2 + \sum_{\alpha=2}^{m}\lambda_\alpha x^{2\alpha}, \tag{1}$$

where $p = (-i\,d/dx)$ and $\lambda_\alpha > 0$ are the anharmonicity constants. This has been done in Sec. II. The essential feature of our approach is to approximate the AHO wave function, say ϕ_n, by the basic states $|n\rangle$ of a HO with a suitably chosen renormalized frequency $\omega_0(n, \lambda)$. With this $\omega_0(n, \lambda)$ the matrix H, when considered in the basic states $|n\rangle$ of the above HO gets diagonalized and E_n closely approximates $\langle n|H|\rangle$. The use of a scaled HO basis for obtaining the eigenvalues of an AHO has been recognized earlier.[2,9,20,21] After obtaining the general eigenenergy expression for the oscillator, we have, for the purpose of obtaining accurate results, modified it by introducing suitable correcting terms. Further to verify the accuracy of our main result, viz., the eigenenergy expression, we particularize it in Sec. III to the case of a doubly anharmonic oscillator with quartic and sextic anharmonicities. We then compare our results with the corresponding (accurate) results already available in the literature. Our results are found to be quite good.

In Sec. IV, our general expression has been particularized to the case of a generalized $\lambda_m\, x^{2m}$ AHO. It may be of interest to note that the procedure adopted by us has already been used by Mathews *et al.*[21,21] for deriving expressions of $E_n(\lambda)$ for the quartic, sextic, and octic oscillators. These authors have checked the accuracy of their relations by comparing them with the corresponding values available in the literature. Finally Sec. V is devoted to a discussion of our results.

II. DERIVATION OF GENERAL EIGENENERGY EXPRESSION

The Schrödinger equation for the quantum mechanical AHO with polynomial interaction defined by (1) can be written as

$$-\frac{d^2\psi}{dx^2} + \left(\omega^2x^2 + \sum_{\alpha=2}^{m}\lambda_\alpha x^{2\alpha} - E\right)\psi = 0 \tag{2}$$

(in units $\hbar = 2\,m = 1$).

Here $m \geqslant 2$, and $\lambda_\alpha > 0$ are the coupling parameters. We consider the matrix representation of H in the eigenstates $|n\rangle$ of the following harmonic oscillator with renormalized frequency ω_0:

 0022-2488/84/102947-06$02.50

$$H_0 = -\frac{d^2\psi}{dx^2} + \omega_0^2 x^2, \quad H_0|n\rangle = (2n+1)\omega_0|n\rangle. \tag{3}$$

The matrix elements of H in the basic states of H_0 are easily determined to be the following:

$$\langle n|H|n\rangle \equiv H_{nn} = \omega_0\left[\left(n+\frac{1}{2}\right)(1+\nu^2) + \sum_{\alpha=2}^{m} g_\alpha^2 \nu^{\alpha+1} \frac{2\alpha!}{2^{2\alpha}} \sum_{r=0}^{\alpha} \frac{1}{(\alpha-r)!} 2^\alpha \frac{{}^n c_r}{r!}\right], \tag{4}$$

and

$$\langle n+2|H|n\rangle = \langle n|H|n+2\rangle \equiv H_{n,n+2} = \omega_0\left(-\frac{1}{2}(1-\nu^2) + \sum_{\alpha=2}^{m} g_\alpha^2 \nu^{\alpha+1} \frac{2\alpha!}{2^{2\alpha+1}} \times \sum_{r=1}^{\alpha} \frac{1}{(\alpha-r)!} {}^n C_{r-1} \frac{2^{r+1}}{(r+1)!}\right) \times \{(n+2)(n+1)\}^{1/2}. \tag{5}$$

Here $\nu = \omega/\omega_0$ is a positive quantity, ${}^n c_r = n!/r!(n-r)!$, and g_α^2 is the coupling parameter defined by

$$g_\alpha^2 = \lambda_\alpha/\omega^{\alpha+1}. \tag{6}$$

Other matrix elements $\langle n|H|n'\rangle$ are obtained similarly; indeed, $\langle n|H|n'\rangle = 0$, when $|n'-n| > 2m$.

Now the renormalized frequency $\omega_0(n, \lambda)$ is suitably chosen to diagonalized the matrix H so that $\langle n|H|n\rangle$ closely approximates E_n. To accomplish this (for approximating E_n by H_{nn}) we choose ν such that the nearest off-diagonal terms (viz., $H_{n,n\pm2}$ and $H_{n\pm2,n}$) are numerically as small as possible. On substitution of ν so obtained, H_{nn} is expected to closely approximate E_n. That is,

TABLE I. Energy eigenvalues of $\omega^2x^2 + \lambda_2x^4 + \lambda_3x^6$ oscillator ($\omega^2 = \lambda_2 = \lambda_3 = 1$). Here n is the excitation number of the level, E_C is the result obtained by our calculations, E_E is the exact eigenvalue, δ_C is the relative error (defined) as $\delta_C = (E_E - E_C)/E_C$ in our calculated results, δ_E is the corresponding error obtained by Datta *et al.*[22]

n	E_C	E_E[a]	δ_C	δ_E
1	5.756	5.656	-0.177×10^{-1}	0.336×10^{-2}
2	11.144*	11.107	-0.333×10^{-2}	0.522×10^{-2}
3	17.607	17.637	0.170×10^{-2}	0.578×10^{-2}
4	25.014	25.068	0.215×10^{-2}	0.538×10^{-2}
5	33.227	33.293	0.198×10^{-2}	0.486×10^{-2}
6	42.163	42.236	0.173×10^{-2}	0.438×10^{-2}
7	51.763	51.841	0.150×10^{-2}	0.397×10^{-2}
8	61.981	62.062	0.131×10^{-2}	0.361×10^{-2}
9	72.778	72.861	0.114×10^{-2}	0.332×10^{-2}
10	84.124*	84.209	0.101×10^{-2}	0.305×10^{-2}
20	223.206	223.295	0.399×10^{-3}	0.169×10^{-2}
30	400.285	400.371	0.215×10^{-3}	0.116×10^{-2}
40	608.270	608.349	0.130×10^{-3}	0.882×10^{-3}
50	843.009	843.079	0.830×10^{-4}	0.710×10^{-3}
60	1101.681	1101.739	0.526×10^{-4}	0.595×10^{-3}
70	1382.208	1382.253	0.326×10^{-4}	0.510×10^{-3}
80	1682.973	1683.005	0.190×10^{-4}	0.447×10^{-3}
90	2002.676	2002.692	0.799×10^{-5}	0.398×10^{-3}
100	2340.237*	2340.237	0	0.368×10^{-3}

[a] See Ref. 22.

TABLE II. Same as Table I but with $\omega^2 = \lambda_2 = 1, \lambda_3 = 5$.

n	E_C	E_E[a]	δ_C	δ_E
1	7.466	7.279	-0.257×10^{-1}	-0.331×10^{-1}
2	14.806	14.731	-0.509×10^{-2}	-0.104×10^{-1}
3	23.799	23.837	0.159×10^{-2}	-0.513×10^{-2}
4	34.233	34.303	0.204×10^{-2}	-0.295×10^{-2}
5	45.884	45.965	0.176×10^{-2}	-0.186×10^{-2}
6	58.623	68.709	0.146×10^{-2}	-0.124×10^{-2}
7	72.358	72.447	0.123×10^{-2}	-0.870×10^{-3}
8	87.021	87.111	0.103×10^{-2}	-0.628×10^{-3}
9	102.553	102.644	0.887×10^{-3}	-0.462×10^{-3}
10	118.909	118.999	0.756×10^{-3}	-0.346×10^{-3}
20	320.873	320.960	0.271×10^{-3}	-0.478×10^{-5}
30	579.751	579.833	0.141×10^{-3}	0.357×10^{-4}
40	884.895	884.971	0.859×10^{-4}	0.415×10^{-4}
50	1230.077	1230.147	0.569×10^{-4}	0.401×10^{-4}
60	1611.072	1611.134	0.385×10^{-4}	0.373×10^{-4}
70	2024.764	2024.818	0.267×10^{-4}	0.343×10^{-4}
80	2468.735	2468.780	0.182×10^{-4}	0.315×10^{-4}
90	2941.035	2941.071	0.122×10^{-4}	0.290×10^{-4}
100	3440.049	3440.074	0.727×10^{-5}	0.268×10^{-4}

[a] See Ref. 22.

$$E_n \simeq \frac{1}{\nu}\left[\left(n+\frac{1}{2}\right)(1+\nu^2) + \sum_{\alpha=2}^{m} g_\alpha^2 \nu^{\alpha+1} \frac{2\alpha!}{2^{2\alpha}} \times \sum_{r=0}^{\alpha} \frac{1}{(\alpha-r)!} 2^r \frac{{}^n c_r}{r!}\right]\omega, \tag{7}$$

where ν is the value between 0 and 1 which satisfies the following algebraic equation:

$$\sum_{\alpha=2}^{m} g_\alpha^2 \nu^{\alpha+1} \frac{2\alpha!}{2^{2\alpha}} \sum_{r=1}^{\alpha} \frac{1}{(\alpha-r)!} ({}^n C_{r-1} + {}^{n-2} C_{r-1}) \times \frac{2^r}{(r+1)!} + \nu^2 = 1. \tag{8}$$

Though the terms other than the nearest off-diagonal ones left uncompensated so far also contribute to E_n but if n is not very small, the contribution from the terms $H_{mm'}$ with $m + m' < 2n$ is largely counterbalanced by the contribution

TABLE III. Same as Table I but with $\omega^2 = \lambda_2 = 1, \lambda_3 = 10$.

n	E_C	E_E[a]	δ_C	δ_E
1	8.537	8.346	-0.229×10^{-1}	-0.368×10^{-1}
2	17.139	17.046	-0.546×10^{-2}	-0.119×10^{-1}
3	27.682	27.726	0.159×10^{-2}	-0.614×10^{-2}
4	39.948	40.027	0.197×10^{-2}	-0.368×10^{-2}
5	53.663	53.754	0.169×10^{-2}	-0.243×10^{-2}
6	68.674	68.769	0.138×10^{-2}	-0.171×10^{-2}
7	84.871	84.969	0.115×10^{-2}	-0.126×10^{-2}
8	102.173	102.271	0.958×10^{-3}	-0.965×10^{-3}
9	120.510	120.608	0.813×10^{-3}	-0.759×10^{-3}
10	139.827	139.924	0.693×10^{-3}	-0.609×10^{-3}
20	378.705	378.794	0.235×10^{-3}	-0.129×10^{-3}
30	685.298	685.381	0.121×10^{-3}	-0.446×10^{-4}
40	1046.930	1047.005	0.716×10^{-4}	-0.176×10^{-4}
50	1456.187	1456.256	0.476×10^{-4}	-0.649×10^{-5}
60	1908.040	1908.102	0.325×10^{-4}	-0.116×10^{-5}
70	2398.784	2398.838	0.225×10^{-4}	0.160×10^{-5}
80	2925.547	2925.586	0.133×10^{-4}	0.310×10^{-5}
90	3485.991	3486.027	0.103×10^{-4}	0.392×10^{-5}
100	4078.212	4078.239	0.662×10^{-5}	0.437×10^{-5}

[a] See Ref. 22.

TABLE IV. Same as Table I but with $\omega^2 = \lambda_2 = 1, \lambda_3 = 20$.

n	E_C	E_E[a]	δ_C	δ_E
1	9.960	9.679	-0.290×10^{-1}	-0.387×10^{-1}
2	20.018	19.904	0.573×10^{-2}	-0.126×10^{-1}
3	32.443	32.493	0.154×10^{-2}	-0.657×10^{-2}
4	46.923	47.014	0.194×10^{-2}	-0.398×10^{-2}
5	63.129	63.233	0.164×10^{-2}	-0.266×10^{-2}
6	80.878	80.987	0.135×10^{-2}	-0.189×10^{-2}
7	100.041	100.150	0.109×10^{-2}	-0.142×10^{-2}
8	120.517	120.626	0.904×10^{-3}	-0.109×10^{-2}
9	142.227	142.334	0.752×10^{-3}	-0.874×10^{-3}
10	165.101	165.207	0.642×10^{-3}	-0.712×10^{-3}
20	448.233	448.326	0.207×10^{-3}	-0.176×10^{-4}
30	811.921	812.006	0.104×10^{-3}	-0.745×10^{-4}
40	1241.077	1241.153	0.612×10^{-4}	-0.395×10^{-4}
50	1726.879	1726.948	0.340×10^{-4}	-0.236×10^{-4}
60	2263.345	2263.406	0.270×10^{-4}	-0.153×10^{-4}
70	2846.066	2846.118	0.183×10^{-4}	-0.104×10^{-4}
80	3471.618	3471.660	0.121×10^{-4}	-0.728×10^{-5}
90	4137.240	4137.273	0.798×10^{-4}	-0.528×10^{-5}
100	4840.653	4840.673	0.413×10^{-5}	-0.381×10^{-5}

[a] See Ref. 22.

from the terms with $m + m' > 2n$. However for accurate values of $E_n(\lambda)$ one should take into account this contribution as well. Now as the exact value of E_n is independent of ω_0 (and hence of ν) we utilize this extra information to obtain a more accurate expression for E_n by modifying the equation which defines ν. The modified form can be written as

$$g_m^2 \nu^{m+1}\left[\frac{2m!}{2^{2m}}\sum_{r=1}^{m}\frac{1}{(m-r)!}({}^nC_{r-1} + {}^{n-2}C_{r-1}) \times\frac{2^r}{(r+1)!} - A_m\right] + \sum_{\alpha=2}^{m-1} g_\alpha^2 \nu^{\alpha+1}\frac{2\alpha!}{2^{2\alpha}}\sum_{r=1}^{\alpha}\frac{1}{(\alpha-r)!} \times({}^nC_{r-1} + {}^{n-2}C_{r-1})\frac{2^r}{(r+1)!}(1-A_\alpha) + A_1\nu^2 = A_0 + B/(n+\tfrac{1}{2})^2. \quad (9)$$

Here $A_\alpha(\alpha = 0, 1,\ldots,m)$ and B are arbitrarily chosen constants. To obtain these constants, we choose some particular coefficients λ_α and level number n, then find those values of ν (between 0 and 1) which make H_{nn} closely approximate the corresponding value of E_n (assumed known from some other computation). With these values of ν and corresponding n and λ_α we obtain linear algebraic equations equal in number to the number of constants involved in (9). The values of the constants can easily be obtained by solving these equations.

Expression (7) with relevant ν obtained from Eq. (9) constitutes our expression for E_n. This simple generalized expression is applicable for a wide range of n and λ_α. In the following section we shall examine the accuracy of our eigenenergy expression.

TABLE V. Same as Table I but with $\omega^2 = \lambda_2 = 1, \lambda_3 = 50$.

a	E_C	E_E[a]	δ_C	δ_E
1	12.298	11.913	-0.323×10^{-1}	-0.399×10^{-1}
2	24.798	24.651	-0.596×10^{-2}	-0.129×10^{-1}
3	40.312	40.374	0.154×10^{-2}	-0.679×10^{-2}
4	58.418	58.528	0.188×10^{-2}	-0.413×10^{-2}
5	78.696	78.821	0.159×10^{-2}	-0.277×10^{-2}
6	100.917	101.046	0.128×10^{-2}	-0.198×10^{-2}
7	124.918	125.046	0.102×10^{-2}	-0.149×10^{-2}
8	150.570	150.697	0.843×10^{-3}	-0.116×10^{-2}
9	177.774	177.899	0.703×10^{-3}	-0.928×10^{-3}
10	206.444	206.567	0.595×10^{-3}	-0.759×10^{-3}
20	561.570	561.672	0.182×10^{-3}	-0.196×10^{-3}
30	1018.024	1018.112	0.864×10^{-4}	-0.875×10^{-4}
40	1556.816	1556.893	0.495×10^{-4}	-0.489×10^{-4}
50	2166.849	2166.916	0.309×10^{-4}	-0.310×10^{-4}
60	2840.595	2840.653	0.204×10^{-4}	-0.213×10^{-4}
70	3572.511	3572.558	0.132×10^{-4}	-0.154×10^{-4}
80	4358.288	4358.324	0.826×10^{-5}	-0.116×10^{-4}
90	5194.457	5194.481	0.462×10^{-5}	-0.911×10^{-5}
100	6078.141	6078.155	0.230×10^{-5}	-0.727×10^{-5}

[a] See Ref. 22.

TABLE VI. Same as Table I but with $\omega^2 = \lambda_2 = 1, \lambda_3 = 100$.

n	E_C	E_E[a]	δ_C	δ_E
1	14.488	14.023	-0.332×10^{-1}	-0.403×10^{-1}
2	29.285*	29.109	-0.605×10^{-2}	-0.130×10^{-1}
3	47.677	47.749	0.151×10^{-2}	-0.686×10^{-2}
4	69.155	69.283	0.185×10^{-2}	-0.418×10^{-2}
5	93.219	93.364	0.155×10^{-2}	-0.280×10^{-2}
6	119.594	119.743	0.124×10^{-2}	-0.201×10^{-2}
7	148.086	148.234	0.998×10^{-3}	-0.151×10^{-2}
8	178.544	178.689	0.811×10^{-3}	-0.117×10^{-2}
9	210.847	210.989	0.673×10^{-3}	-0.942×10^{-3}
10	244.895	245.033	0.563×10^{-3}	-0.771×10^{-3}
20	666.759	666.871	0.168×10^{-3}	-0.201×10^{-3}
30	1209.143	1209.237	0.777×10^{-4}	-0.906×10^{-4}
40	1849.455	1849.536	0.438×10^{-4}	-0.511×10^{-4}
50	2574.496	2574.563	0.260×10^{-4}	-0.327×10^{-4}
60	3375.310	3375.364	0.160×10^{-4}	-0.227×10^{-4}
70	4245.302	4245.364	0.989×10^{-5}	-0.166×10^{-4}
80	5179.351	5179.378	0.521×10^{-5}	-0.126×10^{-4}
90	6173.326	6173.341	0.243×10^{-5}	-0.999×10^{-5}
100	7223×813*	7223.813	0	-0.806×10^{-5}

[a] See Ref. 22.

III. NUMERICAL ANALYSIS—APPLICATION TO DOUBLY AHO

To check the accuracy of our expression we consider the special case of a doubly AHO. Choosing $m = 3$, Eq. (2) reduces to

$$H\psi = E\psi, \quad \text{with} \quad H = -\frac{d^2\psi}{dx^2} + \omega^2x^2 + \lambda_2x^4 + \lambda_3x^6. \quad (10)$$

Equation (10) is the Schrödinger equation for a doubly anharmonic oscillator. Here the values chosen for λ_α are given in the table captions. For this AHO, (7) transforms to the following form:

TABLE VII. Same as Table I but with $\omega^2 = 1, \lambda_2 = -1, \lambda_3 = 100$.

n	E_C	E_E[a]	δ_C	δ_E
2	16.368	16.221	-0.906×10^{-2}	-0.13×10^{-1}
4	38.440	38.533	0.241×10^{-2}	-0.42×10^{-2}
6	66.491	66.604	0.170×10^{-2}	-0.20×10^{-2}
8	99.318	99.435	0.118×10^{-2}	-0.11×10^{-2}
10	136.299	136.417	0.865×10^{-3}	-0.73×10^{-3}

[a] See Ref. 22.

TABLE VIII. Same as Table I but with $\omega^2 = 1, \lambda_2 = 0, \lambda_3 = 1$.

n	E_C	E_E[a]	δ_C	δ_E
2	10.032	9.966	-0.662×10^{-2}	-0.13×10^{-1}
4	22.843	22.910	0.292×10^{-2}	-0.41×10^{-2}
6	38.962	39.059	0.248×10^{-2}	-0.20×10^{-2}
8	57.745	57.845	0.173×10^{-2}	-0.12×10^{-2}
10	78.851	78.958	0.136×10^{-2}	-0.76×10^{-3}

[a] See Ref. 22.

$$E_n = (1/\nu)\{(n+\tfrac{1}{2})(1+\nu^2) + \tfrac{5}{2}g_3^2\nu^4(n+\tfrac{1}{2})[(n+\tfrac{1}{2})^2+\tfrac{5}{4}] + \tfrac{3}{2}g_2^2\nu^3[(n+\tfrac{1}{2})^2+\tfrac{1}{4}]\}\omega. \quad (11)$$

For this case the constants A and B of Eq. (9) are obtained according to the procedure mentioned earlier. Values of E_n marked by an asterisk in Tables I and VI are used in obtaining these constants. Equation (9) now reduces to

$$g_3^2\nu^4\{\tfrac{15}{4}[(n+\tfrac{1}{2})^2+\tfrac{7}{4}] - A_3\} + 2g_2^2\nu^3(n+\tfrac{1}{2})(1-A_2) + A_1\nu^2 = A_0 + B/(n+\tfrac{1}{2})^2. \quad (12)$$

Here

$$g_\alpha^2 = \lambda_\alpha/\omega^{\alpha+1}, \quad A_0 = 0.286\,851\,297\,3, \quad A_1 = 0.086\,163\,746, \quad A_2 = 0.425\,555\,777, \quad A_3 = 13.045\,265\,93, \quad B = 0.062\,535. \quad (13)$$

In the evaluation of these constants, the chosen values of E_n either equal or closely approximate the accurate values calculated from the truncated Hill determinants generated by scaled basis functions.[22] For these known values of E_n, n, and λ, we obtain the corresponding values of ν (lying between zero and one) which satisfy Eq. (11). Corresponding to these values of E_n, n, and λ_α, Eq. (12) gives five linear algebraic equations. Solving these equations yields the constants mentioned in Eq. (13).

Equation (11), with ν determined from Eqs. (12) and (13), expresses our results for the doubly AHO. These results, valid for a wide range of n and λ_α, are displayed in Tables I–VIII. The accuracy of our results is examined by comparing them with the known numerical results in the literature.[22] For $n<2$, our eigenenergy expression provides less satisfactory results than for $n>2$. This discrepancy could be explained by remembering that for $n<2$ there exists no nonvanishing $H_{mm'}$ with $m+m'<2n$ to counterbalance the effect of $H_{mm'}$ with $m+m'>2n$. For $n=1$, however, we must add a correction term which represents the contribution from the nearest off-diagonal terms. Then the eigenenergy for $n=1$ becomes the lower eigenvalue of the matrix

$$\begin{pmatrix} H_{11} & H_{13} \\ H_{31} & H_{33} \end{pmatrix}$$

and this is

$$E_1 = \frac{H_{11}+H_{33}}{2} - \frac{1}{2}\sqrt{[(H_{11}+H_{33})^2 - 4(H_{11}H_{33}-H_{13}H_{31})]}. \quad (14)$$

In the above equation the values of H_{11}, H_{33}, H_{13}, and H_{31} with the relevant ν substituted from Eq. (7) are obtained from Eqs. (4) and (5). Equation (14) has been used in obtaining the results for $n<2$.

Finally we compute the relative error in our results and compare it with the corresponding error obtained by Datta *et al.*[22] An overall improvement in the results is observed.

IV. AHO WITH SINGLE GENERALIZED ANHARMONIC TERM

In this section we examine the accuracy of the eigenenergy expression for an AHO with a single generalized anharmonic term. For $\alpha = m$, Eq. (7) reduces to that for a $\lambda_m x^{2m}$ oscillator, viz.,

$$E_n = \frac{1}{\nu}\left[\left(n+\frac{1}{2}\right)(1+\nu^2) + g_m^2\nu^{m+1}\frac{2m!}{2^{2m}} \times \sum_{r=0}^{m}\frac{1}{(m-r)!}2^r\frac{{}^nc_r}{r!}\right], \quad (15)$$

with relevant ν substituted from

$$g_m^2\nu^{m+1}\left[\frac{2m!}{2^{2m}}\sum_{r=1}^{m}\frac{1}{(m-r)!}({}^nc_{r-1} + {}^{n-2}c_{r-1}) \times\frac{2^r}{(r+1)!} - A_m\right] = A_0 - A_1\nu^2 + B/(n+\tfrac{1}{2})^2. \quad (16)$$

We now simplify the above expressions for large n and λ_m and compare it with the known WKB–large-n and λ_m results. The solution of Eq. (16) for large n and λ_m can be expanded as

$$\nu = (A_0/\rho_1 g_m^2)^{1/(m+1)}(n+\tfrac{1}{2})^{(1-m)/(1+m)} \times\left[1 - \frac{\rho_2}{(m+1)(n+\tfrac{1}{2})^2} + \frac{1}{(m+1)}\frac{B}{A_0(n+\tfrac{1}{2})^2} - \frac{1}{(m+1)^2}\frac{B\rho_2}{A_0(n+\tfrac{1}{2})^4} - \frac{1}{(m+1)}\frac{A_1}{A_0}\left(\frac{A_0}{\rho_1 g_m^2}\right)^{2/(m+1)} \times(n+\tfrac{1}{2})^{2(1-m)/(1+m)}\left(1 - \frac{3\rho_2}{(1+m)(n+\tfrac{1}{2})^2} + \frac{2\rho_2}{(1+m)^2(n+\tfrac{1}{2})^4}\right) + \cdots\right]. \quad (17)$$

Here

$$\rho_1 = \frac{2m!}{(m+1)!(m-1)!}\frac{1}{2^{m-1}}, \quad \rho_2 = \frac{(m-1)(m-2)}{2!2^2}\frac{2}{3}\left(m+\frac{15}{2}\right), \quad g_m^2 = \lambda_m/\omega^{m+1}. \quad (18)$$

If we substitute this value of ν in Eq. (15), and we set $\omega = 1$, then E_n becomes

$$E_n = \lambda_m^{1/(1+m)}\left[X\left(n+\frac{1}{2}\right)^{2m/(1+m)}\left(1+\frac{Y}{(n+\tfrac{1}{2})^2}\right) + Z\left(\frac{(n+\tfrac{1}{2})}{\lambda_m}\right)^{2/(1+m)} + \cdots\right], \quad (19)$$

where

$$X=\left(\frac{\rho_1}{A_0}\right)^{1/(m+1)}\left(1+\frac{(m+1)A_0}{2m}\right),$$

$$Y=B\left(\frac{-1}{(m+1)A_0}+\frac{(m+1)}{[2m+(m+1)A_0]}\right)+\frac{(m^2-1)(5-2m)A_0}{4[2m+(m+1)A_0]}+\frac{\rho_2}{(m+1)},$$

and

$$Z=\left(\frac{A_0}{\rho_1}\right)^{1/(m+1)}\left[1+A_1\left(\frac{1}{A_0(m+1)}-\frac{1}{2}\right)\right]. \tag{20}$$

The WKB approximation for the general anharmonic oscillator with Hamiltonian $H=p^2/2m+\frac{1}{2}m\omega^2x^2+\lambda_m x^{2m}$ yields the following expansion, for large n, of the nth eigenvalue[3,23]:

$$E_n^{\mathrm{WKB}}=2^{m/(1+m)}\lambda_m^{1/(m+1)}\times\left\{x_m\left[\left(n+\frac{1}{2}\right)+\frac{\delta_m}{(n+\frac{1}{2})}\right]^{2m/(1+m)}+2^{2/(1+m)}y_m\left(\frac{(n+\frac{1}{2})}{\lambda_m}\right)^{2/(1+m)}+\cdots\right\}, \tag{21}$$

where

$$x_m=2^{(m-2)/(m+1)}\left(\frac{\pi(m+1)\Gamma(1/m)}{\Gamma^2(1/2m)}\right)^{2m/(1+m)},$$

$$y_m=2^{(2-m)/(1+m)}\left(\frac{\pi(m+1)\Gamma(1/m)}{\Gamma^2(1/2m)}\right)^{2/(m+1)}\times\frac{\Gamma(1/m)\Gamma^2(3/2m)}{\Gamma(3/m)\Gamma^2(1/2m)},$$

$$\delta_m=\frac{2m-1}{12\pi(m+1)}\cot\left(\frac{\pi}{2m}\right), \tag{22}$$

$$\hbar=2m=\omega/2=1.$$

Equation (21) can be further simplified to

$$E_n^{\mathrm{WKB}}=(\lambda_m)^{1/(m+1)}\left[X_m\left(n+\frac{1}{2}\right)^{2m/(1+m)}\times\left(1+\frac{\delta_m 2m/(1+m)}{(n+\frac{1}{2})^2}\right)+Y_m\left(\frac{(n+\frac{1}{2})}{\lambda_m}\right)^{2/(1+m)}+\cdots\right], \tag{23}$$

where

$$X_m=2^{2[(m-1)/(m+1)]}\left(\frac{\pi(m+1)\Gamma(1/m)}{\Gamma^2(1/2m)}\right)^{2m/(1+m)},$$

and

$$Y_m=2^{4/(1+m)}\left(\frac{\pi(m+1)\Gamma(1/m)}{\Gamma^2(1/2m)}\right)^{2/(m+1)}\times\frac{\Gamma(1/m)\Gamma^2(3/2m)}{\Gamma(3/m)\Gamma^2(1/2m)}.$$

With the proper choice of A_0, A_1, and B, expression (19) obtained by us is in complete agreement to that obtained by Hioe and Montroll.[3] Further with a slightly different choice of above parameters [$A_m=0$, $A_0=a(1-\nu^2)$, $A_1=b(1-\nu^2)$, and $B=c(1-\nu^2)$] our eigenenergy expression becomes the generalization of the results obtained earlier by Mathews *et al.*[20,21] for the cases of $m=2$, 3, and 4.

V. CONCLUSION

To our knowledge the generalized eigenenergy expression obtained by us with such a wide validity has been derived for the first time. It may be noted that we do not impose any restriction (except that $\lambda_\alpha>0$) on the relative amplitude of the coupling parameters which describes the relative strength of the corresponding anharmonic terms. As a special case of our generalized formulation we study the doubly AHO with Hamiltonian defined by Eq. (9). Results obtained for this case are quite good. Also we have examined the accuracy of our results for $\lambda_\alpha<0$. If $\lambda_3=100$ then results are quite satisfactory for $\lambda_2=-1$ but not so good for large negative values of λ_2. A similar observation for low-lying levels was also made by Datta *et al.*[22] For $\lambda_2=0$ the Hamiltonian in Eq. (9) reduces to that for a λ_3x^6 oscillator and as shown in Table VIII, the accuracy remains comparable to that when $\lambda_2\neq0$.

It may also be mentioned that a fresh choice of the parameters appearing in Eq. (12) is required for calculating $E_n(\lambda)$ with values of n and λ much different from what have been considered in the present work.

In Sec. IV we have used large-n and -λ WKB approximations for obtaining the generalized expression for $\lambda_m x^{2m}$ anharmonicity. The resulting expression produces uniformly accurate results over a large parameter domain (including smaller values of n and λ).

Using five eigenvalues computed numerically, our expression (12) provides a simple way of obtaining uniformly good results. Numerical computation of $E_n(\lambda)$ requires repeated calculation of the roots.[22] For $n=100$ and $\lambda_3=100$ ($\lambda_1=\lambda_2=1$) these repeated calculations of roots, involving a 250th-order polynomial, must be done as much as 250 times. The order of polynomial and repeated calculation of roots go on increasing as the values of n and λ increase further. In our case however, the calculation of only a single root of a fourth-order polynomial is needed. The present method therefore accomplishes results with relatively less computational labor.

[1] C. M. Bender and T. T. Wu, Phys. Rev. Lett. **37**, 117 (1976).
[2] C. E. Reid, J. Mol. Spectrosc. **36**, 183 (1970).
[3] F. T. Hioe and E. W. Montroll, J. Math. Phys. **16**, 1945 (1975).
[4] F. T. Hioe, D. MacMillen, and E. W. Montroll, J. Math. Phys. **17**, 1320 (1976).
[5] C. M. Bender and T. T. Wu, Phys. Rev. **184**, 1231 (1969).
[6] B. Simon, Ann. Phys. **58**, 76 (1970).
[7] S. N. Biswas, K. Datta, R. P. Saxena, P. K. Srivastava, and V. S. Varma, J. Math. Phys. **14**, 1190 (1973).
[8] F. T. Hioe, D. MacMillen, and E. W. Montroll, Phys. Rep. C **43**, 305 (1978).
[9] K. Banerjee, S. P. Bhatnagar, V. Choudhary, and S. S. Kanwal, Proc. R. Soc. London Ser. A **360**, 575 (1978).
[10] K. Banerjee, Proc. R. Soc. London Ser. A **364**, 265 (1978).
[11] F. T. Hioe, J. Chem. Phys. **69**, 204 (1978).
[12] F. T. Hioe, Phys. Rev. D **15**, 488 (1977).
[13] V. Singh, S. N. Biswas, and K. Datta, Phys. Rev. D **18**, 1901 (1978).

[14]P. M. Mathews and M. Lakshmanan, Ann. Phys. **79**, 171 (1973).
[15]M. Lakshmanan, Lett. Nuovo Cimento **8**, 743 (1973).
[16]G. P. Flessas, Phys. Lett. A **72**, 289 (1979).
[17]G. E. Sobelman, Phys. Rev. D **19**, 3754 (1979).
[18]F. T. Hioe, J. Math. Phys. **19**, 1307 (1978).
[19]V. Singh, A. Rampal, S. N. Biswas, and K. Datta, Lett. Math. Phys. **4**, 131 (1980).
[20]P. M. Mathews, M. Seetharaman, S. Raghavan, and V. T. A. Bhargava, Phys. Lett. A **83**, 118 (1981).
[21]P. M. Mathews, M. Seetharaman, S. Raghavan, and V. T. A. Bhargava, Pramana **17**, 121 (1981).
[22]K. Datta and A. Rampal, Phys. Rev. D **23**, 2875 (1981).
[23]J. Pasupathy and V. Singh, University of Texas (Austin), Preprint DOE-ER-03992-423, 1980.

The shift operator technique for SO(7) in an $[SU(2)]^3$ basis. I. Theory

J. Van der Jeugt[a] and P. De Wilde[b]
Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit-Gent, Krijgslaan 281-S9, B-9000 Gent, Belgium

(Received 20 December 1983; accepted for publication 2 March 1984)

Shift operators for the chain of Lie groups $SO(7) \supset [SU(2)]^3$ are constructed, and several properties are examined. Quite a number of relations connecting quadratic products of shift operators are built up. Finally, the connection between matrix elements of the shift operators and reduced matrix elements of the tensor operator, which forms part of the SO(7) generator basis, is explained.

PACS numbers: 03.65.Fd

I. INTRODUCTION

In a previous paper[1] the reduction chain $SO(7) \supset [SU(2)]^3$ was examined. The SO(7) generator basis was composed of the $[SU(2)]^3$ generators and the components of a tensor $T^{[1/2\ 1/2\ 1]}$, of rank $(\frac{1}{2}, \frac{1}{2}, 1)$ with respect to $[SU(2)]^3$. In the same reference, all the reduced matrix elements of the tensor $T^{[1/2\ 1/2\ 1]}$ were determined in the case of the symmetric representations $[v, 0, 0]$.

The application of the technique used in Ref. 1 to other representations than $[v, 0, 0]$, leads to extremely difficult recursion relations between reduced matrix elements of $T^{[1/2\ 1/2\ 1]}$. The solution of such relations would be very hard.

Therefore, another, more general method is introduced in the present paper: the shift operator technique. This technique was originally introduced by Hughes,[2,3] and used by himself and others to solve state labeling problems.[4–19]

In Sec. III, the $[SU(2)]^3$ shift operators for SO(7) are constructed. Some properties, derived from general relations for SU(2) shift operators,[20] are mentioned. Section IV contains the construction of relations connecting quadratic products of shift operators. Several kinds of formal transformation rules, by which relations turn into each other, are discussed. Finally, the connection between matrix elements of the shift operators and reduced matrix elements of the tensor $T^{[1/2\ 1/2\ 1]}$ is presented in Sec. V. The relations (4.2)–(4.12) will enable us to determine explicit expressions for the reduced matrix elements of $T^{[1/2\ 1/2\ 1]}$ for representations different from $[v, 0, 0]$. These applications are shown in a forthcoming paper.[21]

II. THE SO(7) GENERATORS, NOTATIONS AND CONVENTIONS

Since we are studying the chain $SO(7) \supset [SU(2)]^3$, it is convenient to make the subgroup structure appear explicitly in the SO(7) generator basis. The three commuting SU(2) subgroups are then generated by the sets $\{s_{-1}, s_0, s_{+1}\}$, $\{t_{-1}, t_0, t_{+1}\}$ and $\{u_{-1}, u_0, u_{+1}\}$, satisfying the well-known commutation relations

$$[s_0, s_{\pm 1}] = \pm s_{\pm 1}, \tag{2.1}$$

$$[s_{-1}, s_{+1}] = s_0, \tag{2.2}$$

and the same relations for $t_{0,\pm 1}$ and $u_{0,\pm 1}$. The remaining generators of SO(7) form a bispinor-vector $T^{[1/2\ 1/2\ 1]}_{\alpha\ \beta\ \gamma}$ $(\alpha, \beta = \pm\frac{1}{2}, \gamma = 0, \pm 1)$ under the $[SU(2)]^3$ subgroup. The subscripts on the bispinor-vector are s, t, u ordered, that is

$$[s_\mu, T^{[1/2\ 1/2\ 1]}_{\alpha\ \beta\ \gamma}] = \tfrac{1}{2}\sqrt{3} \langle \tfrac{1}{2}\alpha 1\mu | \tfrac{1}{2}\alpha + \mu \rangle T^{[1/2\ 1/2\ 1]}_{\alpha+\mu\ \beta\ \gamma}, \tag{2.3}$$

$$[t_\mu, T^{[1/2\ 1/2\ 1]}_{\alpha\ \beta\ \gamma}] = \tfrac{1}{2}\sqrt{3} \langle \tfrac{1}{2}\beta 1\mu | \tfrac{1}{2}\beta + \mu \rangle T^{[1/2\ 1/2\ 1]}_{\alpha\ \beta+\mu\ \gamma}, \tag{2.4}$$

$$[u_\mu, T^{[1/2\ 1/2\ 1]}_{\alpha\ \beta\ \gamma}] = \sqrt{2} \langle 1\gamma 1\mu | 1\ \gamma + \mu \rangle T^{[1/2\ 1/2\ 1]}_{\alpha\ \beta\ \gamma+\mu}. \tag{2.5}$$

Finally, the commutators of the components of the bispinor-vector among themselves are

$$\begin{aligned}
&[T^{[1/2\ 1/2\ 1]}_{\alpha_1\ \beta_1\ \gamma_1}, T^{[1/2\ 1/2\ 1]}_{\alpha_2\ \beta_2\ \gamma_2}] \\
&= (1/\sqrt{2})\delta_{\beta_1,-\beta_2}\delta_{\gamma_1,-\gamma_2}(-1)^{3/2-\beta_1-\gamma_1}\langle \tfrac{1}{2}\alpha_1\tfrac{1}{2}\alpha_2 | 1\ \alpha_1+\alpha_2\rangle s_{\alpha_1+\alpha_2} \\
&\quad + (1/\sqrt{2})\delta_{\alpha_1,-\alpha_2}\delta_{\gamma_1,-\gamma_2}(-1)^{3/2-\alpha_1-\gamma_1}\langle \tfrac{1}{2}\beta_1\tfrac{1}{2}\beta_2 | 1\ \beta_1+\beta_2\rangle t_{\beta_1+\beta_2} \\
&\quad + (1/\sqrt{2})\delta_{\alpha_1,-\alpha_2}\delta_{\beta_1,-\beta_2}(-1)^{1-\alpha_1-\beta_1}\langle 1\gamma_1 1\gamma_2 | 1\ \gamma_1+\gamma_2\rangle u_{\gamma_1+\gamma_2}.
\end{aligned} \tag{2.6}$$

[a] Research Assistant N.F.W.O. (Belgium).
[b] Research Assistant I.W.O.N.L. (Belgium).

 0022-2488/84/102953-05$02.50

Clearly, the SO(7) algebra is completely determined by (2.1)–(2.6).

The second-order Casimir operator of SO(7) has been constructed by Vanden Berghe *et al.*,[1] and reads (for reasons of simplicity, the tensor components $T^{[1/2\ 1/2\ 1]}_{\alpha\ \beta\ \gamma}$ are henceforeward denoted as $T_{\mathrm{sgn}(\alpha)\mathrm{sgn}(\beta)\mathrm{sgn}(\gamma)}$)

$$\begin{aligned} I_2 = & -2T_{+++}T_{---} + 2T_{++0}T_{--0} - 2T_{++-}T_{--+} \\ & + 2T_{+-+}T_{-+-} - 2T_{+-0}T_{-+0} + 2T_{+--}T_{-++} \\ & + 3s_0 - s_0(s_0-1) \\ & + 2s_{+1}s_{-1} + 2t_{+1}t_{-1} - t_0(t_0-1) + u_{+1}u_{-1} \\ & - \tfrac{1}{2}u_0(u_0-1). \end{aligned} \tag{2.7}$$

When the irreducible representations (irreps) of SO(7) are denoted by the Dynkin labels $[v_1, v_2, v_3]$, the I_2 eigenvalue is[22]

$$\begin{aligned} \langle I_2 \rangle = & -\frac{1}{2}\left\{ v_1(v_1+5) + 2v_2(v_2+4) \right. \\ & \left. + 3\frac{v_3}{2}\left(\frac{v_3}{2}+3\right) + 2v_1v_2 + v_1v_3 + 2v_2v_3 \right\}. \end{aligned} \tag{2.8}$$

It is known that one needs nine internal labels to classify the states of any SO(7) irrep $[v_1, v_2, v_3]$ unambiguously. The subgroup $[SU(2)]^3$ provides six label generating operators: S^2, s_0, T^2, t_0, U^2, and u_0. The expression of S^2 in terms of the basis generators is

$$S^2 = -2s_{+1}s_{-1} + s_0^2 - s_0; \tag{2.9}$$

analogous expressions are valid for T^2 and U^2. The eigenvalues of the above-mentioned list of six operators are, respectively, $s(s+1), \lambda, t(t+1), \mu, u(u+1)$, and ν. Then the basis states of an irrep of SO(7) can be written as

$$|v_1, v_2, v_3; \gamma\ s\ t\ u\ \lambda\ \mu\ \nu\rangle, \tag{2.10}$$

or, if confusion is excluded

$$|\gamma\ s\ t\ u\ \lambda\ \mu\ \nu\rangle. \tag{2.11}$$

Herein, γ denotes a set of three missing labels. If there is no degeneracy in the reduction of the SO(7) irrep into $[SU(2)]^3$ irreps (which is for instance the case for the representations $[v, 0, 0]$, $[v, 0, 1]$ and $[0, 0, v]$), we simply use the notation $|s\ t\ u\ \lambda\ \mu\ \nu\rangle$. Moreover, if λ, μ, and ν are irrelevant labels, we summarily denote the kets as $|s\ t\ u\rangle$. Formulas containing these shorthand notations should be understood as being valid for all λ, μ, and ν satisfying $\lambda \in \{-s, -s+1, \ldots, +s\}$, $\mu \in \{-t, \ldots, +t\}$ and $\nu \in \{-u, \ldots, +u\}$.

III. THE SHIFT OPERATORS $O^{i\ j\ k}_{s\ t\ u}$

A general analysis of SU(2) shift operators has been given by Hughes and Yadegar.[20] They have set up a formula for a shift operator $O^k_l (k = -j, -j+1, \ldots, +j)$ in terms of the SU(2) generators $l_{0,\pm}$ and the components of an SU(2) tensor $T^{[j]}_\mu (\mu = -j, \ldots, +j)$. The extension of the theory for $[SU(2)]^n$ shift operators has been studied elsewhere.[19,23] The main result of that study was that the expression for $[SU(2)]^n$ shift operators can be derived immediately from the corresponding expressions for ordinary SU(2) shift operators.

Consequently, in the present case the expressions for the shift operators $O^{i\ j\ k}_{s\ t\ u}$ built up with the tensor components $T^{[1/2\ 1/2\ 1]}_{\alpha\ \beta\ \gamma}$ can be deduced from the SU(2) shift operators for tensors $T^{[1/2]}_\mu$ and $T^{[1]}_\mu$. The latter forms are given by[20]

$$O_l^{+1/2} = -T^{[1/2]}_{-1/2}\,1_+ - (l+m+1)T^{[1/2]}_{1/2}, \tag{3.1}$$

$$O_l^{-1/2} = -T^{[1/2]}_{1/2}\,1_- + (l+m)T^{[1/2]}_{-1/2}, \tag{3.2}$$

$$\begin{aligned} O_l^{+1} = & (l-m+1)T^{[1]}_{-1}l_+ + (l+m+1)T^{[1]}_{+1}l_- \\ & + \sqrt{2}(l-m+1)(l+m+1)T^{[1]}_0, \end{aligned} \tag{3.3}$$

$$O_l^0 = T^{[1]}_{-1}l_+ - T^{[1]}_{+1}l_- + \sqrt{2}mT^{[1]}_0, \tag{3.4}$$

$$\begin{aligned} O_l^{-1} = & -(l+m)T^{[1]}_{-1}l_+ - (l-m)T^{[1]}_{+1}l_- \\ & + \sqrt{2}(l+m)(l-m)T^{[1]}_0. \end{aligned} \tag{3.5}$$

In Eqs. (3.1)–(3.5), $\{l_-, l_0, l_+\}$ is the generator basis of SU(2), where $l_\pm = \mp\sqrt{2}l_{\pm 1}$, and the commutation relations of the triple $\{l_{-1}, l_0, l_{+1}\}$ are determined in (2.1) and (2.2). The labels m and $l(l+1)$ are the eigenvalues of l_0 and L^2. The tensor $T^{[j]}_\mu$ satisfies

$$[l_0, T^{[j]}_\mu] = \mu T^{[j]}_\mu, \tag{3.6}$$

$$[l_\pm, T^{[j]}_\mu] = [(j \mp \mu)(j \pm \mu + 1)]^{1/2} T^{[j]}_{\mu \pm 1}. \tag{3.7}$$

If we define

$$s_\pm = \mp\sqrt{2}s_{\pm 1},\ t_\pm = \mp\sqrt{2}t_{\pm 1},\ \text{and}\ u_\pm = \mp\sqrt{2}u_{\pm 1}, \tag{3.8}$$

the expressions (3.1)–(3.5) can be used explicitly. Now formula (5.9) of Ref. 19 is applied to derive from (3.1)–(3.5) the expressions for the shift operators $O^{i\ j\ k}_{s\ t\ u}$ $(i, j \in \{-\frac{1}{2}, \frac{1}{2}\}$, $k \in \{-1, 0, +1\})$. We show, as an example, the operator $O^{1/2\ 1/2\ 1}_{s\ t\ u}$

$$\begin{aligned} O^{1/2\ 1/2\ 1}_{s\ t\ u} = & (s+\lambda+1)(t+\mu+1)\left[(u-\nu+1)T_{++-}u_+ + (u+\nu+1)T_{+++}u_- + \sqrt{2}(u+\nu+1)(u-\nu+1)T_{++0}\right] \\ & + (s+\lambda+1)\left[(u-\nu+1)T_{+--}u_+ + (u+\nu+1)T_{+-+}u_- + \sqrt{2}(u+\nu+1)(u-\nu+1)T_{+-0}\right] \\ & + (t+\mu+1)\left[(u-\nu+1)T_{-+-}u_+ + (u+\nu+1)T_{-++}u_- + \sqrt{2}(u+\nu+1)(u-\nu+1)T_{-+0}\right] \\ & + \left[(u-\nu+1)T_{---}u_+ + (u+\nu+1)T_{--+}u_- + \sqrt{2}(u+\nu+1)(u-\nu+1)T_{--0}\right]. \end{aligned} \tag{3.9}$$

The action of a shift operator upon an $[SU(2)]^3$ state is given by[20]

$$O^{i\ j\ k}_{s\ t\ u}|\gamma\ s\ t\ u\ \lambda\ \mu\ \nu\rangle = \sum_{\gamma'} c_{\gamma'}\,|\gamma'\ s+i\ t+j\ u+k\ \lambda+i\ \mu+j\ \nu\rangle. \tag{3.10}$$

Remark that the projection ν remains unchanged. This cannot be required for λ and μ, since s and t are altered by half-integer values. The expressions (3.1)–(3.2) are chosen in such way that λ (respectively, μ) changes by the same amount as s (respectively, t).

Let us finally recall some properties of SU(2) shift operators, which can be transferred immediately to the $[SU(2)]^3$

case. If j is an integer, the shift operator O_l^{-k} (constructed from the tensor $T^{[jl]}_{\mu}$) is obtained from the expression of O_l^{+k} by changing in the latter operator every l into $-l-1$; formally:

$$O_l^{-k} = O^{k}_{-l-1}. \tag{3.11}$$

If the rank of the tensor is half-integral, this rule is somewhat more complicated. For the operators (3.1)–(3.2), we deduce

$$O^{1/2}_{-l-1} = -\frac{l_+ O_l^{-1/2}}{(l+m)} = -\frac{O_l^{-1/2} l_+}{(l+m+1)},$$
$$O^{-1/2}_{-l-1} = \frac{l_- O^{1/2}}{(l+m+1)} = \frac{O_l^{1/2} l_-}{(l+m)}. \tag{3.12}$$

Evidently, because of the symmetry between the $SU(2)_s$ and $SU(2)_t$ subalgebras in SO(7), the shift operator $O^{j\ k}_{s\ t\ u}$ follows from the operator $O^{i\ j\ k}_{s\ t\ u}$ by performing the transformation

$$(s, \lambda, t, \mu) \rightarrow (t, \mu, s, \lambda) \tag{3.13}$$

in the expression of $O^{i\ j\ k}_{s\ t\ u}$.

IV. RELATIONS CONNECTING QUADRATIC PRODUCT OPERATORS

In order to derive reduced matrix elements of the tensor $T^{[1/2\ 1/2\ 1]}_{\alpha\ \beta\ \gamma}$, relations connecting quadratic product operators of the type

$$O^{i'\ j'\ k'}_{s+i\ t+j\ u+k} O^{i\ j\ k}_{s\ t\ u} \tag{4.1}$$

will be set up in this section. The operator (4.1) with $i'+i=\sigma, j'+j=\tau$, and $k'+k=\eta$ will be called "of the type (σ, τ, η)," or "belonging to the class (σ, τ, η)."

It is known that there exist certain relations between product operators of the same (σ, τ, η) class.[4,7–9,12,14,16,19] (The way to construct such relations is clearly explained in Ref. 19.) In the present case there are 45 nonzero (σ, τ, η) classes. However, due to transformation rules derived from (3.12)–(3.13), the relations between objects of some classes may be obtained from the relations between shift operators of another class, only by simple substitutions. Then, it is easy to see that the explicit construction of relations may be restricted to the eight classes listed in Table I. In the latter table, $\#(\sigma, \tau, \eta)$ indicates the number of quadratic products of shift operators of type (4.1) belonging to the class (σ, τ, η). Explicit calculations show that the minimum number of product operators belonging to a certain (σ, τ, η) class which occur in one relation, is a fixed number $R(\sigma, \tau, \eta)$, depending upon the (σ, τ, η)-values. This seems to be a general property of relations connecting shift operator products.[19] With every (σ, τ, η) class corresponds a maximum number $N(\sigma, \tau, \eta)$ of linearly independent relations, that is, every other possible relation between elements of the same (σ, τ, η) class is necessarily a linear combination (with s-, t-, and u-dependent coefficients) of the basis set of independent relations. All this information has been gathered in Table I.

The properties and transformation rules in Table I not only lead the relations within one class into the relations within another class, but can sometimes perform transformations within the same (σ, τ, η) class. This implies that it is not always necessary to construct $N(\sigma, \tau, \eta)$ relations, since the missing ones are derived from the given ones merely by an obvious substitution. Taking into account this remark, we conclude that the following basis set generates the complete set of relations between quadratic product operators

$(1°)\ (\sigma, \tau, \eta) = (-1, -1, -1)$

$$O^{-1/2\ -1/2\ -1}_{s-1/2\ t-1/2\ u} O^{-1/2\ -1/2\ 0}_{s\ t\ u} - O^{-1/2\ -1/2\ 0}_{s-1/2\ t-1/2\ u-1} O^{-1/2\ -1/2\ -1}_{s\ t\ u} = 0, \tag{4.2}$$

$(2°)\ (\sigma, \tau, \eta) = (-1, -1, 0)$

$$u^2 \frac{O^{-1/2\ -1/2\ -1}_{s-1/2\ t-1/2\ u+1} O^{-1/2\ -1/2\ 1}_{s\ t\ u}}{(u-\nu+1)(u+\nu+1)} - (u+1)^2 \frac{O^{-1/2\ -1/2\ 1}_{s-1/2\ t-1/2\ u-1} O^{-1/2\ -1/2\ -1}_{s\ t\ u}}{(u-\nu)(u+\nu)} - (2u+1) O^{-1/2\ -1/2\ 0}_{s-1/2\ t-1/2\ u} O^{-1/2\ -1/2\ 0}_{s\ t\ u} = 0, \tag{4.3}$$

$(3°)\ (\sigma, \tau, \eta) = (-1, 0, -2)$

$$\frac{O^{-1/2\ -1/2\ -1}_{s-1/2\ t+1/2\ u-1} O^{-1/2\ 1/2\ -1}_{s\ t\ u}}{(t+\mu+1)} - \frac{O^{-1/2\ 1/2\ -1}_{s-1/2\ t-1/2\ u-1} O^{-1/2\ -1/2\ -1}_{s\ t\ u}}{(t+\mu)} = 0, \tag{4.4}$$

$(4°)\ (\sigma, \tau, \eta) = (-1, 0, -1)$

$$2(t+1)u \frac{O^{-1/2\ 1/2\ -1}_{s-1/2\ t-1/2\ u} O^{-1/2\ -1/2\ 0}_{s\ t\ u}}{(t+\mu)} - (2t+1)(u+1) \frac{O^{-1/2\ -1/2\ 0}_{s-1/2\ t+1/2\ u-1} O^{-1/2\ 1/2\ -1}_{s\ t\ u}}{(t+\mu+1)} + (2t-u+1) \frac{O^{-1/2\ -1/2\ -1}_{s-1/2\ t+1/2\ u} O^{-1/2\ 1/2\ 0}_{s\ t\ u}}{(t+\mu+1)} = 0, \tag{4.5}$$

$(5°)\ (\sigma, \tau, \eta) = (-1, 0, 0)$

$$(2t+1)(u+1)^2 \frac{O^{-1/2\ 1/2\ 1}_{s-1/2\ t-1/2\ u-1} O^{-1/2\ -1/2\ -1}_{s\ t\ u}}{(t+\mu)(u-\nu)(u+\nu)} - (2t+1)u^2 \frac{O^{-1/2\ -1/2\ -1}_{s-1/2\ t+1/2\ u+1} O^{-1/2\ 1/2\ 1}_{s\ t\ u}}{(t+\mu+1)(u+\nu+1)(u-\nu+1)} + (u+1)[(u+1)(2u-1)+2(t+1)] \times \frac{O^{-1/2\ 1/2\ 0}_{s-1/2\ t-1/2\ u} O^{-1/2\ -1/2\ 0}_{s\ t\ u}}{(t+\mu)} - u[u(2u+3)-2t] \frac{O^{-1/2\ -1/2\ 0}_{s-1/2\ t+1/2\ u} O^{-1/2\ 1/2\ 0}_{s\ t\ u}}{(t+\mu+1)} = 0, \tag{4.6}$$

TABLE I. Properties of relations connecting quadratic product operators.

(σ, τ, η)	$\#(\sigma, \tau, \eta)$	$R(\sigma, \tau, \eta)$	$N(\sigma, \tau, \eta)$
$(-1, -1, -1)$	2	2	1
$(-1, -1, 0)$	3	3	1
$(-1, 0, -2)$	2	2	1
$(-1, 0, -1)$	4	3	2
$(-1, 0, 0)$	6	4	3
$(0, 0, -2)$	4	3	2
$(0, 0, -1)$	8	5	4
$(0, 0, 0)$	12	6	7

$$(2u+3)\frac{O^{-1/2\ 1/2\ 1}_{s-1/2\ t-1/2\ u-1}O^{-1/2\ -1/2\ -1}_{s\ t\ u}}{(t+\mu)(u-\nu)(u+\nu)}$$
$$-(2u-1)\frac{O^{-1/2\ -1/2\ -1}_{s-1/2\ t+1/2\ u+1}O^{-1/2\ 1/2\ 1}_{s\ t\ u}}{(t+\mu+1)(u+\nu+1)(u-\nu+1)}$$
$$+(2u-1)\frac{O^{-1/2\ 1/2\ -1}_{s-1/2\ t-1/2\ u+1}O^{-1/2\ -1/2\ 1}_{s\ t\ u}}{(t+\mu+1)(u+\nu+1)(u-\nu+1)}$$
$$-(2u+3)\frac{O^{-1/2\ -1/2\ 1}_{s-1/2\ t+1/2\ u-1}O^{-1/2\ 1/2\ -1}_{s\ t\ u}}{(t+\mu+1)(u-\nu)(u+\nu)}=0, \quad (4.7)$$

$(6°)\ (\sigma,\tau,\eta)=(0,0,-2)$

$$t(2s+1)\frac{O^{-1/2\ -1/2\ -1}_{s+1/2\ t+1/2\ u-1}O^{1/2\ 1/2\ -1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu+1)}$$
$$-(s+1)(2t+1)\frac{O^{1/2\ 1/2\ -1}_{s-1/2\ t-1/2\ u-1}O^{-1/2\ -1/2\ -1}_{s\ t\ u}}{(s+\lambda)(t+\mu)}$$
$$+(s+t+1)\frac{O^{-1/2\ 1/2\ -1}_{s+1/2\ t-1/2\ u-1}O^{1/2\ -1/2\ -1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu)}=0, \quad (4.8)$$

$(7°)\ (\sigma,\tau,\eta)=(0,0,-1)$

$$[2(s+1)(t+1)(u+1)(2u-1)-(s+1)u(u-1)$$
$$-(t+1)(u^2+2u-1)]\frac{O^{-1/2\ -1/2\ -1}_{s+1/2\ t+1/2\ u}O^{1/2\ 1/2\ 0}_{s\ t\ u}}{(s+\lambda+1)(t+\mu+1)}$$
$$-(2s+1)(t+1)(u+1)(2u-1)$$
$$\times\frac{O^{-1/2\ -1/2\ 0}_{s+1/2\ t+1/2\ u-1}O^{1/2\ 1/2\ -1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu+1)}$$
$$+(2s+1)(t+1)(u-1)u$$
$$\times\frac{O^{-1/2\ 1/2\ -1}_{s+1/2\ t-1/2\ u}O^{1/2\ -1/2\ 0}_{s\ t\ u}}{(s+\lambda+1)(t+\mu)}$$
$$+(2s+1)(t+1)u(u+1)$$
$$\times\frac{O^{-1/2\ 1/2\ 0}_{s+1/2\ t-1/2\ u-1}O^{1/2\ -1/2\ -1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu)}$$
$$+[-4(s+1)(t+1)u^2+(s+1)u(u-1)$$
$$+(t+1)u(u+1)]$$
$$\times\frac{O^{1/2\ -1/2\ -1}_{s-1/2\ t+1/2\ u}O^{-1/2\ 1/2\ 0}_{s\ t\ u}}{(s+\lambda)(t+\mu+1)}=0; \quad (4.9)$$

$(8°)\ (\sigma,\tau,\eta)=(0,0,0)$

$$u\frac{O^{-1/2\ -1/2\ -1}_{s+1/2\ t+1/2\ u+1}O^{1/2\ 1/2\ 1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu+1)(u+\nu+1)(u-\nu+1)}$$
$$+(2u+1)\frac{O^{-1/2\ -1/2\ 0}_{s+1/2\ t+1/2\ u}O^{1/2\ 1/2\ 0}_{s\ t\ u}}{(s+\lambda+1)(t+\mu+1)}$$
$$+(u+1)\frac{O^{-1/2\ -1/2\ 1}_{s+1/2\ t+1/2\ u-1}O^{1/2\ 1/2\ -1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu+1)(u+\nu)(u-\nu)}$$
$$-u\frac{O^{-1/2\ 1/2\ -1}_{s+1/2\ t-1/2\ u+1}O^{1/2\ -1/2\ 1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu)(u+\nu+1)(u-\nu+1)}$$
$$-(2u+1)\frac{O^{-1/2\ 1/2\ 0}_{s+1/2\ t-1/2\ u}O^{1/2\ -1/2\ 0}_{s\ t\ u}}{(s+\lambda+1)(t+\mu)}$$
$$-(u+1)\frac{O^{-1/2\ 1/2\ 1}_{s+1/2\ t-1/2\ u-1}O^{1/2\ -1/2\ -1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu)(u+\nu)(u-\nu)}$$
$$-2u(u+1)(2u+1)(s+1)(2t+1)$$
$$\times[I_2+s(s+4)+t(t+1)+\tfrac{1}{2}u(u+1)]=0, \quad (4.10)$$

$$[(t+1)(2u+3)+t]$$
$$\times\frac{O^{-1/2\ -1/2\ -1}_{s+1/2\ t+1/2\ u+1}O^{1/2\ 1/2\ 1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu+1)(u+\nu+1)(u-\nu+1)}$$
$$+(2t+1)(2u+3)\frac{O^{-1/2\ -1/2\ 0}_{s+1/2\ t+1/2\ u}O^{1/2\ 1/2\ 0}_{s\ t\ u}}{(s+\lambda+1)(t+\mu+1)}$$
$$-2(t+1)(u+1)$$
$$\times\frac{O^{-1/2\ 1/2\ -1}_{s+1/2\ t-1/2\ u+1}O^{1/2\ -1/2\ 1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu)(u+\nu+1)(u-\nu+1)}$$
$$-[(t+1)(2u+3)+t]$$
$$\times\frac{O^{1/2\ -1/2\ -1}_{s-1/2\ t+1/2\ u+1}O^{-1/2\ 1/2\ 1}_{s\ t\ u}}{(s+\lambda)(t+\mu+1)(u+\nu+1)(u-\nu+1)}$$
$$-(2t+1)(2u+3)\frac{O^{1/2\ -1/2\ 0}_{s-1/2\ t+1/2\ u}O^{-1/2\ 1/2\ 0}_{s\ t\ u}}{(s+\lambda)(t+\mu+1)}$$
$$+2(t+1)(u+1)\frac{O^{1/2\ 1/2\ -1}_{s-1/2\ t-1/2\ u+1}O^{-1/2\ -1/2\ 1}_{s\ t\ u}}{(s+\lambda)(t+\mu)(u+\nu+1)(u-\nu+1)}$$
$$-2(2s+1)(2t+1)(t+1)(u+1)^2(2u+3)$$
$$\times[I_2+s(s+1)+t(t+2)+\tfrac{1}{2}u(u+5)]=0, \quad (4.11)$$

$$(u+1)(2u+1)[(2t+1)(2u+1)+2(2s+1)]$$
$$\times\frac{O^{-1/2\ -1/2\ 0}_{s+1/2\ t+1/2\ u}O^{1/2\ 1/2\ 0}_{s\ t\ u}}{(s+\lambda+1)(t+\mu+1)}$$
$$+(u+1)(2u+1)[2s(u+2)+2(t+1)(u+1)+1]$$
$$\times\frac{O^{-1/2\ -1/2\ 1}_{s+1/2\ t+1/2\ u-1}O^{1/2\ 1/2\ -1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu+1)(u+\nu)(u-\nu)}$$
$$-u(2t+1)[2(s+1)+u(2u+3)]$$
$$\times\frac{O^{-1/2\ 1/2\ -1}_{s+1/2\ t-1/2\ u+1}O^{1/2\ -1/2\ 1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu)(u+\nu+1)(u-\nu+1)}$$
$$-(2u+1)[4(s+1)(t+1)(2u+3)$$
$$-2(s+1)+2(t+1)(2u^2+u-2)-u(2u+3)]$$
$$\times\frac{O^{-1/2\ 1/2\ 0}_{s+1/2\ t-1/2\ u}O^{1/2\ -1/2\ 0}_{s\ t\ u}}{(s+\lambda+1)(t+\mu)}$$
$$-(u+1)[2(2s+1)(t+1)(u+2)(2u+1)$$
$$+(2s+1)(2t+1)+(2t+1)(u+1)(2u+1)]$$
$$\times\frac{O^{-1/2\ 1/2\ 1}_{s+1/2\ t-1/2\ u-1}O^{1/2\ -1/2\ -1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu)(u+\nu)(u-\nu)}$$
$$+2u(2t+1)(s+1)(u+1)(2u+1)$$
$$\times\frac{O^{1/2\ -1/2\ -1}_{s-1/2\ t+1/2\ u+1}O^{-1/2\ 1/2\ 1}_{s\ t\ u}}{(s+\lambda)(t+\mu+1)(u+\nu+1)(u-\nu+1)}$$
$$-2(s+1)(2t+1)u(u+1)(2u+1)$$
$$\times\{[(2t+1)u(2u+3)-2(s-t)]$$
$$\times[I_2+s(s+1)+t(t+1)+\tfrac{1}{2}u(u+1)]$$
$$-4(t+1)u^2(2s+1)(u+t-s+1)$$
$$-10(2s+1)(t+1)(t-s)u-6(t+1)u(-s+u+1)$$
$$-3su(2u+3)-6(t-s)(2st+s+t+1)\}=0. \quad (4.12)$$

V. THE RELATION BETWEEN MATRIX ELEMENTS OF THE SHIFT OPERATORS AND REDUCED MATRIX ELEMENTS OF THE TENSOR

In their general analysis[20] Hughes and Yadegar have given the connection between the actions of the shift operators and the reduced matrix elements of the tensor of which the shift operators are composed. For the operators (3.1)–(3.5), we derive

$$\frac{O_l^{1/2}}{[l+m+1]^{1/2}}|\gamma, l, m\rangle = -\left(\frac{2l+1}{2l+2}\right)^{1/2}\sum_{\gamma'}\langle\gamma', l+\tfrac{1}{2}\|T^{[1/2]}\|\gamma, l\rangle|\gamma', l+\tfrac{1}{2}, m+\tfrac{1}{2}\rangle; \tag{5.1}$$

$$\frac{O_l^{-1/2}}{[l+m]^{1/2}}|\gamma, l, m\rangle = \left(\frac{2l+1}{2l}\right)^{1/2}\sum_{\gamma'}\langle\gamma', l-\tfrac{1}{2}\|T^{[1/2]}\|\gamma, l\rangle|\gamma', l-\tfrac{1}{2}, m-\tfrac{1}{2}\rangle; \tag{5.2}$$

$$\frac{O_l^{+1}}{[(l+m+1)(l-m+1)]^{1/2}}|\gamma, l, m\rangle = \left[\frac{(2l+2)(2l+1)}{(2l+3)}\right]^{1/2}\sum_{\gamma'}\langle\gamma', l+1\|T^{[1]}\|\gamma, l\rangle|\gamma', l+1, m\rangle; \tag{5.3}$$

$$O_l^0|\gamma, l, m\rangle = \left[\frac{2l(l+1)}{2l+1}\right]^{1/2}\sum_{\gamma'}\langle\gamma', l\|T^{[1]}\|\gamma, l\rangle|\gamma', l, m\rangle; \tag{5.4}$$

$$\frac{O_l^{-1}}{[(l+m)(l-m)]^{1/2}}|\gamma, l, m\rangle = -\left[\frac{(2l+1)(2l)}{(2l-1)}\right]^{1/2}\sum_{\gamma'}\langle\gamma', l-1\|T^{[1]}\|\gamma, l\rangle|\gamma', l-1, m\rangle. \tag{5.5}$$

Here again, γ denotes a set of labels needed to specify the states completely.

It is now straightforward to deduce from (5.1)–(5.5) the similar properties for the shift operators $O_{s\ t\ u}^{i\ j\ k}$, defined in Sec. III. We give, as an example

$$\begin{aligned} O_{s\ t\ u}^{1/2\ 1/2\ 1}/[(s+\lambda+1)(t+\mu+1)(u+\nu+1)(u-\nu+1)]^{1/2}|\gamma\, s\, t\, u\, \lambda\, \mu\, \nu\rangle \\ = \left[\frac{(2s+1)}{(2s+2)}\frac{(2t+1)}{(2t+2)}\frac{(2u+2)(2u+1)}{(2u+3)}\right]^{1/2}\sum_{\gamma'}\langle\gamma'\, s+\tfrac{1}{2}\, t+\tfrac{1}{2}\, u+1\|T^{[1/2\ 1/2\ 1]}\|\gamma\, s\, t\, u\rangle \\ \times|\gamma'\, s+\tfrac{1}{2}\, t+\tfrac{1}{2}\, u+1\, \lambda+\tfrac{1}{2}\, \mu+\tfrac{1}{2}\, \nu\rangle. \end{aligned} \tag{5.6}$$

For some SO(7) representations, there is no degeneracy in the reduction to $[SU(2)]^3$. Then the basis states of this representation are completely labeled by the $[SU(2)]^3$ labels s, t, u, λ, μ, and ν. In this case, Eq. (5.6) reads

$$\begin{aligned} \frac{\langle s+\tfrac{1}{2}\, t+\tfrac{1}{2}\, u+1\, \lambda+\tfrac{1}{2}\, \mu+\tfrac{1}{2}\, \nu|O_{s\ t\ u}^{1/2\ 1/2\ 1}|s\, t\, u\, \lambda\, \mu\, \nu\rangle}{[(s+\lambda+1)(t+\mu+1)(u+\nu+1)(u-\nu+1)]^{1/2}} \\ = \left[\frac{(2s+1)}{(2s+2)}\frac{(2t+1)}{(2t+2)}\frac{(2u+2)(2u+1)}{(2u+3)}\right]^{1/2}\langle s+\tfrac{1}{2}\, t+\tfrac{1}{2}\, u+1\|T^{[1/2\ 1/2\ 1]}\|s\, t\, u\rangle. \end{aligned} \tag{5.7}$$

Relations such as (5.7) will prove to be very useful in a forthcoming paper,[21] where all possible reduced matrix elements of the tensor $T^{[1/2\ 1/2\ 1]}$ will be determined in cases of no degeneracy.

ACKNOWLEDGMENTS

We gratefully acknowledge H. De Meyer and G. Vanden Berghe (Rijksuniversiteit-Gent) for many profitable conversations, as well as Professor C. Grosjean for his continuous interest.

[1]G. Vanden Berghe, H. De Meyer, and P. De Wilde, J. Phys. A: Math. Gen. **15**, 2677 (1982).
[2]J. W. B. Hughes, J. Phys. A: Math. Gen. **6**, 48 (1973).
[3]J. W. B. Hughes, J. Phys. A: Math. Gen. **6**, 281 (1973).
[4]G. Vanden Berghe and H. De Meyer, J. Math. Phys. **21**, 1902 (1980).
[5]H. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 1906 (1980).
[6]H. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 1913 (1980).
[7]G. Vanden Berghe and H. De Meyer, J. Math. Phys. **21**, 1967 (1980).
[8]H. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 1973 (1980).
[9]G. Vanden Berghe and H. De Meyer, J. Math. Phys. **21**, 2632 (1980).
[10]H. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 2635 (1980).
[11]G. Vanden Berghe and H. De Meyer, J. Math. Phys. **22**, 2326 (1981).
[12]G. Vanden Berghe, H. De Meyer, and J. Van der Jeugt, J. Math. Phys. **22**, 2332 (1981).
[13]H. De Meyer, G. Vanden Berghe, and J. Van der Jeugt, J. Math. Phys. **23**, 1223 (1982).
[14]J. Van der Jeugt, H. De Meyer, and G. Vanden Berghe, J. Math. Phys. **24**, 1339 (1983).
[15]J. Van der Jeugt, G. Vanden Berghe, and H. De Meyer, J. Math. Phys. **24**, 1345 (1983).
[16]H. De Meyer, G. Vanden Berghe, and J. W. B. Hughes, J. Math. Phys. **22**, 2360 (1981).
[17]H. De Meyer, G. Vanden Berghe, and J. W. B. Hughes, J. Math. Phys. **22**, 2366 (1981).
[18]J. Van der Jeugt, H. De Meyer, and G. Vanden Berghe, J. Math. Phys. **24**, 1025 (1983).
[19]J. Van der Jeugt, H. De Meyer, and G. Vanden Berghe, Ann. Phys. **147**, 85 (1983).
[20]J. W. B. Hughes and J. Yadegar, J. Math. Phys. **19**, 2068 (1978).
[21]P. De Wilde and J. Van der Jeugt, J. Math. Phys. **25**, 2958 (1984).
[22]B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill, New York, 1963).
[23]J. W. B. Hughes, J. Math. Phys. **24**, 1015 (1983).

The shift operator technique for SO(7) in an [SU(2)]³ basis. II. Applications

P. DeWilde[a] and J. Van der Jeugt[b]
Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit-Gent, Krijgslaan 281-S9, B-9000 Gent, Belgium

(Received 20 December 1983; accepted for publication 2 March 1984)

Reduced matrix elements of the tensor operator which forms part of the SO(7) generator basis are calculated for the three respective classes of SO(7) representations $[v,0,0]$, $v\in\mathbb{N}_0$, $[v,0,1]$, $v\in\mathbb{N}_0$, $[0,0,v]$, $v\in\mathbb{N}_0$. The calculation is based on the relations between quadratic products of shift operators, established in a previous paper by the same authors.

PACS numbers: 03.65.Fd

I. INTRODUCTION

The generators of SO(7) can be grouped into the generators of the three SU(2) subgroups and into a bispinor-vector $T^{[1/2\ 1/2\ 1]}$, as was demonstrated by Vanden Berghe *et al.*[1] The reduced matrix elements of the generators of $[SU(2)]^2$ are well known. Those of the bispinor-vector have only been calculated in the case of the symmetric SO(7) representations $[v,0,0]$, $v\in\mathbb{N}_0$, in Ref. 1.

The three classes of SO(7) representations $[v,0,0]$, $v\in\mathbb{N}_0$, $[v,0,1]$, $v\in\mathbb{N}_0$, and $[0,0,v]$, $v\in\mathbb{N}_0$ reduce without degeneracy into $[SU(2)]^3$ representations (s,t,u). This will allow us to calculate the matrix elements of the bispinor-vector between $[SU(2)]^3$ states $|stu\lambda\mu\nu\rangle$ that form a basis for the three respective above-mentioned classes of SO(7) representations. We thus distinguish three cases and in each case the reduced matrix elements that have to be considered will depend on the specific $SO(7)\to[SU(2)]^3$ branching rule for that class of SO(7) representations.

The reduced matrix elements are calculated by a method entirely based on the use of the relations between quadratic products of shift operators. These relations have been established in the previous paper.[3]

In the second section of this paper the reduced matrix elements for the $[v,0,0]$, $v\in\mathbb{N}_0$ representations are concisely calculated as a verification. In the third section the matrix elements for the $[v,0,1]$, $v\in\mathbb{N}_0$ representations are calculated and particular features of the solution by means of relations between shift operators are illustrated. In the fourth section the matrix elements for the $[0,0,v]$, $v\in\mathbb{N}_0$ representations are calculated. We end with a remark about symmetry properties of the reduced matrix elements.

II. REDUCED MATRIX ELEMENTS IN THE $[v,0,0]$, $v\in\mathbb{N}_0$, REPRESENTATION

In the present article, frequent use will be made of the matrix elements of the products of shift operators occurring in Eqs. (4.10)–(4.12) of Van der Jeugt and De Wilde[3] (to be referred to as I), between $[SU(2)]^3$ states $\langle stu\lambda\mu\nu|$ and $|stu\lambda\mu\nu\rangle$. We therefore introduce the following notations:

$$
\begin{aligned}
A &= \left\langle stu\lambda\mu\nu \left| \frac{O^{-1/2\ \ -1/2\ \ -1}_{s+1/2\ t+1/2\ u+1}\, O^{1/2\ 1/2\ 1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu+1)(u+\nu+1)(u-\nu+1)} \right| stu\lambda\mu\nu \right\rangle,\\
B &= \left\langle stu\lambda\mu\nu \left| \frac{O^{-1/2\ \ -1/2\ \ 0}_{s+1/2\ t+1/2\ u}\, O^{1/2\ 1/2\ 0}_{s\ t\ u}}{(s+\lambda+1)(t+\mu+1)} \right| stu\lambda\mu\nu \right\rangle,\\
C &= \left\langle stu\lambda\mu\nu \left| \frac{O^{-1/2\ \ -1/2\ \ 1}_{s+1/2\ t+1/2\ u-1}\, O^{1/2\ 1/2\ -1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu+1)(u+\nu)(u-\nu)} \right| stu\lambda\mu\nu \right\rangle,\\
D &= \left\langle stu\lambda\mu\nu \left| \frac{O^{-1/2\ \ 1/2\ \ -1}_{s+1/2\ t-1/2\ u+1}\, O^{1/2\ -1/2\ 1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu)(u+\nu+1)(u-\nu+1)} \right| stu\lambda\mu\nu \right\rangle,\\
E &= \left\langle stu\lambda\mu\nu \left| \frac{O^{-1/2\ \ 1/2\ \ 0}_{s+1/2\ t-1/2\ u}\, O^{1/2\ -1/2\ 0}_{s\ t\ u}}{(s+\lambda+1)(t+\mu)} \right| stu\lambda\mu\nu \right\rangle,\\
F &= \left\langle stu\lambda\mu\nu \left| \frac{O^{-1/2\ \ 1/2\ \ 1}_{s+1/2\ t-1/2\ u-1}\, O^{1/2\ -1/2\ -1}_{s\ t\ u}}{(s+\lambda+1)(t+\mu)(u+\nu)(u-\nu)} \right| stu\lambda\mu\nu \right\rangle,\\
G &= \left\langle stu\lambda\mu\nu \left| \frac{O^{1/2\ \ -1/2\ \ -1}_{s-1/2\ t+1/2\ u+1}\, O^{-1/2\ 1/2\ 1}_{s\ t\ u}}{(s+\lambda)(t+\mu+1)(u+\nu+1)(u-\nu+1)} \right| stu\lambda\mu\nu \right\rangle,\\
H &= \left\langle stu\lambda\mu\nu \left| \frac{O^{1/2\ \ -1/2\ \ 0}_{s-1/2\ t+1/2\ u}\, O^{-1/2\ 1/2\ 0}_{s\ t\ u}}{(s+\lambda)(t+\mu+1)} \right| stu\lambda\mu\nu \right\rangle,
\end{aligned}
\tag{2.1}
$$

[a] Research Assistant I.W.O.N.L. (Belgium).
[b] Research Assistant N.F.W.O. (Belgium).

 0022-2488/84/102958-08$02.50

$$I=\left\langle stu\lambda\mu\nu\left|\frac{O^{\,1/2\ \ -1/2\ \ 1}_{\,s-1/2\ t+1/2\ u-1}\,O^{-1/2\ 1/2\ -1}_{\,s\ \ t\ \ u}}{(s+\lambda)(t+\mu+1)(u+\nu)(u-\nu)}\right|stu\lambda\mu\nu\right\rangle,$$

$$J=\left\langle stu\lambda\mu\nu\left|\frac{O^{\,1/2\ \ 1/2\ \ -1}_{\,s-1/2\ t-1/2\ u+1}\,O^{-1/2\ -1/2\ 1}_{\,s\ \ t\ \ u}}{(s+\lambda)(t+\mu)(u+\nu+1)(u-\nu+1)}\right|stu\lambda\mu\nu\right\rangle,$$

$$K=\left\langle stu\lambda\mu\nu\left|\frac{O^{\,1/2\ \ 1/2\ \ 0}_{\,s-1/2\ t-1/2\ u}\,O^{-1/2\ -1/2\ 0}_{\,s\ \ t\ \ u}}{(s+\lambda)(t+\mu)}\right|stu\lambda\mu\nu\right\rangle,$$

$$L=\left\langle stu\lambda\mu\nu\left|\frac{O^{\,1/2\ \ 1/2\ \ 1}_{\,s-1/2\ t-1/2\ u-1}\,O^{-1/2\ -1/2\ -1}_{\,s\ \ t\ \ u}}{(s+\lambda)(t+\mu)(u+\nu)(u-\nu)}\right|stu\lambda\mu\nu\right\rangle.$$

Due to the branching rule[2] for the symmetric representations $[v,0,0]$, $v\in\mathbb{N}_0$, of SO(7) into representations of the subgroup $[SU(2)]^3$, only the following reduced matrix elements of generators of SO(7) are nontrivial:

$$\langle s+\tfrac{1}{2}\ u+1\|T^{[1/2\ 1/2\ 1]}\|su\rangle,$$
$$\langle s+\tfrac{1}{2}\ u-1\|T^{[1/2\ 1/2\ 1]}\|su\rangle.$$

We will denote them, respectively, by $A(s,u)$ and $B(s,u)$. The symbol $T^{[1/2\ 1/2\ 1]}$ is the bispinor-vector, introduced in Vanden Berghe *et al.*[1]

The connection between the symbols introduced in Eq. (2.1) and the reduced matrix elements can be established by means of (I5.1)–(I5.5), as illustrated in (I5.6)–(I5.7). We obtain

$$A=-(2u+2)|A(s,u)|^2, \tag{2.2}$$
$$C=-2u|B(s,u)|^2. \tag{2.3}$$

As a result of the above-mentioned branching rule, $B=D=E=F=G=H=I=K=0$.

By considering the matrix elements between $[SU(2)]^3$ states $\langle stu\lambda\mu\nu|$ and $|stu\lambda\mu\nu\rangle$ of the expressions occurring in the right- and left-hand sides of Eq. (I4.12) we obtain, in the notation (2.1),

$$\begin{aligned}&(2s+1)(u+1)(2u+1)\\&\quad\times(2u+3)C-2(s+1)(2s+1)u(u+1)(2u+1)\\&\quad\times\{(2s+1)u(2u+3)[\langle I_2\rangle+2s(s+1)+\tfrac{1}{2}u(u+1)]\\&\quad-2s(s+1)u(2u-1)(2u+3)+2(s+1)\\&\quad\times u(u+1)(2u+3)+3su(2u+3)\}=0.\end{aligned} \tag{2.4}$$

For the representations $[v,0,0]$, $v\in\mathbb{N}_0$ of SO(7), $\langle I_2\rangle=-\frac{1}{2}v(v+5)$, as is easily seen from (I2.8). Substituting this result and (2.3) in (2.4) we obtain

$$\begin{aligned}&|B(s,u)|^2\\&\quad=\tfrac{1}{4}(2s+1)(2s+2)u(v+2s-u+4)(v-2s+u+1).\end{aligned} \tag{2.5}$$

As a result of (I3.11), A can be obtained by the formal substitution $u\to -u-1$ in C. On account of (2.2) and (2.3) this implies (formally)

$$|A(s,u)|^2=-|B(s,u-1)|^2. \tag{2.6}$$

In this way we find

$$\begin{aligned}|A(s,u)|^2=\tfrac{1}{4}&(2s+1)(2s+2)(u+1)\\&\times(v+2s+u+5)(v-2s-u).\end{aligned} \tag{2.7}$$

The expressions (2.5) and (2.7) for the reduced matrix elements are in agreement with those calculated in Vanden Berghe *et al.*[1]

III. REDUCED MATRIX ELEMENTS IN THE $[v,0,1]$, $v\in\mathbb{N}_0$, REPRESENTATION

In order to choose a set of independent reduced matrix elements of $T^{[1/2\ 1/2\ 1]}$ in the $[SU(2)]^3$ basis one needs to consider the SO(7)→$[SU(2)]^3$ branching rule for $[v,0,1]$, $v\in\mathbb{N}_0$ representations of SO(7) into $[SU(2)]^3$ representations (s,t,u). It reads

$$[v,0,1]\to\sum_{s,t,u}(s,t,u),$$

with

$$u=v+\tfrac{1}{2},\ v-\tfrac{1}{2},\ldots,\tfrac{1}{2},$$
$$|s-t|=\tfrac{1}{2},$$
$$s+t=v+1-u,\ v-u,\ldots,\tfrac{1}{2}.$$

This rule is considered as a special case of a more general branching rule in De Wilde *et al.*[4]

On account of symmetry operations with respect to the first two labels and of the property

$$\begin{aligned}&\langle s't'u'\|T^{[1/2\ 1/2\ 1]}\|stu\rangle^*\\&\quad=(-1)^{s'-s+t'-t+u'-u}\langle stu\|T^{[1/2\ 1/2\ 1]}\|s't'u'\rangle,\end{aligned} \tag{3.1}$$

all other reduced matrix elements of $T^{[1/2\ 1/2\ 1]}$ can be related to

$$\begin{aligned}A(s,u)&=\langle s+\tfrac{1}{2}\,s+1\ u+1\|T^{[1/2\ 1/2\ 1]}\|s\,s+\tfrac{1}{2}\,u\rangle,\\B(s,u)&=\langle s+\tfrac{1}{2}\,s+1\ u\ \ \|T^{[1/2\ 1/2\ 1]}\|s\,s+\tfrac{1}{2}\,u\rangle,\\C(s,u)&=\langle s+\tfrac{1}{2}\,s+1\ u-1\|T^{[1/2\ 1/2\ 1]}\|s\,s+\tfrac{1}{2}\,u\rangle,\\D(s,u)&=\langle s+\tfrac{1}{2}\,s\ u+1\|T^{[1/2\ 1/2\ 1]}\|s\,s+\tfrac{1}{2}\,u\rangle,\\E(s,u)&=\langle s+\tfrac{1}{2}\,s\ u\ \ \|T^{[1/2\ 1/2\ 1]}\|s\,s+\tfrac{1}{2}\,u\rangle.\end{aligned} \tag{3.2}$$

Equations (I5.1)–(I5.5) can be used to establish the relations

$$A=-(2u+2)|A(s,u)|^2, \tag{3.3}$$
$$B=-\frac{2u(u+1)}{2u+1}|B(s,u)|^2, \tag{3.4}$$
$$C=-2u|C(s,u)|^2, \tag{3.5}$$
$$D=(2u+2)|D(s,u)|^2, \tag{3.6}$$
$$E=\frac{2u(u+1)}{2u+1}|E(s,u)|^2, \tag{3.7}$$
$$F=2u|D(s,u-1)|^2, \tag{3.8}$$
$$J=-(2u+2)|C(s-\tfrac{1}{2},u+1)|^2, \tag{3.9}$$
$$L=-2u|A(s-\tfrac{1}{2},u-1)|^2. \tag{3.10}$$

As a result of the condition $|s-t|=\frac{1}{2}$ in the branching rule, $G=H=I=0$ if $t=s+\frac{1}{2}$.

In contrast with the reduced matrix elements in the $[v,0,0]$, $v\in\mathbb{N}_0$ representation, no single equation containing only one reduced matrix element can be obtained from the set (I4.2)–(I4.12) by taking matrix elements, even with the use of the properties (I3.11)–(I3.12). It is possible, however, to obtain a set of recursion relations containing only $A(s,u)$ and $C(s,u)$.

Substitution of $s\to -s-1$ and $t\to -t-1$ in Eq. (I4.10), taking into account the property (I3.12), and considering matrix elements, gives

$$uG+(2u+1)H+(u+1)I-uJ-(2u+1)K-(u+1)L$$
$$+2s(2t+1)u(u+1)(2u+1)[\langle I_2\rangle+(s-3)(s+1)+t(t+1)+\tfrac{1}{2}u(u+1)]=0. \tag{3.11}$$

Substitution of $s\to -s-1$ in (I4.12) gives

$$2s(2t+1)u(u+1)(2u+1)A+[2(2s+1)-(2t+1)(2u+1)](u+1)(2u+1)H$$
$$+[2(s+1)(u+2)-2(t+1)(u+1)-1](u+1)(2u+1)I-[2s-u(2u+3)](2t+1)uJ+[2s-4s(t+1)(2u+3)$$
$$+2(t+1)(2u^2+u+2)-u(2u+3)](2u+1)K-[(2s+1)(2t+1)+2(2s+1)(t+1)(u+2)(2u+1)$$
$$-(2t+1)(u+1)(2u+1)](u+1)L-2s(2t+1)u(u+1)(2u+1)\{[2(s+t+1)+(2t+1)u(2u+3)]$$
$$\times[\langle I_2\rangle+s(s+1)+t(t+1)+\tfrac{1}{2}u(u+1)]+2(2s+1)(s+t+1)(t+1)u(2u+5)+2(s+1)(t+1)u(2u-1)(2u+3)$$
$$-2(t+1)u(u+1)(2u+3)+3(s+1)u(2u+3)+6(s+t+1)(2st+s+t)\}=0. \tag{3.12}$$

If we let $t=s+\frac{1}{2}$, take into account $G=H=I=0$, (3.3), (3.9), (3.10), the expectation value (I2.8) of I_2, which equals $-\frac{1}{8}(4v^2+24v+21)$ for the $[v,0,1]$, $v\in\mathbb{N}_0$, representations of SO(7), and eliminate K from (3.11) and (3.12), we get

$$(2s+1)(2s+3)u(2u+3)|A(s-\tfrac{1}{2},u-1)|^2-2s(2s+2)(u+1)(2u+1)|A(s,u)|^2-(2s+1)(2s+3)(2u+2)|C(s-\tfrac{1}{2},u+1)|^2$$
$$=-s(2s+1)(2s+2)(2s+3)(2u+1)(2u+3)[\tfrac{1}{8}(4v^2+24v+21)-2s^2-2s+\tfrac{3}{4}-2su-\tfrac{3}{2}u(u+2)]. \tag{3.13}$$

Clearly another recursion relation is required to obtain $A(s,u)$ and $C(s,u)$. It is obtained by performing the transformation $(s,\lambda,t,\mu)\to(t,\mu,s,\lambda)$ on Eq. (I4.8). We can then consider the matrix elements between the states $\langle s\,t\,u-2\,\lambda\,\mu\,\nu|$ and $|stu\lambda\mu\nu\rangle$ and apply the relations (I5.1) – (I5.5) between the matrix elements of the shift operators and the reduced matrix elements. After multiplication of the equation with its complex conjugate, taking into account property (3.1), we obtain

$$[2s(2s+2)]^2|A(s,u)|^2|C(s,u+2)|^2$$
$$=[(2s+1)(2s+3)]^2|A(s-\tfrac{1}{2},u+1)|^2|C(s-\tfrac{1}{2},u+1)|^2. \tag{3.14}$$

In manipulating recursion relations between reduced matrix elements, it is often convenient to use a graphic representation. We plot all $[SU(2)]^3$ representations (s,t,u) allowed by the branching rule as points in a three-dimensional space. In the case of the $[v,0,1]$, $v\in\mathbb{N}_0$, SO(7) representations, the points (s,t,u) lie in the planes $t=s+\frac{1}{2}$ and $t=s-\frac{1}{2}$. We represent a reduced matrix element by an arrow pointing from the point (s,t,u) towards the point (s',t',u') if (s,t,u), respectively, (s',t',u') are the labels appearing in the right-, respectively, left-hand side of the reduced matrix element, written in one of the forms (3.2) or their complex conjugates.

Figure 1 shows the projection of the points (s,t,u) in the plane $t=s+\frac{1}{2}$ on the plane $t=0$. The SO(7) representation considered is [9,0,0]. Odd-numbered arrows represent the reduced matrix elements $C(s,u)$, even-numbered arrows $A(s,u)$. It is easily seen that Eq. (3.13) gives a relation between the reduced matrix elements represented by arrows numbered by $2n$, $2n+1$, and $2n+2$, $n\in\{0,1,...,[v/2]\}$. If the number of an arrow is not contained in the set $\{1,2,...,v\}$, the corresponding reduced matrix element is zero. This is a consequence of the definitions (3.2) and the branching rule. Equation (3.14) gives a relation between the reduced matrix elements represented by arrows numbered by $2n-1$, $2n$, and $2n+1$, $n\in\{1,2,...,[v/2]\}$.

Figure 1 illustrates that in order to solve the recursion relations (3.13) and (3.14) all we need is the value of the reduced matrix elements $C(\frac{1}{2}(v-u+\frac{1}{2}),u)$ represented by the arrows numbered 1. To calculate them we can make use of

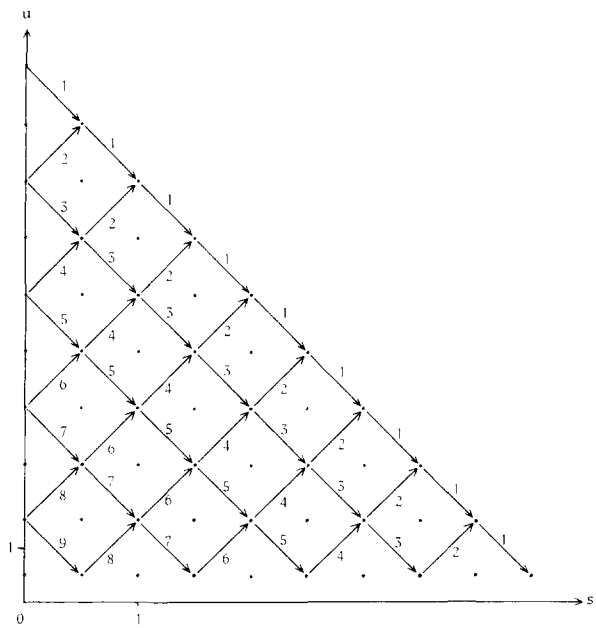


FIG. 1. Graph of the points (s,u), where $(s,s+\frac{1}{2},u)$ are $[SU(2)]^3$ representations appearing in the reduction of the SO(7) representation [9,0,0]. The reduced matrix element $\langle s's'+\frac{1}{2}\,u'\|T^{[1/2\ 1/2\ 1]}\|s\,s+\frac{1}{2}\,u\rangle$ is represented by an arrow pointing from the point (s,u) to (s',u') if $\frac{19}{2}-2s-u$ is even.

the fact that if $t = s + \frac{1}{2}$ and s has its maximum value, i.e., $s = s_M = \frac{1}{2}(v - u + \frac{1}{2})$, not only $G = H = I = 0$, but also $A = B = D = 0$. This allows us to obtain J at once from Eq. (I4.11):

$$J = -\tfrac{1}{2}(u+1)(2u+1)(2u+3)(v-u+\tfrac{1}{2}) \times (v-u+\tfrac{3}{2})(v-u+\tfrac{5}{2}), \tag{3.15}$$

and, with the use of (2.16)

$$|C(s_M - \tfrac{1}{2}, u+1)|^2 = \tfrac{1}{4}(2u+1)(2u+3)(v-u+\tfrac{1}{2})(v-u+\tfrac{3}{2})(v-u+\tfrac{5}{2}). \tag{3.16}$$

Now $|A(s_M - \frac{1}{2}, u-1)|^2$ can be calculated from (3.13), because $A(s_M, u)^2 = 0$. Then $|C(s_M - \frac{1}{2}, u-1)|^2$ can be calculated from (3.14). So, by using alternately Eqs. (3.13) and (3.14), all values of $|A(s,u)|^2$ and $|C(s,u)|^2$ are obtained, with one restriction, however. It can be easily seen from Fig. 1 that only those reduced matrix elements are obtained for which $v - 2s - u + \frac{1}{2}$ is even. In this case, we will use the subscript E, else we will use the subscript $\mathcal{O}$. From now on in this paragraph no subscript is used if $v - 2s - u + \frac{1}{2}$ can be either even or odd.

Consideration of a set of values for $|A_E(s,u)|^2$ and $|C_E(s,u)|^2$ obtained from the recursion relations leads us to propose

$$|A_E(s,u)|^2 = \tfrac{1}{16}(2s+1)(2s+3)[(2u+1)(2u+3)/(u+1)] \times (v+2s+u+\tfrac{13}{2})(v-2s-u+\tfrac{1}{2}), \tag{3.17}$$

$$|C_E(s,u)|^2 = \tfrac{1}{16}(2s+1)(2s+3)[(2u-1)(2u+1)/u] \times (v+2s-u+\tfrac{9}{2})(v-2s+u+\tfrac{1}{2}), \tag{3.18}$$

which are indeed solutions of Eqs. (3.13), (3.14) and satisfy the condition (3.16).

As in paragraph 2, we can use the fact that, as a consequence of (I3.11), A can be obtained by the formal substitution $u \to -u-1$ in C and vice versa. This substitution changes the parity of $v - 2s - u + \frac{1}{2}$ so that we can conclude with Eqs. (3.3) and (3.4) that, formally,

$$|A_{\mathcal{O}}(s,u)|^2 = -|C_E(s,-u-1)|^2, \tag{3.19}$$

$$|C_{\mathcal{O}}(s,u)|^2 = -|A_E(s,-u-1)|^2. \tag{3.20}$$

In this way we obtain at once from Eqs. (3.17) and (3.18) that

$$|A_{\mathcal{O}}(s,u)|^2 = \tfrac{1}{16}(2s+1)(2s+3)[(2u+1)(2u+3)/(u+1)] \times (v+2s+u+\tfrac{11}{2})(v-2s-u-\tfrac{1}{2}), \tag{3.21}$$

$$|C_{\mathcal{O}}(s,u)|^2 = \tfrac{1}{16}(2s+1)(2s+3)[(2u-1)(2u+1)/u] \times (v+2s-u+\tfrac{11}{2})(v-2s-u+\tfrac{3}{2}). \tag{3.22}$$

To calculate $B(s,u)^2$ we need a relation between $|A(s,u)|^2$, $|B(s,u)|^2$, and $|C(s,u)|^2$ or, equivalently, between A, B, C, J, and L. This relation we obtain by taking matrix elements of (I4.10), performing the transformation $(s,\lambda,t,\mu) \to (t,\mu,s,\lambda)$ and taking into account that $G = H = I = 0$ for $t = s + \frac{1}{2}$:

$$uA + (2u+1)B + (u+1)C = (2s+1)(2s+3)u(u+1)(2u+1) \times [-\tfrac{1}{8}(4v^2+24v+21) + s(s+1) + (s+\tfrac{1}{2})(s+\tfrac{9}{2}) + \tfrac{1}{2}u(u+1)],$$

or

$$|A(s,u)|^2 + |B(s,u)|^2 + |C(s,u)|^2 = -\tfrac{1}{2}(2s+1)(2s+3)(2u+1) \times [-\tfrac{1}{8}(4v^2+24v+21) + s(s+1) + (s+\tfrac{1}{2})(s+\tfrac{9}{2}) + \tfrac{1}{2}u(u+1)]. \tag{3.23}$$

We can now calculate $|B_E(s,u)|^2$ by substituting the results (3.17) and (3.18) in this equation

$$|B_E(s,u)|^2 = \tfrac{1}{16}(2s+1)(2s+3)\frac{2u+1}{u(u+1)} \times (v+2s-u+\tfrac{9}{2})(v-2s-u+\tfrac{1}{2}). \tag{3.24}$$

As a result of (I3.11), B is invariant for the transformation $u \to -u-1$. It then follows from Eq. (3.4) that, formally,

$$|B_{\mathcal{O}}(s,u)|^2 = -|B_E(s,-u-1)|^2. \tag{3.25}$$

From Eqs. (3.24) and (3.25) we can then deduce

$$|B_{\mathcal{O}}(s,u)|^2 = \tfrac{1}{16}(2s+1)(2s+3)[(2u+1)/u(u+1)] \times (v+2s+u+\tfrac{11}{2})(v-2s+u+\tfrac{3}{2}). \tag{3.26}$$

It is easy now to calculate the values of $|D(s,u)|^2$ and $|E(s,u)|^2$ by considering an appropriate scalar-type equation of I, taking matrix elements, and substitution of (3.3)–(3.10). In this way, we can derive from Eq. (I4.11) that

$$\begin{aligned}&[2s+1+(2s+3)(2u+3)](2u+1)|A(s,u)|^2 \\ &+ (2s+2)2u(2u+3)|B(s,u)|^2 \\ &+ (2s+3)(2u+1)(2u+2)|D(s,u)|^2 \\ &+ (2s+3)(2u+1)(2u+2)|C(s-\tfrac{1}{2},u+1)|^2 \\ &= (2s+1)(2s+2)(2s+3)(2u+1)(u+1)(2u+3) \\ &\times [\tfrac{1}{8}(4v^2+24v+21) - 2s^2 - 4s - \tfrac{5}{4} - \tfrac{1}{2}u(u+5)].\end{aligned} \tag{3.27}$$

Substitution of the known reduced matrix elements results in

$$|D_E(s,u)|^2 = \tfrac{1}{16}[(2u+1)(2u+3)/(u+1)] \times (v-2s+u+\tfrac{5}{2})(v-2s-u+\tfrac{1}{2}), \tag{3.28}$$

$$|D_{\mathcal{O}}(s,u)|^2 = \tfrac{1}{16}[(2u+1)(2u+3)/(u+1)] \times (v+2s-u+\tfrac{7}{2})(v+2s+u+\tfrac{11}{2}). \tag{3.29}$$

In a similar way we can deduce from (I4.10) that

$$\begin{aligned}&|A(s,u)|^2 + |B(s,u)|^2 + |C(s,u)|^2 \\ &+ |D(s,u)|^2 + |E(s,u)|^2 + |D(s,u-1)|^2 \\ &= -\tfrac{1}{2}(2s+2)^2(2u+1)[-\tfrac{1}{8}(4v^2+24v+21) \\ &+ 2s^2 + 6s + \tfrac{3}{4} + \tfrac{1}{2}u(u+1)].\end{aligned} \tag{3.30}$$

This relation allows us to obtain

$$|E_E(s,u)|^2 = \frac{1}{16}\frac{2u+1}{u(u+1)}\left(v+2s+4su+3u+\frac{9}{2}\right)^2, \tag{3.31}$$

and with the formal equality $|E_{\mathscr{O}}(s,u)|^2 = -|E_E(s,u)|^2$,

$$|E_{\mathscr{O}}(s,u)|^2 = \frac{1}{16}\frac{2u+1}{u(u+1)}\left(v-2s-4su-3u+\frac{3}{2}\right)^2. \tag{3.32}$$

This concludes the calculation of the reduced matrix elements in the $[v,0,1]$, $v\in\mathbb{N}_0$ representation of SO(7).

IV. REDUCED MATRIX ELEMENTS IN THE [0,0,v], $v\in\mathbb{N}_0$, REPRESENTATION

First of all we need the SO(7)→[SU(2)]^3 branching rule for the $[0,0,v]$, $v\in\mathbb{N}_0$ representations of SO(7). It can be shown, for instance by using the method outlined in De Meyer *et al.*,[2] that if we denote the [SU(2)]^3 representations by (s,t,u), the branching rule reads

$$[0,0,v]\rightarrow\sum_{s,t,u,}(s,t,u)$$

with

$$\begin{aligned} &u=\frac{v}{2},\frac{v}{2}-1,\ldots,\tfrac{1}{2}\text{ or }0,\\ &|s-t|=u,u-1,\ldots,\tfrac{1}{2}\text{ or }0,\\ &s+t=\frac{v}{2},\frac{v}{2}-1,\ldots,u. \end{aligned} \tag{4.1}$$

A graphical illustration for $v=6$ of this branching rule is given in Fig. 2. The [SU(2)]^3 representations (s,t,u) are considered as points in the three-dimensional space. The tetrahedron with smallest content and containing all points (s,t,u) is represented as a solid body. The planes S,T,U, and V will be of interest later on. Their equations are

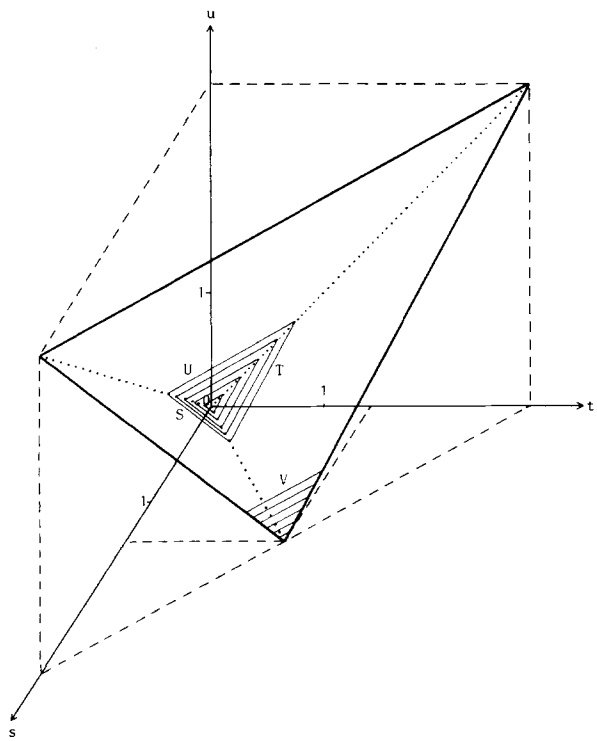


FIG. 2. Graph of the tetrahedron with smallest content and containing all points (s,t,u), where (s,t,u) are the [SU(2)]^3 representations occurring in the reduction of the SO(7) representation [0,0,6].

$$S: -s+t+u=0, \tag{4.2}$$

$$T: s-t+u=0, \tag{4.3}$$

$$U: s+t-u=0, \tag{4.4}$$

$$V: s+t-v/2=0. \tag{4.5}$$

If v is even, there exist on all edges of the tetrahedron points (s,t,u) corresponding to [SU(2)]^3 representations, if v is odd, no such points lie on the edge $s-t=u=0$, but this is not relevant to our reasoning for calculating the reduced matrix elements.

We define the following reduced matrix elements:

$$\begin{aligned} A(s,t,u) &= \langle s+\tfrac{1}{2}\,t+\tfrac{1}{2}\,u+1\|T^{[1/2\ 1/2\ 1]}\|stu\rangle,\\ B(s,t,u) &= \langle s+\tfrac{1}{2}\,t+\tfrac{1}{2}\,u\ \ \|T^{[1/2\ 1/2\ 1]}\|stu\rangle,\\ C(s,t,u) &= \langle s+\tfrac{1}{2}\,t+\tfrac{1}{2}\,u-1\|T^{[1/2\ 1/2\ 1]}\|stu\rangle,\\ D(s,t,u) &= \langle s+\tfrac{1}{2}\,t-\tfrac{1}{2}\,u+1\|T^{[1/2\ 1/2\ 1]}\|stu\rangle,\\ E(s,t,u) &= \langle s+\tfrac{1}{2}\,t-\tfrac{1}{2}\,u\ \ \|T^{[1/2\ 1/2\ 1]}\|stu\rangle,\\ F(s,t,u) &= \langle s+\tfrac{1}{2}\,t-\tfrac{1}{2}\,u-1\|T^{[1/2\ 1/2\ 1]}\|stu\rangle. \end{aligned} \tag{4.6}$$

The labels s, t, and u are subject to the conditions (4.1). All other reduced matrix elements can be related to those defined in (4.6) by means of the property (3.1).

To calculate the reduced matrix elements (4.6) we proceed in the following way: first $E(s,t,u)$ and $F(s,t,u)$ are calculated in the plane V (i.e., for [SU(2)]^3 representations (s,t,u), where s, t, and u satisfy Eq. (4.5) of the plane V). Then $E(s,t,u)$ is calculated in the plane U and $F(s,t,u)$ in the plane T. Combining these results and taking into account properties such as (I3.11) and (I3.12) will yield all reduced matrix elements.

The following results can be established with the aid of Eqs. (I5.1)–(I5.5):

$$D=(2u+2)|D(s,t,u)|^2, \tag{4.7}$$

$$E=[2u(u+1)/(2u+1)]\,|E(s,t,u)|^2, \tag{4.8}$$

$$F=(2u+2)|F(s,t,u)|^2, \tag{4.9}$$

$$H=[2u(u+1)/(2u+1)]\,|E(s-\tfrac{1}{2},t+\tfrac{1}{2},u)|^2, \tag{4.10}$$

$$L=-2u|A(s-\tfrac{1}{2},t-\tfrac{1}{2},u-1)|^2. \tag{4.11}$$

We first calculate $E(s,t,u)$ on the line $s+t=v/2=u$ in the plane V. On that line $A=B=C=D=G=J=K=0$. We can transform Eq. (I4.11) by performing the substitution $t\rightarrow -t-1$ and taking matrix elements, into an equation between A,D,E,G,J, and K. From this equation E can be solved, and we obtain, taking into account (4.8) and the expectation value (I2.8) of I_2, which equals $-\frac{1}{8}v(v+6)$ for the $[0,0,v]$, $v\in\mathbb{N}_0$ representations of SO(7):

$$\begin{aligned} &|E(s,t,u)|^2\\ &\quad=\tfrac{1}{4}(s+1)(2s+1)t\,(2t+1)[(2u+1)/u(u+1)]\\ &\qquad\times(v+2)^2,\ s+t=v/2=u. \end{aligned} \tag{4.12}$$

With the knowledge of this special value for $E(s,t,u)$, we can calculate $D(s,t,u)$, $E(s,t,u)$, and $F(s,t,u)$ in the plane V in much the same way as was done for $A(s,u)$, $B(s,u)$, and $C(s,u)$ in the previous paragraph.

In the plane V, $A = B = C = 0$, so that we can deduce from Eq. (I4.10), with the aid of (4.7) and (4.9) that

$$|D(s,t,u)|^2 + |E(s,t,u)|^2 + |F(s,t,u)|^2 = (s+1)(2t+1)(2u+1) \times [\tfrac{1}{8}v^2 + \tfrac{7}{4}v - 2s^2 - 3s + vs - \tfrac{1}{2}u(u+1)], \quad (4.13)$$

$s + t = v/2$.

This is essentially an equation between $E(s,t,u)$, $F(s,t,u)$, and $F(t-\frac{1}{2}, s+\frac{1}{2}, u+1)$ because

$$|D(s,t,u)|^2 = |F(t-\tfrac{1}{2}, s+\tfrac{1}{2}, u+1)|^2, \quad (4.14)$$

as can be seen from the definitions (4.6) by interchanging s and t and applying the property (3.1).

We now need a relation between $E(s,t,u)$ and an $E(s',t',u-1)$. Why we need precisely this one can be seen by drawing a sketch similar to Fig. 1, but for the plane V. It follows that we have to consider a relation between products of shift operators of nonscalar type. We take matrix elements of both sides of Eq. (I4.2), after performing the substitution $s \to -s-1$. By multiplying this equation with its complex conjugate and taking into account (I5.1)–(I5.5) we obtain

$$\frac{u-1}{2u-1}\left|E\left(s+\frac{1}{2}, t-\frac{1}{2}, u-1\right)\right|^2 |F(s,t,u)|^2 = \frac{u+1}{2u+1}\left|F\left(s+\frac{1}{2}, t-\frac{1}{2}, u\right)\right|^2 |E(s,t,u)|^2. \quad (4.15)$$

Equations (4.13)–(4.15) together with the boundary value (4.12) are sufficient to calculate recursively all values of $E(s,t,u)$ and $F(s,t,u)$ in the plane V. Consideration of these values leads to the following general solutions in the plane V:

$$|E(s,t,u)|^2 = \tfrac{1}{2}(s+1)(2t+1)[(2u+1)/u(u+1)] \times (s-t+u+1)(-s+t+u)(v+2)^2, \quad s+t = v/2, \quad (4.16)$$

$$|F(s,t,u)|^2 = \tfrac{1}{2}(s+1)(2t+1)(-s+t+u-1) \times (-s+t+u)(1/u) \times (v/2-u+1)(v/2+u+1), \quad (4.17)$$

$s + t = v/2$.

To calculate $E(s,t,u)$ in the plane U, we remark that in this plane $D = G = J = K = 0$. After performing the substitutions $s \to -s-1$ and $t \to -t-1$ in Eq. (I4.10) we deduce from it the following relation on the plane U:

$$(u+1)L - (2u+1)H - (u+1)I = s(2t+1)2u(u+1)(2u+1) \times [-\tfrac{3}{8}v(v+6) + (s+1)(s-3) + t(t+1) + \tfrac{1}{2}u(u+1)], \quad (4.18)$$

$s + t = u$.

As H and L are related to $E(s,t,u)$ and $A(s,t,u)$ by (4.10) and (4.11), we need another relation between L, H, and I in order to eliminate I. After performing the substitutions $s \to -s-1$ and $t \to -t-1$ we deduce from (I4.12) that

$$\begin{aligned}&[2(s+1)(u+2) + 2t(u+1) - 1](u+1)(2u+1)L \\ &+ [2s - 2t(2u^2+u-2) + 4st(2u+3) \\ &- u(2u+3)](2u+1)H + [(2s+1)(2t+1) \\ &+ 2(2s+1)t(u+2)(2u+1) \\ &- (2t+1)(u+1)(2u+1)](u+1)I \\ &+ 2s(2t+1)u(u+1)(2u+1)\{[2(s-t) \\ &- (2t+1)u(2u+3)][-\tfrac{3}{8}v(v+6) \\ &+ s(s+1) + t(t+1) + \tfrac{1}{2}u(u+1)] \\ &- (2s+1)(s-t)2tu(2u+5) \\ &- 2(s+1)tu(2u-1)(2u+3) + 2tu(u+1)(2u+3) \\ &+ 3(s+1)u(2u+3) - 6(s-t)(2st+s+t+1)\} = 0,\end{aligned} \quad (4.19)$$

$s + t = u$.

We can now eliminate I from (4.18) and (4.19). Taking into account (4.10) and (4.11) we obtain

$$2t(u+1)^2|A(s,t-1,u-1)|^2 + (2t-1)u(u+1)|E(s,t,u)|^2 + (2s+1)t(2t-1)(u+1)^2(2u+1) \times [-\tfrac{3}{8}v(v+6) + 2s^2 + s - 2su + \tfrac{3}{2}u^2 + \tfrac{1}{2}u - 3] = 0, \quad (4.20)$$

$s + t = u$.

Again we need an equation between products of shift operators of nonscalar type. As can be inferred from a sketch of the plane U, we need an equation between products with a total shift (1,0,1) or $(-1,0,-1)$. We therefore consider (I4.5). To avoid the occurrence of as yet unknown matrix elements, we perform the substitution $t \to -t-1$. In a way analogous to derivation of Eq. (4.15), we obtain

$$(2t)^2(u+1)^2\frac{u-1}{2u-1} \times \left|E\left(s-\frac{1}{2}, t-\frac{1}{2}, u-1\right)\right|^2 |A(s,t-1,u-1)|^2 = (2t-1)^2u^2\frac{u+1}{2u+1} \times \left|A\left(s-\frac{1}{2}, t-\frac{1}{2}, u-1\right)\right|^2 |E(s,t,u)|^2, \quad (4.21)$$

$s + t = u$.

We can now calculate recursively $E(s,t,u)$ and $A(s,t,u)$ from (4.20) and (4.21). Indeed, a boundary value of $E(s,t,u)$ is known from (4.12). Only the result for $E(s,t,u)$ is of interest here. We obtain

$$|E(s,t,u)|^2 = \frac{1}{8}(2s+1)2t\frac{(2u+1)(2u+2)}{u} \times \left(\frac{v}{2}+s-t+2\right)\left(\frac{v}{2}-s+t+1\right), \quad (4.22)$$

$s + t = u$.

We now derive the general result for $E(s,t,u)$ from the values of $E(s,t,u)$ on the planes V and U. From the definition (4.6) of $E(s,t,u)$ and the branching rule (4.1) it is clear that $|E(s,t,u)|^2$ is proportional to

$$(s-t+u+1)^a(-s+t+u)^b, \quad a,b \in \mathbb{R}^+. \quad (4.23)$$

Comparison with the partial results (4.16) and (4.22) leads to $a = b = 1$. As a result we have to look for an expression $\epsilon(s,t,u)$ that satisfies

$$\epsilon(s,t,u) = (2s+2)(2t+1)\frac{2u+1}{u(u+1)}(v+2)^2, \quad s+t=\frac{v}{2},$$

$$\epsilon(s,t,u) = \left(\frac{v}{2}+s-t+2\right)\left(\frac{v}{2}-s+t+1\right) \times \frac{(2u+1)(2u+2)}{u}, \quad s+t=u. \tag{4.24}$$

A careful consideration of this problem learns that the only solution is

$$\epsilon(s,t,u) = 2\frac{2u+1}{u(u+1)}\left(\frac{v}{2}+s-t+2\right) \times \left(\frac{v}{2}-s+t+1\right)(s+t+1)^2. \tag{4.25}$$

Thus, combining (4.23) and (4.25), we obtain

$$|E(s,t,u)|^2 = \frac{1}{4}(s+t+1)^2(s-t+u+1)(-s+t+u)\frac{2u+1}{u(u+1)} \times \left(\frac{v}{2}+s-t+2\right)\left(\frac{v}{2}-s+t+u\right). \tag{4.26}$$

If $u = 0$, this expression is indefinite, because in that case $s = t$, as a result of the branching rule (4.1). However, it follows from the definition (4.6) of $E(s,t,u)$ that $|E(s,s,0)|^2 = 0$.

As a result of (I3.12), B is obtained by making the formal substitution $t \to -t-1$ in E. In this way we obtain from (4.26) that

$$|B(s,t,u)|^2 = \frac{1}{4}(s-t)^2(s+t+u+2)(s+t-u+1)\frac{2u+1}{u(u+1)} \times \left(\frac{v}{2}+s+t+3\right)\left(\frac{v}{2}-s-t\right). \tag{4.27}$$

If $u = 0$, then $s = t$ and $\lim_{u\to 0}|B(s,t,u)|^2 = 0$. We therefore expect $|B(s,s,0)|^2 = 0$. This follows indeed from Eq. (4.27) after the substitution $u \to -u-1$ and taking into account that if $u = 0$, $C = E = F = H = I = L = 0$.

To obtain the general expression for $F(s,t,u)$, we first calculate $F(s,t,u)$ in the plane T. To that end, we perform the transformations $s \to -t-1$, $t \to -s-1$, and $u \to -u-1$ in Eq. (I4.11). This allows us to deduce from it

$$(2s+1)(u+1)(2u-1)|B(s-\tfrac{1}{2},t-\tfrac{1}{2},u)|^2 + 2su(u+1)|F(s,t,u)|^2 + s(2s+1)(2t+1)u(2u-1)(2u+1) \times \left[-\tfrac{3}{8}v(v+6) + 2s^2 + s - 3 + 2su + \tfrac{3}{2}u^2 - \tfrac{1}{2}u\right] = 0, \quad s-t+u=0. \tag{4.28}$$

Because $B(s,t,u)$ is already known, we obtain

$$|F(s,t,u)|^2 = \tfrac{1}{2}(2s+1)(2t+1)(2u-1)\left(\frac{v}{2}+s-t+2\right) \times \left(\frac{v}{2}-s+t+1\right), \quad s-t+u=0. \tag{4.29}$$

In a way similar to that of the derivation of the general result (4.26) for $E(s,t,u)$, we can deduce from (4.17) and (4.29) that

$$|F(s,t,u)|^2 = \tfrac{1}{4}(s+t+u+1)(s+t-u+1) \times(-s+t+u-1)(-s+t+u) \times\frac{1}{u}\left(\frac{v}{2}+s-t+2\right)\left(\frac{v}{2}-s+t+1\right). \tag{4.30}$$

From (4.14) we obtain at once

$$|D(s,t,u)|^2 = \tfrac{1}{4}(s+t+u+2)(s+t-u) \times(s-t+u+2)(s-t+u+1) \times\frac{1}{u+1}\left(\frac{v}{2}+s-t+2\right)\left(\frac{v}{2}-s+t+1\right). \tag{4.31}$$

To calculate $A(s,t,u)$ we use the property that A can be obtained from D by making the formal substitution $t \to -t-1$. As a result we have the formal equality $|A(s,t,u)|^2 = -|D(s,-t-1,u)|^2$. Consequently,

$$|A(s,t,u)|^2 = \tfrac{1}{4}(s+t+u+2)(s+t+u+3) \times(s-t+u+1)(-s+t+u+1) \times\frac{1}{u+1}\left(\frac{v}{2}+s+t+3\right)\left(\frac{v}{2}-s-t\right). \tag{4.32}$$

In an analogous way the formal equality $|C(s,t,u)|^2 = -|A(s,t,-u-1)|^2$ leads to

$$|C(s,t,u)|^2 = \tfrac{1}{4}(s+t-u+2)(s+t-u+1) \times(s-t+u)(-s+t+u) \times\frac{1}{u}\left(\frac{v}{2}+s+t+3\right)\left(\frac{v}{2}-s-t\right). \tag{4.33}$$

As a result of the definitions of $F(s,t,u)$ and $C(s,t,u)$, we find $|F(s,s,0)|^2 = |C(s,s,0)|^2 = 0$.

V. CONCLUSION

We have calculated all matrix elements of the bispinorvector $T^{[1/2\ 1/2\ 1]}$ between $[SU(2)]^3$ states which constitute a basis for the respective classes of SO(7) representations $[v,0,0]$, $v \in \mathbb{N}_0$, $[v,0,1]$, $v \in \mathbb{N}_0$ and $[v,0,0]$, $v \in \mathbb{N}_0$. Indeed, due to the property (3.1) and the Wigner–Eckart theorem, all matrix elements of $T^{[1/2\ 1/2\ 1]}$ between states that constitute a basis for the SO(7) representations $[v,0,0]$, $v \in \mathbb{N}_0$ are known if (2.5) and (2.7) are given. For the $[v,0,1]$, $v \in \mathbb{N}_0$ representations, it is sufficient to know the results (3.17), (3.18), (3.21), (3.22), (3.24), (3.26), (3.28), (3.29), (3.31), and (3.32). For the $[0,0,v]$, $v \in \mathbb{N}_0$ representations, all matrix elements of $T^{[1/2\ 1/2\ 1]}$ are known once we are given (4.26), (4.27), (4.30)–(4.33).

Finally we want to mention some additional symmetry

properties of the reduced matrix elements. Equation (5.4) of Vanden Berghe *et al.*[1] can be generalized for arbitrary SO(7) representations that reduce without degeneracy into $[SU(2)]^3$ representations. We obtain, denoting $\langle s't'u'\|T^{[1/2\ 1/2\ 1]}\|stu\rangle$ by $\langle s't'u'\|stu\rangle$,

$$\begin{aligned}
&\langle I_2\rangle \\
&= -[(2s+1)(2t+1)(2u+1)]^{-1} \\
&\quad\times[|\langle s+\tfrac{1}{2}\,t+\tfrac{1}{2}\,u+1\|stu\rangle|^2+|\langle s+\tfrac{1}{2}\,t+\tfrac{1}{2}\,u\|stu\rangle|^2 \\
&\quad+|\langle s+\tfrac{1}{2}\,t+\tfrac{1}{2}\,u-1\|stu\rangle|^2+|\langle s+\tfrac{1}{2}\,t-\tfrac{1}{2}\,u+1\|stu\rangle|^2 \\
&\quad+|\langle s+\tfrac{1}{2}\,t-\tfrac{1}{2}\,u\|stu\rangle|^2+|\langle s+\tfrac{1}{2}\,t-\tfrac{1}{2}\,u-1\|stu\rangle|^2 \\
&\quad+|\langle s-\tfrac{1}{2}\,t+\tfrac{1}{2}\,u+1\|stu\rangle|^2+|\langle s-\tfrac{1}{2}\,t+\tfrac{1}{2}\,u\|stu\rangle|^2 \\
&\quad+|\langle s-\tfrac{1}{2}\,t+\tfrac{1}{2}\,u-1\|stu\rangle|^2+|\langle s-\tfrac{1}{2}\,t-\tfrac{1}{2}\,u+1\|stu\rangle|^2 \\
&\quad+|\langle s-\tfrac{1}{2}\,t-\tfrac{1}{2}\,u\|stu\rangle|^2+|\langle s-\tfrac{1}{2}\,t-\tfrac{1}{2}\,u-1\|stu\rangle|^2 \\
&\quad-s(s+1)-t(t+1)-\tfrac{1}{2}u(u+1)]. \qquad (5.1)
\end{aligned}$$

This equation points out that invariances in $\langle I_2\rangle$ must somehow be manifest in the matrix elements themselves.

Indeed, for the $[v,0,0]$, $v\in\mathbb{N}_0$ representations of SO(7), $\langle I_2\rangle = -\frac{1}{2}v(v+5)$ is invariant under the transformation $v\to -v-5$, as are the reduced matrix elements (2.5) and (2.7). For the $[0,0,v]$, $v\in\mathbb{N}_0$, representations, $\langle I_2\rangle = -\frac{3}{8}v(v+6)$ is invariant under the transformation $v\to -v-6$, and so are the reduced matrix elements (4.26), (4.27), (4.30)–(4.33).

For the $[v,0,1]$, $v\in\mathbb{N}_0$, representations, $\langle I_2\rangle = -\frac{1}{8}(4v^2+24v+21)$ is again invariant under the transformation $v\to -v-6$, but the reduced matrix elements don't exhibit this property.

From the expressions (3.17), (3.18), (3.21), (3.22), (3.24), (3.26), (3.28), (3.29), (3.31), and (3.32) it is apparent, however, that reduced matrix elements with index E are transformed under the transformation $v\to -v-6$ into reduced matrix elements with index $\mathcal{O}$, and vice versa. For example

$$|A_E(s,u)|^2 \xrightarrow[v\to -v-6]{} |A_{\mathcal{O}}(s,u)|^2.$$

If we make the choice $t = s+\frac{1}{2}$, Eq. (5.1) can be written for $[v,0,1]$, $v\in\mathbb{N}_0$, representations as

$$\begin{aligned}
\tfrac{1}{8}(4v^2+24v+21) &= [(2s+1)(2s+2)(2u+1)]^{-1} \\
&\quad\times[|A(s,u)|^2+|B(s,u)|^2+|C(s,u)|^2 \\
&\quad+|A(s-\tfrac{1}{2},u-1)|^2+|B(s-\tfrac{1}{2},u)|^2 \\
&\quad+|C(s-\tfrac{1}{2},u+1)|^2+|D(s,u)|^2 \\
&\quad+|E(s,u)|^2+|D(s,u-1)|^2] \\
&\quad-2s^2-3s-\tfrac{3}{4}-\tfrac{1}{2}u(u+1). \qquad (5.2)
\end{aligned}$$

We now consider this equation for $v-2s-u+\frac{1}{2}$ even and perform the transformation $v\to -v-6$. As a result of the above-mentioned transformation properties of the individual matrix elements, we obtain the same expression in the right-hand side, but for $v-2s-u+\frac{1}{2}$ odd. By this argument the right-hand side of (5.2) is invariant under the transformation $v\to -v-6$ if it is the same function of s and u for $v-2s-u+\frac{1}{2}$ even or odd. This is clear at once, because (5.2) is valid as well for $v-2s-u+\frac{1}{2}$ even or odd, and the left-hand side of (5.2) is independent of s and u.

ACKNOWLEDGMENTS

We gratefully acknowledge G. Vanden Berghe and H. De Meyer (Rijksuniversiteit-Gent) for many profitable conversations, as well as Professor C. Grosjean for his continuous interest.

[1] G. Vanden Berghe, H. De Meyer, and P. De Wilde, J. Phys. A: Math. Gen. **15**, 2677 (1982).
[2] H. De Meyer, P. De Wilde, and G. Vanden Berghe, J. Phys. A: Math. Gen. **15**, 2665 (1982).
[3] J. Van der Jeugt and P. De Wilde, J. Math. Phys. **25**, 2953 (1984).
[4] P. De Wilde, H. De Meyer, J. Van der Jeugt, and G. Vanden Berghe, J. Math. Phys. **25**, 1195 (1984).

Formal relations between classical superalgebras and fermion–boson creation and annihilation operators

Dun-Sang Tang
Physics Department, University of Texas, Austin, Texas 78712

(Received 29 December 1983; accepted for publication 24 February 1984)

Classical Lie superalgebras $A(m-1, n-1)$, $B(m, n)$, $C(n)$, and $D(m, n)$ are realized by sets of fermionic and bosonic creation and annihilation operators. Some aspects of the physical representations are discussed.

PACS numbers: 03.65.Fd

I. INTRODUCTION

It has already been shown that supersymmetry is potentially relevant in elementary particle physics and gravity.[1,2] It has also been demonstrated to a certain extent that supersymmetry plays some roles in the determination of the energy spectra of odd and even nuclei.[3,4] The idea of supersymmetry may even find its applicability to a broader context, namely, general many-body (fermion–boson) problems. The questions are what the relations of second quantization and supersymmetry are and how to formulate the many-body problems in terms of Lie superalgebras. In this paper we will answer the former.

The setup of the paper is as follows. In Sec. II, we will review the basic ideas of the classical Lie superalgebras. Some defining relations and definitions which are used in the later sections are introduced. Section III is a study of the type 1 Lie superalgebra $A(m, n)$ realized by the fermion and boson creation and annihilation operators. Section IV is about the type 2 Lie superalgebra $D(m, n)$ and Sec. V is on type 2 Lie superalgebra $B(m, n)$ in relation to the second quantization of fermions and bosons. Section VI contains discussions on the physical representations and in Sec. VII are some concluding remarks. The Appendix is a discussion of the type 1 Lie superalgebra $C(n)$ for completeness.

II. CLASSICAL LIE SUPERALGEBRAS

Our discussion will follow the articles of Kac[5] and Hurni and Morel.[6] We will discuss only $A(m, n)$, $B(m, n)$, and $D(m, n)$ which are found to be closely related to second quantization.

A Lie superalgebra L is a $Z(2)(=\{0, 1\})$ graded linear vector space which is a direct sum of an even L_0 and an odd L_1 vector space,

$$L = L_0 + L_1. \tag{1}$$

(Henceforth, we will call L_0 the even part and L_1 the odd part.) Furthermore, all elements X, Y, and Z in L satisfy the following axioms of the bracket operation $[\,,\,]: L \times L \to L$:

$$(1)[X, Y] = -(-1)^{g(X)g(Y)}[Y, X]; \tag{2}$$

(2) graded Jacobi identity,

$$[X, [Y, Z]] + (-1)^{g(X)(g(Y)+g(Z))}[Y, [Z, X]] + (-1)^{g(Z)(g(X)+g(Y))}[Z, [X, Y]] = 0. \tag{3}$$

Here, $g(X)$ is the grading (or degree) of X. If $X \in L_0$, then $g(X) = 0$. If $x \in L_1$, then $g(X) = 1$. The grading of a product of elements is the sum of the gradings of all elements in the product modulo 2[i.e., $Z(2)$ graded]. Equation (1) says that the bracket is an anticommutatior if X and Y have odd gradings; otherwise the bracket is a commutator. Moreover, together with the second axiom, L_0 forms an ordinary Lie algebra and L_1 behaves as a representation of L_0.

(1) $A(m-1, n-1) = \mathrm{su}(m/n)$ is a type 1 Lie superalgebra defined by

$$A(m-1, n-1) = A_{-1}(m-1, n-1) + A_0(m-1, n-1) + A_1(m-1, n-1), \tag{4}$$

in which the ordinary Lie algebra $A_0(m-1, n-1)$ is reducible and is given by

$$A_0(m-1, n-1) = A(m-1) + A(n-1) + K, \tag{5}$$

where $A(m-1) = \mathrm{su}(m)$, $A(n-1) = \mathrm{su}(n)$, and K is a real number which corresponds to the abelian group U(1). This $A_0(m-1, n-1)$ forms the even part of $A(m-1, n-1)$.

The odd part is composed of two pieces: $A_{-1}(m-1, n-1)$ corresponds to the irreducible representation $(m, \bar{n})$ of $\mathrm{su}(m) + \mathrm{su}(n)$ and $A_1(m-1, n-1)$ corresponds to $(\bar{m}, n)$. The dimension of $A(m-1, n-1)$ is $(m+n)^2 - 1 - \delta_{m,n}$.

Let H be the Cartan subalgebra of L_0. If $\alpha(h)$ is not zero in the relation

$$[h, X] = \alpha(h)X, \tag{6}$$

for $h \in H$ and $X \in L$, then α is called a root of X. It is an even root if $X \in L_0$ and an odd root if $X \in L_1$. A collection of all roots is denoted by Δ. The set of all even (odd) roots is denoted by $\Delta_0(\Delta_1)$.

The root system of $A(m-1, n-1)$ is given by

$$\Delta_0 = \{\epsilon_i - \epsilon_j; \delta_i - \delta_j\}, \tag{7}$$

$$\Delta_1 = \{\pm(\epsilon_i - \delta_j), \quad i \neq j, \quad 1 \leqslant i \leqslant m, \quad 1 \leqslant j \leqslant n\}, \tag{8}$$

where ϵ_i and δ_j are basis vectors.

We choose a simple root system as

$$\Pi = \{\epsilon_1 - \epsilon_2, \epsilon_2 - \epsilon_3, \ldots, \epsilon_m - \delta_1, \delta_1 - \delta_2, \ldots, \delta_{n-1} - \delta_n\}, \tag{9}$$

in which just one odd root $\epsilon_m - \delta_1$ is included. We use the word "choose" because it has been shown that unlike the classical Lie algebra in which the simple root is unique, the simple root systems in Lie superalgebra are not unique.[5]

(2) $B(m, n) = \mathrm{osp}(2m + 1/2n)$ is a type 2 Lie superalgebra defined by

$$B(m, n) = B_0(m, n) + B_1(m, n), \tag{10}$$

 0022-2488/84/102966-08$02.50

where $B_0(m, n)$ is an ordinary reducible Lie algebra given by

$$B_0(m, n) = B(m) + C(n), \tag{11}$$

where $B(m) = \mathrm{so}(2m + 1)$ and $C(n) = \mathrm{sp}(2n)$.

Here $B_1(m, n)$ is the odd part which corresponds to the irreducible representation $(2m + 1, 2n)$ of $B(m) + C(n)$. The root system of $B(m, n)$ is given by

$$\Delta_0 = \{ \pm\epsilon_i \pm \epsilon_j;\ \pm 2\delta_i;\ \pm\epsilon_i;\ \pm\delta_i \pm \delta_j \}, \quad i \neq j, \tag{12}$$

$$\Delta_1 = \{ \pm\delta_i;\ \pm\epsilon_i \pm \delta_j \}. \tag{13}$$

We choose a system of simple roots as

$$\Pi = \{\delta_1 - \delta_2, \ldots, \delta_n - \epsilon_1, \epsilon_1 - \epsilon_2, \ldots, \epsilon_{m-1} - \epsilon_m, \epsilon_m\}, \tag{14}$$

in which an odd root $\delta_n - \epsilon_1$ is included.

(3) $D(m, n) = \mathrm{osp}(2m/2n)$ is a type 2 Lie algebra defined by

$$D(m, n) = D_0(m, n) + D_1(m, n), \tag{15}$$

where $D_0(m, n)$ is a reducible Lie algebra

$$D_0(m, n) = D(m) + C(n), \tag{16}$$

where $C(n)$ is the symplectic algebra defined as in $B(m, n)$ and $D(m) = \mathrm{so}(2m)$.

The odd part $D_1(m, n)$ transforms as the $(2m, 2n)$ irreducible representation of $D_0(m, n)$. The root system of $D(m, n)$ is given by

$$\Delta_0 = \{ \pm\epsilon_i \pm \epsilon_j;\ \pm 2\delta_i;\ \pm\delta_i \pm \delta_j \quad i \neq j\}, \tag{17}$$

$$\Delta_1 = \{ \pm\epsilon_i \pm \delta_j \}. \tag{18}$$

We choose a system of simple roots as

$$\Pi = \{\delta_1 - \delta_2, \ldots, \delta_{n-1} - \delta_n, \delta_n - \epsilon_1, \epsilon_1 - \epsilon_2, \ldots, \epsilon_{m-1} - \epsilon_m, \epsilon_{m-1} + \epsilon_m\}, \tag{19}$$

in which one odd root $\delta_n - \epsilon_1$ is included.

Our study of the connection between supersymmetry and second quantization relies heavily on the root systems in the subsequent sections. Basically, it is a construction of these Lie superalgebras by the fermionic and bosonic creation and annihilation operators, along the same line of thought as in the realization of the classical Lie algebras[7] by second quantization.

III. $A(m-1, n-1)$ AND SECOND QUANTIZATION

$A(m-1, n-1)$ has already been employed in the study of nuclei. Its structure in relation to fermionic and bosonic operators has been known for a time. We are going to write down the defining relations in the most general form in anticipation of any other possible applications in other branch of physical interests.

Let $\{a_i, a_i^+, i = 1, \ldots, m\}$ be a set of fermionic operators which satisfy the usual anticommutation relations

$$\{a_i, a_j^+\} = \delta_{i,j}, \quad \{a_i, a_j\} = 0, \quad \{a_i^+, a_j^+\} = 0. \tag{20}$$

The index i stands for the ordered complete quantum label of a state. Then the set

$$\{a_i^+ a_j, a_i^+ a_i - \tfrac{1}{2}, \quad i \neq j, \quad 1 \leqslant i, j \leqslant m\} \tag{21}$$

constitutes the classical Lie algebra $A(m-1)$ associated with the group $\mathrm{SU}(m)$. The particle number operator

$$N_a = \sum a_i^+ a_i \tag{22}$$

is a Casimir invariant. The notation Σ means to sum over all values of i. The Cartan subalgebra is

$$\{a_i^+ a_i - \tfrac{1}{2}, \quad i = 1, \ldots, m\}, \tag{23}$$

and the set of raising and lowering operators is

$$\{a_i^+ a_j, \quad i \neq j, \quad 1 \leqslant i, j \leqslant m\}. \tag{24}$$

Let $\{b_k, b_k^+, \quad k = 1, \ldots, n\}$ be a set of bosonic operators which satisfy the commutation relations

$$[b_k, b_l^+] = \delta_{kl}, \quad [b_k, b_l] = 0, \quad [b_k^+, b_l^+] = 0. \tag{25}$$

The set of bilinear terms

$$\{b_k^+ b_l, \quad k \neq l, \quad b_k^+ b_k + \tfrac{1}{2}, \quad 1 \leqslant k, l \leqslant n\} \tag{26}$$

also satisfy the commutation relations of the classical Lie algebra $A(n-1)$ associated with the group $\mathrm{SU}(n)$. The particle number operator

$$N_b = \sum b_k^+ b_k \tag{27}$$

is a Casimir invariant. The Cartan subalgebra is

$$\{b_k^+ b_k + \tfrac{1}{2}, \quad k = 1, \ldots, n\} \tag{28}$$

and the set of raising and lowering operators is

$$\{b_k^+ b_l, \quad k \neq l, \quad 1 \leqslant k, l \leqslant n\}. \tag{29}$$

Note that if we replace $a_i^+ a_i - \frac{1}{2}$ by $a_i^+ a_i$, $b_k^+ b_k + \frac{1}{2}$ by $b_k^+ b_k$, all commutators of the algebra $A(m-1) + A(n-1)$ remain unchanged. For the present treatment, we assume $m \neq n$ (see the discussion of Hurni and Morel[6] for the case $m = n$).

Now we add the set of elements

$$S(m, n) = \{a_i^+ b_k, \quad b_k^+ a_i; \quad 1 \leqslant i \leqslant m, \quad 1 \leqslant k \leqslant n\} \tag{30}$$

to $A_0(m-1, n-1)$. It is easy to verify that $A_0(m-1, n-1) + S(m, n)$ satisfy axioms (1) and (2) if $A_0(m-1, n-1)$ has even grading and $S(m, n)$ has odd grading. Therefore they form a Lie superalgebra.

To establish its association with the Lie superalgebra $A(m-1, n-1)$, we need only to show the structure constants (specifically, the Cartan matrix[8]) of $A_0(m-1, n-1) + S(m, n)$ are equal to that of $A(m-1, n-1)$. We will first establish the association of the roots with the generators. Consider the properties of the following element in the Cartan subalgebra H of $A_0(m-1, n-1)$

$$H = \sum \epsilon_i a_i^+ a_i + \sum \delta_k b_k^+ b_k, \tag{31}$$

where ϵ_i and δ_k are the standard basis of the dual of H. The commutator of H with elements in $A(m-1)$ and $A(n-1)$ show that $a_i^+ a_j (i \neq j)$ corresponds to the root $\epsilon_i - \epsilon_j$ and $b_k^+ b_l (k \neq l)$ to the root of $\delta_k - \delta_l$. Moreover, the commutator of H with elements in $S(m, n)$ show that $a_i^+ b_k$ corresponds to $\epsilon_i - \delta_k$ and the Hermitian conjugate $b_k^+ a_i$ to $\delta_k - \epsilon_i$. Therefore, $A_0(m-1, n-1) + S(m, n)$ reproduces all roots of $A(m-1, n-1)$ in Eqs. (7) and (8). Here, the even part is $A_0(m-1, n-1)$ and the odd part is $S(m, n)$.

Now let us define the following notations. Simple positive roots are denoted by

$$\alpha_i = \begin{cases} \epsilon_i - \epsilon_{i+1}, & 1 \leqslant i \leqslant m-1, \\ \epsilon_{m+1} - \delta_1, & i = m, \\ \delta_k - \delta_{k+1}, & i = m+k, \quad 1 \leqslant k \leqslant n. \end{cases} \tag{32}$$

The simple negative roots are $-\alpha_i$. The corresponding generators are

$$E_{\alpha_i} = \begin{cases} a_i^+ a_{i+1}, & 1 \leqslant i \leqslant m-1, \\ a_m^+ b_1, & i = m, \\ b_k^+ b_{k+1}, & i = m+k, \quad 1 \leqslant k \leqslant n. \end{cases} \tag{33}$$

Immediately it follows from Eq. (33) that

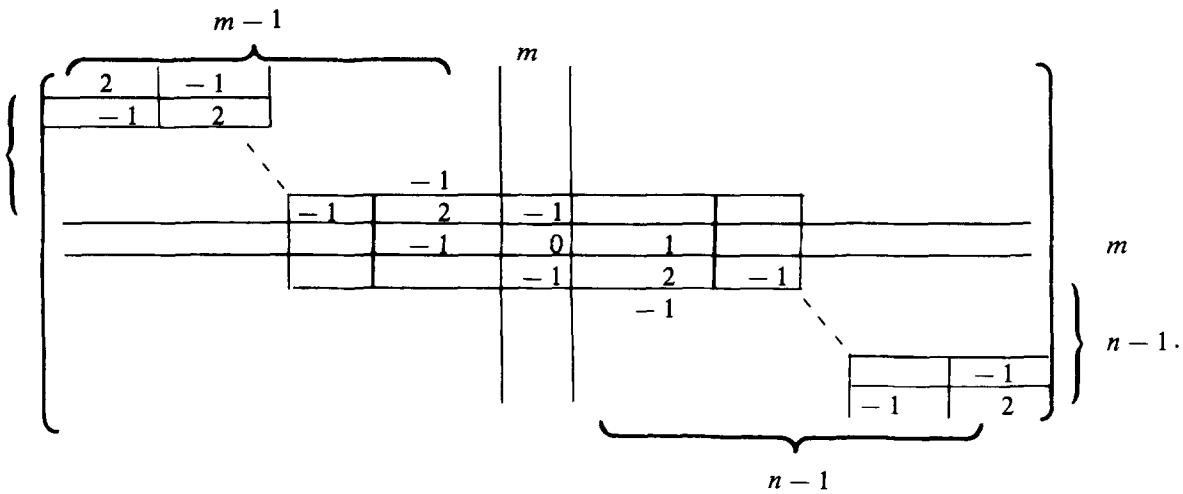
$$\begin{aligned} &[E_{\alpha_i}, E_{-\alpha_j}] = \delta_{ij} h_i, \quad i,j = 1,\ldots,m-1, \\ &\qquad\qquad m+1,\ldots,m+n-1, \\ &\{E_{\alpha_m}, E_{-\alpha_m}\} = h_m, \\ &[E_{\alpha_m}, E_{-\alpha_i}] = [E_{\alpha_i}, E_{-\alpha_m}] = 0 \quad \text{for } i \neq m, \\ &[h_i, E_{\pm\alpha_j}] = \pm a_{ij} E_{\pm\alpha_j} \quad \text{for all } i \text{ and } j, \end{aligned} \tag{34}$$

where

$$h_i = \begin{cases} a_i^+ a_i - a_{i+1}^+ a_{i+1}, & 1 \leqslant i \leqslant m-1, \\ a_m^+ a_m + b_1^+ b_1, & i = m, \\ b_k^+ b_k - b_{k+1}^+ b_{k+1}, & i = m+k, \quad 1 \leqslant k \leqslant n \end{cases}$$

are the elements in the Cartan subalgebra of the Lie superalgebra and (a_{ij}) is exactly the Cartan matrix of $A(m-1, n-1)$:

$(a_{ij}) =$



All other elements in the superalgebra $A(m-1, n-1)$ can be generated from the above generators as is customarily done in classical Lie algebras.[9] Inspection of the generators reveals the following.

(1) Total particle number (fermion and boson) is conserving in $A(m-1, n-1)$. Fermion particle number and boson particle number may vary separately because of the odd generators.

(2) The b_k^+ transform as (n) and the b_k transform as $(\bar{n})$ in $A(n-1)$. The a_j^+ transform as (m) and the a_j transform as $(\bar{m})$ in $A(m-1)$. Therefore, the $a_i^+ b_k$ transform as $(m, \bar{n})$, and the $a_j b_k^+$ as (m, n) in $A(m-1) + A(n-1)$. Hence $A_1(m-1, n-1) = \{a_i^+ b_k^+, i = 1,\ldots,m, k = 1,\ldots,n\}$,

and

$$A_{-1}(m-1, n-1) = \{a_i b_k^+, \quad i = 1,\ldots,m, \quad k = 1,\ldots,m\},$$

and

$$S(m, n) = A_1(m-1, n-1) + A_{-1}(m-1, n-1).$$

A discussion of the physical representation of $A(m-1, n-1)$ is in Sec. VI.

IV. $D(m, n)$ AND SECOND QUANTIZATION

As discussed in Sec. II the even part of $D(m, n)$ consists of so(2m) and sp(2n). It is long known that all bilinear combinations of the fermionic creation and annihilation operators of m states close under commutation operation and form the classical Lie algebra so(2m) (see Ref. 7). Furthermore it is also known that the symplectic algebra sp(2n) can be constructed from the bilinear combinations of bosonic creation and annihilation operators.[10] To form $D(m, n)$ additional vectors have to be constructed. Since a_i^+ and a_i transform as (2m) in so(2m) and b_k^+ and b_k transform as (2n) in sp(2n), it is tempting to use bilinear combinations of these operators to form the odd part of $D(m, n)$. The following sets serve the purpose:

$$D_1 = \{a_i b_k, a_i b_k^+, a_i^+ b_k, a_i^+ b_k^+, \quad 1 \leqslant i \leqslant m, \quad 1 \leqslant k \leqslant n\}. \tag{35}$$

Here, $D(m)$ is given by

$$\{a_i^+ a_i - \tfrac{1}{2}, a_i^+ a_j, a_i a_j, a_i^+ a_j^+, \quad i \neq j, \quad 1 \leqslant i, j \leqslant m\}, \tag{36}$$

and its Cartan subalgebra is

$$\{a_i^+ a_i - \tfrac{1}{2}, \quad i = 1,\ldots,m\}, \tag{37}$$

and $C(n)$ is given by

$$\{b_k^+ b_k + \tfrac{1}{2}, b_k^+ b_l, b_k b_l, b_k^+ b_l^+, \quad k \neq l, \quad 1 \leqslant k, l \leqslant n\}, \tag{38}$$

and its Cartan subalgebra is

$$\{b_l^+ b_l + \tfrac{1}{2}, \quad l = 1,\ldots,n\}. \tag{39}$$

Now we prove that the union of the sets in (35), (36), and (38) contribute to the classical Lie superalgebra $D(m, n)$. We proceed in a manner analogous to that in Sec. III. It is simple to show that if D_1 has odd grading and $D(m)$ and $C(n)$ have even grading, they satisfy axioms (1) and (2) and therefore they form a Lie superalgebra. To make the identification with $D(m, n)$, consider the operator

$$H = \sum \epsilon_i (a_i^+ a_i - \tfrac{1}{2}) + \sum \delta_k (b_k^+ b_k + \tfrac{1}{2}), \tag{40}$$

which is an element in the Cartan subalgebra of $D(m) + C(n)$. It is straightforward to calculate the commutators of H with all elements in Eqs. (35), (36), and (38). It is found that $a_i^+ a_j$, $a_i a_j$, and $a_i^+ a_j^+$ correspond to the roots $\epsilon_i - \epsilon_j$, $-\epsilon_i - \epsilon_j$, and $\epsilon_i + \epsilon_j$, respectively; $b_k^+ b_l$, $b_k b_l$, and $b_k^+ b_l^+$ correspond to the roots $\delta_k - \delta_l$, $-\delta_k - \delta_l$, and $\delta_k + \delta_l$, respectively; and $b_k b_k$ and $b_k^+ b_k^+$ correspond to the roots $-2\delta_k$ and $2\delta_k$, respectively. Moreover, the elements in D_1 correspond to all the odd roots, that is, $a_i b_k$, $a_i b_k^+$, $a_i^+ b_k$, and $a_i^+ b_k^+$ correspond to $-\epsilon_i - \delta_k$, $-\epsilon_i + \delta_k$, $\epsilon_i - \delta_k$, and $\epsilon_i + \delta_k$, respectively. As a rule, creation operators provide a positive basis vectors ϵ_i for fermions and δ_k for bosons. Destruction operators provide the negative basis vectors $-\epsilon_i$ for fermions and $-\delta_k$ for bosons. In other words, Hermitian conjugation changes the sign of the basis vectors.

We introduce the simple root notation

$$\alpha_i = \begin{cases} \delta_1 - \delta_2, & i = 1,...,n-1, \\ \delta_n - \epsilon_1, & i = n, \\ \epsilon_k - \epsilon_{k+1}, & i = n+k, \quad k = 1,...,m-1, \\ \epsilon_{m-1} + \epsilon_m, & i = n+m. \end{cases} \tag{41}$$

The corresponding generators are

$$E_{\alpha_i} = \begin{cases} b_i^+ b_{i+1}, & i = 1,...,n-1, \\ b_n^+ a_1, & i = n, \\ a_k^+ a_{k+1}, & i = n+k, \quad k = 1,...,m-1, \\ a_{m-1}^+ a_m^+, & i = n+m. \end{cases} \tag{42}$$

It follows from Eqs. (42) that

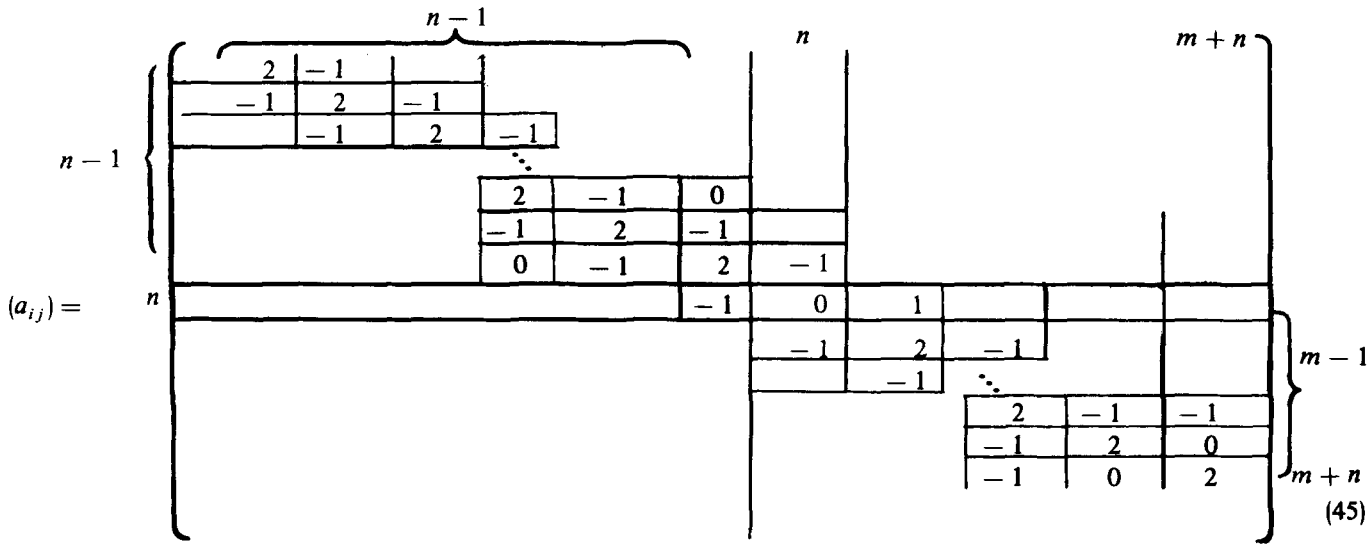
$$\begin{aligned} &[E_{\alpha_i}, E_{-\alpha_j}] = \delta_{ij} h_i, \\ &\{E_{\alpha_{m+1}}, E_{-\alpha_{m+1}}\} = h_{m+1}, \\ &[E_{\alpha_{m+1}}, E_{-\alpha_i}] = [E_{\alpha_i}, E_{-\alpha_{m+1}}] = 0, \quad \text{for } i \neq m+1 \\ &[h_i, E_{\pm\alpha_j}] = \pm a_{ij} E_{\pm\alpha_j}, \quad \text{for all } i \text{ and } j, \end{aligned} \tag{43}$$

where

$$h_i = \begin{cases} b_i^+ b_i - b_{i+1}^+ b_{i+1}, & i = 1,...,n-1, \\ (b_n^+ b_n + \tfrac{1}{2}) + (a_1^+ a_1 - \tfrac{1}{2}), & i = n, \\ a_k^+ a_k - a_{k+1}^+ a_{k+1}, & i = n+k, \quad k = 1,...,m-1, \\ (a_{m-1}^+ a_{m-1} - \tfrac{1}{2}) + (a_m^+ a_m - \tfrac{1}{2}) & i = m+n \end{cases} \tag{44}$$

are the elements in the Cartan subalgebra of the Lie superalgebra and (a_{ij}) is given by

$(a_{ij}) =$



(45)

The $D(m, n)$ Lie superalgebra realized in this construction has properties different from that of $A(m-1, n-1)$. We note the following.

(1) Total particle number no longer conserves. Instead, variation of particle number in even integral values is allowed.

(2) The odd part of the superalgebra contains elements which change the fermion number and the boson number by one unit simultaneously.

(3) This $D(m, n)$ Lie superalgebra is the largest superalgebra that accommodates only bilinear combinations of the fermionic and bosonic operators. We will discuss the representation in Sec. VI.

V. $B(m, n)$ AND SECOND QUANTIZATION

The even part of $B(m, n)$ is composed of a so$(2m+1)$, which is constructed from all possible linear and bilinear combinations of the fermionic creation and destruction operators,[7] and a sp$(2n)$ which is the one in $D(m, n)$. The odd part corresponds to the irreducible representation $(2m+1, 2n)$ of $B(m)+C(n)$. The possible candidates are products of a bosonic operator with a fermionic operator. To proceed with the search for the odd generators consider an element

$$H=\sum\epsilon_i\left(a_i^+a_i-\frac{1}{2}\right)+\sum\delta_k\left(b_k^+b_k+\frac{1}{2}\right) \tag{46}$$

of the Cartan subalgebra of $B(m)+C(n)$. In the usual way of deriving roots the commutator of H with the bilinear products $a_i^+b_j$, $a_ib_j^+$, a_ib_j, and $a_i^+b_j^+$ produce odd roots $\epsilon_i-\delta_j$, $-\epsilon_i+\delta_j$, $-\epsilon_i-\delta_j$, and $\epsilon_i+\delta_j$, respectively. To search for the operator that associates with $-\delta_j=-\epsilon_i+(\epsilon_i-\delta_j)$ the commutator

$$[a_i, a_i^+b_j]=(1-2a_i^+a_i)b_j=-2h_ib_j \tag{47}$$

shows that h_ib_j corresponds to $-\delta_j$. As a matter of fact, the commutator of H with h_ib_j just produces the root $-\delta_j$. Therefore it is tempting to say that the union of the set

$$\text{so}(2m+1)=\begin{cases}a_i^+,\ a_i,\ a_i^+a_i-\frac{1}{2},\ a_i^+a_j^+,a_i^+a_j,\ a_ia_j,\\ i\neq j,\quad 1\leqslant i,j\leqslant m,\end{cases} \tag{48}$$

the set

$$\text{sp}(2n)=\begin{cases}b_k^+b_k+\frac{1}{2},\ b_k^+b_l(k\neq l),\ b_kb_l,\ b_k^+b_l^+,\\ 1\leqslant l,\ k\leqslant n,\end{cases} \tag{49}$$

and the odd part

$$\{a_i^+b_k,\ a_ib_k^+,\ a_ib_k,\ a_i^+b_k^+,\ h_ib_k,\ h_ib_k^+,$$
$$i=1,...,m,\quad k=1,...,n\} \tag{50}$$

constitute $B(m, n)$. However, it is not true because (1) there are too many operators (m of them) corresponding to $-\delta_j$. They are h_ib_k for $i=1,...,m$. (2) h_ib_k is not fermionic (i.e., odd). It is a bosonic operator even though it reproduces the odd root $-\delta_j$. (3) The anticommutator

$$\{h_ib_k, a_jb_k^+\}=a_j(h_i-1-b_k^+b_k+2h_ib_k^+b_k)$$

does not close, and (4) h_i in h_ib_k do not give the correct representation $(2m+1, 2n)$. The way out of the difficulty is culty is indicated by argument (2) above. The symplectic Lie algebra sp$(2n)$ and the odd roots $\pm\delta_k$ imply that b_k does to a certain extent correspond to $-\delta_k$. However, it is bosonic in character. To make it fermionic, let us multiply a Clifford element θ (its grading is 1) to b_k to form θb_k (grading is then 1). We demand θ satisfies the following relations. It anticommutes with any a_i^+ and a_i, $\{\theta, a_i^+\}=\{\theta, a_i\}=0$, $i=1,...,m$.

It commutes with any b_k and b_k^+,

$$[\theta, b_k^+]=[\theta, b_k]=0,\quad k=1,...,n;$$

moreover, $\theta^2=-1$ and $\theta^+=-\theta$. Then, if we replace h_ib_k by θb_k, $h_ib_k^+$ by θb_k^+, a_i by θa_i, and a_i^+ by θa_i^+ in (50) and (48), respectively, all four difficulties listed above immediatley disappear. It is easy to show that all bracket operations close to form the $B(m, n)$ superalgebra. Because we have introduced θ, the even part contains, instead of Eq. (48), the set

$$\{\theta a_i,\ \theta a_i^+,\ a_i^+a_j,\ a_ia_j,\ a_i^+a_j^+,\quad i\neq j,\quad a_i^+a_i-\tfrac{1}{2}\}, \tag{51}$$

which is also a $B(m)=\text{so}(2m+1)$.

