JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 13, NUMBER 5 MAY 1972

On the Evaluation of the Multiplicity-Free Wigner Coefficients of $U(n)$*

E. Chacón,† Mikael Ciftan,‡ and L. C. Biedenharn
Duke University, Durham, North Carolina 27706
(Received 30 June 1971)

An explicit algebraic expression [containing the minimal number $(n-1)$ sums] for the general reduced Wigner coefficients associated with the multiplicity free Kronecker product $[h_1 \cdots h_n] \times [p0 \cdots 0]$ of irreducible representations of $U(n)$, is determined. The calculation employs a combined use of recursive relations derived for the Wigner coefficients, and matrix elements (with respect to Gel'fand basis states) of a generator of $U(n)$ raised to an arbitrary power. We also give an alternative procedure using the techniques of the "pattern calculus." The method is illustrated first for the case of $U(3)$ and then generalized to arbitrary $U(n)$. It is found that the results can be expressed succinctly in terms of a new algebraic function S_{nm} — thereby elucidating the structural details of the underlying algebra.

1. INTRODUCTION

The applications of a symmetry group in physics are almost entirely subsumed under the reduction of the Kronecker (direct) product of unitary irreducible representations (irreps). This reduction forms the well-known Clebsch–Gordan series (in the space of irrep labels) and the explicit matrices (in the vector spaces of the irreps) which effect this reduction are termed Wigner coefficients, to accord with Wigner's original derivation[1] of these coefficients (in 1931) for $SU(2)$.

When the group $SU(3)$ became important in many fields of physics, Moshinsky[2] and later several authors[3–5] derived algebraic expressions for the special $SU(3)$ Wigner coefficients associated with the multiplicity-free Kronecker product

$$[h_1' h_2' h_3'] \times [p\ 0\ 0] \tag{1}$$

of irreps of $U(3)$.[6] Throughout this paper we shall speak of $U(n)$ and not distinguish $SU(n)$ except as the subset having $h_3' = 0$. We shall use Young frames $[h_1 h_2 \cdots h_n]$ to denote the irreducible representations (IR) of the group $U(n)$. Multiplicity-free Wigner coefficients are unique to within over-all phase. For $SU(3)$ there exists a canonical definition of the *general* (not necessarily multiplicity-free) Wigner coefficient, which is unique to within phase and equivalence under the Weyl group (this is the meaning we ascribe to "canonical").

The methods employed in the derivations of the above mentioned special $SU(3)$ Wigner coefficients were varied: in some cases the realization of the corresponding states by boson operators and the evaluation of a scalar product of boson polynomials[2]; in other cases, the solution of recursion relations.[3] The results obtained are also of a varied degree of complexity, ranging from expressions containing two sums to expressions with six summations for the $SU(3)$ reduced Wigner coefficient, also called isoscalar factor. [It is known[2] that the $U(n)$ Wigner coefficient can be factorized into a product of a $U(n-1)$ Wigner coefficient times a $U(n)$ reduced Wigner coefficient (RWC). (In $SU(3)$ the RWC is called "isoscalar factor" corresponding to isospin group ($SU2$) contained in $SU3$.) As it is supposed that the analysis proceeds in successive steps, once we know the $U(n-1)$ Wigner coefficients, it is the $U(n)$ RWC which should next be evaluated.]

In this paper we shall obtain an algebraic expression for the $U(3)$ RWC associated with the Kronecker product in (1) which is comparable in simplicity to other expressions hitherto published. Furthermore, the present analysis gives rise to simple rules that make the formula comprehensible. This clarification of structural details and the resulting algorithm allows one to readily generalize the results obtained for $U(3)$ to $U(n)$; in fact, in Secs. 3 and 4 of the present paper, we give algebraic expressions for the general Wigner coefficients of $U(n)$ associated with the Kronecker product

$$[h_1 \cdots h_n] \times [p0 \cdots 0].$$

In a notation which is an adaptation of the usual $SU(2)$ notation to the case when we use Gel'fand labels for the states, the coefficient we shall evaluate is

$$\left\langle \begin{array}{ccccccc} h_1' & & h_2' & & h_3' & p\ \ 0\ \ 0 & \\ & q_1' & & q_2' & ; & q\ \ 0 & \\ & & q_1' & & & 0 & \end{array} \middle| \begin{array}{ccc} h_1 & & h_2 \quad h_3 \\ & q_1 & q_2 \\ & & q_1' \end{array} \right\rangle . \tag{2}$$

Since this expression denotes a complete $U(3)$ Wigner coefficient, the general $U(3)$ RWC will be the quotient of the coefficient in (2) by the $U(2)$ Wigner coefficient

$$\left\langle \begin{matrix} q_1' & q_2' \\ q_1' \end{matrix} \; ; \; \begin{matrix} q & 0 \\ 0 \end{matrix} \middle| \begin{matrix} q_1 & q_2 \\ q_1' \end{matrix} \right\rangle = \sqrt{\frac{(q_1 - q_2 + 1)\cdot(q_1' - q_2')!\,q!}{(q_1 - q_2' + 1)!\,(q_2 - q_2')!}}\,(\delta_{q_1+q_2,q_1'+q_2'+q}) \qquad (3)$$

It greatly helps toward clarifying the content of our expressions if we make use of the operator notation discussed by Louck in his review paper.[7] In this notation the Wigner coefficient (2) is written as

$$\left\langle \begin{matrix} h_1 & h_2 & h_3 \\ q_1 & q_2 \\ q_1' \end{matrix} \middle| \left\langle \begin{matrix} & t & \\ s & & 0 \\ p & 0 & 0 \\ q & 0 & \\ & 0 & \end{matrix} \right\rangle \middle| \begin{matrix} h_1' & h_2' & h_3' \\ q_1' & q_2' \\ q_1' \end{matrix} \right\rangle,$$

where $t = h_1 - h_1'$, $s - t = h_2 - h_2'$, $p - s = h_3 - h_3'$, or in more general notation

$$\left\langle \begin{matrix} \overline{m}_{13} & \overline{m}_{23} & \overline{m}_{33} \\ \overline{m}_{12} & \overline{m}_{22} \\ \overline{m}_{12} \end{matrix} \middle| \left\langle \begin{matrix} & \Gamma_{11} & \\ \Gamma_{12} & & 0 \\ M_{13} & 0 & 0 \\ M_{12} & & 0 \\ & 0 & \end{matrix} \right\rangle \middle| \begin{matrix} m_{13} & m_{23} & m_{33} \\ m_{12} & m_{22} \\ m_{12} \end{matrix} \right\rangle$$

The symbol between the initial and final states denotes a "Wigner operator" which carries two patterns sharing common IR labels $[M_{13}\ 0\ 0]$. The lower pattern is the ordinary Gel'fand pattern, while the upper one (written in inverted form) is formally identical to another Gel'fand pattern. The weight components Γ_{11}, $\Gamma_{12} - \Gamma_{11}$, $M_{13} - \Gamma_{12}$ of the upper pattern denote the changes that the operator effects upon the initial state representation labels m_{13}; furthermore the changes that the Wigner operator produces upon the initial $U(2)$ and $U(1)$ labels are obtained using the projection

$$\left\langle \begin{matrix} & \Gamma & \\ M_{13} & 0 & 0 \\ & M_{12} & 0 \\ & & 0 \end{matrix} \right\rangle = \sum_{\gamma} \left[\begin{matrix} & \Gamma & \\ M_{13} & 0 & 0 \\ & M_{12} & 0 \\ & \gamma & \end{matrix} \right] \left\langle \begin{matrix} & \gamma & \\ M_{12} & 0 & \\ & 0 & \end{matrix} \right\rangle.$$

The operator

$$\left\langle \begin{matrix} & \gamma & \\ M_{12} & 0 & \\ & 0 & \end{matrix} \right\rangle$$

is the $U(2)$ Wigner operator, and the operator in square brackets is the $U(3)$ reduced Wigner operator. If this $U(3)$ Wigner operator is taken between *maximal* initial and final $U(2)$ state labels, then $\gamma = 0$, $\overline{m}_{12} = m_{12}$, $\overline{m}_{22} = m_{22} + M_{12}$. The condition $\gamma = 0$ eliminates the summation.

A distinctive feature of our approach is that the coefficient in (2) appears as a linear combination of coefficients with $q = 0$, these latter which have a very simple (monomial) structure are easy to evaluate either by a recursive process or by the arrow pattern calculus[5]; we shall discuss both methods. To obtain the Wigner coefficient (2) from the $q = 0$ case, one needs the matrix elements (ME) of powers of certain generators of $U(3)$; we shall again discuss two alternative methods to obtain these ME, one using the Wigner–Eckart theorem, and the other employing the pattern calculus rules. The detailed derivation of the coefficient in (2) will be carried out in Sec. 2. It will be seen that the analysis of Sec. 2 is generalizable to $U(n)$. In Sec. 4 we discuss this generalization and show that the algebraic formula for the RWC associated with the Kronecker product $[h_1' h_2' \ldots h_n'] \times [p0 \ldots 0)$ of IR of $U(n)$ contains $(n-1)$ sums. In Sec. 3 we discuss the structure of those special coefficients in (2) which have $q = 0$.

2. DERIVATION OF THE $U(3)$ RWC

A very fruitful technique in the evaluation of Wigner operators—for both practical as well as theoretical purposes—is the method of embedding in larger unitary groups; this general procedure has a long history of success in both the Frobenius and the Young treatment of the symmetric group. Similar methods were first given by Louck[8] and by Brody, Moshinsky, and Renero,[9] and also independently by Biedenharn, Giovannini, and Louck.[10] The latter paper showed in particular that the embedding of $U(n)$ in $U(n^2)$ is optimal in two senses: (1) All other cases can be exhibited as special cases, and (2) there exists a *factorization lemma*, explicitly involving the properties *of the canonical* splitting of the multiplicity.

It follows from Refs. 9 and 10 that the Wigner coefficient in (2) is identical to the scalar product

$$\left\langle 0 \left\| \begin{pmatrix} & h_1' & \\ h_1' & & h_2' \\ & q_1' & \end{pmatrix}^{\dagger} \begin{pmatrix} & 0 & \\ 0 & & 0 \\ p & & 0 \\ & q & \end{pmatrix}^{\dagger} \begin{pmatrix} & h_1' & \\ h_1 & & h_2 \\ & q_1 & \end{pmatrix} \right\| 0 \right\rangle, \tag{4}$$

where each symbol

$$\left\langle \begin{pmatrix} & z & \\ x & & y \\ a & b & c \\ & t & \end{pmatrix} \right\| 0 \rangle$$

stands for a state belonging to the symmetric IR $[a+b+c, 0 \cdots 0]$ of $U(9)$, and classified by a subgroup $U^u(3) \times U^l(3)$ of $U(9)$; the state being a Gel'fand state

$$\left| \begin{matrix} a & b & c \\ & x & y \\ & & z \end{matrix} \right\rangle$$

with respect to transformations by the (upper) $U^u(3)$ group with generators $C^{\mu\nu}$, and also a Gel'fand state

$$\left| \begin{matrix} a & b & c \\ & s & t \\ & & r \end{matrix} \right\rangle$$

with respect to transformations by the (lower) $U^l(3)$ group with generators $\mathcal{C}_{\mu\nu}$. [For the time being we set $h_3' = 0$ in (4). There is no loss of generality in doing this; at the end of this section we shall mention how to obtain the general result with $h_3' \neq 0$ from the result of the case $h_3' = 0$].

Explicit realizations of the $U(3) \times U(3)$ states, as well as of the generators $C^{\mu\nu}, \mathcal{C}_{\mu\nu}$ in terms of boson creation and annihilation operators

$$a_\mu^\nu, \ \bar{a}_\mu^\nu \quad \mu, \nu = 1, 2, 3 \tag{5}$$

are well known.[11] In particular we shall quote the following result that is needed later:

$$\begin{pmatrix} & 0 & \\ 0 & & 0 \\ p & & 0 \\ q & & 0 \\ & 0 & \end{pmatrix} = \sqrt{\frac{(p-q)!}{p!q!}}\, \mathcal{C}_{23}^{q} \begin{pmatrix} & 0 & \\ 0 & & 0 \\ p & & 0 \\ 0 & & 0 \\ & 0 & \end{pmatrix}. \tag{6}$$

Now, let us denote provisionally any three operators like those appearing in the scalar product (4) by B_1, B_2, B_3, respectively. Then, if $\mathcal{L}$ is a $U(3)$ generator, we have the identity

$$\langle 0 | B_1^\dagger (\mathcal{L} B_2)^+ B_3 | 0 \rangle = \langle 0 | B_1^\dagger B_2^\dagger \mathcal{L}^+ B_3 | 0 \rangle - \langle 0 | (\mathcal{L} B_1)^+ B_2^\dagger B_3 | 0 \rangle;$$

and iterating $(q-1)$ times this result we arrive at

$$\langle 0 | B_1^\dagger (\mathcal{L}^q B_2)^+ B_3 | 0 \rangle = \sum_\alpha (-)^\alpha \binom{q}{\alpha} \langle 0 | \mathcal{L}^\alpha B_1)^+ B_2^\dagger (\mathcal{L}^+)^{q-\alpha} B_3 | 0 \rangle. \tag{7}$$

When we use this formula in Eq. (4), with the expression given in (6) substituted for the middle state, we come to the conclusion that the scalar product in (4) is equal to

$$\binom{p}{q}^{-1/2} \sum_\alpha \frac{(-)^\alpha}{\alpha!(q-\alpha)!} \times \left\langle 0 \left\| \left\{ \mathcal{C}_{23}^{\alpha} \begin{pmatrix} & h_1' & \\ h_1' & & h_2' \\ & q_1' & \end{pmatrix}^{\dagger} \right\} \begin{pmatrix} & 0 & \\ 0 & & 0 \\ p & & 0 \\ 0 & & 0 \\ & 0 & \end{pmatrix}^{\dagger} \right\| \mathcal{C}_{32}^{q-\alpha} \begin{pmatrix} & h_1' & \\ h_1' & & h_2' \\ h_1 & & h_2 \\ & q_1 & \end{pmatrix} \right\| 0 \right\rangle. \tag{8}$$

The matrix elements (ME) of the generators of $U(3)$ with respect to Gel'fand states are well known.[12] Using repeatedly these ME we can calculate the ME of any power of a generator. This, however, is not a very practical procedure; so we prefer to use the fact[13] that

$$T_\tau^k = \sqrt{\frac{(2k)!}{(k+\tau)!(-\tau+k)!}}\, (\mathcal{C}_{13})^{\tau+k} (\mathcal{C}_{23})^{k-\tau}, \qquad \tau = k, k-1, \ldots, -k \tag{9}$$

are the components of an $SU(2)$ irreducible tensor of rank k; using the Wigner–Eckart theorem we can calculate its ME with respect to $U(3)$ Gel'fand states, and specialize the result to $\tau = -k$.

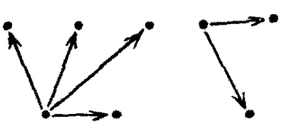
The techniques of the pattern calculus afford a very economical derivation of the desired result for the operator $(\mathcal{C}_{23})^\alpha$.

Consider the operator $\mathcal{C}_{23}$ itself. Using the known matrix elements one finds that

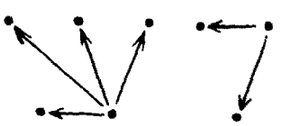
$$\mathcal{C}_{23} = \left\langle \Delta = \begin{matrix} 0 & 0 & 0 \\ & 1 & 0 \\ & & 0 \end{matrix} \right\rangle + \left\langle \Delta = \begin{matrix} 0 & 0 & 0 \\ & 0 & 1 \\ & & 0 \end{matrix} \right\rangle, \tag{10}$$

where the two operators given correspond to the patterns

$$\Delta: \begin{pmatrix} 0 & 0 & 0 \\ & 1 & 0 \\ & & 0 \end{pmatrix} \Rightarrow \tag{11a}$$



$$\Delta: \begin{pmatrix} 0 & 0 & 0 \\ & 0 & 1 \\ & & 0 \end{pmatrix} \Rightarrow \tag{11b}$$



with each operator giving $U(3) : U(2)$ and $U(2) : U(1)$ contributions. The algebraic factors corresponding to the patterns of the first operator are

$$\left(\frac{(p_{12}+1-p_{13})(p_{12}+1-p_{23})(p_{12}+1-p_{33})}{(p_{12}+1-p_{22})} \times \frac{(p_{12}-p_{11})}{(p_{12}-p_{22})}\right)^{1/2}, \quad (12a)$$

while those of the second operator are

$$\left(\frac{(p_{22}+1-p_{13})(p_{22}+1-p_{23})(p_{22}+1-p_{33})}{(p_{22}+1-p_{12})} \times \frac{(p_{22}-p_{12})}{(p_{22}-p_{11})}\right)^{1/2}. \quad (12b)$$

Using the fact that the numerators[14] of the two operators

$$\Delta = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & \\ & 0 & \end{pmatrix} \text{ and } \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & \\ & 0 & \end{pmatrix}$$

commute, we find

$$(\mathcal{C}_{23})^{\alpha} = \sum_{\beta} \frac{\alpha!}{\beta!(\alpha-\beta)!} \left\langle \Delta = \begin{matrix} 0 & 0 & 0 \\ \alpha-\beta & \beta & \\ & \alpha & \end{matrix} \right\rangle_R D_{\alpha-\beta,\beta}, \quad (13)$$

where

$$D_{\alpha-\beta,\beta} = [(p_{12}-p_{22})(p_{12}-p_{22}+\alpha-2\beta)]^{1/2} \times \frac{\Gamma(p_{12}-p_{22}-\beta)}{\Gamma(p_{12}-p_{22}+\alpha-\beta+1)}. \quad (14)$$

(This and other similar results will be discussed in more detail in a future paper.)

$(D_{\alpha-\beta,\beta})^2$ is a sum of reciprocals of rational factors, the sum being expressed in Eq. (14) as a monomial in a closed algebraic form; it arises from the denominator contributions.[14] Similarly,

$$\left\langle \begin{matrix} 0 & 0 & 0 \\ \Delta: \quad \alpha-\beta & \beta & \\ & 0 & \end{matrix} \right\rangle_R$$

arises from numerator contributions[14] and it is the analog of the $U(2)$ renormalized operators,[14] the ME of which can be obtained directly by the use of the arrow pattern calculus rules.

By any of the two alternative methods, we find

$$\mathcal{C}_{23}^{\alpha} \begin{pmatrix} & h_1' & \\ h_1' & & h_2' \\ h_1' & h_2' & 0 \\ q_1' & & q_2' \\ & q_1' & \end{pmatrix} = \alpha! \sum_{\beta} G_{\beta} \begin{pmatrix} & h_1' & \\ h_1' & & h_2' \\ h_1' & & h_2' & 0 \\ q_1'+\beta & & q_2'+\alpha-\beta \\ & q_1' & \end{pmatrix} \quad (15)$$

with

$$G_{\beta} = \frac{\sqrt{(q_1'-q_2'+1)(q_1'-q_2'+2\beta-\alpha+1)}}{\beta!(\alpha-\beta)!} \times \frac{(q_1'-q_2'+\beta-\alpha)!}{(q_1'-q_2'+1+\beta)!} \frac{\sqrt{\dfrac{(h_1'-q_1')!(h_2'-q_2')!(h_1'-q_2'+1)!}{(q_1'-h_2')!(q_2')!(q_1'+1)!}}}{\sqrt{\dfrac{(h_1'-q_1'-\beta)!(h_2'-q_2'-\alpha+\beta)!(h_1'-q_2'-\alpha+\beta+1)!}{(q_1'-h_2'+\beta)!(q_2'+\alpha-\beta)!(q_1'+1+\beta)!}}} \frac{\sqrt{\dfrac{\beta!}{(q_1'-q_2'-\alpha+\beta)!}}}{\sqrt{\dfrac{1}{(q_1'-q_2')!}}}, \quad (16)$$

$$\mathcal{C}_{32}^{q-\alpha} \begin{pmatrix} & h_1' & \\ h_1' & & h_2' \\ h_1 & h_2 & h_3 \\ q_1 & & q_2 \\ & q_1' & \end{pmatrix} = (q-\alpha)! \sum_{\gamma} E_{\gamma} \begin{pmatrix} & h_1' & \\ h_1' & & h_2' \\ h_1 & & h_2 & & h_3 \\ q_1-\gamma & & q_2-q+\alpha+\gamma \\ & q_1' & \end{pmatrix}, \quad (17)$$

with

$$E_{\gamma} = \frac{\sqrt{(q_1-q_2+1)(q_1-q_2+1+q-2\gamma-\alpha)}}{\gamma!(q-\alpha-\gamma)!} \times \frac{(q_1-q_2-\gamma)!}{(q_1-q_2+1+q-\alpha-\gamma)!} \times \frac{\sqrt{\dfrac{(q_1-q_1')!}{(q_1'-q_2)!}}}{\sqrt{\dfrac{(q_1-q_1'-\gamma)!}{(q_1'-q_2+q-\alpha-\gamma)!}}} \times \frac{\sqrt{\dfrac{(h_1-q_1+\gamma)!(h_2-q_2+q-\alpha-\gamma)!(h_1-q_2+q+1-\alpha-\gamma)!}{(q_1-h_2-\gamma)!(q_2-h_3-q+\alpha+\gamma)!(q_1-h_3+1-\gamma)!}}}{\sqrt{\dfrac{(h_1-q_1)!(h_2-q_2)!(h_1-q_2+1)!}{(q_1-h_2)!(q_2-h_3)!(q_1-h_3+1)!}}}. \quad (18)$$

In formulas (16) and (18), as well as in all the basic formulas in the rest of the paper, we can achieve a great deal of simplicity in notation if we use two algebraic functions that will be defined now. Let $(h_1 \cdots h_n)$ be a set of Gel'fand labels of an IR of $U(n)$, and let $(q_1 \cdots q_m)$ be another set of Gel'fand labels of an IR of $U(m)$, with $m = n$ or $m = n-1$, these being the only two cases that interest us. Furthermore, let us suppose that the labels h_s, q_s obey the "betweenness" relations: $h_s \geq q_s \geq h_{s+1}$ for all s. Then make the following definition:

$$S_{nm}(h_1 \cdots h_n; q_1 \cdots q_m) \equiv \sqrt{\frac{\prod_{k=1}^{m}\prod_{s=1}^{k}(h_s - q_k + k - s)!}{\prod_{k=1}^{n-1}\prod_{s=k+1}^{n}(q_k - h_s + s - k - 1)!}}, \quad m = n \text{ or } n-1. \tag{19}$$

As a particular case of this definition, notice that $[S_{nn}(h_s; h_s)]^2$ is related to the dimension $\Delta_n(h_1 \cdots h_n)$ of the IR $(h_1 \cdots h_n)$ of $U(n)$; in fact

$$[S_{nn}(h_1 \cdots h_n; h_1 \cdots h_n)]^2 = \prod_{i<j=2}^{n}(h_i - h_j + j - i) = \prod_{k=1}^{n} k!\,\Delta_n(h_1 \cdots h_n). \tag{20}$$

It will be found that all the basic results of this paper can be expressed very neatly in terms of S_{nm} functions and their squares. This was the reason for writing Eq. (16) and (18) in that peculiar way, as it is easy to verify that each square root in those formulas is an S_{3m} or S_{2m} function, and the rational part is an $[S_{22}]^2$, so we can rewrite G_β and E_γ as

$$G_\beta = \frac{S_{22}(q'_s; q'_s)S_{22}(\bar{q}'_s; \bar{q}'_s)}{[S_{22}(\bar{q}'_s; q'_s)]^2} \frac{S_{21}(\bar{q}'_s; q'_1)}{S_{21}(q'_s; q'_1)} \frac{S_{32}(h'_s; q'_s)}{S_{32}(h'_s; \bar{q}'_s)} \tag{21}$$

with $\bar{q}'_1 \equiv q'_1 + \beta$, $\bar{q}'_2 \equiv q'_2 + \alpha - \beta$ and

$$E_\gamma = \frac{S_{22}(q_s; q_s)S_{22}(\bar{q}_s; \bar{q}_s)}{[S_{22}(q_s; \bar{q}_s)]^2} \frac{S_{21}(q_s; q'_1)}{S_{21}(\bar{q}_s; q'_1)} \frac{S_{32}(h_s; \bar{q}_s)}{S_{32}(h_s; q_s)} \tag{22}$$

with $\bar{q}_1 \equiv q_1 - \gamma$, $\bar{q}_2 \equiv q_2 - q + \alpha + \gamma$.

We shall next describe a rule by means of which one can write very easily the different terms that occur in the definition (19) of the S_{nm} functions. We shall state the rule for the case $n = 3$, the generalization to arbitrary n should be obvious.

Draw a row of $n = 3$ dots, and assign to each dot the labels h_1, h_2, h_3, respectively. Below this row of dots draw a second row, of either $m = 3$ dots [case (a)] or $m = 2$ dots [case (b)], and assign to them the labels q_1, q_2, q_3 in case (a) or q_1, q_2 in case (b). As it is supposed that $h_s \geq q_s \geq h_{s+1}$, we shall draw the dots in the second row in such a way that the set of labels resemble a Gel'fand pattern, i.e., the dot associated to q_s should be drawn between the dots corresponding to h_s and h_{s+1}. (See the two diagrams below.) Next, draw an arrow from every dot

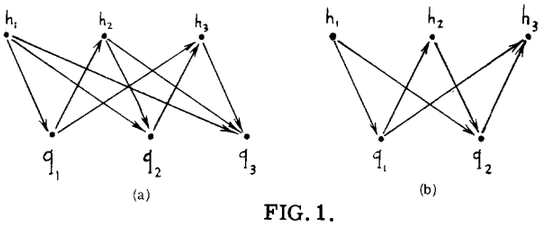


FIG. 1.

in one row to every dot in the other row, the direction of the arrow being such that the final dot is always to the right of the initial dot. To each arrow associate a factorial $(t - h + \epsilon)!$, where t is the label at the tail of the arrow, h is the label at the head of the arrow, and ϵ is the number of dots, in any single row, which are located to the right of the initial dot and to the left of the final dot. Multiply together all these factorials and distribute them in a fraction, with arrows going downwards giving numerator contributions, and arrows going upwards giving denominator contributions. Finally, take the positive square root of the fraction.

Following this rule one can easily verify that with Fig. 1(a) is associated the function

$$\sqrt{\frac{(h_1 - q_1)!(h_1 - q_2 + 1)!(h_1 - q_3 + 2)!(h_2 - q_2)!(h_2 - q_3 + 1)!(h_3 - q_3)!}{(q_1 - h_2)!(q_1 - h_3 + 1)!(q_2 - h_3)!}},$$

and with Fig. 1(b) is associated the function

$$\sqrt{\frac{(h_1 - q_1)!(h_1 - q_2 + 1)!(h_2 - q_2)!}{(q_1 - h_2)!(q_1 - h_3 + 1)!(q_2 - h_3)!}},$$

which are precisely $S_{33}(h_s; q_s)$ and $S_{32}(h_s; q_s)$ according to definition (19).

Substituting the last results in (8) we see that the scalar product in (4) is equal to

$$\binom{p}{q}^{-1/2} \sum_{\alpha\beta\gamma} (-)^\alpha G_\beta E_\gamma \times \left\langle 0 \left| \begin{pmatrix} & h'_1 & \\ h'_1 & & h'_2 \\ q'_1+\beta & & q'_2+\alpha-\beta \\ & q'_1 & \end{pmatrix}^{\dagger} \begin{pmatrix} & 0 & \\ 0 & & 0 \\ p & 0 & 0 \\ & 0 & \end{pmatrix}^{\dagger} \begin{pmatrix} & h'_1 & \\ h'_1 & & h'_2 \\ h_1 & h_2 & h_3 \\ q_1-\gamma & & q_2-q+\alpha+\gamma \\ & q'_1 & \end{pmatrix} \right| 0 \right\rangle. \tag{23}$$

But now the scalar product in (23) is nothing else but a Wigner coefficient of type

$$\left\langle \begin{matrix} h_1' & h_2' & 0 & p & 0 & 0 \\ m_1' & m_2' & ; & 0 & 0 & \\ r_1 & & & 0 & & \end{matrix} \middle| \begin{matrix} h_1 & h_2 & h_3 \\ m_1 & m_2 & \\ r_1 & & \end{matrix} \right\rangle . \tag{24}$$

In the next section we shall obtain the algebraic expression of the Wigner coefficient (24); the result appears in Eq. (48). From this result we see that this type of coefficients are independent of the value of r_1, and furthermore they are proportional to $\delta m_1 m_1'$, $\delta m_2 m_2'$. This means that in Eq. (23) we have $\delta_{q_1-\gamma,q_1'+\beta} \times \delta_{q_2-q+\alpha+\gamma,q_2'+\alpha-\beta}$, or equivalently: $\delta_{q_1+q_2,q_1'+q_2'+q}\delta_{\gamma,q_1-q_1'-\beta}$ and the last Kronecker delta allows us to eliminate the summation over γ leaving only two sums in (23).

Hence, substituting Eq. (16), (18), and (48) into Eq. (23), dividing by the $SU(2)$ coefficient in (3) in order to have the $U(3)$ RWC, and subtracting h_3' from every h_s, h_s', q_s, q_s' in order to have the result valid for $h_3' \neq 0$, we arrive at the following algebraic expression for the special $U(3)$ RWC we are discussing,

$$\left\langle \begin{matrix} h_1' & h_2' & h_3' & p & 0 & 0 \\ q_1' & q_2' & ; & q & 0 & \end{matrix} \middle\| \begin{matrix} h_1 & h_2 & h_3 \\ q_1 & q_2 & \end{matrix} \right\rangle = \delta_{h_1+h_2+h_3,\, h_1'+h_2'+h_3'+p}\delta_{q_1+q_2,q_1'+q_2'+q}$$

$$\times \sqrt{(p-q)!(q_1'-q_2'+1)(h_1-h_2+1)(h_1-h_3+2)(h_2-h_3+1)}$$

$$\times \frac{\sqrt{\dfrac{(q_1-q_1')!(q_2-q_2')!(q_1-q_2'+1)!}{(q_1'-q_2)!}}}{\sqrt{\dfrac{(h_1-h_1')!(h_2-h_2')!(h_3-h_3')!(h_1-h_2'+1)!(h_1-h_3'+2)!(h_2-h_3'+1)!}{(h_1'-h_2)!(h_2'-h_3)!(h_1'-h_3+1)!}}}$$

$$\times \frac{\sqrt{\dfrac{(h_1'-q_1')!(h_2'-q_2')!(h_1'-q_2'+1)!}{(q_1'-h_2')!(q_2'-h_3')!(q_1'-h_3'+1)!}}}{\sqrt{\dfrac{(h_1-q_1)!(h_2-q_2)!(h_1-q_2+1)!}{(q_1-h_2)!(q_2-h_3)!(q_1-h_3+1)!}}} \sum_{p_1p_2} (-)^{p_1+p_2}(q_1'-q_2'+p_1-p_2+1)$$

$$\times \frac{(q_1'-h_2'+p_1)!(q_1'-h_3'+1+p_1)!(q_1'-q_2+p_1)!(h_1-q_1'-p_1)!}{p_1!(q_1'-q_2'+1+p_1)!(h_1'-q_1'-p_1)!(q_1-q_1'-p_1)!(q_1'-h_2+p_1)!(q_1'-h_3+1+p_1)!}$$

$$\times \frac{(q_1'-q_2'-p_2)!(q_2'-h_3'+p_2)!(h_1-q_2'+1-p_2)!(h_2-q_2'-p_2)!}{p_2!(h_1'-q_2'+1-p_2)!(h_2'-q_2'-p_2)!(q-q_1+q_1'-p_2)!(q_1'-q_2+q+1-p_2)!(q_2'-h_3+p_2)!}$$

$$= \delta_{h_1+h_2+h_3,h_1'h_2'h_3'}\delta_{q_1+q_2,q_1'+q_2'+q}\sqrt{(p-q)!}\;\; S_{22}(q_s';q_s')S_{22}(q_s;q_s')\frac{S_{33}(h_s;h_s)S_{32}(h_s';q_s')}{S_{33}(h_s;h_s')S_{32}(h_s;q_s)}$$

$$\times \sum_{p_1p_2} (-)^{p_1+p_2}\left(\frac{S_{32}(h_s;q_s'+p_s)S_{22}(q_s'+p_s;q_s'+p_s)}{S_{32}(h_s';q_s'+p_s)S_{22}(q_s'+p_s;q_s')S_{22}(q_s;q_s'+p_s)}\right)^2. \tag{25}$$

3. THE MONOMIAL WIGNER COEFFICIENT $\left\langle \begin{matrix} \gamma \\ p\ 0\ 0 \\ 0\ 0 \\ 0 \end{matrix} \right\rangle$

To complete the proof of the validity of formula (25) for the $U(3)$ RWC, we have yet to obtain the value of the special coefficient given in Eq. (24). As the derivation of this type of coefficients for $U(n)$ is just as easy as for $U(3)$, we shall discuss the general case, i.e., we shall obtain an algebraic expression for the $U(n)$ Wigner coefficient,

$$\left\langle \begin{matrix} h_1' \ldots h_{n-1}'0 & p0 \ldots 0 \\ m_1' \ldots m_{n-1}' & ; \ 0 \ldots 0 \\ r_1 \ldots r_{n-2} & 0 \ldots 0 \\ \max & \max \end{matrix} \middle| \begin{matrix} h_1 \ldots h_{n-1}h_n \\ m_1 \ldots m_{n-1} \\ r_1 \ldots r_{n-2} \\ \max \end{matrix} \right\rangle, \tag{26}$$

where the word "max" means that the corresponding Gel'fand state is of highest weight (maximal) in $U(n-2)$.

To begin with, if we factorize the coefficient (26) into a product [$U(n-1)$ Wigner coeff] × [$U(n)$RWC], the first factor is

$$\left\langle \begin{matrix} m'_1 \cdots m'_{n-1} & 0 \cdots 0 \\ r_1 \cdots r_{n-2} & 0 \cdots 0 \\ \max & \max \end{matrix} \middle| \begin{matrix} m_1 \cdots m_{n-1} \\ r_1 \cdots r_{n-2} \\ \max \end{matrix} \right\rangle = \prod_{s=1}^{n=1} \delta_{m_s m'_s} \tag{27}$$

and as the RWC does not depend on r_s, we see that the whole coefficient (26) is independent of r_s.

The derivation of the coefficient (26) will be done in two steps. In the first step, we shall express (26) as a multiple of a Wigner coefficient of the same type as (26) but specialized to $m_s = m'_s = h'_s$, $s = 1, 2, \ldots, n-1$; the proportionality factor will be found by solving a recursion formula for Wigner coefficients. In the second step, the specialized coefficient that appeared in the first step, will be determined by combining some results of Refs. 9 and 10.

The method to obtain recursion relations for Wigner coefficients is well known[15]; let us briefly recall in what it consists, taking the case of $SU(2)$ for the sake of simplicity. One considers

$$\Psi^{j_1 j_2}_{JM} = \sum_{m_1 m_2} \langle j_1 m_1; j_2 m_2 | JM \rangle \psi^{(1)}_{j_1 m_1} \psi^{(2)}_{j_2 m_2} \tag{28}$$

and apply on it a generator of $SU(2)$, for instance $J_+ \equiv J^{(1)}_+ + J^{(2)}_+$, to obtain

$$J_+ \Psi^{j_1 j_2}_{JM} = \sum_{m_1 m_2} \langle j_1 m_1; j_2 m_2 | JM \rangle \{ \psi^{(2)}_{j_2 m_2} J^{(1)}_+ \psi^{(1)}_{j_1 m_1} + \psi^{(1)}_{j_1 m_1} J^{(2)}_+ \psi^{(2)}_{j_2 m_2} \}. \tag{29}$$

But it is known that

$$J_+ \Psi^{j_1 j_2}_{JM} = N(JM) \Psi^{j_1 j_2}_{JM+1},$$

where $N(JM)$ is an algebraic factor whose precise value does not concern us for the moment. So the left-hand side of (29) becomes

$$\sum_{m_1 m_2} \{ N(JM) \langle j_1 m_1; j_2 m_2 | JM + 1 \rangle \} \psi^{(1)}_{j_1 m_1} \psi^{(2)}_{j_2 m_2} \tag{29a}$$

and the right-hand side of (29) becomes, after changes of dummy indices,

$$\sum_{m_1 m_2} \{ N(j_1, m_1 - 1) \langle j_1 m_1 - 1; j_2 m_2 | JM \rangle + N(j_2, m_2 - 1) \langle j_1 m_1; j_2 m_2 - 1 | JM \rangle \} \psi^{(1)}_{j_1 m_1} \psi^{(2)}_{j_2 m_2}. \tag{29b}$$

The coefficient of $\psi^{(1)}_{j_1 m_1} \psi^{(2)}_{j_2 m_2}$ in (29a) must be equal to the coefficient of the same term in (29b), and this equality gives the desired recursion formula for Wigner coefficients.