We list below some of the crucial relations. Introduce the simple roots

$$\alpha_i=\begin{cases}\delta_i-\delta_{i+1}, & i=1,...,n-1,\\ \delta_n-\epsilon_1, & i=n,\\ \epsilon_k-\epsilon_{k+1}, & i=n+k,\quad k=1,...,m-1,\\ \epsilon_m, & i=n+m.\end{cases} \tag{52}$$

The corrresponding generators are

$$E_{\alpha_i}=\begin{cases}b_i^+b_{i+1}, & i=1,...,n-1,\\ b_n^+a_1, & i=n,\\ a_k^+a_{k+1}, & i=n+k,\quad k=1,...,m-1,\\ \theta a_m^+, & i=n+m.\end{cases} \tag{53}$$

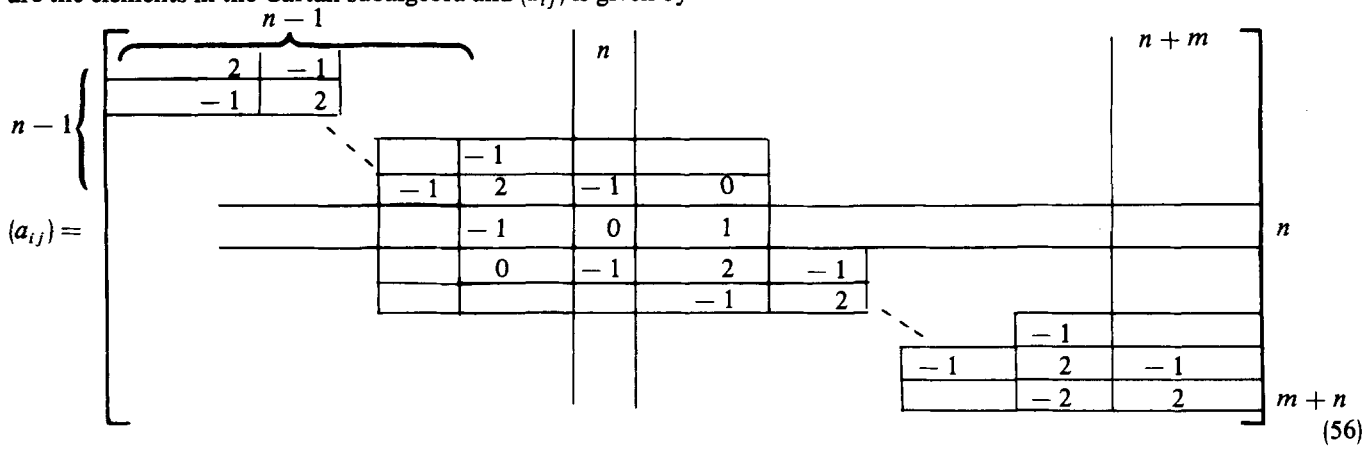
We have

$$\begin{aligned}&[E_{\alpha_i}, E_{-\alpha_j}]=\delta_{ij}h_i,\\ &\{E_{\alpha_n}, E_{-\alpha_n}\}=h_n,\\ &[E_{\alpha_n}, E_{-\alpha_i}]=[E_{\alpha_i}, E_{-\alpha_n}]=0,\quad \text{for } i\neq n\\ &[h_i, E_{\pm\alpha_j}]=\pm a_{ij}E_{\pm\alpha_j},\quad \text{for all } i \text{ and } j,\end{aligned} \tag{54}$$

where

$$h_i=\begin{cases}b_i^+b_i-b_{i+1}^+b_{i+1}, & i=1,...,n-1,\\ (b_n^+b_n+\frac{1}{2})+(a_1^+a_1-\frac{1}{2}), & i=n,\\ a_k^+a_k-a_{k+1}^+a_{k+1}, & i=n+k,\quad k=1,...,m-1,\\ 2(a_m^+a_m-\frac{1}{2}), & i=n+m\end{cases} \tag{55}$$

are the elements in the Cartan subalgebra and (a_{ij}) is given by



(56)

Note that the above formulation simply says that all fermion and boson destruction and creation operators are multiplied by a Clifford element such that the resulting fermion operators have even grading and the boson operators have odd grading. This formulation applies also to $A(m-1, n-1)$ and $D(m, n)$ with virtually nothing changed.

VI. REPRESENTATIONS

The Lie superalgebras realized by explicit construction through fermionic and bosonic creation and annihilation operators are important in the sense that it opens potential application of supersymmetry even to many-body (fermion–boson) problems. However, to make contact with quantum mechanics, the physical representations of the classical Lie superalgebras have to be identified. This is not yet completely solved since the infinite dimensional unitary irreducible representations[5] of the classical Lie superalgebras are not known. The discussion in this section is rather indicative than conclusive. Many facets[5,11] of the theory are to remain explored.

Let Fa denote the Fock space of the fermoionic operators, and Fb the Fock space of the bosonic operators. Consider the action of the elements of $A(m-1, n-1)$, $D(m, n)$, and $B(m, n)$ on the direct product $Fa \times Fb$. In the case of $A(m-1, n-1)$, the even parts $A(m-1)$ and $A(n-1)$ do not change the particle numbers of bosons and fermions. The physical irreducible representations are the totally antisymmetric representation (1^M) for $A(m-1)$ and the totally symmetric representation (N) for $A(n-1)$. Here M and N are the particle numbers of fermion and boson, respectively. This implies that the physical state space $Fa \times Fb$ decomposes into

$$\sum \oplus (1^M, N) \tag{57}$$

summing over M and N. The odd parts of $A(m-1, n-1)$, however, mix the totally antisymmetric representations labeled by different M integers with the totally symmetric representations labeled by different N integers and at the same time keep the total particle number (fermion and boson) constant, i.e.,

$$M + N = \text{constant positive integer } R. \tag{58}$$

Therefore, the irreducible representations of $A(m-1, n-1)$ can be labeled by R. Hence, the $A(m-1, n-1)$ irreducible representations restrict $Fa \times Fb$ to a subspace of states which belong to

$$\sum \oplus (1^M, N), \quad 0 \leqslant M \leqslant m, \quad 0 \leqslant N \tag{59}$$

summing over M and N with restriction (58).

In the case of $D(m, n)$, the even part consists of a so$(2m)$ which has the spinor representation as the physical representation. Furthermore, the elements in so$(2m)$ realized by the fermionic creation and destruction operators are bilinear and therefore spinor representations of odd and even fermion particle number do not mix. Hence, for so$(2m)$, we have irreducible representations either

$$(1/2^m) \quad \text{or} \quad (1/2^{m-1} - 1/2), \tag{60}$$

where $(1/2^m) = (1/2 \ldots m \text{ times} \ldots 1/2)$ is the spinor representation label in which each entry is the highest weight state associated with an element of the so$(2m)$ Cartan subalgebra in (36). The spinor representation is finite dimensional.

The sp$(2n)$, which also constitutes the even part of $D(m, n)$ is made up of bilinear terms from the boson creation and annihilation operators. Therefore, representations of odd and even boson particle numbers do not mix. Hence, we have irreducible representations[10] either

$$\langle 1/2^n \rangle \quad \text{or} \quad \langle 1/2^{n-1}\ 3/2 \rangle, \tag{61}$$

where $\langle 1/2 \rangle = \langle 1/2 \ldots n \text{ times} \ldots 1/2 \rangle$ in which each entry is the lowest weight state associated with an element of the sp$(2n)$ Cartan subalgebra in (39). Note that the representation is infinite dimensional (physically, each bosonic state can accommodate any number of boson quanta) and thus is consistant with the fact that the group of real sp$(2n)$ is noncompact. This is obvious if one consider the case $n = 1$ whence SP(2) = U(1, 1), a noncompact group which has two continuous and two discrete irreducible representations. The physical one is that discrete irreducible representation which is bounded from below.[7]

The odd part of $D(m, n)$ is composed of elements each of which is a product of one fermionic and one bosonic operator. It mixes the even particle number and odd particle number representations of both so$(2m)$ and sp$(2n)$. However, it preserves the overall (sum of fermion and boson particle

number) oddness and evenness of so(2m) and sp(2n) together. Therefore, the physical irreducible representations of $D(m, n)$ decompose $Fa \times Fb$ into subspaces of states in which the overall oddness or evenness is a constant. In other words, the physical irreducible representations of $D(m, n)$ have the $A(m-1) + A(n-1)$ subalgebra decomposition either

$$\sum \oplus (1^M N), \tag{62a}$$

in which $(N+M) = 0 \bmod 2$ and summing over M and N with $0 \leqslant N$, $0 \leqslant M \leqslant m$ or

$$\sum \oplus (1^M N), \tag{62b}$$

in which $(N+M) = 1 \bmod 2$ and summing over M and N with $0 \leqslant N$, $0 \leqslant M \leqslant m$.

In the case of $B(m, n)$, the above restriction is completely removed. This is explained below. The even part of $B(m, n)$ has the same sp(2n) as $D(m, n)$ and therefore the same physical sp(2n) representations (61). It also consists of so($2m+1$) which is formed by linear and bilinear terms of the fermion creation and annihilation operators. The linear terms θa_i and θa_i^+ mix the evenness and oddness of the fermion particle numbers. Therefore, the physical irreducible representation is the spinor representation

$$[1/2^M],$$

which has the so(2m) subgroup decomposition

$$(1/2^m) + (1/2^{m-1} - 1/2).$$

The odd part of $B(m, n)$ has terms linear in bosonic creation and annihilation operators, namely, θb_k and θb_k^+, which also mix the sp(2n) odd and even particle number representations (61). Hence, the whole physical space $Fa \times Fb$ belongs to the physical irreducible representation of $B(m, n)$, which has the following $A(m-1) + A(n-1)$ decomposition:

$$\sum \oplus (1^M N),$$

summing over N and M with $0 \leqslant M \leqslant m$, $0 \leqslant N$.

The above discussion is by no means rigorous mathematically. However, it brings out the essential points in the construction of the physical representations for a classical Lie superalgebra, namely, the evenness and oddness in the particle (overall, fermion, boson) number and the roles played by the even and odd parts of the superalgebra, the former (even part) supplying the irreducible representations and the later (odd part) mixing up these irreducible representations.

VII. CONCLUDING REMARKS

We have accomplished so far realizations of the classical Lie superalgebras $A(m-1, n-1)$, $B(m, n)$, $\mathbb{C}(n)$, and $D(m, n)$ and indicated to some extent properties of the corresponding physical representations. These suggest that a Hamiltonian of many fermion–boson interaction is an element of the enveloping algebra of the classical superalgebras. The discussion in the previous sections are so far limited to the study of "kinematic symmetries" of the basic fermion–boson creation and annihilation operators. The dynamics aspect has not yet been touched. It will be crucial for future development to construct physical model Hamiltonians of ordinary fermion–boson interactions that possess dynamical supersymmetry.

APPENDIX: LIE SUPERALGEBRA $\mathbb{C}(n)$

$\mathbb{C}(n)$ can be realized by one fermionic and $n-1$ bosonic creation and annihilation operators. The standard basis is

$$\{\epsilon_1, \delta_i, \quad i = 1,\dots,n-1\}.$$

The simple roots are

$$\alpha_i = \begin{cases} \epsilon_1 - \delta_1, & i = 1, \\ \delta_k + \delta_{k+1}, & i = k+1, \quad k = 1,\dots,n-2, \\ 2\delta_{n-1}, & i = n. \end{cases}$$

The corresponding generators are

$$E_{\alpha_i} = \begin{cases} a_1^+ b_1, & i = 1, \\ b_k^+ b_{k+1}, & i = k+1, \quad k = 1,\dots,n-2, \\ \frac{1}{2} b_{n-1}^+ b_{n-1}^+, & i = n. \end{cases}$$

The commutators and anticommutators are

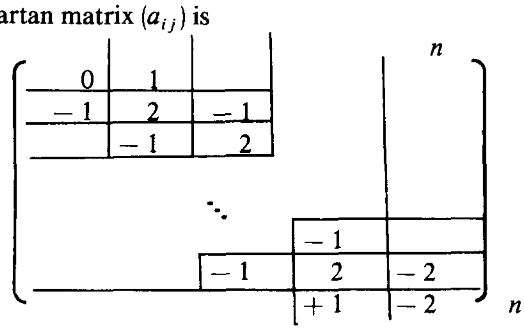
$$[E_{\alpha_i}, E_{-\alpha_i}] = \delta_{ij} h_i, \quad i = 2,\dots,n,$$
$$\{E_{\alpha_1}, E_{-\alpha_1}\} = h_1,$$
$$[E_{\alpha_1}, E_{-\alpha_i}] = [E_{\alpha_i}, E_{-\alpha_1}] = 0, \quad i \neq 1,$$
$$[h_i, E_{\pm\alpha_j}] = \pm a_{ij} E_{\pm\alpha_j}, \quad \text{for all } i \text{ and } j,$$

where

$$h_i = \begin{cases} (a_1^+ a_1 - \frac{1}{2}) + (b_1^+ b_1 + \frac{1}{2}), & i = 1, \\ b_k^+ b_k - b_{k+1}^+ b_{k+1}, & i = k+1, \; k = 1,\dots,n-2, \\ -(b_{n-1}^+ b_{n-1} + \frac{1}{2}), & i = n. \end{cases}$$

The Cartan matrix (a_{ij}) is

$(a_{ij}) =$



The even part is $a_1^+ a_1 - \frac{1}{2}$ with sp($2(n-1)$) as in (49) and the odd part is

$$\{a_1 b_j, a_1 b_j^+, a_1^+ b_j, a_1^+ b_j^+, \quad j = 1,\dots,n-1\}.$$

The physical irreducible representation is similar to that of $D(m, n)$ in which the evenness or oddness of the particle (fermion and boson) number is preserved. The subspace of $Fa \times Fb$ in a supermultiplet of the physical irreducible representation of $\mathbb{C}(n)$ can only belong to either

$$\sum \oplus (1^M N),$$

with the restriction $(M+N) = 0 \bmod 2$ and summing over M and N, or

$$\sum \oplus (1^M N),$$

with the restriction $(M+N)=1 \bmod 2$ and summing over M and N, where M can only be 0 or 1, in the unitary group decomposition.

[1]P. Fayet and S. Ferrera, Phys. Rep. **32**, 249 (1977).
[2]P. van Nieuwenhuizen, Phys. Rep. **68**, 189 (1981).
[3]F. Iachello, Phys. Rev. Lett. **44**, 772 (1980).
[4]B. Balantekin and I. Bars, J. Math. Phys. **22**, 1149, 1810 (1981); B. Balantekin, I. Bars, and F. Iachello, Yale Preprints YTP81-07.
[5]V. Kac, Adv. Math. **26**, 8 (1977); Commun. Math. Phys. **53**, 31 (1977); *Lecture Notes in Mathematics* (Springer, New York, 1978), Vol. 676.
[6]J. -P. Hurni and B. Morel, J. Math. Phys. **23**, 2236 (1982); J. Math. Phys. **24**, 157 (1983).
[7]B. Wybourne, *Classical Groups for Physicists* (Wiley–Interscience, New York, 1974).
[8]H. Z. Sun and Q. Z. Han, Progr. Phys. **3**, 81 (1983).
[9]N. Jacobson, *Lie Algebras* (Dover, New York, 1962).
[10]M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971).
[11]T. D. Palev, J. Math. Phys. **21**, 1293 (1980).

Exact solutions of the Schrödinger equation for a class of three-dimensional isotropic anharmonic oscillators

P. G. L. Leach
Department of Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa

(Received 26 September 1983; accepted for publication 20 January 1984)

A mathematical explanation of how the Schrödinger equation for a class of three-dimensional isotropic anharmonic oscillators possesses exact solutions is given. The class of potentials is shown to be wider than at present found in the literature.

PACS numbers: 03.65.Ge, 33.10.Cs

I. INTRODUCTION

The determination of exact solutions to the time-independent Schrödinger equation has been an object of study for many years and in earlier studies[1] classes of potentials have been enumerated. Trivially the Schrödinger equation

$$(-\tfrac{1}{2}\nabla^2 + V)\psi = E\psi \tag{1.1}$$

($\hbar$ and m are taken as one throughout this paper) is always satisfied by $\psi = \psi_0$, where ψ_0 is a real-valued C^2 function when $E = 0$ and $V = \frac{1}{2}\nabla^2\psi_0/\psi_0$ provided that the potential V makes physical sense. However, if ψ_0 is the only explicit solution which this potential provides the result is of little or no practical use. What is useful is a potential for which more than one, preferably many more than one, explicit solutions can be obtained. The value of such solutions is that, if the potential can be treated by perturbation methods, the exact solutions provide an excellent check of the perturbation expansion for those states for which they exist. If the potential cannot be treated by perturbation methods, then the exact solutions do at least give some information about the energy levels.

In the particular instance of anharmonic oscillators there has been some recent interest in those potentials for which an exact solution to the time-independent Schrödinger equation exists. Examples of such potentials occur in the study of the spectra of molecules for which the potential has been modeled as a double-minimum well. This is found even in fairly early studies of simple molecules such as ammonia and hydrogen-bonded solids.[2] Exact solutions have been obtained for both one-dimensional systems[3,4] and isotropic multidimensional systems. The usual method of determination of potential and state function adopted is independent of the dimension and is illustrated by the simplest example. The Hamiltonian is

$$H = \tfrac{1}{2}p^2 + V(q), \tag{1.2}$$

with

$$V(q) = aq^2 + bq^4 + cq^6, \tag{1.3}$$

where a, b, and c are constants and c is necessarily positive. The time-independent Schrödinger equation is

$$\left(-\frac{1}{2}\frac{d^2}{dq^2} + V(q)\right)\psi(q) = E\psi(q). \tag{1.4}$$

A solution of the form

$$\psi(q) = \exp(-\tfrac{1}{2}\alpha q^2 - \tfrac{1}{4}\beta q^4) \tag{1.5}$$

(the normalization factor has been ignored), where α and β are constants and β is necessarily positive, is assumed. The substitution of $\psi(q)$ (1.5) into (1.4) yields the following set of conditions on the coefficients a, b, c, α, and β when in the resulting expression the coefficient of each separate power of q is set equal to zero:

$$E = \tfrac{1}{2}\alpha, \quad \alpha = \tfrac{1}{2}(\alpha^2 - 3\beta), \quad b = \alpha\beta, \quad c = \tfrac{1}{2}\beta^2. \tag{1.6}$$

This state is a ground state and the ground state energy level is positive or negative according to whether α is positive or negative. Eliminating α and β from the expressions for a, b, and c, a potential of the form (1.3) will give rise to a solution of the form (1.5) if a, b, and c satisfy the relation

$$a = b^2/c \pm \tfrac{3}{2}\sqrt{2}c. \tag{1.7}$$

As stated above a single explicit solution for a given potential is of little value. It is possible to obtain more than one explicit solution (corresponding to different energy levels) for a particular potential by assuming that $\psi(q)$ has the form

$$\psi(q) = f(q)\exp(-\tfrac{1}{2}\alpha q^2 - \beta q^4), \tag{1.8}$$

where $f(q)$ is a polynomial of form $qg(q^2)$ or $g(q^2)$ for odd and even states, respectively. In this case the coefficients in the potential are not the same as given in (1.6) and the coefficients of $g(q^2)$ are fixed by the values of α and β. Potentials containing further anharmonic terms may be obtained by increasing the degree of the polynomial in the exponent of (1.5) [or (1.8)]. A curious feature is that this method worked only for potentials of the form

$$V(q) = \sum_{n=0}^{2N} \alpha_n q^{2n+2}. \tag{1.9}$$

Leach[4] has provided a mathematical basis for what had been hitherto a solution obtained by ansatz. The time-independent Schrödinger equation (1.4) may be written as

$$(D^2 + \lambda - X)\psi = 0, \tag{1.10}$$

where $D \equiv d/dq$, λ is the eigenvalue, and $X(q)$ the potential (to within a constant multiplier). Assuming that the differential operator in (1.10) may be factored,

$$D^2 + \lambda - X = (D - \alpha)(D - \beta), \tag{1.11}$$

it follows that

$$\beta = -\alpha, \quad \alpha^2 - \alpha' = X - \lambda. \tag{1.12}$$

Equation (1.10) is now

 0022-2488/84/102974-05$02.50

$$(D-\alpha)(D+\alpha)\psi = 0, \tag{1.13}$$

with solution

$$\psi(q) = A\exp\left(-\int^q \alpha(u)du\right) + B\exp\left(\int^q \alpha(u)du\right) \times \int^q \exp\left(2\int^u \alpha(v)dv\right)du. \tag{1.14}$$

In the context of an oscillator potential, $\alpha(q)$ is a polynomial and either A or B will be zero according to the behavior of $\alpha(q)$ as $|q|\to\infty$. The solution for $\psi(q)$ in (1.14) is a ground state solution. To obtain higher states a factorization of the form

$$(D-\alpha)(fD-\beta) = f(D^2+\lambda-X) \tag{1.15}$$

is assumed. By an analysis similar to that above,

$$X-\lambda = \alpha^2-\alpha'-(2\alpha f'-f'')/f, \tag{1.16}$$

$$\psi(q) = Af(q)\exp\left(-\int^q \alpha(u)du\right) + Bf(q)\exp\left(-\int^q \alpha(u)du\right) \times \int^q f^{-2}(u)\exp\left(2\int^u \alpha(v)dv\right)du, \tag{1.17}$$

with the same qualification on A and B.

The differential equations (1.12) and (1.16) are both Riccati equations. However, in the context of anharmonic oscillators it is not necessary to solve them. In (1.12) the choice of a polynomial for $\alpha(q)$ will give a polynomial $X(q)$. For the wave function to be square integrable the degree of $\alpha(q)$ must be odd, i.e., $2N+1$, whence the degree of $X(q) = 4N+2$, which is in accordance with (1.9). There is, however, no requirement that $X(q)$ be a polynomial in q^2. The polynomial

$$\alpha(q) = \sum_{n=0}^{2N+1} a_n q^n \tag{1.18}$$

will provide both a square integrable wave function and a polynomial potential. An oscillator potential is often obtained as a Maclaurin polynomial approximation to a potential about an equilibrium point. In such a case the coefficient of q^1 in $X(q)$ is zero, i.e., to (1.18) we add

$$a_0a_1 + a_2 = 0 \tag{1.19}$$

as a constraint on the coefficients of $\alpha(q)$. More generally the origin and equilibrium point need not coincide and the constraint (1.17) is not necessary. In (1.16) to obtain a polynomial $X(q)$ it is necessary for the term $(2\alpha f'-f'')/f$ to be a polynomial. This imposes constraints on the coefficients of the polynomial function $f(q)$ as was discussed in Ref. 4.

II. THREE-DIMENSIONAL ISOTROPIC OSCILLATOR

The multidimensional isotropic oscillator has also been treated by a method which is similar to that outlined by Eqs. (1.2)–(1.9) (see Ref. 5). Provided the system is isotropic its dimension is immaterial. The potentials which have been treated are of the form

$$V(r) = \sum_{n=0}^{2N} a_n r^{2n+2}, \tag{2.1}$$

where r is the radial variable and the a_n are not all arbitrary. We shall construct a broader class of potentials which includes the examples (2.1).

The three-dimensional time-independent Schrödinger equation

$$(-\tfrac{1}{2}\nabla^2 + V)\psi = E\psi \tag{2.2}$$

is reduced to the radial equation

$$[D^2+\lambda-X-l(l+1)r^{-2}]u = 0, \tag{2.3}$$

where $D \equiv d/dr$ and X is V to within a constant multiplier, by means of the substitution

$$\psi = [u(r)/r]Y_{lm}(\theta,\phi). \tag{2.4}$$

As for the one-dimensional problem we assume that the operator in (2.3) can be factored to give

$$(D-\alpha)(D-\beta)u = 0. \tag{2.5}$$

This requires that

$$\beta = -a, \quad \alpha\beta-\beta' = \lambda-X-l(l+1)r^{-2}. \tag{2.6}$$

Hence X, λ, and α are related according to

$$\alpha^2-\alpha' = X+l(l+1)r^{-2}-\lambda. \tag{2.7}$$

In terms of α the solution of (2.3) is

$$u(r) = A\exp\left(-\int^r \alpha(x)dx\right) + B\exp\left(-\int^r \alpha(x)dx\right) \times \int^r \exp\left(2\int^x \alpha(y)dy\right)dx, \tag{2.8}$$

where A and B are constants. If $\alpha(r) > 0$ as $r\to\infty$, $B = 0$ and, if $\alpha(r) < 0$ as $r\to\infty$, $A = 0$. In the following discussion we will consider the former case only as it is the simpler. The latter case, which corresponds to a different potential, could be discussed in a similar fashion.

As we are treating an anharmonic oscillator our choice of $\alpha(r)$ for substitution into (2.7) and (2.8) must be such that $X(r)$ is a polynomial in r and that the coefficients of the highest powers in $X(r)$ and $\alpha(r)$ must be positive. If we take

$$\alpha(r) = \sum_{n=-1}^{N} a_n r^n, \quad N > 1 \tag{2.9}$$

from (2.7) we see that $X(r)$ is a polynomial of degree $2N$, $a_N > 0$, $a_0 = 0$, and $\lambda = a_1(1-2a_{-1})$. The wave function is required to be square integrable and so

$$\int_0^\infty u^2(r)dr < \infty. \tag{2.10}$$

As

$$u(r) = r^{-a_{-1}}\exp\left(-\sum_{n=1}^{N}\frac{a_n}{n+1}r^{n+1}\right), \tag{2.11}$$

the integral is well-behaved as $r\to\infty$ since $a_N > 0$. It will be well-behaved as $r\to 0$ provided we make the choice $a_{-1} = -(l+1)$. Hence the eigenvalue is

$$\lambda = a_1(2l+3). \tag{2.12}$$

The function $\alpha(r)$ is

$$\alpha(r) = -(l+1)r^{-1} + \sum_{n=1}^{N} a_n r^n, \tag{2.13}$$

and the potential is, to within a constant multiplier,

$$X(r)=\left(\sum_{n=1}^{N}a_n r^n\right)^2-\sum_{n=1}^{N}(2l+n+3)a_{n+1}r^n, \quad (2.14)$$

in which, for a given value of the angular momentum eigenvalue, there are N independent constants. Alternatively we could write

$$X(r)=\sum_{n=1}^{2N}b_n r^n, \quad (2.15)$$

in which the $2N$ coefficients b_n are subject to N constraints.

The potential given in (2.14) describes an oscillator for which $r=0$ is not an equilibrium point. If $r=0$ is an equilibrium point the coefficient of r^1 in (2.14) must be zero. This requires that $a_2=0$. In the case $N=2$ the above results are the same as those obtained by Flessas and Watt[6] when due allowance is made for the presence of angular momentum. In general one can use any function $\alpha(r)$. The choice

$$\alpha(r)=-(l+1)r^{-1}+\sum_{n=1}^{N}a_n r^{i_n/j_n}$$

could be of some use in the study of the fractionally anharmonic oscillator which has been treated recently by Znojil[7] from a different viewpoint.

III. HIGHER STATES

The wave function and eigenvalue obtained in Sec. 2 are for the lowest energy level for a given value of the angular momentum. To obtain wave functions and eigenvalues for higher energy states we make a modification of the factorization as was done at (1.15) and write

$$(D-\alpha)(fD-\beta)\equiv f[D^2+\lambda-X-l(l+1)r^{-2}], \quad (3.1)$$

so that

$$\beta=-\alpha f+f', \quad (3.2)$$

$$X+l(l+1)r^{-2}-\lambda=\alpha^2-\alpha'-(2\alpha f'-f'')f^{-1}. \quad (3.3)$$

The function $u(r)$ is given by

$$u(r)=Af(r)\exp\left(-\int^r\alpha(x)dx\right)+Bf(r)\exp\left(-\int^r\alpha(x)dx\right)\times\int^r f^{-2}(x)\exp\left[2\int^x\alpha(y)dy\right]dx. \quad (3.4)$$

For X to represent an oscillator potential, the constant B will be zero if $\alpha(r)\to\infty$ as $r\to\infty$ and A will be zero if $\alpha(r)\to-\infty$ as $r\to\infty$. The latter case is computationally more complex and for clarity we shall confine our discussion to the former.

The angular momentum term on the left-hand side of (3.3) is provided for by allowing $\alpha(r)$ to contain a term in r^{-1}. We obtain higher states by taking $f(r)$ to be a polynomial in r. The coefficients of $f(r)$ for an oscillator potential are then determined by the requirement that $X(r)$ be a polynomial in r, i.e., the right-hand side of (3.3) may not contain fractions with polynomials in r in the denominator with the exception of a term in r^{-2}. We illustrate the process with some simple examples. They must of necessity be simple due to computational complexity even in the simplest cases. Let

$$\alpha(r)=\sum_{n=-1}^{N}a_n r^n, \quad (3.5)$$

$$f(r)=r-k. \quad (3.6)$$

Substituting these into (3.3) we find that

$$a_{-1}(a_{-1}+1)=l(l+1), \quad a_{-1}a_0+b_{-1}=0,$$

$$\lambda=a_1-a_0^2-2a_{-1}a_1-2b_0,$$

$$X(r)=\left(\sum_{n=1}^{N}a_n r^n\right)^2+\sum_{n=2}^{N}(2a_{-1}-n)a_n r^{n-1}+2a_0\sum_{n=1}^{N}a_n r^n+2\sum_{n=1}^{N-1}b_n r^n, \quad (3.7)$$

where

$$b_n=k^{-n-1}\sum_{m=-1}^{n}a_m k^m \quad (3.8)$$

and

$$\sum_{m=-1}^{N}a_m k^m=0. \quad (3.9)$$

The first three of (3.7) come from equating the coefficients of r^{-2}, r^{-1}, and r^0 of each side of (3.3), respectively, and the last is what remains. Equations (3.8) and (3.9) come from the requirement that f factor α. From (3.7) we see that

$$a_{-1}=-(l+1), \quad ka_0=-1, \quad \lambda=(2l+3)(a_0^2+a_1), \quad (3.10)$$

and (3.9) becomes

$$\sum_{m=1}^{N}a_m k^{m+1}=l+2. \quad (3.11)$$

Since $a_N>0$, there is always one positive root of (3.11). Because of the connection between k and a_0 different roots correspond to different potentials and so we have only one explicit solution to the Schrödinger equation for a given $\alpha(r)$.

Instead of using the $f(r)$ given in (3.6) we could use

$$f(r)=\prod_{j=1}^{J}(r-k_j). \quad (3.12)$$

Equations (3.7) are almost replicated. The only change is that the terms containing the b_n are more complicated. We may define these terms recursively by

$$b_n^{j+1}=k_j^{-n-1}\sum_{m=-1}^{n}b_m^j k_j^m, \quad j=1,J-1, \quad (3.13)$$

with the coefficients b_n^1 being given by equating the coefficients of like powers of r from the two sides of

$$\sum_{n=1}^{N+J}b_n^1 r^n=\alpha(r)f'(r)-f''(r). \quad (3.14)$$

Each term containing a b_n would also be multiplied by a factor $(-1)^{J-1}$ compared with (3.7) and the roots of $f(r)$ are determined from the equations

$$\sum_{m=-1}^{N+J-j}b_m^j k_j^m=0. \quad (3.15)$$

However, this is not a fruitful path to follow as the potential depends on the values of the roots of $f(r)$ and so only one eigenstate is found per potential.

To obtain multiple eigenstates it is necessary to place some restrictions on the forms of the functions $\alpha(r)$ and $f(r)$.

We consider some examples.

Example (i):

$$\alpha(r) = -(l+1)r^{-1} + ar + br^3, \quad b>0,$$
$$f(r) = r^2 - k.$$

Then

$$(2\alpha f' - f'')f^{-1} = 4br^2 + 4(a+kb), \tag{3.16}$$

provided that

$$4(a+kb)k = 4l+6. \tag{3.17}$$

The roots of (3.17) are

$$k = (2b)^{-1}[-a \pm \sqrt{a^2 + b(4l+6)}]. \tag{3.18}$$

The potential is determined to within a constant factor by

$$X(r) = [a^2 - b(2l+9)]r^2 + 2abr^4 + b^2r^6. \tag{3.19}$$

For this potential there are two explicit eigenfunctions with eigenvalues

$$\lambda_0 = a(2l+5) - 2\sqrt{a^2 + b(4l+6)}, \tag{3.20}$$

$$\lambda_1 = a(2l+5) + 2\sqrt{a^2 + b(4l+6)}, \tag{3.21}$$

corresponding to the two values of k. For the lower eigenvalue, λ_0, k is negative and $f(r)$ is never zero and so the state is a ground state. For $\lambda_1, f(n)$ has a single zero for positive r and is the first excited state. The result when $f(r) \equiv 1$ may be obtained from (2.12) and (2.14) using the particular expression for $\alpha(r)$ in this example. It is

$$\lambda = a(2l+3),$$
$$X(r) = [a^2 - (2l+5)b]r^2 + 2abr^3 + b^2r^6. \tag{3.22}$$

We see that the potential differs from that given by (3.19). This is a general feature for anharmonic potentials as was noted also for the one-dimensional problem.

To obtain a greater number of explicit solutions to the Schrödinger equation, we merely increase the degree of the polynomial $f(r)$. We shall look at the cases when $f(r)$ is a quartic and a sextic polynomial in r.

Example (ii):

$$\alpha(r) = -(l+1)r^{-1} + ar + br^3,$$
$$f(r) = r^4 - kr^2 + c.$$

In the course of doing the algebra it is convenient to introduce new variables. They are defined by

$$b = 2\sigma^2, \quad a = 2\sigma\nu,$$
$$a + kb = 2\sigma\eta, \quad \rho^2 = \sigma r^2, \quad \sigma > 0. \tag{3.23}$$

In terms of the new notation,

$$c = \tfrac{1}{4}(2l+3)(\eta - \nu)(\eta + \nu)^{-1}, \tag{3.24}$$

$$f = \sigma^{-2}[\rho^4 - (\eta - \nu)\rho^2 + \tfrac{1}{4}(2l+3)(\eta - \nu)(\eta + \nu)^{-1}], \tag{3.25}$$

$$\alpha = \sigma^{1/2}[-(l+1)\rho^{-1} + 2\nu\rho + 2\rho^3]. \tag{3.26}$$

We find that the potential is given by

$$X = 2\sigma\{[2\nu^2 - (2l+13)]\rho^2 + 4\nu\rho^4 + 2\rho^6\}, \tag{3.27}$$

and the eigenvalues by

$$\lambda = 2\sigma[(2l+7)\nu + 4\eta], \tag{3.28}$$

where η is a root of the equation

$$\eta^3 - [\nu^2 + (2l+4)]\eta - \nu = 0. \tag{3.29}$$

The three roots of (3.29) are all real and obey the inequalities

$$\eta_1 < -|\nu| < \eta_2 < |\nu| < \eta_3. \tag{3.30}$$

The potential X is independent of the value of η and so possesses three eigenstates. The ground state is given by η_1 and f has no real roots. For η_2, f has one real root for $\rho > 0$ and, for η_3, f has two real roots for $\rho > 0$.

Example (iii): Using the same α as in example (ii) and introducing new variables as in (3.23), as a sextic polynomial f has the form

$$f = \sigma^{-3}\Big(\rho^6 - \tfrac{1}{2}(\eta - 3\nu)\rho^4 + \frac{(\eta - 3\nu)(\eta + 3\nu)(2l+5)}{2[(\eta+\nu)(\eta+3\nu) - 3(2l+5)]}\rho^2 - \frac{(\eta - 3\nu)(2l+5)(2l+3)}{4[(\eta+\nu)(\eta+3\nu) - 3(2l+5)]}\Big). \tag{3.31}$$

In this notation the potential is given by

$$X = 2\sigma\{[2\nu^2 - (2l+17)]\rho^2 + 4\nu\rho^4 + 2\rho^6\}, \tag{3.32}$$

and the eigenvalues by

$$\lambda = 2\sigma[(2l+9)\nu + 2\eta], \tag{3.33}$$

where η is a solution of

$$[\eta^2 - (\nu^2 + 2l + 5)][\eta^2 - 9(\nu^2 + 2l + 5)] - 48\eta\nu - 36 = 0. \tag{3.34}$$

Again the potential is independent of η. The four real roots of (3.34) satisfy the inequalities

$$\eta_1 < -\kappa\sqrt{5} < \eta_2 < 0 < \eta_3 < \kappa < 3\kappa < \eta_4, \tag{3.35}$$

in the case $\nu > 0$, and

$$\eta_1 < -3\kappa < -\kappa < \eta_2 < 0 < \eta_3 < \kappa\sqrt{5} < \eta_4, \tag{3.36}$$

in the case $\nu < 0$, where

$$\kappa^2 = \nu^2 + 2l + 5, \quad \kappa > 0. \tag{3.37}$$

The number of zeros of f for $\rho > 0$ increases from zero with η_1 to three with η_4.

IV. DISCUSSION

In this note we have demonstrated that a factorization method can be used to determine exact solutions of the time-dependent Schrödinger equation for a three-dimensional isotropic anharmonic oscillator. It should be evident that the method is applicable to an arbitrary number of dimensions. In the process we have extended the class of potentials for which an exact solution is known to exist. Indeed any potential which can be written in the form

$$X(r) = \alpha^2(r) - \alpha'(r) + \lambda - l(l+1)r^{-2} \tag{4.1}$$

will have a wave function of the form

$$\psi(r,\theta,\phi) = \exp\Big(-\int^r \alpha(u)du\Big)\, Y_{lm}(\theta,\phi). \tag{4.2}$$

To attempt to solve (4.1) for $\alpha(r)$ is just the equivalent of attempting to solve the original time-independent Schrödinger equation (2.3). This is why a constructive procedure has been used here. It will have been observed that the number of parameters in a given potential is fewer than the number of terms in the potential. This means that not every an-

harmonic potential can be treated by the methods described here even though the powers in the potential are appropriate. Given a potential for a particular problem, it would first be necessary to test to see whether it is a member of the class described here. By way of example suppose that we have a potential of the form

$$X(r) = Ar^2 + Br^3 + Cr^6, \quad C > 0. \tag{4.3}$$

It will be of the desired form if there exists an $\alpha(r)$ and an $f(r)$ which satisfy (3.3) when X from (4.3) is included in the left-hand side. Taking $\alpha(r)$ and $f(r)$ to be given by

$$\alpha(r) = -(l+1)r^{-1} + ar + br^3,$$
$$f(r) = \sum_{n=1}^{N} c_n r^{2n}, \quad c_N = 1, \tag{4.4}$$

we require

$$Ar^2 + Br^4 + Cr^6 = (ar + br^3)^2 - (2l + 4N + 5)br^2. \tag{4.5}$$

Equating coefficients of like powers of r on each side of (4.5)

$$b^2 = C, \quad 2ab = B, \quad a^2 - (2l + 4N + 5)b = A. \tag{4.6}$$

Eliminating the a and b, we require a positive integer N such that

$$N = \tfrac{1}{4}(\tfrac{1}{4}B^2 C^{-3/2} - A C^{-1/2} - 2l - 5). \tag{4.7}$$

Should this be the case we will then have an explicit wave function for each of $N+1$ eigenstates with eigenvalues given by

$$\lambda = \tfrac{1}{2}(2l + 4N + 1)B C^{-1/2} + 4c_{N-1} C^{1/2}, \tag{4.8}$$

where c_{N-1} is one of the real roots of a polynomial equation of degree $N+1$.

ACKNOWLEDGMENT

It is a pleasure to thank the referee for constructive comments and suggestions.

[1]For an example, see M. F. Manning, Phys. Rev. **48**, 161 (1935); L. P. Eisenhart, Phys. Rev. **74**, 87 (1948).

[2]For an example, see D. M. Dennison and G. E. Uhlenbeck, Phys. Rev. **41**, 313 (1932); F. T. Wall and G. Glockler, J. Chem. Phys. **5**, 314 (1937); R. L. Somorjai and D. F. Hornig, J. Chem. Phys. **36**, 1980 (1962).

[3]G. P. Flessas, Phys. Lett. A **81**, 17 (1981); G. P. Flessas, J. Phys. A: Math. Gen. **14**, L209 (1981); E. Magyari, Phys. Lett. A **81**, 116 (1981); A. Khare, Phys. Lett. A **83**, 237 (1981).

[4]P. G. L. Leach, "Exact quantum mechanical solutions for a class of anharmonic oscillators" (submitted to J. Phys. A: Math. Gen.).

[5]G. P. Flessas and K. P. Das, Phys. Lett. A. **78** 19 (1980).

[6]G. P. Flessas and A. Watt, J. Phys. A: Math. Gen. **14**, L315 (1981).

[7]A. Znojil, J. Math. Phys. **24**, 1136 (1983).

Chain models and their effective Hamiltonians

M. Znojil
Institute of Nuclear Physics, Czechoslovak Academy of Sciences, 250 68 Řež, Czechoslovakia

(Received 1 August 1983; accepted for publication 9 March 1984)

We develop a new formalism suitable for diagonalization of Hamiltonians with $2t+1$ nonzero diagonals (chain models). In a systematic perturbation-like way, we get the expansions of Green and/or wave functions. They are generated by solution of a sequence of "fixed-point" quadratic equations for $(t\times t)$-dimensional matrices. For $t=1$ and $t=2$, this solution is feasible by elementary means, so that the respective tridiagonal and pentadiagonal chain models may be considered exactly solvable in this context. In an alternative first-order variational-perturbation formulation, the method may provide the simultaneous upper and lower energy bounds for any t.

PACS numbers: 03.65.Ge, 03.65.Db, 02.60. + y

I. INTRODUCTION

The concept of a strong interaction between neighboring "orbitals" has inspired the introduction and detailed study of the so-called chain models in solid-state physics,[1] in scattering or nuclear theory,[2] in the recent papers on the anharmonic oscillators,[3,4] etc. These models are all characterized by the Schrödinger equation

$$H\psi = E\psi, \tag{1.1}$$

with some infinite-dimensional band-matrix Hamiltonian H.

By definition, the off-diagonal matrix elements of H are not small and the standard perturbation theory may converge only after a careful choice of the unperturbed operator H_0.[4,5] Moreover, there is a conflict between an easy diagonalization of this H_0 and its "realistic" character, i.e., between the simplicity of the perturbation expansions and their convergence. Therefore, the nonsmallness of the perturbation $H-H_0$ usually requires the nonperturbative treatment of H.

For the simplest tridiagonal Hamiltonians H, the nonperturbative techniques of their diagonalization[1] range from the analytic theory of continued fractions[6] and analogies with the Mathieu functions to the general theory of moments[7] and Padé approximants.[8]

Any generalization of this mathematics to band matrices with $2t+1$ diagonals, $t>1$, encounters immediately a number of specific methodical difficulties.[3,4,9] Our present intention is to show that most of these difficulties disappear at least in the $t=2$ case of five diagonals.

The material is organized as follows. In Sec. II, we review the "matrix continued fraction" (MCF) approach[9] to diagonalization of the band matrices and emphasize that the MCF algorithm need not be employed in the numerical context only. Indeed, for the particular $t=2$ anharmonic-oscillator (AHO) example,[3] we are able to obtain also the explicit asymptotic form of the effective Hamiltonian (Appendix A) and/or to analyze its MCF convergence still by purely analytic means (Appendix B).

When applied to the general MCF quantity, this geometric "fixed-point" (FP) approach[10] generates the natural FP approximants. In the chain-model context, they define algebraically the approximate effective Hamiltonians in a large model space (Sec. III).

In Sec. IV, a systematic FP treatment of the higher-order corrections is suggested. For the Schrödinger equation (1.1), this enables one to define exactly the effective Hamiltonian by an infinite series in an arbitrary model space. This general FP expansion resembles perturbation theory but its algebraic construction has a more complicated character (Appendix C), especially for $t\geqslant 1$.

The particular $t=1$ (Sec. V A) and $t=2$ (Sec. V B) versions of the FP expansions are exceptional and simple: Their construction remains fully elementary. This makes the pentadiagonal chain model "exactly solvable" and formally close to the current tridiagonal Hamiltonians.

II. BAND-MATRIX HAMILTONIANS AND THE MCF ALGORITHM

A. Effective Hamiltonians and bound states

Let us start by a reformulation of the MCF formalism initiated by Graffi and Grecchi[3,9] and applicable to any Schrödinger-type eigenvalue problem:

$$\mathscr{H}(E)\psi = 0, \tag{2.1}$$

with the band or, equivalently,[9] block-tridiagonal "Hamiltonian"

$$\mathscr{H}(E) = \begin{pmatrix} A_0 & B_0 & & \\ C_1 & A_1 & B_1 & \\ & C_2 & A_2 & B_2 \\ & & & \ddots \end{pmatrix}, \tag{2.2}$$

$$\dim A_0 = M_0 \geqslant 1, \quad \dim A_k = M_k = t \geqslant 1, \quad k = 1,2,\ldots .$$

First, we introduce a model-space projector

$$P = P_{(M)} = |0\rangle\langle 0| + |1\rangle\langle 1| + \cdots + |d-1\rangle\langle d-1|,$$
$$d = M_0 + M_1 + \cdots + M_M.$$

In the spirit of Feshbach,[11] we may rewrite (2.1), i.e.,

$$P\mathscr{H}P\psi + P\mathscr{H}Q\psi = 0, \quad Q = 1-P,$$
$$Q\mathscr{H}P\psi + Q\mathscr{H}Q\psi = 0,$$

in the equivalent form

$$\mathscr{H}^{\text{eff}}_{(M)}P_\psi = 0, \quad \mathscr{H}^{\text{eff}}_{(M)} = P\mathscr{H}P - P\mathscr{H}Q(1/Q\mathscr{H}Q)Q\mathscr{H}P, \tag{2.3}$$
$$Q_\psi = -(1/Q\mathscr{H}Q)Q\mathscr{H}P_\psi,$$

where, obviously, the complexity of solution concentrates in

 0022-2488/84/102979-09$02.50

the construction of the effective matrix $\mathscr{H}^{\rm eff}_{(M)}$ itself.

In this formulation, the main MCF idea stems from the M-dependence of the exact effective Hamiltonians $\mathscr{H}^{\rm eff}_{(M)}(E)$,

$$\mathscr{H}^{\rm eff}_{(0)}(E)=G_0,\quad \mathscr{H}^{\rm eff}_{(1)}(E)=\begin{pmatrix}A_0 & B_0\\ C_1 & G_1\end{pmatrix},$$
$$\mathscr{H}^{\rm eff}_{(2)}(E)=\begin{pmatrix}A_0 & B_0 & \\ C_1 & A_1 & B_1\\ & C_2 & G_2\end{pmatrix},\ldots . \tag{2.4}$$

Indeed, when we choose the various dimensions $d=M_0+tM$ of the model space, the different $(t\times t)$-dimensional "effective" submatrices $G_0,G_1,\ldots$ become related precisely by the MCF recurrences

$$G_k=A_k-B_k(1/G_{k+1})C_{k+1},\quad k=0,1,\ldots . \tag{2.5}$$

These recurrent relations are to be initialized in accord with the standard variational technique [truncation of $\mathscr{H}(E)$], i.e., by $G_N=A_N$, $N\to\infty$. All the effective Hamiltonians (2.4) become exact in the MCF limit $N\to\infty$.

The dimension M_0 is variable. Its change may be used as a "regularization procedure" whenever we encounter the random zeros of $\det G_{k+1}$ at some $k=k_0\geqslant 1$ in (2.5) (an example may be found, e.g., in Ref. 12). In the "regular" cases, we may fix $M_0=1$ and get the simplest MCF form of the secular equation

$$(A_0)_{11}-\sum_{i,j=1}^{t}(B_0)_{1i}(G_1^{-1})_{ij}(C_1)_{j1}=0, \tag{2.6}$$

which is extremely suitable for a numerical search for the binding energies E. Quite formally, the inverse of the left-hand side expression (or $\det G_0$ in general) may be interpreted also as the MCF Green function.[3,9]

Concerning the wave functions ψ, we may notice that their MCF form

$$\mathbf{X}_k=D_kD_{k-1}\cdots D_1\mathbf{X}_0,\quad D_k=-(1/G_k)C_k, \tag{2.7}$$
$$(\mathbf{X}_k)_m=\langle kt+m-1|\psi\rangle,\quad m=1,2,\ldots,t,\quad k=1,2,\ldots,$$

follows from (2.3) with $k=M\neq 0$. It must be accompanied by the remaining $M=0$ requirement (2.3)

$$G_0\mathbf{X}_0=0. \tag{2.8}$$

Then, the projections (2.7) satisfy (2.1) at the physical energies only.

Vice versa, any initialization G_0 and $\mathbf{X}_0$ of the respective definitions (2.5) and (2.7) leads to the formal solution $\langle n|\psi\rangle$ provided that (2.8) holds. Then, the requirement $\|\psi\|<\infty$ must be imposed in a way illustrated again in Ref. 12 or in Sec. V A below (see alternative proof of Theorem 2).

B. Examples of the band-matrix Hamiltonians

Anharmonic oscillators (AHO) possess some interesting physical interpretations (e.g., in zero-dimensional field theory) and are often employed due to their technical simplicity.[3–5] In the quantum-mechanical context, we may represent their Hamiltonians

$$H=-\frac{d^2}{dr^2}+\frac{l(l+1)}{r^2}+r_2+\lambda r^{2t},$$
$$\lambda>0,\quad t=2,3,\ldots, \tag{2.9}$$

in the harmonic-oscillator (HO) basis $|n\rangle$ and get the Schrödinger equation (1.1) or (2.1) in the band-matrix form:

$$0=\sum_{n=m-t}^{m+t}[(2a_m-E)\delta_{mn}+\lambda\langle m|r^{2t}|n\rangle]\langle n|\psi\rangle, \tag{2.10}$$
$$a_m=2m+l+\tfrac{3}{2},\ m=0,1,\ldots .$$

The matrix elements are easily obtainable since $\langle m|r^{2t}|n\rangle=(T^t)_{mn}$, where

$$T=\begin{pmatrix}a_0 & b_0 & & \\ b_0 & a_1 & b_1 & \\ & b_1 & a_2 & b_2\\ & & \ddots & \end{pmatrix},\quad b_m=[(m+1)(m+l+\tfrac{3}{2})]^{1/2},$$

with $l=-1,0$ (parity in one dimension), $l=0,1,\ldots$ (angular momentum in three dimensions), or $l>-\frac{1}{2}$ (noninteger parameter simulating the r^{-2} component in the potential[12]).

The trivial choice of $t=1$ corresponds merely to the harmonic oscillator represented in a "wrong" basis. With

$$A_k=(2+\lambda)a_k-E,\quad B_k=\lambda b_k,\quad t=1,\quad k\geqslant 0, \tag{2.11}$$

this simplest tridiagonal chain model may be used as a methodical guide. In the present MCF context, we may immediately notice that the HO formula (2.5),

$$G_k=(2+\lambda)a_k-E-\cfrac{\lambda^2b_k^2}{(2+\lambda)a_{k+1}-E-\cdots}, \tag{2.12}$$

represents a classical continued fraction. It converges if and only if[6] $\lambda>-1$, i.e., in all cases of interest.

In the simplest nontrivial $t=2$ example, the matrix $\mathscr{H}(E)$ may be given the form (2.2) with $2k+M_0-2=n$ and

$$A_k=\begin{pmatrix}(2a_n-E)/\lambda+b_{n-1}^2+a_n^2+b_n^2 & (a_n+a_{n+1})b_n\\ (a_n+a_{n+1})b_n & (2a_{n+1}-E)/\lambda+b_n^2+a_{n+1}^2+b_{n+1}^2\end{pmatrix},$$
$$B_k=C_{k+1}^T=\begin{pmatrix}b_nb_{n+1} & 0\\ (a_{n+1}+a_{n+2})b_{n+1} & b_{n+1}b_{n+2}\end{pmatrix},\quad k\geqslant 0. \tag{2.13}$$

Similary, for any $t\geqslant 2$, the MCF numerical algorithm is easily implemented to produce the AHO energies.[3]

With $t<0$, the singularly anharmonic Hamiltonians (2.9) may also be converted into the $(2|t|+1)$-diagonal matrices $\mathscr{H}(E)$: It suffices to use the nonorthogonal basis functions $\langle r|n\rangle r^{|t|}$ suppressed near the origin. The simplest example with

$$V(r)=r^2+a^2r^{-4},\quad t=-2, \tag{2.14}$$

will be considered below.

The $(2t+1)$-diagonal forms of $\mathscr{H}(E)$ may be derived also from some nonpolynomial anharmonicities—a nontrivial example (fractionally anharmonic potentials) may be found in Ref. 12 or 13. The band-matrix form of $\mathscr{H}(E)$ may also be postulated *a priori*, as a phenomenological (possibly, nonlocal) interaction model.[13]

III. ASYMPTOTICALLY SMOOTH CHAIN MODELS

A. Large model space and the AHO example

Notice that the k-dependence of A_k and $B_k = C_{k+1}^T$ in (2.2) is often smooth. In general, the $k \gg 1$ asymptotic AHO estimates

$$A_k = \lambda k^t a(t)(1 + O(1/k)),$$
$$B_k = C_{k+1}^T = \lambda k^t b(t)(1 + O(1/k)) \tag{3.1}$$
$$a_{ij}(t) = \binom{2t}{t+i-j}, \quad b_{ij}(t) = \binom{2t}{i-j},$$
$$i,j = 1,2,\ldots,t, \quad t \geqslant 2,$$

contain the k-dependence in the common scalar factor only. Thus, putting $G_{N+1-k} = \lambda N^t g_k$, the nonlinear matrix mapping $G_{N+1-k} \to G_{N-k}$ may be approximated by its leading-order form

$$g_{k+1} = a - b(1/g_k)b^T, \quad g_1 = a. \tag{3.2}$$

As a consequence, the AHO MCF expansion

$$G_k = A_k - B_k\{A_{k+1} - B_{k+1} \times [1/(A_{k+2} - \cdots)]B_{k+1}^T\}^{-1}B_k^T \tag{3.3}$$

of the effective Hamiltonian (2.4) in a large model space may be replaced by its asymptotic modification $G_k^{(A)} = \lambda k^t g_\infty$, where

$$g_\infty = a - b\{a - b[1/(a - \cdots)]b^T\}^{-1}b^T \tag{3.4}$$

is a new, simplified MCF expansion with constant coefficients. This leads to results of the following type.

Theorem 1: In the asymptotically leading-order approximation, the effective $t = 2$ AHO Hamiltonians have the form

$$\mathscr{H}_{(M)}^{(A)} = \begin{pmatrix} A_0 & B_0 & & & \\ & \ddots & & & \\ & & C_{M-1} & A_{M-1} & B_{M-1} \\ & & & C_M & G_M^{(A)} \end{pmatrix}, \quad M \gg 1, \tag{3.5}$$

where $M_0 = t$ and

$$G_M^{(A)} = \lambda M^2 \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} + \tilde{\lambda} M \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + O(M), \quad 1 \ll \tilde{\lambda} \ll M. \tag{3.6}$$

Proof: Formula (3.4) is reducible to the quadratic equation

$$g_\infty = a - b g_\infty^{-1} b^T, \tag{3.7}$$

for $(t \times t)$-dimensional matrices g_∞. It defines the stationary value (fixed point, FP) of the mapping $g_k \to g_{k+1}$. One of the fixed points g_∞ must become the point of accumulation of the sequence g_k, $k \gg 1$ based on the physical (variational) initialization $g_1 = a$.

For $t = 2$, Eq. (3.7) may be solved easily and its solution is unique (Appendix A). The form of the correction term follows from Appendix B. Q.E.D.

An interesting feature of the preceding result is its independence of the MCF convergence. Although its proof may also be delivered (for g_k and $t = 2$, see Appendix B), the semistable character of the fixed point g_∞ indicates a possibility of divergence of the MCF matrices G_k, caused by the second-order corrections. This is the main short-coming and limitation of the original MCF method.[9,14]

B. Fixed-point approximation, its sign ambiguity, and perturbation-variational reinterpretation

For the general band-matrix Hamiltonian, we may choose $M \gg 1$ and expect that the effective Hamiltonian $\mathscr{H}_{(M)}^{\text{eff}}(E)$ is again a slowly varying function of the parameter M. In particular, if $A_k \sim A_{k+1}$ and $B_k = C_{k+1}^T \sim B_{k+1}$ in a long interval of indices $k \gg 1$, the fairly reliable fixed-point (FP) approximation F_k to $G_k \sim G_{k+1}$ may be generated along the same lines as in the AHO example, namely, via the quadratic equations of the type (3.7),

$$F_k = A_k - B_k(1/F_k)C_{k+1}. \tag{3.8}$$

These t^2 coupled polynomial equations for the t^2 matrix elements of the approximate effective Hamiltonian $\mathscr{H}_{(k)}^{\text{eff}}$ may be given also the symmetrized form

$$F_k\left(\frac{1}{B_k} + \frac{1}{C_{k+1}}\right)F_k - F_k\frac{1}{B_k}A_k - A_k\frac{1}{C_{k+1}}F_k + B_k + C_{k+1} = 0, \tag{3.9}$$

and solved in accord with the general method described here in Appendix C.

The algebraic root of the quadratic equation (3.9) need not be symmetric and contains an arbitrary pseudo-orthogonal $(t \times t)$-dimensional matrix S. Vice versa, the symmetric root $F = F^T$ contains only the discrete sign-ambiguity $S(t)$. Its removal is to be based on the MCF origin of the method in a way illustrated on the $t = 1$ example in Ref. 12 [$S(t) = \pm 1$] and on the simplest nontrivial $t = 2$ example with $S(t) = (\pm 1, \pm 1)$ in Appendices A and B. For $t \gg 1$, similar constructions become rather complicated—it is simpler to return to the original MCF iteration itself.[15]

For the reasonable values of t, an appropriate elimination of ambiguities will not be difficult whenever we succeed in finding an analog of (3.6) and dominant parameter $\tilde{\lambda}$ simulating the MCF origin of $F_k \sim G_k$. Of course, such an analysis depends on $\mathscr{H}(E)$. For a numerical illustration, we have chosen here the particular example (2.14) with $a = 0$, $t = 2$. The following results were obtained.

(1) The proof of Theorem 2 remains essentially the same. The parameter $\tilde{\lambda}$ may be treated as a variable accelerating the convergence in a trial and error way.

(2) The numerical test (Table I) shows that the practical acceleration of convergence is achieved in a large interval of $\tilde{\lambda} \in (-1, +\infty)$. Similar phenomena may be expected also in the more general cases.

TABLE I. Energy convergence for the various effective interaction corrections [for the harmonic-oscillator potential (2.14) with $a = 0$ and $t = 2$].

$\tilde{\lambda}$	-1.55[d]	-1.50[d]	-1.00[c]	0.00[b]	$\pm\infty$[a]	-2.00[c]
N						
10	2.542	2.711	3.143	3.291	3.422	
20	2.576	2.742	3.099	3.205	3.293	
30	2.629	2.775	3.080	3.167	3.237	
40	2.671	2.799	3.069	3.144	3.204	
50	2.703	2.818	3.061	3.129	3.182	
100	2.790	2.869	3.043	3.091	3.128	3.505
200	2.854	2.907	3.030	3.064	3.090	3.316
300	2.883	2.924	3.024	3.052	3.073	3.248
400	2.899	2.935	3.021	3.045	3.063	3.210
500	2.910	2.942	3.019	3.040	3.057	3.185

[a] $(N \times N)$-dimensional truncation.
[b] Leading-order effective Hamiltonian.
[c] Trial acceleration of convergence.
[d] The "antivariational" energy estimates.
[e] Bad choice—deceleration of convergence.

(3) There exists an interval $[\tilde{\lambda} \in (-1.55, -1.50)]$ such that the corresponding energies converge to the exact value $E = 3$ from below with the increasing cutoff parameter N. Such an "antivariational" phenomenon seems also quite general—it was observed in the $t = 1$ case in Ref. 12. As a perturbation-variational algorithm, it would give the best results, but its rigorous foundation (proof of convergence) as well as practical implementation (location of the corresponding interval of $\tilde{\lambda}$) seems an open problem at present.

IV. FIXED-POINT EXPANSION OF THE EXACT EFFECTIVE HAMILTONIAN

Higher-order corrections to the FP approximation $G_k \sim F_k^{(0)}$ need not always be negligible in the applications. An occurrence of large matrix elements in the corrections $G_k - F_k$ may indicate that the band matrix $\mathscr{H}(E)$ is not smooth enough outside of the chosen model space. Then, we must consider the new quantities defined by the subtraction of $F_k = F_k^{(0)}$,

$$G_k^{(1)} = G_k - F_k^{(0)}, \tag{4.1}$$

and treat them in a recurrent way again.

In analogy with the preceding section, the specification of mapping $G_{k+1}^{(1)} \to G_k^{(1)}$ [combination of Eqs. (2.5), (3.8), and (4.1)] and of its fixed points $F_k^{(1)}$ (common values of $G_{k+1}^{(1)}$ and $G_k^{(1)}$) may be repeated. This leads to the higher and higher corrections

$$G_k^{(i)} = G_k - F_k^{(0)} - F_k^{(1)} - \cdots - F_k^{(i-1)}, \tag{4.2}$$

mappings

$$G_k^{(i)} = A_k^{(i)} - B_k [1/(G_{k+1}^{(i)} + D_k^{(i)})] C_{k+1}, \tag{4.3}$$
$$A_k^{(i)} = A_k - D_{k-1}^{(i)},$$
$$D_k^{(i)} = F_{k+1}^{(0)} + F_{k+1}^{(1)} + \cdots + F_{k+1}^{(i-1)},$$

and definitions

$$F_k^{(i)} = A_k^{(i)} - B_k [1/(F_k^{(i)} + D_k^{(i)})] C_{k+1} \tag{4.4}$$

of the fixed points with $i = 1,2,\ldots$. For the reasonably smooth matrices $\mathscr{H}$, we may expect that such a scheme is convergent, $G_k^{(i)} \to 0$ for $i \to \infty$, in full analogy with the $t = 1$ example as investigated thoroughly in Ref. 10.

We may symmetrize Eq. (4.4) (cf. the $i = 0$ discussion). Without any difficulties, we may omit the lower as well as upper indices and get the equation

$$F(B^{-1} + C^{-1})F + F(C^{-1}D - B^{-1}A) + (DB^{-1} - AC^{-1})F = DB^{-1}A + AC^{-1}D - B - C, \quad B = C^T, \quad i \geqslant 0. \tag{4.5}$$

In accord with Appendix C, we may generate its algebraic solution whenever t is reasonably small.

For $i = 0$, the solution of (4.5) or (3.9) was made unique in the way indicated in Appendix B. Indeed, via a geometric interpretation of the mapping $G_{k+1} \to G_k$ in a vicinity of its fixed points $F_k = F_k[S(t)], S(t) = (\pm 1,\ldots, \pm 1)$, we may always specify the physical choice of $S(t)$ and $F_k^{\mathrm{phys}(0)}$ as corresponding to the variational initialization $G_N = A_N$, $N \gg 1$ of the auxiliary sequence G_k.

A priori, the unique specification of the $i > 0$ fixed points $F_k^{\mathrm{phys}(i)}$ also should be easy since their matrix elements may be expected to be small. More rigorously, this expectation is supported not only by the AHO-type examples, but also by the general structure of the method. When we rewrite (4.5) in the form

$$F_k^{(i)} = A_k^{(i)} \frac{1}{B_k^T} (\Delta_k^{(i)} + F_k^{(i)}) \frac{1}{F_k^{(i)} + D_k^{(i)}} B_k^T, \quad i \geqslant 1, \tag{4.6}$$

where $\Delta_k^{(i)} = F_{k+1}^{(i-1)} - F_k^{(i-1)}$, or

$$F(B^{-1} + C^{-1})F + F(C^{-1}D - B^{-1}A) + (DB^{-1} - AC^{-1})F = AC^{-1}\Delta + \Delta B^{-1}A, \quad i \geqslant 1, \tag{4.7}$$

we see that we may always choose the sets of signs $S(t) = S^{(i)}(t), i \geqslant 1$, in such a way that the resulting matrix $F^{(i)}$ is "minimal," $\|\Delta_k^{(i)}\| = O(\|F_k^{\mathrm{phys}(i)}\|^2)$.

Obviously, the norm of Δ measures the "smoothness" of $\mathscr{H}(E)$ and determines the rate of convergence of our final formula, namely, of the FP expansion

$$G_k = \sum_{i=0}^{\infty} F_k^{\mathrm{phys}(i)}. \tag{4.8}$$

Thus, we may conclude that this expansion will converge very quickly whenever the k-dependence of the submatrices A_k, B_k, and C_{k+1} happens to be monotonic or at least sufficiently "smooth." In all such cases we get an estimate

$$\|\Delta_k^{(i)}\| \sim (1/k)\|\Delta_k^{(i-1)}\|, \quad k \gg 1, \tag{4.9}$$

exhibiting a perturbation-type character of convergence in the dimension parameter $1/k$. Our general construction—diagonalization of the band-matrix Hamiltonians—is completed.

When we insert (4.8) into (2.4) with $k = M \gg 1$, we may easily generate the energies as well as the projections of wave functions. The details of this algebraic exercise may be found in Ref. 9.

V. APPLICATIONS

A. The simplest illustration-tridiagonal HO Hamiltonian

Theorem 2: For $i = 0$, the physical HO FP approximant to G_k reads

$$F_k^{(0)} = \tfrac{1}{2}[(2+\lambda)a_k - E + D_k], \quad k > k_0(E),$$

$$D_k = \{[(2+\lambda)a_k - E]^2 - 4\lambda^2 b_k^2\}^{1/2} > 0, \quad \lambda > -1.$$

Proof: The $i = 0$ quadratic equation (4.5) has two roots $F_{k(\pm)}^{(0)} = (1+\lambda/2)a_k - E/2 \pm D_k/2$. They are both real for $k > k_0(E)$ since

$$D_k^2 = (2+\lambda)^2 a_k^2 + O(k) - 4\lambda^2 k^2$$
$$= 16(\lambda+1)k^2 + O(k), \quad k \gg 1.$$

One of them is to be eliminated as unphysical.

In the $k \gg 1$ asymptotic region the mapping $G_{k+1} \to G_k$ with $G_k = \lambda k \gamma_k$ has the asymptotic form

$$\gamma_k = 2 + 4\lambda^{-1} - 1/\gamma_{k+1} + O(1/k), \quad k \gg 1.$$

Its geometric interpretation shows that this mapping is contractive for almost all initializations including $\gamma_N = A_N/\lambda N$ (Fig. 1). This implies that after a sufficient number of iterations, the sequence γ_k accumulates near the stable fixed point

$$\gamma_k \sim \gamma_k^{(+)} = 1 + 2\lambda^{-1}(1 + (1+\lambda)^{1/2}), \quad \lambda > -1$$

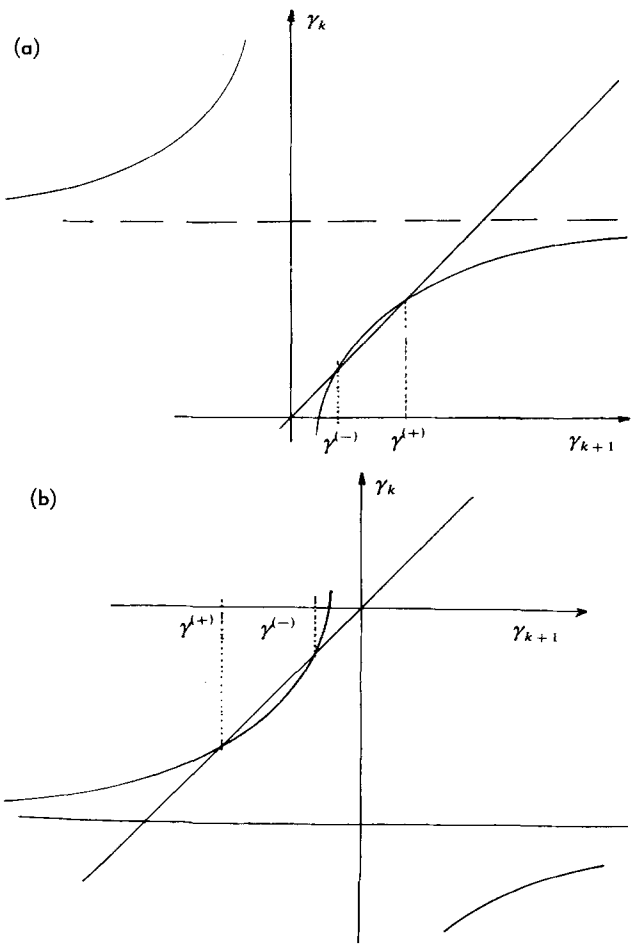


FIG. 1. Geometric interpretation of the HO mapping $G_{k+1} \to G_k$. (a) $\lambda = 1/\sinh^2 x > 0$, $\gamma^{(\pm)} = \exp(\pm 2x)$. (b) $\lambda = -1/\cosh^2 x > -1$, $\gamma^{(\pm)} = -\exp(\pm 2x)$.

(see Fig. 1 again). This establishes the convergence of the MCF expansion (2.12) and eliminates the $\gamma^{(-)}$ root (sign ambiguity) on the variational grounds. Q.E.D.

Alternative proof of theorem 2: We may omit the variational argument and start from the explicit solution

$$G_0 = 0, \quad \mathbf{X}_0 \neq 0$$

of the model-space condition (2.8) with $t = M_0 = 1$. When interpreted as a pair of initializations of the recurrences (2.5) and (2.7), respectively, it fixes the formal projections $\langle n|\psi\rangle$. The independent physical requirement

$$\|\psi\| < \infty$$

should fix the binding energies. Among the two roots $\gamma^{(\pm)}$ of the $i = 0$ equation (4.5), we have to pick out the one compatible with the convergence requirement

$$|C_k/G_k| = |1/\gamma_k| < 1, \quad k \gg 1$$

which eliminates the non-MCF root $\gamma^{(-)}$ as spurious again. Q.E.D.

In the $i = 1$ approximation it is sufficient to evaluate the quantities

$$\Delta_k^{(1)} = F_{k+1}^{(0)} - F_k^{(0)} = 2 + \lambda + [4(2+\lambda)^2(2k+l+\tfrac{3}{2})$$
$$- 4E(2+\lambda) - 4\lambda^2(2k+l+\tfrac{7}{2})]/(D_{k+1} + D_k)$$
$$= 2 + \lambda + 4(1+\lambda)^{1/2} + O(1/k),$$

and find the $i = 1$ solution of (4.5). After the appropriate insertions, we obtain the result

$$F_{k(\pm)}^{(1)} = +2k(1+\lambda)^{1/2}[-1 \pm (1 + O(1/k))],$$

$$F_k^{(1)\text{phys}} = F_{k(+)}^{(1)} = O(1), \quad k \gg 1.$$

It is compatible with the smallness requirement $F^{\text{phys}} \sim \Delta^{1/2}$ and does not increase with the increasing model space. The next corrections

$$\Delta_k^{(n+1)} = O(1/k^n), \quad n = 1,2,...,$$

become negligible for the large dimensions of the model space.

Our FP expansion converges as a geometric series in the HO example. This feature will be preserved also in the more complicated systems of a similar character, in spite of more complicated technicalities. For $t = 2$, all the related discussion preserves even a non-numerical character—this will be proved in the following final section.

B. Pentadiagonal Hamiltonians and their diagonalization by algebraic means

Theorem 3: A Schrödinger-type equation (2.1) with the pentadiagonal "Hamiltonian" $\mathscr{H}(E)$ may be represented by its finite-dimensional equivalents (2.3) + (2.4) with

$$G_k = Z^{-1} U \left[\sum_{i=0}^{\infty} \begin{pmatrix} x^{(i)} & y^{(i)} \\ y^{(i)} & z^{(i)} \end{pmatrix} \right] U Z^{-1}, \quad k \gg 1, \qquad (5.1)$$

where the diagonal Z and unitary matrix $U = U^T$, as well as the matrix elements $x^{(i)}$, $y^{(i)}$, and $z^{(i)}$ may be all defined by the elementary algebraic expressions.

Proof: First, let us notice that all the quadratic equa-

tions (4.5) contain the common i-independent factor $B^{-1}+C^{-1}$ ($\equiv$ matrix A of Appendix C). It is convenient to write $B=C^T$ and $1/B+1/B^T=ZRZ$ with the arbitrary diagonal matrix Z and real and symmetric R. Since the secular equation $\det(R-E)=0$ (ordinary quadratic equation) admits the algebraic solution, we may write

$$R=\begin{pmatrix}\cos\varphi & \sin\varphi\\ -\sin\varphi & \cos\varphi\end{pmatrix}\begin{pmatrix}E_+ & \\ & E_-\end{pmatrix}\begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix}.$$

To avoid the transpositions, we may insert here the identity matrices

$$\begin{pmatrix}1&0\\0&1\end{pmatrix}=\begin{pmatrix}0&1\\1&0\end{pmatrix}\begin{pmatrix}0&1\\1&0\end{pmatrix},$$

and get

$$B^{-1}+C^{-1}=ZU\begin{pmatrix}E_- & \\ & E_+\end{pmatrix}UZ, \tag{5.2}$$

$$U=\begin{pmatrix}\sin\varphi & \cos\varphi\\ \cos\varphi & -\sin\varphi\end{pmatrix}.$$

Using (5.2) and denoting

$$F=Z^{-1}U\begin{pmatrix}x&y\\y&z\end{pmatrix}UZ^{-1}, \tag{5.3}$$

we may simplify the structure of Eq. (4.5) when we insert $I=ZZ^{-1}$ and $I=UU$ whenever necessary. With respect to the possible changes of signs of $E_\pm$, we have to distinguish between the four cases, with $E_-=\pm g_1^2$ and $E_+=\pm g_2^2$. For the sake of simplicity, we shall consider here only one of them, say, $E_-=-g_1^2$ and $E_+=+g_2^2$, which gives

$$\begin{pmatrix}x&y\\y&z\end{pmatrix}\begin{pmatrix}-g_1^2&0\\0&g_2^2\end{pmatrix}\begin{pmatrix}x&y\\y&z\end{pmatrix}+\begin{pmatrix}x&y\\y&z\end{pmatrix}\begin{pmatrix}a&b\\c&d\end{pmatrix}$$
$$+\begin{pmatrix}a&c\\b&d\end{pmatrix}\begin{pmatrix}x&y\\y&z\end{pmatrix}+\begin{pmatrix}h_1&h_2\\h_2&h_3\end{pmatrix}=0, \tag{5.4}$$

in place of (4.5). In the AHO case, the coefficients are real (Appendix D).

The latter equation has obviously the form of three independent scalar relations

$$\begin{aligned}&-g_1^2x^2+2ax+g_2^2y^2+2cy+h_1=0,\\&-g_1^2y^2+2by+g_2^2z^2+2dz+h_3=0,\\&-g_1^2xy+bx+g_2^2yz+cz+(a+d)y+h_2=0.\end{aligned} \tag{5.5}$$

When we denote

$$\begin{aligned}h_4&=h_1+(a/g_1)^2-(c/g_2)^2,\\h_5&=h_3+(b/g_1)^2-(d/g_2)^2,\\h_6&=h_2+(ab/g_1^2)-(cd/g_2^2),\end{aligned}$$

and, in analogy with Appendix A,

$$\begin{aligned}&\alpha+\beta=g_1x-a/g,\quad \alpha-\beta=g_2y+c/g_2,\\&\gamma+\delta=g_1y-b/g_1,\quad \gamma-\delta=g_2z+d/g_2,\end{aligned} \tag{5.6}$$

we may distinguish again the various sign combinations. Omitting this discussion here, we obtain the two definitions

$$\beta=h_4/4\alpha,\quad \delta=h_5/4\gamma$$

accompanied by the third ordinary quadratic equation

$$\alpha\delta+\beta\gamma=h_6/2,$$

i.e., an equation analogous to (A5),

$$h_5\rho+h_4/\rho=2h_6,\quad \rho=\alpha/\gamma,$$

which defines the two values of the ratio

$$\rho=\rho_\pm=(1/h_5)[h_6\pm(h_6-h_4h_5)^{1/2}]. \tag{5.7}$$

Finally, an analog of (A6),

$$(g_2-g_1\rho_\eta)\gamma^2+(bg_2/g_1+cg_1/g_2)\gamma$$
$$+\tfrac{1}{4}(g_2h_5+g_1h_4/\rho_\eta)=0,\quad \eta=\pm1,$$

represents a compatibility of the two independent definitions of y in (5.6) and defines the four values of γ:

$$\gamma=\gamma_{\epsilon\eta}=\frac{1}{2(g_2-g_1\rho_\eta)}\Big(-\frac{bg_2}{g_1}-\frac{ag_1}{g_2}+\epsilon[a^2+d^2$$
$$+2bc+g_1^2h_1-g_2^2h_3+2g_1g_2\eta(h_6^2-h_4h_5)^{1/2}\Big)^{1/2}\Big],$$
$$\epsilon,\eta=\pm1. \tag{5.8}$$

The backward insertions are then trivial and give

$$z=z_{\epsilon\eta}=-d/g_2^2+\gamma_{\epsilon\eta}/g_2-h_5/4g_2\gamma_{\epsilon\eta}, \tag{5.9}$$

etc. We may summarize that the two-dimensional quadratic equation (4.5) is completely reducible to the sequence of a few one-dimensional ones. Q.E.D.

The algebraic construction in the preceding proof is straightforward and elementary. It is a little bit cumbersome and lengthy in practice so that we may assume that the energy- and coupling-dependence of the FP expansion (5.1) will be generated by an appropriate symbolic-manipulation procedure in most cases. In a large model space, we may use also the asymptotic expansions as a guide to the simplifications.

VI. SUMMARY

We have considered the class of physical models based on the chaining hypothesis[1] and described by the multidiagonal Hamiltonians. We have shown that at least some of these models may admit a reasonably simple mathematical treatment. Our conclusions are of the following three types.

(1) As a reformulation of the relevant physical information (interaction between the neighboring basis states only), our FP method is an alternative to the standard perturbation theories. As a guide how to construct the effective Hamiltonians in a systematic way, it may even be given sometimes the direct perturbative interpretation. In our anharmonic oscillator example, the small parameter characterizing the convergence of our FP series may be identified with the inverse dimension $1/d$.

(2) As a somewhat nonstandard algebraic formalism the present method leads to the new explicit, recurrent, or asymptotic relations for the wave function and/or for the effective interaction. It is based on the special block-tridiagonal structure of the Hamiltonian and on its geometric fixed-point reinterpretation. This may specify, e.g., the limitations and/or possible extensions of the classical perturbative expansions. Furthermore, in contrast to the standard numerical or continued fractional methods (and their results equivalent to the present $i=0$ approximation), the corrections to the effective Hamiltonian have an additive character. This is another analogy with the perturbation theory.

(3) As a practical numerical or seminumerical algorithm, the present FP approach enables one to save significantly the computer time. Indeed, besides the acceleration of convergence of the ordinary truncation eigenvalue method, the complementary "antivariational" lower bounds for energies were obtained in the illustrative example. The existence of such a "sandwiching of eigenvalues" seems a more general phenomenon[12] to be investigated more thoroughly in the future.

APPENDIX A: ASYMPTOTIC UNIQUENESS OF THE AHO FP APPROXIMANT

Lemma 1: The fixed point of the $t=2$ mapping (3.2) is unique.

Proof: The mapping $g_k(=1/f_k)\to g_{k+1}(=1/f_{k+1})$, i.e.,

$$f_{k+1}=\left[\begin{pmatrix}6&4\\4&6\end{pmatrix}-\begin{pmatrix}1&0\\4&1\end{pmatrix}f_k\begin{pmatrix}1&4\\0&1\end{pmatrix}\right]^{-1} \tag{A1}$$

is assumed to have a fixed point $f=f_k=f_{k+1}$. Putting

$$f=U\eta U^T,\quad U=\frac{1}{\sqrt{2}}\begin{pmatrix}1&1\\1&-1\end{pmatrix},\quad \eta=\begin{pmatrix}A&B\\B&D\end{pmatrix},$$

we obtain its symmetrized ($f=f^T$) algebraic definition

$$\eta\begin{pmatrix}6&0\\0&-2\end{pmatrix}\eta+\eta\begin{pmatrix}10&-20\\4&-6\end{pmatrix}+\begin{pmatrix}10&4\\-20&-6\end{pmatrix}\eta+\begin{pmatrix}-2&0\\0&6\end{pmatrix}=0,$$

which is equivalent to the three coupled scalar quadratic equations

$$\begin{aligned}&(3A+5)^2=3(B-2)^2+16,\\&(3B-10)^2=3(D+3)^2+64,\\&(3A+5)(3B-10)=3(B-2)(D+3)-32.\end{aligned} \tag{A2}$$

In terms of the linear combinations of the four new (primed) variables

$$\begin{aligned}&3A+5=\epsilon_1(A'+B'),\quad \sqrt{3}(B-2)=\epsilon_2(A'-B'),\\&3B-10=\epsilon_3(D'+C'),\quad \sqrt{3}(D+3)=\epsilon_4(D'-C'),\\&\epsilon_i=\pm 1,\quad i=1,2,3,4,\end{aligned} \tag{A3}$$

we may reduce the first two items in (A2) to the definitions

$$B'=4/A',\quad C'=16/D'. \tag{A4}$$

Considering the third condition in (A2) with

$$\epsilon_1=\epsilon_2=\epsilon_3=\epsilon_4=+1$$

for the time being, we get

$$2A'/D'+D'/2A'=-2, \tag{A5}$$

with the unique and doubly degenerate real solution $D'=-2A'$.

When we compare the second and the third item in (A3), we see that the original parameter B is defined in two ways now. Their compatibility requires

$$\sqrt{3}(A'-4/A')=4-(2A'+8/A'), \tag{A6}$$

and fixes the remaining freedom—this is again a quadratic equation with the degenerate solution. Hence, we have

$$\begin{aligned}&A'=4-2\sqrt{3},\quad B'=4+2\sqrt{3},\\&D'=-8+4\sqrt{3},\quad C'=-8-4\sqrt{3}.\end{aligned} \tag{A7}$$

Now, let us vary the signs ϵ_2 (by an interchange of A' and B', i.e., $A'\leftrightarrow B'$), ϵ_4 (interchanging $D'\leftrightarrow C'$), ϵ_1 (by the transformation $A'\to -B'$, $B'\to -A'$), or ϵ_3 (changing $D'\to -C'$ and $C'\to -D'$). Obviously, we get

$$\begin{aligned}&A'=4\epsilon_1-2\epsilon_2\sqrt{3},\quad B'=4\epsilon_1+2\epsilon_2\sqrt{3},\\&C'=-8\epsilon_3-4\epsilon_4\sqrt{3},\quad D'=-8\epsilon_3+4\epsilon_4\sqrt{3}.\end{aligned} \tag{A8}$$

Finally, the insertion into (A2) implies that $3A+5=8\epsilon_1^2=8$, etc., so that the resulting solution is unique,

$$A=1,\quad B=-2,\quad D=5. \tag{A9}$$

Q.E.D.

APPENDIX B: MCF CONVERGENCE IN THE AHO-TYPE CASES

Lemma 2: The matrix continued fraction

$$f=\left\{\begin{pmatrix}6&4\\4&6\end{pmatrix}-\begin{pmatrix}1&0\\4&1\end{pmatrix}\left[\begin{pmatrix}6&4\\4&6\end{pmatrix}-\cdots\right]^{-1}\begin{pmatrix}1&4\\0&1\end{pmatrix}\right\}^{-1} \tag{B1}$$

is convergent and equal to

$$\begin{pmatrix}1&-2\\-2&5\end{pmatrix}.$$

Its finite approximants have the form

$$\frac{1}{2}\begin{pmatrix}1&1\\1&-1\end{pmatrix}\left[\begin{pmatrix}5&2\\2&1\end{pmatrix}+\begin{pmatrix}\alpha&-\xi\\-\xi&\gamma\end{pmatrix}\right]^{-1}\begin{pmatrix}1&1\\1&-1\end{pmatrix}, \tag{B2}$$

with $5\geqslant\alpha\geqslant\xi\geqslant\gamma>0$.

Proof: Equation (B1) or (3.2) may be given the form

$$\begin{aligned}&\tilde{\varphi}=\begin{pmatrix}10&0\\0&2\end{pmatrix}-\begin{pmatrix}3&2\\-2&-1\end{pmatrix}\varphi^{-1}\begin{pmatrix}3&-2\\2&-1\end{pmatrix},\\&\tilde{\varphi}=Ug_{k+1}U=\begin{pmatrix}5&2\\2&1\end{pmatrix}+\begin{pmatrix}\tilde{\alpha}&-\tilde{\xi}\\-\tilde{\xi}&\tilde{\gamma}\end{pmatrix},\\&\varphi=Ug_kU=\begin{pmatrix}5&2\\2&1\end{pmatrix}+\begin{pmatrix}\alpha&-\xi\\-\xi&\gamma\end{pmatrix}.\end{aligned} \tag{B3}$$

This coincides with the rational mapping

$$\begin{aligned}&\tilde{\alpha}=(\alpha+8\xi+16\gamma+5\rho)/(1+\epsilon),\\&\tilde{\xi}=(\xi+4\gamma+2\rho)/(1+\epsilon),\\&\tilde{\gamma}=(\gamma+\rho)/(1+\epsilon),\\&\tilde{\rho}=\rho/(1+\epsilon),\\&\tilde{\epsilon}=(\alpha+12\xi+37\gamma+19\rho)/(1+\epsilon),\end{aligned} \tag{B4}$$

with the two additional auxiliary variables

$$\rho=\alpha\gamma-\xi^2=\det\left(\varphi-\begin{pmatrix}5&2\\2&1\end{pmatrix}\right),\ \epsilon=\alpha+4\xi+5\gamma+\rho.$$

All the relations (B4) have the same form

$$\tilde{x}=(x+v_x)/(U_xx+V_x+1), \tag{B5}$$

with

$$v_\alpha = 8\xi + 16\gamma + 5\rho, \quad v_\xi = 4\gamma + \rho, \quad v_\gamma = 0 = v_\rho,$$
$$v_\epsilon = 8\xi + 32\gamma + 18\rho,$$
$$U_\alpha = 1, \quad U_\xi = 4, \quad U_\gamma = 5, \quad U_\rho = 1 = U_\epsilon,$$
$$V_\alpha = 4\xi + 5\gamma + \rho, \quad V_\xi = \alpha + 5\gamma + \rho,$$
$$V_\rho = \alpha + 4\xi + 5\gamma,$$
$$V_\gamma = \alpha + 4\xi + \rho, \quad V_\epsilon + 0.$$

They all possess also a geometric interpretation similar to that given in Fig. 1. We may observe the following.

(i) $\alpha = \xi = \gamma = 0$ is a fixed point of the mapping (B3).

(ii) The numerical test (Table II) confirms the convergence to the zero fixed point. In particular, the trivial initialization of the original quantities f is reflected here by the initial choice of a "correction" $\alpha = 5$, $\xi = 2$, and $\gamma = 1$.

(iii) From any initial set of the positive values α, ξ, and γ such that also $\rho > 0$, we get $\epsilon > 0$ and $\tilde{\alpha} > 0$, $\tilde{\xi} > 0$, $\tilde{\gamma} > 0$ and $\tilde{\rho} > 0$: The positivity of parameters is preserved by the mapping [Eq. (B4)].

(iv) We see from Fig. 2 that each sequence $x = (\alpha, \xi, \gamma, \rho,$ or $\epsilon)$ should independently accumulate near its stable fixed point

$$x_{(+)} = [-V_x + (V_x^2 + 2U_x v_x)^{1/2}]/2U_x, \tag{B6}$$

the value of which depends on the remaining parameters.

(v) In more detail, Eqs. (B3) and (B6) imply the following.

(a) The sequence of ρ's is monotonic and decreases to zero.

(b) For a sufficiently small ρ, we may partially "linearize" Eq. (B3) by omitting ρ's and see that γ starts to approach $\gamma_{(+)}$ in a monotonic way [Fig. 2(a)].

(c) Similar procedure ("linearization") may be then repeated for ξ and also for α or ϵ. From Eq. (B6) we see that the stationary values $x_{(+)}$ become strictly ordered as soon as $\rho \ll 1$. We get

$$\gamma \sim \gamma_{(+)} \sim [-V_\gamma + (V_\gamma^2 + 20\rho)^{1/2}]/10 \gtrsim (\rho/5)^{1/2} \gg \rho,$$
$$\xi \sim \xi_{(+)} \sim [-V_\xi + (V_\xi^2 + 64\gamma)^{1/2}]/8 \gtrsim \sqrt{\gamma} \gg \gamma, \tag{B7}$$
$$\alpha \sim \alpha_{(+)} [-V_\alpha + (V_\alpha^2 + 4v_\alpha)^{1/2}]/2 \gtrsim (8\xi)^{1/2} \gg \xi.$$

(d) The same type of ordering takes place from the very beginning, even for large ρ's (cf. Table II); and vice versa, if the ordering of the deviations is violated by our choice of another initialization, the overall convergence would be slowed down. Hence, the nonlinearity of the original relations (B3) is essential and a rate of the overall convergence is controlled by the convergence of α's or ϵ's,

$$\tilde{\alpha} = \alpha/(1+\alpha), \quad \tilde{\epsilon} = \epsilon/(1+\epsilon), \tag{B8}$$

in the simplest approximation. This completes the proof.

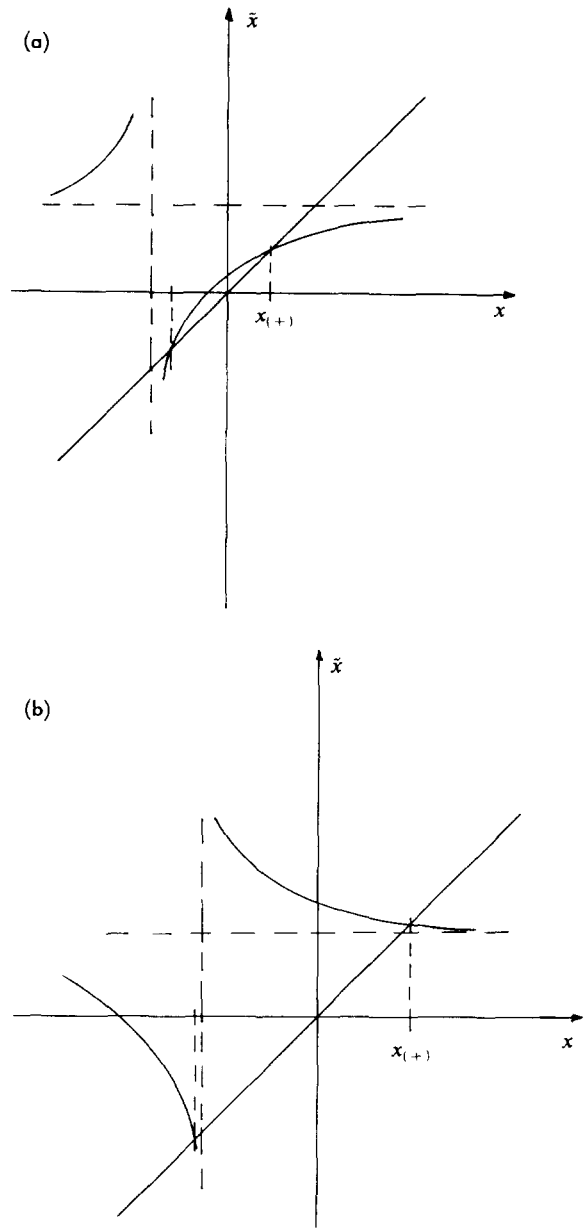


FIG. 2. Two possible patterns of convergence $x \to x_{(+)}$ in Appendix B. (a) $v < (V+1)/U$. (b) $v > (V+1)/U$.

TABLE II. Sample of the MCF convergence [Eq. (B4)].

M	γ_M	ξ_M	α_M	M	γ_M	ξ_M	α_M
0	1.000	2.000	5.000	5	0.003 4	0.0419	0.6701
1	0.100	0.400	2.100	10	0.000 5	0.0124	0.3642
2	0.029	0.171	1.362	15	0.000 1	0.0058	0.2501
3	0.012	0.095	1.012	20	0.000 08	0.0034	0.1905
4	0.006	0.061	0.806	25	0.000 04	0.0022	0.1538

APPENDIX C: QUADRATIC EQUATION FOR MATRICES AND ITS SOLUTION

For the $(t\times t)$-dimensional matrix F, let us consider the equation

$$FAF+\tfrac{1}{2}(BF+FB^{T})+C=0, \tag{C1}$$

where $A=A^{T}$, $C=C^{T}$, and the symmetric root $F=F^{T}$ is to be determined by the algebraic means.

We shall assume that there exists a matrix D such that $B=2DA$. Then, obviously, we may write

$$(F+D)A(F+D^{T})=DAD^{T}-C, \tag{C2}$$

which is equivalent to (C1).

Next, the real and symmetric matrices may be diagonalized by some orthogonal matrices U and W,

$$\begin{aligned} &A=UaU^{T},\quad DAD^{T}-C=WbW^{T},\\ &a=\epsilon\alpha^{2},\quad b=\eta\beta^{2}, \end{aligned} \tag{C3}$$

$$\epsilon=\begin{pmatrix}\pm 1 & & \\ & \pm 1 & \\ & & \ddots\end{pmatrix},\quad \eta=\begin{pmatrix}\pm 1 & & \\ & \pm 1 & \\ & & \ddots\end{pmatrix}.$$

Hence, with an arbitrary matrix S which satisfies the pseudo-orthogonality relations

$$S\epsilon S^{T}=\eta, \tag{C4}$$

and may be constructed in accord with the underlying group theory in various ways, we obtain the general solution of (C1) in the form

$$F=-D+W\beta S\alpha^{-1}U^{T}. \tag{C5}$$

Finally, up to the possible changes of signs, the $t(t-1)/2$ independent matrix elements of S become fixed by the $t(t-1)/2$ symmetry requirements $F=F^{T}$.

The above construction may be illustrated by the $t=2$ example of Appendix A once more. We have

$$\begin{aligned} &U=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\quad a=\begin{pmatrix}6 & 0\\ 0 & -2\end{pmatrix},\\ &W=\frac{1}{\sqrt{5}}\begin{pmatrix}1 & 2\\ -2 & 1\end{pmatrix},\quad b=\begin{pmatrix}\frac{160}{3} & 0\\ 0 & 0\end{pmatrix}, \end{aligned} \tag{C6}$$

so that the general solution F has the form (C5),

$$\begin{aligned} F=&\begin{pmatrix}-\frac{5}{3} & 2\\ \frac{10}{3} & -3\end{pmatrix}\\ &+8\begin{pmatrix}\rho_1\cosh\xi/6, & \rho_2\sinh\xi/2\sqrt{3}\\ -\rho_1\cosh\xi/3, & -\rho_2\sinh\xi/\sqrt{3}\end{pmatrix}, \end{aligned} \tag{C7}$$

$$\rho_{1,2}=\pm 1,\quad \xi\in(0,\infty).$$

It is to be complemented by the only symmetry requirement,

$$2\rho_1\cosh\xi+\rho_2\sqrt{3}\sinh\xi=1, \tag{C8}$$

which eliminates the variable ξ.

The solution of (C8) is straightforward and gives

$$e^{\xi}=1/(2\rho_1+\rho_2\sqrt{3}).$$

Hence, the resulting formula $\cosh\xi=2\rho_1$ implies that $\rho_1=1$ and, finally, $\rho_2=-\operatorname{sgn}\xi$. After the insertion into (C7), we see that $F=F^{T}$ is unique and equal to the solution obtained in Appendix A.

APPENDIX D: AHO EXAMPLE AND EQ. (5.4)

(i) The AHO $t=2$ example may be given an explicit form without any computer—using (2.13) together with the definition (5.2) and

$$Z=b_{n+1}^{-1/2}\begin{pmatrix}1/b_n & 0\\ 0 & 1/b_{n+2}\end{pmatrix},$$

$$R=\begin{pmatrix}2b_n & -(a_{n+1}+a_{n+2})\\ -(a_{n+1}+a_{n+2}) & 2b_{n+2}\end{pmatrix}.$$

The simple prescriptions $[n=n(k)=2k+M_0-2]$

$$\begin{aligned} &E_{\pm}=b_n+b_{n+2}\\ &\qquad\pm[(b_{n+2}-b_n)^2+(a_{n+1}+a_{n+2})^2]^{1/2},\\ &\tan 2\varphi=-(a_{n+1}+a_{n+2})/(b_{n+2}-b_{n-1}),\\ &\vdots \end{aligned}$$

specify the explicit AHO form of the FP series (5.1).

(ii) We may easily verify that $E_{+}>0$ and $E_{-}<0$ since the corresponding expression

$$(2n+l+\tfrac{9}{2})^4-(n+1)(n+3)(n+l+\tfrac{3}{2})(n+l+\tfrac{7}{2})$$

is a positive and monotonically increasing function of k and l in the whole interval of interest (say, for $k\geqslant 0$ and $l\geqslant -\frac{3}{2}$). Hence, g_1 and g_2 of Eq. (5.4) are both real. The value of the angle φ is also a monotonically decreasing function of k and l such that $\pi/4<\varphi<\pi/2$. On the boundaries, $\tan\varphi=(1+\sqrt{7})/\sqrt{6}$ for $k=0$ and $l=-\frac{3}{2}$, and $\varphi\to\pi/4$ for $k\to\infty$ and/or $l\to\infty$.

(iii) The AHO solution admits the alternative forms of denotation, e.g.,

$$\begin{aligned} &U=\sigma_1\cos\psi+\sigma_2\sin\psi,\quad \psi=\varphi-\pi/4,\\ &\sigma_1=\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix},\quad \sigma_2=\frac{1}{\sqrt{2}}\begin{pmatrix}1 & -1\\ -1 & -1\end{pmatrix},\\ &\sigma_1^2=\sigma_2^2\simeq I,\quad \sigma_1\sigma_2=-\sigma_2\sigma_1. \end{aligned}$$

This will not be discussed here.

[1] D. W. Bullett, R. Haydock, V. Heine, and M. J. Kelly, *Solid State Physics*, edited by H. Ehrenreich, F.Seitz, and D. Turnbull (Academic, New York, 1980), Vol. 35, and references contained therein.
[2] H. Feshbach, A. Kerman, and S. Koonin, Ann. Phys. (NY) **125**, 429 (1980).
[3] S. Graffi and V. Grecchi, Lett. Nuovo Cimento **12**, 425 (1975).
[4] M. Znojil, Phys. Rev. D **24**, 903 (1981).
[5] I. G. Halliday and P. Suranyi, Phys. Rev. D **21**, 1529 (1980).
[6] H. S. Wall, *Analytic Theory of Continued Fractions* (Van Nostrand, New York, 1948).
[7] N. I. Akhiezer, *The Classical Moment Problem* (Oliver and Boyd, London, 1955).
[8] G. A. Baker, *Essential of Padé Approximants* (Academic, New York, 1975).
[9] M. Znojil, J. Math. Phys. **21**, 1629 (1980).
[10] M. Znojil, J. Phys. A: Math. Gen. **13**, 2375 (1980), Appendix.
[11] H. Feshbach, Ann. Phys. (NY) **5**, 357 (1958).
[12] M. Znojil, J. Phys. A: Math. Gen. **16**, 279 (1983): 293 (1983).
[13] R. R. Whitehead, A. Watt, G. P. Flessass, and M. A. Nagarajan, J. Phys. A: Math. Gen. **15**, 1217 (1982).
[14] M. Znojil, Phys. Rev. D **26**, 3750 (1982).
[15] D. F. Scofield, Phys. Rev. Lett. **29**, 811 (1972).

A rigorous derivation of the "miracle" identity of three-dimensional inverse scattering

Margaret Cheney
Department of Mathematics, Stanford University, Stanford, California 94305

(Received 13 February 1984; accepted for publication 4 May 1984)

The large-energy asymptotic behavior of scattering solutions of the three-dimensional time-dependent Schrödinger equation is investigated. The second term of the expansion leads to the "miracle" of Newton's three-dimensional inverse scattering theory.

PACS numbers: 03.65.Nk

I. INTRODUCTION

We consider the time-independent Schrödinger equation in three dimensions

$$\Delta\psi + k^2\psi - V\psi = 0, \tag{1.1}$$

where Δ is the Laplacian, k is a real scalar (k^2 is the energy), and the potential $V(x)$ is a real-valued function that decays at infinity. Scattering solutions of (1.1) are solutions of the Lippmann–Schwinger equation[1]

$$\psi(k,e,x) = \exp(ike\cdot x) - \int (4\pi|x-y|)^{-1} \times \exp(ik|x-y|)V(y)\psi(k,e,y)dy, \tag{1.2}$$

where x and y are vectors in R^3 and e is a unit vector in R^3. We shall investigate the large-k behavior of these scattering solutions.

The chief goal of previous work on high-energy Schrödinger scattering has been to find approximate but tractable expressions for the scattering amplitude. The most popular approximations are the Born, the WKB, and the eikonal (Glauber) approximations. These are discussed in Ref. 1.

Our objective here is to find and use a particular high-energy (large-k) asymptotic expansion of the scattering wave function ψ. It is well known[2] that for large energy ψ behaves like a plane wave; we find that the second term in the expansion is (Theorem)

$$\psi(k,e,x) = \exp(ike\cdot x) \times\left[1 + (2ik)^{-1}\int_0^\infty V(x-re)dr + o(k^{-1})\right]. \tag{1.3}$$

Fourier transformation in k of the solution ψ gives us a solution u of the plasma wave equation

$$\Delta u - \left(\frac{\partial^2}{\partial t^2}\right)u - Vu = 0. \tag{1.4}$$

Fourier transformation of the asymptotic expansion (1.3), moreover, tells us (Corollary) that

$$u(t,e,x) = \delta(t - e\cdot x) - \frac{1}{2}\int_0^\infty V(x-re)dr\, H(t-e\cdot x) + h(t,e,x). \tag{1.5}$$

where H is the Heaviside function and h is continuous. Equation (1.5) is the frequently used[3,4] "progressing wave expansion" for solutions of (1.4). We will show that (1.5) is valid under mild hypotheses.

The main point is that (1.5) gives a rigorous derivation of a key equation appearing in Newton's three-dimensional inverse scattering theory.[2,5,6] We shall denote by $\eta(t,e,x)$ the Fourier transform in k of $\psi(k,e,x)\exp(ike\cdot x) - 1$. The function η arises in Newton's theory as the solution of the Marchenko equation, which is an integral equation whose input is the measured scattering data. The potential V can then be obtained from η with the help of Eq. (1.5) as follows. As we will see (Corollary), Eq. (1.5) tells us that the limit $\eta(0^+,e,x)$ is precisely

$$-\frac{1}{2}\int_0^\infty V(x-re)dr.$$

We then use the fact that $e\cdot\nabla_x V(x-re) = -(\partial/\partial r)V(x-re)$ to obtain the equation

$$V(x) = -2e\cdot\nabla_x\eta(0^+,e,x). \tag{1.6}$$

This equation is called the "miracle" because the right side of (1.6) depends on e, whereas the left side does not; that the solution η of Newton's Marchenko equation should have this property does indeed seem miraculous. In Newton's theory, this miracle is actually a consistency requirement on the scattering data; it is a characterization of admissible scattering data.[5]

II. THEOREMS AND PROOFS

We shall write $\gamma(k,e,x) = \psi(k,e,x)\exp(-ike\cdot x)$. We denote by L^1 and L^2 the spaces of integrable and square-integrable functions, respectively.

Lemma 1: Suppose that V is bounded and integrable with $|x-y|^{-1}V$ uniformly bounded in L^1. Then for k sufficiently large, $\nabla\gamma$ is uniformly bounded.

Proof: We multiply the Lippmann–Schwinger equation [Eq. (1.2)] by $\exp(-ike\cdot x)$, obtaining

$$\gamma(k,e,x) = 1 - \int (4\pi|x-y|)^{-1}\exp(ik[|x-y| - e\cdot(x-y)])V(y)\gamma(k,e,y)dy. \tag{2.1}$$

Next we differentiate (2.1) with respect to x. Mollifier techniques similar to those used, for example, in Ref. 7 allow us to differentiate under the integral sign. We note that the x and y derivatives of

$$(|x-y|)^{-1}\exp(ik[|x-y| - e\cdot(x-y)])$$

differ only by a minus sign; accordingly, we replace ∇_x by

 0022-2488/84/102988-03$02.50

$-\nabla_y$ and integrate by parts. We therefore have

$$\nabla\gamma(k,e,x) = \int (-4\pi|x-y|)^{-1}\exp(ik\,[|x-y| - e\cdot(x-y)])\nabla V(y)\gamma(k,e,y)dy - \int (-4\pi|x-y|)^{-1}\exp(ik\,[|x-y| - e\cdot(x-y)])V(y)\nabla\gamma(k,e,y)dy. \quad (2.2)$$

After multiplying (2.2) by $|V(x)|^{1/2}$, we can write it as

$$|V|^{1/2}\nabla\gamma = (I-K)^{-1}\alpha, \quad (2.3)$$

where

$$Kf(x) = \int (4\pi|x-y|)^{-1} \times \exp(ik\,[|x-y| - e\cdot(x-y)])|V(x)|^{1/2} \times V(y)|V(y)|^{-1/2}f(y)dy, \quad (2.4)$$

and

$$\alpha(k,e,x) = |V(x)|^{1/2}\int (-4\pi|x-y|)^{-1} \times \exp(ik\,[|x-y| - e\cdot(x-y)]) \times \nabla V(y)\gamma(k,e,y)dy.$$

Our hypotheses imply that α is an L^2 function whose norm is bounded uniformly in e and in k for k large. Because $(I-K)$ is invertible for large k (see Ref. 2), the function $|V|^{1/2}\nabla\gamma$ is also an L^2 function whose norm is uniformly bounded for large k. Moreover, an application of the Schwarz inequality to (2.2) shows that for large k, $\nabla\gamma$ is uniformly bounded. Q.E.D.

Lemma 2: Suppose V belongs to $L^1\cap L^2$, and suppose that for some x_0, both $|V|$ and $|\nabla V|$ are bounded by $F(|x-x_0|)$, where F is a positive function satisfying $\int_0^\infty F(t)dt < \infty$. Then for sufficiently large k, there is a constant c such that the following estimate holds:

$$|\gamma(k,e,x) - 1| \leqslant ck^{-1}. \quad (2.5)$$

Proof: We write the integral in Eq. (2.1) in polar coordinates with $x-y = (r,\theta,\phi)$:

$$\gamma = 1 - (4\pi)^{-1}\int_0^\infty\int_0^\pi \exp[ikr(1-\cos\theta)] \times \int_0^{2\pi} V(x-(r,\theta,\phi))\gamma(k,e,x-(r,\theta,\phi))d\phi \times \sin\theta\, d\theta\, r\, dr. \quad (2.6)$$

We make the substitution $u = 1-\cos\theta$, and integrate by parts in the u integral. We thus obtain

$$\gamma = 1 - (2ik)^{-1}\int_0^\infty \exp(2ikr)V(x-re)\gamma(k,e,x-re)dr + (2ik)^{-1}\int_0^\infty V(x-re)\gamma(k,e,x-re)dr + (4\pi ik)^{-1}\int_0^\infty\int_0^2 \exp(ikru)\left(\frac{\partial}{\partial u}\right) \times \int_0^{2\pi} V(x-(r,u,\phi))\gamma(k,e,x-(r,u,\phi))d\phi\, du\, dr. \quad (2.7)$$

The hypotheses imply that each of the integrals on the right side of (2.7) is uniformly bounded. Q.E.D.

Lemma 3: Suppose V is bounded and integrable, and suppose that for some x_0, the three functions $|V|$, $|\nabla V|$, and $|\nabla V|$ are all bounded by $F(|x-x_0|)$, where F is a positive function satisfying $\int_0^\infty F(t)dt < \infty$. Then for k sufficiently large, $\Delta\gamma$ is uniformly bounded.

Proof: The proof is similar to the proof of Lemma 1.

Theorem: Suppose that V is bounded and integrable, and suppose that for some x_0, the three functions $|V|$, $|\nabla V|$, and $|\nabla V|$ are all bounded by $F(|x-x_0|)$, where F is a positive function satisfying $\int_0^\infty F(t)dt < \infty$. Then

$$\gamma(k,e,x) - 1 = (2ik)^{-1}\int_0^\infty V(x-re)dr + g(k,e,x), \quad (2.8)$$

where for k sufficiently large and for ϵ with $0<\epsilon<\frac{1}{2}$, g satisfies $|g| \leqslant ck^{-1-\epsilon}$.

Proof: We write (2.7) as $\gamma - 1 = I_1 + I_2 + I_3$.

In I_1, we use (2.5), obtaining

$$2ikI_1 = -\int_0^\infty \exp(2ikr)V(x-re)dr + O(k^{-1}). \quad (2.9)$$

The first term of (2.9) we integrate by parts; this shows that $I_1 = O(k^{-2})$. Similarly, in I_2 we also use (2.5), obtaining

$$I_2 = (2ik)^{-1}\int_0^\infty V(x-re)dr + O(k^{-2}). \quad (2.10)$$

In I_3 we interchange the order of integration; for $0<\epsilon<\frac{1}{2}$ we write

$$4\pi ikI_3 = \int_0^2 \left(\int_0^\infty \cdots dr\right)^\epsilon \left(\int_0^\infty \cdots dr\right)^{1-\epsilon} du. \quad (2.11)$$

The first factor of (2.11) we integrate by parts; our hypotheses and the result of Lemma 3 assure us that the resulting integrals are bounded. This shows that $I_3 = O(k^{-1-\epsilon})$. Q.E.D.

Corollary: Let V satisfy the hypotheses of the above theorem. Then $\eta(t,e,x)$, the Fourier transform of $\psi(k,e,x) \times \exp(-ike\cdot x) - 1$, can be written

$$\eta(t,e,x) = -\frac{1}{2}\int_0^\infty V(x-re)drH(t) + h(t,e,x), \quad (2.12)$$

where H is the Heaviside function and h is continuous and, when V gives rise to no bound states, vanishes for $t\leqslant 0$. We therefore have

$$V(x) = -2e\cdot\nabla\eta(0^+,e,x). \quad (2.13)$$

Proof: We compute η using the above theorem:

$$(2\pi)^{-1}\int_{-\infty}^\infty \exp(-ikt)[\psi(k,e,x)\exp(-ike\cdot x) - 1]dk = -\frac{1}{2}\int_0^\infty V(x-re)dr\left(\frac{1}{2}\operatorname{sgn} t\right) + (2\pi)^{-1}\int_{|k|>1}\exp(-ikt)g(k,e,x)dk + (2\pi)^{-1}\int_{-1}^1 \exp(-ikt)g(k,e,x)dk. \quad (2.14)$$

The second term on the right side of (2.14) is the Fourier transform of an L^1 function; this second term is therefore a continuous function that decays at infinity. The third term on the right side of (2.14), however, is

$$(2\pi)^{-1}\int_{-1}^{1}\exp(-ikt)g(k,e,x)dk$$
$$=(2\pi)^{-1}\int_{-1}^{1}\exp(-ikt)[\psi(k,e,x)$$
$$\times\exp(-ike\cdot x)-1]dk$$
$$+(2\pi)^{-1}\int_{-1}^{1}\exp(-ikt)(2ik)^{-1}dk$$
$$\times\int_{0}^{\infty}V(x-re)dr. \tag{2.15}$$

The first term on the right side of (2.15) is again a Fourier transform of an L^1 function. The second term is

$$(2\pi)^{-1}\int_{0}^{\infty}V(x-re)dr$$
$$\times\left[\int_{-1}^{1}(2ik)^{-1}\cos kt\,dk-\int_{-1}^{1}(2k)^{-1}\sin kt\,dk\right]. \tag{2.16}$$

The first term of (2.16) is the integral of an odd function over an interval symmetric about the origin and is therefore zero. The second term of (2.16), with the change of variables $s=kt$, is

$$(2\pi)^{-1}\int_{0}^{\infty}V(x-re)dr\int_{0}^{t}s^{-1}\sin s\,ds.$$

We note that $\int_0^\infty s^{-1}\sin s\,ds=\pi/2$.

We have thus shown that the Fourier transform of $\psi(k,e,x)\exp(-ike\cdot x)-1$ is

$$-\frac{1}{2}\int_{0}^{\infty}V(x-re)dr\left(\frac{1}{2}\operatorname{sgn}t-\pi^{-1}\int_{0}^{t}s^{-1}\sin s\,ds\right)$$
$$+l(t,e,x),$$

where $l(t,e,x)$ is a continuous function that decays at infinity in t. However, since $\psi(k,e,x)\exp(-ike\cdot x)-1$ is analytic in the upper half k-plane, its Fourier transform must be zero for t negative. Q.E.D.

ACKNOWLEDGMENTS

I am grateful to Michael Wienstein for suggesting the use of Eq. (2.16). I would also like to thank Joseph Keller for reading the manuscript and making a number of helpful comments.

This work was supported by the National Science Foundation, the Office of Naval Research, the Army Research Office, and the Air Force Office of Scientific Research.

[1] R. G. Newton, *Scattering Theory of Waves and Particles* (Springer, New York, 1982), 2nd ed.
[2] R. G. Newton, J. Math. Phys. **21**, 1698 (1980).
[3] C. S. Morawetz, Comp. Math. Appl. 7, 319 (1981); W. W. Symes, SIAM J. Math. Anal. **12**, 421 (1981); P. D. Lax, *Lectures on Hyperbolic Partial Differential Equations*, Courant Institute Lecture Notes (unpublished); G. N. Balanis, J. Math. Phys. **13**, 1001 (1972).
[4] D. Ludwig, Commun. Pure Appl. Math. **13**, 473 (1960).
[5] R. G. Newton, J. Math. Phys. **22**, 2191 (1981).
[6] R. G. Newton, J. Math. Phys. **23**, 594 (1982).
[7] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order* (Springer, New York, 1977).

Remarks on inverse scattering in one dimension

Roger G. Newton
Physics Department, Indiana University, Bloomington, Indiana 47405

(Received 23 November 1983; accepted for publication 25 May 1984)

This paper answers the following questions: (1) what are the consequences in the matrix-Marchenko inversion scheme if a given S matrix lacks forward analyticity; and (2) in particular, does the condition known as the miracle depend on forward analyticity, and if not, what properties of S does it depend on? The answers are (1) if the input S matrix lacks forward analyticity then the output S matrix has it anyway, and (2) integrability of $kR_{l,r}$ is sufficient for the miracle to occur. It is also found that the matrix–Marchenko procedure simultaneously constructs the potentials for two scattering problems which differ only by the signs of their reflection coefficients.

PACS numbers: 03.65.Nk, 11.20. – e

I. INTRODUCTION

The inverse scattering problem for the Schrödinger equation in one dimension, as is well known, is essentially solved: If one of the reflection coefficients R_l or R_r is given as a function of the wave number, and there are no point eigenvalues (bound states), then necessary and sufficient conditions are known[1–3] for a unique potential to exist, and one has well-established construction procedures. If there are point eigenvalues then these and their associated norming constants must also be provided.

From the perspective of higher-dimensional inverse scattering problems, however, it is of interest to formulate the problem in one dimension alternatively as one in which not only R_l or R_r are given, but *both* are known, and so is the transmission coefficient T. In other words the entire 2×2 S matrix

$$S=\begin{pmatrix} T & R_r \\ R_l & T \end{pmatrix} \tag{1.1}$$

is given. This problem is also solved,[4] and the construction procedure leads to a consistency condition known as the *miracle*: The solution of a matrix generalization of the Marchenko equation must have a certain specified structure in order for an underlying potential to exist.

The following properties are known[1–3] to be necessary conditions for $S(k)$ (where k is the wave number) to be an admissible S matrix.[5]

(i) $S(-k)=S^*(k)$.

(ii) $\tilde{S}=qSq, \quad q=\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

(iii) $SS^\dagger=\mathbf{1}$.

(iv) $[S(k)-\mathbf{1}]\in L^2(-\infty,\infty)$.

(v) $T(k)$ is the boundary value of an analytic function that is meromorphic in $\mathbb{C}^+$, with simple poles at the points $k=i\kappa$ if $-\kappa^2$ is an eigenvalue (bound state), and

$$\lim_{|k|\to\infty} T(k)=1.$$

These properties also allow a unique determination (by quadrature) of all of S from a knowledge of either R_l or R_r, together with the point eigenvalues.[2] Therefore the miracle *must* happen whenever R_l or R_r is in the class of functions that satisfy the Faddeev–Deift–Trubowitz sufficient conditions[2,3] for inversion, which include conditions (i)–(v).

There is a sixth condition that should be mentioned.

(vi) Either (a) $S(0)=-q$ or (b) $T(0)\neq 0$.

If $|x|V\in L^1$ this property follows, possibility (a) being generic and (b) indicating an exceptional "half-bound" state at $k=0$. It is part of the sufficient condition of Faddeev–Deift–Trubowitz.[2–3] On the other hand, if $|x|V\notin L^1$ examples are known[6] for which (vi) fails.

We now wish to study the consequences in the inversion via the matrix–Marchenko equation[4] if property (v), which is generally referred to as *forward analyticity*, is not satisfied. Does the miracle depend on this property, and if not, what goes wrong? We note that without (v) it is impossible to construct all of S from one of the reflection coefficients. Therefore the reconstruction of the potential from R_l or R_r alone cannot work. Indeed, the derivation of the Marchenko equation that utilizes only R_l or R_r, depends on (v). What is the status of (vi)?

The questions to be answered here have analogs in higher dimensions, where the miracle plays a crucial but rather obscure role. The alternative of solving the inverse problem from partial data by an analog of the inversion from R_l or R_r in one dimension is not known to exist in higher dimensions, and any clarification of the miracle and of forward analyticity in $\mathbb{R}$ may lead to a better understanding of their roles in $\mathbb{R}^n$, $n>1$. While forward analyticity is known to be necessary for admissibility of S, it never appears to be utilized in the higher-dimensional generalization of the Marchenko procedure.

Before we begin the study it is important to realize that the S matrix plays two different roles in frequency-domain scattering theory: The first comes from the asymptotics of the "physical" solution of the Schrödinger at large $|x|$; let us call this the *asymptotic* S matrix. Its definition allows the transmission and reflection coefficients to be related to observations. The second is to connect the "outgoing wave" solutions to the "incoming wave" solutions; let us call this the *functional* S matrix. It is the second role that makes the inversion procedure possible. The Marchenko equation and its generalizations utilize this, and only this, role of the S matrix. Therefore it is important to check that a solution of the inverse problem in which the functional S matrix enters

 0022-2488/84/102991-04$02.50

as initial information, has asymptotics at large $|x|$ that also agree with this S matrix, i.e., that the asymptotic S matrix of the output equals the functional S matrix that served as input.

With these remarks in mind we pose the following problem: Given an S matrix that satisfies properties (i)–(iv) but not (v) or (vi), what can we say about the solution of the matrix–Marchenko equation in relation to a Schrödinger equation? Can the miracle occur? Are the asymptotic and functional S matrices equal? For simplicity, we shall study this problem only for the case of no bound states.

The answers we find are as follows. If $kR_{l,r} \in L^1(-\infty,\infty)$ then the miracle will necessarily occur. Thus the miracle is independent of forward analyticity and (in one dimension) is brought about by a simple integrability condition. Consequently a potential in a Schrödinger equation that leads to the given S matrix will exist, even without (v). What is more, the matrix solution of the matrix–Marchenko equation with a given S matrix that satisfies (vi),(a) simultaneously leads to the solution of a problem with an S matrix that violates it. Therefore it is somewhat misleading to call (vi),(a) generic; it is generic only for a rather artifically restricted class of potentials.

We also find that if the functional S matrix does not satisfy (v) then the asymptotic S matrix of the constructed solution of the Schrödinger equation is not equal to it. The two S matrices differ by a factor that makes the asymptotic S matrix satisfy (v), even though the functional S matrix does not.

In Sec. II we show the connection between integrability of $kR_{l,r}$ and the miracle. We also show there how a solution of the matrix–Marchenko equation for a problem with (vi),(a) leads to a solution of a problem that violates it. In Sec. III we examine the asymptotics of the solution of the matrix–Marchenko equation if (v) fails.

II. INTEGRABILITY AND THE MIRACLE

The physical solution of the Schrödinger equation in $\mathbb{R}$,

$$\psi'' + k^2\psi = V\psi, \tag{2.1}$$

is defined by the Lippmann–Schwinger equation

$$\psi(k,x) = \exp(iIkx)\hat{1} - \frac{i}{2k}\int_{-\infty}^{\infty} dy\, e^{ik|x-y|}V(y)\psi(k,y), \tag{2.2}$$

where ψ is a two-component column vector, $\hat{1} = \binom{1}{1}$, and $I = \left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)$. If $|x|V \in L^1(-\infty,\infty)$ the function $\gamma(k,x) = \exp(-iIkx)\psi(k,x)$ is known to have the following properties.

(1) For each fixed x, and all $k \in \mathbb{R}$, γ is a continuous function of k, with

$$\gamma(k,x) = \hat{1} + O(1/k),$$

$k \to \pm\infty$.

(2) γ is the boundary value of an analytic function that is meromorphic in $\mathbb{C}^+$, with simple poles at $k = i\kappa$ if $-\kappa^2$ is a bound-state eigenvalue.

We shall assume that there are *no bound states*; in that case the analytic continuation of γ is holomorphic in $\mathbb{C}^+$. Furthermore, $\lim_{|k|\to\infty}\gamma(k,x) = \hat{1}$.

The solutions $\psi(k,x)$ and $\psi(-k,x) = \psi^*(k,x)$ of (2.1) are related to one another by

$$\psi(-k,x) = S^{-1}(k)q\psi(k,x), \tag{2.3}$$

in which the 2×2 matrix S is the *functional S matrix*. Equivalently we have

$$\gamma(-k,x) = S_x{}^{-1}(k)q\gamma(k,x), \tag{2.3'}$$

where

$$S_x(k) = e^{iIkx}S(k)e^{-iIkx}.$$

It is convenient to replace (2.3') by the the equation

$$\Gamma_x(-k) = S_x{}^{-1}(k)q\Gamma_x(k)q, \tag{2.4}$$

where $\Gamma_x(k)$ is a 2×2 matrix such that $\gamma(k,x) = \Gamma_x(k)\hat{1}$ and with analogous asymptotic and analytic properties. A simple rewriting of (2.4) gives, by the use of (i) and (iii),

$$\Gamma_x(k) - \mathbb{1} = [\tilde{S}_x(k) - \mathbb{1}]q[\Gamma_x(k) - \mathbb{1}]q + \tilde{S}_x(k) - \mathbb{1} + q[\Gamma_x(-k) - \mathbb{1}]q. \tag{2.4'}$$

Equation (2.4) on the real axis, together with the analytic and asymptotic properties of $\Gamma_x(k)$ in $\mathbb{C}^+$, constitutes a Riemann–Hilbert problem. It is solved by defining the Fourier transforms

$$g_x(\alpha) = \left(\frac{1}{2\pi}\right)\int_{-\infty}^{\infty} dk\, e^{-ik\alpha}[\tilde{S}_x(k) - \mathbb{1}], \tag{2.5}$$

$$\zeta_x(\alpha) = \left(\frac{1}{2\pi}\right)\int_{-\infty}^{\infty} dk\, e^{-ik\alpha}[\Gamma_x(k) - \mathbb{1}] \tag{2.5'}$$

and replacing (2.4') by its Fourier transform

$$\zeta_x(\alpha) = g_x(\alpha) + q\zeta_x(-\alpha)q + \int_{-\infty}^{\infty} d\beta\, g_x(\alpha+\beta)q\zeta_x(\beta)q.$$

The analyticity of $\Gamma_x(k)$ implies that $\zeta_x(\alpha) = 0$ for $\alpha < 0$. Therefore for $\alpha > 0$

$$\zeta_x(\alpha) = g_x(\alpha) + \int_0^{\infty} d\beta\, g_x(\alpha+\beta)q\zeta_x(\beta)q \tag{2.6}$$

and also

$$q\zeta_x(\alpha)q = -g_x(-\alpha) - \int_0^{\infty} d\beta\, g_x(\beta-\alpha)q\zeta_x(\beta)q.$$

It follows that

$$\lim_{\alpha\downarrow 0}\frac{d[\zeta_x(\alpha) + q\zeta_x(\alpha)q]}{dx} = \lim_{\alpha\downarrow 0}\frac{d[g_x(\alpha) - g_x(-\alpha)]}{dx}.$$

If $dg_x(\alpha)/dx$ is continuous at $\alpha = 0$ then the right-hand side vanishes and we have

$$\frac{d\zeta_x(0+)}{dx} = \lim_{\alpha\downarrow 0}\frac{d\zeta_x(\alpha)}{dx} = -q\left(\frac{d\zeta_x(0+)}{dx}\right)q. \tag{2.7}$$

On the other hand, continuity of $dg_x(\alpha)/dx$ follows from integrability of

$$\frac{d(S_x - \mathbb{1})}{dx} = 2ik\begin{pmatrix} 0 & R_r e^{2ikx} \\ -R_l e^{-2ikx} & 0\end{pmatrix}.$$

Therefore, if kR_l and kR_r are in $L^1(-\infty,\infty)$ then dg_x/dx is a continuous function of α and hence we have (2.7).

Now the use of the matrix–Marchenko equation (2.6) and of (2.5') and (2.7) leads to a Schrödinger-like equation for the square matrix $\Psi(k,x) = \exp(iIkx)\Gamma_x(k)$,

$$\Psi'' + k^2\Psi = \Psi\mathscr{V}, \tag{2.8}$$

where $\mathscr{V}$ is the 2×2 matrix

$$\mathscr{V}(x) = -2I\frac{d}{dx}\zeta_x(0+), \tag{2.9}$$

and by (2.7),

$$q\mathscr{V}q = \mathscr{V}. \tag{2.10}$$

This implies that $\mathscr{V}$ is of the form

$$\mathscr{V} = \begin{pmatrix} \lambda_1 & \lambda_2 \\ \lambda_2 & \lambda_1 \end{pmatrix},$$

and such a $\mathscr{V}$ has the eigenvectors $\hat{1}$ (with eigenvalue $V = \lambda_1 + \lambda_2$) and $I\hat{1} = \mathring{1}$ (with eigenvalue $U = \lambda_1 - \lambda_2$). Therefore, multiplication of (2.8) by $\hat{1}$ leads to the Schrödinger equation for $\psi = \Psi\hat{1}$ with the potential $V(x) = \lambda_1 + \lambda_2$ while (2.9) becomes

$$V(x)\hat{1} = -2I\frac{d}{dx}\zeta_x(0+)\hat{1}. \tag{2.9'}$$

The equality of the two components on the rhs of (2.9') implied by that of those on the lhs, is the miracle.[4] Thus absolute integrability of $kR_{l,r}$ together with properties (i)–(iv), is a sufficient condition for the miracle to occur.[7]

On the other hand, multiplication of (2.8) by $\mathring{1}$ leads to the Schrödinger equation with the potential $U = \lambda_1 - \lambda_2$ and (2.9) becomes

$$U(x)\mathring{1} = -2\frac{d}{dx}\zeta_x(0+)\mathring{1}. \tag{2.9''}$$

Again a miracle occurs automatically because of (2.10). The pair of functions that are the components of $\xi = I\Psi\mathring{1}$, which satisfy the Schrödinger equation with the potential U given by (2.9''), obey the relation

$$\xi(-k,x) = [IS(k)I]^{-1}q\xi(k,x). \tag{2.3''}$$

This means the potential U is associated with a functional S matrix in which the two reflection coefficients R_l and R_r differ by factors of -1 from those associated with V of (2.9'), and the transmission coefficient is the same. If $T(0)\neq 0$, then both U and V are exceptional, but if $T(0) = 0$ then the reflection coefficients for U have the "wrong" value $+1$ at $k = 0$ if those for V have the "right" value -1, or vice versa. It follows that the functions $\lambda_1(x)$ and $\lambda_2(x)$ must be such that $|x|\lambda_1 \notin L^1$ and $|x|\lambda_2 \notin L^1$.

III. THE ASYMPTOTIC S MATRIX

We now assume that a continuous functional S matrix with properties (i)–(iv), and such that $(S-\mathbb{1}) \in L^1$, is given and the Riemann–Hilbert problem posed by (2.4) is solved, so that $\Gamma_x(k)$ satisfies (2.4'). The analyticity and asymptotics of $\Gamma_x(k)$ imply that

$$\Gamma_x(k) - \mathbb{1} = \frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{dk'[\Gamma_x(k')-\mathbb{1}]}{k'-k-i\epsilon}$$

in the limit as $\epsilon\downarrow 0$. Therefore by (2.4')

$$\Gamma_x(k) = \mathbb{1} + \frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{dk'[\tilde{S}_x(k')-\mathbb{1}]q\Gamma_x(-k')q}{k'-k-i\epsilon}$$

or

$$\Psi(k,x) = e^{ikIx} + \frac{1}{2\pi i}\times\int_{-\infty}^{\infty}\frac{dk'\,e^{i(k-k')Ix}[\tilde{S}(k')-\mathbb{1}]q\Psi(-k',x)q}{k'-k-i\epsilon}. \tag{3.1}$$

In order to examine the limit of Ψ as $x\to\pm\infty$ we use the fact that if $f(k)$ is continuous and integrable then

$$\lim_{\epsilon\downarrow 0}\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{dk'}{k'-k-i\epsilon}f(k')e^{ik'x} = \begin{cases} f(k)e^{ikx} + o(1), & \text{as } x\to\infty, \\ o(1), & \text{as } x\to-\infty. \end{cases}$$

As a solution of (2.8) Ψ must have the asymptotic form

$$\Psi(k,x) = A_{\pm}e^{ikx} + B_{\pm}e^{-ikx} + o(1), \quad \text{as } x\to\pm\infty \tag{3.2}$$

and (2.4) implies that

$$A_{\pm}(-k) = S^{-1}(k)qB_{\pm}(k)q. \tag{3.3}$$

By equating coefficients of $\exp(ikx)$ and $\exp(-ikx)$ we find from (3.1) that

$$(\mathbb{1}-L)B_+ = 0, \tag{3.4}$$

$$B_+(k) = L + \frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{dk'\,L[\mathbb{1}-S(-k')]B_+(k')}{k'-k-i\epsilon}, \tag{3.5}$$

where $L = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

Let us call $\mathscr{A}^+$ ($\mathscr{A}^-$) the class of complex-valued functions $f(k)\in L^1(-\infty,\infty)$ that are boundary values of functions holomorphic in $\mathbb{C}^+$ ($\mathbb{C}^-$) and such that $f(k)\to 0$ as $|k|\to\infty$; $\mathscr{H}^{\pm}$ are the classes of 2×2 matrices with entries in $\mathscr{A}^{\pm}$. Then (3.5) implies that

$$L(B_+ - \mathbb{1})\in\mathscr{H}^+, L[S(-k)B_+(k) - \mathbb{1}]\in\mathscr{H}^-. \tag{3.6}$$

By (3.4) B_+ has the form

$$B_+ = \begin{pmatrix} 0 & 0 \\ f & g \end{pmatrix}.$$

Using (1.1) and (i) we conclude from (3.6) that $f\in\mathscr{A}^+, f_- = T^*f\in\mathscr{A}^+, (g-1)\in\mathscr{A}^+, (g_- - 1) = (T^*g - 1)\in\mathscr{A}^-$. Let us assume first that $T(0)\neq 0$, i.e., we have an exceptional case. Then the pair g, g_- solves a standard Riemann–Hilbert problem, which has a unique solution if and only if the index of T is equal to zero.[8] Furthermore, if that is the case then the only solution of the problem which f and f_- must solve is the trivial solution $f = f_- = 0$.

The index n of T is obtained from its phase. If $T = |T|e^{i\delta}$ then $\delta(0) - \delta(\infty) = \pi n$. But the properties (ii) and (iii) of S imply that the determinant of S is given by $\det S = T/T^*$. Therefore $\det S = e^{2i\delta}$, and the index equals the number of bound states, by the Levinson theorem.[9] Thus the assumption that there are no point eigenvalues is equivalent to the assumption that the index of T is equal to zero. We may therefore conclude that g is uniquely determined by T, and $f = 0$: $B_+ = gL$. If $(T-1)\in\mathscr{A}^+$ then $g = 1$.

In the generic case we have $T(0) = 0$. We then define $T' = (1 + i\alpha/k)T$ with $\alpha > 0$ and pose the standard Riemann–Hilbert problem with the boundary condition g'_-

$= T'^{*}g'$. Again the Levinson theorem implies that the index vanishes and there is a unique solution pair g', g'_-. The pair of functions that solves the original problem is then given by $g = g', g_- = g'_-/(1 - i\alpha/k)$.

This solution is unique, even though T', g', and g'_- depend on the arbitrary positive constant α. Suppose there were two zero-free solutions, $g_{1-} = T^{*}g_1, g_{2-} = T^{*}g_2$. Then $g_{1-}/g_{2-} = g_1/g_2$, from which we can conclude by Liouville's theorem that $g_1 = g_2$. (The singularity at $k = 0$ is removable.) That the functions g_- and g that emerge from T' are zero-free (i.e, g vanishes nowhere in $\mathbb{C}^+$ and g_- nowhere in $\mathbb{C}^-$) follows from the fact that the functions g' and g'_- which uniquely solve the standard Riemann–Hilbert problem are necessarily zero-free.[10]

Thus we find that in all cases $B_+ = gL$.

In the same manner we find from (3.1) that $A_- = g(1 - L)$, i.e., $A_- = qB_+q$. Finally, (3.3) is used to obtain A_+ and B_-. With the use of these results (3.2) may be rewritten in the form

$$\Psi(x) = \begin{cases} \begin{pmatrix} Tg^{*}e^{ikx} & 0 \\ R_r g^{*}e^{ikx} & ge^{-ikx} \end{pmatrix} + o(1), & \text{as } x \to \infty, \\ \begin{pmatrix} ge^{ikx} & R_l g^{*}e^{-ikx} \\ 0 & Tg^{*}e^{-ikx} \end{pmatrix} + o(1), & \text{as } x \to -\infty. \end{cases}$$

A vector solution Ψ of the Schrödinger equation with the potential $V = \mathscr{V}\hat{1}$ is obtained from Ψ by allowing it to act on the vector $\hat{1}$: $\psi = \Psi\hat{1}$. Thus ψ has the asymptotics

$$\psi(x) = \begin{cases} \begin{pmatrix} Tg^{*}e^{ikx} \\ R_r g^{*}e^{ikx} \end{pmatrix} + ge^{-ikx} + o(1), & \text{as } x \to \infty, \\ \begin{pmatrix} ge^{ikx} \\ Tg^{*}e^{-ikx} \end{pmatrix} + R_l g^{*}e^{-ikx} + o(1), & \text{as } x \to -\infty. \end{cases}$$

The asymptotic S matrix is therefore given by

$$S_a = (g^{*}/g)S \tag{3.7}$$

and ψ differs from the physical solution by a factor of g.

We note that the diagonal elements of S_a are $Tg^{*}/g = g^{*}_{-}/g$. Since the unique solution g of the Riemann–Hilbert problem posed by $g_- = T^{*}g$ is necessarily free of zeros in $\mathbb{C}^+$, the diagonal elements of S_a are boundary values of a function in $\mathscr{A}^+$. Thus the asymptotic S matrix that emerges from the inversion has property (v), even though the functional S matrix that served as input does not have it. The function ψ that satisfies (2.3) with the given functional S matrix differs from the "physical" solution of the Schrödinger equation by a factor $g(k)$, so that $\psi(-k,x)/g(-k) = q\{[g(-k)/g(k)]S(k)\}^{*}\,\psi(k,x)/g(k)$ and Sg^{*}/g satisfies (v). In other words, the function ψ that satisfies (2.3) does not satisfy the correct boundary conditions as a function of x; the function that does is ψ/g.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under Grant No. PHY 80-20457.

[1] I. Kay and H. E. Moses, Nuovo Cimento **3**, 276 (1956).
[2] L. D. Faddeev, Trudy Mat. Inst. Stekl. **73**, 314 (1964) [Am. Math. Soc. Transl. **2**, 139 (1967)].
[3] P. Deift and E. Trubowitz, Commun. Pure Appl. Math. **32**, 121 (1979).
[4] R. G. Newton, J. Math. Phys. **21**, 493 (1980).
[5] An asterik denotes the complex conjugate, a tilde the transpose, and a dagger the Hermitian adjoint.
[6] P. B. Abraham, B. DeFacio, and H. E. Moses, Phys. Rev. Lett. **46**, 1657 (1981); K. R. Brownstein, Phys. Rev. D **25**, 2704 (1982). The potentials in these examples, however, are distributions.
[7] In three dimensions the miracle consists of a condition similar to (2.9'), namely the θ-independence of Eq. (13) in R. G. Newton, Phys. Rev. Lett. **43**, 541 (1979); or of Eq. (4.3) in R. G. Newton, J. Math. Phys. **21**, 1698 (1980). See also Eq. (7.3) of C. S. Morawetz, Comp. Math. Appl. **7**, 319 (1981). It was pointed out below Eq. (3.6) of R. G. Newton, J. Math. Phys. **23**, 594 (1982), that the analog of Eq. (2.10) holds in that case, but in constrast to the one-dimensional case, this does not imply the miracle.
[8] N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Leyden, 1977).
[9] See Ref. 4 as corrected by R. G. Newton, J. Math. Phys. **24**, 2152 (1983), footnote 20.
[10] See Ref. 8 or R. G. Newton, J. Math. Phys. **23**, 2257 (1982), Lemma 2.

The connection between time- and frequency-domain three-dimensional inverse scattering methods

James H. Rose
Ames Laboratory, USDOE, Iowa State University, Ames, Iowa 50011

Margaret Cheney
Department of Mathematics, Stanford University, Stanford, California 94305

Brian DeFacio[a)]
Institute for Theoretical Physics, Chalmers Tekniska Högskola, S 412 96 Göteborg, Sweden

(Received 22 February 1984; accepted for publication 11 May 1984)

We relate three-dimensional scattering theory for the time-independent Schrödinger equation without spherical symmetry to scattering theory for the plasma-wave equation (PWE). We review a recent inverse scattering method for the PWE and find the corresponding method for the Schrödinger equation. We then review Newton's three-dimensional Marchenko method for the Schrödinger equation and transform it to the corresponding PWE method. The resulting time-domain hyperbolic method clarifies the role of causality in Newton's important recent work.

PACS numbers: 03.65.Nk

I. INTRODUCTION

The inverse scattering problem is of fundamental importance to physics because it gives an explicit, nonparametric way of inferring the properties of an object (described by an interaction potential) from a scattering experiment. The most complete and systematic study of inverse scattering has been made for the scattering of particles governed by the time-independent Schrödinger equation. Exact methods for solving the inverse problem were found for the case of spherically symmetric potentials by Gel'fand and Levitan,[1] Marchenko,[2] Kay and Moses[3] and others. For nonspherically symmetric potentials in three dimensions there are two results that are relevant to this paper. One, an exact solution that has been available for some time, uses the Born approximation for high-energy scattering in the nearly forward direction. A second, more recent result is a generalization by Newton[4] of Marchenko's one-dimensional method to three dimensions.

For one-dimensional quantum inverse scattering, the connection between the Schrödinger equation and a related hyperbolic wave equation, the plasma-wave equation (PWE) has been noted.[5] Recently, Morawetz,[6] Callias and Uhlmann,[7] and DeFacio and Rose[8] have proposed exact inverse methods for the PWE with a three-dimensional potential. These methods give considerable physical insight into the general features of inverse scattering methods; in particular, the hyperbolic methods clarify the role of causality.

People working on the PWE have remarked that the PWE could be used to study the quantum-mechanical inverse scattering problem. The purpose of this paper is to establish the connections between the Schrödinger and the PWE inverse scattering methods. The connection we use can also be applied to the direct problem, although we will not pursue this aspect of the problem.

The structure of the paper is as follows. Section II contains the basic facts about scattering theory for both the Schrödinger equation and the PWE; connections between the two are established. The well-known completeness relation for the Schrödinger equation is translated to the plasma-wave case and is found to result in a generalization of the Radon transform. In Sec. III an exact near-field plasma-wave inverse method proposed by DeFacio and Rose and by Callias and Uhlmann is reviewed, and the corresponding inverse method is developed for Schrödinger scattering. The derivation of the Newton–Marchenko equation for Schrödinger scattering is reviewed in Sec. IV. This is followed by a time-domain version and an explanation of correspondences between time- and frequency-domain quantities. Finally, we conclude the paper with a discussion and summary of our results.

II. BASIC FACTS ABOUT SCATTERING THEORY

Three-dimensional quantum scattering theory starts from the time-independent Schrödinger equation

$$(\Delta + k^2 - V)\psi = 0. \tag{2.1}$$

Here Δ is the Laplacian and V is a real-valued function, which in this paper we will take to be smooth, positive, and of compact support. Thus, we assume that V gives rise to no bound states. Scattering solutions of Eq. (2.1) are determined by the Lippmann–Schwinger equation

$$\psi^{\pm}(k,\hat{e},\mathbf{x}) = \exp(ik\hat{e}\cdot\mathbf{x}) + \int G_0^{\pm}(k,|\mathbf{x}-\mathbf{y}|)V(\mathbf{y})\psi^{\pm}(k,\hat{e},\mathbf{y})d\mathbf{y}, \tag{2.2}$$

where $\hat{e}$ is a unit vector denoting the direction of incidence

$$G_0^{\pm}(k,r) = -(4\pi r)^{-1}\exp(\pm ikr). \tag{2.3}$$

The $+$ sign corresponds to the outgoing radiation condition, while the $-$ sign corresponds to the incoming radiation condition. Examination of Eq. (2.2) shows that ψ^+ and ψ^- are related via

$$\psi^-(k,\hat{e},\mathbf{x}) = \psi^+(-k,-\hat{e},\mathbf{x}) \tag{2.4}$$

and

[a)] Permanent address: Department of Physics and Astronomy, University of Missouri, Columbia, Missouri 65211.

 0022-2488/84/102995-06$02.50

$$\psi(k,\hat{e},\mathbf{x}) = \psi^*(-k,\hat{e},\mathbf{x}), \tag{2.5}$$

where the star indicates complex conjugate. We write the outgoing wave function as

$$\psi^+(k,\hat{e}_i,\mathbf{x}) = \exp(ik\hat{e}_i\cdot\mathbf{x}) + |x|^{-1}A(k,\hat{e}_s,\hat{e}_i) \times \exp(ik|\mathbf{x}|) + h(k,\hat{e}_i,\mathbf{x}), \tag{2.6}$$

where $\mathbf{x} = |\mathbf{x}|\hat{e}_s$ and A is the scattering amplitude

$$A(k,\hat{e}_s,\hat{e}_i) = -(4\pi)^{-1}\int \exp(-ik\hat{e}_s\cdot\mathbf{x})V(\mathbf{x})\psi(k,\hat{e}_i,\mathbf{x})d\mathbf{x}. \tag{2.7}$$

It can be shown that for each k, the remainder $h(k,\hat{e}_i,\mathbf{x})$ is a uniformly square-integrable function of $\mathbf{x}$.

The wave functions $\psi^\pm$ form a complete set in the sense that they give rise to an eigenfunction expansion, which is a decomposition of an L^2 function f in terms of the ψ's. If the potential has no bound states, the expansion takes the form[9]

$$f(\mathbf{x}) = (2\pi)^{-3}\int_0^\infty \int_{S^2} \psi^\pm(k,\hat{e},\mathbf{x}) \times \int \psi^{\pm *}(k,\hat{e},\mathbf{y})f(\mathbf{y})d\mathbf{y}\, d^2\hat{e}\, k^2\, dk, \tag{2.8}$$

where S^2 denotes the unit sphere in R^3. If f is real valued, then using the fact that $\psi^\pm(k,\hat{e},\mathbf{x}) = \psi^{\pm *}(-k,\hat{e},\mathbf{x})$, we can write the complex conjugate of (2.8) as

$$f(\mathbf{x}) = (2\pi)^{-3}\int_{-\infty}^0 \int_{S^2} \psi^\pm(k,\hat{e},\mathbf{x}) \times \int \psi^{\pm *}(k,\hat{e},\mathbf{y})f(\mathbf{y})d\mathbf{y}\, d^2\hat{e}\, k^2\, dk. \tag{2.9}$$

Equations (2.8) and (2.9) together give

$$f(\mathbf{x}) = (16\pi^3)^{-1}\int_{-\infty}^\infty \int_{S^2} \psi^\pm(k,\hat{e},\mathbf{x}) \times \int \psi^{\pm *}(k,\hat{e},\mathbf{y})f(\mathbf{y})d\mathbf{y}\, d^2\hat{e}\, k^2\, dk. \tag{2.10}$$

A time-independent equation such as (2.1) can be related to equations in the time domain. We shall use the Fourier transform

$$\hat{f}(t) = (2\pi)^{-1}\int_{-\infty}^\infty \exp(-ikt)f(k)dk \tag{2.11}$$

to obtain from (2.1) the plasma-wave equation

$$\left[\Delta - \left(\frac{\partial^2}{\partial t^2}\right) - V(\mathbf{x})\right]u = 0, \tag{2.12}$$

where $u = \hat{\psi}$. Note that we are using the circumflex to denote both Fourier transforms and unit vectors. We have thus transformed (2.1) into a hyperbolic equation. The solutions of (2.12) that interest us are those that correspond to Schrödinger's scattering solutions, namely those defined by the formal analog of the Lippmann–Schwinger equation (2.2)

$$u^\pm(t,\hat{e},\mathbf{x}) = \delta(t-\hat{e}\cdot\mathbf{x}) + \int\int \hat{G}_0{}^\pm(t-t',|\mathbf{x}-\mathbf{y}|) \times V(\mathbf{y})u^\pm(t',\hat{e},\mathbf{y})dt'\, d\mathbf{y}, \tag{2.13}$$

where

$$\hat{G}_0{}^\pm(t,r) = -\delta(r\mp t)(4\pi r)^{-1} \tag{2.14}$$

and

$$u^\pm(t,\hat{e},\mathbf{x}) = (2\pi)^{-1}\int_{-\infty}^\infty \exp(-ikt)\psi^\pm(k,\hat{e},\mathbf{x})dk. \tag{2.15}$$

In writing (2.15), we have extended the wave function ψ to negative values of k via (2.5). As in the frequency domain, the outgoing solution u^+ and the incoming solution u^- are related by

$$u^-(t,\hat{e},\mathbf{x}) = u^+(-t,-\hat{e},\mathbf{x}). \tag{2.16}$$

The time-domain integral equation can be used to describe a scattering experiment as follows. At large negative times the system is prepared with an incident pulse $u^0(t,e,\mathbf{x}) = \delta(t-\hat{e}\cdot\mathbf{x})$, propagating along the $\hat{e}$ direction. Because the potential has compact support, the (distributional) solution of (2.12) for large negative times is identically equal to $\delta(t-\hat{e}\cdot\mathbf{x})$. This stipulation takes the place of both initial conditions for (2.12), so that $(\partial/\partial t)u^0 = \delta'$ for large negative times. The incident field then collides with the target in the neighborhood of the origin and scatters in an outgoing spherical wave plus the incident pulse propagating in the forward direction. After the collision of u^0 with the potential, the field is measured, typically on a large sphere enclosing the origin. These measurements are the scattering data from which we will infer the potential.

The far-field scattered wave can be described by the impulse response function R which is defined to be

$$R(\hat{e}_i,\hat{e}_s,\tau) = \lim_{\substack{t,|\mathbf{x}|\to\infty \\ \tau = t-|\mathbf{x}|}} |\mathbf{x}|(u(t,\hat{e}_i,\mathbf{x}) - \delta(t-\hat{e}_i\cdot\mathbf{x})), \tag{2.17}$$

where the scattering direction is given by $\hat{e}_s = \mathbf{x}/|\mathbf{x}|$. The impulse response function can be expressed explicitly in terms of the potential and the field u by using Eq. (2.13). We carry out the t-integration and use

$$|\mathbf{x}-\mathbf{y}| = |\mathbf{x}| - \hat{e}_s\cdot\mathbf{y} + O(|\mathbf{x}|^{-1}),$$

obtaining

$$R(\hat{e}_i,\hat{e}_s,t) = -(4\pi)^{-1}\int u^+(t+\hat{e}_s\cdot\mathbf{y},\hat{e}_i,\mathbf{y})V(\mathbf{y})d\mathbf{y}. \tag{2.18}$$

This formula shows that the impulse response is precisely the Fourier transform of the scattering amplitude

$$R(\hat{e}_i,\hat{e}_s,t) = (2\pi)^{-1}\int_{-\infty}^\infty \exp(-ikt)A(k,\hat{e}_s,\hat{e}_i)dk. \tag{2.19}$$

The Fourier transform (2.11) can also be used to convert the frequency-domain eigenfunction expansion (2.8) into a formal time-domain one

$$\begin{aligned} f(\mathbf{x}) &= (16\pi^3)^{-1}\int_{-\infty}^\infty \int_{S^2} \psi^\pm(k,\hat{e},\mathbf{x}) \int\int_{-\infty}^\infty \exp(-ikt) \\ &\quad \times u^\pm(t,\hat{e},\mathbf{y})dt\, f(\mathbf{y})d\mathbf{y}\, d^2\hat{e}\, k^2\, dk \\ &= -(8\pi^2)^{-1}\int_{-\infty}^\infty \int_{S^2} u^\pm(t,\hat{e},\mathbf{x}) \\ &\quad \times \left(\frac{\partial^2}{\partial t^2}\right)\int u^\pm(t,\hat{e},\mathbf{y})f(\mathbf{y})d\mathbf{y}\, d^2\hat{e}\, dt, \end{aligned} \tag{2.20}$$

u being real because of the extension of $\psi(k,\hat{e},\mathbf{x})$ to negative frequencies, Eq. (2.5).

This is a generalization of the Radon transform inver-

sion formula[10] in the following sense. We can write (2.20) as the pair of equations

$$\tilde{f}^{\pm}(t,\hat{e}) = \int u^{\pm}(t,\hat{e},\mathbf{y}) f(\mathbf{y}) d\mathbf{y}, \tag{2.21}$$

$$f(\mathbf{x}) = -(8\pi^2)^{-1} \int_{-\infty}^{\infty} \int_{S^2} u^{\pm}(t,\hat{e},\mathbf{x}) \left(\frac{\partial^2}{\partial t^2}\right) \tilde{f}^{\pm}(t,\hat{e}) d^2\hat{e}\, dt. \tag{2.22}$$

If the u's are chosen to be the free space wave functions [i.e., $u^{\pm} = \delta(t - \hat{e}\cdot\mathbf{x})$], then the pair (2.1.) and (2.22) reduce to

$$\tilde{f}(t,\hat{e}) = \int \delta(t - \hat{e}\cdot\mathbf{y}) f(\mathbf{y}) d\mathbf{y}, \tag{2.23}$$

$$f(x) = -(8\pi^2)^{-1} \int_{S^2} \left(\frac{\partial^2}{\partial t^2}\right) \tilde{f}(t = \hat{e}\cdot\mathbf{x}, \hat{e}) d^2\hat{e}. \tag{2.24}$$

Equation (2.23) is precisely the Radon transform, and (2.24) is the Radon transform inversion formula.

III. EXACT NEAR-FIELD INVERSE SCATTERING

Recently, Defacio and Rose have introduced an exact inverse method for the time-domain PWE which depends on near-field data. We will review this result and then derive the equivalent result for the Schrödinger equation. We also discuss the connection between the near-field results and the related Born approximation for solutions of the Schrödinger equation.

The review of the time-domain near-field inverse method starts with a discussion of the "fundamental identity." This result was first used in one-dimensional inverse scattering by Balanis[5] in 1972. Later the corresponding result for the three-dimensional case was found by Morawetz,[6] Callias and Uhlmann,[7] and by DeFacio and Rose.[8]

The fundamental identity is based on the progressing wave expansion

$$u(t,\hat{e},\mathbf{x}) = \delta(t - \hat{e}\cdot\mathbf{x}) + B(\hat{e},\mathbf{x}) H(t - \hat{e}\cdot\mathbf{x}) + D(\hat{e},\mathbf{x}) E(t - \hat{e}\cdot\mathbf{x}) + F(t,\hat{e},\mathbf{x}). \tag{3.1}$$

Where H is the Heaviside function [$H(y) = 1$ if $y > 0$], $E(s)$ is $sH(s)$, F is a continuously differentiable function that is zero for $t < \hat{e}\cdot\mathbf{x}$, and B and D are as yet undetermined. The heuristic principle underlying this expansion is that high-frequency signals are not scattered much by the potential.

The transport equations for the PWE are then determined by substituting Eq. (3.1) into (2.12) and equating orders of singularity. The terms proportional to δ'' and δ' are trivially zero. The term proportional to δ yields

$$V(\mathbf{x}) = -2\hat{e}\cdot\nabla B(\hat{e},\mathbf{x}), \tag{3.2}$$

which can also be written

$$V(\mathbf{x}) = -2\hat{e}\cdot\nabla \lim_{t \to \hat{e}\cdot\mathbf{x}^+} [u(t,\hat{e},\mathbf{x}) - \delta(t - \hat{e}\cdot\mathbf{x})].$$

This simple equation will be referred to as the fundamental identity. It is an expression of the fact that Eq. (2.12) is independent of the variable $\hat{e}$, which we have introduced in the boundary conditions. The fundamental identity relates the potential to the jump in the scattered field at the wave front; if we knew this jump in the field for all points $\mathbf{x}$ (including points within the support of the potential) we could reconstruct V from Eq. (3.2). Normally, however, we are unable to make measurements within the support of V. The fundamental identity can be used to construct a more practical exact near-field inversion method as follows.

Let us suppose that the support of the potential is contained in a sphere S_b of radius b. Further suppose that the jump in the scattered field on the characteristic surface, $B(\hat{e},\mathbf{x})$, is measured for all points on S_b. Then the potential can be determined from a simple linear inverse method involving the Radon transform. We proceed by constructing the Radon transform of V from $B(\hat{e},\mathbf{x})$. In particular we shall obtain the integral of V over the plane $\mathbf{x}\cdot\hat{\omega} = t$, where $\hat{\omega}$ is a unit vector normal to the plane. We choose the incident direction $\hat{e}_i$ so that $\hat{e}_i$ is perpendicular to $\hat{\omega}$. The vectors $\hat{e}_i$ and $\hat{e}_i \times \hat{\omega}$ are then both perpendicular to $\hat{\omega}$ and are therefore both parallel to the plane $\mathbf{x}\cdot\hat{\omega} = t$. As r varies with s and t held fixed, the vector $t\hat{\omega} + s(\hat{e}_i \times \hat{\omega}) + r\hat{e}_i$ thus sweeps out a line $L(t,\hat{\omega},s)$ that is contained in the plane $\mathbf{x}\cdot\hat{\omega} = t$ and that has direction $\hat{e}_i$. This line L intersects the sphere S_b at two points, namely those given by

$$\mathbf{p}_{\pm}(t,\hat{\omega},s) = t\hat{\omega} + s(\hat{e}_i \times \hat{\omega}) \pm (b^2 - t^2 - s^2)^{1/2} \hat{e}_i. \tag{3.3}$$

We integrate (3.2) along the section of L contained in the sphere S_b; this results in

$$\int_{-(b^2 - t^2 - s^2)^{1/2}}^{(b^2 - t^2 - s^2)^{1/2}} V(t\hat{\omega} + s(\hat{e}_i \times \hat{\omega}) + r\hat{e}_i) dr = 2[B(\hat{e}_i, p_-(t,\hat{\omega},s)) - B(\hat{e}_i, p_+(t,\hat{\omega},s))]. \tag{3.4}$$

The first term on the right side of (3.4), moreover, is zero; this follows from integrating (3.2) along L from $-\infty$ to p_-. Integration over the s variable then gives us

$$\int_{-(b^2 - t^2)^{1/2}}^{(b^2 - t^2)^{1/2}} ds \int_{-(b^2 - t^2 - s^2)^{1/2}}^{(b^2 - t^2 - s^2)^{1/2}} dr\, V(t\hat{\omega} + s(\hat{e}_i \times \hat{\omega}) + r\hat{e}_i) = 2 \int_{-(b^2 - t^2)^{1/2}}^{(b^2 - t^2)^{1/2}} B(\hat{e}_i, p_+(t,\hat{\omega},s)) ds. \tag{3.5}$$

The left side of (3.5) is precisely the integral of V over the section of the plane $\mathbf{x}\cdot\hat{\omega} = t$ that is contained in the sphere S_b. Since V is zero outside S_b, this is the Radon transform $\tilde{V}(t,\hat{\omega}) = \int \delta(t - \mathbf{x}\cdot\hat{\omega}) V(\mathbf{x}) d\mathbf{x}$. To obtain the potential V itself, we merely invert the Radon transform using Eq. (2.24).

$$V(\mathbf{x}) = (2\pi)^{-2} \int_{s^2} \left(\frac{\partial^2}{\partial t^2}\right) \times \int_{-(b^2 - t^2)^{1/2}}^{(b^2 - t^2)^{1/2}} B(\hat{e}_i, p_+(t,\hat{\omega},s)) ds \,|_{t = \hat{\omega}\cdot\mathbf{x}}\, d^2\hat{\omega}. \tag{3.6}$$

The vector $\hat{e}_i$ appearing on the right side of (3.6) depends on $\hat{\omega}$, so it will be removed by the $\hat{\omega}$ integration. Because the incident direction $\hat{e}_i$ can be any vector perpendicular to $\hat{\omega}$, $\hat{e}_i$ need only vary in a single half-plane to provide enough information to effect the reconstruction (3.6) exactly.

We can translate this method into a frequency-domain one for the Schrödinger equation as follows. We apply the Fourier transform to Eq. (3.1), obtaining the large-k asymptotic expansion

$$\psi(k,\hat{e},\mathbf{x}) = \exp(ik\hat{e}\cdot\mathbf{x})[1 + B(\hat{e},\mathbf{x})(ik)^{-1} + o(k^{-1})]. \tag{3.7}$$

The foregoing inversion method leading to Eq. (3.6) can

therefore also be applied to the Schrödinger equation case. The quantity $B(\hat{e},\mathbf{y})$ must here be extracted from high-energy measurements. This method, in fact, depends entirely on high-energy data. It is related to another high-energy inversion method, one which is based on the Born approximation.

We recall that the Born approximation of the scattering amplitude $A(k,\hat{e}_s,\hat{e}_i)$ is

$$-(4\pi)^{-1}\int V(\mathbf{x})\exp[ik(\hat{e}_i-\hat{e}_s)\cdot\mathbf{x}]d\mathbf{x}. \tag{3.8}$$

This approximation is valid for large k in the sense that the difference between (3.8) and $A(k,\hat{e}_s,\hat{e}_i)$ vanishes as k becomes infinite. Expression (3.8) is precisely the Fourier transform of $-V(2\pi^2)$. We therefore denote expression (3.8) by $-\hat{V}(q\hat{q})(2\pi^2)$, where $\mathbf{q}=q\hat{q}=k(\hat{e}_s-\hat{e}_i)$. Because approximation (3.8) of $A(k,\hat{e}_s,\hat{e}_i)$ is exact in the $k\to\infty$ limit, we obtain $\hat{V}$ from the scattering amplitude via

$$-(2\pi^2)\hat{V}(\mathbf{q})=\lim_{\substack{k\to\infty\\ \mathbf{q}=k(\hat{e}_s-\hat{e}_i)}} A(k,\hat{e}_s,\hat{e}_i). \tag{3.9}$$

Inversion of the Fourier transform then gives us the potential itself.

To understand the relation between the Born inversion method and the near-field inversion method given by (3.6), let us consider a fixed direction of incidence $\hat{e}_i$. Equation (3.9) allows us to construct the Fourier transform $\hat{V}(\mathbf{q})$ for any finite $\mathbf{q}$ perpendicular to $\hat{e}_i$ (see Fig. 1). We then Fourier transform over $q=|\mathbf{q}|$. This gives

$$\tilde{V}(t,\hat{q})=(2\pi)^{-1}\int_{-\infty}^{\infty}dq\,\exp(iqt)\int V(\mathbf{x})\exp(-iq\hat{q}\cdot\mathbf{x})d\mathbf{x}, \tag{3.10}$$

which is precisely the Radon transform (2.23). From the Born approximation for a single angle of incidence $\hat{e}_i$, we can thus construct the integral of V over any plane whose normal is perpendicular to $\hat{e}_i$.

This is precisely the same information we obtain from the near-field method discussed above. In fact, Eq. (3.5) also gives us $\tilde{V}(t,\hat{\omega})$, where $\hat{\omega}$ is perpendicular to $\hat{e}_i$.

The Born and near-field inverse methods are similar in the following respects. First, they both depend exclusively on high-frequency data. Second, both depend entirely on forward or near-forward scattering data. Third, measurements at a single angle of incidence provide the same information for both. Last, both can be expressed in terms of the Radon transform.[11]

The two methods are different however, in that one (the Born method) uses far-field data, whereas the other uses near-field data.

Finally, we note that the asymptotic expansion (3.7) can be computed entirely in the frequency domain. It is valid under less restrictive hypotheses on V than we use in this paper. Explicitly, the result[12] is as follows.

Theorem: Let V be bounded and integrable with $|V(\mathbf{x})|$, $|\nabla V(\mathbf{x})|$ and $|\Delta V(\mathbf{x})|$, all bounded by $F(|\mathbf{x}-\mathbf{x}_0|)$ for some $\mathbf{x}_0$, where F is a positive function satisfying $\int_0^\infty F(t)dt<\infty$. Then $\psi(k,\hat{e},\mathbf{x})=\exp(ik\hat{e}\cdot\mathbf{x})\quad[1+(2ik)^{-1}\int_0^\infty V(x-r\hat{e})dr+g(k,\hat{e},\mathbf{x})]$, where for $|k|>1$ and any ϵ with $0<\epsilon<\frac{1}{2}$, g satisfies $|g|<c|k|^{-1-\epsilon}$.

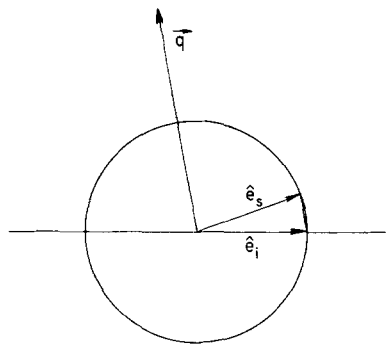


FIG. 1. In the limit as $k\to\infty$, the relation $\mathbf{q}=k(\hat{e}_s-\hat{e}_i)$ forces $\hat{e}_s$ to approach $\hat{e}_i$ and forces $\mathbf{q}$ and $\hat{e}_i$ to become perpendicular.

IV. NEWTON'S MARCHENKO EQUATION IN THE FREQUENCY AND TIME DOMAINS

We review briefly the derivation of the Newton–Marchenko equation for the time-independent Schrödinger equation. We then interpret the elements of the derivation in the time domain, thus obtaining a Marchenko equation for the plasma-wave equation.

A. Frequency-domain derivation of Newton's Marchenko equation

The Newton–Marchenko method rests on the following relation between the scattering amplitude and the wave functions

$$\psi^+(k,\hat{e},\mathbf{x})=\psi^-(k,\hat{e},\mathbf{x})-(2\pi i)^{-1}k\int_{S^2}A(k,\hat{e}',\hat{e})\psi^-(k,\hat{e}',\mathbf{x})d\hat{e}'. \tag{4.1}$$

Equation (4.1) can be derived as follows.[13] The relation

$$(-\Delta-E)^{-1}=(-\Delta+V-E)^{-1}+(-\Delta-E)^{-1}V(-\Delta+V-E)^{-1} \tag{4.2}$$

for $E=k^2\pm i\epsilon$ can be written

$$(-\Delta+V-k^2\mp i\epsilon)^{-1}=-(I-G_0(\pm k+i\epsilon)V)^{-1}G_0(\pm k+i\epsilon), \tag{4.3}$$

where $G_0(k)$ is the integral operator with kernel $G_0(k,|\mathbf{x}-\mathbf{y}|)$. Application of (4.3) to $V(\mathbf{x})\exp(ik\hat{e}\cdot\mathbf{x})$ and use of (2.2) results in

$$\psi^{\pm}(k,\hat{e},\mathbf{x})=\exp(ik\hat{e}\cdot\mathbf{x})-\lim_{\epsilon\to0^+}(-\Delta+V-k^2\pm i\epsilon)^{-1}\times(V(\mathbf{x})\exp(ik\hat{e}\cdot\mathbf{x})). \tag{4.4}$$

Subtraction of the $+$ and $-$ equations of (4.4) gives

$$\psi^+(k,\hat{e},\mathbf{x})-\psi^-(k,\hat{e},\mathbf{x})=\lim_{\epsilon\to0^+}[(-\Delta+V-k^2-i\epsilon)^{-1}-(-\Delta+V-k^2+i\epsilon)^{-1}](V(\mathbf{x})\exp(ik\hat{e}\cdot\mathbf{x})), \tag{4.5}$$

which by Stone's formula[14] is equal to $2k\times P_{\{k^2\}}(V(\mathbf{x})\exp(ik\hat{e}\cdot\mathbf{x}))$, where $P_{\{k^2\}}$ is the spectral projection. The eigenfunction expansion (2.8), however, gives us an explicit formula for the spectral projection; it is merely (2.8) without the integration over k. Accordingly, we apply the inner integral of (2.8) to the function $f(\mathbf{x})=V(\mathbf{x})\exp(ik\hat{e}\cdot\mathbf{x})$.

Equation (2.8) actually contains two eigenfunction expansions, one in terms of ψ^+ and the other in terms of ψ^-. We choose the ψ^- expansion for our present purpose. We can now write Eq. (4.5) as

$$\psi^+(k\hat{e},\mathbf{x}) - \psi^-(k,\hat{e},\mathbf{x}) = -ik\,2^{-3}\pi^{-2} \times \int_{S^2} \psi^-(k,\hat{e}',\mathbf{x}) \int \psi^{-*}(k,\hat{e}',\mathbf{y}) \times V(\mathbf{y})\exp(ik\hat{e}\cdot\mathbf{y})d\mathbf{y}\, d^2\hat{e}', \quad (4.6)$$

which is precisely (4.1) because $A(k,-\hat{e}_s,-\hat{e}_i) = A(k,\hat{e}_i,\hat{e}_s)$ (reciprocity).

The Marchenko integral equation is derived[4] as follows from (4.1). First we subtract off the large-k asymptotic form of ψ and then take the Fourier transforms in k. We use identity (2.4) to write (4.1) as

$$\psi^+(k,\hat{e},\mathbf{x}) = \psi^+(-k,-\hat{e},\mathbf{x}) - (2\pi i)^{-1}k \times \int_{S^2} A(k,\hat{e}',\hat{e})\psi^+(-k,-\hat{e}',\mathbf{x})d^2\hat{e}'. \quad (4.7)$$

We multiply (4.7) by $\exp(-ik\hat{e}\cdot\mathbf{x})$ and write

$$\beta(k,\hat{e},\mathbf{x}) = \psi^+(k,\hat{e},\mathbf{x})\exp(-ik\hat{e}\cdot\mathbf{x})$$

obtaining

$$\beta(k,\hat{e},\mathbf{x}) = \beta(-k,-\hat{e},\mathbf{x}) - (2\pi i)^{-1}k \int_{S^2} A(k,\hat{e}',\hat{e}) \times \exp(ik(\hat{e}'-\hat{e})\cdot\mathbf{x})\beta(-k,-\hat{e}',\mathbf{x})d^2\hat{e}'. \quad (4.8)$$

We subtract 1 from both sides of (4.8)

$$\beta(k,\hat{e},\mathbf{x}) - 1 = \beta(-k,-\hat{e},\mathbf{x}) - 1 - (2\pi i)^{-1}k \times \int_{S^2} A(k,\hat{e}',\hat{e})\exp(ik(\hat{e}'-\hat{e})\cdot\mathbf{x})d^2\hat{e}' - (2\pi i)^{-1}k \int_{S^2} A(k,\hat{e}',\hat{e})\exp(ik(\hat{e}'-\hat{e})\cdot\mathbf{x}) \times [\beta(-k,-\hat{e}',\mathbf{x}) - 1]d^2\hat{e}'. \quad (4.9)$$

We now write

$$\eta(\alpha,\hat{e},\mathbf{x}) = (2\pi)^{-1}\int_{-\infty}^{\infty} \exp(-ik\alpha)(\beta(k,\hat{e},\mathbf{x}) - 1)dk, \quad (4.10)$$

$$M(\alpha,\hat{e},\hat{e}',\mathbf{x}) = (2\pi)^{-2}\int_{-\infty}^{\infty} \exp[-ik(\alpha + (\hat{e}-\hat{e}')\cdot\mathbf{x})] \times ikA(k,\hat{e}',\hat{e})dk. \quad (4.11)$$

In this notation the Fourier transform of Eq. (4.9) is

$$\eta(\alpha,\hat{e},\mathbf{x}) = \eta(-\alpha,-\hat{e},\mathbf{x}) + \int_{S^2} M(\alpha,\hat{e},\hat{e}',\mathbf{x})d^2\hat{e}' + \int_{-\infty}^{\infty}\int_{S^2} M(\alpha-\delta,\hat{e},\hat{e}',\mathbf{x}) \times \eta(-\delta,-\hat{e}',\mathbf{x})d^2\hat{e}'\,d\delta. \quad (4.12)$$

Analyticity of β in the upper half k-plane[4] implies that

$$\eta(\alpha,\hat{e},\mathbf{x}) = 0 \quad \text{for } \alpha < 0, \quad (4.13)$$

which in turn implies that Eq. (4.12) for $\alpha > 0$ is

$$\eta(\alpha,\hat{e},\mathbf{x}) = \int_0^{\infty}\int_{S^2} M(\alpha+\delta,\hat{e},\hat{e}',\mathbf{x})\eta(\delta,-\hat{e}',\mathbf{x})d^2\hat{e}'\,d\delta + \int_{S^2} M(\alpha,\hat{e},\hat{e}',\mathbf{x})d^2\hat{e}'. \quad (4.14)$$

This is Newton's Marchenko equation. The potential can be recovered from the solution η of (4.14) via

$$V(\mathbf{x}) = -2\hat{e}\cdot\nabla_x\eta(0^+,\hat{e},\mathbf{x}). \quad (4.15)$$

This is the "miracle" of Newton.[4] When this equation appears as a requirement on the solution of (4.14), it certainly appears miraculous because the right side involves $\hat{e}$ while the left side is independent of $\hat{e}$. Newton has shown,[4] however, that it is actually a characterization of admissible scattering data in the following sense. Roughly, if the scattering amplitude is such that Eq. (4.14) has a unique solution, and if this solution is miraculous, then the solution corresponds via (4.10) to a solution of the Schrödinger equation that has the correct (given) scattering amplitude. In this context, (4.15) is a complicated constraint on the data, a constraint that must be applied after the most difficult and expensive part of the reconstruction [namely, solving (4.14)] has been carried out.

B. Time domain analog of Newton's Marchenko equation

We Fourier transform Eq. (4.1):

$$u^+(t,\hat{e},\mathbf{x}) = u^-(t,\hat{e},\mathbf{x}) - (2\pi)^{-1}\int_{S^2}\int_{-\infty}^{\infty} u^-(\tau,\hat{e}',\mathbf{x}) \times\left(\frac{\partial}{\partial t}\right)R(\hat{e},\hat{e}',t-\tau)d\tau\, d^2\hat{e}'. \quad (4.16)$$

In (4.16) we use (2.16) and write $u^+(t,\hat{e},\mathbf{x}) = \delta(t-\hat{e}\cdot\mathbf{x}) + u^{sc}(t,\hat{e},\mathbf{x})$, obtaining

$$u^{sc}(t,\hat{e},\mathbf{x}) = u^{sc}(-t,-\hat{e},\mathbf{x}) - (2\pi)^{-1}\int_{S^2}\left(\frac{\partial}{\partial t}\right)R(\hat{e},\hat{e}',t-\hat{e}\cdot\mathbf{x})d^2\hat{e}' - (2\pi)^{-1}\int_{S^2}\int_{-\infty}^{\infty} u^{sc}(\tau,-\hat{e}',\mathbf{x})\left(\frac{\partial}{\partial t}\right)R(\hat{e},\hat{e}',t+\tau)d\tau\, d^2\hat{e}'. \quad (4.17)$$

We then use the fact that

$$u(t,\hat{e},\mathbf{x}) = 0 \text{ for } t < \hat{e}\cdot\mathbf{x} \quad (4.18)$$

(causality!) to obtain for $t > \hat{e}\cdot\mathbf{x}$

$$u^{sc}(t,\hat{e},\mathbf{x}) = -(2\pi)^{-1}\int_{S^2}\left(\frac{\partial}{\partial t}\right)R(\hat{e},\hat{e}',t-\hat{e}'\cdot\mathbf{x})d\hat{e}' - (2\pi)^{-1} \times \int_{S^2}\int_{\hat{e}'\cdot\mathbf{x}}^{\infty} u^{sc}(\tau,-\hat{e}',\mathbf{x})\left(\frac{\partial}{\partial t}\right)R(\hat{e},\hat{e}',t+\tau)d\tau\, d^2\hat{e}'. \quad (4.19)$$

Equation (4.19) is precisely Eq. (4.14), with $\alpha = t - \hat{e}\cdot\mathbf{x}$ and

$$\eta(\alpha,\hat{e},\mathbf{x}) = u^+(\alpha+\hat{e}\cdot\mathbf{x},\hat{e},\mathbf{x}) - \delta(\alpha) = u^{sc}(t,\hat{e},\mathbf{x}), \quad (4.20)$$

$$M(\alpha,\hat{e},\hat{e}',\mathbf{x}) = -(2\pi)^{-1}\left(\frac{\partial}{\partial\alpha}\right)R(\hat{e},\hat{e}',\alpha+(\hat{e}-\hat{e}')\cdot\mathbf{x}) = -(2\pi)^{-1}\left(\frac{\partial}{\partial t}\right)R(\hat{e},\hat{e}',t-\hat{e}'\cdot\mathbf{x}). \quad (4.21)$$

Correspondences (4.20) and (4.21) allow us to apply the Newton–Marchenko inversion method to the PWE, and in doing so we obtain insight into the workings of the method.

Application of the Newton–Marchenko method to the

PWE proceeds as follows. First, we measure the impulse response. The Marchenko equation [Eq. (4.19)] is then solved for the scattered field; finally, the potential is recovered via the fundamental identity (3.2). Correspondence (4.20), in fact, shows that the "miracle" (4.15) is precisely the "fundamental identity" (3.2).

The time-domain theory helps us understand the workings of Newton's Marchenko method. We see that the Marchenko equation (4.14) or (4.19) allows us to compute from far-field data the entire wave field at all points in space. The crucial "triangularity" property (4.13) is precisely causality (4.18). Moreover, we know from Newton's characterization result[4] that solving the Marchenko equation makes use of all the data.

Once the solution has been computed, however, the potential is obtained via (4.15) from the high-frequency asymptotic expansion of the wave field. In the time domain this corresponds to using only values of the scattered field on the wavefront $t = \hat{e}_i \cdot \mathbf{x}$. In fact, the fundamental identity allows recovery of the potential from $u^{sc}(\hat{e}_i \cdot \mathbf{x}, \hat{e}_i, \mathbf{x})$ for a single direction of incidence. The Newton–Marchenko method is therefore not economical because its intermediate steps involve computation of the entire scattered field for all $\mathbf{x}, t$, and $\hat{e}_i$. However, because the Marchenko equation makes use of all the data, this method is presumably less sensitive to high-frequency error than are the methods of Sec. III.

V. DISCUSSION AND SUMMARY

We have seen that the Schrödinger equation in quantum mechanics is related to the PWE by means of the Fourier transform. As a consequence, inverse scattering methods developed for one equation may be applied to the other. In this paper we have transformed an exact near-field inversion method for the PWE to one for the Schrödinger equation. Similarly, we have Fourier transformed the Schrödinger Newton–Marchenko method to the time domain. In doing so we discovered that Newton's "miracle" is precisely the time-domain "fundamental identity" which relates the potential to the jump in the scattered field at the wavefront. In the frequency domain, we found that the miracle and the fundamental identity are asymptotic high-energy results. It is important to note, though, that they are high-energy limits of the entire solution of the Schrödinger equation, and not just of the scattering amplitude.

In general, the time-domain results seem intuitively more understandable. This is because the hyperbolicity and the straight characteristics of the PWE allow us to express causality in a natural way.

In this paper we have emphasized the inverse problem when discussing the connection between the two wave equations. However, the same connection is also valid for direct scattering. For example, numerical and approximate solutions of one problem can be used to construct corresponding solutions for the other problem. Recently, so-called "marching-in-time" methods,[15] which take advantage of causality, have been developed for hyperbolic wave equations. These methods can therefore also be used to obtain numerical solutions to the Schrödinger equation.

Numerical implementation of the inverse scattering methods, however, has not yet been investigated, but the derivatives appearing in Eqs. (2.24) and (4.19) indicate possible stability problems due to high-frequency error. This suggests that prior data and regularization techniques such as those used in tomography may be of considerable use.

ACKNOWLEDGMENTS

We would like to thank Jim Corones for bringing the three of us together at the spring 1983 meeting of the Ames Lab inverse problems group. This group is supported by Navy Contract number NOOO-14-83-KOO38. MC, in particular, is grateful for being invited to participate. We would also like to thank Mark Davison and Harry McMaken for stimulating discussions at this meeting.

We are grateful to Roger Newton and Joe Keller for reading the manuscript and making helpful comments.

BDF would like to thank the Chalmers Institute of Theoretical Physics for their Jubilee Professorship.

This work was supported by the Air Force Office of Scientific Research, the National Science Foundation, the Office of Naval Research, and the Army Research Office.

[1] J. M. Gel'fand and B. M. Levitan, Am. Math. Soc. Transl., Ser. 2 **1**, 253 (1955).
[2] V. A. Marchenko, Math. Rev. **17**, 740 (1956).
[3] I. Kay, and H. Moses, Nuovo Cimento **2**, 917 (1955); Nuovo Cimento **3**, 66 (1956); Nuovo Cimento **3**, 276 (1956).
[4] R. G. Newton, J. Math. Phys. **21**, 1698 (1980); J. Math Phys. **22**, 2191 (1981); J. Math. Phys. **23**, 594 (1982), "The Marchenko and Gel'Fand–Levitan Methods in the Inverse Scattering Problem in One and Three Dimensions," in *Conference on Inverse Scattering: Theory and Application*, edited by J. B. Bednar, R. Redner, E. Robinson, and A. Weglein (SIAM, Philadelphia, 1983).
[5] G. N. Balanis, J. Math. Phys. **13**, 1001 (1972).
[6] C. S. Morawetz, Comput. Math. Appl. **7**, 319 (1981).
[7] C. J. Callias and G. A. Uhlmann, "Asymptotics of the Scattering Amplitude for Singular Potentials," technical report MSRI 034-83, Mathematical Sciences Research Institute, Berkeley.
[8] B. DeFacio and J. H. Rose, "Exact Inverse Scattering Theory for the Non-Spherically Symmetric Three-Dimensional Plasma Wave Equation," unpublished.
[9] S. Agmon, Ann. Scuola Norm. Supl. Piso. Ser. IV **2**, 151 (1975).
[10] I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions, Integral Geometry and Representation Theory* (Academic, New York, 1966), Vol. 5.
[11] S. R. Deans, *The Radon Transform and Some of Its Applications* (Wiley, New York, 1983).
[12] M. Cheney, J. Math. Phys. **25**, 2988 (1984).
[13] R. G. Newton, *Scattering Theory of Waves and Particles*, 2nd ed. (Springer, New York, 1982).
[14] M. Reed and B. Simon, *Functional Analysis* (Academic, New York, 1972), p. 237.
[15] H. Gerard, J. Acoust. Soc. Am. **71**, 264 (1982).

Twenty-four optimal inequalities and several new representations for the Coulomb T matrices in momentum space

H. van Haeringen
Department of Mathematics and Informatics, Delft University of Technology, Delft, The Netherlands

(Received 29 December 1983; accepted for publication 2 March 1984)

We derive new series and integral representations for the Coulomb transition matrix in momentum space, $\langle \mathbf{p}|T_c|\mathbf{p}'\rangle$, and for its partial-wave projections, $\langle p|T_{cl}|p'\rangle$ $(l=0,1,...)$, to be denoted by T_c and T_{cl}, respectively. We also consider hypergeometric-function representations for T_c and T_{cl} and discuss their analytic continuation to the whole complex k plane (k^2 is the energy). The new integrals are essentially $\int_0^\pi \cosh \gamma t\ (\rho - \cos t)^{-1}\, dt$ for T_c and $\int_0^\pi \cosh \gamma t \times Q_l(uu' + vv' \cos t)dt$ for T_{cl}, where γ is Sommerfeld's parameter and ρ,u,u',v, and v' are variables depending on the energy and the momenta; related integrals follow from these. A well-known and convenient series representation for T_c consists essentially of the sum $\Sigma_n y^n(n^2+\gamma^2)^{-1}$, where y depends on the energy and the momenta. We derive its analog for T_{cl}, the corresponding sum being $\Sigma_n(n^2+\gamma^2)^{-1}Q_l^n(u)P_l^{-n}(u')$, $1<u'<u$. This sum is a new member of the family of sums of products of Legendre functions that can be evaluated in a relatively simple closed form; other members of this family have been recently obtained by the author. With the new representations for T_c and T_{cl} we derive a set of twenty-four optimal inequalities (containing two conjectured inequalities) for these Coulomb T matrices, presumably covering all cases relevant for physics. For the proof of these inequalities several different representations, and in particular the newly derived ones, would appear to be indispensable. Many of the inequalities are new. They are valid for fixed real Coulomb strength and fixed real energy $\neq 0$. Because of the complexity of exact closed forms for T_c and T_{cl}, approximations are needed for numerical calculations; the most natural one consists of replacing the Coulomb T matrix by the Coulomb potential. Our inequalities are useful for estimating the accuracy of this approximation.

PACS numbers: 03.65.Nk, 11.20.Dj, 02.30. + g

1. INTRODUCTION AND SUMMARY

A. Background

Almost two centuries ago C. A. Coulomb[1] presented experimental evidence for the existence of the inverse-square law of forces for repulsive (in 1785) and attractive (in 1787) electrical charges. Since then, Coulomb's law and, more generally, interactions between charged particles have been studied extensively, in the context of classical mechanics and quantum mechanics,[2–37] respectively. With the advent of quantum theory the Schrödinger equation became a major object of study. Its two-particle solution with the Coulomb potential consists of the Coulomb wave functions, in position space (r space, or coordinate representation).

In the quantum-theoretical description of interactions between more than two particles, the two-particle off-shell transition (T) operator is important. It can be defined by $T(E) = V + V(E-H)^{-1}V$, where E is the energy, V the potential operator, and H the Hamilton operator. Accordingly, much research has been devoted to this subject, especially for the Coulomb potential. The full Coulomb T operator is denoted by T_c, and its partial-wave projection by T_{cl}, $l=0,1,...$. The momentum representation (p space) has certain advantages over the coordinate representation. For instance, the Coulomb Hamiltonian in the momentum representation exhibits a hidden special (additional) symmetry, which can be described by an o_4 algebra.[6,7,9–11] By using this special symmetry, Schwinger[9] derived an integral representation for T_c. Hostler[8] obtained the same representation in a different way.

For research on charged-particle scattering and the Coulomb T matrix the reader is referred to in particular those references cited in Refs. 17–20. Recently much work has been done on inequalities[21–24] for the Coulomb T matrix in momentum representation, and on explicit closed expressions for various scattering quantities related to the Coulomb T matrix, especially those associated with Coulomb-plus short-range potentials.[25–33]

The use of T_c in exact form in numerically solving three- or more particle equations is cumbersome and time consuming, and hence expensive. This is due to the relative complexity of T_c (in p space: $\langle \mathbf{p}|T_c|\mathbf{p}'\rangle$), so that satisfactory approximations have been sought. The most obvious approximation consists of replacing T_c by the Coulomb potential operator V_c, which considerably reduces the complexity since $\langle \mathbf{p}|V_c|\mathbf{p}'\rangle$ has a very simple form. The accuracy of this approximation can be estimated well by some ratio of T_c and V_c. In this paper we shall derive a large number of inequalities (mostly optimal) for R_c and R_{cl}, covering all relevant cases; here R_c is the ratio of T_c and V_c in momentum space, and R_{cl} is the ratio of their partial-wave projections. Somewhat surprisingly, it turned out that *new* representations for R_c and R_{cl} are necessary to prove these inequalities. In several cases a specific representation is needed to prove each inequality. Therefore, we shall begin (in Sec. 2) by deriving these new representations.

 0022-2488/84/103001-32$02.50

B. Introductory remarks

The two major subjects of this paper are (i) inequalities and (ii) new representations for the full- and partial-wave projected Coulomb T matrices in momentum space, denoted by $\langle \mathbf{p}|T_c|\mathbf{p}'\rangle$ and $\langle p|T_{cl}|p'\rangle$ $(l=0,1,...)$, respectively. Units are taken such that $\hbar = 1 = 2m$, where m is the reduced mass; the energy variable $E \equiv k^2$ is suppressed. It is more convenient to study the closely related Coulomb ratios R_c and R_{cl}. These dimensionless ratios are defined by

$$R_c := R_{cl} := 1 \quad \text{for } s = 0, \tag{1.1a}$$

and

$$R_c := \langle \mathbf{p}|T_c|\mathbf{p}'\rangle / \langle \mathbf{p}|V_c|\mathbf{p}'\rangle, \quad \mathbf{p} \neq \mathbf{p}', \tag{1.1b}$$

$$R_{cl} := \langle p|T_{cl}|p'\rangle / \langle p|V_{cl}|p'\rangle, \quad p \neq p', \tag{1.1c}$$

for $s \neq 0$. Here s is the Coulomb strength parameter, and V_c and V_{cl} are the full- and partial-wave projected Coulomb potential operators, respectively. The denominators of these ratios are real positive and have a simple closed form, see Eq. (1.2)

The study of R_c started in Ref. 18, p. 96; cf. Ref. 21. This R_c study was substantially extended in Ref. 22, where many numerical values of $|R_c|$ and $|R_{cl}|$ have been displayed; it is interesting to note that some conjectures could be "derived" from these graphic computer representations.

For the Coulomb potential one has the well-known expressions[18]

$$V_c(r) \equiv V_{cl}(r) = 2k\gamma/r \equiv -2s/r, \tag{1.2a}$$

$$\langle \mathbf{p}|V_c|\mathbf{p}'\rangle = -s\pi^{-2}|\mathbf{p}-\mathbf{p}'|^{-2}, \tag{1.2b}$$

and

$$\langle p|V_{cl}|p'\rangle = -2s(\pi pp')^{-1} Q_l((p^2+p'^2)/2pp'), \tag{1.2c}$$

where γ is Sommerfeld's parameter, which depends on the energy k^2 according to $\gamma \equiv -s/k$. For $s>0$ the Coulomb potential is attractive, and for $s<0$ repulsive.

The momenta p and p' will be taken real positive throughout this paper; the energy k^2 is in general complex with $k \neq 0$, Re $k \geqslant 0$, Im $k \geqslant 0$, and Im $k \downarrow 0$ whenever necessary. For positive energy we have $k>0$ and for negative energy we take $\kappa := -ik > 0$. Analytic continuation of R_c and R_{cl} into the whole complex k plane will be considered in Sec. 12 only.

Complex energy plays a role in the various *representations* for R_c and R_{cl}. In contrast, the *inequalities* for R_c and R_{cl} to be discussed apply to either *negative or positive* energy; in this case we take s,k, and l to be fixed ($s \in \mathbb{R}$, either $\kappa > 0$ or $k>0$, and $l \in \mathbb{N}$), whereas p, p', and the angle θ between $\mathbf{p}$ and $\mathbf{p}'$ $(0 \leqslant \theta \leqslant \pi)$ are supposed to vary on their respective domains.

At $k=0$, $k=p$, and $k=p'$, R_c and R_{cl} are singular. These singular points may be considered as boundary points; they separate the various regions to which the respective inequalities apply. The positive-energy case is thus split up into two different cases (S and A, see Secs. 23 and 24); together with the negative-energy case (N) three essentially different cases result. Further, each one of the dichotomies (i) attraction/repulsion, (ii) R_c/R_{cl}, and (iii) upper bound/lower bound yields a factor of two, thus generating a total of twenty-four different inequalities. Despite the fact that these inequalities are optimal, the given boundary functions have—with a few exceptions—a simple form.

As part of the study of these inequalities we shall study the zeros of R_c and of R_{cl} for positive energy (Secs. 18–20), and various limits (Secs. 21 and 22) to prove their optimality.

The most interesting results obtained on the twenty-four *inequalities* for R_c and R_{cl} are summarized in Tables I and II and Figs. 3 and 4, see Secs. 23 and 24. An intriguing unsolved problem concerns the conjectured inequalities (c1) and (c2) (cf. Sec. 17),

$$|R_c| \overset{(c1)}{\leqslant} |B_{c0}(i\gamma)|, \tag{1.3}$$

$$|R_{cl}| \overset{(c2)}{<} |B_{cl}(i\gamma)|, \tag{1.4}$$

where $\gamma > 0$, $(p,p') \in S$ (see Sec. 23), and B_{cl} is a relatively simple function (see Sec. 13). We contend that Eqs. (1.3) and (1.4) are valid. If valid, they are optimal. When their right-hand members are replaced by 1, they are easy to prove, but not optimal.

The most interesting *new representations* for R_c and R_{cl} are given by Eqs. (2.7), (3.14), and (4.2). The new integrals reduce to

$$\text{for } R_c: \int_0^\pi \cosh\ \gamma t\,(\rho - \cos t)^{-1}\,dt, \tag{1.5}$$

$$\text{for } R_{cl}: \int_0^\pi \cosh \gamma t\, Q_l(uu' + vv' \cos t)dt, \tag{1.6}$$

and the new series for R_{cl} to

$$\sum_{n=0}^{\infty} (n^2+\gamma^2)^{-1} Q_l^n(u) P_l^{-n}(u'), \quad 1 < u' < u; \tag{1.7}$$

for the notation see Sec. 1.C. Note that no such simple series representation for R_{cl} has been known until now, other than the so-called Sturm series, or Weinberg series,[34–39] which converges only for negative energy (see Sec. 10 and cf. Sloan[34]).

Various Special Functions and their properties will be used frequently in this paper. For these we refer to Refs. 40–43, which may be considered as standard references for the special functions of mathematical physics.

We note that the sum (1.7) constitutes a new member of a family of sums of products of Legendre functions that can be evaluated in a relatively simple closed form.[44,43]

Further details and several proofs and derivations are given in Appendixes A–Z of Ref. 32. A survey of Coulomb formulas may be found in Ref. 33.

Finally we point out that for Coulomb plus separable potentials the associated T matrix T_{csl} is easily expressed in terms of T_{cl} and simple functions; further, for T_{csl} inequalities similar to those for T_{cl} can be derived.[45]

C. Conventions and notations

We shall use the same conventions and notations as in previous related work[18,22,27]; the most relevant of these as well as new notations will be given below.

The sets of the complex, real, integer, and natural numbers are denoted by $\mathbb{C}$, $\mathbb{R}$, $\mathbb{Z}$, and $\mathbb{N}$, respectively; a prime indicates that zero is omitted: $\mathbb{C}' := \mathbb{C} \setminus \{0\}$, $\mathbb{R}' := \mathbb{R} \setminus \{0\}$, $\mathbb{Z}' := \mathbb{Z} \setminus \{0\}$, $\mathbb{N}' := \mathbb{N} \setminus \{0\}$, where $:=$ means "is defined

by." Further, $B(\ ,\ )$ is the beta function,

$$B(y,z) := \Gamma(y)\Gamma(z)/\Gamma(y+z),$$

and $\psi(\ )$ the logarithmic derivative of the gamma function $\Gamma(\ )$,

$$\psi(z) := \Gamma'(z)/\Gamma(z).$$

$F_{i\gamma}$ and $F^{l,i\gamma}$ are defined by

$$F_{i\gamma}(z) := {}_2F_1(1,i\gamma;1+i\gamma;z)$$

and

$$F^{l,i\gamma}(z) := (l+1+i\gamma)^{-1}\,{}_2F_1(1,i\gamma-l;i\gamma+l+2;z),$$

respectively, where ${}_2F_1$ is Gauss' hypergeometric function.[40–43] ${}_1F_1(a;c;z)$ is the confluent hypergeometric function, satisfying the confluent hypergeometric differential equation, and $U(a,c,z)$ is a second solution of this equation.[40] C_n^{λ} is Gegenbauer's polynomial, and $P_n^{(\alpha,\beta)}$ Jacobi's polynomial. $P_\nu^\mu(z)$ and $Q_\nu^\mu(z)$ $(\nu,\mu,z\in\mathbb{C})$ are the (associated) Legendre functions[40–43] "off the branch cut," i.e., $z\notin(-\infty,+1]$.

When $\nu\in\mathbb{N}$, both P_ν^μ and Q_ν^μ have the cut $[-1,+1]$; when $\mu=0$, $P_\nu^0\equiv P_\nu$ has the cut $(-\infty,-1]$; and when $\mu=0$ and $\nu\in\mathbb{N}$, P_ν is Legendre's polynomial so that it has no branch cut at all.

Further, $C\approx 0.5772$ is Euler's constant, and

$$\epsilon_0 := 1, \quad \epsilon_n := 2, \quad n\in\mathbb{N}',$$

$$(z)_n := \Gamma(z+n)/\Gamma(z), \quad n\in\mathbb{Z},\ z\notin\mathbb{Z},$$

$$(1-z)_n = (-1)^n\Gamma(z)/\Gamma(z-n), \quad n\in\mathbb{Z},\ z\notin\mathbb{Z},$$

$$(2n)!! := 2^n n!, \quad n\in\mathbb{N},$$

$$(2n-1)!! := (2n)!/(2n)!!, \quad n\in\mathbb{N}.$$

Some special symbols and abbreviations in connection with the Coulomb potential are

$$\eth := 1 \text{ for } p>k;\ := \exp(\pi\gamma) \text{ for } p<k;$$

$$\eth' := 1 \text{ for } p'>k;\ := \exp(\pi\gamma) \text{ for } p'<k,$$

$$C_0^2 := e^{-\pi\gamma}\Gamma(1+i\gamma)\Gamma(1-i\gamma), \quad i\gamma\notin\mathbb{Z}'$$

$$= e^{-\pi\gamma}\pi\gamma/\sinh\pi\gamma, \quad i\gamma\notin\mathbb{Z},$$

$$c_{l\gamma} := \prod_{n=1}^{l}\left(1+\frac{\gamma^2}{n^2}\right)^{-1}, \quad i\gamma\notin\mathbb{Z}',$$

$$c_{\infty\gamma} := \lim_{l\to\infty} c_{l\gamma} = \prod_{1}^{\infty}\left(1+\frac{\gamma^2}{n^2}\right)^{-1}$$

$$= \pi\gamma/\sinh\pi\gamma, \quad i\gamma\notin\mathbb{Z}.$$

In connection with the momentum variables $\mathbf{p}$, $\mathbf{p}'$, p, and p' it turns out to be practical to use several interrelated although redundant variables because (i) the notation is simplified and offers typographical advantages and (ii) it is preferable to work with real variables. We shall use as "standard" notation

$$a := (p-k)/(p+k), \quad a' := (p'-k)/(p'+k),$$

$$u := (p^2+k^2)/2pk, \quad u' := (p'^2+k^2)/2p'k,$$

$$v := (p^2-k^2)/2pk, \quad v' := (p'^2-k^2)/2p'k,$$

$$w := (p^2+p'^2)/2pp'; \quad u^2-v^2\equiv 1, \quad uu'-vv'\equiv w;$$

$$q^2 := |\mathbf{p}-\mathbf{p}'|^2 = p^2+p'^2-2pp'\cos\theta,$$

$$x^2 := 1+4q^{-2}pp'vv', \quad \mathrm{Re}\,x\leqslant 0,$$

$$y := (x+1)/(x-1), \quad x = (y+1)/(y-1),$$

$$y = -e^{i\varphi}, \quad x = i\tan(\varphi/2), \quad 0\leqslant\varphi<\pi,$$

$$\delta := -\ln(-y)\geqslant 0 \text{ when } -1\leqslant y<0,$$

$$\zeta := -\ln y>0 \text{ when } 0<y<1,$$

$$\delta\leftrightarrow\zeta-i\pi, \quad \rho=\cosh\delta \text{ or } \rho=-\cosh\zeta.$$

For convenience some of these definitions will be occasionally repeated where applied.

Remark: We shall consider R_c and R_{cl} as functions of various (independent) variables, which may differ on different occasions, depending on the purpose we have in mind. For instance, in Sec. 12 we shall consider $R_c(s,k)$ and $R_{cl}(s,k)$, in Sec. 15, $R_c(\gamma)$ and $R_{cl}(\gamma)$, and in Sec. 16, $R_c(\rho;\gamma)$ and $R_{cl}(p'/k;w;\gamma)$. Although not strictly mathematically correct this notation is convenient and should not give rise to misunderstanding if restricted to the proper context.

D. Summary

Representations (mostly new) for R_c and R_{cl} are derived in Secs. 2–5 and 7–12. Inequalities for R_c and R_{cl} are discussed and partly derived in Secs. 6 and 17–24. In connection with these inequalities certain auxiliary functions, viz., the Coulomb boundary function B_{cl}, $|R_c|$, $|R_{cl}|$, and the Coulomb modulus functions M_c and M_{cl} are discussed (Secs. 13–16). As part of the inequalities, the zeros of R_c and R_{cl} for positive energy are derived in Secs. 18–20, and limits from which the optimality of the inequalities follows are derived in Secs. 21 and 22.

The most important results obtained in this paper are given by the following equations: (2.7)–(2.9), (3.14)–(3.16), (4.2), (4.4), (6.1), (7.5)–(7.8), (8.2); especially *convergence* of the series in (9.1) and (10.1)–(10.3); (11.11), (11.15), (12.1), (12.2), (12.8)–(12.19), (12.21), (12.24), (13.1), (13.10), (13.15)–13.23), (14.7), (15.7), (16.10), (16.12), (16.15)–(16.24), (17.1), (17.2), (17.8)–(17.10), (17.16), (19.4), (19.10), (19.16), (19.24), (19.30), (20.14), (20.30), (20.46)–(20.49), (20.69)–(20.71), (21.9), (21.11)–(21.13), (21.17), (21.20), (21.21), (22.5), (22.8), (22.11), (22.13), (22.14), (22.21), (22.26)–(22.30), and (22.42)–(22.44).

A concise synopsis of inequalities for and limits of R_c and R_{cl} is given in Tables I and II, cf. Figs. 3 and 4 (Secs. 23 and 24).

2. NEW INTEGRAL REPRESENTATION FOR R_c

In this section we shall derive a new integral representation for R_c [Eqs. (2.7)–(2.9)] by using[18]

$$R_c = 1 - i\gamma(\rho+1)\int_0^\infty e^{-i\gamma t}$$

$$\times(\rho+\cosh t)^{-1}\,dt, \quad \mathrm{Re}\,i\gamma>-1, \tag{2.1}$$

where

$$\rho := (1+y^2)/(-2y); \quad -1\leqslant y<0: \rho\geqslant 1;$$

for convenience we shall take $\gamma\in\mathbb{R}'$ and $\rho>1$. Putting $\rho\equiv\cosh\delta$ with $\delta>0$ we get

$$\cosh\delta+\cosh t=0 \Leftrightarrow$$

$$t = \pm\delta+i\pi(2n+1), \quad n\in\mathbb{Z}. \tag{2.2}$$

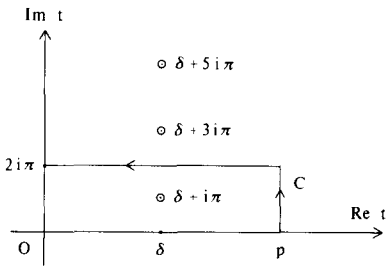


FIG. 1. Contour for R_c integral representation (Sec. 2).

Let us consider in the complex t plane the rectangular contour that runs from the origin O to $p > \delta$, then to $p + 2i\pi$, next to $2i\pi$, and finally from $2i\pi$ back to the origin (Fig. 1). Denoting this contour by C and taking $p \to \infty$ we find

$$\int_C = \int_0^{\infty} - \int_0^{2\pi i} - \int_{2\pi i}^{2\pi i + \infty}. \tag{2.3}$$

Inside C the integrand on the right-hand side of Eq. (2.1) is analytic except for a simple pole at $t = \delta + i\pi$. Its residue is easily evaluated, and with Cauchy's theorem we have

$$\int_C e^{-i\gamma t}/(\rho + \cosh t)dt = -2\pi i e^{\pi\gamma} e^{-i\gamma\delta}/\sinh\delta. \tag{2.4}$$

Furthermore,

$$\int_{2\pi i}^{2\pi i + \infty} e^{-i\gamma t}/(\rho + \cosh t)dt = e^{2\pi\gamma}\int_0^{\infty} e^{-i\gamma t}/(\rho + \cosh t)dt. \tag{2.5}$$

From (2.3)–(2.5) we find directly

$$\int_0^{\infty} e^{-i\gamma t}/(\rho + \cosh t)dt = [1 - e^{2\pi\gamma}]^{-1}\left[-2\pi i e^{\pi\gamma} e^{-i\gamma\delta}/\sinh\delta + \int_0^{2\pi i} e^{-i\gamma t}/(\rho + \cosh t)dt\right]. \tag{2.6}$$

By inserting (2.6) into (2.1) we obtain

$$R_c = 1 + C_0^2 e^{\pi\gamma} e^{-i\gamma\delta}(1 + \cosh\delta)/\sinh\delta = \left(\frac{C_0^2}{2\pi}\right)(1 + \cosh\delta)\int_0^{2\pi} \frac{e^{\gamma t}\,dt}{(\cosh\delta + \cos t)}, \tag{2.7}$$

where $C_0^2 := 2\pi\gamma/(e^{2\pi\gamma} - 1)$. Noting that

$$0 < e^{-\delta} = -y < 1;$$

$$(1 + \cosh\delta)/\sinh\delta = (1 - y)/(1 + y),$$

we reduce (2.7) to

$$R_c = 1 + C_0^2 e^{\pi\gamma}(-y)^{i\gamma}(1 - y)/(1 + y) - \frac{\gamma(1 - y)^2}{\sinh\pi\gamma}\int_0^{\pi} \frac{\cosh\gamma t\,dt}{1 + y^2 + 2y\cos t}, \quad -1 < y < 0. \tag{2.8}$$

Equations (2.7) and (2.8) give useful and interesting new representations for R_c. By analytic continuation and by further simple operations one can obtain other representations which are, however, not essentially different. In particular, representations for R_c valid for $0 < y < 1$ are arrived at by replacing

$$e^{\pi\gamma}(-y)^{i\gamma} \to y^{i\gamma} \tag{2.9}$$

in Eq. (2.8) and in related representations. The proof of Eq. (2.9) follows by deriving the sign of the infinitesimally small imaginary part of $\pm y$, using the convention Im $k \downarrow 0$.

3. NEW INTEGRAL REPRESENTATION FOR R_{cl}

In this section we shall derive a new integral representation for R_{cl} [Eqs. (3.14)–(3.16)] by using[18]

$$R_{cl}Q_l(w) = Q_l(w) - i\gamma\int_0^1 t^{i\gamma - 1}Q_l(\zeta)dt, \quad \text{Re } i\gamma > -l - 1, \tag{3.1}$$

where

$$\zeta := uu' - vv'(t + 1/t)/2; \tag{3.2}$$

for convenience we take $k > 0$. Just as in Sec. 2 we shall apply contour deformation.

For reducing (3.1) it seems natural to use the well-known expansion

$$Q_\nu(uu' - (u^2 - 1)^{1/2}(u'^2 - 1)^{1/2}\cos\psi) = \sum_{n=0}^{\infty}(-1)^n\epsilon_n Q_\nu^n(u)P_\nu^{-n}(u')\cos n\psi, \tag{3.3}$$

$$1 < u' < u, \quad \psi \in \mathbb{R}, \quad -\nu \notin \mathbb{N}'.$$

However, the condition that ψ be real prevents us from directly inserting (3.3) into (3.1). Therefore, we first deform the contour of integration in (3.1). This will lead to a new integral representation for R_{cl}, from which we shall derive in Sec. 4 a new series representation by using (3.3).

Let us consider the region $|t| \leqslant 1$ in the complex t plane. We need to know the "nonanalyticities" of $Q_l(\zeta)$, expressed in terms of t. As is well known, Q_l is analytic except for the branch cut $[-1,1]$. We use

$$a := (p - k)/(p + k), \quad a' := (p' - k)/(p' + k),$$

and assume that

$$0 < p' < k < p \;\&\; k^2 < pp', \tag{3.4a}$$

which implies

$$-1 < a' < 0 < a < 1; 0 < -a'/a < 1. \tag{3.4b}$$

Then the condition $-1 \leqslant \zeta \leqslant 1$ is equivalent to

$$-2 \leqslant -vv'(t - a'/a)(t - a/a')/(2t) \leqslant 0, \tag{3.4c}$$

from which we derive the following two branch cuts in the complex t plane:

$$1/(aa') \leqslant t \leqslant a/a' < -1, \tag{3.5a}$$

$$-1 < a'/a \leqslant t \leqslant a'a < 0. \tag{3.5b}$$

When $pp' < k^2$ (*ceteris paribus*), we only have to interchange a/a' and a'/a, so that the branch cuts in this case are given by

$$1/(aa') \leqslant t \leqslant a'/a < -1, \tag{3.5c}$$

$$-1 < a/a' \leqslant t \leqslant a'a < 0. \tag{3.5d}$$

Let $f(z)$ be analytic on $|z| \leqslant 1$ except for a branch cut, $b \leqslant z \leqslant c$, $-1 < b < c < 0$.

We shall reduce the integral $\int_0^1 z^{\mu - 1}f(z)dz$ by contour defor-

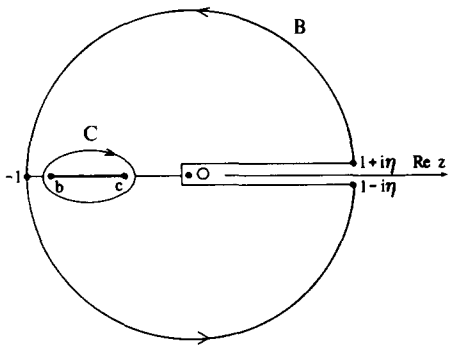


FIG. 2. Contour for R_{cl} integral representation (Sec. 3).

mation. The branch cut and the contours are displayed in Fig. 2, where $\eta \downarrow 0$ is understood. We have, on the one hand

$$\int_{\substack{1+i\eta \\ \text{along O}}}^{1-i\eta} (-z)^{\mu-1} f(z)dz = -2i \sin \pi\mu \int_0^1 z^{\mu-1} f(z)dz, \tag{3.6}$$

and on the other hand,

$$= \int_{\substack{1+i\eta \\ \text{big circle B}}}^{1-i\eta} (-z)^{\mu-1} f(z)dz + \int_C (-z)^{\mu-1} f(z)dz. \tag{3.7}$$

The last integral is obviously equal to

$$2i\pi D := \lim_{\epsilon \downarrow 0} \int_b^c (-z)^{\mu-1} [f(z+i\epsilon) - f(z-i\epsilon)]dz; \tag{3.8}$$

by putting

$$z = e^{i\varphi}, \quad 0 < \varphi < 2\pi \Rightarrow -z = e^{i(\varphi-\pi)}, \quad -\pi < \varphi - \pi < \pi,$$

the next to last integral is reduced to

$$\int_{\varphi=0}^{2\pi} [e^{i(\varphi-\pi)}]^{\mu-1} f(e^{i\varphi}) d(e^{i\varphi}) = -ie^{-i\pi\mu} \int_0^{2\pi} e^{i\mu\varphi} f(e^{i\varphi}) d\varphi. \tag{3.9}$$

By combining Eqs. (3.6)–(3.9) we obtain

$$\sin \pi\mu \int_0^1 z^{\mu-1} f(z)dz = -\pi D + \frac{1}{2} e^{-i\pi\mu} \int_0^{2\pi} e^{i\mu\varphi} f(e^{i\varphi}) d\varphi. \tag{3.10}$$

Now we take $\mu = i\gamma, f(t) = Q_l(uu' - \frac{1}{2}vv'(t + 1/t))$, insert (3.10) into (3.1), and obtain (putting for convenience $\varphi \to 2\pi - \varphi$)

$$R_{cl}Q_l(w) = Q_l(w) + C_0^2 e^{\pi\gamma} D - C_0^2 (2\pi)^{-1} \int_0^{2\pi} e^{\gamma\varphi} Q_l(uu' - vv' \cos\varphi) d\varphi. \tag{3.11}$$

By using the well-known equality[40]

$$Q_l(x + i0) - Q_l(x - i0) = -i\pi P_l(x), \quad -1 < x < 1,$$

we have been able to derive D in closed form from Eq. (3.8), where $b = a'/a$ and $c = a'a$ must be taken, so that

$$2D = \int_{a'/a}^{a'a} (-t)^{i\gamma-1} P_l(uu' - vv'(t + 1/t)/2)dt. \tag{3.12}$$

We have found that

$$D = e^{\pi\gamma} Q_l^{i\gamma}(u) P_l^{-i\gamma}(u'), \quad 1 < u' < u, \tag{3.13}$$

see Appendix D of Ref. 32. Substituting (3.13) in (3.11), and using $t := \varphi - \pi$ in the integral, we derive

$$R_{cl}Q_l(w) = C_0^2 e^{2\pi\gamma} Q_l^{i\gamma}(u) P_l^{-i\gamma}(u') + Q_l(w) - C_0^2 e^{\pi\gamma} \pi^{-1} \int_0^{\pi} \cosh \gamma t Q_l(uu' + vv' \cos t)dt,$$

$$vv' < 0; \quad 0 < p' < k < p \ \& \ k^2 < pp'; \quad 1 < u' < u. \tag{3.14}$$

The condition $k^2 < pp'$ used in the derivation is essential. When $pp' < k^2$ we have $0 < -a/a' < 1$ so that the branch cut of $Q_l(\zeta)$, see (3.1), inside $|t| \leqslant 1$ now becomes: $[a/a', aa']$. In this case we find the same expression as in (3.14) but with u and u' interchanged, i.e.,

$$R_{cl}Q_l(w) = C_0^2 e^{2\pi\gamma} Q_l^{i\gamma}(u') P_l^{-i\gamma}(u) + Q_l(w) - C_0^2 e^{\pi\gamma} \pi^{-1} \int_0^{\pi} \cosh \gamma t Q_l(uu' + vv' \cos t)dt, \tag{3.15}$$

$$vv' < 0; \quad 0 < p' < k < p \ \& \ pp' < k^2; \quad 1 < u < u'.$$

This follows from (3.14) by applying the transformation (t_{12}), see Sec. 23. It is clear that for $pp' \to k^2$, hence $u' \to u$, either of these two expressions may be taken, so that (3.14) and (3.15) are also valid for $pp' = k^2$, $u' = u$. Note that always $p \neq p'$, because $p = p'$ implies $w = 1$ which means that $Q_l(w)$ is not defined in this case. Note further that $k > 0$.

By analytic continuation (cf. Sec. 12) we obtain the desired general expression for R_{cl} from (3.14) and (3.15). In particular, when $k < p' < p$, hence $vv' > 0$ and $u > u'$, we obtain the expression given on the right-hand side of Eq. (3.14), but with one difference, viz., the factor $e^{2\pi\gamma}$ in the first term on the right-hand side must be replaced by $e^{\pi\gamma}$, i.e.,

$$R_{cl}Q_l(w) = C_0^2 e^{\pi\gamma} Q_l^{i\gamma}(u) P_l^{-i\gamma}(u') + Q_l(w) - C_0^2 e^{\pi\gamma} \pi^{-1} \int_0^{\pi} \cosh \gamma t \, Q_l(uu' + vv' \cos t)dt, \tag{3.16}$$

$$vv' > 0; \quad 0 < k < p' < p; \quad k^2 < pp'; \quad 1 < u' < u.$$

To prove this we first consider the factor $Q_l(uu' + vv' \cos t)$ in the integrand, where $-1 \leqslant \cos t \leqslant 1$. Since

$$uu - vv' = w = (p^2 + p'^2)/(2pp') > 1, \quad p \neq p', \tag{3.17}$$

$$uu' + vv' = 1 + (pp' - k^2)^2/(2k^2 pp') \geqslant 1, \tag{3.18}$$

the argument of Q_l in the integral is, when the energy is positive, always $\geqslant 1$, so that no problems can arise from its branch cut $[-1,1]$.

Second we consider $P_l^{-i\gamma}(u')$:

$$P_l^{-i\gamma}(u') = \left(\frac{u'-1}{u'+1}\right)^{i\gamma/2} \frac{1}{\Gamma(1+i\gamma)} \times {}_2F_1(-l, l+1; 1+i\gamma; (1-u')/2). \tag{3.19}$$

Since $(u' - 1)/(u' + 1) = (a')^2$, we get either a factor $(a')^{i\gamma}$ or $(-a')^{i\gamma}$ corresponding to $a' > 0$ or $-a' > 0$, respectively. From the definition, $a' := (p' - k)/(p' + k)$, we see that $\mathrm{Im}(-a') \downarrow 0$ when $\mathrm{Im}\, k \downarrow 0$. Hence,

$$(-a')^{i\gamma}, \quad -a' > 0 \leftrightarrow e^{-\pi\gamma}(a')^{i\gamma}, \quad a' > 0, \tag{3.20}$$

and therefore,

$$P_l^{-i\gamma}(u'), \quad -a' > 0 \leftrightarrow e^{-\pi\gamma} P_l^{-i\gamma}(u'), \quad a' > 0, \tag{3.21}$$

which completes the proof.

Related integral representations follow easily with the help of symmetries of R_{cl}, see Sec. 23 and Appendix S of Ref. 32.

4. NEW SERIES REPRESENTATION FOR R_{cl}

In Eqs. (3.14)–(3.16) we have given an interesting new integral representation for R_{cl}. These expressions are useful for further investigations of R_{cl}. In particular, one can easily derive new series representations for R_{cl} as we shall show in this section. The argument of Q_l in Eqs. (3.14)–(3.16) now has a form which makes application of Eq. (3.3) possible, as discussed after Eq. (3.2).

Under the conditions of Eq. (3.14), including the case $pp' = k^2$,

$$vv' < 0; \quad 0 < p' < k < p \ \& \ k^2 \leqslant pp'; \quad 1 < u' \leqslant u,$$

we have

$$v = (u^2 - 1)^{1/2} > 0, \quad v' = -(u'^2 - 1)^{1/2} < 0.$$

We substitute Eq. (3.3) in the integral on the right-hand side of Eq. (3.14), and use

$$\int_0^\pi \cosh \gamma t \cos nt \, dt = \frac{(-1)^n \gamma \sinh \pi\gamma}{(n^2 + \gamma^2)}, \quad n \in \mathbb{N}. \tag{4.1}$$

Then we obtain

$$R_{cl} Q_l(w) = C_0^2 e^{2\pi\gamma} Q_l^{i\gamma}(u) P_l^{-i\gamma}(u') + Q_l(w) - \gamma^2 \sum_{n=0}^{\infty} \epsilon_n (n^2 + \gamma^2)^{-1} Q_l^n(u) P_l^{-n}(u'), \tag{4.2}$$

$$vv' < 0; \quad 0 < p' < k < p \ \& \ k^2 \leqslant pp'; \quad 1 < u' \leqslant u$$
$$[u' = u \text{ only if } pp' = k^2].$$

Here $\epsilon_0 := 1$ and $\epsilon_n := 2$ for $n = 1,2,...$; note that $u - u' = (p - p')(pp' - k^2)/(2kpp')$. Related series representations for this "symmetric" case ($vv' < 0$) follow easily from the symmetries of R_{cl}, see Sec. 23 and Appendix S of Ref. 32.

The proof of the convergence of the infinite sum in Eq. (4.2), in particular for the borderline case $u = u'$, is arrived at by considering the asymptotic behavior (i.e., for $n \to \infty$) of the terms of this sum; see Appendix C of Ref. 32.

A series representation for the "asymmetric" case, when $vv' > 0$, may be derived either from Eq. (4.2) by analytic continuation, or from Eq. (3.16). In the latter case one needs a slightly modified version of Eq. (3.3), viz.,

$$Q_\nu(uu' + vv' \cos \varphi) = \sum_{n=0}^{\infty} \epsilon_n Q_\nu^n(u) P_\nu^{-n}(u') \cos n\varphi, \tag{4.3}$$

$$v = (u^2 - 1)^{1/2} > 0, \quad v' = (u'^2 - 1)^{1/2} > 0; \quad 1 < u' < u.$$

By using Eq. (4.1) we obtain

$$R_{cl} Q_l(w) = C_0^2 e^{\pi\gamma} Q_l^{i\gamma}(u) P_l^{-i\gamma}(u') + Q_l(w) - \gamma^2 \sum_{n=0}^{\infty} (-1)^n \epsilon_n (n^2 + \gamma^2)^{-1} Q_l^n(u) P_l^{-n}(u'), \tag{4.4}$$

$$vv' > 0; \quad 0 < k < p' < p; \quad 1 < u' < u.$$

It is interesting to observe that the right-hand side of Eq. (4.4) is obtained from the right-hand side of Eq. (4.2) by substituting

$$P_l^{-i\gamma}(u') \to e^{-\pi\gamma} P_l^{-i\gamma}(u'),$$
$$P_l^{-n}(u') \to (-1)^n P_l^{-n}(u').$$

Other series representations for this "asymmetric" case ($vv' > 0$) follow easily from the symmetries of R_{cl}; see Sec. 23 and Appendix S of Ref. 32.

5. SECOND DERIVATION OF THE NEW SERIES REPRESENTATION FOR R_{cl}

In this section we shall give a second proof of Eqs. (4.2) and (4.4). In this case we use the hypergeometric-function representation of the full Coulomb T matrix in three dimensions[18]:

$$\langle \mathbf{p} | T_c | \mathbf{p}' \rangle = k\gamma\pi^{-2} q^{-2} [1 + x^{-1} I(y)],$$

$$I(y) = -C_0^2 y^{i\gamma} + \sum_{n=0}^{\infty} \epsilon_n \gamma^2 y^n / (n^2 + \gamma^2), \tag{5.1}$$

$$x < -1, \quad 0 < y < 1; \quad y = (x + 1)/(x - 1).$$

The corresponding expression for $-1 < y < 0$ is obtained by putting

$$y^{i\gamma} = e^{\pi\gamma} (-y)^{i\gamma} \quad \text{since } \operatorname{Im} y \uparrow 0. \tag{5.2}$$

The condition $0 < y < 1$ given in (5.1) is satisfied when

$$0 < k < p' < p; \quad 0 < a'a < a'/a < 1. \tag{5.3}$$

According to the definition of the partial-wave projection T_{cl} of T_c we have

$$R_{cl} = 2\pi \{ \langle p | V_{cl} | p' \rangle \}^{-1} \int_{-1}^{1} \langle \mathbf{p} | T_c | \mathbf{p}' \rangle P_l(\cos\theta) d(\cos\theta), \tag{5.4}$$

where θ is the angle between $\mathbf{p}$ and $\mathbf{p}'$. By inserting

$$\langle p | V_{cl} | p' \rangle = 2k\gamma (\pi pp')^{-1} Q_l(w),$$

and (5.4) into (5.1) we get

$$R_{cl} Q_l(w) = Q_l(w) + \frac{1}{2} \int_{-1}^{1} x^{-1} \times I(y) P_l(\cos\theta) (w - \cos\theta)^{-1} d(\cos\theta), \tag{5.5}$$

where

$$w := (p^2 + p'^2)/(2pp'), \qquad w = uu' - vv',$$
$$q^2 := |\mathbf{p} - \mathbf{p}'|^2 = 2pp'(w - \cos\theta).$$

By using [recall (5.3)]

$$\cos\theta = w - 2vv'(x^2 - 1)^{-1} = uu' - vv'(y + 1/y)/2; \tag{5.6}$$

$$\cos\theta = 1 \Leftrightarrow y = a'/a,$$
$$\cos\theta = -1 \Leftrightarrow y = a'a < a'/a,$$

we derive from (5.5)

$$R_{cl} Q_l(w) = Q_l(w) - \frac{1}{2} \int_{a'a}^{a'/a} P_l\left(uu' - \frac{1}{2} vv' \left(y + \frac{1}{y}\right)\right) \times I(y) y^{-1} \, dy. \tag{5.7}$$

Since $I(y)$ contains the variable y in the form y^μ, where $\mu = i\gamma$ and $\mu = 0,1,...$, respectively, it suffices to evaluate the

integral

$$2D(\mu) := \int_{a'a}^{a'/a} P_l\left(uu' - \frac{1}{2}vv'\left(y + \frac{1}{y}\right)\right)y^{\mu-1}\,dy. \tag{5.8}$$

In Appendix D of Ref. 32 we obtain [cf. Eqs. (3.12) and (3.13) with $p' \to k^2/p'$]

$$D(\mu) = e^{-i\pi\mu}Q_l^{\mu}(u)P_l^{-\mu}(u'), \quad 1 < u' < u. \tag{5.9}$$

From Eqs. (5.1) and (5.7)–(5.9) we easily find

$$\begin{aligned} R_{cl}Q_l(w) &= Q_l(w) + C_0^2 D(i\gamma) - \gamma^2 \sum_{n=0}^{\infty} \frac{\epsilon_n D(n)}{(n^2+\gamma^2)} \\ &= Q_l(w) + C_0^2 e^{\pi\gamma} Q_l^{i\gamma}(u)P_l^{-i\gamma}(u') \\ &\quad - \gamma^2 \sum_{n=0}^{\infty} (-1)^n \epsilon_n (n^2+\gamma^2)^{-1} Q_l^n(u)P_l^{-n}(u'), \\ vv' &> 0, \quad 1 < u' < u, \quad 0 < k < p' < p, \end{aligned} \tag{5.10}$$

which coincides with Eq. (4.4). Bear in mind that

$$u - u' = (p - p')(pp' - k^2)/(2kpp'), \tag{5.11}$$

so that $k < p' < p$ implies $u' < u$.

6. THE SUPREMA 1 AND $C_0^2 c_{i\gamma}^{-1/2}$ OF $|R_{cl}|$

In this section we shall prove the inequalities

$$\begin{aligned} |R_{cl}| &< C_0^2 \left|\binom{l+i\gamma}{l}\right|, \quad \gamma < 0, \\ |R_{cl}| &< 1, \quad \gamma > 0, \end{aligned} \tag{6.1}$$

valid for $k < p' < p$; note that

$$\left|\binom{l+i\gamma}{l}\right| = c_{i\gamma}^{-1/2}.$$

In view of the transformation invariances of R_{cl} (see Sec. 23) the region of validity of (6.1) is easily extended to

$$k < p' \,\&\, k < p \text{ or } p < k \,\&\, p' < k, \text{ i.e., } (p,p') \in A.$$

Equation (5.10) is very convenient for proving (6.1). We distinguish a real part R_{cl}^r and a remaining complex part R_{cl}^c of R_{cl}:

$$R_{cl} = R_{cl}^c + R_{cl}^r, \quad \gamma \in \mathbb{R}', \tag{6.2}$$

$$R_{cl}^c := C_0^2 e^{\pi\gamma} Q_l^{i\gamma}(u)P_l^{-i\gamma}(u')/Q_l(w), \tag{6.3}$$

$$\begin{aligned} R_{cl}^r := 1 - [Q_l(w)]^{-1}\gamma^2 \sum_{n=0}^{\infty} (-1)^n \epsilon_n (n^2+\gamma^2)^{-1} \\ \times Q_l^n(u)P_l^{-n}(u'). \end{aligned} \tag{6.4}$$

Since u, u', and w are real and greater than 1 we have

$$\begin{aligned} &Q_l(w) > 0, \quad P_l^{-n}(u) > 0, \\ &(-1)^n Q_l^n(u) > 0, \quad n = 0,1,2,\ldots. \end{aligned} \tag{6.5}$$

Hence all terms of the infinite sum in Eq. (6.4) are positive, so that $R_{cl}^r < 1$. By separating off the $n = 0$ term we have the stronger inequality

$$R_{cl}^r < 1 - Q_l(u)P_l(u')/Q_l(w) < 1, \tag{6.6}$$

which will be useful below. Inserting the equality (Ref. 40, pp. 178 and 179)

$$\sum_{n=0}^{\infty} (-1)^n \epsilon_n Q_l^n(u)P_l^{-n}(u') = Q_l(w), \quad 1 < u' < u, \tag{6.7}$$

into (6.4) we obtain

$$\begin{aligned} R_{cl}^r = 2[Q_l(w)]^{-1} \sum_{n=1}^{\infty} n^2(n^2+\gamma^2)^{-1}(-1)^n \\ \times Q_l^n(u)P_l^{-n}(u') > 0; \end{aligned} \tag{6.8}$$

hence, by (6.6) and (6.8),

$$0 < R_{cl}^r < 1 - Q_l(u)P_l(u')/Q_l(w) < 1. \tag{6.9}$$

We use the following two basic inequalities (see Appendix B of Ref. 32)

$$|\Gamma(1+i\gamma)P_\nu^{i\gamma}(z)| \leqslant P_\nu(z), \quad z > 1, \ \nu \geqslant 0, \ \gamma \in \mathbb{R}', \tag{6.10}$$

and

$$\begin{aligned} &\Gamma(1+\nu)e^{\pi\gamma}|Q_\nu^{i\gamma}(z)| < |\Gamma(1+\nu+i\gamma)|Q_\nu(z), \\ &z > 1, \ \nu > -1, \ \gamma \in \mathbb{R}', \end{aligned} \tag{6.11}$$

where $\mathbb{R}' := \mathbb{R} \setminus \{0\}$; equality holds in (6.10) if and only if $\nu = 0$, cf.

$$\Gamma(1-i\gamma)P_0^{i\gamma}(z) = (z+1)^{i\gamma/2}(z-1)^{-i\gamma/2}.$$

From Eqs. (6.2), (6.3), and (6.9)–(6.11) we have

$$\begin{aligned} |R_{cl}| &\leqslant R_{cl}^r + C_0^2 e^{\pi\gamma}|Q_l^{i\gamma}(u)P_l^{-i\gamma}(u')|/Q_l(w) \\ &< R_{cl}^r + C_0^2 |\Gamma(l+1+i\gamma)\{l!\Gamma(1+i\gamma)\}^{-1}| \\ &\quad \times Q_l(u)P_l(u')/Q_l(w) \\ &< 1 + \eta_l\left[C_0^2\left|\binom{l+i\gamma}{l}\right| - 1\right], \end{aligned} \tag{6.12}$$

where η_l is defined by

$$\eta_l := Q_l(u)P_l(u')/Q_l(w). \tag{6.13}$$

From (6.9) it follows that $0 < \eta_l < 1$. Furthermore, we have

$$C_0^2\left|\binom{l+i\gamma}{l}\right| \begin{cases} < 1, & \gamma > 0 \\ > 1, & \gamma < 0. \end{cases} \tag{6.14}$$

Hence, by (6.12)–(6.14)

$$|R_{cl}| < 1, \quad \gamma > 0, \tag{6.15}$$

$$|R_{cl}| < C_0^2\left|\binom{l+i\gamma}{l}\right|, \quad \gamma < 0, \tag{6.16}$$

which completes the proof of (6.1). Both these upper bounds are *suprema* so that these inequalities are optimal; see Sec. 22 and Ref. 23.

7. HYPERGEOMETRIC-FUNCTION REPRESENTATIONS FOR R_c

In this and the following section we shall briefly discuss a number of representations for R_c and R_{cl}, expressed in terms of the Gaussian hypergeometric function ${}_2F_1$.

First of all, from Ref. 18 we have

$$R_c = 1 - (1-y)(1+y)^{-1}[F_{i\gamma}(y) - F_{i\gamma}(1/y)] \tag{7.1}$$

$$= y(1+y)^{-1}F^{0,i\gamma}(y) + (1+y)^{-1}F^{0,i\gamma}(1/y), \tag{7.2}$$

$$y \neq -1, \ y \notin [0,\infty).$$

Here

$$F_{i\gamma}(y) := {}_2F_1(1,i\gamma;1+i\gamma;y),$$

and $F^{0,i\gamma}$ is a particular case (viz., for $l = 0$) of the function

$F^{l,i\gamma}$:

$$F^{l,i\gamma}(y) := (l+1+i\gamma)^{-1}\,{}_2F_1(1,i\gamma-l;i\gamma+l+2;y). \tag{7.3}$$

By applying the equality

$$F_{-i\gamma}(y)+F_{i\gamma}(1/y)=1+\Gamma(1+i\gamma)\Gamma(1-i\gamma)(-y)^{i\gamma} \tag{7.4}$$

to Eq. (7.1) we obtain

$$\begin{aligned} R_c &= C_0^2 e^{\pi\gamma}(-y)^{i\gamma}(1-y)/(1+y)+1 \\ &\quad -[(1-y)/(1+y)]\,[F_{i\gamma}(y)+F_{-i\gamma}(y)-1] \qquad (7.5)\\ &= [(1-y)/(1+y)]\,[C_0^2 e^{\pi\gamma}(-y)^{i\gamma}+2/(1-y) \\ &\quad -F_{i\gamma}(y)-F_{-i\gamma}(y)], \qquad (7.6)\\ &-1<y<0; \quad vv'<0[\,p'<k<p; \quad pp'\neq k^2 \text{ if } \theta=0]; \end{aligned}$$

by analytic continuation:

y complex, $y\neq -1$, $y\notin[0,\infty)$.

Further,

$$\begin{aligned} R_c &= C_0^2 y^{i\gamma}(1-y)/(1+y)+1 \\ &\quad -[(1-y)/(1+y)]\,[F_{i\gamma}(y)+F_{-i\gamma}(y)-1] \qquad (7.7)\\ &= [(1-y)/(1+y)]\,[C_0^2 y^{i\gamma}+2/(1-y)-F_{i\gamma}(y) \\ &\quad -F_{-i\gamma}(y)],\ 0<y<1; \quad vv'>0[k<p'<p]; \qquad (7.8) \end{aligned}$$

by analytic continuation:

y complex, $y\notin(-\infty,0]$, $y\notin[1,\infty)$.

8. HYPERGEOMETRIC-FUNCTION REPRESENTATIONS FOR R_{cl}

Analogous representations for R_{cl} follow from Ref. 27, p. 1267. The analog of Eq. (7.1) is

$$\begin{aligned} &2R_{cl}Q_l(w) = -c_{l\gamma}\,[\mathscr{F}_l/(i\gamma)+\mathscr{E}_l+2\mathscr{L}_l Q_0(w)],\\ &c_{l\gamma}^{-1} := \binom{l+i\gamma}{l}\binom{l-i\gamma}{l}=\prod_{n=1}^{l}\left(1+\frac{\gamma^2}{n^2}\right),\\ &2Q_0(w)=\ln\left(\frac{w+1}{w-1}\right)=\ln\left(\frac{p+p'}{p-p'}\right)^2,\\ &\mathscr{F}_l := F_{i\gamma}(aa')P_l^{(-i\gamma,i\gamma)}(u)P_l^{(-i\gamma,i\gamma)}(u') \qquad (8.1)\\ &\quad +F_{i\gamma}(1/(aa'))P_l^{(i\gamma,-i\gamma)}(u)P_l^{(i\gamma,-i\gamma)}(u')\\ &\quad -F_{i\gamma}(a'/a)P_l^{(i\gamma,-i\gamma)}(u)P_l^{(-i\gamma,i\gamma)}(u')\\ &\quad -F_{i\gamma}(a/a')P_l^{(-i\gamma,i\gamma)}(u)P_l^{(i\gamma,-i\gamma)}(u'), \end{aligned}$$

and $\mathscr{E}_l$ and $\mathscr{L}_l$ are simple rational functions, given explicitly in Ref. 27. In particular, $\mathscr{E}_0=\mathscr{L}_0\equiv 0$, and for real k and γ we have $\operatorname{Im}\mathscr{L}_l=0$ and

$$\operatorname{Im}\mathscr{E}_l=-2\gamma^{-1}\operatorname{Im}P_l^{(i\gamma,-i\gamma)}(u)\operatorname{Im}P_l^{(i\gamma,-i\gamma)}(u').$$

From Eq. (8.1) we obtain an analog of (7.5) or (7.6), by using Eq. (7.4). Just as in Sec. 6, it is convenient to distinguish a complex part R_{cl}^{c} and a remaining *real* part R_{cl}^{r} (for $k>0$ and γ real)

$$\begin{aligned} &R_{cl}\equiv R_{cl}^{c}+R_{cl}^{r},\\ &-\gamma c_{l\gamma}^{-1}Q_l(w)R_{cl}^{c}\\ &\quad = C_0^2 e^{\pi\gamma}(-a')^{i\gamma}\\ &\qquad \times P_l^{(i\gamma,-i\gamma)}(u')\operatorname{Im}\{a^{i\gamma}P_l^{(i\gamma,-i\gamma)}(u)\},\\ &-\gamma c_{l\gamma}^{-1}Q_l(w)R_{cl}^{r}\\ &\quad = (\gamma/2)\operatorname{Re}\mathscr{E}_l+\gamma Q_0(w)\mathscr{L}_l\\ &\qquad +\operatorname{Im}P_l^{(i\gamma,-i\gamma)}(u)\operatorname{Re}P_l^{(i\gamma,-i\gamma)}(u') \qquad (8.2)\\ &\qquad +\operatorname{Im}\{P_l^{(-i\gamma,i\gamma)}(u')[P_l^{(-i\gamma,i\gamma)}(u)F_{i\gamma}(a'a)\\ &\qquad -P_l^{(i\gamma,-i\gamma)}(u)F_{i\gamma}(a'/a)]\},\\ &\quad a'<0<a; \quad vv'<0; \quad 0<p'<k<p; \quad \gamma\in\mathbb{R}. \end{aligned}$$

Note that

$$\lim_{a'\to 0}R_{cl}^{r}=0. \tag{8.3}$$

A representation analogous to (7.7) or (7.8) is obtained by replacing, in the expression for R_{cl}^{c}, $e^{\pi\gamma}(-a')^{i\gamma}$ by $(a')^{i\gamma}$; the resulting representation is valid for

$$0<a'<a; \quad vv'>0; \quad 0<k<p'<p; \quad \gamma\in\mathbb{R}. \tag{8.4}$$

In this case Eq. (8.3) is still valid.

9. SERIES REPRESENTATIONS FOR R_c VALID FOR NEGATIVE ENERGY

In this section we shall discuss some series representations for R_c that are valid for negative energy [see the remark on the allowed values of κ, at the end of this section]. In Sec. 10 analogous series representations for R_{cl} will be considered. The series to be discussed are known in the literature as Sturmian series, and also as Weinberg series,[38,39] see the references cited by Sloan.[34] Some of these series for the Coulomb T matrix mentioned in the literature are divergent, whereas for others "the convergence is conditional at best" (Ref. 34, p. 1016; see also p. 1020). As it does not seem to be known whether these series, in certain cases, are divergent or convergent—and if convergent, to which values—it is worthwhile to consider them here again. To be more specific: We shall prove, by using Abel's theorem on power series, that the series in Eqs. (9.1)–(9.3) and (10.1)–(10.3) do converge, and moreover to the correct values. For the Coulomb ratio R_c we have

$$R_c = 1+\frac{4}{1-x^2}\sum_{n=1}^{\infty}\frac{s/\kappa}{n-s/\kappa}C_{n-1}^{1}\left(\frac{x^2+1}{x^2-1}\right) \tag{9.1}$$

$$= \frac{4}{1-x^2}\lim_{t\uparrow 1}\sum_{n=1}^{\infty}t^n\frac{n}{n-s/\kappa}C_{n-1}^{1}\left(\frac{x^2+1}{x^2-1}\right) \tag{9.2}$$

$$= 1+2(1+\cos\varphi)\sum_{n=1}^{\infty}\left(\frac{s}{\kappa}\right)\left(n-\frac{s}{\kappa}\right)^{-1} \times C_{n-1}^{1}(-\cos\varphi), \tag{9.3}$$

where (9.3) is merely (9.1) in another notation, and

$$\begin{aligned} &\kappa\equiv -ik>0, \quad x^2<0,\\ &x^2 := -(p^2p'^2+\kappa^4+2\kappa^2pp'\cos\theta)/(\kappa^2q^2),\\ &q^2 := p^2+p'^2-2pp'\cos\theta,\\ &(x^2+1)/(x^2-1) =: -\cos\varphi, \quad 0<\varphi<\pi,\\ &-s/\kappa\equiv i\gamma\in\mathbb{R}, \quad -1+s/\kappa\notin\mathbb{N}. \end{aligned}$$

The connection with Eq. (2.1) is obtained by taking

$$x = i\tan\varphi/2; \quad y = (x+1)/(x-1) = -e^{i\varphi}, \quad 0<\varphi<\pi;$$

bear in mind that

$$p>0, \quad p'>0, \text{ and } q>0$$

is assumed throughout. Note further that

$$C^{\lambda}_{m}(-z) = (-1)^m C^{\lambda}_{m}(z), \tag{9.4}$$

and

$$C^{1}_{n-1}(\cos\varphi) = \sin n\varphi/\sin\varphi, \quad n = 1,2,..., \tag{9.5}$$

so that (9.3) with (9.5) inserted reminds us of the well-known formula

$$\sum_{n=1}^{\infty}(-1)^{n-1}\left(\frac{2}{n}\right)\sin n\varphi = \varphi, \quad -\pi<\varphi<\pi. \tag{9.6}$$

Clearly the limit $\varphi\to 0$ ($x\to 0$) may not be taken after the summation symbol in (9.3).

To prove Eqs. (9.1) and (9.3) one may use Eq. (2.1), which is easily rewritten as

$$R_c = 1 - 2i\gamma(1+\cos\varphi)\int_0^1 (t^2+1+2t\cos\varphi)^{-1}t^{i\gamma}\,dt. \tag{9.7}$$

To ensure convergence here we take

$$-i\gamma \equiv s/\kappa < 1. \tag{9.8}$$

Further we use the well-known formula

$$(t^2+1+2t\cos\varphi)^{-\lambda} = \sum_{n=1}^{\infty} C^{\lambda}_{n-1}(-\cos\varphi)t^{n-1}, \quad |t|<1, \ \lambda\neq 0. \tag{9.9}$$

Then it is straightforward to obtain[34]

$$R_c = 1 + 2\left(\frac{s}{\kappa}\right)(1+\cos\varphi)\lim_{t\uparrow 1}\sum_{n=1}^{\infty} t^n\left(n-\frac{s}{\kappa}\right)^{-1} \times C^{1}_{n-1}(-\cos\varphi). \tag{9.10}$$

It takes only one further step to derive (9.3) from (9.10). With this aim we invoke *Abel's theorem on power series.*

(i) If the power series $\Sigma_0^\infty a_n z^n$ converges for $z=c$, it converges absolutely for $|z|<|c|$.

(ii) If $\Sigma_0^\infty a_n z^n$ converges for $z=r>0$, then S, defined by

$$S(z) := \sum_{0}^{\infty} a_n z^n, \quad 0\leqslant z\leqslant r, \tag{9.11}$$

is continuous on $[0,r]$. This is most explicitly designated by "Abel's theorem on continuity up to the circle of convergence."

With this theorem and Eqs. (9.5) and (9.6) one easily verifies that the limit and the sum in Eq. (9.10) may be interchanged, which proves Eq. (9.3). Clearly the condition $s/\kappa<1$ of (9.8) may be relaxed in (9.10) and in (9.3). One must only avoid the values $s/\kappa = 1,2,3,...$, which correspond to the Coulomb bound-state poles of the T matrix. It is also clear that φ must be real with $0<\varphi<\pi$, cf. Eq. (9.5); this implies that either κ must be real or

$$\kappa^2 = pp'e^{i\alpha} (-\pi<\alpha<\pi) \,\&\, \cos\alpha + \cos\theta > 0. \tag{9.12}$$

Equation (9.2) follows similarly; here it is not allowable to interchange lim and Σ because the resulting infinite sum is clearly divergent.

10. SERIES REPRESENTATIONS FOR R_{cl} VALID FOR NEGATIVE ENERGY

In this section we shall discuss some Sturm-type series representations for R_{cl}, which are valid for negative energy only. They are the natural analogs of the series for R_c given in Sec. 9; there we have given some preliminary remarks about these Sturmian series. First we give the formulas and then their convergence and validity will be proved.

The partial-wave analogs of Eqs. (9.1)–(9.3) are

$$R_{cl} = 1 + (l!)^2[2Q_l(w)]^{-1}(-vv'/4)^{-l-1} \times \sum_{n=l+1}^{\infty}\frac{s/\kappa}{n-s/\kappa}\frac{(n-l-1)!}{(n+l)!} \times C^{l+1}_{n-l-1}\left(\frac{u}{v}\right)C^{l+1}_{n-l-1}\left(\frac{u'}{v'}\right), \tag{10.1}$$

$$R_{cl} = (l!)^2[2Q_l(w)]^{-1}(-vv'/4)^{-l-1} \times \sum_{n=l+1}^{\infty}\frac{n}{n-s/\kappa}\frac{(n-l-1)!}{(n+l)!} \times C^{l+1}_{n-l-1}\left(\frac{u}{v}\right)C^{l+1}_{n-l-1}\left(\frac{u'}{v'}\right), \tag{10.2}$$

which is essentially Eq. (12) of Ref. 27, p. 1274, and

$$R_{cl} = 1 + (l!)^2[2Q_l(w)]^{-1}(4\sin\chi\sin\chi')^{l+1} \times \sum_{n=l+1}^{\infty}\frac{s/\kappa}{n-s/\kappa}\frac{(n-l-1)!}{(n+l)!}C^{l+1}_{n-l-1}(-\cos\chi) \times C^{l+1}_{n-l-1}(\cos\chi'), \tag{10.3a}$$

where (10.3a) is just (10.1) rewritten, and

$$-e^{i\chi} = a = (p-i\kappa)/(p+i\kappa),$$
$$-\cos\chi = u/v = (p^2-\kappa^2)/(p^2+\kappa^2),$$
$$\sin\chi = 2p\kappa/(p^2+\kappa^2),$$
$$-\cos\chi' = u'/v' = (p'^2-\kappa^2)/(p'^2+\kappa^2),$$
$$\sin\chi' = 2p'\kappa/(p'^2+\kappa^2),$$
$$-1/vv' = \sin\chi\sin\chi' > 0,$$
$$0<\chi<\pi, \quad 0<\chi'<\pi, \quad \chi\neq\chi', \quad p\neq p', \quad u\neq u',$$
$$\kappa \equiv -ik > 0, \quad -s/\kappa \equiv i\gamma\in\mathbb{R}, \quad -l-1+s/\kappa\notin\mathbb{N}.$$

For $l=0$, Eq. (10.3a) reduces to

$$R_{c0} = 1 + \left(\frac{2}{Q_0(w)}\right)\sum_{n=1}^{\infty}\left(\frac{s}{n\kappa}\right)\left(n-\frac{s}{\kappa}\right)^{-1}\sin n\chi\sin n\chi'. \tag{10.3b}$$

As noted by Sloan,[34] it is convenient to apply the addition theorem for Gegenbauer polynomials.[41] Then it is easy to derive from, e.g., the integral representation (2.1), Eqs. (10.1)–(10.3) at an earlier stage, viz., these equations with the substitution

$$\sum_{n=l+1}^{\infty}\cdots \longrightarrow \lim_{t\uparrow 1}\sum_{n=l+1}^{\infty} t^n\cdots. \tag{10.4}$$

So it suffices to prove that limit and sum may be interchanged here. In other words, one has to prove the (conditional) convergence of the series in (10.1)–(10.3), and further that the series converge to the correct values; this problem was left open by Sloan.[34]

(i) We shall first prove that the infinite series in Eqs.

(10.1)–(10.3) are convergent. In fact, the sum in (10.2) is only conditionally convergent, whereas the sums in (10.1) and (10.3) are absolutely convergent [note that the terms of the sum in (10.2) are just a factor n bigger]. To this end we use a formula for the asymptotic behavior, for large n, of the Gegenbauer polynomials,[41]

$$C^{l+1}_{n-l-1}(\cos\chi) = (n/2)^l(l!)^{-1}(\sin\chi)^{-l-1}\sin(n\chi - \pi l/2) + O(n^{l-1}), \quad n\to\infty, \tag{10.5}$$

valid uniformly for $0<\epsilon\leqslant\chi\leqslant\pi-\epsilon<\pi$. By using Eqs. (9.4) and (10.5) we obtain from Eq. (10.2)

$$R_{cl}Q_l(w) = \sum_{n=l+1}^{\infty}\left(n-\frac{s}{\kappa}\right)^{-1} \times 2\sin\left(n\chi-\frac{\pi l}{2}\right)\sin\left(n\chi'-\frac{\pi l}{2}\right)\left[1+O\left(\frac{1}{n}\right)\right] \tag{10.6}$$

$$= \sum_{n=l+1}^{\infty}\left(n-\frac{s}{\kappa}\right)^{-1}[\cos(n\chi-n\chi') - (-1)^l\cos(n\chi+n\chi')]\left[1+O\left(\frac{1}{n}\right)\right]. \tag{10.7}$$

These series converge since $0<\chi<\pi, 0<\chi'<\pi$, and $\chi\neq\chi'$, cf.

$$\sum_{n=1}^{\infty} n^{-1}\cos(2n\varphi) = -\ln(2\sin\varphi), \quad 0<\varphi<\pi. \tag{10.8}$$

This proves that the sum in Eq. (10.2) is conditionally (but not absolutely) convergent. It follows easily that the sums in Eqs. (10.1) and (10.3) are absolutely convergent; their terms behave like n^{-2} for large n.

(ii) Second we shall prove that the sums in Eqs. (10.1)–(10.3) converge to the correct values, i.e., that the right members are equal to their respective left members. To this end we recall Abel's theorem on power series, which has been set out in Sec. 9. We denote the infinite sums including the factor t^n, see Eq. (10.4), in shorthand notation by $\Sigma A_n t^n$. In (i) we have proved that ΣA_n converges; with Abel's theorem we then have

$$\lim_{t\uparrow 1}\sum A_n t^n = \sum A_n. \tag{10.9}$$

Since the left member is equal to the left members of Eqs. (10.1), (10.2), and (10.3a), respectively, it follows that limit and sum may be interchanged.

This completes the proof of Eqs. (10.1)–(10.3).

For more details on convergence properties of the Sturmian expansion of the Coulomb T matrices, in particular near the special points $p=0, p'=0$, and $p=p'$, see Sloan[34] and references cited therein; for numerical results showing bad convergence of these expansions, see Chen and Ishihara.[36,37]

The residues of R_{cl} at $\kappa=s/n$, $n>l$, follow directly from Eq. (10.3a):

$$\lim_{\kappa\to s/n}(s/\kappa-n)R_{cl} = -n(l!)^2(n-l-1)![2(n+l)!Q_l(w)]^{-1} \times(4\sin\chi\sin\chi')^{l+1}C^{l+1}_{n-l-1}(\cos\chi) \times C^{l+1}_{n-l-1}(\cos\chi'), \tag{10.10}$$

$$\kappa=s/n, \quad n=l+1,\ l+2,\ldots.$$

Since $0<\chi<\pi$ and $0<\chi'<\pi$, this residue is different from zero when $n=l+1$ so that R_{cl} has always a pole at $\kappa=s/(l+1)$. However, the residue may vanish when $n>l+1$; this occurs at the well-known zeros of the Gegenbauer polynomial C^{l+1}_{n-l-1} on $(-1,1)$. For simplicity we shall consider the $l=0$ case: Then we have

$$\lim_{\kappa\to s/n}\left(\frac{s}{\kappa}-n\right)R_{c0} = \frac{-2\sin n\chi\sin n\chi'}{Q_0(w)}. \tag{10.11}$$

Defining

$$\chi_{m,n} := \pi m/n; \quad n>1; \quad m=1,2,\ldots,\ n-1, \tag{10.12}$$

we see that for $\chi=\chi_{m,n}$, and for $\chi'=\chi_{m,n}$, the residues of R_{c0} at $\kappa=s/n, s/2n, s/3n, \ldots$ are zero. At these points R_{c0} is continuous in κ. This remarkable fact indicates that no simple inequalities exist, apart from those to be enumerated in Sec. 24, for R_{cl} for negative energy in the case of attraction.

11. INTEGRAL REPRESENTATIONS FOR R_c VALID FOR NEGATIVE ENERGY

The integral representation for R_c given by Eq. (2.8) is not directly suitable for negative energy. In this section we shall modify this equation such that analytic continuation in particular to negative energy becomes possible; see Eq. (11.11). This representation may be useful for deriving inequalities for R_c for negative energy and $s>\kappa$, in the future; in the collection of inequalities in Sec. 24 we consider only $s<\kappa$ (note that R_c has a simple pole at $s=n\kappa$, $n=1,2,\ldots$).

We begin by inserting

$$C_0^2 e^{\pi\gamma} = \pi\gamma/\sinh\pi\gamma, \tag{11.1}$$

$$y=-e^{-\delta}, \ \delta>0$$

into Eq. (2.8), which yields

$$R_c\sinh\pi\gamma = \sinh\pi\gamma + \pi\gamma(-y)^{i\gamma}(1-y)/(1+y) - \gamma(1-y)^2\int_0^{\pi}(1+y^2+2y\cos t)^{-1} \times\cosh\gamma t\,dt\ (-1<y<0) \tag{11.2}$$

$$= \sinh\pi\gamma + \pi\gamma e^{-i\gamma\delta}(e^{\delta}+1)/(e^{\delta}-1) - \gamma(1+\cosh\delta)\int_0^{\pi}(\cosh\delta-\cos t)^{-1} \times\cosh\gamma t\,dt\ (\delta>0). \tag{11.3}$$

We shall use the auxiliary formulas[41,42]

$$\gamma\int_0^{\pi}\cosh\gamma t\,dt = \sinh\pi\gamma, \tag{11.4}$$

$$\int_0^{\pi}(\cosh\delta\pm\cos t)^{-1}dt = \frac{\pi}{\sinh\delta}, \quad \delta>0, \tag{11.5}$$

and

$$\int_0^\pi (\cosh\delta - \cos t)^{-1}(1+\cos t)dt = \frac{2\pi}{(e^\delta - 1)}, \quad \delta > 0. \tag{11.6}$$

Then we obtain from Eq. (11.3)

$$R_c = 1 + \gamma(\sinh \pi\gamma)^{-1}(1+\cosh\delta) \times \int_0^\pi (\cosh\delta - \cos t)^{-1}(e^{-i\gamma\delta} - \cosh\gamma t)dt, \tag{11.7}$$

and

$$R_c \sinh\pi\gamma = \gamma\int_0^\pi (\cosh\delta - \cos t)^{-1}(1+\cos t) \times [e^{-i\gamma\delta}(e^\delta + 1)/2 - \cosh\gamma t]dt. \tag{11.8}$$

We wish to perform analytic continuation to arbitrary negative energy. Then δ becomes purely imaginary, so that the integrand in Eq. (11.8) becomes singular (nonintegrable). We remove this singularity beforehand, by adding the term $\cos\gamma\delta - \cos\gamma\delta \equiv 0$ inside the square brackets on the right-hand side of Eq. (11.8). By using Eq. (11.6) we obtain in this way

$$\begin{aligned} R_c(\pi\gamma)^{-1}\sinh\pi\gamma &= (e^\delta - 1)^{-1}(e^{\delta - i\gamma\delta} - e^{i\gamma\delta}) \\ &\quad + \pi^{-1}\int_0^\pi (1+\cos t)(\cos\gamma\delta - \cosh\gamma t) \\ &\quad \times(\cosh\delta - \cos t)^{-1}\,dt. \end{aligned} \tag{11.9}$$

Now we are in a position to perform analytic continuation to negative energy: $\kappa \equiv -ik > 0$. Then x and δ become purely imaginary, and $|y| = 1$. Putting (cf. Sec. 9)

$$\begin{aligned} &\alpha := i\gamma \in \mathbb{R}\setminus\mathbb{Z}, \quad x = i\tan\varphi/2, \\ &y = (x+1)/(x-1) = -e^{i\varphi} = -e^{-\delta}, \\ &\delta = -\ln(-y) = -i\varphi, \quad 0 < \varphi < \pi, \end{aligned} \tag{11.10}$$

we get at once from (11.9)

$$R_c = \frac{\pi\alpha}{\sin\pi\alpha}\left[\frac{\sin(\frac{1}{2}-\alpha)\varphi}{\sin\frac{1}{2}\varphi} + \frac{1}{\pi}\int_0^\pi (1+\cos t)\frac{\cos\alpha\varphi - \cos\alpha t}{\cos\varphi - \cos t}\,dt\right], \tag{11.11}$$

$0 < \varphi < \pi.$

This is a very convenient representation valid for negative energy; only those values of κ for which $\alpha \equiv i\gamma \equiv -s/\kappa$ is an integer are to be excluded. Note that the integrand in (11.11) is continuous when the isolated singularity at $t = \varphi$ is removed. As is well known, R_c has bound-state poles (only) at $s/\kappa = 1,2,3,\ldots$. So we expect that the singularities of the representation (11.11) at $\alpha \equiv -s/\kappa = 0,1,2,\ldots$ can be removed. It can be seen upon inspection that

$$\lim_{\alpha\to 0} R_c = 1, \tag{11.12}$$

as expected. Now let $n \in \mathbb{N}' := \{1,2,\ldots\}$; then we have to prove that

$$\frac{\pi\sin(\frac{1}{2}-n)\varphi}{\sin\frac{1}{2}\varphi} + \int_0^\pi (1+\cos t) \times(\cos n\varphi - \cos nt)/(\cos\varphi - \cos t)dt = 0, \quad n\in\mathbb{N}'. \tag{11.13}$$

One easily verifies that (11.13) is equivalent to

$$\int_0^\pi \frac{(\cos n\varphi - \cos nt)}{(\cos\varphi - \cos t)}dt = \pi\sin n\varphi/\sin\varphi, \; n\in\mathbb{N}, \; 0<\varphi<\pi, \tag{11.14}$$

which equality is proved in Appendix E of Ref. 32.

It is interesting to observe that the equality (11.13) can be used to derive the residue of R_c at $\alpha = -n$ from Eq. (11.11):

$$\begin{aligned} \lim_{s/\kappa\to n}\left(\frac{s}{\kappa} - n\right)R_c &= \lim_{\alpha\to -n} -(\alpha+n)R_c \\ &= (-1)^n n\left[\frac{\sin(\frac{1}{2}+n)\varphi}{\sin\frac{1}{2}\varphi} + \pi^{-1}\int_0^\pi \cdots\right] \\ &= (-1)^n n[\sin(\tfrac{1}{2}+n)\varphi - \sin(\tfrac{1}{2}-n)\varphi]/\sin\tfrac{1}{2}\varphi \\ &= (-1)^n 2n\sin n\varphi\cot\tfrac{1}{2}\varphi. \end{aligned} \tag{11.15}$$

The same expression for this residue is easily obtained from Eq. (9.3), with the aid of Eqs. (9.4) and (9.5)

$$C^1_{n-1}(-\cos\varphi) = (-1)^{n-1}\sin n\varphi/\sin\varphi.$$

The residue of R_c given by (11.15) is different from zero when $n = 1$ so that R_c has always a pole at $\kappa = s$. However, the residue may vanish when $n > 1$. Defining

$$\varphi_{m,n} := \pi m/n; \quad n > 1; \; m = 1,2,\ldots, n-1, \tag{11.16}$$

we have from Eq. (9.3),

$$R_c(\varphi = \varphi_{m,n}) = 1 - 2\cot(\pi m/2n) \times \sum_{N=1}^{\infty}{}'(-1)^N s(\kappa N - s)^{-1}\sin\left(\frac{\pi N m}{n}\right), \tag{11.17}$$

where the prime indicates that the terms for $N = n,2n,3n,\ldots$ are to be omitted. Clearly R_c is continuous in κ at $\kappa = s/n$, $s/2n$, $s/3n,\ldots$ in this case. This remarkable fact indicates that no simple inequalities exist, apart from those to be given in Sec. 24, for R_c for negative energy in the case of attraction.

12. ANALYTIC CONTINUATION OF REPRESENTATIONS FOR R_c AND R_{cl}

In this section we shall discuss the analytic continuation with respect to k of representations for R_c and R_{cl}. Several representations given in preceding sections are valid for $\mathrm{Re}\, i\gamma > -1$ and $\mathrm{Re}\, i\gamma > -l-1$, respectively (and $\mathrm{Im}\, k \geqslant 0$). Further analytic continuation is in the first instance prevented by the Coulomb bound-state poles at $i\gamma = -n$, where $n = 1,2,\ldots$ for R_c and $n = l+1, l+2,\ldots$ for R_{cl}. These poles are easily recognizable in the appropriate formulas below, where the residues at these poles will also be given. So in the complex k plane we have the following.

(i) Poles at $k = is/n$; for R_c: $n = 1,2,\ldots$, for R_{cl}: $n = l+1, l+2\ldots$.

The remaining singularities of *both* R_c and R_{cl} in the complex k plane consist of the following.

(ii) An essential singularity at $k=0$.

(iii) Branch points at $k=p$, $k=-p$, $k=p'$, and $k=-p'$.

(iv) Branch cuts connecting these branch points.

When $0<p'<p$ the most convenient choice for these branch cuts is $(-\infty,-p]$, $[-p',0]$, $[0,p']$, and $[p,\infty)$.

Throughout this section we assume that $0<p'<p$ and we shall denote the four branch cuts, excluding the branch points, by $\delta_{\pm 1}$ and $\delta_{\pm 2}$ according to:

$$\begin{aligned} &\delta_{-2}: k\in(-\infty,-p), \quad && \delta_{-1}: k\in(-p',0), \\ &\delta_{2}: k\in(p,\infty), && \delta_{+1}: k\in(0,p'). \end{aligned} \tag{12.1}$$

(Note that $p=p'$ is allowed only for R_c, provided $\theta>0$.) The discontinuities of R_c and R_{cl} across these branch cuts will be given below.

This section is the only one where we take the *lower* half of the complex k plane into account, Im $k<0$. In all other sections we restrict ourselves to Im $k\geqslant 0$, Re $k\geqslant 0$, and $k\neq 0$, and for $k>0$ the limit Im $k\downarrow 0$ is implicitly understood.

Remark: We indicate those variables of the Coulomb ratios that are useful for the discussion in question. For instance, below we shall discuss $R_c\equiv R_c(s,k)$, whereas in Sec. 16 we shall deal with $R_c(\rho;\gamma)$. Although this notation is not consistent, it is the most convenient, and in the right context no misunderstanding is possible. Note further that $\gamma\equiv -s/k$ where s and k are the basic, independent variables.

Analytic continuation of representations for R_c and R_{cl} follow easily from their hypergeometric-function representations; see Secs. 7 and 8, respectively. A particularly useful equality in this connection is

$$\begin{aligned} &F_{i\gamma}(z)+F_{-i\gamma}(1/z) \\ &\quad =1+(\pi\gamma/\sinh\pi\gamma)(-z)^{-i\gamma}, \quad z\notin[0,\infty). \end{aligned}$$

Alternatively one can use the series representations given by Eqs. (5.1) and (4.4) [cf. Eq. (3.16)], where one should also be aware of the branch cuts of $Q_l^{i\gamma}(u)$ and $P_l^{-i\gamma}(u')$, which are given by $|k|=p$ and $|k|=p'$, respectively. Next we are going to give some interesting results that were derived in Appendix G of Ref. 32, to which the reader is referred for more details.

For R_c we have

$$\begin{aligned} R_c(s)-R_c(-s)=(\pi\gamma/\sinh\pi\gamma) \\ \times(1-y)(1+y)^{-1}[(-y)^{i\gamma}-(-y)^{-i\gamma}], \end{aligned} \tag{12.2}$$

which expression has a branch cut for $0<y<1$, i.e., on $\delta_{\pm 1}$ and $\delta_{\pm 2}$. When $-1<y<0$ (the "symmetric" case, cf. Sec. 23) R_c has no branch cut and we have

$$R_c(s)-R_c(-s)=R_c(s)-R_c^{*}(s), \quad -1<y<0. \tag{12.3}$$

When the energy is *negative* we have $|y|=1$ so that we may put

$$y=-e^{i\varphi}, \quad 0<\varphi<\pi. \tag{12.4}$$

Putting further

$$\alpha:=i\gamma=-s/\kappa\in\mathbb{R}\setminus\mathbb{Z}, \tag{12.5}$$

we have from Eq. (12.2)

$$R_c(s)-R_c(-s)=-2\pi\alpha(\sin\pi\alpha)^{-1}\sin\varphi\alpha\cot(\varphi/2). \tag{12.6}$$

This result also follows easily from Eq. (11.11). The residues at the poles of R_c follow directly from Eq. (12.6) [cf. (11.15)]

$$\lim_{\kappa\to s/n}\left(\frac{s}{\kappa}-n\right)R_c=(-1)^n 2n\sin n\varphi\cot\left(\frac{\varphi}{2}\right), \tag{12.7}$$

$\kappa=s/n$, $n=1,2,\ldots$.

The discontinuities across the branch cuts δ_i also follow from the hypergeometric-function representations. We define

$$\begin{aligned} &\Delta(s,k):=2\pi\gamma x^{-1}|y|^{i\gamma}, \\ &k\in\mathbb{R},\ k\neq 0,\ k\neq\pm p,\ k\neq\pm p'. \end{aligned} \tag{12.8}$$

Then Δ is invariant under $p\leftrightarrow p'$ and

$$\begin{aligned} &\lim_{\epsilon\downarrow 0}[R_c(s,k+i\epsilon)-R_c(s,k-i\epsilon)] \\ &\quad=\begin{cases} \Delta(s,k) & \text{for } k\in\delta_1 \text{ or } k\in\delta_{-2}, \\ -\Delta(s,k) & \text{for } k\in\delta_{-1} \text{ or } k\in\delta_2. \end{cases} \end{aligned} \tag{12.9}$$

Further, for $p'<k<p$ we have

$$\begin{aligned} &\Delta(s,k)+\Delta(-s,k) \\ &\quad=2\pi\gamma x^{-1}(|y|^{i\gamma}-|y|^{-i\gamma}) \\ &\quad=-2\sinh\pi\gamma[R_c(s)-R_c(-s)], \quad p'<k<p. \end{aligned} \tag{12.10}$$

It may be noted that

$$\Delta(-s,-k)=\Delta(s,k). \tag{12.11}$$

For the Coulomb ratios R_{cl} we have obtained[32] [in analogy with (12.2)]

$$\begin{aligned} &R_{cl}(s)-R_{cl}(-s) \\ &\quad=-2i\gamma Q_l^{i\gamma}(u)Q_l^{-i\gamma}(u')/Q_l(w), \quad p'<|k|<p \\ &\quad\equiv -2i\gamma Q_l^{-i\gamma}(u)Q_l^{i\gamma}(u')/Q_l(w). \end{aligned} \tag{12.12}$$

(The identity here is valid whenever these functions are defined.) The condition $p'<|k|<p$ stems from the branch cuts of $Q_l^{\pm i\gamma}(u)$ and $Q_l^{\pm i\gamma}(u')$, which are $|k|=p$ and $|k|=p'$, respectively. An expression with broader validity can be given in terms of the Jacobi polynomials $P_l^{(i\gamma,-i\gamma)}$, see Appendix G of Ref. 32. When $p'<k<p$, R_{cl} has no branch cut and we get the interesting relation

$$R_{cl}(s)-R_{cl}(-s)=R_{cl}(s)-R_{cl}^{*}(s), \quad p'<k<p, \tag{12.13}$$

which is the natural analog of (12.3).

The residues at the poles follow from (12.12), but more readily from Eqs. (10.1) or (10.2) [cf. (10.10)]

$$\begin{aligned} \lim_{\kappa\to s/n}\left(\frac{s}{\kappa}-n\right)R_{cl}=&-n(l!)^2(n-l-1)![2(n+l)!Q_l(w)]^{-1} \\ &\times(-4/vv')^{l+1}C_{n-l-1}^{l+1}(u/v) \\ &\times C_{n-l-1}^{l+1}(u'/v'), \quad \kappa=s/n, \\ &n=l+1,l+2,\ldots . \end{aligned} \tag{12.14}$$

The discontinuities of R_{cl} across the branch cuts δ_i follow from the hypergeometric-function representations given in Sec. 8. We define

$$\begin{aligned} &\Delta_l(s,k,p,p'):=-2\pi\gamma e^{\pi\gamma}P_l^{-i\gamma}(u')Q_l^{i\gamma}(u)/Q_l(w), \\ &k\in\mathbb{R},\ k\neq 0,\ k\neq\pm p,\ k\neq\pm p',\ p>0,\ p'>0. \end{aligned} \tag{12.15}$$

Then Δ_l is *not* invariant under $p \leftrightarrow p'$; it is even in k,

$$\Delta_l(s,k,p,p') = \Delta_l(s,-k,p,p'), \tag{12.16}$$

and for $l=0$ we find

$$\Delta_0(s,k,p,p') = 2\pi[Q_0(w)]^{-1}|a'|^{i\gamma}\sin(\gamma \ln|a|). \tag{12.17}$$

Across the δ_i we have [bear in mind (12.1) and $0<p'<p$]

$$\begin{aligned}\lim_{\epsilon \downarrow 0}&[R_{cl}(s,k+i\epsilon) - R_{cl}(s,k-i\epsilon)] \\ &= \Delta_l(s,k,p,p') \text{ for } k\in\delta_1; \\ &= -\Delta_l(-s,k,p,p') \text{ for } k\in\delta_{-1}; \\ &= -\Delta_l(s,k,p',p) \text{ for } k\in\delta_2; \\ &= \Delta_l(-s,k,p',p) \text{ for } k\in\delta_{-2}.\end{aligned} \tag{12.18}$$

Interestingly enough, the partial-wave analog of Eq. (12.10) has a very similar form:

$$\begin{aligned}&\Delta_l(s,k,p,p') + \Delta_l(-s,k,p,p') \\ &\quad = -2\sinh\pi\gamma[R_{cl}(s) - R_{cl}(-s)], \quad p'<k<p.\end{aligned} \tag{12.19}$$

Finally we give some interesting limits:

$$\lim_{p\to\infty} \Delta(s,k) = \pm 2\pi\left(\frac{s}{p'}\right)|a'|^{\mp is/k}, \quad k\gtrless 0, \tag{12.20}$$

and

$$\lim_{p'\to\infty} p' \lim_{p\to\infty} \Delta(s,k) = 2\pi s \operatorname{sgn}(k). \tag{12.21}$$

The factor $\operatorname{sgn}(k)$ arises from the convention[18] that $\operatorname{Re} x \leqslant 0$. It would disappear if we were to follow the convention

$$x \lesseqgtr 0 \text{ if } k \gtrless 0, \tag{12.22}$$

in which case $\Delta(s,k)$ would be *even* in k, just as $\Delta_l(s,k,p,p')$.

For Δ_l we have obtained [cf. (22.18) and (22.39)]

$$\begin{aligned}\lim_{p\to\infty} &\Delta_l(s,k,p,p') \\ &= -2\pi\gamma(k/p')^{l+1}P_l^{-i\gamma}(u')\Gamma(l+1+i\gamma)/l!\end{aligned} \tag{12.23}$$

and

$$\lim_{p'\to\infty} p' \lim_{p\to\infty} \Delta_l(s,k,p,p') = 2\pi s 4^{-l}\binom{2l}{l}, \tag{12.24}$$

which is the natural partial-wave analog of Eq. (12.21).

13. THE COULOMB BOUNDARY FUNCTION B_{cl}

The Coulomb boundary function B_{cl} is a limiting case of the Coulomb ratios R_{cl} (for all l) and R_c (for $l=0$). We expect that it will play an important role in inequalities for R_c and R_{cl} [see the conjectures (c1) and (c2) in Sec. 17]. In this section we shall briefly discuss the function B_{cl}. For a more extensive study of B_{cl} the reader is referred to Appendix H of Ref. 32.

We define the function B_{cl}, $l=0,1,\ldots$ by

$$B_{cl}(z) = 1 - 2z\int_0^\infty e^{-2zt}(\cosh t)^{-2l-2}\,dt,$$

$$\operatorname{Re} z > -l-1. \tag{13.1}$$

Analytic continuation with respect to z will be considered below. Several other representations for B_{cl} can be derived from Eq. (13.1); for example, through integration by parts we get

$$B_{cl}(z) = (2l+2)\int_0^\infty e^{-2zt}(\cosh t)^{-2l-3}\sinh t\,dt,$$

$$\operatorname{Re} z > -l-1.$$

According to Eq. (22.11) we have

$$B_{cl}(i\gamma) = \lim_{p\to\infty}\lim_{p'\to 0} R_{cl}, \tag{13.2}$$

where p and p' may be interchanged, and the order of the limits may be reversed. For $l=0$ we have [see Eqs. (21.9) and (21.10)]

$$B_{c0}(i\gamma) = \lim_{y\downarrow -1} R_c, \tag{13.3}$$

where $y\downarrow -1$ corresponds to either

$$pp' \to k^2 \& \theta \to 0,$$

or

$$p\to 0 \& p'\to\infty \text{ or } p'\to 0 \& p\to\infty, \quad \text{any } \theta. \tag{13.4}$$

It is noteworthy that B_{c0} is a limiting case of R_c as well as $R_{cl=0}$. We mention further the integral representations

$$\begin{aligned}B_{c0}(i\gamma) &= 1 - (\pi\gamma/\sinh\pi\gamma)\Big[2i\gamma + (2/\pi) \\ &\quad\times\int_0^\pi(\cosh\gamma t - 1)(1-\cos t)^{-1}\,dt\Big]\end{aligned} \tag{13.5}$$

$$\begin{aligned}&= (\pi\gamma/\sinh\pi\gamma)\Big[-2i\gamma + 1 \\ &\quad - \frac{4}{\pi}\int_0^{\pi/2}\cot^2 t\sinh^2\gamma t\,dt\Big].\end{aligned} \tag{13.6}$$

These results follow from Eq. (2.8) by evaluating the limit in (13.3), and also from Eqs. (3.14) and (3.15) by performing the limit in (13.2), for $l=0$. Application of the limit in (13.2) to Eq. (3.14) seems impossible for $l>0$. However, we have been able to derive the general $-l$ analogs of Eqs. (13.5) and (13.6) in a different way; see Eq. (13.22) below.

Interesting series representations for B_{c0} are

$$\begin{aligned}B_{c0}(i\gamma) &= 1 - 2i\gamma + 4\gamma^2\sum_{n=1}^\infty(-1)^n(n+i\gamma)^{-1} \\ &= -2i\gamma\pi\gamma/\sinh\pi\gamma + 1 - 4\gamma^2\end{aligned} \tag{13.7}$$

$$\begin{aligned}&\times\sum_{n=1}^\infty[2n(2n-1)-\gamma^2] \\ &\times[\{(2n-1)^2+\gamma^2\} \\ &\times(4n^2+\gamma^2)]^{-1}.\end{aligned} \tag{13.8}$$

By using these expressions we have been able to prove the conjecture (c2)′ for $l=0$, i.e.,

$$\pi\gamma/\sinh\pi\gamma < |B_{c0}(i\gamma)|, \quad \gamma>0.$$

Here we shall give certain relevant information about B_{cl} for $l=0,1,\ldots$. Proofs can be found in Ref. 32.

Several hypergeometric-function representations follow from Eq. (13.1), e.g.,

$$\begin{aligned}B_{cl}(z) &= 1 - 2z(z+l+1)^{-1}{}_2F_1(1,z-l;z+l+2;-1) \\ &= (2l+2)(z+l+1)^{-1}(z+l+2)^{-1} \\ &\quad\times{}_2F_1(2,z-l;z+l+3;-1).\end{aligned} \tag{13.9}$$

From these and similar relations one can easily obtain

various series and integral representations, and the equality

$$B_{cl}(-z) = B_{cl}(z) + 4^{l+1} z B(l+1+z, l+1-z), \tag{13.10}$$

which is useful for analytic continuation. From this relation one finds

$$\operatorname{Im} B_{cl}(i\gamma) = -2\gamma 4^l \Gamma(l+1+i\gamma)\Gamma(l+1-i\gamma)/(2l+1)!, \quad \gamma \in \mathbb{R}. \tag{13.11}$$

The asymptotic behavior of $B_{cl}(i\gamma)$ follows from Eq. (13.1) by integrating by parts

$$B_{cl}(i\gamma) = -(l+1)(2\gamma^2)^{-1} - (l+1)(1+3l/4)\gamma^{-4} + O(\gamma^{-6}), \quad \gamma \to \infty. \tag{13.12}$$

Further,

$$\operatorname{Re} B_{cl}(i\gamma) = 1 - 4b_l\gamma^2 + O(\gamma^4), \quad \gamma \to 0,$$

$$b_0 = \ln 2, \quad \lim_{l\to\infty} 2lb_l = 1, \tag{13.13}$$

and b_l is strictly decreasing in l. The numbers b_l can be calculated from the recursive relation

$$(2l+1)b_l = -(2l)^{-1} + 2lb_{l-1}, \quad l = 1,2,..., \tag{13.14}$$

which follows from the recursive relation for B_{cl}

$$(2l+1)B_{cl}(i\gamma) = 1 + 2(l+\gamma^2/l)B_{c,l-1}(i\gamma), \quad l = 1,2,... . \tag{13.15}$$

The quantity $|\Gamma(l+1+i\gamma)|^{-2} \operatorname{Re} B_{cl}(i\gamma)$ is strictly decreasing in γ on $[0,\infty)$, from $(l!)^{-2}$ to $-\infty$. Hence, $\operatorname{Re} B_{cl}(i\gamma)$ possesses one and only one positive zero γ, which we denote by γ_l,

$$\operatorname{Re} B_{cl}(i\gamma_l) = 0, \quad l = 0,1,... . \tag{13.16}$$

We have found that

$$\gamma_0 \lesssim 0.90, \tag{13.17}$$

$$(2l+1)/4 < \gamma_l^2 < l+1, \quad l = 0,1,..., \tag{13.18}$$

and

$$\lim_{l\to\infty} l^{-1}\gamma_l^2 = z_\infty \gtrsim 0.855, \tag{13.19}$$

where z_∞ is the only positive zero z of the function

$${}_1F_1(-\tfrac{1}{2};\tfrac{1}{2};z) \equiv e^z[1 - 2z\,{}_1F_1(1;\tfrac{3}{2};-z)]. \tag{13.20}$$

Further,

$$\lim_{l\to\infty} \operatorname{Re} B_{cl}(i\sqrt{zl}) = e^{-z}\,{}_1F_1(-\tfrac{1}{2};\tfrac{1}{2};z), \quad z > 0. \tag{13.21}$$

The generalization to $l = 0, 1,...$ of the integral representation (13.6) reads

$$B_{cl}(i\gamma)/B(l+1+i\gamma, l+1-i\gamma) = (2l+1)!(l!)^{-2} - 4^l\left[2i\gamma + \frac{4}{\pi}\int_0^{\pi/2} (\sinh\gamma t/\sin t)^2(\cos t)^{2l+2}\,dt\right], \quad l = 0,1,... . \tag{13.22}$$

Representations for $|B_{cl}(i\gamma)|^2$ can be given analogously with those for $|R_{cl}|^2$. In this connection we mention the interesting equality

$$\int_0^\infty |B_{cl}(i\gamma)|^2\,d\gamma = \pi(l+1)^2 B\left(2l+2, \frac{3}{2}\right). \tag{13.23}$$

14. REPRESENTATIONS FOR $|R_c|$

In this section we shall give some representations for $|R_c|^2$, valid for positive energy in the "symmetric" case (cf. Sec. 23)

$$k > 0, \quad \gamma \in \mathbb{R}, \qquad -1 \leqslant y < 0; \quad \rho := (1+y^2)/(-2y) \geqslant 1. \tag{14.1}$$

In this case $|R_c|$ is even in γ. The representations given below may be useful for deriving inequalities for $|R_c|$.

By using Eq. (2.1),

$$R_c = 1 - i\gamma(\rho+1)\int_0^\infty \frac{e^{-i\gamma t}dt}{(\rho+\cosh t)}, \tag{14.2}$$

and after performing integration by parts we obtain

$$R_c = \int_0^\infty g(t)e^{-i\gamma t}\,dt, \quad \gamma \in \mathbb{R}, \tag{14.3}$$

where

$$g(t) := (\rho+1)\sinh t\,(\rho+\cosh t)^{-2}, \quad t \geqslant 0; \tag{14.4}$$

we define $g(t) := 0$ for $t < 0$. Clearly g is a real and continuous function; so it can be seen by inspection that $|R_c|$ is even in γ. It follows that (cf. Appendix F of Ref. 32)

$$g(t) = (2\pi)^{-1}\int_{-\infty}^\infty R_c(\gamma)e^{i\gamma t}\,d\gamma, \quad t \in \mathbb{R}, \tag{14.5}$$

where $R_c(\gamma) \equiv R_c$, and that

$$|R_c|^2 = 2a\int_0^\infty \cos(a\gamma x)\int_0^\infty g(t)g(t+ax)dt\,dx = 2a\int_0^\infty g(t)\int_0^\infty \cos(a\gamma x)g(t+ax)dx\,dt, \tag{14.6}$$

$a > 0$, $\gamma \in \mathbb{R}$, $g(t) \in \mathbb{R}$.

Here a is an arbitrary positive parameter. Taking $a = 1$ and inserting Eq. (14.4) we obtain

$$|R_c|^2 = 2(\rho+1)^2\int_0^\infty \cos\gamma x\int_0^\infty \sinh t \times \sinh(t+x)(\rho+\cosh t)^{-2} \times [\rho+\cosh(t+x)]^{-2}\,dt\,dx. \tag{14.7}$$

By integrating by parts we obtain some related representations for $|R_c|^2$. Note in particular that

$$2(\rho+1)^2\int_0^\infty \sinh t\,(\rho+\cosh t)^{-2} \times \int_0^\infty \sinh(t+x)[\rho+\cosh(t+x)]^{-2}\,dx\,dt = 1, \tag{14.8}$$

which corresponds to $R_c = 1$ for $\gamma = 0$. By using (14.8) we obtain from (14.7)

$$|R_c|^2 = 1 - 4(\rho+1)^2 \int_0^\infty \sin^2\left(\frac{\gamma x}{2}\right) \int_0^\infty \sinh t \sinh(t+x) \times(\rho+\cosh t)^{-2}[\rho+\cosh(t+x)]^{-2}\, dt\, dx \quad (14.9)$$

$$= -1 + 4(\rho+1)^2 \int_0^\infty \cos^2\left(\frac{\gamma x}{2}\right) \times \int_0^\infty \sinh t \sinh(t+x)\,(\rho+\cosh t)^{-2} \times[\rho+\cosh(t+x)]^{-2}\, dt\, dx. \quad (14.10)$$

Further we have derived from Eq. (14.7)

$$|R_c|^2 = 1 - 2\gamma(\rho+1)^2 \int_0^\infty \sin\gamma x \int_0^\infty \sinh t \times(\rho+\cosh t)^{-2}[\rho+\cosh(t+x)]^{-1}\, dt\, dx, \quad (14.11)$$

and

$$|R_c|^2 = 1 - 2\gamma(\rho+1)\int_0^\infty \sin\gamma x(\rho+\cosh x)^{-1}\, dx + 2\gamma^2(\rho+1)^2 \int_0^\infty \cos\gamma x \int_0^\infty (\rho+\cosh t)^{-1} \times[\rho+\cosh(t+x)]^{-1}\, dt\, dx. \quad (14.12)$$

All these representations are valid under the conditions given by Eq. (14.1).

15. REPRESENTATIONS FOR $|R_{cl}|$

This section runs parallel to Sec. 14. Here we shall give some representations for $|R_{cl}|^2$, valid for positive energy in the "symmetric" case [cf. Sec. 23]

$$k>0, \quad \gamma\in\mathbb{R}, \quad vv'<0. \quad (15.1)$$

In this case $|R_{cl}|$ is even in γ. The representations given below may be useful for deriving inequalities for $|R_{cl}|$.

We begin by Eq. (3.1),

$$R_{cl}Q_l(w) = Q_l(w) - i\gamma\int_0^1 \tau^{i\gamma-1}Q_l(\zeta)d\tau, \quad (15.2)$$

$$\zeta := uu' - vv'(\tau+1/\tau)/2 = uu' - vv'\cosh t,$$

where $\tau = e^{-t}$. With the new variable of integration t, we obtain through integration by parts,

$$R_{cl} = \int_0^\infty g_l(t)e^{-i\gamma t}\, dt, \quad \gamma\in\mathbb{R}, \quad (15.3)$$

where

$$g_l(t) := vv'[Q_l(w)]^{-1}\sinh tQ_l'(uu' - vv'\cosh t), \quad t>0; \quad (15.4)$$

we define $g_l(t) := 0$ for $t<0$. Note that only in Q_l' the prime indicates the derivative; further, $uu' - vv' = w$, and $\zeta > 1$ when $vv' < 0$ so that $Q_l(\zeta)$ is *real* in this case. It is clear that g_l is a real and continuous function; so it can be seen by inspection that $|R_{cl}|$ is even in γ. It follows that (cf. Appendix F of Ref. 32)

$$g_l(t) = (2\pi)^{-1}\int_{-\infty}^\infty R_{cl}(\gamma)e^{i\gamma t}\, d\gamma, \quad t\in\mathbb{R}, \quad (15.5)$$

where

$$R_{cl}(\gamma)\equiv R_{cl},$$

and that

$$|R_{cl}|^2 = 2a\int_0^\infty \cos(a\gamma x)\int_0^\infty g_l(t)g_l(t+ax)dt\, dx = 2a\int_0^\infty g_l(t)\int_0^\infty \cos(a\gamma x)g_l(t+ax)dx\, dt, \quad (15.6)$$

$$a>0, \quad \gamma\in\mathbb{R}, g_l(t)\in\mathbb{R},$$

where a is an arbitrary positive parameter. Taking $a = 1$ and inserting Eq. (15.4) we obtain

$$|R_{cl}|^2 = 2(vv')^2[Q_l(w)]^{-2}\int_0^\infty \cos\gamma x \int_0^\infty \sinh t \sinh(t+x) \times Q_l'(uu' - vv'\cosh t) \times Q_l'(uu' - vv'\cosh(t+x))dt\, dx. \quad (15.7)$$

Just as in Sec. 14, related representations follow by performing integration by parts. In particular we have

$$2(vv')^2[Q_l(w)]^{-2}\int_0^\infty \sinh tQ_l'(uu' - vv'\cosh t) \times\int_0^\infty \sinh(t+x)Q_l'((uu' - vv'\cosh(t+x))dx\, dt = 1, \quad (15.8)$$

which corresponds to $R_{cl} = 1$ for $\gamma = 0$. By using (15.8) we obtain from (15.7)

$$|R_{cl}|^2 = 1 - 4\left[\frac{vv'}{Q_l(w)}\right]^2\int_0^\infty \sin^2\left(\frac{\gamma x}{2}\right)\int_0^\infty \sinh t \sinh(t+x)Q_l'(uu' - vv'\cosh t)Q_l'(uu' - vv'\cosh(t+x))dt\, dx \quad (15.9)$$

$$= -1 + 4\left[\frac{vv'}{Q_l(w)}\right]^2\int_0^\infty \cos^2\left(\frac{\gamma x}{2}\right)\int_0^\infty \sinh t \sinh(t+x)Q_l'(uu' - vv'\cosh t)Q_l'(uu' - vv'\cosh(t+x))dt\, dx. \quad (15.10)$$

Two different representations following from Eq. (15.7) are

$$|R_{cl}|^2 = 1 - 2\gamma vv'[Q_l(w)]^{-2}\int_0^\infty \sin\gamma x\int_0^\infty \sinh t \times Q_l'(uu' - vv'\cosh t) \times Q_l(uu' - vv'\cosh(t+x))dt\, dx, \quad (15.11)$$

and

$$|R_{cl}|^2 = 1 - 2\gamma[Q_l(w)]^{-1}\int_0^\infty \sin\gamma x Q_l(uu' - vv'\cosh x)dx + 2\gamma^2[Q_l(w)]^{-2}\int_0^\infty \cos(\gamma x)\int_0^\infty Q_l(uu' - vv'\cosh t) \times Q_l(uu' - vv'\cosh(t+x))dt\, dx. \quad (15.12)$$

All these representations are valid under the conditions given by Eq. (15.1).

16. THE COULOMB MODULUS FUNCTIONS M_c AND M_{cl}

In this section we shall introduce and discuss the so-called Coulomb modulus functions M_c and M_{cl}. We assume that

$$0<p'<k<p. \tag{16.1}$$

Then $vv'<0$, and $|R_c|$ and $|R_{cl}|$ are even in γ. It is convenient to regard γ as an *independent* variable (note that k is fixed). A second independent variable of R_c is ρ:

$$\rho:=\cosh\delta:=(1+y^2)/(-2y)\geqslant 1, \tag{16.2}$$

where

$$\rho=1\Leftrightarrow x=0\Leftrightarrow pp'=k^2\&\theta=0. \tag{16.3}$$

The following two representations are basic to this section (cf. Secs. 14 and 15, and Appendix F of Ref. 32):

$$R_c(\rho;\gamma)=\int_0^\infty g(t)e^{-i\gamma t}\,dt,\quad \gamma\in\mathbb{R}, \tag{16.4}$$

$$R_{cl}(p'/k;w;\gamma)=\int_0^\infty g_l(t)e^{-i\gamma t}\,dt,\quad \gamma\in\mathbb{R}, \tag{16.5}$$

where

$$g(t):=(\rho+1)(\rho+\cosh t)^{-2}\sinh t \tag{16.6}$$

and

$$g_l(t):=vv'[Q_l(w)]^{-1}Q_l'(uu'-vv'\cosh t)\sinh t \tag{16.7}$$

for $t\geqslant 0$; further we define $g(t):=g_l(t):=0$ for $t<0$. Equation (16.4) follows from Eq. (2.1), and (16.5) from (3.1) by integration by parts. (We might take $\mathrm{Re}\,i\gamma>-1$ and $\mathrm{Re}\,i\gamma>-l-1$, respectively; however, the realness of γ is essential below.) The choice of the variables p'/k and w in Eq. (16.5) is convenient for later derivations. For the moment we note that this choice is allowable since p, and hence u and v, can be expressed in terms of p' of w according to [note that $p'<p$]

$$p=p'w+p'(w^2-1)^{1/2}. \tag{16.8}$$

The functions g and g_l are *real* and continuous as can be seen by inspection. It follows from the theory of Fourier transforms that

$$g(t)=(2\pi)^{-1}\int_{-\infty}^{\infty}R_c(\rho;\gamma)e^{i\gamma t}\,d\gamma,\quad t\in\mathbb{R}, \tag{16.9}$$

$$g_l(t)=(2\pi)^{-1}\int_{-\infty}^{\infty}R_{cl}(p'/k;w;\gamma)e^{i\gamma t}\,d\gamma,\quad t\in\mathbb{R}. \tag{16.10}$$

Moreover, the norms of Fourier-transformed functions are equal. Defining the Coulomb modulus functions M_c and M_{cl} by

$$M_c(\rho):=\pi^{-1}\int_0^\infty |R_c(\rho;\gamma)|^2\,d\gamma, \tag{16.11}$$

$$M_{cl}(p'/k;w):=\pi^{-1}\int_0^\infty |R_{cl}(p'/k;w;\gamma)|^2\,d\gamma, \tag{16.12}$$

we find directly (cf. Appendix F of Ref. 32)

$$M_c(\rho)=\int_0^\infty g^2(t)\,dt, \tag{16.13}$$

$$M_{cl}\left(\frac{p'}{k};w\right)=\int_0^\infty g_l^2(t)\,dt. \tag{16.14}$$

These two equations are very convenient since they offer an opportunity to evaluate the right-hand members of Eqs. (16.11) and (16.12). We have derived many interesting relations involving M_c and M_{cl}, respectively, see Ref. 33. Here we mention only

$$M_c(\rho)=(\tfrac{4}{15})_2F_1(-\tfrac{1}{2},2;\tfrac{7}{2};(\rho-1)/(\rho+1)), \tag{16.15}$$

$$M_c(\rho)\text{ is strictly decreasing in }\rho\text{ on }[1,\infty), \tag{16.16}$$

$$M_c(1)=\pi^{-1}\int_0^\infty |B_{c0}(i\gamma)|^2\,d\gamma, \tag{16.17}$$

$$M_c^{\max}=M_c(1)=\tfrac{4}{15}, \tag{16.18}$$

$$M_c^{\inf}=\lim_{\rho\to\infty}M_c(\rho)=\tfrac{1}{6}, \tag{16.19}$$

and for M_{cl}, $l=0,1,\ldots$

$$M_{cl}\left(\frac{p'}{k};w\right)=\int_0^\infty\left[\frac{Q_l'(t+w)}{Q_l(w)}\right]^2(t^2-2tvv')^{1/2}\,dt, \tag{16.20}$$

$$M_{cl}(1^-;\infty)=\pi^{-1}\int_0^\infty\left(\frac{\pi\gamma}{\sinh\pi\gamma}\right)^2\left|\binom{l+i\gamma}{l}\right|^2 d\gamma$$
$$=(l+1)/(4l+6), \tag{16.21}$$

$$M_{cl}(0^+;\infty)=\pi^{-1}\int_0^\infty |B_{cl}(i\gamma)|^2\,d\gamma$$
$$=(l+1)^2B(2l+2,\tfrac{3}{2}), \tag{16.22}$$

$$M_{cl}(\tau;\infty):=\lim_{p'\downarrow 0}M_{cl}\left(\frac{p}{k};w\right)=M_{cl}(\tau)\left(\tau:=\frac{p}{k}\right), \tag{16.23}$$

$$M_{cl}(\tau):=\lim_{a\to\infty}M_{cl}(\tau^{-1}e^{-a};\cosh a)=(l+1)^2$$
$$\times B(2l+2,\tfrac{3}{2})_2F_1(-\tfrac{1}{2},2l+2;2l+\tfrac{7}{2};\tau^2),$$
$$0<\tau<1. \tag{16.24}$$

17. FIVE CONJECTURES

The Coulomb modulus functions M_c and M_{cl} introduced in the preceding section are useful, among other things, for supporting the following two conjectured inequalities:

$$\text{(c1):}\quad |R_c|\leqslant|B_{c0}(i\gamma)|,\quad \gamma\in\mathbb{R}',\ (\mathrm{p},\mathrm{p}')\in S, \tag{17.1}$$

where equality holds iff $pp'=k^2\&\theta=0$, and

$$\text{(c2):}\quad |R_{cl}|<|B_{cl}(i\gamma)|,\quad \gamma\in\mathbb{R}',\ (p,p')\in S. \tag{17.2}$$

In this section we shall discuss these two conjectures, one conjecture (c2)′ related to (c2), and two conjectures concerning M_{cl}. Just as in Sec. 16 we take $p'<k<p$.

As stated in (16.16), $M_c(\rho)$ is strictly decreasing in ρ for $1\leqslant\rho<\infty$. Let us suppose that a certain $\rho_0\in[1,\infty)$ exists such that

$$|R_c(\rho;\gamma)|\leqslant|R_c(\rho_0;\gamma)| \tag{17.3}$$

for all $\rho\in[1,\infty)$ and $\gamma\in\mathbb{R}$. Then clearly

$$M_c(\rho)\leqslant M_c(\rho_0) \tag{17.4}$$

for all ρ, according to Eq. (16.11). From the monotonicity of $M_c(\rho)$ it then follows that ρ_0 must be equal to 1.

Hence, *if* (17.3) holds with ρ_0 independent of γ, it follows that $\rho_0 = 1$ and that the supremum (with respect to ρ) of $|R_c(\rho;\gamma)|$ is the maximum given by

$$|R_c(1;\gamma)| = |B_{c0}(i\gamma)|, \quad \gamma\in\mathbb{R}. \tag{17.5}$$

It follows that, under the above condition, the inequality (17.1) is *valid and optimal.*

Now we are going to discuss the connection between (17.2) and the function M_{cl} introduced in Sec. 16. We have proved[33] that for fixed w (hence: varying p' *and* p)

$$\operatorname*{Max}_{0<p'<k} M_{cl}\left(\frac{p'}{k};w\right) = M_{cl}\left(\frac{k}{p};w\right), \quad w \text{ fixed}, \tag{17.6}$$

and

$$\operatorname*{Inf}_{0<p'<k} M_{cl}\left(\frac{p'}{k};w\right) = M_{cl}(1^-;w), \quad w \text{ fixed}. \tag{17.7}$$

Equation (17.7) leads to our third conjecture,

(c3): $M_{cl}(1^-;w)$ is strictly increasing in w on $(1,\infty)$. (17.8)

Continuing with (17.6) we observe that $p'/k = k/p$ implies that

$$w = \cosh a, \quad a := 2\ \ln(p/k) > 0,$$

so that the right member of (17.6) may be rewritten as

$$M_{cl}(k/p;w) = M_{cl}(e^{-a/2};\cosh a), \quad a>0.$$

This leads to our fourth conjecture,

(c4): $M_{cl}(e^{-a/2};\cosh a)$ is strictly increasing in a on $(0,\infty)$. (17.9)

A direct corollary of (17.6) and (17.9) is

$$M_{cl}(p'/k;w) < M_{cl}^{\sup} := M_{cl}(0^+;\infty), \tag{17.10}$$

for all p and p' subject to $0<p'<k<p$. Here

$$M_{cl}(0^+;\infty) := \lim_{a\to\infty} M_{cl}(e^{-a/2};\cosh a). \tag{17.11}$$

Let us suppose that a pair (p_0,p_0') exists such that

$$|R_{cl}(p'/k;w;\gamma)| \leqslant |R_{cl}(p_0'/k;w_0;\gamma)| \tag{17.12}$$

for all γ, p', and w [where $p'<k<p$ and $w_0 := (p_0^2 + p_0'^2)/2p_0p_0'$]. Then clearly

$$M_{cl}(p'/k;w)\leqslant M_{cl}(p_0'/k;w_0). \tag{17.13}$$

It follows directly from (17.10) that we must have

$$p_0' \downarrow 0 \text{ and } w_0 \to +\infty.$$

By using further the equality (Sec. 16)

$$M_{cl}^{\sup} = \pi^{-1}\int_0^\infty |B_{cl}(i\gamma)|^2\ d\gamma, \tag{17.14}$$

we obtain: *If* (17.12) holds with p_0 and p_0' independent of γ, then the supremum (with respect to p and p') of $|R_{cl}(p'/k;w;\gamma)|$ is given by

$$|R_{cl}(0^+;+\infty;\gamma)| = |B_{cl}(i\gamma)|. \tag{17.15}$$

This clarifies the connection between the conjectures (c4) and (c2). The inequality given by (17.2) is clearly optimal, if valid.

Finally we point out that a consequence of (c2) is

$$\text{(c2)}'\text{: } C_0^2 e^{\pi\gamma} c_{l\gamma}^{-1/2} < |B_{cl}(i\gamma)|, \quad \gamma\in\mathbb{R}', \tag{17.16}$$

where

$$c_{l\gamma}^{-1/2} = \left|\binom{l+i\gamma}{l}\right|, \tag{17.17}$$

cf. Eq. (22.23).

We have been able to prove each one of the conjectures (c2)', (c3), and (c4) for $l=0$ (only).

18. THE ZERO FUNCTIONS Z_c, Z_c^s, AND Z_c^a, AND THE ZEROS OF R_c AND R_{cl} FOR POSITIVE ENERGY

In this section we shall introduce the zero functions Z_c, Z_c^s, and Z_c^a, and give some preliminary information about the zeros of the ratios R_c and R_{cl} for positive energy ($k>0$). A more detailed study will be undertaken in Secs. 19 (for R_c) and 20 (for R_{cl}). We take k and γ fixed, and consider p, p', $\cos\theta \equiv \hat{p}\cdot\hat{p}'$, and l as variables. A remarkable fact to be proved in the following sections is that R_{cl} possesses an infinite number of zeros ($\forall l = 0, 1, \ldots, \forall k>0$) $\forall\gamma\in\mathbb{R}' := \mathbb{R}\setminus\{0\}$, whereas in contrast R_c possesses zeros only for a discrete set of values of γ. We define, for $k>0$ and any real γ,

$$Z_c(\gamma) := \inf_{p,p',\theta} |R_c(p,p';\theta;\gamma)|, \tag{18.1}$$

$$Z_c^s(\gamma) := \inf_{p,p',\theta} |R_c(p,p';\theta;\gamma)|, \quad (p,p')\in S, \tag{18.2}$$

$$Z_c^a(\gamma) := \inf_{p,p',\theta} |R_c(p,p';\theta;\gamma)|, \quad (p,p')\in A, \tag{18.3}$$

where "inf" denotes the greatest lower bound, and the regions S and A are defined in Sec. 23. Clearly

$$Z_c(\gamma) = \min(Z_c^s(\gamma), Z_c^a(\gamma)); \tag{18.4}$$

further, since $R_c\equiv 1$ for $\gamma = 0$ (Sec. 1),

$$Z_c(0) = Z_c^s(0) = Z_c^a(0) = 1. \tag{18.5}$$

These zero functions play a role in the discussion of (i) zeros of R_c and (ii) inequalities for R_c. In Ref. 24 we have proved that the inequality

$$|R_c| > 1, \quad \gamma<0, \quad (p,p')\in A, \tag{18.6}$$

holds and that it is optimal. This implies that

$$Z_c^a(\gamma) = 1, \quad \gamma\leqslant 0. \tag{18.7}$$

Since $|R_c|$ is even in γ when $(p,p')\in S$, we have

$$Z_c^s(\gamma) = Z_c^s(-\gamma), \quad \gamma\in\mathbb{R}.$$

However, in the "asymmetric" case, $(p,p')\in A$, $|R_c|$ is not even in γ, so that Z_c^a is not even in γ.

In Sec. 23 subregions S_i and A_i (i = 1,2,3,4) are introduced. The S_i can be transformed into one another, and similarly the A_i, by the transposition (t_1): $p\leftrightarrow p'$ and the inversion (t_2): $p\to k^2/p \,\&\, p'\to k^2/p'$. Since R_c is invariant under these transformations, we may in many cases restrict ourselves to one particular subregion S_i, and A_i, respectively. For example, if (p,p') were to be restricted to any S_i instead of to S, on the right-hand side of Eq. (18.2), the result would be the same. Note in particular that the borderline S_0, correspond-

ing to $pp' = k^2$ (see Sec. 23) need not be taken into account here: It can be reached by a limiting process, so that the *same* infimum in Eq. (18.2) is obtained, whether or not we include the curve $pp' = k^2$.

In Sec. 19, we shall first study the symmetric case and then [Eq. (19.13) ff] the asymmetric case. The zeros of R_c to be derived there, $\overset{\circ}{\gamma}{}^s_n$ and $\overset{\circ}{\gamma}{}^a_n$, are also zeros of $Z^s_c(\gamma)$ and $Z^a_c(\gamma)$, respectively. The reverse statement is also true since the infima are in fact minima; see Eqs. (19.12b) and (19.32b).

In Sec. 20 a similar study will be made for deriving the zeros of R_{cl}, $l = 0, 1,...$, for positive energy. As stated earlier, R_{cl} has zeros for *all* $\gamma \neq 0$; therefore, introducing zero functions Z_{cl} for the partial-wave case would make no sense (they would be identically zero). Although the approach in Sec. 20 is in principle the same as in Sec. 19, the formulas are more complicated and less transparent. Therefore, we consider first the symmetric case $S_0 \cup S_1$ for $l = 0$ and next for general l [starting at Eq. (20.30)], and finally the asymmetric case A_1 for general l [Eqs. (20.50)–(20.71)].

The main results of Secs. 19 and 20 are given by Eqs. (19.4), (19.10), (19.24), (19.30), (20.45)–(20.49), and (20.69)–(20.80).

19. THE ZEROS OF R_C FOR POSITIVE ENERGY

In this section we shall derive the zeros of R_c for positive energy. Preliminary remarks have been made in Sec. 18.

For the derivations in this section, the new representations given in Secs. 2 and 11 are most helpful. We distinguish (i) the symmetric case and (ii) the asymmetric case (cf. Sec. 23), and we begin with the symmetric case, which is relatively simple.

(i) For R_c in the *symmetric* case we use Eq. (11.9), slightly rewritten as

$$\begin{aligned} R_c(\pi\gamma)^{-1}\sinh\pi\gamma = & -i\coth(\delta/2)\sin\gamma\delta + \cos\gamma\delta \\ & - \pi^{-1}\int_0^\pi (1+\cos t)(\cosh\gamma t - \cos\gamma\delta) \\ & \times(\cosh\delta - \cos t)^{-1}\,dt, \\ \delta = & -\ln(-y) > 0; \quad -1 < y < 0. \end{aligned}$$

Here γ is real; since $|R_c|$ is even in γ it suffices to consider positive values of γ only. It can be seen upon inspection that R_c can be zero only if $\sin\gamma\delta = 0 \& \delta > 0$. Further,

$$(\cos\gamma\delta = -1) \Rightarrow (R_c < 0). \tag{19.2}$$

Thus we must have $\cos\gamma\delta = +1$, and so (note that $R_c = 1$ for $\gamma = 0$)

$$\gamma\delta = 2n\pi, \quad n = 1,2,\ldots . \tag{19.3}$$

Hence a *necessary* condition for R_c to be zero is

$$\delta = \delta_n := 2\pi n/\gamma, \quad n = 1,2,\ldots . \tag{19.4}$$

We insert (19.4) into (19.1) and denote the left-hand side of the resulting equation by $h^s_n(\gamma)$,

$$h^s_n(\gamma) := (\pi\gamma)^{-1}\sinh\pi\gamma R_c\,[\delta = \delta_n := 2\pi n/\gamma], \tag{19.5}$$

$$\gamma > 0, \ n = 1,2,...,$$

where the superscript s stands for "symmetric." Then

$$\begin{aligned} h^s_n(\gamma) = & 1 - \pi^{-1}\int_0^\pi (1+\cos t) \\ & \times(\cosh\gamma t - 1)[\cosh(2\pi n/\gamma) - \cos t\,]^{-1}\,dt, \end{aligned} \tag{19.6}$$

$$\gamma > 0, \ n = 1,2,\ldots .$$

Defining $h^s_n(0) := 1 (\forall n)$ we see that $h^s_n(\gamma)$ is continuous in γ on $[0,\infty)$. Further, $h^s_n(\gamma)$ is strictly decreasing in γ and strictly increasing in n when $\gamma > 0$, and

$$\begin{aligned} & \lim_{\gamma\downarrow 0} h^s_n(\gamma) = 1, \quad \lim_{\gamma\to\infty} h^s_n(\gamma) = -\infty, \quad n = 1,2,..., \\ & \lim_{n\to\infty} h^s_n(\gamma) = 1, \quad \gamma \geqslant 0. \end{aligned} \tag{19.7}$$

Clearly $h^s_n(\gamma)$ has just *one* zero γ, which we denote by $\overset{\circ}{\gamma}{}^s_n$ [again s stands for "symmetric"], for each $n = 1,2,\ldots$. These zeros are also the zeros of R_c, occurring when $\delta = \delta_n = 2\pi n/\overset{\circ}{\gamma}{}^s_n$, according to Eq. (19.4). In this way we have determined the infinite number of zeros of R_c in the symmetric case: There exists an infinite sequence of pairs, $\{(\gamma,\delta)\}_n$, for which $R_c = 0$; the γ's for which $R_c = 0$ accumulate at $\gamma\to\infty$ and at $\gamma\to -\infty$.

An approximation formula for the zeros $\overset{\circ}{\gamma}{}^s_n$ is easily derived: First we have from Eq. (19.6) the approximate equation

$$\pi\cosh\left(\frac{2\pi n}{\gamma}\right) \approx \int_0^\pi (1+\cos t)\cosh\gamma t\,dt \tag{19.8a}$$

$$= \gamma^{-1}(1+\gamma^2)^{-1}\sinh\pi\gamma, \tag{19.8b}$$

where (19.8a) holds provided

$$\pi^{-1} \lesssim \gamma \lesssim 2\pi n. \tag{19.8c}$$

Further approximation gives

$$\exp(\pi\gamma - 2\pi n/\gamma) \approx \pi\gamma(1+\gamma^2) \approx \pi\gamma^3, \quad n = 1,2,\ldots . \tag{19.9}$$

The approximate solutions of Eq. (19.9) for $n = 1,2,...$ are given by [cf. Ref. 24]

$$\gamma \approx \overset{\circ}{\gamma}{}^s_n \approx (2n)^{1/2} + (4\pi)^{-1}\ln(8\pi^2 n^3) + o(1), \quad n\to\infty . \tag{19.10}$$

In Ref. 24 a comparison has been made between these approximate values and the exact values for $n = 1,2,3$, and 4, which shows that Eq. (19.10) gives accurate values even for the first few zeros.

The zero function $Z^s_c(\gamma)$ introduced in Sec. 18 is zero when $R_c = 0$ in the symmetric case. Including the negative zeros, we have

$$Z^s_c(\gamma) = 0 \quad \text{for} \quad \gamma = \pm\overset{\circ}{\gamma}{}^s_n, \quad n = 1,2,\ldots . \tag{19.11}$$

We recall that $\overset{\circ}{\gamma}{}^s_n$ are the zeros of $h^s_n(\gamma)$; see Eqs. (19.5) and (19.6). To prove Eq. (19.11) we observe that

$$Z^s_c(\gamma) = \inf_{\delta>0}|R_c|, \quad \gamma \text{ fixed}. \tag{19.12a}$$

In fact, the infimum here is a minimum, i.e., it is attained; consequently the zeros $\pm\overset{\circ}{\gamma}{}^s_n$ of R_c *coincide* with the zeros of Z^s_c.

To prove that

$$Z^s_c(\gamma) = \min_{\delta>0}|R_c|, \quad \gamma \text{ fixed}, \tag{19.12b}$$

we take $\gamma > 0$ and note that $|R_c|$ is a continuous function of δ; hence, it suffices to prove that the infimum in Eq. (19.12a) does not originate from $\delta \to \infty$. Since

$$|R_c| \to \pi\gamma/\sinh \pi\gamma \quad \text{for } \delta \to \infty\ (y \uparrow 0),$$

it is sufficient to prove that, for any fixed $\gamma > 0$,

$$\exists \delta > 0 : |R_c| < \pi\gamma/\sinh \pi\gamma.$$

Taking (once again) $\delta = \delta_n = 2\pi n/\gamma$, we see that it is sufficient to prove that [cf. Eq. (19.5)]

$$\exists n : |h_s^n(\gamma)| < 1, \tag{19.12c}$$

which is easy since $h_n^s(\gamma)$ is strictly increasing in n and $h_n^s(\gamma) \uparrow 1$ as $n \to \infty\ (\forall \gamma > 0)$ according to Eq. (19.7).

(ii) In the remaining part of this section we shall determine the zeros of R_c for positive energy in the *asymmetric* case (cf. Sec. 23). This task is quite similar to the corresponding one for the symmetric case which has just been completed.

First we reduce Eq. (19.1), which holds for the symmetric case; by using

$$\int_0^\pi (1 + \cos t)(\cosh \delta - \cos t)^{-1}\, dt = 2\pi(e^\delta - 1)^{-1}, \quad \delta > 0,$$

we obtain [cf. Eq. (11.9)]

$$R_c(\pi\gamma)^{-1} \sinh \pi\gamma = e^{-i\gamma\delta} \coth(\delta/2) - \pi^{-1} \int_0^\pi \cosh \gamma t\, (1 + \cos t)(\cosh \delta - \cos t)^{-1}\, dt, \tag{19.13}$$

$$\delta = -\ln(-y) > 0; \quad -1 < y < 0.$$

A suitable representation valid for the asymmetric case, when $0 < y < 1$, can be obtained from Eq. (19.13) by analytic continuation. As stated in (2.9), the correct replacement in this process is

$$(-y)^{i\gamma} \to e^{-\pi\gamma} y^{i\gamma}, \tag{19.14}$$

which one finds by determining the imaginary part of $\pm y$, bearing in mind that Im $k \geqslant 0$ by convention. With

$$\zeta := -\ln y\, [> 0 \text{ when } 0 < y < 1] \tag{19.15}$$

we find that (19.14) comes down to

$$\delta \to \zeta - i\pi. \tag{19.16}$$

Further, we shall use $c_{\infty\gamma}$ for notational simplicity. From

$$c_{l\gamma}^{-1} = \binom{l + i\gamma}{l}\binom{l - i\gamma}{l} = \prod_{n=1}^{l}\left(1 + \frac{\gamma^2}{n^2}\right), \tag{19.17}$$

we find

$$c_{\infty\gamma} := \lim_{l \to \infty} c_{l\gamma} = \prod_{n=1}^{\infty}\left(1 + \frac{\gamma^2}{n^2}\right)^{-1} \tag{19.18}$$

$$= \pi\gamma/\sinh \pi\gamma = \Gamma(1 + i\gamma)\Gamma(1 - i\gamma) = C_0^2 e^{\pi\gamma}. \tag{19.19}$$

From Eq. (19.13), (19.15), (19.16), and (19.19) it is straightforward to derive

$$c_{\infty\gamma}^{-1} R_c = \exp(-i\gamma\zeta - \pi\gamma)\tanh(\zeta/2) + \pi^{-1} \int_0^\pi \cosh \gamma t\, (1 + \cos t)(\cosh \zeta + \cos t)^{-1}\, dt, \tag{19.20}$$

$$\zeta = -\ln y > 0; \quad 0 < y < 1.$$

This representation for R_c, valid for the asymmetric case and for arbitrary γ, is convenient to determine the zeros of R_c. In this case $|R_c|$ is not symmetric in γ so that we have to distinguish positive and negative values of γ. From Ref. 24 we have

$$|R_c| > 1, \quad \gamma < 0, \quad (p,p') \in A, \tag{19.21}$$

so that R_c has no zero for $\gamma < 0$ (Coulomb attraction). Therefore, we shall assume in the remainder of this section that $\gamma > 0$ (Coulomb repulsion).

From Eq. (19.20) it is clear that R_c can be zero only if $e^{-i\gamma\zeta} = \pm 1$. Since $c_{\infty\gamma} > 0$, $\forall \gamma \in \mathbf{R}$, we have

$$(e^{-i\gamma\zeta} = 1) \Rightarrow (R_c > 0). \tag{19.22}$$

In this case R_c has no zero, so we put $e^{i\gamma\zeta} = -1$,

$$\gamma\zeta = (2n - 1)\pi, \quad n = 1,2,\ldots . \tag{19.23}$$

A *necessary* condition for R_c to be zero is clearly

$$\zeta = \zeta_n := (2n - 1)\pi/\gamma, \quad n = 1,2,\ldots . \tag{19.24}$$

We insert Eq. (19.24) into Eq. (19.20) and denote the left-hand side of the resulting equation by $-h_n^a(\gamma)$,

$$-h_n^a(\gamma) := c_{\infty\gamma}^{-1} R_c\, [\zeta = \zeta_n := (2n - 1)\pi/\gamma], \quad \gamma > 0,\ n = 1,2,\ldots, \tag{19.25}$$

where the superscript a stands for "asymmetric". The reason for the minus sign is that by this convention $h_n^a(\gamma)$ has properties similar to those of $h_n^s(\gamma)$; see Eq. (19.7); it follows that

$$h_n^a(\gamma) = \exp(-\pi\gamma)\tanh[(n - \tfrac{1}{2})\pi/\gamma] - \pi^{-1} \int_0^\pi \cosh \gamma t\, (1 + \cos t) \times \left[\cosh\left\{\frac{(2n - 1)\pi}{\gamma}\right\} + \cos t\right]^{-1} dt, \tag{19.26}$$

$$\gamma > 0,\ n = 1,2,\ldots .$$

Defining $h_n^a(0) := 1\ (\forall n)$ we see that $h_n^a(\gamma)$ is continuous in γ on $[0, \infty)$. Further, $h_n^a(\gamma)$ is strictly decreasing in γ and strictly increasing in n when $\gamma > 0$, and

$$\lim_{\gamma \downarrow 0} h_n^a(\gamma) = 1, \quad \lim_{\gamma \to \infty} h_n^a(\gamma) = -\infty, \quad n = 1,2,\ldots, \tag{19.27}$$

$$\lim_{n \to \infty} h_n^a(\gamma) = e^{-\pi\gamma}, \quad \gamma \geqslant 0.$$

Clearly these properties of $h_n^a(\gamma)$ are similar to those of $h_n^s(\gamma)$ mentioned below Eq. (19.6).

The discussion of the symmetric case following Eq. (19.7) can be repeated here, *mutatis mutandis*: For each $n = 1,2,\ldots$, $h_n^a(\gamma)$ has just *one* zero γ, which we denote by $\mathring{\gamma}_n^a$. These zeros are also the zeros of R_c, occurring when $\zeta = \zeta_n = (2n - 1)\pi/\mathring{\gamma}_n^a$, according to Eq. (19.24), and they accumulate at $\gamma \to +\infty$.

An approximation formula for the zeros $\mathring{\gamma}_n^a$ is obtained

by approximating Eq. (19.26),

$$h_n^a(\gamma)=0$$
$$\Rightarrow \exp(-\pi\gamma)\tanh[(n-\tfrac{1}{2})\pi/\gamma]\pi\cosh[(2n-1)\pi/\gamma]$$
$$\approx \int_0^\pi \cosh\gamma t\,(1+\cos t)\,dt$$
$$=\gamma^{-1}(1+\gamma^2)^{-1}\sinh\pi\gamma; \tag{19.28}$$

hence,

$$\exp[2\pi\gamma-(2n-1)\pi/\gamma]\approx\pi\gamma(1+\gamma^2)$$
$$\approx\pi\gamma^3, \quad n=1,2,..., \tag{19.29a}$$

provided

$$1\lesssim\gamma\lesssim(2n-1)\pi. \tag{19.29b}$$

The approximate solutions of Eq. (19.29a) for $n=1,2,...$ are given by (cf. Ref. 24)

$$\gamma\approx\mathring{\gamma}_n^a\approx(n-\tfrac{1}{2})^{1/2}+(8\pi)^{-1}\ln(\pi^2n^3)+o(1), \quad n\to\infty. \tag{19.30}$$

According to Ref. 24, Eq. (19.30) gives accurate values even for the first few zeros.

The zero function $Z_c^a(\gamma)$ introduced in Sec. 18 is zero when $R_c=0$ in the asymmetric case,

$$Z_c^a(\gamma)=0 \quad \text{for } \gamma=\mathring{\gamma}_n^a, \quad n=1,2,.... \tag{19.31}$$

The proof follows by observing that

$$Z_c^a(\gamma)=\inf_{\zeta>0}|R_c|, \quad \gamma \text{ fixed}; \tag{19.32a}$$

cf. Eqs. (19.11) and (19.12). More precisely, the zeros $\mathring{\gamma}_n^a$ of R_c coincide with the zeros of Z_c^a, since the infimum in Eq. (19.32a) is in fact a minimum:

$$Z_c^a(\gamma)=\min_{\zeta>0}|R_c|, \quad \gamma \text{ fixed}; \tag{19.32b}$$

cf. Eq. (19.12b). To prove this, we note that $|R_c|$ is a continuous function of ζ; hence, it suffices to prove that the infimum does not originate from either $\zeta\to0$ or $\zeta\to\infty$. Now we have

$$|R_c|\to C_0^2 \quad \text{for } \zeta\to\infty\,(\gamma\downarrow0);$$
$$\to1 \quad \text{for } \zeta\downarrow0\,(\gamma\uparrow1).$$

Since $C_0^2<1$ for $\gamma>0$, it is sufficient to prove that, for any fixed $\gamma>0$,

$$\exists\zeta>0:|R_c|<C_0^2.$$

Taking (again) $\zeta=\zeta_n=(2n-1)\pi/\gamma$, we see that it is sufficient to prove that [cf. Eqs. (19.19) and (19.25)]

$$\exists n:|h_n^a(\gamma)|<e^{-\pi\gamma}, \tag{19.32c}$$

which is easy since $h_n^a(\gamma)$ is strictly increasing in n and $h_n^a(\gamma)\uparrow e^{-\pi\gamma}$ as $n\to\infty\,(\forall\gamma>0)$ according to Eq. (19.27).