Let us now derive a recursion formula for Wigner coefficients of $U(n)$. From the known[12] ME of the generator $\mathcal{C}_{n-1n}$ of $U(n)$ with respect to $U(n)$ Gel'fand states, we have

$$\mathcal{C}_{n-1n} \left| \begin{matrix} h_1 \cdots h_{n-1} & h_n \\ m_1 \cdots m_{n-1} & \\ r_1 \cdots r_{n-2} & \\ \max & \end{matrix} \right\rangle = \sum_{\tau=1}^{n-1} N_\tau(h_s, m_s, r_s) \times \left| \begin{matrix} h_1 & h_2 \cdots h_{n-1} & h_n \\ m_1 + \delta_{\tau 1} & m_2 + \delta_{\tau 2} \cdots m_{n-1} + \delta_{\tau n-1} & \\ & r_1 \cdots r_{n-2} & \\ & \max & \end{matrix} \right\rangle \tag{30}$$

with

$$N_\tau(h_s, m_s, r_s) = \sqrt{\frac{-\prod_{s=1}^{n} (h_s - m_\tau + \tau - s) \prod_{s=1}^{n-2} (r_s - m_\tau + \tau - s - 1)}{\prod_{\substack{s=1 \\ s \neq t}}^{n-1} (m_s - m_\tau + \tau - s - 1)(m_s - m_\tau + \tau - s)}}. \tag{31}$$

From here, by a method entirely similar to that described above for the case of $SU(2)$, we obtain the following recursion formula for the particular Wigner coefficients of $U(n)$ that we are considering in this paper:

$$\begin{aligned} &\sum_{\tau=1}^{n-1} N_\tau(h_s, m_s, r_s) \left\langle \begin{matrix} m'_1 \cdots m'_{n-1} & q0 \cdots 0 \\ r'_1 \cdots r'_{n-2} \ ; & r0 \cdots 0 \\ \max & \max \end{matrix} \middle| \begin{matrix} m_1 + \delta_{\tau 1} & m_2 + \delta_{\tau 2} \cdots m_{n-1} + \delta_{\tau n-1} \\ r_1 & r_2 \cdots r_{n-2} \\ & \max \end{matrix} \right\rangle \\ &= \sum_{\tau=1}^{n-1} N_\tau(h'_s, m'_s - \delta_{\tau s}, r'_s) \left\langle \begin{matrix} m'_1 - \delta_{\tau 1} & m'_2 - \delta_{\tau 2} \cdots m'_{n-1} - \delta_{\tau n-1} & q0 \cdots 0 \\ r'_1 & r'_2 \cdots r'_{n-2} \ ; & r0 \cdots 0 \\ & \max & \max \end{matrix} \middle| \begin{matrix} m_1 \cdots m_{n-1} \\ r_1 \cdots r_{n-2} \\ \max \end{matrix} \right\rangle \\ &+ N_1(p; q-1, r) \left\langle \begin{matrix} m'_1 \cdots m'_{n-1} & q-1\,0 \cdots 0 \\ r'_1 \cdots r'_{n-2} \ ; & r\,0 \cdots 0 \\ \max & \max \end{matrix} \middle| \begin{matrix} m_1 \cdots m_{n-1} \\ r_1 \cdots r_{n-2} \\ \max \end{matrix} \right\rangle. \end{aligned} \tag{32}$$

[We have suppressed in (32) all the $U(n)$ labels of the Wigner coefficients, namely: $(h'_1 \cdots h'_{n-1} 0)$, $(p0 \cdots 0)$ and $(h_1 \cdots h_{n-1} h_n)$.]

Now let us select a fixed number: k, $1 \leq k \leq n-1$. When we specialize formula (32) to the case $q = r = 0$, $m'_s = m_s + \delta_{sk}$, $s = 1, 2, \cdots, n-1$, then all but two terms in (32) vanish because the Wigner coefficients in them do not meet the condition expressed on the right side of (27). The only two surviving terms are the term with $\tau = k$ on the left of (32), and the term with $\tau = k$ on the right of (32), which give

$$N_k(h_s, m_s, r_s) \times \left\langle \begin{matrix} h'_1 & \cdot & \cdot & \cdot & h'_{n-1} & 0 \\ m_1 \cdots & m_{k-1} & m_k + 1 & m_{k+1} \cdots & m_{n-1}; & 0 \cdots 0 \end{matrix} \right| \left. \begin{matrix} h_1 & \cdot & \cdot & \cdot & h_{n-1} & h_n \\ m_1 \cdots & m_{k-1} & m_k + 1 & m_{k+1} \cdots & m_{n-1} & \end{matrix} \right\rangle$$

$$= N_k(h'_s, m_s, r_s) \times \left\langle \begin{matrix} h_1 & \cdot & \cdot & \cdot & h'_{n-1} & 0 \\ m_1 \cdots & m_{k-1} & m_k & m_{k+1} \cdots & m_{n-1}; & 0 \cdots 0 \end{matrix} \right| \left. \begin{matrix} h_1 & \cdot & \cdot & \cdot & h_{n-1} & h_n \\ m_1 \cdots & m_{k-1} & m_k & m_{k+1} \cdots & m_{n-1} & \end{matrix} \right\rangle . \quad (33)$$

In writing (33) we have availed ourselves of the independence of the coefficients (26) with respect to r_s to choose for the $r_s = r'_s$ in the Wigner coefficients, values such that all states involved are maximal in $U(n-1)$

Formula (33) is clearly a recursion relation for the index m_k, which allows one to express the Wigner coefficient with arbitrary m_k as a multiple of the Wigner coefficient with $m_k = h'_k$ and same values of the other m_s, $s \neq k$. Noticing that, according to (31),

$$\frac{N_k(h_s, m_s, r_s)}{N_k(h'_s, m_s, r_s)} = \sqrt{\frac{\prod_{s=1}^{k} (h_s - m_k + k - s) \prod_{s=k+1}^{n} (m_k - h_s + s - k)}{\prod_{s=1}^{k} (h'_s - m_k + k - s) \prod_{s=k+1}^{n} (m_k - h'_s + s - k)}} \quad (34)$$

(where we have split $\Pi_{s=1}^{n}$ in the form indicated above in order to have all factors positive), we obtain from (33), by iteration, the following result;

$$\left\langle \begin{matrix} h'_1 & \cdot & \cdot & \cdot & h'_{n-1} & 0 \\ m_1 \cdots & m_{k-1} & m_k & m_{k+1} \cdots & m_{n-1}; & 0 \cdots 0 \end{matrix} \right| \left. \begin{matrix} h_1 & \cdot & \cdot & \cdot & h_{n-1} & h_n \\ m_1 \cdots & m_{k-1} & m_k & m_{k-1} \cdots & m_{n-1} & \end{matrix} \right\rangle$$

$$= F_k \left\langle \begin{matrix} h'_1 & \cdot & \cdot & \cdot & h'_{n-1} & 0 \\ m_1 \cdots & m_{k-1} & h'_k & m_{k+1} \cdots & m_{n-1}; & 0 \cdots 0 \end{matrix} \right| \left. \begin{matrix} h_1 & \cdot & \cdot & \cdot & h_{n-1} & h_n \\ m_1 \cdots & m_{k-1} & h'_k & m_{k+1} \cdots & m_{n-1} & \end{matrix} \right\rangle \quad (35)$$

with

$$F_k = \sqrt{\frac{\prod_{s=1}^{k} (h_s - m_k + k - s)! \prod_{s=k+1}^{n} (h'_k - h_s + s - k - 1)! \prod_{s=1}^{k} (h'_s - h'_k + k - s)! \prod_{s=k+1}^{n} (m_k - h'_s + s - k - 1)!}{\prod_{s=1}^{k} (h_s - h'_k + k - s)! \prod_{s=k+1}^{n} (m_k - h_s + s - k - 1)! \prod_{s=1}^{k} (h'_s - m_k + k - s)! \prod_{s=k+1}^{n} (h'_k - h'_s + s - k - 1)!}} . \quad (36)$$

This result is valid for any k, $k = 1, 2, \ldots, n-1$; so by combination of the $n-1$ cases we arrive at

$$\left\langle \begin{matrix} h'_1 \ldots & h'_{n-1} & 0 & p0 \cdots 0 \\ m_1 \ldots & m_{n-1}; & 0 \cdots 0 & \end{matrix} \right| \left. \begin{matrix} h_1 \cdots & h_{n-1} & h_n \\ m_1 \cdots & m_{n-1} & \end{matrix} \right\rangle$$

$$= \prod_{k=1}^{n-1} F_k \left\langle \begin{matrix} h'_1 \cdots & h'_{n-1} & 0 & p0 \cdots 0 \\ h'_1 \cdots & h'_{n-1}; & 0 \cdots 0 & \end{matrix} \right| \left. \begin{matrix} h_1 \cdots & h_{n-1} & h_n \\ h'_1 \cdots & h'_{n-1} & \end{matrix} \right\rangle . \quad (37)$$

Let us note that according to the definition of the $S_{nn-1}(h; m)$ given in Sec. II,

$$S_{nn-1}(h_s; m_s) = \sqrt{\frac{\prod_{k=1}^{n-1} \prod_{s=1}^{k} (h_s - m_k + k - s)!}{\prod_{k=1}^{n-1} \prod_{s=k+1}^{n} (m_k - h_s + s - k - 1)!}} . \quad (38)$$

Then, if we denote the four products that appear in the numerator of (36) by $N_1(k)$, $N_2(k)$, $N_3(k)$, and $N_4(k)$ in the order in which they are written, and similarly denote the four products in the denominator of (36) by $D_1(k)$, $D_2(k)$, $D_3(k)$, and $D_4(k)$, we have

$$\sqrt{\prod_{k=1}^{n-1} \frac{N_1(k)}{D_2(k)}} = S_{nn-1}(h_s; m_s),$$

$$\sqrt{\prod_{k=1}^{n-1} \frac{D_1(k)}{N_2(k)}} = S_{nn-1}(h_s; h'_1 \cdots h'_{n-1}),$$

$$\sqrt{\prod_{k=1}^{n-1} \frac{N_3(k)}{D_4(k)}} = S_{nn-1}(h'_1 \cdots h'_{n-1} 0; h'_1 \cdots h'_{n-1}),$$

$$\sqrt{\prod_{k=1}^{n-1} \frac{D_3(k)}{N_4(k)}} = S_{nn-1}(h'_1 \cdots h'_{n-1} 0; m_s), \quad (39)$$

which shows that the algebraic factor on the right-hand side of (37) is

$$\prod_{k=1}^{n-1} F_k = \frac{S_{nn-1}(h_1 \cdots h_n; m_1 \cdots m_{n-1})}{S_{nn-1}(h'_1 \cdots h'_{n-1}0; m_1 \cdots m_{n-1})} \times \frac{S_{nn-1}(h'_1 \cdots h'_{n-1}0; h'_1 \cdots h'_{n-1})}{S_{nn-1}(h_1 \cdots h_n; h'_1 \cdots h'_{n-1})}. \quad (40)$$

It only remains to determine the last Wigner coefficient in (37). This can be done by means of the following reasoning. In Ref. 8 it has been shown that the Wigner coefficient on the right-hand side of (37) is equal to the scalar product

$$\left\langle 0 \left| \left\langle \begin{matrix} \max \\ h'_1 \cdots h'_{n-1} \\ h'_1 h'_2 \cdots h'_{n-1}\, 0 \\ h'_1 \cdots h'_{n-1} \\ \max \end{matrix} \right\rangle^{\dagger} \left\langle \begin{matrix} \max \\ 0 \cdots 0 \\ p0 \cdots 0 \\ 0 \cdots 0 \\ \max \end{matrix} \right\rangle^{\dagger} \right. \times \left\langle \begin{matrix} \max \\ h'_1 \cdots h'_{n-1} \\ h_1 h_2 \cdots h_{n-1}\, h_n \\ h'_1 \cdots h'_{n-1} \\ \max \end{matrix} \right| \left. 0 \right\rangle, \quad (41)$$

of $U(n^2) \supset U(n) \times U(n)$ states, of the type defined in Sec. II. But according to the "factorization lemma" of Ref. (9), the scalar product (41) is, at the same time, equal to

$$\sqrt{\frac{\mathcal{M}([h_1 \cdots h_n])}{\mathcal{M}([h'_1 \cdots h'_{n-1}0])\mathcal{M}([p0 \cdots 0])}} \times \left\langle \begin{matrix} h'_1 \cdots h'_{n-1}\, 0 & p0 \cdots 0 \\ h'_1 \cdots h'_{n-1}\,; & 0 \cdots 0 \\ \max & \max \end{matrix} \right| \left. \begin{matrix} h_1 \cdots h_{n-1}\, h_n \\ h'_1 \cdots h'_{n-1} \\ \max \end{matrix} \right\rangle^2, \quad (42)$$

where $\mathcal{M}([h_s])$ is the so called[7] measure of the highest weight tableau, and is given by

$$\mathcal{M}([h_1 \cdots h_n]) = \frac{\prod_{s=1}^{n} (h_s + n - s)!}{\prod_{r<s}^{r} (h_r - h_s + s - r)}. \quad (43)$$

Hence we deduce that

$$\left\langle \begin{matrix} h'_1 \cdots h'_{n-1}\, 0 & p0 \cdots 0 \\ h'_1 \cdots h'_{n-1} & 0 \cdots 0 \\ \max & \max \end{matrix} \right| \left. \begin{matrix} h_1 \cdots h_{n-1}\, h_n \\ h'_1 \cdots h'_{n-1} \\ \max \end{matrix} \right\rangle = \sqrt{\frac{\mathcal{M}([h'_1 \cdots h'_{n-1}0])\mathcal{M}([p0 \cdots 0])}{\mathcal{M}([h_1 \cdots h_n])}}. \quad (44)$$

We shall write the Wigner coefficient (44) in terms of the S_{km} functions defined in Sec. 2. With this purpose in mind, let us note the following relations among the S_{km} and the measure of the highest weight tableau given in (43):

$$[\mathcal{M}([h_1 \cdots h_n])]^{1/2} = \frac{S_{nn}(h_1 \cdots h_n; h'_1 \cdots h'_{n-1}0)}{S_{nn-1}(h_1 \cdots h_n; h'_1 \cdots h'_{n-1})S_{nn}(h_s; h_s)}, \quad (45)$$

$$[\mathcal{M}([h_1 \cdots h_{n-1}0])]^{-1/2} = S_{nn-1}(h_1 \cdots h_{n-1}0; h_1 \cdots h_{n-1}). \quad (46)$$

In formula (45) the h'_s are arbitrary, except for the fact that they must obey the betweenness relations $h_s \geq h'_s \geq h_{s+1}$. Note incidentally, that $S_{nn}(h_s; h_s)$ is related to the dimension $\Delta_n(h_1 \cdots h_n)$ of the IR $[h_1 \cdots h_n]$ of $U(n)$; in fact

$$[S_{nn}(h_s; h_s)]^2 = \prod_{i<j}^{n} (h_i - h_j + j - i) = \Delta_n(h_1 \cdots h_n) \prod_{k=1}^{n-1} k!. \quad (47)$$

Using the previous formulas and (40) to express all the right-hand side of (37) in terms of S_{km}, we have the final result:

$$\left\langle \begin{matrix} h'_1 \cdots h'_{n-1}\, 0 & p0 \cdots 0 \\ m'_1 \cdots m'_{n-1} & ; \quad 0 \cdots 0 \\ \max & \max \end{matrix} \right| \left. \begin{matrix} h_1 \cdots h_{n-1}\, h_n \\ m_1 \cdots m_{n-1} \\ \max \end{matrix} \right\rangle = \delta_{m_1 m'_1} \cdots \delta_{m_{n-1} m'_{n-1}} \sqrt{p!}\, \frac{S_{nn-1}(h_1 \cdots h_n; m_1 \cdots m_{n-1})S_{nn}(h_s; h_s)}{S_{nn-1}(h'_1 \cdots h'_{n-1}0; m_1 \cdots m_{n-1})S_{nn}(h_1 \cdots h_n; h'_1 \cdots h'_{n-1}0)}. \quad (48)$$

Specializing to $n = 3$, we have the result used in Sec. 2 to obtain the algebraic expression (25) of the $U(3)$ RWC.

One may also obtain this same result from the pattern calculus. The boson polynomial,

$$B\left(\begin{matrix} & & 0 & & \\ & 0 & & 0 & \\ p & & 0 & & 0 \\ & 0 & & 0 & \\ & & 0 & & \end{matrix} \right) = (a_3^3)^p$$

has two different formulations, using the factorization lemma. The first formulation is

$$B\left(\begin{matrix} & & 0 & & \\ & 0 & & 0 & \\ p & & 0 & & 0 \\ & 0 & & 0 & \\ & & 0 & & \end{matrix} \right) = \mathcal{M}^{-1/2} \times \sum_{\Gamma} \left\langle \begin{matrix} & & \Gamma & & \\ & p & & 0 & \\ 0 & & & & 0 \\ & 0 & & 0 & \\ & & 0 & & \end{matrix} \right\rangle \left\langle \begin{matrix} & & \Gamma & & \\ & p & & 0 & \\ 0 & & & & 0 \\ & 0 & & 0 & \\ & & 0 & & \end{matrix} \right\rangle \mathcal{M}^{1/2} \quad (49)$$

(using the lemma directly) and the second formulation

[obtained by applying the lemma to a_3^3 itself which is

$$B\left(\begin{array}{ccc} & 0 & \\ 0 & & 0 \\ 1 & 0 & \\ & 0 & \\ 0 & & \end{array}\right)\Bigg]$$

is

$$B(p00) = \left(\hbar^{-1/2}\sum_{\gamma}\left\langle\begin{array}{ccc} & \gamma & \\ 1 & 0 & 0 \\ 0 & & 0 \\ & 0 & \end{array}\right\rangle\left\langle\begin{array}{ccc} & \gamma & \\ 1 & 0 & 0 \\ 0 & & 0 \\ & 0 & \end{array}\right\rangle\hbar^{1/2}\right). \tag{50}$$

One then uses the pattern calculus rules[5,14] to obtain an explicit monomial expression for the Wigner coefficient

$$\left\langle\left|\left\langle\begin{array}{ccc} & \Gamma & \\ p & 0 & 0 \\ 0 & & 0 \\ & 0 & \end{array}\right\rangle\right|\right\rangle.$$

Let us note parenthetically the correspondences

$$\gamma_1 \Rightarrow \begin{array}{cc} 1 & \\ 1 & 0 \end{array} \qquad \gamma_2 \Rightarrow \begin{array}{cc} 0 & \\ 1 & 0 \end{array} \qquad \gamma_3 \Rightarrow \begin{array}{cc} 0 & \\ 0 & 0 \end{array}$$

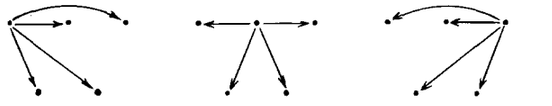
$$\Downarrow \qquad\qquad \Downarrow \qquad\qquad \Downarrow$$

$$\begin{bmatrix} & 1 & \\ 1 & & 0 \\ 1 & 0 & \\ & 0 & \end{bmatrix} \qquad \begin{bmatrix} & 0 & \\ 1 & & 0 \\ 1 & 0 & \\ & 0 & \end{bmatrix} \qquad \begin{bmatrix} & 0 & \\ 0 & & 0 \\ 1 & 0 & \\ & 0 & \end{bmatrix}$$

$$\Downarrow \qquad\qquad \Downarrow \qquad\qquad \Downarrow$$

$$\Delta_1: \begin{pmatrix} 1 & 0 & \\ & 0 & \end{pmatrix} \quad \Delta_2: \begin{pmatrix} 0 & 1 & 0 \\ & 0 & \end{pmatrix} \quad \Delta_3: \begin{pmatrix} 0 & 0 & 1 \\ & 0 & \end{pmatrix} \tag{51}$$

$$\Downarrow \qquad\qquad \Downarrow \qquad\qquad \Downarrow$$



Since these operators have, with respect to each other, opposing arrows (indicating associated noncommutativity) in the $U(3)$ "denominator" only, we find

$$\begin{bmatrix} & \Gamma & \\ p & 0 & 0 \\ 0 & & 0 \\ & 0 & \end{bmatrix} = \frac{p!}{\alpha!\beta!\gamma!}\begin{bmatrix} & \alpha & \\ \alpha+\beta & & 0 \\ p & 0 & \\ & 0 & \end{bmatrix}_R \cdot D^{(3)}_{\alpha\beta\gamma}, \tag{52}$$

where $\gamma \equiv p - \alpha - \beta$, and the $U(3)$ monomial "denominator" $D^{(3)}_{\alpha\beta\gamma}$ is known explicitly. (This and other similar results will be discussed in more detail in a future paper.)

4. THE $U(n)$ CASE

All the steps followed in the last two sections for the derivation of the $U(3)$ RWC can be repeated in the derivation of the $U(n)$ RWC associated with the Kronecker product $[h'_1 \cdots h'_n] \times [p0 \cdots 0]$ of IR of $U(n)$. We proceed now to give the results of this case.

As the analysis is so similar to that of the $U(3)$ case, we shall skip all the unessential details.

We first evaluate the Wigner coefficient

$$\left\langle\begin{array}{ccc|c} h'_1h'_2\cdots h'_{n-1}\,0 & p0\cdots 0 & & h_1h_2\cdots h_{n-1}\,h_n \\ q'_1\cdots q'_{n-1} & ; & q0\cdots 0 & q_1q_2\cdots q_{n-1} \\ q'_1\cdots q'_{n-2} & & 0\cdots 0 & q'_1\cdots q'_{n-2} \\ \max & & \max & \max \end{array}\right\rangle$$

$$= \left\langle 0\left|\left(\begin{array}{c} \max \\ h'_1\cdots h'_{n-1} \\ h'_1h'_2\cdots h'_{n-1}\,0 \\ q'_1\cdots q'_{n-1} \\ q'_1\cdots q'_{n-2} \\ \max \end{array}\right)^{\dagger}\left(\begin{array}{c} \max \\ 0\cdots 0 \\ p0\cdots 00 \\ q0\cdots 0 \\ 0\cdots 0 \\ \max \end{array}\right)^{\dagger}\right.\right.$$

$$\times \left.\left.\left(\begin{array}{c} \max \\ h'_1\cdots h'_{n-1} \\ h_1h_2\cdots h_{n-1}\,h_n \\ q_1\cdots q_{n-1} \\ q'_1\cdots q'_{n-2} \\ \max \end{array}\right)\right|0\right\rangle. \tag{53}$$

The $U(n)$ RWC is then the quotient of the coefficient (53) by the $U(n-1)$ Wigner coefficient

$$\left\langle\begin{array}{ccc|c} q'_1\cdots q'_{n-2}\,q'_{n-1} & q0\cdots 0 & & q_1\cdots q_{n-2}\,q_{n-1} \\ q'_1\cdots q'_{n-2} & ; & 0\cdots 0 & q'_1\cdots q'_{n-2} \\ \max & & \max & \max \end{array}\right\rangle, \tag{54}$$

whose algebraic expression was obtained in Sec. 3. The result is given by Eq. (48) making the appropriate changes; in particular, we have to subtract q'_{n-1} from every q'_s and q_s, but noticing that for any fixed number α,

$$S_{n\,m}(h_1-\alpha,\ldots,h_n-\alpha;q_1-\alpha,\ldots,q_m-\alpha) \equiv S_{n\,m}(h_1,\ldots,h_n;q_1,\ldots,q_m), \tag{55}$$

we find that the $U(n-1)$ Wigner coefficient in Eq. (54) has the value

$$\sqrt{q!}\,\frac{S_{n-1\,n-2}(q_s;q'_1\cdots q'_{n-2})S_{n-1\,n-1}(q_s;q_s)}{S_{n-1\,n-2}(q'_s;q'_1\cdots q'_{n-2})S_{n-1\,n-1}(q_s;q'_s)}. \tag{56}$$

For the middle state in (53) we have an expression similar to (6), namely

$$\left(\begin{array}{c} \max \\ 0\cdots 0 \\ p0\cdots 0 \\ q0\cdots 0 \\ 0\cdots 0 \\ \max \end{array}\right) = \sqrt{\frac{(p-q)!}{p!q!}}\,(\mathcal{C}_{n-1\;n})^q \left(\begin{array}{c} \max \\ 0\cdots 0 \\ p0\cdots 0 \\ 0\cdots 0 \\ \max \end{array}\right) \tag{57}$$

and using formula (7) again, we obtain this expression for the Wigner coefficient in (53):

$$\binom{p}{q}^{-1/2}\sum_{\alpha}\frac{(-)^{\alpha}}{\alpha!\,(q-\alpha)!}\left\langle 0\left|\left\{\mathcal{C}^{\alpha}_{n-1\,n}\begin{pmatrix}\max\\ h'_1\cdots h'_{n-1}\\ h'_1h'_2\cdots h'_{n-1}\;0\\ q'_1\cdots q'_{n-1}\\ q'_1\cdots q'_{n-2}\\ \max\end{pmatrix}\right\}^{\dagger}\times\begin{pmatrix}\max\\ 0\cdots 0\\ p0\cdots 0\\ 0\cdots 0\\ \max\end{pmatrix}^{\dagger}\mathcal{C}^{q-\alpha}_{n\,n-1}\begin{pmatrix}\max\\ h'_1\cdots h'_{n-1}\\ h_1h_2\cdots h_{n-1}\;h_n\\ q_1\cdots q_{n-1}\\ q'_1\cdots q'_{n-2}\\ \max\end{pmatrix}\right|0\right\rangle. \tag{58}$$

In the appendix we shall prove that

$$\mathcal{C}^{k}_{n-1\,n}\left|\begin{matrix}h_1\cdots h_{n-1}\;h_n\\ q_1\cdots q_{n-1}\\ r_1\cdots r_{n-2}\\ \max\end{matrix}\right\rangle=k!\sum_{\substack{\rho_1\cdots\rho_{n-1}\\(\rho_1+\cdots+\rho_{n-1}=k)}}\frac{S_{n-1\,n-1}(q_s+\rho_s;q_s+\rho_s)S_{n-1\,n-1}(q_s;q_s)}{[S_{n-1\,n-1}(q_s+\rho_s;q_s)]^2}$$
$$\times\frac{S_{n-1\,n-2}(q_s+\rho_s;r_s)S_{nn-1}(h_s;q_s)}{S_{n-1\,n-2}(q_s;r_s)S_{nn-1}(h_s;q_s+\rho_s)}\left|\begin{matrix}h_1\cdots h_{n-1}\;h_n\\ q_1+\rho_1\cdots q_{n-1}+\rho_{n-1}\\ r_1\cdots r_{n-2}\\ \max\end{matrix}\right\rangle \tag{59}$$

and by Hermitian conjugation we deduce from here that

$$\mathcal{C}^{k}_{n\,n-1}\left|\begin{matrix}h_1\cdots h_{n-1}\;h_n\\ q_1\cdots q_{n-1}\\ r_1\cdots r_{n-2}\\ \max\end{matrix}\right\rangle=k!\sum_{\substack{\rho_1\cdots\rho_{n-1}\\(\rho_1+\cdots+\rho_{n-1}=k)}}\frac{S_{n-1\,n-1}(q_s;q_s)S_{n-1\,n-1}(q_s-\rho_s;q_s-\rho_s)}{[S_{n-1\,n-1}(q_s;q_s-\rho_s)]^2}$$
$$\times\frac{S_{n-1\,n-2}(q_s;r_s)S_{nn-1}(h_s;q_s-\rho_s)}{S_{n-1\,n-2}(q_s-\rho_s;r_s)S_{nn-1}(h_s;q_s)}\left|\begin{matrix}h_1\cdots h_{n-1}\;h_n\\ q_1-\rho_1\cdots q_{n-1}-\rho_{n-1}\\ r_1\cdots r_{n-2}\\ \max\end{matrix}\right\rangle. \tag{60}$$

Substitution of the last two results into Eq. (58) leads us to the conclusion that the Wigner coefficient in (53) is equal to

$$\binom{p}{q}^{-1/2}\sum_{\substack{\alpha\rho_1\cdots\rho_{n-1}\sigma_1\cdots\sigma_{n-1}\\(\rho_1+\cdots+\rho_{n-1}=\alpha)\\(\sigma_1+\cdots+\sigma_{n-1}=q-\alpha)}}(-)^{\alpha}\left\langle 0\left|\begin{pmatrix}\max\\ h'_1\cdots h'_{n-1}\\ h'_1h'_2\cdots h'_{n-1}\;0\\ q'_1+\rho_1\cdots q'_{n-1}+\rho_{n-1}\\ q'_1\cdots q'_{n-2}\\ \max\end{pmatrix}^{\dagger}\begin{pmatrix}\max\\ 0\cdots 0\\ p0\cdots 0\\ 0\cdots 0\\ \max\end{pmatrix}^{\dagger}\begin{pmatrix}\max\\ h'_1\cdots h'_{n-1}\\ h_1h_2\cdots h_{n-1}\;h_n\\ q_1-\sigma_1\cdots q_{n-1}-\sigma_{n-1}\\ q'_1\cdots q'_{n-2}\\ \max\end{pmatrix}\right|0\right\rangle$$
$$\times\frac{S_{n-1\,n-1}(q_s;q_s)S_{n-1\,n-1}(q_s-\sigma_s;q_s-\sigma_s)S_{n-1\,n-2}(q_s;q'_1\cdots q'_{n-2})S_{n\,n-1}(h_s;q_s-\sigma_s)}{[S_{n-1\,n-1}(q_s;q_s-\sigma_s)]^2S_{n-1\,n-2}(q_s-\sigma_s;q'_1\cdots q'_{n-2})S_{n\,n-1}(h_s;q_s)}$$
$$\times\frac{S_{n-1\,n-1}(q'_s+\rho_s;q'_s+\rho_s)S_{n-1\,n-1}(q'_s;q'_s)S_{n-1\,n-2}(q'_s+\rho_s;q'_1\cdots q'_{n-2})S_{nn-1}(h'_1\cdots h'_{n-1}0;q'_s)}{[S_{n-1\,n-1}(q'_s+\rho_s;q'_s)]^2S_{n-1\,n-2}(q'_s;q'_1\cdots q'_{n-2})S_{n\,n-1}(h'_1\cdots h'_{n-1}0;q'_s+\rho_s)}. \tag{61}$$

Now the ME appearing on (61) is nothing else but the particular $U(n)$ Wigner coefficient evaluated in Sec. 3 and given in Eq. (48). [Recall that it was shown that this coefficient is independent of the $U(n-2)$ labels.] When we substitute the value of this coefficient into Eq. (61), it brings in $(n-1)$ Kronecker deltas: $\delta_{q'_s+\rho_s,q_s-\sigma_s}$, $s=1,2,\ldots,n-1$. The two conditions $\rho_1+\cdots+\rho_{n-1}=\alpha$, $\sigma_1+\cdots+\sigma_{n-1}=q-\alpha$, are equivalent to multiplication of (61) by $\delta_{\alpha,\rho_1+\cdots+\rho_{n-1}}$, $\delta_{q-\alpha,\sigma_1+\cdots+\sigma_{n-1}}$. Hence we can use the first $(n-1)$ Kronecker deltas to get rid of all the summations over σ_s in (61), then use $\delta_{\alpha,\rho_1+\cdots+\rho_{n-1}}$ to eliminate the summation over α, and the last Kronecker delta then becomes $\delta_{q+q'_1+\cdots+q'_{n-1},q_1+\cdots+q_{n-1}}$.

Following the mentioned steps, dividing (61) by the $U(n-1)$ Wigner coefficient in (56), and furthermore subtracting h'_n from every h'_s, h_s, q'_s, and q_s, we have the final result for the $U(n)$ RWC associated with the Kronecker product $[h'_1\cdots h'_n]\times[p0\cdots 0]$ of IR of $U(n)$:

$$\left\langle \begin{matrix} h'_1 \cdots h'_{n-1}\, h'_n \\ q'_1 \cdots q'_{n-1} \end{matrix} ; \begin{matrix} p0 \cdots 0 \\ q0 \cdots 0 \end{matrix} \middle\| \begin{matrix} h_1 \cdots h_{n-1}\, h_n \\ q_1 \cdots q_{n-1} \end{matrix} \right\rangle$$

$$= \delta_{h'_1+\cdots+h'_n+p,\,h_1+\cdots+h_n}\,\delta_{q'_1+\cdots+q'_{n-1}+q,\,q_1+\cdots+q_{n-1}}$$
$$\times \sqrt{(p-q)!}\,\frac{S_{nn}(h_s;h_s)S_{n\,n-1}(h'_s;q'_s)}{S_{nn}(h_s;h'_s)S_{n\,n-1}(h_s;q_s)}$$
$$\times S_{n-1\,n-1}(q_s;q'_s)S_{n-1\,n-1}(q'_s;q'_s)$$
$$\times \sum_{\rho_1\cdots\rho_{n-1}} (-)^{\rho_1+\cdots+\rho_{n-1}}$$
$$\times \left[\frac{S_{nn-1}(h_s;\bar{q}_s)S_{n-1\,n-1}(\bar{q}_s;\bar{q}_s)}{S_{nn-1}(h'_s;\bar{q}_s)S_{n-1\,n-1}(q_s;\bar{q}_s)S_{n-1\,n-1}(\bar{q}_s;q'_s)}\right]^2 \tag{62}$$

with $\bar{q}_s = q'_s + \rho_s$, $s = 1, 2, \ldots, n-1$.

5. COMMENTS

We have derived in this paper an explicit algebraic formula [Eq. (62)] for the reduced Wigner coefficients involved in the Kronecker product of IR of $U(n)$:

$$[h'_1 \cdots h'_n] \times [p0 \cdots 0].$$

It should be noted that this is the most general multiplicity-free Kronecker product on $U(n)$, so the case we have analyzed is the *only one in which the Wigner coefficients are uniquely defined* (except for a phase factor) *by the group itself.* In the case of $U(3)$ there exists also a canonical definition of Wigner coefficients[10] even for general (nonmultiplicity-free) Kronecker products; to our knowledge this is the only case in which such an extension has been achieved.

It has been brought to our attention that a result equivalent to our formula (62) was obtained earlier by Jucys.[16a] The analysis of Jucys is based on properties of the symmetric group $S(n)$, and his result is expressed in terms of functions over the symmetric group; whereas our result appears in terms of the functions $S_{n\,m}$, introduced in the present paper, which are expressed directly in terms of the quantum labels of the initial and final Gel'fand states.

The authors of this paper want to thank the referee for having pointed out to them Ref. 16b. In that paper results identical to our equations (48), (59), and (62) are obtained by methods slightly different to those followed by us.

ACKNOWLEDGMENTS

One of us (E.C.) wants to acknowledge Professor H. Fairbank, chairman of the Physics Department at Duke University, for the hospitality extended to him.

APPENDIX

In this appendix we shall give the proof of Eq. (59) in the text, concerning the ME of a power of the $U(n)$ generator $\mathcal{C}_{n-1\,n}$ with respect to $U(n)$ Gel'fand States.

Using the Gel'fand–Zetlin results[12] for the ME of $\mathcal{C}_{n-1\,n}$, we have

$$\left\langle \begin{matrix} h_s \\ q_s \\ r_s \\ \max \end{matrix} \middle| \mathcal{C}^{k+1}_{n-1\,n} \middle| \begin{matrix} h_s \\ q_s - \rho_s \\ r_s \\ \max \end{matrix} \right\rangle$$
$$= \sum_{\tau=1}^{n-1} N_\tau \left\langle \begin{matrix} h_s \\ q_s \\ r_s \\ \max \end{matrix} \middle| \mathcal{C}^{k}_{n-1\,n} \middle| \begin{matrix} h_s \\ q_s - \rho_s + \delta_{\tau s} \\ r_s \\ \max \end{matrix} \right\rangle \tag{A1}$$

with

$$N_\tau = \frac{\sqrt{-\prod_{s=1}^{n}(h_s - q_\tau + \rho_\tau + \tau - s)\prod_{s=1}^{n-2}(r_s - q_\tau + \rho_\tau + \tau - s - 1)}}{\sqrt{\prod_{\substack{s=1 \\ s\neq\tau}}^{n-1}(q_s - q_\tau + \rho_\tau - \rho_s + \tau - s - 1)(q_s - q_\tau + \rho_\tau - \rho_s + \tau - s)}}. \tag{A2}$$

In the Gel'fand states above, h_s, q_s, and r_s are a generic designation for the IR labels of $U(n)$, $U(n-1)$, and $U(n-2)$, respectively; and in order to obtain nonvanishing results, the nonnegative integers ρ_s must satisfy the condition[12]

$$\sum_{s=1}^{n-1} \rho_s = k + 1. \tag{A3}$$

Equation (A1) is a recursion formula for the ME of $\mathcal{C}^k_{n-1\,n}$; we shall obtain the solution of this recursion by the induction method. We make the following conjecture:

$$\left\langle \begin{matrix} h_s \\ q_s \\ r_s \\ \max \end{matrix} \middle| \mathcal{C}^{k}_{n-1\,n} \middle| \begin{matrix} h_s \\ q'_s \\ r_s \\ \max \end{matrix} \right\rangle = \delta_{q'_1+\cdots+q'_{n-1}+k,\,q_1+\cdots+q_{n-1}}$$
$$\times k!\,\frac{S_{n-1\,n-1}(q_s;q_s)S_{n-1\,n-1}(q'_s;q'_s)}{[S_{n-1\,n-1}(q_s;q'_s)]^2}$$
$$\times \frac{S_{n-1\,n-2}(q_s;r_s)S_{nn-1}(h_s;q'_s)}{S_{n-1\,n-2}(q'_s;r_s)S_{nn-1}(h_s;q_s)}; \tag{A4}$$

then substitute this (with $q'_s = q_s - \rho_s + \delta_{\tau s}$) on the rhs of (A1) and show that, after some simplifications, what is obtained for the rhs of (A1) is a formula like the rhs of (A4) but with $(k+1)$ instead of k, and $q'_s = q_s - \rho_s$. This proves that the "ansatz" (A4) satisfies the recursion formula (A1), and since the ansatz is true for $k = 0$, it will be true in general. [The definition of the S_{nm} functions in (A4) was given in Sec. 2].