20. THE ZEROS OF R_{cl} FOR POSITIVE ENERGY

In this section we shall derive the zeros of R_{cl} for positive energy. We shall use the new integral representations derived in Sec. 3. It is also possible to use the new series representations derived in Secs. 4 and 5, but this would give no additional new results (cf. Appendix Z of Ref. 32).

The discussion in this section runs parallel to the one in Sec. 19. However, the formulas below turn out to be much more complicated.

First we shall investigate the *symmetric* case; then $|R_{cl}|$ is even in γ so that it is sufficient to consider $\gamma>0$ only. As usual we have $k>0$ and $l\in\mathbb{N}=\{0,1,...\}$. According to the transformation invariances of R_{cl} to be discussed in Sec. 23 we may restrict ourselves to any one of the subregions S_i ($i=1,2,3,$ or 4), combined with a part of S_0; we choose $i=1$:

$$0<p'<k<p\&k^2\leqslant pp'.$$

Then $v>0$, $v'<0$, $1<u'\leqslant u$, and [cf. Eq. (3.4)]

$$-1<a'<0<a<1, \quad 0<-a'/a\leqslant1,$$
$$(u-1)/(u+1)=a^2, \quad (u'-1)/(u'+1)=(-a')^2,$$

where $a:=(p-k)/(p+k)$, $a':=(p'-k)/(p'+k)$. We shall use the integral representation given by Eq. (3.14), which also holds for $pp'=k^2$:

$$R_{cl}Q_l(w)=C_0^2e^{2\pi\gamma}Q_l^{i\gamma}(u)P_l^{-i\gamma}(u')+Q_l(w)$$
$$-C_0^2e^{\pi\gamma}\pi^{-1}\int_0^\pi\cosh\gamma tQ_l(uu'+vv'\cos t)\,dt, \tag{20.1}$$
$$0<p'<k<p\&k^2\leqslant pp'; \ 1<u'\leqslant u; \ vv'<0.$$

The integral here is real since

$$uu'+vv'\cos t\geqslant uu'+vv'$$
$$=1+(pp'-k^2)^2/(2k^2pp')\geqslant1, \tag{20.2}$$

and $Q_l(z)$ is real for $z>1$. So the first term on the right-hand side of Eq. (20.1) is the only complex-valued term. We shall use the equalities

$$2\Gamma(l+1-i\gamma)\sinh\pi\gamma e^{\pi\gamma}Q_l^{i\gamma}(u)$$
$$=i\pi l!\,[a^{i\gamma}P_l^{(i\gamma,-i\gamma)}(u)-a^{-i\gamma}P_l^{(-i\gamma,i\gamma)}(u)], \tag{20.3}$$
$$0<a<1,$$

$$\Gamma(l+1+i\gamma)P_l^{-i\gamma}(u')$$
$$=l!(-a')^{i\gamma}P_l^{(i\gamma,-i\gamma)}(u'), \quad -1<a'<0 \tag{20.4}$$
$$=l!(a')^{i\gamma}P_l^{(i\gamma,-i\gamma)}(u'), \quad 0<a'<1. \tag{20.5}$$

In this way we obtain from Eq. (20.1)

$$R_{cl}Q_l(w)=c_{l\gamma}C_0^2e^{\pi\gamma}(-a')^{i\gamma}P_l^{(i\gamma,-i\gamma)}(u')$$
$$\times i(2\gamma)^{-1}[a^{i\gamma}P_l^{(i\gamma,-i\gamma)}(u)-a^{-i\gamma}P_l^{(-i\gamma,i\gamma)}(u)]$$
$$+Q_l(w)-C_0^2e^{\pi\gamma}\pi^{-1}\int_0^\pi\cosh\gamma t$$
$$\times Q_l(uu'+vv'\cos t)\,dt, \tag{20.6}$$
$$0<a<1, \ -1<a'<0; \ 1<u'\leqslant u; \ vv'<0.$$

It is convenient to rewrite the last two terms on the right-hand side of Eq. (20.6); performing integration by parts we find

$$Q_l(w)-C_0^2e^{\pi\gamma}\pi^{-1}\int_0^\pi\cosh\gamma t\,Q_l(uu'+vv'\cos t)\,dt$$
$$=-vv'(\sinh\pi\gamma)^{-1}\int_0^\pi\sinh\gamma t\sin t\,Q_l'(uu'+vv'\cos t)\,dt. \tag{20.7}$$

Note that only in Q_l' does the prime indicate a derivative. For notational convenience we introduce some abbrevia-

tions, which will be used in this section only. First,

$$q(\gamma) := -\int_0^\pi \sinh \gamma t \sin t\, Q_l'(uu' + vv' \cos t)dt; \quad (20.8)$$

since $Q_l'(z) < 0$ for $z > 1$, $q(\gamma)$ is positive for $\gamma > 0$. Next we define

$$\begin{aligned} \Re_l &:= \operatorname{Re} P_l^{(i\gamma,\,-i\gamma)}(u), \quad u\in\mathbb{R},\ \gamma\in\mathbb{R}, \\ \Re_l' &:= \operatorname{Re} P_l^{(i\gamma,\,-i\gamma)}(u'), \quad u'\in\mathbb{R},\ \gamma\in\mathbb{R}, \\ \Im_l &:= \operatorname{Im} P_l^{(i\gamma,\,-i\gamma)}(u), \quad u\in\mathbb{R},\ \gamma\in\mathbb{R}, \\ \Im_l' &:= \operatorname{Im} P_l^{(i\gamma,\,-i\gamma)}(u'), \quad u'\in\mathbb{R},\ \gamma\in\mathbb{R}. \end{aligned} \quad (20.9)$$

Finally we put, for $\gamma\in\mathbb{R}$,

$$f(\gamma;a) := \Re_l \sin(\gamma \ln a) + \Im_l \cos(\gamma \ln a), \quad (20.10)$$

$$\begin{aligned} g(\gamma;-a') &:= \Re_l' \cos\{\gamma \ln(-a')\} - \Im_l' \sin\{\gamma \ln(-a')\} \\ &\quad + i[\Re_l' \sin\{\gamma \ln(-a')\} + \Im_l' \cos\{\gamma \ln(-a')\}]. \end{aligned} \quad (20.11)$$

With these notations we obtain from Eqs. (20.6) and (20.7),

$$\begin{aligned} &\sinh \pi\gamma Q_l(w) R_{cl} \\ &\quad = -\pi c_{l\gamma} f(\gamma;a) g(\gamma;-a') + vv' q(\gamma), \quad (20.12) \\ &\quad 0 < a < 1,\ -1 < a' < 0;\ 1 < u' \leqslant u;\ -a' \leqslant a. \end{aligned}$$

One easily verifies that $\Re_l$ and $\Re_l'$ are even in γ, and that $\Im_l$ and $\Im_l'$ are odd in γ; hence $f(\gamma;a)$ is odd in γ, and

$$g^*(\gamma;-a') = g(-\gamma;-a'), \quad \gamma\in\mathbb{R},$$

so that

$$R_{cl}^*(\gamma) = R_{cl}(-\gamma), \quad \gamma\in\mathbb{R}, \quad (20.13)$$

which shows that $|R_{cl}|$ is even in γ.

As we are looking for zeros of R_{cl}, we set its imaginary part equal to zero. This means that we must have either $f(\gamma;a) = 0$ or $\operatorname{Im} g(\gamma;a) = 0$. Since $q(\gamma) > 0$ the first alternative gives no zero, so that a *necessary* condition for R_{cl} to be zero is

$$\Re_l' \sin\{\gamma \ln(-a')\} + \Im_l' \cos\{\gamma \ln(-a')\} = 0. \quad (20.14)$$

The derivation of the zeros of R_{cl} for the general $l = 0,1,...$ case is quite complicated because of the complexity of the above formulas. Therefore, we shall first consider the simpler $l = 0$ case, which will pave the way for the general case.

For $l = 0$ we have $\Re_l \equiv \Re_l' \equiv 1$ and $\Im_l \equiv \Im_l' \equiv 0$ so that Eq. (20.14) reduces to

$$\sin\{\gamma \ln(-a')\} = 0, \quad (20.15)$$

which is a *necessary* condition for R_{c0} to be zero. We denote the solutions a' of Eq. (20.15) by

$$\begin{aligned} a' &= -a_m'(\gamma) := -\exp(-m\pi/\gamma), \quad (20.16) \\ &\gamma > 0,\ m = 1,2,...; \end{aligned}$$

then

$$\cos\{\gamma \ln(a_m'(\gamma))\} = (-1)^m. \quad (20.17)$$

Next we define the functions $H_m^s(\gamma;a)$, $m = 1,2,...$, by

$$H_m^s(\gamma;a) := \sinh \pi\gamma Q_0(w) R_{c0}\,[a' = -a_m'(\gamma)], \quad (20.18)$$

where the superscript "s" stands for "symmetric." Clearly the zeros of $\{H_m^s(\gamma;a)\}_{m=1}^\infty$ coincide with the zeros of R_{c0}. From Eq. (20.12) we have

$$H_m^s(\gamma;a) = (-1)^{m+1}\pi \sin(\gamma \ln a) + vv' q(\gamma), \quad (20.19)$$

subject to [cf. Eq. (3.4)]

$$a' = -a_m'(\gamma); \quad 0 < a < 1; \quad 0 < -a'/a \leqslant 1. \quad (20.20)$$

Now we claim that R_{c0} has, for *any fixed* $\gamma > 0$, an infinite sequence of zeros. To prove this we first observe that we can take $a_m'(\gamma)$ arbitrarily close to zero, for *any fixed* $\gamma > 0$, by taking m sufficiently large, say $m > M$. For such large m we have approximately

$$u' \approx 1, \quad v' \approx -2a_m'(\gamma) \approx 0, \quad (20.21)$$

and from Eq. (20.8),

$$\begin{aligned} q(\gamma) &\approx -Q_l'(u)\int_0^\pi \sinh \gamma t \sin t\, dt \\ &= -Q_l'(u)(1+\gamma^2)^{-1} \sinh \pi\gamma. \quad (20.22) \end{aligned}$$

Below we shall need the relations (recall that $u^2 - v^2 \equiv 1$)

$$\begin{aligned} Q_l(u) &= -\tfrac{1}{2}\ln(u-1) + O(1), \quad u \downarrow 1, \\ Q_l'(u) &= (1-u^2)^{-1} - \tfrac{1}{4}l(l+1)\ln(u-1) + O(1), \quad u \downarrow 1, \\ Q_0'(u) &= (1-u^2)^{-1} = -v^{-2}. \end{aligned} \quad (20.23)$$

The zeros of $H_m^s(\gamma;a)$, for fixed $\gamma > 0$, are the solutions a of the equation

$$(-1)^m \pi \sin(\gamma \ln a) = vv' q(\gamma), \quad (20.24)$$

which is, for sufficiently large m, approximated by the equation

$$\begin{aligned} &(-1)^{m+1}\pi v \sin(\gamma \ln a)(1+\gamma^2)/\sinh \pi\gamma \\ &\quad \approx 2a_m'(\gamma) = 2\exp(-m\pi/\gamma) \downarrow 0, \quad m \to \infty. \quad (20.25) \end{aligned}$$

Clearly the approximate solutions are given roughly by

$$a = a_{m;n}^s(\gamma) \approx \exp(-n\pi/\gamma), \quad n = 1,2,... . \quad (20.26)$$

It is important to note that an additional condition must be met here: When n/γ is large, a and hence v is very small according to

$$v \approx 2a_{m;n}^s(\gamma) \approx 0.$$

Then Eq. (20.25) can be rewritten as

$$\begin{aligned} &(-1)^{m+1}\pi \sin\{\gamma \ln a_{m;n}^s(\gamma)\}(1+\gamma^2)/\sinh \pi\gamma \\ &\quad \approx \exp[(n-m)\pi/\gamma]. \quad (20.27) \end{aligned}$$

Consequently we must have $n < m$ such that

$$0 < a_m'(\gamma) < a_{m;n}^s(\gamma) \approx 0. \quad (20.28)$$

Recall for comparison that $0 < -a' \leqslant a < 1$ [cf. Eqs. (20.12) and (20.20)].

Summarizing: We have proved that R_{c0} has, for any fixed $\gamma > 0$, an infinite number of zeros. These are attained for

$$\begin{aligned} &a' = -a_m'(\gamma) := -\exp(-m\pi/\gamma); \\ &m = M^s(\gamma),\ M^s(\gamma)+1,..., \\ &\& \quad (20.29) \\ &a = a_{m;n}^s(\gamma);\ n = 1,2,...,N^s(\gamma;m), \end{aligned}$$

for certain natural numbers $M^s(\gamma)$ and $N^s(\gamma;m)$. In particular, $N^s(\gamma;m) \to \infty$ when $m \to \infty$. The approximation (20.26) is valid roughly for sufficiently large m, such that (20.21) holds.

The zeros of R_{c0} accumulate at $a'=0$, i.e., at $p'=k$, and at $a=0$, i.e., at $p=k$. From (20.16) and (20.19) it follows that R_{c0} has no zero on S_0, i.e., for $pp'=k^2 \Leftrightarrow a+a'=0$.

Now we are in a position to derive the zeros of R_{cl} for general $l=0,1,\ldots$. A *necessary* condition is

$$a'=-a'_{l;m}(\gamma), \quad \gamma>0, \quad m=1,2,\ldots. \tag{20.30}$$

These functions $a'_{l,m}$ are the solutions of Eq. (20.14). It is easily seen that these solutions form a denumerably infinite set: In particular, for sufficiently large m we have $a'\approx 0$ and hence [note that $\Re'_l$ and $\Im'_l$ depend on a']

$$\Re'_l \approx \operatorname{Re} P_l^{(i\gamma,-i\gamma)}(1)=\operatorname{Re}\binom{l+i\gamma}{l}, \tag{20.31}$$

$$\Im'_l \approx \operatorname{Im} P_l^{(i\gamma,-i\gamma)}(1)=\operatorname{Im}\binom{l+i\gamma}{l},$$

so that

$$a'_{l;m}(\gamma)\approx \exp[-(m\pi+\xi)/\gamma], \quad m\to\infty, \tag{20.32}$$

where $\xi=\xi_l(\gamma)$ and

$$\tan\xi=\operatorname{Im}\binom{l+i\gamma}{l}/\operatorname{Re}\binom{l+i\gamma}{l}. \tag{20.33}$$

From Eq. (20.11) we obtain

$$g(\gamma;a'_{l;m}(\gamma))=\Re'_l/\cos[\gamma\ln(-a')], \quad a'=-a'_{l;m}(\gamma). \tag{20.34}$$

We define the functions $H^s_{l;m}(\gamma;a)$, $m=1,2,\ldots$, by

$$H^s_{l;m}(\gamma;a):=\sinh\pi\gamma Q_l(w)R_{cl}[a'=-a'_{l;m}(\gamma)]. \tag{20.35}$$

Then we have from Eq. (20.12)

$$H^s_{l;m}(\gamma;a)=vv'q(\gamma) - \pi c_{l\gamma} f(\gamma;a)\Re'_l/\cos[\gamma\ln(-a')], \tag{20.36}$$

subject to

$$a'=-a'_{l;m}(\gamma); \quad 0<a<1; \quad 0<-a'/a\leqslant 1, \tag{20.37}$$

where $f(\gamma;a)$ is given by (20.10) and $q(\gamma)$ by (20.8). The zeros a of

$$H^s_{l;m}(\gamma;a), \quad m=1,2,\ldots, \tag{20.38}$$

give precisely all the zeros of R_{cl} we are seeking. For sufficiently large m these zeros can be derived explicitly in good approximation. We use Eqs. (20.22) and (20.34); further, $v'\approx 2a'$, so that

$$(\Re'_l)^2+(\Im'_l)^2\approx\binom{l+i\gamma}{l}\binom{l-i\gamma}{l}=c_{l\gamma}^{-1}, \tag{20.39}$$

and ξ can be taken such that

$$\cos\xi=c_{l\gamma}^{1/2}\operatorname{Re}\binom{l+i\gamma}{l}. \tag{20.40}$$

It follows that for $m\to\infty$ the zeros of R_{cl} are approximately given by the solutions a of the equation

$$(-1)^{m+1}\pi f(\gamma;a)c_{l\gamma}^{1/2}(1+\gamma^2)/\sinh\pi\gamma = -2\exp[-(m\pi+\xi)/\gamma]vQ'_l(u), \quad m\to\infty, \tag{20.41}$$

which should be compared with the corresponding one for $l=0$, Eq. (20.25). Obviously the solutions a of Eq. (20.41) are approximately given by the zeros a of the function $f(\gamma;a)$ given by Eq. (20.10). Denoting the zeros a of R_{cl} by $a^s_{l;m;n}(\gamma)$, $n=1,2,\ldots$, we easily find that

$$a^s_{l;m;n}(\gamma)\approx\exp[-(n\pi+\xi)/\gamma], \quad m\to\infty, \tag{20.42}$$

cf. Eqs. (20.32) and (20.33). As before [Eq. (20.26) ff] an additional condition must be met here: When n/γ is large, a and v are very small and $u\downarrow 1$. Inserting Eq. (20.23) into Eq. (20.41) we get

$$(-1)^{m+1}\pi f(\gamma;a)c_{l\gamma}^{1/2}(1+\gamma^2)/\sinh\pi\gamma \approx[a^{-1}+2l(l+1)a\ln a] \times\exp[-(m\pi+\xi)/\gamma], \quad m\to\infty, \ a\downarrow 0. \tag{20.43}$$

It follows that the approximate solutions a given by Eq. (20.42) are only accurate if the right-hand member of Eq. (20.43) is sufficiently close to zero. By inserting Eq. (20.42) we get for this right-hand member approximately $\exp\{(n-m)\pi/\gamma\}$, so that the right-hand member of Eq. (20.43) is sufficiently close to zero if $n<m$ and more generally

$$n\leqslant N^s_l(\gamma;m), \tag{20.44}$$

for a certain $N^s_l(\gamma;m)\in\mathbb{N}$.

Summarizing: We have proved that in the subregion S_1, combined with part of S_0:

$$0<p'<k<p\&k^2\leqslant pp', \quad -1<a'<0, \ 0<a<1, \ 0<-a'/a\leqslant 1, \tag{20.45}$$

the ratio R_{cl}, $l=0,1,\ldots$, has, for *any fixed* $\gamma\neq 0$, an infinite number of zeros. These occur for

$$a'=-a'_{l;m}(\gamma), \quad m=M^s_l(\gamma), \ M^s_l(\gamma)+1,\ldots, \tag{20.46}$$

&

$$a=a^s_{l;m;n}(\gamma), \quad n=1,2,\ldots,N^s_l(\gamma;m). \tag{20.47}$$

Here $a'_{l;m}(\gamma)$ and $a^s_{l;m;n}(\gamma)$ are *even* functions of γ. The zeros of R_{cl} accumulate at $a'=0(p'=k)$ and at $a=0(p=k)$. Approximate expressions are

$$a'_{l;m}(\gamma)\approx\exp\{-(m\pi+\xi)/|\gamma|\}, \quad m\to\infty, \tag{20.48}$$

$$a^s_{l;m;n}(\gamma)\approx\exp\{-(n\pi+\xi)/|\gamma|\}, \quad m\to\infty, \tag{20.49}$$

where $\xi=\xi_l(\gamma)$ follows from Eq. (20.33). Further, $M^s_l(\gamma)$ and $N^s_l(\gamma;m)$ are natural numbers, and $N^s_l(\gamma;m)\to\infty$ when $m\to\infty$, for any fixed $\gamma\neq 0$. From Eqs. (20.10), (20.14), and (20.36) it follows that R_{cl} has *no* zero on S_0, i.e., for $pp'=k^2\Leftrightarrow a+a'=0$.

The zeros of R_{cl} in the remaining subregions S_i, i = 2,3, and 4, follow easily with the aid of the transformations to be discussed in Sec. 23. The remaining task in this section is to determine the zeros of R_{cl} for positive energy in the "*asymmetric*" case. In view of Sec. 23 we may restrict (p,p') to any one of the subregions A_i, i = 1,2,3,4. We take $(p,p')\in A_1$ which means that

$$0<k<p'<p \text{ and } 0<a'<a<1. \tag{20.50}$$

In this case $|R_{cl}|$ is not symmetric in γ, so that we have to consider $\gamma<0$ and $\gamma>0$ separately. Nevertheless, many of the formulas and derivations given for the symmetric case will be useful below.

We shall use the representation for R_{cl} given by Eq.

(3.16),

$$R_{cl}Q_l(w) = C_0^2 e^{\pi\gamma} Q_l^{i\gamma}(u) P_l^{-i\gamma}(u') + Q_l(w)$$
$$- C_0^2 e^{\pi\gamma}\pi^{-1}\int_0^\pi \cosh \gamma t Q_l(uu' + vv'\cos t)dt, \quad (20.51)$$
$$0<k<p'<p;\ 0<a'<a<1;\ \gamma\in\mathbb{R},$$

which follows from Eq. (20.1) by applying Eq. (3.21). After some manipulations, using in particular Eq. (20.7) we obtain

$$R_{cl}Q_l(w) = c_{l\gamma}C_0^2(a')^{i\gamma}P_l^{(i\gamma,-i\gamma)}(u')$$
$$\times i(2\gamma)^{-1}[a^{i\gamma}P_l^{(i\gamma,-i\gamma)}(u) - a^{-i\gamma}P_l^{(-i\gamma,i\gamma)}(u)]$$
$$- vv'(\sinh\pi\gamma)^{-1}\int_0^\pi \sinh\gamma t\sin t$$
$$\times Q_l'(uu' + vv'\cos t)dt, \quad (20.52)$$
$$0<a'<a<1;\ 1<u'<u.$$

By using the notations introduced in Eqs. (20.8)–(20.10), and defining

$$h(\gamma;a') := \Re_l'\cos(\gamma\ln a') - \Im_l'\sin(\gamma\ln a')$$
$$+ i[\Re_l'\sin(\gamma\ln a') + \Im_l'\cos(\gamma\ln a')], \quad (20.53)$$

we obtain from Eq. (20.52)

$$\sinh\pi\gamma Q_l(w)R_{cl}$$
$$= -\pi c_{l\gamma}e^{-\pi\gamma}f(\gamma;a)h(\gamma;a') + vv'q(\gamma), \quad (20.54)$$
$$0<a'<a<1;\ \gamma\in\mathbb{R},$$

which may be compared with Eq. (20.12). Setting the imaginary part of R_{cl} equal to zero we get [note that $q(\gamma)>0$]

$$\Re_l'\sin(\gamma\ln a') + \Im_l'\cos(\gamma\ln a') = 0. \quad (20.55)$$

This is a *necessary* condition for R_{cl} to be zero. The solutions a' of Eq. (20.55) for fixed $\gamma\neq 0$ are

$$a' = a_{l;m}'(\gamma), \quad \gamma\in\mathbb{R}' := \mathbb{R}\setminus\{0\}, \quad m = 1,2,\ldots. \quad (20.56)$$

Several remarks should be made here

(i) We exclude $\gamma = 0$ since $R_{cl}\equiv 1$ for $\gamma = 0$.

(ii) Since $\Re_l'$ is even and $\Im_l'$ odd in γ, the left-hand member of Eq. (20.55) is *odd* in γ; therefore, the solutions $a_{l;m}'(\gamma)$ are *even* in γ.

(iii) For $l = 0$ we have

$$a_{0;m}'(\gamma) = \exp(-m\pi/|\gamma|), \quad \gamma\in\mathbb{R}',\ m = 1,2,\ldots. \quad (20.57)$$

(iv) The solutions of Eq. (20.55) are [for $\gamma>0$] exactly equal to the negatives of the solutions of Eq. (20.14), cf. Eq. (20.30). This follows by observing that the argument of the Jacobi polynomial $P_l^{(i\gamma,-i\gamma)}$ [see Eq. (20.9)] is

$$u' \equiv (1 + a'^2)/(1 - a'^2), \quad (20.58)$$

which is even in a', so that $\Re_l'$ and $\Im_l'$ are also *even* in a'. This justifies our using the same notation $a_{l;m}'(\gamma)$ as in Eq. (20.30). From Eq. (20.32) we have

$$a_{l;m}'(\gamma)\approx\exp\{-(m\pi+\xi)/|\gamma|\}, \quad m\to\infty, \quad (20.59)$$
$$\tan\xi = \mathrm{Im}\binom{l+i\gamma}{l}\Big/\mathrm{Re}\binom{l+i\gamma}{l}.$$

Further, $h(\cdot;\cdot)$ is in fact the same function as $g(\cdot;\cdot)$ introduced by Eq. (20.11), so

$$h(\gamma;a_{l;m}'(\gamma)) = \Re_l'/\cos[\gamma\ln a'], \quad a' = a_{l;m}'(\gamma). \quad (20.60)$$

We define the functions $H_{l;m}(\gamma;a)$, $m = 1,2,\ldots$, by

$$H_{l;m}(\gamma;a) := \sinh\pi\gamma Q_l(w)R_{cl}[a' = a_{l;m}'(\gamma)]. \quad (20.61)$$

To avoid cumbersome notations, a superscript "a" denoting the asymmetric case is omitted here. Then we have from Eq. (20.54)

$$H_{l;m}(\gamma;a) = vv'q(\gamma)$$
$$- \pi e^{-\pi\gamma}c_{l\gamma}f(\gamma;a)\Re_l'/\cos(\gamma\ln a'), \quad (20.62)$$

subject to

$$a' = a_{l;m}'(\gamma);\quad 0<a'<a<1;\quad \gamma\in\mathbb{R}', \quad (20.63)$$

where $f(\gamma;a)$ is given by (20.10) and $q(\gamma)$ by (20.8). The zeros a of

$$H_{l;m}(\gamma;a),\quad m = 1,2,\ldots \quad (20.64)$$

give exactly all the zeros of R_{cl} we are seeking. With the aid of Eq. (20.40) we derive from Eq. (20.62) that the zeros of R_{cl} are approximately given for $m\to\infty$ by the solutions a of the equation

$$(-1)^{m+1}\pi f(\gamma;a)c_{l\gamma}^{1/2}e^{-\pi\gamma}(1+\gamma^2)/\sinh\pi\gamma$$
$$= 2\exp\{-(m\pi+\xi)/|\gamma|\}vQ_l'(u),\quad m\to\infty,\ \gamma\in\mathbb{R}'. \quad (20.65)$$

The solutions a of Eq. (20.65) are approximately given by the zeros a of $f(\gamma;a)$. Denoting the zeros a of R_{cl} by $a_{l;m;n}(\gamma)$, $n = 1,2,\ldots$, $\gamma\in\mathbb{R}'$, we have

$$a_{l;m;n}(\gamma)\approx\exp\{-(n\pi+\xi)/|\gamma|\},\quad m\to\infty; \quad (20.66)$$

cf. Eq. (20.42). As before, an additional condition must be met: For $n/|\gamma|$ large we get, inserting Eq. (20.23) into Eq. (20.65),

$$(-1)^m\pi f(\gamma;a)c_{l\gamma}^{1/2}e^{-\pi\gamma}(1+\gamma^2)/\sinh\pi\gamma$$
$$\approx[a^{-1} + 2l(l+1)a\ln a]$$
$$\times\exp\{-(m\pi+\xi)/|\gamma|\},\quad m\to\infty,\ a\downarrow 0,$$
$$\gamma\in\mathbb{R}'. \quad (20.67)$$

It follows that the approximation in Eq. (20.66) is only accurate if

$$n\leqslant N_l^a(\gamma;m) \quad (20.68)$$

for a certain $N_l^a(\gamma;m)\in\mathbf{N}$.

Summarizing: In the subregion A_1, i.e.,

$$0<k<p'<p;\quad 0<a'<a<1,$$

$R_{cl}, l = 0,1,\ldots$, has, for *any fixed* $\gamma\neq 0$, an infinite number of zeros. These occur for

$$a' = a_{l;m}'(\gamma),\quad m = M_l^a(\gamma),\quad M_l^a(\gamma)+1,\ldots, \quad (20.69)$$

&

$$a = a_{l;m;n}(\gamma),\quad n = 1,2,\ldots,N_l^a(\gamma;m). \quad (20.70)$$

Here $a_{l;m}'(\gamma)$ [the same function as used in Eq. (20.46)] is *even* in γ, but $a_{l;m;n}(\gamma)$ is *not even* in γ. The zeros of R_{cl} accumulate at $a' = 0$ ($p' = k$) and at $a = 0$ ($p = k$). Approximate expressions are

$$a_{l;m}'(\gamma)\approx\exp\{-(m\pi+\xi)/|\gamma|\},\quad m\to\infty,$$

which coincides with Eq. (20.48), and

$$a_{l;m;n}(\gamma)\approx\exp\{-(n\pi+\xi)/|\gamma|\},\quad m\to\infty; \tag{20.71}$$

cf. Eqs. (20.49) and (20.33). Further, $M_l^a(\gamma)$ and $N_l^a(\gamma;m)$ are natural numbers, and $N_l^a(\gamma;m)\to\infty$ when $m\to\infty$, for any fixed $\gamma\neq 0$.

The zeros of R_{cl} in the remaining subregions A_i, i = 2,3, and 4, follow easily with the aid of the transformations to be discussed in Sec. 23.

In Appendix Z of Ref. 32 we have derived some rough estimates, for $l=0$ only, for the cases (i)$|\gamma|\to\infty$, (ii)$|\gamma|=1$, and (iii) $\gamma\to 0$. The results are: (i) for $|\gamma|\to\infty$:

$$\text{Symmetric case } S_1\text{: } M^s(\gamma)\approx\gamma^2, \tag{20.72}$$

$$N^s(\gamma;\gamma^2)\approx(2/\pi)\gamma\ln\gamma\ (\gamma>0),$$

$$\text{Asymmetric case } A_1\text{: } M^a(\gamma)\approx 2\gamma^2,\ \gamma>0 \tag{20.73}$$

$$\approx -\gamma,\ \gamma<0;$$

(ii) for $|\gamma|=1$:

$$\text{Symmetric case } S_1(\gamma=1)\text{: } a'=-a'_m(1)=e^{-m\pi}, \tag{20.74}$$

$$a=a^s_{m;n}(1)\approx e^{-n\pi}[1+(-1)^{n+1-m}$$

$$\times\exp\{(n+1-m)\pi\}(4\pi)^{-1}], \tag{20.75}$$

$$m=2,3,\ldots;\ n=1,2,\ldots,m-1;$$

$$\text{Asymmetric case } A_1\text{: } a'=a'_m(1)=e^{-m\pi}; \tag{20.76}$$

$$\gamma=+1\text{: } a_{m;n}(1)\approx e^{-n\pi}[1+(-1)^{n-m}$$

$$\times\exp\{(n+2-m)\pi\}(4\pi)^{-1}], \tag{20.77}$$

$$m=3,4,\ldots;\ n=1,2,\ldots,m-2;$$

$$\gamma=-1\text{: } a_{m;n}(-1)\approx e^{-n\pi}[1+(-1)^{n-m}$$

$$\times\exp\{(n-m)\pi\}(4\pi)^{-1}], \tag{20.78}$$

$$m=1,2,\ldots;\ n=1,2,\ldots,\lesssim m;$$

(iii) for $\gamma\to 0$:

$$a'_m(\gamma)\approx\exp\{-m\pi/|\gamma|\},\quad m=1,2,\ldots, \tag{20.79}$$

$$a^{(s)}_{m;n}(\gamma)\approx\exp\{-n\pi/|\gamma|\},\quad n=1,2,\ldots,m-1, \tag{20.80}$$

for both S_1 and A_1.

No essentially new problems are expected in deriving analogous rough estimates for the zeros of R_{cl} for $l>0$.

21. LIMITS OF R_c

In this section we shall evaluate various interesting limits of R_c. Some of these will serve for proving that the inequalities for R_c to be given in Sec. 24 are *optimal*.

We shall use Eq. (2.1), i.e.,

$$R_c=1-i\gamma(\rho+1)\int_0^\infty e^{-i\gamma t}/(\rho+\cosh t)\,dt,$$

$$\text{Re } i\gamma>-1. \tag{21.1}$$

When the energy is negative, $i\gamma\equiv -s/\kappa$ is real; further,

$$\rho=(1+x^2)/(1-x^2),$$

and

$$x^2=1-q^{-2}\kappa^{-2}(p^2+\kappa^2)(p'^2+\kappa^2)$$

$$=-q^{-2}\kappa^{-2}[p^2p'^2+2\kappa^2pp'\cos\theta+\kappa^4],$$

so that $x^2\leqslant 0$ for $\kappa>0$, and $-1<\rho\leqslant 1$. In Ref. 24 it has been proved that

$$R_c\to 1\quad \text{for } \rho\to -1, \text{ i.e., } x^2\to-\infty. \tag{21.2}$$

Clearly this limit corresponds to

$$\text{either } q\to 0 \text{ or } (p\to\infty\ \&\ p'\to\infty, \text{ any } \theta). \tag{21.3}$$

Since $q^2=p^2+p'^2-2pp'\cos\theta$ we have

$$(q\to 0)\Leftrightarrow(p-p'\to 0\ \&\ \theta\to 0).$$

When $\rho=1$ we get

$$R_c=B_{c0}(-s/\kappa)$$

$$=1+2(s/\kappa)\int_0^\infty e^{ts/\kappa}/(1+\cosh t)\,dt,\quad s<\kappa. \tag{21.4}$$

Clearly $\rho=1$ corresponds to $x=0$, i.e.,

$$pp'=\kappa^2\ \&\ \theta=\pi. \tag{21.5}$$

For the *limit* $x\to 0$ the only other possibility is

$$p\to 0\ \&\ p'\to\infty \text{ or } p\to\infty\ \&\ p'\to 0 \text{ (any } \theta). \tag{21.6}$$

Now we are going to investigate the positive-energy case. Then $x^2>0$; as usual we take $x<0$. Then we have [cf. Sec. 23]

$$(p,p')\in S\text{: } -1<x\leqslant 0;\quad -1\leqslant y<0;\quad 1\leqslant\rho<\infty; \tag{21.7}$$

$$(p,p')\in A\text{: } x<-1;\quad 0<y<1;\quad -\infty<\rho<-1. \tag{21.8}$$

One easily verifies that

$$\lim_{\rho\downarrow 1}R_c=B_{c0}(i\gamma). \tag{21.9}$$

This limit corresponds to $x\uparrow 0$ or $y\downarrow -1$, which can only occur if either

$$pp'\to k^2\ \&\ \theta\to 0 \tag{21.10a}$$

or

$$p\to 0\ \&\ p'\to\infty \text{ or } p\to\infty\ \&\ p'\to 0 \text{ (any } \theta). \tag{21.10b}$$

The limit of R_c for $x\to-\infty$ ($\rho\uparrow -1$; $y\uparrow 1$) is 1, see Ref. 24. In terms of p,p', and θ we get the same result as in (21.3). The remaining point of interest is $x=-1$. We need Eqs. (2.8) and (2.9) to derive the appropriate limits of R_c. For the symmetric case S we obtain

$$\lim_{y\uparrow 0}(-y)^{-i\gamma}R_c=C_0^2e^{\pi\gamma}=\lim_{y\uparrow 0}|R_c|, \tag{21.11}$$

and for the asymmetric case A

$$\lim_{y\downarrow 0}y^{-i\gamma}R_c=C_0^2=\lim_{y\downarrow 0}|R_c|. \tag{21.12}$$

These limits correspond to either $p\to k$ or $p'\to k$ or $p\to k\ \&\ p'\to k$, provided q does not go to zero.

Remark: In view of the limit of Eq. (21.11), a *necessary* condition for the validity of the conjecture (c1) [Eq. (17.1)]

$$|R_c|\overset{(c1)}{\leqslant}|B_{c0}(i\gamma)|,$$

with equality iff $pp'=k^2\ \&\ \theta=0$, is that

$$C_0^2e^{\pi\gamma}<|B_{c0}(i\gamma)|. \tag{21.13}$$

For a proof of this inequality the reader is referred to Appendix I of Ref. 32.

Finally we shall evaluate

$$\lim_{p\to\infty} p \lim_{p'\to\infty} (R_c - 1). \tag{21.14}$$

(i) For negative energy we have

$$\rho + 1 = 2/(1-x^2) \underset{p'\to\infty}{\to} 2\kappa^2/(p^2+\kappa^2) \underset{p\to\infty}{\approx} 2\kappa^2/p^2.$$

By using Eq. (21.1) it is easily seen that, in order to evaluate (21.14), the interval of integration may be replaced by $[0,\epsilon]$ for any $\epsilon > 0$. Consequently the exponential function may be set equal to 1, and $\cosh t$ may be replaced by $1 + t^2/2$. Then (21.14) becomes

$$\lim_{p\to\infty} 4p\kappa s \int_0^{\epsilon} \frac{dt}{(4\kappa^2 + p^2 t^2)} = 2s \lim_{p\to\infty} \int_0^{\epsilon p/2\kappa} \frac{d\tau}{(1+\tau^2)} = \pi s, \tag{21.15}$$

where $\tau := pt/2\kappa$. Since R_c is invariant under the inversion (t_3) (see Sec. 23),

$$(t_3): p \to \kappa^2/p \;\&\; p' \to \kappa^2/p', \tag{21.16}$$

it follows that

$$\lim_{p\downarrow 0}\left(\frac{\kappa^2}{p}\right) \lim_{p'\downarrow 0} (R_c - 1) = \pi s. \tag{21.17}$$

(ii) For positive energy Eq. (21.1) cannot be applied here, since $\rho + 1 \approx -2k^2/p^2 < 0$. Instead we use the representation given by Eqs. (2.8) and (2.9)

$$R_c = 1 + C_0^2 y^{i\gamma}(1-y)/(1+y) - \frac{\gamma(1-y)^2}{\sinh \pi\gamma} \int_0^{\pi} \frac{\cosh \gamma t\, dt}{1 + y^2 + 2y\cos t}, \quad 0 < y < 1. \tag{21.18}$$

Since $\lim_{p'\to\infty} y = a = (p-k)/(p+k)$ and $\lim_{p\to\infty} p(1-a) = 2k$, we obtain for (21.14) the expression

$$kC_0^2 - 2k\gamma[\sinh \pi\gamma]^{-1} \times \lim_{y\uparrow 1} (1-y) \int_0^{\pi} (1 + y^2 + 2y\cos t)^{-1} \cosh \gamma t\, dt.$$

Since only $t \approx \pi$ contributes to the limit, we may replace $\cosh \gamma t$ by $\cosh \gamma\pi$. Using further (e.g., Ref. 42, p. 366)

$$\int_0^{\pi} (1 + y^2 + 2y\cos t)^{-1} dt = \frac{\pi}{(1-y^2)}, \quad y^2 < 1, \tag{21.19}$$

we easily find

$$\lim_{p\to\infty} p \lim_{p'\to\infty} (R_c - 1) = -\pi k\gamma = \pi s. \tag{21.20}$$

We point out that this result can also be obtained from Eqs. (4.1) and (4.3) of Ref. 24, or from Eq. (7.8). In view of the inversion (t_2) (see Sec. 23), we have

$$\lim_{p\downarrow 0}\left(\frac{k^2}{p}\right) \lim_{p'\downarrow 0} (R_c - 1) = \pi s. \tag{21.21}$$

22. LIMITS OF R_{cl}

In this section we shall evaluate various interesting limits of R_{cl}, $l = 0,1,\ldots$. Some of these will serve for proving that the inequalities for R_{cl} to be given in Sec. 24 are *optimal*.

First we shall consider the negative-energy case; then the following representation is useful:

$$R_{cl} = 1 + s[\kappa Q_l(w)]^{-1} \int_0^1 t^{-1-s/\kappa} \times Q_l(w - vv'(t - 2 + 1/t)/2)dt, \tag{22.1}$$

$\kappa > 0$, $s \in \mathbb{R}'$, $w := (p^2 + p'^2)/2pp' > 1$,

$-vv' = (p^2 + \kappa^2)(p'^2 + \kappa^2)/4\kappa^2 pp' > 0.$

It is easy to see that[24]

$$\lim_{w\downarrow 1} R_{cl} = 1. \tag{22.2}$$

Since $Q_l(z)$ is decreasing in z, and

$$\lim_{z\to\infty} z^{l+1} Q_l(z) = \frac{l!}{(2l+1)!!}, \tag{22.3}$$

the only other possibility for $R_{cl} \to 1$ is $-vv'$ and/or $w \to \infty$. Therefore, we shall investigate the cases p or $p' \to 0$ or ∞. From Eqs. (22.1) and (22.3) we obtain

$$\lim_{p'\downarrow 0} R_{cl} = 1 + \left(\frac{s}{\kappa}\right) \int_0^1 t^{-1-s/\kappa} \times [1 + (1 + \kappa^2/p^2)(1-t)^2/4t]^{-l-1} dt. \tag{22.4}$$

(i) The limit of the right-hand member of Eq. (22.4) for $p \to \infty$ is elementary; one has

$$\lim_{p\to\infty} \lim_{p'\downarrow 0} R_{cl} = 1 + \left(\frac{s}{\kappa}\right) 4^{l+1} \times \int_0^1 t^{l-s/\kappa} (1+t)^{-2l-2} dt,$$

which is just one of the expressions for B_{cl} (see Sec. 13). In view of the symmetries of R_{cl} (Sec. 23) we therefore have

$$\lim_{p\to\infty} \lim_{p'\downarrow 0} R_{cl} = \lim_{p'\downarrow 0} \lim_{p\to\infty} R_{cl} = B_{cl}\left(\frac{-s}{\kappa}\right), \tag{22.5}$$

where p and p' may be interchanged. In Ref. 24 we have proved, by using monotonicity properties of R_{cl}, that $B_{cl}(-s/\kappa)$ is the supremum of R_{cl} when $s > 0$, and its infimum when $s < 0$; it was derived there that $pp' = \kappa^2$ gives a relative extremum and that

$$\lim_{w\to\infty} R_{cl}(pp' = \kappa^2) = B_{cl}\left(\frac{-s}{\kappa}\right). \tag{22.6}$$

We note that Eqs. (22.5) and (22.6) give the partial-wave analogs of (21.6) and (21.5), respectively.

(ii) The limit of the right-hand member of Eq. (22.4) for $p \downarrow 0$ is somewhat more complicated. This is due to the behavior of the integrand at $t = 1$, which apparently interferes with its behavior at $p = 0$. In fact we may, for all $l = 0,1,\ldots$, replace the interval of integration by $[1-\epsilon,1]$ for any $\epsilon \in (0,1)$. Taking $\epsilon \downarrow 0$ we may set $t^{-1-s/\kappa}$ equal to 1, so that

$$\lim_{p\downarrow 0}\left(\frac{\kappa^2}{p}\right)\lim_{p'\downarrow 0}(R_{cl}-1)$$
$$=\lim_{\epsilon\downarrow 0}\lim_{p\downarrow 0}\left(\frac{\kappa s}{p}\right)\times\int_{1-\epsilon}^{1}\left[1+\frac{(1+\kappa^2/p^2)(1-t)^2}{4}\right]^{-l-1}dt.$$

Introducing $\tau:=(1-t)\kappa/2p$ we get for this limit

$$\lim_{\epsilon\downarrow 0}\lim_{p\downarrow 0}2s\int_0^{\epsilon\kappa/2p}\left[1+\left(1+\frac{p^2}{\kappa^2}\right)\tau^2\right]^{-l-1}dt$$
$$=2s\int_0^{\infty}(1+\tau^2)^{-l-1}\,d\tau.$$

Since (e.g., Ref. 42, p. 295)

$$\int_0^{\infty}(1+\tau^2)^{-l-1}\,d\tau=\pi 2^{-2l-1}\binom{2l}{l},\tag{22.7}$$

we have proved in this way that

$$\lim_{p\downarrow 0}\left(\frac{\kappa^2}{p}\right)\lim_{p'\downarrow 0}(R_{cl}-1)=\pi s 4^{-l}\binom{2l}{l},\tag{22.8a}$$

which should be compared with Eq. (21.17). In view of the inversion (t_3) [see Sec. 23] we also have

$$\lim_{p\to\infty}p\lim_{p'\to\infty}(R_{cl}-1)=\pi s 4^{-l}\binom{2l}{l}.\tag{22.8b}$$

Obviously these results imply that

$$\lim_{p\downarrow 0}\lim_{p'\downarrow 0}R_{cl}=\lim_{p\to\infty}\lim_{p'\to\infty}R_{cl}=1;\tag{22.9}$$

note that p and p' may be interchanged in Eqs. (22.8) and (22.9).

Now we are going to investigate the positive-energy case. Zeros of R_{cl} have been derived in Sec. 20; see also the inequalities (i13), (i15), (i21), and (i23) in Sec. 24. The upper bounds given by (i22) and (i24) have been derived in Sec. 6.

First we consider the symmetric case (Sec. 23). Then $vv'<0$, $|R_{cl}|$ is even in γ, and the following representation,

$$R_{cl}=1-i\gamma[Q_l(w)]^{-1}\int_0^1 t^{i\gamma-1}\times Q_l(uu'-vv'(t+1/t)/2)dt,\quad vv'<0,\tag{22.10}$$

is useful. This is merely Eq. (22.1) rewritten, so we can take over some of the results obtained for the negative-energy case. Note, however, that $vv'<0$ implies that either $p'<k<p$ or $p<k<p'$. Therefore, the analog of Eq. (22.2) does not hold here: $w\to 1$ can be obtained only for $p\to k$ & $p'\to k$ such that $vv'\uparrow 0$; it is easily seen that the evaluation of a limit, if any, from Eq. (22.10) is difficult in this case. The limit of R_{cl} for $p'\downarrow 0$ is obtained in the same way as before; see Eq. (22.4). Taking then the limit for $p\to\infty$ we obtain the analog of Eq. (22.5), i.e.,

$$\lim_{p\to\infty}\lim_{p'\downarrow 0}R_{cl}=\lim_{p'\downarrow 0}\lim_{p\to\infty}R_{cl}=B_{cl}(i\gamma),\tag{22.11}$$

where p and p' may be interchanged.

To study the behavior of R_{cl} at $p'\uparrow k$ we use the representation given by Eq. (20.6), which holds for $p'<k<p$. Since $w\to u$, $u'\downarrow 1$ and $v'\uparrow 0$ for $p'\uparrow k$, the combination of the last two terms on the right-hand side of Eq. (20.6) tends to 0 in this limit. Then it is easily verified that

$$\lim_{p'\uparrow k}R_{cl}Q_l(w)(-a')^{-i\gamma}=-c_{l\gamma}C_0^2e^{\pi\gamma}\times\binom{l+i\gamma}{l}\gamma^{-1}\,\mathrm{Im}\{a^{i\gamma}P_l^{(i\gamma,\,-i\gamma)}(u)\},$$
$$p>k,\ a>0,\ \gamma\neq 0.\tag{22.12}$$

Taking now $p\downarrow k$, i.e., $a\downarrow 0$, we see that

$$\lim_{p\downarrow k}\lim_{p'\uparrow k}|R_{cl}|Q_l(w)$$

does not exist; $|R_{cl}|\,Q_l(w)$ keeps oscillating between zero (see Sec. 20) and

$$\pi|P_l^{(i\gamma,\,-i\gamma)}(u)|\cdot\left|\binom{l+i\gamma}{l}\sinh\pi\gamma\right|^{-1};\tag{22.13a}$$

the limit superior follows directly from this expression

$$\limsup_{p\downarrow k}\lim_{p'\uparrow k}|R_{cl}|Q_l(w)=\frac{\pi}{|\sinh\pi\gamma|},\quad\gamma\neq 0.\tag{22.13b}$$

Here we have used

$$c_{l\gamma}^{-1}=\binom{l+i\gamma}{l}\binom{l-i\gamma}{l},\quad C_0^2\,e^{\pi\gamma}=\pi\gamma/\sinh\pi\gamma,$$

and

$$\max_{\alpha}|x\sin\alpha+y\cos\alpha|=(x^2+y^2)^{1/2},$$

where $\alpha=\gamma\ln a$. Since $Q_l(w)\to\infty$ for $w\downarrow 1$ it is clear that

$$\lim_{p\downarrow k}\lim_{p'\uparrow k}R_{cl}=0.\tag{22.14}$$

Now we are going to investigate the behavior of R_{cl} for $p'\uparrow k$ and $p\to\infty$. Then Eq. (22.12) is not convenient; instead we use the representation of Eq. (4.2), i.e.,

$$R_{cl}Q_l(w)=C_0^2e^{2\pi\gamma}Q_l^{i\gamma}(u)P_l^{-i\gamma}(u')+Q_l(w)-\gamma^2\sum_{n=0}^{\infty}\epsilon_n(n^2+\gamma^2)^{-1}Q_l^n(u)P_l^{-n}(u'),\tag{22.15}$$

$vv'<0$; $0<p'<k<p$ & $k^2\leqslant pp'$; $1<u'\leqslant u$.

By using

$$\lim_{p'\uparrow k}(-a')^{-i\gamma}P_l^{-i\gamma}(u')=1/\Gamma(1+i\gamma),\tag{22.16}$$

we get from (22.15) [cf. (22.12)],

$$\lim_{p'\uparrow k}(-a')^{-i\gamma}R_{cl}=\frac{\Gamma(1-i\gamma)e^{\pi\gamma}Q_l^{i\gamma}(u)}{Q_l(u)},\quad p>k.\tag{22.17}$$

By using further (e.g., Ref. 40, p. 197)

$$\lim_{z\to\infty}z^{l+1}e^{-i\pi\mu}Q_l^{\mu}(z)=\frac{2^l l!\Gamma(l+1+\mu)}{(2l+1)!},\tag{22.18}$$

we find

$$\lim_{p\to\infty}\lim_{p'\uparrow k}(-a')^{-i\gamma}R_{cl}=\frac{\Gamma(1-i\gamma)\Gamma(l+1+i\gamma)}{l!}=\Gamma(1+i\gamma)\Gamma(1-i\gamma)\binom{l+i\gamma}{l};\tag{22.19}$$

hence,

$$\lim_{p\to\infty}\lim_{p'\downarrow k}|R_{cl}| = C_0^2 e^{\pi\gamma}\left|\binom{l+i\gamma}{l}\right|. \tag{22.20}$$

It is not difficult to verify that the order in which the two limits are taken may be reversed in (22.19) as well as in (22.20). In view of the transformation invariances of R_{cl}, see Sec. 23, Eq. (22.20) holds also when p and p' are interchanged, and furthermore we have, by using the inversion (t_2): $p\to k^2/p$ & $p'\to k^2/p'$,

$$\lim_{p\downarrow 0}\lim_{p'\downarrow k}|R_{cl}| = \lim_{p'\downarrow k}\left|\lim_{p\downarrow 0} R_{cl}\right| = C_0^2 e^{\pi\gamma}\left|\binom{l+i\gamma}{l}\right|. \tag{22.21}$$

Remark: In view of these limits, a *necessary* condition for the validity of the conjecture (c2) [see Sec. 24],

$$|R_{cl}| \overset{(c2)}{<} |B_{cl}(i\gamma)| \tag{22.22}$$

for the symmetric case, is that

$$C_0^2\, e^{\pi\gamma}\left|\binom{l+i\gamma}{l}\right| < |B_{cl}(i\gamma)| \tag{22.23}$$

in this case. Note that in (22.22) as well as in (22.23), both members are even in γ. We have been able to prove (22.23) only for $l=0$; cf. Eq. (21.13). So (22.23) is our fifth conjecture (cf. Sec. 17), which we denote by (c2)', since (c2)$\Rightarrow$(c2)'.

Next we consider the asymmetric case. Then $vv'>0$ so we use the representation given by Eq. (4.4) rather than the one of Eq. (4.2), and alternatively, Eq. (20.52) instead of Eq. (20.6). We have to modify previous derivations, from Eq. (22.12) on, only slightly. With $p'>k$ & $p>k$ $(a'>0, a>0)$, the principal difference from preceding formulas is that we have $(a')^{i\gamma}$ instead of $(-a')^{i\gamma}$ and C_0^2 instead of $C_0^2 e^{\pi\gamma}$. So we get from (20.52) the analog of Eq. (22.12)

$$\lim_{p'\downarrow k} R_{cl}Q_l(w)(a')^{i\gamma} = -c_{l\gamma}C_0^2 \times\binom{l+i\gamma}{l}\gamma^{-1}\,\mathrm{Im}\{a^{i\gamma}P_l^{(i\gamma,\,-i\gamma)}(u)\},$$

$$p>k,\ a>0,\ \gamma\neq 0. \tag{22.24}$$

Hence

$$\lim_{p\downarrow k}\lim_{p'\downarrow k}|R_{cl}|Q_l(w) \tag{22.25}$$

does not exist, and

$$\limsup_{p\downarrow k}\lim_{p'\downarrow k} R_{cl}Q_l(w) = 2\pi|e^{2\pi\gamma}-1|^{-1},\quad \gamma\neq 0, \tag{22.26}$$

$$\lim_{p\downarrow k}\lim_{p'\downarrow k} R_{cl} = 0. \tag{22.27}$$

The analog of Eq. (22.17) is

$$\lim_{p'\downarrow k}(a')^{-i\gamma}R_{cl} = \frac{\Gamma(1-i\gamma)Q_l^{i\gamma}(u)}{Q_l(u)},\quad p>k, \tag{22.28}$$

and the analogs of (22.19) and (22.20) are[23]

$$\lim_{p\to\infty}\lim_{p'\downarrow k}(a')^{-i\gamma}R_{cl} = C_0^2\binom{l+i\gamma}{l}, \tag{22.29}$$

and

$$\lim_{p\to\infty}\lim_{p'\downarrow k}|R_{cl}| = C_0^2\left|\binom{l+i\gamma}{l}\right|, \tag{22.30}$$

respectively. Related results follow by applying the transformations (t_1) and (t_2) (Sec. 23).

Remark: In relation to the inequalities (i22) and (i24), see Sec. 24, we point out that

$$C_0^2\left|\binom{l+i\gamma}{l}\right| \begin{cases} >1, & \gamma<0,\\ <1, & \gamma>0. \end{cases} \tag{22.31}$$

The proof follows easily from (cf. Ref. 23)

$$C_0^2 = 2\pi\gamma/\{\exp(2\pi\gamma)-1\} = \pi\gamma\exp(-\pi\gamma)/\sinh\pi\gamma,$$

$$\frac{\pi\gamma}{\sinh\pi\gamma} = \prod_1^\infty\left(1+\frac{\gamma^2}{n^2}\right)^{-1} < 1, \tag{22.32}$$

$$\left|\binom{l+i\gamma}{l}\right|^2 = \prod_1^l\left(1+\frac{\gamma^2}{n^2}\right) > 1, \tag{22.33}$$

$$\frac{\left|\binom{l+i\gamma}{l}\right|^2\pi\gamma}{\sinh\pi\gamma} = \prod_{l+1}^\infty\left(1+\frac{\gamma^2}{n^2}\right)^{-1} < 1. \tag{22.34}$$

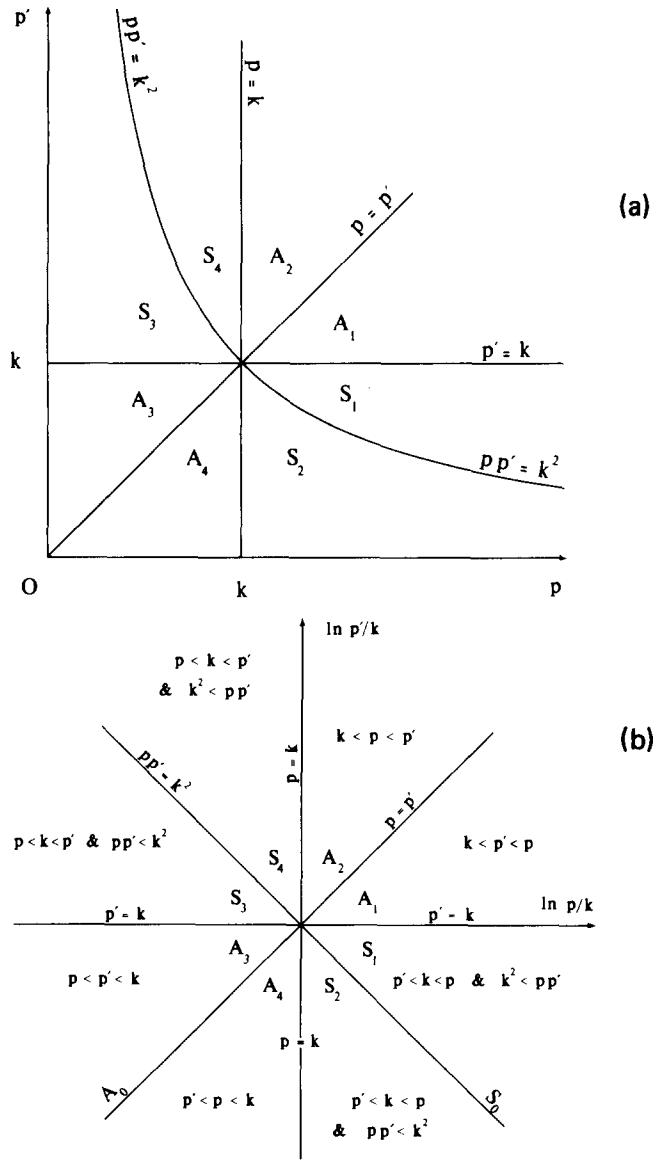


FIG. 3. The subregions S_i and A_i (Sec. 23): (a) in the (p, p') plane; (b) in the $(\ln p/k, \ln p'/k)$ plane.

When $\gamma<0$, the right-hand member of Eq. (22.30) is the supremum of $|R_{cl}|$, but for $\gamma>0$ it is not: In this case the supremum is 1. The corresponding inequalities (see Sec. 24),

$$|R_{cl}| \overset{\text{(i22)}}{<} 1, \quad \gamma>0,$$

$$|R_{cl}| \overset{\text{(i24)}}{<} C_0^2\left|\binom{l+i\gamma}{l}\right|, \quad \gamma<0,$$

have been proved in Sec. 6. In view of Eqs. (22.30) and (22.31) we only have to prove that the value 1 is approached arbitrarily closely, which is easy. By using, e.g., the representation of Eq. (3.16), which holds for $k<p'<p$, we have

$$\lim_{w\downarrow 1} R_{cl}=1, \quad \gamma\in\mathbb{R}, \tag{22.35}$$

as can be seen upon inspection. The limit $w\downarrow 1$, implying $p-p'\to 0$, has to be taken such that

$$(p-k)(p'-k)\geqslant\epsilon>0. \tag{22.36}$$

Finally we shall evaluate

$$\lim_{p\to\infty} p \lim_{p'\to\infty}(R_{cl}-1) \tag{22.37}$$

for positive energy. Using Eq. (22.18) and the representation given by Eq. (3.16), with p and p' interchanged, we obtain

$$\lim_{p'\to\infty}(R_{cl}-1)=C_0^2\left(\frac{k}{p}\right)^{l+1}\left[\frac{P_l^{-i\gamma}(u)\Gamma(l+1+i\gamma)}{l!}-e^{\pi\gamma}\pi^{-1}\int_0^{\pi}(u+v\cos t)^{-l-1}\cosh\gamma t\,dt\right]. \tag{22.38}$$

Further we use (e.g., Ref. 40, p. 197)

$$\lim_{z\to\infty} z^{-l}P_l^{\mu}(z)=\frac{(2l-1)!!}{\Gamma(l+1-\mu)}. \tag{22.39}$$

In this way we find for the expression (22.37)

$$C_0^2\,k\,4^{-l}\binom{2l}{l}-C_0^2e^{\pi\gamma}\pi^{-1}2^{l+1}\lim_{p\to\infty} p\times\int_0^{\pi}\left[1-\cos t+\left(\frac{p}{k}\right)^2(1+\cos t)\right]^{-l-1}\cosh\gamma t\,dt. \tag{22.40}$$

One easily verifies that the interval of integration may be replaced by $[\pi-2\epsilon,\pi]$ where $\epsilon\downarrow 0$ is taken after $p\to\infty$. Then $\cosh\gamma t$ may effectively be replaced by $\cosh\gamma\pi$, and it turns out to be convenient to introduce $\tau:=(p/k)\cot(t/2)$. In this way the second term of (22.40) is reduced to

$$-2kC_0^2e^{\pi\gamma}\pi^{-1}\cosh\pi\gamma\lim_{\epsilon\downarrow 0}\lim_{p\to\infty}\times\int_0^{(p/k)\tan\epsilon}(1+\tau^2)^{-l-1}\left(1+\frac{\tau^2k^2}{p^2}\right)^l d\tau. \tag{22.41}$$

The double limit of the integral here is equal to the integral given by Eq. (22.7). Combining the resulting expression with the first term of (22.40) we obtain

$$\lim_{p\to\infty} p\lim_{p'\to\infty}(R_{cl}-1)=\pi s4^{-l}\binom{2l}{l}. \tag{22.42}$$

In view of the inversion (t_2) [see Sec. 23] we have

$$\lim_{p\downarrow 0}\left(\frac{k^2}{p}\right)\lim_{p'\downarrow 0}(R_{cl}-1)=\pi s4^{-l}\binom{2l}{l}; \tag{22.43}$$

TABLE I. Twenty-four inequalities for R_c and R_{cl}; see Sec. 24 and cf. Table II. In case S, $|R_c|$ is even in γ, and in case S_l, $|R_{cl}|$ is even in γ. Hence (i9) is identical to (i11), (i10) to (i12), (i13) to (i15), and (i14) to (i16). The inequalities (i1)–(i24) are all valid. We conjecture that (i10) and (i12) should each be replaced by (c1), and (i14) and (i16) by (c2). These conjectured inequalities (c1) and (c2) are optimal if valid. All other inequalities are optimal.

Classification		Repulsion: $s<0$: $k\gamma>0$	Attraction: $s>0$: $k\gamma<0$
	Negative energy: $k^2<0$: $-ik=\kappa>0$		
N	$vv'<0$	(i1) (i2) $B_{c0}(-s/\kappa)\leqslant R_c<1$	(i3) (i4) $1<R_c\leqslant B_{c0}(-s/\kappa)$, $0<s/\kappa<1$
N_l	$vv'<0$	(i5) (i6) $B_{cl}(-s/\kappa)<R_{cl}<1$	(i7) (i8) $1<R_{cl}<B_{cl}(-s/\kappa)$, $0<s/\kappa<l+1$
	Positive energy: $k^2>0$: $k>0$	Conjectures: (c1) $\lvert R_c\rvert\leqslant\lvert B_{c0}(i\gamma)\rvert$ and (c2) $\lvert R_{cl}\rvert<\lvert B_{cl}(i\gamma)\rvert$	
S	$vv'<0$	(i9) (i10) $Z_c^s\leqslant\lvert R_c\rvert<1$	(i11) (i12) $Z_c^s\leqslant\lvert R_c\rvert<1$
S_l	$vv'<0$	(i13) (i14) $0\leqslant\lvert R_{cl}\rvert<1$	(i15) (i16) $0\leqslant\lvert R_{cl}\rvert<1$
A	$vv'>0$	(i17) (i18) $Z_c^a\leqslant\lvert R_c\rvert<1$	(i19) (i20) $1<\lvert R_c\rvert<C_0^2\equiv e^{-\pi\gamma}\pi\gamma/\sinh\pi\gamma$
A_l	$vv'>0$	(i21) (i22) $0\leqslant\lvert R_{cl}\rvert<1$	(i23) (i24) $0\leqslant\lvert R_{cl}\rvert<C_0^2\lvert\binom{l+i\gamma}{l}\rvert$

cf. Eqs. (22.8), which give almost the same results for the negative-energy case. As before we have

$$\lim_{p\downarrow 0}\lim_{p'\downarrow 0} R_{cl} = \lim_{p\to\infty}\lim_{p'\to\infty} R_{cl} = 1, \tag{22.44}$$

where p and p' may be interchanged.

23. TRANSFORMATION INVARIANCES OF R_c AND R_{cl}

The Coulomb T matrices $\langle \mathbf{p}|T_c|\mathbf{p}'\rangle$ and $\langle p|T_{cl}|p'\rangle$, hence $\langle \mathbf{p}|V_c|\mathbf{p}'\rangle$ and $\langle p|V_{cl}|p'\rangle$, and consequently R_c and R_{cl} are invariant under each one of the transformations:

(t_0): $\hat{p} \leftrightarrow \hat{p}'$,

(t_1): $p \leftrightarrow p'$,

(t_2): $p \to k^2/p \,\&\, p' \to k^2/p'$ $(E = k^2 > 0)$,

(t_3): $p \to \kappa^2/p \,\&\, p' \to \kappa^2/p'$ $(-E = \kappa^2 > 0)$;

the proofs follow easily from the appropriate representations; cf. Refs. 6, 10, and 22. Clearly this invariance is maintained under any combination of $(t_0) - (t_3)$, e.g.,

$(t_{01}) := (t_0)\,\&\,(t_1)$: $\mathbf{p} \leftrightarrow \mathbf{p}'$,

$(t_{12}) := (t_1)\,\&\,(t_2)$: $p \to k^2/p' \,\&\, p' \to k^2/p$;

TABLE II. Limits of R_c and R_{cl} in relation to the inequalities of Table I; see Sec. 24. The ratios R_c and R_{cl} are invariant for $p \leftrightarrow p'$ and for $p \to |E|/p \,\&\, p' \to |E|/p'$, $E \in \mathbf{R}'$. Superscripts r indicate Coulomb repulsion, and a attraction. N indicates negative energy. For positive energy, A indicates the asymmetric case, and S the symmetric case, in which $|R_c|$ and $|R_{cl}|$ are even in γ; consequently (i9) and (i11) are identical, and further (i10) and (i12), (i13) and (i15), and (i14) and (i16) are pairwise identical. N.B. In (i21) the infimum 0 is obtained for $p\downarrow k \,\&\, p'\downarrow k$ if these two limits are evaluated in succession. *Any* value in [0, 1] (and no other value) can be obtained by evaluating a particular combined limit: For $p - k = k^{1-\alpha}(p - p')^{\alpha}$ $(0 < \alpha < 1)$, $R_{cl} \to 1 - \alpha$ when $p' \uparrow p$.

Class	Ineq	Type	Approached/attained for	Ineq	Type	Approached/attained for
N^r	i1	MIN INF:	$pp' = \kappa^2 \& \theta = \pi$; for $p\to 0 \& p'\to\infty$ or $p\to\infty \& p'\to 0$, any θ	i2	SUP	$q\to 0$ or $p\to\infty \& p'\to\infty$, any θ
N^a	i3	INF	$q\to 0$ or $p\to\infty \& p'\to\infty$, any θ	i4	MAX SUP:	$pp' = \kappa^2 \& \theta = \pi$; for $p\to 0 \& p'\to\infty$ or $p\to\infty \& p'\to 0$, any θ
N_l^r	i5	INF	$p\to 0 \& p'\to\infty$ or $p\to\infty \& p'\to 0$; relative MIN for $pp' = \kappa^2$	i6	SUP	$w\to 1$ or $p\to 0 \& p'\to 0$ or $p\to\infty \& p'\to\infty$
N_l^a	i7	INF	$w\to 1$ or $p\to 0 \& p'\to 0$ or $p\to\infty \& p'\to\infty$	i8	SUP	$p\to 0 \& p'\to\infty$ or $p\to\infty \& p'\to 0$; relative MAX for $pp' = \kappa^2$
S^r	i9	MIN	$(p-k)(p'-k) \lesssim 0$?	i10	NO SUP	
S^a	i11	idem	idem	i12	idem	
S	—	NO EXTR	$\lvert R_c\rvert \uparrow\downarrow C_0^2 e^{\pi\gamma}$ for $(p-k)(p'-k)\uparrow 0$: $\theta \geqslant \epsilon > 0$	c1	MAX? SUP?	$pp' = k^2 \& \theta = 0$; for $p\to 0 \& p'\to\infty$ or $p\to\infty \& p'\to 0$, any θ
S_l^r	i13	MIN INF:	$(p-k)(p'-k) \lesssim 0$; $p\downarrow k \,\&\, p'\uparrow k$ or $p\uparrow k \,\&\, p'\downarrow k$	i14	NO SUP	
S_l^a	i15	idem	idem	i16	idem	
S_l	—	NO EXTR	$\lvert R_{cl}\rvert \to C_0^2 e^{\pi\gamma} c_{l\gamma}^{-1/2}$ for $p\to\infty \& p'\uparrow k$, $p\to 0 \& p'\downarrow k$, $p'\to\infty \& p\uparrow k$, or $p'\to 0 \& p\downarrow k$	c2	SUP?	$p\to 0 \& p'\to\infty$ or $p\to\infty \& p'\to 0$
A^r	i17	MIN	$(p-k)(p'-k) \gtrsim 0$?	i18	SUP	$q\to 0$ or $p\to\infty \& p'\to\infty$, any θ
A^a	i19	INF	$q\to 0$ or $p\to\infty \& p'\to\infty$, any θ	i20	SUP	$(p-k)(p'-k)\downarrow 0$: $\theta \geqslant \epsilon > 0$
A	—	$\gamma > 0$: NO EXTR	$\lvert R_c\rvert \to C_0^2$ for $(p-k)(p'-k)\downarrow 0$: $\theta \geqslant \epsilon > 0$			
A_l^r	i21	MIN INF:	$(p-k)(p'-k) \gtrsim 0$; $p\downarrow k \,\&\, p'\downarrow k$ or $p\uparrow k \,\&\, p'\uparrow k$ if successive limits	i22	SUP	$w\to 1$: $(p-k)(p'-k) \geqslant \epsilon > 0$; or $p\to 0 \& p'\to 0$ or $p\to\infty \& p'\to\infty$
A_l^a	i23	idem	idem	i24	SUP	$p\to\infty \& p'\downarrow k$, $p\to 0 \& p'\uparrow k$, $p'\to\infty \& p\downarrow k$, or $p'\to 0 \& p\uparrow k$
A_l	—	$\gamma > 0$: NO EXTR	$\lvert R_{cl}\rvert \to C_0^2 c_{l\gamma}^{-1/2}$ for $p\to\infty \& p'\downarrow k$, $p\to 0 \& p'\uparrow k$, $p'\to\infty \& p\downarrow k$, or $p'\to 0 \& p\uparrow k$	—	$\gamma < 0$: NO EXTR	$R_{cl}\to 1$ for $w\to 1$: $(p-k)(p'-k) \geqslant \epsilon > 0$; or $p\to 0 \& p'\to 0$ or $p\to\infty \& p'\to\infty$

one can combine (t_2) and (t_3) by

$$(t_E): p \to |E|/p \;\&\; p' \to |E|/p' \;(E \in \mathbb{R}').$$

One may call (t_0), (t_1), and (t_{01}) transpositions, and (t_{12}), (t_2), (t_3), and (t_E) inversions. Let D be the domain of the allowable values of p and p' when $k > 0$:

$$D := \{(p,p') \mid p > 0, p' > 0, p \neq k, p' \neq k\}; \tag{23.1}$$

note that the case $p = p'$ is excluded for R_{cl}; it applies to R_c nly if $\theta > 0$. It is convenient to divide D as follows:

$$\mathrm{D} = \mathrm{S} \cup \mathrm{A} \text{ (disjoint)}, \tag{23.2}$$

$$\mathrm{S} := \{(p,p') \mid p' < k < p \text{ or } p < k < p'\}, \tag{23.3}$$

$$\mathrm{A} := \{(p,p') \mid k < p \;\&\; k < p' \text{ or } p < k \;\&\; p' < k\}. \tag{23.4}$$

In S, the moduli of the Coulomb ratios, $|R_c|$ and $|R_{cl}|$, are even (symmetric) in γ for real γ; in A they are not even (asymmetric). Let us partition S and A into the disjoint subsets $\mathrm{S_i}$ and $\mathrm{A_i}$ according to

$$\begin{aligned}
&\mathrm{S} = \bigcup_{i=0}^{4} \mathrm{S_i}, && \mathrm{A} = \bigcup_{i=0}^{4} \mathrm{A_i},\\
&\mathrm{S_0}: pp' = k^2, && \mathrm{A_0}: p = p',\\
&\mathrm{S_1}: p' < k < p \;\&\; k^2 < pp', && \mathrm{A_1}: k < p' < p,\\
&\mathrm{S_2}: p' < k < p \;\&\; pp' < k^2, && \mathrm{A_2}: k < p < p',\\
&\mathrm{S_3}: p < k < p' \;\&\; pp' < k^2, && \mathrm{A_3}: p < p' < k,\\
&\mathrm{S_4}: p < k < p' \;\&\; k^2 < pp', && \mathrm{A_4}: p' < p < k,
\end{aligned} \tag{23.5}$$

see Fig. 3(a) and 3(b). These subsets or subregions $\mathrm{S_i}$ and $\mathrm{A_i}$ play a role in representations and inequalities for positive energy. Such a division for the negative-energy case does not seem to be useful. Let N and Z indicate negative and zero energy, respectively; then S, A, N, and Z cover all real-energy cases. Note that

$$\begin{aligned}
&vv' < 0 \quad \text{for S and N},\\
&vv' > 0 \quad \text{for A},
\end{aligned} \tag{23.6}$$

so that the sign of vv' is a useful indicator.

Each one of the regions S and A is invariant under the transposition (t_1) and also under the inversion (t_2). One easily verifies that the subregions are transformed as follows:

$$\begin{aligned}
(t_1): \quad & \mathrm{S_0} \to \mathrm{S_0}, && \mathrm{S_1} \leftrightarrow \mathrm{S_4}, && \mathrm{S_2} \leftrightarrow \mathrm{S_3},\\
& \mathrm{A_0} \to \mathrm{A_0}, && \mathrm{A_1} \leftrightarrow \mathrm{A_2}, && \mathrm{A_3} \leftrightarrow \mathrm{A_4},
\end{aligned} \tag{23.7}$$

$$\begin{aligned}
(t_2): \quad & \mathrm{S_0} \to \mathrm{S_0}, && \mathrm{S_1} \leftrightarrow \mathrm{S_3}, && \mathrm{S_2} \leftrightarrow \mathrm{S_4},\\
& \mathrm{A_0} \to \mathrm{A_0}, && \mathrm{A_1} \leftrightarrow \mathrm{A_3}, && \mathrm{A_2} \leftrightarrow \mathrm{A_4}.
\end{aligned} \tag{23.8}$$

Clearly we may often restrict (p,p') to one appropriate subregion $\mathrm{S_i}$ or $\mathrm{A_i}$ $(i = 1,2,3,4)$; the other subregions of S or A are obtained by applying (t_1) and (t_2), under which R_c and R_{cl} are invariant. $\mathrm{S_0}$ is a borderline: the curve $pp' = k^2$, which may be considered as a limit case. Similarly $\mathrm{A_0}$ $(p = p')$ is a limit case for R_c if $\theta = 0$ and for R_{cl}; it may be considered as a limit case for R_c if $\theta > 0$.

24. TWENTY-FOUR OPTIMAL INEQUALITIES

In this section we shall present and discuss twenty-four inequalities, denoted by (i1)–(i24), for the Coulomb ratios R_c and R_{cl}; see Table I. We assume that the energy is either positive or negative: $k^2 \in \mathbb{R}'$, and that $\gamma \in \mathbb{R}'$.

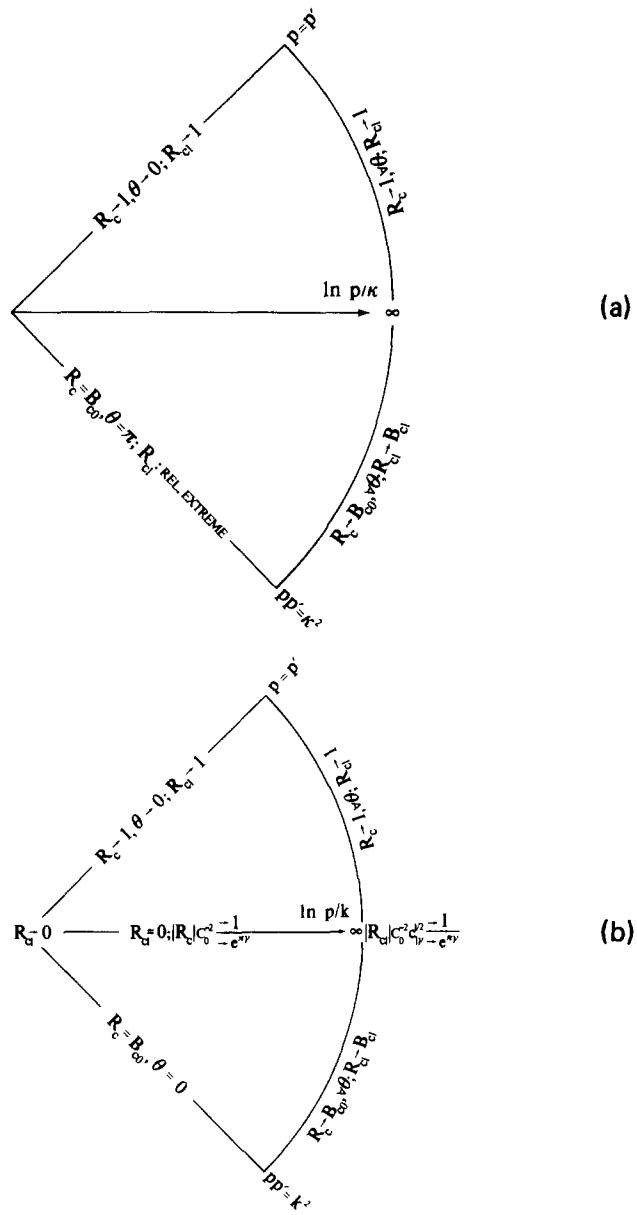


FIG. 4. Limits of R_c and R_{cl}: (a) for negative energy in the $(\ln p/\kappa, \ln p'/\kappa)$ plane; (b) for positive energy in the $(\ln p/k, \ln p'/k)$ plane. Only the quadrant $-\pi/4 < \arg z < \pi/4$ is displayed in both cases. Due to the invariance of R_c and of R_{cl} under the transformations (t_1), (t_2), and (t_3), these figures can be extended to the whole plane simply by rotation on the axis $p = p'$, and on the axes $pp' = \kappa^2$ and $pp' = k^2$, for (a) and (b), respectively. [Cf. Fig. 3(b) in Sec. 23 and Tables I and II in Sec. 24.] N. B. Fig. 4(b): On the $p' = k$ axis $R_{cl} \approx 0$ indicates that R_{cl} has an infinite number of zeros accumulating at $p' = k$. Further, $R_{cl} \to 0$ for $p \to k$ & $p' \to k$ if these two limits are evaluated in succession. In S the limit of R_{cl} for $p \to k$ & $p' \to k$ always yields 0. However, in A *any* value in [0,1] (and no other value) can be obtained by evaluating a particular combined limit, e.g., $R_{cl} \to 1$ if first $p' \to p$ and next $p \;\&\; p' \to k$. For $p - k = k^{1-\alpha}(p - p')^{\alpha}$ $(0 < \alpha < 1)$, $R_{cl} \to 1 - \alpha$ when $p' \uparrow p$.

We conjecture that (i10) and (i12) should each be replaced by (c1), and (i14) and (i16) by (c2). These inequalities (c1) and (c2) are optimal if valid; see below and cf. Sec. 17. All other inequalities are valid and optimal. In Sec. 18 we have

defined the functions Z_c^s and Z_c^a as infima of $|R_c|$; however, in Sec. 19 we have proved that these infima are in fact minima; consequently the associated inequalities: (i9), (i11), and (i17), are optimal. Optimality proofs of the inequalities follow by considering the appropriate limits of R_c and R_{cl} for which the boundary values indicated are attained or approached arbitrarily closely. These limits have been discussed in Secs. 21 and 22; the results are briefly summarized in Table II and Fig. 4(a) and 4(b).

Next we shall briefly discuss how (i1)–(i24) are proved. The negative-energy inequalities (i1)–(i8) have been proved in Ref. 24. For (i9), (i11), and (i17) the reader is referred to Secs. 18 and 19 and for (i13), (i15), (i21), and (i23) to Secs. 18 and 20. Further, (i19) has been proved in Ref. 24, and (i22) and (i24) in Sec. 6. We are going to prove the remaining inequalities: (i18), (i20), and (i10), (i12), (i14), (i16).

The proof of (i18) and (i20) is analogous to the one of (i22) and (i24) given in Sec. 6. By using Eqs. (4.1) and (4.2) of Ref. 24 we have

$$R_c = R_c^c + R_c^r, \tag{24.1}$$

$$R_c^c := \eta\, C_0^2\, y^{i\gamma}, \tag{24.2}$$

$$R_c^r := 1 - \eta\gamma^2 \sum_{n=0}^{\infty} \frac{\epsilon_n\, y^n}{(n^2+\gamma^2)}$$

$$= 2\eta \sum_{n=1}^{\infty} \frac{y^n n^2}{(n^2+\gamma^2)}, \tag{24.3}$$

$$\eta := (1-y)/(1+y) = -x^{-1}. \tag{24.4}$$

In the asymmetric case we have $0<y<1$; it can be seen upon inspection that

$$0<\eta<1, \quad 0<R_c^r<1. \tag{24.5}$$

Hence, for $\gamma\in\mathbb{R}'$,

$$|R_c| \leqslant |R_c^c| + R_c^r$$

$$< \eta C_0^2 + 1 - \eta - 2\eta\gamma^2 \sum_{n=1}^{\infty} \frac{y^n}{(n^2+\gamma^2)}$$

$$< 1 + \eta(C_0^2 - 1), \tag{24.6}$$

which may be compared with Eq. (6.12). Since

$$C_0^2 \lesseqgtr 1 \text{ for } \gamma \gtreqless 0, \tag{24.7}$$

we have clearly

$$|R_c| < 1, \quad \gamma > 0,$$

$$|R_c| < C_0^2, \quad \gamma < 0, \tag{24.8}$$

which are just (i18) and (i20).

To prove (i10) and (i12) we use the representation

$$R_c = (\rho+1)\int_0^{\infty} (\rho+\cosh t)^{-2} \sinh t\, e^{-i\gamma t}\, dt,$$

$$\operatorname{Re} i\gamma > -1, \tag{24.9}$$

which follows from Eq. (2.1) after integration by parts. Since $\rho\geqslant 1$ in the symmetric case it follows easily that $|R_c|<1$. More generally, if

$$R_c = \int_0^{\infty} g(t) e^{-i\gamma t}\, dt,$$

with $\gamma\in\mathbb{R}'$, $g(t)\geqslant 0$, and g not almost everywhere zero, then

$$|R_c| < \int_0^{\infty} g(t)\, dt. \tag{24.10}$$

In particular, the $<$ sign holds here according to the theorem of Cauchy–Schwarz,

$$g_1(t) := (g(t))^{1/2} \exp(i\gamma t/2), \quad g_2(t) := g_1^*(t),$$

$$|(g_1,g_2)| < |(g_1,g_1)(g_2,g_2)|^{1/2} = (g_1,g_1), \tag{24.11}$$

since g_1 and g_2 are not proportional when $\gamma\neq 0$. [With these formulas we have proved that $Q_\nu(z)$ and $\ln Q_\nu(z)$ are, for $\nu > -1$ and $z>1$, decreasing and convex functions of ν and of z; see Appendix Q of Ref. 32; these results were used for the proof of (i5) and (i8).]

Finally, (i14) and (i16) follow in the same way from the representation

$$R_{cl} Q_l(w) = vv' \int_0^{\infty} e^{-i\gamma t} \sinh t$$

$$\times Q_l'(uu' - vv' \cosh t)\, dt,$$

$$vv' < 0, \ \operatorname{Re} i\gamma > -l-1, \tag{24.12}$$

which follows from Eq. (3.1) after integration by parts; note that $vv'<0$ in the symmetric case and that $uu' - vv' \equiv w$.

[1] C. A. Coulomb, Mém. Acad. R. Sci. Inst. Fr. 1785 (Paris, 1788), 569–577, and 578–611, respectively.
[2] E. Rutherford, Philos. Mag. **21**, 669 (1911).
[3] W. Pauli, Z. Phys. **36**, 336 (1926).
[4] G. Wentzel, Z. Phys. **40**, 590 (1926).
[5] J. Meixner, Math. Z. **36**, 677 (1933).
[6] V. Fock, Z. Phys. **98**, 145 (1935).
[7] V. Bargmann, Z. Phys. **99**, 576 (1936).
[8] L. Hostler, J. Math. Phys. **5**, 591, 1235 (1964).
[9] J. Schwinger, J. Math. Phys. **5**, 1606 (1964).
[10] M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330, 346 (1966).
[11] A. M. Perelomov and V. S. Popov, Zh. Eksp. Teor. Fiz. **50**, 179 (1966) [Sov. Phys. JETP **23**, 118 (1966)].
[12] L. Hostler, J. Math. Phys. **8**, 642 (1967).
[13] J. Nuttall and R. W. Stagat, Phys. Rev. A **3**, 1355 (1971).
[14] E. Prugovečki and J. Zorbas, J. Math. Phys. **14**, 1398 (1973).
[15] S. N. Banerjee, Nuovo Cimento A **25**, 199 (1974).
[16] A. C. Chen, Am. J. Phys. **47**, 1073 (1979).
[17] J. C. Y. Chen and A. C. Chen, in *Advances in Atomic and Molecular Physics*, edited by D. R. Bates and I. Esterman (Academic, New York, 1972), Vol. 8, pp. 71–129.
[18] H. van Haeringen, "The Coulomb Potential in Quantum Mechanics and Related Topics," Ph. D. thesis, Free University, Amsterdam, 1978, 272 p.
[19] C. Chandler, Nucl. Phys. A **353**, 129c (1981).
[20] L. P. Kok, Nucl. Phys. A **353**, 171c (1981).
[21] H. van Haeringen, Nucl. Phys. A **327**, 77 (1979).
[22] L. P. Kok and H. van Haeringen, Phys. Rev. C **21**, 512 (1980).
[23] H. van Haeringen and L. P. Kok, Delft Prog. Rep. **7**, 3 (1982).
[24] H. van Haeringen and L. P. Kok, "New Inequalities for the Coulomb T matrix in momentum space," J. Math. Phys. (to be published).
[25] L. P. Kok and H. van Haeringen, Phys. Rev. Lett. **46**, 1257 (1981).
[26] H. van Haeringen, Phys. Lett. A **82**, 359 (1981).
[27] H. van Haeringen, J. Math. Phys. **24**, 1054, 1152, 1157, 1267, and 1274 (1983).
[28] H. van Haeringen, J. Math. Phys. **24**, 2467 (1983).
[29] L. P. Kok, J. W. de Maag, H. H. Brouwer, and H. van Haeringen, Phys. Rev. C **26**, 2381 (1982).
[30] J. W. de Maag, L. P. Kok, and H. van Haeringen, J. Math. Phys. **25**, 684 (1984).
[31] H. van Haeringen, Delft Prog. Rep. **8**, 110 (1983); Report 82 10 (Delft, 1982).

[32]H. van Haeringen, "Coulomb T matrix: Representations, Inequalities, Limits, and Zeros," Report 82 07 (Delft, 1982); 2nd ed.: Report 83 07 (Delft, 1983).
[33]H. van Haeringen, "Interactions for charged and neutral particles: Theory and Formulas," Report 82 12 (Delft, 1982); 2nd ed. Report 83 12 (Delft, 1983).
[34]I. H. Sloan, Phys. Rev. A **7**, 1016 (1973).
[35]A. C. Chen and J. C. Y. Chen, J. Phys. B **4**, L 102 (1971).
[36]J. C. Y. Chen and T. Ishihara, J. Phys. B **2**, 12 (1969).
[37]J. C. Y. Chen and T. Ishihara, Phys. Rev. **186**, 25 (1969).
[38]M. Rotenberg, Ann. Phys. (N. Y.) **19**, 262 (1962).
[39]S. Weinberg, Phys. Rev. **131**, 440 (1963).
[40]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966).
[41]A. Erdélyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vols. I and II; *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vols. I and II.
[42]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, corrected and enlarged edition (Academic, New York, 1980).
[43]E. R. Hansen, *A Table of Series and Products* (Prentice-Hall, Englewood Cliffs, NJ, 1975).
[44]H. van Haeringen, Delft Prog. Rep. **6**, 236 (1981); J. Math. Phys. **23**, 964 (1982); **24**, 1054 (1983).
[45]H. van Haeringen, "Inequalities for T matrices associated with Coulomb-plus-short-range potentials," in preparation.

New inequalities for the Coulomb T matrix in momentum space

H. van Haeringen
Department of Mathematics and Informatics, Delft University of Technology, Delft, The Netherlands

L. P. Kok
Institute for Theoretical Physics, P. O. Box 800, Groningen, The Netherlands

(Received 8 November 1983; accepted for publication 9 March 1984)

We present and prove new inequalities for the momentum representation (i) of the full Coulomb transition operator T_c, and (ii) of its partial-wave projections, T_{cl}, for all $l = 0, 1, \ldots$. Also, a previously conjectured inequality is proved: $|T_c/V_c| > 1$, under certain conditions. These inequalities are useful for gaining insight into the—very convenient, and often proposed in the context of three-particle calculations—approximation in which the Coulomb transition operator is replaced by the Coulomb potential operator. Such an approximation is obviously not accurate at any zero of the T matrix. We investigate the zeros of $\langle \mathbf{p}|T_c|\mathbf{p}'\rangle$ for positive energy, and we give simple approximation formulas for these zeros.

PACS numbers: 03.65.Nk, 02.30. + g, 11.20.Dj

I. INTRODUCTION

The two-particle off-shell Coulomb transition (T) matrix plays an important role in equations describing reactions of charged particles. The pure Coulomb T matrix is known in (complicated) closed form, in coordinate space as well as in momentum space. Because of the complexity of these closed forms, suitable and accurate approximations are sometimes desirable, especially for numerical calculations. An obvious and elementary approximation consists in replacing the Coulomb T operator T_c by the Coulomb potential operator V_c at some stage in the formalism. The calculational advantages of this approximation are considerable since V_c has a very simple closed form.

To gain some insight in the accuracy of this approximation one can study the ratio R_c defined by

$$R_c := \langle \mathbf{p}|T_c|\mathbf{p}'\rangle/\langle \mathbf{p}|V_c|\mathbf{p}'\rangle, \tag{1.1}$$

and its partial-wave-projection analog, defined by

$$R_{cl} := \langle p|T_{cl}|p'\rangle/\langle p|V_{cl}|p'\rangle, \tag{1.2}$$

where T_{cl} and V_{cl} are the partial-wave projected Coulomb T operator and potential operator, respectively. Here and henceforth, the notation of Refs. 1 and 2 is used. In particular, γ is Sommerfeld's parameter, the Coulomb potential is $V_{cl}(r) = 2k\gamma/r \equiv -2s/r$, $\forall l = 0,1,2,\ldots$; θ is the angle between the momenta $\mathbf{p}$ and $\mathbf{p}'$,

$$w := (p^2 + p'^2)/2pp', \quad v := (p^2 - k^2)/2pk,$$
$$v' := (p'^2 - k^2)/2p'k.$$

The energy is $k^2 \equiv -\kappa^2$ with $k > 0$ for positive energy and $\kappa > 0$ for negative energy; Q_l is Legendre's function of the second kind.

In Refs. 3 and 4 more inequalities and many new representations for R_c and for R_{cl} are derived. We point out that the least upper bound of $|R_{cl}|$ that is conjectured in Ref. 1, Eq. (5.5′) is derived in Ref. 3.

$$1 \leqslant |R_c|, \ p > k \ \&p' > k \ \text{ or } \ p < k \ \&p' < k;$$

$$k^2 > 0, \ s > 0 (\gamma < 0), \tag{1.3}$$

has been found numerically to hold for many values of p, p', and θ. Its validity could not be proved analytically but it was conjectured; see Eq. (4.4) of Ref. 1.

In this paper we shall extend the results of Ref. 1 on inequalities for R_c and R_{cl}. In particular, we shall give two different proofs of a sharpened version of Eq. (1.3).

In Secs. II and III we shall derive new *optimal* inequalities for R_c and for R_{cl}, respectively, valid for negative energy. In Sec. IV we prove Eq. (1.3), and in Sec. V we give a second proof of this remarkable inequality. In Sec. VI we shall discuss zeros of R_c and of R_{cl} for positive energy. It is clear that at such zeros the approximation in which the Coulomb T operator is replaced by the Coulomb potential, which amounts to replacing R_c or R_{cl} by 1, is not satisfactory.

In Refs. 3 and 4 more inequalities and many new representations for R_c and for R_{cl} are derived. We point out that the least upper bound of $|R_{cl}|$ that is conjectured in Ref. 1, Eq. (5.5′), is derived in Ref. 3.

II. INEQUALITIES FOR R_c VALID FOR NEGATIVE ENERGY

In this section we shall prove four inequalities for R_c, to be denoted by (i1)–(i4) [see Eqs. (2.3) and (2.4)], which are valid for fixed negative energy: $k^2 < 0$, $-ik \equiv \kappa > 0$. Then $i\gamma \equiv -s/\kappa$ is positive for repulsion ($s < 0$), and negative for attraction ($s > 0$). First we define a function B_{c0}, which may be considered as a boundary function for the Coulomb ratio R_c, by

$$B_{c0}(z) := 4\int_0^1 t^z(1-t)(1+t)^{-3}\,dt, \ \operatorname{Re} z > -1. \tag{2.1}$$

It can be seen by inspection that the function B_{c0} is positive and strictly decreasing on $(-1, \infty)$. Furthermore,

$$B_{c0}(0) = 4\int_0^1 [2(1+t)^{-3} - (1+t)^{-2}]dt = 1,$$

$$\lim_{z\to\infty} B_{c0}(z) = 0, \quad \lim_{z\downarrow -1} B_{c0}(z) = \infty. \tag{2.2}$$

We shall prove

 0022-2488/84/103033-06$02.50

$$B_{c0}(-s/\kappa) \overset{(i1)}{\leqslant} R_c \overset{(i2)}{<} 1, \quad s<0, \tag{2.3}$$

$$1 \overset{(i3)}{<} R_c \overset{(i4)}{\leqslant} B_{c0}(-s/\kappa), \quad 0<s<\kappa. \tag{2.4}$$

These inequalities are *optimal*. For the proof we use the following representation[2-4] for R_c:

$$R_c = 1 + (s/\kappa)I,$$

$$I := \int_0^1 \frac{(w-\cos\theta)t^{-1-s/\kappa}\,dt}{w-\cos\theta - vv'(t-2+1/t)/2}, \quad w>\cos\theta, \tag{2.5}$$

where

$$w := (p^2+p'^2)/(2pp'),$$

$$vv' := -(p^2+\kappa^2)(p'^2+\kappa^2)/(4\kappa^2 pp') < 0.$$

Since $w>\cos\theta$ and $vv'<0$ it is clear that $I>0$. Hence,

$$R_c > 1 \quad \text{if } s>0, \quad R_c < 1 \quad \text{if } s<0. \tag{2.6}$$

To prove that (i2) and (i3) are *optimal*, we shall prove that $I\to 0$, hence, $R_c\to 1$, for $w-\cos\theta\to 0$, i.e., for $\cos\theta = 1$, $p-p'\to 0$. It is clear that the integrand in (2.5) is *minimal* with respect to θ for $\theta = 0$. Inserting $\theta = 0$, and $w-1 = (p-p')^2/(2pp')$, we get

$$I(\theta=0) = \int_0^1 t^{-s/\kappa}[t+\alpha^2(1-t)^2]^{-1}\,dt,$$

$$\alpha^2 := (p^2+\kappa^2)(p'^2+\kappa^2)/[4\kappa^2(p-p')^2]. \tag{2.7}$$

So we have to prove that $I(\theta=0)\to 0$ for $\alpha\to\infty$. It is easily seen that the region $t\approx 1$ (only) deserves some care. In fact it is sufficient to prove that

$$\lim_{\alpha\to\infty}\int_{1-c}^1 t^{-s/\kappa}[t+\alpha^2(1-t)^2]^{-1}\,dt = 0$$

for any $c\in(0,1)$. As c may be arbitrarily close to 1 it is allowable to replace $t^{-s/\kappa}$ and the isolated term t by 1. Then we have

$$\lim_{\alpha\to\infty}\int_{1-c}^1 [1+\alpha^2(1-t)^2]^{-1}\,dt$$

$$= \lim_{\alpha\to\infty}\int_0^c \frac{dt}{(1+\alpha^2 t^2)} = \lim_{\alpha\to\infty}\alpha^{-1}\arctan(\alpha c) = 0.$$

Hence,

$$\lim_{p-p'\to 0} R_c = 1. \tag{2.8}$$

This completes the proof of the inequalities (i2) and (i3).

To prove (i1) and (i4) we return to Eq.(2.5). Since $vv'<0$ and $t-2+1/t = (1-t)^2/t \geqslant 0$ it is clear that the integrand is *maximal* with respect to θ for $\theta = \pi$. Inserting $\cos\theta = -1$ and $w+1 = (p+p')^2/(2pp')$ into (2.5) we obtain

$$I(\theta=\pi) = \int_0^1 t^{-1-s/\kappa}[1+A(t-2+1/t)]^{-1}\,dt,$$

$$A := (p^2+\kappa^2)(p'^2+\kappa^2)/[4\kappa^2(p+p')^2]. \tag{2.9}$$

Furthermore, $\partial A/\partial p = 0$ for $p = \kappa^2/p'$; then $A = \frac{1}{4}$ which is its minimal value under variation of p and p'. Substitution of $A = \frac{1}{4}$ into (2.9) gives the *maximal* value of I:

$$I^{\max} = \int_0^1 4t^{-s/\kappa}(1+t)^{-2}\,dt. \tag{2.10}$$

Integration by parts finally yields

$$1+\left(\frac{s}{\kappa}\right)I^{\max} = 4\int_0^1 t^{-s/\kappa}(1-t)(1+t)^{-3}\,dt$$

$$= B_{c0}\left(\frac{-s}{\kappa}\right), \tag{2.11}$$

which completes the proof of (i1) and (i4).

III. INEQUALITIES FOR R_{cl} VALID FOR NEGATIVE ENERGY

In this section we shall prove four *optimal* inequalities for R_{cl}, to be denoted by (i5)–(i8) [see Eqs. (3.5) and (3.6)], valid for fixed negative energy. The plan of the proof is similar to that of the preceding section. First we define a boundary function B_{cl}, $l = 0,1,...$, for the Coulomb ratio R_{cl}, by

$$B_{cl}(z) := (l+1)4^{l+1}\int_0^1 t^{z+l}(1-t)(1+t)^{-2l-3}\,dt,$$

$$\operatorname{Re} z > -l-1. \tag{3.1}$$

Integration by parts gives

$$B_{cl}(z) = 1 - 4^{l+1}z\int_0^1 t^{z+l}(1+t)^{-2l-2}\,dt. \tag{3.2}$$

It is obvious that the function B_{cl} is positive and strictly decreasing on $(-l-1,\infty)$. Furthermore,

$$B_{cl}(0) = 1, \quad \forall l \text{ [from (3.2)]}, \tag{3.3}$$

$$\lim_{z\to\infty} B_{cl}(z) = 0, \quad \lim_{z\downarrow -l-1} B_{cl}(z) = \infty, \quad \forall l. \tag{3.4}$$

We shall prove

$$B_{cl}(-s/\kappa) \overset{(i5)}{<} R_{cl} \overset{(i6)}{<} 1, \quad s<0, \ l = 0,1,..., \tag{3.5}$$

$$1 \overset{(i7)}{<} R_{cl} \overset{(i8)}{<} B_{cl}(-s/\kappa), \quad 0<s/\kappa<l+1, \ l = 0,1,... . \tag{3.6}$$

These inequalities are optimal. For the proof we use the following representation[2-4] for R_{cl},

$$R_{cl} = 1 + (s/\kappa)I_l, \tag{3.7}$$

$$I_l := [Q_l(w)]^{-1}\int_0^1 t^{-1-s/\kappa} \times Q_l(w - vv'(t-2+1/t)/2)dt, \tag{3.8}$$

$$s/\kappa < l+1, \ l = 0,1,... .$$

Let us recall some well-known facts about the Legendre function Q_l. We have

$$Q_l(z) = -\tfrac{1}{2}\ln(z-1) + O(1), \quad z\downarrow 1, \tag{3.9}$$

$$\lim_{z\downarrow 1} Q_l(z) = \infty;$$

$$\lim_{z\to\infty} z^{l+1}Q_l(z) = \frac{2^l(l!)^2}{(2l+1)!} = \frac{l!}{(2l+1)!!}, \tag{3.10}$$

$Q_l(z)$ is positive and strictly decreasing for $1<z<\infty$.

Since $w>1$ and $vv'<0$ it is clear that $I_l>0$. Hence,

$$R_{cl} > 1 \text{ if } s>0, \quad R_{cl} < 1 \text{ if } s<0. \tag{3.11}$$

To prove that (i6) and (i7) are *optimal*, we observe that $w\downarrow 1$

and $vv' \to -(p^2+\kappa^2)^2/(4\kappa^2 p^2) < 0$ when $p - p' \to 0$. Thus we obtain

$$\lim_{w \downarrow 1} I_l = \lim_{w \downarrow 1} [Q_l(w)]^{-1} \times \int_c^1 t^{-1-s/\kappa} Q_l(w - vv'(t-2+1/t)/2)dt, \quad \text{any } c \in (0,1)$$

$$= \lim_{w \downarrow 1} [\ln(w-1)]^{-1} \times \int_c^1 t^{-1-s/\kappa} \ln\left[\frac{-vv'(t-2+1/t)}{2}\right] dt = 0. \tag{3.12}$$

Hence,

$$\lim_{p-p' \to 0} R_{cl} = 1. \tag{3.13}$$

This completes the proof of the inequalities (i6) and (i7).

To prove (i5) and (i8) we return to Eq. (3.8) and we are going to derive the least upper bound, or supremum (sup), of I_l under variation of p and p'. We first keep w fixed; we put $w = \cosh\alpha$, α fixed, and we have $p' = pe^{\alpha}$. Under variation of p(or p'), the argument of Q_l in Eq. (3.8) is minimal for $p^2 = \kappa^2 e^{-\alpha}$, i.e., for $pp' = \kappa^2$. This minimal value is $b + cw$, where we use for the moment the variables

$$b := (1-t)^2/4t \geqslant 0, \quad c := (1+t)^2/4t \geqslant 1. \tag{3.14}$$

Since $Q_l(z)$ is decreasing in z, we have the *maximum* of I_l under the condition that w be fixed,

$$I_l \leqslant I_l(pp' = \kappa^2) = [Q_l(w)]^{-1} \int_0^1 t^{-1-s/\kappa} Q_l(b+cw)dt. \tag{3.15}$$

The remaining freedom we have is to vary w. In Appendix R of Ref. 3 it is proved that q_l defined by[5]

$$q_l(w) := Q_l(b+cw)/Q_l(w), \quad b \geqslant 0, \quad c \geqslant 1, \quad w > 1, \tag{3.16}$$

is a strictly increasing function of w on $(1,\infty)$ for fixed b and c, except for $b = 0$, $c = 1$ when clearly $q_l(w) \equiv 1$. Therefore, we get the supremum of the expression on the right-hand side of (3.15) by taking its limit for $w \to \infty$. This limit is equal to the supremum of I_l which we denote by $I_l^{\sup}$. From Eq. (3.10) we have

$$\lim_{w \to \infty} q_l(w) = c^{-l-1},$$

hence, by (3.14),

$$I_l^{\sup} = 4^{l+1} \int_0^1 t^{l-s/\kappa}(1+t)^{-2l-2} dt, \tag{3.17}$$

and according to Eq. (3.2) we have

$$B_{cl}(-s/\kappa) = 1 + (s/\kappa) I_l^{\sup}. \tag{3.18}$$

Recalling Eq. (3.7) we see that this completes the proof of the inequalities (i5) and (i8) given in (3.5) and (3.6).

IV. FIRST PROOF OF $|R_c| > 1$ ($0 < y < 1$, $\gamma < 0$)

We shall use the series representations[2–4]

$$R_c = \left(\frac{1-y}{1+y}\right) C_0^2 y^{i\gamma} + 1 - \left(\frac{1-y}{1+y}\right) \gamma^2 \sum_{n=o}^{\infty} \frac{\epsilon_n y^n}{n^2+\gamma^2} \tag{4.1}$$

$$= \left(\frac{1-y}{1+y}\right)\left[C_0^2 y^{i\gamma} + 2\sum_{n=1}^{\infty} \frac{y^n n^2}{n^2+\gamma^2}\right], \tag{4.2}$$

where y is defined by

$$y := (x+1)/(x-1), \quad x < 0,$$

$$x^2 := 1 + k^{-2}|\mathbf{p} - \mathbf{p}'|^{-2}(p^2 - k^2)(p'^2 - k^2),$$

and

$$\epsilon_0 := 1, \quad \epsilon_n := 2, \quad n > 0,$$

$$C_0^2 := 2\pi\gamma/(e^{2\pi\gamma} - 1), \tag{4.3}$$

$$C_0^2 = -\pi\gamma + \pi\gamma \coth \pi\gamma = 1 - \pi\gamma + 2\gamma^2 \sum_{n=1}^{\infty} \frac{1}{n^2+\gamma^2}.$$

An essential feature of the proof in this section is to consider $R_c y^{-i\gamma}$ in place of R_c. It is obvious that

$$\mathrm{Re}(R_c y^{-i\gamma}) > 1, \quad \forall \gamma < 0, \ 0 < y < 1, \tag{4.4}$$

implies $|R_c| > 1$, $\forall \gamma < 0$, $0 < y < 1$. In fact we shall prove a stronger inequality, viz., (4.4) with C_0^2 replaced by $C_0^2 + \pi\gamma = \pi\gamma \coth \pi\gamma < C_0^2$, which amounts to

$$\mathrm{Re}(R_c y^{-i\gamma}) > 1 - \pi\gamma(1-y)/(1+y) > 1, \quad \forall \gamma < 0, \ 0 < y < 1. \tag{4.5}$$

This seems to be the sharpest modification, in a certain sense, of the inequality $|R_c| > 1$.

To begin with the proof of (4.5) now, we combine (4.1), (4.2), and (4.3) and use

$$\mathrm{Re}\, y^{-i\gamma} = \cos(\gamma \ln y) = 1 - 2\sin^2(\tfrac{1}{2}\gamma \ln y).$$

Then we obtain, denoting for convenience $\frac{1}{2}\gamma \ln y$ by α,

$$\mathrm{Re}(R_c\, y^{-i\gamma}) = \frac{C_0^2(1-y)}{(1+y)} + \cos(2\alpha)\left[1 - \left\{\frac{1-y}{1+y}\right\}\gamma^2 \sum_{n=0}^{\infty} \frac{\epsilon_n y^n}{n^2+\gamma^2}\right]$$

$$= 1 + \left\{\frac{1-y}{1+y}\right\}\left[C_0^2 - \gamma^2 \sum_{n=0}^{\infty} \frac{\epsilon_n y^n}{n^2+\gamma^2} - 4\sin^2\alpha \sum_{n=1}^{\infty} \frac{y^n n^2}{n^2+\gamma^2}\right]$$

$$= 1 + \left\{\frac{1-y}{1+y}\right\}\left[-\pi\gamma + \gamma^2 \sum_{n=1}^{\infty} \frac{y^n\{-2 + 2y^{-n} - 4n^2\gamma^{-2}\sin^2\alpha\}}{n^2+\gamma^2}\right]$$

$$= 1 + \left\{\frac{1-y}{1+y}\right\}\left[-\pi\gamma + \gamma^2 \sum_{n=1}^{\infty} \frac{y^n f_n(y)}{n^2+\gamma^2}\right], \tag{4.6}$$

where

$$f_n(y) := -2 + 2y^{-n} - 4n^2\gamma^{-2}\sin^2\alpha, \quad n = 1,2,3,\ldots.$$

Since $\sin^2\alpha < \alpha^2$, $\alpha \neq 0$, we have [recall $\alpha := (\gamma/2)\ln y$]

$$f_n(y) > g_n(y) := -2 + 2y^{-n} - n^2 \ln^2 y.$$

Now we introduce the function h_n,

$$h_n(y) := -(y/2n)g_n'(y) = y^{-n} + n\ln y, \quad n = 1,2,3,\ldots.$$

Then $h_n(y) \geqslant 1$ for $0 < y \leqslant 1$ since $h_n(1) = 1$ and

$$h_n'(y) = n(1-y^{-n})/y < 0, \quad 0 < y < 1, \ n = 1,2,3,\ldots.$$

Hence $g_n'(y) \leqslant -2n$ so that $g_n(y)$ is decreasing, and so we have

$$f_n(y) > g_n(y) > g_n(1) = 0, \quad n = 1,2,3,..., \ 0 < y < 1.$$

By inserting $f_n(y) > 0$ into (4.6) we obtain

$$\mathrm{Re}(R_c y^{-i\gamma}) > 1 - \pi\gamma(1-y)/(1+y) > 1,$$
$$\forall \gamma < 0, \ 0 < y < 1,$$

which completes the proof of (4.5) and hence the proof of $|R_c| > 1$.

V. SECOND PROOF OF $|R_c| > 1$ ($0 < y < 1$, $\gamma < 0$)

Here we use again Eqs. (4.1)–(4.3). First we observe that

$$|R_c| \geqslant \left\{\frac{1-y}{1+y}\right\}\left[C_0^2 - 2\sum_{n=1}^{\infty}\frac{y^n n^2}{n^2+\gamma^2}\right]$$
$$> \left\{\frac{1-y}{1+y}\right\}\left[C_0^2 - 2\sum_{n=1}^{\infty} y^n\right]$$
$$= \left\{\frac{1-y}{1+y}\right\}\left[C_0^2 - \frac{2y}{1-y}\right],$$

for $\gamma < 0$, $0 < y < 1$. Hence *if*

$$C_0^2 \geqslant (1+3y)/(1-y), \tag{5.1}$$

we have at once

$$|R_c| > (1+3y)/(1+y) - 2y/(1+y) = 1.$$

One easily verifies that (5.1) is equivalent with $0 < y \leqslant y_0$, where

$$y_0 := (C_0^2 - 1)/(C_0^2 + 3). \tag{5.2}$$

So we have to prove $|R_c| > 1$ only under the condition

$$y_0 < y < 1. \tag{5.3}$$

Since $|R_c| \geqslant \mathrm{Re}\, R_c$ it is *sufficient* to prove that [cf. (4.1)]

$$\frac{1-y}{1+y} C_0^2 \cos(\gamma \ln y) + 1 - \frac{1-y}{1+y}\gamma^2$$
$$\times \sum_{n=0}^{\infty}\frac{\epsilon_n y^n}{n^2+\gamma^2} > 1,$$

which is equivalent with

$$\cos(\gamma \ln y) > C_0^{-2}\gamma^2 \sum_{n=0}^{\infty}\frac{\epsilon_n y^n}{n^2+\gamma^2}. \tag{5.4}$$

Now we shall use

$$0 < \gamma \ln y < \gamma \ln y_0 < 2/\pi, \quad \forall \gamma < 0, \ y_0 < y < 1. \tag{5.5}$$

Only the last inequality here is difficult; it will be proved in Appendix A. From (5.5) it is clear that the infimum with respect to y of the left-hand side of (5.4) is $\cos(\gamma \ln y_0)$. Furthermore, the supremum with respect to y of the right-hand side of (5.4) is obtained by taking $y = 1$. Hence, for the proof of (5.4) it is sufficient to prove the inequality in

$$\cos(\gamma \ln y_0) > C_0^{-2}\gamma^2 \sum_{n=0}^{\infty}\frac{\epsilon_n}{n^2+\gamma^2}$$
$$= C_0^{-2}\pi\gamma \coth \pi\gamma = \frac{e^{2\pi\gamma}+1}{2}. \tag{5.6}$$

When

$$\gamma < \ln(-1 + 2\cos(2/\pi))/(2\pi) = -0.07913...,$$

the right-hand side of (5.6) is smaller than $\cos(2/\pi) = 0.8041...$ so that (5.6) holds in this case, by (5.5). Therefore, it is sufficient to restrict γ to $-0.07914 < \gamma < 0$. Now we shall use the inequalities $\cos z > 1 - z^2/2$ and

$$1 + 2\gamma > (e^{2\pi\gamma} + 1)/2, \quad -0.1 < \gamma < 0. \tag{5.7}$$

To prove (5.7) we define the function f by

$$f(\gamma) := 1 + 4\gamma - e^{2\pi\gamma}. \tag{5.8}$$

Since f is concave, $f(0) = 0$, and $f(-0.1) = 0.0665\cdots > 0$, it follows that $f(\gamma) > 0$ for $-0.1 < \gamma < 0$ which proves (5.7). Returning to (5.6) we see that it is sufficient to prove the question-marked inequalities in

$$\cos(\gamma \ln y_0) > 1 - \tfrac{1}{2}\gamma^2 \ln^2 y_0 \overset{?}{>} 1 + 2\gamma > (e^{2\pi\gamma}+1)/2$$
$$\Leftarrow \ln(1/y_0) = \ln[(C_0^2+3)/(C_0^2-1)] \overset{?}{<} 2(-\gamma)^{-1/2}. \tag{5.9}$$

Since $C_0^2 > 1 - \pi\gamma$ [cf. Eq. (4.3)] we have

$$\ln[(C_0^2+3)/(C_0^2-1)] < \ln(1 - 4/\pi\gamma).$$

Introducing $\alpha := (-\gamma)^{-1/2}$ and

$$g(\alpha) := 2\alpha - \ln(1 + 4\alpha^2/\pi),$$

we have to prove that

$$g(\alpha) > 0 \text{ for } \alpha > 3.555 \approx (0.07914)^{-1/2}.$$

One easily verifies that $g(0) = 0$, and

$$g'(\alpha) > 0 \Rightarrow g(\alpha) > 0, \quad \forall \alpha > 0,$$

which completes the proof.

VI. A NOTE ON ZEROS OF R_c

In this section we shall consider the zeros of R_c for positive energy. We shall give a simple and elegant proof of the fact that R_c possesses at least one zero, for a certain value of γ. Elsewhere[3] it has been proved that R_c and R_{cl} possess an infinite number of zeros which correspond to values of γ accumulating at $|\gamma| = \infty$.

We distinguish two cases: (i) $-1 < y < 0$, and (ii) $0 < y < 1$.

(i) When $-1 < y < 0$ we have $-1 < x < 0$ and $vv' < 0$. In this case $|R_c|$ is even in γ; therefore, we shall call this the "symmetric" case. Clearly we may take $\gamma > 0$ without loss of generality. We shall use the following representation for R_c, valid for $-1 < y < 0$,

$$R_c = \left\{\frac{1-y}{1+y}\right\}\left\{\frac{(-y)^{i\gamma}\pi\gamma}{\sinh \pi\gamma} + 2\sum_{n=1}^{\infty}\frac{y^n n^2}{n^2+\gamma^2}\right\}, \tag{6.1}$$

which is the analog of Eq. (4.2) (valid for $0 < y < 1$). We are looking for solutions of the equation $R_c = 0$. It is clear that we must have $(-y)^{i\gamma} = \pm 1$. In Ref. 3 it has been proved that $(-y)^{i\gamma} = -1$ gives no solution at all. So we shall put $(-y)^{i\gamma} = 1$. We define a sequence of values for y,

$$y_n := -\exp(-2\pi n/\gamma), \quad n = 1,2,...(\gamma > 0). \tag{6.2}$$

Restricting ourselves to $n = 1$, we define a function h by

$$h(\gamma) := R_c(y = y_1). \tag{6.3}$$

Then h is a *continuous* function of γ. One easily obtains

$$h(1) \approx 0.27, \quad h(3) \approx -0.02. \tag{6.4}$$

Hence, h possesses a zero, which we denote by ${}^{s}\gamma_1$, where the superscript s stands for "symmetric." Therefore,

$$R_c = 0 \quad \text{when} \quad \gamma = {}^{\circ}\gamma_1^s \quad \text{and} \quad y = -\exp(-2\pi/{}^{\circ}\gamma_1^s). \tag{6.5}$$

Computer calculations suggest that for each n just one and only one zero can be found. In Ref. 3 this has been proved with the help of a new representation for R_c. An interesting approximation formula for the zeros ${}^{\circ}\gamma_n^s$ can be derived, viz.

$${}^{\circ}\gamma_n^s \approx (2n)^{1/2} + (4\pi)^{-1}\ln(8\pi^2 n^3) + o(1), \quad n\to\infty. \tag{6.6}$$

It follows from Eq. (6.1) by taking $y \approx 0$ and solving the resulting equation,

$$\exp(\pi\gamma - 2\pi n/\gamma) = \pi\gamma^3. \tag{6.7}$$

In Table I, we give exact (from computer calculation) and approximate [from Eq.(6.6)] values for the first four zeros ${}^{\circ}\gamma_n^s$:

$$R_c = 0 \quad \text{when} \quad \gamma = {}^{\circ}\gamma_n^s \quad \text{and} \quad y = -\exp(-2\pi n/{}^{\circ}\gamma_n^s), \quad n = 1,2,\ldots. \tag{6.8}$$

It follows that the accuracy of the approximation formula (6.6) is good: For $n = 4$ the approximate value is only 8% below the exact value.

(ii) When $0 < y < 1$ we have $x < -1$ and $vv' > 0$. In this case $|R_c|$ is not even in γ; we shall call this the "asymmetric" case. When $\gamma < 0$ we have $|R_c| > 1$ (see Secs. IV and V) so that R_c has no zero. Therefore, we take γ positive.

From Eqs. (4.1) and (4.2), which are valid in this case, it follows that $R_c = 0$ implies that $y^{i\gamma} = \pm 1$. From Eq. (4.2) it can be seen by inspection that R_c is positive when $y^{i\gamma} = +1$. So we must have $y^{i\gamma} = -1$ and we define

$$y_n := \exp(-(2n-1)\pi/\gamma), \quad n = 1,2,\ldots(\gamma > 0). \tag{6.9}$$

Restricting ourselves to $n = 1$, we define a function f by

$$f(\gamma) := R_c(y = y_1). \tag{6.10}$$

Then f is continuous, and we have

$$f(0) := \lim_{\gamma\downarrow 0} f(\gamma) = -1, \qquad f(1) \approx 0.03. \tag{6.11}$$

Hence, f possesses a zero, which we denote by ${}^{\circ}\gamma_1^a$, where the superscript a stands for "asymmetric." Hence,

$$R_c = 0 \quad \text{when} \quad \gamma = {}^{\circ}\gamma_1^a \quad \text{and} \quad y = \exp(-\pi/{}^{\circ}\gamma_1^a). \tag{6.12}$$

By using Eq.(4.2) we see that the zeros of R_c follow from

$$\frac{\pi\gamma}{(e^{2\pi\gamma}-1)} = \sum_{m=1}^{\infty} \frac{y^m m^2}{m^2+\gamma^2}, \quad y = \exp\left(\frac{-(2n-1)\pi}{\gamma}\right). \tag{6.13}$$

Computer calculations suggest that for each n just one solution exists. (For a proof, see Ref. 3.) By taking $y \approx 0$ in (6.13) we get an approximate equation,

$$\exp[2\pi\gamma - (2n-1)\pi/\gamma] \approx \pi\gamma^3, \tag{6.14}$$

from which we derive approximate values for the zeros of R_c in the "asymmetric" case,

$${}^{\circ}\gamma_n^a \approx (n - \tfrac{1}{2})^{1/2} + (8\pi)^{-1}\ln(\pi^2 n^3) + o(1), \quad n\to\infty. \tag{6.15}$$

In Table II, we give exact (from computer calculation) and approximate [from Eq. (6.15)] values for the first five zeros ${}^{\circ}\gamma_n^a$:

$$R_c = 0 \quad \text{when} \quad \gamma = {}^{\circ}\gamma_n^a \quad \text{and} \quad y = \exp[-(2n-1)\pi/{}^{\circ}\gamma_n^a], \quad n = 1,2,\ldots. \tag{6.16}$$

Clearly Eq. (6.15) gives excellent values, lying less than 5% below the exact values.

TABLE I. The first four zeros ${}^{\circ}\gamma_n^s$ of R_c in the "symmetric" case.

${}^{\circ}\gamma_n^s$	"Exact"	Approximate	Quotient
$n = 1$	2.132	1.762	1.210
$n = 2$	2.821	2.514	1.122
$n = 3$	3.343	3.060	1.093
$n = 4$	3.775	3.507	1.077

TABLE II. The first five zeros ${}^{\circ}\gamma_n^a$ of R_c in the "asymmetric" case.

${}^{\circ}\gamma_n^a$	"Exact"	Approximate	Quotient
$n = 1$	0.833	0.798	1.044
$n = 2$	1.454	1.399	1.039
$n = 3$	1.862	1.803	1.033
$n = 4$	2.187	2.127	1.028
$n = 5$	2.464	2.405	1.025

APPENDIX: $\gamma \ln[(C_0^2 - 1)/(C_0^2 + 3)] < 2/\pi$

In this appendix we shall prove that

$$0 < \gamma \ln y_0 < 2/\pi, \quad \forall\gamma < 0, \tag{A1}$$

where y_0 is defined by

$$y_0 := (C_0^2 - 1)/(C_0^2 + 3),$$
$$C_0^2 := 2\pi\gamma/(e^{2\pi\gamma} - 1).$$

Defining the function f by

$$f(z) := z\ln[(z + 3 - 3e^{-z})/(z - 1 + e^{-z})], \quad z > 0, \tag{A2}$$

where $z = -2\pi\gamma$, we easily see that (A1) is equivalent with

$$0 < f(z) < 4 \quad \forall z > 0. \tag{A3}$$

First we note that (A3) is optimal, because of

$$\lim_{z\downarrow 0} f(z) = 0, \quad \lim_{z\to\infty} f(z) = 4.$$

It is clear that $f(z) > 0$ and that

$$f(z) < z\ln[(z+3)/(z-1)], \quad \forall z > 1.$$

We shall use now

$$g(z) := \ln[(z+3)/(z-1)] - 4/z < 0, \quad \forall z \geqslant \tfrac{3}{2}. \tag{A4}$$

The proof of $g(z) < 0$ follows from: $g(3/2) = \ln 9 - \frac{8}{3}$, $\lim_{z\to\infty} g(z) = 0$, and

$$g'(z) = 4(2z-3)/[z^2(z-1)(z+3)] > 0, \quad \forall z > \tfrac{3}{2}.$$

From (A4) we have at once

$$0 < f(z) < 4, \quad \forall z \geqslant \tfrac{3}{2}. \tag{A5}$$

To prove the remaining part of (A3): $f(z) < 4$ for $0 < z < \frac{3}{2}$ we use the elementary fact that

$$h(z) := e^{-z} - 1 + z - z^2/2 + z^3/6 > 0, \quad \forall z > 0.$$

The proof follows by observing successively that $h'''(z)$, $h''(z)$, $h'(z)$, and $h(z)$ are positive $\forall z > 0$. Using

$$z - 1 + e^{-z} > z^2/2 - z^3/6, \quad \forall z > 0$$

in the denominator, and $1 - e^{-z} < z, \forall z > 0$ in the numerator of the argument of ln in (A2), we easily obtain

$$f(z) < z \ln\{24/(3z - z^2)\}, \quad 0 < z < 3. \tag{A6}$$

Denoting the right member of (A6) by $f_1(z)$, we now claim that

$$0 < f_1(z) \leqslant c := f_1(\tfrac{3}{2}) = (\tfrac{3}{2})\ln(\tfrac{32}{3}) = 3.5506..., \quad 0 < z \leqslant \tfrac{3}{2}. \tag{A7}$$

To prove this we define the function f_2,

$$f_2(z) := [f_1(z) - f_1(\tfrac{3}{2})]/z = \ln[24/(3z - z^2)] - c/z, \quad 0 < z \leqslant \tfrac{3}{2}.$$

Then

$$f_2'(z) = [2z^2 - (3 + c)z + 3c]/(z^2(3 - z)) > 0, \quad 0 < z \leqslant \tfrac{3}{2}$$

(note: negative discriminant), hence $f_2(z)$ increases from $-\infty$ to 0 when z increases from 0 to $\frac{3}{2}$. Combining (A6) and (A7) we have

$$0 < f(z) < f_1(z) \leqslant f_1(\tfrac{3}{2}) = 3.5506\cdots < 4, \quad 0 < z \leqslant \tfrac{3}{2},$$

which, together with (A5), completes the proof of (A3), and hence of (A1).

[1] L. P. Kok and H. van Haeringen, Phys. Rev. C **21**, 512 (1980).

[2] H. van Haeringen, "The Coulomb Potential in Quantum Mechanics and Related Topics," Ph. D. thesis, Free University, Amsterdam, 1978 (unpublished).

[3] H. van Haeringen, "Coulomb T matrix: Representations, Inequalities, Limits, and Zeros," Report 83 07, Delft, 1983, 2nd ed., p. 214; H. van Haeringen, "Twenty-four optimal inequalities and several new representations for the Coulomb T matrices in momentum space," J. Math. Phys. **25**, 3001 (1984).

[4] H. van Haeringen, "Interactions for charged and neutral particles," Report 83 12, Delft, 1983, 2nd ed., p. 140 (to be published).

[5] See also Appendixes Q and L of Report 83 07 (Ref. 3).

Kruskal-like representation of the de Sitter metric

Johanna Friedl Beck[a] and Akira Inomata
Department of Physics, State University of New York at Albany, Albany, New York 12222

(Received 3 November 1983; accepted for publication 15 March 1984)

A new representation of the de Sitter metric is proposed, which is analogous to Kruskal's coordinates for the Schwarzschild metric. The counterpart of Wheeler's tortoise coordinate is also discussed.

PACS numbers: 04.20. – q

I. INTRODUCTION

The Schwarzschild solution for the metric around a mass of the gravitational radius a

$$ds^2 = -(1-a/r)c^2\,dt^2 + (1-a/r)^{-1}\,dr^2 + r^2\,d\omega^2, \quad (1)$$

and the de Sitter solution for the metric of a uniformly curved space-time of radius R

$$ds^2 = -(1-r^2/R^2)c^2\,dt^2 + (1-r^2/R^2)^{-1}\,dr^2 + r^2\,d\omega^2, \quad (2)$$

are different in content but have some similarities in structure. Both are spherically symmetric with

$$d\omega^2 = d\theta^2 + \sin^2\theta\,d\phi^2 \quad (3)$$

and static. As the former has an apparent singularity at $r=a$, so does the latter at $r=R$. The spherical singularities in both cases are removable. The removable character of the Schwarzschild singularity can be best demonstrated in the Kruskal representation.[1] In doing so, Kruskal could exhibit the maximal singularity-free extension of the Schwarzschild solution. The de Sitter singularity can also be seen as the horizon of an observer at the origin $r=0$ and a moving object can cross through the singular sphere $r=R$. Furthermore, the maximal extension of the de Sitter solution can be realized as a pseudosphere in five dimensions.[2] In this regard, one can add nothing to these well-studied solutions. However, in the present paper, placing an emphasis on their similarities in structure, we propose a nonsingular representation for the de Sitter metric (2), which is analogous to Kruskal's for the Schwarzschild metric (1).

II. KRUSKAL-LIKE COORDINATES

Consider the following transformation:

$$u = \{(R-r)/(R+r)\}^{1/2}\cosh(ct/R), \quad (4)$$

$$v = \{(R-r)/(R+r)\}^{1/2}\sinh(ct/R). \quad (5)$$

This transformation maps the de Sitter interior world ($0\leqslant r<R$, $-\infty<t<\infty$) into the domain of $u^2-v^2=1$ in the quadrant $u>|v|$ (sector I shown in Fig. 1). From (4) and (5), we readily obtain the expressions,

$$u^2 - v^2 = (R-r)/(R+r), \quad (6)$$

$$v/u = \tanh(ct/R). \quad (7)$$

The de Sitter metric (1) is then expressed as

$$ds^2 = f^2(u,v)(du^2 - dv^2) + r^2(u,v)d\omega^2 \quad (8)$$

with

$$f(u,v) = R + r = 2R/(u^2 - v^2 + 1). \quad (9)$$

The metric (8) is very similar in form to the Kruskal's representation of the Schwarzschild metric. The functional factor $f(u,v)$ in (9) is an elementary function of (u^2-v^2) while the corresponding factor in the Kruskal case is transcendental. The whole line $r=R$ in the (r,t) plane corresponds to the origin $u=v=0$ as is obvious from (4) and (5). The two one-dimensional families of ideal limit points with $r\to R$ and $t\to\pm\infty$ correspond to the remaining boundary points $u=|v|>0$. The line $r=0$ in the (r,t) plane is represented by the two branches of the hyperbola $u^2-v^2=1$ in the (u,v) plane. It is evident that the metric is entirely regular on the lines, $u=\pm v$ and $u^2-v^2=1$, and inside the region of sector I (the shaded area in Fig. 1).

The line element (8) is regular even in the other three quadrants of the (u,v) plane up to the two branches of the hyperbola, $u^2-v^2=-1$, which correspond to $r=\infty$. This means that the de Sitter coordinates with $r<R$ form a local coordinate patch which covers only part of the whole manifold. In much the same way that the limited Schwarzschild solution is extended to the entire Kruskal manifold, the de Sitter solution can be analytically extended to all sectors in the (u,v) manifold. The extension of sector I to sectors II, III, and IV can be achieved by choosing u and v as follows:

Sector II: $u = \{(r-R)/(r+R)\}^{1/2}\sinh(ct/R)$ (10)

$v = \{(r-R)/(r+R)\}^{1/2}\cosh(ct/R)$, (11)

Sector III: $u = -\{(R-r)/(R+r)\}^{1/2}\cosh(ct/R)$ (12)

$v = -\{(R-r)/(R+r)\}^{1/2}\sinh(ct/R)$, (13)

Sector IV: $u = -\{(r-R)/(R+r)\}^{1/2}\sinh(ct/R)$ (14)

$v = -\{(r-R)/(R+r)\}^{1/2}\cosh(ct/R)$.] (15)

The form of the first equation (6) of the inverse relations remains the same for all sectors, but the second equation (7)

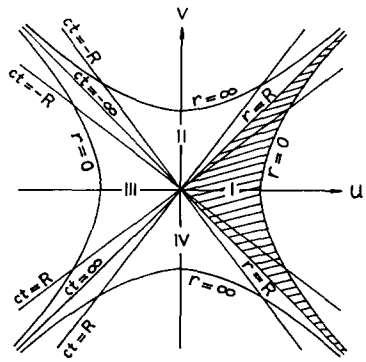


FIG. 1. A u–v diagram for the de Sitter metric.

[a] Present address: Pfaarkirchen Gymnasium, Pfaarkirchen, Federal Republic of Germany.

 0022-2488/84/103039-02$02.50

is valid only for sectors I and III. For sectors II and IV, we have

$$u/v = \tanh(ct/R). \tag{16}$$

Let us examine the extended de Sitter manifold shown as four sectors limited by the hyperbolas $u^2 - v^2 = \pm 1$. The regions of constant r are hyperbolas with asymptotes $u = \pm v$, and the regions of constant t are straight lines through the origin. All radial null geodesics are 45-degrees lines $du = \pm dv$. The signs of the (u, v) variables in each sector have been so chosen that $dv > 0$ for the propagation of light rays. The diagram for the extended de Sitter manifold is very similar in many respects to the Kruskal diagram for the Schwarzschild manifold. The appearance of the family of limited points with $r = 0$ on the two branches of the hyperbolas $u^2 - v^2 = 1$ in sector II and IV is in sharp contrast to the Schwarzschild case. A light ray originating a point P in sector I travels either toward $r = 0$ or $r = R$ depending on the initial direction of motion. If the light ray comes to the boundary at $r = R$, it will continue its way in sector II until it reaches the hyperbola which corresponds to $r = \infty$. The light ray coming toward the center $r = 0$ does not have to stop at $r = 0$. Since the center point $r = 0$ is regular, it may be reflected on the hyperbola and directed to the way to infinity. All geodesics passing through P are confined within the timelike angle between these two possible paths (null geodesics). A test particle can therefore cross the regular boundary at $r = R$ and advance to sector II. A particle, if it starts from a point in sector II, has no choice for its destination; it goes to infinity. A signal sent from sector III can enter sector II, ending up at infinity, but it will never reach sector I. There is no way to establish a communication between sectors I and III. It is interesting that a signal originating in sector IV can get into sector I and III. The signal arriving at the hyperbola of $r = 0$ in sector I or III may be reflected and find its way to infinity. Thus it is clear that the extended manifold is complete; all geodesics emanating from a point in this manifold have infinite length in both directions.

III. TORTOISE COORDINATES

Drawing an analogy between the approach to the Schwarzschild singularity and the famous paradox that Achilles could never pass the tortoise, Wheeler used a tortoise coordinate which goes all the way down to minus infinity while the Schwarzschild variable remains outside the singular surface.[3] Correspondingly, we can define for the de Sitter case a tortoise coordinate by

$$r^* = \tfrac{1}{2}R \ln\{(R + r)/(R - r)\}, \tag{17}$$

which varies from 0 to ∞ according as r changes from 0 to R. The de Sitter metric (2) is then expressed in the form

$$ds^2 = -\operatorname{sech}^2(r^*/R)(c^2\,dt^2 - dr^{*2}) + R^2 \tanh^2(r^*/R)\,d\omega^2. \tag{18}$$

It may also be instructive to introduce the coordinates analogous to those of Eddington and Finkelstein[4]:

$$U = ct - r^*, \tag{19}$$

$$V = ct + r^*. \tag{20}$$

These coordinates label outgoing and ingoing geodesics of free-falling photons, respectively. Outgoing radial null geodesics are given by $U = \text{const}$, while ingoing null geodesics are given by $V = \text{const}$. The situation is again very similar to the Schwarzschild case. Using the outgoing coordinates (U, r), we can exhibit that an observer at $r = 0$ received ingoing signals released only from points inside the singular surface while the outgoing signals go through the singularity. Similarly, in the ingoing coordinates, we see that outgoing photons cannot cross the singular surface. Finally, we point out that the tortoise, outgoing and ingoing variables are related to the (u, v) variables as follows:

$$u = \exp(-r^*/R)\cosh(ct/R), \tag{21}$$

$$v = \exp(-r^*/R)\sinh(ct/R), \tag{22}$$

or

$$u = \tfrac{1}{2}\{\exp(U/R) + \exp(-V/R)\}, \tag{23}$$

$$v = \tfrac{1}{2}\{\exp(U/R) - \exp(-V/R)\}. \tag{24}$$

ACKNOWLEDGMENT

We would like to thank Professor H. Salecker for his hospitality at the University of Munich where part of the work was done.

[1] M. D. Kruskal, Phys. Rev. **119**, 1743 (1960).

[2] See, e.g., A. Trautmann, in *Brandeis Lectures on General Relativity* (Prentice-Hall, Englewood Cliffs, NJ, 1965).

[3] See, e.g., C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[4] A. S. Eddington, Nature, **133**, 192 (1924); D. Finkelstein, Phys. Rev. **110**, 965 (1958).

Static spherically symmetric solutions in a general class of scalar tensor theories

A. Banerjee, S. B. Dutta Choudhury, and N. Banerjee
Department of Physics, Jadavpur University, Calcutta-700032, India

(Received 25 October 1983; accepted for publication 16 February 1984)

The problem of the gravitational field due to a point mass, charged or uncharged, is completely solved within the framework of a general class of scalar tensor theories proposed by Nordtvedt. In Nordtvedt's theory the parameter ω can be an arbitrary function of the scalar field. Different exact solutions are obtained for different functional forms of ω both in the presence and absence of charge. All these solutions including a few already existing in the literature are given in a systematic manner and their asymptotic behavior is studied.

PACS numbers: 04.50. + h

I. INTRODUCTION

The generalization of the Brans–Dicke theory[1] was obtained in Nordtvedt's work[2] by allowing the parameter ω to be a function of the scalar field. This general class includes the theories of Jordan[3] and Brans–Dicke[1] as special cases. It has been claimed by Nordtvedt that an accurate light deflection experiment and also the data on the rate of advance of the perihelion of Mercury could require $\omega' \neq 0$ to fit the data. It is worthwhile in this context to obtain exact solutions of field equations in the generalized scalar tensor theory for different choices of ω as functions of the scalar field ψ, the choices which were introduced from different physical interests. The spherically symmetric static gravitational field in the generalized scalar tensor theory was first discussed by Dutta Choudhury and Bhattacharya[4] for a special choice of ω as a function of the scalar field ψ, introduced by Barkar[5] in order to make the Newtonian gravitational constant G independent of space and time. Rao and Reddy[6] extended the calculations in the presence of an electromagnetic field in the same theory, but with a geometric restriction such as the Weyl tensor vanishing.

Our purpose in this paper is primarily to completely solve the spherically symmetric gravitational field of a charged particle in Nordtvedt's general scalar tensor theory, where ω is an arbitrary function of the scalar field. These solutions include the class obtained previously by Raychaudhuri and Bandyopadhyay[7] or Buchdahl[8] in the context of the Brans–Dicke theory, where ω is constant. Field equations are written in "particle units," that is in units where the particle masses are fixed with G varying. The asymptotic behavior of the solutions are in general studied and some examples are given for different choices of ω as functions of the scalar field. It may be worthwhile to mention in this context that some general relationship between the electric potential and the scalar field were previously given independent of any symmetry by Banerjee and Dutta Choudhury.[9]

In a separate section, the solutions are obtained in the absence of an electromagnetic field. Some of these are already given by Van den Bergh[10] in different notations by using a technique to generate them from the Brans–Dicke metric.

There are some classes of solutions in the presence or absence of an electromagnetic field, which are conformally flat and those in the Brans–Dicke theory were previously obtained by Banerjee and Santos[11] and also by Reddy.[12]

II. STATIC SPHERICALLY SYMMETRIC METRIC ABOUT A CHARGED MASS POINT IN THE GENERAL SCALAR TENSOR THEORY

We consider the line element in the isotropic form

$$ds^2 = e^{\nu}\,dt^2 - e^{\mu}(dr^2 + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2), \quad (2.1)$$

where μ and ν are functions of the radial coordinate r alone. The coupled Nordtvedt–Maxwell field equations in matter-free space may be written as

$$\frac{\mu'^2}{4} + \frac{\mu'\nu'}{2} + \frac{\mu' + \nu'}{r} = -\frac{e^{-\nu}\phi'^2}{\psi} + \frac{\omega\psi'^2}{2\psi^2} - \left(\mu' + \frac{\nu'}{2} + \frac{2}{r}\right)\frac{\psi'}{\psi}, \quad (2.2)$$

$$\frac{\mu'' + \nu''}{2} + \frac{\nu'^2}{4} + \frac{\mu' + \nu'}{2r} = \frac{e^{-\nu}\phi'^2}{\psi} - \frac{\omega\psi'^2}{2\psi^2} - \left(\frac{\nu'}{2} + \frac{1}{r}\right)\frac{\psi'}{\psi} - \frac{\psi''}{\psi}, \quad (2.3)$$

$$\mu'' + \frac{\mu'^2}{4} + \frac{2\mu'}{r} = -\frac{e^{-\nu}\phi'^2}{\psi} - \frac{\omega\psi'^2}{2\psi^2} - \left(\frac{\mu'}{2} + \frac{2}{r}\right)\frac{\psi'}{\psi} - \frac{\psi''}{\psi}, \quad (2.4)$$

where a prime indicates differentiation with respect to r. Further, the wave equation for the scalar field ψ and the Maxwell's equation for the electric potential ϕ are given, respectively, by

$$\left(\frac{\mu' + \nu'}{2} + \frac{2}{r}\right)\psi' + \psi'' = -\frac{\omega'\psi'}{(2\omega + 3)}, \quad (2.5)$$

and

$$[r^2 e^{(\mu - \nu)/2}\phi']' = 0. \quad (2.6)$$

However, the wave equation (2.5) does not give any additional information in view of the fact that it is a consequence of the other field equations for electrovac in a scalar tensor theory. The only nonvanishing component of the elecromag-

 0022-2488/84/103041-06$02.50

netic field tensor for a radial electric field is given by $F_{01} = \phi'$. Now multiplying (2.3) by 2 and adding the result with the equation obtained by subtracting (2.4) from (2.2) we get

$$\left(\nu' + \frac{\psi'}{\psi}\right)\left(\frac{\mu'}{2} + \frac{\nu'}{2} + \frac{2}{r}\right) + \nu'' + \frac{\nu'\psi'}{\psi} + \frac{\psi''}{\psi} = \frac{2e^{-\nu}\phi'^2}{\psi}. \tag{2.7}$$

The relation (2.6) may be written in the form

$$\tfrac{1}{2}\mu'\phi' = \tfrac{1}{2}\nu'\phi' - 2\phi'/r - \phi''. \tag{2.8}$$

Using (2.8) in (2.7) to eliminate μ' we have

$$(e^{\nu}\psi)''\phi' - (e^{\nu}\psi)'\phi'' = 2\phi'^3, \tag{2.9}$$

which in turn yields on integration

$$e^{\nu}\psi = \phi^2 + a\phi + b, \tag{2.10}$$

with a and b as arbitrary constants. Again adding (2.2) and (2.3) we obtain

$$\frac{1}{2}(\mu'' + \nu'') + \frac{1}{2}(\mu' + \nu')\frac{\psi'}{\psi} + \frac{\psi''}{\psi} + \left(\frac{\mu' + \nu'}{2} + \frac{\psi'}{\psi}\right)\left(\frac{\mu' + \nu'}{2} + \frac{3}{r}\right) = 0, \tag{2.11}$$

which can be integrated to yield

$$r^2 e^{(\mu+\nu)/2}\psi = mr^2 + n, \tag{2.12}$$

where m and n are arbitrary constants. The integration of Eq. (2.6) yields

$$r^2 e^{(\mu-\nu)/2}\phi' = q, \tag{2.13}$$

q is an arbitrary integration constant, which is seen to be related with the total charge of the source. In view of the relations (2.10), (2.12), and (2.13) one can write the following three relations:

$$e^{\mu} = (m + n/r^2)^2(\phi^2 + a\phi + b)^{-1}\psi^{-1}, \tag{2.14}$$

$$e^{\nu} = (\phi^2 + a\phi + b)\psi^{-1}, \tag{2.15}$$

$$\phi'(\phi^2 + a\phi + b)^{-1} = q(mr^2 + n)^{-1}. \tag{2.16}$$

Using (2.14)–(2.16) to eliminate ν and the derivatives of μ, ν, and ϕ from the field equations it is possible to arrive at the equation

$$(2\omega + 3)(\psi'^2/\psi^2) = -[q^2(a^2 - 4b) + 16mn](mr^2 + n)^{-2}. \tag{2.17}$$

It is evident from the above relations that one can determine the metric explicitly provided one integrates (2.16) and (2.17) for ϕ and ψ, respectively. It is, however, necessary to know the exact functional form of ω in order to integrate Eq. (2.17).

III. SOLUTIONS FOR THE ELECTRIC POTENTIAL

The integration of Eq. (2.16) yields the following different solutions for ϕ depending on the magnitudes of the constants a, b, m, and n: (1) $a^2 > 4b$, $m = 0$, $n \neq 0$,

$$\phi = \left(\frac{a^2}{4} - b\right)^{1/2}\left[\frac{1 + a_1\exp\{(q/n)(a^2 - 4b)^{1/2}r\}}{1 - a_1\exp\{(q/n)(a^2 - 4b)^{1/2}r\}}\right] - \frac{a}{2}; \tag{3.1}$$

(2) $a^2 > 4b$, $m \neq 0$, $n = 0$,

$$\phi = \left(\frac{a^2}{4} - b\right)^{1/2}\left[\frac{1 + a_2\exp\{-(q/m)(a^2 - 4b)^{1/2}r^{-1}\}}{1 - a_2\exp\{-(q/m)(a^2 - 4b)^{1/2}r^{-1}\}}\right]; \tag{3.2}$$

(3) $a^2 > 4b$, $m > 0$, $n < 0$,

$$\phi = \left(\frac{a^2}{4} - b\right)^{1/2}\left[\frac{1 + a_3((\sqrt{m}r - \sqrt{-n})/(\sqrt{m}r + \sqrt{-n}))^{q((b - a^2/4)/mn)^{1/2}}}{1 - a_3((\sqrt{m}r - \sqrt{-n})/(\sqrt{m}r + \sqrt{-n}))^{q((b - a^2/4)/mn)^{1/2}}}\right] - \frac{a}{2}; \tag{3.3}$$

(4) $a^2 > 4b$, $m < 0$, $n > 0$,

$$\phi = \left(\frac{a^2}{4} - b\right)^{1/2}\left[\frac{1 + a_4((\sqrt{n} + \sqrt{-m}r)/(\sqrt{n} - \sqrt{-m}r))^{q((b - a^2/4)/mn)^{1/2}}}{1 - a_4((\sqrt{n} + \sqrt{-m}r)/(\sqrt{n} - \sqrt{-m}r))^{q((b - a^2/4)/mn)^{1/2}}}\right] - \frac{a}{2}; \tag{3.4}$$

(5) $a^2 > 4b$, $m > 0$, $n > 0$,

$$\phi = \left(\frac{a^2}{4} - b\right)^{1/2}\left[\frac{1 - a_5\exp\{q((a^2 - 4b)/mn)^{1/2}\tan^{-1}(\sqrt{m/n}r)\}}{1 + a_5\exp\{q((a^2 - 4b)/mn)^{1/2}\tan^{-1}(\sqrt{m/n}r)\}}\right] - \frac{a}{2}; \tag{3.5}$$

(6) $a^2 = 4b$, $m = 0$, $n \neq 0$,

$$\phi = [a_6 - (q/n)r]^{-1} - a/2; \tag{3.6}$$

(7) $a^2 = 4b$, $m \neq 0$, $n = 0$,

$$\phi = [a_7 + (q/m)(1/r)]^{-1} - a/2; \tag{3.7}$$

(8) $a^2 = 4b$, $m > 0$, $n < 0$,

$$\phi = \frac{1}{2}\left[\ln\left\{a_8\left(\frac{\sqrt{m}r - \sqrt{-n}}{\sqrt{m}r + \sqrt{-n}}\right)^{-(q/4)(-mn)^{-1/2}}\right\}\right]^{-1} - \frac{a}{2}; \tag{3.8}$$

(9) $a^2 = 4b$, $m < 0$, $n > 0$,

$$\phi = \frac{1}{2}\left[\ln\left\{a_9\left(\frac{\sqrt{n}+\sqrt{-m}r}{\sqrt{n}-\sqrt{-m}r}\right)^{(-q/4)(-mn)^{-1/2}}\right\}\right] - \frac{a}{2}; \tag{3.9}$$

(10) $a^2 = 4b$, $m > 0$, $n > 0$,

$$\phi = [a_{10} - q(mn)^{-1/2}\tan^{-1}(\sqrt{m/n}\, r)]^{-1} - a/2; \tag{3.10}$$

(11) $a^2 < 4b$, $m = 0$, $n \neq 0$,

$$\phi = \left(b - \frac{a^2}{4}\right)^{1/2}\tan\left\{\frac{q\left(b - \frac{a^2}{4}\right)^{1/2}}{n}r + a_{11}\right\} - \frac{a}{2}; \tag{3.11}$$

(12) $a^2 < 4b$, $m \neq 0$, $n = 0$,

$$\phi = \left(b - \frac{a^2}{4}\right)^{1/2}\tan\left[a_{12} - \left(\frac{q}{m}\right)\left(b - \frac{a^2}{4}\right)^{1/2} r^{-1}\right] - \frac{a}{2}; \tag{3.12}$$

(13) $a^2 < 4b$, $m > 0$, $n < 0$,

$$\phi = \left(b - \frac{a^2}{4}\right)^{1/2}\tan\left[\ln\left\{a_{13}\left(\frac{\sqrt{m}r - \sqrt{-n}}{\sqrt{m}r + \sqrt{-n}}\right)^{(1/2)q((a^2/4 - b)/mn)^{1/2}}\right\}\right] - \frac{a}{2}; \tag{3.13}$$

(14) $a^2 < 4b$, $m < 0$, $n > 0$,

$$\phi = \left(b - \frac{a^2}{4}\right)^{1/2}\tan\left[\ln\left\{a_{14}\left(\frac{\sqrt{n}+\sqrt{-m}r}{\sqrt{n}-\sqrt{-m}r}\right)^{(1/2)q((a^2/4 - b)/mn)^{1/2}}\right\}\right] - \frac{a}{2}; \tag{3.14}$$

(15) $a^2 < 4b$, $m > 0$, $n > 0$,

$$\phi = \left(b - \frac{a^2}{4}\right)^{1/2}\tan\left[q\left(\frac{b - a^2/4}{mn}\right)^{1/2}\tan^{-1}\left(\sqrt{\frac{m}{n}}\, r\right) + a_{15}\right] - \frac{a}{2}; \tag{3.15}$$

where, $a_1, a_2, \ldots, a_{15}$ are arbitrary constants. Both m and n cannot be negative in view of Eq. (2.12) and the corresponding solution is therefore excluded.

In the above mentioned solutions all are not completely different and independent of one another. The solutions for $mn = 0$ that is either for $m = 0$, $n \neq 0$ or $m \neq 0$, $n = 0$ are equivalent, as can be shown by the fact that one can be obtained from the other by a coordinate transformation such as $r \to 1/r$. It can be shown in the following manner.

In view of (2.14)–(2.17) one can write for $m = 0$, $n \neq 0$

$$ds^2 = \frac{1}{\psi}\left[(\phi^2 + a\phi + b)dt^2 - \frac{n^2}{r^4}(\phi^2 + a\phi + b)^{-1}(dr^2 + r^2 d\Omega^2)\right], \tag{3.16}$$

$$\frac{(d\phi/dr)}{(\phi^2 + a\phi + b)} = \frac{q}{n}, \tag{3.17}$$

and

$$(2\omega + 3)\frac{1}{\psi^2}\left(\frac{d\psi}{dr}\right)^2 = -\frac{q^2(a^2 - 4b)}{n^2}. \tag{3.18}$$

For $n = 0$, $m \neq 0$ one has

$$ds^2 = (1/\psi)[(\phi^2 + a\phi + b)dt^2 - m^2(\phi^2 + a\phi + b)^{-1}(dr^2 + r^2 d\Omega_\delta^2)], \tag{3.19}$$

$$\frac{(d\phi/dr)}{(\phi^2 + a\phi + b)} = \frac{q}{mr^2}, \tag{3.20}$$

and

$$(2\omega + 3)\frac{1}{\psi^2}\left(\frac{d\psi}{dr}\right)^2 = -\frac{q^2(a^2 - 4b)}{m^2 r^4}. \tag{3.21}$$

Now it is evidently possible to obtain the solutions in the latter case from those in the former by a coordinate transformation such as $r \to 1/r$ and by suitably changing the constants.

It is further noted that the solutions given in (2.14)–(2.17) are equivalent in two cases, that is, in (a) $m > 0$, $n < 0$ and (b) $m < 0$, $n > 0$. Thus the solutions (3.2), (3.4), (3.5), (3.6), (3.8), (3.10), (3.12), (3.14), and (3.15) may be considered to be nine independent solutions. These solutions include all cases in which $(2\omega + 3)$ is greater than, equal to, or less than zero. In all of the above-mentioned solutions $\phi' \to 0$ as $r \to \infty$. From (2.17) $\psi' \to 0$ as $r \to \infty$ provided ψ remains finite. By taking derivatives of both sides in (2.14) and (2.15) it can be shown easily that both μ' and ν' vanish asymptotically so that the solutions are asymptotically flat in such cases. Some of these solutions are also conformally flat, that is, the Weyl tensor vanishes in each case. We consider, for example, the solution (3.6). Putting $a_6 = 0$ we have in this case, that is for $a^2 = 4b$, $m = 0$, $n \neq 0$

$$\phi = -n/q \cdot 1/r - a/2, \tag{3.22}$$

and

$$\phi' = n/q \cdot 1/r^2. \tag{3.23}$$

Now in view of (2.14)–(2.16) one finds

$$e^{(\nu - \mu)/2} = (\phi'/q)r^2 = n/q^2, \tag{3.24}$$

which is sufficient to ensure that the Weyl tensor vanishes.[13] In this case $(2\omega + 3) = 0$.

Similarly in the solution (3.4) with $a^2 > 4b$, $m < 0$, $n > 0$, if one puts $a_4 = 1$ and assumes the arbitrary constants appearing in the solution satisfy $n = q^2$, $a^2 - 4b = -4m$, one finds that

$$\phi = -(q/r + a/2).$$

Thus (3.4) is conformally flat. Here $(2\omega + 3) > 0$. In the solution (3.15) that is for $a^2 < 4b$, $m > 0$, $n > 0$ assuming that $a^2 - 4b = -4m$ and $n = q^2$ it is possible to obtain

$$\phi = -(q/r + a/2).$$

Here also one can calculate $e^{(\nu-\mu)/2}$ in the manner shown in (3.24) and find that it is a constant with the result that Weyl tensor vanishes. In this case $(2\omega + 3) < 0$.

In all the above three cases it is not difficult to find ψ and solve for the metric utilizing (2.14)–(2.17). The solutions in the B–D theory are found to be exactly identical with those given earlier by Banerjee and Santos[11] and those in the Barkar's theory were previously given by Rao and Reddy.[6]

Nine independent solutions mentioned above for a charged particle in the scalar tensor theory contain five solutions previously discussed by Raychaudhuri and Bandyopadhyay[7] in the special case of the Brans–Dicke theory. For example, solutions (3.8), (3.10), (3.12), (3.14), and (3.15) are the same solutions as were given by them for $(2\omega + 3) > 0$. The remaining four solutions are for $(2\omega + 3) < 0$ and for $(2\omega + 3) = 0$ and were not mentioned by Raychaudhuri and Bandyopadhyay. The detailed calculations for establishing the equivalence of our solutions and those done previously are omitted.

IV. SOLUTIONS FOR THE SCALAR FIELD

The integration of Eq. (2.17) yields the value of ψ and for this one should use the explicit functional form of $\omega(\psi)$. Different forms of ω as functions of the scalar field are proposed by different workers from different points of view (Van den Bergh Ref. 10). We find explicit solutions for ψ in some of these different theories without mentioning the physical justifications of our choices. Appropriate references are, however, mentioned in each case.

Case 1 (Brans–Dicke theory) [1]: $\omega = \text{const.}$

(i) $m = 0$, $n \neq 0$,

$$\psi = \psi_0 \exp\left[\frac{q}{n}\left(\frac{4b - a^2}{2\omega + 3}\right)^{1/2} r\right]; \tag{4.1}$$

(ii) $m \neq 0$, $n = 0$,

$$\psi = \psi_0 \exp\left[-\frac{q}{m}\left(\frac{4b - a^2}{2\omega + 3}\right)^{1/2} \frac{1}{r}\right]; \tag{4.2}$$

(iii) $m > 0$, $n < 0$,

$$\psi = \psi_0 \left[\frac{\sqrt{m}r - \sqrt{-n}}{\sqrt{m}r + \sqrt{-n}}\right]^{(1/2)\sqrt{K_1}},$$

where

$$K_1 = \frac{q^2(a^2 - 4b) + 16mn}{mn(2\omega + 3)}; \tag{4.3}$$

(iv) $m < 0$, $n > 0$,

$$\psi = \psi_0 \left[\frac{\sqrt{n} + \sqrt{-m}r}{\sqrt{n} - \sqrt{-m}r}\right]^{(1/2)\sqrt{K_1}}; \tag{4.4}$$

and

(v) $m > 0$, $n > 0$,

$$\psi = \psi_0 \exp[(-K_1)^{1/2} \tan^{-1}(\sqrt{m/n}\, r)]. \tag{4.5}$$

The integration constant in each case is ψ_0. Among the above solutions for reasons mentioned earlier (4.1) and (4.2), (4.3), and (4.4) are equivalent and we have three independent solutions for ψ.

Case 2 (Barkar's Theory) [5]: Here $\omega = (4 - 3\psi)/2(\psi - 1)$ and the consequence of this choice is that the so-called Newtonian gravitational constant becomes independent of space coordinates and is no longer a variable quantity as in the Brans–Dicke theory. The three independent solutions for the scalar field ψ in this case are given by (i) $m \neq 0$, $n = 0$,

$$\psi = 1 - \left[\frac{1 - B_1 \exp\{-(q/m)(4b - a^2)^{1/2} 1/r\}}{1 + B_1 \exp\{-(q/m)(4b - a^2)^{1/2} 1/r\}}\right]^2; \tag{4.6}$$

(ii) $m > 0$, $n < 0$,

$$\psi = \sec^2 \ln\left[B_2\left(\frac{\sqrt{m}r - \sqrt{-n}}{\sqrt{m}r + \sqrt{-n}}\right)^{(1/4)\sqrt{K_2}}\right], \tag{4.7}$$

with

$$K_2 = \frac{q^2(a^2 - 4b) + 16mn}{mn};$$

(iii) $m > 0$, $n > 0$,

$$\psi = 1 + \left[\frac{1 - B_3 \exp\{\sqrt{K_2} \tan^{-1}(\sqrt{(m/n)}r)\}}{1 + B_3 \exp\{\sqrt{K_2} \tan^{-1}(\sqrt{(m/r)}r)\}}\right], \tag{4.8}$$

where B_1, B_2, and B_3 are arbitrary integration constants.

For the following three theories see Van den Bergh[10] and references therein.

Case 3 (Schwinger's theory): In this case $\omega = (1 - 3\alpha\psi)/2\alpha\psi$, where $\alpha = \text{const.}$ Solutions for ψ are given by (i) $m \neq 0$, $n = 0$,

$$\psi = [(q/2m)\{-\alpha(a^2 - 4b)\}^{1/2}\, 1/r + C_1]^{1/2}, \tag{4.9}$$

(ii) $m > 0$, $n < 0$,

$$\psi = \left[C_2 - \frac{K_3^{1/2}}{4} \ln\left(\frac{\sqrt{m}r - \sqrt{-n}}{\sqrt{m}r + \sqrt{-n}}\right)\right]^{-2}, \tag{4.10}$$

where

$$K_3 = \frac{\alpha\{q^2(a^2 - 4b) + 16mn\}}{mn};$$

and (iii) $m > 0$, $n > 0$,

$$\psi = [C_3 - \tfrac{1}{2}(-K_3)^{1/2} \tan^{-1}(\sqrt{m/n}\, r)]^{-2}. \tag{4.11}$$

In the above C_1, C_2, C_3 are arbitrary constants.