Let us proceed to implement the above described steps. When we substitute (A4) with $q'_s = q_s - \rho_s +$

$\delta_{\tau s}$ into the rhs of (A1), there will appear some $S_{n\,m}$ functions containing a Kronecker delta $\delta_{\tau s}$ in their arguments. We can get rid of this $\delta_{\tau s}$ by a method illustrated by the following example:

$$S_{n-1\,n-1}(q_s-\rho_s+\delta_{\tau s};q_s-\rho_s+\delta_{\tau s})=\left(\prod_{i<j}^{n-1}(q_i-q_j+\rho_j-\rho_i+\delta_{\tau i}-\delta_{\tau j}+j-i)\right)^{1/2}$$

$$=\left(\prod_{\substack{i<j\\ i,j\neq\tau}}^{n-1}(q_i-q_j+\rho_j-\rho_i+j-i)\prod_{\substack{s=1\\ s\neq\tau}}^{n-1}(q_s-q_\tau+\rho_\tau-\rho_s+\tau-s-1)\right)^{1/2}$$

$$=\left(\frac{\prod_{i<j}^{n-1}(q_i-q_j+\rho_j-\rho_i+j-i)\prod_{s=1(s\neq\tau)}^{n-1}(q_s-q_\tau+\rho_\tau-\rho_s+\tau-s-1)}{\prod_{s\neq\tau}^{n-1}(q_s-q_\tau+\rho_\tau-\rho_s+\tau-s)}\right)^{1/2}$$

$$=S_{n-1\,n-1}(q_s-\rho_s;q_s-\rho_s)\left(\frac{\prod_{s=1(s\neq\tau)}^{n-1}(q_s-q_\tau+\rho_\tau-\rho_s+\tau-s-1)}{\prod_{\substack{s=1\\ s\neq\tau}}^{n-1}(q_s-q_\tau+\rho_\tau-\rho_s+\tau-s)}\right)^{1/2}. \tag{A5}$$

By a similar method, we obtain

$$[S_{n-1\,n-1}(q_s;q_s-\rho_s+\delta_{\tau s})]^{-2}$$

$$=[S_{n-1\,n-1}(q_s;q_s-\rho_s)]^{-2}\prod_{s=1}^{\tau}(q_s-q_\tau+\rho_\tau+\tau-s)\prod_{s=\tau+1}^{n-1}(q_\tau-q_s-\rho_\tau+s-\tau); \tag{A6}$$

$$[S_{n-1\,n-2}(q_s-\rho_s+\delta_{\tau s},r_s)]^{-1}$$

$$=[S_{n-1\,n-2}(q_s-\rho_s;r_s)]^{-1}\left(\prod_{s=1}^{\tau-1}(r_s-q_\tau+\rho_\tau+\tau-s-1)\prod_{s=\tau}^{n-2}(q_\tau-r_s-\rho_\tau-\tau+s+1)\right)^{-1/2} \tag{A7}$$

$$S_{n\,n-1}(h_s;q_s-\rho_s+\delta_{\tau s})=S_{n\,n-1}(h_s;q_s-\rho_s)\left(\prod_{s=1}^{\tau}(h_s-q_\tau+\rho_\tau+\tau-s)\prod_{s=\tau+1}^{n}(q_\tau-h_s-\rho_\tau-\tau+s)\right)^{-1/2}. \tag{A8}$$

We substitute all the previous results on the rhs of (A1) and obtain, after several cancellations

$$\text{rhs (A1)}=k!\,\frac{S_{n-1\,n-1}(q_s;q_s)S_{n-1\,n-1}(q_s-\rho_s;q_s-\rho_s)S_{n-1\,n-2}(q_s;r_s)S_{n\,n-1}(h_s;q_s-\rho_s)}{[S_{n-1\,n-1}(q_s;q_s-\rho_s)]^2S_{n-1\,n-2}(q_s-\rho_s;r_s)S_{n\,n-1}(h_s;q_s)}$$

$$\times\sum_\tau\rho_\tau\prod_{\substack{s=1\\ s\neq\tau}}^{n-1}\frac{(q_s-q_\tau+\rho_\tau+\tau-s)}{(q_s-q_\tau+\rho_\tau-\rho_s+\tau-s)}. \tag{A9}$$

Now, let us write $q_s-\rho_s-s=X_s$; then we see that the last sum in (A9) is equal to

$$\sum_{\tau=1}^{n-1}\rho_\tau\prod_{\substack{s=1\\ s\neq\tau}}^{n-1}\left(1+\frac{\rho_s}{(X_s-X_\tau)}\right)$$

$$=\sum_{k=1}^{n-1}\sum_{\substack{i_1,i_2,\ldots,i_k=1\\ (i_1<i_2<\cdots<i_k)}}^{n-1}\rho_{i_1}\rho_{i_2}\cdots\rho_{i_k}\sum_\nu\prod_{\substack{\mu\\ \mu\neq\nu}}\frac{1}{(X_\mu-X_\nu)}, \tag{A10}$$

where μ,ν are numbers in the set $\{i_1,i_2,\ldots,i_k\}$. If we relabel the indices $i_s\to s$, then the sum over ν in (A10) becomes

$$\sum_{s=1}^{k}\prod_{\substack{\tau=1\\ \tau\neq s}}^{k}\frac{1}{(X_s-X_\tau)} \tag{A11}$$

and it has been shown by Louck and Biedenharn[17] that this sum vanishes identically for $k\geq 2$. So in (A10) only the term $k=1$ is different from zero, and this is equal to $\rho_1+\cdots+\rho_{n-1}=k+1$, according to (A3). This factor $(k+1)$ then converts the $k!$ in (A9) into $(k+1)!$, and then the rhs of (A1) agrees with the conjecture (A4) but with k shifted to $(k+1)$, and with $q'_s=q_s-\rho_s$. As mentioned before, this then shows that the conjecture (A4) is true in general.

* Supported in part by the National Science Foundation and the Army Research Office (Durham).
† On Leave of absence from Universidad Nacional de México, Mexico City, and Comisión Nacional de Energia Nuclear, México.
‡ Supported by ARO(D).

1 E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959), p. 191.
2 M. Moshinsky, Rev. Mod. Phys. 34, 813 (1962).
3 L. Banyai, N. Marinescu, I. Raszillier, and V. Rittenberg, Phys. Letters 14, 156 (1965).
4 S. Alisauskas, Lietuvos fizikos rinkinys 9, 641 (1969).
5 L. C. Biedenharn and J. D. Louck, Commun. Math. Phys. 8, 89 (1968).
6 Professor G. Racah, in an unpublished work communicated in a letter to Professor M. Moshinsky (1963), also derived a formula for these coefficients. We thank Professor Moshinsky for bring-

ing that letter to our attention. Racah's result contains two sums and is very similar (though not quite identical) to the result we obtain in this paper.
[7] J. D. Louck, Amer. J. Phys. 38, 3 (1970).
[8] J. D. Louck, J. Math. Phys. 6, 1786 (1965).
[9] T. A. Brody, M. Moshinsky, and I. Renero, J. Math. Phys. 6, 1540 (1965).
[10] L. C. Biedenharn, A. Giovannini, and J. D. Louck, J. Math. Phys. 8, 691 (1967).
[11] J. G. Nagel and M. Moshinsky, J. Math. Phys. 6, 682 (1965).
[12] I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk SSSR 71, 825 (1950).
[13] Reference 2, formula 3.11.
[14] L. C. Biedenharn, *Spectroscopic and Group Theoretical Methods in Physics* (North Holland, Amsterdam, 1968), Racah Memorial Volume, p. 59.
[15] E. U. Condon and G. H. Shortley, *Theory of Atomic Spectra* (Cambridge U.P., Cambridge, 1935).
[16a] A. A. Jucys, Lietuvos fizikos rinkinys 10, 5 (1970).
[16b] S. J. Ališauskas, A.-A. A. Jucys, and A. P. Jucys, "On the Symmetric Tensor Operators of the Unitary Group," J. Math Phys. 13, (1972) (to be published).
[17] J. D. Louck and L. C. Biedenharn, J. Math. Phys. 11, 2368 (1970).

Scintillation of Randomized Electromagnetic Fields*

Leonard S. Taylor
Electrical Engineering Department, University of Maryland, College Park, Maryland 20742
(Received 17 November 1971; Revised Manuscript Received 4 January 1971)

The propagation of a randomized electromagnetic field in a uniform medium is considered and expressions are derived for the autocorrelation and spectrum of intensity fluctuations in terms of the fourth-order coherence function on the initial plane.

INTRODUCTION

A variety of phenomena of current interest involve the propagation through uniform media of fields which have been randomized by previous passage through stochastic regions. Particular examples include the radio-astronomical observations of the solar plasma and the contribution of the turbulent upper layers of the atmosphere to the twinkling of starlight. A striking example of the complexity of the phenomenon may sometimes be observed on the bottom of a swimming pool when the surface is lightly rippled; at both the deep and shallow ends the light on the floor of the pool is uniform, but in the center, rippling bands of light and shadow are observed. This phenomena and others may be crudely understood on a ray basis, when it is recognized that even small phase variations introduced by the stochastic element are converted into intensity fluctuations during passage through the uniform medium by focusing. This is the principle used in shadowgraph photography of turbulent (and uniform) flows.

Although the ray-optical theory has been well known, a general electromagnetic solution using the Helmholtz–Kirchhoff integral has been lacking. The difficulty has been that eighth-order integrations are involved. In this paper, we employ a transformation derived in a previous paper by Torrieri and Taylor[1] to reduce the integrals to lower orders. Compact general expressions are obtained for the autocorrelation and spectrum of intensity fluctuations in terms of the fourth-order coherence function on the initial plane. From these general expressions the restricted formulas previously obtained by Mercier,[2] Salpeter,[3] and Jokipii[4] for various special cases are readily deduced.

THEORY

We consider the propagation of a wave $\psi(\mathbf{r})$, obeying the scalar wave equation

$$(\nabla^2 + k^2)\psi(\mathbf{r}) = 0,$$

where as usual $k^2 = \epsilon k_0^2 = \epsilon(\omega/c)^2$. It is assumed that boundary values are given over the $z = 0$ plane. It has been shown[1] that the Helmholtz–Kirchhoff integral reduces in this case to the form

$$\psi(\mathbf{r}) = -\frac{1}{2\pi}\frac{\partial}{\partial z}\int\int \psi(\mathbf{r}')\frac{e^{jkR}}{R}\,dS', \qquad z > 0, \tag{1}$$

where $R = |\mathbf{r} - \mathbf{r}'|$ and the integration is over the plane $z' = 0$. We use Eq. (1) to obtain the following expression for the fourth-order coherence function $M(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \langle\psi(\mathbf{r}_1)\psi^*(\mathbf{r}_2)\psi(\mathbf{r}_3)\psi^*(\mathbf{r}_4)\rangle$ in the space $z > 0$ in terms of the fourth-order coherence on the boundary plane:

$$M(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \left(\frac{1}{2\pi}\right)^4 \frac{\partial}{\partial z_1}\cdots\frac{\partial}{\partial z_4}\int_{P_1}\int_{P_2}\int_{P_3}\int_{P_4} dS_1'\cdots$$
$$\times\, dS_4' M(\mathbf{r}_1', \mathbf{r}_2', \mathbf{r}_3', \mathbf{r}_4')\frac{e^{jk(R_1 - R_2 + R_3 - R_1)}}{R_1R_2R_3R_4}. \tag{2}$$

We shall assume that M is stationary over the $z = 0$ plane: It may be written as a function of the differences between coordinates. Thus

$$M(\mathbf{r}_1', \mathbf{r}_2', \mathbf{r}_3', \mathbf{r}_4') = M_0(x_1' - x_2', \ldots, y_3' - y_4')$$
$$= M_0(\xi_1', \ldots, \xi_6') = M_0(\boldsymbol{\xi}'), \tag{3}$$

where

$$\xi_1' = x_1' - x_2', \quad \xi_2' = x_2' - x_3', \quad \xi_3' = x_3' - x_4'$$
$$\xi_4' = y_1 - y_2', \quad \xi_5' = y_2' - y_3', \quad \xi_6' = y_3' - y_4' \tag{4}$$

and $\boldsymbol{\xi}'$ is a vector in 6-space. It should be noted that the differences $x_4' - x_1'$ and $y_4' - y_1'$ can be expressed in terms of the other differences and have not been included explicitly and that we have attached a zero subscript to M as a reminder that this indicates the value over the $z = 0$ plane. We introduce the sixfold fourier transform

$$M_0(\boldsymbol{\xi}') = \left(\frac{1}{2\pi}\right)^6 \int^{(6)} d\mathbf{k}_\xi \exp(j\mathbf{k}_\xi\cdot\boldsymbol{\xi}')\hat{M}_0(\mathbf{k}_\xi), \tag{5}$$

where $k_{\xi_1}, \ldots, k_{\xi_6}$ are the spatial frequencies corresponding to $\xi_1', \ldots, \xi_6'$ and $d\mathbf{k}_\xi = dk_{\xi_1}\cdots dk_{\xi_6}$. The integration is sixfold, as indicated by the numeral in parentheses and each integral is taken from $-\infty$ to $+\infty$. We employ this convention throughout unless otherwise indicated. Substituting Eq. (5) into Eq. (2) we obtain upon exchanging order of integration

$$M(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \left(\frac{1}{2\pi}\right)^{10} \int^{(6)} d\mathbf{k}_\xi M_0(\mathbf{k}_\xi)\frac{\partial}{\partial z_1}\cdots\frac{\partial}{\partial z_4}$$

ing that letter to our attention. Racah's result contains two sums and is very similar (though not quite identical) to the result we obtain in this paper.
[7] J. D. Louck, Amer. J. Phys. 38, 3 (1970).
[8] J. D. Louck, J. Math. Phys. 6, 1786 (1965).
[9] T. A. Brody, M. Moshinsky, and I. Renero, J. Math. Phys. 6, 1540 (1965).
[10] L. C. Biedenharn, A. Giovannini, and J. D. Louck, J. Math. Phys. 8, 691 (1967).
[11] J. G. Nagel and M. Moshinsky, J. Math. Phys. 6, 682 (1965).
[12] I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk SSSR 71, 825 (1950).
[13] Reference 2, formula 3.11.
[14] L. C. Biedenharn, *Spectroscopic and Group Theoretical Methods in Physics* (North Holland, Amsterdam, 1968), Racah Memorial Volume, p. 59.
[15] E. U. Condon and G. H. Shortley, *Theory of Atomic Spectra* (Cambridge U.P., Cambridge, 1935).
[16a] A. A. Jucys, Lietuvos fizikos rinkinys 10, 5 (1970).
[16b] S. J. Ališauskas, A.-A. A. Jucys, and A. P. Jucys, "On the Symmetric Tensor Operators of the Unitary Group," J. Math Phys. 13, (1972) (to be published).
[17] J. D. Louck and L. C. Biedenharn, J. Math. Phys. 11, 2368 (1970).

Scintillation of Randomized Electromagnetic Fields*

Leonard S. Taylor
Electrical Engineering Department, University of Maryland, College Park, Maryland 20742
(Received 17 November 1971; Revised Manuscript Received 4 January 1971)

The propagation of a randomized electromagnetic field in a uniform medium is considered and expressions are derived for the autocorrelation and spectrum of intensity fluctuations in terms of the fourth-order coherence function on the initial plane.

INTRODUCTION

A variety of phenomena of current interest involve the propagation through uniform media of fields which have been randomized by previous passage through stochastic regions. Particular examples include the radio-astronomical observations of the solar plasma and the contribution of the turbulent upper layers of the atmosphere to the twinkling of starlight. A striking example of the complexity of the phenomenon may sometimes be observed on the bottom of a swimming pool when the surface is lightly rippled; at both the deep and shallow ends the light on the floor of the pool is uniform, but in the center, rippling bands of light and shadow are observed. This phenomena and others may be crudely understood on a ray basis, when it is recognized that even small phase variations introduced by the stochastic element are converted into intensity fluctuations during passage through the uniform medium by focusing. This is the principle used in shadowgraph photography of turbulent (and uniform) flows.

Although the ray-optical theory has been well known, a general electromagnetic solution using the Helmholtz–Kirchhoff integral has been lacking. The difficulty has been that eighth-order integrations are involved. In this paper, we employ a transformation derived in a previous paper by Torrieri and Taylor[1] to reduce the integrals to lower orders. Compact general expressions are obtained for the autocorrelation and spectrum of intensity fluctuations in terms of the fourth-order coherence function on the initial plane. From these general expressions the restricted formulas previously obtained by Mercier,[2] Salpeter,[3] and Jokipii[4] for various special cases are readily deduced.

THEORY

We consider the propagation of a wave $\psi(\mathbf{r})$, obeying the scalar wave equation

$$(\nabla^2 + k^2)\psi(\mathbf{r}) = 0,$$

where as usual $k^2 = \epsilon k_0^2 = \epsilon(\omega/c)^2$. It is assumed that boundary values are given over the $z = 0$ plane. It has been shown[1] that the Helmholtz–Kirchhoff integral reduces in this case to the form

$$\psi(\mathbf{r}) = -\frac{1}{2\pi}\frac{\partial}{\partial z}\int\int \psi(\mathbf{r}')\frac{e^{jkR}}{R}\,dS', \qquad z > 0, \tag{1}$$

where $R = |\mathbf{r} - \mathbf{r}'|$ and the integration is over the plane $z' = 0$. We use Eq. (1) to obtain the following expression for the fourth-order coherence function $M(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \langle\psi(\mathbf{r}_1)\psi^*(\mathbf{r}_2)\psi(\mathbf{r}_3)\psi^*(\mathbf{r}_4)\rangle$ in the space $z > 0$ in terms of the fourth-order coherence on the boundary plane:

$$M(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \left(\frac{1}{2\pi}\right)^4 \frac{\partial}{\partial z_1}\cdots\frac{\partial}{\partial z_4}\int_{P_1}\int_{P_2}\int_{P_3}\int_{P_4} dS_1'\cdots$$
$$\times\, dS_4' M(\mathbf{r}_1', \mathbf{r}_2', \mathbf{r}_3', \mathbf{r}_4')\frac{e^{jk(R_1 - R_2 + R_3 - R_1)}}{R_1R_2R_3R_4}. \tag{2}$$

We shall assume that M is stationary over the $z = 0$ plane: It may be written as a function of the differences between coordinates. Thus

$$M(\mathbf{r}_1', \mathbf{r}_2', \mathbf{r}_3', \mathbf{r}_4') = M_0(x_1' - x_2', \ldots, y_3' - y_4')$$
$$= M_0(\xi_1', \ldots, \xi_6') = M_0(\boldsymbol{\xi}'), \tag{3}$$

where

$$\xi_1' = x_1' - x_2', \quad \xi_2' = x_2' - x_3', \quad \xi_3' = x_3' - x_4'$$
$$\xi_4' = y_1 - y_2', \quad \xi_5' = y_2' - y_3', \quad \xi_6' = y_3' - y_4' \tag{4}$$

and $\boldsymbol{\xi}'$ is a vector in 6-space. It should be noted that the differences $x_4' - x_1'$ and $y_4' - y_1'$ can be expressed in terms of the other differences and have not been included explicitly and that we have attached a zero subscript to M as a reminder that this indicates the value over the $z = 0$ plane. We introduce the sixfold fourier transform

$$M_0(\boldsymbol{\xi}') = \left(\frac{1}{2\pi}\right)^6 \int^{(6)} d\mathbf{k}_\xi \exp(j\mathbf{k}_\xi\cdot\boldsymbol{\xi}')\hat{M}_0(\mathbf{k}_\xi), \tag{5}$$

where $k_{\xi_1}, \ldots, k_{\xi_6}$ are the spatial frequencies corresponding to $\xi_1', \ldots, \xi_6'$ and $d\mathbf{k}_\xi = dk_{\xi_1}\cdots dk_{\xi_6}$. The integration is sixfold, as indicated by the numeral in parentheses and each integral is taken from $-\infty$ to $+\infty$. We employ this convention throughout unless otherwise indicated. Substituting Eq. (5) into Eq. (2) we obtain upon exchanging order of integration

$$M(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \left(\frac{1}{2\pi}\right)^{10} \int^{(6)} d\mathbf{k}_\xi M_0(\mathbf{k}_\xi)\frac{\partial}{\partial z_1}\cdots\frac{\partial}{\partial z_4}$$

$$\times \int^{(8)} \frac{\exp[j\mathbf{k}_\xi \cdot \boldsymbol{\xi}' + jk(R_1 - R_2 + R_3 - R_4)]}{R_1 R_2 R_3 R_4} \times dx_1' \cdots dy_4'. \quad (6)$$

Expressing the ξ_i' in terms of x_i', y_i' and regrouping

$$M(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \left(\frac{1}{2\pi}\right)^6 \int^{(6)} d\mathbf{k}_\xi \hat{M}_0(\mathbf{k}_\xi) \left(\frac{\partial}{\partial z_1} T(\mathbf{r}_1; k_{\xi_1}, k_{\xi_4})\right) \times \left(\frac{\partial}{\partial z_2} T^*(\mathbf{r}_2; k_{\xi_1} - k_{\xi_2}, k_{\xi_4} - k_{\xi_5})\right) \times \left(\frac{\partial}{\partial z_3} T(\mathbf{r}_3; k_{\xi_3} - k_{\xi_2}, k_{\xi_6} - k_{\xi_5})\right) \left(\frac{\partial}{\partial z_3} T^*(\mathbf{r}_4; k_{\xi_3}, k_{\xi_6})\right), \quad (7)$$

where by definition

$$T(\mathbf{r}_i; k_{\xi m}, k_{\xi n}) = \frac{1}{2\pi} \int^{(2)} dx_i' dy_i' \times \frac{\exp[j(k_{\xi_m} x_i' + k_{\xi_n} y_i' + kR_i)]}{R_i}. \quad (8)$$

Torrieri and Taylor[1] evaluated $T(\mathbf{r}_i; k_{\xi_m}, k_{\xi_n})$ using a representation of a spherical wave as a sum of elementary cylindrical waves. The result leads directly to the formula

$$\frac{\partial}{\partial z_i} T(\mathbf{r}_i; k_{\xi_m}, k_{\xi_n}) = -\exp\{j[k_{\xi_m} x_i + k_{\xi_n} y_i - z_i(k^2 - k_{\xi_m}^2 - k_{\xi_n}^2)^{1/2}]\}. \quad (9)$$

Using Eq. (9), we obtain from Eq. (7)

$$\begin{aligned} M(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \left(\frac{1}{2\pi}\right)^6 \int^{(6)} d\mathbf{k}_\xi \hat{M}_0(\mathbf{k}_\xi) \\ \times \exp(j[\mathbf{k}_{\xi_1} x_1 + k_{\xi_4} y_1 - z_1(k^2 - k_{\xi_1}^2 - k_{\xi_4}^2)^{1/2}] \\ - j\{(k_{\xi_1} - k_{\xi_2})x_2 + (k_{\xi_4} - k_{\xi_5})y_2 \\ - z_2[k^2 - (k_{\xi_1} - k_{\xi_2})^2 - (k_{\xi_4} - k_{\xi_5})^2]^{1/2}\} \\ + j\{(k_{\xi_3} - k_{\xi_2})x_3 + (k_{\xi_6} - k_{\xi_5})y_3 \\ - z_3[k^2 - (k_{\xi_3} - k_{\xi_2})^2 - (k_{\xi_6} - k_{\xi_5})^2]^{1/2}\} \\ - j[k_{\xi_3} x_4 + k_{\xi_6} y_4 - z_4(k^2 - k_{\xi_3}^2 - k_{\xi_6}^2)^{1/2}]). \end{aligned} \quad (10)$$

Again regrouping, we obtain the formal result

$$\begin{aligned} M(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \left(\frac{1}{2\pi}\right)^6 \int^{(6)} d\mathbf{k}_\xi \hat{M}_0(\mathbf{k}_\xi) \\ \times \exp(j\mathbf{k}_\xi \cdot \boldsymbol{\xi} - j\{z_1(k^2 - k_{\xi_1}^2 - k_{\xi_4}^2)^{1/2} \\ - z_2[k^2 - (k_{\xi_1} - k_{\xi_2})^2 - (k_{\xi_4} - k_{\xi_5})^2]^{1/2} \\ + z_3[k^2 - (k_{\xi_3} - k_{\xi_2})^2 - (k_{\xi_6} - k_{\xi_5})^2]^{1/2} \\ - z_4(k^2 - k_{\xi_3}^2 - k_{\xi_6}^2)^{1/2}\}). \end{aligned} \quad (11)$$

We can assume that $\hat{M}_0$ is nonzero only for $\mathbf{k}_\xi$ such that each of the square roots appearing in Eq. (11) is real. This restriction on the spectrum can be interpreted in terms of the angular spectrum of plane waves to mean[1] that evanescent waves are not included: These waves would rapidly decay in any case. In order to evaluate the integral we shall make a stronger assumption, however. We shall expand each of the square roots in series and retain only the first two terms of each expansion. If l is the transverse scale of the wave fluctuations on the initial plane, it is readily seen that this approximation requires

$$l > \lambda = 2\pi/k, \quad z\lambda^3/l^4 \ll 1. \quad (12)$$

These requirements appear to be readily satisfied for all cases involving waves which have been randomized by terrestrial and solar atmospheric effects.

Proceeding with the expansions and limiting ourselves to the important case of the transverse coherence function by taking $z_1 = z_2 = z_3 = z_4 = z$, we obtain

$$\begin{aligned} M(\boldsymbol{\xi}, z) &= \left(\frac{1}{2\pi}\right)^6 \int^{(6)} d\mathbf{k}_\xi \hat{M}_0(\mathbf{k}_\xi) \exp[j\mathbf{k}_\xi \cdot \boldsymbol{\xi} + (jz/k) \\ &\quad \times (k_{\xi_1} k_{\xi_2} - k_{\xi_3} k_{\xi_2} + k_{\xi_4} k_{\xi_5} - k_{\xi_6} k_{\xi_5})] \\ &= F^{-1}[\hat{M}_0(\mathbf{k}_\xi) h_x h_y] = M_0(\boldsymbol{\xi}) \stackrel{\star}{\star} F^{-1}\{h_x\} F^{-1}\{h_y\}, \end{aligned} \quad (13)$$

where we have employed the convolution theorem and defined

$$\begin{aligned} h_x &= \exp[(jz/k)(k_{\xi_1} k_{\xi_2} - k_{\xi_3} k_{\xi_2})], \\ h_y &= \exp[(jz/k)(k_{\xi_4} k_{\xi_5} - k_{\xi_6} k_{\xi_5})] \end{aligned} \quad (14)$$

We proceed to calculate the inverse transforms as follows

$$\begin{aligned} F^{-1}\{h_x\} &= \left(\frac{1}{2\pi}\right)^3 \int^{(3)} dk_{\xi_1} dk_{\xi_2} dk_{\xi_3} \\ &\quad \times \exp[j(k_{\xi_1}\xi_1 + k_{\xi_2}\xi_2 + k_{\xi_3}\xi_3) + \left(\frac{jz}{k}\right) \\ &\quad \times (k_{\xi_1} k_{\xi_2} - k_{\xi_3} k_{\xi_2})] \\ &= \left(\frac{1}{2\pi}\right)^2 \int^{(2)} dk_{\xi_1} dk_{\xi_3} \delta\left(\left(\frac{z}{k}\right)(k_{\xi_1} - k_{\xi_3}) + \xi_2\right) \\ &\quad \times \exp[j(k_{\xi_1}\xi_1 + k_{\xi_3}\xi_3)]. \end{aligned} \quad (15)$$

It is convenient to introduce the transformation

$$u = (z/k)(k_{\xi_1} - k_{\xi_3}) + \xi_2, \quad v = \tfrac{1}{2}(k_{\xi_1} + k_{\xi_3}), \quad (16)$$

which has the Jacobian (k/z). Consequently,

$$\begin{aligned} F^{-1}\{h_x\} &= \left(\frac{1}{2\pi}\right)\left(\frac{k}{z}\right) \exp\left[j(\xi_3 - \xi_1)\left(\xi_2 \frac{k}{2z}\right)\right] \int^{(2)} du\, dv \\ &\quad \times \delta(u) \exp\left\{j\left[(\xi_1 + \xi_3)v - (\xi_3 - \xi_1)\left(\frac{ku}{2z}\right)\right]\right\} \\ &= \left(\frac{k}{2\pi z}\right) \delta(\xi_1 + \xi_3) \exp\left[j(\xi_3 - \xi_1)\left(\frac{\xi_2 k}{2z}\right)\right]. \end{aligned} \quad (17)$$

Thus

$$\begin{aligned} M(\boldsymbol{\xi}, z) &= M_0(\boldsymbol{\xi}) \stackrel{\star}{\star} (k/2\pi z)^2 \delta(\xi_1 + \xi_3)\delta(\xi_4 + \xi_6) \\ &\quad \times \exp\{(jk/2z)[\xi_2(\xi_3 - \xi_1) + \xi_5(\xi_6 - \xi_4)]\} \\ &= (k/2\pi z)^2 \int^{(6)} d\boldsymbol{\xi}' M_0(\boldsymbol{\xi}') \delta(\xi_1 - \xi_1' + \xi_3 - \xi_3') \\ &\quad \times \delta(\xi_4 - \xi_4' + \xi_6 - \xi_6') \\ &\quad \times \exp\{(jk/2z)[(\xi_2 - \xi_2')(\xi_3 - \xi_3' - \xi_1 + \xi_1') \\ &\quad + (\xi_5 - \xi_5')(\xi_6 - \xi_6' - \xi_4 + \xi_4')]\}. \end{aligned} \quad (18)$$

The quantity of primary interest is the autocorrelation of irradiance $\langle I(p,q,z) I(0,0,z)\rangle$. Thus, taking $\xi_1 = \xi_3 = \xi_4 = \xi_6 = 0$, $\xi_2 = p$, $\xi_5 = q$, we obtain

$$\begin{aligned} &\langle I(p,q,z) I(0,0,z)\rangle \\ &\quad = (k/2\pi z)^2 \int^{(6)} d\boldsymbol{\xi}' M_0(\boldsymbol{\xi}') \delta(-\xi_1' - \xi_3') \delta(-\xi_4' - \xi_6') \\ &\quad \times \exp\{(jk/2z)[(p - \xi_2')(\xi_1' - \xi_3') \\ &\quad + (q - \xi_5')(\xi_4' - \xi_6')]\}. \end{aligned} \quad (19)$$

Integrating over ξ_3', ξ_6' and dropping the primes on the variables of integration

$$\begin{aligned}\langle I(p,q,z)I(0,0,z)\rangle &= (k/2\pi z)^2 \int^{(4)} d\xi_1 d\xi_2 d\xi_4 d\xi_5 \\ &\times M_0(\xi_1, \xi_2, -\xi_1, \xi_4, \xi_5, -\xi_4) \\ &\times \exp\{(jk/z)[\xi_1(p-\xi_2) + \xi_4(q-\xi_5)]\} \\ &= (k/2\pi z)^2 \int^{(4)} d\xi_1 d\xi_2 d\xi_4 d\xi_5 \\ &\times M_0(\xi_1, p-\xi_2, -\xi_1, \xi_4, q-\xi_5, -\xi_4') \\ &\times \exp\{(jk/z)(\xi_1\xi_2 + \xi_4\xi_5)\}. \end{aligned} \tag{20}$$

Experimental investigations and previous theoretical work in astronomy has been concerned with the spectrum of irradiance fluctuations,[3,4]

$$\hat{M}_{2I}(f_x, f_y; z) = F_{p,q}\{\langle I(p,q,z)I(0,0,z)\rangle - 1\}, \tag{21}$$

where $F_{p,q}$ indicates the Fourier transform with respect to p, q. Thus, from Eq. (20),

$$\begin{aligned}\hat{M}_{2I} &= (k/2\pi z)^2 \int^{(6)} dp\,dq\; d\xi_1 d\xi_2 d\xi_4 d\xi_5 \\ &\times M_0(\xi_1, p-\xi_2, -\xi_1, \xi_4, q-\xi_5, -\xi_4) \\ &\times \exp\{j[k/z)(\xi_1\xi_2 + \xi_4\xi_5) - pf_x - qf_y]\} \\ &- (2\pi)^2 \delta(f_x)\delta(f_y). \end{aligned} \tag{22}$$

We interchange the order of integration and set $p - \xi_2 = \alpha$, $q - \xi_5 = \beta$ in the inner integral. Thus

$$\begin{aligned}\hat{M}_{2I} &= (k/2\pi z)^2 \int^{(4)} d\xi_1 d\xi_2 d\xi_4 d\xi_5 \\ &\times \exp\{j[k/z)(\xi_1\xi_2 + \xi_4\xi_5) - \xi_2 f_x - \xi_5 f_y]\} \\ &\times \int^{(2)} d\alpha d\beta\, M(\xi_1, \alpha, -\xi_1, \xi_4, \beta, -\xi_4) \\ &\times \exp[-j(\alpha f_x + \beta f_y)] - (2\pi)^2 \delta(f_x)\delta(f_y). \end{aligned} \tag{23}$$

We again reorder the integration

$$\begin{aligned}\hat{M}_{2I} &= (k/2\pi z)^2 \int^{(4)} d\xi_1 d\xi_4 d\alpha d\beta \\ &\times M_0(\xi_1, \alpha, -\xi_1, \xi_4, \beta, -\xi_4) \exp[-j(\alpha f_x + \beta f_y)] \\ &\times \int^{(2)} d\xi_2 d\xi_5 \exp j\left[\left(\frac{k\xi_1}{z - f_x}\right)\xi_2 + \left(\frac{k\xi_4}{z - f_y}\right)\xi_5\right] \\ &- (2\pi)^2 \delta(f_x)\delta(f_y). \end{aligned} \tag{24}$$

The inner integrals are now recognizable as delta functions and we obtain the remarkably simple result

$$\begin{aligned}\hat{M}_{2I} &= \int^{(2)} d\alpha d\beta \left[M_0\left(\frac{zf_x}{k}, \alpha, \frac{-zf_x}{k}, \frac{zf_y}{k}, \beta, \frac{-zf_y}{k}\right) - 1\right] \\ &\times \exp[-j(\alpha f_x + \beta f_y)] \\ &= F_{\alpha,\beta}\left[M_0\left(\frac{zf_x}{k}, \alpha, \frac{-zf_x}{k}, \frac{zf_y}{k}, \beta, \frac{-zf_y}{k}\right) - 1\right]. \end{aligned} \tag{25}$$

Equation (25) may be written in terms of the original variables in the form

$$\begin{aligned}\hat{M}_{2I} = F_{\alpha,\beta}[M_0(&x_1 - x_2 = x_4 - x_3 = zf_x/k, x_2 - x_3 \\ &= x_1 - x_4 = \alpha,\; y_1 - y_2 = y_4 - y_3 = zf_y/k, y_2 - y_3 \\ &= y_1 - y_4 = \beta) - 1]. \end{aligned} \tag{26}$$

Equations (20) and (25) [or (26)] are the basic general formulas. Before continuing on to obtain some general conclusions, we digress to consider the case of a wave which has weak random phase fluctuations across the initial plane. This "weak phase screen" case has been partially analyzed directly from the eightfold integral. We show that our general formulas lead to the same results and, additionally, obtain some new results for the autocorrelation of irradiance fluctuations.