Case 4 (Model with curvature coupling): Here

$$\omega = 3\psi/2(1 - \psi). \tag{4.12}$$

The following independent solutions exist. All of these solutions remain finite at $r \to \infty$.

(i) $m \neq 0$, $n = 0$,

$$\psi = \sec^2\left[D_1 - \frac{q}{m}\left(\frac{a^2 - 4b}{3}\right)^{1/2} \cdot \frac{1}{r}\right]. \tag{4.13a}$$

(ii) $m > 0$, $n < 0$,

$$\psi = 1 - \left[\frac{1 - D_2\left(\frac{\sqrt{m}r - \sqrt{-n}}{\sqrt{m}r + \sqrt{-n}}\right)^{K_4}}{1 + D_2\left(\frac{\sqrt{m}r - \sqrt{-n}}{\sqrt{m}r + \sqrt{-n}}\right)^{K_4}}\right]^2, \tag{4.13b}$$

where

$$K_4 = \left[\frac{q^2(a^2-4b)+16mn}{12mn}\right]^{1/2}.$$

(iii) $m>0$, $n>0$,

$$\psi = \sec^2[K_4 \tan^{-1}(\sqrt{m/n}\,r) + D_3]. \tag{4.13c}$$

D_1, D_2, D_3 are arbitrary constants.

Case 5 (Models with ψ^l coupling): Here $\omega = \omega_0\psi^l$, with l being a constant power. ψ is given by the following relations in three different cases.

(i) $m\neq 0$, $n=0$,

$$[-(2\omega_0\psi^l+3)]^{1/2} - \sqrt{3}\tan^{-1}[-(1+\tfrac{2}{3}\omega_0\psi^l)]^{1/2} = A_1 - (lq/2m)(a^2-4b)^{1/2}\cdot 1/r. \tag{4.14}$$

(ii) $m>0$, $n<0$,

$$\left[\frac{(2\omega_0\psi^l+3)^{1/2}-\sqrt{3}}{(2\omega_0\psi^l+3)^{1/2}+\sqrt{3}}\right]\exp[(2\omega_0\psi^l+3)^{1/2}] = A_2\left[\frac{\sqrt{m}r-\sqrt{-n}}{\sqrt{m}r+\sqrt{-n}}\right]^{lK/2}, \tag{4.15}$$

with

$$K = \left[\frac{q^2(a^2-4b)+16mn}{4mn}\right]^{1/2}.$$

(iii) $m>0$, $n>0$,

$$[-(2\omega_0\psi^l+3)]^{1/2}p\sqrt{3}\tan^{-1}[-(2\omega_0\psi^l+3)/3]^{1/2} = (lK/2)\tan^{-1}(\sqrt{m/n}\,r) + A_3. \tag{4.16}$$

In the above A_1, A_2, and A_3 are arbitrary constants. The solutions for ψ mentioned above are the independent solutions which remain finite as $r\to\infty$ and thus the metric determined from (2.14) and (2.15) using ϕ and ψ given here will exhibit asymptotic flatness.

V. THE SPHERICALLY SYMMETRIC STATIC GRAVITATIONAL FIELD IN THE GENERAL SCALAR TENSOR THEORY FOR AN UNCHARGED POINT PARTICLE

The gravitational field due to a mass particle without any charge is treated here separately due to the fact that one cannot arrive at these solutions by simply putting $\phi' = 0$ in the previous results. The reason for this is that in the present case there is no electric field and relations like (2.8), (2.9) are trivially satisfied. Following the same procedure as shown in Sec. II one gets, instead of the relation (2.7), the following:

$$(\nu''\psi + \nu'\psi' + \psi'') + (\nu'\psi + \psi')((\mu' + \nu')/2 + 2/r) = 0,$$

which on integration yields

$$r^2 e^{(\mu+\nu)/2}(\nu'\psi + \psi') = p, \tag{5.1}$$

and also the relation

$$r^2 e^{(\mu+\nu)/2}\psi = mr^2 + n, \tag{5.2}$$

m, n, p being arbitrary integration constants. By using (5.1) and (5.2) it is possible to eliminate μ' and ν' from any of the field equations written in the absence of electromagnetic field and one thus obtains the equation

$$(2\omega+3)(\psi'^2/\psi^2) + (p^2+16mn)(mr^2+n)^{-2} = 0. \tag{5.3}$$

In view of Eqs. (2.5), (5.1), and (5.2) the following independent set of relations between the metric components and the scalar field can be obtained in different cases.

(i) $p=n=0$, $m\neq 0$,

$$e^{\nu}\psi = a_1, \quad e^{\mu}\psi = m^2/a_1, \tag{5.4}$$

a_1 being an arbitrary constant and $(2\omega+3)=0$ as is evident from (5.3).

(ii) $n=0$, p, $m\neq 0$,

$$e^{\nu}\psi = a_2\, e^{-p/mr}, \quad e^{\mu}\psi = (m^2/a^2)e^{p/mr}. \tag{5.5}$$

Here a_2 is an arbitrary constant and $(2\omega+3)<0$.

(iii) $p\neq 0$, $m>0$, $n<0$.

In this case $(2\omega+3)>0$ and

$$e^{\nu}\psi = a_3\left(\frac{\sqrt{m}r-\sqrt{-n}}{\sqrt{m}r+\sqrt{-n}}\right)^{p/2\sqrt{-mn}},$$

$$e^{\mu}\psi = a_3^{-1}\, r^{-4}(mr^2+n)^2\left(\frac{\sqrt{m}r-\sqrt{-n}}{\sqrt{m}r+\sqrt{-n}}\right)^{-p/2\sqrt{-mn}}, \tag{5.6}$$

a_3 being an arbitrary constant.

(iv) $p\neq 0$, $m>0$, $n>0$.

Here $(2\omega+3)>0$,

$$e^{\nu}\psi = a_4\exp\left[\frac{p}{\sqrt{mn}}\tan^{-1}\left(\sqrt{\frac{m}{n}}\,r\right)\right],$$

$$e^{\mu}\psi = a_4^{-1}\, r^{-4}(mr^2+n)^2\exp\left[-\frac{p}{\sqrt{mn}}\tan^{-1}\left(\sqrt{\frac{m}{n}}\,r\right)\right]. \tag{5.7}$$

Here a_4 is an arbitrary constant. We see that in all these four cases mentioned above the solutions for the metric are given in terms of the scalar field and the radial coordinate. The scalar field itself is a function of the radial coordinate in the static spherically symmetric case and can be obtained by solving Eq. (5.3) in different situations provided $\omega(\psi)$ is given in terms of ψ. We can use different functional forms of $\omega(\psi)$ in different theories mentioned in the previous section and can solve for ψ. The solutions for ψ will be structurally identical with the solutions for ψ in the presence of electromagnetic field, which is expected because of the identical nature of the differential equations (2.17) and (5.3). Once ψ is known, the metric can be found from Eqs. (5.4)–(5.7). It may be noted that the metric corresponding to (5.6) for different functional forms of $\omega(\psi)$ satisfy the condition $(2\omega+3)>0$ and they can be found identical with those mentioned earlier by Van den Bergh[10] in different notations. The metric in the latter case was however, generated from the Brans–Dicke metric by a suitable technique. Other solutions with $(2\omega+3)=0$ (Brans–Dicke theory) or $(2\omega+3)<0$ in our paper are new. Further it can be shown by arguments mentioned in Sec. III that the solutions corresponding to (5.4)–(5.7) are all asymptotically flat.

ACKNOWLEDGMENT

One of the authors (N.B.) is grateful to the C.S.I.R. (India) for the award of a Junior Research Fellowship.

[1]C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).
[2]K. Nordtvedt, Jr., Astrophys. J. **161**, 1059 (1970).
[3]P. Jordan, Z. Phys. **157**, 112 (1959).
[4]S. B. Dutta Choudhury and D. Bhattacharya, J. Math. Phys. **23**, 1366 (1982).
[5]B. M. Barkar, Astrophys. J. **219**, 5 (1978).
[6]V. U. M. Rao and D. R. K. Reddy, Gen. Relativ. Gravit. **14**, 1017 (1982).
[7]A. K. Raychaudhuri and N. Bandyopadhyay, Prog. Theor. Phys. **59**, 414 (1978).
[8]H. A. Buchdahl, Nuovo Cimento B **12**, 269 (1972).
[9]A. Banerjee and S. B. Dutta Choudhury, J. Math. Phys. **21**, 2799 (1980).
[10]N. Van den Bergh, Gen. Relativ. Gravit. **14**, 17 (1982).
[11]A. Banerjee and N. O. Santos, J. Math. Phys. **22**, 1075 (1981).
[12]D. R. K. Reddy, J. Math. Phys. **20**, 23 (1979).
[13]A. Banerjee and M. M. Som, Int. J. Theor. Phys. **20**, 349 (1981).

Aspects of high-dimensional theories in embedding spaces

Marcos D. Maia and W. Mecklenburg[a)]
Universidade de Brasília, Departamento de Matemática, 70.910-Brasília, DF, Brazil

(Received 6 February 1984; accepted for publication 24 February 1984)

In this paper we concern ourselves with the question of whether physical meaning may be attributed to the extra dimensions provided by embedding procedures applied to physical space-times. We also comment on similarities and differences to conventional Kaluza–Klein pictures.

PACS numbers: 04.50. + h, 12.10.En

I. INTRODUCTION

This paper is motivated by the recent revival of interest in high-dimensional field theories, in particular those of the Kaluza–Klein type (cf. Refs. 1–7 for a list of relevant papers on the latter subject).

The central novel feature is the suggestion to introduce high-dimensional spaces as the (local and isometric) embedding spaces of four-dimensional space-times. The physical motivation for doing so is that by the very nature of the embedding process the extra dimensions find their origin in the presence of gravity. Moreover, if the embedding is required to be a minimal one then the dimension of the high-dimensional space is fixed.

For technical reasons we will confine our attention in this note to flat embedding spaces $M_p(m,n)$ of dimension p and signature $(m(+),n(-))$. We will make occasional remarks concerning different choices, however.

We will then touch in particular the following point. The metric of the embedding space can be given a form where it closely resembles the Kaluza–Klein ansatz.[1–7] We find however upon investigating further constraints inherent in the embedding formalism that dynamically the two pictures have quite different features.

It can also be argued that embedding has some bearing on the problem of mixing internal and space-time symmetries. This aspect has been dealt with in a different paper.[8]

We analyze in particular the limit of vanishing gravity. We find that the gauge-field type objects in our Kaluza–Klein-like picture disappear in this limit. (We note in passing that in this limit the nontrivial mixing of internal and space-time symmetries is lost.[8])

The plan of the paper is as follows. In Sec. II we present the formalism of embedding inasmuch as we need it here; in particular we will introduce the Gauss–Codazzi–Ricci (GCR) equations. In Sec. III we comment on similarities and differences from the Kaluza–Klein picture. Further discussion of questions of dynamics involves solving the GCR equations in a special case. Section IV summarizes our conclusions.

II. THE FORMALISM OF EMBEDDING

Consider a four-dimensional pseudo-Riemannian space V_4 whose points are labeled by local coordinates x^i. We will look for local isometric embeddings[9–11] into $M_p(m,n)$, a Minkowskian space of dimension p and signature $(m(+),m(-)), p=(m+n)$. We note in passing that principally the embedding problem can be formulated using quite arbitrary Riemannian spaces V_p rather than M_p as embedding spaces. The present restriction to M_p is mainly for technical reasons. Powerful theorems[11] restrict the minimal dimension p of M_p for embedding of a Riemannian V_4 to be maximally 14 under suitable differentiability conditions.[12] Of course, under special circumstances, p may be less than 14. In this way embedding provides a natural upper limit on the number of extra dimensions. Also, the signature $(m(+),m(-))$ is fixed. Note that even though the embedding is specified locally only, still the embedding space $M_p(m,n)$ remains the same for all regular points of V_4.

We will in this paper, not touch the question whether there could be a p-dimensional theory that determines M_p. We merely note in passing that principally two possibilities exist. One is where the p-dimensional space is simply M_p (or some other conveniently chosen, but fixed, embedding space), and all we are really concerned with are the four-dimensional Einstein equations and the GCR equations (see below). Another possibility would be to start off with a p-dimensional version of general relativity. Then M_p would presumably correspond to some "ground-state" solution of this theory. However, discussion of such points is beyond the scope of the present paper.

Let us now turn to the formalism of embedding. We will embed V_4 into $M_p(m,n)$. The numbers p,m,n are determined through the choice of V_4. Attach now to every point $(x^i)\in V_4$ an orthonormal system of vectors $N_A(x^i)$, $A=5,\ldots,p$, which are orthogonal to V_4. They span a space which we will call $\mathcal{N}$. The N_A together with the vectors tangent to V_4 at (x^i) span the space $M_p(m,n)$. Points in M_p are labeled by Cartesian coordinates Z^μ, $\mu=1,\ldots,p$. Let, in particular, $X^\mu(x^i)$ be the Cartesian coordinates in M_p of $(x^i)\in V_4$. Then for an arbitrary point $Z^\mu\in M_p$ we can find a point $(x^i)\in V_4$ such that the following decomposition holds:

$$Z^\mu = Z^\mu(x^i,y^A) = X^\mu(x^i) + y^A N^\mu_A(x^i). \tag{2.1}$$

By (2.1), the parameters y^A are associated with directions N_A in $\mathcal{N}$. The set of coordinates $z^\alpha=(x^i,y^A)$ defines a *Gaussian coordinate system* in M_p. Physical space-time, V_4 is a subspace of M_p and singled out through $y^A=0$. Thus the space-time projection of a generic function $\phi(z_\alpha)$, to be denoted by $\phi(z_\alpha)|_{V_4}$, is given by $\phi(z_\alpha)|_{V_4}=\phi(z_\alpha)|_{y^A=0}$.

Let now $\eta_{\mu\nu}$ denote the Cartesian components of the

[a)] Address after 1 June, 1983: 19, Ditmar-Koel-Strasse, D-02 Hamburg 11, Federal Republic of Germany.

 0022-2488/84/103047-04$02.50

metric tensor of $M_p(m,n)$. The vectors tangent to V_4 at (x^i) have components $X^\mu_{,i}(x^j)$. The embedding is determined through the equations

$$X^\mu_{,i}X^\nu_{,j}\eta_{\mu\nu} = \bar{g}_{ij}, \tag{2.2}$$

$$X^\mu_{,i}N^\nu_A\eta_{\mu\nu} = 0, \tag{2.3}$$

$$N^\mu_A N^\nu_B\eta_{\mu\nu} = \eta_{AB}, \tag{2.4}$$

where $\bar{g}_{ij}$ is the metric of V_4. Equations (2.3) and (2.4) are just orthonormality relations for tangent and normal vectors, respectively, whereas (2.2) states an *isometry condition*. In fact, the line element in V_4 is $ds^2 = \bar{g}_{ij}\,dx^i\,dx^j$. Rewritten in terms of Cartesian coordinates it becomes $ds^2 = \eta_{\mu\nu}\,dX^\mu(x^i)dX^\nu(x^i) = \eta_{\mu\nu}X^\mu_{,i}X^\nu_{,j}\,dx^i\,dx^j$. The requirement that the embedding be an isometric one then means that we have to equate these two expressions for ds^2. This gives indeed (2.2).

A metric $\gamma_{\alpha\beta}$ in the Gaussian system in M_p is specified through a coordinate transformation,

$$\gamma_{\alpha\beta} = Z^\mu_{,\alpha}Z^\nu_{,\beta}\eta_{\mu\nu}. \tag{2.5}$$

This relation expresses again an isometry of M_p, at least formally, since

$$ds^2 = \gamma_{\alpha\beta}\,dx^\alpha\,dx^\beta = \eta_{\mu\nu}Z^\mu_{,\alpha}Z^\nu_{,\beta}\,dx^\alpha\,dx^\beta = \eta_{\mu\nu}\,dZ^\mu\,dZ^\nu.$$

(The Gaussian metric $\gamma_{\alpha\beta}$ is induced by the Cartesian metric $\eta_{\mu\nu}$.)

Using (2.1) we can spell out $\gamma_{\alpha\beta}$ in a more explicit way. One finds

$$\gamma_{ij} = \bar{g}_{ij} + (y^A N^\mu_{A,i}X^\nu_{,j} + y^A N^\mu_{A,j}X^\nu_{,i} + y^A y^B N^\mu_{A,i}N^\nu_{B,j})\eta_{\mu\nu}, \tag{2.6}$$

$$\gamma_{iA} = -y^B N^\mu_{A,i}N^\nu_B\eta_{\mu\nu}, \tag{2.7}$$

$$\gamma_{AB} = N^\mu_A N^\nu_B\eta_{\mu\nu} = \eta_{AB}, \tag{2.8}$$

where η_{AB} is the Cartesian metric of $\mathcal{N} = M_{p-4}$.

It is now convenient to introduce the notations

$$b_{ijA} = N^\mu_{A,i}X^\nu_{,j}\eta_{\mu\nu}, \tag{2.9}$$

and

$$A_{iAB} = N^\mu_{A,i}N^\nu_B\eta_{\mu\nu}. \tag{2.10}$$

From (2.9) and (2.10) we obtain the symmetry properties $b_{ijA} = b_{jiA}$ and $A_{iAB} = -A_{iBA}$ as well as the identity

$$N^\mu_{Aj} = -\bar{g}^{ik}b_{jiA}X^\mu_{,k} + \eta^{BC}A_{jAB}N^\mu_C, \tag{2.11}$$

which is readily checked by contracting both sides with $\eta_{\lambda\mu}X^\lambda_{,i}$ and $\eta_{\lambda\mu}N^\lambda_D$, respectively. The metric tensors $\bar{g}^{ij}$ and η^{AB} are introduced as matrix inverses of $\bar{g}_{ij}$ and η_{AB}. In terms of b_{ijA} and A_{iAB}, we can rewrite $\gamma_{\alpha\beta}$ as

$$\gamma_{\alpha\beta} = \begin{bmatrix} g_{ij} + y^A y^B \eta^{CD}A_{iAC}A_{jBD} & y^D A_{iAD} \\ y^D A_{jBD} & \eta_{AB} \end{bmatrix}, \tag{2.12}$$

with

$$g_{ij} = \bar{g}_{ij} - 2y^A b_{ijA} + y^A y^B \bar{g}^{kl}b_{ikA}b_{jlB}. \tag{2.13}$$

One notices that, except for the appearance of b_{ijA} [cf. Eq. (2.13)], (2.12) is nothing but the well-known Kaluza–Klein ansatz[1–7] over $M_p = M_4 \times M_{p-4}$.

In fact, for pseudo-orthogonal transformations of $\mathcal{N}(= M_{p-4})$ which leave the metric η_{AB} invariant, the Killing fields are $K_{AB}{}^C(y)\partial_C = (y_A\delta^C_B - y_B\delta^C_A)\partial_C = y_A\partial_B - y_B\partial_C$, which, of course, are nothing but the conventional expressions for the generalized angular momentum operators. We can now write $\gamma_{iA} = y^B A_{iAB} = -\frac{1}{2}K_A{}^{BC}(y)A_{iBC}$. From this observation (and recalling similar arguments for conventional Kaluza–Klein theories) we conclude in particular that rotations of $\mathcal{N}$ are accompanied by gauge transformations of $A_{iAB}(x^i)$, the gauge group being the group of pseudo-orthogonal transformations of $\mathcal{N}$. That is, as far as their transformation behavior is concerned, the A_{iAB} behave like gauge fields.

To complete the discussion of the formalism of embedding we note that M_p being flat implies that the curvature tensor of M_p vanishes, $R^\alpha{}_{\beta\gamma\delta} = 0$. An observer confined to V_4 finds that the equations $R^\alpha{}_{\beta\gamma\delta}|_{V_4} = R^\alpha{}_{\beta\gamma\delta}|_{y^A=0} = 0$ have to be satisfied. These latter conditions are the Gauss–Codazzi–Ricci (GCR) equations. They provide a criterion intrinsic to V_4 for the consistency of the embedding; they may be looked upon as an integrability condition for (2.2)–(2.4).

In order to spell out the GCR equations it is convenient to introduce some notations. Let Γ^A be a set of generalized Dirac matrices, i.e., $\{\Gamma^A,\Gamma^B\} = 2\eta^{AB}$, and set $\Sigma^{AB} = \frac{1}{4}[\Gamma^A,\Gamma^B]$. Then we denote $b_{ij} = b_{ijA}\Gamma^A$ and $A_i = A_{iAB}\Sigma^{AB}$. Define a Riemann-covariant d_j in V_4 through

$$d_j b_{ikA} = \partial_j b_{ikA} + \bar{\Gamma}^m{}_{ij}b_{mkA} + \bar{\Gamma}^m{}_{kj}b_{imA}, \tag{2.14}$$

where the $\bar{\Gamma}^m{}_{ij}$ are the Christoffel symbols with respect to $\bar{g}_{ij}$. Finally, set

$$D_i = d_i + \tfrac{1}{2}A_i. \tag{2.15}$$

Then the GCR equations can be written in the form[13]

$$\text{(G)}\quad \bar{R}^i{}_{jkl} = \{\bar{g}^{im}/2^{[(p-4)/2]}\}\,\mathrm{tr}(b_{jk}b_{lm} - b_{jl}b_{km}), \tag{2.16}$$

$$\text{(C)}\quad [D_j,b_{ik}] - [D_i,b_{jk}] = 0, \tag{2.17}$$

$$\text{(R)}\quad [D_i,D_j] = \tfrac{1}{4}\bar{g}^{mn}[b_{mj},b_{ni}]. \tag{2.18}$$

In (2.16) $[d]$ denotes the integer part of d.

From the GCR equations we conclude in particular that there is no reason to expect that the four-dimensional observer can discard b_{ijA} or A_{iAB} — these quantities appear to be inevitably coupled to physical gravity, $\bar{g}_{ij}$.

The reason why this could happen is the following. From the p-dimensional point of view, (2.12) is just the result of a coordinate transformation, $z^\alpha \to Z^\mu(z^\alpha)$. Now, among the isometries of M_p are, for example, x^i-dependent rotations of the y^A; to a four-dimensional observer these will look like gauge transformations affecting A_i. The existence of such gauge transformations, which cannot be compensated by four-dimensional coordinate transformations, implies the existence of gauge fields for the four-dimensional observer. The coupling of gravity to these fields (and to b_{ij}) thus provides an indirect way for the four-dimensional observer to see the extra dimensions. All this is similar to the conventional Kaluza–Klein picture.

There are important differences to the latter, however, and it is to these that we will turn in the next section.

III. DYNAMICS AND RELATION TO THE KALUZA–KLEIN PICTURE

In this section we will compare the present picture with conventional Kaluza–Klein theories. Also we will study some dynamical aspects of the present formalism.

In the conventional Kaluza–Klein (KK) picture one encounters indeed a metric of the form $\gamma_{\alpha\beta}$ [cf. Eq. (2.12)]. A first difference however is the presence of b_{ijA}. A second is that the Killing fields $K_A{}^{BC}(y)$ and the metric tensor η_{AB}, here relating to $\mathcal{N} = M_{p-4}$, are conventionally replaced by the corresponding objects on $B_{p-4} = G/H$, a coset space. This latter difference however would probably disappear if we would choose a space of the form $M_4 \times B_{p-4} = M_4 \times G/H$ as the embedding space.

Actually, the latter choice would give us more options as far as possible devices for dimensional reduction are concerned. Here we have chosen to implement dimensional reduction effectively by projection, $y^A = 0$. (We note in passing that this does *not* mean that all y dependence is neglected. In fact, if we would just neglect all y dependence, A_i and b_{ij} would never appear in the GCR equations.) As $y^A = 0$ projects out physical space-time this is sensible physically and also, dimensional reduction by integrating (averaging) over y does not really seem feasible for flat embedding spaces. Conversely, y integration is the conventional dimensional reduction procedure for compact internal spaces like G/H. However, apart from noticing that other options are at least principally open to us we will (mainly for reasons of convenience) keep confining our attention to flat embedding spaces.

The most important difference to the conventional KK picture now is that instead of a set of Einstein equations in p dimensions like $G_{\alpha\beta} \equiv R_{\alpha\beta} - \frac{1}{2}\gamma_{\alpha\beta}\mathrm{R} = T_{\alpha\beta}$, we have here equations that involve the Riemann tensor $R^{\alpha}{}_{\beta\gamma\delta}$ rather than just the Ricci tensor $R^{\alpha}{}_{\beta}$. In fact, a typical equation is the one given above, $R^{\alpha}{}_{\beta\gamma\delta} = 0$. We therefore have to deal with more equations than in the conventional KK picture.

We may ask however, if the contracted equations $G_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2}\gamma_{\alpha\beta}R = 0$ (which follow from $R^{\alpha}{}_{\beta\gamma\delta} = 0$) correspond to a conventional set of field equations. The equations $G_{\alpha\beta} = 0$ are just the field equations of the conventional KK picture which after dimensional reduction give the field equations for an Einstein–Yang–Mills system. In the present case, with a dimensional reduction implemented through $y^A = 0$, the picture looks different. The reason for this is the following. The curvature scalar corresponding to $\gamma_{\alpha\beta}$ is (the calculation being as in the conventional KK picture)

$$R_p(\gamma_{\alpha\beta}) = R_4(g_{ij}) - \tfrac{1}{4}K_{\mathrm{C}}{}^{AB}(y)K_D{}^{A'B'}(y) \times \eta^{CD}(F_{ABij}F_{A'B'}{}^{ij}), \tag{3.1}$$

with $F_{ij} = [D_i D_j] = F_{ABij}\Sigma^{AB}$. From (3.1) we conclude that the energy-momentum tensor for the Yang–Mills field contains a factor $y^A y^B$ and therefore will vanish for $y^A = 0$. Therefore the contracted GCR equations $R^{\alpha}{}_{\beta}|_{y^A=0} = 0$ are not expected to be a conventional set of field equations for a gravity–Yang–Mills system.

This can also be seen more directly. In fact, from the Gauss equation we conclude

$$2^{[(p-4)/2]}\overline{R}_{jk} = \mathrm{tr}(b_{jk}b^{i}{}_{i} - b_{ji}b^{i}{}_{k}), \tag{3.2}$$

and

$$2^{[(p-4)/2]}\overline{R} = \mathrm{tr}(b_j{}^j b_i{}^i - b_{ij}b^{ij}), \tag{3.3}$$

whence we infer

$$\begin{aligned} &2^{[(p-4)/2]}(\overline{R}_{jk} - \tfrac{1}{2}\bar{g}_{jk}\overline{R}) \\ &\quad = 2^{[(p-4)/2]}\overline{G}_{jk} = \{\mathrm{tr}(b_{jk}b_i{}^i - b_j{}^i b_k{}^i) \\ &\qquad - \tfrac{1}{2}g_{jk}\,\mathrm{tr}(b_j{}^j b_i{}^i - b_{ij}b^{ij})\} \\ &\quad \equiv 2^{[(p-4)/2]}t_{ij}(b). \end{aligned} \tag{3.4}$$

We see indeed that there is no directly related energy-momentum tensor for the gauge field A_i.

All these observations together mean that apart from the form of the metric tensor $\gamma_{\alpha\beta}$ the resemblance to the KK picture does not persist entirely on a dynamical level. Even though this situation might improve if we would choose an embedding space of the form $M_4 \times G/H$ rather than $M_4 \times M_{p-4}$ we would even for the choice $M_4 \times G/H$ have the problem of too many equations. Therefore the present construction cannot replace the usual KK picture. (Actually there is one more point supporting this assertion to which we will come shortly.)

One could envisage a full dynamical setting as follows. The space V_4 may be looked upon as a solution of Einstein's equations,

$$\overline{G}_{ij} = T_{ij}. \tag{3.5}$$

Consistency with (3.4) requires

$$T_{ij} = t_{ij}(b). \tag{3.6}$$

Optimistically, one would look upon (3.6) as giving a geometrical meaning to the energy-momentum tensor. In addition to (3.5) and (3.6), we would have the GCR equation. One now feels that inasmuch as a dynamical meaning can be attributed to the quantities A_i and b_{ij}, physical reality can also be attributed to the extra dimensions provided by the embedding procedure. In fact, it is certainly clear that the b_{ij} and A_i in general do not vanish. For the b_{ij} this follows from the nonvanishing of $\overline{R}^{i}{}_{jke}$ for curved spaces.[14]

We conclude this section by discussing the solution of the GCR equations in the limit of vanishing (four-dimensional) gravity, $\bar{g}_{ij} = \eta_{ij}$. This is to be understood asymptotically; if $\bar{g}_{ij}$ would be exactly equal to η_{ij} in a finite region then the (minimal) embedding space would be just M_4. We will see that in the limit $\bar{g}_{ij} \to \eta_{ij}$, A_i and b_{ij} disappear. In this sense A_i is a "gravity-induced" gauge field. Note that its disappearance in the limit of vanishing gravity is a phenomenon that has no parallel in the conventional KK picture.

To prove our assertion notice first that the Gauss equation (2.16) implies for vanishing gravity, $\bar{g}_{ij} = \eta_{ij}$,

$$\mathrm{tr}(b_{jk}b_{lm}) = \mathrm{tr}(b_{jl}b_{km}). \tag{3.7}$$

If $\eta_{AB} \sim \delta_{AB}$ (and only for this case our proof holds true), then (3.9) implies that all b_{ij} are equal. The Ricci equation (2.18) then implies

$$F_{ij} = [D_i, D_j] = 0, \tag{3.8}$$

whence we conclude that the gauge field A_i vanishes (up to a gauge transformation).

Actually, Eq. (3.9) also implies that all b_{ij} are equal in any Lorentz frame. This can happen only for $b_{ij} = 0$. Thus our assertion is proved.

IV. CONCLUSIONS

The present paper may be looked upon as an investigation of the question whether the extra dimensions provided by the embedding formalism can be given physical relevance.

At the end of this paper we feel that the answer should be affirmative, even though many difficult questions remain open. Our reasons to believe that physical reality might be attributed to the extra dimensions are the following: (1) the appearance of the fields b_{ij}, A_i which appear to be coupled to physical gravity and thus make the extra dimensions visible indirectly; and (2) the existence of a "combined symmetry" which seems to indicate a mixing of space-time with internal coordinates.[8]

Our analysis concerning the relation between the present picture and that of conventional Kaluza–Klein theories has produced a number of negative results. However we like to emphasize that a certain similarity (which to our knowledge has not been noted before) is there, and might still deserve further investigation. (See note added in proof.)

There have been occasional references in the literature concerning the use of embedding spaces, for example, for the construction of combined symmetrics[15] or for possible derivations of the equations of geometrodynamics.[16] Even though our paper is partially motivated by this work, we do feel that we ask, and partially answer, quite different questions.

Note added in proof: Some of these results were further investigated by one of us. It is now clear that the presence of the fields b_{ij} is responsible for the compactification of the coordinate space generated by y^A. Taking such space to be the internal space (instead of $\mathcal{N}$), dimensional reduction can be performed in the usual way and a picture closer to that of Kaluza–Klein theory emerges (M. D. Maia, "Geometry of Kaluza–Klein Theory," preprint, Universidade de Brasilia, Departamento de Matematica, 1984).

ACKNOWLEDGMENTS

On of us (W.M.) wishes to thank the other (M.D.M.) for the invitation to Brasilia. He also wishes to express his gratitude to all the members of the Departamento de Matemática for the hospitality extended to him.

We wish to thank Senhor José Pereira dos Santos (Garibaldo) for the speedy typing of the manuscript. The visit of W. M. to Brasilia was made possible through grants from the GMD and the CNPq.

[1] E. Witten, Nucl. Phys. B **195**, 481 (1982).
[2] A. Salam and J. Strathdee, Ann. Phys. (NY) **141**, 316 (1982).
[3] Y. M. Cho and P. G. O. Freund, Phys. Rev. D **12**, 1711 (1975).
[4] L. N. Chang, K. I. Macrae, and F. Mansouri, Phys. Rev. D **13**, 235 (1976).
[5] E. Cremmer and J. Scherk, Nucl. Phys. B **103**, 399 (1976); Nucl. Phys. B **108**, 409 (1976).
[6] R. Casalbuoni, G. Domokos, and S. Kövesi-Domokos, Phys. Rev. D **17**, 2048 (1978).
[7] A. Trautmann, Rep. Math. Phys. **16**, 2026 (1975).
[8] M. D. Maia, Trieste preprint IC/82/116; "Combined symmetries in curved space-times," J. Math. Phys. **25**, 2090 (1984).
[9] L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, NJ, 1966), p. 143 ff.
[10] Hubert F. Goenner, "Local isometric embeddings of Riemannian manifolds and Einstein's theory of gravitation," in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980).
[11] Robert E. Greene, "*Isometric Embeddings of Riemannian and Pseudo-Riemannian Manifolds*," *Memoirs of the American Mathematical Society* (AMS, Providence, RI, 1970), No. 97.
[12] A typical theorem is the following (cf. Ref. 11, p. 49). Let U be an open set in V_d and $u \in U$. Let g be a C^∞ pseudo-Riemannian metric defined on U. Suppose that g has signature (p_1, n_1) at u, $(p_1 + n_1) = d$. Then there is an open set V, with $u \in V$ such that there is an embedding of V in some $V_{p,n}$, $n + p = \frac{1}{2}d(d+3)$ which is isometric for $g|_V$. If g is Riemannian at u then n may be taken equal to 0. More generally, n and p may be chosen subject only to the restriction $n \geqslant n_1$, $p \geqslant p_1$, $n + p = \frac{1}{2}d(d+3)$.
[13] In more detail, the GCR equations read

(G) $\bar{R}^{i}{}_{jkl} = 2\bar{g}^{im}\eta^{AB}b_{j[kA}b_{l]mB}$,

(C) $b_{i[kA,j]} + \bar{\Gamma}^{m}_{i[k}b_{mj]A} = \eta^{CD}b_{i[jD}A_{k]AC}$,

(R) $A_{[jAB,i]} + \bar{g}^{mn}b_{m[jA}b_{i]nB} + \eta^{DE}A_{[jDB}A_{i]AE} = 0$.

Square brackets denote antisymmetrization for the indices next to it.
[14] As a somewhat singular example consider the case of a five-dimensional embedding space. In this case, V_4 has constant curvature and $\bar{R}^{i}{}_{jke} \sim (\bar{g}_{jk}\bar{g}^{i}{}_{e} - \bar{g}_{je}\bar{g}_{k}{}^{i})$. The quantities A_{iAB} vanish because of antisymmetry in the indices A and B. Only one b_{ijA} is left, $b_{ij} \equiv b_{ij5}$. On account of the Gauss equation, $b_{ij} \sim \bar{g}_{ij}$. Because of the Ricci lemma this solves also the Codazzi equation, and the Ricci equation is trivially satisfied. We also notice that $t_{ij}(b) \sim \bar{g}_{ij}$ [cf. Eq. (3.5)]. This corresponds to the fact that a constant curvature V_4 is just (anti-) de Sitter space, i.e., a solution of the vacuum-Einstein equations with cosmological constant.
[15] C. Fronsdal, Nuovo Cimento **13**, 2560 (1959); Y. Ne'eman and J. Rosen, Ann. Phys. (NY) **31**, 391 (1965); N. Ne'eman, Rev. Mod. Phys. **37**, 227 (1965).
[16] S. Gallone, Nuovo Cimento B **49**, 149 (1979).

Irreversible random and cooperative processes on lattices: Spatial correlations

J. W. Evans, D. R. Burgess,[a] and D. K. Hoffman
Ames Laboratory and Department of Chemistry, Iowa State University, Ames, Iowa 50011

(Received 31 October 1983; accepted for publication 11 May 1984)

For processes where "filling" events occur irreversibly and, in general, cooperatively at the sites of a lattice, the minimal closed hierarchy of rate equations involves only probabilities for (effectively) connected subconfigurations of empty sites. Extended hierarchies of equations for (effectively) disconnected empty subconfigurations couple back to these. Here we consider a solution to the latter via previously developed exact and approximate truncation schemes based on a shielding property of empty sites. Numerical results for several processes are presented for correlation behavior in both autocatalytic and autoinhibitory rate regimes. The asymptotic large separation behavior of the spatial correlations is analyzed most easily by z-transforming the equations with respect to separations and is fundamentally different from that of equilibrium distributions.

PACS numbers: 05.50. + q, 05.70.Ln, 05.40. + j

I. INTRODUCTION

There are many processes in physics and chemistry where events occur irreversibly (on the time scale of interest) "filling" the sites of a lattice (the terminology of adsorption is adopted here for unification). These could be reactions on [one-dimensional (1-D)] polymer chains, chemisorption or reaction of attached groups at specific [two-dimensional (2-D)] surface sites, or solid-state reactions at the sites of a [three-dimensional (3-D)] crystalline structure. In general, these processes are cooperative, i.e., the "filling" rates (which are the input to our model) depend on the occupancy state of sites in the neighborhood of those being filled.

We have previously analyzed the hierarchial form of the rate equations which govern the time evolution of probabilities of configurations of the subsets of sites.[1] The minimal closed hierarchy involves only "effectively connected" empty subconfigurations (which include single empty sites and configurations such that any site specified empty is within the cooperative range of some other such site). "Effectively connected" reduces to "connected" in the usual sense for nearest-neighbor (NN) cooperative effects. Exact and approximate truncation and solution techniques for this minimal hierarchy have been developed previously motivated by the following observation[1]: Suppose a wall of sites specified empty separates the lattice into disconnected regions and is sufficiently thick that an event is not simultaneously affected by the state of sites on both sides; then this wall shields sites on one side from the effect of those on the other.

Several workers[2-6] have previously noted (for some simple 1-D systems) that determination of probabilities of connected empty configurations does not allow one to calculate probabilities of clusters of >2 filled sites or of (separated) two-point correlations. For these, it is necessary to consider extended hierarchies for probabilities of "effectively disconnected" empty subconfigurations. The only such previous analyses have been for simple processes on infinite, uniform 1-D lattices. Plate *et al.*[5] considered filling of single sites with NN cooperative effects and calculated probabilities of several smaller disconnected empty configurations and filled clusters. Wolf[6] has given an elegant expression for the two-point correlation function for random dimer filling.

In Sec. II, we apply the above-mentioned exact and approximate truncation based methods of solution to the "effectively disconnected" empty hierarchies. In this work, we concentrate on determining various two-cluster correlations rather than probabilities for clusters of filled sites. We indicate that when using standard approximate truncation techniques, there is a fundamental restriction on the range of separations for which reliable results can be obtained. A modified truncation scheme appropriate to the determination of longer-range correlations is indicated. Several irreversible processes are considered including reaction on a 1-D, infinite, uniform polymer chain with NN cooperative effects, and with NN blocking and second NN cooperative effects; random dimer filling of NN sites on a 1-D infinite, uniform lattice; both monomer adsorption with NN cooperative effects and random dimer filling of NN sites on a 2-D infinite, uniform square lattice.

The large separation asymptotic structure of the correlations is of particular interest. Consequently in Sec. III and IV we develop z-transform techniques to elucidate this behavior (analogous to the treatment in Ref. 6 of edge effects for a 1-D semi-infinite system). Here equations for correlations are transformed with respect to the separation between clusters. Transformed quantities could be thought of as generating functions. For a system amenable to exact truncation and solution, the exact "fast decay" asymptotic form of the two-cluster correlations is obtained. The z-transform method suggests this behavior is characteristic of all irreversible processes. Results should be compared with the fundamentally slower "exponential decay" of equilibrium distributions. A brief comparison of corresponding (exactly determined) 1-D and (approximately determined) 2-D behavior for two processes is made in Sec. V.

The following notation is used here. The probabilities of configurations σ of subsets of sites specified either empty (unreacted) "o" or filled (reacted) "a" are denoted by $f[\sigma]$. Thus we can decompose σ as $\{n\}_a + \{m\}_o$, where $\{n\}$ ($\{m\}$)

[a] Present address: Marycrest College, Davenport, Iowa 52804.

 0022-2488/84/103051-13$02.50

represents the subset of $n(m)$ sites specified "a" ("o"). Consider now a processes where filling $o \to a$ occurs at single sites j and the rate of filling, denoted τ^j_σ, depends on the configuration σ^j of the influencing environment (assumed finite here). If σ equals $\{m\}_o$, then intuitively one has[1]

$$\frac{d}{dt} f[\{m\}_o] = -\sum_{j \in \{m\}} \sum_{\sigma^j_*} \tau_\sigma f[\{m\}_o + \sigma^j_*]. \tag{1.1}$$

Here σ^j_* denotes the part of σ^j not overlapping $\{m\}$ and the sum over σ^j_* is required to take account of all possible configurations of the influencing environment of site j. Here probabilities on the right-hand side are converted to those involving only empty sites using

$$f[\{l\}_a + \{p\}_o] = \sum_{\{l'\} \subset \{l\}} (-1)^{l'} f[\{l'\}_o + \{p\}_o]. \tag{1.2}$$

Equations (1.1) and (1.2) or the corresponding equations for irreversible dimer filling of NN sites to $oo \to aa$ are used for several specific processes here. They incorporate the above-mentioned connectivity and shielding features. Often we will refer to the conditional probability $q[\sigma\,\overline{\sigma'}] = f[\sigma + \sigma']/f[\sigma']$ of σ given σ'. Empty conditioning sites $\bar{o}$ will often be denoted by "ϕ" for typographic convenience and $\theta \equiv f[a]$ will denote the coverage (conversion).

II. SOLUTION VIA TRUNCATION OF (EFFECTIVELY) DISCONNECTED HIERARCHIES

A. Exact truncation and solution

Exact truncation and solution in closed form of the infinite hierarchies is only possible for certain processes on 1-D[7] or Bethe[8] lattices (except for random or "almost" random filling in higher dimensions[1,9]). The most general process $o \to a$ at single sites on 1-D infinite, uniform lattices for which the minimal hierarchy is exactly solvable is where events are blocked by filled sites within a range r and influenced by sites within a range $2r + 1$.[6,7] It is also possible to exactly solve the effectively disconnected hierarchies for such processes. We do not discuss the general case here but give two examples which demonstrate the basic features involved. The terms "empty" and "unreacted" will be used interchangeably here (as will "filled" and "reacted").

Example (*i*): Reaction on an initially completely unreacted infinite, uniform polymer chain; NN cooperative effects.

The rates of reaction $o \to a$ for sites with 0, 1 left (right) hand and 2 reacted NN are denoted by $\tau_{o.o}$, $\tau_{a.o}(\tau_{o.a})$, and $\tau_{a.a}$, respectively. If $\mathbf{o}_n$ denotes an n-tuple of empty sites, the minimal closed subhierarchy becomes

$$\begin{aligned} -\frac{d}{dt} f[o] &= \tau_{o.a} f[ooo] + \tau_{a.a} f[aoo] \\ &\quad + \tau_{o.a} f[ooa] + \tau_{a.a} f[aoa] \\ &= \tau_{a.a} f[o] + (\tau_{a.o} + \tau_{o.a} - 2\tau_{a.a}) f[oo] \\ &\quad + (\tau_{o.o} - \tau_{a.o} - \tau_{o.a} + \tau_{a.a}) f[ooo], \end{aligned} \tag{2.1a}$$

$$\begin{aligned} -\frac{d}{dt} f[\mathbf{o}_n] &= (n-2)\tau_{o.a} f[\mathbf{o}_n] + 2\tau_{o.a} f[\mathbf{o}_{n+1}] \\ &\quad + \tau_{a.a} f[a\mathbf{o}_n] + \tau_{o.a} f[\mathbf{o}_n a] \\ &= \{(n-2)\tau_{o.o} + \tau_{a.o} + \tau_{o.a}\} f[\mathbf{o}_n] \\ &\quad + (2\tau_{o.o} - \tau_{a.o} - \tau_{o.a}) f[\mathbf{o}_{n+1}], \\ n &\geqslant 2. \end{aligned} \tag{2.1b}$$

One can also obtain, from (1.1) and (1.2), an infinite closed subhierarchy for probabilities of singly disconnected empty configurations $\mathbf{o}_p \overset{m}{\text{---}} \mathbf{o}_r$ of an empty p-tuple and r-tuple separated by "m" unspecified sites. Specifically, for $l \geqslant 1$; $r \geqslant 2$, one has

$$\begin{aligned} -\frac{d}{dt} f\left[o \overset{l}{\text{---}} o\right] &= \tau_{o.a} f\left[ooo \overset{l-1}{\text{---}} o\right] \\ &\quad + \tau_{a.a} f\left[aoo \overset{l-1}{\text{---}} o\right] + \tau_{o.a} f\left[ooa \overset{l-1}{\text{---}} o\right] \\ &\quad + \tau_{a.a} f\left[aoa \overset{l-1}{\text{---}} o\right] + \text{reflected terms} \\ &= \tau_{a.a} f\left[o \overset{l}{\text{---}} o\right] + (\tau_{a.o} - \tau_{a.a}) f\left[oo \overset{l-1}{\text{---}} o\right] \\ &\quad + (\tau_{o.a} - \tau_{a.a}) f\left[oo \overset{l}{\text{---}} o\right] \\ &\quad + (\tau_{o.o} - \tau_{o.a} - \tau_{a.o} + \tau_{a.a}) \\ &\quad \times f\left[ooo \overset{l-1}{\text{---}} o\right] + \text{reflected terms}, \end{aligned} \tag{2.2a}$$

$$\begin{aligned} -\frac{d}{dt} f\left[o \overset{l}{\text{---}} \mathbf{o}_r\right] &= \tau_{a.a} f\left[o \overset{l}{\text{---}} \mathbf{o}_r\right] + (\tau_{a.o} - \tau_{a.a}) f\left[oo \overset{l-1}{\text{---}} \mathbf{o}_r\right] \\ &\quad + (\tau_{o.a} - \tau_{a.a}) f\left[oo \overset{l}{\text{---}} \mathbf{o}_r\right] \\ &\quad + (\tau_{o.o} - \tau_{o.a} - \tau_{a.o} + \tau_{a.a}) \\ &\quad \times f\left[ooo \overset{l-1}{\text{---}} \mathbf{o}_r\right] + \{(r-2)\tau_{o.o} + \tau_{a.o} + \tau_{o.a}\} \\ &\quad \times f\left[o \overset{l}{\text{---}} \mathbf{o}_r\right] + (\tau_{o.o} - \tau_{a.o}) f\left[o \overset{l-1}{\text{---}} \mathbf{o}_{r+1}\right] \\ &\quad + (\tau_{o.o} - \tau_{o.a}) f\left[o \overset{l}{\text{---}} \mathbf{o}_{r+1}\right]. \end{aligned} \tag{2.2b}$$

The quantity $-d/dt f\left[\mathbf{o}_r \overset{l}{\text{---}} o\right]$ is obtained by reflecting terms on the right-hand side of (2.2b). Furthermore, for $l \geqslant 1$; $p, r \geqslant 2$, one has

$$\begin{aligned} -\frac{d}{dt} f\left[\mathbf{o}_p \overset{l}{\text{---}} \mathbf{o}_r\right] &= \{(p+r-4)\tau_{o.o} + 2\tau_{a.o} + 2\tau_{o.a}\} f\left[\mathbf{o}_p \overset{l}{\text{---}} \mathbf{o}_r\right] \\ &\quad + (\tau_{o.o} - \tau_{a.o})\left(f\left[\mathbf{o}_{p+1} \overset{l}{\text{---}} \mathbf{o}_r\right] + f\left[\mathbf{o}_p \overset{l-1}{\text{---}} \mathbf{o}_{r+1}\right]\right) \\ &\quad + (\tau_{o.o} - \tau_{o.a})\left(f\left[\mathbf{o}_{p+1} \overset{l-1}{\text{---}} \mathbf{o}_r\right] + f\left[\mathbf{o}_p \overset{l}{\text{---}} \mathbf{o}_{r+1}\right]\right). \end{aligned} \tag{2.2c}$$

The form of (1.1) and (1.2) for probabilities of multiply disconnected empty configurations can also be made explicit.

Now we invoke the shielding property of an adjacent pair of empty sites to exactly solve (2.1) and (2.2). From (2.1b) one can readily show that, if ϕ_n denotes an n-tuple of empty conditioning sites, then one has[5,10]

$$q[o\phi_n] \equiv q[\phi_n o] \equiv f[o_{n+1}]/f[o_n]$$
$$\equiv q[o\phi\phi] \equiv q[\phi\phi o] \equiv \exp(-\tau_{o.o}t)$$
$$(\equiv q_2, \text{ say}), \quad \text{for } n \geqslant 2. \tag{2.3}$$

This allows exact truncation of (2.1) using $f[o_n] = f[oo]q_2^{n-2}$ for $n \geqslant 2$, and results in a closed set of equations for $f[o], f[oo]$, and q_2.[5,10] Similarly (2.2) implies that (cf. Refs. 1 and 7)

$$q\left[o\phi_n \overset{l}{\text{——}} \phi_m\right] \equiv f\left[o_{n+1} \overset{l}{\text{——}} o_m\right]/f\left[o_n \overset{l}{\text{——}} o_m\right]$$
$$\equiv q_2 \equiv f\left[o_m \overset{l}{\text{——}} o_{n+1}\right]/f\left[o_m \overset{l}{\text{——}} o_n\right]$$
$$\equiv q\left[\phi_m \overset{l}{\text{——}} \phi_n o\right], \quad \text{for } n \geqslant 2 \text{ and all } l, m. \tag{2.4}$$

This allows truncation of (2.2) using

$$f\left[o \overset{l}{\text{——}} o_r\right] = f\left[o \overset{l}{\text{——}} o_{r-1}\right] q\left[\phi \overset{l}{\text{——}} \phi_{r-1} o\right]$$
$$= f\left[o \overset{l}{\text{——}} o_{r-1}\right] q_2 = f\left[o \overset{l}{\text{——}} oo\right] q_2^{r-2},$$
$$f\left[o_r \overset{l}{\text{——}} o\right] = f\left[oo \overset{l}{\text{——}} o\right] q_2^{r-2},$$
$$f\left[o_r \overset{l}{\text{——}} o_r\right] = q_2^{p+r-4} f\left[oo \overset{l}{\text{——}} oo\right], \quad \text{for } p, r \geqslant 2. \tag{2.5}$$

From (2.2) and (2.5) one can immediately obtain a closed set of equations for $f[o\text{-}o], f[o\text{-}oo], f[oo\text{-}o]$, and $f[oo\text{-}oo]$ (where - represents a single unspecified site). Similarly, equations for corresponding f's with a larger separation couple back to those with (one) smaller and can therefore be integrated. Plate *et al.*[5] have previously recognized this structure for the case where $\tau_{a.o} = \tau_{o.a}$ and consequently one has $f\left[oo \overset{m}{\text{——}} o\right] = f\left[o \overset{m}{\text{——}} oo\right]$. However, they did not provide a systematic analysis of correlation behavior or comparison with that of equilibrium distributions as given here. When $\tau_{a.o} = \tau_{o.a}$ the appropriate equations are[5]

$$-\frac{d}{dt} f\left[o \overset{l}{\text{——}} o\right] = 2\tau_{a.a} f\left[o \overset{l}{\text{——}} o\right]$$
$$+ 2(\tau_{o.a} - \tau_{a.a}) f\left[oo \overset{l}{\text{——}} o\right] + 2\{(\tau_{o.a} - \tau_{a.a}) + (\tau_{o.o} - 2\tau_{o.a} + \tau_{a.a})q_2\} f\left[oo \overset{l-1}{\text{——}} o\right], \tag{2.6a}$$

$$-\frac{d}{dt} f\left[oo \overset{l}{\text{——}} o\right] = \{(2\tau_{o.a} + \tau_{a.a}) + (\tau_{a.a} - \tau_{o.a})q_2\} f\left[oo \overset{l}{\text{——}} o\right]$$
$$+ (\tau_{o.a} - \tau_{a.a}) f\left[oo \overset{l}{\text{——}} oo\right] + (\tau_{o.o} - \tau_{o.a}) q_2 f\left[oo \overset{l-1}{\text{——}} o\right]$$
$$+ \{(\tau_{o.a} - \tau_{a.a}) + (\tau_{o.o} - 2\tau_{o.a} + \tau_{a.a})q_2\} f\left[oo \overset{l-1}{\text{——}} oo\right], \tag{2.6b}$$

$$-\frac{d}{dt} f\left[oo \overset{l}{\text{——}} oo\right]$$
$$= \{4\tau_{o.a} + 2(\tau_{o.o} - \tau_{o.a})q_2\} f\left[oo \overset{l}{\text{——}} oo\right]$$
$$+ 2(\tau_{o.o} - \tau_{o.a}) q_2 f\left[oo \overset{l-1}{\text{——}} oo\right]. \tag{2.6c}$$

Quantities of prime interest are the two-point correlation functions

$$c\left[o \overset{m}{\text{——}} o\right] \equiv f\left[o \overset{m}{\text{——}} o\right] - f[o]^2$$
$$= f\left[a \overset{m}{\text{——}} a\right] - f[a]^2 \equiv c\left[a \overset{m}{\text{——}} a\right]. \tag{2.7}$$

The identity which is the middle equality of (2.7), is proved using only conservation of probability[1] and thus holds for any distribution. All results given here will correspond to a choice of rates $\tau_{o.a}/\tau_{o.o} = \tau_{a.o}/\tau_{o.o} = \tau_{a.a}/\tau_{o.a} \equiv \alpha$. (This relationship is valid for an Arrhenius form of the rates with pairwise additive activation energies.) In Fig. 1 (Fig. 2) $c\left[o \overset{m}{\text{——}} o\right]$ is plotted as a function of "m" for an autocatalytic (autoinhibitory) choice of rates. These are compared with equilibrium (Ising model) lattice–gas correlations where, for correspondence, we choose the Ising model pairwise interaction J to satisfy $e^{-\beta J} = \alpha$, where $\beta = (kT)^{-1}$. The structure of equilibrium correlations is reviewed in Appendix A. The behavior of $c\left[oo \overset{m}{\text{——}} o\right] \equiv f\left[oo \overset{m}{\text{——}} o\right] - f[oo]f[o]$ and $c\left[oo \overset{m}{\text{——}} oo\right] \equiv f\left[oo \overset{m}{\text{——}} oo\right] - f[oo]^2$ regarded as functions of "m" is similar and thus not shown. The large separation (m) decay of the irreversible correlations is fundamentally faster than that of the corresponding equilibrium (Ising model) correlations as shown in detail in Sec. III. This is most apparent for the coverage $\theta \equiv f[a] = \frac{1}{2}$ in the regime of highly autoinhibitory rates (corresponding to highly repulsive NN Ising model interactions), where the correlation length[11] for the equilibrium distribution approaches infinity

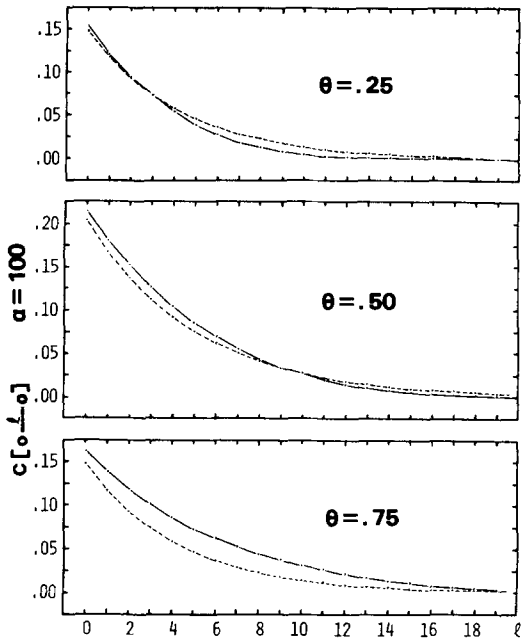


FIG. 1. $c\left[\overset{l}{o\text{—}o}\right] \equiv f\left[\overset{l}{o\text{—}o}\right] - f[o]^2$ for reaction $o \to a$ on an infinite, uniform polymer chain; NN cooperative effects with rates $\tau_{o\cdot o} : \tau_{o\cdot a} : \tau_{a\cdot o} : \tau_{a\cdot a} = 1 : \alpha : \alpha : \alpha^2$ and $\alpha = 100$ [the dashed line shows the corresponding Ising model correlation with pairwise interaction J satisfying $e^{-\beta J} = \alpha, \beta = (kT)^{-1}$].

but the irreversible distribution $c\left[\overset{m}{o\text{——}o}\right]$ has fast decay behavior (see Fig. 3). In Fig. 4, we plot $c[o\text{-}o]$ as a function of θ for various α. Notice the asymmetry about $\theta = \frac{1}{2}$ in contrast to the equilibrium theory.

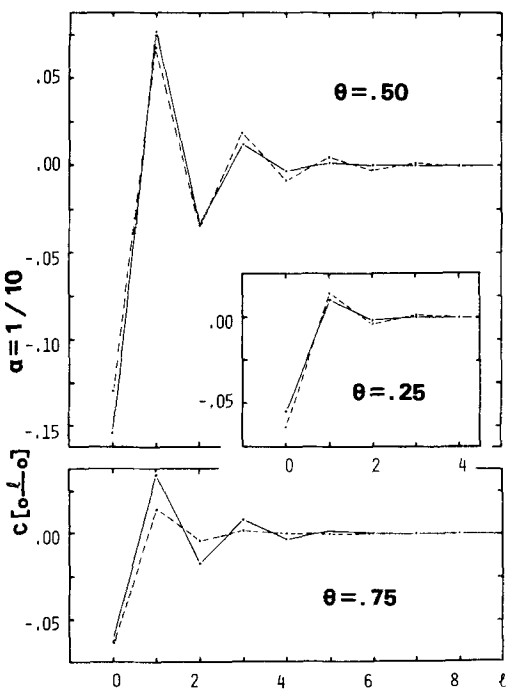


FIG. 2. $c\left[\overset{l}{o\text{—}o}\right]$ for the process of Fig. 1 with $\alpha = \frac{1}{10}$ (the dashed line shows corresponding Ising correlations).

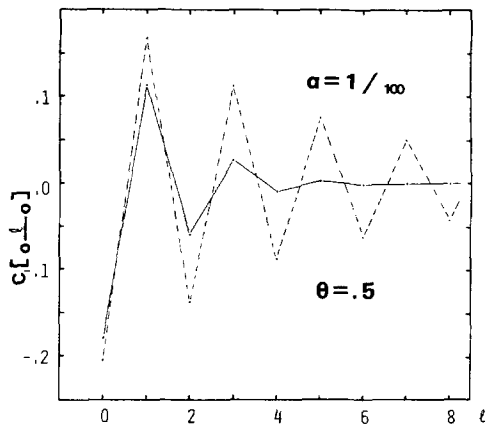


FIG. 3. $c\left[\overset{l}{o\text{—}o}\right]$ for the process of Fig. 1 with $\alpha = \frac{1}{100}$ (the dashed line shows corresponding Ising correlations).

Probabilities of multiply disconnected empty configurations can also be obtained exactly using shielding to truncate the corresponding hierarchial equations. The behavior of $f\left[oo\overset{l}{\text{——}}o\overset{k}{\text{——}}oo\right]$ is analyzed in Sec. IV. Sometimes these quantities can be directly expressed in terms of simpler ones using shielding. For example, one has

$$f\left[\mathbf{o}_m\overset{l}{\text{——}}\mathbf{o}_n\overset{k}{\text{——}}\mathbf{o}_p\right] = \frac{q_2^{n-2} f\left[\mathbf{o}_m\overset{l}{\text{——}}oo\right] f\left[oo\overset{k}{\text{——}}\mathbf{o}_p\right]}{f[oo]}, \quad \text{for } n \geqslant 2, \tag{2.8a}$$

which can be further simplified using (2.5) if m or $p \geqslant 2$, and

$$f[ooo\text{--}o\text{-}ooo\text{---}oo\text{-}o] = \frac{q_2^2 f[oo\text{--}o\text{-}oo] f[oo\text{---}oo] f[oo\text{-}o]}{f[oo]^2}. \tag{2.8b}$$

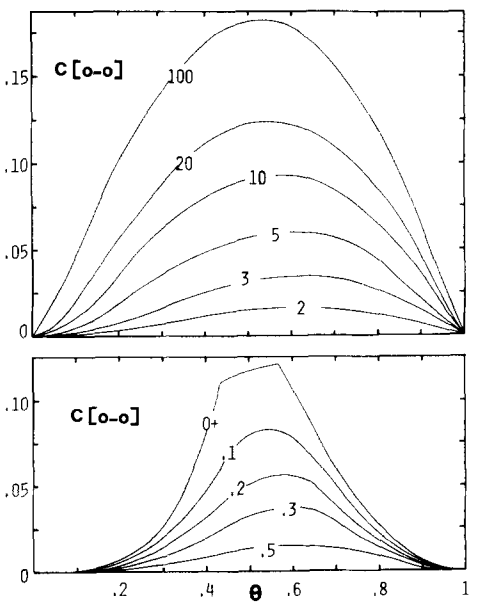


FIG. 4. $c[o\text{-}o]$ as a function of coverage θ for the process of Fig. 1 for various α (shown).

It should be clear that, using shielding, the probability of any disconnected empty configuration can be written in terms of $f[oo]$, q_2 and probabilities of the type

$$f\left[\sigma\overset{l}{\text{—}}o\overset{m}{\text{—}}o\text{—}\cdots\text{—}o\overset{r}{\text{—}}\sigma'\right], \text{ where } \sigma,\ \sigma' = o \text{ or } oo.$$

Example (ii): Reaction of an initially completely unreacted infinite, uniform polymer chain; NN blocking and second NN cooperative effects.

Because of the NN blocking, the minimal closed hierarchy here involves only n-tuples of empty sites[12,13] (cf. general range two cooperative effects where it also includes certain disconnected configurations[13]). The shielding property of a 4-tuple of empty sites guarantees that[1,12,13]

$$q[\phi_n o] \equiv q[o\phi_n] \equiv q[o\phi\phi\phi\phi] \equiv q[\phi\phi\phi\phi o] = \exp(-\tau_{oo.oo}t)\ (\equiv q_4, \text{ say}), \quad \text{for } n \geqslant 4, \tag{2.8}$$

where $\tau_{oo.oo}$ is the rate for reaction with all first and second NN unreacted. Thus the minimal hierarchy can be exactly truncated using (2.8) to obtain a finite closed set of equations for $f[o], f[oo], f[ooo], f[oooo]$, and q_4.[12,13] Similarly, using[1]

$$q\left[o\phi_n\overset{l}{\text{—}}\phi_m\right] \equiv q_4 \equiv q\left[\phi_m\overset{l}{\text{—}}\phi_n o\right], \quad \text{for } n \geqslant 4, \tag{2.9}$$

one can obtain a closed set of equations for $f[o\text{-}o], f[o\text{-}oo]$, $f[o\text{-}ooo]$, $f[o\text{-}oooo]$, $f[oo\text{-}oo]$, $f[oo\text{-}ooo], f[oo\text{-}oooo]$, $f[ooo\text{-}ooo], f[ooo\text{-}oooo], f[oooo\text{-}oooo]$ (as well as the corresponding reflected quantities if the rates are not reflection invariant). Equations of the corresponding f's for larger separations couple back to those for smaller separations (after truncation using shielding) and thus may be integrated exactly. In Fig. 5, $c\left[o\overset{m}{\text{—}}o\right]$ is plotted as a function of m for the range two blocking case.

Again, probabilities of multiply disconnected empty configurations can also be obtained exactly and sometimes

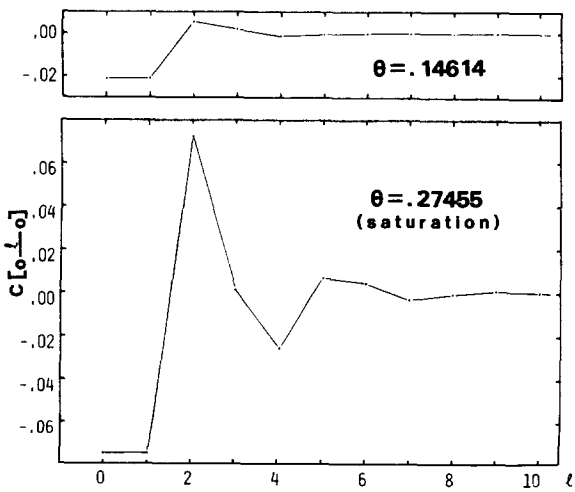


FIG. 5. $c\left[o\overset{l}{\text{—}}o\right]$ for reaction $o \to a$ on an infinite, uniform polymer chain with range two blocking (i.e., reacted first and second NN block reaction).

these can be reduced in terms of simpler quantities using shielding. For example, one has

$$f[ooooo\text{-}oooo\text{-}o] = \frac{q_4 f[oooo\text{-}oooo]f[oooo\text{-}o]}{f[oooo]},$$
$$f[o\text{-}ooooo\text{--}oooo\text{-}o] = \frac{q_4 f[o\text{-}oooo]f[oooo\text{--}oooo]f[oooo\text{-}o]}{f[oooo]^2}. \tag{2.10}$$

These identities also hold for a process with arbitrary range two cooperative effects.[13]

Example (iii): Random dimer filling of NN sites on a 1-D infinite, uniform lattice with rate k.

The minimal closed hierarchy here is given by[14,15]

$$-\frac{d}{dt}f[\mathbf{o}_n] = (n-1)kf[\mathbf{o}_n] + 2kf[\mathbf{o}_{n+1}], \quad n \geqslant 1. \tag{2.11}$$

The shielding property of a single empty site[1] guarantees that[14,15]

$$q[o\phi_n] \equiv e^{-kt}\ (\equiv q_1, \text{ say}), \quad n \geqslant 1, \tag{2.12}$$

which allows truncation and solution of (2.11) to obtain $f[o] = \exp(2q_1 - 2)$. Wolf[6] has analyzed the singly disconnected hierarchial equations

$$-\frac{d}{dt}f\left[o\overset{l}{\text{—}}o\right] = 2k\left\{f\left[oo\overset{l-1}{\text{—}}o\right] + f\left[oo\overset{l}{\text{—}}o\right]\right\}, \quad \text{for } l \geqslant 1, \tag{2.13}$$

which after using the shielding property

$$q\left[o\phi\overset{m}{\text{—}}\phi\right] \equiv q\left[\phi\overset{m}{\text{—}}\phi o\right] \equiv q_1$$

can be written as

$$-\frac{d}{dt}f\left[o\overset{l}{\text{—}}o\right] = 2kq_1\left\{f\left[o\overset{l-1}{\text{—}}o\right] + f\left[o\overset{l}{\text{—}}o\right]\right\}. \tag{2.14}$$

Noting that $f[oo] = q_1 f[o]$, one can show that[6]

$$f\left[o\overset{l}{\text{—}}o\right] = f[o]\left\{\sum_{k=0}^{l}\frac{x^k}{k!} + \frac{\frac{1}{2}x^{l+1}}{(l+1)!}\right\}, \tag{2.15a}$$

so

$$c\left[o\overset{l}{\text{—}}o\right] = f\left[o\overset{l}{\text{—}}o\right] - f[o]^2 = -f[o]\left\{\frac{\frac{1}{2}x^{l+1}}{(l+1)!} + \sum_{k=l+2}^{\infty}\frac{x^k}{k!}\right\}, \tag{2.15b}$$

where $x = \ln f[o]$. This result can be obtained simply using the techniques described in the next section. Note that at saturation $t = \infty$, $f[oo] = 0$ so $x = -2$ and also $f[o\text{-}o] \equiv f[ooo] = 0$. It is also true (but less obvious) that $f[o\text{---}o] = 0$ at saturation. In Table I, $c\left[o\overset{l}{\text{—}}o\right]$ is listed for $0 \leqslant l \leqslant 7$ at full, and approximately half, saturation.

This process has the unusual feature that, since a single site shields, all empty (and thus all) subconfiguration probabilities can be expressed in terms of $f[o]$, q_1 and the singly disconnected $f\left[o\overset{l}{\text{—}}o\right]$ [cf. (2.8) and (2.10)].

TABLE I. $c\left[o\overset{l}{\text{——}}o\right]\equiv c\left[a\overset{l}{\text{——}}a\right]$ for random dimer filling NN sites of an infinite, uniform 1-D lattice and 2-D square lattice.

	1-D			2-D	
	$\theta=$ 0.43	$\theta=1-e^{-2}$ ≈ 0.86467 (saturation)		$\theta=$ 0.45	$\theta\approx 0.9063$ (saturation)
l			l		
0	0.8490×10^{-1}	-0.1832×10^{-1}	0	0.3617×10^{-1}	-0.8772×10^{-2}
1	-0.3028×10^{-1}	-0.1832×10^{-1}	1	-0.6830×10^{-2}	-0.1977×10^{-2}
2	0.6309×10^{-2}	0.2680×10^{-1}	2	0.7308×10^{-3}	0.1431×10^{-2}
3	-0.9421×10^{-3}	-0.1832×10^{-1}	3	-0.5701×10^{-4}	-0.5647×10^{-3}
4	0.1102×10^{-3}	0.8751×10^{-2}	4	0.3507×10^{-5}	0.1556×10^{-3}
5	-0.1062×10^{-4}	-0.3278×10^{-2}	5	-0.1783×10^{-6}	-0.3375×10^{-4}
6	0.8702×10^{-6}	0.1018×10^{-2}	6	0.7733×10^{-8}	0.6096×10^{-5}
7	-0.6213×10^{-7}	-0.2709×10^{-3}	7	-0.2924×10^{-9}	-0.9473×10^{-6}

B. Approximate truncation and solution

Starting with (1.1) and (1.2), one can readily obtain an infinite, closed hierarchy for conditional probabilities where all conditioned (o-) and conditioning (ϕ-) sites are empty.[1] The approximate truncation techniques described in Ref. 1 operate directly on this hierarchy. We either factorize the q's in terms of those with a single o-site, then truncate ϕ-sites further than "n" times the cooperative range from this site (a scheme denoted by $FT.n$) or truncate ϕ-sites further than this distance from all o-sites, then factorize in terms of q's with a single o-site (denoted by $T.nF$). Consequently, these schemes are only appropriate for the calculation of "small separation" correlations, e.g.,

$$c[o^j\text{——}o^k]=f[o^k](q[o^j\text{——}\phi^k]-q[o^j]), \quad (2.16)$$

where o^j indicates that site j is empty, etc., and where k is within the truncation range of j. (Otherwise, truncation replaces $q[o^j\text{——}\phi^k]$ with $q[o^j]\equiv f[o^j]$ and thus $c[o^j\text{——}o^k]$ by zero.) Two processes are considered below. To determine larger separation two-cluster correlations, a more sophisticated truncation scheme must be developed where the truncated equations preserve the coupling structure with respect to separation. One such example is presented.

Example (*iv*): Monomer filling of an initially empty infinite, uniform, 2-D square lattice; NN cooperative effects with rotation invariance.

One quantity amenable to calculation from $FT.n$ and $T.nF$ truncation schemes (which truncate "ϕ" sites further than "n" lattice vectors from "o" sites) for $n\geqslant 2$ is $f[o\text{-}o]=q[o\text{-}\phi]f[o]$ which involves a singly disconnected configuration. We consider only the $FT.2$ and $T.3F$ cases.

For the $FT.2$ case, the minimal closed set of equations involves q's for the nine configurations[1]

$$o,\ o\phi\overset{o}{\phi}\phi,\ o\phi\phi,\ \phi\overset{o}{\phi}\phi,\ o\overset{\phi}{\phi}\phi,$$

$$\overset{\phi}{o}\phi\phi,\ \phi\overset{oo}{\phi},\ o\overset{\phi\phi}{\phi}\phi.$$

The truncated q-equations for $q[o\text{-}\phi]$ and $q[\phi o\phi]$ close with this set and equations for these eleven q's may be integrated to obtain $f[o\text{-}o]$ as a function of t or θ. Similarly, the $FT.2$ equations for $q[\overset{\phi}{o}\cdot]$, $q[\overset{\phi}{\phi}\overset{\phi}{o}]$, and $q[\overset{\phi\phi}{\phi}\overset{o}{\cdot}]$ close with the original set of nine, allowing calculation of $f[\overset{o}{o}\cdot]$. Several other disconnected f's may be similarly calculated.

It turns out that $f[o\text{-}o]$ can be immediately calculated from the 128 $T.3F$ equations.[1,16] Although this set was obtained from a connected q-hierarchy, truncation has the effect of introducing $q[o\,\overline{o\sigma}]$ for disconnected $o\sigma$ and σ. In particular, q's for configurations $\phi\text{-}\overset{o}{\phi}$, $\phi o\overset{\phi}{\phi}$, $\phi o\phi$, $o\phi\text{-}\phi$, and $\phi\phi o\phi$ are included.[16] Since one has

$$f[o\text{-}o]\equiv f[ooo]/q[\phi o\phi]\equiv f[oooo]/(q[o\phi\text{-}\phi]q[\phi\phi o\phi])$$

$$\equiv f[o\overset{o}{o}o]/(q[\phi\text{-}\overset{o}{\phi}]q[\phi o\overset{\phi}{\phi}]), \quad (2.17)$$

and $f[ooo]$, $f[oooo]$, $[o\overset{o}{o}o]$ can readily be calculated from the $T.3F$ equations, it follows that $f[o\text{-}o]$ can also be calculated ["product consistency" of q's for our truncation schemes[1] guarantees all three expressions of (2.17) give the same result]. The probabilities of several other disconnected configurations including $oo\text{-}o$, $oo\overset{o}{\cdot}$, $\overset{o}{o}\text{-}o$, and $o\overset{o}{\cdot}$ can be calculated directly from the $T.3F$ equations.

In Fig. 6, we plot $c[o\text{-}o]\equiv f[o\text{-}o]-f[o]^2$ for the case where the rate of adsorption with "i" filled NN is given by $k\alpha^i$; $i=0,1,\ldots,4$ (for various α). The $FT.2$ equations appear to give a reasonable estimate of $c[o\text{-}o]$ for $0.2\lesssim\alpha\lesssim 5$. However, the $T.3F$ equations are only good for $0.3\lesssim\alpha\lesssim 2$. Behavior should be compared with $c[o\text{-}o]$ in Fig. 4 for the analogous 1-D process.

Next we consider the series of processes where filling of sites with $\leqslant m$ occupied NN is random but of those with $>m$ is blocked ($m=0,1,2,3$). In Fig. 7, $c[o\text{-}o]$ for these processes is plotted as a function of θ and, in Table II, corresponding saturation values are given for $f[o\text{-}o]$. For $m=3$, i.e., random filling of all sites except those with four occupied NN, exact results can be obtained.[9]

Example (*v*): Random dimer filling of NN sites on an initially empty infinite, uniform 2-D square lattice with rate k.

Here we can obtain an infinite closed hierarchy for the $q[o\bar{\sigma}]$, where σ contains only empty sites and $o\sigma$ is connected.[15] Truncating "ϕ" sites further than two lattice vectors from "o" yields a minimal closed set of equations for q's with configurations o, $o\phi$, $o\overset{\phi}{\phi}$, $o\phi\phi$, $o\overset{\phi}{\phi}\phi$, $\phi\overset{o}{\phi}\phi$, $\phi\overset{o}{\underset{\phi}{\phi}}\phi$, and $\overset{o\phi}{\phi\phi}$.[15] In addition, in this approximation it is possible to obtain estimates of probabilities of a few smaller disconnected configu-

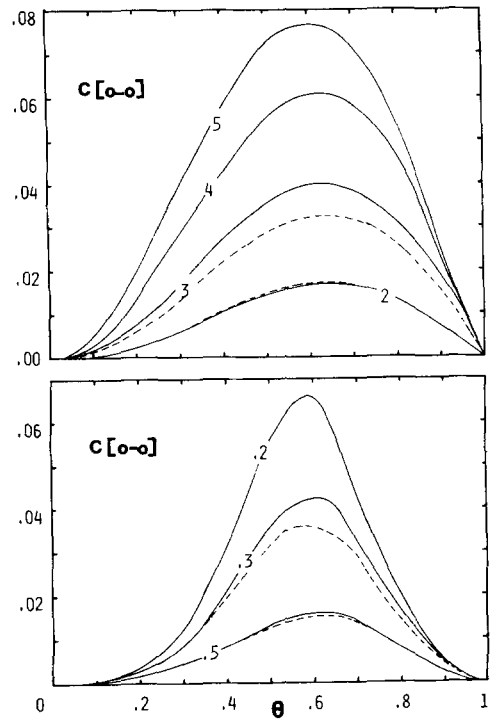


FIG. 6. $c[o\text{-}o]$ as a function of coverage θ for monomer filling $o \to a$ of sites on an infinite, uniform, 2-D square lattice where each filled NN changes the rate by a factor of α (shown); the solid (dashed) lines give $FT.2$ ($T.3F$) values.

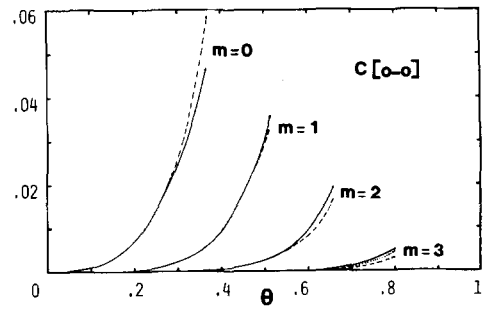


FIG. 7. $c[o\text{-}o]$ for random monomer filling $o \to a$ of sites on an infinite, uniform 2-D square lattice with $\leq m$ filled NN and blocking with $> m$; the solid (dashed) lines give $T.3F$ ($FT.2$) values and the dotted line for $m = 3$ gives the exact value.

rations, namely, $o\overset{o}{-}$, $o\text{-}o$, and $o\overset{o}{-}o$ since truncated equations for $q[{}_{o}\overset{\phi}{-}]$ and $q[{}_{\phi}\overset{\phi}{o}]$, in this approximation, close with those of the above eight q's and truncated equations for $q[o\overset{\phi}{-}\phi]$ and $q[\phi\overset{\phi}{o}\phi]$ close with this extended set of ten. Also, truncated equations for $q[o\text{-}\phi]$ and $q[\phi o\phi]$ close with the original set of eight. These f's may be calculated immediately from the 114 three lattice vector truncation equations.[17] In Fig. 8, $c[o\text{-}o] \equiv f[o\text{-}o] - f[o]^2$ and $c[{}_{o}\text{-}o] \equiv f[{}_{o}\text{-}o] - f[o]^2$ are plotted as functions of θ. The deviation of the third- from second-order is much more significant for these disconnected quantities than for previously examined, connected f's.[15,17] The third-order truncation predicts saturation values of 0.68×10^{-2} for $f[o\text{-}o]$, 0.83×10^{-2} for $f[{}_{o}\text{-}o]$ and 0.45×10^{-3} for $f[o\overset{o}{-}o]$. In Fig. 9, the three point correlation $c[o\overset{o}{-}o] = f[o\text{-}{}^{o}o] - 2f[o]f[o\text{-}{}^{o}] - f[o]f[o\text{-}o] + f[o]^3$ is plotted against θ. As might be expected, there is only qualitative similarity between second- and third-order truncations.

Finally we given an example of a crude technique for truncating and closing the hierarchy equations, for certain two-cluster correlations, which retains their coupling structure. In the next section we show this scheme guarantees the type of fast asymptotic decay seen in examples (i) and (ii) rather than exponential equilibrium-type decay. The equations for the correlations $c\left[o\overset{l}{-\!-}o\right]$ with l "horizontal" separating sites have the form

$$-k^{-1}\frac{d}{dt}c\left[o\overset{l}{-\!-}o\right] = 2\left(c\left[oo\overset{l}{-\!-}o\right] + 2c\left[\overset{o}{o}\overset{l}{-\!-}o\right] + c\left[o\overset{l-1}{-\!-}oo\right]\right), \quad \text{for } l \geq 1, \tag{2.18}$$

where $c\left[oo\overset{l}{-\!-}o\right] \equiv f\left[oo\overset{l}{-\!-}o\right] - f[oo]f[o]$, etc.. One can write

$$c\left[oo\overset{l}{-\!-}o\right] \equiv q[o\phi]c\left[o\overset{l}{-\!-}o\right] + \left(q\left[o\phi\overset{l}{-\!-}\phi\right] - q[o\phi]\right)f\left[o\overset{l}{-\!-}o\right] \tag{2.19}$$

and here we neglect the second term for $l \geq 2$. This is not too unreasonable since there will be some shielding from the left "ϕ" site in $q\left[o\phi\overset{l}{-\!-}\phi\right]$. Similarly we replace $c\left[\overset{o}{o}\overset{l}{-\!-}o\right]$ by $q[o\phi]c\left[o\overset{l}{-\!-}o\right]$ neglecting $\left(q\left[\overset{o}{\phi}\overset{l}{-\!-}\phi\right] - q\left[\overset{o}{\phi}\right]\right) \times f\left[o\overset{l}{-\!-}o\right]$ for $l \geq 2$ (which is, no doubt, a more severe approximation).

Thus we integrate the equations

$$-k^{-1}\frac{d}{dt}c[o\text{--}o] \simeq 2(3q[o\phi]c[o\text{--}o] + c[oo\text{-}o]),$$

$$-k^{-1}\frac{d}{dt}c\left[o\overset{l}{-\!-}o\right] \simeq 2q[o\phi]\left(3c\left[o\overset{l}{-\!-}o\right] + c\left[o\overset{l-1}{-\!-}o\right]\right), \quad l \geq 3 \tag{2.20}$$

for "l" up to some arbitrary integer (see Table I.). Here $q[o\phi]$ and $c[oo\text{--}o]$ are obtained from the third-order truncated

TABLE II. Saturation values for $f[o\text{-}o]$ for random monomer filling of sites with $\leq m$ filled NN and blocking with $> m$ on an infinite, uniform 2-D square lattice (the exact value for $m = 3$ is $2/45 = 0.0444\cdot$).

m	0	1	2	3	4
$T.3F$	0.449	0.2739	0.1364	0.0451	0.000
$FT.2$	0.454	0.2738	0.1345	0.0431	0.000

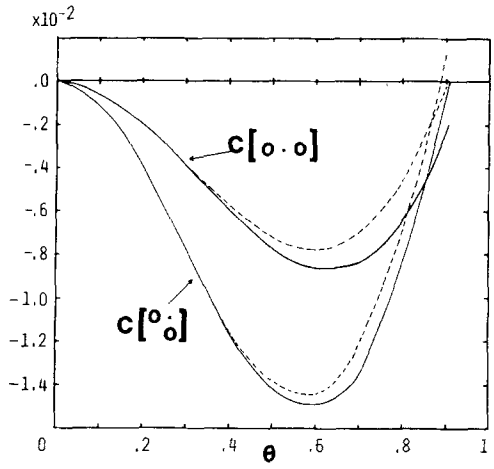


FIG. 8. $c[o\text{-}o]$ and $c[o\text{-}^{o}]$ for random dimer filling of NN sites on an infinite, uniform 2-D square lattice, the solid (dashed) lines give third- (second-) order truncation values.

equations.[17] We remark that $c[o\text{--}o]$ from (2.20) then agrees with the third-order truncation value. The resulting values for correlations of larger separation appear reasonable.

III. BASIC z-TRANSFORM TECHNIQUES FOR TWO-CLUSTER CORRELATIONS

The large separation behavior of the spatial correlations for the 1-D processes of examples (i) and (ii) clearly depends on the coupling structure of the disconnected hierarchy (correlations with a certain separation are coupled through their rate equations to those for a range of smaller separations). The same structure is apparent for processes on higher-dimensional lattices (and we have indicated that suitable approximate truncation schemes should preserve this structure). Here we present a technique, appropriate for treating equations with such structure, which is directed towards elucidating the asymptotic form of the large separation behavior of the correlations. It is first applied to the analysis of exact equations for two-cluster correlations for processes with NN cooperative effects in 1-D [exactly solvable example (i)] and 2-D [example (iv)]. Next we treat the approximately truncated Eq. (2.20) for random dimer filling on a 2-D square lattice. In the next section some extensions to longer-range cooperative effects and multicluster correlations are discussed.

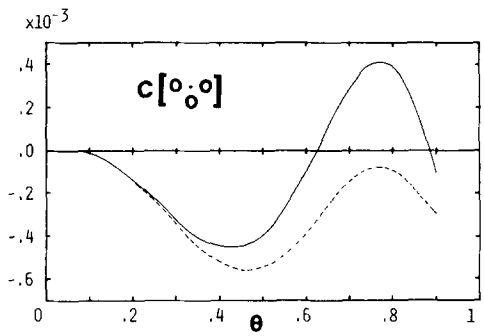


FIG. 9. $c[\overset{o}{o}\text{-}o]$ for the process of Fig. 8; the solid (dashed) line gives third- (second-) order truncation values.

Example (i) revisited: Reaction on an initially unreacted infinite, uniform polymer chain; NN cooperative effects, with $\tau_{a\cdot o} = \tau_{o\cdot a}$.

We consider the large "m" behavior of $c\left[oo\overset{m}{\text{——}}oo\right]$, $c\left[oo\overset{m}{\text{——}}o\right]$, and $c\left[o\overset{m}{\text{——}}o\right]$. Since the f's reduce to factored form (i.e., the c's vanish) in the large "m" limit, equations for these factored forms are simply obtained from (2.2) in the $m\to\infty$ limit. It thus readily follows that the c's obey the same equations as the corresponding f's, i.e., (2.2), for $m\geqslant 1$. Given the coupling structure of (2.2), it is natural to consider the z-transformed quantities

$$\hat{c}_z[oo\text{——}oo] = \sum_{l=0}^{\infty} z^l c\left[oo\overset{l+1}{\text{——}}oo\right],$$
$$\hat{c}_z[oo\text{——}o] = \sum_{l=0}^{\infty} z^l c\left[oo\overset{l+1}{\text{——}}o\right],$$
$$\hat{c}_z[o\text{—}o] = \sum_{l=0}^{\infty} z^l c\left[o\overset{l+1}{\text{——}}o\right]. \tag{3.1}$$

Applying the z-transform to (2.2) for the c's and dividing by $(d/dt)q_2 = -\tau_{o\cdot o}q_2$ to obtain d/dq_2 equations (since we are not directly concerned with time dependence here), one obtains firstly for $\hat{c}_z[oo\text{——}oo]$,

$$\frac{d}{dq_2}\hat{c}_z[oo\text{——}oo]$$
$$= \{4\alpha q_2^{-1} + 2(1+z)(1-\alpha)\}\hat{c}_z[oo\text{——}oo]$$
$$+ 2(1-\alpha)(f[oooo] - f[oo]^2), \tag{3.2}$$

where the reduced rate $\alpha \equiv \tau_{o\cdot a}/\tau_{o\cdot o} (\equiv 1-\gamma)$. For convenience set $p_2 \equiv 1 - q_2$. Then (3.2) may be integrated after substituting for $f[oo] = q_2^{2\alpha}\exp(-2\gamma p_2)$, $f[oooo] = q_2^2 f[oo]$ (see Refs. 5 and 10), and taking the inverse z-transform,

$$c\left[oo\overset{l}{\text{——}}oo\right] = \frac{1}{(l-1)!}\frac{d^{l-1}}{dz^{l-1}}\hat{c}_z[oo\text{——}oo]\Big|_{z=0} \tag{3.3}$$

to obtain

$$c\left[oo\overset{l}{\text{——}}oo\right]$$
$$= q_2^{4\alpha}\exp(-2\gamma p_2)\frac{(-2\gamma)^l}{(l-1)!}\left\{+\int_0^{p_2} dp(1-p)^{2\gamma}(p_2-p)^{l-1}\right.$$
$$\left. - \int_0^{p_2} dp\, e^{-2\gamma p}(p_2-p)^{l-1}\right\}. \tag{3.4}$$

The first integral in (3.4) equals $p_2^l l^{-1}F(-2\gamma, 1, l+1, p_2)$ and the second equals $p_2^l l^{-1}M(1, l+1, -2\gamma p_2)$, where $F(\ )$ and $M(\ )$ are hypergeometric and confluent hypergeometric functions, respectively.[18] Inserting standard series expansions for these functions, (3.4) becomes

$$q_2^{-4\alpha}\exp(2\gamma p_2)(-2\gamma p_2)^{-l}c\left[oo\overset{l}{\text{——}}oo\right]$$
$$= \sum_{k=0}^{\infty}\left\{\frac{(2\gamma)!(-p_2)^k}{(2\gamma-k)!(k+l)!} - \frac{(-2\gamma p_2)^k}{(k+l)!}\right\}$$
$$\sim \frac{-2\gamma p_2^{\,2}}{(l+2)!} \text{ as } l\to\infty. \tag{3.5}$$

From either (3.4) or (3.5) the "fast decay" behavior of this

correlation is clear. An alternative approach involving iterative solution rather than z-transform of the coupled-l equations is given in Appendix B.

Treatment of the equation for $c\left[oo\overset{l}{\text{——}}o\right]$ is analogous but more complicated because of the coupling to $c\left[oo\overset{l'}{\text{——}}oo\right]$, $l' = l, l-1$. Integrating the $(d/dq_2)\hat{c}_z[oo\text{——}o]$ equation and taking the inverse z-transform, one obtains

$$
\begin{aligned}
&A(p_2)c\left[oo\overset{l}{\text{——}}o\right] \\
&= -\frac{\gamma^{l-1}}{(l-1)!}\int_0^{p_2} dp\, A(p)(p-p_2)^{l-1}\eta(p) \\
&\quad - \int_0^{p_2} dp\, A(p)\zeta(p)\sum_{k=0}^{l-1}\frac{1}{k!}\gamma^k(p-p_2)^k c\left[oo\overset{l-k}{\text{——}}oo\right] \\
&\quad - \int_0^{p_2} dp\, A(p)\psi(p)\sum_{k=0}^{l-2}\frac{1}{k!}\gamma^k(p-p_2)^k c\left[oo\overset{l-k-1}{\text{——}}oo\right],
\end{aligned}
\tag{3.6}
$$

where $A(p) = (1-p)^{-2\alpha-\hat{\alpha}}e^{\gamma p}$,

$$
\begin{aligned}
\eta(p) &= (1-\alpha)(f[ooo] - f[oo]f[o]) + \{(\alpha-\hat{\alpha}) \\
&\quad \times (1-p)^{-1} + 1 - 2\alpha + \hat{\alpha}\}(f[oooo] - f[oo]^2),
\end{aligned}
$$

and $\zeta(p) = (\alpha-\hat{\alpha})(1-p)^{-1}$ and $\psi(p) = (\alpha-\hat{\alpha})(1-p)^{-1} + 1 - 2\alpha + \hat{\alpha}$. Also $\hat{\alpha} \equiv \tau_{a.a}/\tau_{o.o}$ and all f's and c's appearing on the right-hand side are to be evaluated at $q_2 = 1-p$. The fast decay behavior of the first term, similar to (3.4) and (3.5), is clear, however that of the remaining terms is less transparent. To clarify this, it is convenient to substitute from (3.4) for $c\left[oo\overset{l'}{\text{——}}oo\right]$. One may explicitly perform the sum over "k" which, in the case of the second term of (3.6), has the form

$$
\begin{aligned}
&\sum_{k=0}^{l-1}\frac{1}{k!}\gamma^k(p-p_2)^k \cdot \frac{1}{(l-k-1)!}(-2\gamma)^{l-k}(p-p')^{l-k-1} \\
&\quad = \frac{-2\gamma^l}{(l-1)!}(2p'-p_2-p)^{l-1}.
\end{aligned}
\tag{3.7}
$$

Thus the decay has the same form as the first term. Analysis of the third term is similar, except a contribution with a $1/(l-2)!$ factor results.

One may continue to analyze the behavior of $c\left[o\overset{l}{\text{——}}o\right]$ in the same fashion. The same type of "fast decay" is again observed with $1/(l-k)!$ factors, $k = 1, 2, 3$, for $c\left[o\overset{l}{\text{——}}o\right]$ the integral form.

The appearance of $1/(l-k)! \sim (e/2\pi)^{1/2}\exp\{-(l-k+\frac{1}{2})[\ln(l-k+1)-1]\}$ "fast decay" factors is markedly different from the corresponding equilibrium, here Ising model, behavior which is characterized by exponential decay. More generally, we note that any nth-order spatially Markovian distribution (the Ising model distribution is first-order) has spatial correlations exhibiting exponential decay (see Appendix A). Faster decay in irreversible (compared with equilibrium) distributions is anticipated since events occuring irreversibly have only one chance to correlate (during the transformation $o \rightarrow a$) rather than through continual interaction.

Example (iv) specialized: (Monomer filling of an initially empty infinite, uniform square lattice, with NN blocking and rate k.)

In general, one could consider probabilities for configurations consisting of two clusters separated by various distances and directions (or the two-cluster correlations obtained from these by subtracting the product of probabilities for the individual clusters). However, only "horizontal" separations are considered here. Denote the two cluster correlations for

$$
\underbrace{o\text{---}\ldots\text{-}o}_{l} \text{ by } c\left[o\overset{l}{\text{——}}o\right], \quad o\overset{o}{\underset{o}{o}}\underbrace{o\text{---}\ldots\text{-}o}_{l} \text{ by } c\left[o\overset{o}{\underset{o}{o}}o\overset{l}{\text{——}}o\right],
$$

etc.

Now from (2.1),

$$
\begin{aligned}
&-\frac{d}{dt}c[oo] = 2kf\left[o\overset{o}{\underset{o}{o}}o\right](1-f[o]), \\
&-\frac{d}{dt}c\left[o\overset{l}{\text{——}}o\right] = 2kc\left[o\overset{l-1}{\text{——}}o\overset{o}{\underset{o}{o}}o\right], \quad l \geqslant 1, \\
&-\frac{d}{dt}c\left[o\overset{l}{\text{——}}o\overset{o}{\underset{o}{o}}o\right] \\
&\quad = k\Big\{c\left[o\overset{l}{\text{——}}o\overset{o}{\underset{o}{o}}o\right] + c\left[o\overset{l-1}{\text{——}}o\overset{oo}{\underset{oo}{oo}}o\right] + c\left[o\overset{l}{\text{——}}o\overset{oo}{\underset{oo}{oo}}o\right] \\
&\qquad + 2c\left[o\overset{l}{\text{——}}\overset{\overset{o}{ooo}}{\underset{o}{ooo}}\right] + c\left[o\overset{o}{\underset{o}{o}}o\overset{l-1}{\text{——}}o\overset{o}{\underset{o}{o}}o\right]\Big\}, \quad l \geqslant 1, \\
&\qquad\vdots
\end{aligned}
\tag{3.8}
$$

The c's have zero initial values so the nonzero "source" for these comes from the $l = 0$ equations. The structure of the equations (for either f's or c's) suggests application of a z-transform

$$
\begin{aligned}
&\hat{c}_z[o\text{—}o] = \sum_{l=0}^{\infty} z^l c\left[o\overset{l}{\text{—}}o\right], \\
&\hat{c}_z\left[o\text{—}o\overset{o}{\underset{o}{o}}o\right] = \sum_{l=0}^{\infty} z^l c\left[o\overset{l}{\text{—}}o\overset{o}{\underset{o}{o}}o\right], \cdots,
\end{aligned}
\tag{3.9}
$$

so

$$
c\left[o\overset{l}{\text{——}}o\right] = \frac{1}{l!}\frac{d^l}{dz^l}\hat{c}_z[o\text{——}o]\Big|_{z=0}, \quad \text{etc.} \tag{3.10}
$$

Applying the transform to the $c\left[o\overset{l}{\text{——}}o\right]$ equations gives

$$
\begin{aligned}
-\frac{d}{dt}\hat{c}_z[o\text{——}o] &= 2k\Big\{z\hat{c}_z\left[o\text{——}o\overset{o}{\underset{o}{o}}o\right] \\
&\quad + f\left[o\overset{o}{\underset{o}{o}}o\right](1-f[o])\Big\}.
\end{aligned}
\tag{3.11}
$$

The "higher-order" transformed equations all have a similar

structure linear in z and the $\hat{c}$'s with z-independent source terms. Of course, Eq. (3.11) cannot be solved exactly, but to indicate the nature of the decay of the correlation, we first extract a term $\hat{c}_z[o\!-\!\!-\!\!-\!o]q\left[\phi\begin{smallmatrix}o\\o\\o\end{smallmatrix}\right]$ from $\hat{c}_z\left[o\!-\!\!-\!\!-\!o\begin{smallmatrix}o\\o\\o\end{smallmatrix}\right]$, then integrate and invert the transform to obtain

$$c\left[o\overset{l}{-\!\!-\!\!-}o\right] = -\frac{2k}{l!}\int_0^t dt' f\left[o\begin{smallmatrix}o\\o\\o\end{smallmatrix}\right]' \times\{1-f[o]'\}\left\{-2k\int_{t'}^t dt''\, q\left[\phi\begin{smallmatrix}o\\o\\o\end{smallmatrix}\right]'\right\}^l - 2k\int_0^t dt'\sum_{k=0}^{l-1}\frac{1}{k!}\left\{-2k\int_{t'}^t dt''q\left[\phi\begin{smallmatrix}o\\o\\o\end{smallmatrix}\right]''\right\}^k \times\left\{c\left[o\overset{l-1-k}{-\!\!-\!\!-}o\begin{smallmatrix}o\\o\\o\end{smallmatrix}\right]' - c\left[o\overset{l-1-k}{-\!\!-\!\!-}o\right]'q\left[\phi\begin{smallmatrix}o\\o\\o\end{smallmatrix}\right]'\right\}, \tag{3.12}$$

where the prime (double prime) indicates that the functions are evaluated at $t'(t'')$. Note that

$$c\left[o\overset{m}{-\!\!-\!\!-}o\begin{smallmatrix}o\\o\\o\end{smallmatrix}\right] - c\left[o\overset{m}{-\!\!-\!\!-}o\right]q\left[\phi\begin{smallmatrix}o\\o\\o\end{smallmatrix}\right] \equiv f\left[o\overset{m}{-\!\!-\!\!-}o\right]\left(q\left[\phi\overset{m}{-\!\!-\!\!-}\phi\begin{smallmatrix}o\\o\\o\end{smallmatrix}\right] - q\left[\phi\begin{smallmatrix}o\\o\\o\end{smallmatrix}\right]\right).$$

Our attempt to demonstrate the basic form of the asymptotic behavior of this correlation through the first term in (3.12) has met with mixed success. Although the first term is anticipated to make a significant contribution, this is also true of the second ($q\left[\phi\overset{l'}{-\!\!-\!\!-}\phi\begin{smallmatrix}o\\o\\o\end{smallmatrix}\right]$ and $q\left[\phi\begin{smallmatrix}o\\o\\o\end{smallmatrix}\right]$ differ significantly for small l' because of the limited shielding ability of the right-hand "ϕ" in the first q). A rigorous proof of the asymptotic "fast decay" would require an explicit sum over "k" to exhibit the behavior of the second term (the $k=l-1$ term by itself has $1/(l-1)!$ behavior). However, heuristically, we expect the same type of fast decay behavior in two (and higher) dimensions as discovered in the above 1-D process. The inability to propagate correlation through continual interaction suggests faster decay than that of the corresponding 2-D equilibrium distributions which is exponential (to infinite separation value).[19] Further indication of general "fast decay" behavior comes from the formal density expansion for correlations with separation l which has a lead term with $1/(l-c)!$ type behavior.[20]

Example (v) revisited: (Random dimer filling of an infinite, uniform 2-D square lattice with rate k.)

If we apply to the truncated closed equations (2.20) for $c\left[o\overset{l}{-\!\!-\!\!-}o\right]$ for $l\geqslant 3$, the transform $\hat{c}_z[o\!-\!\!-\!\!-\!o] = \sum_{l=0}^{\infty} z^l c\left[o\overset{l+3}{-\!\!-\!\!-}o\right]$, integrate with respect to time, and then invert the transform, we obtain for $l\geqslant 3$,

$$c\left[o\overset{l}{-\!\!-\!\!-}o\right] = -\frac{(-2)^{l-3}}{(l-3)!}\int_0^s ds'\left(\int_{s'}^s ds''\, q[o\phi\,]''\right)^{l-3} \times\exp\left(-6\int_{s'}^s ds'\, q[o\phi\,]'\right)c[o\text{--}o]', \tag{3.13}$$

where $s=kt$. Thus our truncation, preserving the coupling structure of the $c\left[o\overset{l}{-\!\!-\!\!-}o\right]$ equations, has ensured a "fast" asymptotic decay of these two-point correlations.

IV. EXTENDED z-TRANSFORM TECHNIQUES FOR MULTICLUSTER CORRELATIONS AND LONGER-RANGE COOPERATIVE EFFECTS

In examples (i) and (iv) involving NN cooperative effects, two cluster correlations with separation "l" are coupled directly through their rate equations to ones with "$l-1$," but not shorter, separations. Correspondingly, in expressions for these, $1/(l-c)!$ type factors, where $c=0(1)$, naturally appear. Not surprisingly, the random dimer filling problems exhibits similar behavior. For such correlations in more general cooperative processes, we expect the decay to be characterized most naturally in terms of separation measured in units of the cooperative range (cf. Fig. 5). Consider, for example, a 1-D system with range R cooperative effects. To clearly exhibit the above-mentioned decay behavior, we modify the z-transform as follows. Set

$$\hat{c}_z\left[o\overset{i}{-\!\!-\!\!-}o\right] \equiv \sum_{k=0}^{\infty} z^k c\left[o\overset{i+Rk}{-\!\!-\!\!-}o\right], \quad \text{where } i=0,1,\ldots,R-1, \tag{4.1}$$

and define similar transforms for other two-cluster correlations. Using (4.1), the resulting transformed c-equations can be written with the right-hand side linear in z and the $\hat{c}$'s. Integrating and inverting the z-transforms, (4.1) produces the desired $1/k!$ factor. If applied to example (ii), one sees immediately that a closed coupled pair of equations result for $\hat{c}_z\left[oooo\overset{0}{-\!\!-\!\!-}oooo\right]$ and $\hat{c}_z\left[oooo\overset{1}{-\!\!-\!\!-}oooo\right]$ which when integrated and inverted describe the exact large "l" asymptotic behavior of $c\left[oooo\overset{l}{-\!\!-\!\!-}oooo\right]$. Next one would consider the $\hat{c}_z\left[oooo\overset{i}{-\!\!-\!\!-}ooo\right]$ equations and then continue to analyze c's for successively smaller clusters.

The appropriate form of the z-transform for general cooperative processes should be clear from the above discussion. Consider for example a process on a 2-D square or 3-D cubic lattice where an event is cooperatively influenced by the state of surrounding sites that can be reached in R or less steps of a single lattice vector. Then a z-transform of the form (4.1) is appropriate for examination of two-cluster correlations in any principal lattice direction.

Finally, we consider the analysis of "effectively" multiply disconnected configurations where it is natural to implement multiple z-transforms (with one transform variable for each separating distance). We illustrate this technique with a simple example.

Example (i) revisited: Reaction on an initially unreacted, infinite, uniform polymer chain; NN cooperative effects and $\tau_{a,o}=\tau_{o,a}$.

The simplest nontrivially doubly disconnected quantity that one can consider here is $f\left[oo\overset{l}{-\!\!-\!\!-}o\overset{k}{-\!\!-\!\!-}oo\right]$. Starting with (2.1) and (2.2) for these f's, and implementing the shielding property of adjacent pairs of empty sites, specifically (2.8a), one obtains for l, $k\geqslant 1$,

$$-\frac{d}{dt}f\left[oo\overset{l}{\text{—}}o\overset{k}{\text{—}}oo\right]$$

$$=\{\tau_{a.a}+4\tau_{a.o}+2(\tau_{o.o}-\tau_{a.o})q_2\}f\left[oo\overset{l}{\text{—}}o\overset{k}{\text{—}}oo\right]+(\tau_{o.o}-\tau_{a.o})q_2\left(f\left[oo\overset{l-1}{\text{—}}o\overset{k}{\text{—}}oo\right]\right.$$

$$\left.+f\left[oo\overset{l}{\text{—}}o\overset{k-1}{\text{—}}oo\right]\right)+(\tau_{o.o}-2\tau_{a.o}+\tau_{a.a})f\left[oo\overset{l-1}{\text{—}}oo\right]f\left[oo\overset{k-1}{\text{—}}oo\right]f[oo]^{-1}q_2$$

$$+(\tau_{a.o}-\tau_{a.a})\left(f\left[oo\overset{l-1}{\text{—}}oo\right]f\left[oo\overset{k}{\text{—}}oo\right]+f\left[oo\overset{l}{\text{—}}oo\right]f\left[oo\overset{k-1}{\text{—}}oo\right]\right)f[oo]^{-1}. \tag{4.2}$$

Instead of this f, it is more convenient to consider

$$c\left[oo\overset{l}{\text{—}}o\overset{k}{\text{—}}oo\right]$$

$$\equiv f\left[oo\overset{l}{\text{—}}o\overset{k}{\text{—}}oo\right]-f\left[oo\overset{l}{\text{—}}o\right]f[oo]$$

$$-f[oo]f\left[o\overset{k}{\text{—}}oo\right]+f[oo]^2f[o]. \tag{4.3}$$

This is not the standard Ursell–Mayer form for a three-cluster correlation which, for $c\left[oo\overset{l}{\text{—}}o\overset{k}{\text{—}}oo\right]$, differs from the right-hand side of (4.3) by $\left\{-f\left[oo\overset{l+k+1}{\text{—}}oo\right]f[o]+f[oo]^2f[o]\right\}$. Use of correlations in the form (4.3) is motivated by the factorized form of the latter terms in (4.2) and leads naturally to rate equations for these c's with $l,\ k\geqslant 1$ obtained from (4.2) by simply replacing $f\left[oo\overset{m}{\text{—}}o\overset{n}{\text{—}}oo\right]$ with $c\left[oo\overset{m}{\text{—}}o\overset{n}{\text{—}}oo\right]$ and $f\left[oo\overset{m}{\text{—}}oo\right]$ with $c\left[oo\overset{m}{\text{—}}oo\right]$.

To these we apply the double z-transform

$$\hat{c}_{z_1,z_2}\equiv\hat{c}_{z_1,z_2}[oo\text{—}o\text{—}oo]\equiv\sum_{l,k>0}z_1^l z_2^k c\left[oo\overset{l+1}{\text{—}}o\overset{k+1}{\text{—}}oo\right] \tag{4.4}$$

using such results as

$$\sum_{l,k>0}z_1^l z_2^k c\left[oo\overset{l}{\text{—}}o\overset{k+1}{\text{—}}oo\right]$$

$$=z_1\hat{c}_{z_1,z_2}+q_2\hat{c}_{z_2}[oo\text{——}oo]+f[oo]\hat{c}_{z_2}[oo\text{——}o], \tag{4.5}$$

with $\hat{c}_z[oo\text{——}oo]$ and $\hat{c}_z[oo\text{——}o]$ as defined in Sec. III, to obtain

$$-\frac{d}{dt}\hat{c}_{z_1,z_2}$$

$$=\{\tau_{a.a}+4\tau_{a.o}+(\tau_{o.o}-\tau_{a.o})(2+z_1+z_2)q_2\}\hat{c}_{z_1,z_2}$$

$$+(\tau_{o.o}-\tau_{a.o})q_2(q_2\hat{c}_{z_2}[oo\text{——}oo]$$

$$+f[oo]\hat{c}_{z_2}[oo\text{——}o]+z_1\leftrightarrow z_2)$$

$$+(\tau_{o.o}-2\tau_{a.o}+\tau_{a.a})(z_1\hat{c}_{z_1}[oo\text{——}oo]$$

$$+f[oooo]-f[oo]^2)$$

$$\times(z_1\leftrightarrow z_2)f[oo]^{-1}q_2$$

$$+(\tau_{a.o}-\tau_{a.a})(\{z_1\hat{c}_{z_1}[oo\text{——}oo]+f[oooo]$$

$$-f[oo]^2\}\hat{c}_{z_2}[oo\text{——}oo]$$

$$+z_1\leftrightarrow z_2)f[oo]^{-1}. \tag{4.6}$$

Finally dividing by $-(d/dt)q_2=\tau_{o.o}q_2$, integrating with respect to q_2, and inverting the double z-transform, one obtains

$$c\left[oo\overset{l}{\text{—}}o\overset{k}{\text{—}}oo\right]$$

$$=-q_2^{\hat{\alpha}+4\alpha}\exp(-2\gamma p_2)\int_0^{p_2}dp(1-p)^{-(\hat{\alpha}+4\alpha)}e^{2\gamma p}$$

$$\times\left\{\frac{\gamma^l(p-p_2)^{l-1}}{(l-1)!}\sum_{\hat{k}=0}^{k-1}\gamma^{\hat{k}}(p-p_2)^{\hat{k}}\left(c\left[oo\overset{k-\hat{k}}{\text{—}}oo\right]\right.\right.$$

$$\times(1-p)+c\left[o\overset{k-\hat{k}}{\text{—}}oo\right]f[oo]\Big)/\hat{k}!$$

$$+\tfrac{1}{2}(1-2\alpha+\hat{\alpha})B(l)B(k)f[oo]^{-1}$$

$$+(\alpha-\hat{\alpha})B(l)\left(\sum_{\hat{k}=0}^{k-1}\gamma^{\hat{k}}(p-p_2)^{\hat{k}}c\left[oo\overset{k-\hat{k}}{\text{—}}oo\right]/\hat{k}!\right)$$

$$\left.\times f[oo](1-p)^{-1}+l\leftrightarrow k\right\}, \tag{4.7}$$

where

$$B(l)=\sum_{\hat{l}=0}^{l-2}\gamma^{\hat{l}}(p-p_2)^{\hat{l}}c\left[oo\overset{l-\hat{l}-1}{\text{—}}oo\right]/\hat{l}!$$

$$+\frac{(f[oooo]-f[oo]^2)\gamma^{l-1}(p-p_2)^{l-1}}{(l-1)!}$$

and p_2, $\alpha\equiv 1-\gamma$, $\hat{\alpha}$ are as previously defined and all f's on the right-hand side are evaluated at $q_2=1-p$.

If the sums over l and k are evaluated explicitly [cf. the $c\left[oo\overset{l}{\text{—}}o\right]$ calculation (3.7)], then it becomes clear that each term in $c\left[oo\overset{l}{\text{—}}o\overset{k}{\text{—}}oo\right]$ has fast decay behavior dominated by $(1/(l-c)!)\cdot(1/(k-c')!)$ as $l,k\to\infty$ [where $c,c'=0(1)$].

V. DISCUSSION

The aim of this work is to elucidate the coupling structure of the hierarchy equations for (effectively) disconnected empty configurations as well as the behavior of the corresponding probabilities, particularly two-cluster correlations. For exactly solvable 1-D processes, the latter exhibit large separation asymptotic "fast" decay behavior quite different from the slower exponential equilibrium-type decay.

Similar behavior is anticipated for general irreversible processes on 1-, 2-, and 3-D lattices. We have remarked that 1-D distributions satisfying an nth-order spatial Markovian property will have spatial correlations exhibiting exponential decay. Consequently in developing approximate truncation schemes for 1-D processes not amenable to exact solution, one must ensure that these do not impose such a Markovian property leading to spurious exponential decay of correlations. Similar considerations apply in higher dimension [note our treatment leading to (2.20) and (3.13)].

It is instructive to compare our results for corresponding 1- and 2-D systems. Consider first random dimer filling on the 1-D linear and 2-D square lattices. In both cases $c\left[a\overset{l}{\text{---}}a\right] \equiv c\left[o\overset{l}{\text{---}}o\right]$ alternates in sign with l, specifically, $c[aa] = f[aa] - f[a]^2 > 0$, $c[a\text{-}a] = f[a\text{-}a] - f[a]^2 < 0$, etc. (except that $c[aa] = f[oo] - f[o]^2$ becomes negative near saturation). To understand these results, suppose that some specific site is filled. Then for low coverage: (a) this enhances the (conditional) probability (above $f[a]$) that the site to the right is filled (by the same dimer). This constitutes one of 2(4) possible dimer orientations is 1-D (2-D). Thus the effect is more pronounced in 1-D, i.e., $c[aa] > 0$ is larger in 1-D than 2-D, at least away from saturation.

(b) If the dimer does cover the specified site and one to the right, then this reduces the number of ways a dimer can cover the site two to the right, i.e., from 2 to 1 (4 to 3) in 1-D (2-D). Thus $c[a\text{-}a] < 0$ and is much larger in 1-D than 2-D, at least away from saturation.

(c) One may continue this argument demonstrating the alternating sign.

Accurate 2-D truncations verify these conclusions for $c[aa]$ and $c[a\text{-}a]$. Solution of the crudely truncated equations (2.20) supports the anticipated trend that $c\left[a\overset{l}{\text{---}}a\right]$, $l \geqslant 2$ are much larger in 1-D than 2-D (see Table I).

For comparison, consider a process corresponding to monomer filling with NN cooperative effects on a 1-D linear and 2-D square lattice. Suppose a filled NN inhibits the adsorption rate. If some specific site is filled than this clearly inhibits the (conditional) probability (below $f[a]$) that a neighbor is occupied, which in turn enhances the probability that a site two away is occupied, etc.. Thus $c[aa] < 0$, $c[a\text{-}a] > 0$, etc.. The dimensional dependence of these correlations should clearly be much less significant than for the random dimer filling problem. Our results support this conclusion for $c[aa]$ [Ref. (1)] and $c[a\text{-}a]$ (compare Figs. 4 and 5).

Finally, we remark on some extensions and alternative applications of the methods developed here. First, a similar treatment could be given for irreversible cooperative processes on lattices involving several different competing types of events (where exact results are again available for certain 1-D processes).[21] Second, for processes on lattices which are, e.g., semi-infinite or have defective clusters, probabilities of subconfigurations are naturally labeled by the displacement from the edge or defects. Coupling structure with respect to these labels is similar to that observed here and analogous techniques are appropriate.[6] Such analyses will be presented in later work.

ACKNOWLEDGMENTS

The authors would like to thank R. S. Nord for assistance in calculating the two-point correlations for random dimer filling.

Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract No. W-7405-ENG-82. This work was supported by the Office of Basic Energy Sciences.

APPENDIX A: SPATIALLY MARKOVIAN DISTRIBUTIONS

For any distribution, one can write

$$f\left[o\overset{l}{\text{---}}o\right] = \sum_{\sigma_i = o,\, a} f[o\sigma_1\sigma_2\cdots\sigma_l o]$$

$$= \sum_{\sigma_i = o,\, a} q\left[o\,\overline{\sigma_1\sigma_2\cdots\sigma_l o}\right] \times q\left[\sigma_1\,\overline{\sigma_2\cdots\sigma_l o}\right]\cdots q[\sigma_l \bar{o}] f[o]. \tag{A1}$$

If the distribution is (first-order) spatially Markovian, then $q[\sigma\,\overline{\sigma'\boldsymbol{\sigma}'}] = q[\sigma\,\overline{\sigma'}]$, where σ' can be either filled or empty, and thus

$$f\left[o\overset{l}{\text{---}}o\right] = (\mathsf{T}^{l+1})_{o,\,o} f[o], \tag{A2}$$

where $(\mathsf{T})_{\sigma,\,\sigma'} = q[\sigma\,\overline{\sigma'}]$ all of which can be expressed in terms of θ and $f[oo]$. Furthermore, T has eigenvalues

$$\lambda^+ = 1, \lambda^- = 1 - (1 - \theta - f[oo])\theta^{-1}(1-\theta)^{-1} (|\lambda^-| \leqslant 1),$$

and

$$g(\mathsf{T}) = g(1)\begin{pmatrix}1\\1\end{pmatrix}(1-\theta, \theta) + g(\lambda^-)\begin{pmatrix}\theta\\ \theta - 1\end{pmatrix}(1, -1), \tag{A3}$$

where $f[oo] = q[o\phi\,]f[o]$ and $g(\ )$ is arbitrary. Thus,

$$c\left[o\overset{l}{\text{---}}o\right] = \theta(1-\theta)(\lambda^-)^{l+1}. \tag{A4}$$

General two-cluster correlations are calculated similarly and also have exponential decay.

The equilibrium, Ising model, lattice gas distribution is, in fact, spatially Markovian and[22]

$$(1 - \theta - f[oo])\theta^{-1}(1-\theta)^{-1} = 2/(\{1 + 4\theta(1-\theta)(e^{-\beta J} - 1)\}^{1/2} + 1), \tag{A5}$$

where J is the pairwise interaction and $\beta = (kT)^{-1}$. Note that $c\left[o\overset{l}{\text{---}}o\right]$ has the same value at $\theta = \theta^*$ and $1 - \theta^*$.

If the distribution is nth-order spatially Markovian, i.e., $q[\sigma\,\overline{\sigma_1\sigma_2\cdots\sigma_n\boldsymbol{\sigma}}] = q[\sigma\,\overline{\sigma_1\sigma_2\cdots\sigma_n}]$, an analogous calculation shows that $f\left[o\overset{l}{\text{---}}o\right]$ contains a factor involving a component of the $[l/n]$th power of a "matrix" $\tilde{\mathsf{T}}$, where

$$(\tilde{\mathsf{T}})_{\sigma_1\sigma_2\cdots\sigma_n,\,\sigma_{n+1}\cdots\sigma_{2n}} = q[\sigma_1\,\overline{\sigma_2\sigma_3\cdots\sigma_{n+1}}]\, q[\sigma_2\,\overline{\sigma_3\sigma_4\cdots\sigma_{n+2}}] \cdots q[\sigma_n\,\overline{\sigma_{n+1}\cdots\sigma_{2n}}]. \tag{A6}$$

Here $[k]$ is the largest integer $\leqslant k$. Now since the q's are positive and

$$\sum_{\substack{\sigma_i = o,\, a \\ i = 1 \text{ to } n}} (\tilde{\mathsf{T}})_{\sigma_1\sigma_2\cdots\sigma_n,\,\sigma_{n+1}\cdots\sigma_{2n}} = 1,$$

it follows that the matrix norm[23]

$$\|\tilde{\mathsf{T}}\|_1 = \max_{\sigma_i'}\left(\sum_{\sigma_i} |(\tilde{\mathsf{T}})_{\sigma_1\sigma_2\cdots\sigma_n,\,\sigma_1'\sigma_2'\cdots\sigma_n'}|\right) = 1. \tag{A7}$$

Consequently all the eigenvalues satisfy $|\lambda| \leqslant 1$. Further, one can immediately check that $\tilde{\mathsf{T}}$ has an equal component dual eigenvector with eigenvalue 1. Analogous to the simple Markovian case, this eigenvalue generates the nonzero asymptotic part of $f\left[o\overset{l}{\text{—}}o\right]$ and the others describe the exponential decay of $c\left[o\overset{l}{\text{—}}o\right]$. Finally, we note that an equilibrium lattice-gas distribution with range "n" interactions is nth-order spatially Markovian.

APPENDIX B: ITERATIVE SOLUTION OF (2.6c)

Here we present an iterative solution of (2.6c) for $f\left[oo\overset{l}{\text{—}}oo\right]$ (cf. the treatment in Ref. 6 of processes on semi-infinite lattices). Defining

$$\chi_l = q_2^{-4\alpha} e^{2\gamma p_2} f\left[o\,o\overset{l}{\text{—}}oo\right], \tag{B1}$$

(2.6c) may be written as

$$\frac{d}{dp_2}\chi_l = -2\gamma\chi_{l-1}. \tag{B2}$$

Noting $\chi_0 = q_2^{2\gamma}$, iterative solution of (B2) yields

$$\chi_l = \sum_{k=0}^{l-1} \frac{(-2\gamma p_2)^k}{k!} + (-2\gamma)^l \int_0^{p_2} dp^1 \int_0^{p^1} dp^2 \cdots \int_0^{p^{l-1}} dp^l (1-p^l)^{2\gamma}. \tag{B3}$$

Evaluating the multiple integral shows the second term equals

$$(-2\gamma p_2)^l \sum_{k=0}^{\infty} \frac{(2\gamma)!(-p_2)^k}{(k+l)!(2\gamma-k)!}$$

in agreement with (3.5).

[1] J. W. Evans, D. R. Burgess, and D. K. Hoffman, J. Chem. Phys. **79**, 5011 (1983).
[2] E. Klesper, W. Gronski, and V. Barth, Macromol. Chem. **150**, 223 (1971); **160**, 167 (1972).
[3] E. A. Boucher, J. Chem. Phys. **59**, 3848 (1973).
[4] T. H. K. Barron, R. J. Bawden, and E. A. Boucher, J. Chem. Soc. **70**, 651 (1974).
[5] N. A. Plate, A. D. Litmanovich, O. V. Noah, A. L. Toam, and N. B. Vasilyev, J. Polym. Sci. **12**, 2165 (1974).
[6] N. O. Wolf, Ph.D. thesis, Iowa State University, 1979.
[7] N. O. Wolf, J. W. Evans, and D. K. Hoffman, J. Math. Phys. **25**, 2519 (1984).
[8] J. W. Evans, J. Math. Phys. **25**, 2527 (1984).
[9] J. W. Evans and D. K. Hoffman, J. Stat. Phys. **36**, 65 (1984).
[10] N. O. Wolf, D. R. Burgess, and D. K. Hoffman, Surf. Sci. **100**, 453 (1980).
[11] J. M. Ziman, *Models of Disorder* (Cambridge U.P., Cambridge, 1979).
[12] E. A. Boucher, Faradays Trans. II **69**, 1839 (1973); J. J. Gonzalez, P. C. Hemmer, and J. S. Høye, Chem. Phys. **3**, 228 (1974).
[13] J. W. Evans and D. R. Burgess, J. Chem. Phys. **79**, 5023 (1983).
[14] E. R. Cohen and H. Reiss, J. Chem. Phys. **38**, 680 (1963).
[15] K. J. Vette, T. W. Orent, D. K. Hoffman, and R. S. Hansen, J. Chem. Phys. **60**, 4854 (1974).
[16] D. R. Burgess, Ph.D. thesis, Iowa State University, 1982.
[17] R. S. Nord and J. W. Evans, J. Chem. Phys. (submitted), which corrects the three lattice vector truncation of Ref. 15.
[18] *The Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1970).
[19] B. M. McCoy and T. T. Wu, *The Two-Dimensional Ising Model* (Harvard U.P., Cambridge, 1973).
[20] J. W. Evans, Physica A **123**, 297 (1984); see also D. K. Hoffman, J. Chem. Phys. **65**, 95 (1976).
[21] J. W. Evans, D. K. Hoffman, and D. R. Burgess, J. Chem. Phys. **80**, 936 (1984).
[22] R. Fowler and E. A. Guggenheim, *Statistical Thermodynamics* (Cambridge U.P., London, 1965).
[23] A. E. Taylor, *Introduction to Functional Analysis* (Wiley, New York, 1958).

A new class of Wick powers of generalized free fields

Erwin Brüning[a)]
Research Institute of Mathematical Sciences, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606, Japan

(Received 30 November 1983; accepted for publication 24 February 1984)

In the frame of generalized Wick powers of generalized free fields a new realization of locality is presented in which locality of the field is not automatically ensured by locality of the two-point function and the well-known combinatorics of Fock space but instead an additional constraint expressing locality of the higher n-point functions has to be satisfied for locality of the field. By example it is shown that this class of models is fairly large.

PACS numbers: 11.10. – z

I. INTRODUCTION

Though there has been much effort the locality condition in general quantum field theory is far from being well understood. The implications of locality together with the other linear constraints of QFT (Poincaré covariance, spectral condition) have been analyzed rather completely.[1] But very little is known about the interplay of locality and positivity (and covariance and the spectral condition). Some problems of the joint realization of locality and positivity and the other linear constraints have been pointed out in Ref. 2. In particular there is a proof of the well-known independence of the locality condition from the other defining constraints of QFT. This is done by showing that the "number" of possible "quantum field theories" is reduced drastically by the locality condition (in special cases from uncountably many to exactly one or even none!). This clearly indicates one important aspect of the locality condition in QFT. An important step towards a thorough understanding of the interplay of locality and positivity is presented by Yngvason[3] who has shown by means of spacelike symmetrization and a subtle approximation procedure that there always exists a separating family of states on the Borchers algebra, which are local and are thus good candidates for quantum field theories. But we think that this result is too "general" in order to account for the subtleties in the interplay of locality and the other constraints of QFT as indicated by our remarks above.

A straightforward way to get an idea about how the locality condition is realized together with the Poincaré covariance, the spectral condition and the positivity condition, is simply to look at known models of QFT's. But unfortunately on physical space-time we know only very few of such models (generalized free fields, Wick products of free and generalized free fields). An obvious fact which is rarely realized is that in these models locality is essentially implied by the other constraints: On the level of the twofold vacuum-expectation values (VEV), locality is indeed implied by Poincaré covariance and positivity[1] but see Sec. V for a generalization. For the VEV's of higher order, locality is implied by the locality of the twofold VEV and the well-known combinatorics (symmetrization with respect to arguments) of Fock space.

Without realizing this obvious fact, we started in Ref. 4 the construction of a class of models of QFT's as "products" of known models. The starting point of this construction is a generalized free field A and a contraction map F from the space of test functions $\mathscr{S} = \mathscr{S}(\mathbb{R}^4)$ (the Schwartz-space of rapidly decreasing $\mathscr{C}^\infty$-functions on $\mathbb{R}^4$) into a space $\mathscr{F}_2$ of symmetric functions of two variables and to show that

$$A^F(f) = :A^2:(F(f))$$

is a well-defined Jacobi field. The question arises which contraction maps yield relativistic quantum fields. This question has been answered in Ref. 4 on the basis of the following assumptions.

(i) A^F is a relativistic quantum field for every generalized free (gf) field A.

(ii) The various notions of locality are equivalent for this analysis (but see Sec. IV).

The result is that A^F is a certain linear combination of derivatives of the Wick product of the gf field A.

In this paper we will drop these assumptions and want to show that for each gf field A a huge class of contraction maps $\mathscr{C}_{i,c}(A)$ can be constructed such that A^F is a relativistic quantum field for each $F \in \mathscr{C}_{i,c}(A)$. For the resulting models it is not obvious that these are not only Poincaré-covariant fields with physical energy momentum spectrum but also local fields. It turns out that in these models we are going to construct locality of the field is, for the first time, not "automatic" by locality of the twofold VEV and the given combinatorics but instead is so by an extra constraint on the contraction map. This constraint is equivalent to locality of the VEV's of order $n \geqslant 3$. Locality of the twofold VEV however is again implied by the other constraints!

We indicate how this paper is organized. Section II introduces our frame and reduces Poincaré covariance and locality of the field A^F to some conditions for the contraction map. Section III contains the analysis of the Poincaré covariance and the spectral condition and presents our starting point for the realization of locality. In order to get an idea about the possibilities of realizing locality, Sec. IV contains a short discussion of some notions of locality which turn out to be relevant in our context. Some sufficient conditions for locality are given. Finally Sec. V contains the realization $\{A^F | F \in \mathscr{C}_{i,c}(A)\}$ of relativistic quantum fields. Some appendixes contain the proofs of those statements which need some more calculations.

[a)] On leave of absence from Fakultät für Physik, Universität Bielefeld, West Germany. JSPS-Fellow.

 0022-2488/84/103064-12$02.50

II. PRODUCTS OF GENERALIZED FREE FIELDS

Every gf field A over the Schwartz-space $\mathscr{S} = \mathscr{S}(\mathbb{R}^4)$ of rapidly decreasing $\mathscr{C}^\infty$-functions on $\mathbb{R}^4$ is up to unitary equivalence fixed by well-known combinatorics and its two-fold vacuum-expectation value

$$\begin{aligned}\mathscr{W}_2^A(f\otimes g) &= \langle\phi_0, A(f)A(g)\phi_0\rangle \\ &= \int t(dp)\tilde{f}(-p)\tilde{g}(p) \\ &= \int_0^\infty \rho(d\kappa)\langle\delta_\kappa^+(p), \tilde{f}(-p)\tilde{g}(p)\rangle, \end{aligned} \tag{2.1}$$

where $\tilde{f} = \mathscr{F}f$ denotes the Fourier transform of $f\in\mathscr{S}$, and t is a positive tempered Lorentz-invariant measure with support $\Sigma\subseteq\overline{V}_+$, and ρ is a positive tempered measure on $(0,\infty)$. There is $L\in\mathbb{N}$ such that

$$\int \frac{t(dp)}{(1+p^0)^{2L}} = N_L^2 < \infty. \tag{2.2}$$

$L^{(\infty)}_{(n)}(\Sigma,t)$ is defined to be the vector space of all measurable functions $f:\Sigma^{xn} = \Sigma\times\cdots\times\Sigma\to\mathbb{C}$, which are bounded almost everywhere with respect to $t^{xn} = t\times\cdots\times t$. As usual, we identify two functions f and g in $L^{(\infty)}_{(n)}(\Sigma,t)$, if $\|f-g\|_\infty = 0$, where

$$\|\cdot\|_\infty = \|\cdot\|_{L^{(\infty)}_{(n)}(\Sigma,t)}.$$

Furthermore, we have to introduce $\hat{\Sigma} = \Sigma\cup(-\Sigma)$, $\hat{t} = t + \check{t}$, where $\check{t}(dp) = t(-dp)$, and then $L^\infty_{(2)}(\hat{\Sigma},\hat{t})$ in complete analogy to $L^{(\infty)}_{(2)}(\Sigma,t)$. The following space of functions turns out to be of particular importance, it is defined by

$$\begin{aligned}\mathscr{G}_2 = \mathscr{G}_2(A) = \{ f_2\in L^{(\infty)}_{(2)}(\hat{\Sigma},\hat{t}) \mid f_2 \text{ symmetric},\\ q^\pm_{N_1,N_2}(f_2)<\infty,\ Q_{N_3}(f_2)<\infty,\ N_j = 0,1,2,\ldots\},\end{aligned} \tag{2.3}$$

where

$$q^\pm_{N_1,N_2}(f_2) = \left\| \prod_{j=1}^2 (1+p_j^0)^{N_j}\chi_\Sigma(p_j) f_2(\pm p_1, \pm p_2)\right\|_\infty \tag{2.4}$$

and

$$Q_N(f_2) = \|(1+|p-q|)^N \chi_{\Sigma\times\Sigma}(p,q) f_2(p,-q)\|_\infty. \tag{2.5}$$

χ_M denotes the characteristic function of the set M and in (2.4) it is understood that we have either (p_1,p_2) or $(-p_1,-p_2)$ as argument of f_2.

In order to define a product of the field A with itself we introduce the notion of a contraction map.

Definition 2.1: A *contraction map* F is a linear continuous map $F:\mathscr{S}\to\mathscr{G}_2$ such that $F(f)^* = F(f^*)$ holds for every $f\in\mathscr{S}$, where $\mathscr{G}_2$ is topologized by the following system of seminorms:

$$\{\|\cdot\|_\infty,\ q^\pm_{N_1,N_2}(\cdot),\ Q_{N_3}(\cdot) \mid N_j = 0,1,2,\ldots\}.$$

The space $\mathscr{C}(A)$ of all contraction maps for a gf field A is thus

$$\begin{aligned}\mathscr{C}(A) &= \mathscr{L}_h(\mathscr{S},\mathscr{G}_2(A)) \\ &= \{F\in\mathscr{L}(\mathscr{S},\mathscr{G}_2(A)) \mid F(f)^* = F(f^*)\ \ \forall f\in\mathscr{S}\}.\end{aligned}$$

Remark 2.1: In order to define the nth power of A, $n>2$, a corresponding space $\mathscr{G}_n$ of functions on $\mathbb{R}^{4n}$ and a linear continuous map $F:\mathscr{S}\to\mathscr{G}_n$, $F(f)^* = F(f^*)$ have to be introduced. Instead of the seminorm Q_N we have to use a system of seminorms $Q_N^{(j)}$, $j = 2,\ldots,n$,

$$\begin{aligned}Q_N^{(j)}(f_n) = \Big\| \Big(1 + \Big|\sum_{i=1}^{j-1} p_i - \sum_{i=j}^n p_j\Big|\Big)^N \\ \times \chi_{\Sigma^{xn}}(p_1,\ldots,p_n) f_n(p_1,\ldots,p_{j-1},-p_j,\ldots,-p_n)\Big\|_\infty.\end{aligned}$$

We need some more spaces of functions. $\mathscr{F}_2$ is defined to be the completion of $\{\chi_{\Sigma\times\Sigma}\cdot F(f) \mid f\in\mathscr{S}\}$ with respect to the system of seminorms $\{q^+_{N_1,N_2}(\cdot) \mid N_j = 0,1,2,\ldots\}$ and then $\mathscr{F}_0 = \mathbb{C}\phi_0$ and

$$\mathscr{F}_{2n} = S_{2n}\,\tilde{\otimes}^n_\pi\,\mathscr{F}_2 \tag{2.6}$$

is the completed n-fold symmetric tensor product of $\mathscr{F}_2$ with respect to the projective tensor product topology induced by $\mathscr{F}_2$.

According to a well-known construction there is, by positivity of $\mathscr{W}_2 = \mathscr{W}_2^A$, $\mathscr{W}_2(f^*\otimes f)\geqslant 0\ \forall f\in\mathscr{S}$, canonically associated with $\mathscr{W}_2$ a Hilbert space $\mathscr{H}_1$, a scalar product $\langle .,.\rangle$, and a strongly continuous linear map $\phi:\mathscr{S}\to\mathscr{H}_1$ such that $\mathscr{W}_2(f\otimes g) = \langle\phi(f^*),\phi(g)\rangle\ \forall f,g\in\mathscr{S}$. The Fourier transform $\tilde{\phi}$ of ϕ is a polynomially bounded Lorentz-covariant measure with support Σ and values in $\mathscr{H}_1$. The n-fold tensor product $\tilde{\phi}^{\otimes n}$ then takes values in $\mathscr{H}_n$, the n-fold tensor product of the Hilbert space $\mathscr{H}_1$, $\mathscr{H}_n = \otimes^n\mathscr{H}_1$. Clearly any $f\in\mathscr{F}_n$ may be integrated with respect to $\tilde{\phi}^{\otimes n}$ to get vectors $\tilde{\phi}^{\otimes n}(f)$ in the symmetric subspace $\otimes^n_s\mathscr{H}_1$ of $\mathscr{H}_n$.

Now the product of A with itself, associated with a contraction map F, is denoted by A^F. A^F is by definition a Jacobi field,[4] e.g.,

$$\begin{aligned}&A^F = ((A^F_{n,m})_{n,m=0,1,2,\ldots}) \text{ on} \\ &\mathscr{D} = \bigoplus_{n=0}^\infty \mathscr{D}^n \quad \text{(direct sum)}, \\ &\mathscr{D}^0 = \mathbb{C}\phi_0, \quad \mathscr{D}^n = \{\tilde{\phi}^{\otimes 2n}(f_{2n}) \mid f_{2n}\in\mathscr{F}_{2n}\}.\end{aligned} \tag{2.7}$$

Its state space is $\mathscr{K} = \oplus_{n=0}^\infty\mathscr{K}_n$ (Hilbert sum). $\mathscr{K}_n$ = closure of $\mathscr{D}^n$ in $\mathscr{H}_{2n}$. It is defined according to the following formulas:

$$\begin{aligned}&A^F_{00}(f)\phi_0 = 0, \quad A^F_{10}(f)\phi_0 = \tilde{\phi}^{\otimes 2}(F(f)), \\ &A^F_{01}(f)\tilde{\phi}^{\otimes 2}(f_2) = \langle\phi^{\otimes 2}(F(f)^*), \tilde{\phi}^{\otimes 2}(f_2)\rangle\phi_0 \\ &\quad \forall f\in\mathscr{S}\,\forall f_{2n}\in\mathscr{F}_{2n} \\ &A^F_{m,n}(f)\tilde{\phi}^{\otimes 2n}(f_{2n}) = \tilde{\phi}^{\otimes 2m}(H_{m,n}(f;f_{2n})) \\ &\quad 1\leqslant n,m<\infty,\ |n-m|\leqslant 1,\end{aligned} \tag{2.8}$$

$$\begin{aligned}&H_{n-1,n}(f;f_{2n})(p_1,\ldots,p_{2n-2}) \\ &\quad = \sqrt{2n(2n-1)}\int_\Sigma\int_\Sigma F(f)(-q_2,-q_1) \\ &\qquad \times f_{2n}(p_1,\ldots,p_{2n-2},q_1,q_2)t(dq_1)t(dq_2), \\ &H_{n,n}(f;f_{2n})(p_1,\ldots,p_{2n}) \\ &\quad = \sqrt{2}\sum_{j=1}^{2n}\int_\Sigma \chi_\Sigma(p_j)F(f)(p_j,-p) \\ &\qquad \times f_{2n}(p_1,\ldots,p_{j-1},p,p_{j+1},\ldots,p_{2n})t(dp), \\ &H_{n+1,n}(f;f_{2n})(p_1,\ldots,p_{2n+2}) = [(2n+1)(2n+2)]^{-1/2} \\ &\quad \times \sum_{\substack{i,j=1\\ i\neq j}}^{2n+2}\chi_{\Sigma\times\Sigma}(p_i,p_j) \\ &\quad \times f_{2n}(p_1,\ldots,\check{p}_i,\ldots,\check{p}_j,\ldots,p_{2n+2})F(f)(p_i,p_j).\end{aligned} \tag{2.9}$$

These definitions really work according to

Proposition 2.1: A^F, as specified above, is, for every contraction map F, a well-defined Jacobi field over $\mathscr{S}$.

Proof: See Appendix A.

The linear contraints of QFT will impose some further restrictions on the contraction maps F, which express the fact that A^F, as defined above, is a relativistic quantum field. In order to guarantee Poincaré covariance of A^F we have to specify a strongly continuous unitary representation V of $P^{\uparrow}_{+}$ on $\mathscr{K}$ such that

$$V(\alpha)A^F(f)V(\alpha)^{-1} = A^F(f_\alpha)\forall\alpha\in P^{\uparrow}_{+}\ \forall f\in\mathscr{S} \tag{2.10}$$

holds ($f_\alpha(x) = f(\alpha^{-1}x)$). We now show that $V = U\restriction\mathscr{K}$ will do the job, where U is the unitary strongly continuous representation of $P^{\uparrow}_{+}$ associated with the generalized free field A in

$$\mathscr{H} = \bigoplus_{n=0}^{\infty}\mathscr{H}_n.$$

By definition U acts on $\tilde{\phi}^{\otimes 2n}(f_{2n}), f_{2n}\in\mathscr{F}_{2n}$, according to

$$U(\alpha)\tilde{\phi}^{\otimes 2n}(f_{2n}) = \tilde{\phi}^{\otimes 2n}((f_{2n})_\alpha),\quad \alpha = \{a,\Lambda\}, \tag{2.11}$$

$$(f_{2n})_\alpha(p_1,\dots,p_{2n}) = e^{ia(p_1+\cdots+p_{2n})}f_{2n}(\Lambda^{-1}p_1,\dots,\Lambda^{-1}p_{2n})$$

and thus leaves all $\mathscr{K}_n$ and therefore, $\mathscr{K} = \oplus_{n=0}\mathscr{K}_n$ invariant.

Now by $V = U\restriction\mathscr{K}$ relation (2.10) is equivalent to

$$\tilde{\phi}^{\otimes 2m}((H_{m,n}(f;f_{2n}))_\alpha) = \tilde{\phi}^{\otimes 2m}(H_{m,n}(f_\alpha;(f_{2n})_\alpha)) \tag{2.10'}$$

for all $f\in\mathscr{S}, f_{2n}\in\mathscr{F}_{2n}$, $\alpha\in P^{\uparrow}_{+}$ and $n,m = 0,1,2,\dots$. By Eq. (2.9) and some combinatorics this is equivalent to $F(p)_\alpha(p_1,p_2) = F(f_\alpha)(p_1,p_2)$, for $\hat{t}\times\hat{t}$ almost all $(p_1,p_2)\in\Sigma\times\Sigma$, and $(p_1,p_2)\in\Sigma\times(-\Sigma)$. Because of the Hermiticity relation $F(f)^* = F(f^*)$ this is equivalent to

$$F(f)_\alpha(p_1,p_2) = F(f_\alpha)(p_1,p_2), \tag{2.12}$$

for $\hat{t}\times\hat{t}$ almost all $(p_1,p_2)\in\hat{\Sigma}\times\hat{\Sigma}$. Therefore, A^F is a Poincaré covariant field according to (2.10) and $V = U\restriction\mathscr{K}$ if and only if the contraction map satisfies relation (2.12). It is quite obvious that (A^F,V) then satisfy the spectral condition. The spectrum of the generators P^F of the space-time translation in the representation V can be calculated. The result is

$$\Sigma^F = \sigma(P^F) = \overline{\bigcup_{n=1}^{\infty}\Sigma^F_n},$$

$$\Sigma^F_n = \{k = p_1+\cdots+p_n\,|\,p_j\in\Sigma^1_F\},$$

$$\Sigma^1_F = \{p = q_1+q_2\,|\,(q_1,q_2)\in\Sigma\times\Sigma\cap\ \text{supp}\ F(f),\ \text{for some}\ f\in\mathscr{S}\}.$$

To analyze locality needs some more calculations. To begin with notice that the only contributions $(A^F(f)A^F(g))_{m,n}$ of $A^F(f)A^F(g)$ which do not vanish identically are those with $m = n-2, n-1,\dots,n+2$. But as $A^F(f^*)\subseteq(A^F(f))^*$ implies $(A^F(g^*)A^F(f^*))_{nm}\subseteq((A^F(f)A^F(g))_{m,n})^*$ we only need to consider the cases $m = n+2, n+1, n$. These contributions have to be calculated according to (2.8) and (2.9):

$$(A^F(f)A^F(g))_{n+2,n}\tilde{\phi}^{\otimes 2n}(f_{2n}) = \tilde{\phi}^{\otimes 2(n+2)}(H_{n+2,n+1}(f;H_{n+1,n}(g;f_{2n}))), \tag{2.13a}$$

$$(A^F(f)A^F(g))_{n+1,n}\tilde{\phi}^{\otimes 2n}(f_{2n}) = \tilde{\phi}^{\otimes 2(n+1)}(H_{n+1,n+1}(f;H_{n+1,n}(f;f_{2n})) + H_{n+1,n}(f;H_{n,n}(g;f_{2n}))), \tag{2.13b}$$

$$(A^F(f)A^F(g))_{n,n}\tilde{\phi}^{\otimes 2n}(f_{2n}) = \tilde{\phi}^{\otimes 2n}(H_{n,n}(f;H_{n,n}(g;f_{2n})) + H_{n,n-1}(f;H_{n-1,n}(g;f_{2n})) + H_{n,n+1}(f;H_{n+1,n}(g;f_{2n}))). \tag{2.13c}$$

As $(f,g)\mapsto H_{n+2,n+1}(f;H_{n+1,n}(g;f_{2n}))$ is symmetric with respect to permutations of f and g, the contribution $(A^F(f)A^F(g))_{n+2,n}$ is always local. In Appendix B, it is shown that locality of the two other contributions is equivalent to locality of the following three functionals of $f, g\in\mathscr{S}$:

$$L^F_1(f\otimes g) = \mathscr{W}^F_2(f\otimes g) = \int_\Sigma\int_\Sigma t(dp)t(dq)F(f)(-p,-q)F(g)(q,p) = \int_\Sigma t(dp)K^F(f,g)(-p,p), \tag{2.14a}$$

$$L^F_2(f\otimes g;g_2) = \int_\Sigma\int_\Sigma t(dp_1)t(dp_2)g_2(p_1,p_2)K^F_S(f,g)(p_1,p_2), \tag{2.14b}$$

$$L^F_3(f\otimes g;g_2,f_2) = \int_\Sigma\int_\Sigma t(dp_1)t(dp_2)g_2(p_1,p_2) \times\int_\Sigma t(dq)K^F_S(f,g)(p_1,-q)f_2(q,p_2), \tag{2.14c}$$

for any $g_2, f_2\in\mathscr{F}_2$, where

$$K^F(f,g)(p_1,p_2) = \int_\Sigma t(dq)F(f)(p_1,-q)F(f)(q,q_2), \tag{2.15}$$

$$K^F_S(p,g)(p_1,p_2) = \tfrac{1}{2}\{K^F(f,g)(p_1,p_2)+K^F(f,g)(p_2,p_1)\}.$$

This then proves:

Proposition 2.2: For any gf field A and any contraction map F, the Jacobi field A^F, as defined according to Proposition 2.1, is a relativistic quantum field if and only if the constraints (2.12) and (2.14) hold.

Remark 2.2: It is only the locality condition which fixes the three constants in (2.9) uniquely (up to a common positive factor). By definition 2.1, the space of all contraction maps for a gf field A is just

$$\mathscr{C}(A) = \mathscr{L}_h(\mathscr{S},\mathscr{G}_2(A)).$$

According to relation (2.12) Poincaré covariance is expressed by a linear constraint on the contraction map. Therefore, the set of contraction maps defining Poincaré covariant fields A^F is a subspace of $\mathscr{C}(A)$. It is denoted by

$$\mathscr{C}_i(A) = \{F\in\mathscr{C}(A)\,|\,A^F\ \text{is Poincaré covariant}\}.$$

However, the set

$$\mathscr{C}_c(A) = \{F\in\mathscr{C}(A)\,|\,A^F\ \text{is a local field}\}$$

is not a subspace of $\mathscr{C}(A)$, as the locality constraint (2.14) is nonlinear in F.

Notice that the spaces $\mathscr{G}_2$ and $\mathscr{F}_2$ depend on the measure t defining the gf field A under consideration. This implies that the locality constraint (2.14) is highly nonlinear in t. This is a major problem in the analysis of this constraint. Using additional assumptions a class $\mathscr{C}^0(A)$ of contraction maps such that

$$\mathscr{C}^0(A)\subset \bigcap_{\{A'\}} \mathscr{C}_{i,c}(A'), \quad \mathscr{C}_{i,c}(A)=\mathscr{C}_i(A)\cap\mathscr{C}_c(A),$$

($\{A'\}$: set of all generalized free fields over $\mathscr{S}$) has been determined in Ref. 4. This class $\mathscr{C}^0(A)$ consists of those contraction maps for which A^F is the Wick product of A with itself or appropriate derivatives thereof. In Sec. V, a class of contraction maps is exhibited which goes far beyond these models and which uses a much more refined version of combining locality and positivity.

III. REALIZATION OF POINCARÉ COVARIANCE AND PHYSICAL SPECTRUM

A first lemma characterizes contraction maps in general. A second lemma takes into account the covariance constraint (2.12) and presents a characterization thereof. This then allows a discussion of our "ansatz" for contraction maps which will be the starting point for incorporating the locality condition (2.14).

Lemma 3.1: (a) For any contraction map F there is a function $\Gamma=\Gamma_F:\hat{\Sigma}\times\hat{\Sigma}\to\mathscr{S}'(\mathbb{R}^4)$ and a $\hat{t}\times\hat{t}$ null set $W=W_F\subset\hat{\Sigma}\times\hat{\Sigma}$ such that

$$F(f)(p_1,p_2)=\langle\Gamma(p;p_1,p_2),\tilde{f}(p)\rangle\ \forall(p_1,p_2)\in W^c\forall f\in\mathscr{S}. \tag{3.1}$$

The following sets of distributions are equicontinuous

(i) $\{\Gamma(\cdot;p_1,p_2)|(p_1,p_2)\in\hat{\Sigma}\times\hat{\Sigma}\}$.

(ii) $\{(1+p_1^0)^{N_1}(1+p_2^0)^{N_2}\Gamma(\cdot;\pm p_1,\pm p_2)|(p_1,p_2)\in\Sigma\times\Sigma\}$, $N_1,N_2\in\mathbb{N}$ fixed.

(iii) $\{(1+|p_1-p_2|)^{N_3}\Gamma(\cdot;p_1,-p_2)|(p_1,p_2)\in\Sigma\times\Sigma\}$ $N_3\in\mathbb{N}$ fixed. (3.2)

Furthermore, we have the following.

(iv) $\Gamma^*(\cdot;p_1,p_2)=\Gamma(\cdot;-p_2,-p_1)$.

(v) $\Gamma(\cdot;p_1,p_2)=\Gamma(\cdot;p_2,p_1)$.

(b) Conversely any function $\Gamma:\hat{\Sigma}\times\hat{\Sigma}\to\mathscr{S}'(\mathbb{R}^4)$ which satisfies the properties (i)–(v) of (3.2) defines according to (3.1) a contraction map.

Proof: According to the definition of a contraction map F there is for any $f\in\mathscr{S}$ a $\hat{t}\times\hat{t}$ null set $W_f\subset\hat{\Sigma}\times\hat{\Sigma}$ such that $(p_1,p_2)\mapsto F(f)(p_1,p_2)$ is a bounded function on W_f^c:

$$|F(f)(p_1,p_2)|\leqslant\|F(f)\|_\infty\leqslant\|\tilde{f}\|_{\mathscr{S}},$$

where $\|\cdot\|_{\mathscr{S}}$ is some continuous norm on $\mathscr{S}$. Because of the separability of $\mathscr{S}$ there is a common $\hat{t}\times\hat{t}$ null set $W\subset\hat{\Sigma}\times\hat{\Sigma}$ such that $(p_1,p_2)\mapsto F(f)(p_1,p_2)$ is bounded on W^c for all $f\in\mathscr{S}$. This implies that for any fixed $(p_1,p_2)\in W^c$, $f\mapsto F(f)(p_1,p_2)$ defines a continuous linear functional on $\mathscr{S}$, e.g., there is $\Gamma(\cdot;p_1,p_2)\in\mathscr{S}'(\mathbb{R}^4)$ such that Eq. (3.1) holds. On W we define $\Gamma(\cdot;p_1,p_2)$ to be zero. This proves the first part of (a). The rest of part (a) is now an easy consequence of the definition taking into account the specific continuity properties of a contraction map, respectively, the symmetry relations $F(f)^*=F(f^*)$ and $F(f)(p_1,p_2)=\overline{F(f)(p_2,p_1)}$. As usual for $T\in\mathscr{S}'$ we define T^* by $T^*(f)=\overline{T(f^*)}\forall f\in\mathscr{S}$. The proof of part (b) starts by defining a map F according to (3.1). The properties (3.2) then imply that F is a contraction map.

Lemma 3.2: (a) The "kernel" $\Gamma=\Gamma_F$ of any contraction map F, which satisfies the covariance condition (2.12) has the following form:

$$\Gamma(p;p_1,p_2)=\sum_{m=0}^{M}(-1)^m\gamma^{(m)}(p_1,p_2)\cdot\partial^{\otimes m}\delta_{p_1+p_2}(p),$$

e.g., (3.3)

$$F(f)(p_1,p_2)=\sum_{m=0}^{M}\gamma^{(m)}(p_1,p_2)\cdot(\partial^{\otimes m}\tilde{f})(p_1+p_2),$$

where $\gamma^{(m)}$ is a function from $\hat{\Gamma}\times\hat{\Gamma}$ into the space of symmetric tensors of degree m,

$$\gamma^{(m)}(p_1,p_2)=(\gamma_{j_1\cdots j_m}(p_1,p_2),\quad j_1,\ldots,j_m=0,1,2,3),$$

such that for all $(p_1,p_2)\in\hat{\Sigma}\times\hat{\Sigma}$ the following applies.

(i) $\gamma^{(m)}(p_1,p_2)=\gamma^{(m)}(p_2,p_1)$.

(ii) $\overline{\gamma^{(m)}(-p_2,-p_1)}=(-1)^m\gamma^{(m)}(p_1,p_2)$. (3.4)

(iii) $\gamma^{(m)}(\Lambda p_1,\Lambda p_2)=\Lambda^{\otimes m}\gamma^{(m)}(p_1,p_2)\forall\Lambda\in L^\uparrow_+$.

[$\cdot$ in (2.3) indicates contraction of Lorentz tensors.]

(b) Conversely, any finite system of tensor functions $\{\gamma^{(m)}\}$ which satisfies the constraints (3.4) and has in addition the following growth property (3.5):

(i) on $(-\Sigma)\times\Sigma$ and $\Sigma\times(-\Sigma)$ $\gamma^{(m)}$ is bounded,

(ii) there are integers K,k and numbers C_K,d_k such that for all $(p_1,p_2)\in\Sigma\times\Sigma$ and all $j_\nu\in\{0,1,2,3\}$ and $m=0,..,M$

$$|\gamma_{j_1\cdots j_m}(\pm p_1,\pm p_2)|\leqslant C_K(1+p_1^0)^K(1+p_2^0)^K\quad(\text{either } ++ \text{ or } --), \tag{3.5}$$

$$|\gamma_{j_1\cdots j_m}(p_1,-p_2)|\leqslant d_k(1+|p_1-p_2|)^k,$$

defines by Eq. (3.3) a contraction map which satisfies the covariance constraint (2.12).

Proof: The covariance constraint (2.12) implies first that $(\Lambda p_1,\Lambda p_2)\in W^c$ whenever $(p_1,p_2)\in W^c$ for all $\Lambda\in L^\uparrow_+$. Therefore, fix $(p_1,p_2)\in W^c$ arbitrarily and express condition (2.12) in terms of the kernel Γ of F. The result is

$$e^{ia\cdot(p_1+p_2)}\Gamma(\Lambda p;\Lambda p_1,\Lambda p_2)=e^{ia\cdot p}\Gamma(p;p_1,p_2)\quad(+)$$

for all $\alpha=(a,\Lambda)\in P^\uparrow_+$ and all $p_j\in\Sigma$. Now fix $\Lambda=1_4$ and let a vary in $\mathbb{R}^4$. This then implies

$$\operatorname{supp}\Gamma(\cdot;p_1,p_2)\subseteq\{p\in\mathbb{R}^4|p=p_1+p_2\}.$$

Therefore, the tempered distribution $\Gamma(\,;p_1,p_2)$ is a sum of derivatives of delta distributions concentrated in $p=p_1+p_2$:

$$\Gamma(p;p_1,p_2)=\sum_{|\alpha|<M}(-1)^{|\alpha|}\gamma_\alpha(p_1,p_2)D_p^\alpha\delta_{p_1+p_2}(p).$$

This can be rewritten in the form (3.3). As $\partial^{\otimes m}\tilde{f}=(\partial_{j_1,\ldots,j_m}\tilde{f};\ j_1..j_m=0,1,2,3)$ is a symmetric tensor we can and will take $\gamma^{(m)}$ to be a symmetric tensor too. The covariance constraint $(+)$ reads for $\forall\Lambda\in L^\uparrow_+$, $\forall(p_1,p_2)\in\hat{\Sigma}\times\hat{\Sigma}\forall f\in\mathscr{S}$:

$$\sum_{m=0}^{M}\{\Lambda^{\otimes m}\gamma^{(m)}(\Lambda^{-1}p_1,\Lambda^{-1}p_2)-\gamma^{(m)}(p_1,p_2)\}\cdot(\partial^{\otimes m}\tilde{f})(p_1+p_2)=0.$$

Choosing now for arbitrary but fixed $(p_1,p_2)\in\hat{\Sigma}\times\hat{\Sigma}$ the test functions $\tilde{f}$ appropriately we get (3.4) (iii). Similarly (i) and (ii)

of (3.4) follow from (v) and (iv) of (3.2). This proves part (a). In order to prove (b) note first that the growth conditions (3.5) are sufficient for (i)–(iii) of (3.2); this allows us to define Γ, respectively, F by Eq. (3.3). The rest of the proof now is a straightforward calculation. Certainly, the tensor functions $\gamma^{(m)}$ of formula (3.3) with the properties (3.4) and (3.5) could be analyzed further. But, since realization of Lorentz covariance is not our main concern, we proceed instead with an "ansatz" for a contraction map which is quite obvious by Lemma 3.2:

$$F(f)(p_1,p_2)=\sum_{\nu=0}^{N}\gamma_\nu((p_1+p_2)^2;p_1^2,p_2^2)\times\mathscr{F}(L_\nu f)(p_1+p_2)\forall p_j\in\hat{\Sigma}\forall f\in\mathscr{S} \tag{3.6}$$

with the following specifications:

(i) L_ν is a local $L^\uparrow_+$-invariant differential operator on $\mathscr{S}$, e.g.,

$$(L_\nu f)(x)=(Q_\nu(D)f)(x),\quad D=\left(\frac{\partial}{\partial x^0},\ldots,\frac{\partial}{\partial x^3}\right),$$

$$L_\nu(f_\Lambda)=(L_\nu f)_\Lambda\ \forall f\in\mathscr{S}\forall\Lambda\in L^\uparrow_+,$$

Q_ν is a polynomial in D.

(ii) γ_ν are measurable functions $\mathbb{R}\times\mathbb{R}_+\times\mathbb{R}_+\to\mathbb{R}$, and there is a constant $0\leqslant C<\infty$ and there are integers $k,K\in\mathbb{N}$ such that for all $(r,t,s)\in\mathbb{R}\times\mathbb{R}_+\times\mathbb{R}_+$ and all $\nu=0,1,..,N$

$$\gamma_\nu(r;s,t)=\gamma_\nu(r;t,s),$$

$$|\gamma_\nu(r;s,t)|\leqslant C(1+|r|)^{K/2}(1+s)^{K/2}(1+t)^{K/2}.$$

We will show that essentially all contraction maps F of the form (3.6) will define local fields. This is quite obvious for $\gamma_\nu=$ const.[4]

IV. SOME REMARKS ON LOCALITY

In order to analyze locality of A^F, $\mathscr{F}\in\mathscr{C}(A)$, we need to know a kind of test for locality of various distributions. As these distributions are given as the composition of a given (local) distribution with a linear continuous map on $\mathscr{S}$, we have to know locality of such maps and we have to test given distributions whether they are local or not. This section recalls the basic definitions and gives some results which turn out to be useful in the following section.

Two test functions $f,g\in\mathscr{S}=\mathscr{S}(\mathbb{R}^4)$ are said to have spacelike separated supports, denoted by $f\times g$, iff

$$\operatorname{supp}f\otimes g\subseteq:\mathscr{K}=\{(x_1,x_2)\in\mathbb{R}^4\times\mathbb{R}^4|(x_2-x_1)^2<0\}.$$

A distribution $T_2\in\mathscr{S}'_2=\mathscr{S}'(\mathbb{R}^4\times\mathbb{R}^4)$ is said to be local iff

$$\bigwedge_{f,g\in\mathscr{S}}f\times g\Rightarrow T_2(f\otimes g)=T_2(g\otimes f).$$

A continuous linear map $l:\mathscr{S}\to\mathscr{S}$ is called (i) local iff $\operatorname{supp}l\phi\subseteq\operatorname{supp}\phi\ \forall\phi\in\mathscr{S}$, (ii) spacelike separation preserving (ssp) iff

$$\bigwedge_{f,g\in\mathscr{S}}f\times g\Rightarrow lf\times lg,$$

(iii) local with respect to a local distribution $T_2\in\mathscr{S}'_2$ iff $l'\otimes l'T_2$ is again local ($l':\mathscr{S}'\to\mathscr{S}'$ the dual of l).

Just by definition it is easy to see: Every local map is ssp and every ssp map is local with respect to any local destribution.

Remark 4.1: The notation of locality with respect to a local distribution may be generalized and this version mostly occurs in applications. A map l on $\mathscr{S}$ is called local with respect to $T_2\in\mathscr{S}'_{2,c}$

$$\mathscr{S}'_{2,c}=\{T_2\in\mathscr{S}'_2|T_2\ \text{is local}\}$$

iff there is a locally convex topological vector space $\mathscr{F}$ such that (i) $\mathscr{S}$ is a dense subspace of $\mathscr{F}$, (ii) T_2 admits a continuous linear extension on $\mathscr{F}\otimes_\pi\mathscr{F}$, (iii) $l:\mathscr{S}\to\mathscr{F}$ is linear and continuous, (iv) $l'\otimes l'T_2\in\mathscr{S}'_{2,c}$ ($l':\mathscr{F}'\to\mathscr{S}'$ dual map). Typically $\mathscr{F}$ is the completion of $\mathscr{S}$ with respect to a continuous seminorm p on $\mathscr{S}$ such that $|T_2|\leqslant p\otimes_\pi p$.

The general form of a local map is well known (Peetre): A continuous linear map $l:\mathscr{S}\to\mathscr{S}$ is local if and only if it is a differential operator on $\mathscr{S}$, e.g.,

$$(l\phi)(x)=\sum_{|\alpha|<M}a_\alpha(x)(D^\alpha\phi)(x),\quad a_\alpha\in\mathscr{O}_M,$$

and the coefficients a_α are constant iff l commutes with all translations.

The general form of a ssp map l is much harder to analyze. As this is not in the scope of this section we only mention that any map of the form

$$(l\phi)(x)=\phi(\lambda\cdot\Lambda^{-1}(x-a)),\quad \lambda>0,\ \Lambda\in L,\ a\in\mathbb{R}^4$$

is ssp and the conjecture is that every ssp map is the composition of such a map with a local map.

The following example which is due to Borchers[5] shows that the set of maps which are local with respect to given $T_2\in\mathscr{S}'_{2,c}$ is indeed much larger than the set of ssp maps. Suppose $T_2\in\mathscr{S}'_{2,c}$ is the twofold VEV of a relativistic quantum field, e.g., for every $f,g\in\mathscr{S}$:

$$T_2(f\otimes g)=\int t(dp)\tilde{f}(-p)\tilde{g}(p)$$

$$=\int_0^\infty\rho(d\kappa)\langle\delta_\kappa^+(p),\tilde{f}(-p)\tilde{g}(p)\rangle,$$

where $t(dp)(\rho(d\kappa))$ are positive tempered measures on $\overline{V}_+$ [$[0,\infty)$, respectively]. Suppose now $h:[0,\infty)\to\mathbb{R}$ to be a polynomially bounded function in $L^2_{\mathrm{loc}}([0,\infty),d\rho)$ and define

$$(l\phi)(x)=\mathscr{F}(h(p^2)\tilde{\phi}(p))(x)\equiv(h(-\Box)\phi)(x).$$

Then l is local with respect to T_2:

$$(l'\otimes l'T_2)(f\otimes g)=\int_0^\infty h(\kappa)^2\rho(d\kappa)\langle\delta_\kappa^+(p),\tilde{f}(-p)\tilde{g}(p)\rangle$$

because $\rho_1(d\kappa)=h(\kappa)^2\rho(d\kappa)$ has the same properties as ρ and in case h is not a polynomial this map l is not ssp.

As we are mainly interested in maps which are local with respect to a local distribution we present a simple analysis of this notion. For any $f_2\in\mathscr{S}_2=\mathscr{S}(\mathbb{R}^4\times\mathbb{R}^4)$, let us denote by $f_2^\pm$ the symmetric (antisymmetric) part of f_2:

$$f_2^\pm(x,y)=\tfrac{1}{2}(f_2(x,y)\pm f_2(y,x)).$$

For $T_2\in\mathscr{S}'_2$ $T_2^\pm$ are defined in the same way. First recall

$$T_2\in\mathscr{S}'_2\ \text{is local}\Leftrightarrow\operatorname{supp}T_2^-\subseteq\mathscr{K}^c. \tag{4.1}$$

Proof: If T_2 is local we have by definition

$$T_2^-(f\otimes g)=0\ \text{for any}\ f,g\in\mathscr{S}\ \text{such that}$$

$$\operatorname{supp}f\otimes g\subseteq\mathscr{K}.$$

By continuous linear extension we get $T_2^-(f_2)=0$ for any $f_2\in\mathscr{S}$ with $\operatorname{supp}f_2\subseteq\mathscr{K}$ that is $\operatorname{supp}T_2^-\subseteq\mathscr{K}^c$. If $T_2\in\mathscr{S}'_2$

satisfies supp $T_2^- \subseteq \mathscr{K}^c$ we use the decomposition $T_2 = T_2^+ + T_2^-$ to conclude for all f,g such that $f \times g$:

$$T_2(f \otimes g) = T_2^+(f \otimes g) + T_2^-(f \otimes g)$$
$$= T_2^+(f \otimes g) = T_2^+(g \otimes f) = T_2(g \otimes f),$$

e.g., $T_2 \in \mathscr{S}'_{2,c}$.

Now it is easy to prove:

Lemma 4.2: For any continuous linear map $l: \mathscr{S} \to \mathscr{S}$ the following conditions are equivalent.

(1) $\forall f_2 \in \mathscr{S}_2$: supp $f_2^- \subseteq \mathscr{K} \Rightarrow$ supp $l^{\otimes 2}(f_2^-) \subseteq \mathscr{K}$.

(2) $\forall f_2 \in \mathscr{S}_2$: supp $f_2 \subseteq \mathscr{K} \Rightarrow$ supp $(l^{\otimes 2} f_2) \subseteq \mathscr{K}$.

(3) $\forall T_2 \in \mathscr{S}'_{2,c}: l' \otimes l' T_2' \in \mathscr{S}'_{2,c}$.

Proof: (1)$\Rightarrow$(2): If $f_2 \in \mathscr{S}_2$ satisfies supp $f_2 \subseteq \mathscr{K}$ it satisfies supp $f_2^- \subseteq \mathscr{K}$ too and therefore, by (1) supp $l^{\otimes 2}(f_2^-) \subseteq \mathscr{K}$, but

$$l^{\otimes 2}(f_2^-) = (l^{\otimes 2}(f_2))^-.$$

(2)$\Rightarrow$(3): Again we have $(l' \otimes l' T_2)^- = l' \otimes l' T_2^-$ and thus for all $f_2 \in \mathscr{S}_2$, supp $f_2 \subseteq \mathscr{K}$:

$$(l' \otimes l' T_2)^-(f_2) = T_2^-(l^{\otimes 2} f_2^-) = 0$$

by assumption (2) and supp $T_2^- \subseteq \mathscr{K}^c$; therefore supp $(l' \otimes l' T_2)^- \subseteq \mathscr{K}^c$ and thus $l' \otimes l' T_2 \in \mathscr{S}'_{2,c}$.

(3)$\Rightarrow$(1): Take $f_2 \in \mathscr{S}_2$ with supp $f_2^- \subseteq \mathscr{K}$ arbitrarily, but fixed. By assumption (3): $(l' \otimes l' T_2)^-(f_2^-) = 0 \forall T_2 \in \mathscr{S}'_{2,c}$. As $(l' \otimes l' T_2)^-(f_2^-) = T_2^-(l^{\otimes 2}(f_2^-))$ this implies

$$T_2^-(l^{\otimes 2} f_2^-) = 0 \forall T_2 \in \mathscr{S}'_2 \text{ supp } T_2^- \subseteq \mathscr{K}^c,$$

that is supp $l^{\otimes 2}(f_2^-) \subseteq \mathscr{K}$.

This simple analysis shows that a map which is local with respect to every local distribution is not necessarily a ssp map and thus not necessarily local. This distinction will be used for a new realization of locality in our models A^F, $F \in \mathscr{C}(A)$.

Now we want to point out by example that at least certain ssp maps can be obtained as limits of local maps. It is, in particular, in this repect that we are going to enlarge the set of local contraction maps. [It is quite obvious that a contraction map which is ssp in both variables separately realizes the locality constraint (2.14) and thus defines a local field A^F. This class of contraction maps has been determined in Ref. 4]. The translations τ_a, $a \in \mathbb{R}^4$, $(\tau_a f)(x) = f(x + a)$, are ssp maps but are not local. But nevertheless it is known that the translations are limits of local maps $L_n(a)$, $n \to +\infty$:

$$(L_n(a)f)(x) := \sum_{\nu=0}^{n} \frac{1}{n!} \langle (D^\nu f)(x), a^{\otimes \nu} \rangle,$$

$$\tau_a f = \lim_{n \to \infty} L_n(a) f$$

(for instance with respect to uniform convergence on $\mathbb{R}^4$ for any fixed $f \in \mathscr{S}$ whose Fourier transform $\tilde{f}$ has compact support).

This section is finished by some sufficient conditions for locality. The statement roughly is that $L^\uparrow_+$-invariance and temperedness of a distribution of *one* 4-vector imply locality.

Proposition 4.3: Every $L^\uparrow_+$-invariant tempered distribution is local, e.g., $T \in \mathscr{S}'(\mathbb{R}^4)$, T $L^\uparrow_+$-invariant $\Rightarrow T(-\xi) = T(\xi) \forall \xi \in \mathbb{R}^4$, $\xi^2 < 0$.

Proof: The Fourier transform $\tilde{T} = \mathscr{F} T$ of T is $L^\uparrow_+$-invariant too, thus by a slight modification of Theorem 5 of Güttinger and Rieckers,[6] $\tilde{T}$ has the form

$$\tilde{T}(p) = T_1(p^2) + \epsilon(p^0) T_2(p^2) + \sum_{n=0}^{N} C_n \Box^n \delta(p).$$

$T_2 \in \mathscr{S}'(\mathbb{R}_+)$ and T_1 can be chosen to be in $\mathscr{S}'(\mathbb{R})$. For the definition of $T_1(p^2)$ and $\epsilon(p^0) T_2(p^2)$ see Ref. 6.

This representation implies the following.

(i) $\tilde{T}(\check{p}) = \tilde{T}(p)$, where $\check{p} = (p^0, -\mathbf{p})$ (space-reflection symmetry).

(ii) There is $N \in \mathbb{N}$ such that $\phi_N(p) = [1 + (p^2)^2]^{-N}$ is an admissible test function for $\tilde{T}$ (that is $\tilde{T}$ has a continuous linear extension which is $L^\uparrow_+$-invariant and is defined on ϕ_N).

Therefore, $T_N(\xi) = \langle \tilde{T}(p), e^{ip \cdot \xi} \phi_N(p) \rangle$ is a well-defined polynomially bounded continuous function of $\xi \in \mathbb{R}^4$. The corresponding properties of $\tilde{T}$ and ϕ_N imply that $T_N(\xi)$ is $L^\uparrow_+$- and space-reflection invariant. In particular, the restriction of T_N to the hyperplane $\{\xi_0 = 0\}$ is space-reflection invariant, e.g., $T_N(0, -\xi) = T_N(0, \xi) \forall \xi \in \mathbb{R}^3$. Now any spacelike point $\eta \in \mathbb{R}^4$ is the image under a $\Lambda \in L^\uparrow_+$ of a spacelike point ξ of the form $\xi = (0, \xi)$. Therefore, by $L^\uparrow_+$-invariance of T_N we get for such η: $T_N(\eta) = T_N(-\eta)$.

Observe now that T itself is obtained by applying the $L^\uparrow_+$-invariant differential operator $(1 + \Box^2)^N$ to the function T_N: $T(\xi) = (1 + \Box^2)^N T_N(\xi)$, e.g.,

$$\langle T(\xi), f(\xi) \rangle = \int d^4\xi \, T_N(\xi)(1 + \Box^2)^N f(\xi) \forall f \in \mathscr{S}.$$

Thus we get $\langle T(-\xi), f(\xi) \rangle = \langle T(\xi), f(\xi) \rangle$ for all $f \in \mathscr{S}$ such that supp $f \subseteq \{\xi \mid \xi^2 < 0\}$ which proves locality of T.

Remarks 4.4: This proposition is a simple consequence of the following lemma which is a certain generalization of the BEG lemma.[7]

Lemma: For every $L^\uparrow_+$-invariant tempered distribution $T \in \mathscr{S}'(\mathbb{R}^4)$, there is $N \in \mathbb{N}$ and a continuous polynomially bounded function T_N such that (i) $T = (1 + \Box^2)^N T_N$, and (ii) T_N is $L^\uparrow_+$-invariant.

Obviously these results do not generalize to the case of $L^\uparrow_+$-invariant tempered distributions on $\mathbb{R}^{4n}$, $n > 1$. But the following related result turns out to be useful.

Proposition 4.5: Suppose $\tau \in \mathscr{S}'(\mathbb{R}^{4n})$ is a tempered space-reflection- and $L^\uparrow_+$-invariant distribution with supp $\tau \subseteq \overline{V}_+^{xn}$. Suppose further that τ has a representation

$$\tau(p_1, \ldots, p_n) = \prod_{j=1}^{n} \Box_j^{N_j} \tau_{\{N_j\}}(p_1, \ldots, p_n),$$

with $\tau_{\{N_j\}}$ a $L^\uparrow_+$- and space-reflection-invariant continuous polynomially bounded function with support in $\overline{V}_+^{xn}$ such that there is $0 < C_0 < \infty$ and $M_{ij} \in \mathbb{N}$ such that

$$|\tau_{\{N_j\}}(p_1, \ldots, p_n)| \leq C_0 \prod_{i,j=1}^{n} [1 + p_i \cdot p_j]^{M_{ij}} \forall p_i \in \overline{V}_+.$$

Then

$$T(\xi) = \int e^{i\xi \cdot \sum_{j=1}^{n} p_j} \tau(p_1, \ldots, p_n) d^4p_1 \ldots d^4p_n$$

is a well-defined tempered distribution on $\mathbb{R}^4$. It is $L^\uparrow_+$-invariant and local.

Remark: By the BEG lemma every space reflection and $L^{\uparrow}_{+}$-invariant tempered distribution with support in $\overline{V}^{n}_{+}$ has such a representation. Therefore, our assumption essentially reduces to the assumption that $\tau_{\{N_j\}}$ can be dominated by the above polynomial of the invariants $p_i \cdot p_j$.

Proof: By support properties of $\tau_{\{N_j\}}$ a tempered distribution on $\mathbb{R}^4$ is well defined by

$$T(f) = \int \tau_{\{N_j\}}(p_1,\dots,p_n) \prod_{j=1}^{n} \Box_j^{N_j} \tilde{f}(p_1 + \cdots + p_n) d^4p_1 \dots d^4p_n .$$

T is space reflection and $L^{\uparrow}_{+}$-invariant. For $p_i, p_j \in \overline{V}_+$ the following estimate holds:

$$p_i \cdot p_j = \frac{1}{2}\{(p_i + p_j)^2 - p_i^2 - p_j^2\}$$

$$\leqslant \frac{1}{2}(p_i + p_j)^2 \leqslant \frac{1}{2}\left(\sum_{j=1}^{n} p_j\right)^2 .$$

Therefore, $\tau_{\{N_j\}}$ is dominated by

$$C_1 \left\{1 + \left(\sum_{j=1}^{n} p_j\right)^2\right\}^m .$$

If $M \geqslant m$ is chosen appropriately $\psi_M(p_1,\dots,p_n)$ $= [1 + (\Sigma_{j=1}^{n} p_j)^2]^{-M}$ is an admissible test function for τ. Therefore,

$$T_M(\xi) = \langle \tau(p_1,\dots,p_n), e^{i\xi \cdot \sum_{j=1}^{n} p_j} \psi_M(p_1,\dots,p_n) \rangle$$

is a well-defined continuous polynomially bounded function of $\xi \in \mathbb{R}^4$. It is again space reflection and $L^{\uparrow}_{+}$-invariant. The argument of Proposition 4.3 implies locality of T_M and by $T = (1 - \Box)^M T_M$ locality of T follows as well as $L^{\uparrow}_{+}$-invariance.

V. A REALIZATION OF LOCALITY

In this section the contraction map F is supposed to be of the form (3.6). By Proposition 2.2 locality of the field A^F is equivalent to locality of the functionals L_i^F, $i = 1,2,3$. A first lemma ensures that L_1^F is local for every contraction map of the form (3.6).

Lemma 5.1: For any $F \in \mathscr{C}(A)$ of the form (3.6).

$$\mathscr{W}_2^F(f \otimes g) = L_1^F(f \otimes g)$$

$$= \int_\Sigma \int_\Sigma t(dp) t(dq) F(f)(-p,-q) F(g)(p,q)$$

is a local distribution.

Proof: We write L_1^F as

$$L_1^F(f \otimes g) = \sum_{\nu,\mu=0}^{N} \int_0^\infty \int_0^\infty \rho(d\kappa_1)\rho(d\kappa_2)$$

$$\times \langle D_{\nu\mu}((y-x);\kappa_1,\kappa_2); (L_\nu f)(x)(L_\mu g)(y) \rangle ,$$

$$D_{\nu\mu}(\xi,\kappa_1,\kappa_2) = \langle \delta^+_{\kappa_1}(p) \delta^+_{\kappa_2}(q), e^{i\xi(p+q)} \gamma_\nu((p+q)^2; \kappa_1^2,\kappa_2^2) \gamma_\mu((p+q)^2;\kappa_1^2,\kappa_2^2) \rangle .$$

By assumption on the $\{\gamma_\nu\}$, Proposition 4.5 applies and ensures locality of $D_{\nu\mu}(\cdot;\kappa_1,\kappa_2)$ for fixed κ_1,κ_2. Locality of L_1^F now follows easily.

This lemma implies: Within the class of contraction maps as specified by (3.6) locality of the field A^F is controlled by locality of the functionals L_2^F, L_3^F. By combinatorics locality of these functionals is a condensed way of expressing locality of all VEV of order $n \geqslant 3$.

A first step reduces locality of L_2^F, L_3^F to that of two other functionals which are somewhat simpler. By Eqs. (2.14b) and (2.14c) locality of L_2^F, L_3^F is seen to be equivalent to that of $(f,g) \to K_S^F(f,g)(p_1,p_2)$ for $t \times t$ almost all $(p_1, \pm p_2) \in \Sigma \times \Sigma$. By *-symmetry this is equivalent to locality of $K_s^F(f,g)(p_1,p_2)$ for $\hat{t} \times \hat{t}$ almost all $(p_1,p_2) \in \hat{\Sigma} \times \hat{\Sigma}$, which we prefer to express as locality of

$$K_F^{\pm}(f \otimes g; g_2)$$

$$= \int_{\Sigma \times 3} t(dp_1) t(dp_2) t(dp_3) g_2(p_1,p_2)$$

$$\times \{F(f)(p_1,-p_3) F(g)(p_3,\pm p_2)$$

$$+ F(f)(\pm p_2,-p_3) F(g)(p_3,p_1)\}, \tag{5.1}$$

for all $g_2 \in \mathscr{F}_2^0 = \{g_2 \in \mathscr{F}_2 | \operatorname{supp} g_2 \text{ compact}\}$.

A first realization of locality is obtained by specifying the contraction map according to the following assumptions. Suppose F is of the form (3.6) but with coefficients γ_ν of the following type:

$$\gamma_\nu(r,s,t) = \sum_{\mu=0}^{n_\nu} (-r)^\mu \gamma_{\nu,\mu}(s,t) \forall (r,s,t) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ , \tag{5.2}$$

where $\gamma_{\nu,\mu} : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ are measurable functions which are symmetric and polynomially bounded.

Evaluation of $\mathscr{K}_F^{\pm}(f,g)(p_1,p_2)$ yields

$$\mathscr{K}_F^{\pm}(f,g)(p_1,p_2)$$

$$= \int_0^\infty \rho(d\kappa) \langle \Delta_\kappa^+(y-x),\ e^{ip_1 x} \mathscr{L}(f;p_1^2,\kappa^2)(x)$$

$$\times \mathscr{L}(g;\kappa_1^2,p_2^2)(y) e^{ip_2 y} \rangle ,$$

where

$$\mathscr{L}(f;s,\kappa^2)(x) = \sum_{\nu=0}^{N} \sum_{\mu=0}^{n_\nu} \gamma_{\nu\mu}(s,\kappa^2)(\Box^\mu L_\nu f)(x)$$

obviously is a local operator on $\mathscr{S}$. Now as

$$\mathscr{K}_F^{\pm}(f \otimes g, g_2)$$

$$= \int_{\Sigma \times 2} t(dp_1) t(dp_2) g_2(p_1,p_2)$$

$$\times \{K^F(f,g)(p_1,\pm p_2) + K^F(f,g)(\pm p_2,p_1)\} \tag{5.1'}$$

holds, locality of $(f,g) \to \mathscr{K}_F^{\pm}(f \otimes g; g_2)$ follows for any $g_2 \in \mathscr{F}_2^0$.

The set of "local" contractions obtained this way is denoted by $\mathscr{C}_{i,c}^0(A)$, e.g.,

$$\mathscr{C}_{i,c}^0(A)$$

$$= \{F \in \mathscr{C}(A) | F \text{ is of the form (3.6)}$$

$$\text{with coefficients according to (5.2)}\}$$

and we can summarize:

Proposition 5.2: Any contraction map $F \in \mathscr{C}_{i,c}^0(A)$ defines a local relativistic quantum field A^F according to Proposition 2.1.

In order to enlarge the set of contraction maps which define local fields we want to take limits of sequences of contraction maps in $C_{i,c}^0(A)$. This requires some topological

considerations. It is well known[8] that the set of states on the Borchers–Uhlmann-algebra of local relativistic quantum fields is a weakly closed convex set. Therefore, if we take the closure of the set of states of fields characterized by Proposition 5.2 we still get local relativistic quantum fields. The problem now is to exhibit conditions on sequences of contraction maps of class $\mathscr{C}^0_{i,c}(A)$ which imply weak convergence of the associated sequence of states. In our approach it is quite natural to proceed as follows:

Definition 5.3: A sequence of contraction maps $\{F_j\}_{j\in\mathbb{N}} \subset \mathscr{C}(A)$ (a) *converges* to a contraction F iff $F_j(f)\to_{j\to\infty} F(f)$ in $\mathscr{G}_2$, (b) is a *Cauchy sequence* iff $\{F_j(f)\}_{j\in\mathbb{N}}$ is a Cauchy sequence in $\mathscr{G}_2$ for any $f\in\mathscr{S}(\mathbb{R}^4)$.

This is a convenient definition because of

Lemma 5.4: $\mathscr{C}(A)$ is complete.

Proof: As $L^\infty_{(2)}(\hat{\Sigma},\hat{t})$ is complete the space $\mathscr{G}_2$ too is complete. Remember that $\mathscr{C}(A)$ was defined to be the subspace of Hermitian elements in $\mathscr{L}(\mathscr{S},\mathscr{G}_2)$, the space of continuous linear functions $\mathscr{S}\to\mathscr{G}_2$, in the sense of $F(f)^* = F(f^*)\ \forall f\in\mathscr{S}$. By continuity of $f_1\to f_1^*$ in $\mathscr{S}$ and $f_2\to f_2^*$ in $\mathscr{G}_2$, $\mathscr{C}(A)$ is a closed subspace of $L(\mathscr{S},\mathscr{G}_2)$. As $\mathscr{S} = \mathscr{S}(\mathbb{R}^4)$ is barrelled $\mathscr{L}(\mathscr{S},\mathscr{G}_2)$ is weakly sequential complete which proves the lemma.

Furthermore, it turns out that the functionals L_i^F, respectively, $\mathscr{K}_F^\pm$ are continuous with respect to this notion of convergence:

Lemma 5.5: The functionals L_i^F, $i = 1,2,3$ and $\mathscr{K}_F^\pm$ are continuous in F, e.g., for any fixed $f,g\in\mathscr{S}$, $g_2,f_2\in\mathscr{F}_2$ the relation $F = \lim F_j$ implies $L_i^F(f\otimes g,\dots)$ $= \lim_{j\to\infty} L_i^{F_j}(f\otimes g,\dots)$, $i = 1,2,3$ and $\mathscr{K}_F^\pm(f\otimes g;g_2)$ $= \lim_{j\to\infty} \mathscr{K}_{F_j}^\pm(f\otimes g;g_2)$.

Proof: A first step establishes the basic estimates for these functionals. Using arguments which are similar to those of Appendix A we get by Eqs. (2.14), respectively, (5.1):

$$|L_1^F(f\otimes g)| \leqslant N_L^4 q_{L,L}^-(F(f))q_{L,L}^+(F(g)),$$

$$|L_2^F(f\otimes g;g_2)| \vee |\mathscr{K}_F^+(f\otimes g;g_2)| \leqslant 2N_L^6 q_{2L,2L}^+(g_2)Q_0(F(f))q_{2L,0}^+(F(g)),$$

$$|L_3^F(f\otimes g;g_2,f_2)| \leqslant \tfrac{1}{2}N_L^8\{q_{3L,L}^+(g_2)q_{L,3L}^+(f_2)Q_L(F(f))Q_L(F(g)) + q_{2L,L}^+(g_2)q_{L,2L}^+(f_2)q_{0,L}^-(F(f))q_{L,0}^+(F(g))\},$$

$$|\mathscr{K}_F^-(f\otimes g;g_2)| \leqslant N_L^6\{q_{3L,3L}^+(g_2)Q_L(F(f))Q_L(F(g)) + q_{2L,2L}^+(g_2)q_{0,L}^-(F(f))q_{L,0}^+(F(g))\}.$$

The second step is based on the fact that the functionals under consideration depend linearly on the "product" $F(f)\cdot F(g)$. Thus we get for instance:

$$|L_+^F(f\otimes g) - L_1^{F'}(f\otimes g)| \leqslant N_L^4\{q_{L,L}^-(F(f)-F'(f))q_{L,L}^+(F'(f)) + q_{L,L}^-(F'(f))q_{L,L}^+(F(g)-F'(g))\}$$

and therefore $F = \lim_{j\to\infty} F_j$ implies $L_1^F(f\otimes g)$ $= \lim_{j\to\infty} L_1^{F_j}(f\otimes g)$.

An immediate consequence of these continuity properties is

Corollary 5.6: The sets $\mathscr{C}_i(A)$ and $\mathscr{C}_c(A)$ of contraction maps defining Poincaré covariant, respectively, local fields are closed.

Proof: $P^\uparrow_+$ acts continuously on $\mathscr{G}_2$ and thus for all $a\in P^\uparrow_+$ $F_j(f)_a = F_j(f_a)\forall_j\in\mathbb{N}$ implies $F(f)_a$ $= \lim_{j\to\infty} F_j(f)_a = F(f_a)$. Therefore, if $F = \lim_{j\to\infty} F_j$ and $F_j\in\mathscr{C}_i(A)$ then $F\in\mathscr{C}_i(A)$. By Lemma 5.5 the functionals controlling locality of A^F are continuous in F. Therefore, $\mathscr{C}_c(A)$ is closed too.

By these closure properties the class $\mathscr{C}^0_{i,c}(A)$ of contraction maps defining relativistic quantum fields is easily extended.

Proposition 5.7: The fields A^F associated with contraction maps F in the "closure" $\overline{\mathscr{C}^0_{i,c}(A)}$ $= \{F\in\mathscr{C}(A)\,|\,F = \lim_{j\to\infty} F_j, \{F_j\}\subset\mathscr{C}^0_{i,c}(A)\}$ of $\mathscr{C}^0_{i,c}(A)$ are local relativistic quantum fields.

Proof: As the spectral condition is ensured by construction Proposition 2.2 and Corollary 5.6 prove Proposition 5.7.

Remark 5.8: Notice the following additional stability property of the set $\mathscr{C}_{i,c}(A)$ of all contraction maps defining local relativistic quantum fields: For every $F\in\mathscr{C}_{i,c}(A)$ and every local operator $l:\mathscr{S}\to\mathscr{S}$ which commutes with the action of $P^\uparrow_+$ on $\mathscr{S}$ we have

$$F\circ l\in\mathscr{C}_{i,c}(A). \tag{5.3}$$

For the proof recall $\Theta\in\mathscr{C}_{i,c}(A)$ iff $\Theta\in\mathscr{C}(A)$ and (i) $\Theta(f)_a$ $= \Theta(f_a)\forall f\in\mathscr{S}\forall a\in P^\uparrow_+$, (ii) L_i^Θ, $i = 1,2,3$ are local. For $\Theta = F\circ l$ (i) is obvious by $(lf)_a = l(f_a)$. To prove (ii) note $L_i^{F\circ l}(f\otimes g;\dots) = L_i^F(lf\otimes lg;\dots)$.

If we weaken the notion of convergence we have a chance to get more limit points and thus in the case under consideration more contraction maps defining local fields. The way this is done is modeled by the way we get the nonlocal (but still spacelike separation preserving) translation operators on $\mathscr{S}$ as limits of local operators. But for the case at hand we encounter the problem of localization of test functions both in coordinate and in momentum space. In order to meet this difficulty an additional argument has to be used.

Let us denote by $\mathscr{S}_2^0 = \text{lin}\{f\otimes g\,|\,g,f\in\mathscr{S}$, supp $\tilde{f}$ and supp $\tilde{g}$ compact$\}$. Then we will use the following notion of "weak" convergence.

Definition 5.9: A sequence of contraction maps is said to *converge* to a contraction map F *in the weak sense* iff for any $f_2\in\mathscr{S}_2^0$ and any $g_2\in\mathscr{F}_2^0$ $\mathscr{K}_{F_j}^\pm(f_2;g_2)\underset{j\to\infty}{\to}\mathscr{K}_F^\pm(f_2;g_2)$.

Note that for those elements in $\mathscr{S}_2$ for which we have convergence of $\{\mathscr{K}_{F_j}^\pm(\cdot;g_2)\}$ by definition the locality condition cannot be formulated. Nevertheless, locality of $\mathscr{K}_F^\pm(\cdot;g_2)$ is preserved.

Proposition 5.10: Suppose $F\in\mathscr{C}(A)$ is the limit in the weak sense of a sequence $\{F_j\}_{j\in\mathbb{N}}\subset\mathscr{C}_c(A)$. Then F too defines a local field A^F, e.g., $F\in\mathscr{C}_c(A)$.

Proof: (a) Take $g\in\mathscr{F}_2^0$ arbitrary but fixed and define $D = D(g_2) = \{f_2\in\mathscr{S}_2\,|\,\{\mathscr{K}_{F_j}^\pm(f_2;g_2)\}_{j\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{C}\}$. D is a subspace of $\mathscr{S}_2$ and a linear functional $\mathscr{K}_*^\pm$ is well defined on D by $\mathscr{K}_*^\pm(f_2;g_2)$ $= \lim_{j\to\infty}\mathscr{K}_{F_j}^\pm(f_2;g_2), f_2\in D$.

By assumption we know for all $f_2\in\mathscr{S}_2^0$: $\mathscr{K}_F^\pm(f_2;g_2)$ $= \lim_{j\to\infty}\mathscr{K}_{F_j}^\pm(f_2;g_2)$. This implies: (i) $\mathscr{S}_2^0\subset D$, and (ii) $\mathscr{K}_*^\pm(\cdot;g_2)\restriction\mathscr{S}_2^0 = \mathscr{K}_F^\pm\restriction\mathscr{S}_2^0$.

Thus $\mathscr{K}_{*}^{\pm}(\cdot;g) \restriction \mathscr{S}_2^0$ is continuous and admits a unique continuous linear extension to all of $\mathscr{S}_2$ and this extension is $\mathscr{K}_F^{\pm}$ as $\mathscr{S}_2^0$ is dense in $\mathscr{S}_2$. Therefore, we know

$$\mathscr{K}_F^{\pm} \restriction D = \mathscr{K}_{*}^{\pm}.$$

(b) Now take $f,g \in \mathscr{S}$ with spacelike separated supports. $F_j \in \mathscr{C}_c(A)$ implies $0 = \mathscr{K}_{F_j}^{\pm}([f,g];g_2)\ \forall j \in \mathbb{N}$, $[f,g] = f\otimes g - g\otimes f \in \mathscr{S}\otimes\mathscr{S}$, therefore, $[f,g] \in D$ and $\mathscr{K}_{*}^{\pm}([f,g];g_2) = 0$. By (a) we get

$$\mathscr{K}_F^{\pm}([f,g];g_2) = \mathscr{K}_{*}^{\pm}([f,g];g_2) = 0.$$

Thus $\mathscr{K}_F^{\pm}(\cdot;g_2)$ is local for any $g_2 \in \mathscr{F}_2^0$. By Proposition 2.2 we get $F \in \mathscr{C}_c(A)$. An immediate consequence of this proposition is

Corollary 5.11: Any contraction map F in $\mathscr{C}_i(A)$ which is the limit in the weak sense of a sequence $\{F_j\}_{j\in\mathbb{N}} \subset \mathscr{C}_{i,c}(A)$ belongs to $\mathscr{C}_{i,c}(A)$ and thus defines a local relativistic quantum field. In particular, we have $\mathscr{C}_i(A) \cap \overline{\mathscr{C}_{i,c}^0(A)}^{w} \subseteq \mathscr{C}_{i,c}(A)$, where $\overline{M}^{w}$ denotes the "closure" in sense of convergence of Definition 5.9.

VI. CONCLUSIONS

Proposition 5.7 and Corollary 5.11 tell us which contraction maps at least give rise to a relativistic quantum field. Clearly one could be interested in a more explicit characterization of the elements F in $\mathscr{C}_{i,c}(A)$ than that given by these propositions in terms of the property of being certain limits. Instead of doing this we prefer to indicate a class of examples which show explicitly that $\mathscr{C}_{i,c}(A)$ contains quite a lot of elements which at least for the first moment have a surprisingly general form. As we know (Remark 5.8) that $F \in \mathscr{C}_{i,c}(A)$ implies $F\circ L \in \mathscr{C}_{i,c}(A)$ for any L $\uparrow_+$-invariant constant coefficient differential operator L we only discuss contraction maps of the form

$$F(f)(p_1,p_2) = \gamma((p_1+p_2)^2;p_1^2,p_2^2)\tilde{f}(p_1+p_2), \tag{6.1}$$

where γ is specified according to Condition (3.6) (ii). Suppose that for every $\lambda \in \mathbb{R}_+ = [0,\infty]$ there are a real symmetric measurable function γ_λ on $\mathbb{R}_+^2$ which is polynomially bounded, a real measurable function c on $\mathbb{R}_+$, and a positive Borel measure σ on $\mathbb{R}_+$ such that

(a) for any compact subset $K \subset \mathbb{R}_+^3$ there is a measurable function ψ_K on $\mathbb{R}_+$ such that ($[\lambda]$ = largest $n \in \mathbb{N}$ such that $n \leqslant \lambda$)

$$\text{(i)}\quad \sup_{(r^2,s,t)\in K} |\gamma_\lambda(s,t)c(\lambda)(r^2)^{[\lambda]}| \leqslant \chi_K(\lambda), \tag{6.2}$$

$$\text{(ii)}\quad \int_{\mathbb{R}^+} \sigma(d\lambda)\chi_K(\lambda) < \infty.$$

$$\text{(b)}\quad \gamma(r;s,t) = \int_{\mathbb{R}_+} \sigma(d\lambda)\gamma_\lambda(s,t)\Gamma_\lambda(r^2),\quad \Gamma_\lambda(x) = c(\lambda)x^{[\lambda]}.$$

Lemma 6.1: Any contraction map F of the form (6.1) with γ satisfying condition (3.6) (ii) defines a relativistic quantum field, e.g., $F \in \mathscr{C}_{i,c}(A)$, whenever γ allows a representation according to (6.2).

Proof: In order to prove $F \in \mathscr{C}_{i,c}(A)$ we have to show that $\mathscr{K}_F^{\pm}(\cdot;g_2)$ is a local distribution for any $g_2 \in \mathscr{F}_2^0$. Choose $f,g \in \mathscr{S}$ arbitrarily, supp $\tilde{f}$ compact. As supp $\tilde{f}$ and supp g_2 are compact the domain of integration in $\mathscr{K}_F^{\pm}(f\otimes g;g_2)$ with respect to (p_1,p_2,p_3) is a compact set. Therefore, there is a compact set $K \subset \mathbb{R}_+^3$ such that for all such (p_1,p_2,p_3) $([(p_1-p_3)^2]^2,p_1^2,p_3^2) \in K$, $([(p_3 \pm p_2)^2]^2,p_2^2,p_3^2) \in K$. If we choose now χ_K according to (6.2) (i) we have the following estimate:

$$\begin{aligned}
&\int \sigma(d\lambda_1)\sigma(d\lambda_2)\left\{\int t(dp_1)t(dp_2)t(dp_3)|g_2(p_1,p_2)\gamma_{\lambda_1}(p_1^2,p_3^2)\gamma_{\lambda_2}(p_2^2,p_3^2)\right.\\
&\quad\left.\times\Gamma_{\lambda_1}([(p_1-p_3)^2]^2)\Gamma_{\lambda_2}([(p_3\pm p_2)^2]^2)\tilde{f}(p_1-p_3)\tilde{g}(p_3\pm p_2)|\right\}\\
&\leqslant \int \sigma(d\lambda_1)\sigma(\lambda_2)\left\{\int t(dp_1)t(dp_2)t(dp_3)\chi_K(\lambda_1)\chi_K(\lambda_2)|g_2(p_1,p_2)\tilde{f}(p_1-p_3)\tilde{g}(p_3\pm p_2)|\right\}
\end{aligned}$$

and this is finite because of (6.2) (ii).
Therefore, Fubini's theorem applies and by (6.2) (b) we get

$$\begin{aligned}
&\int \sigma(d\lambda_1)\sigma(d\lambda_2)\int t(dp_1)t(dp_2)t(dp_3)g_2(p_1,p_2)\gamma_{\lambda_1}(p_1^2,p_3^2)\gamma_{\lambda_2}(p_2^2,p_3^2)\\
&\quad\times\Gamma_{\lambda_1}([(p_1-p_3)^2]^2)\Gamma_{\lambda_2}([(p_3\pm p_2)^2]^2)\tilde{f}(p_1-p_3)\tilde{g}(p_3\pm p_2)\\
&= \int t(dp_1)t(dp_2)t(dp_3)g_2(p_1,p_2)\gamma((p_1-p_3)^2;p_1^2,p_3^2)\gamma((p_3\pm p_2)^2;p_2^2,p_3^2)\tilde{f}(p_1-p_3)\tilde{g}(p_3\pm p_2).
\end{aligned}$$

The second summand of $\mathscr{K}_F^{\pm}$ is treated in the same way and this then implies for all $f\otimes g \in \mathscr{S}_2^0$ and all $g_2 \in \mathscr{F}_2^0$

$$\begin{aligned}
\mathscr{K}_{\pm}^{F}(f\otimes g;g_2) &= \int \sigma(d\lambda_1)\sigma(d\lambda_2)\int t(dp_1)t(dp_2)t(dp_3)g_2(p_1,p_2)\\
&\quad\times\{\gamma_{\lambda_1}(p_1^2,p_3^2)\Gamma_{\lambda_1}([(p_1-p_3)^2]^2)\tilde{f}(p_1-p_3)\gamma_{\lambda_2}(p_2^2,p_3^2)\Gamma_{\lambda_2}([(p_3\pm p_2)^2])\tilde{g}(p_3\pm p_2)\\
&\quad+\gamma_{\lambda_1}(p_2^2,p_3^2)\Gamma_{\lambda_1}([(\pm p_2-p_3)^2]^2)\tilde{f}(\pm p_2-p_3)\gamma_{\lambda_2}(p_1^2,p_3^2)\Gamma_{\lambda_2}([(p_1+p_3)^2]^2)\tilde{g}(p_1+p_3)\}\\
&= \int \sigma(d\lambda_1)\sigma(d\lambda_2)\int \rho(d\kappa_1)\rho(d\kappa_2)\rho(d\kappa_3)\gamma_{\lambda_1}(\kappa_1^2,\kappa_3^2)\gamma_{\lambda_2}(\kappa_2^2,\kappa_3^2)T_{\pm}(f\otimes g;g_2;\kappa_j,\lambda_i),
\end{aligned}$$

where after changing $\lambda_1 \leftrightarrow \lambda_2$ in the second summand, T is given by the following expression:

$$T_{\pm}(f\otimes g;g_2;\kappa_j,\lambda_i)=\int dx\,dy\{G_{\kappa_1\kappa_2}(x,\pm y)f_{\lambda_1}(x)g_{\lambda_2}(y)+G_{\kappa_1\kappa_2}(y,\pm x)f_{\lambda_2}(x)g_{\lambda_1}(y)\}\Delta^{+}_{\kappa_3}(y-x),$$

$$G_{\kappa_1\kappa_2}(x,y)=\int\delta^{+}_{\kappa_1}(dp_1)\delta^{+}_{\kappa_2}(dp_2)e^{ixp_1+iyp_2}g_2(p_1,p_2),$$

$$f_\lambda(x)=(\Gamma_\lambda(\Box^2)f)(x)=c(\lambda)((\Box^2)^{[\lambda]}f)(x).$$

As $f\to f_\lambda$ is a local operator locality of $\Delta^{+}_{\kappa_3}$ implies that of $T(\cdot;g_2,\kappa_j,\lambda_i)$. Now we use Lemma 6.2 to conclude that $\mathscr{K}^{\pm}_F(\cdot;g_2)$ is a local distribution for any $g_2\in\mathscr{S}^0_2$. Our estimate above ensures that this lemma applies. Therefore, we have $F\in\mathscr{C}_{i,c}(A)$.

Lemma 6.2: Suppose $\Lambda\subset\mathbb{R}^m$ is a Borel subset and σ is a Borel measure on Λ, and for every $\lambda\in\Lambda$ we are given $\mathscr{K}_\lambda\in\mathscr{S}'_{2,c}$. Suppose furthermore that there is $\mathscr{K}\in\mathscr{S}'_2$ such that for all $f_2\in\mathscr{S}^0_2$

(i) $\int_\Lambda|\sigma(d\lambda)|\,|\mathscr{K}_\lambda(f_2)|<\infty$,

(ii) $\int_\Lambda\sigma(d\lambda)\mathscr{K}_\lambda(f_2)=\mathscr{K}(f_2)$.

Then $\mathscr{K}$ is a local distribution: $\mathscr{K}\in\mathscr{S}'_{2,c}$.

Proof: The proof is essentially the same as that of Proposition 5.10 when we start with the definition

$$D=\left\{f_2\in\mathscr{S}_2\,\middle|\,\int_\Lambda|\sigma(d\lambda)|\,|\mathscr{K}_\lambda(f_2)|<\infty\right\}.$$

We give now some particular cases of Proposition 5.10, respectively, Lemma 6.1. Suppose g is an entire function which is polynomially bounded on $\mathbb{R}_+$ and γ_0 and α are two symmetric continuous functions on $\mathbb{R}^2_+$ such that $\alpha\geqslant 0$ and α,γ_0 are polynomially bounded. Then by Eq. (6.1)

$$\gamma(r;s,t)=\gamma_0(s,t)g(\alpha(s,t)r^2) \tag{6.3}$$

gives rise to a contraction map F in $\mathscr{C}_{i,c}(A)$. For the proof one uses Proposition 5.10 directly or one chooses the building blocks σ, c, γ_λ in Lemma 6.1 appropriately.

Remark 6.3: Obviously Lemma 6.1 can be generalized to include more functions Γ_λ, in particular if more information about our basic measure $t(dp)$ is taken into account. Also it is possible to construct explicitly other classes of functions γ than those of (6.3).

A contraction map F in $\mathscr{C}_{i,c}(A)$ as described by Propositions 5.10 and 5.7 may be thought of as a special pseudodifferential operator and accordingly an appropriate suggestive notation is

$$F(f)(p_1,p_2)=\mathscr{F}(F(\Box_x;p_1^2,p_2^2)f)(p_1+p_2). \tag{6.4}$$

In order to give an idea about the models A^F of relativistic quantum fields constructed this way we write down the VEV's of A^F up to order 4 ($f_i\in\mathscr{S}$):

$$\mathscr{W}^F_2(f_1\otimes f_2)=\int_0^\infty\int_0^\infty\rho(d\kappa_1)\rho(d\kappa_2)\langle\Delta^{+}_{\kappa_1}(y-x)\Delta^{+}_{\kappa_2}(y-x),(F(\Box;\kappa_1^2,\kappa_2^2)f_1)(x)(F(\Box;\kappa_1^2,\kappa_2^2)f_2)(y)\rangle,$$

$$\begin{aligned}\mathscr{W}^F_3(f_1\otimes f_2\otimes f_3)&=2\sqrt{2}\int_{\mathbb{R}^3_+}\rho(d\kappa_1)\rho(d\kappa_2)\rho(d\kappa_3)\\&\times\langle\Delta^{+}_{\kappa_1}(x_1-x_2)\Delta^{+}_{\kappa_2}(x_1-x_3)\Delta^{+}_{\kappa_3}(x_3-x_2),(F(\Box;\kappa_1^2,\kappa_2^2)f_1)(x_1)(F(\Box;\kappa_1^2,\kappa_3^2)f_2)(x_2)\\&\times(F(\Box;\kappa_2^2,\kappa_3^2)f_3)(x_3)\rangle.\end{aligned}$$

Unfortunately the 4-point-function is a bit more involved. We write $\mathscr{W}^F_4$ as

$$\mathscr{W}^F_4(f_1\otimes\cdots\otimes f_4)=\mathscr{W}^F_2(f_1\otimes f_2)\mathscr{W}^F_2(f_3\otimes f_4)+\mathscr{W}^1_{2,2}(f_1\otimes\cdots\otimes f_4)+\mathscr{W}^2_{2,2}(f_1\otimes\cdots\otimes f_4),$$

where

$$\begin{aligned}\mathscr{W}^1_{2,2}(f_1\otimes\cdots\otimes f_4)&=\langle A^F_{11}(f_2^*)A^F_{10}(f_1^*)\phi_0,A^F_{11}(f_3)A^F_{10}(f_4)\phi_0\rangle\\&=2\int_{\mathbb{R}^4_+}\prod_{j=1}^4\rho(d\kappa_j)\langle\{\Delta^{+}_{\kappa_1}(x_3-x_2)\Delta^{+}_{\kappa_2}(x_4-x_1)+\Delta^{+}_{\kappa_1}(x_4-x_2)\Delta^{+}_{\kappa_2}(x_3-x_1)\\&+\Delta^{+}_{\kappa_1}(x_3-x_1)\Delta^{+}_{\kappa_2}(x_4-x_2)+\Delta^{+}_{\kappa_1}(x_4-x_1)\Delta^{+}_{\kappa_2}(x_3-x_2)\}\Delta^{+}_{\kappa_3}(x_2-x_1)\Delta^{+}_{\kappa_4}(x_4-x_3),\\&\times(F(\Box;\kappa_2^2,\kappa_3^2)f_1)(x_1)(F(\Box;\kappa_1^2,\kappa_3^2)f_2)(x_2)(F(\Box;\kappa_1^2,\kappa_4^2)f_3)(x_3)(F(\Box;\kappa_2^2,\kappa_4^2)f_4)(x_4)\rangle.\end{aligned}$$

For shortness the explicit formula for $\mathscr{W}^2_{2,2}$ is only written in momentum space:

$$\begin{aligned}\mathscr{W}^2_{2,2}(f_1\otimes\cdots\otimes f_4)&=\langle A^F_{21}(f_2^*)A^F_{10}(f_1^*)\phi_0,A^F_{21}(f_3)A^F_{10}(f_4)\phi_0\rangle\\&=\frac{1}{6}\int_{\Sigma^{*4}}\prod_{j=1}^4t(dp_j)\sum_{i,j}{}'F(f_1)(-p_{i_1},-p_{i_2})F(f_2)(-p_{i_3},-p_{i_4})F(f_3)(p_{j_1},p_{j_2})\\&\times F(f_4)(p_{j_3},p_{j_4}),\end{aligned}$$

where $\Sigma^1_{i,j}$ indicates summation over all $\mathbf{i} = (i_1,...,i_4)$, $\mathbf{j} = (j_1,...,j_4)$ such that $\{i_1,...,i_4\} = \{1,2,3,4\}$, $i_1 < i_2$, $i_3 < i_4$, and the same for $\mathbf{j}$.

As the class of models described above allow a fairly general choice of contraction maps a large number of models of relativistic quantum fields is provided which allow specifications of the contraction maps involved according to quite different purposes. So it might be worthwhile to try to formulate equations of motions for some of these models (possible by using pseudodifferential operators) or to consider the question of arranging all freedom in this construction in such a way that the question for a nontrivial scattering matrix allows a sensitive answer (eventually for higher powers of the field A).

ACKNOWLEDGMENTS

This paper was started during my stay at the Institut für Theoretische Physik, Georg-August Universität Göttingen, and finished at RIMS, Kyoto University. I would like to thank the members of both institutes for their kind hospitality. To Professor H. Araki and Professor H. J. Borchers I am indebted for their kind hospitality and for some stimulating discussions. Then I wish to express my thanks to JSPS for their financial support of my stay in Kyoto.

APPENDIX A

Proof of Proposition 2.1: We have to show for all $n,m = 0,1,2,...$: (a) $A^F_{mn}(f)$ maps $\mathscr{D}^n$ into $\mathscr{D}^m$ for all $f \in \mathscr{S}$; (b) $f \to A^F_{mn}(f)\phi_n$ is a continuous map from $\mathscr{S}$ into $\mathscr{K}_m$ for all $\phi_n \in \mathscr{D}^n$; (c) $A^F_{nm}(f^*) \subseteq (A^F_{mn}(f))^*$ for all $f \in \mathscr{S}$; (d) ϕ_0 is cyclic with respect to $\{A^F(f) | f \in \mathscr{S}\}$;

(a) We prove (a) by showing $H_{m,n}(f; f_{2n}) \in \mathscr{F}_{2m}$ $\forall f \in \mathscr{S} \forall f_{2n} \in \mathscr{F}_{2n}$, and $n,m = 0,1,2,...$. Now $H_{n+1,n}(f; f_{2n})$ $= \text{const}\, \mathscr{S}_{2(n+1)}(\chi_{\Sigma\times\Sigma} F(f) \otimes f_{2n})$ and $\chi_{\Sigma\times\Sigma} F(f) \in \mathscr{F}_2$. Therefore, $H_{n+1,n}(f; f_{2n}) \in \mathscr{F}_{2n+2}$ for $n = 0,1,2,...$. $H_{n-1,n}(f; f_{2n}) \in \mathscr{F}_{2n-2}$ for $n = 1,2,...$ follows from $H_{0,1}(f; f_2) \in \mathscr{F}_0$. But this is immediate because of

$$\iint_{\Sigma\times\Sigma} F(f)(-p_1,-p_2) f_2(p_1,p_2)\, t(dp_1)\, t(dp_2)$$
$$= \iint_{\Sigma\times\Sigma} F(f)(-p_1,-p_2)(1+p_1^0)^{2L}$$
$$\times (1+p_2^0)^{2L} f_2(p_1,p_2) \cdot (1+p_1^0)^{-2L}$$
$$\times (1+p_2^0)^{-2L}\, t(dp_1)\, t(dp_2)$$

and thus

$$|H_{0,1}(f, f_2)| \leqslant N_L^4\, q^-_{2L,2L}(F(f))\, q^+_{2L,2L}(f_2). \tag{A1}$$

The case $H_{n,n}(f; f_{2n})$ is similar. Notice first that

$(1+p_1^0)^{N_1}(1+p_2^0)^{N_2} H_{1,1}(f; f_2)(p_1,p_2)$ equals

$$\int_\Sigma (1+p^0)^{-2L}\, t(dp) \{ [(1+p^0)(1+|p_1-p|)$$
$$\times (1+p_1^0)^{-1}]^{-N_1} (1+|p_1-p|)^{N_1}$$
$$\times \chi_\Sigma(p_1) F(f)(p_1,-p)(1+p^0)^{2L+N_1}(1+p_2^0)^{N_2} f_2(p,p_2)$$
$$+ [(1+p^0)(1+|p_2-p|)(1+p_2^0)^{-1}]^{-N_2}$$
$$\times (1+|p_2-p|)^{N_2} \chi_\Sigma(p_2) F(f)(p_2,-p)$$
$$\times (1+p^0)^{2L+N_2}(1+p_1^0)^{N_1} f_2(p,p_1)\}$$

and this implies

$$q^+_{N_1,N_2}(H_{1,1}(f; f_2)) \leqslant \{ Q_{N_1}(F(f)) q^+_{N_1+2L,N_2}(f_2)$$
$$+ Q_{N_2}(F(f)) q^+_{N_2+2L,N_1}(f_2)\} N_L^2. \tag{A2}$$

Now $H_{n,n}(f; f_{2n}) \in \mathscr{F}_{2n}$ for all $f \in \mathscr{S}$, all $f_{2n} \in \mathscr{F}_{2n}$, and $n = 1,2,...$ follows immediately.

(b) In order to prove (b) notice first that

$$\|\tilde{\phi}^{\otimes 2n}(f_{2n})\|^2 = \int \prod_{j=1}^{2n} t(dp_j) |f_{2n}(p_1,...,p_n)|^2$$
$$\leqslant (N_L^2)^{2n} q^+_{2L,...,2L}(f_{2n})^2 \tag{A3}$$

holds and then

$$q^+_{N_1...N_{2n-2}}(H_{n-1,n}(f; f_{2n}))$$
$$\leqslant \{2n(2n-1)\}^{1/2} N_L^4\, q^-_{0,0}(F(f)) q^+_{N_1,...,N_{2n-n},2L,2L}(f_{2n}), \tag{A4}$$

$$q^+_{N_1,...,N_{2n}}(H_{n,n}(f; f_{2n}))$$
$$\leqslant 2N_L^2 \sum_{j=1}^{2n} Q_{N_j}(F(f)) q^+_{N_1,...,N_j+2L;N_{j+1},...,N_{2n}}(f_{2n}), \tag{A5}$$

$$q^+_{N_1,...,N_{2n+2}}(H_{n+1,n}(f; f_{2n}))$$
$$\leqslant \{(2n+1)(2n+2)\}^{-1/2} \sum_{\substack{i,j=1 \\ i\neq j}}^{2n+2} q^+_{N_i,N_j}(F(f))$$
$$\times q^+_{N_1,...,N_i,...,N_j,...,N_{2n+2}}(f_{2n}). \tag{A6}$$

By definition of a contraction map these inequalities together show that $f \to \|A^F_{m,n}(f)\tilde{\phi}^{2n}(f_{2n})\|$ is indeed a continuous seminorm on $\mathscr{S}$.

(c) is a simple consequence of the following equation:

$$\langle \tilde{\phi}^{\otimes 2m}(g_{2m}), A^F_{mn}(f)\tilde{\phi}^{\otimes 2n}(f_{2n})\rangle$$
$$= \langle A^F_{n,m}(f^*)\tilde{\phi}^{\otimes 2m}(g_{2m}), \tilde{\phi}^{\otimes 2n}(f_{2n})\rangle,$$

which in turn follows immediately from the definition and the assumption $F(f)^*(p_1,p_2) = \overline{F(f)(-p_2,-p_1)}$ $= F(f^*)(p_1,p_2)$.

(d) The statespace $\mathscr{K}$ of A^F is defined this way. Notice that by definition of $\mathscr{F}_{2n}$, $n = 1,2,...$, $\{H^{(n)}(h_1,...,h_n) | h_i \in \mathscr{S}\}$ is dense in $\mathscr{F}_{2n}$, where $H^{(n)}(h_1,...,h_n)$ $= H_{n,n-1}(h_i, H^{(n-1)}(h_2,...,h_n)), H^{(1)}(h) = 2^{-1/2}\chi_{\Sigma\times\Sigma} \cdot F(h)$.

APPENDIX B

In order to derive the locality constraint we have according to (2.13) to evaluate

$$\mathscr{K}_{n+1,n}(f,g; f_{2n}) = H_{n+1,n}(f; H_{n,n}(g, f_{2n}))$$
$$+ H_{n+1,n+1}(f; H_{n+1,n}(g; f_{2n})),$$
$$\mathscr{K}_{n,n}(f,g; f_{2n}) \tag{B1}$$
$$= H_{n,n}(f; H_{n,n}(g; f_{2n})) + H_{n,n-1}(f; H_{n-1,n}(g; f_{2n}))$$
$$+ H_{n,n+1}(f; H_{n+1,n}(g, f_{2n})).$$

By definition (2.9) and some lengthy calculation we get

$$\mathscr{K}_{n+1,n}(f,g; f_{2n})$$
$$= S_{n+1,n}(f,g; f_{2n}) + C_{n+1,n}(f,g; f_{2n}),$$
$$\mathscr{K}_{n,n}(f,g; f_{2n}) \tag{B2}$$
$$= 2\mathscr{W}_2^F(f \otimes g) f_{2n} + S_{n,n}(f,g; f_{2n}) + C_{n,n}(f,g; f_{2n}).$$

Here $S_{n+1,n}$ and $S_{n,n}$ are symmetric with respect to permutations of f and g and are thus "local". The remaining terms are

$$C_{n+1,n}(f,g;f_{2n}) = 4\cdot\{(2n+1)(2n+2)\}^{-1/2}\sum_{\substack{j,k=1\\ j\neq k}}^{2n+2}\int_{\Sigma} t(dp)\chi_{\Sigma}(p_k)F(f)(p_k,-p)$$
$$\times F(g)(p,p_j)\cdot\chi_{\Sigma\times\Sigma}(p_j,p)\cdot f_{2n}(p_1,\ldots,\hat{p}_k,\ldots,\hat{p}_j,\ldots,p_{2n+2}), \tag{B3}$$
$$C_{n,n}(f,g;f_{2n}) = 2\sum_{j=1}^{2n}\chi_{\Sigma}(p_j)\int_{\Sigma} t(dp)\{K^F(f,g)(p_j,-p)+K^F(f,g)(-p,p_j)\}f_{2n}(p_1,\ldots,p,p_{j+1},\ldots,p_{2n}).$$

Therefore, locality of A^F is equivalent to locality of

$(f,g)\rightarrow$

(i) $\mathscr{W}_2^F(f\otimes g)$,

(ii) $\tilde{\phi}^{\otimes(2n+2)}(C_{n+1,n}(f,g;f_{2n}))$, and (B4)

(iii) $\tilde{\phi}^{\otimes 2n}(C_{n,n}(f,g;f_{2n}))$,

for all $f_{2n}\in\mathscr{F}_{2n}$ and $n=0,1,2,\ldots$.

By some combinatorics locality of (B4) (ii) can be seen to be equivalent to that of $\tilde{\phi}^{\otimes 2}(C_{1,0}(f,g))$ and locality of (B4) (iii) to be equivalent to that of $\tilde{\phi}^{\otimes 2}(C_{1,1}(f,g;f_2))$ for all $f_2\in\mathscr{F}_2$. In order not to deal with vector-valued functions we take scalar products with $\tilde{\phi}^{\otimes 2}(g_2^*)$, $g_2\in\mathscr{F}_2$, and thus arrive at the conclusion that locality of A^F is equivalent to locality of the following functionals:

(i) $\mathscr{W}_2^F(f\otimes g)=L_1^F(f\otimes g)$;

(ii) $\langle\tilde{\phi}^{\otimes 2}(g_2^*),\ \tilde{\phi}^{\otimes 2}(C_{1,0}(f,g))\rangle = 4\sqrt{2}L_2^F(f\otimes g;g_2)$; (B5)

(iii) $\langle\tilde{\phi}^{\otimes 2}(g_2^*),\ \tilde{\phi}^{\otimes 2}(C_{1,1}(f,g;f_2))\rangle = 8L_3^F(f\otimes g;f_2,g_2)$;

for all $f_2,g_2\in\mathscr{F}_2$.

[1] N. N. Bogolubov, A. A. Logunov, and I. T. Todorov, *Introduction to Axiomatic Field Theory* (Benjamin, London, 1975); R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society, Providence, RI, 1965); R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).

[2] E. Brüning, "An analysis of some non-linear aspects of QFT," Bielefeld preprint Bi-TP 81/40 and in *Mathematical Problems in Theoretical Physics*, Proceedings: Berlin (West) 1981, edited by R. Schrader, R. Seiler, and D. A. Uhlenbrock (Springer-Verlag, Berlin, 1981), Lecture Notes in Physics No. 153.

[3] J. Yngvason, "On the locality ideal in the algebra of test functions for quantum fields," preprint RH-82-04 of the Science Institute, Department of Mathematics, University of Iceland, Reykjavik.

[4] E. Brüning, "On a generalization of Wick-powers for generalized free fields in terms of Jacobi-fields," Bielefeld preprint BI-TP 7918; Rep. Math. Phys. **19**, 2 (1983).

[5] H. J. Borchers, (private communication, Göttingen, 1982).

[6] W. Güttinger and A. Rieckers, "Spectral representation of Lorentz-invariant distributions and scale transformation," Commun. Math. Phys. **7**, 190–217 (1968).

[7] J. Bros, H. Epstein, and V. Glaser, "On the connection between analyticity and Lorentz-covariance of Wightman functions," Commun. Math. Phys. **6**, 77–100 (1967).

[8] H. J. Borchers, "Algebraic aspects of Wightman field theory," in *Statistical Mechanics and Field Theory*, edited by R. N. Sen and C. Weil, Haifa Summer School 1971 (Israel University, Jerusalem, 1972).

An equivalence class of quantum field theories at finite temperature

H. Matsumoto, Y. Nakano, and H. Umezawa
Department of Physics, The University of Alberta, Edmonton, Alberta, Canada T6G 2J1

(Received 25 October 1983; accepted for publication 11 May 1984)

It is shown that there exists a certain equivalence class of quantum field theories at finite temperature each of which produces the same statistical averages. The theories in this equivalent class are classified by multiplicities of field degrees of freedom and have a one to one correspondence with the choices of the path in the real-time path-ordered formulation of the statistical average. Among them, thermo field dynamics is found to be the most convenient theory.

PACS numbers: 11.10. − z, 03.70. + k, 05.30. − d

I. INTRODUCTION

Thermo field dynamics (TFD)[1–6] has been developed in the last ten years as an extension of the quantum field theory at zero temperature to finite temperature. It preserves many properties of the zero-temperature quantum field theory, especially the operator formalism and time-ordered formulation of the Green's functions (i.e., the Feynman diagram method in real time). It has an advantage of being able to include information regarding the states through the temperature-dependent vacuum, so that various requirements arising from operator relations (such as Ward–Takahashi relations) can be treated consistently even when the phase transition induces a spontaneous breakdown of symmetry. Furthermore, since it is formulated in real space-time, no complications other than those which appear in the case of zero temperature arise according to, for example, the boundary conditions of the symmetry transformation as in the supersymmetry.[7] Also, since the time variable in the TFD is real, all quantities calculated are directly related to the frequency-momentum dependent observables. These constitute the main differences of TFD from the Matsubara Green's function method[8–10] in which the imaginary time (u) in the finite interval ($0 \leqslant u \leqslant \beta$) is used and a certain analytic continuation is needed in order to obtain time-dependent physical quantities.

Thermo field dynamics has been applied to practical computations in many subjects. The perturbation theory in thermo field dynamics is developed in Refs. 3, 4, and 6. The Feynman rules in TFD with propagators of 2×2 matrics are slightly different from those with one-component propagators used in the recent literature.[11,12] It was pointed out that the Feynman rules of TFD ensure that the Kubo–Martin–Schwinger (KMS) condition is satisfied in each order of perturbation.[13]

It is also shown that the ultraviolet divergences in TFD are renormalized by temperature-independent counterterms in all orders of loop corrections, provided the theory is renormalized at zero temperature. The Ward–Takahashi relations can be employed to develop new techniques in finite temperature calculations.[4,14]

There have been many attempts to include real-time in the calculation of statistical averages.[15–20] As an extension of the Matsubara method, the path along the imaginary time axis from 0 to $-i\beta$ is analytically continuated and the trace formula is rewritten in terms of the path-ordered products in the complex-time plane.[20] Though the choice of the path along the real-time axis gives a real-time formulation of the statistical average, it has not been elevated to the status of the *operator formulation* of quantum field theory because of the complications arising from the fact that operators appearing in the path-ordered formalism have the same number of degrees of freedom as that in the trace formula, contrary to the situation in thermo field dynamics in which the number of degrees of freedom is doubled.

Among axiomatic field theorists, it has been recognized for some time now that the effect of temperature can be included in a free field theory by doubling the field degrees of freedom.[21,22] This approach has been axiomatized by the used of the C^*-algebra and the Kubo–Martin–Schwinger(KMS) condition,[23,24] and is now called the axiomatic statistical mechanics.[25] Recently the correspondence and equivalence between the thermo field dynamics and the axiomatic statistical mechanics has been clarified.[5]

As an immediate extension of the work of Ref. 2, the existence of a certain freedom in the choice of the thermal vacuum was pointed out in the paper in Ref. 26, the main purpose of which was to clarify the relation between the Heisenberg equation (or the Schrödinger equation) in TFD and the Liouville equation of the density matrix and to point out the fact that the time-ordered two-point Green's function obtained in Ref. 3 may be regarded as a mixing of two Green's functions with opposite time directions.

It is the purpose of this paper to present the relation between the path-ordered real-time formalism and the quantum field theoretical formulation at finite temperature in its most complete and widest sense. In a recent short paper,[27] the authors have presented the relation between the Matsubara method and the thermo field dynamics. Extending this formalism, we will show that a certain class of the quantum field theoretical formulations at finite temperature forms an equivalence class and each member of this class has a one to one correspondence to each member of a class of choices of path in the path-ordered formalism. The structure of the quantum field theories in this equivalence class will be discussed. It will be found that each theory in this class is characterized by the multiplicity of field degrees of freedom and by a metric. Furthermore, the metric can be chosen to be unity only in a doublet formalism, which turns out to be the thermo field dynamics and which corresponds to a special

 0022-2488/84/103076-10$02.50

choice of the path in the path–ordered formalism. Thus we conclude that thermo field dynamics is the most convenient formalism.

The analysis in this paper supplies us with a simple rule for building a quantum field theory corresponding to each choice of path in the path-ordered formalism. It also provides us with a rule whereby one member of the equivalence class of quantum field theories at finite temperature may be related to another member.

Our consideration takes the following step. According to the Wightman theorem,[28] one can reconstruct a Hilbert space when a complete set of vacuum expectation values is given. Therefore, a possible way of constructing the quantum field theory with real time at finite temperature is to regard the statistical averages of operators as the vacuum expectation values referring to a suitably defined vacuum. One may try to perform this task in the framework of perturbative expansions. However, it is quite difficult to see if these statistical averages in the perturbative expansion satisfy the Wightman axioms for the vacuum expectation values. We, therefore, take a short route looking for a quantum field theory in the form of a perturbative expansion in which the vacuum expectation values agree with the statistical average of operators with real times. In the course of this study, the doubling (more generally $2N$-multiplying) of the field degrees of freedom (the thermal multiplication of fields) naturally takes place in order that the perturbation theory acquires a causal formulation of a quantum field theory. When we cover a variety of choices of the path, we find a certain equivalence class of the quantum field theories at finite temperature.

In those quantum field theoretical formalisms, the physically observable results are given by matrix elements only of the first component of doubled (or multiplied) fields; other matrix elements act as hidden variables which reflect the presence of an indeterminate heat bath. The physical equivalence among the different formalism is required only for physical matrix elements.

In the next section, we briefly summarize a real-time formulation of the qunatum statistical average. The formalism follows the method of Mills of Ref. 20. A brief summary of the perturbation theory in thermo field dynamics is also presented. In Sec. III, the doublet representation of the quantum field theory at finite temperature is related to the real-time formulation. It will be shown that the most natural finite temperature extension of a conventional field theory is the thermo field dynamics presented in Refs. 1–6, where the Hermiticity of operators is preserved. In Sec. IV, it is shown that a real-time formulation requires an even number of field multiplication in a corresponding field theory (a $2N$-component field theory). It is pointed out also that the doublet theory gives the simplest form of quantum field theory at finite temperature. In Sec. V, the results of the preceding sections in the form of perturbative expansion (i.e., the interaction representation) will be translated into the language of the Heisenberg picture. It will be shown that the equilibrium condition and the KMS condition are summarized as the tilde-substitution rule. From this follows the physical equivalence of the different formalisms. Then we can conclude that there exists a certain equivalence class of the field theoretical formulation at finite temperature which has a one to one correspondence with a certain class of choices of path in the path-ordered formalism. The inclusion of quantum mechanical operators such as spin operators in the formalism is also discussed. Section VI is devoted to the concluding remarks.

II. PRELIMINARY

A. Real time formulation of statistical average

For simplicity, we consider a complex scalar field $\phi(x)$ whose dynamics is determined by a Hamiltonian H. Hereafter when the space coordinates are irrelevant, we will retain only the time variables. Let us define the Heisenberg operator with a complex time z by

$$A(z) = \exp\{izH\}A(0)\exp\{-izH\}. \tag{2.1}$$

We consider the thermal average of a product of Heisenberg operators:

$$W(z_n,\ldots,z_1) = \mathrm{tr}[e^{-\beta H}A_n(z_n)\cdots A_1(z_1)]/\mathrm{tr}[e^{-\beta H}], \tag{2.2}$$

where $\beta = 1/k_B T$. A perturbation theory is given by separating H into an unperturbed part H_0 and an interaction part H_I as follows. It is well known that we can write $\exp\{-iH(z_2 - z_1)\}$ as

$$\exp\{-iH(z_2 - z_1)\} = \exp\{-iH_0 z_2\}U(z_2,z_1)\exp\{iH_0 z_1\} \tag{2.3}$$

with

$$U(z_2,z_1) = T_c \exp\left\{-i\int_{z_1}^{z_2} dz\, H_I(z)\right\}, \tag{2.4}$$

where

$$H_I(z) = \exp\{iH_0 z\}H_I(0)\exp\{-iH_0 z\}, \tag{2.5}$$

T_c denotes a path ordering product along a certain path connecting the complex points z_2 and z_1, and an integration is defined along this path. Then the thermal trace of the product of the Heisenberg operators is obtained by

$$\begin{aligned}\mathrm{tr}[e^{-\beta H}A_n(z_n)\cdots A_1(z_1)] &= \mathrm{tr}[e^{-i(\tau - i\beta)H}A_n(z_n)\cdots A_1(z_1)e^{i\tau H}] \\ &= \mathrm{tr}[e^{-\beta H_0}U(\tau - i\beta,z_n)\mathscr{A}_n(z_n)U(z_n,z_{n-1}) \\ &\quad \times \mathscr{A}_{n-1}(z_{n-1})\cdots\mathscr{A}_1(z_1)U(z_1,\tau)],\end{aligned} \tag{2.6}$$

where

$$\mathscr{A}(z) = e^{iH_0 z}A(0)e^{-iH_0 z}. \tag{2.7}$$

Therefore if $(z_n,\ldots z_1)$ are ordered in the appropriate manner along a path C starting from a complex number τ and ending at $\tau - i\beta$, we can write[19,20]

$$W(z_n,\ldots,z_1) = \frac{\langle T_c U(\tau - i\beta,\tau)\mathscr{A}_n(z_n)\cdots\mathscr{A}_1(z_1)\rangle_0}{\langle T_c U(\tau - i\beta,\tau)\rangle_0}. \tag{2.8}$$

Here $\langle\cdots\rangle_0$ denotes $\mathrm{tr}[e^{-\beta H_0}\cdots]/\mathrm{tr}[e^{-\beta H_0}]$. The calculation of $W(z_n,\ldots z_1)$ is reduced to a calculation of a path-ordered product.

Defining a path C starting from point τ and ending at $\tau - i\beta$ in the complex time plane and assuming that $(z_1,\ldots z_n)$ lie on the path C, we can define a statistical average of the path-ordered product on C by

$$G(z_1,...,z_n) = \mathrm{tr}[e^{-\beta H} T_c A_1(z_1)\cdots A_n(z_n)]/\mathrm{tr}[e^{-\beta H}], \tag{2.9}$$

where T_c is the path ordering operator. Then $G(z_1,...,z_n)$ is obtained in the interaction picture as

$$G(z_1,...,z_n) = \frac{\langle T_c U(\tau - i\beta,\tau)\mathscr{A}_1(z_1)\cdots\mathscr{A}_n(z_n)\rangle_0}{\langle T_c U(\tau - i\beta,\tau)\rangle_0}. \tag{2.10}$$

When the path C is chosen along the imaginary axis (i.e., $\tau = 0$ and $z_1,...z_n$ are pure imaginary), (2.10) gives the perturbation formula in the Matsubara Green's function method. In this sense (2.10) can be considered as a generalization of the Matsubara method; it may be considered as a result of suitable analytic continuation of time variables.

There is one comment regarding the choice of the path C (see Ref. 20). The expectation value of two operators A and B, $\langle A(z)B(z')\rangle$, is shown to be analytic in the domain

$$-\beta < \mathrm{Im}(z - z') < 0. \tag{2.11}$$

In order to ensure simultaneous analyticity with respect to $z_1,...,z_n$ in (2.10) in the same region, one must guarantee that the path ordering and (2.11) are not mutually incompatible. This means that a path C must be chosen with a monotonically decreasing imaginary part.

In order to apply the above formalism to the calculation of the statistical average of dynamical quantities, $\{z_1,...z_n\}$ should be real time $\{t_1,...,t_n\}$. In this case the path C should run on the real axis to cover the set $\{t_1,...,t_n\}$. Because of (2.11), the path along the real axis is considered as the limit with an infinitesimally small slope. We are thus led to the choices of the path C illustrated in Fig. 1; the path runs along the real axis $2N$ fold and the turning points of the path, t_A and t_B, must be taken sufficiently large ($t_A \to -\infty$, $t_B \to +\infty$). These choices of the path C lead to real-time formulations for the statistical average.

It should be noted that the Matsubara frequency method may be employed only when the path C is along the imaginary axis. Therefore, the Matsubara frequency is not applicable when $z_1,...,z_n$ are real.

B. Thermo field dynamics

In thermo field dynamics, the perturbation theory is given as follows.[2–4,6,13] Suppose that a Lagrangian $\mathscr{L}(\phi^\dagger(x),\phi(x))$ consisting of complex scalar fields $\phi(x)$ and $\phi^\dagger(x)$ is given. A finite temperature field theory requires that the degrees of freedom of the field theory is doubled by introducing tilde fields $\tilde{\phi}(x)$ and $\tilde{\phi}^\dagger(x)$. They form thermal doublets $\phi^\alpha(x)$ and $\phi^{\dagger\alpha}(x)$ as

$$\phi^\alpha(x) = \begin{cases} \phi(x) \\ \tilde{\phi}^\dagger(x), \end{cases} \quad \phi^{\dagger\alpha}(x) = \begin{cases} \phi^\dagger(x) \\ \tilde{\phi}(x), \end{cases} \quad \alpha = \begin{cases} 1 \\ 2. \end{cases} \tag{2.12}$$

The dynamics of $\phi^\alpha(x)$ and $\phi^{\dagger\alpha}(x)$ is determined by the thermal Lagrangian $\hat{\mathscr{L}}(x)$ defined by

$$\hat{\mathscr{L}}(x) = \mathscr{L}(\phi^\dagger(x),\phi(x)) - \widetilde{\mathscr{L}(\phi^\dagger(x),\phi(x))}, \tag{2.13}$$

where $\widetilde{\mathscr{L}(\phi^\dagger(x),\phi(x))}$ is given by

$$\widetilde{\mathscr{L}(\phi^\dagger(x),\phi(x))} = \mathscr{L}^*(\tilde{\phi}^\dagger(x),\tilde{\phi}(x)). \tag{2.14}$$

Here * denotes a complex conjugate. In general, the tilde operation rule is defined for an operator O (or O_1 and O_2) and c-numbers c_1 and c_2 as

$$\begin{aligned} &(1)\quad \widetilde{O_1 O_2} = \tilde{O}_1\tilde{O}_2, \\ &(2)\quad \widetilde{c_1 O_1 + c_2 O_2} = c_1^*\tilde{O}_2 + c_2^*\tilde{O}_2, \\ &(3)\quad \tilde{\tilde{O}} = \eta_0 O, \end{aligned} \tag{2.15}$$

$$\eta_0 = \begin{cases} +1, \text{ for a boson-like operator,} \\ -1, \text{ for a fermion-like operator.} \end{cases}$$

We denote the unperturbed part of $\hat{\mathscr{L}}$ as $\hat{\mathscr{L}}_0$ and the interaction part as $\hat{\mathscr{L}}_I$. The field $\phi^\alpha(x)(\phi^{\dagger\alpha}(x))$ in the interaction picture is denoted by $\varphi^\alpha(x)(\varphi^{\dagger\alpha}(x))$. The free Lagrangian $\hat{\mathscr{L}}_0$ is assumed to lead to a free field equation,

$$\left[-\frac{\partial^2}{\partial t^2} - \omega^2(-i\nabla)\right]\varphi^\alpha(x) = 0. \tag{2.16}$$

Then $\varphi^\alpha(x)$ is given by

$$\varphi^\alpha(x) = \int \frac{d^3k}{\sqrt{(2\pi)^3 2\omega(\mathbf{k})}} \times [a^\alpha(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x} - i\omega(\mathbf{k})t} + b^{\dagger\alpha}(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{x} + i\omega(\mathbf{k})t}], \tag{2.17}$$

with

$$a^\alpha(\mathbf{k}) = U_B^{\alpha\gamma}(\omega(\mathbf{k}))a_\beta^\gamma(\mathbf{k}), \quad b^\alpha(\mathbf{k}) = U_B^{\alpha\gamma}(\omega(\mathbf{k}))b_\beta^\gamma(\mathbf{k}), \tag{2.18}$$

$$U_B(\omega(\mathbf{k})) = \begin{pmatrix} \cosh\theta(\mathbf{k}) & \sinh\theta(\mathbf{k}) \\ \sinh\theta(\mathbf{k}) & \cosh\theta(\mathbf{k}) \end{pmatrix}. \tag{2.19}$$

The annihilation operators $a_\beta(\mathbf{k})$, $b_\beta(\mathbf{k})$, $\tilde{a}_\beta(\mathbf{k})$, and $\tilde{b}_\beta(\mathbf{k})$ annihilate the temperature-dependent vacuum $|0,\beta\rangle$ in the interaction picture:

$$a_\beta(\mathbf{k})|0,\beta\rangle = b_\beta(\mathbf{k})|0,\beta\rangle = \tilde{a}_\beta(\mathbf{k})|0,\beta\rangle = \tilde{b}_\beta(\mathbf{k})|0,\beta\rangle = 0. \tag{2.20}$$

Such a Bogoliubov transformation is required by the thermal instability of quanta created by $a^\dagger(\mathbf{k})$ and $b^\dagger(\mathbf{k})$, and the Bogoliubov parameter is chosen as

$$\sinh^2\theta(\mathbf{k}) = 1/(e^{\beta\omega(\mathbf{k})} - 1) \tag{2.21}$$

to reproduce the thermal distribution function. A statistical average is given as an expectation value between the temperature-dependent vacuums. The free propagator of the thermal doublet can then be obtained by direct calculation with the use of (2.17), (2.18), and (2.20) as[3,4]

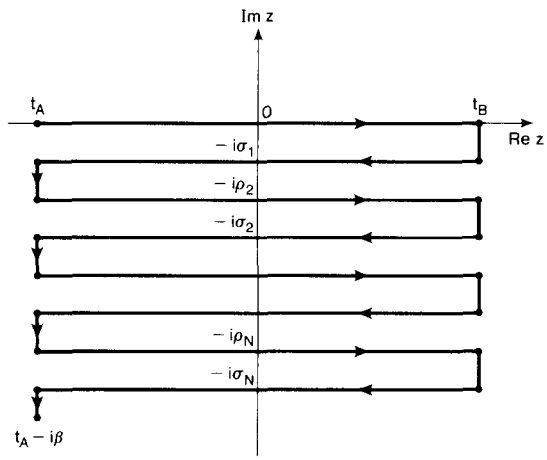


FIG. 1. Choice of the path in the real time formulation.

$$\langle 0,\beta|T\varphi^{\alpha}(x)\varphi^{\dagger\gamma}(y)|0,\beta\rangle = i\int\frac{d^4k}{(2\pi)^4}e^{-ik(x-y)}\Delta^{\alpha\gamma}(k), \tag{2.22}$$

with

$$\Delta(k) = \frac{1}{k_0^2-\omega(\mathbf{k})^2+i\epsilon}\frac{1}{e^{\beta\omega(\mathbf{k})}-1}\begin{pmatrix} e^{\beta\omega(\mathbf{k})} & e^{\beta\omega(\mathbf{k})/2} \\ e^{\beta\omega(\mathbf{k})/2} & 1\end{pmatrix} - \frac{1}{k_0^2-\omega(\mathbf{k})^2-i\epsilon}\frac{1}{e^{\beta\omega(\mathbf{k})}-1}\begin{pmatrix} 1 & e^{\beta\omega(\mathbf{k})/2} \\ e^{\beta\omega(\mathbf{k})/2} & e^{\beta\omega(\mathbf{k})}\end{pmatrix}, \tag{2.23}$$

which can also be rewritten as[4,6]

$$\Delta(k) = U_B(\omega(\mathbf{k}))\frac{\tau}{k_0^2-\omega(\mathbf{k})^2+i\tau\epsilon}U_B(\omega(\mathbf{k})) = \frac{\tau}{k_0^2-\omega(\mathbf{k})^2+i\tau\epsilon} - 2\pi i\delta(k_0^2-\omega(\mathbf{k})^2)\frac{1}{e^{\beta\omega(\mathbf{k})}-1}\times\begin{pmatrix} 1 & e^{\beta\omega(\mathbf{k})/2} \\ e^{\beta\omega(\mathbf{k})/2} & 1\end{pmatrix}, \tag{2.24}$$

with

$$\tau = \begin{pmatrix} 1 & 0 \\ 0 & -1\end{pmatrix}. \tag{2.25}$$

Then the time-ordered product is given by

$$\langle O(\beta)|T\phi^{\alpha_1}(x_1)\cdots\phi^{\alpha_l}(x_l)\phi^{\dagger\gamma_1}(x_1')\cdots\phi^{\dagger\gamma_m}(x_m')|O(\beta)\rangle = \frac{\langle 0,\beta|T\exp\{i\int d^4x\,\hat{\mathscr{L}}_I(x)\}\varphi^{\alpha_1}(x_1)\cdots\varphi^{\alpha_l}(x_l)\varphi^{\dagger\gamma_1}(x_1')\cdots\varphi^{\dagger\gamma_m}(x_m')|0,\beta\rangle}{\langle 0,\beta|T\exp\{i\int d^4x\,\hat{\mathscr{L}}_I(x)\}|0,\beta\rangle}, \tag{2.26}$$

where $|O(\beta)\rangle$ is the temperature-dependent vacuum in the Heisenberg picture. The Feynman rules are constructed with the use of the propagator (2.23) [or (2.24)] and the vertices given by $\hat{\mathscr{L}}_I(x)$. It has been shown that the above Feynman rules are consistent with the KMS condition[13] and that calculations performed with only the 1–1 component of (2.24) upset the KMS condition.

III. THERMAL DOUBLET FORMALISM (TWO COMPONENT EXTENSION)

In this section, we will discuss the relation between a real-time formulation of statistical average and a double formalism of quantum field theory. As was pointed out in Sec. II A, a perturbation theory of a real-time formulation is obtained by a suitable choice of the path C in the complex-time plane. Since we are interested in the situation in which the times $z_1,\ldots,z_n$ are real, we chose the path C along the real axis. The simplest such path is presented in Fig. 2, in which the path goes from one real-time t_A to another real-time t_B along the real axis, drops vertically from t_B to $t_B - i\sigma$ $(0<\sigma\leqslant\beta)$, returns parallel to the real axis from $t_B - i\sigma$ to $t_A - i\sigma$ and ends at $t_A - i\beta$. We take the limit $t_A\to-\infty$, $t_B\to+\infty$. Then the formula in Eq. (2.10) is rewritten as

$$G(t_1,\ldots,t_n) = \lim_{\substack{t_A\to-\infty\\ t_B\to+\infty}}\left\langle T_c\exp\left\{-i\int_{-i\sigma}^{-i\beta}dz\,H_I(t_A+z)\right\}\times\exp\left\{-i\int_{t_B}^{t_A}dt\,H_I(t-i\sigma)\right\}\times\exp\left\{-i\int_0^{-i\sigma}dz\,H_I(t_B+z)\right\}\times\exp\left\{-i\int_{t_A}^{t_B}dt\,H_I(t)\right\}\mathscr{A}_1(t_1)\cdots\mathscr{A}_n(t_n)\right\rangle_0^{\mathrm{con}}, \tag{3.1}$$

where the notation "con" means the connected part defined such that the vacuum diagrams are excluded. The interaction Hamiltonian at the vertical region of the path C, $H_I(t_{A,B}+z)$, always contains t_A or t_B as a time variable. The contributions of this section of the path to (3.1) carry $t_i - t_{A,B}$ as time variables. Therefore when we consider the limit $t_{A,B}\to\mp\infty$, any contribution from $H_I(t_{A,B}+z)$ damps. Finally we have

$$G(t_1,\ldots,t_n) = \left\langle T_c\exp\left\{+i\int_{-\infty}^{\infty}dt\,H_I(t-i\sigma)\right\}\exp\left\{-i\int_{-\infty}^{\infty}dt\,H_I(t)\right\}\mathscr{A}_1(t_1)\cdots\mathscr{A}_n(t_n)\right\rangle_0^{\mathrm{con}} \tag{3.2}$$

$$= \frac{\langle T_c\exp\{-i\int_{-\infty}^{\infty}dt\,[H_I(t)-H_I(t-i\sigma)]\}\mathscr{A}_1(t_1)\cdots\mathscr{A}_n(t_n)\rangle_0}{\langle T_c\exp\{-i\int_{-\infty}^{\infty}dt\,[H_I(t)-H_I(t-i\sigma)]\}\rangle_0}. \tag{3.3}$$

We consider a complex scalar model where the Lagrangian density is given by

$$\mathscr{L}(x) = \mathscr{L}_0(x) + \mathscr{L}_I(x) \tag{3.4}$$

with

$$\mathscr{L}_0(x) = \frac{\partial\phi^{\dagger}(x)}{\partial t}\frac{\partial\phi(x)}{\partial t} - \phi^{\dagger}(x)\omega^2(-i\nabla)\phi(x). \tag{3.5}$$

Then Hamiltonian density is given by

$$\mathscr{H}(x) = \mathscr{H}_0(x) + \mathscr{H}_I(x) \tag{3.6}$$

with

$$\mathscr{H}_0(x) = \frac{\partial\phi^{\dagger}(x)}{\partial t}\frac{\partial\phi(x)}{\partial t} + \phi^{\dagger}(x)\omega^2(-i\nabla)\phi(x), \tag{3.7}$$

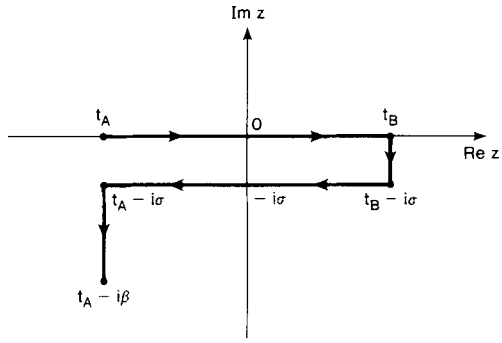


FIG. 2. Choice of the path in the doublet formalism.

$$\mathcal{H}_I(x) = -\mathcal{L}_I(x). \tag{3.8}$$

Free field $\varphi(x)$ in the interaction picture is given by

$$\varphi(x) = \int \frac{d^3k}{\sqrt{(2\pi)^3 2\omega(\mathbf{k})}} \times [a(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x} - i\omega(\mathbf{k})t} + b^\dagger(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{x} + i\omega(\mathbf{k})t}], \tag{3.9}$$

with

$$[a(\mathbf{k}),a^\dagger(\mathbf{k}')] = [b(\mathbf{k}),b^\dagger(\mathbf{k}')] = \delta(\mathbf{k}-\mathbf{k}'). \tag{3.10}$$

The free Hamiltonian H_0 is given by

$$H_0 = \int d^3k\, \omega(\mathbf{k})[a^\dagger(\mathbf{k})a(\mathbf{k}) + b^\dagger(\mathbf{k})b(\mathbf{k})]. \tag{3.11}$$

Propagators appearing in (3.3) can be evaluated using the definition of the trace $\langle\cdots\rangle_0$. The result is summarized as (see also Ref. 27 for the nonrelativistic case)

$$\begin{aligned}
&\left\langle T_c\begin{pmatrix}\varphi(\mathbf{x},t)\\ \varphi(\mathbf{x},t-i\sigma)\end{pmatrix}(\varphi^\dagger(\mathbf{x}',t')\varphi^\dagger(\mathbf{x}',t'-i\sigma))\right\rangle_0 \\
&= i\int\frac{d^4k}{(2\pi)^4}e^{-ik(x-x')}\frac{1}{2\omega(\mathbf{k})} \\
&\times\left\{\frac{1}{k_0-\omega(\mathbf{k})+i\epsilon}\frac{1}{e^{\beta\omega(\mathbf{k})}-1}\right. \\
&\times\begin{pmatrix} e^{\beta\omega(\mathbf{k})} & e^{(\beta/2+\gamma)\omega(\mathbf{k})}\\ e^{(\beta/2-\gamma)\omega(\mathbf{k})} & 1\end{pmatrix} \\
&-\frac{1}{k_0+\omega(\mathbf{k})-i\epsilon}\frac{1}{e^{\beta\omega(\mathbf{k})}-1} \\
&\times\begin{pmatrix} e^{\beta\omega(\mathbf{k})} & e^{(\beta/2-\gamma)\omega(\mathbf{k})}\\ e^{(\beta/2+\gamma)\omega(\mathbf{k})} & 1\end{pmatrix} \\
&-\frac{1}{k_0-\omega(\mathbf{k})-i\epsilon}\frac{1}{e^{\beta\omega(\mathbf{k})}-1} \\
&\times\begin{pmatrix} 1 & e^{(\beta/2+\gamma)\omega(\mathbf{k})}\\ e^{(\beta/2-\gamma)\omega(\mathbf{k})} & e^{\beta\omega(\mathbf{k})}\end{pmatrix} \\
&+\frac{1}{k_0+\omega(\mathbf{k})+i\epsilon}\frac{1}{e^{\beta\omega(\mathbf{k})}-1} \\
&\left.\times\begin{pmatrix} 1 & e^{(\beta/2-\gamma)\omega(\mathbf{k})}\\ e^{(\beta/2+\gamma)\omega(\mathbf{k})} & e^{\beta\omega(\mathbf{k})}\end{pmatrix}\right\},
\end{aligned} \tag{3.12}$$

where γ is defined by

$$\sigma = \beta/2 + \gamma \quad (-\beta/2 < \gamma \leqslant \beta/2). \tag{3.13}$$

Note that this structure of the propagator is the same as the one [cf. (2.23)] in thermo field dynamics of Sec. II B in the case of $\gamma = 0$. [See also (3.41).] The $(-i\epsilon)$-term corresponds to the returning path as was pointed out in Ref. 26. The constant γ specifies the intersection of the returning path on the imaginary axis. For quantized free fields, the statistical average of field products can be expressed by a sum of products of two point functions according to Wick's theorem. Therefore, the T_c-propagator (3.12) and the interaction Hamiltonian $[H_I(t) - H_I(t-i\sigma)]$ determine the perturbation expansion of (3.3). The result is the real-time formulation of the statistical average.

In order to express the above real-time formulation in terms of the Feynman diagram method, we need one more step; that is we have to rewrite the T_c-product in terms of the usual T-product. To do this we must use only the real-time parameter although we have the complex time $t - i\sigma$. Since σ is not a variable, but a constant, we can regard $\varphi(t-i\sigma)$ $(\varphi^\dagger(t-i\sigma))$ as a function of t, calling it $\tilde{\varphi}^\ddagger(t)$ $(\tilde{\varphi}(t))$. We now have the doublet field; $\varphi(\mathbf{x},t)$ and $\tilde{\varphi}^\ddagger(\mathbf{x},t) = \varphi(\mathbf{x},t-i\sigma)[\varphi^\ddagger(\mathbf{x},t) = \varphi^\dagger(\mathbf{x},t)$ and $\tilde{\varphi}(\mathbf{x},t) = \varphi^\dagger(\mathbf{x},t-i\sigma)]$. Although $\varphi(\mathbf{x},t)$ and $\varphi(\mathbf{x},t-i\sigma)$ are the same fields with different complex times, we consider $\varphi(\mathbf{x},t)$ *and* $\tilde{\varphi}^\ddagger(\mathbf{x},t)$ *as two independent fields* when we identify $\varphi(\mathbf{x},t-i\sigma)$ as $\tilde{\varphi}^\ddagger(\mathbf{x},t)$. This is an important step to elevate a real-time formalism to the quantum field theoretical operator formalism. Recalling the fact that σ contains a constant γ, we make use of the notation $\varphi(\mathbf{x},t;\gamma)$, $\tilde{\varphi}^\ddagger(\mathbf{x},t;\gamma)$, $\varphi^\ddagger(\mathbf{x},t;\gamma)$, and $\tilde{\varphi}^\ddagger(\mathbf{x},t;\gamma)$.

Summarizing, we introduce mutually commuting fields $\varphi(x;\gamma)$ and $\tilde{\varphi}(x;\gamma)$ satisfying equal time commutation relations,

$$\left[\varphi(\mathbf{x},t;\gamma),\frac{\partial}{\partial t}\varphi^\ddagger(\mathbf{x}',t;\gamma)\right] = i\delta(\mathbf{x}-\mathbf{x}'), \tag{3.14a}$$

$$\left[\tilde{\varphi}(\mathbf{x},t;\gamma),\frac{\partial}{\partial t}\tilde{\varphi}^\ddagger(\mathbf{x}',t;\gamma)\right] = -i\delta(\mathbf{x}-\mathbf{x}'), \tag{3.14b}$$

otherwise commute, and make the following correspondence:

$$\begin{aligned}
&\varphi(\mathbf{x},t)\to\varphi(\mathbf{x},t;\gamma), \quad \varphi^\dagger(\mathbf{x},t)\to\varphi^\ddagger(\mathbf{x},t;\gamma), \\
&\varphi(\mathbf{x},t-i\sigma)\to\tilde{\varphi}^\ddagger(\mathbf{x},t;\gamma), \quad \varphi^\dagger(\mathbf{x},t-i\sigma)\to\tilde{\varphi}(\mathbf{x},t;\gamma).
\end{aligned} \tag{3.15}$$

Here the symbol ‡ refers to the double-dagger conjugate which will be defined later and is reduced to the Hermitian conjugate (†) when $\gamma = 0$. We require that the thermal averages are given by the vacuum expectation values. The temperature-dependent vacuum $|0,\beta\rangle$ is introduced and physical particle creation and annihilation operators are denoted by

$$\begin{aligned}
&a_\beta^\dagger(\mathbf{k}),\quad \tilde{a}_\beta^\dagger(\mathbf{k}),\quad b_\beta^\dagger(\mathbf{k}),\quad \tilde{b}_\beta^\dagger(\mathbf{k}); \\
&a_\beta(\mathbf{k}),\quad \tilde{a}_\beta(\mathbf{k}),\quad b_\beta(\mathbf{k}),\quad \tilde{b}_\beta(\mathbf{k}); \\
&a_\beta(\mathbf{k})|0,\beta\rangle = \tilde{a}_\beta(\mathbf{k})|0,\beta\rangle = b_\beta(\mathbf{k})|0,\beta\rangle = \tilde{b}_\beta(\mathbf{k})|0,\beta\rangle = 0.
\end{aligned} \tag{3.16}$$

The free field operators $\varphi(x;\gamma)$, $\varphi^\ddagger(x;\gamma)$, $\tilde{\varphi}(x;\gamma)$,and $\tilde{\varphi}^\ddagger(x;\gamma)$ are expressed in terms of these operators as

$$\varphi(x;\gamma) = \int\frac{d^3k}{\sqrt{(2\pi)^3 2\omega(\mathbf{k})}} \times [a(\mathbf{k};\gamma)e^{i\mathbf{k}\cdot\mathbf{x} - i\omega(\mathbf{k})t} + b^\ddagger(\mathbf{k};\gamma)e^{-i\mathbf{k}\cdot\mathbf{x} + i\omega(\mathbf{k})t}], \tag{3.17a}$$

$$\varphi^{\ddagger}(x;\gamma)=\int\frac{d^3k}{\sqrt{(2\pi)^3 2\omega(\mathbf{k})}}\times[a^{\ddagger}(\mathbf{k};\gamma)e^{-i\mathbf{k}\cdot\mathbf{x}+i\omega(\mathbf{k})t}+b(\mathbf{k};\gamma)e^{i\mathbf{k}\cdot\mathbf{x}-i\omega(\mathbf{k})t}], \tag{3.17b}$$

$$\tilde{\varphi}(x;\gamma)=\int\frac{d^3k}{\sqrt{(2\pi)^3 2\omega(\mathbf{k})}}\times[\tilde{a}(\mathbf{k};\gamma)e^{-i\mathbf{k}\cdot\mathbf{x}+i\omega(\mathbf{k})t}+\tilde{b}^{\ddagger}(\mathbf{k},\gamma)e^{i\mathbf{k}\cdot\mathbf{x}-i\omega(\mathbf{k})t}], \tag{3.17c}$$

$$\tilde{\varphi}^{\ddagger}(x;\gamma)=\int\frac{d^3k}{\sqrt{(2\pi)^3 2\omega(\mathbf{k})}}\times[\tilde{a}^{\ddagger}(\mathbf{k};\gamma)e^{i\mathbf{k}\cdot\mathbf{x}-i\omega(\mathbf{k})t}+\tilde{b}(\mathbf{k};\gamma)e^{-i\mathbf{k}\cdot\mathbf{x}+i\omega(\mathbf{k})t}], \tag{3.17d}$$

with

$$a(\mathbf{k};\gamma)=e^{\gamma\omega(\mathbf{k})/2}a(\mathbf{k}),\quad a^{\ddagger}(\mathbf{k};\gamma)=e^{-\gamma\omega(\mathbf{k})/2}a^{\dagger}(\mathbf{k}), \tag{3.18}$$

$$\tilde{a}(\mathbf{k};\gamma)=e^{\gamma\omega(\mathbf{k})/2}\tilde{a}(\mathbf{k}),\quad \tilde{a}^{\ddagger}(\mathbf{k};\gamma)=e^{-\gamma\omega(\mathbf{k})/2}\tilde{a}^{\dagger}(\mathbf{k}),$$

$$\begin{pmatrix}a(\mathbf{k})\\ \tilde{a}^{\dagger}(\mathbf{k})\end{pmatrix}=\begin{pmatrix}\cosh\theta(\mathbf{k}) & \sinh\theta(\mathbf{k})\\ \sinh\theta(\mathbf{k}) & \cosh\theta(\mathbf{k})\end{pmatrix}\begin{pmatrix}a_B(\mathbf{k})\\ \tilde{a}_B^{\dagger}(\mathbf{k})\end{pmatrix}, \tag{3.19a}$$

$$\begin{pmatrix}a^{\dagger}(\mathbf{k})\\ \tilde{a}(\mathbf{k})\end{pmatrix}=\begin{pmatrix}\cosh\theta(\mathbf{k}) & \sinh\theta(\mathbf{k})\\ \sinh\theta(\mathbf{k}) & \cosh\theta(\mathbf{k})\end{pmatrix}\begin{pmatrix}a_B^{\dagger}(\mathbf{k})\\ \tilde{a}_B(\mathbf{k})\end{pmatrix}. \tag{3.19b}$$

[The same relations as (3.18) and (3.19) hold for b, $b^{\dagger}$, $\tilde{b}$, and $\tilde{b}^{\dagger}$.] Here

$$\sinh^2\theta(\mathbf{k})=\frac{1}{e^{\beta\omega(\mathbf{k})}-1},\quad \cosh^2\theta(\mathbf{k})=\frac{e^{\beta\omega(\mathbf{k})}}{e^{\beta\omega(\mathbf{k})}-1}. \tag{3.20}$$

Note that, for example, $a(\mathbf{k})$ and $a^{\dagger}(\mathbf{k})$ are Hermitian conjugates of one another but $a(\mathbf{k};\gamma)$ and $a^{\ddagger}(\mathbf{k};\gamma)$ are not when $\gamma\neq 0$. This is the reason why we distinguished the double-dagger conjugate ($\ddagger$) from Hermitian conjugate ($\dagger$). It is easy to show that

$$\left\langle T_c\begin{pmatrix}\varphi(\mathbf{x},t)\\ \varphi(\mathbf{x},t-i\sigma)\end{pmatrix}(\varphi^{\dagger}(\mathbf{x}',t')\varphi^{\dagger}(\mathbf{x}',t'-i\sigma))\right\rangle_0 = \langle 0,\beta|T\begin{pmatrix}\varphi(\mathbf{x},t;\gamma)\\ \tilde{\varphi}^{\ddagger}(\mathbf{x},t;\gamma)\end{pmatrix}(\varphi^{\ddagger}(\mathbf{x}',t';\gamma)\tilde{\varphi}(\mathbf{x}',t';\gamma))|0,\beta\rangle. \tag{3.21}$$

In this way the statistical average of the T_c-product becomes the vacuum expectation value of the T-product of the quantum fields φ, $\varphi^{\ddagger}$, $\tilde{\varphi}$, and $\tilde{\varphi}^{\ddagger}$. We insist here again that the T_c-product is rewritten in terms of the T-product because of the introduction of the two commuting fields.

Through the correspondence (3.15), the interaction Hamiltonian density is identified as follows:

$$\mathscr{H}_I(\varphi^{\dagger}(\mathbf{x},t),\varphi(\mathbf{x},t))\to\mathscr{H}_I(\varphi^{\ddagger}(\mathbf{x},t;\gamma),\varphi(\mathbf{x},t;\gamma)), \tag{3.22a}$$

$$\mathscr{H}_I(\varphi^{\dagger}(\mathbf{x},t-i\sigma),\varphi(\mathbf{x},t-i\sigma))\to\mathscr{H}_I(\tilde{\varphi}(\mathbf{x},t;\gamma),\tilde{\varphi}^{\ddagger}(\mathbf{x},t;\gamma)). \tag{3.22b}$$

Since $\mathscr{H}_I$ is Hermitian, it satisfies

$$\{\mathscr{H}_I(\varphi^{\dagger}(x),\varphi(x))\}^{\dagger}=\mathscr{H}(\varphi^{\dagger}(x),\varphi(x)) \tag{3.23}$$

which leads to a formal operator relation

$$\mathscr{H}_I^{*}(\varphi(x),\varphi^{\dagger}(x))=\mathscr{H}_I(\varphi^{\dagger}(x),\varphi(x)). \tag{3.24}$$

Then (3.22b) becomes

$$\mathscr{H}_I(\varphi^{\dagger}(\mathbf{x},t-i\sigma),\varphi(\mathbf{x},t-i\sigma))\to\mathscr{H}_I^{*}(\tilde{\varphi}^{\ddagger}(\mathbf{x},t;\gamma),\tilde{\varphi}(\mathbf{x},t;\gamma)) \tag{3.25}$$

$$=\widetilde{\mathscr{H}_I(\varphi^{\ddagger}(\mathbf{x},t;\gamma),\varphi(\mathbf{x},t;\gamma))}, \tag{3.26}$$

that is, the Hamiltonian for the tilde fields is obtained by the tilde operation. With this identification, the Feynman rules become identical to those in (3.3). In summary, we have proved the correspondence between the path-ordered formalism with the choice of the path given in Fig. 2 and the quantum field theoretical formulation of doublet representation:

$$\left\langle T_c\exp\left\{-i\int_{-\infty}^{\infty}dt\,[H_I(t)-H_I(t-i\sigma)]\right\}\varphi(x_1)\cdots\varphi(x_l)\varphi^{\dagger}(x_1')\cdots\varphi^{\dagger}(x_m')\right\rangle_0 = \langle 0,\beta|T\exp\left\{-i\int_{-\infty}^{\infty}dt\hat{H}_I(t)\right\}\varphi(x_1;\gamma)\cdots\varphi(x_l;\gamma)\varphi^{\ddagger}(x_1';\gamma)\cdots\varphi^{\ddagger}(x_m';\gamma)|0,\beta\rangle, \tag{3.27}$$

where

$$\hat{H}_I(t;\gamma)=H_I(t;\gamma)-\widetilde{H_I(t;\gamma)}, \tag{3.28}$$

with

$$H_I(t;\gamma)=\int d^3x\,\mathscr{H}_I(\varphi^{\ddagger}(x;\gamma),\varphi(x;\gamma)). \tag{3.29}$$

In particular the case $\gamma=0$ reduces to the thermo field dynamics presented in Sec. II B (note $\mathscr{H}_I=-\mathscr{L}_I$).

We will now show that the field theory with arbitrary γ in (3.28) ($-\beta/2\leqslant\gamma\leqslant\beta/2$) can be identified as a field theory with a certain metric. In fact, we see from (3.18) that $\{a(\mathbf{k};\gamma),a^{\ddagger}(\mathbf{k};\gamma),b(\mathbf{k};\gamma),b^{\ddagger}(\mathbf{k};\gamma)\}$ and $\{a(\mathbf{k}),a^{\dagger}(\mathbf{k}),b(\mathbf{k}),b^{\dagger}(\mathbf{k})\}$ are related to each other through the operator transformation

$$(a(\mathbf{k};\gamma),a^{\ddagger}(\mathbf{k};\gamma),b(\mathbf{k};\gamma),b^{\ddagger}(\mathbf{k};\gamma)) = \eta_{\gamma}(a(\mathbf{k}),a^{\dagger}(\mathbf{k}),b(\mathbf{k}),b^{\dagger}(\mathbf{k}))\eta_{\gamma}^{-1}, \tag{3.30}$$

where

$$\eta_{\gamma}=\exp\{-\tfrac{1}{2}\gamma H_0\}, \tag{3.31}$$

$$H_0=\int d^3k\,\omega(\mathbf{k})[a^{\dagger}(\mathbf{k})a(\mathbf{k})+b^{\dagger}(\mathbf{k})b(\mathbf{k})]. \tag{3.32}$$

Similar relations hold for the tilde fields. Therefore $\varphi(x)$ and $\varphi^{\dagger}(x)$ are defined by

$$\varphi(x)=\int\frac{d^3k}{\sqrt{(2\pi)^3 2\omega(\mathbf{k})}}\times[a(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}-i\omega(\mathbf{k})t}+b^{\dagger}(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{x}+i\omega(\mathbf{k})t}], \tag{3.33a}$$

$$\varphi^{\dagger}(x) = \int \frac{d^3k}{\sqrt{(2\pi)^3 2\omega(\mathbf{k})}} \times [a^{\dagger}(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x} + i\omega(\mathbf{k})t} + b(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x} - i\omega(\mathbf{k})t}]; \quad (3.33b)$$

and are related to $\varphi(x;\gamma)$ and $\varphi^{\ddagger}(x;\gamma)$ as

$$\varphi(x,\gamma) = \eta_\gamma \varphi(x) \eta_\gamma^{-1}, \quad \varphi^{\ddagger}(x,\gamma) = \eta_\gamma \varphi^{\dagger}(x) \eta_\gamma^{-1}. \quad (3.34)$$

The tilde fields also satisfy similar relations:

$$\tilde{\varphi}(x;\gamma) = \tilde{\eta}_\gamma \tilde{\varphi}(x) \tilde{\eta}_\gamma^{-1}, \quad \tilde{\varphi}^{\ddagger}(x;\gamma) = \tilde{\eta}_\gamma \tilde{\varphi}^{\dagger}(x) \tilde{\eta}_\gamma^{-1}. \quad (3.35)$$

Note that $\eta_\gamma = 1$ for $\gamma = 0$ according to (3.31). Now operators carrying (a constant) γ are related with those with $\gamma = 0$ through the relations

$$O(\gamma) = \hat{\eta}_\gamma O \hat{\eta}_\gamma^{-1}, \quad O(\gamma)^{\ddagger} = \hat{\eta}_\gamma O^{\dagger} \hat{\eta}_\gamma^{-1}, \quad (3.36)$$

where

$$\hat{\eta}_\gamma = \eta_\gamma \tilde{\eta}_\gamma. \quad (3.37)$$

Therefore, when $\gamma \neq 0$, the double-dagger conjugate assumes the role of the Hermitian conjugate:

$$(O_1(\gamma) O_2(\gamma))^{\ddagger} = O_2(\gamma)^{\ddagger} O_1(\gamma)^{\ddagger}, \quad (3.38a)$$

$$(c_1 O_1(\gamma) + c_2 O_2(\gamma))^{\ddagger} = c_1^* O_1(\gamma)^{\ddagger} + c_2^* O_2(\gamma)^{\ddagger}, \quad (3.38b)$$

$$(O(\gamma)^{\ddagger})^{\ddagger} = O(\gamma). \quad (3.38c)$$

The relation between the Hermitian conjugate and double-dagger conjugate is

$$O^{\dagger} = \hat{\eta}_\gamma^{-2} O^{\ddagger} \hat{\eta}_\gamma^2. \quad (3.39)$$

Note that, from (3.36), we have $O(\gamma)^{\ddagger} = O(\gamma)$ if $O^{\dagger} = O$. Therefore, the self-adjointness is modified as

$$O(\gamma)^{\ddagger} = O(\gamma) \quad \text{or} \quad \{O(\gamma)\hat{\eta}_\gamma^2\}^{\dagger} = O(\gamma)\hat{\eta}_\gamma^2. \quad (3.40)$$

In particular when $\gamma = 0$, we have $\hat{\eta}_\gamma = 1$ and the double-dagger conjugate and the Hermitian conjugate become identical to one another and the conventional definition of Hermiticity is preserved in this case. For example, the Hamiltonian density becomes Hermitian when $\gamma = 0$.

Since the physical results are given by the expectation values of nontilde fields, these expectation values should be γ-independent. Indeed we can prove this γ-independence as follows.[27] Taking advantage of the on-shell condition of the temperature dependent part of the thermal propagator, we can rewrite the Fourier component of the rhs of (3.12) as

$$\begin{pmatrix} 1 & 0 \\ 0 & e^{-\gamma k_0} \end{pmatrix} U_B(\omega(\mathbf{k})) \frac{\tau}{k_0^2 - \omega(\mathbf{k})^2 + i\tau\epsilon} \times U_B(\omega(\mathbf{k})) \begin{pmatrix} 1 & 0 \\ 0 & e^{\gamma k_0} \end{pmatrix}, \quad (3.41)$$

by which we can interpret that γ-dependent factors appear on the propagators terminated at $\tilde{\varphi}$ or $\tilde{\varphi}^{\ddagger}$. Diagrams involving an expectation value of nontilde fields do not possess external lines of $\tilde{\varphi}$ or $\tilde{\varphi}^{\ddagger}$, therefore these factors are assigned to $\tilde{H}_I$-vertices. If an $\tilde{H}_I$-vertex terminates in a $\tilde{\varphi}(\tilde{\varphi}^{\ddagger})$ field, it acquires a factor $e^{\gamma k_0}(e^{-\gamma k_0})$ for each $\tilde{\varphi}(\tilde{\varphi}^{\ddagger})$. But these γ-dependent factors are completely canceled at each $\tilde{H}_I$-vertex owing to the k_0-conservation. One of the immediate results of the γ-independence is

$$\langle 0,\beta | T \exp\left\{ -i \int_{-\infty}^{\infty} dt \hat{H}_I(t;\gamma) \right\} | 0,\beta \rangle = 1, \quad (3.42)$$

since the left-hand side for $\gamma = -\beta/2$ (i.e., $\sigma = 0$) is identity according to the corresponding path-ordered form. This means that the vacuum fluctuation does not induce a change of normalization.

The above analysis shows that corresponding to each choice of the path given in Fig. 1 one can construct a quantum field theory of equilibrium at finite temperature. The different choice of the path is related to the change of metric in the corresponding field theory, which is characterized by γ. The tilde field works as a hidden field which manifests the thermal effects. The arbitrariness for the choice of γ disappears when we require that the Hamiltonian density is Hermitian which gives $\gamma = 0$. As will be presented in the next section, we can extend the analysis of this section into a $2N$-component field theory.

IV. GENERALIZATION OF THERMO FIELD DYNAMICS—2N COMPONENT FIELD THEORY

In this section we investigate how far the quantum field theoretical formulation at finite temperature can be generalized. In order to avoid the appearance of imaginary time, the choice of the path C shown in Fig. 1 is the most general one; the path runs back and forth $2N$ times along the real axis ($0 = \rho_1 < \sigma_1 < \rho_2 < \sigma_2 \cdots < \rho_N < \sigma_N \leqslant \beta$). The points t_A and t_B are taken as infinite; $t_A \to -\infty$, $t_B \to +\infty$. For other choices of path, we need to introduce the imaginary time.

Let us consider the thermal average of the path-ordered operators:

$$G(t_1,\ldots,t_n) = \mathrm{tr}[e^{-\beta H} T_c A_1(x_1) \cdots A_n(x_n)] / \mathrm{tr}[e^{-\beta H}]. \quad (4.1)$$

By an argument similar to that presented in Sec. III, (4.1) is given in the interaction picture as

$$G(x_1,\ldots,x_n) = \frac{\langle T_c \exp\{-i\int_{-\infty}^{\infty} dt \sum_{i=1}^{N} [H_I(t-i\rho_i) - H_I(t-i\sigma_i)]\} \mathscr{A}_1(t_1) \cdots \mathscr{A}_n(t_n) \rangle_0}{\langle T_c \exp\{-i\int_{-\infty}^{\infty} dt \sum_{i=1}^{N} [H_I(t-i\rho_i) - H_I(t-i\sigma_i)]\} \rangle_0}. \quad (4.2)$$

The notation is the same as that in Sec. III. In order to obtain the perturbation series from (4.2), we need $4N^2$ types of two point functions. Those take a compact form in a $2N \times 2N$ matrix:

$$\left\langle T_c \begin{pmatrix} \varphi(\mathbf{x},t-i\rho_i) \\ \varphi(\mathbf{x},t-i\sigma_i) \end{pmatrix} (\varphi^{\dagger}(\mathbf{x}',t'-i\rho_j) \varphi^{\dagger}(\mathbf{x}',t'-i\sigma_j)) \right\rangle_0$$

$$= i \int \frac{d^4k}{(2\pi)^4} \frac{e^{-ik(x-x')}}{2\omega(\mathbf{k})} \left\{ \frac{F_+(k_0)}{k_0 - \omega(\mathbf{k}) + i\epsilon} - \frac{F_-(k_0)}{k_0 - \omega(\mathbf{k}) - i\epsilon} + \frac{G_+(k_0)}{k_0 + \omega(\mathbf{k}) + i\epsilon} - \frac{G_-(k_0)}{k_0 + \omega(\mathbf{k}) - i\epsilon} \right\}_{ij}, \quad (4.3)$$

with

$$F_+(\omega)=\begin{pmatrix}e^{-\rho\omega} & 0\\ 0 & e^{-\sigma\omega}\end{pmatrix}\begin{pmatrix}C & S\\ C & S\end{pmatrix}\begin{pmatrix}e^{\rho\omega} & 0\\ 0 & e^{\sigma\omega}\end{pmatrix}, \quad (4.4a)$$

$$F_-(\omega)=\begin{pmatrix}e^{-\rho\omega} & 0\\ 0 & e^{-\sigma\omega}\end{pmatrix}\begin{pmatrix}S & S\\ C & C\end{pmatrix}\begin{pmatrix}e^{\rho\omega} & 0\\ 0 & e^{\sigma\omega}\end{pmatrix}, \quad (4.4b)$$

$$G_+(\omega)=\begin{pmatrix}e^{-\rho\omega} & 0\\ 0 & e^{-\sigma\omega}\end{pmatrix}\begin{pmatrix}S^T & C^T\\ S^T & C^T\end{pmatrix}\begin{pmatrix}e^{\rho\omega} & 0\\ 0 & e^{\sigma\omega}\end{pmatrix}, \quad (4.5a)$$

$$G_-(\omega)=\begin{pmatrix}e^{-\rho\omega} & 0\\ 0 & e^{-\sigma\omega}\end{pmatrix}\begin{pmatrix}C^T & C^T\\ S^T & S^T\end{pmatrix}\begin{pmatrix}e^{\rho\omega} & 0\\ 0 & e^{\sigma\omega}\end{pmatrix}, \quad (4.5b)$$

where $e^{\pm\rho\omega}$, $e^{\pm\sigma\omega}$, S, and C are $N\times N$ matrices:

$$e^{\rho\omega}=\begin{pmatrix}e^{\rho_1\omega} & & 0\\ & \ddots & \\ 0 & & e^{\rho_N\omega}\end{pmatrix}, \quad e^{\sigma\omega}=\begin{pmatrix}e^{\sigma_1\omega} & & 0\\ & \ddots & \\ 0 & & e^{\sigma_N\omega}\end{pmatrix}, \quad (4.6)$$

$$S=\frac{1}{e^{\beta\omega}-1}\begin{pmatrix}1 & \cdots & 1\\ e^{\beta\omega} & \cdot & \cdot\\ \vdots & \cdot & \cdot\\ e^{\beta\omega} & \cdots e^{\beta\omega} & 1\end{pmatrix}, \quad (4.7)$$

$$C=\frac{1}{e^{\beta\omega}-1}\begin{pmatrix}e^{\beta\omega} & 1\cdots & 1\\ \vdots & \ddots & \vdots\\ e^{\beta\omega} & \cdots & e^{\beta\omega}\end{pmatrix}.$$

It is easy to see that

$$C-S=I\equiv\begin{pmatrix}1 & & 0\\ & \ddots & \\ 0 & & 1\end{pmatrix}. \quad (4.8)$$

Therefore we have the relation,

$$F_+(k_0)-F_-(k_0)=G_-(k_0)-G_+(k_0)=\tau\equiv\begin{pmatrix}I & 0\\ 0 & -1\end{pmatrix}. \quad (4.9)$$

In order to look for a field theoretical formulation which leads to the same perturbation rules, we introudce the $2N$-component field $\psi(x)$ and $\psi^\ddagger(x)$ and identify as

$$\begin{pmatrix}\varphi(\mathbf{x},t-ip_i)\\ \varphi(\mathbf{x},t-i\sigma_i)\end{pmatrix}\to\psi(x)=\begin{pmatrix}\psi_{\rho_i}(x)\\ \psi^\ddagger_{\sigma_i}(x)\end{pmatrix}, \quad (4.10a)$$

$$(\varphi^\dagger(\mathbf{x},t-i\rho_j)\varphi^\dagger(\mathbf{x},t-i\sigma_j))\to\psi^\ddagger(x)=(\psi^\ddagger_{\rho_j}(x)\psi_{\sigma_j}(x)). \quad (4.10b)$$

Their canonical commutation relations are

$$[\psi(\mathbf{x},t),\dot\psi^\ddagger(\mathbf{y},t)]=i\tau\delta(\mathbf{x}-\mathbf{y}). \quad (4.11)$$

We assume that $\psi(x)$ and $\psi^\ddagger(x)$ are related to the physical fields $\psi_\beta(x)$ and $\psi^\dagger_\beta(x)$, which are mutually Hermitian conjugate and satisfy

$$\langle 0,\beta|T\psi_\beta(x)\psi^\dagger_\beta(y)|0,\beta\rangle = i\int\frac{d^4k}{(2\pi)^4}e^{-ik(x-y)}\frac{\tau}{k_0^2-\omega(\mathbf{k})^2+i\tau\epsilon}, \quad (4.12)$$

through the following transformation:

$$\psi(x)=A(-i\partial_0)\psi_\beta(x), \quad (4.13a)$$

$$\psi^\ddagger(x)=\psi^\dagger_\beta(x)B(-i\overleftarrow{\partial}_0). \quad (4.13b)$$

The requirements for the matrices A and B are

$$F_+(k_0)=A[(1+\tau)/2]B=G_-(k_0), \quad (4.14a)$$

$$F_-(k_0)=A[(1-\tau)/2]B=G_+(k_0). \quad (4.14b)$$

First we ask if there exists a choice of (ρ,σ) in which $\psi(x)$ and $\psi^\ddagger(x)$ become mutually Hermitian conjugate. This requirement leads to the condition $B^\dagger=A$, that is,

$$\tau=A\tau A^\dagger \quad (4.15)$$

and

$$F_+(k_0)+F_-(k_0)=AA^\dagger=G_-(k_0)+G_+(k_0). \quad (4.16)$$

From the explicit form of (F_++F_-) and (G_-+G_+), we can see easily that the condition (4.16) is never satisfied unless $N=1$. Therefore two-component theory is the only possible theory which preserves the Hermiticity of operators at zero temperature ($\gamma=0$ in Sec. III). In other cases, the double-dagger ($\ddagger$) conjugate, which needs the introduction of a certain metric, requires a generalization of the Hermitian conjugate.

As an example, we can choose A and B as

$$A=\begin{pmatrix}e^{-\rho\omega} & 0\\ 0 & e^{-\sigma\omega}\end{pmatrix}\begin{pmatrix}I & S\\ I & C\end{pmatrix}, \quad B=\begin{pmatrix}C & S\\ I & I\end{pmatrix}\begin{pmatrix}e^{\rho\omega} & 0\\ 0 & e^{\sigma\omega}\end{pmatrix}. \quad (4.17)$$

Then the thermal average is given by

$$G(x_1,\ldots,x_l,y_1,\ldots,y_m)=\langle 0,\beta|T\exp\left\{-i\int dt\,\hat H_I(t)\right\}\times\psi(x_1)\cdots\psi(x_l)\psi^\ddagger(y_1)\cdots\psi^\ddagger(y_m)|0,\beta\rangle, \quad (4.18)$$

where

$$\hat H_I(t)=\sum_{i=1}^N[H^i_I(t)-\tilde H^i_I(t)]=\sum_{i=1}^N\int d^3x[\mathscr{H}_I(\psi^\ddagger_{\rho_i}(x),\psi_{\rho_i}(x))-\mathscr{H}_I(\psi^\ddagger_{\sigma_i}(x),\psi_{\sigma_i}(x))]. \quad (4.19)$$

In a similar way as shown in Sec. III, the rhs of (4.18) is proved to be independent of the parameters (ρ_i,σ_i); the proof heavily depends on the energy conservation at each vertex.

The analysis of this section shows that there is a variety of choices for the causal formulation of finite temperature field theory, which forms an equivalence class of the quantum field theory at finite temperature. However, in order to include thermal effects, one must increase the field degrees of freedom by a factor two or $2N$ ($N\geqslant 1$) in general. Since the first component is identified as the physical fields, other $2N-1$ fields are considered as hidden variables. If we require that Hermiticity of operators in the original theory is preserved, the only possible theory is the doublet theory which is the thermo field dynamics of Refs. 1–6. It is due to this reason that we do not plan any furhter study of the $2N$-component formalism with $N>1$ in this paper.

V. OPERATOR FORMALISM IN HEISENBERG PICTURE AND KMS CONDITION

In previous sections, we have identified the perturbation theory in a real-time path-ordered formulation of multipoint functions and their causal formulation familiar in ordinary quantum field theories. In this section, we reformulate the results of the previous sections in terminology of the Heisenberg picture. We consider only the thermal doublet representation (i.e., $N=1$) although we do not assume $\gamma=0$.

We consider a model of a complex field $\phi(x)$ and assume that its canonical conjugate is $\dot{\phi}^\dagger(x)$. The Hamiltonian density is given by $\mathscr{H}(\phi^\dagger(x),\phi(x))$. The results of the previous section show that the thermal effect can be taken into account by doubling the freedom of the field. The correspondence is

$$\begin{pmatrix}\phi(\mathbf{x},t)\\ \phi(\mathbf{x},t-i\sigma)\end{pmatrix}\to\Psi_\gamma(\mathbf{x},t)=\begin{pmatrix}\psi_\gamma(\mathbf{x},t)\\ \tilde{\psi}^\ddagger_\gamma(\mathbf{x},t)\end{pmatrix}, \tag{5.1a}$$

$$(\phi^\dagger(\mathbf{x},t)\phi^\dagger(\mathbf{x},t-i\sigma))\to\Psi^\ddagger_\gamma(\mathbf{x},t)=(\psi^\ddagger_\gamma(\mathbf{x},t)\tilde{\psi}_\gamma(\mathbf{x},t)). \tag{5.1b}$$

The fields $\Psi_\gamma(x)$ and $\Psi^\ddagger_\gamma(\mathbf{x})$ satisfy the equal-time commutation relation,

$$[\Psi_\gamma(\mathbf{x},t),\dot{\Psi}^\ddagger_\gamma(\mathbf{y},t)]=i\tau\delta(\mathbf{x}-\mathbf{y}). \tag{5.2}$$

The Hamiltonian is given by

$$\hat{H}=H-\tilde{H}, \tag{5.3}$$

with

$$H=\int d^3x\,\mathscr{H}(\psi^\ddagger_\gamma(x),\psi_\gamma(x)), \tag{5.4a}$$

$$\tilde{H}=\int d^3x\,\widetilde{\mathscr{H}(\psi^\ddagger_\gamma(x),\psi_\gamma(x))}. \tag{5.4b}$$

Therefore the time development is expressed as

$$\Psi_\gamma(\mathbf{x},t)=\exp\{i\hat{H}t\}\Psi_\gamma(\mathbf{x},0)\exp\{-i\hat{H}t\}. \tag{5.5}$$

The thermal average of $\phi(x)$ and $\phi^\dagger(x)$ corresponds to the vacuum expectation value of $\psi_\gamma(x)$ and $\psi^\ddagger_\gamma(x)$:

$$\begin{aligned}&\langle\phi(x_1)\cdots\phi(x_l)\phi^\dagger(y_1)\cdots\phi^\dagger(y_m)\rangle\\ &\quad=\langle O(\beta)|\psi_\gamma(x_1)\cdots\psi_\gamma(x_l)\psi^\ddagger_\gamma(y_1)\cdots\psi^\ddagger(y_m)|O(\beta)\rangle.\end{aligned} \tag{5.6}$$

Here $\langle\cdots\rangle$ denotes the thermal average. The correspondences of (5.1) and (5.6) indicate that

$$\begin{aligned}\langle A(t-i\sigma)\cdots\rangle&=\langle O(\beta)|A_\gamma(t-i\sigma)\cdots|O(\beta)\rangle\\ &=\langle O(\beta)|\tilde{A}^\ddagger_\gamma(t)\cdots|O(\beta)\rangle,\end{aligned} \tag{5.7}$$

where A is an operator consisting of ϕ and $\phi^\dagger$. Therefore we should have the operator relation

$$\langle O(\beta)|A_\gamma(t-i\beta/2-i\gamma)=\langle O(\beta)|\tilde{A}^\ddagger_\gamma(t). \tag{5.8}$$

Here we used the definition $\sigma=\beta/2+\gamma$.

First, we consider $\gamma=0$ case. In this case the double-dagger conjugate is identical with the Hermitian conjugate. Then we have

$$\langle O(\beta)|A(t-i\beta/2)=\langle O(\beta)|\tilde{A}^\dagger(t), \tag{5.9a}$$

its conjugate is

$$A^\dagger(t+i\beta/2)|O(\beta)\rangle=\tilde{A}(t)|O(\beta)\rangle. \tag{5.9b}$$

The relations (5.9) are the tilde-substitution rules. These relations indicate that we are very close to the KMS condition. In particular when A is the Hamiltonian H, (5.9) indicates that

$$(H-\tilde{H})|O(\beta)\rangle=0. \tag{5.10}$$

In this way the tilde-substitution rule specifies the equilibrium state.

When $\gamma\neq0$, we require that there exists a metric $\hat{\eta}_\gamma$ which relates A_γ and A at $\gamma=0$:

$$A_\gamma(x)=\hat{\eta}_\gamma A(x)\hat{\eta}_\gamma^{-1}, \tag{5.11a}$$

$$A^\ddagger_\gamma(x)=\hat{\eta}_\gamma A^\dagger(x)\hat{\eta}_\gamma^{-1}. \tag{5.11b}$$

Also we require that $\hat{\eta}_\gamma$ is invariant under the tilde conjugation:

$$\tilde{\hat{\eta}}_\gamma=\hat{\eta}_\gamma. \tag{5.12}$$

From (5.8) and (5.11), we have

$$\langle O(\beta)|\hat{\eta}_\gamma A(t-i\beta/2-i\gamma)=\langle O(\beta)|\hat{\eta}_\gamma\tilde{A}^\dagger(t), \tag{5.13a}$$

which becomes

$$\langle O(\beta)|\hat{\eta}_\gamma e^{\gamma H}A(t-i\beta/2)=\langle O(\beta)|\hat{\eta}_\gamma e^{\gamma H}\tilde{A}^\dagger(t). \tag{5.13b}$$

Here we have used the fact that A consists only of the non-tilde fields (and therefore $\tilde{A}$ consists of tilde fields only). Comparing (5.9a) and (5.12) and using (5.10), we see that the above condition is satisfied by

$$\hat{\eta}_\gamma=\exp\{-\gamma(H+\tilde{H})/2\}. \tag{5.14}$$

Note that the Hamiltonian density $\mathscr{H}(\psi^\ddagger_\gamma,\psi_\gamma)$ is not necessarily Hermitian when $\gamma\neq0$ but that H is Hermitian by this choice of metric $\hat{\eta}_\gamma$:

$$H_\gamma=\hat{\eta}_\gamma H\hat{\eta}_\gamma^{-1}=H. \tag{5.15}$$

We notice also that the double-dagger conjugate and Hermitian conjugate are related through

$$A^\ddagger_\gamma(x)=\hat{\eta}^2_\gamma A^\dagger(x)\hat{\eta}_\gamma^{-2}. \tag{5.16}$$

This means also that

$$\{|O(\beta)\rangle\}^\ddagger=\langle O(\beta)|\hat{\eta}_\gamma^{-2}, \tag{5.17a}$$

$$\{\langle O(\beta)|\}^\ddagger=\hat{\eta}^2_\gamma|O(\beta)\rangle. \tag{5.17b}$$

From (5.8), we have

$$A^\ddagger_\gamma(t+i\beta/2+i\gamma)\hat{\eta}^2_\gamma|O(\beta)\rangle=\tilde{A}_\gamma(t)\hat{\eta}^2_\gamma|O(\beta)\rangle.$$

Multiplying $\hat{\eta}_\gamma^{-2}$ from the left, we get

$$A^\ddagger_\gamma(t+i\beta/2)|O(\beta)\rangle=\tilde{A}_\gamma(t+i\gamma)|O(\beta)\rangle. \tag{5.18}$$

The substitution rule for the case of $\gamma\neq0$ is summarized by

$$\langle O(\beta)|A_\gamma(t-i\beta/2-i\gamma)=\langle O(\beta)|\tilde{A}^\ddagger_\gamma(t), \tag{5.19a}$$

$$A^\ddagger_\gamma(t+i\beta/2-i\gamma)|O(\beta)\rangle=\tilde{A}_\gamma(t)|O(\beta)\rangle. \tag{5.19b}$$

When (5.19) are satisfied, we have

$$\begin{aligned}&\langle O(\beta)|A_\gamma(t)B_\gamma(t')|O(\beta)\rangle\\ &\quad=\langle O(\beta)|\tilde{A}^\ddagger_\gamma(t+i\beta/2+i\gamma)B_\gamma(t')|O(\beta)\rangle\\ &\quad=\langle O(\beta)|B_\gamma(t')\tilde{A}^\ddagger_\gamma(t+i\beta/2+i\gamma)|O(\beta)\rangle\\ &\quad=\langle O(\beta)|B_\gamma(t')A_\gamma(t+i\beta)|O(\beta)\rangle,\end{aligned} \tag{5.20}$$

which is the KMS condition. Also by the use of (5.11) and (5.10), the γ-independence can be shown easily, since

$$\langle O(\beta)|A_\gamma(x)\cdots B_\gamma(y)|O(\beta)\rangle$$
$$= \langle O(\beta)|\hat{\eta}_\gamma A(x)\cdots B(y)\hat{\eta}_\gamma^{-1}|O(\beta)\rangle$$
$$= \langle O(\beta)|e^{-\gamma\bar{H}}A(x)\cdots B(y)e^{\gamma\bar{H}}|O(\beta)\rangle$$
$$= \langle O(\beta)|A(x)\cdots B(y)|O(\beta)\rangle. \quad (5.21)$$

In this proof, the equilibrium condition (5.10) plays an important role. In Sec. III, we showed that the energy conservation[27] is the basis for the γ-independence. As a matter of fact, (5.10) is intimately related to the energy conservation; the condition (5.10) guarantees that there are no energy flows between tilde and nontilde systems in the equilibrium state. This property is not preserved in the nonequilibrium, which will be discussed elsewhere.

As can be seen from the analysis in this section, the tilde-substitution rule (5.9) [or (5.19)] is the relation which determines the equilibrium properties of thermo field dynamics and is equivalent to the KMS condition. This relation and the commutativity between tilde and nontilde operators enable us to rewrite the T_c-product into the causal formulation, which leads us to a quantum field theoretical formulation. The members of the equivalence class in this formulation are classified by γ. The parameter γ is identified with the choice of the path and is amalgamated in the metric η_γ. Even when quantum mechanical operators such as spin operators $S_i(n)$ $(n = 1,2,...)$ are considered, the inclusion of these operators in thermo field dynamics is straightforward in the light of the above consideration. We introduce tilde operators $\tilde{S}_i(n)$ which commute with nontilde operators $S_i(n)$ and require that the tilde-substitution rules (5.9) [or (5.19)] are satisfied. Then quantum mechanical operators are also included in the causal formulation. An important difference is that they do not simply satisfy Wick's theorem in the perturbation calculation.

VI. CONCLUDING REMARKS

In this paper, we have studied the relation between a real-time path-ordered formulation of statistical average and a quantum field theoretical formulation. There exists a natural correspondence between the path-ordered formulation and the prescription in quantum field theory at finite temperature with multiplied field degrees of freedom with a thermal Hamilton $\hat{H}$. Those quantum field theoretical formulations form a equivalence class of the quantum field theory at finite temperature and are classified by the even number of the multiplicities of field degrees of freedom and by the metric. This, on one hand, gives a simple rule for identifying a particular quantum field theory with a particular path-ordered formalism and, on the other hand, leads to a clear perspective over a wide class of equivalent field theories at finite temperature. It was pointed out that the tilde-substitution rules are the essential relations which enable us to rewrite the path-ordered formulation in the causal formulation.

The tilde-substitution rule plays an important role in specifying the properties of the equilibrium state as pointed out in Sec. V. It is interesting, therefore, to consider how these relations are modified in the case of nonequilibrium phenomena. Such studies are now in progress.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council, Canada and the Dean of Science, Faculty of Science, the University of Alberta, Edmonton, Alberta, Canada.

[1] L. Leplae, F. Mancini, and H. Umezawa, Phys. Rep. C **10**, 154 (1974).
[2] Y. Takahashi and H. Umezawa, Collect. Phenom. **2**, 55 (1975).
[3] H. Matsumoto, Fortschr. Phys. **25**, 1 (1977).
[4] H. Umezawa, H. Matsumoto, and M. Tachiki, *Thermo Field Dynamics and Condensed States* (North-Holland, Amsterdam, 1982).
[5] I. Ojima, Ann. Phys. (NY) **137**, 1 (1981).
[6] H. Matsumoto, I. Ojima, and H. Umezawa, Ann. Phys. (NY) **152**, 348 (1984).
[7] L. Girardello, M. Grisaru, and P. Salomonson, Nucl. Phys. B **178**, 331 (1981).
[8] T. Matsubara, Prog. Theor. Phys. **14**, 351 (1955).
[9] H. Ezawa, Y. Tomozawa, and H. Umezawa, Nuovo Cimento **5**, 810 (1957).
[10] A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics* (Pergamon, Oxford, 1965); D. A. Kirzhnits, *Field Theoretical Method in Many Body Systems* (Pergamon, Oxford, 1965).
[11] H. A. Weldon, Phys. Rev. D **26**, 1394 (1982).
[12] J. L. Cambier, J. R. Primack, and M. Sher, Nucl. Phys. B **209**, 372 (1982).
[13] G. W. Semenoff and H. Umezawa, Nucl. Phys. B **220** [FS8], 196 (1983).
[14] J. Whitehead, H. Matsumoto, and H. Umezawa, Phys. Rev. B **25**, 4737 (1982); *ibid.* **29**, 423 (1984).
[15] J. Schwinger, J. Math. Phys. **2**, 407 (1961).
[16] E. S. Fradkin, Sov. Phys. JETP **9**, 912 (1952).
[17] L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics: Green's Function Methods in Equilibrium and Nonequilibrium Problems* (Benjamin, New York, 1962).
[18] L. V. Keldysh, Sov. Phys. JETP **20**, 1018 (1965).
[19] R. A. Craig, J. Math. Phys. **9**, 605 (1968).
[20] R. Mills, *Propagators for Many-particle Systems* (Gordon and Breach, New York, 1969).
[21] H. Araki and E. J. Woods, J. Math. Phys. **4**, 637 (1963).
[22] H. Araki and W. Wyss, Helv. Phys. Acta **37**, 136 (1964).
[23] R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957).
[24] P. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959).
[25] *Constructive Statistical Mechanics and Structural Field Theory* in *Cargèse Lecture in Physics Vol. 4*, edited by D. Kastler (Gordon and Breach, New York, 1970).
[26] M. Schmutz, Z. Phys. B **30**, 97 (1978).
[27] H. Matsumoto, Y. Nakano, H. Umezawa, F. Mancini, and M. Marinaro, Prog. Theor. Phys. **70**, 599 (1983).
[28] A. S. Wightman, Phys. Rev. **101**, 860 (1956).

Nonlinear Schrödinger-type field equation for the description of dissipative systems. III. Frictionally damped free motion as an example for an aperiodic motion

D. Schuch, K.-M. Chung, and H. Hartmann
Institut für Physikalische und Theoretische Chemie der Johann Wolfgang Goethe-Universität, D-6000 Frankfurt (Main), Federal Republic of Germany

(Received 12 October 1983; accepted for publication 18 May 1984)

A new theory for the description of dissipative systems by nonlinear Schrödinger-type field equations (NLSE's) with logarithmic nonlinearity, which has been recently developed by the authors, is applied to investigate the frictionally damped free motion and similar spatially unrestricted aperiodic problems. Wave-packet solutions as well as time-dependent wave-function solutions are derived and discussed. In the limit of vanishing friction (friction constant $\gamma \to 0$) these solutions turn into the well-known solutions of the respective linear Schrödinger field equation. The same applies to the mean values of position, momentum, and energy, as well as to the uncertainty product of position and momentum. For $\gamma \neq 0$, however, interesting new effects appear. In contrast to the linear theory the uncertainty product of position and momentum does not diverge any more for infinitely long times, $t \to \infty$, but asymptotically approaches a definite constant value which depends on characteristic parameters of the system like its mass, initial width, and friction constant γ. Another effect, the faster spreading of the Gaussian wave-packet solution compared to the linear theory, can be explained with the help of a special property of our *nonlinear* differential equation. In a way similar to what is usually only known for *linear* differential equations, the wave-packet solutions of our NLSE can be obtained by *superposition* of the wave functions which are individually also solutions of the same NLSE. The properties of the time-dependent superposition coefficients appearing in this connection are discussed. The extension to the corresponding three-dimensional problem as well as the differences arising in the investigation of the NLSE's of the free fall and the motion in a constant electric field are given. Concluding, some differences are discussed which appear applying our nonlinear field theory to describe periodic or aperiodic motions, respectively.

PACS numbers: 11.10.Lm, 03.65. – w

I. INTRODUCTION

In parts I[1] and II[2] of our work, the derivation of our nonlinear Schrödinger-type field equation (NLSE) and the characteristic properties of our logarithmic nonlinearity are discussed in detail and our NLSE is compared with similar approaches of other authors. Up to now we only regarded spatially restricted, periodic motions like the motion in the harmonic oscillator potential, $V = (m/2)\omega_0^2 x^2$, or the motion in a magnetic field. Subject of the present work is the investigation of spatially unrestricted, aperiodic motions damped by a linearly velocity-dependent frictional force. As an example, first the NLSE for the one-dimensional free motion including friction, i.e., for the potential $V = 0$, is solved and the properties of the solutions as well as the properties of mean values evaluated with the help of these solutions are investigated. Important limits like, e.g., the behavior for vanishing friction, i.e., $\gamma \to 0$, and for the times $t = 0$ and $t \to \infty$, respectively, are regarded.

Relations between the different existing solutions of our NLSE are shown, where a special property of our NLSE, which is usually only attributed to linear differential equations, becomes evident. After a discussion of the consequences resulting from this property, the extension to the corresponding three-dimensional problem is given briefly. As further examples for aperiodic motions the free fall, $V = mgx$ (g is the constant of gravity), and the motion in a constant electric field, i.e., $V = -e\xi x$ (e is the elementary charge, ξ is the absolute value of electric field strength), as well as the corresponding NLSE's are considered and differences in comparison to the previous problem, $V = 0$, are specified.

Finally, essential differences between the results obtained from the NLSE of an aperiodic motion and those obtained from the NLSE of a periodic motion or from the linear Schrödinger field theory, respectively, are discussed in a short résumé.

II. WAVE-PACKET SOLUTIONS

A. Gaussian wave packets

The one-dimensional motion of a system exclusively affected by a linearly velocity-dependent frictional force in classical particle mechanics can be described by the following Newton's equation of motion:

$$m\ddot{x}_{kl} + m\gamma\dot{x}_{kl} = \dot{p}_{kl} + \gamma p_{kl} = 0. \tag{1}$$

The nonlinear field equation of Schrödinger-type (NLSE) corresponding to this classical corpuscular equation of motion is obtained in a way described in detail in our previous works,[1–4] and for this particular problem it is given by[5]

 0022-2488/84/103086-07$02.50

$$i\hbar \frac{d}{dt}\Psi_{\mathrm{NL}} = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\Psi_{\mathrm{NL}} - i\hbar\gamma(\ln \Psi_{\mathrm{NL}} - \langle \ln \Psi_{\mathrm{NL}}\rangle)\Psi_{\mathrm{NL}}. \quad (2)$$

For the solution of this NLSE an ansatz of the form

$$\Psi_{\mathrm{WP,NL}}(x,t) = N\exp\{-\tilde{x}^2/2a(t) + iS\}, \quad (3)$$

with

$$\tilde{x} = x - \eta(t), \quad S = K(t) + L(t)\tilde{x}, \quad a(t) = a_R(t) + ia_I(t)$$

is chosen.

With the help of the density function

$$\Psi^*_{\mathrm{WP,NL}}\cdot\Psi_{\mathrm{WP,NL}} = \rho_{\mathrm{WP,NL}}(x,t) = \left(\frac{a+a^*}{2\pi a a^*}\right)^{1/2}\exp\left\{-\tilde{x}^2\left(\frac{a+a^*}{2aa^*}\right)\right\} = \frac{1}{\sqrt{\pi}}\frac{1}{\sqrt{2\langle\tilde{x}^2\rangle}}\exp\left\{-\frac{\tilde{x}^2}{2\langle\tilde{x}^2\rangle}\right\}, \quad (4)$$

which has to obey the Fokker–Planck equation (FPE)

$$\dot{\rho}_{\mathrm{NL}} + \mathrm{div}(\rho_{\mathrm{NL}}\mathbf{v}) - D\Delta\rho_{\mathrm{NL}} = 0, \quad (5)$$

the normalization constant N can be obtained from

$$N\cdot N^* = [(a+a^*)/2\pi a a^*]^{1/2} \quad (6)$$

in the form[6]

$$N = [(a+a^*)/2\pi a^2]^{1/4}, \quad (7)$$

and the mean value of $\tilde{x}^2$ can be determined to be

$$\langle\tilde{x}^2\rangle = (a_R/2)[1 + (a_I/a_R)^2]. \quad (8)$$

Inserting this ansatz for Ψ_{NL} into the NLSE (2) leads to an equation which can be arranged in terms of powers of $\tilde{x}$.

From the term linear in $\tilde{x}$ the relations

$$L(t) = (m/\hbar)\dot{\eta}(t), \quad (9)$$

$$m\ddot{\eta} + m\gamma\dot{\eta} = 0 \quad (10)$$

result. The second of these conditions shows that $\eta(t)$ has to fulfill Newton's equation of motion (1).

Regarding the $\tilde{x}^2$ term, we find that $a(t)$ has to obey the differential equation

$$\dot{a} - \gamma a - i(\hbar/m) = 0, \quad (11)$$

which turns into the Bernoulli-type differential equation

$$\dot{\tilde{y}} + \gamma\tilde{y} + (2\hbar/m)\tilde{y}^2 = 0 \quad (12)$$

using the substitution $\tilde{y} = (i/2)[1/a(t)]$.

By the aid of the solution of this differential equation,

$$\tilde{y}(t) = \left\{\left(\frac{2\hbar}{m\gamma} + \frac{1}{\tilde{y}_0}\right)e^{\gamma t} - \frac{2\hbar}{m\gamma}\right\}^{-1} \quad (13)$$

with

$$\tilde{y}_0 = \tilde{y}(t=0),$$

the quantity $a(t) = a_R(t) + ia_I(t)$ can be determined:

$$a_R(t) = a_{R,0}\cdot e^{\gamma t}, \quad (14)$$

$$a_I(t) = a_{I,0}\cdot e^{\gamma t} + (\hbar/m\gamma)(e^{\gamma t} - 1). \quad (15)$$

For $t = 0$ we obtain

$$a_R(0) = a_{R,0}, \quad a_I(0) = a_{I,0}, \quad (16)$$

for $t\to\infty$,

$$a_R(t\to\infty) = \infty, \quad a_I(t\to\infty) = \infty. \quad (17)$$

The limit of vanishing friction, i.e., $\gamma\to 0$, leads to

$$\lim_{\gamma\to 0} a(t) = \lim_{\gamma\to 0} a_{R,0} + i\left[\lim_{\gamma\to 0} a_{I,0} + (\hbar/m)t\right]. \quad (18)$$

As $a_R(t)$, $a_I(t)$, and quantities depending thereon, like, e.g., $\langle\tilde{x}^2\rangle$, in the limit of vanishing friction have to turn into the corresponding quantities of the linear Schrödinger theory,[7] the allowed possibilities for the correct choice of the constants $a_{R,0}$ and $a_{I,0}$, respectively, are confined. The most simple choice which is in accordance with these limiting conditions, $a_{R,0} = a_0$ and $a_{I,0} = 0$, shall be discussed in the following.[8]

After the determination of $a(t)$ the mean value $\langle\tilde{x}^2\rangle$ can be stated immediately in the form

$$\langle\tilde{x}^2\rangle = (a_0/2)\{e^{\gamma t} + [2\hat{S}\sinh((\gamma/2)t)]^2\} \quad (19)$$

with

$$\hat{S} = \hbar/m\gamma a_0, \quad (20)$$

and is a measure for the width of the Gaussian function $\rho_{\mathrm{WP,NL}}(x,t)$.

In particular for

$$t = 0, \quad \langle\tilde{x}^2\rangle(0) = \tfrac{1}{2}a_0,$$
$$t\to\infty, \quad \langle\tilde{x}^2\rangle(\infty) = \infty \quad (21)$$

are obtained, and for $\gamma\to 0$,

$$\lim_{\gamma\to 0}\langle\tilde{x}^2\rangle = (a_0/2)\{1 + [(\hbar/ma_0)t]^2\} = \langle\tilde{x}^2\rangle_{\mathrm{L}}. \quad (22)$$

As required, the limit $\gamma\to 0$ exactly yields the result of the linear Schrödinger field theory, for $\gamma\neq 0$, however, the width of $\rho_{\mathrm{WP,NL}}(x,t)$ as well as of $\Psi_{\mathrm{WP,NL}}(x,t)$ increases faster than in the linear theory; i.e., the Gaussian function spreads faster than in the case without friction.

Regarding the change of $\langle\tilde{x}^2\rangle$ with respect to time,

$$\frac{d}{dt}\langle\tilde{x}^2\rangle = \gamma(a_0/2)\{e^{\gamma t} + 2\hat{S}^2\sinh\gamma t\}, \quad (23)$$

in the limit $\gamma\to 0$ this also turns into the corresponding expression of the linear theory,

$$\lim_{\gamma\to 0}\frac{d}{dt}\langle\tilde{x}^2\rangle = \left(\frac{\hbar}{m}\right)^2\frac{1}{a_0}t, \quad (24)$$

but in contrast to the linear theory, already for $t = 0$ it has a finite positive value,

$$\frac{d}{dt}\langle\tilde{x}^2\rangle(t=0) = \gamma\frac{a_0}{2}, \quad (25)$$

which vanishes for $\gamma\to 0$, i.e., in the absence of friction. An explanation, which makes physically plausible the fact that our Gaussian function spreads even faster than the Gaussian function of the linear theory, will be given in Sec. III A.

With the help of $\rho_{\mathrm{WP,NL}}(x,t)$ the "diffusion constant" D of the FPE (5) can now be easily determined in different ways, e.g., simply by inserting $\rho_{\mathrm{WP,NL}}(x,t)$ into the FPE or using the condition of separability (see parts I and II)

$$-D(\Delta\rho_{\mathrm{NL}}/\rho_{\mathrm{NL}}) = \gamma(\ln\rho_{\mathrm{NL}} - \langle\ln\rho_{\mathrm{NL}}\rangle). \quad (26)$$

The result

$$D = (\gamma/2)\langle\tilde{x}^2\rangle \quad (27)$$

shows that in this case due to the time dependence of $\langle \tilde{x}^2 \rangle$ also the "diffusion constant" D is time dependent (but *not* coordinate dependent).

Now taking another look at the "diffusion term" of the FPE, we find

$$-D\Delta\rho_{\mathrm{NL}} = (\gamma/2)[(\langle \tilde{x}^2 \rangle - \tilde{x}^2)/\langle \tilde{x}^2 \rangle]\rho_{\mathrm{NL}}, \tag{28}$$

i.e., on the one hand this time-reversibility breaking term is proportional to the friction constant γ, on the other hand, however, it is also proportional to the (relative) deviation of the quantity $\tilde{x}^2$ from its mean value; that means, this term does not only vanish in the limit $\gamma \to 0$, but also in the absence of fluctuations of $\tilde{x}^2$. As the relation

$$\tfrac{1}{2}[(\langle \tilde{x}^2 \rangle - \tilde{x}^2)/\langle \tilde{x}^2 \rangle] = \ln \rho_{\mathrm{NL}} - \langle \ln \rho_{\mathrm{NL}} \rangle \tag{29}$$

shows, this can be reached, if ρ_{NL} does not depend on the coordinates (whereas it still may be time dependent).

A stationary solution of the FPE is obtained only for $t \to \infty$ with the corresponding limits $\lim_{t\to\infty} \rho_{\mathrm{WP,NL}} = 0$ and $\lim_{t\to\infty} \langle \tilde{x}^2 \rangle = \infty$, respectively; i.e., $\rho_{\mathrm{WP,NL}}(x,t)$ asymptotically approaches a stationary final state which is characterized by an infinitely broad and at the same time completely flat density function (in analogy to the linear theory).

Finally, the $\tilde{x}$-independent term of the NLSE supplies us with an equation for the determination of $\dot{K}(t)$,

$$\dot{K}(t) = \frac{1}{\hbar}\frac{m}{2}\dot{\eta}^2 + \frac{\gamma}{4}\frac{a_I}{a_R}, \tag{30}$$

from which the coefficient $K(t)$ might be easily obtained.

B. Mean values and uncertainty relation

With the aid of the wave-packet solution the mean value of energy can be determined to be

$$\left\langle i\hbar\frac{d}{dt}\right\rangle = \frac{m}{2}\dot{\eta}^2 + \frac{\hbar^2}{4ma_R} = \frac{m}{2}\dot{\eta}^2 + \frac{\hbar^2}{4ma_0}e^{-\gamma t}. \tag{31}$$

We obtain for $t = 0$,

$$\left\langle i\hbar\frac{d}{dt}\right\rangle(t=0) = \frac{m}{2}\dot{\eta}^2(0) + \frac{\hbar^2}{4ma_0} \tag{32}$$

for $t \to \infty$,

$$\left\langle i\hbar\frac{d}{dt}\right\rangle(t\to\infty) = 0, \tag{33}$$

and find that again in the limit $\gamma \to 0$ this result turns into the corresponding quantity of the undamped problem,

$$\lim_{\gamma\to 0}\left\langle i\hbar\frac{d}{dt}\right\rangle = \frac{m}{2}\dot{\eta}_0^2 + \frac{\hbar^2}{4ma_0}. \tag{34}$$

What is remarkable is that taking into account friction the mean value $\langle i\hbar(d/dt)\rangle = \langle H \rangle$ for $t \to \infty$ completely vanishes [because of $(\hbar^2/4ma_0)\cdot e^{-\gamma t}$], whereas without friction, even for $\dot{\eta}_0 = 0$ always a constant, contribution $\hbar^2/4ma_0$ exists. This phenomenon is closely connected to another one concerning the time dependence of the uncertainty product of position and momentum. Therefore, we will discuss both phenomena together after the determination of this uncertainty product.

For this purpose we need the relations

$$\overline{\Delta p^2} = \langle p^2 \rangle - \langle p \rangle^2 = \hbar^2/2a_R, \tag{35}$$

$$\overline{\Delta x^2} = \langle x^2 \rangle - \langle x \rangle^2 = \langle \tilde{x}^2 \rangle. \tag{36}$$

In our nonlinear theory this leads to the uncertainty product

$$U_{\mathrm{NL}} = \overline{\Delta p^2}\cdot\overline{\Delta x^2} = \frac{\hbar^2}{4}\left[1 + \left(\frac{a_I}{a_R}\right)^2\right]$$

$$= \frac{\hbar^2}{4}[1 + \hat{S}^2(1 - e^{-\gamma t})^2] \geq \frac{\hbar^2}{4}, \tag{37}$$

with a *minimum* at $t = 0$,

$$U_{\mathrm{NL},0} = \hbar^2/4 = U_{\mathrm{NL,min}}, \tag{38}$$

and a *maximum* at $t \to \infty$,

$$U_{\mathrm{NL},\infty} = (\hbar^2/4)(1 + \hat{S}^2) = U_{\mathrm{NL,max}}. \tag{39}$$

The limit of vanishing friction again yields the expression of the linear Schrödinger theory,

$$\lim_{\gamma\to 0} U_{\mathrm{NL}} = (\hbar^2/4)[1 + ((\hbar/ma_0)t)^2] = U_{\mathrm{L}}, \tag{40}$$

and hence for the *minimum* at $t = 0$,

$$U_{\mathrm{L},0} = \hbar^2/4 = U_{\mathrm{L,min}} = U_{\mathrm{NL,min}}, \tag{41}$$

and for the *maximum* at $t \to \infty$,

$$U_{\mathrm{L},\infty} = \infty = U_{\mathrm{L,max}}. \tag{42}$$

It is striking that without friction for $t \to \infty$ the product $\overline{\Delta p^2}\cdot\overline{\Delta x^2}$ diverges, whereas taking friction into consideration this product approaches a finite *constant* value. At that point again in our nonlinear theory the dimensionless quantity $\hat{S} = \hbar/m\gamma a_0$ appears, which is a *pure number* characterizing the *system* also with respect to its *maximum uncertainty*.

We list some consequences.

(1) If m or a_0 are very large (like in macroscopic systems), $\hat{S}$, and hence $U_{\mathrm{NL,Max}}$, are very small, and vice versa.

(2) For $\hbar \to 0$ $\hat{S}$ also vanishes.

(3) If $m\gamma$ is very small, $\hat{S}$, and hence the effect on a microscopic level, is very large, however, the effect on a macroscopic level, like, e.g., that represented by the friction term $-m\gamma\mathbf{v}$ in Newton's equation of motion, is very small, and vice versa. Due to this combined appearance of m and γ, the same quantity γ is appropriate for a microscopic as well as for a macroscopic description.

(4) If γ is very large, which corresponds to frequent interactions between the system and the surrounding. $\hat{S}$, and hence $U_{\mathrm{NL,Max}}$, are very small; however, if the period of time between two interactions, $\tau = 1/\gamma$, is very large, $\hat{S}$, and hence $U_{\mathrm{NL,Max}}$, also increase, and in the limit of complete absence of interaction (i.e., for $\tau \to \infty$, or $\gamma \to 0$, respectively) we finally end up with divergency.

(5) Even in the limit of "permanent interaction" or "permanent observation," respectively, i.e., for $\tau \to 0$, it is not possible to come below the lower bound $U_{\mathrm{NL,Min}} = U_{\mathrm{L,Min}} = \hbar^2/4$.

(6) As $\overline{\Delta x^2} = \langle \tilde{x}^2 \rangle$ diverges for $t \to \infty$, a finite limit $U_{\mathrm{NL,Max}}$ for $t \to \infty$ can only be reached if for $t \to \infty$ the value of $\overline{\Delta p^2}$ tends to zero in such a way that the product of both

factors yields a finite value. But from $\overline{\Delta p^2}(t\to\infty)=0$ also follows that $\langle p^2\rangle(t\to\infty)=0$ and therefore $\langle p^2/2m\rangle=\langle i\hbar(d/dt)\rangle(t\to\infty)=0$; i.e., for $t\to\infty$ the kinetic energy completely vanishes. This is the above-mentioned connection between the behavior of the mean value of energy and the uncertainty product of position and momentum. So, the divergency of U_L for $t\to\infty$ is due to the constant finite value accepted by $\langle i\hbar(d/dt)\rangle$ and hence also accepted by $\overline{\Delta p^2}$. This fact together with the divergency of $\overline{\Delta x^2}=\langle\tilde{x}^2\rangle$ causes the product of these two quantities to diverge in the linear theory also.