WEAK PHASE SCREEN

Assuming unit amplitude and random phase S on the $z = 0$ plane

$$\begin{aligned}M_0 &= \langle \psi_1\psi_2^*\psi_3\psi_4^*\rangle = \langle \exp j[(S_1 + S_3) - (S_2 + S_4)]\rangle \\ &= \exp -\tfrac{1}{2}\langle S_1^2 + S_2^2 + S_3^2 + S_4^2 \\ &\quad + 2(S_1S_3 + S_2S_4 - S_1S_2 - S_1S_4 - S_2S_3 - S_3S_4)\rangle, \end{aligned} \tag{27}$$

where we assume Gaussian variables. We define the phase autocorrelation function ρ_{ij},

$$\langle S_iS_j\rangle = S_0^2\rho_{ij}, \tag{28}$$

where $\rho_{ii} = 1$ and

$$\rho_{ij} = \rho([(x_i - x_j)^2 + (y_i - y_j)^2]^{1/2}). \tag{29}$$

Thus

$$M_0 = \exp[-S_0^2(2 + \rho_{13} + \rho_{24} - \rho_{12} - \rho_{23} - \rho_{34} - \rho_{41})]. \tag{30}$$

It should be observed that

$$\rho_{ij} = \rho([(x_i - x_j)^2 + (y_i - y_j)^2]^{1/2}). \tag{31}$$

Thus using our definitions of ξ, and noting that we must take $\xi_3 = -\xi_1$, $\xi_6 = -\xi_4$, we find

$$\begin{aligned}\rho_{13} &= \rho([(\xi_2 + \xi_1)^2 + (\xi_5 + \xi_4)^2]^{1/2}), \\ \rho_{24} &= \rho([(\xi_2 - \xi_1)^2 + (\xi_5 - \xi_4)^2]^{1/2}), \\ \rho_{12} &= \rho_{34} = \rho((\xi_1^2 + \xi_4^2)^{1/2}), \\ \rho_{23} &= \rho_{14} = \rho((\xi_2^2 + \xi_5^2)^{1/2}). \end{aligned} \tag{32}$$

Whence

$$M_0 = \exp[S_0^2(-2 + 2\rho_{12} + 2\rho_{23} - \rho_{13} - \rho_{24})]. \tag{33}$$

For a weak phase screen $S_0 < 1$, we write

$$M_0 \approx 1 + S_0^2(-2 + 2\rho_{12} + 2\rho_{23} - \rho_{13} - \rho_{24}). \tag{34}$$

and obtain for the intensity fluctuations, using (34) with $\xi_2 \to -\xi_2$, $\xi_5 \to -\xi_5$, $p = q = 0$,

$$\begin{aligned}\langle I^2(z)\rangle &= (k/2\pi z)^2 S_0^2 \int^{(4)} d\xi_1 d\xi_2 d\xi_4 d\xi_5 \\ &\times \exp[(jk/z)(\xi_1\xi_2 + \xi_4\xi_5)]\{(S_0^2 - 2) \\ &+ 2\rho((\xi_1^2 + \xi_4^2)^{1/2}) + 2\rho((\xi_2^2 + \xi_5^2)^{1/2}) \\ &- \rho([(\xi_1 - \xi_2)^2 + (\xi_4 - \xi_5)^2]^{1/2}) \\ &- \rho([(\xi_1 + \xi_2)^2 + (\xi_4 + \xi_5)^2]^{1/2})\}. \end{aligned} \tag{35}$$

The five fourfold integrals which result from the five terms in brackets are labeled $J_1, \ldots, J_5$ and can be evaluated as follows:

$$\begin{aligned}J_1 &= (1 - 2S_0^2)(k/2\pi z)^2 \int^{(4)} d\xi_1 d\xi_2 d\xi_4 d\xi_5\, e^{(jk/z)(\xi_1\xi_2 + \xi_4\xi_5)} \\ &= (1 - 2S_0^2 \int^{(2)} d\xi_2 d\xi_5 \delta(\xi_2)\delta(\xi_5) = 1 - 2S_0^2; \end{aligned} \tag{36}$$

$$J_2 = 2(k/2\pi z)^2 S_0^2 \int^{(4)} d\xi_1 d\xi_2 d\xi_4 d\xi_5 \rho((\xi_1^2 + \xi_4^2)^{1/2}) \times \exp[(jk/z)(\xi_1\xi_2 + \xi_4\xi_5)] = 2S_0^2 \int^{(2)} d\xi_1 d\xi_4 \rho((\xi_1^2 + \xi_4^2)^{1/2})\delta(\xi_1)\delta(\xi_4) = 2S_0^2. \tag{37}$$

Similarly,

$$J_3 = 2(k/2\pi z)^2 S_0^2 \int^{(4)} d\xi_1 d\xi_2 d\xi_4 d\xi_5 \rho((\xi_2^2 + \xi_5^2)^{1/2}) \times \exp[(jk/z)(\xi_1\xi_2 + \xi_4\xi_5)] = 2S_0^2 \int^{(2)} d\xi_2 d\xi_5 \rho((\xi_2^2 + \xi_5^2)^{1/2})\delta(\xi_2)\delta(\xi_5) = 2S_0^2. \tag{38}$$

Next, setting $A = \xi_1 - \xi_2$, $B = \xi_1 + \xi_2$, $C = \xi_4 - \xi_5$, $D = \xi_4 + \xi_5$,

$$\begin{aligned} J_4 &= -(k/2\pi z)^2 S_0^2 \int^{(4)} d\xi_1 d\xi_2 d\xi_4 d\xi_5 \rho([(\xi_1 - \xi_2)^2 + (\xi_4 - \xi_5)^2]^{1/2}) \exp[(jk/z)(\xi_1\xi_2 + \xi_4\xi_5)] \\ &= -\tfrac{1}{4}(k/2\pi z)^2 S_0^2 \int^{(2)} dA dC\ \rho((A^2 + C^2)^{1/2}) \times \exp[-(jk/4z)(A^2 + C^2)] \\ &\quad \times \int^{(2)} dB dD\ \exp[(jk/4z)(B^2 + D^2)] \\ &= \tfrac{1}{2}(k/2\pi z) j S_0^2 \int^{(2)} dA dC\, \rho((A^2 + C^2)^{1/2}) \times \exp[-(jk/4Z)(A^2 + C^2)]. \end{aligned} \tag{39}$$

Similarly,

$$J_5 = -\tfrac{1}{4}(k/2\pi z)^2 S_0^2 \int^{(2)} dB dD\, \rho((B^2 + D^2)^{1/2}) \times \exp[(jk/4z)(B^2 + D^2)] \times \int^{(2)} dA dC\ \exp[-(jk/4z)(A^2 + C^2)] = J_4^*. \tag{40}$$

Thus, transforming to polar coordinates in the (A, C) plane,

$$\begin{aligned} J_4 + J_5 &= -(k/z)S_0^2\, \mathrm{Re}\left[j \int_0^\infty r dr\ \rho(r)\ \exp\left(\frac{-jkr^2}{4z}\right)\right] \\ &= -(k/z)S_0^2 \int_0^\infty r dr\ \rho(r)\ \sin\left(\frac{kr^2}{4z}\right). \end{aligned} \tag{41}$$

Combining and setting $t = kr^2/4z$, we obtain the general result

$$\langle I^2(z)\rangle = 1 + 2S_0^2\left(1 - \int_0^\infty \rho(2\sqrt{zt/k})\ \sin t\, dt\right). \tag{42}$$

For the Gaussian model $\rho(r) = \exp(-r^2/l^2)$, we obtain

$$\langle I^2(z)\rangle = 1 + \frac{2S_0^2(4z/kl^2)^2}{[1 + (4z/kl^2)^2]}, \tag{43}$$

which was obtained by Mercier[2] from the eightfold integral.

To obtain the intensity autocorrelation $\langle I(p,q,z) I(0,0,z)\rangle$ in the transverse plane we need only recalculate Eq. (35), substituting $\xi_2 - p$ and $\xi_5 - q$ for ξ_2 and ξ_5, respectively, wherever they appear in the argument of ρ in the integral in Eq. (35). Thus

$$\begin{aligned} \langle I(p,q,z)I(0,0,z)\rangle &= (k/2\pi z)^2 S_0^2 \int^{(4)} d\xi_1 d\xi_2 d\xi_4 d\xi_5 \\ &\quad \times \exp[(jk/z)(\xi_1\xi_2 + \xi_4\xi)] \\ &\quad \times \{(S_0^{-2} - 2) + 2\rho((\xi_1^2 + \xi_4^2)^{1/2}) \\ &\quad + 2\rho([(\xi_2 - p)^2 + (\xi_5 - q)^2]^{1/2}) \\ &\quad - \rho([(\xi_1 - \xi_2 - p)^2 + (\xi_4 - \xi_5 + q)^2]^{1/2}) \\ &\quad - \rho([(\xi_1 + \xi_2 - p)^2 + (\xi_4 + \xi_5 - q)^2]^{1/2})\}. \end{aligned} \tag{44}$$

The five fourfold integrals which result from the five terms in brackets are labeled $J_{1pq}, \ldots, J_{5pq}$ and are evaluated as follows: Clearly

$$\begin{aligned} J_{1pq} &= J_1 = 1 - 2S_0^2, \\ J_{2pq} &= J_2 = 2S_0^2, \\ J_{3pq} &= 2S_0^2 \rho((p^2 + q^2)^{1/2}). \end{aligned} \tag{45}$$

Comparing the terms in Eq. (44) with our previous expressions we find, using Eq. (39),

$$J_{4pq} = \frac{-j}{2}\left(\frac{k}{2\pi z}\right) S_0^2 \int^{(2)} dA dC\, \rho\{[(A + p)^2 + (C + q)^2]^{1/2}\} \exp\left[-\left(\frac{jk}{4z}\right)(A^2 + C^2)\right] \tag{46}$$

$$J_{5pq} = +\frac{j}{2}\left(\frac{k}{2\pi z}\right) S_0^2 \int^{(2)} dB dD\, \rho([(B - p)^2 + (D - q)^2]^{1/2}) \exp\left[\left(\frac{jk}{4z}\right)(B^2 + D^2)\right], \tag{47}$$

Thus, continuing

$$\langle I(p,q,z)I(0,0,z)\rangle = 1 + 2S_0^2\left[\rho((p^2 + q^2)^{1/2}) - \frac{1}{2}\frac{k}{2\pi z} \times \int^{(2)} dA dC\ \rho([(A + p)^2 + (C + q)^2]^{1/2}) \times \sin\left(\frac{k}{4z}(A^2 + C^2)\right)\right]. \tag{48}$$

For a Gaussian model, the integrals involved in Eq. (48) may be readily evaluated using the following formula which may be deduced from the tables (Ref. 5, p. 485):

$$\int_{-\infty}^{\infty} e^{-x^2} \begin{matrix}\sin\\ \cos\end{matrix} [\xi(x - \alpha)^2] dx = \pi^{1/2}(1 + \xi^2)^{-1/4} \times \exp\left(-\frac{\xi^2\alpha^2}{1 + \xi^2}\right) \begin{matrix}\sin\\ \cos\end{matrix} \left(\frac{1}{2}\tan^{-1}\xi + \frac{\xi\alpha^2}{1 + \xi^2}\right). \tag{49}$$

Thus with $\rho = \exp\{-[(A + p)^2 + (C + q)^2]/l^2\}$, we obtain for the weak phase screen

$$\begin{aligned} \langle I(p,q,z)I(0,0,z)\rangle &= 1 + 2S_0^2\left[\exp\left(\frac{-(p^2 + q^2)}{l^2}\right)\right. \\ &\quad - \xi(1 + \xi^2)^{-1/2} \exp\left(\frac{-\xi}{1 + \xi^2}\cdot\frac{(p^2 + q^2)}{l^2}\right) \\ &\quad \left.\times \sin\left(\tan^{-1}\xi + \frac{\xi}{1 + \xi^2}\cdot\frac{(p^2 + q^2)}{l^2}\right)\right], \end{aligned} \tag{50}$$

where we have introduced the notation $\xi = kl^2/4z$. For $p = q = 0$, we obtain the special result given by Eq. (43). It may be noted in passing that except for the special results given in Eqs. (41) and (43), our formulation is correct for anisotropic weak phase screens. That is, Eq. (48) may be used with $\rho(p,q)$ instead of $\rho((p^2 + q^2)^{1/2})$ and $\rho(A + p, C + q)$ instead of $\rho((A + p)^2 + (C + q)^2)$. For the anisotropic model $\rho(x,y) = \exp(-x^2/l_x^2 - y^2/l_y^2)$, we obtain, using Eq. (48), the result

$$\langle I(p,q,z)I(0,0,z)\rangle = 1 + 2S_0^2\left[\exp\left(\frac{-p^2}{l_x^2} - \frac{q^2}{l_y^2}\right) - \left(\frac{\xi_x^2\xi_y^2}{(1 + \xi_x^2)(1 + \xi_y^2)}\right)^{-1/4} \exp\left(-\frac{\xi_x^2 p^2}{l_x^2(1 + \xi_x^2)} - \frac{\xi_y^2 q^2}{l_y^2(1 + \xi_y^2)}\right)\right.$$

$$\times \sin\left(\tfrac{1}{2}\tan^{-1}\xi_x + \tfrac{1}{2}\tan^{-1}\xi_y + \frac{\xi_x p^2}{l_x^2(1+\xi_x^2)} + \frac{\xi_y^2 q_y^2}{l_y^2(1+\xi_y^2)}\right)\Bigg], \tag{51}$$

where we have set $\xi_x = kl_x^2/4z$, $\xi_y = kl_y^2/4z$. Letting $l_y \to \infty$, we obtain a result for the one-dimensional Gaussian model

$$\langle I(p,z)I(0,z)\rangle = 1 + 2S_0^2\left[\exp\left(-\frac{p^2}{l_x^2}\right)\left(\frac{\xi_x^2}{(1+\xi_x^2)}\right)^{1/4} \times \exp\left(\frac{-\xi_x^2 p^2}{l_x^2(1+\xi_x^2)}\right)\sin\left(\tfrac{1}{2}\tan^{-1}\xi + \frac{\pi}{4} + \frac{\xi_x p^2/l_x^2}{(1+\xi_x^2)}\right)\right]. \tag{52}$$

The formulas given by (50)–(52) have not to our knowledge appeared previously.

The spectrum of irradiance fluctuations may be obtained from Eq. (25). Thus,using Eq. (33), we write

$$\hat{M}_{2I} = S_0^2 \int^{(2)} d\alpha d\beta \left\{\left[-2 + 2\rho\left(\frac{zf_x}{k}, \frac{zf_y}{k}\right) + 2\rho(\alpha,\beta) - \rho\left(\frac{zf_x}{k} - \alpha, \frac{zf_y}{k} - \beta\right) - \rho\left(\frac{zf_x}{k} + \alpha, \frac{zf_y}{k} + \beta\right)\right] \times \exp[-j(\alpha f_x + \beta f_y)]\right\}. \tag{53}$$

The first two terms cancel and we readily obtain, using the shift theorem,

$$\hat{M}_{2I}(f_x,f_y,z) = 4S_0^2\hat{\rho}(f_x,f_y)\sin^2[z(f_x^2+f_y^2)/2k]. \tag{54}$$

Equation (54) is referred to in numerous recent radio-astronomy papers. The formula was first stated by Salpeter[3] for the case $z \gg \lambda$. In the only published derivation, Jokipii[4] obtained (54) directly from the fourfold integral for the one-dimensional phase screen, assuming that $\hat{\rho}$ is cut off for wavenumbers greater than order kL/z, where L is the correlation length of the phase fluctuations. In Ref. 4 a result is also obtained for the spectrum of irradiance fluctuations for a strong phase screen. We shall also obtain the generalization of this result in a later section.

LIMITING CASE: GEOMETRICAL OPTICS

We consider the case of large k/z in which we assume $(zf_x/k) \to 0$ in the limit. We immediately obtain from Eq. (25), upon applying the Taylor expansion about $z = 0$ and using the symmetry of M_0,

$$\hat{M}_{2I} = \int^{(2)} d\alpha d\beta \left[M_0(0,\alpha,0,0,\beta,0) - 1 + \left(\frac{\partial^2 M_0}{\partial\xi_1^2} - \frac{\partial^2 M_0}{\partial\xi_1\partial\xi_3}\right)_0\left(\frac{zf_x}{k}\right)^2 + \left(\frac{\partial^2 M_0}{\partial\xi_4^2} - \frac{\partial^2 M_0}{\partial\xi_4\partial\xi_6}\right)_0\left(\frac{zf_y}{k}\right)^2\right]\exp[-j(\alpha f_x + \beta f_y)]$$

$$= \hat{M}_{2I}(z=0) + \left(\frac{z}{k}\right)^2 F\left[\left(\frac{\partial^2 M_0}{\partial\xi_1^2} - \frac{\partial^2 M_0}{\partial\xi_1\partial\xi_3}\right)_0 f_x^2 + \left(\frac{\partial^2 M_0}{\partial\xi_4^2} - \frac{\partial^2 M_0}{\partial\xi_4\partial\xi_6}\right)_0 f_y^2\right]. \tag{55}$$

[In Eq. (55) the zero subscript applied to the parentheses implies $\xi_1 = \xi_3 = \xi_4 = \xi_6 = 0$].

Because the Fourier transform in the second term in Eq. (55) does not contain z or k, we can assert the general result that $\hat{M}_{2I}$ will increase as $(z/k)^2$ in the initial region. For the weak phase screen, this result can be obtained directly from Eq. (54) as noted by Salpeter. The geometrical optics z^2 dependence has also been well known in shadowgraph theory.[6,7]

LIMITING CASE: THE DIFFRACTION REGION

For a phase-screen, as z increases the fourth-order coherence function of Eq. (33) as it appears in the integral of Eq. (25), can be approximated by

$$M_0\left(\frac{zf_x}{k}, \alpha, \frac{-zf_x}{k}, \frac{zf_y}{k}, \beta, \frac{-zf_y}{k}\right) \approx e^{-2S_0^2}\left\{e^{2S_0^2\rho(\alpha,\beta)} + \exp\left[-S_0^2\rho\left(\alpha + \frac{zf_x}{k}, \beta + \frac{zf_y}{k}\right)\right] + \exp\left[-S_0^2\rho\left(\alpha - \frac{zf_x}{k}, \beta - \frac{zf_y}{k}\right)\right] - 2\right\}. \tag{56}$$

This approximation is valid when $e^{-2S_0^2\rho(zf_x/k, zf_y/k)} \approx 1$, and we can consider Eq. (33) to be a product of "noninteracting" factors. Using a Gaussian model to simplify the expression of this condition, we find that we require $z \gg (kl/f)\ln S_0$. This condition clearly restricts our approximation to the diffraction region. Applying Eq. (56) to Eq. (25), we readily obtain, upon application of the shift theorem and some rearrangement of terms,

$$\hat{M}_{2I}(z \gg (kl/f)\ln S_0) \approx F(e^{2S_0^2[\rho(\alpha,\beta)-1]} + 2e^{-S_0^2\rho(\alpha,\beta)} - 2e^{-2S_0^2} - 1) - 4\sin^2[(f_x^2+f_y^2)z/2k]F(e^{-S_0^2\rho(\alpha,\beta)}) \approx \hat{M}_{2I}(f=0) - 4\sin^2[(f_x^2+f_y^2)z/2k]F(e^{-S_0^2\rho(\alpha,\beta)}). \tag{57}$$

The zero spatial frequency term in this equation will not be exact since it cannot fall within the range of our approximation. Thus we write instead

$$M_{2I}[z \gg (kl/f)\ln S_0] \approx -4\sin^2[(f_x^2+f_y^2)z/2k] \times F(e^{-S_0^2\rho(\alpha,\beta)}). \tag{58}$$

Equation (58) is the generalization of the Salpeter formula, Eq.(54), of the general (not-weak) phase screen. We point out that, although we have restricted our discussion here to a phase screen, it is possible to extend the formula to an amplitude-phase screen by expressing the wave in terms of a complex amplitude μ and writing the coherence function in terms of the autocorrelation of the μ, as done by Mercier. It is clear that this procedure will again lead to a separation of the effects of distance z from the initial plane into the Fresnel filter $4\sin^2[(f_x^2 + f_y^2)z/2k]$ for high spatial frequencies.

A result similar to that derived here was obtained by Jokipii[4] for a one-dimensional phase screen subject to a number of special assumptions. Equation (58) suggests that the most convenient procedure to follow in the analysis of irradiance fluctuations created by distant randomization is to restrict attention to the high frequency portion of spectrum, $[f \gg (kl/z)\ln S_0]$ in which the effects of distance are most simply treated.

CONCLUSION

Equations (20) and (25) express the autocorrelation and spectrum of intensity fluctuations for a ran-

domized field as transforms of the fourth-order coherence on the initial planes. (These results have been generalized to higher order moments of the field.[8]) The special formulas given by Eqs. (58) and (55) are useful in radio-astronomical observations of the solar plasma and in optical phenomena due to high altitude turbulence, respectively. The intermediate case (corresponding to the Fresnel distance $z = kl/f$, or the middle of the swimming pool described earlier) requires numerical integration for each particular statistical model of the initial field.

* Research supported in part by the Atmospheric Sciences Section, National Science Foundation, NSF Grant GA-20776.

1 D. J. Torrieri and L. S. Taylor, J. Math Phys. **12**, 1211 (1971).
2 R. P. Mercier, Proc. Cambridge Phil. Soc. A **58**, 382 (1962).
3 E. E. Salpeter, Astrophys. J. **147**, 433 (1967).
4 J. R. Jokipii, Astrophys. J. **161**, 1147 (1970).
5 I. S. Gradshteyn and I. M. Rhyzhik *Table of Integrals, Series and Products*, transl. by A. Jeffrey (Academic, New York, 1965).
6 Weyl, F. J., "Analytical Methods in Optical Examination of Supersonic Flow," Naval Ordnance Laboratory, Rept. No. 211045 (1945).
7 L. S. Taylor, Amer. Inst. Aeronaut. J. 8, 1284 (1970).
8 D. Lekhyananda, thesis (University of Maryland, 1972).

Exact Solutions to the Yang-Mills Field Equations*

Gerald Rosen
Drexel University, Philadelphia, Pennsylvania 19104
(Received 6 July 1971)

New spatially localized solutions to the Yang–Mills classical field equations are reported.

Manifesting singularities of the Coulomb form, the spatially localized "particlelike" solutions to the Yang–Mills classical field equations obtained several years ago[1–3] yield infinite values for the total field energy. A systematic search for particlelike solutions of finite energy has been performed by the present author with methods developed previously for essentially nonlinear field theories.[4] Although the investigation has disclosed the new and interesting solutions reported here, no particlelike solution of finite energy has been found. Our results support the conjecture that spatially localized solutions of finite energy do not exist in a classical Yang–Mills field theory.

We work with the Yang–Mills classical field equations in the form[5]

$$\partial F^{\mu\nu}/\partial x^{\nu} = ie[A_{\nu}, F^{\mu\nu}], \tag{1}$$

$$F_{\mu\nu} \equiv \partial A_{\mu}/\partial x^{\nu} - \partial A_{\nu}/\partial x^{\mu} + ie[A_{\mu}, A_{\nu}], \tag{2}$$

where indices are raised with the Minkowskian metric tensor $g^{\mu\nu} \equiv \text{diag}\,(-1, 1, 1, 1)$, e denotes the fundamental unit of charge in a system of physical units with $\hbar$ and c equal to unity, and the components of $A_{\mu} = A_{\mu}(x)$ and $F_{\mu\nu} = F_{\mu\nu}(x)$ are complex $\mathbf{n} \times \mathbf{n}$ matrices that transform according to the formulas

$$A_{\mu} \xrightarrow{S} A'_{\mu} \equiv S^{-1}A_{\mu}S + ie^{-1}S^{-1}\partial S/\partial x^{\mu}, \tag{3}$$

$$F_{\mu\nu} \xrightarrow{S} F'_{\mu\nu} \equiv S^{-1}F_{\mu\nu}S, \tag{4}$$

where $S = S(x)$ is an element of the general gauge group of all nonsingular complex $\mathbf{n} \times \mathbf{n}$ matrix functions of x, or a subgroup thereof. Under the variation of A_{μ}, the field equations (1) follow from the gauge-invariant Lagrangian density

$$\mathcal{L} = -(1/4n)\,\text{tr}(F^{\mu\nu}F_{\mu\nu}), \tag{5}$$

where tr denotes the trace of the matrix and the numerical prefactor $-(4n)^{-1}$ is implied by the purely electromagnetic field case[5] $F_{\mu\nu} = f_{\mu\nu}\mathbb{1}$, with $\mathbb{1}$ the $\mathbf{n} \times \mathbf{n}$ identity matrix. Associated with (5) we have the canonical total field energy

$$E = \frac{1}{2n}\int \text{tr}(F_{0i}F_{0i} + \tfrac{1}{2}F_{ij}F_{ij})d^3x, \tag{6}$$

a gauge-invariant complex constant of the motion in the general case, but patently real and positive for Hermitian A_{μ} and $F_{\mu\nu}$.

The computation of solutions to Eqs. (1) and (2) is facilitated by first applying a gauge transformation to the fields with

$$S(x) = T\left(\exp ie\int_0^{x^0} A_0(y)dy^0\right)_{\mathbf{y}=\mathbf{x}} \tag{7}$$

in which T is the chronological product ordering operator, so that $A'_0 = 0$ at all x according to (3) with (7). Dropping the primes, with $A_0 = 0$ Eqs. (1) and (2) split up to produce the simpler equations

$$-\ddot{A}_i + \frac{\partial F_{ij}}{\partial x_j} = ie[A_j, F_{ij}], \tag{8}$$

$$\frac{\partial \dot{A}_i}{\partial x_i} = ie[A_i, \dot{A}_i] \tag{9}$$

$$F_{0i} = -\dot{A}_i, \tag{10}$$

$$F_{ij} = \frac{\partial A_i}{\partial x_j} - \frac{\partial A_j}{\partial x_i} + ie[A_i, A_j], \tag{11}$$

where the Latin indices run 1, 2, 3 and a dot denotes differentiation with respect to x^0.

Particlelike solutions to Eqs. (8) and (9) take the form

$$A_i = (J + fx^0K)\frac{\partial \phi}{\partial x_i}, \tag{12}$$

where J and K are constant complex $\mathbf{n} \times \mathbf{n}$ nonzero matrices such that

$$[J, K] = iK, \tag{13}$$

f is a disposable real constant, and $\phi = \phi(\mathbf{x})$ in (12) is a real scalar function independent of x^0. We have $F_{ij} = 0$ from (11) with the form (12), and hence (8) is satisfied identically. Equation (9) is satisfied by (12) with (13) provided that ϕ is a solution to the nonlinear Poisson equation

$$\nabla^2\phi + e|\nabla\phi|^2 = 0. \tag{14}$$

domized field as transforms of the fourth-order coherence on the initial planes. (These results have been generalized to higher order moments of the field.[8]) The special formulas given by Eqs. (58) and (55) are useful in radio-astronomical observations of the solar plasma and in optical phenomena due to high altitude turbulence, respectively. The intermediate case (corresponding to the Fresnel distance $z = kl/f$, or the middle of the swimming pool described earlier) requires numerical integration for each particular statistical model of the initial field.

* Research supported in part by the Atmospheric Sciences Section, National Science Foundation, NSF Grant GA-20776.

1 D. J. Torrieri and L. S. Taylor, J. Math Phys. **12**, 1211 (1971).
2 R. P. Mercier, Proc. Cambridge Phil. Soc. A **58**, 382 (1962).
3 E. E. Salpeter, Astrophys. J. **147**, 433 (1967).
4 J. R. Jokipii, Astrophys. J. **161**, 1147 (1970).
5 I. S. Gradshteyn and I. M. Rhyzhik *Table of Integrals, Series and Products*, transl. by A. Jeffrey (Academic, New York, 1965).
6 Weyl, F. J., "Analytical Methods in Optical Examination of Supersonic Flow," Naval Ordnance Laboratory, Rept. No. 211045 (1945).
7 L. S. Taylor, Amer. Inst. Aeronaut. J. 8, 1284 (1970).
8 D. Lekhyananda, thesis (University of Maryland, 1972).

Exact Solutions to the Yang-Mills Field Equations*

Gerald Rosen
Drexel University, Philadelphia, Pennsylvania 19104
(Received 6 July 1971)

New spatially localized solutions to the Yang–Mills classical field equations are reported.

Manifesting singularities of the Coulomb form, the spatially localized "particlelike" solutions to the Yang–Mills classical field equations obtained several years ago[1–3] yield infinite values for the total field energy. A systematic search for particlelike solutions of finite energy has been performed by the present author with methods developed previously for essentially nonlinear field theories.[4] Although the investigation has disclosed the new and interesting solutions reported here, no particlelike solution of finite energy has been found. Our results support the conjecture that spatially localized solutions of finite energy do not exist in a classical Yang–Mills field theory.

We work with the Yang–Mills classical field equations in the form[5]

$$\partial F^{\mu\nu}/\partial x^{\nu} = ie[A_{\nu}, F^{\mu\nu}], \tag{1}$$

$$F_{\mu\nu} \equiv \partial A_{\mu}/\partial x^{\nu} - \partial A_{\nu}/\partial x^{\mu} + ie[A_{\mu}, A_{\nu}], \tag{2}$$

where indices are raised with the Minkowskian metric tensor $g^{\mu\nu} \equiv \text{diag}\,(-1, 1, 1, 1)$, e denotes the fundamental unit of charge in a system of physical units with $\hbar$ and c equal to unity, and the components of $A_{\mu} = A_{\mu}(x)$ and $F_{\mu\nu} = F_{\mu\nu}(x)$ are complex $\mathbf{n} \times \mathbf{n}$ matrices that transform according to the formulas

$$A_{\mu} \xrightarrow{S} A'_{\mu} \equiv S^{-1}A_{\mu}S + ie^{-1}S^{-1}\partial S/\partial x^{\mu}, \tag{3}$$

$$F_{\mu\nu} \xrightarrow{S} F'_{\mu\nu} \equiv S^{-1}F_{\mu\nu}S, \tag{4}$$

where $S = S(x)$ is an element of the general gauge group of all nonsingular complex $\mathbf{n} \times \mathbf{n}$ matrix functions of x, or a subgroup thereof. Under the variation of A_{μ}, the field equations (1) follow from the gauge-invariant Lagrangian density

$$\mathcal{L} = -(1/4n)\,\text{tr}(F^{\mu\nu}F_{\mu\nu}), \tag{5}$$

where tr denotes the trace of the matrix and the numerical prefactor $-(4n)^{-1}$ is implied by the purely electromagnetic field case[5] $F_{\mu\nu} = f_{\mu\nu}\mathbb{1}$, with $\mathbb{1}$ the $\mathbf{n} \times \mathbf{n}$ identity matrix. Associated with (5) we have the canonical total field energy

$$E = \frac{1}{2n}\int \text{tr}(F_{0i}F_{0i} + \tfrac{1}{2}F_{ij}F_{ij})d^3x, \tag{6}$$

a gauge-invariant complex constant of the motion in the general case, but patently real and positive for Hermitian A_{μ} and $F_{\mu\nu}$.

The computation of solutions to Eqs. (1) and (2) is facilitated by first applying a gauge transformation to the fields with

$$S(x) = T\left(\exp ie\int_0^{x^0} A_0(y)dy^0\right)_{\mathbf{y}=\mathbf{x}} \tag{7}$$

in which T is the chronological product ordering operator, so that $A'_0 = 0$ at all x according to (3) with (7). Dropping the primes, with $A_0 = 0$ Eqs. (1) and (2) split up to produce the simpler equations

$$-\ddot{A}_i + \frac{\partial F_{ij}}{\partial x_j} = ie[A_j, F_{ij}], \tag{8}$$

$$\frac{\partial \dot{A}_i}{\partial x_i} = ie[A_i, \dot{A}_i] \tag{9}$$

$$F_{0i} = -\dot{A}_i, \tag{10}$$

$$F_{ij} = \frac{\partial A_i}{\partial x_j} - \frac{\partial A_j}{\partial x_i} + ie[A_i, A_j], \tag{11}$$

where the Latin indices run 1, 2, 3 and a dot denotes differentiation with respect to x^0.

Particlelike solutions to Eqs. (8) and (9) take the form

$$A_i = (J + fx^0K)\frac{\partial \phi}{\partial x_i}, \tag{12}$$

where J and K are constant complex $\mathbf{n} \times \mathbf{n}$ nonzero matrices such that

$$[J, K] = iK, \tag{13}$$

f is a disposable real constant, and $\phi = \phi(\mathbf{x})$ in (12) is a real scalar function independent of x^0. We have $F_{ij} = 0$ from (11) with the form (12), and hence (8) is satisfied identically. Equation (9) is satisfied by (12) with (13) provided that ϕ is a solution to the nonlinear Poisson equation

$$\nabla^2\phi + e|\nabla\phi|^2 = 0. \tag{14}$$

We obtain

$$\phi = e^{-1} \ln(1 + ar^{-1}) \tag{15}$$

as the immediate spherically symmetric solution to (14), where $r \equiv (x_1^2 + x_2^2 + x_3^2)^{1/2}$ and a denotes a disposable real constant of integration in (15). By putting (12) with (15) into (10), we find that

$$F_{0i} = fe^{-1}a(r + a)^{-1}r^{-2}x_i K, \tag{16}$$

and the energy (6) is then obtained by explicit integration as

$$E = \frac{2\pi f^2 a}{ne^2} \operatorname{tr}(K^2) \tag{17}$$

for $a > 0$. In the case $n = 2$, an obvious representation for matrices J and K satisfying (13) is

$$J = \begin{pmatrix} i & 0 \\ 0 & 0 \end{pmatrix}, \quad K = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \tag{18}$$

and (17) vanishes. More generally, for any finite-dimensional[6] matrix K that satisfies (13) we have

$$\operatorname{tr}(K^2) = -i \operatorname{tr}(K[J, K]) = 0, \tag{19}$$

and the energy (17) is zero. Equation (19) also shows that K cannot be Hermitian with our assumption that the dimension n is finite.[6] Finally, we note that Eq. (14) also admits a broad class of solutions without spherical symmetry.

The A_i may be Hermitian in particlelike solutions to Eqs. (8) and (9) of the form

$$A_i = \epsilon_{ijk} I_j x_k \chi, \tag{20}$$

where ϵ_{ijk} is the Levi-Civita symbol, the I_i are constant complex $\mathbf{n} \times \mathbf{n}$ nonzero matrices such that

$$[I_i, I_j] = i\epsilon_{ijk} I_k, \tag{21}$$

and $\chi = \chi(r)$ is a real scalar function of $r \equiv (x_1^2 + x_2^2 + x_3^2)^{1/2}$. Since the form (20) is wholly independent of x^0, Eq. (9) is satisfied identically and (10) states that $F_{0i} = 0$. The right side of (11) is evaluated with (20) and (21) to yield

$$\begin{aligned} F_{ij} &= [-2\epsilon_{ijk}\chi + (\epsilon_{ikl}x_j - \epsilon_{jkl}x_i)x_l r^{-1}\chi' \\ &\quad - e\epsilon_{ijl}x_k x_l \chi^2] I_k \\ &= [-\epsilon_{ijk}(2\chi + r\chi') + \epsilon_{ijl}x_k x_l (r^{-1}\chi' - e\chi^2)] I_k, \end{aligned} \tag{22}$$

from which it follows by straightforward computation that Eq. (8) is satisfied if χ is a solution to the nonlinear ordinary differential equation

$$\chi'' + 4r^{-1}\chi' - 3e\chi^2 - e^2 r^2 \chi^3 = 0. \tag{23}$$

With the introduction of the dimensionless variable $\xi \equiv 1 + er^2\chi$, Eq. (23) becomes

$$r^2\xi'' + \xi - \xi^3 = 0, \tag{24}$$

an equation that has been discussed in a generic context in the literature.[7] The obvious particular solutions to (24), $\xi \equiv 0$ and $\xi \equiv -1$, produce the nontrivial special solutions to Eq. (23),

$$\chi = -Ne^{-1}r^{-2} \quad \text{for } N = 1, 2, \tag{25}$$

and (22) then yields

$$F_{ij} = (2N - N^2)e^{-1}r^{-4}\epsilon_{ijl}x_l x_k I_k \tag{26}$$

$$= \begin{cases} e^{-1}r^{-4}\epsilon_{ijl}x_l x_k I_k & \text{for } N = 1 \\ 0 & \text{for } N = 2. \end{cases} \tag{27}$$

Hence, the total canonical field energy (6) is infinite for the solution (20), (25) with $N = 1$, while the solution with $N = 2$ must be discarded as pure gauge. Now in terms of the dimensionless variable $\xi = 1 + er^2\chi$, (22) is expressed as

$$F_{ij} = e^{-1}\{-\epsilon_{ijk}r^{-1}\xi' + \epsilon_{ijl}x_k x_l[r^{-3}\xi' + r^{-4}(1 - \xi^2)]\} I_k, \tag{28}$$

and thus we find

$$\operatorname{tr}(F_{ij}F_{ij}) = \tfrac{2}{3}nI(I + 1)e^{-2}[2r^{-2}(\xi')^2 + r^{-4}(1 - \xi^2)^2] \tag{29}$$

for an n-dimensional irreducible representation of the I_i in (21), because we then have $\operatorname{tr}(I_k I_l) = \frac{1}{3}nI(I + 1)\delta_{kl}$, with $2I$ a positive integer.

Thus, the energy (6),

$$E = \tfrac{2}{3}\pi I(I + 1)e^{-2}(\mathcal{I}_1 + \mathcal{I}_2), \tag{30}$$

would be finite if it were possible to find a solution to (24) such that the quantities

$$\mathcal{I}_1 \equiv 2\int_0^\infty (\xi')^2 dr \quad \text{and} \quad \mathcal{I}_2 \equiv \int_0^\infty r^{-2}(1 - \xi^2)^2 dr \tag{31}$$

are finite. But the differential equation (24) follows from the variational principle

$$\frac{\delta}{\delta\xi(r)}(\mathcal{I}_1 + \mathcal{I}_2) = 0, \tag{32}$$

which implies[8] [by making the specific variation $\xi(r) \to \xi(\lambda r)$ with λ a constant parameter] that a necessary condition for the existence of a solution with $\mathcal{I}_1$ and $\mathcal{I}_2$ finite is $\mathcal{I}_1 + \mathcal{I}_2 = 0$. Consequently, since the quantities (31) are both positive, no solution exists to Eq. (24) with $\mathcal{I}_1$ and $\mathcal{I}_2$ both finite. The general nature of solutions to Eq. (24) is discussed briefly in the Appendix.

APPENDIX

Excluding from consideration the obvious solutions $\xi \equiv 0, \pm 1$, solutions to Eq. (24) that are analytic at $r = 0$ have the (slowly convergent) expansion about $r = 0$

$$\xi = \pm[1 + ar^2 + \tfrac{3}{10}a^2 r^4 + \tfrac{1}{10}a^3 r^6 + O(a^4 r^8)], \tag{A1}$$

with $a(\neq 0)$ a real constant, while the solutions that are analytic at $r = \infty$ have the (slowly convergent) expansion about $r = \infty$

$$\xi = \pm[1 + br^{-1} + \tfrac{3}{4}b^2 r^{-2} + \tfrac{11}{20}b^3 r^{-3} + O(b^4 r^{-4})], \tag{A2}$$

with $b(\neq 0)$ a real constant. It can be shown that for both positive and negative values of the parameters a and b, the solutions (A1) and (A2) become infinite at a finite positive r, say at $r = r_*$, with the asymptotic form admitted by (24),

$$\xi \doteq \sqrt{2} r_*(r_* - r)^{-1} \quad \text{as } r \to r_*. \tag{A3}$$

The character of (A3) is suggestive of the Schwarzschild singularity for a mass-point in general relativity.

* Work supported by a National Science Foundation grant.

1 M. Ikeda and Y. Miyachi, Progr. Theoret. Phys. **27**, 474 (1962).

2 R. P. Treat, Nuovo Cimento **50**, 871 (1967).

3 H. G. Loos, J. Math. Phys. **8**, 1870 (1967).

4 G. Rosen, J. Math. Phys. **9**, 996 (1968); **9**, 999 (1968); and works cited therein.

5 Physical identification of the electromagnetic part of the Yang-Mills field must await contact of the theory with experiment; alternative and perhaps more interesting possible identifications of the electromagnetic part have been discussed by J. Schwinger, Rev. Mod. Phys. **36**, 609 (1964) and H. G. Loos, Nuovo Cimento **58A**, 365 (1968), for which the solutions reported in the present paper relate classical electromagnetic structure.

6 On the other hand, if an infinite-dimensional representation were admissible on physical grounds, then self-adjoint operators J and K could be found, and

$$\mathrm{tr}(K^2) \equiv \sum_{i=1}^{\infty} \langle i|K^2|i\rangle = \sum_{i=1}^{\infty} |K|i\rangle|^2$$

would be positive for any complete orthonormal basis $\{|i\rangle\}$ in the representation space [with the gauge group restricted to unitary operators $S = S(x)$].

7 See the works cited by E. Kamke, *Differentialgleichungen Lösungsmethoden und Lösungen* (Akademische Verlags., Leipzig, 1967), p. 564, Eq. 6.82.