III. WAVE-FUNCTION SOLUTIONS

A. Wave-packet solution as a superposition of wave-function solutions

A shortcoming attributed to nonlinear differential equations, and thus also to nonlinear Schrödinger-type field equations, is that, in contrast to linear differential equations, wave packets cannot be constructed in the usual way by superposition of wave-function solutions of the respective differential equation (see, e.g., Refs. 9–12). This problem shall be investigated now in further detail with regard to our NLSE.

The Gaussian wave-packet solution $\Psi_{\mathrm{WP,L}}$ which is obtained from the solution $\Psi_{\mathrm{WP,NL}}$ of our NLSE in the limit of vanishing friction, is a general solution of the linear SE of free motion. As the *superposition principle* is valid in the *linear* theory, this wave-packet solution can be represented by means of the particular solutions

$$\Psi_{k,\mathrm{L}}(x,t)=(1/\sqrt{2\pi})\cdot\exp\{i[kx-\omega_k t]\} \tag{43}$$

in the form

$$\Psi_{\mathrm{WP,L}}(x,t)=\frac{1}{\sqrt{2\pi}}\int dk\,A_{\mathrm{L}}(k)\Psi_{k,\mathrm{L}}(k,t)$$
$$=\frac{1}{\sqrt{2\pi}}\int dk\,A_{\mathrm{L}}(k)e^{i(kx-\omega_k t)}, \tag{44}$$

where the functions $\Psi_{k,\mathrm{L}}$ describe plane waves traveling with the constant group velocity $v_g=(\hbar/m)k$ ($v_g=2v_p$; v_p is the phase velocity) and $\omega_k=\hbar k^2/2m$ is also constant for given k.

The expansion coefficients $A_{\mathrm{L}}(k)$ can be determined from

$$A_{\mathrm{L}}(k)=\frac{1}{\sqrt{2\pi}}\int dx\,\Psi_{\mathrm{WP,L}}(x,t)\cdot\Psi^*_{k,\mathrm{L}}(x,t). \tag{45}$$

As these coefficients $A_{\mathrm{L}}(k)$ are not time-dependent, it is practical to solve this integral for the most simple case, i.e., for $t=0$,

$$A_{\mathrm{L}}(k)=\frac{1}{\sqrt{2\pi}}\int dx\,\Psi_{\mathrm{WP,L}}(x,0)\cdot e^{-ikx}, \tag{46}$$

which means that the Fourier transform of the initial state $\Psi_{\mathrm{WP,L}}(0,t)$ has to be determined.

For the Gaussian wave packet under consideration this leads to

$$A_{\mathrm{L}}(k)=\left(\frac{a_0}{\pi}\right)^{1/4}\exp\{-(a_0/2)(k_{\mathrm{WP}}-k)^2\} \tag{47}$$

with k_{WP} being the constant k-value of the wave-packet solution.

In order to investigate whether it is possible to find a representation of the wave-packet-type solution of our *nonlinear* problem comparable with the one of the linear theory which is based on the *superposition principle*, it first has to be analyzed whether there also exist wave-function solutions of our NLSE comparable with the plane waves of the linear problem.

Such functions $\Psi_{k,\mathrm{NL}}(x,t;k(t))$ can actually be determined. They have the form

$$\Psi_{k,\mathrm{NL}}(x,t;k(t))=(1/\sqrt{2\pi})\cdot\exp\{i[k(t)(x-a_I k(t))$$
$$+(a_I/4)k^2(t)(e^{\gamma t}+1)]\}, \tag{48}$$

with

$$k(t)=k_0\cdot e^{-\gamma t}$$

and for $\gamma\to 0$ they turn into the plane wave solutions of the linear SE.

The mean value of energy, given by

$$\left\langle i\hbar\frac{d}{dt}\right\rangle=\frac{m}{2}\dot{\eta}^2(t)=\frac{\hbar^2k_0^2}{2m}\cdot e^{-2\gamma t}, \tag{49}$$

for $t\to\infty$ approaches zero.

An expansion of the wave-packet solution of our NLSE, $\Psi_{\mathrm{WP,NL}}(x,t;k_{\mathrm{WP}}(t))$, appropriately written in the form

$$\Psi_{\mathrm{WP,NL}}=\left(\frac{1}{a}\right)^{1/2}\left(\frac{a_R}{\pi}\right)^{1/4}\exp\left\{-\frac{1}{2a}(x-a_I k_{\mathrm{WP}})^2\right.$$
$$\left.+i\left[k_{\mathrm{WP}}(x-a_I k_{\mathrm{WP}})+\frac{a_I}{4}k_{\mathrm{WP}}^2(e^{\gamma t}+1)+\chi(t)\right]\right\}, \tag{50}$$

with

$$\dot{\chi}(t)=(\gamma/4)(a_I/a_R), \tag{51}$$

has the form

$$\Psi_{\mathrm{WP,NL}}(x,t;k_{\mathrm{WP}}(t))$$
$$=\int dk(t)\,A_{\mathrm{NL}}(k)\Psi_{k,\mathrm{NL}}(x,t;k(t)). \tag{52}$$

The coefficients $A_{\mathrm{NL}}(k)$ are determined with the help of the relation

$$A_{\mathrm{NL}}(k)=\int dx\,\Psi_{\mathrm{WP,NL}}(x,t)\cdot\Psi^*_{k,\mathrm{NL}}(x,t), \tag{53}$$

taking into account that in contrast to the linear problem now these coefficients A_{NL} are time dependent due to the time dependence of $k=k(t)$. Therefore, it is not sufficient to determine the coefficients for the most simple case $t=0$, but one has to use the explicitly time-dependent functions $\Psi_{\mathrm{WP,NL}}(x,t)$ and $\Psi_{k,\mathrm{NL}}(x,t)$ for the calculation. In this way, we finally obtain

$$A_{\mathrm{NL}}(k)=\left(\frac{a_R}{\pi}\right)^{1/4}\exp\left\{-\frac{a_R}{2}(k-k_{\mathrm{WP}})^2\right\}$$
$$\times\exp\left\{i\left[\frac{a_I}{4}(k_{\mathrm{WP}}^2-k^2)(e^{\gamma t}-1)+\chi(t)\right]\right\}. \tag{54}$$

For $t=0$ these coefficients are identical with those of the linear theory,

$$A_{NL}(k(0)) = A_{NL}(k_0), \tag{55}$$

which are also obtained in the limit $\gamma \to 0$. For $t \neq 0$, however, due to the time-dependence of the $A_{NL}(k)$ they are different from those of the linear theory.

The variation in time of the coefficients is given by

$$\dot{A}_{NL}(k) = f(k) \cdot A_{NL}(k), \tag{56}$$

with

$$f(k) = (\gamma/2)\{\tfrac{1}{2} + a_R(k_{WP} - k)^2 + ia_I(k_{WP}^2 - k^2) + i\tfrac{1}{2}a_I/a_R\}, \tag{56'}$$

and in particular for $t = 0$ by

$$\dot{A}_{NL}(t=0) = \gamma\{(a_0/2)(k_{WP,0} - k_0)^2 + \tfrac{1}{4}\}A_L(k_0). \tag{57}$$

In the limit of vanishing friction the variation in time also vanishes, which is in agreement with the coefficients of the undamped problem being constant in time.

So the problem just discussed represents an example for a nonlinear differential equation for which a Gaussian-type solution function exists which can be expanded in terms of wave functions. As a special feature in this case, however, the fact has to be regarded that each of these expansion functions is also a solution of this nonlinear differential equation. That means we can construct a solution of a *nonlinear* differential equation by *superposition* of other solutions of the *same* equation which is usually valid only for linear differential equations. In this sense, in analogy to the linear case, our Gaussian function can actually be denoted as a wave-packet, constructed by superposition of wave-function solutions provided with appropriate coefficients.

From the explicit form of the components $\Psi_{k,NL}$ and $A_{NL}(k)$, respectively, another feature of the wave-packet solution of our NLSE also becomes transparent, namely its faster spreading compared to the corresponding wave packet of the linear SE.

In the *linear* case, for $t = 0$ the Gaussian wave packet is constructed from conponents $\Psi_{k,NL}$ with *different* but *constant* k-values, where the contributions of the individual components are determined by the *constant* coefficients $A_L(k)$.

For $t > 0$ a dephasing, and hence spreading, of the Gaussian function starts, as the individual components, although having constant ceofficients $A_L(k)$, have different k-values and therefore different group and phase velocities. However, the *differences in velocity* remain *constant* for all times t.

In the *nonlinear* case, for $t = 0$ the situation is analogous to the linear case, only now the *k-values* and *coefficients* $A_{NL}(k)$ are *time dependent*. For $t > 0$ this entails the following consequences.

(1) In addition to the effect of the linear case based on the different initial k-values, in the nonlinear case because of $k(t) = k_0 e^{-\gamma t}$ a different change in time of these different initial k-values takes place, wherefrom *velocity differences* result which are *variable in time*.

(2) Due to the time dependence of the coefficients $A_{NL}(k)$ the contributions of the individual *components* $\Psi_{k,NL}$ to the wave-packet are also *variable in time*.

That means, the reason for the faster dephasing, and hence spreading, of the wave-packet compared to the linear case is twofold. It is not only a consequence of the *variable velocity differences* between the components of the wave packet, caused by diferent changes of the initial velocities during the progress of time, but also a consequence of the *time dependence* of the *contributions* $A_{NL}(k)$ of the components.

B. Further properties of the coefficients $A_{NL}(k)$

A further subject of interest, e.g., with respect to the behavior of the density function $\rho_{WP,NL} = \Psi^*_{WP,NL}\Psi_{WP,NL}$, are the properties of the absolute values of the coefficients, $|A_{NL}(k)|$:

$$|A_{NL}(k)| = (a_R/\pi)^{1/4} \cdot \exp\{-(a_R/2)\hat{k}^2\} \tag{58}$$

with

$$|\hat{k}| = |k_{WP} - k|.$$

This quantity depends on the difference between the k-value concerned and the fixed constant value k_{WP}.

Regarding the change in time also for this quantity, we find the following relation

$$\frac{d}{dt}|A_{NL}(k)| = \gamma\left\{\frac{a_R}{2}\hat{k}^2 + \frac{1}{4}\right\}\left(\frac{a_R}{\pi}\right)^{1/4} \cdot \exp\left\{-\frac{a_R}{2}\hat{k}^2\right\}. \tag{59}$$

[Note: $(d/dt)|A_{NL}(k)| \neq |(d/dt)A_{NL}(k)|$.

Therefrom we see that the change in time of $|A_{NL}(k)|$ on the one hand is proportional to the friction constant γ and vanishes for $\gamma \to 0$ and on the other hand depends on $|A_{NL}|$ itself and on $|\hat{k}|$. The dependence on $|\hat{k}|$ is of such a kind, that one term exists which is proportional to $\hat{k}^2 e^{-(a/2)R\hat{k}^2}$, therefore corresponding to a Maxwell–Boltzmann-distribution law, and a second term which is proportional to $e^{-(a/2)R\hat{k}^2}$, thus describing an exponential decay. This second term warrants that also for $\hat{k} = 0$, i.e., for $k = k_{WP}$, the absolute value $|A_{NL}|$ changes in time, despite the fact that the first term vanishes in this case; i.e., *all* coefficients are time dependent. A rough qualitative draft of the change in time of $|A_{NL}(k)|$ as a function of $|\hat{k}|$ shows a curve starting at $|\hat{k}| = 0$ with the positive value $\frac{1}{4}[\gamma((a/\pi)R)^{1/4}]$, running through a maximum at $|\hat{k}_{Max}| = (3/2a_R)^{1/2}$ with the maximum value of $e^{-3/4}[\gamma((a/\pi)R)^{1/4}]$ and finally approaching the $|\hat{k}|$-axis asymptotically (see Fig. 1).

The maximum value itself as well as its position, i.e., $|\hat{k}_{Max}|$, are time dependent, whereby the position of the maximum approaches $|\hat{k}| = 0$ with the progress of time.

IV. THREE-DIMENSIONAL FREE MOTION INCLUDING FRICTION

With the aid of the product ansatz

$$\Psi(\mathbf{r},t) = \Psi_x(x,t) \cdot \Psi_y(y,t) \cdot \Psi(z,t), \tag{60}$$

the corresponding NLSE

$$\left\{i\hbar\frac{d}{dt} + \frac{\hbar^2}{2m}\Delta + i\hbar\gamma(\ln\Psi - \langle\ln\Psi\rangle)\right\}\Psi(\mathbf{r},t) = 0 \tag{61}$$

can be separated into three NLSE's of the same kind,

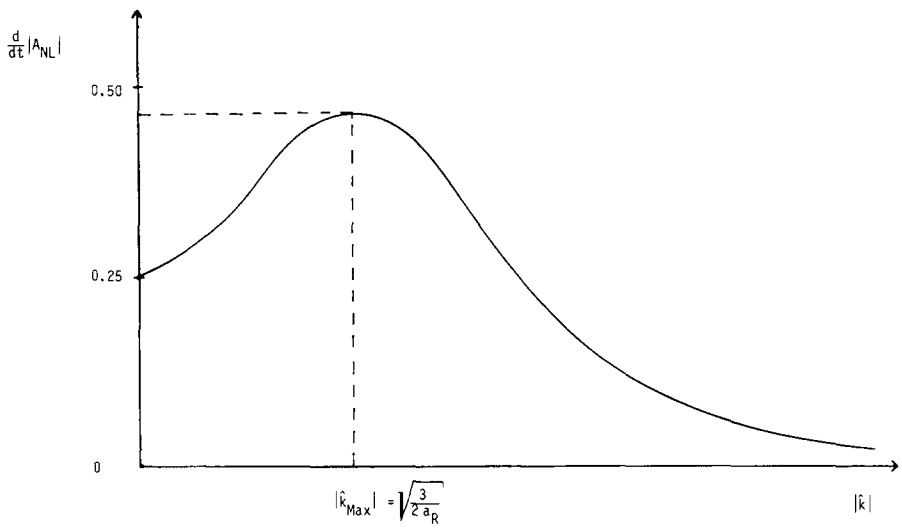


FIG. 1. Rough qualitative draft of $(d/dt)|A_{NL}(k)|$ as a function of $|\hat{k}|$. Ordinate in units of $[\gamma(a_R/\pi)^{1/4}]$.

$$\left\{i\hbar\frac{d}{dt} + \frac{\hbar^2}{2m}\frac{d^2}{dx_i^2} + i\hbar\gamma(\ln\Psi_{x_i} - \langle\ln\Psi_{x_i}\rangle)\right\}\Psi_{x_i}(x_i,t) = 0, \tag{62}$$

which is the NLSE of the one-dimensional free motion including friction, already discussed in the previous sections of this work.

V. FURTHER APERIODIC MOTIONS

A. Free fall including friction

Inserting the potential $V = mgx$, wherefrom the constant force of gravity $K = -mge_x$ (with e_x = unit vector in the x-direction) can be derived, into the NLSE and again arranging the terms in powers of $\tilde{x} = x - \eta(t)$, in comparison to the NLSE of the free motion we obtain an additional contribution to the $\tilde{x}$- and $\tilde{x}^0$-terms, respectively; the $\tilde{x}^2$-term remains unchanged.

The additional contribution to the $\tilde{x}$-term entails that $\eta(t)$ now has to obey the classical equation of motion

$$m\ddot{\eta} = -mg - m\gamma\dot{\eta}. \tag{63}$$

Due to the additional $\tilde{x}^0$-term $\dot{K}(t)$ for the wave-function solution in this case has the form

$$\dot{K} = \frac{1}{\hbar}\left(\frac{m}{2}\dot{\eta}^2 - mg\eta\right) = \frac{1}{\hbar}\mathcal{L}(\eta,\dot{\eta};t), \tag{64}$$

with

$$\mathcal{L}(\eta,\dot{\eta};t) = (m/2)\dot{\eta}^2 - mg\eta = T(\dot{\eta}) - V(\eta)$$

and

$$\left\langle i\hbar\frac{d}{dt}\right\rangle = \frac{m}{2}\dot{\eta}^2 + mg\eta = T(\dot{\eta}) + V(\eta), \tag{65}$$

and for the wave-packet solution

$$\dot{K} = \frac{1}{\hbar}\left(\frac{m}{2}\dot{\eta}^2 - mg\eta\right) + \frac{\gamma}{4}\frac{a_I}{a_R}, \tag{66}$$

respectively.

B. Constant electric field including friction

As in this case the potential $V = e\xi x$ also linearly depends on x, just as in the preceding problem we get an additional $\tilde{x}$- and $\tilde{x}^0$-term, respectively.

The $\tilde{x}$-term leads to the classical equation of motion

$$m\ddot{\eta} = e\xi - m\gamma\dot{\eta}. \tag{67}$$

From the $\tilde{x}^0$-term results

$$\dot{K} = \frac{1}{\hbar}\left(\frac{m}{2}\dot{\eta}^2 + e\xi\eta\right) = \frac{1}{\hbar}\mathcal{L}(\eta,\dot{\eta};t) \tag{68}$$

for the wave-function solution and

$$\dot{K} = \frac{1}{\hbar}\left(\frac{m}{2}\dot{\eta}^2 + e\xi\eta\right) + \frac{\gamma}{4}\frac{a_I}{a_R} \tag{69}$$

for the wave-packet solution, respectively.

VI. CONCLUSION

We would like to conclude with a few remarks comparing periodic and aperiodic motions in the linear Schrödinger field theory and our nonlinear field theory, respectively.

Describing a spatially restricted periodic motion like the motion in the potential of a harmonic oscillator by the aid of both methods mentioned above, it is found that in both theories a positive ground state energy has to exist in order to fulfill the uncertainty relation of position and momentum. In this ground state, however, the system has a momentum and thus a velocity larger than zero. Therefore, in the nonlinear theory which takes into account a velocity-dependent frictional force, also in this energetically lowest state there still exists an interaction with the surroundings (for details see, e.g., Refs. 1–3).

Regarding now a spatially unrestricted aperiodic motion like the free motion, in the framework of the linear theory, i.e., without friction, we find that always $U_L \geqslant \hbar^2/4$ is fulfilled.

(1) $\overline{(\Delta p^2)_L}$ always has a constant positive value $\hbar^2/2a_0$. This also manifests in the fact that for every choice of $\dot{\eta}_{kl}$, even for $\dot{\eta}_{kl} = 0$, the energy of the system contains a constant positive contribution $E_{L,0} = \hbar^2/4ma_0 > 0$, in a way a kind of "ground state" energy.

(2) $\overline{(\Delta x^2)_L}$ is certainly time dependent, but it only accepts values between $a_0/2 \leqslant \overline{(\Delta x^2)_L} \leqslant \infty$.

But $\overline{\Delta p^2}$ does not necessarily have to have a positive

value different from zero for all times t in order that $U \geqslant \hbar^2/4$ is valid. It only has to be warranted that the product fulfills $\overline{\Delta x^2} \cdot \overline{\Delta p^2} \geqslant \hbar^2/4$, and in particular that it does not completely vanish. In this connection it is quite possible that the factor $\overline{\Delta p^2}$ changes in time in a way that it becomes zero under certain conditions (e.g., for $t \to \infty$), if only at the same time $\overline{\Delta x^2}$ changes in a way that always the product of both factors fulfills the uncertainty relation.

It is exactly this case that we find investigating our NLSE for the frictionally damped free motion. Therefore, in spite of satisfying the uncertainty relation, $\overline{(\Delta p^2)_{\mathrm{NL}}} = 0$ is nevertheless possible for $t \to \infty$ and consequently the complete dissipation of the energy (also of the contribution $E_{\mathrm{L},0}$, mentioned above) can be reached. Thereby any further interaction with the surroundings (due to $\mathbf{p} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$, respectively) also comes to an end.

One more consequence of this behavior of $\overline{(\Delta x^2)_{\mathrm{NL}}}$ and $\overline{(\Delta p^2)_{\mathrm{NL}}}$ is the effect that in contrast to the linear theory U_{NL} does not diverge for $t \to \infty$, but approaches a constant value which is characterized by the system-specific quantity $\hat{S} = \hbar/m\gamma a_0$. A more detailed discussion of this fact has already been given in Sec. II B.

Another important result of the present investigation concerns the construction of a wave-packet solution of a differential equation by superposition of solution functions of this equation.

In a theory based on a linear differential equation (in this case the SE) it is possible, for periodic problems as well as for aperiodic problems, to obtain solutions of this equation by superposition of other already known solutions of this equation due to the mathematical properties of linear differential equations. Moreover, every arbitrary function can be represented as a superposition by the solutions of such a linear differential equation, if these solutions form a complete basis set. So for the harmonic oscillator the corresponding properties of the Hermite polynomials can be applied and for the free motion we can apply the properties of the plane waves.

Investigating the periodic problem of the undercritically damped harmonic oscillator[1-3] it was certainly also possible to represent the wave-packet-like solution with the help of Hermite polynomials containing the correct reduced frequency $\Omega = (\omega_0^2 - \gamma^2/4)^{1/2}$, but this was only a consequence of the just-mentioned completeness of this basis set. The individual functions of the representation, however are *no* solutions of the NLSE (except for the function of the stationary final state), a fact which is not unusual for nonlinear differential equations.

On the contrary, the result obtained for the here-investigated damped free motion is quite unusual. In this case we also found a wave-packet solution which can be expanded with the aid of a set of functions similar to the plane waves of the linear theory and even turning into these for $\gamma \to 0$. But now the innovation lies in the fact that *each* of these *wave functions* by itself is also a *solution of this NLSE*. So we can here in fact speak of a real wave packet. This phenomenon is quite untypical for nonlinear differential equations (for some works concerned with superposition in other NLSE's see, e.g., Refs. 13–16) and resulting consequences will be considered in forthcoming works.

ACKNOWLEDGMENTS

The authors express their thanks to the "Deutsche Forschungsgemeinschaft" for financial support and would also like to thank the "Akademie der Wissenschaften und der Literatur zu Mainz" for supporting this work.

[1] D. Schuch, K.-M. Chung, and H. Hartmann, J. Math. Phys. **24**, 1652 (1983).
[2] D. Schuch, K.-M. Chung, and H. Hartmann, Int. J. Quantum Chem. **25**, 391 (1984).
[3] D. Schuch, Ph.D. thesis, Department of Physical and Theoretical Chemistry, University of Frankfurt (Main), 1982.
[4] D. Schuch, K.-M. Chung, and H. Hartmann, "Ion Cyclotron Resonance Spectrometry II, in *Lecture Notes in Chemistry Vol. 31* (Springer-Verlag, Berlin, 1982), p. 514.
[5] The symbol $\langle A \rangle$ denotes the mean value of a quantity A, evaluated according to $\langle A \rangle = \int_{-\infty}^{+\infty} \Psi^* A \Psi \, dx$, where the functions Ψ are assumed to be normalized.
[6] Note, that the normalization constant is *not* given by $N = (N \cdot N^*)^{1/2} = N^*$, as this form would not yield the correct normalization constant of the undamped problem in the limit $\gamma \to 0$.
[7] See, e.g., S. Flügge, *Practical Quantum Mechanics* (Springer-Verlag, New York 1974), p. 31. Our constant $a_{R,0}$ corresponds to Flügge's a^2.
[8] This is not the only choice of the constants $a_{R,0}$ and $a_{I,0}$ which in accordance with the limiting conditions yields a solution with physical significance. Another solution of this kind is discussed in the context of a different investigation in a forthcoming paper.
[9] R. W. Hasse, J. Math. Phys. **16**, 2005 (1975).
[10] B. Remaud and E. S. Hernandez, J. Phys. A **13**, 2013 (1980).
[11] J. D. Immele, K.-K. Kan, and J. J. Griffin, Nucl. Phys. A **241**, 47 (1975).
[12] B. Remaud and E. S. Hernandez, Physica A **103**, 35 (1980).
[13] I. Bialynicki-Birula and J. Mycielski, Ann. Phys. (N.Y.) **100**, 62 (1976).
[14] N. Gisin, Ph.D. thesis, Department of Theoretical Physics, University of Geneva, 1982.
[15] N. Gisin, "A model of irreversible deterministic quantum dynamics," UGVA-DPT 1983/06-399.
[16] N. Gisin, J. Math. Phys. **24**, 1779 (1983).

Higgs fields from symmetric connections—The bundle picture

L. B. Hudson and R. Kantowski
Department of Physics and Astronomy, The University of Oklahoma, Norman, Oklahoma 73019

(Received 30 November 1983; accepted for publication 24 February 1984)

The derivation (Forgács and Manton) of a classical gauge plus Higgs field theory from a purely symmetric gauge theory by dimensional reduction is translated into the pictorial language of symmetric connections on principal bundles. The bundle description provides a better understanding of symmetric gauge theories; its global nature excludes some gauge group-symmetry group combinations, and shows that the reduced gauge group is larger than the centralizer C_H as previously thought. In general it is a certain factor group, N_K/K, involving both the gauge and symmetry groups, and contains C_H as a normal subgroup. The commonly used F^2 Lagrangian of symmetric gauge theories will not prescribe the dynamics of fields associated with this larger group, unless these fields are suitably constrained. It is shown that a constrained Kaluza–Klein metric is required to construct the F^2 Lagrangian, thus pointing the way to include dynamics for the N_K/K group.

PACS numbers: 11.15. − q, 02.40. + m, 11.30. − j

I. INTRODUCTION

Several attempts have been made to "derive" the electroweak theory,[1–6] however, the approach of Forgács and Manton[7] seems to be the most promising (we will refer to the work of these authors as FM). This approach has been successfully used to derive the bosonic sector of the theory[8] as well as to produce the fermion–Higgs couplings.[9,10] The approach is simply to start with a gauge field (group G) on an extended space-time $\mathbf{M}$ (4 + extra dimensions) with metric and require that the gauge and metric fields be invariant under a symmetry group S. The symmetry removes the extra dimensions of extended space-time as being true independent variables and also turns some components of the vector potentials on $\mathbf{M}$ into Higgs fields corresponding to a reduced gauge group (see Secs. II and III) on reduced space-time $\bar{\mathbf{M}}$ (four dimensions). The remaining vector potential components on $\mathbf{M}$ transform as vector potentials on $\bar{\mathbf{M}}$ but under a smaller gauge group.

Parts of the construction seem somewhat arbitrary and others difficult to follow when presented in the usual vector potential approach of gauge theories. However, to those familiar with principal bundles it becomes extremely transparent when stated in terms of a symmetric connection. In particular a careful bundle analysis clearly shows that N_K/K (not C_H) is the correct reduced gauge group before spontaneous symmetry breaking (see Sec. II), and also gives explicitly which representations of N_K/K can appear as Higgs fields (see Sec. III). Some of these results have been independently found by Jadczyk and Pilch.[11]

In general the resulting gauge field on $\bar{\mathbf{M}}$ corresponding to N_K/K contains fields from both the symmetric metric and the symmetric gauge field (group G) on $\mathbf{M}$. However, in FM the metric on $\mathbf{M}$ was constrained so that only part of the N_K/K connection was nonintegrable, i.e., the part corresponding (isomorphic) to the centralizer group[12] $C_H \subset G$ [see Sec. II, Defs. (f) and (g)]. Consequently, with a proper choice of gauge (cross section) the nonvanishing vector potentials of N_K/K take their values only in the Lie algebra of C_H and are completely determined by the original symmetric gauge field on $\mathbf{M}$. In the conventional presentation, the dynamics is given by taking $\int -\frac{1}{4}F^2\, dV_m$ as the action for the symmetric gauge field on $\mathbf{M}$. Because of the invariance of both the gauge field and the metric on $\mathbf{M}$, this integral reduces to an action on $\bar{\mathbf{M}}$ for the reduced gauge plus Higgs fields, including a Higgs potential, which for the proper gauge-group–symmetry-group combination is exactly the potential of the GWS theory[8] (see 5.28). In this conventional approach the construction of the F^2 Lagrangian appears to depend on a somewhat arbitrary choice of metric on extended space-time. However, in the bundle description the metric on $\mathbf{M}$ is seen to be the metric of a reduced Kaluza–Klein space.[13]

In the next three sections we define several needed mathematical structures. In Sec. II we give those definitions necessary to understand symmetric principal bundles, and in Sec. III those needed to understand invariant connections on such bundles along with their related Higgs fields. Also in Sec. III we elucidate the global construction by looking at the familiar Higgs fields of rank 2 gauge groups.[8] In Sec. IV we define Kaluza–Klein spaces and reduced Kaluza–Klein spaces after pointing out that structures beyond a symmetric connection are required in the FM construction. We are then able to give the full geometrical description of dimensionally reduced symmetric gauge theories. In Sec. V we express the commonly used F^2 action on extended space-time in a gauge-independent form and use the S symmetry to reduce it to an action on space-time for the reduced gauge plus Higgs fields. For completeness we then apply the reduced action to the examples of Sec. III and derive the bosonic sector of the GWS theory.[8] Even though these rank 2 models are incorrect when fermions are included,[9] no currently known model properly includes both.

II. THE SYMMETRIC PRINCIPAL BUNDLE

If a simple Lie group S acts as a group of bundle automorphisms of a principal bundle $\mathcal{Q} = [\mathbf{Q}, \mathbf{M}, \pi, G, \Psi]$ we call $\mathcal{Q}$ symmetric.[14–18] As usual we take $\mathbf{Q}$ to be the bundle space, $\mathbf{M}$ the base space (extended space-time), G the (simple compact) Lie group, π the projection of $\mathbf{Q}$ onto $\mathbf{M}$, and Ψ the right action of G on $\mathbf{Q}$ (see Fig. 1). Symmetry means there exists a map Φ: $\mathbf{Q}\times S \to \mathbf{Q}$ such that

 0022-2488/84/103093-12$02.50

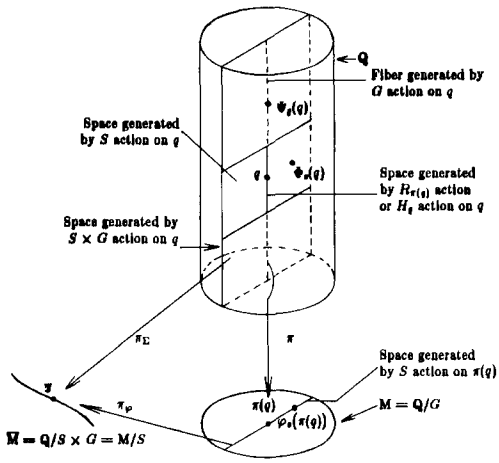


FIG. 1. The symmetric principal bundle $\mathcal{Q} = [\mathbf{Q}, \mathbf{M}, \pi, G, \Psi]$ and associated quotient space $\overline{\mathbf{M}}$.

Φ_s is a diffeomorphism,

$$\Phi_{s_1} \circ \Phi_{s_2} = \Phi_{s_2 s_1}, \quad \Phi_s \circ \Psi_g = \Psi_g \circ \Phi_s. \tag{2.1}$$

We will denote the induced action of S on $\mathbf{M}$ by φ, i.e., φ: $\mathbf{M} \times S \to \mathbf{M}$, where

$$\varphi_s \circ \pi = \pi \circ \Phi_s. \tag{2.2}$$

We can also define an action $\mathbf{\Sigma}$ of $S \times G$ on $\mathbf{Q}$, i.e., $\mathcal{\Sigma}$: $\mathbf{Q} \times (S \times G) \to \mathbf{Q}$ by

$$\mathbf{\Sigma}_{(s,g)} = \Psi_g \circ \Phi_s = \Phi_s \circ \Psi_g. \tag{2.3}$$

Several essential structures can now be defined

(a) $\overline{\mathbf{M}}$ as reduced space-time

$$\overline{\mathbf{M}} \equiv \mathbf{Q}/(S \times G) \simeq \mathbf{M}/S,$$

where factorization is by the action of $\mathcal{\Sigma}$ and φ, respectively (see Fig. 1). We *assume* the S and G actions are such that the factor spaces are all well-defined differentiable manifolds, and that the projections π_φ: $\mathbf{M} \to \overline{\mathbf{M}}$, π_Σ: $\mathbf{Q} \to \overline{\mathbf{M}}$ have locally trivial sections with fibers generated by homogeneous actions of S and $S \times G$, respectively. Here π_φ and π_Σ are defined by

$$\begin{aligned} &\pi_\varphi(m) \equiv \varphi_S(m) \equiv \{\varphi_s(m) \in \mathbf{M} \,|\, s \in S\} \in \mathbf{M}/S, \\ &\pi_\Sigma(q) \equiv \mathcal{\Sigma}_{S \times G}(q) \equiv \{\mathcal{\Sigma}_{(s,g)}(q) \in \mathbf{Q} \,|\, (s,g) \in S \times G\}, \end{aligned} \tag{2.4}$$

and since $m \equiv \Psi_G(q) \equiv \{\Psi_g(q) \in \mathbf{Q} \,|\, g \in G\} \in \mathbf{M} = \mathbf{Q}/G$, we consequently have $\varphi_S(\Psi_G(q)) = \mathcal{\Sigma}_{S \times G}(q) = \pi_\Sigma(q)$, or equivalently $\pi_\Sigma = \pi_\varphi \circ \pi$. In Manton's examples[8] $\overline{\mathbf{M}}$ is just Minkowski space and $\mathbf{M}$ is $\overline{\mathbf{M}}$ with a two-sphere, S^2, attached to every point. The φ action of S is the simultaneous rigid rotation of all the two-spheres into themselves.

(b) $K_q \subset S \times G$ as the isotropy group of $q \in \mathbf{Q}$, i.e., $(r, h) \in K_q \Leftrightarrow \mathcal{\Sigma}_{(r,h)} q = q$.

(c) $R_{\pi(q)} \subset S$ as the isotropy group of $\pi(q)$, i.e., $r \in R_{\pi(q)} \Leftrightarrow \varphi_r(\pi(q)) = \pi(q)$.

It then follows that K_q and $R_{\pi(q)}$ are isomorphic [i.e., $(r, h) \leftrightarrow r$]. The proof is short. Let $(r, h) \in K_q \Rightarrow \mathcal{\Sigma}_{(r,h)}(q) = q \Rightarrow \pi \circ \mathcal{\Sigma}_{(r,h)}(q) = \pi(q)$ but the left-hand side simplifies $\pi \circ \mathcal{\Sigma}_{(r,h)} q = \pi \circ \Psi_h \circ \Phi_r(q) = \pi \circ \Phi_r(q) = \varphi_r \circ \pi(q)$ and hence $r \in R_{\pi(q)}$. Conversely if $r \in R_{\pi(q)} \Rightarrow \pi(q) = \varphi_r(\pi(q)) = \pi \circ \Phi_r(q) \Rightarrow q$ and $\Phi_r(q)$ are on the same fiber and hence there exists a unique $\mathfrak{h}(r) \in G$ such that $\Psi_{\mathfrak{h}(r)}(\Phi_r(q)) = q$, i.e., $(r, \mathfrak{h}(r)) \in K_q$.

From this proof we see that K_q defines a homomorphism of $R_{\pi(q)}$ onto a subgroup $H_q \subset G$,

$$r \to \mathfrak{h}(r) \Leftrightarrow (r, \mathfrak{h}(r)) \in K_q.$$

(d) $\mathbf{Q}^K$ as the reduced bundle space of the reduced gauge group before spontaneous symmetry breaking (see Fig. 2). It is the subspace of $\mathbf{Q}$ which has a common isotropy group K,[15]

$$\mathbf{Q}^K = \{q \in \mathbf{Q} \,|\, K_q = K \subseteq S \times G\}. \tag{2.5}$$

What we *assume* here is that K is a Lie subgroup and that $\overline{\mathbf{M}}$ is covered by local sections of π_Σ: $\mathbf{Q} \to \overline{\mathbf{M}} = \mathbf{Q}/S \times G$ contained in $\mathbf{Q}^K$ (see Fig. 2). These are quite restrictive but are tacit assumptions made in FM. The latter amounts to assuming that when π_Σ is restricted to $\mathbf{Q}^K$ (call it $\pi_{\overline{\Sigma}}$) then

$$\pi_{\overline{\Sigma}}\colon\ \mathbf{Q}^K \to \overline{\mathbf{M}}, \tag{2.6}$$

is a locally trivial fiber bundle. If $R \subset S$ is defined as the projection of K into S by $R \equiv \mathrm{pr}_1(K)$ we can define $\mathbf{M}^R \supseteq \pi(\mathbf{Q}^K)$ by

$$\mathbf{M}^R \equiv \{m \in \mathbf{M} \,|\, \varphi_r(m) = m, \quad \forall\, r \in R\}, \tag{2.7}$$

and consequently there is a set of sections of π_φ: $\mathbf{M} \to \overline{\mathbf{M}} = \mathbf{M}/S$ covering $\overline{\mathbf{M}}$ and contained in $\mathbf{M}^R$ (see Fig. 2). When π_Σ and π are restricted to $\mathbf{Q}^K$ we call them $\pi_{\overline{\Sigma}}$ and $\overline{\pi}$, and when π_φ is restricted to $\mathbf{M}^R$ we call it $\pi_{\overline{\varphi}}$. From the above it follows that

$$\pi_{\overline{\Sigma}} = \pi_{\overline{\varphi}} \circ \overline{\pi}, \tag{2.8}$$

and all are projections for locally trivial fiber bundles. At the end of this section we argue that they are in fact principal bundles.

(e) Z_R, Z_H, and Z_K are the centers of R, H, and K, respectively [$H = \mathrm{pr}_2(K) \subset G$]. Here

$$Z_R \equiv \{r \in R \,|\, rr' = r'r, \quad \forall r' \in R\} \tag{2.9}$$

is the subgroup of R which commutes with *every* element of R. Z_H and Z_K are similarly defined and $Z_K = C_K \cap K$, etc.

(f) C_R, C_H, and $C_K = (C_R, C_H)$ are the centralizers of R, H, and K in S, G, and $S \times G$, respectively. Here,

$$C_R \equiv \{c \in S \,|\, cr = rc, \quad \forall r \in R\}, \tag{2.10}$$

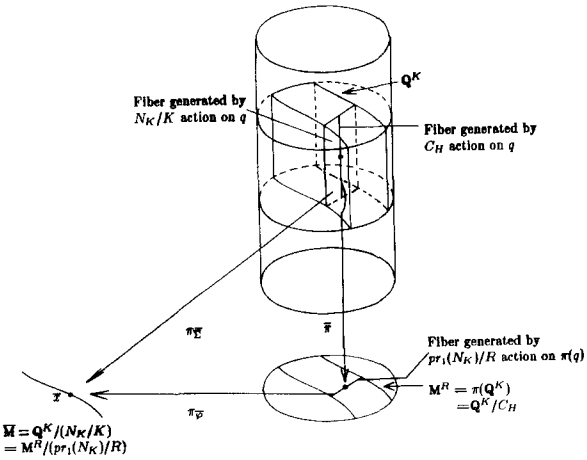


FIG. 2. The reduced symmetric principal bundles $\mathcal{Q}^K = [\mathbf{Q}^K, \overline{\mathbf{M}}, \pi_{\overline{\Sigma}}, N_K/K, \overline{\Sigma}]$, $\mathcal{M}^R = [\mathbf{M}^R, \overline{\mathbf{M}}, \pi_{\overline{\varphi}}, N_R/R, \overline{\varphi}]$, and $\overline{\mathcal{Q}} = [\mathbf{Q}^K, \pi(\mathbf{Q}^K), \overline{\pi}, C_H, \overline{\Psi}]$.

is the subgroup of S which commutes with *every* element of R. C_H and C_K are similarly defined.

(g) N_R, N_H, and N_K are the normalizers of R, H, and K in S, G, and $S\times G$, respectively. Here,

$$N_K\equiv\{n\in S\times G\,|\,nK=Kn\} \tag{2.11}$$

is the largest subgroup of $S\times G$ which contains K as a normal subgroup. Similar definitions are given for N_R and N_H. Notice that N_K also contains C_K as a normal subgroup, and the connected components of the identities are related by $N_K^0=C_K^0\cdot K^0$ (see Appendix).

(h) A_R, A_H, and A_K are the connected normal subgroups of C_R, C_H, and C_K, respectively, generated by the subalgebras of C_R', C_H', and C_K' which are Killing orthogonal to Z_R', Z_H', and Z_K', respectively. Killing orthogonal means orthogonal with respect to the negative definite Killing metrics on S, G, and $S\times G$. For the connected components of the identities, it follows that $C_R^0=A_R\cdot Z_R^0$, $C_H^0=A_H\cdot Z_H^0$, and $C_K^0=A_K\cdot Z_K^0$. Since S and G are compact and Z_R is closed in A_R the intersections $A_R\cap Z_R$, $A_H\cap Z_H$, and $A_K\cap Z_K$ all have trivial Lie algebras.

(i) W_R, W_H, and W_K are the connected normal subgroups of R, H, and K, respectively, generated by the subalgebras of R', H', and K', which are Killing orthogonal to Z_R', Z_H', and Z_K', respectively. Again we have $R^0=W_R\cdot Z_R^0$, $H^0=W_H\cdot Z_H^0$, and $K^0=W_K\cdot Z_K^0$, with finite intersections $W_R\cap Z_R$, $W_H\cap Z_H$, and $W_K\cap Z_K$.

We conclude this section by showing that if

$$\pi_{\bar\Sigma}\colon\ \mathbf{Q}^K\to\bar{\mathbf{M}}=\mathbf{Q}/(S\times G) \tag{2.12}$$

is a locally trivial bundle then it is a principal bundle with group N_K/K. The proof is done in two steps. The first is to show that N_K is the subgroup of $S\times G$ which maps $\mathbf{Q}^K$ onto itself under the action of Σ and second that N_K/K acts freely. First

$$\begin{aligned}&\Sigma_n(q)\in\mathbf{Q}^K,\quad \text{for all } q\in\mathbf{Q}^K\\ &\Leftrightarrow\Sigma_k[\Sigma_n(q)]=\Sigma_n(q),\quad \text{for all } k\in K,\ q\in\mathbf{Q}^K\\ &\Leftrightarrow\Sigma_{n^{-1}}\{\Sigma_k[\Sigma_n(q)]\}=q\\ &\Leftrightarrow\Sigma_{nkn^{-1}}(q)=q,\quad \text{for all } k\in K,\ q\in\mathbf{Q}^K\\ &\Leftrightarrow nKn^{-1}=K\\ &\Leftrightarrow n\in N_K.\end{aligned}$$

Second we start by defining the action $\bar\Sigma$ of N_K/K on $\mathbf{Q}^K$ by

$$\bar\Sigma_{\{nK\}}(q)=\Sigma_{nk}(q)=\Sigma_{nk'}(q). \tag{2.13}$$

This equation shows that $\bar\Sigma$ is a well-defined action of cosets. If it were not free then $q=\Sigma_n(q)=\Sigma_{nk}(q)$ for some $q\in\mathbf{Q}^K$, but then n would be in $K_q(=K)$ and hence a member of the identity coset K. We call this bundle

$$\mathcal{Q}^K\equiv[\mathbf{Q}^K,\bar{\mathbf{M}},\pi_{\bar\Sigma},N_K/K,\bar\Sigma] \tag{2.14}$$

(see Fig. 2). From (g) above and the second isomorphism theorem of groups it follows that $N_K^0/K^0=C_K^0\cdot K^0/K^0\simeq C_K^0/(C_K^0\cap K^0)=C_K^0/Z_K^0=A_K/(A_K\cap Z_K^0)$. Exactly the same arguments as above can be used to show that

$$\pi_{\bar\varphi}\colon\ \mathbf{M}^R\to\bar{\mathbf{M}} \tag{2.15}$$

is a principal bundle with group N_R/R and an induced action $\bar\varphi$ on $\mathbf{M}^R$ defined by

$$\bar\varphi_{\{nR\}}(m)=\varphi_{nr}(m)=\varphi_{nr'}(m). \tag{2.16}$$

We call this principal bundle

$$\mathcal{M}^R=[\mathbf{M}^R,\bar{\mathbf{M}},\pi_{\bar\varphi},N_R/R,\bar\varphi]. \tag{2.17}$$

Using arguments as above (2.15) we see $(N_R^0/R^0)\simeq A_R/(A_R\cap Z_R^0)$, and since the homomorphism $N_R^0/R^0\to(N_R/R)^0$ is onto and has finite kernel, the Lie algebra of N_R/R is just A_R'. The typical fiber for

$$\bar\pi\colon\ \mathbf{Q}^K\to\pi(\mathbf{Q}^K)\subset\mathbf{M}^R \tag{2.18}$$

is found by intersecting (e,G) with N_K, i.e., $(e,G)\cap N_K=(e,C_H)$, and since Ψ acts freely it follows that C_H is the group. We call this bundle

$$\bar{\mathcal{Q}}=[\mathbf{Q}^K,\pi(\mathbf{Q}^K),\bar\pi,C_H,\bar\Psi], \tag{2.19}$$

where $\bar\Psi$ is the restriction of Ψ to the action of C_H on $\mathbf{Q}^K$.

It is clear that

$$\tilde C_H\equiv\{(e,c)K\,|\,c\in C_H\} \tag{2.20}$$

is a normal subgroup of N_K/K and is isomorphic to $C_H\subset G$. It is this subgroup that "carries" the gauge field dynamics in the FM construction.[12] The remaining part of N_K/K is related to the symmetry group S,

$$(N_K/K)/\tilde C_H\simeq\mathrm{pr}_1(N_K)/R, \tag{2.21}$$

where $\mathrm{pr}_1(N_K)$ is the projection of N_K into S and satisfies

$$C_R\cdot R\subseteq\mathrm{pr}_1(N_K)\subset N_R. \tag{2.22}$$

The Lie algebra of $\mathrm{pr}_1(N_K)/R$ is the same as the Lie algebra of N_R/R which, from above, is A_R'.

In Manton's example we don't have to worry about gauge dynamics beyond the C_H group because $N_R/R\simeq Z_2$, the cyclic group of order 2 and $A_R=I\Rightarrow A_R'=0$. The boson part of the electroweak theory is derived by taking $\bar{\mathbf{M}}$ = space-time, $S=\mathrm{SU}_2$ (or SO_3), $R\simeq\mathrm{U}_1$ (or SO_2), and $C_H^0=A_H\cdot Z_H$ with $A_H\simeq\mathrm{SU}_2$ and $Z_H=H\simeq\mathrm{U}_1$. It then follows that if G is simple and connected, it must be one of the rank 2 groups SU_3 (or PU_3), SP_2 (or SO_5), or G_2.

III. A HIGGS FIELD ON $\bar{\mathbf{M}}$ FROM AN S-INVARIANT GAUGE FIELD ON $\mathbf{M}$

In the FM construction, the single dynamical field on extended space-time ($\mathbf{M}$) that ultimately produces two dynamical fields (gauge plus Higgs) on reduced space-time ($\bar{\mathbf{M}}$) is an S-invariant gauge field. Since Trautman's lectures of 1967 (see Ref. 19) it has been known that the mathematical structure behind a gauge field on a space $\mathbf{M}$ corresponding to a gauge group G is a connection ω on a principal bundle, $\mathcal{Q}=[\mathbf{Q},\mathbf{M},\pi,G,\Psi]$. We shall assume this now standard description.

We know that a connection ω on $\mathcal{Q}$ is a one-form on $\mathbf{Q}$ of type "ad" (the adjoint representation) which takes its values in G' (the Lie algebra of G), i.e.,

$$\Psi_g(\omega)=\mathrm{ad}_{g^{-1}}\circ\omega. \tag{3.1}$$

Here, ω also satisfies

$$\omega(\Psi_{\mathring g}(q))=\mathring g, \tag{3.2}$$

where $\mathring g\in G'$ and $\Psi_{\mathring g}$ is the Killing field on $\mathbf{Q}$ generated by the right action of $\Psi_{\exp(\lambda\mathring g)}$ on $\mathbf{Q}$, i.e.,

$$\Psi_{\mathring g}(q)\equiv\frac{d}{d\lambda}\,[\Psi_{\exp(\lambda\mathring g)}(q)]\,|_{\lambda=0}. \tag{3.3}$$

Readers more familiar with the conventional treatment of gauge fields are referred to Sec. IV where ω is expressed in terms of vector potentials $\mathscr{A}_i^a(x)$ by choosing a local cross section σ: $U\subset\mathbf{M}\to\mathbf{Q}$, introducing a coordinate chart $\{x^i\}$ on U, and a basis $\{\mathscr{L}_a\}$ of the Lie algebra G'.

For ω to be called invariant under S (see Refs. 14 and 16) it must simply satisfy

$$\Phi_s(\omega)=\omega \quad \text{for all } s\in S. \tag{3.4}$$

The presence of such an impressed symmetry reduces the dynamical freedom of ω. What starts out as a gauge field on $\mathscr{Q}$ reduces, because of the symmetry constraints (3.4), to a gauge field on $\overline{\mathscr{Q}}$ along with some Higgs fields on $\mathscr{Q}^K$. If N_R/R is finite as in Manton's example, ω completely determines a connection on $\mathscr{Q}^K$; otherwise, ω will only determine a connection on $\mathscr{Q}^K$ if the metric on $\mathbf{M}$ is suitably constrained, e.g., as in FM (see Sec. IV).

The differential equivalent to (3.4) can be stated in terms of a Lie derivative of the connection form ω with respect to the Killing fields $\Phi_{\mathring{s}}$ generated by $\mathring{s}\in S'$ (see Ref. 10),

$$\mathscr{L}_{\Phi_{\mathring{s}}}(\omega)=\Phi_{\mathring{s}}\lrcorner d\omega+d(\Phi_{\mathring{s}}\lrcorner\omega)=0, \tag{3.5}$$

where

$$\Phi_{\mathring{s}}(q)\equiv\frac{d}{d\lambda}\left[\Phi_{\exp(\lambda\mathring{s})}(q)\right]\Big|_{\lambda=0}.$$

Since ω is a one-form of type "ad" and $\Phi_{\mathring{s}}$ is G invariant, $\Phi_{\mathring{s}}\lrcorner\omega=\omega(\Phi_{\mathring{s}}(q))$ is a 0-form (a function) on $\mathscr{Q}$ of type "ad" with values in G'. Its dynamical contents are the Higgs fields which spontaneously break the symmetry of the vacuum. The differential constraints (3.5) are used in Sec. V to identify the kinetic and potential energies of these Higgs fields as part of the F^2 Lagrangian. The Higgs fields $\mathscr{H}$ on $\mathscr{Q}^K$ are the irreducible parts of

$$\mathscr{H}\colon\ \mathbf{Q}^K\to S^*\otimes G',$$

where

$$\mathscr{H}(q)(\mathring{s},\overset{\bullet}{g})\equiv\overset{\bullet}{g}[\omega(\Phi_{\mathring{s}}(q))]. \tag{3.6}$$

Remember that Higgs fields are scalars on $\overline{\mathbf{M}}$ but transform under some representation of the internal gauge group (here it is N_K/K), i.e., they take their values in some vector space on which N_K/K acts. In this case the space is a subspace of the tensor product of S^* (the dual space of S') and G'. Such a vector, $\mathscr{H}(q)$, can be thought of as a bilinear map of $S'\otimes G^*$ into the reals as defined by (3.6). As we move up and down the fibers over $\overline{\mathbf{M}}$ by applying N_K/K via $\overline{\Sigma}$, the values of $\mathscr{H}$ will change according to some representation ρ of the reduced gauge group N_K/K, i.e.,

$$\mathscr{H}(\overline{\Sigma}_{\{nK\}}(q))=\rho^{-1}(\{nK\})\mathscr{H}(q). \tag{3.7}$$

Here ρ is determined by the properties of ω [(3.1), (3.2), (3.4)] as well as the definition of $\mathscr{H}$ (3.6). Also, $S^*\otimes G'$ transforms as $\tilde{\mathrm{ad}}_S\otimes\mathrm{ad}_G$ under $S\times G$ and (3.1) along with (3.4) imply that $\mathscr{H}$ transforms under $N_K\subset S\times G$ by the correspondingly induced action ($\tilde{\mathrm{ad}}_S$ is just the coadjoint action of S on S^*). This is proved by the following steps. Let $n=(n_R,n_H)\in N_K$ then

$$\begin{aligned}
&\mathscr{H}(\Sigma_n(q))(\mathring{s},\overset{\bullet}{g})\\
&=\overset{\bullet}{g}\{\omega[\Phi_{\mathring{s}}(\Sigma_n(q))]\}\quad\text{(by def. of }\mathscr{H})\\
&=\overset{\bullet}{g}\{\omega[\Phi_{\mathring{s}}(\Psi_{n_H}\circ\Phi_{n_R}(q))]\}\\
&=\overset{\bullet}{g}\{\omega[\Psi_{n_H*}\Phi_{\mathring{s}}(\Phi_{n_R}(q))]\}\\
&=\overset{\bullet}{g}\{\Psi_{n_H}\omega[\Phi_{\mathring{s}}(\Phi_{n_R}(q))]\}\\
&=\overset{\bullet}{g}\{\mathrm{ad}_{n_H^{-1}}\circ\omega[\Phi_{\mathring{s}}(\Phi_{n_R}(q))]\}\quad[\text{by (3.1)}]\\
&=(\tilde{\mathrm{ad}}_{n_H}\overset{\bullet}{g})\{\omega[\Phi_{\mathring{s}}(\Phi_{n_R}(q))]\}\\
&=(\tilde{\mathrm{ad}}_{n_H}\overset{\bullet}{g})\{\omega[\Phi_{n_R*}\Phi_{n_R^{-1}*}\Phi_{\mathring{s}}(\Phi_{n_R}(q))]\}\\
&=(\tilde{\mathrm{ad}}_{n_H}\overset{\bullet}{g})\{\omega[\Phi_{n_R*}\Phi_{\mathrm{ad}_{n_R}(\mathring{s})}(q)]\}\\
&=(\tilde{\mathrm{ad}}_{n_H}\overset{\bullet}{g})\{\omega[\Phi_{\mathrm{ad}_{n_R}(\mathring{s})}(q)]\}\quad[\text{by (3.4)}]\\
&=\mathscr{H}(q)(\mathrm{ad}_{n_R}\mathring{s},\tilde{\mathrm{ad}}_{n_H}\overset{\bullet}{g})\quad[\text{by (3.6)}]\\
&=(\tilde{\mathrm{ad}}_{n_R^{-1}}\otimes\mathrm{ad}_{n_H^{-1}})\mathscr{H}(q)(\mathring{s},\overset{\bullet}{g}).
\end{aligned}$$

Since $q\in\mathbf{Q}^K$ is a fixed point of the isotropy group $K\subset N_K$, the above also shows that $\mathscr{H}(q)$ is fixed under the induced action of K, i.e., $\mathscr{H}$ takes its values in the subspace $(S^*\otimes G')^K$ on which K acts as the identity,

$$(S^*\otimes G')^K\equiv\{v\in S^*\otimes G'\,|\,\tilde{\mathrm{ad}}_r\otimes\mathrm{ad}_{\mathfrak{h}(r)}(v)=v,\ \forall(r,\mathfrak{h}(r))\in K\}. \tag{3.8}$$

By decomposing the vector spaces S^* and G' into direct sums of subvector spaces which are invariant under the coadjoint and adjoint actions of N_R and N_H, respectively,

$$S^*=N_R^*\oplus X_R^*=\underbrace{W_R^*\oplus\overbrace{Z_R^*}^{C_R^*}\oplus A_R^*}_{R^*}\oplus X_R^*, \tag{3.9}$$

$$G'=N'_H\oplus X'_H=\underbrace{W'_H\oplus\overbrace{Z'_H}^{C'_H}\oplus A'_H}_{H'}\oplus X'_H, \tag{3.10}$$

we can expand $(S^*\otimes G')^K$ as a direct sum of five subspaces (all invariant under N_K and N_K/K)

$$\begin{aligned}
(S^*\otimes G')^K=&(A_R^*\otimes A'_H)\oplus(A_R^*\otimes Z'_H)\oplus(X_R^*\otimes W'_H)^K\\
&\oplus(X_R^*\otimes X'_H)^K\oplus(R^*\otimes H')^K.
\end{aligned} \tag{3.11}$$

In the above N_R^*, W_R^*, Z_R^*, and A_R^* are the dual spaces to the Lie subalgebras of S' corresponding to the subgroups N_R, W_R, Z_R, and A_R respectively, and X_R^* and X'_H are the vector subspaces of S^* and G' which are "Killing" orthogonal to N_R^* and N'_H, respectively, and are invariant under the

action of N_R and N_H (remember S and G are simple compact Lie groups). Equation (3.2) further restricts $\mathscr{H}$ to a subspace of $(S^* \otimes G')^K$. Since $\Sigma_k(q) = \Psi_{\mathfrak{h}(r)}\Phi_r(q) = q$ it follows that $\Psi_{\mathfrak{h}'(\mathring{h})}(q) + \Phi_{\mathring{r}}(q) = 0$ and (3.2) implies

$$\omega(\Phi_{\mathring{r}}(q)) = -\mathfrak{h}'(\mathring{r}). \tag{3.12}$$

Here $\mathfrak{h}'$: $R' \to H'$ is the Lie algebra homomorphism induced by the Lie group homomorphism $\mathfrak{h}$: $R \to H$ [see Sec. II, Def. (c)]. From the definition of $\mathscr{H}$ in (3.6),

$$\mathscr{H}(q)(\mathring{r}, \mathring{h}) = \mathring{h}\,[\omega(\Phi_{\mathring{r}}(q))] = -\mathring{h}\,(\mathfrak{h}'(\mathring{r})), \tag{3.13}$$

i.e., $\mathscr{H}(q)$'s values in $(R^* \otimes H')^K$ are completely determined by $\mathfrak{h}$: $R \to H$ and consequently the dynamical freedom of the Higgs field is carried by the components of $\mathscr{H}(q)$ in the first four invariant subspaces of (3.11).

We will not go further into the general decomposition of $\mathscr{H}(q)$ into irreducible parts but will tackle each case individually. Even though Manton's first examples[8] do not properly include the fermions of the GWS theory,[9] they are simple and familiar enough to clarify the coordinate free definition (3.6) as well as to illustrate how the global structure excludes some symmetry-group–gauge-group combinations. In these examples we have $S = \mathrm{SU}_2$ (or SO_3), $R = \mathrm{U}_1$ ($\simeq \mathrm{SO}_2$) and consequently $C_R = Z_R = R \Rightarrow C_R/R \simeq I$, $A'_R = (N_R/R)' = 0$, $A^*_R = 0$. We also have $C^0_H = H \cdot A_H$ with $Z_H = H \simeq \mathrm{U}_1$ and $A_H \simeq \mathrm{SU}_2$. Since $W_H = I$, $W'_H = (H/Z_H)' \simeq (I)' = 0$ and consequently only the subspace $(X^*_R \otimes X'_H)^K$ remains as the dynamical range of the Higgs field. The most efficient way to analyze this subspace's transformation properties under actions of the reduced group N_K/K is to choose a special basis for S' and G'. If we pick $\{L_3, L_+, L_-\}$ as a basis of S' (SU'_2 or SO'_3) and the dual basis $\{\theta^3, \theta^-, \theta^+\}$ as a basis for S^* [$\theta^-(L_+) = 1$, etc.], we then have $\{L_3\}$ as a basis of R', $\{\theta^3\}$ as a basis of R^*, $\{L_+, L_-\}$ as a basis of X'_R, and $\{\theta^-, \theta^+\}$ as a basis of X^*_R. Here, R acts on X'_R by

$$\mathrm{ad}_{\exp(\theta L_3)}(L_\pm) = e^{\mp i\theta} L_\pm \Rightarrow [L_3, L_\pm] = \mp i L_\pm, \tag{3.14}$$

and on X^*_R by

$$\tilde{\mathrm{ad}}_{\exp(\theta L_3)}(\theta^\mp) = e^{\pm i\theta}\theta^\mp, \tag{3.15}$$

where $0 \leqslant \theta \leqslant \theta_R$, $\theta_R = 2\pi$ for SO_3, $\theta_R = 4\pi$ for SU_2.

For a basis of G' we pick $\{Y_H, I_3, I_+, I_-, e_\alpha\}$, where $\{Y_H\}$ is a basis of H', $\{I_3, I_+, I_-\}$ a basis of $A'_H \simeq \mathrm{SU}'_2$, and $\{Y_H, I_3\}$ a basis of a Cartan subalgebra. That G' is rank 2 follows from $C'_H = \mathrm{U}'_1 \oplus \mathrm{SU}'_2$ (it contains all group generators commuting with U'_1). The e_α are a set of root vectors beyond I_+ and I_-. There are four, six, and ten of them for $G' = \mathrm{SU}'_3(\mathrm{PU}'_3)$, $\mathrm{Sp}'_2(\mathrm{SO}'_5)$, and G'_2, respectively. All basis vectors are chosen to have equal Killing lengths and scaled so that

$$\mathrm{ad}_{\exp(\phi I_3)}(I_\pm) = e^{\mp i\phi} I_\pm \Rightarrow [I_3, I_\pm] = \mp i I_\pm, \tag{3.16}$$

where $0 \leqslant \phi \leqslant \phi_3$ ($\phi_3 = 2\pi$ or 4π), see Fig. 3 for all root dia-

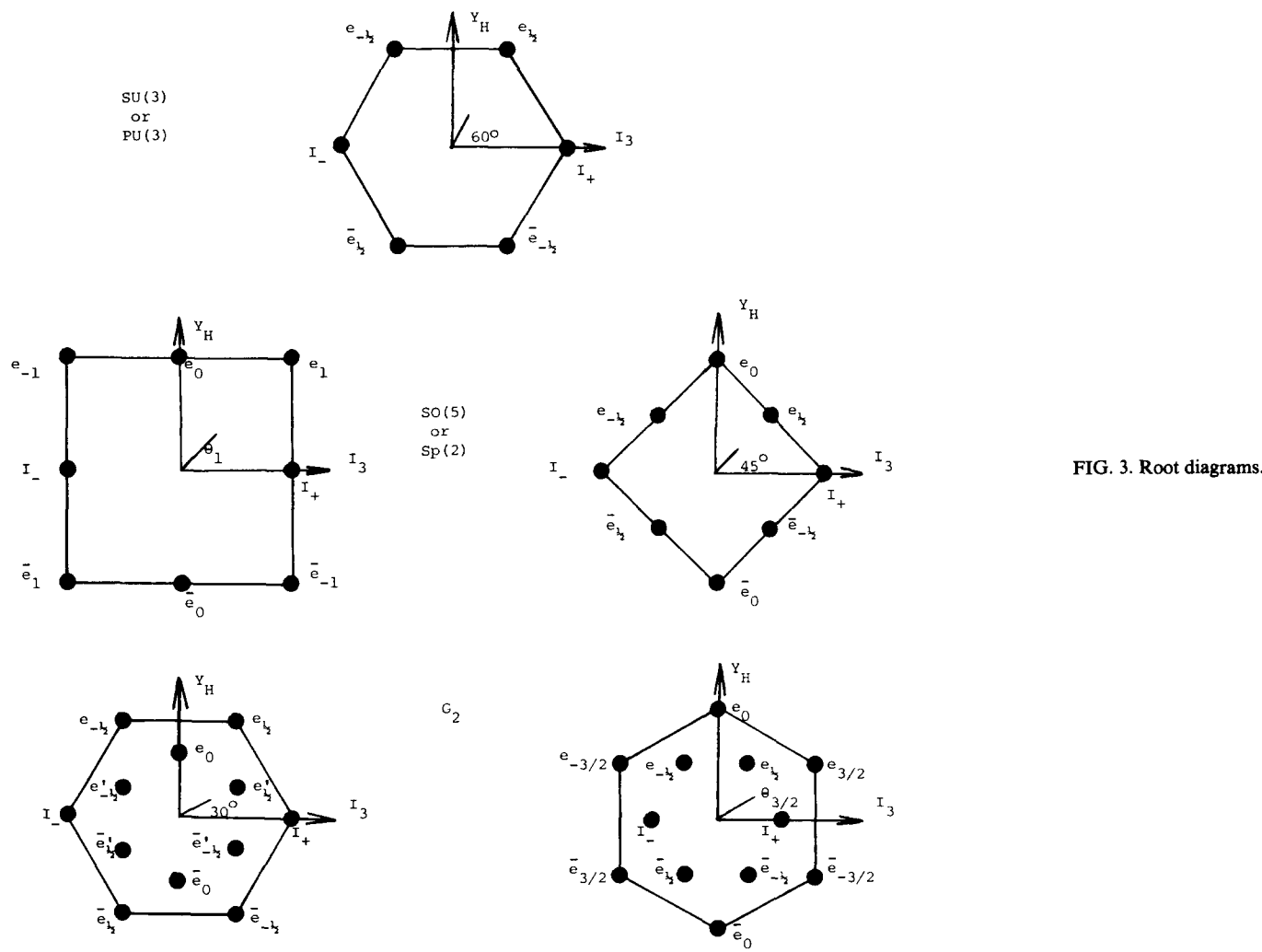


FIG. 3. Root diagrams.

grams and configurations. This scaling makes $\{I_3, I_+, I_-\}$ the usual weak isospin generators but makes Y_H only proportional to the weak hypercharge Y. The homomorphism $\mathfrak{h}$: $R \to H$ is uniquely determined by an integer $n \neq 0$,

$$\exp(\theta L_3) \to \mathfrak{h}(\exp(\theta L_3)) = \exp(\phi Y_H) = \exp[(n\phi_H/\theta_R)\theta Y_H], \quad (3.17)$$

which in turn determines $\mathfrak{h}'$: $R' \to H'$,

$$\mathfrak{h}'(L_3) = (n\phi_H/\theta_R)Y_H. \quad (3.18)$$

Here the domain for θ is $0 \leqslant \theta \leqslant \theta_R$ and for ϕ is $0 \leqslant \phi \leqslant \phi_H$. The ratio ϕ_H/θ_R is fixed by which S and G are used as well as by which SU_2 subgroup of G is picked for A_H. Equations (3.13) and (3.18) fix $\mathscr{H}(q)$'s value in $(R^* \otimes H')^K$ of (3.11) as $-n\phi_H/\theta_R\theta^3 \otimes Y_H$. The hypercharge operator, Y, is the generator of U_1 scaled so that $\exp(2\pi Y) = I$, and consequently

$$Y = (\phi_H/2\pi)Y_H. \quad (3.19)$$

Here U_1 representations will be given by integer values ($y_j = 0, \pm 1, \pm 2, \ldots$) of the hypercharge.

The N_K/K invariant subspace $(X_R^* \otimes X'_H)^K$ is now easy to determine, it is the space spanned by all vectors of the form

$$\{\theta^- \otimes e_m + \theta^+ \otimes \bar{e}_m\} \quad \text{and} \quad i\{\theta^- \otimes e_m - \theta^+ \otimes \bar{e}_m\}, \quad (3.20)$$

where e_m transforms as $\mathrm{ad}_{\mathfrak{h}(r)} e_m = e^{-i\theta} e_m$ to just compensate for $\tilde{\mathrm{ad}}_r(\theta^-) = e^{i\theta}\theta^-$. From (3.17) we have

$$\mathrm{ad}_{\exp \phi Y_H}(e_m) = e^{-i[\theta_R/n\phi_H]\phi} e_m, \quad (3.21)$$

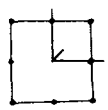
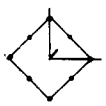
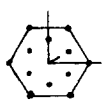
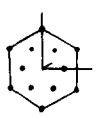
or equivalently

$$[Y_H, e_m] = -i[\theta_R/n\phi_H]e_m, \quad (3.22)$$

i.e., the Y_H components of the root vectors for e_m must match one of the numbers $\theta_R/(n\phi_H)$, where $n = \pm 1, \pm 2, \ldots$; or equivalently the hypercharge y_j as defined by (3.19) for e_m must match one of the values $\theta_R/(2\pi n)$. This is a constraint on G' caused by the group homomorphism $\mathfrak{h}$: $R \to H \subset G$ and excludes several S–G combinations. In particular, the $G = SU_3$ case is excluded altogether (see Table I). Equation (3.22) shows that the permitted e_m's come from a single horizontal (see Fig. 3) string of roots corresponding to a $(2j+1)$-dimensional irreducible representation of the isospin ($\simeq SU_2$) subgroup generated by $\{I_3, I_+, I_-\}$. Consequently the complete Higgs field is

$$\mathscr{H}(q) = \sum_{m=-j}^{m=+j} (\mathscr{H}^m(q)\theta^- \otimes e_m + \text{c.c.}) - \frac{n\phi_H}{\theta_R}\theta^3 \otimes Y_H, \quad (3.23)$$

TABLE I. Parameters associated with various GS combinations.

	Gauge group G	Symmetry group S	$\exp(\theta_R L_3) = I$ θ_R	$\exp(\phi_H Y_H) = I$ ϕ_H	Isospin j	Hypercharge y_j	$\mathfrak{h}$: $R \to H$ rep. integer n $\theta_R/(2\pi y_j)$	Possible case ?
	SU_3	SU_2	4π	$4\pi\sqrt{3}$	$\frac{1}{2}$	3	$\frac{2}{3}$	no
		SO_3	2π				$\frac{1}{3}$	no
	PU_3	SU_2	4π	$4\pi/\sqrt{3}$	$\frac{1}{2}$	1	2	yes
		SO_3	2π				1	yes
	SP_2	SU_2	4π	4π	1	2	1	yes
		SO_3	2π			2	$\frac{1}{2}$	no
	SO_5	SU_2	4π	2π	1	1	2	yes
		SO_3	2π			1	1	yes
	SP_2	SU_2	4π	4π	$\frac{1}{2}$	1	2	yes
					0	2	1	yes
		SO_3	2π		$\frac{1}{2}$	1	1	yes
					0	2	$\frac{1}{2}$	no
	SO_5	SU_2	4π	4π	$\frac{1}{2}$	1	2	yes
					0	2	1	yes
		SO_3	2π		$\frac{1}{2}$	1	1	yes
					0	2	$\frac{1}{2}$	no
	G_2	SU_2	4π	$4\pi\sqrt{3}$	$\frac{1}{2}$	1	2	yes
					0	2	1	yes
					$\frac{1}{2}$	3	$\frac{2}{3}$	no
		SO_3	2π		$\frac{1}{2}$	1	1	yes
					0	2	$\frac{1}{2}$	no
					$\frac{1}{2}$	3	$\frac{1}{3}$	no
	G_2	SU_2	4π	$4\pi/\sqrt{3}$	$\frac{3}{2}$	1	2	yes
					0	2	1	yes
		SO_3	2π		$\frac{3}{2}$	1	1	yes
					0	2	$\frac{1}{2}$	no

where $\mathscr{H}^m(q)$ are complex functions on $\mathbf{Q}^K$ and transform as a $(2j+1)$ representation of isospin and a $y_j = \theta_R/(2\pi n)$ representation of hypercharge Y. This fixes the permitted values of n at ± 1 for SO_3 ($\Rightarrow y_j = \pm 1$) but allows $n = \pm 1, \pm 2$, for SU_2 ($\Rightarrow y_j = \pm 2, \pm 1$). See Table I for the permitted Higgs field and their U_1, SU_2 representations. From this table and Fig. 4 it is also clear that for a given homomorphism $\mathfrak{h}$: $R \to H$, only one set (j, y_j) of Higgs fields appear at a time.

The dynamical components of the symmetric connection ω on $\mathbf{Q}$ beyond those contained in the above mentioned Higgs field on $\mathscr{Q}^K$ are equivalent to a connection $\overline{\omega}$ on $\overline{\mathscr{Q}}$ which is given by pulling ω back to the subspace $\mathbf{Q}^K \subset \mathbf{Q}$. This connection is invariant under the action of $\mathrm{pr}_1(N_K) \subset S$ as given by $\overline{\Phi}$ [Φ restricted to $\mathrm{pr}_1(N_K) \supset C_R$ and $\mathbf{Q}^K$]. We first prove that $\overline{\omega}$ takes its values in C'_H. Let id be the subspace mapping $\mathbf{Q}^K \to \mathbf{Q}$, then $\overline{\omega} = \mathrm{id}\,\omega = \mathrm{id}(\Sigma_k \omega)$ for all $k \in K$. The last equality follows from the definition of $\mathbf{Q}^K$, i.e., $\Sigma_K \circ \mathrm{id} = \mathrm{id}$. But we know

$$\Sigma_k \omega = \Psi_{\mathfrak{h}(r)} \Phi_r \omega = \Psi_{\mathfrak{h}(r)} \omega = \mathrm{ad}_{\mathfrak{h}(r)^{-1}} \circ \omega, \tag{3.24}$$

from (3.1) and (3.4), and consequently

$$\overline{\omega} = \mathrm{ad}_{\mathfrak{h}(r)^{-1}} \circ \overline{\omega} = \mathrm{ad}_{h^{-1}} \circ \overline{\omega}, \tag{3.25}$$

for all $r \in R$ or equivalently for all $h \in H$, i.e., $\overline{\omega}$ takes its values in the subspace of G' which is invariant under the adjoint action of $H \subset G$. This subspace is C'_H, the Lie algebra of C_H [see (f), Sec. II]. The three properties necessary for $\overline{\omega}$ to be a symmetric connection on $\overline{\mathscr{Q}}$,

$$\overline{\Psi}_{ch} \overline{\omega} = \mathrm{ad}_{ch^{-1}} \circ \overline{\omega}, \tag{3.26}$$

$$\overline{\omega}(\overline{\Psi}_{\mathring{c}h}(q)) = \mathring{c}h, \tag{3.27}$$

and

$$\Phi_{cr} \overline{\omega} = \overline{\omega}, \tag{3.28}$$

all follow from (3.1) and (3.4) along with $\mathrm{id} \circ \overline{\Psi}_{ch} = \Psi_{ch} \circ \mathrm{id}$ and $\mathrm{id} \circ \overline{\Phi}_{cr} = \Phi_{cr} \circ \mathrm{id}$.

The restriction of ω to $\mathbf{Q}^K$ in general cannot completely define a connection on $\mathscr{Q}^K$ because of the absence of values in A'_R, or equivalently the horizontal subspaces defined by $\overline{\omega}$ are dimensionally larger than the tangent spaces to $\overline{\mathbf{M}}$ by the dimension of A'_R. However, in the general case it does determine part of the reduced connection ω^K on $\mathscr{Q}^K$. In FM, where there are no dynamic gauge fields associated with $\mathrm{pr}_1(N_K)/R$ [see (2.21)], $\overline{\omega}$ determined "all" of ω^K. In particular, if $C_R = Z_R$ then $A'_R = 0$ and $\omega^K = \overline{\omega}$ is the reduced connection on $\mathscr{Q}^K$. For Manton's special cases, $C'_H = U'_1 \oplus SU'_2$, and the connection ω^K is the gauge field of the intermediate vector bosons of the electroweak theory.

IV. KALUZA–KLEIN SPACES

The Lagrangian density used in FM was the standard $-\frac{1}{4}F^2$ term of the symmetric connection ω and is currently in wide use. Even though the Lagrangian cannot exploit the dynamics of the full reduced gauge group N_K/K we can show its construction requires the presence of a Kaluza–Klein structure thus pointing the way to the Lagrangian which can. Here, F is just the pullback of ω's curvature (Ω) by a cross section (σ, U)

$$F \equiv \sigma \Omega, \tag{4.1}$$

where

$$\Omega \equiv \mathrm{hor}\, d\omega = d\omega + \tfrac{1}{2}[\omega|\omega]. \tag{4.2}$$

Since Ω is a G'-valued two-form of type "ad," construction of the cross section (gauge) independent scalar $-\frac{1}{4}F^2$ requires two, possibly dynamic, fields beyond ω on $\mathscr{Q}$. One is equivalent to a metric $g_{\mathbf{M}}$ on $\mathbf{M}$, and the other is equivalent to a symmetric nondegenerate field g_G of type $\tilde{\mathrm{ad}} \otimes \tilde{\mathrm{ad}}$ on $\mathscr{Q}$ (i.e., g_G has values in the space of inner products on G'). These two fields, along with ω, make $\mathscr{Q}$ a Kaluza–Klein space as can be seen from definitions below. In general, $g_{\mathbf{M}}$ and g_G contain fields whose dynamics are not determined by a Lagrangian density of the form $-\frac{1}{4}F^2$. In the FM construction these fields are externally prescribed, however, by altering the Lagrangian their dynamics could easily be defined.

Kaluza–Klein spaces are of two types, the principal bundle type and the reduced-symmetric principal bundle type[20–23]; we will call them KK and reduced KK, respectively. A KK space $(\mathscr{Q}, g_{\mathbf{Q}})$ is a principal bundle

$$\mathscr{Q} = [\mathbf{Q}, \mathbf{M}, \pi, G, \Psi], \tag{4.3}$$

with a nondegenerate G invariant pseudo-Riemannian metric $g_{\mathbf{Q}}$ on $\mathbf{Q}$. This metric has to induce a nondegenerate metric on the fibers of $\mathscr{Q}$ thus allowing the decomposition of the tangent space $T\mathbf{Q}_q$ at $q \in \mathbf{Q}$ into a direct sum of a vertical subspace V_q ($\equiv$ subspace of $T\mathbf{Q}_q$ tangent to the fibers) and a horizontal subspace H_q ($\equiv$ subspace of $T\mathbf{Q}_q$ orthogonal to V_q by $g_{\mathbf{Q}}$):

$$T\mathbf{Q}_q = V_q \oplus H_q, \tag{4.4}$$

with $V_q \cdot H_q = 0$.

Giving such a G invariant metric $g_{\mathbf{Q}}$ on $\mathbf{Q}$ is equivalent to giving three separate structures:

(1) ω, a connection on $\mathscr{Q}$;
(2) $g_{\mathbf{M}}$, a metric on $\mathbf{M}$; (4.5)

and (3) g_G, a nondegenerate field on $\mathscr{Q}$ of type $\tilde{\mathrm{ad}} \otimes \tilde{\mathrm{ad}}$ with values in $G^* \times G^*$ (i.e., with values in the space of inner products on G'). Now $g_{\mathbf{Q}}$ can be expressed in terms of these three equivalent fields as

$$g_{\mathbf{Q}} = g_G \circ (\omega \otimes \omega) \oplus \pi g_{\mathbf{M}}, \tag{4.6}$$

where g_G and ω can be found from $g_{\mathbf{Q}}$ by choosing a basis $\{\mathscr{L}_\alpha\}$ of G' and from it computing a basis of Killing fields $\{\xi_\alpha\}$ on $\mathbf{Q}$ [by using (3.3)]. Then g_G and ω are given by

$$g_G(q) = g_{\mathbf{Q}}(\xi_\alpha(q), \xi_\beta(q)) \theta^\alpha \otimes \theta^\beta, \tag{4.7}$$

$$\omega = [g_{\mathbf{Q}}(\xi_\alpha(q), \xi_\beta(q))]^{-1} g_{\mathbf{Q}}(\xi_\beta, \_) \mathscr{L}_\alpha, \tag{4.8}$$

where $\{\theta^\alpha\}$ is a basis of G^* dual to $\{\mathscr{L}_\alpha\}$, and $g_{\mathbf{M}}$ is most easily consructed by projecting $g_{\mathbf{Q}}^{-1}$ onto $\mathbf{M}$ (see 4.31).

Various restrictions can be, and are frequently, placed on the range of g_G, e.g., g_G can be restricted to take on only "ad" invariant values, or even further restricted (as in FM) to the fixed Killing form of G. In all but the latter case there is some dynamical freedom in g_G. Because such fields behave as scalars on $\mathbf{M}$, $(\mathscr{Q}, g_{\mathbf{Q}})$ is sometimes called a Jordan–Kaluza–Klein space.[24]

Besides any artificial restriction of g_G as in the FM construction, the underlying KK metric $g_{\mathbf{Q}}$ (or equivalently ω, $g_{\mathbf{M}}$, and g_G) has to be constrained by the group of bundle automorphisms (2.1), (S, Φ), acting as isometries,

$$\Phi_s g_{\mathbf{Q}} = g_{\mathbf{Q}}, \tag{4.9}$$

or (by a not too difficult proof) equivalently,

$$\Phi_s \omega = \omega, \tag{4.10}$$

$$\varphi_s g_{\mathbf{M}} = g_{\mathbf{M}}, \tag{4.11}$$

and

$$\Phi_s g_G = g_G. \tag{4.12}$$

When $g_{\mathbf{Q}}$ is restricted to $\mathbf{Q}^K \subseteq \mathbf{Q}$ [see (2.5) and (2.14)], it induces a metric $g_{\mathbf{Q}^K}$ on $\mathbf{Q}^K$ (which we assume to be nondegenerate). If $g_{\mathbf{Q}^K}$ is also nondegenerate on the fibers of $\mathscr{Q}^K$ it becomes a KK metric for the principal bundle $\mathscr{Q}^K$. The invariance of $g_{\mathbf{Q}^K}$ under the $\bar{\Sigma}$ action of N_K/K follows from the invariance of $g_{\mathbf{Q}}$ under the Σ action of $N_K \subseteq S \times G$. As in (4.5), $g_{\mathbf{Q}^K}$ is equvialent to three fields (ω^K, $g_{\bar{\mathbf{M}}}$, and $g_{N_K/K}$) with no remaining symmetrics such as (4.7),

$$g_{\mathbf{Q}^K} = g_{N_K/K} \circ (\omega^K \otimes \omega^K) \oplus \pi g_{\bar{\mathbf{M}}}. \tag{4.13}$$

The connection ω^K will contain gauge field freedom from $\omega(\bar{\omega})$ which is related to the subgroup $\tilde{C}_H$ of N_K/K and gauge field freedom from $g_{\mathbf{M}}$ which is related to $\mathrm{pr}_1(N_K)/R$ [see (2.21)]. In FM, $g_{\mathbf{M}}$ was constrained to make the gauge field related to $\mathrm{pr}_1(N_K)/R$ integrable and a cross section was chosen so that potentials with values in $(\mathrm{pr}_1(N_K)/R)' \simeq A'_R$ did not appear.

As for the general structure of $g_{\mathbf{M}}$, condition (4.9) (and a nondegeneracy assumption below) makes $\pi_\varphi: \mathbf{M} \to \bar{\mathbf{M}}$ into a *reduced* Kaluza–Klein space. Recall from Sec. II (d) that

$$\pi_\varphi: \mathbf{M} \to \bar{\mathbf{M}} = \mathbf{M}/S, \tag{4.14}$$

has homogeneous fibers (S/R) generated by the non-free φ action of S preventing $\mathbf{M}$ from being the bundle space of a principal bundle and hence ($\mathbf{M}$, $g_{\mathbf{M}}$) from being KK. However, when $\mathfrak{h}$: $R \to H$ of Sec. II, Defs. (c) and (d) are a group isomorphism, $\mathbf{M}$ and its metric $g_{\mathbf{M}}$ can be looked at as the result of factoring (reducing) the larger KK space ($\mathscr{P}$, $g_{\mathbf{Q}}$), where

$$\mathscr{P} = [\mathbf{Q}, \mathbf{Q}/S, \pi_s, S, \Phi], \tag{4.15}$$

by the action of a group of isometric, bundle automorphisms (G, Ψ). What we are now requiring is that the Φ action of S on $\mathbf{Q}$ be free, thus producing another principal fibration $\mathscr{P}$ of $\mathbf{Q}$. The action of G which commutes with the Φ action of S now acts as bundle automorphisms of $\mathscr{P}$. Both bundles $\mathscr{P}$ and $\mathscr{Q}$ have the same bundle space $\mathbf{Q}$ and the same KK metric $g_{\mathbf{Q}}$. $g_{\mathbf{Q}}$ can also be decomposed by using the principal fibration $\mathscr{P}$ of $\mathbf{Q}$ as

$$g_{\mathbf{Q}} = g_s \circ (\omega_s \otimes \omega_s) \oplus \pi_s g_{\mathbf{Q}/S}, \tag{4.16}$$

where ω_s, $g_{\mathbf{Q}/S}$, and g_s are defined analogously to (4.5). When $\mathbf{Q}$ is factored by the Ψ action of G, ($\mathscr{P}$, $g_{\mathbf{Q}}$) is reduced to ($\mathscr{P}/G$, $g_{\mathbf{M}}$) which symbolically stands for the bundle (4.14) and its invariant metric, and which we call a reduced Kaluza–Klein space.

For the simplest example we can take $\mathbf{Q}$ to be the direct product

$$\mathbf{Q} = \bar{\mathbf{M}} \times (K \setminus S \times G), \tag{4.17}$$

where $\bar{\mathbf{M}}$, S, G, and K are as before, and $K \setminus S \times G$ is the space of left cosets. Right multiplication by G and right multiplication by S are separately free when K defines an isomorphism $\mathfrak{h}$: $R \to H$ [Sec. II (c) and (d)]. A KK metric $g_{\mathbf{Q}}$ on $\mathbf{Q}$ for both principal fibrations $\mathscr{Q}$ and $\mathscr{P}$ can be constructed from a KK metric g for the principal bundle

$$\bar{\mathbf{M}} \times (S \times G) \to \bar{\mathbf{M}}, \tag{4.18}$$

if g is constrained to be invariant under the left action of K and nondegenerate on the K fibers. The metric g induces the desired metric $g_{\mathbf{Q}}$ on the factor space $\mathbf{Q}$ in (4.17). This example is not only KK with respect to the two principal fibrations $\mathscr{P}$ and $\mathscr{Q}$ but also reduced KK with respect to the $S \times G$ fibration π_Σ: $\mathbf{Q} \to \bar{\mathbf{M}}$ [see (2.3) and (2.4)]. The left actions of K provide the bundle automorphisms. The KK metric g on $\bar{\mathbf{M}} \times (S \times G)$ is equivalent to (see 4.5) a connection $\omega_{S\times G} = \omega_S + \omega_G$, a metric $g_{\bar{\mathbf{M}}}$ on $\bar{\mathbf{M}}$ and a field $g_{S\times G}$. Here ω_S [not the same as ω_s in (4.16)] must be invariant under left actions of R, ω_G under left actions of H, and $g_{S\times G}$ under left actions of K.

Besides these necessary constraints the FM construction further requires the following.

(1) ω_S must be integrable.

(2) $g_{S\times G}$ at any point in $\bar{\mathbf{M}} \times (S \times G)$ must take on values in the space of inner products on $(S \times G)' = S' \oplus G'$ which when restricted to the G' subspace is the Killing inner product (up to a fixed constant $-G^2$), i.e.,

$$g_{S\times G}(\bar{x}, s, g)(\mathring{g}_1, \mathring{g}_2) = -G^2 K_{G'}(\mathring{g}_1, \mathring{g}_2), \tag{4.19}$$

where

$$K_{G'}(\mathring{g}_1, \mathring{g}_2) \equiv \underset{G}{C}{}^{\alpha}_{\beta\gamma} \underset{G}{C}{}^{\beta}_{\alpha\lambda} \mathring{g}_1^{\gamma} \mathring{g}_2^{\lambda}, \tag{4.20}$$

when $\mathring{g}_1$ and $\mathring{g}_2$ have been expanded in a basis $\{\mathscr{L}_\alpha\}$ of G' in which $[\mathscr{L}_\alpha, \mathscr{L}_\beta] = \underset{G}{C}{}^{\gamma}_{\alpha\beta} \mathscr{L}_\gamma$, i.e.,

$$\mathring{g}_1 = \mathring{g}_1^{\alpha} \mathscr{L}_\alpha \quad \text{and} \quad \mathring{g}_2 = \mathring{g}_2^{\alpha} \mathscr{L}_\alpha.$$

(3) K' and G' must be orthogonal subspaces, i.e.,

$$g_{S\times G}(\bar{x}, s, g)(\mathring{k}, \mathring{g}) = 0, \tag{4.21}$$

for all $\mathring{k} \in K'$ and $\mathring{g} \in G'$.

(4) The Killing inner product $K_{s'}$ on S' defines a subspace $R'_\perp$ orthogonal to R' and at $(\bar{x}, s, g)$ the metric $g_{S\times G}$ defines a subspace $(H' \times G')_\perp$, orthogonal to $H' \times G'$. In the FM construction $R'_\perp$ projects isometrically [up to a scalar $S^2(\bar{x})$] onto $(H' \times G')_\perp$, i.e.,

$$\begin{aligned} &g_{S\times G}(\mathring{s}_1 + \mathring{k}_1 + \mathring{g}_1, \mathring{s}_2 + \mathring{k}_2 + \mathring{g}_2) \\ &\quad = S^2(\bar{x}) K_{S'}(\mathring{s}_1 + \mathring{r}_1, \mathring{s}_2 + \mathring{r}_2), \end{aligned} \tag{4.22}$$

where

$$g_{S\times G}(\mathring{s}_i + \mathring{k}_i + \mathring{g}_i, \mathring{k}) = g_{S\times G}(\mathring{s}_i + \mathring{k}_i + \mathring{g}_i, \mathring{g}) = 0,$$

for all $\mathring{k} \in K'$, $\mathring{g} \in G'$, and $K_{S'}(\mathring{s}_i + \mathring{r}_i, \mathring{r}) = 0$, for all $\mathring{r} \in R'$.

The K invariant KK metric g on $\bar{\mathbf{M}} \times (S \times G)$ that projects by π_K: $\bar{\mathbf{M}} \times (S \times G) \to \bar{\mathbf{M}} \times K \setminus (S \times G)$ to give the FM metric $g_{\mathbf{Q}}$ on $\mathbf{Q}$ (see 4.6) is not unique, however, the simplest such g is

$$\begin{aligned} g = {} & G^2 \delta_{\alpha\beta} \varphi^{\alpha}(g, \bar{x}) \otimes \varphi^{\beta}(g, \bar{x}) \\ & + (-S^2(\bar{x}) \delta_{\Sigma\Lambda} + G^2 \delta_{\alpha\beta} \mathscr{H}^{\alpha}_{\Sigma}(\bar{x}) \mathscr{H}^{\beta}_{\Lambda}(\bar{x})) \\ & \times \mathrm{ad}^{\Sigma}_{s\Pi} \mathrm{ad}^{\Lambda}_{s\Omega} \Phi^{\Pi}(s) \otimes \Phi^{\Omega}(s) \\ & + G^2 \delta_{\alpha\beta} \mathscr{H}^{\beta}_{\Lambda}(\bar{x}) \mathrm{ad}^{\alpha}_{g^{-1}\gamma} \mathrm{ad}^{\Lambda}_{s\Omega} \\ & \times (\varphi^{\gamma}(g, \bar{x}) \otimes \Phi^{\Omega}(s) + \Phi^{\Omega}(s) \otimes \varphi^{\gamma}(g, \bar{x})) \\ & + g_{ij}(\bar{x}) d\bar{x}^i \otimes d\bar{x}^j, \end{aligned} \tag{4.23}$$

where $\varphi^{\alpha}(g,\bar{x})\equiv\Phi^{\alpha}(g)+\mathrm{ad}^{\alpha}_{g^{-1}\beta}\mathscr{A}^{\beta}_{i}(\bar{x})d\bar{x}^{i}$.

In (4.23) $\pi_{S\times G}g_{\bar{\mathbf{M}}}=g_{ij}(\bar{x})d\bar{x}^{i}\otimes d\bar{x}^{j}$ is the lift of the metric $g_{\bar{\mathbf{M}}}$ on $\bar{\mathbf{M}}$ (space-time), written in terms of coordinates $\{\bar{x}^{i}\}$ on $\bar{\mathbf{M}}$. $\{\Phi^{\Sigma}(s)\}$ is a basis of left invariant one-forms on S, corresponding to an ON (orthonormal) basis $\{\mathscr{L}_{\Sigma}\}$ of S'. Here ON is with respect to the negative definite Killing inner product on S', and $\{\Phi^{\alpha}(g)\}$ is similarly defined for the simple compact group G; and $\mathrm{ad}^{\Sigma}_{sII}$ and $\mathrm{ad}^{\alpha}_{g\gamma}$ are the adjoint representations of S and G, respectively, written in terms of the bases $\{\mathscr{L}_{\Sigma}\}$ and $\{\mathscr{L}_{\alpha}\}$.

In (4.23) the connection $\omega_{S\times G}=\omega_{S}+\omega_{G}$ is given by

$$\omega_{S}=\Phi^{\Sigma}(s)\mathscr{L}_{\Sigma}, \tag{4.24}$$

$$\omega_{G}=\varphi^{\alpha}(g,\bar{x})\mathscr{L}_{\alpha}, \tag{4.25}$$

i.e., ω_{S} is integrable but ω_{G} is not. Because g is constrained to be invariant under the left actions of K on $\mathbf{Q}^{K}$, $\mathrm{ad}^{\alpha}_{g^{-1}\beta}\mathscr{A}^{\beta}_{i}(\bar{x})d\bar{x}^{i}\,\mathscr{L}_{\alpha}$ must take its values in C'_{H}, the Lie algebra of the centralizer of H (similar to 4.9→4.10). The $\mathscr{A}^{\beta}_{i}(\bar{x})$ are just the nonvanishing potentials for N_{K}/K, the reduced gauge group on $\mathscr{Q}^{K}$, in a special gauge (see 4.34).

In (4.23) the functions $\mathscr{H}^{\alpha}_{\Sigma}(\bar{x})$ on $\bar{\mathbf{M}}$ contained in $g_{S\times G}$ are nothing more than the Higgs fields of (3.6), in fact

$$\mathscr{H}(q)=\mathrm{ad}^{\beta}_{g^{-1}\alpha}\mathscr{H}^{\alpha}_{\Sigma}(\bar{x})\mathrm{ad}^{\Sigma}_{s\Omega}\theta^{\Omega}\otimes\mathscr{L}_{\beta}, \tag{4.26}$$

where $\{\theta^{\Omega}\}$ is a basis of S^{*} dual to $\{\mathscr{L}_{\Omega}\}$. The K invariance of g implies the K invariance of $g_{S\times G}$ (similar to 4.9→4.12) which produces the K invariance of $\mathscr{H}$ and the decomposition (3.11). Now g of (4.23) can be rewritten in a simpler-looking form as

$$g=G^{2}\delta_{\alpha\beta}\overline{\varphi}^{\alpha}(s,g,\bar{x})\otimes\overline{\varphi}^{\beta}(s,g,\bar{x})-S^{2}(\bar{x})\delta_{\Sigma\Lambda}\Phi^{\Sigma}(s)\otimes\Phi^{\Lambda}(s)+g_{ij}(\bar{x})d\bar{x}^{i}\otimes d\bar{x}^{j}, \tag{4.27}$$

where

$$\overline{\varphi}^{\alpha}(s,g,\bar{x})=\varphi^{\alpha}(g,\bar{x})+\mathrm{ad}^{\alpha}_{g^{-1}\gamma}\mathscr{H}^{\gamma}_{\Sigma}(\bar{x})\mathrm{ad}^{\Sigma}_{s\Omega}\Phi^{\Omega}(s).$$

Here $g_{\mathbf{Q}}$ is more easily constructed from g in this form because $\overline{\varphi}^{\alpha}(\mathring{k})=d\bar{x}^{i}(\mathring{k})=0$, for all $\mathring{k}\in K'$, which follows from (3.13) with $\mathring{k}=\mathring{r}+\mathfrak{h}'(\mathring{r})$ or equivalently from (4.21). By writing

$$g=g_{K}\circ(\omega_{K}\otimes\omega_{K})\oplus\pi_{K}g_{\mathbf{Q}}, \tag{4.28}$$

we see that the part of g not determined by $g_{\mathbf{Q}}$ is, in our simple example,

$$g_{K}\circ(\omega_{K}\otimes\omega_{K})=-S^{2}(\bar{x})\delta_{ab}\mathrm{ad}^{a}_{s\Sigma}\mathrm{ad}^{b}_{s\Lambda}\Phi^{\Sigma}(s)\otimes\Phi^{\Lambda}(s), \tag{4.29}$$

where $\{\mathscr{L}_{a}\}\subset\{\mathscr{L}_{\Sigma}\}$ is an ON basis of R'. Note that this ω_{K} is not the same as ω^{K} in (4.13). Consequently,

$$\pi_{K}g_{\mathbf{Q}}=G^{2}\delta_{\alpha\beta}\overline{\varphi}^{\alpha}(s,g,\bar{x})\otimes\overline{\varphi}^{\beta}(s,g,\bar{x})-S^{2}(\bar{x})[\delta_{\Sigma\Lambda}-\delta_{ab}\delta^{a}_{\Sigma}\delta^{b}_{\Lambda}]\mathrm{ad}^{\Sigma}_{sII}\mathrm{ad}^{\Lambda}_{s\Omega}\times\Phi^{II}(s)\otimes\Phi^{\Omega}(s)+g_{ij}(\bar{x})d\bar{x}^{i}\otimes d\bar{x}^{j}. \tag{4.30}$$

The metric $g_{\mathbf{Q}}$ on $\mathbf{Q}$ is most easily written as $g_{\mathbf{Q}}^{-1}$, it is

$$g_{\mathbf{Q}}^{-1}=g^{ij}(\bar{x})h_{i}\otimes h_{j}-S^{-2}(\bar{x})\delta^{\Sigma\Omega}h_{\Sigma}\otimes h_{\Omega}+G^{-2}\delta^{\alpha\beta}\xi_{\alpha}\otimes\xi_{\beta}, \tag{4.31}$$

where the horizontal vectors h_{i} and h_{Σ} are defined by

$$h_{i}\equiv\frac{\partial}{\partial\bar{x}^{i}}-\mathrm{ad}^{\alpha}_{g^{-1}\beta}\mathscr{A}^{\beta}_{i}(\bar{x})\xi_{\alpha},$$

$$h_{\Sigma}\equiv\xi_{\Sigma}-\mathrm{ad}^{\alpha}_{g^{-1}\beta}\mathscr{H}^{\beta}_{II}(\bar{x})\mathrm{ad}^{II}_{s\Sigma}\xi_{\alpha}. \tag{4.32}$$

Here, ξ_{Σ} and ξ_{α} are just the π_{K} projections onto $\mathbf{Q}$ of the Killing fields l_{Σ} and l_{α} on $\bar{\mathbf{M}}\times(S\times G)$ generated by $\mathscr{L}_{\Sigma}$ and $\mathscr{L}_{\alpha}$, respectively, or in other words the Killing fields on $\mathbf{Q}$ generated respectively by $\mathscr{L}_{\Sigma}$ and $\mathscr{L}_{\alpha}$ [see (3.3) and (3.5)]. The h_{Σ}'s are not all linearly independent. Here ω is not easily written down, but its pullback $\pi_{K}\omega$ from $\mathbf{Q}$ to $\bar{\mathbf{M}}\times(S\times G)$ is just

$$\pi_{K}\omega=\overline{\varphi}^{\alpha}(s,g,\bar{x})\mathscr{L}_{\alpha}, \tag{4.33}$$

where $\overline{\varphi}^{\alpha}$ is given in (4.27). Likewise, the pullback of the reduced connection on $\mathscr{Q}^{K}$, ω^{K}, can be written as

$$\pi_{K}\omega^{K}=\Phi^{\Sigma'}(s)\mathscr{L}_{\Sigma'}+\varphi^{\alpha'}(g,\bar{x})\mathscr{L}_{\alpha'},$$
$$=\Phi^{\Sigma'}(s)\mathscr{L}_{\Sigma'}+\Phi^{\alpha'}(g)\mathscr{L}_{\alpha'}+\mathrm{ad}^{\alpha'}_{g^{-1}\beta}\mathscr{A}^{\beta}_{i}(\bar{x})d\bar{x}^{i}\mathscr{L}_{\alpha'}, \tag{4.34}$$

where $\{\mathscr{L}_{\Sigma'}\}\subset\{\mathscr{L}_{\Sigma}\}$ and $\{\mathscr{L}_{\alpha'}\}\subset\{\mathscr{L}_{\alpha}\}$ are ON bases of A'_{R} and C'_{H}, respectively.

V. THE F^{2} ACTION

In the FM construction the standard gauge field action

$$I_{\mathbf{M}}=\int_{\mathbf{M}}-\frac{1}{4}F^{2}\,dV_{\mathbf{M}} \tag{5.1}$$

was used to give dynamics to the reduced gauge and Higgs fields. In general (5.1) originates as an action integral on $\mathbf{Q}$,

$$I_{\mathbf{Q}}=\int_{\mathbf{Q}}-\frac{1}{4}g_{G}(\Omega\wedge{}^{*}\Omega), \tag{5.2}$$

where ${}^{*}\Omega$ is the dual of Ω computed using the KK metric $g_{\mathbf{Q}}$.[25] The form $\Omega\wedge{}^{*}\Omega$ is proportional to $dV_{\mathbf{Q}}$,[26] has values in $G'\times G'$, and transforms as a field type $\mathrm{ad}\otimes\mathrm{ad}$, i.e.,

$$\Omega\wedge{}^{*}\Omega=(\Omega,\Omega)_{g_{\mathbf{Q}}}\,dV_{\mathbf{Q}}, \tag{5.3}$$

where $(\Omega,\Omega)_{g_{\mathbf{Q}}}$ stands for the inner product of two $p(=2)$ forms induced by the KK metric $g_{\mathbf{Q}}$, and $dV_{\mathbf{Q}}$ is the corresponding volume form. When g_{G}, a field of type $\tilde{\mathrm{ad}}\otimes\tilde{\mathrm{ad}}$, is applied to $\Omega\wedge{}^{*}\Omega$ as in (5.2) the integrand becomes a form proportional to $dV_{\mathbf{Q}}$ which is of type $\rho=1$, i.e., invariant under the Ψ action of G. Equation (5.2) produces (5.1) when evaluated using a local cross section (σ,U) to "coordinatize" $\mathbf{Q}$:

$$I_{\mathbf{Q}}=\int_{\mathbf{M}}-\frac{1}{4}V_{G}\sigma g_{G}(F,F)_{g_{\mathbf{M}}}\,dV_{\mathbf{M}}, \tag{5.4}$$

where

$$V_{G}(m)\equiv\int dV_{G} \tag{5.5}$$

is the volume of the fiber over $m\in\mathbf{M}$ as measured by the vertical part of $g_{\mathbf{Q}}$, see (4.6), and is naturally cross-section-independent. Here $dV_{\mathbf{M}}$ is the volume form on $\mathbf{M}$ defined by $g_{\mathbf{M}}$ and also is defined independently of a cross section. The other term, $\sigma g_{G}(F,F)_{g_{\mathbf{M}}}$, turns out to be cross-section-independent because of the $\rho=1$ nature of (5.2). The action integral (5.2) on the KK space $(\mathscr{Q},g_{\mathbf{Q}})$ can be reduced to an action integral on the KK space $(\mathscr{Q}^{K},g_{\mathbf{Q}^{K}})$. For a general S and G invariant scalar function $\mathscr{L}_{\mathbf{Q}}$ on $\mathbf{Q}$,

$$\int_{\mathbf{Q}}\mathscr{L}_{\mathbf{Q}}\,dV_{\mathbf{Q}}=\int_{\mathbf{Q}^{K}}\left(\frac{V_{G}\circ\bar{\pi}}{V_{N_{K}/K}\circ\pi_{\bar{\Sigma}}}\right)\mathscr{L}_{\mathbf{Q}}\,dV_{\mathbf{Q}^{K}}, \tag{5.6}$$

where V_G was defined in (5.5) and $V_{N_K/K}$ is defined as the volume of the fibers of $\mathcal{Q}^K$ over $\bar{\mathbf{M}}$ as measured by the vertical part of $g_{\mathbf{Q}^K}$. The action (5.2) was constructed from various parts of the KK metric $g_{\mathbf{Q}}$ on $\mathcal{Q}$ (4.5), however, the reduced action on $\mathcal{Q}^K$ contains not only parts of $g_{\mathbf{Q}^K}$, but also parts of $g_{\mathbf{Q}}$ not contained in $g_{\mathbf{Q}^K}$, e.g., the Higgs field of Sec. III which comes from the ω part of $g_{\mathbf{Q}}$. When (5.6) is applied to (5.2), using (5.3) we get

$$I_{\mathbf{Q}} = \int_{\mathbf{Q}^K} \left(\frac{V_G \circ \bar{\pi}}{V_{N_K/K} \circ \pi_{\bar{\Sigma}}} \right) \left\{ -\frac{1}{4} g_G(\Omega, \Omega)_{g_{\mathbf{Q}}} \right\} dV_{\mathbf{Q}^K}, \quad (5.7)$$

and when a cross section σ^K: $\bar{\mathbf{M}} \to \mathbf{Q}^K$ is used to evaluate (5.7) we get

$$I_{\mathbf{Q}} = \int_{\bar{\mathbf{M}}} (V_G \circ \sigma^R) \left\{ -\frac{1}{4} g_G(\Omega, \Omega)_{g_{\mathbf{Q}}} \circ \sigma^K \right\} dV_{\bar{M}}, \quad (5.8)$$

where σ^R: $\bar{\mathbf{M}} \to \mathbf{M}$ is the cross section defined by $\sigma^R \equiv \bar{\pi} \circ \sigma^K$. When this is evaluated we arrive at equation (5.7) of FM (see Ref. 7). $g_{\mathbf{Q}}^{-1}$ as given in (4.31) can be used to decompose $-\frac{1}{4} g_G(\Omega, \Omega)_{g_{\mathbf{Q}}}$ of (5.7) and (5.8) into the sum of three terms,

$$\begin{aligned} &-\tfrac{1}{4} g_G(\Omega, \Omega)_{g_{\mathbf{Q}}} \\ &\quad = -\tfrac{1}{4} g_{N_K/K}(\Omega^K, \Omega^K)_{g_{\mathbf{Q}^K}} \\ &\qquad + \tfrac{1}{2} G^2 S^{-2}(\bar{x}) K_{S'}^{-1} \otimes K_{G'} (D^K \mathscr{H}, D^K \mathscr{H})_{g_{\mathbf{Q}^K}} \\ &\qquad - V(\mathscr{H}), \end{aligned} \quad (5.9)$$

where the Higgs potential $V(\mathscr{H})$ is given by

$$\begin{aligned} V(\mathscr{H}) &= \tfrac{1}{4} S^{-4}(\bar{x}) g_G(\Omega(h_\Sigma, h_\Pi), \Omega(h_\Omega, h_\Lambda)) \delta^{\Sigma\Omega} \delta^{\Pi\Lambda} \\ &= \tfrac{1}{4} S^{-4}(\bar{x}) g_G(\Omega(\xi_\Sigma, \xi_\Pi), \Omega(\xi_\Omega, \xi_\Lambda)) \delta^{\Sigma\Omega} \delta^{\Pi\Lambda} \\ &= \tfrac{1}{4} S^{-4}(\bar{x}) g_G(\omega([\xi_\Sigma, \xi_\Pi]) + [\omega(\xi_\Sigma), \omega(\xi_\Pi)]_G, \\ &\quad \omega([\xi_\Omega, \xi_\Lambda]) + [\omega(\xi_\Omega), \omega(\xi_\Lambda)]_G) \delta^{\Sigma\Omega} \delta^{\Pi\Lambda}. \end{aligned} \quad (5.10)$$

The first step in (5.10), $\Omega(h_\Sigma, h_\Pi) = \Omega(\xi_\Sigma, \xi_\Pi)$, follows from (4.2) and the vertical nature of ξ_α [ξ_α is the Killing field on $\mathbf{Q}$ generated by $\mathscr{L}_\alpha \in G'$ as in (3.3)]. The next step in (5.10) required (4.2) to write

$$\Omega(\xi_\Sigma, \xi_\Pi) = d\omega(\xi_\Sigma, \xi_\Pi) + [\omega(\xi_\Sigma), \omega(\xi_\Pi)]_G, \quad (5.11)$$

and the identity $d\omega(u, v) = u(\omega(v)) - v(\omega(u)) - \omega[u, v]$ along with (3.5) to write

$$d\omega(\xi_\Sigma, \xi_\Pi) = \omega[\xi_\Sigma, \xi_\Pi]. \quad (5.12)$$

Since $\pi_K l_\Omega = \xi_\Omega$ we can evaluate $\omega(\xi_\Omega)$ by applying its pullback $\pi_K \omega$ [of (4.34)] to l_Ω,

$$\omega(\xi_\Omega) = \pi_K \omega(l_\Omega) = \mathrm{ad}^\alpha_{g^{-1}\gamma} \mathscr{H}^\gamma_\Sigma(\bar{x}) \mathrm{ad}^\Sigma_{s\Omega} \mathscr{L}_\alpha, \quad (5.13)$$

from which we can express the Higgs potential as

$$\begin{aligned} V(\mathscr{H}) = \tfrac{1}{4} G^2 S^{-4}(\bar{x}) \delta_{\alpha\beta} &\left\{ \mathscr{H}^\alpha_\Xi(\bar{x}) C^{\Xi}_{S\,\Sigma\Pi} + C^{\alpha}_{G\,\gamma\delta} \mathscr{H}^\gamma_\Sigma(\bar{x}) \mathscr{H}^\delta_\Pi(\bar{x}) \right\} \\ &\times \left\{ \mathscr{H}^\beta_\Delta(\bar{x}) C^{\Delta}_{S\,\Omega\Lambda} + C^{\beta}_{G\,\sigma\lambda} \mathscr{H}^\sigma_\Omega(\bar{x}) \mathscr{H}^\lambda_\Lambda(\bar{x}) \right\} \delta^{\Sigma\Omega} \delta^{\Pi\Lambda}. \end{aligned} \quad (5.14)$$

In this example V_G is constant and we treat (5.9) as the Lagrangian density. The first term in (5.9) is the usual $-\frac{1}{4}F^2$ type term for the free gauge field ω^K, corresponding to the reduced gauge group N_K/K. However, when (4.31) is used in (5.7) this term starts as

$$-\tfrac{1}{4} g^{ij} \Omega(h_i, h_k) \Omega(h_j, h_l) g^{kl}, \quad (5.15)$$

but $\Omega(h_i, h_j) = d\omega(h_i, h_j) = -\omega[h_i, h_j] = -\omega^K[h_i, h_j] = d\omega^K(h_i, h_j) = \Omega^K(h_i, h_j)$. All steps follow because ω_S of (4.24) is integrable or equivalently because the h_i are horizontal in both bundles $\mathcal{Q}$ and $\mathcal{Q}^K$ [i.e., $\omega(h_i) = \omega^K(h_i) = 0$]. Since the h_i contains no ξ_Σ's then it contains no Killing fields generated by $A'_R \subset S'$ [see (4.32)], and we are using a gauge adopted to the partial integrability of ω^K. Applying $g_{N_K/K}(q)$ for $q \in \mathbf{Q}^K$ to (5.14) is exactly the same as applying $g_G(q)$ because (1) $\Omega^K(h_i, h_j)$ takes on values only in $C'_H \subset (N_K/K)' \cap G'$, and (2) both metrics are defined by $g_{\mathbf{Q}}$, i.e., they agree when applied to common directions [see (4.7)]. When (5.15) is evaluated using the vector potentials of (4.32) the first term in (5.9) becomes

$$\begin{aligned} &-\tfrac{1}{4} g_{N_K/K}(\Omega^K, \Omega^K)_{g_{\mathbf{Q}^K}} \\ &\quad = -\tfrac{1}{4} G^2 \delta_{\alpha\beta} g^{ij}(\bar{x}) F^\alpha_{ik}(\bar{x}) F^\beta_{jl}(\bar{x}) g^{kl}(\bar{x}), \end{aligned} \quad (5.16)$$

where

$$F^\alpha_{ik}(\bar{x}) \equiv \frac{\partial}{\partial \bar{x}^i} \mathscr{A}^\alpha_k(\bar{x}) - \frac{\partial}{\partial \bar{x}^k} \mathscr{A}^\alpha_i(\bar{x}) + \mathscr{A}^\gamma_i(\bar{x}) \mathscr{A}^\delta_j(\bar{x}) C^{\alpha}_{G\,\gamma\delta}.$$

The second term in (5.9) is the kinetic energy of the Higgs field. The one-form on $\mathbf{Q}^K$, $D^K\mathscr{H}$, is the ω^K gauge covariant derivative of $\mathscr{H}$, recall from Sec. III that $\mathscr{H}$ transforms under a (in general reducible) representation of N_K/K. Here, $(D^K\mathscr{H}, D^K\mathscr{H})_{g_{\mathbf{Q}^K}}$ is the usual inner product of one-forms and $K_{S'}^{-1} \otimes K_{G'}$ is an invariant field on $\mathcal{Q}^K$ constructed from the Killing inner products on S' and G', respectively, with values in the space of inner products on $S^* \times G'$. When (4.31) is used in (5.7) this term starts as

$$\begin{aligned} &\tfrac{1}{2} G^2 S^{-2}(\bar{x}) K_{S'}^{-1} \otimes K_{G'} (D^K\mathscr{H}, D^K\mathscr{H})_{g_{\mathbf{Q}^K}} \\ &\quad = \tfrac{1}{2} S^{-2}(\bar{x}) g^{ij} g_G(\Omega(h_i, h_\Pi), \Omega(h_j, h_\Lambda)) \delta^{\Pi\Lambda}, \end{aligned} \quad (5.17)$$

however, $\Omega(\_, \xi_\alpha) = 0 \Rightarrow \Omega(h_j, h_\Lambda) = \Omega(h_j, \xi_\Lambda)$ and $\omega(h_j) = 0 \Rightarrow \Omega(h_j, \xi_\Lambda) = d\omega(h_j, \xi_\Lambda)$. From (3.5) $d\omega(h_j, \xi_\Lambda) = (-\xi_\Lambda \lrcorner d\omega)(h_j) = d(\omega(\xi_\Lambda))(h_j) = h_j(\omega(\xi_\Lambda))$. Using (5.13) and recalling the partial integrability of ω^K we get $h_j(\omega(\xi_\Lambda)) = D^K_j(\mathscr{H})(\mathscr{L}_\Lambda)$, where

$$D^K_j(\mathscr{H}) = \mathrm{ad}^\alpha_{g^{-1}\beta}(\mathscr{D}_j \mathscr{H}^\beta_\Sigma) \mathrm{ad}^\Sigma_{s\Lambda} \theta^\Lambda \otimes \mathscr{L}_\alpha, \quad (5.18)$$

with

$$\mathscr{D}_j \mathscr{H}^\beta_\Sigma \equiv \frac{\partial}{\partial \bar{x}^j} \mathscr{H}^\beta_\Sigma(\bar{x}) + C^{\beta}_{G\,\gamma\delta} \mathscr{A}^\gamma_j(\bar{x}) \mathscr{H}^\delta_\Sigma(\bar{x}).$$

Putting this back into (5.17) we have the kinetic energy of the Higgs field with minimal coupling to the reduced gauge group [the second term in (5.9)],

$$\begin{aligned} &\tfrac{1}{2} G^2 S^{-2}(\bar{x}) K_{S'}^{-1} \otimes K_{G'} (D^K\mathscr{H}, D^K\mathscr{H})_{g_{\mathbf{Q}^K}} \\ &\quad = \tfrac{1}{2} G^2 S^{-2}(\bar{x}) g^{ij}(\bar{x}) \delta^{\Pi\Lambda} \delta_{\alpha\beta} (\mathscr{D}_i \mathscr{H}^\alpha_\Pi)(\mathscr{D}_j \mathscr{H}^\beta_\Lambda). \end{aligned} \quad (5.19)$$

For completeness we evaluate (5.9) for the electroweak examples of Sec. III. Choosing a cross section σ^K: $\bar{\mathbf{M}} \to \mathbf{Q}^K$, which is induced by the canonical cross section of $\bar{\mathbf{M}} \to \bar{\mathbf{M}} \times S \times G$, we can write the reduced gauge field as

$$\begin{aligned} \mathscr{A}^K &\equiv \sigma^K \omega^K = d\bar{x}^i \mathscr{A}^{\alpha'}_i(\bar{x}) \mathscr{L}_{\alpha'} \\ &= d\bar{x}^i \{ gB_i(\bar{x}) Y_H + gA^3_i(\bar{x}) I_3 \\ &\quad + gW^+_i(\bar{x}) I_- + gW^-_i(\bar{x}) I_+ \}. \end{aligned} \quad (5.20)$$

The basis of G' used is given in the paragraph containing (3.16) and is only ON up to a constant $k_{G'}^{-2}$, see (4.20),

$$-k_{G'}^{-2}\equiv K_{G'}(Y_H, Y_H) = K_{G'}(I_3, I_3) = K_{G'}(I_+, I_-) = K_{G'}(e_\alpha, \bar{e}_\alpha). \tag{5.21}$$

The constants g, g', μ, and λ are the usual constants of the GWS theory and are introduced at the appropriate places to conform to Abers and Lee.[27] Only two independent parameters, G and S, are present and we choose to rewrite them as functions of g and μ (g' and λ are determined by this theory), i.e., we choose

$$G = k_{G'}/g,\ S^{-1} = \sqrt{-2\mu^2}, \quad \text{and} \quad g_{ij} = \eta_{ij} = (+, -, -, -). \tag{5.22}$$

When the Higgs field is pulled down to $\bar{\mathbf{M}}$ by σ^K it becomes

$$\mathscr{H}^K(\bar{x}) \equiv \sigma^K \mathscr{H} = \sum_{m=-j}^{m=+j} \left(\frac{g}{\sqrt{-\mu^2}} \mathscr{H}^m(\bar{x}) \theta^- \otimes e_m + \text{c.c.} \right) - \cot\theta_j / j\, \theta^3 \otimes Y_H, \tag{5.23}$$

where $n\phi_H/\theta_R$ of (3.23) has been replaced by $\cot\theta_j/j$ for $j \neq 0$ representations. We define $\tan\theta_j$ as the ratio of the Y_H and I_3 root values of e_j (see Fig. 3).

Using (5.20) and (5.22), (5.16) becomes the usual gauge term,

$$-\tfrac{1}{4}G^2 \delta_{\alpha\beta} g^{ij}(\bar{x}) F^\alpha_{ik}(\bar{x}) F^\beta_{jl}(\bar{x}) g^{kl}(\bar{x}) = -\tfrac{1}{4}\eta^{ij}\{B_{ik}B_{jl} + A^3_{ik}A^3_{jl} + 2W^+_{ik}W^-_{jl}\}\eta^{kl}, \tag{5.24}$$

where

$$B_{ik} \equiv \partial_i B_k(\bar{x}) - \partial_k B_i(\bar{x}),$$

$$A^3_{ik} \equiv \partial_i A^3_k - \partial_k A^3_i + ig(W^+_i W^-_k - W^-_i W^+_k),$$

$$W^\pm_{ik} \equiv \partial_i W^\pm_k - \partial_k W^\pm_i \pm ig(A^3_i W^\pm_k - A^3_k W^\pm_i). \tag{5.25}$$

For the gauge (reduced) covariant derivative of the Higgs field, (5.18), we have

$$\begin{aligned}\sigma^K D^K \mathscr{H} = (g\sqrt{-\mu^2}) d\bar{x}^k \sum_{m=-j}^{m=+j} \{ [\partial_i \mathscr{H}^m(\bar{x}) &- i(gj\tan\theta_j B_i(\bar{x}) + gmA^3_i(\bar{x}))\mathscr{H}^m(\bar{x}) - ig\sqrt{(j+m)(j-m+1)/2}\, W^-_i(\bar{x})\mathscr{H}^{m-1}(\bar{x}) \\ &- ig\sqrt{(j-m)(j+m+1)/2}\, W^+_i(\bar{x})\mathscr{H}^{m+1}(\bar{x})]\theta^- \otimes e_m + [\text{c.c.}]\theta^+ \otimes \bar{e}_m \},\end{aligned} \tag{5.26}$$

and for the Higgs kinetic energy, (5.19) becomes

$$\begin{aligned}&\tfrac{1}{2}G^2 S^{-2} g^{ij} \delta^{\Pi\Lambda} \delta_{\alpha\beta} (\mathscr{D}_i \mathscr{H}^\alpha_\Pi)(\mathscr{D}_j \mathscr{H}^\beta_\Lambda) \\ &= \sum_{m=-j}^{m=+j} [\partial_i \mathscr{H}^m(\bar{x}) - ig(j\tan\theta_j B_i(\bar{x}) + m\mathscr{A}^3_i(\bar{x}))\mathscr{H}^m(\bar{x}) \\ &\quad - ig\sqrt{(j+m)(j-m+1)/2}\, W^-_i(\bar{x})\mathscr{H}^{m-1}(\bar{x}) - ig\sqrt{(j-m)(j+m+1)/2}\, W^+_i(\bar{x})\mathscr{H}^{m+1}(\bar{x})] \times [\text{c.c.}].\end{aligned} \tag{5.27}$$

The Higgs potential term, (5.14), is

$$\begin{aligned}V(\mathscr{H}) = \frac{1}{2}\frac{\mu^4}{g^2}\left(\frac{\cot\theta_j}{j}\right)^2 &+ \mu^2 \sum_{m=-j}^{m=+j} \mathscr{H}^m(\bar{x})\mathscr{H}^{m\,*}(\bar{x}) + \frac{1}{2}g^2(j\tan\theta_j)^2\left(\sum_{m=-j}^{m=+j} \mathscr{H}^m \mathscr{H}^{m\,*}\right)^2 \\ &+ \frac{1}{2}g^2\left(\sum_{m=-j}^{m=j} m\mathscr{H}^m \mathscr{H}^{m\,*}\right)^2 + g^2\left(\sum_{m=-j}^{m=j} \sqrt{(j+m)(j-m+1)/2}\,\mathscr{H}^m(\bar{x})\mathscr{H}^{m-1\,*}(\bar{x})\right)(\Sigma\text{c.c.}).\end{aligned} \tag{5.28}$$

For the doublet Higgs this reduces to

$$V(\mathscr{H}) = \frac{1}{4}\frac{\mu^2}{\lambda}\csc^2\theta_{1/2} + \mu^2 \mathscr{H}^\dagger \mathscr{H} + \lambda(\mathscr{H}^\dagger \mathscr{H})^2, \tag{5.29}$$

where $\lambda \equiv \frac{1}{8} g^2 \sec^2\theta_{1/2}$.

As pointed out by Manton,[8] $\theta_{1/2}$ is the weak or Weinberg angle θ_W ($g'/g = \tan\theta_W$), and the Higgs mass $M_{\mathscr{H}} = \sqrt{-2\mu^2}$ is the same as the Z^0 mass ($M_Z = \frac{1}{2}v\sqrt{g^2+g'^2}$). Of the rank 2 gauge group models the bosons are best described by $SU_2 \to G_2$ and $SO_3 \to G_2$ where $j = \frac{1}{2}$, $y_j = 1$ for which $\theta_W = 30°$ agrees reasonably well with current values as does $M_W = \frac{1}{2}g[-\mu^2/\lambda] = 76$ GeV, and $M_Z = 88$ GeV. The subgroup $C_H \subset G_2$ is generated by $\exp\{\varphi Y_H\} \times \exp\{\varphi^3 I_3 + \varphi^+ I_+ + \varphi^- I_-\}$ and is isomorphic to U_2, consequently the reduced gauge group N_K/K is U_2, not $U_1 \times SU_2$ but does have the same algebra.

In conclusion we point out the usefulness of recognizing (1) that N_K/K is the correct reduced gauge group—color can now be included in the symmetry group, i.e., in N_R/R; and (2) that a reduced Kaluza–Klein structure is necessarily present when constructing symmetric gauge Lagrangians—the dynamics of the full reduced gauge group N_K/K can be determined by the Einstein action.

ACKNOWLEDGMENT

We would like to thank Andy Magid for his encouragement and for pointing out several errors in our mathematical arguments.

APPENDIX: PROOF THAT $N^0_R = C^0_R \cdot R^0$

We show here that if S' possesses a definite inner product, $\langle\ ,\ \rangle$ which is ad_S invariant then the connected component of the identity of the normalizer of R, N^0_R, is the product of C^0_R and R^0, i.e.,

$$N^0_R = C^0_R \cdot R^0 \tag{A1}$$

[see Sec. II (f) and (g)].

It is first obvious that R^0 is a connected normal subgroup of N_R^0. Consequently, if $\mathring{n} \in N_R'$, $\mathring{n}$ can be uniquely written as $\mathring{n} = \mathring{r} + \mathring{p}$, where $\mathring{r} \in R'$, $\mathring{p} \in S'$ and where $\langle R', \mathring{p} \rangle = 0$ (i.e., $\mathring{p}$ is orthogonal to R'). Since $\mathring{r} \in N_R'$ it also follows that $\mathring{p} \in N_R'$, i.e., $[\mathring{p}, R'] \subset R'$ and from the ad_S invariance of the inner product

$$0 = \langle \mathring{r}_1, \mathring{p} \rangle = \langle [\mathring{r}_1, \mathring{r}_2], \mathring{p} \rangle + \langle \mathring{r}_1, [\mathring{p}, \mathring{r}_2] \rangle, \tag{A2}$$

for all $\mathring{r}_1$ and $\mathring{r}_2 \in R'$. However, $\langle [\mathring{r}_1, \mathring{r}_2], \mathring{p} \rangle = 0$ because $[\mathring{r}_1, \mathring{r}_2] \in R'$, and hence

$$\langle [\mathring{p}, \mathring{r}_2], R' \rangle = 0, \tag{A3}$$

for all $\mathring{r}_2 \in R'$. Since $\mathring{p} \in N_R'$, $[\mathring{p}, \mathring{r}_2]$ is both in and orthogonal to R', and hence the definiteness of the inner product requires $[\mathring{p}, R'] = 0$. In other words, $\mathring{p} \in C_R'$ or equivalently (A1).

[1] D. B. Fairlie, Phys. Lett. B **82**, 97 (1979).
[2] R. Kerner, J. Math. Phys. **24**, 356 (1983).
[3] M. E. Mayer, "Geometric aspects of symmetry breaking in gauge theories. (Higgs models without Higgs bosons)," in *Lecture Notes in Physics*, (Springer, Berlin, 1980), Vol. 116, pp. 291–295.
[4] Y. Ne'eman, Phys. Lett. B **81**, 190 (1979).
[5] A. S. Schwarz and Yu. S. Tyupkin, Nucl. Phys. B **187**, 321 (1981).
[6] J. G. Taylor, Phys. Lett. B **83**, 331 (1979).
[7] P. Forgács and N. S. Manton, Commun. Math. Phys. **72**, 15 (1980).
[8] N. S. Manton, Nucl. Phys. B **158**, 141 (1979).
[9] N. S. Manton, Nucl. Phys. B **193**, 502 (1981).
[10] G. Chapline and R. Slansky, Nucl. Phys. B **209**, 461 (1982).
[11] A. Jadczyk and K. Pilch, "Geometry of Gauge Fields in a Multidimensional Universe," Preprint ITP-SB (Stony Brook) -83-37.
[12] G. Chapline and N. S. Manton, Nucl. Phys. B **184**, 391 (1981).
[13] R. Coquereaux and A. Jadczyk, Commun. Math. Phys. **90**, 79 (1983).
[14] A. Trautman, Bull. Acad. Pol. Sci. **27**, 7 (1979).
[15] A. S. Schwarz, Commun. Math. Phys. **56**, 79 (1977).
[16] J. Harnad, J. S. Shnider, and L. Vinet, J. Math. Phys. **20**, 931 (1979).
[17] P. G. Bergmann and E. J. Flaherty, J. Math. Phys. **19**, 212 (1978).
[18] H-C. Wang, Nagoya Math. J. **13**, 1 (1958); see reviews by R. Jackiw, Acta Phys. Aust. Suppl. **XXII**, 383 (1980); and M. E. Mayer and A. Trautman, Acta Phys. Aust. Suppl. **XXIII**, 433 (1981).
[19] A. Trautman, Rep. Math. Phys. **1**, 29 (1970).
[20] J. F. Luciani, Nucl. Phys. B **135**, 111 (1978).
[21] R. Percacci and S. Randjbar-Daemi, "Kaluza-Klein theories on bundles with homogeneous fibres–I," ICTP, Trieste, preprint IC/82/18, 1982.
[22] A. Salam and J. Strathdee, "On Kaluza-Klein theory," ICTP, Trieste, preprint IC/81/211, 1981.
[23] E. Witten, Nucl. Phys. B **186**, 412 (1981).
[24] T. Bradfield and R. Kantowski, J. Math. Phys. **23**, 128 (1982).
[25] T. Harnad, S. Shnider, and T. Tafel, Lett. Math. Phys. **4**, 107 (1980).
[26] Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds and Physics* (North-Holland, Amsterdam, 1982), rev. ed.
[27] E. S. Abers and B. W. Lee, Phys. Rep. C **9**, 1 (1973).

An asymptotic behavior of the S-matrix and the inverse scattering problem

Yoshimi Saitō
Department of Mathematics, the University of Alabama in Birmingham, Birmingham, Alabama 35294 and Department of Mathematics, Osaka City University, Osaka, Japan

(Received 1 November 1983; accepted for publication 11 May 1984)

An asymptotic formula of the S-matrix as $k\to\infty$ will be shown for the general short-range scattering in $\mathbf{R}^N (N\geqslant 2)$. The main tool for the proof is the spectral decomposition theorem for the Schrödinger operator. By the use of the asymptotic formula, the uniqueness of the inverse scattering problem will be shown and a reconstruction formula for the potential will be presented for the general short-range scattering.

PACS numbers: 11.20.Fm, 03.65.Nk

I. INTRODUCTION

In a previous paper[1] we gave an asymptotic formula for the S-matrix $S(k)$, $k>0$, associated with the Schrödinger operator $T_3 = -\Delta + Q(y)$ in $\mathbb{R}^3$ with a short-range potential $Q(y) = O(|y|^{-\mu})$ $(|y|\to\infty, \mu>1)$. It was used[1,2] to show the uniqueness for the inverse scattering problem and present a reconstruction formula for the potential $Q(y)$ from the scattering data.

In this paper we shall show that these results can be extended to the Schrödinger operator

$$T = T_N = -\Delta + Q(y) \tag{1.1}$$

in $\mathbb{R}^N$, $N\geqslant 2$, where $Q(y) = O(|y|^{-\mu})$ $(|y|\to\infty, \mu>1)$ is a short-range potential in $\mathbb{R}^N$. First we shall prove an asymptotic formula

$$\lim_{k\to\infty} k^{N-1}(F(k)x_{k,z}, x_{k,z})_{S^{N-1}} = -2\pi \int_{\mathbf{R}^N} \frac{Q(y)dy}{|y-z|^{N-1}} \tag{1.2}$$

in Sec. II. Here

$$F(k) = -2\pi i k^{-(N-2)}(S(k)-I) \quad (k>0), \tag{1.3}$$

I denotes the identity operator on $L^2(S^{N-1})$,

$$x_{k,z}(\omega) = e^{-ikz\omega} \quad (\omega\in S^{N-1}, z\in\mathbb{R}^N), \tag{1.4}$$

and $(\,,\,)_{S^{N-1}}$ is the inner product of $L^2(S^{N-1})$. A necessary and sufficient condition for the solvability of the integral equation

$$g(z) = -2\pi \int_{\mathbf{R}^N} \frac{Q(y)dy}{|y-z|^{N-1}} \tag{1.5}$$

will be given in Sec. III and, by using these results, the uniqueness of the inverse scattering problem in $\mathbb{R}^N$ will be shown with a reconstruction formula which will enable us to reconstruct the potential $Q(y)$ from the scattering data. The mathematically rigorous proof of the main results in Sec. III will be given in Appendices A and B.

Faddeev's work[3] seems to be the first which gave a rigorous proof for the fact that a high energy limit of the scattering amplitude becomes the Fourier transform of the potential. The present work is along the line of Faddeev's work[3] in the sense that it tries to solve the inverse scattering problem through a high energy limit of the S-matrix. Recently Isozaki and Kitada[4] extended Faddeev's result to include long-range scattering as well as short-range scattering. On the other hand Newton[5] gave an extension to $\mathbb{R}^3$ of the Marchenko equation and the Gel'fand–Levitan equation for the Schrödinger operator on $\mathbb{R}^1$, through which he got a reconstruction formula for the potential $Q(y)$. Here $Q(y)$ is assumed to satisfy more restrictive conditions than ours. As for an application of Newton's method to $\mathbb{R}^2$, see Cheney.[6]

II. AN ASYMPTOTIC FORMULA FOR THE S-MATRIX

Let H be a (unique) self-adjoint realization in $L^2(\mathbb{R}^N)$ of the differential operator

$$T = -\Delta + Q(y) \quad (y=(y_1, y_2, \ldots, y_N)\in\mathbb{R}^N), \qquad D(T) = C_0^\infty(\mathbb{R}^N), \tag{2.1}$$

where $D(T)$ is the domain of T and Q is the multiplication operator by a function $Q(y)$ on $\mathbb{R}^N$. The potential $Q(y)$ is assumed to satisfy the following assumption.

Assumption 2.1: $Q(y)$ is a real-valued, continuous function on $\mathbb{R}^N$, $N\geqslant 2$. There exist constants $C_0>0$ and $\mu>1$ such that

$$|Q(y)| \leqslant C_0(1+|y|)^{-\mu} \tag{2.2}$$

for all $y\in\mathbb{R}^N$.

Let $S = W_+^* W_-$ be the scattering operator associated with H, where

$$W_\pm = s-\lim_{t\to\pm\infty} e^{iHt}e^{-iH_0 t} \quad \text{in } L^2(\mathbb{R}^N) \tag{2.3}$$

with a unique self-adjoint extension H_0 of the Laplacian

$$-\Delta = -\sum_{j=1}^N \frac{\partial^2}{\partial y_j^2} \quad \text{on } C_0^\infty(\mathbb{R}^N).$$

It is well known that there exists a family $\{S(k)\}_{k>0}$ of unitary operators on $L^2(S^{N-1})$ such that

$$(\mathcal{F}S\mathcal{F}^*G)(\xi) = \{S(|\xi|)G(|\xi|\cdot)\}(\tilde{\xi}) \qquad (G\in C_0^\infty(\mathbb{R}^N),\ \xi\in\mathbb{R}^N,\ \tilde{\xi}=\xi/|\xi|), \tag{2.4}$$

where $\mathcal{F}$ is the usual Fourier transform from $L^2(\mathbb{R}^N_y)$ to $L^2(\mathbb{R}^N_\xi)$ and $\mathcal{F}^*$ is its adjoint. Here $S(k)$ is called the S-matrix associated with the Schrödinger operator H. For $k>0$, $S(k)$ is a unitary operator on $L^2(S^{N-1})$. For $k>0$, $S(k)-I$ is a compact operator on $L^2(S^{N-1})$ (see Refs. 7 and 8). Let us set

$$F(k) = -2\pi i k^{-(N-2)}(S(k)-I). \tag{2.5}$$

 0022-2488/84/103105-07$02.50

If $\mu > (N+1)/2$ in Assumption 2.1, $F(k)$ is shown to be a Hilbert–Schmidt operator on $L^2(S^{N-1})$, and its Hilbert–Schmidt kernel $F(k;\omega,\omega')$, $\omega, \omega' \in S^{N-1}$, is called the scattering amplitude.[9] By applying the spectral representation theory for the Schrödinger operator H, $S(k)$ can be represented by the eigenoperators associated with H_0 and H, respectively. Let δ be a constant such that

$$\tfrac{1}{2} < \delta < \mu/2\,, \tag{2.6}$$

and set

$$L^2_{\pm\delta}(\mathbb{R}^N) = \left\{ f(x) \Big/ \int_{\mathbb{R}^N} (1+|y|)^{\pm 2\delta} |f(y)|^2\, dy < \infty \right\}. \tag{2.7}$$

Here $L^2_{\pm\delta}$ are Hilbert spaces with their inner products and norms

$$(f,g)_{\pm\delta} = \int_{\mathbb{R}^N} (1+|y|)^{\pm 2\delta} f(y)\,\overline{g(y)}\,dy\,, \tag{2.8}$$

$$\|f\|_{\pm\delta} = (f,g)^{1/2}_{\pm\delta}\,, \tag{2.9}$$

respectively. Then it follows from the limiting absorption principle (Agmon,[7] Saitō,[10] Ikebe–Saitō[11]) that we have the limit

$$R_{\pm}(k) = \lim_{\epsilon \downarrow 0} (H - (k^2 \pm i\epsilon))^{-1} \quad (k>0) \tag{2.10}$$

in $\mathbb{B}(L^2_{\delta}(\mathbb{R}^N), L^2_{-\delta}(\mathbb{R}^N))$, where $\mathbb{B}(X,Y)$ means the Banach space of all linear bounded operators from X to Y. For $r \geqslant 0$ and $k>0$ let us define the operators $\phi^{*}_{0}(r,k)$ and $\phi^{*}_{+}(r,k)$ on $L^2(S^{N-1})$ by

$$\{\phi^{*}_{0}(r,k)x\}(\omega) = (2\pi)^{-N/2} \int_{S^{N-1}} e^{ikr\omega\omega'} x(\omega')\,d\omega', \tag{2.11}$$

$$\{\phi^{*}_{+}(r,k)x\}(\omega) = \{\phi^{*}_{0}(r,k)x\}(\omega) - \{R_{-}(k)f_0\}(r\omega)\,, \tag{2.12}$$

where $x \in L^2(S^{N-1})$, $f_0(y) = Q(y)\{\phi^{*}_{0}(|y|,k)x\}(\tilde{y})$ with $\tilde{y} = y/|y|$, and $\omega\omega'$ means the inner product in $\mathbb{R}^N$ for ω, $\omega' \in S^{N-1}$. Then it can be shown[12] that

$$\int_0^\infty (1+r)^{-2\delta} \|\phi^{*}_{0}(r,k)x\|^2_{L^2(S^{N-1})} r^{N-1}\,dr < \infty\,, \tag{2.13}$$

$$\int_0^\infty (1+r)^{-2\delta} \|\phi^{*}_{+}(r,k)x\|^2_{L^2(S^{N-1})} r^{N-1}\,dr < \infty\,, \tag{2.14}$$

or, equivalently,

$$\{\phi^{*}_{0}(|y|,k)x\}(\tilde{y}), \{\phi^{*}_{+}(|y|,k)x\}(\tilde{y}) \in L^2_{-\delta}(\mathbb{R}^N) \tag{2.15}$$

for all $x \in L^2(S^{N-1})$ and $k>0$. $v(y) = \{\phi^{*}_{+}(|y|,k)x\}(\tilde{y})$ satisfies $(H-k^2)v = 0$ at least in the sense of distributions and $\phi^{*}_{+}(r,k)$ [and its adjoint $\phi_{+}(r,k)$ in $L^2(S^{N-1})$] are used for constructing a spectral decomposition theorem for H.[7,13]

Theorem 2.2:[14] Let Assumption 2.1 be satisfied and let $F(k)$ be as in (2.5). Then we have

$$(F(k)x,x')_{S^{N-1}} = -2\pi^2 \int_{\mathbb{R}^N} Q(y)\{\phi^{*}_{0}(|y|,k)x\}(\tilde{y}) \times \overline{\{\phi^{*}_{+}(|y|,k)x'\}(\tilde{y})}\,dy \tag{2.16}$$

for $x, x' \in L^2(S^{N-1})$ and $k>0$.

In order to show the asymptotic formula (1.2) we need to recall the following.

Lemma 2.3: There exists a constant C_1 such that

$$\left| \int_{S^{N-1}} e^{ir\omega\omega'}\,d\omega' - 2\left(\frac{2\pi}{r}\right)^{(N-1)/2} \cos\left(r - \frac{(N-1)\pi}{2}\right) \right| \leqslant C_1 \min\left\{ \frac{1}{r^{(N-1)/2}}, \frac{1}{r^{(N+1)/2}} \right\} \tag{2.17}$$

for all $r>0$.

Proof: Introducing polar coordinates $(\theta_1,\theta_2,\ldots,\theta_{N-1})$, we get

$$\int_{S^{N-1}} e^{ir\omega\omega'}\,d\omega' = \frac{2\pi^{(N-1)/2}}{\Gamma((N-1)/2)} \int_0^\pi e^{ir\cos\theta_1} (\sin\theta_1)^{N-2}\,d\theta_1\,. \tag{2.18}$$

Then we can show the asymptotic formula

$$\int_{S^{N-1}} e^{ir\omega\omega'}\,d\omega' = 2\left(\frac{2\pi}{r}\right)^{(N-1)/2} \cos\left(r - \frac{(N-1)\pi}{2}\right) + O(r^{-(N-1)/2}) \quad (r \to \infty) \tag{2.19}$$

from the representation formula for the Bessel function $J_\nu(r)$,[15]

$$J_\nu(r) = \frac{(r/2)^\nu}{\sqrt{\pi}\,\Gamma(\nu+\frac{1}{2})} \int_0^\pi e^{ir\cos\theta} \sin^{2\nu}\theta\,d\theta \quad (\nu\text{:real})\,, \tag{2.20}$$

and the well-known asymptotic formula

$$J_\nu(r) = \sqrt{\frac{2}{\pi r}} \cos\left(r - \frac{(2\nu+1)\pi}{2}\right) + O(r^{-3/2}) \quad (r \to \infty)\,. \tag{2.21}$$

Inequality (2.17) is easily obtained from (2.19).

Q.E.D.

The next lemma, which can be easily derived from (2.17), is a key lemma for the proof of (1.2).

Lemma 2.4. Let $h(y)$ be a bounded, measurable function on $\mathbb{R}^N$ such that $h(y) = O(|y|^{-1-\epsilon})$ with $\epsilon>0$ as $|y| \to \infty$. Then we have

$$\lim_{k\to\infty} k^{N-1} \int_{\mathbb{R}^N} h(y) \left| (2\pi)^{-N/2} \int_{S^{N-1}} e^{ik(y-z)\omega}\,d\omega \right|^2 dy = \frac{1}{\pi} \int_{\mathbb{R}^N} \frac{h(y)}{|y-z|^{N-1}}\,dy \tag{2.22}$$

for each $z \in \mathbb{R}^N$, where $(y-z)\omega$ is the inner product of $y-z$ and ω in $\mathbb{R}^N$.

Proof: Setting

$$q(r) = \int_{S^{N-1}} e^{ir\omega\omega'}\,d\omega' - 2\left(\frac{2\pi}{r}\right)^{(N-1)/2} \cos\left(r - \frac{(N-1)\pi}{2}\right), \tag{2.23}$$

we obtain

$$k^{N-1}\int_{\mathbf{R}^N} h(y)\left|(2\pi)^{-N/2}\int_{S^{N-1}} e^{ik(y-z)\omega}\,d\omega\right|^2 dy$$
$$=\frac{1}{\pi}\int_{\mathbf{R}^N}\frac{h(y)\{\cos(2k|y-z|-(N-1)\pi)+1\}}{|y-z|^{N-1}}\,dy+\frac{4k^{(N-1)/2}}{(2\pi)^{(N+1)/2}}\operatorname{Re}\int_{\mathbf{R}^N} h(y)\frac{\cos(k|y-z|-[(N-1)/2]\pi)}{|y-z|^{(N-1)/2}}$$
$$\times q(k|y-z|)dy+\frac{k^{N-1}}{2\pi}\int_{\mathbf{R}^N} h(y)|q(k|y-z|)|^2\,dy\equiv J_1+J_2+J_3\,. \tag{2.24}$$

It follows from the Riemann–Lebesgue theorem that

$$\lim_{k\to\infty} J_1=\frac{1}{\pi}\int_{\mathbf{R}^N}\frac{h(y)}{|y-z|^{N-1}}\,dy\,. \tag{2.25}$$

Let $\alpha>0$. Then

$$|J_2|\leqslant\frac{4k^{(N-1)/2}}{(2\pi)^{(N+1)/2}}\int_{|y-z|<\alpha}\frac{|h(y)||q(k|y-z|)|}{|y-z|^{(N-1)/2}}\,dy$$
$$+\frac{4k^{(N-1)/2}}{(2\pi)^{(N+1)/2}}\int_{|y-z|>\alpha}\frac{|h(y)||q(k|y-z|)|}{|y-z|^{(N-1)/2}}\,dy$$
$$\equiv J_{21}+J_{22}. \tag{2.26}$$

By the use of the boundedness of $h(y)$ and the estimate

$$|q(k|y-z|)|\leqslant C_1 k^{-(N-1)/2}|y-z|^{-(N-1)/2}, \tag{2.27}$$

which is obtained from (2.17), we have $J_{21}=O(\alpha)$ $(\alpha\downarrow 0)$ independent of k. Next it can be easily seen that $J_{22}=O(1/k)$ $(k\to\infty)$ for fixed $\alpha>0$. Thus we have shown that $J_2=o(1)$ as $k\uparrow\infty$. Quite similarly we can show that $J_3=o(1)$ as $k\uparrow\infty$. Q.E.D.

The following formula is an extension to $\mathbf{R}^N$ of the one which was obtained in the case of $\mathbf{R}^3$.[16]

Theorem 2.5 [asymptotic formula for $F(k)$]: Let $Q(y)$ satisfy Assumption 2.1 and let $F(k)$ be as above. For $k>0$ and $z\in\mathbf{R}^N$ set

$$x_{k,z}(\omega)=e^{-ikz\omega}. \tag{2.28}$$

Then we have

$$\lim_{k\to\infty} k^{N-1}(F(k)x_{k,z},x_{k,z})_{S^{N-1}}=-2\pi\int_{\mathbf{R}^N}\frac{Q(y)dy}{|y-z|^{N-1}}. \tag{2.29}$$

Proof: It follows from Theorem 2.2 and (2.11) and (2.12) that

$$k^{N-1}(F(k)x_{k,z},x_{k,z})_{S^{N-1}}$$
$$=-2\pi^2k^{N-1}\int_{\mathbf{R}^N} Q(y)\left|\frac{1}{(2\pi)^{N/2}}\int_{S^{N-1}} e^{ik(y-z)\omega}\,d\omega\right|^2 dy$$
$$+2\pi^2k^{N-1}\int_{\mathbf{R}^N} Q(y)\left\{\frac{1}{(2\pi)^{N/2}}\int_{S^{N-1}} e^{ik(y-z)\omega}\,d\omega\right\}$$
$$\times\overline{(R_-(k)f_1)(y)}\,dy$$
$$\equiv K_1+K_2, \tag{2.30}$$

where $f_1=f_0$ in (2.12) with $x=x_{k,z}$, i.e.,

$$f_1(y)=Q(y)\frac{1}{(2\pi)^{N/2}}\int_{S^{N-1}} e^{ik(y-z)\omega}\,d\omega\,. \tag{2.31}$$

From Lemma 2.4 with $h(y)=Q(y)$ we see that

$$\lim_{k\to\infty} K_1=-2\pi\int_{\mathbf{R}^N}\frac{Q(y)}{|y-z|^{N-1}}\,dy\,. \tag{2.32}$$

Since $(1+|y|)^{2\delta}|Q(y)|\leqslant C_0$ by Assumption 2.1 and the definition of δ, we have

$$|K_2|\leqslant 2\pi^2C_0k^{N-1}\left|\left|\frac{1}{(2\pi)^{N/2}}\int_{S^{N-1}} e^{ik(\cdot-z)\omega}\,d\omega\right|\right|_{-\delta}$$
$$\times\|R_-(k)f_1\|_{-\delta}\,. \tag{2.33}$$

It is known[17] that we have

$$\|R_-(k)\|=O(1/k)\quad(k\to\infty), \tag{2.34}$$

where $\|R_-(k)\|$ is the operator norm of $R_-(k)$ in $\mathbf{B}(L^2_\delta(\mathbf{R}^N),L^2_{-\delta}(\mathbf{R}^N))$. Therefore $\|R_-(k)f_1\|_{-\delta}$ can be estimated as

$$\|R_-(k)f_1\|_{-\delta}\leqslant O\left(\frac{1}{k}\right)\left|\left|\frac{1}{(2\pi)^{N/2}}\int_{S^{N-1}} e^{ik(\cdot-z)\omega}\,d\omega\right|\right|_{-\delta}. \tag{2.35}$$

Thus we obtain from (2.33) and (2.35)

$$K_2=O\left(\frac{1}{k}\right)$$
$$\times k^{N-1}\int_{\mathbf{R}^N}(1+|y|)^{-2\delta}\left|\int_{S^{N-1}} e^{ik(y-z)\omega}\,d\omega\right|^2 dy\,, \tag{2.36}$$

and hence, by using Lemma 2.4 again, the term K_2 is shown to be $O(1/k)$, which completes the proof.

III. THE INVERSE SCATTERING PROBLEM

Let us consider the integral equation

$$g(z)=-2\pi\int_{\mathbf{R}^N}\frac{Q(y)dy}{|z-y|^{N-1}}, \tag{3.1}$$

where $g(z)$ is a given function and we seek the solution $Q(y)$ which is a short-range potential in the sense of Assumption 2.1. In this section we shall give a necessary and sufficient condition on $g(z)$ for the unique solvability of Eq. (3.1), through which we shall show the uniqueness of the inverse scattering problem and a reconstruction formula for the potential $Q(y)$. Since the argument can be done quite similarly to the three-dimensional case[18] the details of the proof will be given in Appendices A and B.

Let us first introduce some notations. Let ϵ be a positive number. Then a function space A_ϵ is defined by

$$A_\epsilon=\{f\in C(\mathbf{R}^N)/f(y)=O(|y|^{-\epsilon})\quad\text{as}\quad|y|\to\infty\}, \tag{3.2}$$

i.e., $f(y)$ is a continuous function on $\mathbf{R}^N$ and satisfies the estimate

$$|f(y)|\leqslant C(1+|y|)^{-\epsilon}\quad(y\in\mathbf{R}^N) \tag{3.3}$$

with a constant $C>0$. Let $\mathcal{S}=\mathcal{S}(\mathbf{R}^N)$ be all rapidly decreasing functions on $\mathbf{R}^N$, and let $\mathcal{S}'=\mathcal{S}'(\mathbf{R}^N)$ be all linear

continuous functionals on $\mathscr{S}$. The pairing between $\mathscr{S}$ and $\mathscr{S}'$ will be denoted by $\langle\ ,\ \rangle$. The Fourier transforms $\mathscr{F}$, $\overline{\mathscr{F}}$, $\mathscr{F}^*$, $\overline{\mathscr{F}}^*$ are defined by

$$(\mathscr{F}f)(\xi)=(2\pi)^{-N/2}\int e^{-i\xi y}f(y)dy\,, \tag{3.4}$$

$$(\overline{\mathscr{F}}f)(\xi)=(\mathscr{F}f)(-\xi)\,, \tag{3.5}$$

$$(\mathscr{F}^*F)(y)=(2\pi)^{-N/2}\int e^{i\xi y}F(\xi)d\xi\,, \tag{3.6}$$

$$(\overline{\mathscr{F}}^*F)(y)=(\mathscr{F}^*F)(-y)\,. \tag{3.7}$$

Here ξy is the inner product of y and ξ in $\mathbb{R}^N$.

Definition 3.1: Let $g\in A_\epsilon$ with $\epsilon>0$ and let $s>0$. Then a linear functional $\Lambda^s g$ on $\mathscr{S}_\xi=\mathscr{S}(\mathbb{R}^N_\xi)$ is defined by

$$\langle\Lambda^s g,G\rangle=\int_{\mathbb{R}^N}g(y)\{\overline{\mathscr{F}}^*(|\xi|^sG)\}(y)dy\quad(G\in\mathscr{S}_\xi)\,, \tag{3.8}$$

where $\overline{\mathscr{F}}^*$ is as in (3.7).

Then we can show that $\Lambda^s g\in\mathscr{S}'_\xi=\mathscr{S}'(\mathbb{R}^N_\xi)$ (see Appendix A).

Let $Q\in A_\mu$ with $1<\mu<N$ and let g be defined by (3.1). Then it can be easily seen that $g\in A_{\mu-1}$, and hence $\Lambda^s g$ is well-defined for any $s>0$. Then, taking the Fourier transforms of the both sides of (3.1), we have

$$\mathscr{F}g=-(2\pi)^{N/2+1}\mathscr{F}(|y|^{-(N-1)})\times\mathscr{F}Q \tag{3.9}$$

in the sense of $\mathscr{S}'$. Here we used the well-known formula

$$\mathscr{F}(f*g)=(2\pi)^{N/2}(\mathscr{F}f)\times(\mathscr{F}g)\,,$$

where $*$ means convolution. On the other hand, we have

$$\mathscr{F}(|y|^{-t})(\xi)=B_t|\xi|^{-(N-t)}\quad(0<t<N) \tag{3.10}$$

in $\mathscr{S}'$ with

$$B_t=2^{N/2-t}\frac{\Gamma((N-t)/2)}{\Gamma(t/2)}\,, \tag{3.11}$$

$\Gamma(t)$ being the Γ-function.[19] It follows from (3.9) and (3.10) with $t=N-1$ that

$$\mathscr{F}g=-\frac{4\pi^{(N+3)/2}}{\Gamma((N-1)/2)}|\xi|^{-1}\mathscr{F}Q\,. \tag{3.12}$$

The well-definedness of $\Lambda g(=\Lambda^{-1}g)$ shows that we can obtain from (3.12)

$$\mathscr{F}Q=-\beta_N|\xi|\mathscr{F}g=-\beta_N\Lambda g\quad\text{in }\mathscr{S}' \tag{3.13}$$

with

$$\beta_N=\frac{\Gamma((N-1)/2)}{4\pi^{(N+3)/2}}\,. \tag{3.14}$$

Thus we have

$$Q(y)=-\beta_N\mathscr{F}^*\Lambda g\,. \tag{3.15}$$

Since the whole process can be reversed, we can see that (3.15) gives a unique solution of Eq. (3.1) under the condition that $\mathscr{F}^*\Lambda g\in A_\mu$. Thus we get the following theorem.

Theorem 3.2: The integral equation (3.1) has a unique solution $Q\in A_\mu$ with $1<\mu<N$ if and only if

$$g\in A_{\mu-1}\quad\text{and}\quad\mathscr{F}^*\Lambda g\in A_\mu\,. \tag{3.16}$$

Then the solution $Q(y)$ has the form (3.15).

Remark 3.3: Let α,λ be constants such that $0<\alpha<N$, $\lambda\in\mathbb{C}$ and let us consider a more general equation

$$g(z)=\lambda\int_{\mathbb{R}^N}\frac{Q(y)}{|y-z|^\alpha}dy\,. \tag{3.17}$$

Then Theorem 3.2 can be easily extended to the case of Eq. (3.17) (see Ref. 20).

This theorem, together with Theorem 2.5, guarantees the uniqueness of the inverse scattering problem and gives a reconstruction formula for the potential $Q(y)$ (see Refs. 3–6 in Sec. I). In Appendix B it will be shown that the formal argument can be verified rigorously.

Using Theorem 3.2, various kinds of reconstruction formulas for $Q(y)$ can be obtained. In the following examples (3.4–3.6) it is assumed that $1<\mu<N$, $Q(y)\in A_\mu$, and $g(z)$ is defined by

$$g(z)=\lim_{k\to\infty}(F(k)x_{k,z},x_{k,z})_{S^{N-1}}\,, \tag{3.18}$$

where $F(k)$ is defined by (2.5) and $x_{k,z}$ is given by (2.27).

Example 3.4: Let g satisfy

$$g\in A_{\mu-1}\cap D(H_0^{1/2})\,, \tag{3.19}$$

$$H_0^{1/2}g\in A_\mu\,, \tag{3.20}$$

where H_0 is a unique self-adjoint extension of $-\Delta$ on $C_0^\infty(\mathbb{R}^N)$ and $D(H_0^{1/2})$ is the domain of $H_0^{1/2}$. Then $Q(y)$ can be represented as

$$Q=-\beta_N H_0^{1/2}g \tag{3.21}$$

with β_N given in (3.14).

In fact in this case we have $\mathscr{F}^*\Lambda g=H_0^{1/2}g$ in $\mathscr{S}'$.

The next example shows that $Q(y)$ can be represented as a form of convolution if $g(z)$ is assumed to behave more nicely.

Example 3.5: Let $g(z)\in A_\epsilon$ with some $\epsilon>0$ and let $g_s=\mathscr{F}^*\Lambda^s g\in A_\mu$ with $1<s<\mu<N$. Then we have

$$Q(y)=-\gamma_{s,N}\int_{\mathbb{R}^N}\frac{g_s(z)}{|y-z|^{N+1-s}}dz\in A_{1+\mu-s} \tag{3.22}$$

with

$$\gamma_{s,N}=\frac{\Gamma((N-s+1)/2)\Gamma((N-1)/2)}{2^{1+s}\pi^{N+3/2}\Gamma((s-1)/2)}\,. \tag{3.23}$$

The proof is essentially the same as in the case of $\mathbb{R}^3$.[21] Therefore only the formal proof will be given. Taking the Fourier transforms of the both sides of (3.1), we get

$$\mathscr{F}g=-2\pi(2\pi)^{N/2}B_{N-1}|\xi|^{-1}\mathscr{F}Q\,, \tag{3.24}$$

whence follows

$$\begin{aligned}Q&=-\frac{1}{(2\pi)^{N/2+1}B_{N-1}}\mathscr{F}^*\left(\frac{1}{|\xi|^{s-1}}\cdot|\xi|^s\mathscr{F}g\right)\\&=-\frac{1}{(2\pi)^{N+1}B_{N-1}}\mathscr{F}^*\left(\frac{1}{|\xi|^{s-1}}\right)*g_s\,.\end{aligned} \tag{3.25}$$

Thus we have (3.22) from (3.25) and (3.10) with $t=s-1$.

A special case of Example 3.5 with $s=2$ is as follows.

Example 3.6: Let $g\in A_\epsilon$ with $\epsilon>0$ and $\Delta g\in A_\mu$ with $\mu>2$, where Δg is defined in the sense of distributions. Then $Q(y)$ can be represented as

$$Q(y)=-\frac{\{\Gamma((N-1)/2)\}^2}{8\pi^{N+2}}\int_{\mathbb{R}^N}\frac{(\Delta g)(z)}{|y-z|^{N-1}}dz\,. \tag{3.26}$$

Theorem 3.2 can be used to characterize the asymptotic behavior of the S-matrix $S(k)$ as $k\to\infty$. Let $\{S_j(k)/k>0\}$, $j=1,2$, be two families of bounded operators on $L^2(S^{N-1})$. We shall say that $S_1(k)$ is asymptotically equal to $S_2(k)$ with respect to $x_{k,z}(\omega)=e^{-ikz\omega}$ if and only if

$$\lim_{k\to\infty} k(\{S_1(k)-S_2(k)\}x_{k,z},x_{k,z})_{S^{N-1}}=0\,. \tag{3.27}$$

Theorem 3.7: Let $\{S(k)/k>0\}$ be a family of bounded operators on $L^2(S^{N-1})$. Then there exists $Q\in A_\mu$ with $\mu>1$ such that $S(k)$ is asymptotically equal to the S-matrix $S_0(k)$ associated with the Schrödinger operator $H=-\Delta+Q$ with respect to $x_{k,z}$ if and only if there exists the limit

$$g(z)=\lim_{k\to\infty} k^{N-1}(F(k)x_{k,z},x_{k,z})_{S^{N-1}} \tag{3.28}$$

for each $z\in\mathbb{R}^N$ and $g(z)$ satisfies

$$g\in A_{\mu-1}\,, \tag{3.29}$$

$$\mathcal{F}^*\Lambda g\in A_\mu\,, \tag{3.30}$$

where $F(k)=-2\pi ik^{-(N-2)}(S(k)-I)$.