8 G. Rosen, Quart. Appl. Math. **27**, 133 (1969).

On the Cauchy Problem for the Coupled Maxwell-Dirac Equations*

John M. Chadam†
The Institute for Advanced Study, Princeton, New Jersey 08540

(Received 24 September 1971; Revised Manuscript Received 4 January 1972)

The Cauchy problem for the coupled Maxwell-Dirac equations is solved within an arbitrary bounded region of space-time. An integral part of the proof is that the Cauchy problem for a cut-off version of these equations has a global solution. The analysis requires that the size of the Cauchy data or the coupling constant be suitably restricted.

1. INTRODUCTION

To date, the Cauchy problem for the coupled Maxwell-Dirac equations[1]

$$(-i\gamma^\mu \partial_\mu + m)\psi = g v^\mu \gamma_\mu \psi, \tag{1a}$$

$$\Box v_\mu = (\Delta - \partial_0^2) v_\mu = g \bar{\psi} \gamma_\mu \psi, \tag{1b}$$

$$\partial^\mu v_\mu = 0 \tag{1c}$$

has only been treated locally.[2] In this paper, we shall prove that this problem can be solved in any bounded region of space-time provided that the size of either the coupling constant g or the Cauchy data is suitably restricted.

The basic idea of the proof is quite natural and simple. First, the coupling constant and the Cauchy data are spatially cut off in a smooth fashion outside the domain of dependence at the initial time (i.e., the time at which the Cauchy data are prescribed) of the given space-time region. The main technical problem then is to show that this cut-off version of the original problem has a global solution when the size of g or the Cauchy data is restricted. That this cut-off solution satisfies the actual Maxwell-Dirac equations (1) in the given space-time region follows from the finite propagation speed of equations of this sort. We remark that this approach with only minimal changes in the computations to follow, also applies to coupled Dirac-Klein-Gordon equations in which the spinor and scalar fields interact through a Yukawa coupling.

In Sec. 2, the basic definitions and notations will be given along with a precise statement of the main results. Some of the less computational proofs will also be presented. The existence of solutions to the cut-off problem will be proved in Sec. 3 using the techniques of Refs. 3-5. Proofs of fundamental but detailed inequalities will be carried out in the Appendix.

2. THE MAIN RESULTS

We begin by describing the solution space and discussing the free equations [i.e., $g = 0$ in Eq. (1)]. More of the details can be found in Ref. 2, Sec. 2, and Ref. 6, Sec. 1. The Dirac equation can be written in a more convenient form

$$\frac{\partial \psi}{\partial t}(x, t) = (\boldsymbol{\alpha} \cdot \nabla + \beta m)\psi(x, t) \tag{2}$$

by choosing $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3) = (-\gamma^0\gamma^1, -\gamma^0\gamma^2, -\gamma^0\gamma^3)$ and $\beta = i\gamma^0$. Because of the anticommutation relations satisfied by the α_i and β, the generator of Eq. (2), $i^{-1}(\boldsymbol{\alpha} \cdot \nabla + \beta m)$, is self-adjoint on the Hilbert space of square-integrable functions from E_3 to spin space. In addition, $[i^{-1}(\boldsymbol{\alpha} \cdot \nabla + \beta m)]^2 = m^2 I - \Delta$, thus suggesting the following definition of the escalated energy spaces of the Dirac equation.

Definition: Let D be the Hilbert space of square integrable functions ψ on E_3 with values in spin space.[7] Denote the diagonal operator $(m^2 I - \Delta)^{1/2}$ by A. Then, for $\alpha \geq 0$, D_α is defined to be $D(A^\alpha) \subset D$ endowed with the norm

$$\| \psi \|_{D_\alpha} = \| A^\alpha \psi \|_2 .^8$$

D_α so defined is a Hilbert space. The operator $(\boldsymbol{\alpha} \cdot \nabla + \beta m)$ is skew-adjoint on D with domain D_1, and hence the propagators of the Dirac equation $D(t) = \exp[t(\boldsymbol{\alpha} \cdot \nabla + \beta m)]$ form a strongly continuous group of unitary transformation on each D_α.

The Hilbert space treatment of the Maxwell equations can most easily be accomplished by rewriting Eq. (1b) with $g = 0$ in its vector-valued form

$$\frac{d}{dt}\begin{pmatrix} v_\mu \\ \dot{v}_\mu \end{pmatrix} = \begin{pmatrix} 0 & I \\ \Delta & 0 \end{pmatrix}\begin{pmatrix} v_\mu \\ \dot{v}_\mu \end{pmatrix}. \tag{3}$$

The solution space can then be described in terms of v_μ and $\dot{v}_\mu$.

$$\xi \doteq \sqrt{2} r_*(r_* - r)^{-1} \quad \text{as } r \to r_*. \tag{A3}$$

The character of (A3) is suggestive of the Schwarzschild singularity for a mass-point in general relativity.

* Work supported by a National Science Foundation grant.

1 M. Ikeda and Y. Miyachi, Progr. Theoret. Phys. **27**, 474 (1962).

2 R. P. Treat, Nuovo Cimento **50**, 871 (1967).

3 H. G. Loos, J. Math. Phys. **8**, 1870 (1967).

4 G. Rosen, J. Math. Phys. **9**, 996 (1968); **9**, 999 (1968); and works cited therein.

5 Physical identification of the electromagnetic part of the Yang-Mills field must await contact of the theory with experiment; alternative and perhaps more interesting possible identifications of the electromagnetic part have been discussed by J. Schwinger, Rev. Mod. Phys. **36**, 609 (1964) and H. G. Loos, Nuovo Cimento **58A**, 365 (1968), for which the solutions reported in the present paper relate classical electromagnetic structure.

6 On the other hand, if an infinite-dimensional representation were admissible on physical grounds, then self-adjoint operators J and K could be found, and

$$\mathrm{tr}(K^2) \equiv \sum_{i=1}^{\infty} \langle i|K^2|i\rangle = \sum_{i=1}^{\infty} |K|i\rangle|^2$$

would be positive for any complete orthonormal basis $\{|i\rangle\}$ in the representation space [with the gauge group restricted to unitary operators $S = S(x)$].

7 See the works cited by E. Kamke, *Differentialgleichungen Lösungsmethoden und Lösungen* (Akademische Verlags., Leipzig, 1967), p. 564, Eq. 6.82.

8 G. Rosen, Quart. Appl. Math. **27**, 133 (1969).

On the Cauchy Problem for the Coupled Maxwell-Dirac Equations*

John M. Chadam†
The Institute for Advanced Study, Princeton, New Jersey 08540

(Received 24 September 1971; Revised Manuscript Received 4 January 1972)

The Cauchy problem for the coupled Maxwell-Dirac equations is solved within an arbitrary bounded region of space-time. An integral part of the proof is that the Cauchy problem for a cut-off version of these equations has a global solution. The analysis requires that the size of the Cauchy data or the coupling constant be suitably restricted.

1. INTRODUCTION

To date, the Cauchy problem for the coupled Maxwell-Dirac equations[1]

$$(-i\gamma^\mu \partial_\mu + m)\psi = g v^\mu \gamma_\mu \psi, \tag{1a}$$

$$\Box v_\mu = (\Delta - \partial_0^2) v_\mu = g \bar{\psi} \gamma_\mu \psi, \tag{1b}$$

$$\partial^\mu v_\mu = 0 \tag{1c}$$

has only been treated locally.[2] In this paper, we shall prove that this problem can be solved in any bounded region of space-time provided that the size of either the coupling constant g or the Cauchy data is suitably restricted.

The basic idea of the proof is quite natural and simple. First, the coupling constant and the Cauchy data are spatially cut off in a smooth fashion outside the domain of dependence at the initial time (i.e., the time at which the Cauchy data are prescribed) of the given space-time region. The main technical problem then is to show that this cut-off version of the original problem has a global solution when the size of g or the Cauchy data is restricted. That this cut-off solution satisfies the actual Maxwell-Dirac equations (1) in the given space-time region follows from the finite propagation speed of equations of this sort. We remark that this approach with only minimal changes in the computations to follow, also applies to coupled Dirac-Klein-Gordon equations in which the spinor and scalar fields interact through a Yukawa coupling.

In Sec. 2, the basic definitions and notations will be given along with a precise statement of the main results. Some of the less computational proofs will also be presented. The existence of solutions to the cut-off problem will be proved in Sec. 3 using the techniques of Refs. 3-5. Proofs of fundamental but detailed inequalities will be carried out in the Appendix.

2. THE MAIN RESULTS

We begin by describing the solution space and discussing the free equations [i.e., $g = 0$ in Eq. (1)]. More of the details can be found in Ref. 2, Sec. 2, and Ref. 6, Sec. 1. The Dirac equation can be written in a more convenient form

$$\frac{\partial \psi}{\partial t}(x, t) = (\boldsymbol{\alpha} \cdot \nabla + \beta m)\psi(x, t) \tag{2}$$

by choosing $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3) = (-\gamma^0\gamma^1, -\gamma^0\gamma^2, -\gamma^0\gamma^3)$ and $\beta = i\gamma^0$. Because of the anticommutation relations satisfied by the α_i and β, the generator of Eq. (2), $i^{-1}(\boldsymbol{\alpha} \cdot \nabla + \beta m)$, is self-adjoint on the Hilbert space of square-integrable functions from E_3 to spin space. In addition, $[i^{-1}(\boldsymbol{\alpha} \cdot \nabla + \beta m)]^2 = m^2 I - \Delta$, thus suggesting the following definition of the escalated energy spaces of the Dirac equation.

Definition: Let D be the Hilbert space of square integrable functions ψ on E_3 with values in spin space.[7] Denote the diagonal operator $(m^2 I - \Delta)^{1/2}$ by A. Then, for $\alpha \geq 0$, D_α is defined to be $D(A^\alpha) \subset D$ endowed with the norm

$$\| \psi \|_{D_\alpha} = \| A^\alpha \psi \|_2 .^8$$

D_α so defined is a Hilbert space. The operator $(\boldsymbol{\alpha} \cdot \nabla + \beta m)$ is skew-adjoint on D with domain D_1, and hence the propagators of the Dirac equation $D(t) = \exp[t(\boldsymbol{\alpha} \cdot \nabla + \beta m)]$ form a strongly continuous group of unitary transformation on each D_α.

The Hilbert space treatment of the Maxwell equations can most easily be accomplished by rewriting Eq. (1b) with $g = 0$ in its vector-valued form

$$\frac{d}{dt}\begin{pmatrix} v_\mu \\ \dot{v}_\mu \end{pmatrix} = \begin{pmatrix} 0 & I \\ \Delta & 0 \end{pmatrix}\begin{pmatrix} v_\mu \\ \dot{v}_\mu \end{pmatrix}. \tag{3}$$

The solution space can then be described in terms of v_μ and $\dot{v}_\mu$.

Definition: Let M be the Hilbert space of square-integrable functions $\binom{v}{\dot{v}}$ from E_3 with values in $\mathbb{R}^4 \oplus \mathbb{R}^4$. Denote the 4×4 diagonal operator $(-\Delta)^{1/2}$ by B. Then, for every $\beta \geq \frac{1}{2}$, $M_\beta = M_\beta^1 \oplus M_\beta^2$ is defined to be $D(B(I + B^{(2\beta-1)/2})) \oplus (I + B^{(2\beta-1)/2})$ endowed with the norm

$$\left\| \begin{pmatrix} v \\ \dot{v} \end{pmatrix} \right\|_{M_\beta} = \left\| \{ \| B(I + B^{(2\beta-1)/2}) v \|_2^2 + \| (I + B^{(2\beta-1)/2}) \dot{v} \|_2^2 \}^{1/2} \right\|$$

M_β so defined is a Hilbert space.[9] It is a straightforward exercise to check that $M(t)$: $M_\beta \to M_\beta$ defined by

$$M(t) \begin{pmatrix} v \\ \dot{v} \end{pmatrix} = \begin{pmatrix} \cos(tB) & B^{-1}\sin(tB) \\ -B\sin(tB) & \cos(tB) \end{pmatrix} \begin{pmatrix} v \\ \dot{v} \end{pmatrix} \tag{4}$$

is a continuous one-parameter group of orthogonal transformations on M_β with skew-adjoint generator

$$\begin{pmatrix} 0 & I \\ -B^2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & I \\ \Delta & 0 \end{pmatrix}$$

and as such is the propagator for Eq. (3).

The Cauchy problem for the coupled Maxwell–Dirac equations in this Hilbert space setting can now be described. The solution space is taken to be the escalated energy space $D_\alpha \oplus M_\beta$ for some fixed α and β.

Given Cauchy data $(\psi^0, \binom{v^0}{\dot{v}^0}) \in D_\alpha \oplus M_\beta$ at some finite time t_0 one must find functions ψ and v, with $t \to (\psi(t), \binom{v(t)}{\dot{v}(t)})$: $(t_0, T) \to D_\alpha \oplus M_\beta$ continuous, which satisfies, for $t_0 < t < T$, (the integrated form of) Eqs. (1)

$$\psi(t) = D(t - t_0)\psi^0 - g \int_{t_0}^{t} D(t - s) V(s) \psi(s)\, ds, \tag{5a}$$

$$\begin{pmatrix} v(t) \\ \dot{v}(t) \end{pmatrix} = M(t - t_0) \begin{pmatrix} v^0 \\ \dot{v}^0 \end{pmatrix} - g \int_{t_0}^{t} M(t - s) \begin{pmatrix} 0 \\ J(s) \end{pmatrix} ds, \tag{5b}$$

where

$$V(s) = i\left(v_0(s) - \sum_{k=1}^{3} v_k(s)\, \alpha_k \right) \tag{6a}$$

and

$$J_\mu(s) = \begin{cases} \psi^\dagger \psi(s), & \mu = 0, \\ \psi^\dagger \alpha_\mu \psi(s), & \mu = 1, 2, 3. \end{cases} \tag{6b}$$

The integrals appearing in Eqs. (5) are to be interpreted in the strong Riemann sense. The solution is said to be global if T can be taken to be $+\infty$.

Within this framework, Gross[2] has proved that for sufficiently small $T - t_0$ (i.e., locally) the above problem has a unique solution in the space $D_1 \oplus M_1$. The interaction term in Eq. (5) is not semi-Lipschitzian on $D_1 \oplus M_1$, thus precluding the use of the Picard theory[10] for obtaining local solutions and requiring instead the observation that the well-known results of Kato[11] are applicable. By using escalated energy spaces (i.e., raising α and β), as will be the case in this work, this problem is avoided and the existence of solutions locally is obtained immediately. As is well known (cf. Ref. 2, Sec. 4, p. 14), however, the likelihood of having the necessary *a priori* energy estimates for extending the solution globally is greatly reduced. In fact, the following results can be roughly viewed as saying that these estimates are available if the discussion is restricted to bounded space–time regions and the coupling constant or Cauchy data is suitably small. We now state the results precisely.

In order to avoid unnecessary measure theoretic complications, the space–time volume R is taken to be a domain with smooth boundary. In addition, there is no loss in generality in taking t_0 to be earlier than the times in R since the problem can be separated into a forward and backward Cauchy problem. The constants α and β are now chosen so that $\psi(t)$ and $v(t)$ lie in $L^p(E_3)$ for a wide range of p's (most importantly, $p = \infty$). In particular for $\alpha = 2$, $\beta = 3/2$, standard techniques[12] apply to show that if $\psi \in D_2$ and $v \in M_{3/2}^1$, then $\psi \in L^p(E_3)$ for $2 \leq p \leq \infty$ and $v \in L^p(E_3)$ for $6 \leq p \leq \infty$. The types of solutions that we shall be discussing, at least in the beginning (cf. also Theorem 2.3 and the subsequent remarks), are described in the following.

Definition: Suppose

$$T_R = \sup_{(x,t) \in R} t \quad \text{and} \quad \left(\psi^0, \begin{pmatrix} v^0 \\ \dot{v}^0 \end{pmatrix} \right) \in D_2 \oplus M_{3/2}.$$

An R-solution of the coupled Maxwell–Dirac equations in $D_2 \oplus M_{3/2}$ with Cauchy data $(\psi^0, \binom{v}{\dot{v}^0})$ at t_0 is a pair ψ and v, with $t \to (\psi(t), \binom{v(t)}{\dot{v}(t)})$: $[t_0, T_R] \to D_2 \oplus M_{3/2}$ continuous, which for each $t_0 < t < T_R$ satisfies Eqs. (5) a.e. in the section $R_t = \{t: (x, t) \in R\}$.

Theorem 2.1: Suppose that $(\psi^0, \binom{v}{\dot{v}^0}) \in D_2 \oplus M_{3/2}$ [i.e., $\alpha = (2\beta + 1)/2 = 2$] and R is a bounded domain in space–time. Then the (integrated form of the) coupled Maxwell–Dirac equations have an R-solution in $D_2 \oplus M_{3/2}$ provided that the coupling constant or the $(D_2 \oplus M_{3/2})$-norm of the Cauchy data is sufficiently small (the size depends on the size of R).

The proof of Theorem 2.1 is carried out by obtaining a global solution to a spatially cut-off problem which will be shown to be an R-solution by hyperbolicity arguments. Specifically, outside the compact spatial region obtained by intersecting the domain of dependence of R with $t = t_0$ (which we shall henceforth call R_{t_0}), cut off the Cauchy data and the coupling constant to obtain $(\psi_c^0, \binom{v_c^0}{\dot{v}_c^0}) \in D_2 \oplus M_{3/2}$ and $g_c \in C_c^\infty(E_3)$. For specificity, we shall take as the cut-off functions those obtained by first multiplying with the characteristic function of R_{t_0}' (R_{t_0} enlarged by one in each direction) and then mollifying the resulting function in such a way as to extend the support by one again; i.e., $f_c(x) = \int \chi_{R_{t_0}'}(y) f(y)\, \phi(x - y)\, dy$, where ϕ is the standard mollifier. The existence of a global solution to the resulting cut-off problem comes from the following theorem, the proof of which is the content of Sec. 3.

Theorem 2.2: Suppose $(\psi_c^0, \binom{v_c^0}{\dot{v}_c^0}) \in D_2 \oplus M_{3/2}$ have compact support and $g_c \in C_c^\infty(E_3)$. In addition, suppose that the associated solutions of the free Dirac equation satisfies the decay condition $\| \psi_c^0(t) \|_\infty \leq C(\psi_c^0)(1 + |t|)^{-\epsilon}$ with $3/4 < \epsilon < 1$. Then the cut-off Maxwell–Dirac equations

$$\psi(t) = \psi_c^0(t) - \int_{t_0}^{t} D(t - s) g_c V(s) \psi(s)\, ds, \tag{7a}$$

$$\begin{pmatrix} v(t) \\ \dot{v}(t) \end{pmatrix} = \begin{pmatrix} v_c^0(t) \\ \dot{v}_c^0(t) \end{pmatrix} - \int_{t_0}^{t} M(t - s) \begin{pmatrix} 0 \\ g_c J(s) \end{pmatrix} ds \tag{7b}$$

have a unique global solution $(\psi_c(t), \binom{v_c(t)}{\dot{v}_c(t)}) \in D_2 \oplus M_{3/2}$

provided that $C(\psi_c^0) + \|\psi_c^0\|_{D_2} + \|\binom{v_c^0}{\dot{v}_c^0}\|_{M_{3/2}}$ or $\|g_c\|_\infty$ is sufficiently small (the size depends upon the size of the smallest set which contains all their supports). In addition the spinor component possesses the same temporal decay as the free solution.

Proof of Theorem 2.1 from Theorem 2.2: First we show that the conditions on $(\psi^0, \binom{v^0}{\dot{v}^0})$ in Theorem 2.1 imply the conditions on their associated cut-off functions required in Theorem 2.2. Recall that $f_c(x) = \int \chi_{R'_{t_0}}(y) f(y)\, \phi(x-y)\, dy$, where ϕ is the standard mollifier. Thus, $\|\partial^\alpha f_c\|_r \le \|\chi_{R'_{t_0}} f\|_p \|\partial^\alpha \phi\|_q$, where $r^{-1} + 1 = p^{-1} + q^{-1}$ which implies that $\|g_c\|_{2,\infty} \le g\|\phi\|_{2,1}$, $\|\psi_c^0\|_{D_2} \le \text{const}\ \|\psi_c^0\|_{2,2} \le \text{const}\ \|\chi_{R'_{t_0}} \psi^0\|_2 \|\phi\|_{2,1} \le \text{const}\ \|\psi^0\|_{2,2} \|\phi\|_{2,1} \le \text{const}\ \|\psi^0\|_{D_2} \|\phi\|_{2,1}$, $\|v_c^0\|_{M^1_{3/2}} \le \|\chi_{R'_{t_0}} v^0\|_2 \|B(I+B)\phi\|_1 \le \|\chi_{R_{t_0}}\|_3 \|v^0\|_6 \|B(I+B)\phi\|_1 \le \text{const}\ \text{vol}(R'_{t_0})^{1/3} \|v^0\|_{M^1_{3/2}} \|\phi\|_{2,1}$, and $\|\dot{v}_c^0\|_{M^2_{3/2}} \le \text{const}\ \|\dot{v}_c^0\|_{1,2} \le \text{const}\ \|\chi_{R'_{t_0}} \dot{v}^0\|_2 \|\phi\|_{1,1} \le \text{const}\ \|\dot{v}^0\|_{M^2_{3/2}} \|\phi\|_{2,1}$. The constant $C(\psi_c^0)$ is usually obtained by means of a decay estimate of the sort (Ref. 6, Proposition 1.1): If $\psi_c^0 \in D_3 \cap W^{3,q}(E_3)$, then for $1 < q \le 2$ $\|\psi_c^0(t)\|_\infty \le \text{const}\ (1 + |t|)^{-(3/q - 3/2)} \|\psi_c^0\|_{3,q}$. Thus, the decay estimate is valid with $C(\psi_c^0) = \text{const}\ \|\psi_c^0\|_{3,6/5}$ by taking $q = 6/5$. As above $\psi_c^0 \in D_3$ with $\|\psi_c^0\|_{D_3} \le \text{const}\ \|\psi^0\|_{D_2} \|\phi\|_{3,1}$ and $\|\psi_c^0\|_{3,6/5} \le \text{const}\ \|\chi_{R'_{t_0}} \psi^0\|_{6/5} \|\phi\|_{3,1} \le \text{const}\ \|\chi_{R'_{t_0}}\|_3 \|\psi^0\|_2 \|\phi\|_{3,1} \le \text{const}\ \text{vol}(R'_{t_0})^{1/3} \|\psi^0\|_{2,2} \|\phi\|_{3,1}$. By assembling the above inequalities it is clear that the hypotheses of Theorem 2.1 imply the conditions required on the cut-off functions in Theorem 2.2.

Thus, Theorem 2.2 provides a solution to the cut-off equations (7). We now show that it is an R-solution of the couple Maxwell–Dirac equations in $D_2 \oplus M_{3/2}$ with Cauchy data $(\psi^0, \binom{v^0}{\dot{v}^0})$ at $t = t_0$. We begin by examining the support properties of the cut-off solution in the standard manner. Using the fundamental inequalities proved in Lemma A.1 in the Appendix (see especially Remark 1 following its proof), the nonlinear term in Eq. (7) is a semi-Lipschitzian map on $D_2 \oplus M_{3/2}$ so that the solution $(\psi_c(t), \binom{v_c(t)}{\dot{v}_c(t)})$ is the limit in $D_2 \oplus M_{3/2}$ of the Picard iterates, defined recursively by

$$\psi_c^m(t) = \psi_c^0(t) - \int_{t_0}^t D(t-s) g_c V_c^{m-1}(s) \psi_c^{m-1}(s)\, ds, \tag{8a}$$

$$\binom{v_c^m(t)}{\dot{v}_c^m(t)} = \binom{v_c^0(t)}{\dot{v}_c^0(t)} - \int_{t_0}^t M(t-s) \binom{0}{g_c J_c^{m-1}(s)} ds, \tag{8b}$$

where $V_c^{m-1}(s)$ and $J_c^{m-1}(s)$ have the obvious meaning. By induction, using the well-known support properties of the propagators D and M (e.g., Ref. 13), it is clear that the iterates have their support in the domain of influence of the support of the Cauchy data and coupling function at $t = t_0$. In addition, as mentioned in the previous remarks, convergence in $D_2 \oplus M_{3/2}$ implies convergence a.e. of a subsequence to the same limit so that the support of the cut-off solution is likewise within the same region. The continuity of $(\psi_c(\cdot), \binom{v_c(\cdot)}{\dot{v}_c(\cdot)}): [t_0, \infty) \to D_2 \oplus M_{3/2}$ and Remark 1 of the Appendix guarantee that $g \int_{t_0}^t D(t-s) V_c(s) \psi_c(s)\, ds$ and $g \int_{t_0}^t M(t-s) \binom{0}{J_c(s)} ds$ exist as D_2 (respectively $M_{3/2}$)-valued integrals for all $t < \infty$. By forming $\psi^0(t) - \psi_c^0(t)$, $\int_{t_0}^t D(t-s)(g-g_c) V_c(s) \psi_c(s)\, ds$, and $\int_{t_0}^t M(t-s)(g-g_c) \binom{0}{J_c(s)} ds$ and using the support properties of the propagators and the cut-off solutions, it is clear that the above quantities for each t are zero a.e. in the sections of the domain of influence of R_{t_0} (which by design contains R). Thus, $(\psi_c, \binom{v_c}{\dot{v}_c})$ is an R-solution of the coupled Maxwell–Dirac equations in $D_2 \oplus M_{3/2}$ with Cauchy data $(\psi^0, \binom{v^0}{\dot{v}^0})$ at time t_0.

Since the Cauchy data is fairly smooth, one can expect that the same is true for the above solution. In particular, we have the following result concerning the differentiability of the solution.

Theorem 2.3: With the hypothesis of Theorem 2.1, the resulting solution $(\psi(t), \binom{v(t)}{\dot{v}(t)})$ as a function into $D_1 \oplus M_{1/2}$ is strongly continuously differentiable and satisfies

$$\psi'(t) = (\boldsymbol{\alpha}\cdot\nabla + \beta m)\psi(t) - gV(t)\psi(t), \tag{9a}$$

$$v_\mu''(t) = \Delta v_\mu(t) - g J_\mu(t), \tag{9b}$$

for each t with $(x,t) \in R$, a.e. in R_t, while $\psi(0) = \psi^0$ and $\binom{v(0)}{\dot{v}(0)} = \binom{v^0}{\dot{v}^0}$ a.e. in the causal shadow of R at $t = t_0$.

Proof: As α, β grow, the energy spaces $D_\alpha \oplus M_\beta$ shrink while the associated norms become larger. Thus, the solution of Theorem 2.1 is a continuous function into $D_1 \oplus M_{1/2}$. The proof of the differentiability will be obtained directly from a general result of Kato (Ref. 11, Theorem 5, p. 211) as follows. The underlying Banach space is taken to be $D_1 \oplus M_{1/2}$ on which the operator $(\boldsymbol{\alpha}\cdot\nabla + \beta m) \oplus \left(\begin{smallmatrix} 0 & I \\ \Delta & 0 \end{smallmatrix}\right)$ generates the continuous one-parameter group $D(t) \oplus M(t)$. Now the domain of the generator is $D_2 \oplus D(B^2) \oplus D(B) \supset D_2 \oplus M_{3/2}$, so that the Cauchy data $(\psi^0, \binom{v^0}{\dot{v}^0})$ [and hence $(\psi_c^0, \binom{v_c^0}{\dot{v}_c^0})$] is in the domain of the generator. As in the proof of the cited result of Kato (Ref. 11, p. 228), if the nonlinear term of Eq. (7) is differentiable, then in $D_1 \oplus M_{1/2}$

$$\frac{d}{dt}\psi(t) = (\boldsymbol{\alpha}\cdot\nabla + \beta m)\psi(t) - g_c V(t)\psi(t), \tag{10a}$$

$$\frac{d}{dt}\binom{v(t)}{\dot{v}(t)} = \begin{pmatrix} 0 & I \\ \Delta & 0 \end{pmatrix}\binom{v(t)}{\dot{v}(t)} - g_c \binom{0}{J(t)}, \tag{10b}$$

with $\psi(0) = \psi_c^0$ and $\binom{v(0)}{\dot{v}(0)} = \binom{v_c^0}{\dot{v}_c^0}$. Thus, for each $t \in (t_0, \infty)$

$$\frac{d}{dt}\psi(t) = (\boldsymbol{\alpha}\cdot\nabla + \beta m)\psi(t) - g_c V(t)\psi(t), \tag{11a}$$

$$\frac{d^2}{dt^2} v(t) = \Delta v(t) - g_c J(t) \tag{11b}$$

are valid in D_1 and $M^2_{1/2} = L^2(E_3)$, respectively, and hence a.e. in E_3. But on R_t for each t with $(x,t) \in R$, $g_c = g$ and in the causal shadow of R at $t = t_0$, $\psi_c^0 = \psi^0$, $v_c^0 = v$, and $\dot{v}_c^0 = \dot{v}^0$ a.e., thus proving the theorem.

All that remains then is the differentiability of the integrals in Eq. (7). As stated in the hypothesis of the general result of Kato, this follows from the fact that $g_c V_c(t)\psi_c(t) \oplus g_c \binom{0}{J_c(t)}$ belongs to $(D_2 \oplus M_{3/2} \subset)$

$D_2 \oplus D(B^2) \oplus D(B)$ and that $[(\boldsymbol{\alpha}\cdot\nabla + \beta m - I) \oplus \begin{pmatrix} I & I \\ \Delta & -I \end{pmatrix}]$ $[g_c V_c(t)\psi_c(t) \oplus g_c \begin{pmatrix} 0 \\ J_c(t) \end{pmatrix}]$ is strongly continuous with respect to t. Both conditions follow in a straightforward way from the fundamental inequality in the Appendix, the second requiring in addition the continuity of $t \to (\psi_c(t)\begin{pmatrix} v_c(t) \\ \dot{v}_c(t) \end{pmatrix})$: $[t_0, \infty) \to D_2 \oplus M_{3/2}$ as given by Theorem 2.2.

To conclude this section, we make an observation that allows the above results to be more easily compared with Gross' basic result (Ref. 2, Theorem 1, p. 4). Note that the space used by Gross, $D_G \oplus M_G$, is equivalent to $D_0 \oplus D[B(I + B)^{-1/2}] \oplus D[(I + B)^{-1/2}]$. The Hilbert Space obtained by completing $D_0 \oplus D(B(I + B)^{-1/2}) \oplus D(I + B)^{-1/2}$ with respect to the norm $\{\|\psi\|_2^2 + \|B(I + B)^{-1/2}v\|_2^2 + \|(I + B)^{-1/2}\dot{v}\|_2^2\}^{1/2}$. But this norm is smaller than the $(D_1 \oplus M_{1/2})$-norm so that our solution in Theorems 2.1 and 2.3 can also be considered as a $(D_G \oplus M_G)$-valued function which is differentiable and satisfies the same differential equation.

3. THE CUTOFF EQUATIONS

This whole section will consist of outlining the proof of Theorem 2.2. The argument will follow, in general, along the lines of that presented in Refs. 3–5, but will require certain refinements and permit some simplifications. It is these modifications that we shall try to stress in this presentation. With no loss of generality, we shall assume that the supports of g_c, ψ_c^0, v_c^0, and $\dot{v}_c^0$ are contained in $B_\rho(0)$, the ball of radius ρ about the origin.

The local existence of solutions to the cut-off problem in $D_2 \oplus M_{3/2}$ follows by the general results of Segal (Ref. 10, Theorem 1, p. 343) from the semi-Lipschitz nature of the nonlinear term as established in Remark 1 of the Appendix. The cited work also guarantees that the solution exists in the maximum interval $(t_0, \bar{t})$ for which $\|(\psi_c(t), \begin{pmatrix} v_c(t) \\ \dot{v}_c(t) \end{pmatrix})\|_{D_2 \oplus M_{3/2}}$ remains bounded. We shall show that with the additional hypotheses of Theorem 2.2, the solution remains bounded for all $t \geq t_0$.

The first component of Eq. (7b) can be written more explicitly as

$$v_c(t) = \cos(tB)v_c^0 + B^{-1}\sin(tB)\dot{v}_c^0 - \int_{t_0}^t B^{-1}\sin(t-s)B[g_c J_c(s)]\,ds, \quad (12)$$

where $J_c(s)$ is given by Eq. (6b) and depends only on $\psi_c(s)$. Substituting Eq. (12) into Eq. (7a) gives the relationship

$$\psi_c(t) = \psi_c^0(t) - \int_{t_0}^t D(t-s)g_c V_c^0(s)\psi_c(s)ds + i\int_{t_0}^t D(t-s)g_c\left(\int_{t_0}^s (B^{-1}\sin(s-\tau)Bg_c J_c(\tau))_0 - \sum_{k=1}^3 (B^{-1}\sin(s-\tau)Bg_c J_c(\tau))_k \alpha_k d\tau\right)\psi_c(s)ds, \quad (13)$$

which depends only on $\psi_c(t)$ and the prescribed quantities $(\psi_c^0, \begin{pmatrix} v_c^0 \\ \dot{v}_c^0 \end{pmatrix})$. Since the propagator $D(t)$ is unitary on D_2,

$$\|\psi_c(t)\|_{D_2} \leq \|\psi_c^0\|_{D_2} + \int_{t_0}^t \|g_c V_c^0(s)\psi_c(s)\|_{D_2} ds + \int_{t_0}^t \left\| g_c \left(\int_{t_0}^s (B^{-1}\sin(s-\tau)Bg_c J_c(\tau))_0 - \sum_{k=1}^3 (B^{-1}\sin(s-\tau)Bg_c J_c(\tau))_k \alpha_k d\tau\right)\psi_c(s)\right\|_{D_2} ds. \quad (14)$$

Now, using part (ii) of Lemma A1 with $\tilde{\chi}$ and $\tilde{\psi} \equiv 0$, χ being either $v_c^0(s)$ or $B^{-1}\sin(s-\tau)Bg_c J_c(\tau)$ and ψ being $\psi_c(s)$ or $\alpha_k\psi_c(s)$, one obtains

$$\|\psi_c(t)\|_{D_2} \leq \|\psi_c^0\|_{D_2} + \text{const}\,\|g_c\|_{2,\infty}\int_{t_0}^t \{\|\cos(sB)v_c^0 + B^{-1}\sin(sB)\dot{v}_c^0\|_{M^1_{3/2}}^{\text{loc}} + \int_{t_0}^s \|B^{-1}\sin(s-\tau)B \times g_c J_c(\tau)\|_{M^1_{3/2}}^{\text{loc}} d\tau\}\|\psi_c(s)\|_{D_2}^{\text{loc}} ds, \quad (15)$$

where the super loc refers to the fact that in computing the norms the integration need only be taken over the support of g_c. This can be reduced to a simpler inequality in ψ_c by using the following particularly convenient form of the result that locally the solutions of the wave equation decay very rapidly.

Lemma 3.1: Suppose $\begin{pmatrix} f \\ h \end{pmatrix} \in M_{3/2}$ and $\text{supp}(f)$, $\text{supp}(h) \subset B_\rho(0)$ (i.e., a ball of radius ρ about the origin in E_3). Then for arbitrary $\sigma \geq 0$

$$\|\cos tBf + B^{-1}\sin tBh\|_{M^1_{3/2}}^{\text{loc}} \leq C(\sigma,\rho)\|\begin{pmatrix} f \\ h \end{pmatrix}\|_{M_{3/2}}(1 + |t|)^{-\sigma}, \quad (16)$$

where loc indicates that in computing the norm the integration is only taken over $B_\rho(0)$.

Proof: To begin, suppose $f, h \in C_c^\infty(B_\rho(0))$. Then

$$B(I + B)B^{-1}\sin tBh(x) = B^{-1}\sin t\,B(B(I + B)h)\,(x) = \frac{\text{const}}{t}\int_{|x-y|=t}[B(I + B)h](y)\,dS_y,$$

which is zero in $B_\rho(0)$ for $t > 2\rho$. On the other hand, for all t and in particular for $|t| \leq 2\rho$

$$\|B^{-1}\sin(tB)h\|_{M^1_{3/2}} = \|B(I + B)B^{-1}\sin(tB)h\|_{M^1_{3/2}} = \|(I + B)\sin(tB)h\|_2 \leq \|(I + B)h\|_2 = \|h\|_{M^2_{3/2}}.$$

Thus, $\|B^{-1}\sin(tB)h\|_{M^1_{3/2}}^{\text{loc}} \leq (1 + 2\rho)^\sigma \|h\|_{M^2_{3/2}}(1 + |t|)^{-\sigma}$. Likewise $\|\cos(tB)f\|_{M^1_{3/2}}^{\text{loc}} \leq (1 + 2\rho)^\sigma \|f\|_{M^1_{3/2}}(1 + |t|)^{-\sigma}$ so that inequality (16) is verified for $f, h \in C_c^\infty(B_\rho(0))$. The result now follows from the denseness of these functions and the continuity, uniformly in t, of the map $(f, h) \to \cos(tB)f + B^{-1}\sin(tB)h : M_{3/2} \to M_{3/2}$ [which in turn follows from the unitarity of $M(t)$ on $M_{3/2}$].

Returning to the proof of the theorem, we may use the above result directly to obtain

$$\|\psi_c(t)\|_{D_2} \leq \|\psi_c^0\|_{D_2} + \text{const}_{\rho,\sigma}\|g_c\|_{2,\infty} \times \left[\left\|\begin{pmatrix} v_c^0 \\ \dot{v}_c^0 \end{pmatrix}\right\|_{M_{3/2}} \int_{t_0}^t (1 + |s|)^{-\sigma}\|\psi_c(s)\|_{D_2}^{\text{loc}} ds + \int_{t_0}^t \left(\int_{t_0}^s (1 + |s-\tau|)^{-\sigma}\|g_c J_c(\tau)\|_{M^1_{3/2}} d\tau\right)$$

$$\times \| \psi_c(s) \|_{D_2}^{\mathrm{loc}} \, ds \Big] . \tag{17}$$

For notational convenience, we define the function

$$]\psi[_t = \sup_{s \in [t_0, t)} \{ \| \psi(s) \|_{D_2} + (1 + |s|)^{\epsilon} \| \psi(s) \|_{\infty} \} . \tag{18}$$

If $t \to \psi(t)$: $[t_0, \bar{t}) \to D_2$ is continuous, as is ψ_c on its interval of existence, then $]\psi[_t$ is continuous as a map from $[t_0, \bar{t})$ into the reals. In fact, it is $]\psi_c[_t$ which we will show is uniformly bounded, thus guaranteeing the same for $\| \psi_c(t) \|_{D_2}$. Using this definition and taking $\sigma > 1$, we have

$$\begin{aligned} \| \psi_c(t) \|_{D_2} &\le \| \psi_c^0 \|_{D_2} \\ &+ \mathrm{const}_{\rho,\sigma} \| g_c \|_{2,\infty} \left[\left\| \begin{pmatrix} v_c^0 \\ v_c^0 \end{pmatrix} \right\|_{M_{3/2}}]\psi[_t \right. \\ &+ \left. \int_{t_0}^{t} \left(\int_{t_0}^{s} (1 + |s - \tau|)^{-\sigma} \| g_c J_c(\tau) \|_{M_{3/2}^2} \, d\tau \right) ds \;]\psi[_t \right] . \end{aligned} \tag{19}$$

Noting that $\| \cdot \|_{M_{3/2}^2} \le \mathrm{const} \, \| \cdot \|_{1,2}$ and using the estimate in Remark 2 of the Appendix, we reduce the above to

$$\begin{aligned} \| \psi_c(t) \|_{D_2} &\le \| \psi_c^0 \|_{D_2} \\ &+ \mathrm{const}_{\rho,\sigma} \| g_c \|_{2,\infty} \left[\left\| \begin{pmatrix} v_c^0 \\ \dot{v}_c^0 \end{pmatrix} \right\|_{M_{3/2}}]\psi[_t \right. \\ &+ \| g_c \|_{1,2} \int_{t_0}^{t} \left(\int_{t_0}^{s} (1 + |s - \tau|)^{-\sigma} \right. \\ &\left. \left. \times (1 + |\tau|)^{-3\epsilon/2} \, d\tau \right) ds \;]\psi_c[_t^3 \right] . \end{aligned} \tag{20}$$

The inner integral can be bounded by a technical lemma of Segal (Ref. 3, Lemma 3.1, p. 467) with a bound const $(1 + |s|)^{-3\epsilon/2}$ by taking $\sigma \ge 3/2$. Thus, the s-integration is bounded because of the choice of ϵ. Summing up, we have

$$\begin{aligned} \| \psi_c(t) \|_{D_2} &\le \| \psi_c^0 \|_{D_2} \\ &+ \mathrm{const}_{\rho,\sigma} \| g_c \|_{2,\infty} \left\| \begin{pmatrix} v_c^0 \\ \dot{v}_c^0 \end{pmatrix} \right\|_{M_{3/2}}]\psi_c[_t \\ &+ \mathrm{const}_{\rho,\sigma} \| g_c \|_{2,\infty} \| g_c \|_{1,2} \;]\psi_c[_t^3 , \end{aligned} \tag{21}$$

We also need a similar estimate for $\| \psi_c(t) \|_{\infty}$ to obtain a nonlinear inequality in $]\psi_c[_t$ from which to deduce its uniform boundedness. To this end, beginning with Eq. (13) and estimating the Dirac propagator as in Proposition 1.1 of Ref. 6, we have for $1 < q < 4/3$

$$\begin{aligned} \| \psi_c(t) \|_{\infty} &\le C(\psi_c^0)(1 + |t|)^{-\epsilon} \\ &+ \mathrm{const} \left[\int_{t_0}^{t} |t - s|^{-1/q} \| g_c V_c^0(s) \psi_c(s) \|_{2,q} \, ds \right. \\ &+ \mathrm{const} \int_{t_0}^{t} |t - s|^{-1/q} \| g_c \\ &\times \left(\int_{t_0}^{s} (B^{-1} \sin(s - \tau) B g_c J_c(\tau))_0 \right. \\ &\left. \left. - \sum_{k=1}^{3} (B^{-1} \sin(s - \tau) B g_c J_c(\tau))_k \, \alpha_k \, d\tau \right) \psi_c(s) \|_{2,q} \, ds \right] . \end{aligned} \tag{22}$$

Again, using part (ii) of Lemma A1 as in equality (15) with q as above, we get

$$\begin{aligned} \| \psi_c(t) \|_{\infty} &\le C(\psi_c^0)(1 + |t|)^{-\epsilon} \\ &+ \mathrm{const} \, \| g_c \|_{2,2q(2-q)^{-1}} \int_{t_0}^{t} \{ \| \cos(sB) v_c^0 \\ &+ B^{-1} \sin(sB) \dot{v}_c^0 \|_{M_{3/2}}^{\mathrm{loc}} \\ &+ \int_{t_0}^{s} \| B^{-1} \sin(s - \tau) B g_c J_c(\tau) \|_{M_{3/2}}^{\mathrm{loc}} \} \\ &\times \| \psi_c(s) \|_{D_2}^{\mathrm{loc}} \, ds . \end{aligned} \tag{23}$$

The argument proceeds as above to give the following analog of inequality (20):

$$\begin{aligned} \| \psi_c(t) \|_{\infty} &\le C(\psi_c^0)(1 + |t|)^{-\epsilon} \\ &+ \mathrm{const}_{\rho,\sigma} \| g_c \|_{2,2q(2-q)^{-1}} \left[\left\| \begin{pmatrix} v_c^0 \\ \dot{v}_c^0 \end{pmatrix} \right\|_{M_{3/2}} \right. \\ &\times \int_{t_0}^{t} |t - s|^{-1/q} (1 + |s|)^{-\sigma} \, ds \;]\psi[_t \\ &+ \left. \| g_c \|_{1,2} \int_{t_0}^{t} |t - s|^{-1/q} (1 + |s|)^{-3\epsilon/2} \, ds \;]\psi_c[_t^3 \right] , \end{aligned} \tag{24}$$

using the cited technical lemma of Segal to treat the τ-integration.

A similar estimate[14] shows the s-integration can be bounded by a function of t which decays faster than $t^{-\epsilon}$. The particular form of the result which is applicable in this case is: If $0 < a < 1$ and $b > 1$, then

$$\int_{-\infty}^{\infty} |t - s|^{-a} (1 + |s|)^{-b} \, ds \le \mathrm{const} \, (1 + |t|)^{-a}$$

for all $t \in R$. Since σ and $3\epsilon/2 > 1$ and $1/q < 1$, the decay property will follow if we can choose $\epsilon < 1/q < 1$ or equivalently $1 < q < 1/\epsilon$. Since $\epsilon < 1$, we can always find such a q. In addition, $\epsilon > 3/4$ automatically guarantees that for this choice $q < 4/3$ as required for the decay of the Dirac propagator. Thus, we have

$$\begin{aligned} \| \psi_c(t) \|_{\infty} &\le (1 + |t|)^{-\epsilon} \left\{ C(\psi_c^0) + \mathrm{const}_{\rho,\sigma} \| g_c \|_{2,2q(2-q)^{-1}} \right. \\ &\times \left. \left[\left\| \begin{pmatrix} v_c^0 \\ v_c^0 \end{pmatrix} \right\|_{M_{3/2}}]\psi_c[_t + \mathrm{const} \, \| g_c \|_{1,2} \;]\psi_c[_t^3 \right] \right\} . \end{aligned} \tag{25}$$

By adding inequalities (21) and (25), we obtain

$$\begin{aligned}]\psi_c[_t &\le \| \psi_c^0 \|_{D_2} + C(\psi_c^0) \\ &+ \mathrm{const}_{\rho,\sigma} (\| g_c \|_{2,\infty} + \| g_c \|_{2,2q(2-q)^{-1}}) \\ &\times \left\| \begin{pmatrix} v_c^0 \\ \dot{v}_c^0 \end{pmatrix} \right\|_{M_{3/2}}]\psi_c[_t + \mathrm{const}_{\rho,\sigma} \| g_c \|_{1,2} \, (\| g_c \|_{2,\infty} \\ &+ \| g_c \|_{2,2q(2-q)^{-1}})]\psi_c[_t^3 . \end{aligned} \tag{26}$$

since g_c has compact support $\| g_c \|_{1,2} \le \mathrm{const} \, \rho^{3/2} \| g_c \|_{2,\infty}$ and $\| g_c \|_{2,2q(2-q)^{-1}} \le \mathrm{const.} \, \rho^{3(2-q)/2q} \| g_c \|_{2,\infty} \le \mathrm{const} \, \rho^{3/2} \| g_c \|_{2,\infty}$ since $1 < q < 4/3$ (assuming $\rho \ge 1$). We can fix $\sigma = 3/2$ (the maximum needed in the proof) and use the explicit representation of $\mathrm{const}_{\rho,\sigma}$ as given in the proof of Lemma 3.1 to obtain

$$]\psi_c[_t \le [\|\psi_c^0\|_{D_2} + C(\psi_c^0)] + \left[C_1\|g_c\|_{2,\infty}\rho^3 \left\|\begin{pmatrix} v_c^0 \\ \dot{v}_c^0 \end{pmatrix}\right\|_{M_{3/2}}\right]]\psi_c[_t + (C_2\rho^{9/2}\|g_c\|_{2,\infty}^2)\ ]\psi_c[_t^3. \quad (27)$$

For fixed ρ, by taking $\|g_c\|_{2,\infty}$ or $\|\binom{v_c^0}{\dot{v}_c^0}\|_{M_{3/2}}$ sufficiently small the coefficient of $]\psi_c[_t$ can be made less than 1/2 so that

$$]\psi_c[_t \le 2[\|\psi_c^0\|_{D_2} + C(\psi_c^0)] + 2(C_2\rho^{9/2}\|g_c\|_{2,\infty}^2)]\psi_c[_t^3, \quad (28)$$

from which the uniform boundedness of $]\psi_c[_t$ follows from an observation of Segal and Strauss[15] by choosing $\|\psi_c^0\|_{D_2} + C(\psi_c^0)$ or $\|g_c\|_{2,\infty}$ sufficiently small. Thus, the theorem is proved, including the decay property of $\|\psi_c(t)\|_\infty$, provided we can also show that $\|\binom{v_c(t)}{\dot{v}_c(t)}\|_{M_{3/2}}$ remains bounded for $t > t_0$. To this end, using Eq. (7b), we get

$$\begin{aligned}\left\|\begin{pmatrix} v_c(t) \\ \dot{v}_c(t) \end{pmatrix}\right\|_{M_{3/2}} &\le \left\|\begin{pmatrix} v_c^0 \\ \dot{v}_c^0 \end{pmatrix}\right\|_{M_{3/2}} + \int_{t_0}^t \left\|\begin{pmatrix} 0 \\ g_c J_c(s) \end{pmatrix}\right\|_{M_{3/2}} ds \\ &\le \left\|\begin{pmatrix} v_c^0 \\ \dot{v}_c^0 \end{pmatrix}\right\|_{M_{3/2}} + \int_{t_0}^t \|g_c J_c(s)\|_{M_{3/2}}\, ds \\ &\le \left\|\begin{pmatrix} v_c^0 \\ \dot{v}_c^0 \end{pmatrix}\right\|_{M_{3/2}} + \text{const}\,\|g_c\|_{1,2} \\ &\quad \times \int_{t_0}^t (1+|s|)^{-3\epsilon/2}\, ds\ ]\psi_c[_t^2, \end{aligned} \quad (29)$$

the last inequality following from Remark 2 in the Appendix. The uniform boundedness of $]\psi_c[_t$ and the fact that $\epsilon > 3/4$ guarantees the uniform boundedness of $\|\binom{v_c(t)}{\dot{v}_c(t)}\|_{M_{3/2}}$, thus proving Theorem 2.2.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. C. Kaysen and the faculty for their kind hospitality while the author was a member at the Institute for Advanced Study.

APPENDIX

In this section, we sketch a proof of three basic inequalities and then from them outline the proof of the properties of the interactions which were used in the main body of the papers: e.g., that the cut-off interaction maps $D_2 \oplus M_{3/2}$ into itself and as such is semi-Lipschitz (needed in Theorems 2.1, 2.2, and 2.3) and that the actual interaction maps $D_2 \oplus M_{3/2}$ into itself (Theorem 2.1), the estimates used in inequalities (20), (23), and (28) in the proof of Theorem 2.2. The crucial thing to notice about the interactions is that, apart from the scalar factor g or g_c, each component is a linear combination of terms $\chi_i\psi_j$, $i,j = 0,1,2,3$, where χ_i may be the components of $\dot{v}$ or ψ.

Lemma A1: Suppose g is a constant, $g_c \in C_c^\infty(E_3)$, ψ_j are the components of $\psi \in D_2$, χ_i are the components of $\chi \in D_2$ or $M_{3/2}$ and similarly for the tilded quantities. Then, with X denoting either χ or $\tilde{\chi}$ and Ψ either ψ or $\tilde{\psi}$,

(i) $$\|g(\chi_i\psi_j - \tilde{\chi}_i\tilde{\psi}_j)\|_{2,2} \le g\ \text{const}\{\|X\|_{M_{3/2}^1}\|\psi - \tilde{\psi}\|_{D_2} + \|\Psi\|_{D_2}\|\chi - \tilde{\chi}\|_{M_{3/2}^1}\}$$

and, if $1 \le q \le 2$,

(ii) $$\|g_c(\chi_i\psi_j - \tilde{\chi}_i\tilde{\psi}_j)\|_{2,q} \le \text{const}\,\|g_c\|_{2,2q(2-q)^{-1}} \times\{\|X\|_{M_{3/2}^1}\|\psi - \tilde{\psi}\|_{D_2} + \|\Psi\|_{D_2}\|\chi - \tilde{\chi}\|_{M_{3/2}^1}\}$$

and

(iii) $$\begin{aligned}\|g_c(\chi_i\psi_j - \tilde{\chi}_i\tilde{\psi}_j)\|_{1,q} \le \text{const}\,(&\|g_c\|_{1,q} + \|g_c\|_{4q(4-q)^{-1}})\{\|\chi - \tilde{\chi}\|_\infty^{1/2}\|\chi - \tilde{\chi}\|_{M_{3/2}}^{1/2}\|\Psi\|_\infty \\ &+ \|X\|_\infty^{1/2}\|X\|_{M_{3/2}^1}^{1/2}\|\psi - \tilde{\psi}\|_\infty \\ &+ \|\chi - \tilde{\chi}\|_\infty\|\Psi\|_\infty^{1/2}\|\Psi\|_{D_2}^{1/2} \\ &+ \|X\|_\infty\|\psi - \tilde{\psi}\|_\infty^{1/2}\|\psi - \tilde{\psi}\|_{D_2}^{1/2}\}.\end{aligned}$$

Proof: We begin by recalling some facts that will be used in the calculation. First $\|f\|_p \le \text{const}\,\|f\|_{2,2} \le \text{const}\,\|f\|_{D_2}$ for $2 \le p \le \infty$, while $\|f\|_p \le \text{const}\,\|f\|_{M_{3/2}^1}$ for $6 \le p \le 6$. In addition, $\|f\|_{M_{3/2}^1} \le \|f\|_{D_2}$, thus permitting us to simultaneously treat the case of X a spinor or a vector by using its $M_{3/2}^1$ norm in the bound. Finally, we shall also make use of the familiar Sobolev inequality $\|D^1 f\|_4 \le \text{const}\,\|D^2 f\|_2^{1/2}\|f\|_\infty^{1/2}$, where

$$\|D^k f\|_p = \sup_{k_1+k_2+k_3=k} \left\|\frac{\partial^k f}{\partial^{k_1}x_1\partial^{k_2}x_2\partial^{k_3}x_3}\right\|_p.$$

We also define

$$\|D^k X\|_p = \sup_{i=1,2,3.} \|D^k\chi_i\|_p.$$

To begin the calculation,

$$\|g(\chi_i\psi_j - \tilde{\chi}_i\tilde{\psi}_j)\|_{2,2} \le g\ \text{const}[\|\chi_i\psi_j - \tilde{\chi}_i\tilde{\psi}_j\|_2^2 + \|D(\chi_i\psi_j - \tilde{\chi}_i\tilde{\psi}_j)\|_2^2 + \|D^2(\chi_i\psi_j - \tilde{\chi}_i\tilde{\psi}_j)\|_2^2]^{1/2}.$$

Now each term will be bounded by the expression appearing on the right-hand side of the expression (i).

The zeroth-order terms can be written as $g(\chi_i - \tilde{\chi}_i)\psi_j + g\tilde{\chi}_i(\psi_j - \tilde{\psi}_j)$. Calling the terms 0*a*i and 0*a*ii (a notation which will become increasingly useful as the argument progresses), we have

$$\|0ai\|_2 \le g\|(\chi_i - \tilde{\chi}_i)\|_\infty\|\psi_j\|_2 \le g\ \text{const}\,\|\chi - \tilde{\chi}\|_{M_{3/2}^1}\|\psi\|_{D_2},$$

thus contributing to the inequality with $\Psi = \psi$. Likewise,

$$\|0aii\|_2 \le g\|\tilde{\chi}_i\|_\infty\|\psi_j - \tilde{\psi}_j\|_2 \le g\ \text{const}\,\|\tilde{\chi}\|_{M_{3/2}^1}\|\psi - \tilde{\psi}\|_{D_2},$$

giving the other term in the bound.

Letting $D^k\chi_i$ denote any of the kth-order derivatives (since, in estimating them, we shall use the fact that they are all simultaneously less than $\|D^k\chi_i\|_p \le \|D^k X\|_p$), we can write the first-order terms as

$$\begin{aligned}&gD(\chi_i\psi_j - \tilde{\chi}_i\tilde{\psi}_j) \\ &\quad = gD\chi_i\psi_j + g\chi_i D\psi_j - gD\tilde{\chi}_i\tilde{\psi}_j - g\tilde{\chi}_i D\tilde{\psi}_j \\ &\quad = g[D(\chi_i - \tilde{\chi}_i)\psi_j + D\tilde{\chi}_i(\psi_j - \tilde{\psi}_j) \\ &\qquad + (\chi_i - \tilde{\chi}_i)D\psi_j + \tilde{\chi}_i D(\psi_j - \psi_j)].\end{aligned}$$

and the terms in the last expression will be referred to as 1*a*i, 1*a*ii, 1*b*i, and 1*b*ii. The estimates proceed as follows:

$$\begin{aligned}\|1ai\|_2 &\le g\,\|D(\chi_i-\tilde\chi_i)\|_4\,\|\psi_j\|_4\\ &\le g\ \text{const}\,\|D^2(\chi_i-\tilde\chi_i)\|_2^{1/2}\,\|\chi_i-\tilde\chi_i\|_\infty^{1/2}\,\|\psi_j\|_4\\ &\le g\ \text{const}\,\|\chi-\tilde\chi\|_{M^1_{3/2}}\|\psi\|_{D_2}.\end{aligned}$$

The other terms are treated similarly, giving

$$\begin{aligned}\|1aii\|_2 &\le g\ \text{const}\,\|\tilde\chi\|_{M^1_{3/2}}\|\psi-\tilde\psi\|_{D_2},\\ \|1bi\|_2 &\le g\ \text{const}\,\|\chi-\tilde\chi\|_{M^1_{3/2}}\|\psi\|_{D_2},\\ \|1bii\|_2 &\le g\ \text{const}\,\|\tilde\chi\|_{M^1_{3/2}}\|\psi-\tilde\psi\|_{D_2}.\end{aligned}$$

The second-order terms can be handled similarly:

$$\begin{aligned}&gD^2(\chi_i\psi_j-\tilde\chi_i\tilde\psi_j)\\ &\quad=g(D^2\chi_i\psi_j+2D\chi_iD\psi_j+\chi_iD^2\psi_j\\ &\qquad-D^2\tilde\chi_i\tilde\psi_j-2D\tilde\chi_iD\tilde\psi_j-\tilde\chi_iD^2\tilde\psi_j)\\ &\quad=g[D^2(\chi_i-\tilde\chi_i)\psi_j+D^2\tilde\chi_i(\psi_j-\tilde\psi_j)\\ &\qquad+2D(\chi_i-\tilde\chi_i)D\psi_j+2D\tilde\chi_i\,D(\psi_j-\tilde\psi_j)\\ &\qquad+(\chi_i-\tilde\chi_i)D^2\psi_j+\tilde\chi_iD^2(\psi_j-\tilde\psi_j)].\end{aligned}$$

The typical estimates are

$$\begin{aligned}\|2ai\|_2 &\le g\,\|D^2(\chi_i-\tilde\chi_i)\|_2\,\|\psi_j\|_\infty\\ &\le g\ \text{const}\,\|\chi-\tilde\chi\|_{M^1_{3/2}}\|\psi\|_{D_2},\\ \|2bi\|_2 &\le g^2\,\|D(\chi_i-\tilde\chi_i)\|_4\,\|D\psi_j\|_4\\ &\le g\ \text{const}\,\|\chi-\tilde\chi\|_{M^1_{3/2}}\|\psi\|_{D_2},\\ \|2ci\|_2 &\le g\,\|\chi_i-\tilde\chi_i\|_\infty\,\|D^2\psi_j\|_2\\ &\le g\ \text{const}\,\|\chi-\tilde\chi\|_{M^1_{3/2}}\|\psi\|_{D_2}.\end{aligned}$$

The second estimate follows along the lines of the first except that terms with the derivative of g_c now appear and that we are now allowed to take other norms of the coupling function besides the sup norm. For example,

$$\begin{aligned}\|0ai\|_q &\le \|g_c\|_{2q(2-q)^{-1}}\,\|\chi_i-\tilde\chi_i\|_\infty\,\|\psi_j\|_2\\ &\le \text{const}\,\|g_c\|_{2q(2-q)^{-1}}\,\|\chi-\tilde\chi\|_{M^1_{3/2}}\,\|\psi\|_{D_2},\end{aligned}$$

which reduces to the previous case with $q=2$. Likewise,

$$\|0aii\| \le \text{const}\,\|g_c\|_{2q(2-q)^{-1}}\,\|\tilde\chi\|_{M^1_{3/2}}\,\|\psi-\tilde\psi\|_{D_2}.$$

In the first-order terms besides the above there are also terms like $0ai$ and $0aii$ with g_c replaced by g_c'. Proceeding as above, one obtains

$$\begin{aligned}&\|D[g_c(\chi_i\psi_j-\tilde\chi_i\tilde\psi_j)]\|_q\\ &\quad\le \text{const}\,(\|g_c\|_{2q(2-q)^{-1}}\\ &\qquad+\|g_c'\|_{2q(2-q)^{-1}})\,(\|\chi-\tilde\chi\|_{M^1_{3/2}}\|\psi\|_{D_2}\\ &\qquad+\|\tilde\chi\|_{M^1_{3/2}}\|\psi-\tilde\psi\|_{D_2}).\end{aligned}$$

Similarly, for the second-order terms,

$$\begin{aligned}\|D^2[g_c(\chi_i\psi_j-\tilde\chi_i\tilde\psi_j)]\|_q &\le \text{const}\,\|g_c\|_{2,2q(2-q)^{-1}}\\ &\quad\times(\|\chi-\tilde\chi\|_{M^1_{3/2}}\|\psi\|_{D_2}+\|\tilde\chi\|_{M^1_{3/2}}\|\psi-\tilde\psi\|_{D_2}).\end{aligned}$$

The third estimate is obtained in the same manner with the emphasis placed on obtaining the maximum power on the sup norms. With the same notations the estimates can be summarized as follows:

$$\begin{aligned}&\|0ai\|_q \text{ and } \|0aii\|_q\\ &\quad\le \|g_c\|_q(\|\chi-\tilde\chi\|_\infty\|\psi\|_\infty+\|\tilde\chi\|_\infty\|\psi-\tilde\psi\|_\infty)\\ &\quad\le \text{const}\,\|g_c\|_q(\|\chi-\tilde\chi\|_\infty\|\psi\|_\infty^{1/2}\|\psi\|_{D_2}^{1/2}\\ &\qquad+\|\chi-\tilde\chi\|_\infty^{1/2}\|\tilde\chi\|_{D_2}^{1/2}\|\psi-\tilde\psi\|_\infty).\end{aligned}$$

In the first-order terms, besides the standard terms $1ai\cdots1bii$, there also appear expressions like $0ai$ and $0aii$ with g_c replaced by g_c'. These are treated as above. On the other hand,

$$\begin{aligned}\|1ai\|_q &\le \|g_c\|_{4q(4-q)^{-1}}\,\|D(\chi_i-\tilde\chi_i)\|_4\,\|\psi_j\|_\infty\\ &\le \text{const}\,\|g_c\|_{4q(4-q)^{-1}}\,\|\chi-\tilde\chi\|_\infty^{1/2}\\ &\quad\times\|\psi\|_\infty\,\|\chi-\tilde\chi\|_{M^1_{3/2}}^{1/2}.\end{aligned}$$

and likewise

$$\begin{aligned}\|1aii\|_q &\le \text{const}\,\|g_c\|_{4q(4-q)^{-1}}\\ &\quad\times\|\tilde\chi\|_\infty^{1/2}\,\|\tilde\chi\|_{M^1_{3/2}}^{1/2}\,\|\psi-\tilde\psi\|_\infty\\ \|1bi\|_q &\le \text{const}\,\|g_c\|_{4q(4-q)^{-1}}\,\|\chi-\tilde\chi\|_\infty\|\psi\|_{D_2}^{1/2}\|\psi\|_\infty^{1/2}\\ \|1bii\| &\le \text{const}\,\|g_c\|_{4q(4-q)^{-1}}\\ &\quad\times\|\tilde\chi\|_\infty\|\psi-\tilde\psi\|_\infty^{1/2}\,\|\psi-\tilde\psi\|_{D_2}^{1/2}.\end{aligned}$$

These estimates can be summed up as suggested in the beginning of the proof to obtain the desired inequalities.

Remark 1: In order to show that the interactions map $D_2\oplus M_{3/2}$ into itself in a semi-Lipschitz fashion, take $G=g$ or g_c, $q=2$ in (ii), and proceed as follows:

$$\begin{aligned}&\left\|G(V\psi-\tilde V\tilde\psi)\oplus G\begin{pmatrix}0\\ J-\tilde J\end{pmatrix}\right\|_{D_2\oplus M_{3/2}}\\ &\quad\le \text{const}\,[\|G(V\psi-\tilde V\tilde\psi)\|_{2,2}^2+\|G(J-\tilde J)\|_{1,2}^2]^{1/2}\\ &\quad\le \text{const}\,\|G\|_{2,\infty}(\|v\|_{M^1_{3/2}}+\|\tilde v\|_{M^1_{3/2}}\\ &\qquad+\|\psi\|_{D_2}+\|\tilde\psi\|_{D_2})(\|v-\tilde v\|_{M^1_{3/2}}+\|\psi-\psi\|_{D_2})\\ &\quad\le \text{const}\,K\,\|G\|_{2,\infty}\left\|\left(\psi,\begin{pmatrix}v\\ \dot v\end{pmatrix}\right)-\left(\tilde\psi,\begin{pmatrix}\tilde v\\ \dot{\tilde v}\end{pmatrix}\right)\right\|_{D_2\oplus M_{3/2}}\end{aligned}$$

if $(\psi,\binom{v}{\dot v})$ and $(\psi,\binom{\tilde v}{\dot{\tilde v}})$ are in a ball of radius K in $D_2\oplus M_{3/2}$. Taking $(\tilde\psi,\binom{\tilde v}{\dot{\tilde v}})\equiv 0$ gives

$$\begin{aligned}&\left\|GV\psi\oplus G\begin{pmatrix}0\\ J\end{pmatrix}\right\|_{D_2\oplus M_{3/2}}\\ &\qquad\le \text{const}\,\|G\|_{2,\infty}\left\|\left(\psi,\begin{pmatrix}v\\ \dot v\end{pmatrix}\right)\right\|_{D_2\oplus M_{3/2}}^2,\end{aligned}$$

which proves most simply that the interactions map $D_2\oplus M_{3/2}$ into itself.

Remark 2: The estimates in inequalities (20), (23), and (28) are obtained from (iii) as follows: Take $\chi=\psi$ and $\tilde\chi,\tilde\psi\equiv 0$. Using the fact that $\|\psi\|_{M^1_{3/2}}\le \text{const}\,\|\psi\|_{D_2}$, we obtain

$$\begin{aligned}\|g_cJ(t)\|_{1,q} &\le \text{const}\,(\|g_c\|_{1,q}\\ &\quad+\|g_c\|_{4q(4-q)^{-1}}\|\psi\|_\infty^{3/2}\|\psi\|_{D_2}^{1/2}.\end{aligned}$$

From the definition in Eq. (18), $\|\psi(t)\|_\infty\le(1+|t|)^{-\epsilon}\,]\psi[_t$ and $\|\psi(t)\|_{D_2}\le\,]\psi[_t$, which when placed in the above gives

$$\begin{aligned}\|g_cJ(t)\|_{1,q} &\le \text{const}\,(\|g_c\|_{1,q}\\ &\quad+\|g_c\|_{4q(4-q)^{-1}})\,(1+|t|)^{-3\epsilon/2}\,]\psi[_t^2,\end{aligned}$$

as required.

* Research supported in part by the National Science Foundation (NSF GP-13627) and the Alfred P. Sloan Foundation.
† On leave from the Department of Mathematics, Indiana University, Bloomington, Indiana 47401.
1 The v^μ are the components of the electromagnetic (real) vector field and ψ is the Dirac spinor field; i.e., ψ is a function from space-time into a four-dimensional vector space called spin space. The positive definite inner product in spin space is denoted by $\psi^\dagger\psi$ and $\bar{\psi}$ denotes $\psi^\dagger\gamma^0$. The γ's are operators in spin space which satisfy $\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}(g^{00} = 1, g^{ii} = -1, g^{\mu\nu} = 0\ \mu \neq \nu)$ and $\gamma^{0*} = \gamma^0, \gamma^{i*} = -\gamma^i$. All representations of operators satisfying these properties are unitarily equivalent. For this work, as usual, no specific choice is required.
2 L. Gross, Commun. Pure Appl. Math. **19**, 1 (1966).
3 I. E. Segal, Ann. Sci. Ecole Norm. Super., Ser 4, **1**, 459 (1968).
4 J. M. Chadam, "Asymptotics for $\Box u = m^2u + G(x, t, u, u_t, u_x)$. I. Global Existence and Decay," Ann. Scuola Norm. Super. Pisa (to be published).
5 J. M. Chadam, Paper II of Ref. 4, Ann. Scuola Norm. Super. Pisa (to be published).
6 J. M. Chadam, "Asymptotic Behavior of Equations Arising in Quantum Theory." J. Appl. Anal. (to be published).
7 In this work it will be notationally convenient to view the four component spinor ψ as a pair of two component spinors.
8 The usual notations for function spaces [e.g., $L^p(E_3)$ with norm $\|\ \|_p$, $\mathcal{S}(E_3)$, $C_c^\infty(E_3)$, etc.] will be used (as above) for the appropriate direct sum analog of these spaces when there is no possibility of confusion.
9 The fact that these M_β are not the standard escalated energy spaces is a reflection of the requirement that the vector potentials must satisfy Eq. (1c).
10 I. E. Segal, Ann. Math. **78**, 339 (1963).
11 T. Kato, J. Math. Soc. Japan **5**, 208 (1953).
12 T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966), pp. 305–06.
13 J. Bjorken and S. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965).
14 N. Shenk and D. Thoe, J. Math. Anal. Appl. **31**, 81 (1970), Lemma 3.1.
15 E.g., W. A. Strauss, J. Functional Anal. **2**, 409 (1968), Lemma 3.7.

A Functional Equation in the Theory of Fluids*

O. Penrose† and J. L. Lebowitz
Belfer Graduate School of Science, Yeshiva University, New York, New York 10033
(Received 4 August 1971)

Two functional equations of the form $\psi^2(s) - E(s)\phi^2(s) = V(s)$, where s is a complex variable and $E(s)$ and $V(s)$ are given even polynomials, are solved for the even entire functions ψ and ϕ which are required to behave like $\cosh[\alpha s + o(s)]$ for large $|Rls|$. Two cases are considered: (i) V of degree zero and E of degree two and (ii) V of degree eight and E of degree six. In the second case the polynomials must satisfy a condition in order for ψ and ϕ to have the right asymptotic behavior. These functional equations arise in solving the Percus–Yevick equation for a mixture of hard spheres with nonadditive diameters.

1. INTRODUCTION

This is a sequel to the paper by Lebowitz and Zomick.[1] That paper will be referred to below as LZ. The problem considered in it was that of solving the Percus–Yevick (PY) integral equation[2] for a binary mixture of particles in which the two-body interaction potential $u_{ij}(r)$ between a particle of species i and one of species j a distance r apart is given by

$$u_{ij}(r) = \begin{cases} +\infty & (0 \leq r \leq R_{ij}) \\ 0 & (r > R_{ij}) \end{cases}, \qquad i, j \in \{1, 2\}, \tag{1.1}$$

where

$$R_{12} = R_{21} = \tfrac{1}{2}(R_{11} + R_{22}) + \alpha, \quad \text{with} \quad 0 \leq \alpha \leq \tfrac{1}{2}(R_{22} - R_{11}). \tag{1.2}$$

It was shown by LZ that the solution of the PY equation for this problem involved a functional equation of the form

$$\psi^2(s) - E(s)\phi^2(s) = V(s), \tag{1.3}$$

where s is a complex variable, E and V are polynomial functions of a given form, but with undetermined coefficients, and ϕ and ψ are the unknown functions (related to the Laplace transforms of certain correlation functions), which are required to be even and entire and to satisfy the asymptotic condition

$$\left.\begin{matrix} ln|\psi(s)| \\ ln|\phi(s)| \end{matrix}\right\} \sim \begin{cases} \alpha Rls & \text{as } Rls \to \infty \\ -\alpha Rls & \text{as } Rls \to -\infty \end{cases}, \tag{1.4}$$

where α is a positive number defined in (1.2).

In the present paper we give a method for solving equations of the form (1) and apply it to the particular equations of this form which arise in LZ. The solution for the particular case which refers to the one-dimensional hard-rod system with potential (1.1) was quoted, without derivation, in LZ.

The basic idea of the method can be seen by considering the case where $E(s) = V(s) = 1$. Here the equation reduces to

$$\psi(s)^2 - \phi(s)^2 = 1, \tag{1.5}$$

which can be written

$$[\psi(s) + \phi(s)][\psi(s) - \phi(s)] = 1. \tag{1.6}$$

This shows that both $\psi + \phi$ and $\psi - \phi$ are entire functions without zeros and hence must be of the form $\exp[w(s)]$ and $\exp[-w(s)]$, respectively, where w is entire; it follows that the solution is

$$\psi(s) = \cosh w(s), \qquad \phi(s) = \sinh w(s).$$

The method we shall describe is based on the same factorization idea, with modifications to allow for the presence of the polynomials $E(s)$ and $V(s)$ and the conditions that the functions ϕ and ψ must be even and satisfy the asymptotic condition (1.4).

In the next stage of the calculation we shall use the following information about the specific form of the polynomial $E(s)$; it is taken from LZ. The details depend on the number of dimensions of the hard-sphere system.

A. One Dimension

Equation (1.3) above corresponds to Eq. (3.57) of LZ with

$$E(s) = s^2 - 4\mu^2 \quad \{= D(s)/[(\rho_1 - \rho_2)^2 a^2] \text{ in LZ notation}\}, \tag{1.7}$$

$$V(s) = A, \tag{1.8}$$

* Research supported in part by the National Science Foundation (NSF GP-13627) and the Alfred P. Sloan Foundation.
† On leave from the Department of Mathematics, Indiana University, Bloomington, Indiana 47401.
1 The v^μ are the components of the electromagnetic (real) vector field and ψ is the Dirac spinor field; i.e., ψ is a function from space-time into a four-dimensional vector space called spin space. The positive definite inner product in spin space is denoted by $\psi^\dagger\psi$ and $\bar{\psi}$ denotes $\psi^\dagger\gamma^0$. The γ's are operators in spin space which satisfy $\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}(g^{00} = 1, g^{ii} = -1, g^{\mu\nu} = 0\ \mu \neq \nu)$ and $\gamma^{0*} = \gamma^0, \gamma^{i*} = -\gamma^i$. All representations of operators satisfying these properties are unitarily equivalent. For this work, as usual, no specific choice is required.
2 L. Gross, Commun. Pure Appl. Math. **19**, 1 (1966).
3 I. E. Segal, Ann. Sci. Ecole Norm. Super., Ser 4, **1**, 459 (1968).
4 J. M. Chadam, "Asymptotics for $\Box u = m^2u + G(x, t, u, u_t, u_x)$. I. Global Existence and Decay," Ann. Scuola Norm. Super. Pisa (to be published).
5 J. M. Chadam, Paper II of Ref. 4, Ann. Scuola Norm. Super. Pisa (to be published).
6 J. M. Chadam, "Asymptotic Behavior of Equations Arising in Quantum Theory." J. Appl. Anal. (to be published).
7 In this work it will be notationally convenient to view the four component spinor ψ as a pair of two component spinors.
8 The usual notations for function spaces [e.g., $L^p(E_3)$ with norm $\|\ \|_p$, $\mathcal{S}(E_3)$, $C_c^\infty(E_3)$, etc.] will be used (as above) for the appropriate direct sum analog of these spaces when there is no possibility of confusion.
9 The fact that these M_β are not the standard escalated energy spaces is a reflection of the requirement that the vector potentials must satisfy Eq. (1c).
10 I. E. Segal, Ann. Math. **78**, 339 (1963).
11 T. Kato, J. Math. Soc. Japan **5**, 208 (1953).
12 T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966), pp. 305–06.
13 J. Bjorken and S. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965).
14 N. Shenk and D. Thoe, J. Math. Anal. Appl. **31**, 81 (1970), Lemma 3.1.
15 E.g., W. A. Strauss, J. Functional Anal. **2**, 409 (1968), Lemma 3.7.