Since the proof is essentially the same as in the case of $\mathbb{R}^3$,[22] we omit it.

ACKNOWLEDGMENTS

This work was done during the author's stay at the University of Alabama in Birmingham for the 1983–84 academic year. The author wishes to express his sincere appreciation to the University of Alabama in Birmingham for the kind hospitality. He thanks Osaka City University also for granting him leave of absence during this period.

APPENDIX A: PRELIMINARY ESTIMATES

Let $\Lambda^s g$ be as in Definition 3.1. In order to show that $\Lambda^s g\in\mathcal{S}'_\xi=\mathcal{S}'(\mathbb{R}^N_\xi)$, let us introduce the following norm $|G|_s$ for $G\in\mathcal{S}_\xi=\mathcal{S}(\mathbb{R}^N_\xi)$:

$$|G|_s=\sum_{|\beta|<N}\sum_{\alpha<\beta}\int_{\mathbb{R}^N}|\xi|^{s-|\beta|+|\alpha|}|D^\alpha G(\xi)|d\xi$$

$$\equiv\sum_{|\beta|<N}\sum_{\alpha<\beta}|G|_{\alpha,\beta}\,, \tag{A1}$$

where $\alpha=(\alpha_1,\alpha_2,\ldots,\alpha_N)$, $\beta=(\beta_1,\beta_2,\ldots,\beta_N)$ are multi-indices with non-negative integers α_j, β_k, $\alpha\leqslant\beta$ means that $\alpha_j\leqslant\beta_j$ for any $j=1,2,\ldots,N$, and

$$D^\alpha=\left(\frac{\partial}{\partial\xi_1}\right)^{\alpha_1}\left(\frac{\partial}{\partial\xi_2}\right)^{\alpha_2},\ldots,\left(\frac{\partial}{\partial\xi_N}\right)^{\alpha_N}. \tag{A2}$$

For any $s>0$ it is easy to see that the topology induced in $\mathcal{S}_\xi$ by the norm $|\ \ |_s$ is weaker than the proper topology of $\mathcal{S}_\xi$.

Lemma A.1: Let $s>0$. Then there exists a positive constant C_s such that

$$|\langle\Lambda^s g,G\rangle|\leqslant C_s\|(1+|y|)^{-N}g\|_{L^1(\mathbb{R}^N)}|G|_s \tag{A3}$$

for any $g\in A_\epsilon$ with $\epsilon>0$ and any $G\in\mathcal{S}_\xi$. Here C_s depends only on $s>0$.

Proof: Let us first assume that N is an even positive integer so that we can write $N=2M$ with a positive integer M. It follows from the identity

$$e^{-iy\xi}=(1+|y|^2)^{-M}(1-\Delta_\xi)^M e^{-iy\xi} \tag{A4}$$

that we have, by integration by parts,

$$\overline{\mathcal{F}}^*(|\xi|^sG(\xi))(y)=(2\pi)^{-N/2}(1+|y|^2)^{-M}$$

$$\times\int_{\mathbb{R}^N}e^{-iy\xi}(1-\Delta_\xi)^M(|\xi|^sG(\xi))d\xi\,. \tag{A5}$$

Since the integral in the right-hand side can be estimated by $|G|_s$, we get

$$|\overline{\mathcal{F}}^*(|\xi|^sG(\xi))(y)|\leqslant C_s(1+|y|)^{-N}|G|_s\,, \tag{A6}$$

whence (A3) follows directly. Let us next consider the case that $N=2M+1$ with a positive integer M. Proceeding as above, we have

$$\mathcal{F}^*(|\xi|^sG)(y)=(2\pi)^{-N/2}(1+|y|^2)^{-M}(iy_j)^{-1}$$

$$\times\int_{\mathbb{R}^N}e^{-iy\xi}\left(\frac{\partial}{\partial\xi_j}\right)(1-\Delta_\xi)^M$$

$$\times\{|\xi|^sG(\xi)\}d\xi\,. \tag{A7}$$

Take $\phi_j\in C_0^\infty(\mathbb{R}^N)$, $j=0,1,2,\ldots,N$, such that $0\leqslant\phi_j(y)\leqslant1$, $\Sigma_{j=0}^N\phi_j(y)=1$, the support of $\phi_0(y)$ is in $B_0=\{y/|y|<2\}$, and the support of $\phi_j(y)$ is contained in $B_j=\{y/\sqrt{N+1}\,|y_j|>|y|,\ |y|>1\}$ for $j=1,2,\ldots,N$. Then we have

$$\langle\Lambda^s g,G\rangle=(2\pi)^{-N/2}\int_{B_0}\phi_0(y)\int_{\mathbb{R}^N}e^{-iy\xi}|\xi|^sG(\xi)d\xi\,dy$$

$$+\sum_{j=1}^N(2\pi)^{-N/2}\int_{B_j}\phi_j(y)(1+|y|^2)^{-M}(iy_j)^{-1}$$

$$\times\int e^{-iy\xi}\left(\frac{\partial}{\partial\xi_j}\right)(1-\Delta_\xi)^M\{|\xi|^sG(\xi)\}d\xi\,dy\,, \tag{A8}$$

from which (A3) is easily obtained.

Q.E.D.

It follows from Lemma A.1 and the remark about the norm $|\ \ |_s$ before Lemma A.1 that $\Lambda^s g$ is well-defined as an element of $\mathcal{S}'_\xi$.

The next lemma about the norm $|\ \ |_s$ will be used to show Theorem 3.2 in Appendix B. Let $\rho(\xi)\in C^\infty(\mathbb{R}^N)$ such that $0\leqslant\rho(\xi)\leqslant1$ and

$$\rho(\xi)=\begin{cases}0, & |\xi|\leqslant\frac{1}{2}\,,\\ 1, & |\xi|\geqslant1\,.\end{cases} \tag{A9}$$

Set $\rho_m(\xi)=\rho(m\xi)$.

Lemma A.2: Let $s>0$ and let $|\ \ |_s$ be as in (A1). Then we have

$$\lim_{m\to\infty}|G-\rho_mG|_s=0 \tag{A10}$$

for any $G\in\mathcal{S}_\xi$.

Proof: Since

$$|G|_s=\sum_{|\beta|<N}\sum_{\alpha<\beta}|G|_{\alpha,\beta}\,,$$

we have only to show that $|G-\rho_mG|_{\alpha,\beta}\to0$ $(m\to\infty)$ for each pair (α,β) such that $|\beta|\leqslant N$, $\alpha\leqslant\beta$. It follows from the estimate

$$|G-\rho_mG|_{0,\beta}=\int_{\mathbb{R}^N}|\xi|^{s-|\beta|}|G(\xi)-\rho_m(\xi)G(\xi)|d\xi$$

$$\leqslant \int_{|\xi|<m^{-1}} |\xi|^{s-|\beta|}|G(\xi)|d\xi \qquad \text{(A11)}$$

that we have $|G-\rho_m G|_{0,\beta}\to 0$ $(m\to\infty)$ for any β such that $|\beta|\leqslant N$. Let us next assume that $|\alpha|\neq 0$. Then we have

$$\begin{aligned}
&|G-\rho_m G|_{\alpha,\beta} \\
&= \int_{\mathbf{R}^N} |\xi|^{s-|\beta|+|\alpha|}|D^\alpha G(\xi)(1-\rho_m(\xi))|d\xi \\
&\quad + \sum_{\substack{\alpha_1+\alpha_2=\alpha \\ \alpha_1\neq 0}} C_{\alpha_1,\alpha_2}\int_{\mathbf{R}^N} |\xi|^{s-|\beta|+|\alpha|} \\
&\quad \times |D^{\alpha_1}\rho_m(\xi)||D^{\alpha_2}G(\xi)|d\xi \\
&\leqslant \int_{|\xi|<m^{-1}} |\xi|^{s-|\beta|+|\alpha|}|D^\alpha G(\xi)|d\xi \\
&\quad + \sum_{\substack{\alpha_1+\alpha_2=\alpha \\ \alpha_1\neq 0}} C_{\alpha_1,\alpha_2}\int_{|\xi|<m^{-1}} |\xi|^{s-|\beta|} \\
&\quad \times |(D^{\alpha_1}\rho)(m\xi)||D^{\alpha_2}G(\xi)|d\xi\,, \qquad \text{(A12)}
\end{aligned}$$

where C_{α_1,α_2} is a constant depending only on α_1 and α_2, and we have used the estimate

$$\begin{aligned}
|D^{\alpha_1}\rho_m(\xi)| &= m^{|\alpha|}|(D^{\alpha_1}\rho)(m\xi)| \\
&\leqslant |\xi|^{-|\alpha|}|(D^{\alpha_1}\rho)(m\xi)| \quad (|\xi|\leqslant m^{-1})\,. \qquad \text{(A13)}
\end{aligned}$$

From (A12) we can easily see that $|G-\rho_m G|_{\alpha,\beta}\to 0$ as $m\to\infty$.

Q.E.D.

APPENDIX B: PROOF OF THEOREM 3.2

A rigorous proof of Theorem 3.2 can be obtained from the following two propositions.

Proposition B.1: Let $Q\in A_\mu$ with $1<\mu<N$. Let g be defined by (3.1), i.e.,

$$g(z) = -2\pi\int_{\mathbf{R}^N}\frac{Q(y)dy}{|z-y|^{N-1}}\,. \qquad \text{(B1)}$$

Then we have $g\in A_{\mu-1}$ and

$$Q = -\beta_N \mathcal{F}^{*}\Lambda g\,, \qquad \text{(B2)}$$

where $\Lambda g = \Lambda^{1}g$ is defined by Definition 3.1 and β_N is given in (3.11).

Proof: Let us first start with approximating $Q(y)$ by smooth functions. Take a sequence $\{Q_n\}$ such that

$$Q_n\in C_0^\infty(\mathbf{R}^N) \quad (n=1,2,...)\,, \qquad \text{(B3)}$$

$$|Q_n(y)|\leqslant C(1+|y|)^{-\mu} \quad (y\in\mathbf{R}^N, n=1,2,...)\,, \qquad \text{(B4)}$$

$$\lim_{n\to\infty} Q_n(y) = Q(y) \quad (y\in\mathbf{R}^N)\,, \qquad \text{(B5)}$$

where $C>0$ is independent of $n=1,2,...$. Set

$$g_n(z) = -2\pi\int_{\mathbf{R}^N}\frac{Q_n(y)}{|z-y|^{N-1}}\,. \qquad \text{(B6)}$$

Then we can easily see that

$$|g_n(z)|\leqslant C'(1+|z|)^{-(\mu-1)} \quad (z\in\mathbf{R}^N, n=1,2,...)\,, \qquad \text{(B7)}$$

$$\lim_{n\to\infty} g_n(z) = g(z) \quad (z\in\mathbf{R}^N)\,, \qquad \text{(B8)}$$

where $C'>0$ is independent of $n=1,2,...$. Further it is shown that

$$\mathcal{F}g_n = -\beta_N^{-1}|\xi|^{-1}\mathcal{F}Q_n \qquad \text{(B9)}$$

in $\mathcal{S}'_\xi$. In fact, we have for any $G\in\mathcal{S}_\xi$

$$\begin{aligned}
\langle \mathcal{F}g_n, G\rangle &= \langle g_n, \overline{\mathcal{F}}^{*}G\rangle \\
&= -2\pi\langle |y|^{-(N-1)}{*}Q_n, \overline{\mathcal{F}}^{*}G\rangle \\
&= -2\pi\langle |y|^{-(N-1)}, \check{Q}_n{*}\overline{\mathcal{F}}^{*}G\rangle \qquad \text{(B10)}
\end{aligned}$$

with $\check{Q}_n(y)=Q_n(-y)$, where we should note that $\check{Q}_n{*}\overline{\mathcal{F}}^{*}G\in\mathcal{S}_y = \mathcal{S}(\mathbf{R}^N_y)$. Since it follows from (3.10) with $t=N-1$ that

$$|y|^{-(N-1)} = B_{N-1}\,\mathcal{F}^{*}(|\xi|^{-1}) \qquad \text{(B11)}$$

with

$$B_{N-1} = \frac{\pi^{1/2}}{2^{N/2-1}\Gamma((N-1)/2)}\,, \qquad \text{(B12)}$$

we have from (B10)

$$\begin{aligned}
\langle \mathcal{F}g_n, G\rangle &= -2\pi B_{N-1}\langle \mathcal{F}^{*}(|\xi|^{-1}), \check{Q}_n{*}\overline{\mathcal{F}}^{*}G\rangle \\
&= -2\pi B_{N-1}\langle |\xi|^{-1}, \overline{\mathcal{F}}(\check{Q}_n{*}\overline{\mathcal{F}}^{*}G)\rangle \\
&= -(2\pi)^{N/2+1}B_{N-1}\langle |\xi|^{-1}, (\overline{\mathcal{F}}\check{Q}_n)G\rangle \\
&= -\beta_N^{-1}\langle |\xi|^{-1}(\mathcal{F}Q_n), G\rangle\,, \qquad \text{(B13)}
\end{aligned}$$

where we have used the formula $\overline{\mathcal{F}}(F{*}G) = (2\pi)^{\pi/2}(\overline{\mathcal{F}}F)\times(\overline{\mathcal{F}}G)$. Equation (B9) is immediate from (B13).

Since $|\xi|\notin\mathcal{S}_\xi$, we cannot multiply both sides of (B9) by $|\xi|$ to get

$$\mathcal{F}Q_n = -\beta_N|\xi|\,\mathcal{F}g_n \qquad \text{(B14)}$$

in $\mathcal{S}'_\xi$. But (B14) holds good. In fact, let $\rho_m(\xi)$ as in Appendix A and let us define $\mathcal{S}_0 = \mathcal{S}_{0\xi}$ by

$$\mathcal{S}_0 = \mathcal{S}_{0\xi} = \{\, G\in\mathcal{S}_\xi / G(\xi)=0 \text{ in a neighborhood of } \xi=0\,\}\,. \qquad \text{(B15)}$$

Then we have $\rho_m(\xi)G\in\mathcal{S}_{0\xi}$ for $G\in\mathcal{S}_\xi$ and we have $|\xi|\rho_m(\xi)G\in\mathcal{S}_{0\xi}\subset\mathcal{S}_\xi$, too. Thus, using (B9), we get

$$\begin{aligned}
-\beta_N\langle \Lambda g_n, \rho_m G\rangle &= -\beta_N\langle g_n, \overline{\mathcal{F}}^{*}(|\xi|\rho_m G)\rangle \\
&= -\beta_N\langle \mathcal{F}g_n, |\xi|\rho_m G\rangle \\
&= \langle |\xi|^{-1}\mathcal{F}Q_n, |\xi|\rho_m G\rangle \\
&= \int_{\mathbf{R}^N}|\xi|^{-1}(\mathcal{F}Q_n)(\xi)|\xi|\rho_m(\xi)G(\xi)d\xi \\
&= \langle \mathcal{F}Q_n, \rho_m G\rangle\,. \qquad \text{(B16)}
\end{aligned}$$

Noting Lemma A.2 and letting $m\to\infty$ in (B16), we arrive at (B14). Equation (B2) is now obtained by letting $n\to\infty$ in (B14).

Q.E.D.

Proposition B.2: Let $g\in A_{\mu-1}$ with $\mu>1$ and let $\mathcal{F}^{*}\Lambda g\in A_\mu$, i.e., there exists $h\in A_\mu$ such that $h=\mathcal{F}^{*}\Lambda g$ in $\mathcal{S}'$. Then

$$Q = -\beta_N\,\mathcal{F}^{*}\Lambda g \quad (= -\beta_N h) \qquad \text{(B17)}$$

is a solution of (B1).

Proof: Let $G\in\mathcal{S}_{0\xi}$, where $\mathcal{S}_{0\xi}$ is given by (B15). Then $|\xi|^{-(N-1)}{*}\overline{\mathcal{F}}^{*}G\in\mathcal{S}$, because $|\xi|^{-1}G\in\mathcal{S}_{0\xi}\subset\mathcal{S}_\xi$ and we have

$$|\xi|^{-(N-1)}*\overline{\mathcal{F}}*G=(2\pi)^{N/2}B_1^{-1}\overline{\mathcal{F}}*(|\xi|^{-1}G)\,. \quad \text{(B18)}$$

As for the definition of $\overline{\mathcal{F}}*$ and B_1, see (3.7) and (3.11), respectively. Then it follows from (B16) and (B18) that

$$\begin{aligned}
&\langle \mathcal{F}(|y|^{-(N-1)}*Q),G\rangle \\
&\quad=\langle Q,|y|^{-(N-1)}*\overline{\mathcal{F}}*G\rangle \\
&\quad=-\beta_N\langle \mathcal{F}*\Lambda g,|y|^{-(N-1)}*\overline{\mathcal{F}}*G\rangle \\
&\quad=-\beta_N(2\pi)^{N/2}B_1^{-1}\langle \mathcal{F}*\Lambda g,\overline{\mathcal{F}}*(|\xi|^{-1}G)\rangle \\
&\quad=-\beta_N(2\pi)^{N/2}B_1^{-1}\langle \Lambda g,|\xi|^{-1}G\rangle \\
&\quad=-(2\pi)^{-1}\langle g,\overline{\mathcal{F}}*(|\xi||\xi|^{-1}G)\rangle \\
&\quad=-(2\pi)^{-1}\langle g,\overline{\mathcal{F}}*G\rangle=(-2\pi)^{-1}\langle \mathcal{F}g,G\rangle\,, \quad \text{(B19)}
\end{aligned}$$

where we should note that $\beta_N(2\pi)^{N/2}B_1^{-1}=(2\pi)^{-1}$. Thus we get

$$\langle \mathcal{F}\{g+2\pi|y|^{-(N-1)}*Q\},G\rangle=0 \quad \text{(B20)}$$

for any $G\in\mathcal{S}_{0\xi}$, which means that the support of $\mathcal{F}\{g+2\pi|y|^{-(N-1)}*Q\}$ is contained in the origin $\xi=0$. Therefore there exists a polynomial $P(y)=P(y_1,y_2,\ldots,y_N)$ such that

$$\mathcal{F}\{g+2\pi|y|^{-(N-1)}*Q\}=P(D)\delta\,, \quad \text{(B21)}$$

where δ is the Dirac δ-function and

$$D=\left(-i\frac{\partial}{\partial\xi_1},\,-i\frac{\partial}{\partial\xi_2},\ldots,\,-i\frac{\partial}{\partial\xi_N}\right)$$

(see Ref. 23). Since $\mathcal{F}*P(D)\delta=P(y)$, we have from (B21)

$$g+2\pi|y|^{-(N-1)}*Q=P(y)\,. \quad \text{(B22)}$$

Here the left-hand side of (B22) is $o(1)$ at infinity by the condition imposed on $g(z)$, and hence we have $P(y)\equiv 0$, which implies that Q is a solution of (B2).

Q.E.D.

[1]Y. Saitō, Osaka J. Math. **19**, 527 (1982).
[2]Y. Saitō, J. Math. Kyoto Univ. **22**, 307 (1982).
[3]L. D. Faddeev, Vestn. Leningr. Univ. **7**, 126 (1956).
[4]H. Isozaki and H. Kitada (in preparation).
[5]R. Newton, J. Math. Phys. **21**, 1698 (1980).
[6]M. Cheney, J. Math. Phys. **25**, 94 (1984).
[7]S. Agmon, Ann. Scuola Nor. Sup. Pisa (4) **2**, 151 (1975).
[8]Y. Saitō, J. Reine Angew. Math. **314**, 99 (1980).
[9]See, e.g., W. D. Amrein, J. M. Jauch, and K. B. Sinha, "Scattering Theory in Quantum Mechanics," in *Lecture Notes and Supplements in Physics* (Benjamin, Reading, MA, 1977).
[10]Y. Saitō, Publ. Res. Inst. Math. Sci. Kyoto Univ. **7**, 581 (1971/72).
[11]T. Ikebe and Y. Saitō, J. Math. Kyoto Univ. **7**, 513 (1972).
[12]Ref. 8, Lemma 2.3 and (i) of Remark 3.3. Here we follow the notations in Ref. 8.
[13]Y. Saitō, "Spectral Representations for Schrödinger Operators with Long-Range Potentials," in *Lecture Notes in Mathematics 727* (Springer, Berlin, 1979).
[14]Ref. 8, Theorem 2.6 and (ii) of Remark 2.3.
[15]E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U. P., Cambridge, 1927), p. 366.
[16]Ref. 1, Theorem 5.3.
[17]Y. Saitō, Publ. Res. Inst. Math. Sci. Kyoto Univ. **9**, 397 (1972). Y. Saitō, Osaka J. Math. **11**, 295 (1974).
[18]Ref. 2.
[19]See, e.g., I. M. Gel'fand and G. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. I. Note that the definition of the Fourier transforms we adopted here is slightly different from the one in the above book.
[20]Ref. 2.
[21]Ref. 2, Theorem 2.1.
[22]Ref. 2, Theorem 3.4.
[23]L. Schwartz, *Théorie des distributions* (Herman, Paris. 1957), Vol. I, p. 100.

Plane-surface deformation in electrohydrostatics with electric double layer[a)]

R. Cade
Department of Mathematics, University of Puerto Rico, Mayagüez, Puerto Rico 00708

(Received 19 March 1984; accepted for publication 4 May 1984)

General equations governing the static deformation of the surface of a conducting liquid under an electric field, when there is an electric double layer at the surface, are here applied to the general situation where the undeformed surface is a horizontal plane, and a number of specific problems are solved. The theory in this case has two features which distinguish it from other situations, studied previously, where the undeformed surface is curved. One feature requires a difference in the practical process of solution, while the other requires some careful theoretical examination in order to justify properly the solutions actually obtained.

PACS numbers: 41.10.Dq, 46.20. + e, 68.10. − m

I. INTRODUCTION

We consider the separating surface S between an electrically conducting liquid A and a nonconducting fluid B (perhaps air or empty space), and suppose that at S, there is an electric double layer. An electrostatic field is applied, which has the effect of maintaining S in a state of static deformation, and our concern is with the problem of determining the shape of S.

The reduced problem in which there is no gravity (as when A and B are of the same density) and no electric field, is that of determining S under just the influence of surface tension, and is the Plateau problem of classical hydrostatics. To generalize the problem so as to include a gravity effect, one must consider a hydrostatic equation containing a gravity term and a nonconstant hydrostatic pressure, and its further generalization to include an electric field is accomplished by replacing the hydrostatic pressure by the Maxwell stress tensor. The validity of this step, long believed in the history of electrostatics, was established rigorously only relatively recently, by Brown.[1]

Brown's justification of the Maxwell stress theory is correct provided there is no electric double layer at the boundary S. Otherwise, a refinement is needed,[2] and this leads to different equations governing the shape of S (see Ref. 3). Since the presence of a double layer is a normal property of matter, this is clearly significant, which is especially true since the double-layer effect has an essentially linear dependence upon the field (in contrast with the essentially quadratic classical effect), and is in terms of a new constitutive constant characterizing the interface (the familiar one being the surface tension).

In Ref. 3, the surface deformation theory was applied to sphere and cylinder problems, by which we mean that the surface in the absence of gravity and an electric field (the "zero-order surface") is (or is part of) a sphere or a cylinder, as the case may be. In the present paper, we apply it to plane-surface problems.

Despite its apparently more fundamental character, the plane-surface theory has peculiar features not shared with the sphere and cylinder theories, features which, in fact, are responsible on the one hand for its being technically more difficult, while on the other hand presenting a difficulty of principle in the theoretical basis of the solution procedure, one which it is important to clarify if the solutions themselves are to be deemed acceptable. It is thus felt that the plane-surface theory requires its own special exposition.

In the next section we review, for easy reference, the general surface-deformation theory. In the following section we make the specialization appropriate to the plane-surface situation. We next solve a specific plane-surface problem in some detail, and thereafter give the solutions of a number of other problems doing little more mathematically than writing them down, the mode of solution in each case being the same. Finally, we discuss the special features, mentioned above, of the present plane-surface theory.

II. GOVERNING EQUATIONS AND SOLUTION PROCEDURE

We assume that the system is either axisymmetric with vertical axis, or two-dimensional with representative section in a vertical plane. In either case, S is generated by a curve C in a vertical plane, which we call the profile, and the equations for determining this are[3]

$$\frac{\epsilon E_B^2}{8\pi} - 2K_m(T + \Phi E_B) + g\omega x - \Gamma = \frac{\partial \nu}{\partial X^i}\frac{\partial X^j}{\partial \nu} T_A{}^i{}_j, \tag{1}$$

$$\Phi \frac{dE_B}{dt} = \frac{\partial \nu}{\partial X^i}\frac{\partial X^j}{\partial t} T_A{}^i{}_j,$$

$$\frac{\partial T^j{}_i}{\partial X^j} + T^h{}_i \begin{Bmatrix} j \\ h\ j \end{Bmatrix} - T^j{}_h \begin{Bmatrix} h \\ i\ j \end{Bmatrix} = \frac{\partial x}{\partial X^i} g\rho, \tag{2}$$

$$(g_{ii}g^{jj})^{1/2} T^i{}_j = (g_{jj}g^{ii})^{1/2} T^j{}_i \quad (i,j \text{ not summed}).$$

The meaning of the notation is as follows.

E_B: the normal component of electric intensity at, and on the B side of, S, with positive direction $A \to B$.

ϵ: the dielectric constant of B.

ω: the density of B.

ρ: the density of A.

T: the surface tension at S.

Φ: a constant which is characteristic of the double layer

[a)] This paper is an amplified version of an address given by the author at the 2nd Latin-American Symposium of Applied Mathematics, Rio de Janeiro, December 1983, and published (in Spanish) in the Proceedings.

 0022-2488/84/103112-08$02.50

(measuring the interactability of the double layer with a field).

X^i: orthogonal curvilinear coordinates, with $i = 0, 1, 2$, such that S under no field and no gravity is a surface $X^0 = \text{const}$.

g_{ij}, g^{ij}: the metric tensors for X^i.

$T^i{}_j$: the stress tensor in A (in mixed form), relative to X^i.

$T_A{}^i{}_j$: $T^i{}_j$ at, and on the A side of, S.

ν: distance measured normally from S and with positive direction $A \to B$, at the point of S concerned.

t: a parameter for the profile C.

K_m: the mean curvature of S [reducing to half the (signed) curvature κ of C in the two-dimensional case], reckoned positive if S, as the boundary of A, is locally convex.

x: a vertical Cartesian coordinate with positive direction upward.

g: the acceleration due to gravity.

Γ: a constant fixed by the solution in a particular case.

As we briefly indicated, we classify problems according to the form of what we call the zero-order surface, S under no gravity and no electric field. For example, a drop of liquid without these influences assumes a spherical form, which is, in fact, the most elementary solution of the classical Plateau problem. But under gravity and a field, there is a departure from the spherical form. It is the object of the present theory to be able to calculate this departure. However, we call all specific problems sphere problems, because the zero-order surface is at least partially (here wholly) a sphere.

The field strength E_B is obtained from the solution of an electrostatic boundary-value problem in which the potential V takes a constant value of S, and this pinpoints the essential difficulty of solving (1) and (2). One needs to know E_B in order to determine S, while at the same time, one has to know S in order to determine E_B. The difficulty is overcome by the use of perturbation theory, regarding the field and gravity effects as both small and represented by perturbation parameters. The first step is to determine E_B for the zero-order surface. Calling the result the first-order field, $E_{B,1}$, one uses it to determine a first-order surface, one represented by a profile which is to the first order in the perturbation parameters. One then redetermines E_B using this first-order surface. The result is the second-order field, $E_{B,2}$, and one uses it to determine the second-order surface, and so on.

III. REDUCTION OF THE EQUATIONS FOR PLANE-SURFACE PROBLEMS

The procedure described has, as we said, already been used for sphere and cylinder problems,[3] when the coordinate system X^i in (1) and (2) is, in the respective cases, spherical polar and cylindrical, and the zero-order profile in both cases a circle or circular arc.

Plane-surface (or free-surface) problems are those for which the zero-order surface is a horizontal plane, and the zero-order profile, a line. Such problems can be two dimensional or three dimensional. Two dimensionally, the appropriate system X^i is Cartesian coordinates (x, y, z) with, say, x as X^0 and the y axis is the representative section. Axisymmetrically, it is cylindrical coordinates (r, ϕ, x), taking x as X^0 (r is X^0 in the cylinder problems). In both cases, x may be considered the same as the x in Eq. (1).

To focus for the time being on the two-dimensional case, we seek to approximate to a profile solution of the form $x = x(y)$, and with X^i as the said Cartesian coordinates, (1) and (2) reduce with a little elementary differential geometry to

$$\frac{\epsilon E_B^2}{8\pi} - \kappa(T + \Phi E_B) + g\omega x - \Gamma = \frac{1}{1+x'^2} T_{A,xx} - \frac{x'}{1+x'^2}(T_{A,xy} + T_{A,yx}) + \frac{x'^2}{1+x'^2} T_{A,yy},$$

$$\Phi \frac{dE_B}{dy} = \frac{x'}{(1+x'^2)^{1/2}} T_{A,xx} + \frac{1}{(1+x'^2)^{1/2}} T_{A,xy} - \frac{x'^2}{(1+x'^2)^{1/2}} T_{A,yx} - \frac{x'}{(1+x'^2)^{1/2}} T_{A,yy}, \tag{3}$$

$$\frac{\partial T_{xx}}{\partial x} + \frac{\partial T_{yx}}{\partial y} = g\rho, \quad \frac{\partial T_{xy}}{\partial x} + \frac{\partial T_{yy}}{\partial y} = 0, \tag{4}$$

$$T_{xy} = T_{yx},$$

where x' means dx/dy.

We conveniently take the origin of coordinates in the undeformed surface, so that the zero-order profile is the line $x = 0$. Then the form of solution we assume, in terms of perturbation theory, is

$$x = x(y) = \sum_{n=1}^{\infty} \lambda_n(y), \tag{5}$$

where $\lambda_n(y)$ is a homogeneous nth-degree polynomial in the perturbation parameters. The nth-order solution, that is to say, the nth approximation to $x(y)$, is

$$x = x_n(y) = \sum_{r=1}^{n} \lambda_r(y). \tag{6}$$

We suppose that we have the first-order solution $E_{B,1}(y)$ of the field problem, and assume the following form of solution for the stress-field equations (4):

$$T_{xx} = -p_0 + M(x)F(y), \quad T_{xy} = T_{yx} = G(y),$$
$$T_{yy} = -p_0 + N(x)H(y), \tag{7}$$

where p_0 is a constant. Working to the first order, the use of (7) in (3) and (4) enables us to determine the functions in (7), except for $N(x)$ which happens, at the said order, to drop out. Then, using the well-known Cartesian formula for the curvature κ, and continuing the due approximation, the first of Eqs. (3) is reduced to the differential equation

$$\frac{d^2x_1}{dy^2} - \frac{g(\rho - \omega)}{T} x_1 = \Delta_1 - \frac{\epsilon E_{B,1}^2}{8\pi T}, \tag{8}$$

where Δ_1 is a constant, and whose solution is the first-order profile. We see that the constant Φ has disappeared, so that, at the first order, the present theory predicts identically with the classical theory.

We regard the magnitude of the applied field as gauged by a parameter P, so that $E_{B,1}(y)$ is of the form $Pf(y)$, where f is a function of y only, containing no hidden parameters. We

choose one perturbation parameter γ as a convenient multiple of P^2, and it is contained implicitly in (8). A second β is a multiple of ΦP and, as we have said, falls out of the first-order problem.

Having the solution of (8), we re-solve the field problem but using the boundary $x = x_1(y)$, thus obtaining the second-order field, which we shall write as $E_{B,1}(y) + e_2(y)$. Now, repeating the process outlined above, but with x_1 known and working to the second order, we obtain a differential equation for the second-order correction $\lambda_2(y)$ to the profile solution:

$$\frac{d^2}{dy^2}\left(\lambda_2 + \frac{\Phi}{T}x_1 E_{B,1}\right) - \frac{g(\rho-\omega)}{T}\left(\lambda_2 + \frac{\Phi}{T}x_1 E_{B,1}\right) = \delta_2 - \frac{g(\rho-\omega)}{T^2}\Phi x_1 E_{B,1} - \frac{\epsilon E_{B,1} e_2}{4\pi T}, \tag{9}$$

where δ_2 is a constant.

In the axisymmetric theory, we have the same equations, (3) and (5)–(7), except for the replacement of y by r, the inclusion of a stress $T^{\phi}{}_{\phi}$ in the analog of (7) and which we equate to T_{rr}, and the retention of the $2K_m$ of (1) [this reducing to the ordinary (signed) curvature of the profile only in the two-dimensional theory]. The analog of (4) is somewhat different, but the axisymmetric departure which takes effect is the different formula for $2K_m$ (Eisenhart,[4] p. 227), leading to analogs of (8) and (9) which differ by containing first-derivative terms. The equations that we obtain are

$$\frac{d^2x_1}{dr^2} + \frac{1}{r}\frac{dx_1}{dr} - \frac{g(\rho-\omega)}{T}x_1 = \Delta_1 - \frac{\epsilon E_{B,1}^2}{8\pi T}, \tag{10}$$

$$\frac{d^2}{dr^2}\left(\lambda_2 + \frac{\Phi}{T}x_1 E_{B,1}\right) + \frac{1}{r}\frac{d}{dr}\left(\lambda_2 + \frac{\Phi}{T}x_1 E_{B,1}\right) - \frac{g(\rho-\omega)}{T}\left(\lambda_2 + \frac{\Phi}{T}x_1 E_{B,1}\right) = \delta_2 - \frac{g(\rho-\omega)}{T^2}\Phi x_1 E_{B,1} - \frac{\epsilon E_{B,1} e_2}{4\pi T}, \tag{11}$$

and again, the Φ-effect enters only at the second order.

The solution of a specific problem to an order not higher than the second involves solving the appropriate one or ones of Eqs. (8)–(11), the problem being distinguished by the function $E_{B,1}(y)$ that is used and, if the order is the second, $e_2(y)$. Each of the equations contains an unknown constant, and two more enter through the integration. It is found that the three constants are fixed by the following two conditions:

$$x(y)\,(x(r)) \to 0, \quad y\,(r) \to \infty, \quad x'(0) = 0, \tag{12}$$

the alternatives in the former applying to the respective two-dimensional and axisymmetric cases. The physical propriety of the condition on x is obvious. That on x' is implied by the smoothness of the profile in the axisymmetric case, but is so in the two-dimensional case only if there is symmetry about the x axis. There will be symmetry in the problems here. If there is not, the condition $x'(0) = 0$ is replaced by $x(y) \to 0$, $y \to -\infty$, one found to be equivalent to the former if there is the symmetry.

IV. THE DIFFERENTIAL PROFILE

Before turning to the solution of specific problems, we shall discuss a point already hinted at and of much significance in electrohydrostatics with the electric double layer, especially so in the plane-surface theory.

Referring to the parameter P which gauges the strength of the applied field and, for ease of explanation, treating specifically the two-dimensional theory, let us express the profile as $x = x(y;P)$. The function

$$x_d = x(y;P) - x(y;-P), \tag{13}$$

we call the *differential profile*, being the difference of the profile from that obtained on field reversal.

There is no differential profile in the classical theory, in which no account is taken of the electric double layer, because the field enters only through E_B^2. It is thus something representing the *qualitative* difference of the predictions of the present theory from those of the classical theory, and is, in essence, the double-layer effect. It has been mentioned before[3] that the probable interest in the present electrohydrostatic theory is its allowing measurement of the constitutive constant Φ, and that the obvious property to use is the differential profile, practical measurement being, no doubt, of the differential maximum height.[5]

Now here, there is no *first-order* differential profile [cf. (8) and the remark following], but by the form of (9), we expect there to be a nonzero second-order differential profile, and this will be given by

$$x_d = \lambda_{d,2}(y;P) = \lambda_2(y;P) - \lambda_2(y;-P), \tag{14}$$

the first-order parts $x_1(y;P), x_1(y;-P)$ being equal and canceling.

The two functions λ_2 on the right of (14) satisfy (9) separately, but of course for the different respective fields. However, the function $E_{B,1}e_2$ is the same, for it is proportional to P^4. To see this, let us, as we may, write $E_{B,1}$ as $P\{f(y) + \psi(y)\}$, where, as before [paragraph following (8)], $Pf(y)$ is $E_{B,1}$, so that $P\psi(y)$ is e_2. Now as we said, $f(y)$ contains no "hidden parameters," but this is not so of $\psi(y)$ as it depends upon the first-order profile. If β had not dropped out, it would, to the required order, be a homogeneous linear expression in β and γ. But β did drop out, leaving γ which is proportional to P^2, a fact which leads at once to the point in question. It follows that when we subtract Eq. (9) for the respective functions λ_2, the last terms cancel, and taking this with the fact that $x_1E_{B,1}$ is a proportional to P^3, we conclude from (9) and (14) that $\lambda_{d,2}$ satisfies the equation

$$\frac{d^2}{dy^2}\left(\lambda_{d,2} + \frac{2\Phi}{T}x_1 E_{B,1}\right) - \frac{g(\rho-\omega)}{T}\left(\lambda_{d,2} + \frac{2\Phi}{T}x_1 E_{B,1}\right) = \delta_{d,2} - [2g(\rho-\omega)/T^2]\Phi x_1 E_{B,1}, \tag{15}$$

where $\delta_{d,2}$ is some new constant.

Likewise, we have in the axisymmetric case,

$$\frac{d^2}{dr^2}\left(\lambda_{d,2} + \frac{2\Phi}{T}x_1 E_{B,1}\right) + \frac{1}{r}\frac{d}{dr}\left(\lambda_{d,2} + \frac{2\Phi}{T}x_1 E_{B,1}\right) - [g(\rho-\omega)/T](\lambda_{d,2} + (2\Phi/T)x_1 E_{B,1}) = \delta_{d,2} - [2g(\rho-\omega)/T^2]\Phi x_1 E_{B,1}. \tag{16}$$

Clearly, if we know the second-order profile, we can obtain the second-order differential profile at once from (14) [or (14) with r in place of y if the problem is axisymmetric]. But if we do not, we can find the latter directly by solving (15)

or (16), and the importance of this alternative is that these equations *do not contain the second-order field.* Now the practical obstacle to carrying specific problems to the second order is always that of finding the second-order field, this being invariably a major undertaking of electrostatics. We therefore have the remarkable situation that the difficult aspect of the theory, in the problem-solving context, does not operate to hinder calculation of the most interesting consequence.

It is a situation holding in any electrohydrostatic problem in which there is no Φ-effect at first order, something always true in our plane-surface problems, and so exceptionally in the sphere and cylinder problems.[3] If there is a Φ-effect at first order, we still do not need the second-order field if, as probable, we are content to know the now nonzero *first-order* differential profile. The special point of the considerations of this section is that, in the plane-surface theory, where the first-order differential profile is always zero, we are not prevented by the difficulty of the field problem from finding the second-order differential profile.

V. THE LINE CHARGE PROBLEM

Perhaps the most fundamental two-dimensional problem is that of a uniform line charge of strength f per centimeter, passing through the point $(a, 0)$ $(a>0)$ of the xy plane.

Taking the potential of the conducting liquid as zero, so that the potential V in the space above becomes zero at the liquid surface,[6] the first-order electrostatic problem is well known as the most elementary of two-dimensional image problems,[7] giving

$$V=\frac{f}{\epsilon}\ln\frac{(x+a)^2+y^2}{(x-a)^2+y^2}, \tag{17}$$

from which, since $E_{B,1}=-(\partial V/\partial x)_{x=0}$,

$$E_{B,1}=-(4f/\epsilon)[a/(a^2+y^2)]. \tag{18}$$

We define the perturbation parameters

$$\beta_f=4f\Phi/\epsilon T,\quad \gamma_f=2f^2/\pi\epsilon T, \tag{19}$$

so that, with (18), the differential equation for the first-order profile (8) becomes

$$\frac{d^2x_1}{dy^2}-q^2x_1=\Delta_1-\frac{\gamma_f a^2}{(a^2+y^2)^2},\quad q^2=\frac{g(\rho-\omega)}{T}. \tag{20}$$

The solution of such equations as this, arising in the electrohydrostatic theory, is most conveniently carried out by the method of variation of parameters (see, e.g., Ince[8]). In the present case, with a view to optimum computational usefulness of the result, we invoke the extensively tabulated functions $Ei\zeta$, $\zeta=\xi+i\eta$, the exponential integral of a complex variable.[9] We define, for complex parameters α and μ, the two real functions

$$\begin{aligned}v(\alpha;\xi;\mu;\pm)=\alpha\{&e^{\xi+\mu}Ei(\xi+\mu)\pm e^{-\xi+\mu}Ei(-\xi+\mu)\}\\&+\bar{\alpha}\{e^{\xi+\bar{\mu}}Ei(\xi+\bar{\mu})\pm e^{-\xi+\bar{\mu}}Ei(-\xi+\bar{\mu})\},\end{aligned} \tag{21}$$

and using complex partial fractions on the right of (20), applying the said method of solution and using the end conditions (12), we obtain for the first-order profile[10]

$$x_1=(\gamma_f/8)v(1+i/qa;qy;iqa;+). \tag{22}$$

The second-order differential profile is easily found by solving (15), but the present problem happens to be the only one of our set for which we can actually calculate the second-order profile itself. The quite difficult electrostatic problem which is for the line charge and zero equipotential given by (22) has been solved (to the required order) before.[10] With the use of its solution, and (22) again, in (9), solving by variation of parameters and using (12), the second-order profile is found to be

$$\begin{aligned}x_2(y)&=x_1(y)+\lambda_2(y),\\
\lambda_2(y)&=\frac{\beta_f\gamma_f}{8}\left\{\frac{av(1+i/qa;qy;iqa;+)}{a^2+y^2}\right.\\
&\quad-\frac{qa}{2}\left(e^{qy}\int_y^\infty+e^{-qy}\int_{-y}^\infty\right)\\
&\quad\left.\times\frac{v(1+i/qa;qt;iqa;+)e^{-qt}}{a^2+t^2}\,dt\right\}\\
&\quad+\frac{\gamma_f^2}{16}\left\{a^2\left(e^{qy}\int_y^\infty+e^{-qy}\int_{-y}^\infty\right)\right.\\
&\quad\times\left[v\left(\frac{1}{qa}-i;qt;iqa;+\right)\right.\\
&\quad\left.\left.+v\left(\frac{1}{qa}-i;qt;iqa;-\right)\right]\frac{e^{-qt}}{(a^2+t^2)^2}\,dt\right.\\
&\quad+\frac{1}{4a}v\left[v\left(1+\frac{i}{qa};0;2iqa;+\right)\left[1+i\left(\frac{1}{qa}-qa\right)\right]\right.\\
&\quad\left.+3+i\left(\frac{3}{qa}-qa\right);qy;iqa;+\right]\\
&\quad\left.+\left[1+v\left(1+\frac{i}{qa};0;2iqa;+\right)\right]\frac{a}{a^2+y^2}\right\}.\end{aligned} \tag{23}$$

VI. THE POINT CHARGE PROBLEM

The three-dimensional analog of the line-charge problem is the axisymmetric one in which a point charge e at $(a, 0)$ replaces the line charge and, to the first order, we solve the differential equation (10), having first found $E_{B,1}$ from the corresponding three-dimensional problem of electrical images.[7]

Again, we use variation of parameters, and the main point of difference, arising from the Bessel-type operator on x_1, on the left of (10), is that we have no well-known function such as $Ei\zeta$ to assist us, so that, unlike (22) [but like (23)], the solution has to be left in terms of integrals. Without giving further details, the result is

$$\begin{aligned}x_1(r)&=\gamma_e a^2\left(I_0(qr)\int_r^\infty K_0(qt)\right.\\
&\quad\left.+K_0(qr)\int_0^r I_0(qt)\right)\frac{t\,dt}{(a^2+t^2)^3},\\
\gamma_e&=e^2/2\pi\epsilon T,\quad q^2=g(\rho-\omega)/T,\end{aligned} \tag{24}$$

where I_0 and K_0 are the usual notation for the modified Bessel functions of order zero.

A graphical comparison of first-order point-charge and line-charge profiles[10] shows that the former are more strongly "peaked," as one would intuitively expect.

We cannot find the second-order profile in the present case, being unable to solve the field problem. We could easily find the second-order differential profile, solving (16), but it is not worth the trouble since, unlike in the problems that follow, field reversal is hardly a physically achievable process.

VII. THE CONDUCTING CYLINDER AND CONDUCTING SPHERE PROBLEMS

While we are inclinded to regard the line and point-charge problems as theoretically fundamental among those for the plane liquid surface, of greater experimental realism are the closely related ones in which, instead of a line charge and a point charge, we have, respectively, a conducting cylinder and a conducting sphere, as shown in Fig. 1, the conductor being in each case at a given potential V_0.

We suppose that the center of the sphere, or of the section of the cylinder, is $(a, 0)$, and that the radius in either case is b $(<a)$. To consider first the cylinder case, the first-order field problem is that of finding a potential V which takes the value V_0 on the cylinder and vanishes on the plane $x = 0$. This is a problem well known in two-dimensional electrostatics (Smythe[11]), in which the field outside the cylinder is found to be the same as if this were replaced by a certain line charge passing through a certain point $(c, 0)$. In fact, we have now, in place of (17),

$$V = \frac{V_0}{2\alpha} \ln \frac{(x+c)^2 + y^2}{(x-c)^2 + y^2}, \qquad (25)$$

$$c = (a^2 - b^2)^{1/2}, \quad \alpha = \cosh^{-1} \frac{a}{b}.$$

It follows at once, defining the perturbation parameters

$$\beta_c = 2\Phi V_0/\alpha T, \quad \gamma_c = \epsilon V_0^2/2\pi\alpha^2 T, \qquad (26)$$

and comparing with the line-charge solution of Sec. V, that the first-order profile is given by

$$x_1(y) = (\gamma_c/8)v(1 + i/qc; qy; iqc; +). \qquad (27)$$

Unlike with the line-charge problem, we are unable to solve the second-order field problem and so cannot find the second-order profile, but from the $\beta_f\,\gamma_f$ term in the second-order line-charge solution (23), we can infer at once what the second-order differential profile will be. It is

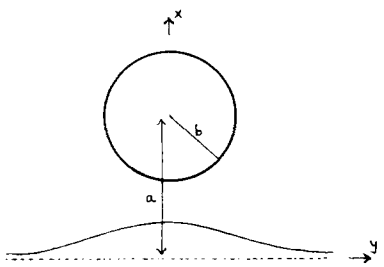


FIG. 1. A normally plane liquid surface deformed under the influence of a charged conducting cylinder or sphere in the space above.

$$x_{d,2}(y) = \frac{\beta_c\gamma_c}{4}\left\{\frac{cv(1 + i/qc; qy; iqc; +)}{c^2 + y^2} - \frac{qc}{2}\left(e^{qy}\int_y^\infty + e^{-qy}\int_{-y}^\infty\right) \times \frac{v(1 + i/qc; qt; iqc, +)e^{-qt}}{c^2 + t^2}\,dt\right\}, \qquad (28)$$

and of course, it is now a relevant, physically achievable attribute, the applied field being supplied by a conductor at a given potential.

In the sphere case, even the first-order field problem is a nontrivial one, and it is convenient to use the option we have of obtaining the general form of solution of (10) by variation of parameters, without using *a priori* a solution for the field. It is, in fact,

$$x_1(r) = \frac{\epsilon}{8\pi T}\left(I_0(qr)\int_r^\infty K_0(qt) + K_0(qr)\int_0^r I_0(qt)\right)[E_{B,1}(t)]^2\,dt. \qquad (29)$$

The first-order field problem is one well known in the theory of "infinite series of images" (see, e.g., Jeans[7] or Smythe[11]), and gives

$$E_{B,1}(r) = -2abV_0\sum_{n=1}^\infty \frac{\sinh\alpha\tanh\alpha\,\mathrm{csch}\,n\alpha\coth n\alpha}{(a^2\tanh^2\alpha\coth^2 n\alpha + r^2)^{3/2}}$$

$$(\alpha = \cosh^{-1}(a/b)). \qquad (30)$$

This form has, however, the drawback of slow convergence, especially if a/b is close to 1, and nowadays one has the advantage of progress that has been made in the solution of problems of the kind by bispherical harmonics. In our case, an immediate adaptation of a solution obtained by Warren and Cuthrell[12] gives

$$E_{B,1}(r) = \frac{-4V_0c^2}{(r^2 + c^2)^{3/2}}\sum_{n=0}^\infty \frac{2n+1}{e^{(2n+1)\alpha} - 1}P_n\left(\frac{r^2 - c^2}{r^2 + c^2}\right)$$

$$(c = (a^2 - b^2)^{1/2}), \qquad (31)$$

where P_n are the Legendre polynomials, and this form converges more rapidly than (30). With either (30) or (31), of course, the use of (29) in practice would present a considerable computational task.

To calculate the second-order differential profile in the present sphere case, we have to appeal to the solution of (16). With the same approach as led to (29), we obtain

$$x_{d,2}(r) = \frac{2\Phi}{T}\left\{q^2\left(I_0(qr)\int_r^\infty K_0(qt) + K_0(qr)\int_0^r I_0(qt)\right)x_1(t)E_{B,1}(t)\,dt - x_1(r)E_{B,1}(r)\right\}, \qquad (32)$$

and, as with (29), we may use either (30) or (31) for the field. We have not introduced perturbation parameters for the sphere problem as this is not done before the explicit use of a field formula.

VIII. THE HALF-SUBMERGED CONDUCTING CYLINDER AND SPHERE

A fascinating problem, which we believe is being studied now for the first time, is where the cylinder or sphere of the last section is half-immersed in the conducting liquid (Fig. 2), forming with it a single conductor at potential zero, and the applied field is a uniform field of strength F, with positive direction the positive x-direction. In what we call the sphere and cylinder problems in the electrohydrostatic context, where the *zero-order liquid* surface is spherical or cylindrical, the problems have been studied of a hemispherical and semicylindrical liquid drop on a plane horizontal solid conducting surface.[3] The problems of the present section are those in which the attachment of "solid" and "liquid" to the zero-order geometrical configuration is inverted.

We assume that, both before and in the deformed state, the liquid surface meets the sphere or cylinder perpendicularly, and that it is precisely a hemisphere or half-cylinder that projects above the surface. This could perhaps be achieved experimentally by having, instead of a sphere (cylinder), a body in the form of a hemisphere (half-cylinder) above, with a flange below, as shown in Fig. 3, and with a facility to adjust the height of the body. In this way, one could force artificially a "contact angle" of $\pi/2$, and the arrangement might be experimentally advantageous. For whereas the measurement of a maximum height of electrohydrostatic displacement would normally be optical, it would now be mechanical, as by a turn of a micrometer screw.

The theoretical implication of the assumption is that, in the present problems, the end conditions (12) are to be replaced by

$$x(y)\,(x(r))\to 0, \quad y\,(r)\to\infty, \quad x'(b)=0. \tag{33}$$

In the cylinder case, the first-order field is the same as for the uniform-field problem in the cylinder theory,[3] except that $E_{B,1}$ is now calculated at the plane liquid surface instead of at the cylinder, and we find that

$$E_{B,1}(y)=F(1-b^2/y^2). \tag{34}$$

We define the perturbation parameters

$$\beta_F=F\Phi/T, \quad \gamma_F=\epsilon F^2/8\pi T, \tag{35}$$

and with the use of (34) in Eq. (8), solving the equation and applying the conditions (33), we obtain for the first-order profile

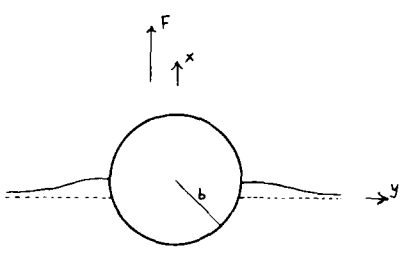


FIG. 2. A half-submerged conducting cylinder of sphere with ambient conducting liquid deformed under a uniform applied field.

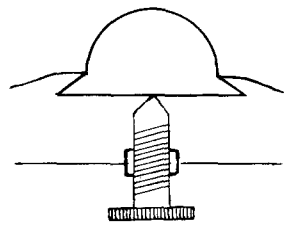


FIG. 3. Practical achievement of the conditions assumed in the half-submerged cylinder or sphere problem.

$$x_1(y)=\frac{1}{2}\gamma_F b^2\left\{\left(1-\frac{q^2b^2}{12}\right)[v(1;qy;0;+) + e^{q(b-y)}v(1;qb;0;-)] - \frac{b^2}{3y^2}+\frac{1}{3}\left(qb-\frac{10}{qb}\right)e^{q(b-y)}\right\}. \tag{36}$$

Using (34) in (15) and writing (36) as

$$x_1(y)=\gamma_F\,p(y), \tag{37}$$

we find for the second-order differential profile

$$x_{d,2}(y)=\beta_F\gamma_F\left\{q\left(e^{qy}\int_y^\infty e^{-qt}+e^{q(2b-y)}\int_b^\infty e^{-qt} + e^{-qy}\int_b^y e^{qt}\right)p(t)\left(1-\frac{b^2}{t^2}\right)dt - 2p(y)\left(1-\frac{b^2}{y^2}\right)\right\}. \tag{38}$$

Some reduction of the integrals here is possible, using (36) and (37), but with no computational advantage since the resulting expression is much longer while still containing integrals that cannot be formally evaluated.

In the sphere case, the first-order field is the same as for the uniform-field problem in the sphere theory,[3] except that again $E_{B,1}$ is taken at the plane liquid surface. Then, the theory is parallel to that above for the cylinder case, solving now the differential equations (9) and (16). In fact, we have in place of (34) and (36)–(38), the respective formulas

$$E_{B,1}(r)=F(1-b^3/r^3), \tag{39}$$

$$x_1(r)=\gamma_F\left(I_0(qr)\int_r^\infty K_0(qt) + \frac{K_0(qr)I_1(qb)}{K_1(qb)}\int_b^\infty K_0(qt) + K_0(qr)\int_b^r I_0(qt)\right)\left(1-\frac{b^3}{t^3}\right)^2 t\,dt, \tag{40}$$

$$x_1(r)=\gamma_F\,\tilde{p}(r), \tag{41}$$

$$x_{d,2}(r)=2\beta_F\gamma_F\left\{q^2\left(I_0(qr)\int_r^\infty K_0(qt) + \frac{K_0(qr)I_1(qb)}{K_1(qb)}\int_b^\infty K_0(qt) + K_0(qr)\int_b^r I_0(qt)\right)\tilde{p}(t)\left(1-\frac{b^3}{t^3}\right)t\,dt - \tilde{p}(r)\left(1-\frac{b^3}{r^3}\right)\right\}, \tag{42}$$

using the same perturbation parameters β_F, γ_F, given by (35).

IX. THE PECULIARITIES OF THE PLANE-SURFACE THEORY

We turn finally to the special features of the plane-surface theory which were mentioned cursorily in Sec. I and subsequently passed over tacitly.

Referring to the last paragraph of Sec. II, where the perturbation method was briefly described, it will be seen that the field effects *and gravity* were to be regarded as perturbations. This is indeed done in the sphere and cylinder theories,[3] but here, we have taken account of gravity exactly in the differential equations, treating only the field effects as perturbations. A consequence of this is that the equations are generally rather more difficult than those of the sphere and cylinder theories.

In fact, if in the present context we had treated gravity as a perturbation, we should have obtained absurd results. The gravity terms in (8) and (10) would be second-order ones and so drop out. We should then be left with equations whose only nonzero solutions are unbounded. In other words, either equation in the absence of both the electrical term and gravity, is *unstable*, so that to introduce both omitted terms as perturbations is a misapplication of perturbation theory.

We have here a sharp difference from what happens in the sphere and cylinder theories,[3] but it is in perfect accord with physical intuition. A drop of liquid in an equal-density liquid environment will have a perfectly spherical form as determined by surface tension. A slight inequality of density introduces a gravity effect small in relation to the surface tension and whose observable consequence is a slight flattening of the sphere. But a plane separating surface between two liquids of equal density is not determined by surface tension, and one can only imagine its being produced artificially. Then, an arbitrarily small disturbance will introduce curvature and so bring surface tension into play, whose effect, in the absence of gravity, will be to cause the disturbance to grow until, ultimately, the surface breaks up into globules. This emphasizes to us the fact that the mathematical instability is in complete correspondence with physical instability.

In the case of liquids of unequal density, a plane separating surface is the only one determined solely by gravity. It is stable, as represented by the dictum of elementary physics, "a liquid seeks its own level." In fact, if the surface is disturbed but the disturbance is not too large, surface tension will come into effect but will not prevail. More reasonable then it would be, perhaps, to treat *surface tension* as a perturbation, so that, with gravity treated exactly, the sphere–cylinder approach is in a sense inverted. But, since the surface-tension effect depends in the main upon the second derivative of the profile, it would be a singular perturbation, and as such, something known through differential-equation theory to be avoided. We think that the proper course in the plane-surface theory is the one we have taken, in which both surface tension and gravity are taken into account exactly, inasmuch as the approximation in the solution process relates only to the electrical perturbation parameters.

The other peculiarity of the plane-surface theory is deeper and more subtle, if not totally unrelated to the one just described. We recall the statement after the stress equations (7), to the effect that we cannot determine the function $N(x)$ in those equations. This, as we said, is of no consequence at the first order. In fact we find, attempting to determine the functions in (7), that

$$T_{A,xy} = T_{A,yx} = \Phi \frac{dE_{B,1}}{dy} + x_1' N(x_1), \tag{43}$$

so that, even if $N(x_1)$ is of order zero, we find when we substitute in the first of Eqs. (3) that, to the first order, the stresses $T_{A,xy}$ and $T_{A,yx}$ disappear.

Working to the second order, we still obtain no information concerning $N(x)$, and to secure its disappearance, we have to make the assumption, perhaps reasonable (but not mentioned), that (when $x = x_1$) it is a first-order quantity. At the third order, we are stopped. Our inability to determine $N(x)$ prevents our proceeding to this order. A completely parallel situation holds in the axisymmetric theory.

The point represents a real difficulty, one quite in contrast with the situation in the sphere and cylinder theories,[3] where, at each order, the stress system is fully determinable. Without its resolution, we must reckon with the possibility of the whole plane-surface perturbation theory being wrong. For it is quite clearly not enough that we are not precluded from going to the second order.

We have to examine closely the situation in which we were obliged to introduce a tensor like (7), which is not a multiple of the fundamental second-rank tensor δ_{ij} (the multiple being minus the familiar hydrostatic pressure). The latter form of tensor is suitable in ordinary hydrostatics and nonviscous hydrodynamics (and also in electrohydrostatics if we have no Φ effect). But the essential requirement of a stress tensor in fluid mechanics is that its divergence should give correctly the force density, and if another has this property and can be used consistently with other, more special, considerations, it will do just as well. This seems rarely to happen. It does, however, in the trivial hydrostatic problem concerning the form of a free liquid surface. This problem is solved by (7), taking $G(y) = 0$ and $N(x)$ as any differentiable function. But if the surface is not free, being constrained, say, by surface tension, this incompletely specified form does not suffice, and we have to take, as usual,

$$T_{xx} = T_{yy} = -p_0 + g\rho x, \quad T_{xy} = T_{yx} = 0. \tag{44}$$

This is what happens in a process of *exact solution*, and herein lies the clue to our difficulty. For in the perturbation theory, the determination of the stress tensor *at any stage* is by modification of that, insofar as it is known, at the *previous stage*. Thus, for example, at the second order, where we begin with the first-order *curved* surface, we are trying to find the second-order stress tensor by modifying a tensor which is insufficiently specified for a surface which is not plane; and yet there is no way out, for the first-order theory itself is one which perturbs the zero-order plane surface, for which, as we have seen, the stress tensor is not unique.

However, the first-order theory is unobjectionable in itself, and we propose an answer to the difficulty in the following way. Having got the first-order surface, we begin a new perturbation process in which this surface is the *zero-order* one. This means, according to the general electrohy-

drostatic theory, taking the coordinate system X^i [cf. Eqs. (1) and (2)] as one such that a surface $X^0 = \text{const}$ gives the new zero-order surface. It now appears that the stress tensor would be fully determinable at each stage, just as in the sphere and cylinder theories, for the essential thing which makes the stress tensor more determinable when the coordinates are "more curvilinear' is the presence of undifferentiated components when we write out in full Eqs. (2).

We are not suggesting that the process just outlined is one practicable to carry out; we are merely trying to establish the point of principle that there exists a solution as a perturbation series whose terms are, in principle, calculable. Assuming that we have done this, there remains one last question: is the assumption that $N(x_1)$ is of first order, tacitly used in the derivation of the second-order solutions in this paper, justified? We can use the suggestion we are making to show that it almost certainly is, which is for the present the only point of a practical nature that it achieves.

For recalling our remark before (44), to the effect that, in elementary hydrostatics, curvature of a surface removes the nonuniqueness of the stress tensor, we may assume that, in the perturbation process starting from our zero-order surface which is curved, the zero-order stress tensor is fixed. In Cartesian form and without a field-dependent term, it would be (44). To T_{yy} there must evidently be added a field term $Q(x, y)$, small to first order, whereby the $N(x)H(y)$ in (7) would, at the surface, be $g\rho x_1 + Q\{x_1(y), y\}$, a first-order quantity. This argument does not purport to be a proof, something beyond the scope of this paper and probably not simple. But pending a formal proof, it seems to make our conjecture highly reasonable.

To summarize, the perturbation method that we used is correct to, and only to, the second order. Strictly speaking, therefore, it is not perturbation theory at all, but is just an approximation method valid to the first order, and to the second order only by a justification which borrows from what would be a real perturbation theory.

[1]W. F. Brown, Jr., Am. J. Phys. **19**, 290 (1951).
[2]R. Cade, Proc. Phys. Soc. London B **67**, 689 (1954).
[3]R. Cade, J. Math. Phys. **23**, 2585 (1982).
[4]L. P. Eisenhart, *An Introduction to Differential Geometry* (Princeton U.P., Princeton, NJ, 1947).
[5]Previously obtained differential maximum-height formulas are (52) and (53) of Ref. 3, but each is wrong through lacking a factor b, as will be obvious on referring to the respective formulas (45) and (50) from which they are obtained. We are taking the present opportunity of making the correction.
[6]In fact, this statement is made in the classical context and makes no allowance for the potential jump which is itself inseparable from the presence of the double layer. However, it has been argued before (Ref. 17 contained in Ref. 3) that, for the purpose of the electrostatic problem, the classical concept is quite adequate.
[7]J. H. Jeans, *Electricity and Magnetism* (Cambridge U.P., Cambridge, 1960), 5th ed.
[8]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1944).
[9]*Tables of the Exponential Integral for Complex Arguments* (NBS, Washington, D.C., 1958).
[10]R. Cade, J. Appl. Phys. **37**, 4396 (1966). This paper contains a previous attempt at the plane-surface theory but which is generally wrong on account of errors in the basic considerations (see Ref. 3, Sec. 1). Fortuitously, however, Eqs. (8) and (10) of the present paper were obtained, so that the first-order results there for the line and point-charge problems are correct, and these problems are dealt with in considerably more detail than they are here. Nevertheless, all of the second-order results are wrong, with the exception of that for the second-order line-charge field, the theory of which depends only upon ordinary electrostatics applied with use of the correct first-order profile.
[11]W. R. Smythe, *Static and Dynamic Electricity* (McGraw–Hill, New York, 1950).
[12]W. E. Warren and R. E. Cuthrell, J. Appl. Phys. **46**, 4597 (1975).

A Lyapunov functional for the evolution of solutions to the porous medium equation to self-similarity. I

William I. Newman
Department of Earth and Space Sciences, University of California, Los Angeles, California 90024

(Received 30 November 1983; accepted for publication 4 May 1984)

We employ the transformation properties and the integral invariants of the porous medium equation in m dimensions to construct a Lyapunov functional which characterizes the evolution of the profile to its self-similar solution.

PACS numbers: 47.55.Mh, 02.30. + g, 44.10. + i, 52.25.Fi

I. INTRODUCTION

Many problems in physics are characterized by density-dependent diffusion. Some of these problems possess a conservation law and a unique, dimensionless combination of the spatial coordinate $\mathbf{x}$, the time t, and the conserved quantity. In such problems, there is substantial experimental, computational, and analytical evidence that the fluid variables lose memory of their initial behavior and relax to a state characterized by an unchanging or "self-similar" profile specified solely by the dimensionless combination present. The existence of this unique dimensionless scaling is the result of certain transformation properties in the origial partial differential equation. We propose to show, using the porous medium problem as an example, that these transformation properties permit the construction of a Lyapunov functional, a quantity which (loosely speaking) provides a measure of the information lost as the profile asymptotically approaches self-similarity.

As a paradigm for density-dependent diffusion, consider the following diffusion equation in the density ρ; namely,

$$\frac{\partial \rho}{\partial t} = \nabla \cdot [D(\rho) \nabla \rho] , \tag{1}$$

where $D(\rho)$ denotes a diffusivity or diffusion coefficient

$$D(\rho) = (N+1)\rho^N , \tag{2}$$

and where $N > 0$. These equations are frequently combined in the form

$$\frac{\partial \rho}{\partial t} = \Delta (\rho)^{1+N} , \tag{3}$$

where Δ denotes the Laplacian operator in m dimensions. From the latter equation, we identify two integral invariants, namely

$$M = \int_\Omega \rho \, d^m x \tag{4}$$

and

$$X = \int_\Omega \rho \, \mathbf{x} \frac{d^m x}{M} , \tag{5}$$

which we associate with the conservation of mass and the first moment (or center of mass) of the distribution. Without loss of generality, we shall assume that the center of mass $\mathbf{X}$ is situated at the origin.

Equation (3) describes the so-called "porous medium problem" [1] and is frequently encountered in hydrology and soil science. In such applications, one considers the flow of a fluid through an m-dimensional porous medium. The mass of the fluid is conserved and the fluid is assumed to obey a polytropic equation of state. Instead of the Euler force equation, the fluid is said to obey Darcy's law which provides for a velocity proportional to the pressure gradient. This in turn gives rise to a power-law dependence on the density by the diffusion coefficient. A suitable scaling of the density then yields the diffusivity (2).

The porous medium equation emerges in other physical contexts where a dissipative process employs a diffusion coefficient which varies as some power of the flow quantities. In plasma physics, the diffusion of heat introduces a diffusivity which varies as the $\frac{5}{2}$ power of temperature. In such instances, Eq. (3) is said to describe the "thermal wave problem." [2]

Density-dependent diffusion introduces a degeneracy into the parabolic character of the problem which manifests in a "boundary layer" near $\rho = 0$ which propagates at a finite speed. This, in turn, guarantees that an initial distribution with compact support (i.e., a spatially confined distribution) will preserve compact support. It has been shown that, if $\rho \geqslant 0$ initially, ρ will remain non-negative. These features are appealing from a physical standpoint, in contrast with the infinite propagation velocities peculiar to linear diffusion. Although these properties of nonlinear diffusion have been known empirically for a long time,[2] their rigorous derivation has been accomplished only recently.[3–5]

The apparent simplicity of the porous medium equation combined with its mass conservative property together with strictly dimensional considerations implies that there is a possible direct relationship between the spatial extent of the mass distribution and the time. Moreover, we might expect the effect of diffusion to erase all memory of the initial conditions (apart from the integral invariants) producing a distribution whose spatial characteristics, apart from a multiplicative scaling, remain unchanged in time. Indeed, such a "self-similar solution" exists and is described by an equation of the form

$$\rho(\mathbf{x},t) = g(t) F[\mathbf{x}/L(t)], \tag{6}$$

where $L(t)$ is a measure of the distribution's length scale which is assumed to vary as a power law in time, namely

$$L(t) = t^\gamma . \tag{7}$$

The presumption of self-similar, symmetric flow reduces the

 0022-2488/84/103120-04$02.50

partial differential equation to an ordinary differential equation in the variable x/L. (We will assume, for the moment, that the initial conditions are angle-independent so we need only explore radial variability in the distribution. Once we have derived the Lyapunov functional, we can trivially demonstrate its validity in the presence of any asymmetries.) Barenblatt and Zel'dovich[5,6] have argued that a self-similar solution is not a singular or degenerate solution but describes what they have called the "intermediate asymptotic" behavior of a system where the solution no longer depends on details of the initial or boundary conditions but is far from equilibrium. The porous medium equation possesses, using the terminology established by Barenblatt and Zel'dovich, a "self-similar solution of the first type," a solution where the index γ is determined from dimensional considerations and conservation laws. Bluman and Cole[7] have shown that the partial differential equation (3) is invariant under a continous Lie group of transformations that is in turn associated with the construction of the self-similar solution. Kamenomostsakaya[8] successfully employed these transformation group properties in demonstrating the evolution of the porous medium solution in one dimension to self-similar form. The existence of these transformation group properties has given rise to the idea that solutions to the porous medium problem are drawn or "attracted" to a self-similar state.

The characterization of self-similar flow as an attractor for porous medium flow has been demonstrated using other approaches. The existence of boundary layers which confine porous medium flow to a region of finite extent necessitates the classification of the solution to (3) as a moving boundary problem. In one dimension, Berryman[9] employed a (Lagrangian) transformation which transformed the original moving boundary problem into a fixed boundary problem. In so doing, he obtained a nonlinear Rayleigh–Ritz quotient and a Lyapunov functional which demonstrated the intermediate asymptotic character of the solution. Berryman's method, unfortunately, cannot be extended to describe asymmetric flows in higher dimensions. In the case where $N = 0$, the porous medium problem in m dimensions simplifies to the linear diffusion equation. Barenblatt and Zel'dovich[5] employed Green's function techniques in order to demonstrate the evolution of the distribution to self-similar form, a Gaussian profile. Most recently, Friedman and Kamin[4] have constructed a subsolution and a supersolution with a suitable comparison principle for the m-dimensional problem, providing the only *complete* demonstration of convergence to self-similar flow in higher dimensions. This approach, although rigorously correct, is especially complex and provides little insight into the mathematical or physical character of the diffusion process.

By visualizing nonlinear diffusion as an irreversible process where memory of the initial conditions are lost, we expect that a monotonically increasing Lyapunov functional (that describes the evolution of the flow to self-similarity) would exist. (The method of Lyapunov is not frequently encountered in application to partial differential equations. For a clear treatment of the Lyapunov functional and examples germane to hydrodynamic stability, the reader should consult Pritchard.[10]) The Lyapunov functional would occupy the same role as the entropy in thermodynamic problems. There, the entropy would invariably increase while the thermodynamic variables would relax to a state characterized by no more information content than that contained within the conserved quantities. We shall demonstrate a general procedure for employing the integral invariants as well as the transformation properties of the partial differential equation (associated with formulating the self-similar solution) in order to construct a Lyapunov functional. We shall show that self-similar flow is an attractor for this nonlinear diffusion process, just as a maximum entropy state is an attractor for irreversible thermodynamic processes. In particular, we shall demonstrate in a heuristic way that the solution to the porous medium equation converges asymptotically to a self-similar form. In the succeeding paper,[11] this demonstration will be performed in a rigorous way.

II. SELF-SIMILAR TRANSFORMATION

In order to understand why self-similarity is special, we must explore the transformation properties of the porous medium equation. It is well known[5,7] that the original partial differential equation must be invariant under a scale or homology transformation in order for a self-similar solution to exist. Defining a new spatial coordinate $\mathbf{x}'$ by

$$\mathbf{x}' = \eta \mathbf{x}\,, \tag{8}$$

where η is a positive, real-valued constant, the new density ρ' must scale according to

$$\rho' = \eta^{-m}\rho \tag{9}$$

in order to preserve the mass (4) in m dimensions under the transformation. The new time variable t' scales as

$$t' = \eta^{mN+2}\, t \tag{10}$$

in order for the transformed partial differential equation to be satisfied. This transformation permits us to generate a family of solutions to the porous medium equation (3) from a given solution, a family of solutions characterized by a "stretching" of the original solution. Under this transformation, only one combination of spatial coordinates and time remains unchanged, the quantity $\mathbf{y}$ defined by

$$\mathbf{y} = \mathbf{x}/t^{\alpha} = \mathbf{x}'/t'^{\alpha}\,, \tag{11}$$

where

$$\alpha = 1/(mN+2)\,. \tag{12}$$

By replacing the role of the variables $\mathbf{x}$ and $\mathbf{x}'$ by $\mathbf{y}$, we eliminate the spatial characterization that distinguished ρ' from ρ. In that sense, we see that the quantity t^{α} defines a natural length scale by which we normalize the spatial variable. Comparing the latter with (7), we associated α with γ and write

$$L(t) = t^{\alpha}\,. \tag{13}$$

In a self-similar solution, the density profile remains unchanged apart from a multiplicative scaling. Recalling Eq. (6), which we now write as

$$\rho(\mathbf{x},t) = g(t)\,F(\mathbf{y})\,, \tag{14}$$

Eq. (10) gives

$$g(t') = g(\eta^{mN+2}t)\,, \tag{15}$$

while Eq. (9) and (14) give

$$g(t') = \eta^{-m} g(t) . \tag{16}$$

Comparing the latter two equations, we obtain

$$g(t) \sim t^{-m/(mN+2)} = t^{-m\alpha} . \tag{17}$$

Assuming that the original distribution is symmetric, the self-similar solution $F(y)$ is now calculable as a solution to an ordinary differential equation, a feature discovered independently by Barenblatt[12] and by Pattle.[13] Since our objective is to obtain a Lyapunov functional for the partial differential equation, we will pursue an alternate approach predicted upon the original partial differential equation.

III. SELF-SIMILAR SOLUTION

As before, we shall assume that the center of mass **X** is situated at the origin (employing a translation of coordinates if necessary). We shall assume that the initial data $\rho(\mathbf{x},0) \geqslant 0$ have compact support. We shall momentarily ignore any angular variation (i.e., asymmetry with respect to the origin). Once we have derived the Lyapunov functional for symmetric flow, it will become apparent that the Lyapunov functional remains strictly monotonically increasing if the distribution is initially asymmetric.

For ease of calculation (with no loss of generality), we shall normalize x, t, and $\rho(x,t)$ so that the porous medium equation becomes

$$\frac{\partial \rho}{\partial t} = \frac{N\alpha}{2} \frac{1}{x^{m-1}} \frac{\partial}{\partial x} \left[x^{m-1} \rho^N \frac{\partial \rho}{\partial x} \right] , \tag{18}$$

and so that the mass (the first invariant M) is numerically equal to

$$M = \int_{|y|<1} (1-y^2)^{1/N} d^m y . \tag{19}$$

From the preceding discussion, we expect the self-similar solution to have the form $t^{-m\alpha} F(x/t^{\alpha})$, where F must be consistent with the mass normalization above. Recalling the definition (11) of y for $t > 0$, we can without loss of generality consider the behavior of $f(y,t)$ defined according to

$$\rho(x,t) = t^{-m\alpha} f(y,t) \tag{20}$$

in place of the density $\rho(x,t)$. The rationale for seeking a solution $f(y,t)$ instead of $\rho(x,t)$ is that, if the self-similar solution is an attractor for this dissipative system, then

$$\lim_{t \to \infty} f(y,t) = F(y) . \tag{21}$$

Since the latter equation is especially simple to test, we replace the roles of x and $\rho(x,t)$ by y and $f(y,t)$, quantities that we have established to be "natural" in the context of self-similar solutions. [Berryman's [9] Lagrangian approach employed another form of "natural" coordinate representation. However, that approach is inapplicable to asymmetrical, higher-dimensional problems.]

For convenience, we define a time variable τ according to

$$\tau = \ln t . \tag{22}$$

Transforming from x, t, and $\rho(x,t)$ to y, τ, and $f(y,\tau)$, Eq. (18) and (19) become (after a substantial amount of algebra)

$$\frac{\partial f}{\partial \tau} = \frac{\alpha}{2} \frac{1}{y^{m-1}} \frac{\partial}{\partial y} \left[y^{m-1} f \frac{\partial}{\partial y} (f^N + y^2) \right] \tag{23}$$

and

$$\int_{\Omega} f(y,\tau)\, d^m y = \int_{|y|<1} (1-y^2)^{1/N} d^m y . \tag{24}$$

Note that the right-hand side of (23) is just the radial term in $\nabla \cdot [f \nabla (f^N + y^2)]$, where f is assumed to depend on the *vector* **y**. This feature will permit the straightforward generalization of the Lyapunov functional for radially symmetric distributions to initially anisotropic distributions.

The self-similar solution for this problem emerges if the left-hand side of (23) vanishes and we replace the role of f by the self-similar solution F. Thus, $y^{m-1} F(d/dy)[F^N + y^2]$ must be constant. Since f and, hence, F have compact support, that constant must be zero (since our expression vanishes when F does). Thus, when F is nonzero,

$$F^N + y^2 = C, \tag{25}$$

where C is a real-valued, positive constant. Since F must be continuous (there is a discontinuity in the derivative of F which provides for a finite propagation speed), the latter equation must be connected to the solution $F = 0$. The normalization constraint (24) selects C to be unity and

$$F(y) = \begin{cases} (1-y^2)^{1/N} & |y| \leqslant 1, \\ 0, & |y| > 1. \end{cases} \tag{26}$$

Having obtained the similarity transformation and solution, we turn our attention to obtaining the Lyapunov functional.

IV. LYAPUNOV FUNCTIONAL

We assume a representation for the Lyapunov functional $H(\tau)$ of the form

$$H(\tau) = -\int_{\Omega} h(f,y)\, d^m y . \tag{27}$$

In particular, we seek a quantity H that is monotonically increasing, i.e.,

$$\frac{dH(\tau)}{d\tau} \geqslant 0 . \tag{28}$$

[In what follows, we shall assume that f has all necessary derivatives, a feature not enjoyed by the similarity solution. In the succeeding paper,[11] these technical difficulties will be addressed.] Evaluating the derivative of H, we see that

$$\frac{dH(\tau)}{d\tau} = -\int_{\Omega} \frac{\partial h(f,y)}{\partial f} \frac{\partial f}{\partial \tau} d^m y , \tag{29}$$

where $\partial h(f,y)/\partial f$ is the functional derivative of h with respect to f. Introducing (23) into (29) and integrating by parts (assuming that f has compact support so that total derivative terms vanish), we find

$$\frac{dH(\tau)}{d\tau} = \frac{\alpha}{2} \int_{\Omega} f \left[\frac{\partial}{\partial y} (f^N + y^2) \right] \left[\frac{\partial}{\partial y} \frac{\partial}{\partial f} h(f,y) \right] d^m y . \tag{30}$$

In order for H to be monotonically increasing, the integrand must be positive. Since $f \geqslant 0$, $(\partial^2/\partial y\, \partial f) h(f,y)$ must have the same sign as $(\partial/\partial y)(f^N + y^2)$. A particularly simple choice of $h(f,y)$ with this property is

$$h(f,y) = [1/(N+1)] f^{N+1}(y,\tau) + f(y,\tau) y^2 , \tag{31}$$

which provides

$$\frac{\partial h(f,y)}{\partial f} = f^N(y,\tau) + y^2 . \tag{32}$$

Then, Eq. (30) becomes

$$\frac{dH(\tau)}{d\tau} = \frac{\alpha}{2}\int_\Omega f\left[\frac{\partial}{\partial y}(f^N + y^2)\right]^2 d^m y \geqslant 0 \tag{33}$$

since the integrand is non-negative. Therefore, the Lyapunov functional

$$H(\tau) = -\int_\Omega \left[\frac{1}{N+1} f^{N+1}(y,\tau) + f(y,\tau)\, y^2\right] d^m y \tag{34}$$

is strictly increasing unless either f vanishes or $f^N + y^2$ is constant over the entire range [subject, of course, to the mass normalization equation (24)]. The latter, as we have already seen, only occurs when f corresponds *exactly* to the self-similar solution. The rate of convergence to the self-similar solution can be determined by evaluating $d \ln H(\tau)/d\tau$. The latter shows that convergence is fastest at early times and is achieved only in an asymptotic sense.

Finally, by generalizing (34) for $N > 0$ to read

$$H(\tau) = -\int_\Omega \left[\frac{1}{N+1} f^{N+1}(\mathbf{y},\tau) + f(\mathbf{y},\tau)\, y^2\right] d^m y \tag{35}$$

and by extending (23) to read

$$\frac{\partial f}{\partial \tau} = \frac{\alpha}{2}\nabla\cdot[f\nabla(f^N + y^2)] , \tag{36}$$

we readily observe that (35) is a Lyapunov functional for the porous medium equation with an initially asymmetric mass distribution.

For the sake of completeness, we note for the linear case ($N = 0$) that the partial differential equation for $f(y,\tau)$ is

$$\frac{\partial f}{\partial \tau} = \frac{1}{4}\nabla\cdot[f\nabla(\ln f + y^2)] \tag{37}$$

with a normalization condition of

$$\int_\Omega F(y)\, d^m y = \sqrt{\pi} . \tag{38}$$

The self-similar solution is observed to be

$$F(\mathbf{y}) = \exp(-y^2), \tag{39}$$

a Gaussian as we expect. The corresponding Lyapunov functional is

$$H(\tau) = -\int_\Omega [f(y,\tau)\ln f(\mathbf{y},\tau) + f(\mathbf{y},\tau)\, y^2]\, d^m y . \tag{40}$$

(Here, the $f \ln f$ term is reminiscent of the entropy of a distribution function where the $\int_\Omega f(y,\tau)\, y^2\, d^m y$ term corresponds to the temperature in the thermodynamic problem. In that case, the temperature would be used as an additional constraint.)

The method developed here for deriving a Lyapunov functional can be adapted for treating other nonlinear diffusion problems which manifest self-similarity. The procedure we have developed consists of three steps: (a) isolate the natural similarity variables; (b) express the partial differential equation in a manner that separates out the self-similar solution in a simple way as well as factors terms so that integration by parts is computationally simple; and (c) isolate a functional that exploits these properties and the factorization in order to provide a quantity which, like entropy, grows monotonically with time. To rigorously show that the monotonic character of the Lyapunov functional implies asymptotic convergence to self-similarity requires functional analytic arguments. These are given in the succeeding paper[11] by Ralston. Indeed, nonlinear diffusion problems possessing traveling wave solutions are amenable to this approach, a consequence of the relationship between self-similar and traveling-wave flows discovered by Barenblatt and Zel'dovich.[5] (In a separate paper,[14] we examine several nonlinear diffusion problems manifesting traveling wave solutions.)

ACKNOWLEDGMENTS

I wish to thank F. J. Dyson for suggesting the Lyapunov functional approach and thank M. D. Kruskal, D. G. Aronson, J. V. Ralston, C. G. Lange, L. Knopoff, F. H. Busse, and W. L. Kath for several useful discussions.

This research was supported by the U.S. Geological Survey Contract 14-08-0001-20529 and N.S.F. Grant EAR 81-16903.

[1] D. G. Aronson, in *Dynamics and Modelling of Reactive Systems*, edited by W. E. Stewart, W. H. Ray, and C. C. Conley (Academic, New York, 1980), pp. 161–176.
[2] Ya. B. Zel'dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Academic, New York, 1967).
[3] O. A. Oleinik, A. S. Kalashnikov, and Y-L. Czou, Izv. Akad. Nauk SSSR Ser. Mat **22**, 667 (1958) (in Russian).
[4] A. Friedman and S. Kamin, Trans. AMS **262**, 551 (1980).
[5] G. I. Barenblatt and Ya. B. Zel'dovich, Ann. Rev. Fluid Mech. **4**, 285 (1972).
[6] G. I. Barenblatt, *Similarity, Self-Similarity and Intermediate Asymptotics*, translated by N. Stein (Consultants Bureau, New York, 1979).
[7] G. W. Bluman and J. D. Cole, *Similarity Methods for Differential Equations* (Springer, New York, 1974).
[8] S. Kamenomostskaya (S. Kamin), Israel J. Math. **14**, 76 (1973).
[9] J. G. Berryman, J. Math. Phys. **21**, 1326 (1980).
[10] A. J. Pritchard, J. Inst. Maths. Appl. **4**, 78 (1968).
[11] J. V. Ralston, J. Math. Phys. **25**, 3124 (1984).
[12] G. I. Barenblatt, Prikl. Mat. Mekh. **16**, 67 (1952) (in Russian).
[13] R. E. Pattle, Quart. J. Mech. Appl. Math. **12**, 407 (1959).
[14] W. I. Newman, J. Theor. Biol. **104**, 473 (1983).

A Lyapunov functional for the evolution of solutions to the porous medium equation to self-similarity. II

James Ralston
Department of Mathematics, University of California, Los Angeles, California 90024

(Received 30 November 1983; accepted for publication 4 May 1984)

This article continues the analysis of the preceding article, showing that the Lyapunov functional introduced by Newman can be used to prove the stability of the Barenblatt–Pattle solutions of the porous medium equation.

PACS numbers: 47.55.Mh, 02.30. + g, 52.25.Fi, 44.10. + i

I. INTRODUCTION

The objective of this article is to show that the Lyapunov function introduced by Newman in the preceding article[1] can be used to prove that finite mass solutions of the porous medium equation tend to self-similar solutions as $t \to \infty$. Precisely stated, given $\phi \in L^1(R^m)$, $\phi \geqslant 0$, let $\rho(x, t)$ be the solution of the porous medium equation with $\rho(x,0) = \phi(x)$ (the precise meaning of "solution" here is explained in Sec. II). Then

$$\lim_{t \to \infty} \int_{R^m} |\rho(x, t) - \rho_s(x, t)| dt = 0, \tag{*}$$

where

$$\rho_s(x, t) = t^{-m\alpha}(a_\phi^2 - t^{-2\alpha}|x|^2)_+^{1/N},$$

a_ϕ is determined by

$$\int_{R^m} \phi \, dx = \int_{R^m} (a_\phi^2 - |x|^2)_+^{1/N} dx,$$

and $\alpha = (Nm + 2)^{-1}$. In the language of dynamical systems this says that similarity solutions are "stable attractors" and in what follows we refer to (*) as the stability theorem. We will follow the notation of Ref. 1 as much as possible.

If one restricts attention to bounded initial distributions ϕ, the stability theorem of Friedman and Kamin[2] gives more information than (*). However, the proof given here does not require the substantial regularity results for the porous medium equation used in Ref. 2.

II. EXISTENCE THEORY

We begin with a few results from the existence theory for quasilinear parabolic equations. Let $\phi(x)$ be a smooth nonnegative function of compact support; we will denote the space of such functions by $C^\infty_{0,+}(R^m)$. Then for $\epsilon > 0$ the initial value problem

$$\frac{\partial \rho}{\partial t} = c\Delta \rho^{1+N}, \quad \rho(x,0) = \phi(x) + \epsilon \tag{1}$$

has a unique classical solution such that ρ and its derivatives in x up to order 2 are bounded on $R^m \times [0,T]$ for all T. These solutions satisfy the maximum principle

$$\phi_1(x) + \epsilon_1 \geqslant \phi_2(x) + \epsilon_2 \Rightarrow \rho_1(x, t) \geqslant \rho_2(x, t).$$

These results can be obtained from [Ref. 3, Chap. V, §8] by first replacing (1) by

$$\frac{\partial \rho}{\partial t} = c\nabla \cdot h(\rho)\nabla \rho, \quad \rho(x,0) = \phi(x) + \epsilon, \tag{1'}$$

where h is a smooth, strictly positive function with $h(\rho) = (1 + N)\rho^N$ for $\rho \geqslant \epsilon$, and then using the maximum principle to conclude solutions of (1') solve (1).

For solutions of (1) one has the estimate of Volpert and Hudjaev [see formula (0.11) of Ref. 4] for $\lambda > 0$, $t, s \geqslant 0$,

$$\int_{R^m} |\rho_1(x, t + s) - \rho_2(x, t)| \omega_\lambda(x) dx \leqslant e^{k_\lambda t} \int_{R^m} |\rho_1(x, s) - \rho_2(x,0)| \omega_\lambda(x) dx, \tag{2}$$

where $\omega_\lambda(x) = \exp(-\lambda\sqrt{1 + |x|^2})$ and

$$k_\lambda = m(\lambda^2 + \lambda)(N + 1) \times \max\{(\sup \rho_1(x, s))^N, (\sup \rho_2(x,0))^N\}.$$

We define $T(t)\phi$ as the limit in $L^1_{\text{loc}}(R^m)$ of the solution $\rho_\epsilon(\cdot, t)$ of (1) as $\epsilon \to 0$. Estimate (2) and the results cited in the first paragraph imply that $T(t)$ extends to a contraction semigroup on $L^1_+(R^m) = \{ g \in L^1(R^m) : g \geqslant 0 \}$, i.e., for $t, s \geqslant 0$, $T(t + s) = T(t)T(s)$, and

$$\|T(t)g_1 - T(t)g_2\|_1 \leqslant \|g_1 - g_2\|_1.$$

It also follows that $g_1(x) \geqslant g_2(x)$, a.e. implies $[T(t)g_1](x) \geqslant [T(t)g_2](x)$, a.e. Moreover, the uniqueness argument of Ref. 4 (see §2 and 3) or the uniqueness theorem of Brezis and Crandall[5] [see Remark (1.22)], implies that for $g(x) = (a^2 - |x|^2)_+^{1/N}$, $T(t)g$ is the similarity solution

$$\rho_s(x, t) = (1 + t)^{-m\alpha}(a^2 - (1 + t)^{-2\alpha}|x|^2)_+^{1/N}.$$

Hence, if ϕ is a function in $C^\infty_{0,+}(R^m)$ and "a" is large enough that $(a^2 - |x|^2)_+^{1/N} \geqslant \phi(x)$ for $x \in R^m$, then $T(t)\phi$ is bounded by the similarity solution with initial data $(a^2 - |x|^2)_+^{1/N}$ for all $t \geqslant 0$. Since the integral of a similarity solution over R^m is constant in time, it follows from the contraction property that for $t > 0$,

$$\int_{R^m} T(t)\phi \, dx = \int_{R^m} \phi \, dx.$$

III. CONSEQUENCES OF EXISTENCE OF THE LYAPUNOV FUNCTIONAL

Since $T(t)$ is a contraction semigroup; and $C^\infty_{0,+}(R^m)$ is dense in $L^1_+(R^m)$, it will suffice to consider $T(t)\phi$ for $\phi \in C^\infty_{0,+}(R^m)$.

Let

$$H(\rho, t) = t^{m\alpha N}(\rho(x))^{1+N}/(1 + N) + t^{-2\alpha}|x|^2\rho(x).$$

Then

 0022-2488/84/103124-04$02.50

$$\int_{R^m} H(\rho, t)dx$$

is Newman's Lyapunov functional, written in terms of the original variables. For $\phi \in C^{\infty}_{0,+}(R^m)$ set $\rho = T(t)\phi$. Since ρ is bounded by a similarity solution, given $\psi \in C^{\infty}_0((0,\infty))$ we can choose $\chi \in C^{\infty}_0(R^m)$, such that $\chi(z) \equiv 1$ on the support of $\psi\rho$. Hence,

$$\int \frac{d\psi}{dt}\left(\int_{R^m} H(\rho, t)dx\right)dt$$
$$= \int_{R^{m+1}} \frac{d\psi}{dt}(t)\chi(x)H(\rho, t)dx\,dt$$
$$= \lim_{\epsilon \to 0} \int_{R^{m+1}} \frac{d\psi}{dt}(t)\chi(x)H(\rho_\epsilon, t)dx\,dt,$$

where ρ_ϵ is the (classical) solution to (1)—note that $\epsilon \leqslant \rho_\epsilon \leqslant \max\phi + \epsilon$ by the maximum principle. Integrating by parts and substituting from (1),

$$\int_{R^{m+1}} \frac{d\psi}{dt}\chi H(\rho_\epsilon, t)dx\,dt$$
$$= -\int_{R^{m+1}} \psi\chi\left(\frac{\partial H}{\partial \rho}(\rho_\epsilon, t)c\Delta\rho_\epsilon^{1+N} + \frac{\partial H}{\partial t}(\rho_\epsilon, t)\right)dx\,dt$$
$$= c\int_{R^{m+1}} \psi(\nabla\chi \cdot \nabla\rho_\epsilon^{1+N})\frac{\partial H}{\partial \rho}(\rho_\epsilon, t)dx\,dt$$
$$+ \int_{R^{m+1}} \psi\chi\left(c\nabla\frac{\partial H}{\partial \rho}\cdot\nabla\rho_\epsilon^{1+N} - \frac{\partial H}{\partial t}\right)dx\,dt$$
$$= -c\int_{R^{m+1}}\left[\psi\left(\Delta\chi\left(t^{m\alpha N}\left(\frac{N+1}{2N+1}\right)\rho_\epsilon^{2N+1}\right)\right.\right.$$
$$\left.\left. + (\nabla\cdot|x|^2\nabla\chi)t^{-2\alpha}\rho_\epsilon^{N+1}\right)\right]dx\,dt$$
$$+ \int \psi\chi(c\nabla(\rho_\epsilon^N t^{m\alpha N} + t^{-2\alpha}|x|^2)\cdot\nabla\rho_\epsilon^{1+N}$$
$$- \frac{m\alpha N}{1+N}t^{m\alpha N-1}\rho_\epsilon^{1+N} + 2\alpha t^{-2\alpha-1}|x|^2\rho_\epsilon)dx\,dt$$
$$= I_1 + I_2. \tag{3}$$

Since $\chi \equiv 1$ on the support of $\psi\rho$, $I_1 \to 0$ as $\epsilon \to 0$. Here I_2 contains the term

$$-\int_{R^{m+1}} \psi\chi\frac{m\alpha N}{1+N}\rho_\epsilon^{1+N}t^{m\alpha N-1}\,dx\,dt$$
$$= -\int_{R^{m+1}} \psi\chi\frac{\alpha N}{1+N}(\nabla\cdot x)\rho_\epsilon^{1+N}t^{m\alpha N-1}\,dx\,dt$$
$$= \int_{R^{m+1}} \psi\chi\frac{\alpha N}{(1+N)}x\cdot\nabla(\rho_\epsilon^{1+N})t^{m\alpha N-1}\,dx\,dt$$
$$+ \int_{R^{m+1}} \psi x\cdot\nabla\chi\frac{\alpha N}{1+N}\rho_\epsilon^{1+N}t^{m\alpha N-1}\,dx\,dt$$
$$= I_3 + I_4. \tag{4}$$

Again $I_4 \to 0$ and $\epsilon \to 0$, and substituting (4) into (3) we have

$$\int \frac{d\psi}{dt}\left(\int_{R^m} H(\rho, t)dx\right)dt$$
$$= \lim_{\epsilon\to 0}\frac{\alpha}{2}\int_{R^{m+1}} \psi\chi\left|\frac{N}{N+\frac{1}{2}}\nabla\rho_\epsilon^{N+1/2}t^{1/2-\alpha}\right.$$
$$\left. + 2t^{-\alpha-1/2}x\rho_\epsilon^{1/2}\right|^2 dx\,dt. \tag{5}$$

Here we have used $c = [N/(2N+2)][1/(mN+2)]$ and $m\alpha N = 1 - 2\alpha$.

Equations (3)–(5) correspond to the more transparent formal computations given by Newman in Ref. 1 to show that

$$\int H(\rho, t)dt$$

is a Lyapunov functional. We will exploit (5) by choosing $\psi \in C^{\infty}_0((\frac{1}{2},\infty))$ such that $\psi(t) \equiv 1$ on $[2,T]$, $|d\psi/dt| \leqslant 1$, and $(d\psi/dt)(t) \leqslant 0$ for $t > T$, and then choosing R so that the support of $\psi\rho$ is contained in $|x| < R$. Then it follows from (5) that

$$\int_{1/2}^2 dt\int_{R^m} H(\rho, t)dx$$
$$\geqslant \frac{\alpha}{2}\limsup_{\epsilon\to\infty}\int_2^T dt\int_{|x|<R}\left|\frac{N}{N+\frac{1}{2}}\right.$$
$$\left.\times\nabla(\rho_\epsilon^{N+1/2})t^{1/2-\alpha} + 2t^{-\alpha-1/2}x\rho_\epsilon^{1/2}\right|^2 dt. \tag{6}$$

Since $\epsilon \leqslant \rho_\epsilon \leqslant \max\phi + \epsilon$, (6) implies that

$$\int_2^T dt\int_{|x|<R}|\nabla\rho_\epsilon^{N+1/2}|^2dx \leqslant C,$$

where C may depend on R and T but not ϵ for $0 < \epsilon < 1$. Hence we can choose $\epsilon_i \downarrow 0$ so that $\nabla\rho_{\epsilon_i}^{N+1/2}$ converges weakly in $L^2(\{|x| < R\}\times[2,T])$. Since $\rho_\epsilon^{N+1/2}$ converges in norm to $\rho^{N+1/2}$ in $L^2(\{|x| < R\}\times[2,T])$, it follows that the weak limit of $\nabla\rho_{\epsilon_i}^{N+1/2}$ is $\nabla\rho^{N+1/2}$ in the sense of distributions. Thus, since norms are lower semicontinuous under weak limits in Hilbert space,

$$\int_{1/2}^2 dt\int_{R^m} H(\rho, t)dx\,dt$$
$$\geqslant \frac{\alpha}{2}\int_2^T dt\int_{|x|<R}\left|\frac{N}{N+\frac{1}{2}}\nabla(\rho^{N+1/2})t^{1/2-\alpha}\right.$$
$$\left. + 2t^{-\alpha-1/2}x\rho^{1/2}\right|^2 dx.$$

Since T is arbitrary and given T we can take R arbitrarily large, it follows that

$$\infty > \int_2^\infty dt\int_{R^m}\left|\frac{N}{N+1/2}\nabla_x(\rho^{N+1/2})t^{1/2-\alpha}\right.$$
$$\left. + 2t^{-\alpha-1/2}x\rho^{1/2}\right|^2 dx. \tag{7}$$

Equation (7) is all that we will need to prove stability.

At this point it is convenient to switch the variables used by Newman. We set $x = t^\alpha y$ and $f(y,t) = t^{m\alpha}\rho(t^\alpha y, t)$. Since ρ is bounded by a similarity solution we have

$$0 \leqslant f(y,t) \leqslant \frac{t^{m\alpha}}{(1+t)^{m\alpha}}\left(a^2 - \frac{t^{2\alpha}}{(1+t)^{2\alpha}}|y|^2\right)_+^{1/N}.$$

Thus for $t \geqslant 2$, $f(y,t)$ has support contained in a fixed ball $|y| < R_0$ and $0 \leqslant f(y,t) \leqslant a^{2/N}$. Moreover

$$\int_{R^m} f(y,t)dy = \int_{R^m}\rho(x,t)dx = \int_{R^m}\phi\,dx.$$

Changing variables in (7) we have

$$\int_2^\infty \frac{dt}{t}\int_{R^m}\left|\frac{N}{N+1/2}\nabla_y f^{N+1/2}+2yf^{1/2}\right|^2 dy<\infty.$$

For almost all $t, \nabla_y f^{N+1/2}(y,t)$ belongs to $L^2(R^m)$ by Fubini's theorem, and one shows directly from the definition of distribution derivative that, for almost all $t, \nabla_y f^{N+1/2}(y,t)$ is the distribution gradient of $f(y,t)$ considered as a function of y alone. Thus we can choose a sequence $t_j \uparrow \infty$ such that

$$\int_{R^m}\left|\frac{N}{N+1/2}\nabla_y f^{N+1/2}(y,t_j)+2yf^{1/2}(y,t_j)\right|^2 dy\to 0, \tag{8}$$

and ∇_y is the gradient to the sense of distributions.

IV. COMPLETION OF THE PROOF OF STABILITY

Setting $f_n(y)=f(y,t_n)$, we can formulate the final step in the proof of stability as a lemma.

Lemma: Suppose $0\leqslant f_n(y)\leqslant M$, support $f_n(y)$ is contained in $|y|<R$ and

$$\int_{R^m} f_n\,dy=c,$$

for all n. Then if

$$\int_{R^m}\left|\frac{N}{N+1/2}\nabla f_n^{1/2+N}+2yf_n^{1/2}\right|^2 dy\to 0,$$

there is a subsequence $\{f_{n_j}\}$ such that

$$\int_{R^m}|f_{n_j}-f_s|dy\to 0, \tag{9}$$

where $f_s=(a_c^2-|y|^2)_+^{1/N}$ and a_c is determined by $\int_{R^m} f_s\,dy=c$.

To see that the lemma implies stability note that

$$\int_{R^m}|\rho(x,t)-t^{-m\alpha}f_s(t^{-\alpha}x)|dx=\int|h(y,t)-f_s(y)|dy.$$

Thus, using the fact that $T(t)$ is a contraction, (9) implies

$$\lim_{t\to\infty}\int_{R^m}|\rho(x,t)-\rho_s(x,t)|dx=0,$$

where $\rho_s=t^{-m\alpha}f_s(t^{-\alpha}x)$.

Proof of Lemma: By the triangle inequality for the L^2-norm we have immediately

$$\int_{R^m}|\nabla_y f_n^{N+1/2}|^2+|f_n^{N+1/2}|^2\,dy<K,$$

for all n. Thus by Rellich's compactness theorem, $\{f_n^{N+1/2}\}$ is precompact in $L^2(y^2<2R^2)$. Thus, there is a subsequence $\{f_{n_j}\}$ such that

$$\int_{R^m}|f_{n_j}^{N+1/2}-g_\infty|^2dy\to 0.$$

Since taking a further subsequence, we may assume $f_{n_j}^{N+1/2}(y)\to g_\infty(y)$, a.e., defining $f_\infty(y)=(g_\infty(y))^{1/(N+1/2)}$, we have $f_{n_j}(y)\to f_\infty(y)$, a.e., and hence

$$\int_{R^m}|f_{n_j}(y)-f_\infty(y)|dy\to 0.$$

Now let $\phi(y)$ be a smooth function of compact support. Then for $1\leqslant k\leqslant m$,

$$0=\lim_{j\to\infty}\int_{R^m}\phi\left(\frac{N}{N+1/2}\frac{\partial f_{n_j}^{N+1/2}}{\partial y_k}+2y_k f_{n_j}^{1/2}\right)dy$$

$$=\lim_{j\to\infty}\int_{R^m}-\frac{\partial\phi}{\partial y_k}\left(\frac{N}{N+1/2}f_{n_j}^{N+1/2}-2y_k\phi f_{n_j}^{1/2}\right)dy.$$

Thus

$$0=\int_{R^m}\left(-\frac{\partial\phi}{\partial y_k}f_\infty^{N+1/2}+y_k\left(2+\frac{1}{N}\right)f_\infty^{1/2}\phi\right)dy. \tag{10}$$

Equation (10) implies that f_∞ is a function of $|y|^2$ alone. To see this let $R(t)$ be a rotation matrix depending smoothly on t and let $\phi(z)$ be a smooth function of compact support:

$$\frac{d}{dt}\left(\int_{R^m}\phi(R(t)y)f_\infty^{N+1/2}(y)dy\right)$$

$$=\int_{R^m}(\nabla_z\phi(R(t)y)\cdot R'(t)y)f_\infty^{N+1/2}(y)dy$$

$$=\int_{R^m}(\nabla_y\phi(R(t)y)\cdot R^T(t)R'(t)y)f_\infty^{N+1/2}(y)dy$$

$$=\int_{R^m}\left(\left(2+\frac{1}{N}\right)y\cdot R^T(t)R'(t)y\right)f_\infty^{1/2}\phi(R(t)y)dy$$

$$-\int_{R^m}(\nabla_y\cdot R^T(t)R'(t)y)f^{N+1/2}\phi(R(t)y)dy$$

$$=0,$$

since $R^T(t)R'(t)$ is skew-symmetric. Thus

$$\int_{R^m}\phi(R(t)y)f_\infty^{N+1/2}(y)dy=\int_{R^m}\phi(z)f_\infty^{N+1/2}(R^T(t)z)dz$$

is independent of t and, modifying f_∞ on a null set, we can assume it is a function of y^2 alone.

Equation (10) implies, taking $\phi=y_k\psi$

$$0=\int_{R^m}\Big(-y\cdot\nabla\psi)f_\infty^{N+1/2}$$

$$+\left(2+\frac{1}{N}\right)|y|^2\psi f_\infty^{1/2}-m\psi f_\infty^{N+1/2}\Big)dy. \tag{11}$$

Setting $\psi(y)=\lambda(|y|^2)$ and $u=|y|^2$, integrating (11) in polar coordinates yields

$$0=\int_0^\infty -\frac{d}{du}(u^{m/2}\lambda(u))f_\infty^{N+1/2}(u)$$

$$+\left(1+\frac{1}{2N}\right)u^{m/2}\lambda(u)f_\infty^{1/2}(u)du.$$

From this one concludes that $f_\infty^{N+1/2}(u)$ is absolutely continuous on $(1/k,k)$ for $k>0$ with derivative $-(1+1/2N)f_\infty^{1/2}(u)$. Since $f_\infty^{N+1/2}$ continuous and nonnegative implies $f_\infty^{1/2}$ continuous, $f_\infty^{N+1/2}$ is continuously differentiable and

$$\frac{df_\infty^{N+1/2}}{du}+\left(1+\frac{1}{2N}\right)f_\infty^{1/2}=0.$$

Hence, if $f_\infty(u_0)\neq 0$, $f_\infty^N(u)=a_0-u$ on $|u-u_0|<\delta$ for some $\delta>0$. Thus, since f_∞^N is continuous and

$$\int_{R^m}f_\infty(y)dy=c,$$

we conclude $f_\infty=f_s$ and the proof of the lemma is complete.

[1]W. I. Newman, J. Math. Phys. **25**, 3120 (1984).
[2]A. Friedman and S. Kamin, Trans. AMS 262, 551 (1980).
[3]O. Ladyzhenskaya, N. Ural'ceva and V. Solonnikov, *Linear and Quasi-linear Equations of Parabolic Type* (AMS Translations, Providence, 1968).
[4]A. I. Volpert and S. I. Hudjaev, Math. USSR-Sb. **7**, 365 (1969).
[5]H. Brezis and M. G. Crandall, J. Math. Appl. **58**, 153 (1979).

Low temperature expansions for the Gibbs states of quantum Ising lattice systems

Lawrence E. Thomas and Zhong Yin
Department of Mathematics, University of Virginia, Charlottesville, Virginia 22903

(Received 29 December 1983; accepted for publication 2 March 1984)

By conditioning on a subalgebra of observables, we develop low temperature expansions for certain quantum Ising models, which include transverse Ising models and anisotropic xy models. These expansions are majorized by the series for functions solving certain nonlinear partial differential equations. Convergence for the majorizing series follows from Cauchy–Kowalevski theory applied to the partial differential equations. The low temperature expansions imply long-range order.

PACS numbers: 64.60.Cn, 05.50. + q, 75.10.Jm, 02.30.Mv

I. INTRODUCTION

The intent of this article is to show that the thermodynamic states for certain quantum Ising models exhibit long-range order at sufficiently low temperature. Let H be a quantum Ising Hamiltonian defined by

$$H_\Lambda(\epsilon) = -\sum_{\substack{\langle i,j\rangle \\ i,j\in\Lambda}} \sigma^x(i)\sigma^x(j) - \epsilon \sum_{A\subset\Lambda} \hat{V}(A)\sigma^z(A). \tag{1.1}$$

Here, $\sigma^x(i)$, $\sigma^z(i)$ are Pauli spin matrices at site $i\in\Lambda\subset\mathbb{Z}^d$, $d\geqslant 2$, and $\sigma^{(\cdot)}(A) = \Pi_{i\in A}\,\sigma^{(\cdot)}(i)$, $A\subset\Lambda$. The first sum on the right-hand side of (1.1) is taken over nearest neighbor bonds and is simply the classical Ising Hamiltonian H_0. (For simplicity we confine attention to the nearest neighbor case; more general classical Ising Hamiltonians as well as other perturbations could be accommodated.) The second sum, which we treat as a perturbation, is a generalization of the transverse Ising term or xy model perturbations considered by numerous authors [see, e.g., Refs. 1–7 and references cited there]. The thermodynamic state at inverse temperature β is defined as a thermodynamic limit of

$$\rho_\Lambda(\cdot,\beta,\epsilon) = (\mathrm{tr}\exp(-\beta H_\Lambda(\epsilon)))^{-1}\,\mathrm{tr}((\cdot)\exp(-\beta H_\Lambda(\epsilon))).$$

We develop a low temperature expansion for the finite volume states ρ_Λ which is uniform in Λ. We then use this expansion to show, in Sec. III, that if the Hamiltonian is isotropic and translation invariant and if the sum for $V(i)$ defined by

$$V(i) \equiv \sum_{A\ni i} \frac{\hat{V}(A)}{|A|}\,\sigma^z(A) \tag{1.2}$$

consists of a finite number of terms, then there exist $\epsilon_c > 0$ and $\beta_c > 0$ such that if $|\epsilon| < \epsilon_c$ and $\beta > \beta_c$ long-range order in the variables $\{\sigma^x(A)\}$ obtains for the infinite volume thermodynamic state.

This result should be compared with that of Ginibre,[1] who also showed long-range order for a variety of quantum Ising models at low temperature by a Peierl's contour argument. The interactions Eq. (1.2) considered here, however, generalize some of those considered by Ginibre; moreover his estimates required $\epsilon\to 0$ as $\beta\to\infty$; no such condition is needed here. Our results should also be compared with those of Frölich and Lieb,[4] who used reflection positivity to establish long-range order at low temperature for a variety of quantum models; see also the work of Kirkwood[7] who showed using reflection positivity that, ϵ need not go to zero, as $\beta\to\infty$. Our methods do not require reflection positivity of the interaction. Finally, we remark that when ϵ is large the quantum fluctuations brought about by the perturbation term of the Hamiltonian can destroy long-range order; in Ref. 6, it is shown that in a variety of transverse Ising-like models, for ϵ sufficiently large, long-range order does not hold for any sufficiently low temperature.

The expansion for ρ_Λ which we develop involves two principal ideas, probabilistic conditioning of the state ρ_Λ, and then some simple nonlinear analysis to estimate the expansion. These same ideas were used in Ref. 6 (see also Ref. 5) with an important difference in the nonlinear analysis described below. Let σ denote a point in $\{-1,1\}^\Lambda$, i.e., a classical spin configuration, $\sigma(i)$ the coordinate of σ at site i. Let $|\sigma\rangle = \otimes_{i\in\Lambda}|\sigma(i)\rangle$, where $|\sigma(i)\rangle$ is an eigenvector of $\sigma^z(i)$ satisfying $\sigma^z(i)|\sigma(i)\rangle = \sigma(i)|\sigma(i)\rangle$. (The set of vectors $|\sigma\rangle$ forms a complete orthonormal set.) Suppressing the Λ subscript, we define $h(\sigma',\sigma,\tau,\epsilon)$ by

$$h(\sigma',\sigma,\tau,\epsilon) = \ln\langle\sigma'|e^{-\tau H(\epsilon)}|\sigma\rangle. \tag{1.3}$$

Then we condition ρ_Λ onto observables at "times" $0,\tau,2\tau,\ldots,(N-1)\tau$, $\tau = \beta/N$ for some suitable integer N; if $X_0(0), X_1(\tau),\ldots,X_{N-1}((N-1)\tau)$, are observables at these times then

$$\begin{aligned}
&\rho_\Lambda(X_0(0)X_1(\tau)\cdots X_{N-1}((N-1)\tau),\beta,\epsilon)\\
&\quad = \frac{1}{\mathbb{Z}}\,\mathrm{tr}(X_0(0)e^{-\tau H(\epsilon)}X_1(\tau)\\
&\qquad \times e^{-\tau H(\epsilon)}\cdots X_{N-1}((N-1)\tau)e^{-\tau H(\epsilon)})\\
&\quad = \frac{1}{\mathbb{Z}}\sum_{\sigma^0,\ldots,\sigma^{N-1}} X_0(0)\\
&\qquad \times e^{h(\sigma^0,\sigma^1,\tau,\epsilon)}\cdots X_{N-1}((N-1)\tau)e^{h(\sigma^{N-1},\sigma^0,\tau,\epsilon)}.
\end{aligned} \tag{1.4}$$

We will show that for τ appropriately large, h_τ can be expanded in a power series and that it has a suitably small density which implies long-range order. (Since $N\geqslant 1$ is arbitrary, all suitably large inverse temperatures, $\beta = \tau N$, give rise to long-range order.)

The power series for the local density of $h(\sigma',\sigma,\tau,\epsilon)$, in fact a double power series in powers of $e^{-2\tau}$ and ϵ, can be obtained in a straightforward manner. To prove that the series is convergent, we majorize the double series by the double series for a function which satisfies a nonlinear first-order

 0022-2488/84/103128-07$02.50

partial differential equation, see Eq. (2.24). The convergence of the double series for the majorizing function is assured by Cauchy–Kowalevski theory applied to this partial differential equation. (This analysis is described in a somewhat abstract setting in Sec. II.) The method for obtaining the majorizing function here should be contrasted with that in Refs. 5 and 6, where the majorizing function was simply the solution to an implicit nonlinear equation, rather than a partial differential equation. We also remark that the partial differential equation we obtain is complicated in appearance, reflecting the fact that the combinatorics of these quantum expansions are complicated. Fortunately, we are not concerned with solving the equation, but only with existence of a solution, in this case an immediate application of Cauchy–Kowalevski theory.

As a last remark, we note that Eq. (1.4) is the expectation for a periodic process of period $\beta = N\tau$.[8] But again our estimates are uniform in N, suggesting that the methods described here could be used to develop expansions for certain infinite particle stochastic processes. This work will be reported elsewhere.

II. INTEGRAL EQUATION FOR THE CONDITIONED CLASSICAL HAMILTONIAN

Define the conditioned classical Hamiltonian h by

$$h(\sigma',\sigma,\tau,\epsilon) = \ln \langle \sigma' | e^{-\tau H(\epsilon)} | \sigma \rangle. \tag{2.1}$$

The purpose of this section is to write out an integral equation for h and to investigate its analytic behavior. By differentiating Eq. (2.1) with respect to ϵ, we obtain

$$\begin{aligned}\frac{\partial}{\partial \epsilon} h(\sigma',\sigma,\tau,\epsilon) &= \exp(-h(\sigma',\sigma,\tau,\epsilon)) \\ &\quad \times \int_0^\tau \langle \sigma' | e^{-(\tau-t)H(\epsilon)} V e^{-tH(\epsilon)} | \sigma \rangle \, dt \\ &= \exp(-h(\sigma',\sigma,\tau,\epsilon)) \sum_{\sigma''} \langle \sigma' | e^{-\tau H(\epsilon)} | \sigma'' \rangle \\ &\quad \times \int_0^\tau \langle \sigma'' | e^{tH(\epsilon)} V e^{-tH(\epsilon)} | \sigma \rangle \, dt \\ &= \sum_{\sigma''} \exp(h(\sigma',\sigma'',\tau,\epsilon) - h(\sigma',\sigma,\tau,\epsilon)) \\ &\quad \times \langle \sigma'' | V_\tau(\epsilon) | \sigma \rangle,\end{aligned} \tag{2.2}$$

with

$$V_\tau(\epsilon) = \int_0^\tau e^{tH(\epsilon)} V e^{-tH(\epsilon)} \, dt. \tag{2.3}$$

Integrating Eq. (2.2) and suppressing the variable σ', we obtain our integral equation

$$\begin{aligned}h(\sigma,\tau,\epsilon) &= h_0(\sigma,\tau) \\ &\quad + \sum_{A \subset \Lambda} \int_0^\epsilon (\exp \partial_A h(\sigma,\tau,\epsilon')) \langle \sigma_A | V_\tau(\epsilon') | \sigma \rangle \, d\epsilon',\end{aligned} \tag{2.4}$$

where

$$h_0(\sigma,\tau) = h_0(\sigma',\sigma,\tau) = h(\sigma',\sigma,\tau,0), \tag{2.5}$$

and we use the notation

$$\sigma_A(i) = \begin{cases} \sigma(i), & \text{if } i \notin A; \\ -\sigma(i), & \text{if } i \in A; \end{cases} \tag{2.6}$$

and the differentiation with respect to the subset A is defined by

$$\partial_A f(\sigma) = f(\sigma_A) - f(\sigma). \tag{2.7}$$

The integral equation (2.4) determines $h(\sigma,\epsilon)$ as a formal power series in ϵ; successive Taylor coefficients of h are determined inductively by the equation. Our task is to investigate the convergence of the series.

We begin by casting the integral equation into a form more convenient for its analysis. Let

$$f(\sigma) = \sum_{A \subset \Lambda} \hat{f}(A) \sigma(A). \tag{2.8}$$

Then define the linear operator Γ by

$$\Gamma f = \sum_{A \neq \phi, \Lambda} |\partial A| \hat{f}(A) \sigma(A). \tag{2.9}$$

Here $|\partial A|$ denotes the surface area of the boundary of A, i.e., the area of the Peierl's contour associated with A. Let $b = \{i,j\}$ with i,j nearest neighboring sites. Then we note the following identity involving Γ and the set of all nearest neighbor bonds:

$$\Gamma f(\sigma) = -\frac{1}{2} \sum_b \partial_b f \tag{2.10}$$

which follows from the identity

$$\partial_b f(\sigma) = -2 \sum_{\substack{A \subset \Lambda \\ |b \cap A| = 1}} \hat{f}(A) \sigma(A). \tag{2.11}$$

Remark: The Hamiltonian which we are considering leaves the states $\{|\sigma\rangle$: an even number of $\sigma(i) = -1\}$ $[\{|\sigma\rangle$: an odd number of $\sigma(i) = -1\}]$ invariant. Consequently $\sigma(A) = \sigma(\Lambda - A)$ [resp. $\sigma(A) = -\sigma(\Lambda - A)$] on these invariant subspaces so that the Fourier coefficients of the functions encountered below are not independent. We do note that of course A and $\Lambda - A$ have the same boundary. For the sake of definiteness, given the choices of representatives $\sigma(A)$, $\sigma(\Lambda - A)$, we choose the one corresponding to the sets A or $\Lambda - A$ with the smaller cardinality; if $|A| = |\Lambda - A|$, either representative can be selected. Henceforth, we will consider ourselves as working on either the even or the odd subspaces (we will suppress a $\pm$ sign); a term such as $\hat{f}(\Lambda)\sigma(\Lambda)$ in Eq. (2.8) will then be absorbed into the $\hat{f}(\phi)$-term.

Using the identity Eq. (2.10), the integral equation can then be written

$$\begin{aligned}-\frac{1}{2} \sum_b \partial_b h(\sigma,\tau,\epsilon) &= \Gamma h_0(\sigma,\tau) \\ &\quad + \Gamma \sum_{A \subset \Lambda} \int_0^\epsilon (\exp \partial_A h(\sigma,\tau,\epsilon')) \langle \sigma_A | V_\tau(\epsilon') | \sigma \rangle \, d\epsilon'.\end{aligned} \tag{2.12}$$

Assume now that $h_0(\sigma,\tau)$ has a formal power series in a parameter λ. Then h has a formal double power series with $\epsilon^m \lambda^n$-coefficient h_{mn} given by

$$-\frac{1}{2}\sum_b \partial_b h_{mn}(\sigma,\tau)$$

$$= \delta_{m0}\Gamma h_{on}(\sigma,\tau) + \Gamma\sum_{i\in\Lambda}\sum_{j=0}^{m-1}\sum_{A\subset\Lambda}\frac{1}{m}P_{jn}(\partial_A h)\times\langle\sigma_A|V_\tau(i,m-1-j)|\sigma\rangle. \tag{2.13}$$

Here we use the notation that if $f(x,y)=\Sigma_{m,n} f_{mn}x^m y^n$ with $f_{00}=0$, $P_{mn}(f)$ is the polynomial in the f_{rs}'s, $0\leqslant r\leqslant m$, $0\leqslant s\leqslant n$, defined by

$$e^{f(x,y)} = \sum_{m,n} P_{mn}(f)x^m y^n. \tag{2.14}$$

Clearly the polynomial P_{mn} has positive coefficients. The operator $V_\tau(i)$ is defined by

$$V_\tau(i) = \int_0^\tau e^{tH(\epsilon)}V(i)e^{-tH(\epsilon)}\,dt; \tag{2.15}$$

$V_\tau(i,m)$ is the coefficient of ϵ^m in the Taylor expansion for $V_\tau(i)$.

Let us begin estimating $\partial_b h_{mn}$ in Eq. (2.13). First we note that by translation invariance and the isotropic property of the interaction,

$$\frac{1}{2}|\Lambda|d\,\|\partial_b h_{mn}(\cdot,\cdot,\tau)\|_0 = \frac{1}{2}\sum_b \big|\,|\partial_b h_{mn}(\cdot,\cdot,\tau)|\,\big|_0 = \Big|\,\Big|\frac{1}{2}\sum_b \partial_b h_{mn}(\cdot,\cdot,\tau)\Big|\,\Big|_0, \tag{2.16}$$

where the $\|\cdot\|_0$ norm of a function $f(\sigma',\sigma)=\Sigma_{A,B}\hat{f}(A,B)\sigma'(A)\sigma(B)$ is defined as $\|f\|_0=\Sigma_{A,B}|\hat{f}(A,B)|$. Recall that we are working on the invariant subspaces, so $\|\cdot\|_0$ may have different values on the two subspaces, but it does not affect our results. The latter equality in Eq. (2.16) follows by the closure property (see Ref. 6), which here means that if $f(\sigma)=\Sigma_A\hat{f}(A)\sigma(A)$ and A is a subset such that b and b' intersect the boundary of A, then the coefficient of the $\sigma(A)$-term in the Fourier expansion for $\partial_b f$ equals the coefficient in the Fourier expansion for $\partial_{b'} f$.

Equations (2.13) and (2.16) give us an estimate on $\partial_b h_{mn}$ for $m\geqslant 1$

$$\|\partial_b h_{mn}(\cdot,\cdot,\tau)\|_0 \leqslant \frac{2}{d}\sum_{A\subset\Lambda}\frac{1}{m}\times\sum_{j=0}^{m-1}\big|\,|\Gamma[P_{jn}(\partial_A h)\langle\sigma_A|V_\tau(i,m-1-j)|\sigma\rangle]\,|\,\big|_0. \tag{2.17}$$

Second we note the following property of the Γ-operation; if $f=f(\sigma',\sigma)$, $g=g(\sigma',\sigma)$ are two functions of σ' and σ, then

$$\|\Gamma(fg)\|_0 \leqslant \|\Gamma f\|_0\|g\|_0 + \|f\|_0\|\Gamma g\|_0. \tag{2.18}$$

Theorem 2.1: Assume there exist positive constants K, C_1, C_2, and δ such that

(i) For $m+n\geqslant 1$,

$$\|\Gamma P_{mn}(\partial_A h)\|_0 \leqslant K(m+n)P_{mn}(\|\partial_A h\|_0),$$

and h_{00} is independent of both σ' and σ. Here $P_{mn}(\|\partial_A h\|_0)$ denotes $P_{mn}(\|\partial_A h_{01}\|_0, \|\partial_A h_{10}\|_0, \ldots, \|\partial_A h_{mn}\|_0)$.

(ii) For each $A\subset\Lambda$, $i\in\Lambda$ and $m\geqslant 0$,

$$\|\Gamma\langle\sigma_A|V_\tau(i,m)|\sigma\rangle\|_0 \leqslant K(m+1)\|\langle\sigma_A|V_\tau(i,m)|\sigma\rangle\|_0.$$

(iii) $\sup_{A\subset\Lambda}\|\langle\sigma_A|V_\tau(i,m)|\sigma\rangle\|_0 \leqslant C_1^{m+1}$, for $m\geqslant 0$.

(iv) Let $\mathscr{A}(i,m)=\{A\subset\Lambda\,|\,\|\langle\sigma_A|V_\tau(i,m)|\sigma\rangle\|_0\neq 0\}$. Then the cardinality of $\mathscr{A}(i,m)$ is less than C_2^{m+1}, for $m\geqslant 0$.

(v) If $\langle\sigma_A|V_\tau(i,m)|\sigma\rangle\neq 0$, then A can be written as $A=b_1\Delta b_2\Delta\cdots\Delta b_r$ with $r\leqslant\delta(m+1)$ and b_j's are nearest neighbor pairs, for $m\geqslant 0$.

(vi) $\Sigma_{n=0}^\infty\|\partial_b h_{0n}\|_0|\lambda|^n<\infty$, for small $|\lambda|>0$.

Then there exists a $\gamma>0$ such that if $|\epsilon|<\gamma/(C_1C_2)$, $|\lambda|<\gamma$ then

$$\sum_{m,n=0}^\infty \partial_b h_{mn}\epsilon^m\lambda^n$$

is an absolutely convergent series in the sense that

$$\sum_{m,n}\|\partial_b h_{mn}\|_0|\epsilon|^m|\lambda|^n<\infty.$$

Proof: Applying Inequality (2.18) and the hypotheses above to the right-hand side of Inequality (2.17), we obtain, for $m>0$,

$$\begin{aligned}\|\partial_b h_{mn}\|_0 &\leqslant \frac{2K}{d}\sum_{A\subset\Lambda}\frac{m+n}{m}\times\sum_{j=0}^{m-1}P_{jn}(\|\partial_A h_{01}\|_0,\ldots,\|\partial_A h_{jn}\|_0)\times\|\langle\sigma_A|V_\tau(i,m-1-j)|\sigma\rangle\|_0\\ &\leqslant \frac{2K}{d}\frac{m+n}{m}\sum_{j=0}^{m-1}\sum_{A\subset\Lambda}P_{jn}(\delta(m-j)\|\partial_b h_{01}\|_0,\ldots,\delta(m-j)\|\partial_b h_{jn}\|_0)\times\|\langle\sigma_A|V_\tau(i,m-1-j)\sigma\rangle\|_0\\ &\leqslant \frac{2K}{d}\frac{m+n}{m}\sum_{j=0}^{m-1}P_{jn}(\delta(m-j)\|\partial_b h_{01}\|_0,\ldots,\delta(m-j)\|\partial_b h_{jn}\|_0)C_2^{m-j}C_1^{m-j}\\ &\leqslant \frac{2K}{d}\frac{m+n}{m}\sum_{j=0}^{m-1}P_{jn}(\delta(m-j)k_{01},\ldots,\delta(m-j)k_{jn})C_2^{m-j}C_1^{m-j}\equiv k_{mn},\end{aligned} \tag{2.19}$$

where k_{0n} is defined by $k_{0n}=\|\partial_b h_{0n}\|_0$ and all other k_{jn}'s are defined inductively by the last identity of (2.19). Then it suffices to show that there exists a $\gamma>0$ such that if $|\epsilon|<\gamma/(C_1C_2)$, $|\lambda|<\gamma$, then

$$\sum_{m,n}k_{mn}|\epsilon|^m|\lambda|^n<\infty. \tag{2.20}$$

Let $C=C_1C_2$. Multiplying λ^n on both sides of Eq. (2.19) and then summing over $0\leqslant n<\infty$, we obtain that for $m\geqslant 1$,

$$\begin{aligned}\sum_{n=0}^\infty\frac{2K}{d}\frac{m+n}{m}\sum_{j=0}^{m-1}P_{jn}(\delta(m-j)k_{00},\ldots,\delta(m-j)k_{jn})C^{m-j}\lambda^n &= \sum_{n=0}^\infty\frac{2K}{d}\Big(1+\frac{1}{m}\lambda\frac{\partial}{\partial\lambda}\Big)\sum_{j=0}^{m-1}P_{jn}(\delta(m-j)k_{00},\ldots,\delta(m-j)k_{jn})C^{m-j}\lambda^n\\ &= \frac{2K}{d}\Big(1+\frac{1}{m}\lambda\frac{\partial}{\partial\lambda}\Big)\sum_{j=0}^{m-1}P_j(\delta(m-j)k_0(\lambda),\end{aligned}$$

$$\ldots,\delta(m-j)k_j(\lambda))C^{m-j}$$
$$\equiv k_m(\lambda), \tag{2.21}$$

where, for all $j\geqslant 0$, $k_j(\lambda)=\sum_{n=0}^{\infty}k_{jn}\lambda^n$ and P_j is defined by

$$\exp\sum_{n=0}^{\infty}f_nx^n=\sum_{j=0}^{\infty}P_j(f_0,\ldots f_j)x^j, \tag{2.22}$$

(cf. Ref. 5). Multiplying ϵ^m on both sides of Eq. (2.21) and summing over $1\leqslant m<\infty$, we obtain

$$\sum_{m=1}^{\infty}k_m(\lambda)\epsilon^m\equiv k(\epsilon,\lambda)-k_0(\lambda)$$
$$=\sum_{m=1}^{\infty}\frac{2K}{d}\sum_{j=0}^{m-1}\Big[\epsilon^j(C\epsilon)^{m-j}P_j(\delta(m-j)k_0(\lambda),$$
$$\ldots,\delta(m-j)k_j(\lambda))+C\lambda\frac{\partial}{\partial\lambda}\left(\int_0^{\epsilon}\epsilon'^j(C\epsilon')^{m-j-1}\,d\epsilon'\right)P_j(\delta$$
$$\times(m-j)k_0(\lambda),\ldots,\delta(m-j)k_j(\lambda))\Big]. \tag{2.23a}$$

Let $n=m-j$; the sum on the right-hand side of Eq. (2.23) can be written as

$$\sum_{n=1}^{\infty}\sum_{j=0}^{\infty}\frac{2K}{d}\left(\epsilon^j(C\epsilon)^nP_j(\delta nk_0(\lambda),\ldots,\delta nk_j(\lambda))\right.$$
$$\left.+C\lambda\frac{\partial}{\partial\lambda}\left(\int_0^{\epsilon}\epsilon'^j(C\epsilon')^{n-1}\,d\epsilon'\right)P_j(\delta nk_0(\lambda),\ldots,\delta nk_j(\lambda))\right)$$
$$=\frac{2K}{d}\sum_{n=1}^{\infty}(C\epsilon)^n\exp\delta nk(\epsilon,\lambda)$$
$$+\frac{2K}{d}C\lambda\frac{\partial}{\partial\lambda}\int_0^{\epsilon}\sum_{n=1}^{\infty}(C\epsilon')^{n-1}\exp\delta nK(\epsilon',\lambda)d\epsilon'$$
$$=\frac{2K}{d}\frac{C\epsilon\exp\delta k(\epsilon,\lambda)}{1-C\epsilon\exp\delta k(\epsilon,\lambda)}$$
$$+\frac{2K}{d}C\lambda\frac{\partial}{\partial\lambda}\int_0^{\epsilon}\frac{\exp\delta k(\epsilon',\lambda)}{1-C\epsilon'\exp\delta k(\epsilon',\lambda)}\,d\epsilon'. \tag{2.23b}$$

Combining Eqs. (2.23a) and (2.23b) and differentiating them with respect to ϵ, we obtain

$$\left(1-\frac{2K}{d}\frac{C\epsilon\exp\delta k(\epsilon,\lambda)}{(1-C\epsilon\exp\delta k(\epsilon,\lambda))^2}\right)\frac{\partial k}{\partial\epsilon}$$
$$=\frac{2K}{d}\frac{C\exp\delta k(\epsilon,\lambda)}{(1-C\epsilon\exp\delta k(\epsilon,\lambda))^2}$$
$$+\frac{2K}{d}\frac{C\lambda\delta\exp\delta k(\epsilon,\lambda)}{(1-C\epsilon\exp\delta k(\epsilon,\lambda))^2}\frac{\partial k}{\partial\lambda}. \tag{2.24}$$

Let $\tilde\epsilon=C\epsilon$, $\tilde k(\tilde\epsilon,\lambda)=k(\epsilon,\lambda)$; then Eq. (2.24) is a differential equation of form

$$\frac{\partial\tilde k}{\partial\tilde\epsilon}=\mathscr{F}\left(\tilde\epsilon,\lambda,\tilde k,\frac{\partial\tilde k}{\partial\lambda}\right),$$

with $\mathscr{F}$ an analytic function of all four variables for small $\tilde\epsilon$ and λ. Then, by Cauchy–Kowalevski theory, it is solvable (cf. Ref. 9), and the solution $\tilde k(\tilde\epsilon,\lambda)$ is an analytic function of $\tilde\epsilon$ and λ at the origin provided that the initial value $k_0(\lambda)$ is analytic in λ. Thus we conclude that there exists a $\gamma>0$ such that if $|\epsilon|<(\gamma/C)$, $|\lambda|<\gamma$, then $\Sigma_{m,n=0}^{\infty}k_{mn}\epsilon^m\lambda^n$ converges absolutely. It follows that $\Sigma\partial_bh_{mn}\epsilon^m\lambda^n$ converges absolutely.

Corollary 2.2: Under the same hypotheses of Theorem 2.1, assume moreover that for some K', $\alpha\geqslant 0$

$$\|\Gamma^2\partial_bh_{mn}\|_0\leqslant K'(m+n)^{\alpha}\|\partial_bh_{mn}\|_0.$$

Then for some $C(\gamma)$ independent of $i,j\in\Lambda$,

$$\sum_{m,n}\|\partial_{\{i,j\}}h_{mn}\|_0|\epsilon|^m|\lambda|^n<C(\gamma),$$

for $C|\epsilon|$, $|\lambda|<\gamma$.

Proof: Let $\{b_1,b_2,\ldots,b_l\}$ be a path, i.e., b_k's are nearest neighbor bonds with $b_1=\{i,i_1\}$, $b_2=\{i_1,i_2\},\ldots,b_l=\{i_{l-1},j\}$ and $l=|i-j|$. Then we have

$$\|\partial_{\{i,j\}}h_{mn}\|_0=2\left\|\left|\sum_{\substack{A\ni i\\A\not\ni j}}\hat h_{mn}(A,\sigma')\sigma(A)\right|\right\|_0$$
$$+2\left\|\left|\sum_{\substack{A\ni j\\A\not\ni i}}\hat h_{mn}(A,\sigma')\sigma(A)\right|\right\|_0$$
$$\leqslant 2\left\|\left|\sum_{k=1}^{l}\sum_{\substack{A\ni i\\|A\cap b_k|=1}}{}'\hat h_{mn}(A,\sigma')\sigma(A)\right|\right\|_0$$
$$+2\left\|\left|\sum_{k=1}^{l}\sum_{\substack{A\ni j\\|A\cap b_{l-k}|=1}}{}'\hat h_{mn}(A,\sigma')\sigma(A)\right|\right\|_0$$
$$\leqslant 2\sum_{k=1}^{l}\left(\left\|\left|\sum_{\substack{A\ni i\\|A\cap b_k|=1}}{}'\frac{\Gamma^2\hat h_{mn}(A,\sigma')\sigma(A)}{|\partial A|^2}\right|\right\|_0\right.$$
$$\left.+\left\|\left|\sum_{\substack{A\ni j\\|A\cap b_{l-k}|=1}}{}'\frac{\Gamma^2h_{mn}(A,\sigma')\sigma(A)}{|\partial A|^2}\right|\right\|_0\right)$$
$$\leqslant\frac{1}{2}\sum_{k=1}^{l}\frac{1}{k^2}\left(\left\|\left|\sum_{\substack{A\ni i\\|A\cap b_k|=1}}{}'\Gamma^2\hat h_{mn}(A,\sigma')(A)\right|\right\|_0\right.$$
$$\left.+\left\|\left|\sum_{\substack{A\ni j\\|A\cap b_{l-k}|=1}}{}'\Gamma^2\hat h_{mn}(A,\sigma')\sigma(A)\right|\right\|_0\right)$$
$$\leqslant\frac{1}{8}\sum_{k=1}^{l}\frac{1}{k^2}\left(\|\Gamma^2\partial_{b_k}h_{mn}\|_0+\|\Gamma^2\partial_{b_k}h_{mn}\|_0\right)$$
$$\leqslant\frac{1}{8}\sum_{k=1}^{l}\frac{1}{k^2}2K'(m+n)^{\alpha}\|\partial_bh_{mn}\|_0$$
$$=K''(m+n)^{\alpha}\|\partial_bh_{mn}\|_0, \tag{2.25}$$

where Σ' denotes that the sum is over all A's such that $i\in A$ and k is the smallest integer for which $|A\cap b_k|=1(|A\cap b_{l-k}|=1)$, and the upper bound is uniform in i and j. Then the corollary follows from the fact that the series $\Sigma_{m,n}\|\partial_bh_{mn}\|_0\epsilon^m\lambda^n$ and the series $\Sigma_{m,n}\|\partial_bh_{mn}\|_0(m+n)^{\alpha}\epsilon^m\lambda^n$ have the same radius of convergence.

This corollary implies long-range order.

Theorem 2.3: Under the hypotheses of Theorem 2.1 and Corollary 2.2 long-range order holds in the sense that for $i,j\in\mathbf{Z}^d$,

$$\rho_{\Lambda}\{\sigma^x(i)\sigma^x(j),\beta,\epsilon\}\geqslant C_0(\beta,\epsilon)>0 \tag{2.26}$$

uniformly in Λ, for some $C_0(\beta,\epsilon)$.

Proof: We have that

$$\rho_\Lambda(\sigma^x(i)\sigma^x(j),\beta,\epsilon)$$

$$= \frac{1}{\mathbb{Z}} \sum_{\sigma^0,\ldots,\sigma^{N-1}} \exp(h(\sigma^0{}_{\{i,j\}},\sigma^1)$$

$$+ h(\sigma^1,\sigma^2) + \cdots + h(\sigma^{N-1},\sigma^0))$$

$$= \frac{1}{\mathbb{Z}} \sum_{\sigma^0,\ldots,\sigma^{N-1}} \exp(\partial_{\{i,j\}} h(\sigma^0,\sigma^1))\exp(h(\sigma^0,\sigma^1)$$

$$+ h(\sigma^1,\sigma^2) + \cdots + h(\sigma^{N-1},\sigma^0))$$

$$\geqslant \exp(-\|\partial_{\{i,j\}} h(\cdot,\cdot)\|_0)$$

$$\equiv C_0(\beta,\epsilon) \geqslant \exp(-C(\gamma)) > 0,$$

where we have used the symmetry of h, i.e., $h(\sigma^0,\sigma^1) = h(\sigma^1,\sigma^0)$.

III. SOME PROPERTIES OF h_0 AND V

A. h_0-term analysis

If $|\sigma'\rangle = \sigma^x(A)|\sigma\rangle$ with A a set with even cardinality, we have

$$e^{h_0(\sigma',\sigma,\tau)} = \langle\sigma'|e^{-\tau H(0)}|\sigma\rangle$$

$$= \langle\sigma|\sigma^x(A)e^{-\tau H(0)}|\sigma\rangle$$

$$= \langle\sigma^x(B)\sigma|\sigma^x(A)e^{-\tau H(0)}\sigma^x(B)|\sigma\rangle$$

$$= 2^{-|\Lambda|}\sum_{B\subset\Lambda}\langle\sigma^x(B)\sigma|\sigma^x(A)e^{-\tau H(0)}\sigma^x(B)|\sigma\rangle$$

$$= 2^{-|\Lambda|}\,\mathrm{tr}\,\sigma^x(A)e^{-\tau H(0)}$$

$$= 2^{-|\Lambda|}\,\mathrm{tr}\,\sigma(A)\exp\tau\sum_{\langle i,j\rangle}\sigma(i)\sigma(j), \tag{3.1}$$

where we have used the facts that $\sigma^x(A)$ is self-adjoint and that $\sigma^x(B)$ commutes with $H(0)$ and $\sigma^x(A)$, and we have made a unitary transformation to obtain the last equality. The latter expression can be written in a Peierl's contour representation;

$$\langle\sigma',e^{-\tau H(0)}\sigma\rangle = 2\cdot 2^{-|\Lambda|}e^{|\Lambda|d\tau}$$

$$\times\sum_{n\geqslant 1}\sum_{\{\Gamma_1,\ldots,\Gamma_n\}}\prod_{i=1}^{n}e^{-2\tau|\Gamma_i|}(-1)^{p(\Gamma_i)}, \tag{3.2}$$

where $p(\Gamma_i) = 0$ or 1 according to whether the number of sites of A inside (or outside) Γ_i is even or odd (recall that $|A|$ is even). If $\sigma(A)\equiv 1$, i.e., $A = \phi$, then, renormalizing $H(0)$ by adding a suitable constant, we have

$$h_0(\sigma,\sigma,\tau) = \ln\langle\sigma|e^{-\tau H(0)}|\sigma\rangle$$

$$= \sum_{n=1}^{\infty}\frac{1}{n!}\sum_{\{\Gamma_1,\ldots,\Gamma_n\}}\Phi^T(\Gamma_1,\ldots,\Gamma_n)\prod_{i=1}^{n}e^{-2\tau|\Gamma_i|}, \tag{3.3}$$

see Ref. 10, whereas if $A\neq\phi$, $e^{-2\tau|\Gamma_i|}$ is replaced by

$$e^{-2\tau|\Gamma_i|}(-1)^{p(\Gamma_i)} = e^{-2\tau|\Gamma_i|}\prod_{\substack{k\in B_i\\ \partial B_i=\Gamma_i}}\sigma(k)\sigma'(k),$$

where B_i is the set of sites inside Γ_i. Thus

$$h_0(\sigma',\sigma,\tau) = \ln\langle\sigma'|e^{-\tau H(0)}|\sigma\rangle$$

$$= \sum_{n=1}^{\infty}\frac{1}{n!}\sum_{\{\Gamma_1,\ldots,\Gamma_n\}}\Phi^T(\Gamma_1,\ldots,\Gamma_n)$$

$$\times\prod_{i=1}^{n}e^{-2\tau|\Gamma_i|}\prod_{\substack{k\in B_i\\ \partial B_i=\Gamma_i}}\sigma(k)\sigma'(k). \tag{3.4}$$

Moreover, the series for $\partial_b h_0(\lambda)$ is absolutely convergent for τ sufficiently large, i.e., for $\lambda\equiv e^{-2\tau}$ sufficiently small, and the convergence is uniform in Λ. In particular, for small λ,

$$\sum_{n=0}^{\infty}\|\partial_b h_{0n}\|_0|\lambda|^n < \infty.$$

We have already proved the following fact.

Lemma 3.1: There exists a $\gamma_1 > 0$, such that for $|\lambda| < \gamma_1$ and $A\subset\Lambda$,

$$\sum_{n=0}^{\infty}\|\partial_A h_{0n}\|_0\lambda^n < \infty,$$

and the convergence is uniform in Λ.

B. V-terms analysis

Now we investigate the properties of the potential V.

Definition 3.1: A collection of sets $B_0,B_1,\ldots,B_n$ is called an $(n+1)$-diagram starting at site i, if

(a) $i\in B_0$;

(b) $\hat{V}(B_j)\neq 0$, for $0\leqslant j\leqslant n$; and

(c) distance $(B_k, \cup_{j\neq k}B_j)\leqslant 1$, $0\leqslant k\leqslant n$.

An $(n+1)$-covering of a set B is an $(n+1)$-*diagram* $\{B_0,\ldots,B_n\}$ such that $B_0\,\Delta\,B_1\,\Delta\cdots\Delta\,B_n = B$.

Remark: We will not be concerned about the order among B_j's.

Lemma 3.2: Let D be the maximum of the diameters of sets B such that $\hat{V}(B)\neq 0$, s the number of sites within a distance $2D+1$ of a given site and t the number of terms in $V(i)$ in Eq. (1.2). Then the number of $(n+1)$-diagrams starting at site i is less than $s^{2n}(t+1)^{2n+1}$.

Proof: For each $(n+1)$-diagram starting at site i, choose $i_k\in B_k$, $1\leqslant k\leqslant n$. Re-enumerating them and perhaps counting some sites more than once, we obtain a new sequence of sites

$$j_0 = i, j_1 j_2,\ldots,j_{2n},$$

such that $|j_l - j_{l+1}| < 2D+1$, $0\leqslant l\leqslant 2n-1$. For $j_l = i_k$, we set

$$C_{j_l} = \begin{cases} B_k, & \text{if } j_m\neq i_k, \quad \text{for } 0\leqslant m\leqslant j_l - 1,\\ \phi, & \text{otherwise.}\end{cases}$$

Thus each $(n+1)$-diagram starting at site i corresponds to such a sequence. Since there are at most $s^{2n}(t+1)^{2n+1}$ such sequences for fixed site i, the number of $(n+1)$-diagrams starting at site i is less than $s^{2n}(t+1)^{2n+1}$.

We define $V_\tau(i,A,B,n)$ by

$$V_\tau(i) = \int_0^\tau e^{tH(\epsilon)}V(i)e^{-tH(\epsilon)}\,dt$$

$$= \sum_{n=0}^{\infty}\sum_{A,B,\subset\Lambda}\epsilon^n V_\tau(i,A,B,n)\sigma^x(A)\sigma^z(B). \tag{3.5}$$

Thus if $A\subset\Lambda$ and $i\in\Lambda$, we have

$$\langle\sigma_A|V_\tau(i,m)|\sigma\rangle = \sum_B V_\tau(i,A,B,m)\sigma(B).$$

Lemma 3.3: There exist positive constants K, C_1, and C_2 such that for each $m\geqslant 0$, the following estimates hold.

(i) For each $A\subset\Lambda$, $i\in\Lambda$, if $V_\tau(i,A,B,m)\neq 0$, then $|\partial B|\leqslant K(m+1)$. In particular

$$\|\Gamma^l\langle\sigma_A|V_\tau(i,m)|\sigma\rangle\|_0 \leqslant K^l(m+1)^l\|\langle\sigma_A|V_\tau(i,m)|\sigma\rangle\|_0, \quad l=1,2,....$$

(ii) $\sup_{A\subset\Lambda}\|\langle\sigma_A|V_\tau(i,m)|\sigma\rangle\|_0\leqslant C_1^{m+1}$.

(iii) Let $\mathscr{A}(i,m)=\{A \mid \|\langle\sigma_A|V_\tau(i,m)|\sigma\rangle\|_0\neq 0\}$. Then the cardinality of $\mathscr{A}(i,m)$ is less than C_2^{m+1}.

(iv) If $\|\langle\sigma_A|V_\tau(i,m)|\sigma\rangle\|_0\neq 0$, then A can be written as $A=b_1\,\Delta\,b_2\,\Delta\cdots\Delta\,b_r$ with $r\leqslant K(m+1)$ and b_j's are nearest neighbor bonds.

Proof: It is easy to see that

$$\int_0^\tau e^{tH(\epsilon)}V(i)e^{-tH(\epsilon)}\,dt = \sum_{m=0}^{\infty}\frac{\tau^{m+1}}{(m+1)!}[H(\epsilon),[...,[H(\epsilon),V(i)]\cdots]]. \tag{3.6}$$

From this expansion one sees that $V_\tau(i,A,B,n)=0$ unless $|B|\leqslant C_0(n+1)$, where C_0 is the maximum of the cardinalities of the terms of $V(i)$, Eq. (1.2).

Let $\mathbf{B}$ be a collection of n sets: $B_1,B_2,...,B_n$. Then we define

$$H(\mathbf{B},\epsilon)=H_0(\mathbf{B})+\epsilon{\sum_j}'\hat{V}(B_j)\sigma(B_j), \tag{3.7}$$

where $H_0(\mathbf{B})$ is the sum of all terms from $H(0)$ not commuting with some $\sigma^z(B_j)$ and Σ' denotes a sum over distinct B_j's in $\mathbf{B}$. Note that $H(\mathbf{B},\epsilon)$ has no more than $(\max_{1\leqslant j\leqslant n}|\partial B_j|+1)n$ terms, hence for $|\epsilon|\leqslant 1$, satisfies an estimate: $\|H(\mathbf{B},\epsilon)\|_0\leqslant a\,n$ with $a=\max_{\hat{V}(B)\neq 0}|\partial B|+\max_B|\hat{V}(B)|$.

Let $\mathscr{B}_{n+1}$ be the collection of all $(n+1)$-coverings of B. From Eqs. (3.5) and (3.6), we have

$$|V_\tau(i,A,B,n)| \leqslant \sum_{B_0\ni i}\sum_{\substack{\mathbf{B}=\{B_1,...,B_n\}\\ \{B_0,B_1,...,B_n\}\in\mathscr{B}_{n+1}}}|V_\tau(i,A,\mathbf{B},B_0,n)|, \tag{3.8}$$

where $V_\tau(i,A,\mathbf{B},B_0,n)$ is the $\epsilon^n\sigma^x(A)\sigma^z(B)$-coefficient of Eq. (3.5) but with $H(\epsilon)$ replaced by $H(\mathbf{B},\epsilon)$ and $V(i)$ replaced by $(\hat{V}(B_0)/|B_0|)\sigma^z(B)$. But by the Cauchy integral formula and the orthogonality of the $\sigma(A)$'s, we have

$$\begin{aligned}
&|V_\tau(i,A,\mathbf{B},B_0,n)|\\
&\equiv\frac{1}{2^{|\Lambda|}}\left|\sum_\sigma\sigma(B)\langle\sigma_A|\frac{1}{2\pi i}\oint_{|\epsilon|=1}\frac{d\epsilon}{\epsilon^{n+1}}\right.\\
&\quad\times\int_0^\tau e^{tH(\mathbf{B},\epsilon)}\frac{\hat{V}(B_0)}{|B_0|}\sigma^z(B_0)\\
&\quad\left.\times e^{-tH(\mathbf{B},\epsilon)}\,dt\,|\sigma\rangle\right|\\
&\leqslant\frac{1}{2\pi}\cdot\frac{1}{2^{|\Lambda|}}\sum_\sigma\oint_{|\epsilon|=1}\left|\frac{d\epsilon}{\epsilon^{n+1}}\right|\frac{|\hat{V}(B_0)|}{2an}e^{2\tau an}\\
&=\frac{|\hat{V}(B_0)|}{2an}e^{2\tau an}.
\end{aligned} \tag{3.9}$$

Thus

$$|V_\tau(i,A,B,n)|\leqslant s^{2n}(t+1)^{2n+1}\cdot t\cdot e^{2\tau an}. \tag{3.10}$$

It is evident that the number of sets B such that $V_\tau(i,A,B,n)\neq 0$ for some A is less than the number of all $(n+1)$-diagrams starting at site i, hence the number of such B's is bounded by $s^{2n}(t+1)^{2n+1}$ according to Lemma (3.2). Therefore,

$$\begin{aligned}
\sum_B|V_\tau(i,A,B,n)|&\leqslant s^{2n}(t+1)^{2n+1}\cdot s^{2n}(t+1)^{2n+1}\cdot te^{2\tau an}\\
&\leqslant(s^2(t+2)^2)^{2(n+1)}e^{2\tau a(n+1)}\equiv C_1^{n+1},
\end{aligned} \tag{3.11}$$

which proves (ii).

For a fixed set B, with $\mathbf{B}$ an $(n+1)$-covering, we have seen that $H_0(\mathbf{B})$ has at most $K(n+1)$ terms, $K=\max_{\hat{V}(B)\neq 0}|\partial B|$. Hence the number of $\sigma^x(A)$-factors which can be formed from these terms is less than $2^{K(n+1)}$. Taking into account the number of $(n+1)$-coverings, we obtain that the number of sets A for which $V_\tau(i,A,B,n)\neq 0$ is less than $C_A^{(n+1)}$, for some constant C_A. Recalling the fact stated above Inequality (3.11), we get (iii).

To get (i), note that if $\sigma(B)$ appears in the expansion of $\langle\sigma_A|V_\tau(i,m)|\sigma\rangle$, then $B=B_0\,\Delta\,B_1\,\Delta\cdots\Delta\,B_m$ for some $(m+1)$-covering of B. Since

$$|\partial B|\leqslant\sum_{j=0}^m|\partial B_j|\leqslant(m+1)\max_{\hat{V}(B)\neq 0}|\partial B|=(m+1)K,$$

(i) follows.

Now we prove (iv). If $\|<\sigma_A|V_\tau(i,m)|\sigma>\|_0$, then there exists a set B such that $V_\tau(i,A,B,m)\neq 0$. From Inequality (3.8) we have that $V_\tau(i,A,\mathbf{B},B_0,m)\neq 0$ for some $(m+1)$-covering of B, $\{B_0,B_1,...,B_m\}$, $\mathbf{B}=\{B_1,...,B_m\}$. Hence A is formed by some nearest neighbor bonds which cross the boundaries of the B_j's. The number of such bonds is bounded by $K(m+1)$.

Lemma 3.4: Let $A\subset\Lambda$ with even cardinality. If $m+n\geqslant 1$ and $P_{mn}(\partial_A h)=\Sigma_B\hat{P}_{mn}(\partial_A h,B)\sigma(B)_\Lambda$ and $\hat{P}_{mn}(\partial_A h,B)\neq 0$, then $|\partial B|\leqslant K(m+n)$. In particular

$$\|\Gamma^l P_{mn}(\partial_A h)\|_0\leqslant K^l(m+n)^l P_{mn}(\|\partial_A h\|_0), \quad \text{for } l=1,2,....$$

Proof: First from Eq. (3.4), we see that if $h_{on}(\sigma)=\Sigma\hat{h}_{on}(B)\sigma(B)$ and $\hat{h}_{on}(B)\neq 0$, then $|\partial B|=n$. Thus if $P_{on}(\partial_A h)=\Sigma_B\hat{P}_{on}(\partial_A h,B)\sigma(B)$ and $\hat{P}_{on}(\partial_A h,B)\neq 0$, then $|\partial B|\leqslant n$.

Now suppose $m\geqslant 1$. From Eq. (2.4), we have

$$h_{mn}(\sigma,\tau)=\sum_{i\in\Lambda}\sum_{j=0}^{m-1}\sum_{A\subset\Lambda}\frac{1}{m}\times P_{jn}(\partial_A h)\langle\sigma_A|V_\tau(i,m-l-j)|\sigma\rangle. \tag{3.12}$$

Assume that if $P_{kn}(\partial_A h)=\Sigma_B\hat{P}_{kn}(\partial_A h,B)\sigma(B)$, and $\hat{P}_{kn}(\partial_A h,B)\neq 0$, then $|\partial B|\leqslant K(k+n)$ for all $0\leqslant k\leqslant m-1$ and n. Then by (i) of Lemma 3.3 and Eq. (3.12) we have that if $h_{mn}(\sigma,\tau)=\Sigma_B\hat{h}_{mn}(B,\tau)\sigma(B)$ and $\hat{h}_{mn}(B,\tau)\neq 0$ then $|\partial B|\leqslant K(m+n)$. Therefore, by the definition of $P_{mn}(\partial_A h)$, if $P_{mn}(\partial_A h)=\Sigma_B\hat{P}_{mn}(\partial_A h,B)\sigma(B)$ and $\hat{P}_{mn}(\partial_A h,B)\neq 0$, then $|\partial B|\leqslant K(m+n)$.

Theorem 3.5: Let $H_\Lambda(\epsilon)=H_\Lambda(0)+\epsilon V$ with V defined by Eq. (1.2). Then $H_\Lambda(\epsilon)$ satisfies the hypotheses of Theorem 2.1, Corollary 2.2 and Theorem 2.3 so that long-range order holds, for $\beta>\beta_c$, $|\epsilon|\leqslant\epsilon_c$.

Proof: This follows from Lemmas 3.1, 3.2, 3.3, 3.4.

ACKNOWLEDGMENT

Work partially supported by NSF-MCS 74-07313-A03.

[1]J. Ginibre, "Existence of phase transitions for quantum lattice systems," Commun. Math. Phys. **14**, 205–234 (1969).
[2]D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
[3]R. J. Elliott, P. Pfeuty, and C. Wood, "Ising model with a transverse field," Phys. Rev. Lett. **25**, 443–446 (1970).
[4]J. Fröhlich and E. Lieb, "Phase transitions in anisotropic spin systems," Commun. Math. Phys. **60**, 233–267 (1978).
[5]J. R. Kirkwood and L. E. Thomas, "Expansions and phase transitions for the ground state of quantum Ising lattice systems," Commun. Math. Phys. **88**, 569–580 (1983).
[6]L. E. Thomas and Z. Yin, "Low temperature expansions for the Gibbs states of weakly interacting quantum Ising lattice systems," Commun. Math. Phys. **91**, 405–417 (1983).
[7]J. R. Kirkwood, "Phase transitions in the Ising model with transverse field," preprint, Institute for Mathematics and its Applications, University of Minnesota (1983).
[8]A. Klein and L. J. Landau, "Stochastic processes associated with KMS states," J. Funct. Anal. **42**, 368–428 (1981).
[9]V. I. Smirnov, *A Course of Higher Mathematics, IV* (Addison-Wesley, Reading, MA, 1964).
[10]C. Cammarota, "Decay of correlations for infinite range interactions in unbounded spin systems," Commun. Math. Phys. **85**, 517–528 (1982).

ERRATA

Erratum: Structure and motion of the Lee–Yang zeros [J. Math. Phys. 24, 2637 (1983)]

H. Nishimori and R. B. Griffiths
Department of Physics, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213

(Received 24 April 1984; accepted for publication 11 May 1984)

PACS numbers: 05.50. + q, 75.10.Hk, 99.10. + g

That not all zeros of the partition function of an Ising ferromagnet need move towards $\hat{z} = 1$ as the temperature is lowered was pointed out by C. Y. Weng in an unpublished thesis (Carnegie-Mellon, 1968) using an example of the sort we considered in Sec. V D with $n = 4$, and we want to point out the priority of his work.

Erratum: Simple calculation of Löwdin's alpha-function [J. Math. Phys. 25, 1133 (1984)]

Noboru Suzuki
Department of Materials Science, The University of Electro-Communications, Chofu-Shi, Tokyo 182, Japan

(Received 25 May 1984; accepted for publication 8 June 1984)

PACS numbers: 31.15. + q, 71.10. + x, 02.30.Mv, 99.10. + g

1. The 2's in front of the summation symbols in Eqs. (2.2), (2.6), and (3.14) and in front of the first summation symbol in Eq. (3.11) do not mean twice the respective sums, but that the summation indices run by step 2. Therefore, these 2's should be replaced by small 2's just in front of the summation symbols.

2. The small symbol 2 with the same meaning as above should be placed just in front of the first summation symbol in Eq. (3.17).

 0022-2488/84/103135-01$02.50