A Functional Equation in the Theory of Fluids*

O. Penrose† and J. L. Lebowitz
Belfer Graduate School of Science, Yeshiva University, New York, New York 10033
(Received 4 August 1971)

Two functional equations of the form $\psi^2(s) - E(s)\phi^2(s) = V(s)$, where s is a complex variable and $E(s)$ and $V(s)$ are given even polynomials, are solved for the even entire functions ψ and ϕ which are required to behave like $\cosh[\alpha s + o(s)]$ for large $|Rls|$. Two cases are considered: (i) V of degree zero and E of degree two and (ii) V of degree eight and E of degree six. In the second case the polynomials must satisfy a condition in order for ψ and ϕ to have the right asymptotic behavior. These functional equations arise in solving the Percus–Yevick equation for a mixture of hard spheres with nonadditive diameters.

1. INTRODUCTION

This is a sequel to the paper by Lebowitz and Zomick.[1] That paper will be referred to below as LZ. The problem considered in it was that of solving the Percus–Yevick (PY) integral equation[2] for a binary mixture of particles in which the two-body interaction potential $u_{ij}(r)$ between a particle of species i and one of species j a distance r apart is given by

$$u_{ij}(r) = \begin{cases} +\infty & (0 \leq r \leq R_{ij}) \\ 0 & (r > R_{ij}) \end{cases}, \qquad i, j \in \{1, 2\}, \tag{1.1}$$

where

$$R_{12} = R_{21} = \tfrac{1}{2}(R_{11} + R_{22}) + \alpha, \quad \text{with} \quad 0 \leq \alpha \leq \tfrac{1}{2}(R_{22} - R_{11}). \tag{1.2}$$

It was shown by LZ that the solution of the PY equation for this problem involved a functional equation of the form

$$\psi^2(s) - E(s)\phi^2(s) = V(s), \tag{1.3}$$

where s is a complex variable, E and V are polynomial functions of a given form, but with undetermined coefficients, and ϕ and ψ are the unknown functions (related to the Laplace transforms of certain correlation functions), which are required to be even and entire and to satisfy the asymptotic condition

$$\left.\begin{matrix} ln|\psi(s)| \\ ln|\phi(s)| \end{matrix}\right\} \sim \begin{cases} \alpha Rls & \text{as } Rls \to \infty \\ -\alpha Rls & \text{as } Rls \to -\infty \end{cases}, \tag{1.4}$$

where α is a positive number defined in (1.2).

In the present paper we give a method for solving equations of the form (1) and apply it to the particular equations of this form which arise in LZ. The solution for the particular case which refers to the one-dimensional hard-rod system with potential (1.1) was quoted, without derivation, in LZ.

The basic idea of the method can be seen by considering the case where $E(s) = V(s) = 1$. Here the equation reduces to

$$\psi(s)^2 - \phi(s)^2 = 1, \tag{1.5}$$

which can be written

$$[\psi(s) + \phi(s)][\psi(s) - \phi(s)] = 1. \tag{1.6}$$

This shows that both $\psi + \phi$ and $\psi - \phi$ are entire functions without zeros and hence must be of the form $\exp[w(s)]$ and $\exp[-w(s)]$, respectively, where w is entire; it follows that the solution is

$$\psi(s) = \cosh w(s), \qquad \phi(s) = \sinh w(s).$$

The method we shall describe is based on the same factorization idea, with modifications to allow for the presence of the polynomials $E(s)$ and $V(s)$ and the conditions that the functions ϕ and ψ must be even and satisfy the asymptotic condition (1.4).

In the next stage of the calculation we shall use the following information about the specific form of the polynomial $E(s)$; it is taken from LZ. The details depend on the number of dimensions of the hard-sphere system.

A. One Dimension

Equation (1.3) above corresponds to Eq. (3.57) of LZ with

$$E(s) = s^2 - 4\mu^2 \quad \{= D(s)/[(\rho_1 - \rho_2)^2 a^2] \text{ in LZ notation}\}, \tag{1.7}$$

$$V(s) = A, \tag{1.8}$$

where μ and A are positive numbers.

Our $\psi(s)$ and $\phi(s)$ correspond to $\psi(s)$ and $[Q(s)(\rho_1 - \rho_2)a]$ in the notation of LZ.

B. Three Dimensions

Equation (1.3) corresponds to Eq. (5.52) of LZ with

$$E(s) = s^6 - 4h(s), \tag{1.9}$$

$$V(s) = -h(s)A(s), \tag{1.10}$$

where

$h(s) = h_1 s^2 + h_2$ [Eq. (5.38) of LZ],

and

$A(s) = a_1 s^6 + a_2 s^4 + a_3 s^2 + a_4$ [Eq. (5.43) of LZ].

Here $h_1, \ldots, a_4$ are real numbers. This time our $\psi(s)$ and $\phi(s)$ correspond to the $\psi(s)$ and $\phi(s)$ of LZ.

2. FACTORIZATION OF THE EQUATION

To factorize $\psi^2 - E\phi^2$ we define $\sqrt{E}(s)$ to be the branch of the many-valued function $[E(s)]^{1/2}$ which is analytic in the complex s plane with suitable finite cuts, and which satisfies

$$\sqrt{E}(s) \sim s^p \quad \text{for large } s, \tag{2.1}$$

where p is the number of dimensions. This is possible because $E(s) = s^{2p} + \cdots$ both for $p = 1$ and for $p = 3$. Before we proceed, it is worth noting that if the cuts are symmetrical about the origins, as we shall choose them to be, then $\sqrt{E}$ is an odd function of s; for, since E itself is even, the ratio $\sqrt{E}(s)/\sqrt{E}(-s)$ when squared gives 1 and must therefore equal + 1 or − 1 throughout the cut plane, and Eq. (2.1) shows that the value is − 1 both for $p = 1$ and for $p = 3$.

We can now factorize the left-hand side of (1.3) by defining

$$f(s) = \psi(s) + \sqrt{E}(s)\phi(s), \tag{2.2}$$

$$\overline{f}(s) = \psi(s) - \sqrt{E}(s)\phi(s) = f(-s) \tag{2.3}$$

since $\sqrt{E}$ is odd and ψ and ϕ are even. Equation (1.3) now becomes

$$f(s)\overline{f}(s) = V(s) \tag{2.4}$$

or

$$f(s)f(-s) = V(s). \tag{2.4'}$$

The behavior of $f(s)$ for large s is determined by the condition (1.4). This condition, with (2.2) and (2.3), gives

$$|\psi(s)| = \tfrac{1}{2}|f(s) + f(-s)| = \exp[\alpha\,|Rls| + o(s)], \quad |Rls| \to \infty, \tag{2.5}$$

so that either $|f(s)|$ or $|f(-s)|$ (or both) grows at least as fast as $\exp[\alpha|Rls| + o(s)]$ as $|Rls| \to \infty$. Suppose for definiteness that it is $f(s)$ that has this property for $Rls \to \infty$; then (2.4) implies that $f(-s)$ decreases exponentially as $Rls \to \infty$ and thus $f(s)$ has to decrease exponentially as $Rls \to -\infty$. Thus we find that in this case

$$|f(s)| \sim \exp[\alpha Rls + o(s)] \quad \text{as } |Rls| \to \infty. \tag{2.6}$$

The alternative possibility gives a similar result, and the two can be combined in the single formula

$$|f(s)| \sim \exp[\pm\, \alpha Rls + o(s)]$$

$$\text{as} \quad Rls \to \infty \ \textit{and} \ \text{as} \ Rls \to \infty. \tag{2.7}$$

We also want to know how f behaves on the boundary of the cut plane, i.e., on the cuts themselves. Each cut has two sides which we designate the + and the − side, respectively. If the cut is horizontal, it is convenient to take the + side as the top. We define, for each function F over the cut plane, two boundary-value functions on each cut,

$$F^{\pm}(t) = \lim_{\epsilon \to 0} F(t + \epsilon), \qquad t \in C, \tag{2.8}$$

the limit being approached from the + side of the cut C for F^+ and the − side for F^-. For example, since $\sqrt{E}$ reverses sign as we cross any cut, it satisfies

$$\sqrt{E}^+(t) = -\sqrt{E}^-(t) \qquad (t \in C) \tag{2.9}$$

Using this in (2.3), we obtain

$$f^+(t) = \overline{f}^-(t) \quad \text{and} \quad \overline{f}^+(t) = f^-(t).$$

It follows, by (2.4), that f satisfies the boundary condition

$$f^+(t)f^-(t) = V(t) \qquad (t \in C). \tag{2.10}$$

This equation, together with (2.7), constitutes a boundary-value problem for determining f, of the type discussed by Muskhelishvili.[3] In the following, we shall apply his methods to it.

3. THE ONE-DIMENSIONAL CASE

In this case $V(t)$ is a positive constant A so that the functional equation reduces to

$$f(s)f(-s) = A, \qquad s \in \text{cut plane}, \tag{3.1}$$

and (2.10) to

$$f^+(t)f^-(t) = A, \qquad t \in [-2\mu, 2\mu]. \tag{3.2}$$

There is just one cut now, since $\sqrt{E}(s) = \sqrt{(s^2 - 4\mu^2)}$, with two branch points.

We shall write (3.2) in logarithmic form, in terms of

$$\Phi(s) = \ln f(s), \tag{3.3}$$

but it is necessary first to show that $\Phi(s)$ is a (single-valued) function. That is to say, we wish to show that $f(s)$ never vanishes, and that $\arg f(s)$ returns to its original value when s describes any closed contour in the cut plane. Both facts follow from (3.1), the first because $f(-s)$ is holomorphic in the cut plane and the second because any closed contour in the cut plane can be deformed either to zero or to a large circular contour; if it deforms to zero there is no problem, and if not, both s and $-s$ traverse the same circle, so that the total change in $\arg f(s)$ equals the total change in $\arg f(-s)$, and since their sum is the total change in $\arg f(s)f(-s) = \arg A$, i.e., zero, the total change in $\arg f(s)$ is also 0. Thus the definition (3.3) makes sense; it is only necessary to specify the

additive multiple of $2\pi i$ and we do this in such a way that

$$\Phi(s) \sim \pm \alpha s + o(s), \qquad |Rls| \to \infty, \tag{3.4}$$

in conformity with the asymptotic condition (2.7).

Putting (3.3) into (3.2), and using (2.7), the boundary condition on the cut takes the form

$$\Phi^+(t) = \Phi^-(t) = \ln A + 2\pi m i, \tag{3.5}$$

where m is some integer.

Boundary-value problems of this form are discussed[3] in Sec. 84 of M. Since $f(s)$ is bounded in any finite region of the cut plane and on its boundaries [by (2.3)] and is therefore also bounded away from zero [by (2.4′)], its logarithm $\Phi(s)$ is bounded. Thus the solutions of (3.4) we are interested in are those which, in the terminology defined on p. 231 of M, belong to the class h $(-2\mu, 2\mu)$ (i.e., they are bounded at both ends of the cut.)

A particular solution of the boundary-value problem satisfying this condition is the function

$$\tfrac{1}{2}(\ln A + 2\pi m i).$$

The general solution of (3.5) is obtained from this particular solution by adding the general solution of the associated homogeneous equation $\Phi^+(t) + \Phi^-(t) = 0$. According to the theory described in Secs. 35 and 79 of M, this general solution, of finite degree at infinity, is of the form

$$X(s)P(s),$$

where P is a polynomial and X is the "fundamental solution" of the homogeneous problem. This fundamental solution is (see p. M232) the one which is nowhere zero in the finite part of the plane, including the cut, except at the two end points -2μ and 2μ. The appropriate solution, given in Eq. (84.3) of M, is $\sqrt{E(s)}$, and so we have for our general solution

$$\Phi(s) = \tfrac{1}{2}(\ln A + 2\pi m i) + \sqrt{E(s)}P(s), \tag{3.6}$$

where $P(s)$ is a polynomial. The condition (3.4) determines this polynomial as $P(s) = \pm \alpha$, and, putting the resulting formula for $\Phi(s)$ into (3.3), we obtain

$$f(s) = \pm \sqrt{A}\, \exp[\pm \alpha \sqrt{E(s)}], \tag{3.7}$$

a formula in which there are two choices of sign.

The solution of the original functional equation for the one-dimensional case is therefore, by (2.3),

$$\psi(s) = \pm \sqrt{A}\, \cosh \alpha \sqrt{E(s)}, \tag{3.8}$$

$$\phi(s) = \pm \sqrt{A}[\sinh \alpha \sqrt{E(s)}]/\sqrt{E(s)}, \tag{3.9}$$

again with two choices of sign. This is the solution given in Eqs. (3.58) and (3.59) of LZ.

4. THE THREE-DIMENSIONAL CASE

To deal with this case, we return to the general form of the boundary-value problem given in (2.10). Now $E(s)$ is a polynomial of degree 6. We denote its roots by $\pm s_1, \pm s_2, \pm s_3$, with s_2 and $s_3 - s_1$ real, and choose the cuts as shown in Fig. 1.

Another new feature is that now $V(t)$ is a polynomial of degree 8, so that $f(s)$ and $f(-s)$ must have zeros in the cut plane or on its boundaries. In fact, if the zeros of $f(s)$ are z_1, z_2, z_3, z_4, then those of $f(-s)$ are $-z_1, -z_2, -z_3, -z_4$ and so those of $V(s)$ are $\pm z_1$, $\pm z_2, \pm z_3, \pm z_4$. We assume that $V(s)$ and $E(s)$ have no zeros in common, so that none of $\pm z_1, \ldots, \pm z_4$ is the end of a cut. Let us define

$$F(s) = f(s) \prod_{i=1}^{4} (s - z_i)^{-1} \tag{4.1}$$

so that, by (2.10),

$$F(s)F(-s) = V(s) \prod_{i=1}^{4} [(s - z_i)(s + z_i)]^{-1} = c, \tag{4.2}$$

where c is the coefficient of s^8 in $V(s)$ (it is $-a_1 h_1$ in the notation of LZ). Thus F satisfies a functional equation like the one we solved before. The boundary conditions are

$$|F(s)| = \exp[\pm \alpha Rls + o(s)] \tag{4.3}$$

for large $|(Rls)|$, and

$$F^+(t)F^-(t) = cJ(t) \qquad (t \in C), \tag{4.4}$$

where

$$J(t) = V(t)[c \prod_{i=1}^{4} (t - z_i)]^{-2} = \prod_{i=1}^{4} \frac{(t + z_i)}{(t - z_i)} \tag{4.5}$$

and C denotes the set of cuts.

Since there are now three cuts, it is not as easy as before to define a function Φ to serve as $\ln F$. By considering the behavior of $F(s)$ as s traverses a large circle, we can show that $\arg F(s)$ returns to its original value if s traverses a contour that encircles all three cuts, and, by considering a symmetrical contour that encloses only the cut through the origin, we can show that $\arg F(s)$ returns to its original value as s

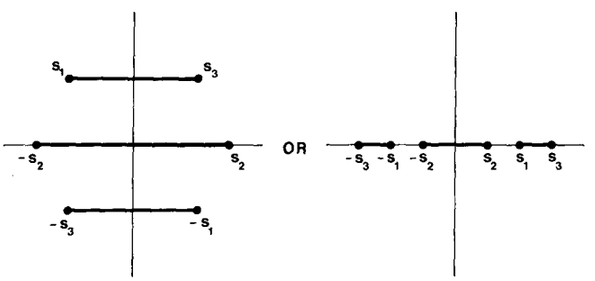


FIG. 1. Cuts C in the s plane.

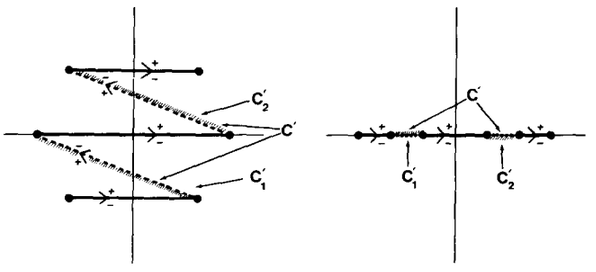


FIG. 2. Cuts C' in the s plane.

traverses any contour enclosing only this cut; but there is as yet no reason to expect $\arg F(s)$ to return to its original value if s traverses a contour that surrounds one of the other cuts. To deal with this, we put two further cuts C' as shown dotted in Fig. 2.

We can now define a function

$$\Phi(s) = \ln F(s), \tag{4.6}$$

which is holomorphic in the cut plane and satisfies

$$\begin{aligned} \Phi^+(t) + \Phi^-(t) &= \ln c + 2\pi m i + \ln J(t) \quad (t \in C), \\ \Phi^+(t) - \Phi^-(t) &= 2\pi n i \quad (t \in C'), \end{aligned} \tag{4.7}$$

where m and n are integers. The reason why we can use the same value of n on both parts of C' is that we know [since $F(s)F(-s) = c$] that if s traverses a symmetrical path enclosing only the part of C containing the origin, then $\arg F$ will return to its original value and so the discontinuities of Φ across the two parts of C' must be the same.

Equation (4.7) is again a Hilbert problem, of the type discussed in Secs. 85–87 of M. As before, its general solution is the sum of a particular solution and the general solution of the associated homogeneous problem. For a particular solution, consider the function [obtained from Eq. M (85.12)]

$$\Phi(s) = \tfrac{1}{2}\ln c + \pi m i + \frac{\sqrt{E(s)}}{2\pi i}\int_C \frac{\ln J(t)dt}{\sqrt{E^+(t)}(t-s)} + \frac{\sqrt{E(s)}}{2\pi i}\int_{C'} \frac{2\pi n i dt}{\sqrt{E^+(t)}(t-s)}, \tag{4.8}$$

where the cuts are traversed in the directions indicated in Fig. 2, and their "plus" sides are at the left of the arrows. The Plemelj formulas (p. M43) show that this is a particular solution of the boundary-value problem (4.7). To get the general solution, we add $X(s)P(s)$ where P is any polynomial and X is the fundamental solution of the class we are looking for. As before, since we want bounded solutions, this fundamental solution is $\sqrt{E(s)}$. Our asymptotic condition (4.3) implies, however, that the polynomial P must be 0, since $\sqrt{E(s)}$ increases like s^3, not s, for large $|(R/s)|$. Thus the particular solution (4.8) itself is the one we are looking for.

To get the correct behavior for large s, we want

$$\Phi(s) \sim \pm \alpha s \quad \text{for large } s,$$

and, since $E(s) \sim s^3$, this implies that the sum of the two integrals must behave like $\pm 2\pi i\alpha / s^2$ for large s. The asympototic behavior of these integrals is obtained from the formula $(t-s)^{-1} = -s^{-1} - ts^{-2} \cdots$, and so we require [writing $\sqrt{E(t)}$ as an abbreviation for ${}^s\sqrt{E^+(t)}$]

$$\frac{1}{2\pi i}\int_C \frac{\ln J(t)dt}{\sqrt{E(t)}} + n\int_{C'} \frac{dt}{\sqrt{E(t)}} = 0 \tag{4.9}$$

(coefficient of s^{-1}), and

$$\frac{1}{2\pi i}\int_C \frac{t\ln J(t)dt}{\sqrt{E(t)}} + n\int_{C'} \frac{tdt}{\sqrt{E(t)}} = \mp\alpha \tag{4.10}$$

(coefficient of s^{-2}).

Now, Eq. (4.5) shows that $\ln J(t)$ is an odd function; hence, the first integral in (4.9) is zero. The second can be written

$$\int_{C_1'} \frac{dt}{\sqrt{E^+(t)}} - \int_{C_2'} \frac{dt}{\sqrt{E^-(t)}}. \tag{4.11}$$

The sides of the contours on which the integrand is evaluated are indicated in Fig. 2, and are symmetrical. Since $\sqrt{E(s)}$ is an odd function, it follows that the two integrals reinforce and so [barring the accident $\int dt/\sqrt{E(t)} = 0$] the condition (4.9) can only be satisfied by taking $n = 0$. The second condition, (4.10), now reduces to

$$\alpha = \mp\frac{1}{2\pi i}\int_C \frac{t\ln J(t)dt}{\sqrt{E(t)}}, \tag{4.12}$$

which imposes a condition on the coefficients $h_i, \ldots, a_4$ in the polynomials $E(s)$ and $V(s)$.

The solution of the functional equation for f is, by (4.6) and (4.1),

$$f(s) = \pm\sqrt{c}\prod_{i=1}^{4}(s-z_i)\left\{\exp\frac{\sqrt{E(s)}}{2\pi i}\int_C \frac{\ln J(t)dt}{\sqrt{E^+(t)}(t-s)}\right\} \tag{4.13}$$

There are several solutions, depending upon the choice of the sign in (4.13) and which four zeros of V we pick to use as $z_1, \ldots, z_4$. We expect that, as in the one-dimensional case, physical requirements will dictate a unique choice. This will certainly be true at low densities where the PY solution can be found from a convergent virial expansion.[4] At higher densities we expect to see a phase transition corresponding to a spatial separation of the two components since α in (1.2) is positive. We plan to investigate this further.

ACKNOWLEDGMENT

We should like to thank Professor J. Clunie for showing us an alternative method of dealing with the one-dimensional case.

---

* Supported by U.S.A.F.O.S.R. Contract No. F44620-71-C-0013.

† Present address: The Open University, Walton, Bletchley, Bucks, England.

1 J. L. Lebowitz and D. Zomick, J. Chem. Phys. **54**, 3335 (1971).

2 J. K. Percus and G. J. Yevick, Phys. Rev. **110**, 1 (1957).

3 N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, 1953). This will be referred to in the text as M.

4 J. Groeneveld, J. Chem. Phys. **53**, 3193 (1970).

Entropy Decomposition and Transfer-Matrix Problems

H. Falk

Department of Physics, City College of the City University of New York, New York, New York 10031

(Received 11 November 1971)

For a system which may be partitioned into M subsystems $A_1, A_2, \ldots, A_M$, such that configurations of subsystems A_i, A_j correspond to realizations of discrete, random variables s_i, s_j, not necessarily isomorphic, and such that the probability for a configuration of the total system is $p(s_1, s_2, \ldots, s_M) = f_1(s_1)\cdots f_M(s_M)g_1(s_1, s_2) \times g_2(s_2, s_3)\cdots g_{M-1}(s_{M-1}, s_M)$, we prove that $p(s_1, s_2, \ldots, s_M) = p_1(s_1|s_2)p_2(s_2|s_3)\cdots p_{M-1}(s_{M-1}|s_M)p_M(s_M)$, where $p_i(s_i|s_{i+1})$ is the conditional probability for s_i given s_{i+1}, and $p_i(s_i)$ is the reduced probability for s_i. This result yields a decomposition of the total entropy into "single-subsystem entropies" $-\sum p_i(s_i)\ln p_i(s_i)$ and "nearest-neighbor" cumulant-like terms $-\sum p_j(s_j, s_{j+1})\ln[p_j(s_j, s_{j+1})/p_j(s_j)p_{j+1}(s_{j+1})]$ only. This Markovian decomposition applies to systems with short-range interactions for which transfer-matrix methods are introduced.

1. INTRODUCTION

This paper deals with an aspect of the decomposition of entropy of physical systems. Attention is directed to the way in which the entropy depends on a very limited subset of correlations which exist in a class of systems described in terms of transfer matrices.[1] The Markovian decomposition demonstrated for chainlike interactions is admittedly so intuitively clear[2] and mathematically direct that to present it in a separate note may seem questionable; however, we believe that the result is not well known in the present context, and the recent appearance [3–6] of entropy and correlation inequalities may suggest some new approximations for unsolved problems.

2. DEFINITIONS AND RESULTS

Consider a system which may be partitioned into M subsystems $A_1, A_2, \ldots, A_M$, such that the configurations of subsystems A_i and A_j correspond to the possible realizations of discrete random variables s_i and s_j, respectively. For example, s_i might have only two realizations ± 1; whereas, s_{i+1} might have 3^{N^2} realizations so that s_i might be associated with a single spin $\frac{1}{2}$ and s_{i+1} might be associated with an entire N^2-spin plane of spin 1 particles. Let the probability $p(s_1, s_2, \ldots, s_M)$ for the joint realization of a set of values $s_1, s_2, \ldots, s_M$ denote the probability for a configuration of the total system and vice versa, and assume that

$$p(s_1, s_2, \ldots, s_M) = f_1(s_1)f_2(s_2)\cdots f_M(s_M)g_1(s_1, s_2) \times g_2(s_2, s_3)\cdots g_{M-1}(s_{M-1}, s_M), \quad (1)$$

where $f_i(s_i)$ and $g_i(s_i, s_{i+1})$ are positive-valued functions. A trivial example of a joint probability with the above structure is $[\exp(-\beta H)]/Z$, where H is the Hamiltonian for a linear chain of spins with nearest-neighbor interactions (Ising); each spin in a magnetic field. The canonical partition function is denoted by Z. The spins need not have the same magnitude and the interaction parameters J_i and the field parameters h_i, $i = 1, 2, \ldots, M$ may change in sign and magnitude as i takes on different values.

Introduce the reduced probabilities

$$p_i(s_i, s_{i+1}) = \sum_{[i,i+1]} p(s_1, s_2, \ldots, s_M), \quad (2)$$

$$p_{i+1}(s_{i+1}) = \sum_{s_i} p_i(s_i, s_{i+1}), \quad (3)$$

where $\sum_{[i,i+1]}$ means sum over all configurations $s_1, s_2, \ldots, s_M$ with s_i, s_{i+1} fixed. In the usual notation the conditional probability for s_i given s_{i+1} is

$$p_i(s_i|s_{i+1}) = p_i(s_i, s_{i+1})/p_{i+1}(s_{i+1}). \quad (4)$$

Theorem: Under the conditions stated above,

$$p(s_1, s_2, \ldots, s_M) = p_1(s_1|s_2)p_2(s_2|s_3)\cdots \times p_{M-1}(s_{M-1}|s_M)p_M(s_M). \quad (5)$$

Proof: Let s_{ik} denote the realizations of s_i and understand that when k is preceded by i, k may take on integer values $1, 2, \ldots,$ through a largest value $L(i)$. Introduce matrices

$$F_i = \mathrm{diag}[f_i(s_{i1}), f_i(s_{i2}), \ldots, f_i(s_{iL(i)})], \quad (6)$$

$$G_i = \begin{bmatrix} g_i(s_{i,1}, s_{i+1,1}) & \cdots & g_i(s_{i,1}, s_{i+1,L(i+1)}) \\ \vdots & & \vdots \\ g_i(s_{i,L(i)}, s_{i+1,1}) & \cdots & g_i(s_{i,L(i)}, s_{i+1,L(i+1)}) \end{bmatrix}. \quad (7)$$

Alternatively, a matrix element of, say, G_i will be denoted by $\langle s_i|G_i|s_{i+1}\rangle$.

Then

$$p_1(s_1, s_2) = \langle s_1|G_1|s_2\rangle\langle s_1|F_1|s_1\rangle\langle s_2|F_2|s_2\rangle \times \sum_{s_M} \langle s_2|G_2F_3G_3F_4\cdots G_{M-1}F_M|s_M\rangle,$$

$$p_2(s_2) = \langle s_2|F_2|s_2\rangle \sum_{s_1} \langle s_1|F_1G_1|s_2\rangle \times \sum_{s_M} \langle s_2|G_2F_3G_3F_4\cdots G_{M-1}F_M|s_M\rangle,$$

$$p_2(s_2, s_3) = \langle s_2|G_2|s_3\rangle\langle s_2|F_2|s_2\rangle\langle s_3|F_3|s_3\rangle \times \sum_{s_1} \langle s_1|F_1G_1|s_2\rangle \sum_{s_M} \langle s_2|G_3F_4G_4F_5\cdots G_{M-1}F_M|s_M\rangle,$$

$$p_3(s_3) = \langle s_3|F_3|s_3\rangle \sum_{s_1} \langle s_1|F_1G_1F_2G_2|s_3\rangle \times \sum_{s_M} \langle s_3|G_3F_4G_4F_5\cdots G_{M-1}F_M|s_M\rangle,$$

$$\vdots$$

$$p_{M-1}(s_{M-1}) = \langle s_{M-1}|F_{M-1}|s_{M-1}\rangle \times \sum_{s_1}\langle s_1|F_1G_1F_2G_2\cdots G_{M-2}|s_{M-1}\rangle \sum_{s_M} \langle s_{M-1}|G_{M-1}F_M|s_M\rangle,$$

$$p_{M-1}(s_{M-1}, s_M) = \langle s_{M-1}|G_{M-1}|s_M\rangle \times \langle s_{M-1}|F_{M-1}|s_{M-1}\rangle\langle s_M|F_M|s_M\rangle \times \sum_{s_1}\langle s_1|F_1G_1F_2G_2\cdots G_{M-2}|s_{M-1}\rangle.$$

Note that Eq. (4) implies that

$$p_{M-1}(s_{M-1}, s_M) = p_{M-1}(s_{M-1}|s_M)p_M(s_M).$$

One finds that with the above relations the product on the right-hand side of Eq. (5) telescopes into the form given in Eq. (1), and the theorem is thus proved.

An equivalent statement of Eq. (5) is

$$p(s_1, s_2, \ldots, s_M) = \prod_{i=1}^{M} p_i(s_i) \prod_{j=1}^{M-1} \left(\frac{p_j(s_j, s_{j+1})}{p_j(s_j) p_{j+1}(s_{j+1})} \right), \quad (8)$$

which implies that the entropy takes the form

$$\begin{aligned} &- \sum_{s_1, \ldots, s_M} p(s_1, s_2, \ldots, s_M) \ln p(s_1, s_2, \ldots, s_M) \\ &\quad = - \sum_i \sum_{s_i} p_i(s_i) \ln p_i(s_i) \\ &\qquad - \sum_j \sum_{s_j, s_{j+1}} p_j(s_j, s_{j+1}) \ln \left(\frac{p_j(s_j, s_{j+1})}{p_j(s_j) p_{j+1}(s_{j+1})} \right). \end{aligned} \quad (9)$$

Thus, for systems under consideration the entropy contains only "single-subsystem entropies" and "nearest neighbor," cumulantlike[4] terms even though the system itself may exhibit statistical dependence between s_l and s_m for $m \neq l + 1$.

Notice that for any Hamiltonian $\mathcal{H}(s_1, s_2, \ldots, s_M)$, which does not necessarily yield to transfer-matrix formalism, the Gibbs–Bogoliubov variational theorem[6] implies that the exact free energy

$$F \leq \sum_{s_1, \ldots, s_M} \tilde{p}(s_1, s_2, \ldots, s_M)[\mathcal{H} + k_B T \ln \tilde{p}(s_1, s_2, \ldots, s_M)], \quad (10)$$

where $\tilde{p}(s_1, s_2, \ldots, s_M)$ is any "trial" probability function; thus, for $\tilde{p}$ one may use the right-hand side of Eq. (8) and determine $\tilde{p}_i(s_i)$ and $\tilde{p}_j(s_j, s_{j+1})$ by functional variation.

The single-subsystem entropies can also be bounded as is indicated in the following example. Consider a case for which s_1 denotes a random variable having 2^N realization corresponding to the states of a chain of N spin-$\frac{1}{2}$ particles each with an associated random variable $\sigma_i = \pm 1$. The entropy contribution of the chain[7] is

$$- \sum_{s_1} p(s_1) \ln p(s_1) \leq - \sum_{i=1}^{N} \sum_{\sigma_i} p_i(\sigma_i) \ln p_i(\sigma_i) \quad (11)$$

but the right-hand side[4]

$$\begin{aligned} &= N \ln 2 - \sum_i [\tfrac{1}{2}(1 + \langle \sigma_i \rangle) \ln(1 + \langle \sigma_i \rangle) \\ &\qquad + \tfrac{1}{2}(1 - \langle \sigma_i \rangle) \ln(1 - \langle \sigma_i \rangle)], \end{aligned} \quad (12)$$

which is a monotone, nonincreasing, convex function of $\langle \sigma_i \rangle$.

---

[1] See, e.g., the lecture notes of E. H. Lieb in *Lectures in Theoretical Physics*, edited by K. T. Mahanthappa and W. E. Brittin (Gordon and Breach, New York, 1969), Vol. XI-D, p. 329.
[2] M. S. Bartlett, *An Introduction to Stochastic Processes* (Cambridge U. P., London, 1962), p. 34.
[3] H. Araki and E. H. Lieb, Commun. Math. Phys. **18**, 160 (1970).
[4] H. Falk and M. Suzuki, Phys. Rev. B **1**, 3051 (1970); B **3**, 228 (1971); H. Falk J. Math. Phys. **12**, 1492 (1971).
[5] C. J. Thompson, "Upper Bounds for Ising Model Correlation Functions," Commun. Math. Phys. (to be published).
[6] See, e.g., H. Falk, Amer. J. Phys. **38**, 858 (1970). The work of George Woodbury, Jr. [J. Chem. Phys. **47**, 270 (1967)] contains a reformulation of various lattice free energy approximations by entropy decomposition methods. I am indebted to Professor Ben Widom for this very relevant reference.
[7] This and other entropy bounds will be found in Refs. 3 and 6.

Nonrelativistic Time-Dependent Scattering Theory and von Neumann Algebras. I. Single Channel Scattering

W. W. Zachary
Naval Research Laboratory, Washington, D.C. 20390
(Received 11 May 1971; Revised Manuscript Received 29 November 1971)

For single-channel nonrelativistic time-dependent scattering we analyze the von Neumann algebras $\mathfrak{M}_\pm$ generated by the wave operators $\Omega_\pm$ and the algebra $\mathfrak{G}$ generated by the spectral projections of both the total and free Hamiltonians. Each of the algebras $\mathfrak{M}_\pm$ is the direct sum of an Abelian algebra and a factor of type I_∞, but both summands need not occur. For a large class of central potentials $\mathfrak{G}$ is the direct sum of a countable number of factors of type I_∞ and its commutant the direct sum of a countable number of finite factors of type I. In the latter case the type of each direct summand is uniquely determined by the angular momentum of the corresponding partial wave. Similar decompositions are obtained for algebras related to $\mathfrak{G}$ and $\mathfrak{G}'$ in a natural way that have certain connections with the absolutely continuous and point spectra of the total Hamiltonian. The structure of some algebras related to $\mathfrak{M}_\pm$ is connected with certain properties of the bound states.

I. INTRODUCTION

A precise formulation of time-dependent single-channel scattering theory was given by Jauch[1] in which necessary conditions were stated for the existence of "simple scattering systems." The central constructs in this theory are the operators defined by the following strong limits on the Hilbert space $L^2(R^3)$:

$$\Omega_\pm^{(M)} = \text{s-lim}_{t \to \mp\infty} e^{itH} e^{-itH_0}, \quad (1)$$

where the self-adjoint operators H and H_0 denote, respectively, the total and free Hamiltonians.

The existence of the limits (1) has been established for short range potentials[2] in which case the corresponding operators are precisely the Møller wave operators. For long range potentials such as the Coulomb potential, the limits in (1) do not exist. In this case it has been shown[3,4] that new operators can be found which have the same physical interpretation as the Møller operators (1). Similar results have been obtained for other long range potentials.[4–8]

Following this idea, several authors[7–10] have defined new asymptotic conditions for the single channel case which replace the Møller operators by operators which exist under more general conditions, i.e., for a wider class of potentials.

One finds that with the above relations the product on the right-hand side of Eq. (5) telescopes into the form given in Eq. (1), and the theorem is thus proved.

An equivalent statement of Eq. (5) is

$$p(s_1, s_2, \ldots, s_M) = \prod_{i=1}^{M} p_i(s_i) \prod_{j=1}^{M-1} \left(\frac{p_j(s_j, s_{j+1})}{p_j(s_j) p_{j+1}(s_{j+1})} \right), \quad (8)$$

which implies that the entropy takes the form

$$\begin{aligned} &- \sum_{s_1, \ldots, s_M} p(s_1, s_2, \ldots, s_M) \ln p(s_1, s_2, \ldots, s_M) \\ &\quad = - \sum_i \sum_{s_i} p_i(s_i) \ln p_i(s_i) \\ &\qquad - \sum_j \sum_{s_j, s_{j+1}} p_j(s_j, s_{j+1}) \ln \left(\frac{p_j(s_j, s_{j+1})}{p_j(s_j) p_{j+1}(s_{j+1})} \right). \end{aligned} \quad (9)$$

Thus, for systems under consideration the entropy contains only "single-subsystem entropies" and "nearest neighbor," cumulantlike[4] terms even though the system itself may exhibit statistical dependence between s_l and s_m for $m \neq l + 1$.

Notice that for any Hamiltonian $\mathcal{H}(s_1, s_2, \ldots, s_M)$, which does not necessarily yield to transfer-matrix formalism, the Gibbs–Bogoliubov variational theorem[6] implies that the exact free energy

$$F \leq \sum_{s_1, \ldots, s_M} \tilde{p}(s_1, s_2, \ldots, s_M)[\mathcal{H} + k_B T \ln \tilde{p}(s_1, s_2, \ldots, s_M)], \quad (10)$$

where $\tilde{p}(s_1, s_2, \ldots, s_M)$ is any "trial" probability function; thus, for $\tilde{p}$ one may use the right-hand side of Eq. (8) and determine $\tilde{p}_i(s_i)$ and $\tilde{p}_j(s_j, s_{j+1})$ by functional variation.

The single-subsystem entropies can also be bounded as is indicated in the following example. Consider a case for which s_1 denotes a random variable having 2^N realization corresponding to the states of a chain of N spin-$\frac{1}{2}$ particles each with an associated random variable $\sigma_i = \pm 1$. The entropy contribution of the chain[7] is

$$- \sum_{s_1} p(s_1) \ln p(s_1) \leq - \sum_{i=1}^{N} \sum_{\sigma_i} p_i(\sigma_i) \ln p_i(\sigma_i) \quad (11)$$

but the right-hand side[4]

$$\begin{aligned} &= N \ln 2 - \sum_i [\tfrac{1}{2}(1 + \langle \sigma_i \rangle) \ln(1 + \langle \sigma_i \rangle) \\ &\qquad + \tfrac{1}{2}(1 - \langle \sigma_i \rangle) \ln(1 - \langle \sigma_i \rangle)], \end{aligned} \quad (12)$$

which is a monotone, nonincreasing, convex function of $\langle \sigma_i \rangle$.

---

[1] See, e.g., the lecture notes of E. H. Lieb in *Lectures in Theoretical Physics*, edited by K. T. Mahanthappa and W. E. Brittin (Gordon and Breach, New York, 1969), Vol. XI-D, p. 329.
[2] M. S. Bartlett, *An Introduction to Stochastic Processes* (Cambridge U. P., London, 1962), p. 34.
[3] H. Araki and E. H. Lieb, Commun. Math. Phys. **18**, 160 (1970).
[4] H. Falk and M. Suzuki, Phys. Rev. B **1**, 3051 (1970); B **3**, 228 (1971); H. Falk J. Math. Phys. **12**, 1492 (1971).
[5] C. J. Thompson, "Upper Bounds for Ising Model Correlation Functions," Commun. Math. Phys. (to be published).
[6] See, e.g., H. Falk, Amer. J. Phys. **38**, 858 (1970). The work of George Woodbury, Jr. [J. Chem. Phys. **47**, 270 (1967)] contains a reformulation of various lattice free energy approximations by entropy decomposition methods. I am indebted to Professor Ben Widom for this very relevant reference.
[7] This and other entropy bounds will be found in Refs. 3 and 6.

Nonrelativistic Time-Dependent Scattering Theory and von Neumann Algebras. I. Single Channel Scattering

W. W. Zachary
Naval Research Laboratory, Washington, D.C. 20390
(Received 11 May 1971; Revised Manuscript Received 29 November 1971)

For single-channel nonrelativistic time-dependent scattering we analyze the von Neumann algebras $\mathfrak{M}_\pm$ generated by the wave operators $\Omega_\pm$ and the algebra $\mathfrak{G}$ generated by the spectral projections of both the total and free Hamiltonians. Each of the algebras $\mathfrak{M}_\pm$ is the direct sum of an Abelian algebra and a factor of type I_∞, but both summands need not occur. For a large class of central potentials $\mathfrak{G}$ is the direct sum of a countable number of factors of type I_∞ and its commutant the direct sum of a countable number of finite factors of type I. In the latter case the type of each direct summand is uniquely determined by the angular momentum of the corresponding partial wave. Similar decompositions are obtained for algebras related to $\mathfrak{G}$ and $\mathfrak{G}'$ in a natural way that have certain connections with the absolutely continuous and point spectra of the total Hamiltonian. The structure of some algebras related to $\mathfrak{M}_\pm$ is connected with certain properties of the bound states.

I. INTRODUCTION

A precise formulation of time-dependent single-channel scattering theory was given by Jauch[1] in which necessary conditions were stated for the existence of "simple scattering systems." The central constructs in this theory are the operators defined by the following strong limits on the Hilbert space $L^2(R^3)$:

$$\Omega_\pm^{(M)} = \text{s-lim}_{t \to \mp\infty} e^{itH} e^{-itH_0}, \quad (1)$$

where the self-adjoint operators H and H_0 denote, respectively, the total and free Hamiltonians.

The existence of the limits (1) has been established for short range potentials[2] in which case the corresponding operators are precisely the Møller wave operators. For long range potentials such as the Coulomb potential, the limits in (1) do not exist. In this case it has been shown[3,4] that new operators can be found which have the same physical interpretation as the Møller operators (1). Similar results have been obtained for other long range potentials.[4–8]

Following this idea, several authors[7–10] have defined new asymptotic conditions for the single channel case which replace the Møller operators by operators which exist under more general conditions, i.e., for a wider class of potentials.

In discussing these new asymptotic conditions certain von Neumann (W^*) and C^*-algebras have made their appearance. In the case of von Neumann algebras[11] among those occurring are the algebras $\mathfrak{A}$ and $\mathfrak{C}$ which are defined, respectively, as the algebras generated by the spectral projections of H and of H_0, the total and free Hamiltonians. The algebra $\mathfrak{C}'$, the commutant of $\mathfrak{C}$, has proved especially useful in the study of the limits for infinite time of certain observables.[8]

These algebras are not adequate, however, for the discussion of the particle interactions which occur at finite times. For this purpose one wants algebras that contain the wave operators. Neither $\mathfrak{A}$ nor $\mathfrak{C}$ have this property except in trivial cases. Our objective in the present paper is to discuss certain von Neumann algebras which contain the wave operators. These are the algebras $\mathfrak{M}_\pm$, generated by $\Omega_\pm$,

$$\mathfrak{M}_\pm = \{\Omega_\pm\}'', \tag{2}$$

which are the smallest von Neumann algebras, respectively, that contain $\Omega_\pm$ and $\Omega_\pm^*$; and the algebra $\mathfrak{B}$ generated by the spectral projections of both H and H_0, i.e., the smallest von Neumann algebra containing both Abelian algebras $\mathfrak{A}, \mathfrak{C}$,

$$\mathfrak{B} = \{\mathfrak{A}, \mathfrak{C}\}''. \tag{3}$$

We will be concerned specifically with the problem of classifying $\mathfrak{M}_\pm$ and $\mathfrak{B}$ and certain related algebras according to type and other properties, and will confine ourselves to the case of single channel scattering. Multichannel scattering will be discussed in a future paper. Our program for time-dependent scattering theory is analogous to certain developments in the algebraic formulations of statistical mechanics and quantum field theory,[12] namely, the analysis of certain algebraic structures which are thought to characterize the theory in some fashion.

In Sec. II we formulate the theory of single-channel time-dependent scattering and also give a brief summary of some required results from the theory of von Neumann algebras. Then in Sec. III we discuss some results concerning the algebras $\mathfrak{B}$ and $\mathfrak{B}'$ which, for rotationally invariant scattering systems, determine their direct sum decompositions into factors of known types. Section IV is devoted to the discussion of certain algebras related to $\mathfrak{M}_+$ and $\mathfrak{B}$. The properties of these algebras are related to certain characteristics of the bound states.

For purposes of convenience in the presentation both signs in $\Omega_\pm$ and $\mathfrak{M}_\pm$ are discussed simultaneously.

II. FORMULATION

The most important case of single channel scattering occurs in the problem of the interaction of two nonrelativistic spinless particles. Consider the Hilbert space $\mathfrak{K} = L^2(R^3)$, then the Hamiltonian operator, after separation of the center of mass, is formally given by

$$H = H_0 + V, \tag{4}$$

where, in suitable units, $H_0 = -\Delta$ and V is an operator corresponding to multiplication by a real-valued local potential. The essential self-adjointness of H was established by Kato[13] for potentials $V(\mathbf{x})$ which are locally square integrable and bounded at infinity,

$$V(x) \in L^2(R^3) + L^\infty(R^3). \tag{5}$$

Hereafter, except for the discussion of Proposition 1 in Sec. III, we will follow the usual practice of replacing H and H_0 by their unique self-adjoint extensions without changing the notation. Simon[6] has discussed self-adjointness proofs under weaker conditions. In this case the definition of the Hamiltonian by an operator sum as in (4) is given up. Simon defines it instead in terms of quadratic forms.

We will be concerned with the isometric operators $\Omega_\pm$, which satisfy

$$\Omega_\pm^* \Omega_\pm = I, \quad \Omega_\pm \Omega_\pm^* = P_\pm, \tag{6}$$

where $P_\pm$ denote the projections onto the respective ranges $R_\pm$ of $\Omega_\pm$, which are subspaces (closed linear manifolds) of $L^2(R^3)$. For short range potentials $\Omega_\pm$ are the Møller wave operators defined by (1). This is true for real-valued $V(\mathbf{x})$ which are

(i) locally square integrable and

(ii) $O(|\mathbf{x}|^{-\beta})$ for some $\beta > 1$ as $|\mathbf{x}| \to \infty$, (7)

i.e., for locally square integrable potentials that decrease at infinity faster than the Coulomb potential.[2]

For long range potentials the limits (1) do not exist but, for certain classes of such potentials, new "renormalized" operators can be defined which have the properties (6).[7-9]

We denote by P the projection onto the subspace $\mathfrak{K}_{ac}$, the absolutely continuous subspace of $\mathfrak{K} = L^2(R^3)$ with respect to H. Hereafter, we will consider the symbols $\Omega_\pm$ to denote the "renormalized" wave operators[7-9] as well as the Møller operators (1) if the context allows the existence of the latter. The equality $P_+ = P_-$ is called the unitarity condition and is equivalent to the unitarity of the S operator,[1] $S = \Omega_-^* \Omega_+$. An important condition which one wants the scattering system to satisfy is that of asymptotic completeness. This requires the equalities $P_+ = P_- = P$ as well as the absence of the singular continuous subspace of $\mathfrak{K}$ relative to H, and eliminates certain pathological states which have the properties of neither scattering nor bound states. Sufficient conditions on the potential for both unitarity and asymptotic completeness have been given.[2,6,8,14]

We now want to summarize a few standard results of the theory of von Neumann algebras[11] in the context of two-body scattering theory. The algebras $\mathfrak{A}$ and $\mathfrak{C}$ have already been defined in the introduction. We now further note that $\mathfrak{A}$(resp. $\mathfrak{C}$) consists of all essentially bounded measurable functions of H (resp. H_0) in the sense of the functional calculus.[15] The algebra $\mathfrak{B}$ is defined by (3), or equivalently (Ref. 11, p. 2), its commutant is given by

$$\mathfrak{B}' = \mathfrak{A}' \cap \mathfrak{C}'. \tag{8}$$

Given a von Neumann algebra $\mathfrak{N}$ on a Hilbert space $\mathfrak{K}$ one can form other von Neumann algebras by the process now to be described (Ref. 11, Section 1.2). For a nonzero projection $G \in \mathfrak{N}$ denote by $\mathfrak{N}_G$ the

restriction of $G\mathfrak{N}G$ to $G\mathcal{K}$. Then, $\mathfrak{N}_G$ and $(\mathfrak{N}')_G = (\mathfrak{N}_G)' \equiv \mathfrak{N}'_G$ are von Neumann algebras. Let $\mathfrak{z}$ denote the center of $\mathfrak{N}$ (i.e., $\mathfrak{z} = \mathfrak{N} \cap \mathfrak{N}'$) then the center of $\mathfrak{N}_G$ is $\mathfrak{z}_G$. In particular, if $\mathfrak{N}$ is a factor ($\mathfrak{z} = \{\lambda I\}$), then so is $\mathfrak{N}_G$. For a projection $G \in \mathfrak{N}$ the central carrier (also central cover or central support) of G is defined as the minimal projection $G' \in \mathfrak{z}$ such that $G \leq G'$. We will frequently use the result that $G \in \mathfrak{N}$ implies that $\mathfrak{N}'_G$ and $\mathfrak{N}'_{G'}$ are *-isomorphic (Ref. 11, p. 18, Proposition 2). In the following two sections we will make use of the classification into types of von Neumann algebras (Ref. 11, Section 1. 8), especially the classification of factors. There are three types of factors, called I, II, and III, of which only type I will occur in this paper. Type I factors are *-isomorphic to $\mathfrak{L}(\mathcal{K})$, the algebra of all bounded operators on some Hilbert space $\mathcal{K}$. If $\mathcal{K}$ is n dimensional, $n < \infty$, then one speaks of type I_n factors, and if it is infinite dimensional (separable) of type I_∞ factors.

We will also make use of the concept of equivalence classes of projections due to Murray and von Neumann (Ref. 11, pp. 215 ff). Two projections E and F belonging to a von Neumann algebra $\mathfrak{N}$ are equivalent relative to $\mathfrak{N}$, $E \sim F$, if there exists a partially isometric operator $U \in \mathfrak{N}$ such that $U^* U = E$ and $UU^* = F$. A projection is infinite if it is equivalent to a proper subprojection of itself; otherwise it is finite. A von Neumann algebra $\mathfrak{N}$ is finite (infinite) if the identity operator is finite (infinite) relative to $\mathfrak{N}$.

It follows from (1) or the corresponding formulation of the asymptotic condition for long range interactions[7-9] and the von Neumann density theorem (Ref. 11, p. 41, Théorème 2) that the wave operators belong to $\mathfrak{B}$. Then, from the definition of $\mathfrak{M}_\pm$ by (2), one easily shows that $\mathfrak{M}_\pm \subset \mathfrak{B}$ or equivalently, $\mathfrak{B}' \subset \mathfrak{M}'_\pm$. From this result and (6) we see that $P_\pm \sim I$ relative to $\mathfrak{M}_\pm$ and $\mathfrak{B}$, the equivalences being implemented by $\Omega_\pm$. It follows that $P_+ \sim P_-$ and that the only projection in the center of $\mathfrak{B}$ (or centers of $\mathfrak{M}_\pm$) that majorizes these projections is the identity. These equivalences also imply that the following pairs of algebras are spatially isomorphic (Ref. 11, p. 9) by means of $\Omega_\pm$: $(\mathfrak{M}_\pm, (\mathfrak{M}_\pm)_{P_\pm})$, $(\mathfrak{B}, \mathfrak{B}_{P_\pm})$, $(\mathfrak{M}'_\pm, (\mathfrak{M}'_\pm)_{P_\pm})$, and $(\mathfrak{B}', \mathfrak{B}'_{P_\pm})$. In view of these spatial isomorphisms we will not discuss the algebras $(\mathfrak{M}_\pm)_{P_\pm}$ and $\mathfrak{B}_{P_\pm}$ or their commutants in Sec. III and IV.

For the short range potentials (7) and for many long range potentials one has the majorizations

$$P_\pm \leq P. \tag{9}$$

In this case it can easily be shown from either of the equivalences $P_\pm \sim I$ that $P \sim I$ relative to $\mathfrak{B}$. In this case the algebras $\mathfrak{B}_{P_\pm}$, $\mathfrak{B}_P$, and $\mathfrak{B}$ are all spatially isomorphic as are also the algebras $\mathfrak{B}'_{P_\pm}$, $\mathfrak{B}'_P$, and $\mathfrak{B}'$. In general we do not know that $P \in \mathfrak{M}_\pm$ unless, of course, $P = P_+$ and/or $P = P_-$. This statement is related to the fact that, in general, $\mathfrak{M}_\pm \neq \mathfrak{B}$, which will be discussed in Sec. IV.

According to (2) and (6) the algebras $\mathfrak{M}_\pm$ are generated by the isometries $\Omega_\pm$. It follows from results of Suzuki[16] that $\mathfrak{M}_\pm$ have the following direct sum decompositions:

$$\mathfrak{M}_\pm = (\mathfrak{M}_\pm)_{G_\pm} \oplus (\mathfrak{M}_\pm)_{G_\pm^\perp}, \tag{10}$$

where $G_\pm$ and $G_\pm^\perp$ are central projections in $\mathfrak{M}_\pm$, $(\mathfrak{M}_\pm)_{G_\pm}$ are Abelian algebras and $(\mathfrak{M}_\pm)_{G_\pm^\perp}$ are factors of type I_∞. Here we have denoted by $G_\pm^\perp \equiv I - G_\pm$ the respective orthocomplements of $G_\pm$. Hereafter we will use this notation also for arbitrary projections. The subspaces onto which $G_\pm$ and $G_\pm^\perp$ project, are defined by[16]

$$\begin{aligned} G_\pm \mathcal{K} &= \bigcap_{n=0}^{\infty} (\Omega_\pm)^n \mathcal{K}, \\ G_\pm^\perp \mathcal{K} &= \sum_{n=0}^{\infty} \oplus (\Omega_\pm)^n R_\pm, \end{aligned} \tag{11}$$

where $R_\pm$ denote the orthogonal complements of the respective ranges of $\Omega_\pm$. We will refer to (10) as the Suzuki decompositions of $\mathfrak{M}_\pm$. It should be emphasized that both summands need not be present in (10). For example, the factor summands are absent if the wave operators $\Omega_\pm$ are unitary. Lavine[14] gives examples of nontrivial repulsive potentials for which $\Omega_\pm^{(M)}$ exist and are unitary.

III. SOME RESULTS CONCERNING $\mathfrak{B}$

We begin with a result which gives a slightly different characterization of $\mathfrak{B}$ than the original definition (3), but which one intuitively expects. In the proof we must distinguish between the formal operators H_0 and H in (4) and their (unique) self-adjoint extensions, which we denote respectively by $\tilde{H}_0$ and $\tilde{H}$.

Proposition 1: Let V be a potential operator in (4) corresponding to a local potential (function) satisfying (5). Then V is self-adioint on the dense domain [in $L^2(R^3)$] $D(V) \supseteq D(\tilde{H}_0)$ and $\mathfrak{B}$ is generated by the spectral projections of $\tilde{H}_0$ and V.

Proof: For the assumed potentials Kato[13] proved the essential self-adjointness of H and, further, that V is self-adjoint with domain $D(V)$ dense in $L^2(R^3)$ and that the respective unique self-adjoint extensions $\tilde{H}, \tilde{H}_0$ of H and H_0 satisfy

$$\tilde{H} = \tilde{H}_0 + V$$

on the dense domain $D(H_0)$ of $L^2(R^3)$ with $D(\tilde{H}) = D(\tilde{H}_0) \subseteq D(V)$. In the present notation (3) states that $\mathfrak{B}$ is generated by the spectral projections of $\tilde{H}_0$ and $\tilde{H}$. Let $\mathfrak{N}$ denote the von Neumann algebra generated by the spectral projections of $\tilde{H}_0$ and V. Then one shows by standard arguments that $\mathfrak{N} = \mathfrak{B}$.

This theorem holds even for constant potentials in which case $D(V)$ is all of $L^2(R^3)$. In these cases, however, one obtains either from (3) or Proposition 1 that $\mathfrak{B}$ coincides with the Abelian algebras $\mathfrak{A}$ and $\mathfrak{C}$. For these potentials the Møller operators (1) do not exist but the renormalized wave operators do and satisfy $\Omega_\pm = I$[8] so that no scattering occurs. For this reason we will exclude these potentials from further consideration.

We now return to the notation used prior to the statement of Proposition 1, wherein H and H_0 denote the unique self-adjoint extensions of the operators in (4) ($\tilde{H}$ and $\tilde{H}_0$ in Proposition 1). Also, for most of the remainder of the paper we will restrict ourselves to the case of rotationally invariant scattering systems so that the potential is central (spherically symmetric),[17] $V(\mathbf{x}) = V(|\mathbf{x}|) = V(r)$. In this case proofs of the existence and unitarity of the Møller operators

have been given for potentials that may be more singular as $r \to 0$ than allowed by the conditions (7).[18] These singularities are of the same type as those considered by Simon[6] for potentials that are not necessarily spherically symmetric.

For central potentials it is well known that the Hilbert space $L^2(R^3)$ can be decomposed in the form

$$\sum_{l=0}^{\infty} \sum_{m=-l}^{l} \oplus \mathcal{K}_{lm}, \tag{12}$$

where each $\mathcal{K}_{lm}$ reduces both H and H_0 and consists of elements of the form

$$r^{-1}u(r)Y_l^m(\theta, \varphi), \tag{13}$$

in which Y_l^m denote the spherical harmonics and $u \in L^2(0, \infty)$. For $l = 0$ one imposes the boundary condition $u(0) = 0$. Let Q_{lm} denote the respective projections onto the subspaces $\mathcal{K}_{lm}$. Then the restrictions of $HQ_{lm} = Q_{lm}H$ and $H_0Q_{lm} = Q_{lm}H_0$ to $\mathcal{K}_{lm}$ denoted hereafter by H_{lm} and $(H_0)_{lm}$, respectively, are unitarily equivalent to the respective ordinary differential operators in $L^2(0, \infty)$:

$$\frac{-d^2}{dr^2} + l(l+1)r^{-2} + V(r) \quad \text{and} \quad \frac{-d^2}{dr^2} + l(l+1)r^{-2}. \tag{14}$$

For the results to follow, we will want to assume that the differential operators (14) have simple spectrum on $(0, \infty)$. For continuous central potentials satisfying (7) this is known to be the case.[19] It also holds for the following class of central potentials introduced by Weidmann.[20] Assume that there exist real numbers $C > 0$, $R > 0$, and $s < \frac{3}{2}$ such that

$$r^s|V(r)| \leq C \quad \text{for } r < R, \tag{15a}$$

$$|V(r)| \leq C \quad \text{for } r \geq R, \tag{15b}$$

and for $r \geq R$,

$$V(r) = V_1(r) + V_2(r),$$

where $V_1(r)$ is of bounded variation and

$$\lim_{r\to\infty} V_1(r) = 0, \qquad V_2(r) \in L^1(R, \infty).$$

We note that for central potentials $V(r)$ satisfying (7), (15a) holds as $r \to 0$ and $V(r) \in L^1(R, \infty)$ for all positive R. Furthermore, Weidmann's class admits certain long range central potentials.

Now we have

Lemma 1: Consider central potentials $V(r) \neq 0$ satisfying (5) such that the operators (14) have simple spectrum and the corresponding wave operators $\Omega_\pm$ exist and satisfy the intertwining relations,

$$e^{iHt}\Omega_\pm = \Omega_\pm e^{iH_0t} \tag{16}$$

for all real finite values of t. Then for each l and m, $l = 0, 1, 2, \cdots$, $|m| \leq l$, $\mathfrak{B}_{Q_{lm}}$ and $\mathfrak{B}'_{Q_{lm}}$ are factors of respective types I_∞ and I_1.

For the short range potentials (7) the intertwining relations (16) are automatically satisfied.[1,2] In the case of long range potentials conditions for the validity of (16) have been given.[7-9] These conditions exclude nonzero constant potentials and admit certain classes of nontrivial long range potentials.[5,7-9]

Proof: Because $V(r)$ satisfies (5), the Hamiltonian is self-adjoint, as we noted in Proposition 1. Since each $\mathcal{K}_{lm}$ reduces both H and H_0 we have, from (8),

$$Q_{lm} \in \mathfrak{B}'. \tag{17}$$

The operators H_{lm} and $(H_0)_{lm}$ generate the respective von Neumann algebras $\mathfrak{A}_{Q_{lm}}$ and $\mathfrak{C}_{Q_{lm}}$ (Ref. 11, p. 16, Proposition 1) and are unitarily equivalent to the respective operators (14) and so, using the hypothesis, have simple spectrum. It follows that $\mathfrak{A}_{Q_{lm}}$ and $\mathfrak{C}_{Q_{lm}}$ are maximal Abelian.[15] From this one easily shows that the algebras $\mathfrak{B}'_{Q_{lm}}$ are Abelian and that

$$\mathfrak{B}'_{Q_{lm}} \subset \mathfrak{A}_{Q_{lm}} \cap \mathfrak{C}_{Q_{lm}}. \tag{18}$$

Now let A be an arbitrary operator in $\mathfrak{B}'_{Q_{lm}}$. Then, from (18),

$$A = f(H_{lm}) = g((H_0)_{lm}) \tag{19}$$

for suitable essentially bounded measurable functions f and g in the sense of the functional calculus.[15] For a given set of values of the pair (l, m) suppose that the spectral projections of H_{lm} and $(H_0)_{lm}$ commute. Then, since these operators generate the respective algebras $\mathfrak{A}_{Q_{lm}}$ and $\mathfrak{C}_{Q_{lm}}$ it follows that these algebras commute and therefore are equal because they are each maximal Abelian. Consequently, $\mathfrak{B}_{Q_{lm}}$ is Abelian. From (14) one finds that if the above commutation property holds for a given value of the pair (l, m), then it holds for all values of l and m, so that $\mathfrak{B}$ is Abelian. We now use (6), (16), and the fact that all operators appearing in (16) belong to $\mathfrak{B}$ to find $e^{iHt} = e^{iH_0t}$ for all real finite t, which is a contradiction because $V(r) \neq 0$. Consequently, the spectral projections of H_{lm} and $(H_0)_{lm}$ do not commute for any values of the set (l, m) and (19) is impossible unless A is a scalar operator. Then, since A is arbitrary in $\mathfrak{B}'_{Q_{lm}}$, we have that $\mathfrak{B}'_{Q_{lm}}$ are factors of type I_1. It follows that $\mathfrak{B}_{Q_{lm}}$ are irreducible type I factors and, since they are each $*$-isomorphic to an algebra of operators in $L^2(0, \infty)$, we conclude that they are of type I_∞.

These results enable us to decompose $\mathfrak{B}$ and $\mathfrak{B}'$ as orthogonal sums of factors. However, it has not been shown that $Q_{lm} \in \mathfrak{B}$, i.e., that the Q_{lm} are central projections. We now generalize Lemma 1 by showing that certain combinations of the Q_{lm} which correspond to the angular momentum multiplets, are central projections of $\mathfrak{B}$.

Theorem 1: Consider the same potentials as in Lemma 1 and write $Q_l = \sum_{m=-l}^{l} Q_{lm}$, $l = 0, 1, 2, \cdots$. Then $Q_l \in \mathfrak{B} \cap \mathfrak{B}'$ and $\mathfrak{B}_{Q_l}$ and $\mathfrak{B}'_{Q_l}$ are factors of respective types I_∞ and I_{2l+1}.

Proof: From (17) we have $Q_l \in \mathfrak{B}'$ for all $l = 0, 1, 2, \cdots$. It was proved in Lemma 1 that $\mathfrak{B}_{Q_0}$ and $\mathfrak{B}'_{Q_0}$ are factors of respective types I_∞ and I_1, so that we restrict our attention to the cases $l \neq 0$. The

algebras $\mathfrak{B}_{Q_l}$ and $\mathfrak{B}'_{Q_l}$ operate on the Hilbert space $\mathfrak{K}_l = \sum_{m=-l}^{l} \oplus \mathfrak{K}_{lm}$ and $\mathfrak{K}_{lm}$ consists of elements of the form (13). Now, the generators H_{lm} and $(H_0)_{lm}$ of each algebra $\mathfrak{B}_{Q_{lm}}$ are unitarily equivalent to the operators (14), which are seen to be independent of m. It follows that, for a fixed value of $l \neq 0$, each algebra $\mathfrak{B}_{Q_{lm}}$ is $*$-isomorphic to the same algebra which, from Lemma 1, can be identified as $\mathfrak{L}(L^2(0,\infty))$, the algebra of all bounded operators on $L^2(0,\infty)$. It follows that $\mathfrak{B}_{Q_l}$ can be represented as a $(2l+1) \times (2l+1)$ diagonal matrix algebra over $\mathfrak{L}(L^2(0,\infty))$ with equal diagonal elements. It then follows that $\mathfrak{B}_{Q_l}$ is $*$-isomorphic to[21] $\mathfrak{L}(L^2(0,\infty))$ and hence is a factor of type I_∞.

Consider the algebra $\mathfrak{B}'_{Q_l}$ for a fixed value $l \neq 0$. By Lemma 1 each Q_{lm} is Abelian relative to $\mathfrak{B}'$ (Ref. 11, p. 123, Définition 3) and hence also relative to $\mathfrak{B}'_{Q_l}$ (Ref. 11, p. 18). From the first part of the proof $\mathfrak{B}_{Q_l}$ is a factor of type I and, therefore, so is $\mathfrak{B}'_{Q_l}$. Consequently, each Q_{lm} is minimal relative to $\mathfrak{B}'_{Q_l}$ (Ref. 11. p. 123), and it follows that $\mathfrak{B}'_{Q_l}$ is of type I_{2l+1} (Ref. 11, p. 126).

The proof of the assertions that $Q_l \in \mathfrak{B}$ for $l = 1, 2, \cdots$ follows from the structure of these projections as diagonal matrix operators with equal diagonal elements. It then follows that also $Q_0 \in \mathfrak{B}$, which completes the proof of the theorem.

This theorem shows that, subject to the stated hypotheses, $\mathfrak{B}$ and $\mathfrak{B}'$ are not factors. We also see that $\mathfrak{B}$ is decomposed as a direct sum of a countable number of I_∞ factors and that $\mathfrak{B}'$ is decomposed as the direct sum of a I_1 factor, a I_3 factor, etc. Incidentally, it can be shown from this that $\mathfrak{B}$ and $\mathfrak{B}'$ are of type I.

A von Neumann algebra is said to be properly infinite if it contains no nonzero finite central projections. It is seen from Theorem 1 that this is not true for $\mathfrak{B}'$ since each Q_l, $l = 0, 1, 2, \cdots$, is central and finite relative to $\mathfrak{B}'$. However, we have the following.

Corollary: Consider the same potentials as in Lemma 1. Then $\mathfrak{B}$ is properly infinite.

Proof: Consider an arbitrary nonzero projection $E \in \mathfrak{B} \cap \mathfrak{B}'$ and suppose that E is finite relative to $\mathfrak{B}$. By Theorem 1, $\mathfrak{B}_{Q_l}$ is a factor and $Q_l \in \mathfrak{B} \cap \mathfrak{B}'$ so that each Q_l, $l = 0, 1, 2, \cdots$, is minimal among the projections of $\mathfrak{B} \cap \mathfrak{B}'$ (Ref. 11, p. 123, Définition 2). Therefore, for each l we have either $EQ_l = 0$ or $EQ_l = Q_l$. However, since $E \neq 0$, E cannot be orthogonal to Q_l for all values of l. Hence, there exists at least one value of l, say l_0, such that $Q_{l_0} \leq E$. Since E is finite by hypothesis it follows (Ref. 11, p. 230, Proposition 2) that Q_{l_0} is also finite. This contradicts Theorem 1 so that E cannot be finite. Since E is an arbitrary nonzero central projection, we conclude that $\mathfrak{B}$ is properly infinite.

Properly infinite algebras of type I have some interesting properties. We will not go into detail concerning these, but only note one such result. Namely, for the potentials under consideration, the above result implies that $\mathfrak{B}$ is isomorphic to $M_n(\mathfrak{B})$, $1 \leq n \leq \infty$, the $n \times n$ matrix algebra with entries from $\mathfrak{B}$.[22]

IV. RELATIONS BETWEEN EXISTENCE OF BOUND STATES AND PROPERTIES OF $\mathfrak{M}_\pm$ and $\mathfrak{B}$

We now want to investigate the algebras $\mathfrak{M}_\pm$, defined by (2), which are generated by the wave operators $\Omega_\pm$. $\mathfrak{M}_\pm$ have the Suzuki decompositions (10) so that they are either Abelian or infinite. If $\Omega_\pm$ are unitary, then $\mathfrak{M}_\pm$ are Abelian, as we noted in Sec. II. Since, from Theorem 1, each $Q_l \in \mathfrak{B} \cap \mathfrak{B}'$ one easily shows in this case that $(\Omega_\pm)_{Q_l}$ are isometries which generate $(\mathfrak{M}_\pm)_{Q_l}$ so that these algebras also have Suzuki decompositions analogous to the expressions (10) for $\mathfrak{M}_\pm$. Hence, $(\mathfrak{M}_\pm)_{Q_l}$ are also either Abelian or infinite.

One easily shows that $\mathfrak{M}_\pm$ are Abelian if and only if all $(\mathfrak{M}_\pm)_{Q_l}$ are Abelian. However, the situation can occur in which $\mathfrak{M}_\pm$ are not Abelian but some of the $(\mathfrak{M}_\pm)_{Q_l}$ are Abelian. For those cases in which bound states exist we will show below that, with certain restrictions on the potential, the number of Abelian $(\mathfrak{M}_\pm)_{Q_l}$ is directly related to certain properties of the bound states.

In deriving the results to follow we will consider the situation in which bound states exist and have finite multiplicity. The fact that the negative eigenvalues of H are isolated with finite multiplicity has been proved under very weak conditions. Thus, it holds when the potential satisfies (5) and the L^∞ part approaches zero as $r = |\mathbf{x}| \to \infty$.[23] Simon[6] has generalized this result to potentials that are not necessarily locally square integrable. The existence of positive eigenvalues can be excluded under fairly mild conditions.[24] Thus, the above assumptions do not place essential restrictions on the potential. The most restrictive of our assumptions in obtaining the following results will be that of asymptotic completeness.

We first state a simple preliminary lemma.

Lemma 2: Consider the same potentials as in Lemma 1. Let E be an arbitrary nonzero projection of $\mathfrak{B}$ or $\mathfrak{B}'$ and E' its central carrier. Then for a given $l = 0, 1, 2, \cdots$ either $EQ_l = 0$ or $Q_l \leq E'$.

Proof: We have $E' \in \mathfrak{B} \cap \mathfrak{B}'$ and, from Theorem 1, Q_l is minimal among the central projections of $\mathfrak{B}$ so that the projection $E'Q_l$ is either 0 or Q_l. Now $E'Q_l = 0$ implies $EQ_l = 0$ and $E'Q_l = Q_l$ is equivalent to $Q_l \leq E'$.

Now we have

Theorem 2: Consider the same potentials as in Lemma 1 and assume that (9) holds. Furthermore, assume that negative-energy bound states exist, that each is isolated with finite multiplicity, and that no positive-energy bound states exist.

(a) Consider a particular value of $l = 0, 1, 2, \cdots$. If $(\mathfrak{M}_\pm)_{Q_l}$ are Abelian, then there are no bound states of angular momentum l. If there are no bound states of angular momentum l and asymptotic completeness holds, then $(\mathfrak{M}_\pm)_{Q_l}$ are Abelian.

(b) Let F denote the projection onto the subspace spanned by the eigenfunctions of H and let F' denote its central carrier relative to $\mathfrak{B}$. Let $G_t^\pm$ denote the projections corresponding to the respective factor summands of the Suzuki decompositions of $\mathfrak{M}_\pm$ [see

(10)]. Then, F' equals the sum of those Q_l such that $(\mathfrak{M}_\pm)_{Q_l}$ are non-Abelian and $F \leq G_\pm^\perp \leq F'$.

Proof: (a) If $(\mathfrak{M}_\pm)_{Q_l}$ are Abelian, then $(\Omega_\pm)_{Q_l}$ are unitary so that $Q_l \leq P_\pm$. From (9) we have that F and $P_\pm$ are orthogonal so that $FQ_l = 0$ and there are no bound states of angular momentum l. Similarly, if there are no bound states of angular momentum l, then $Q_l \leq F^\perp$. From asymptotic completeness we now find $Q_l \leq P_\pm$ so that $(\Omega_\pm)_{Q_l}$ are unitary and $(\mathfrak{M}_\pm)_{Q_l}$ are Abelian.

(b) Since $F \in \mathfrak{B}$ we have from Lemma 2 that either $FQ_l = 0$ or $Q_l \leq F'$ for a given value of l. From the proof of part (a) it is see that $(\mathfrak{M}_\pm)_{Q_l}$ Abelian implies $FQ_l = 0$. Thus, we are interested in those Q_l for which

$$Q_l \leq F'. \tag{20}$$

Denoting the sum of such Q_l by $\sum_l' Q_l$, we must show that

$$F' = \sum_l{}' Q_l. \tag{21}$$

From the above argument we have $F \leq \sum_l' Q_l$. From Theorem 1 each Q_l is central in $\mathfrak{B}$ so that $F' \leq \sum_l' Q_l$. Combining this relation with (20), we obtain (21). Consider the projections $G_\pm^\perp$ corresponding to the respective factor summands of the Suzuki decompositions of $\mathfrak{M}_\pm$. Then we have $G_\pm^\perp \in \mathfrak{M}_\pm \cap \mathfrak{M}_\pm'$ and $Q_l \in \mathfrak{M}_\pm'$. Let $(Q_\pm)_l$ denote the respective central carriers of Q_l relative to $\mathfrak{M}_\pm$. Since $G_\pm^\perp$ are minimal among the projections of $\mathfrak{M}_\pm \cap \mathfrak{M}_\pm'$, we see that the projections $G_\pm^\perp(Q_\pm)_l$ must be either 0 or $G_\pm^\perp$. Consider a value of l, say l_0, such that $(\mathfrak{M}_\pm)_{Q_{l_0}}$ are Abelian. Since $Q_{l_0} \in \mathfrak{M}_\pm'$ implies that $(\mathfrak{M}_\pm)_{Q_{l_0}}$ are respectively $*$-isomorphic to $(\mathfrak{M}_\pm)_{(Q_\pm)_{l_0}}$, we see that the latter algebras are also Abelian and, therefore, that $(Q_\pm)_{l_0}$ are finite relative to $\mathfrak{M}_\pm$.

Assume that $G_\pm^\perp(Q_\pm)_{l_0} = G_\pm^\perp$. This means $G_\pm^\perp \leq (Q_\pm)_{l_0}$ which implies that $G_\pm^\perp$ are finite relative to $\mathfrak{M}_\pm$ (Ref. 11, p. 230, Proposition 2). This is a contradiction because $(\mathfrak{M}_\pm)_{G_\pm^\perp}$ are factors of type I_∞. Thus, we conclude that $G_\pm^\perp(Q_\pm)_l = 0$ for l such that $(\mathfrak{M}_\pm)_{Q_l}$ are Abelian. This implies that

$$G_\pm^\perp \left(\sum_l{}' Q_l\right)^\perp = 0 = G_\pm^\perp F'^\perp$$

or

$$G_\pm^\perp \leq F', \tag{22}$$

which was to be proved. If $(\mathfrak{M}_\pm)_{Q_l}$ are non-Abelian for all values of l, then $F' = I$ so that (22) is trivially satisfied. We complete the proof of Theorem 2 by noting that $F \leq G_\pm^\perp$ follows directly from (11) and the orthogonality of F and $P_\pm$.

For short range potentials we can now give some results relating the properties of the algebras $(\mathfrak{M}_\pm)_{Q_l}$ to the potential range.

Corollary: Consider the same hypotheses as in Theorem 2 and in addition assume that asymptotic completeness holds.

(a) Suppose that

$$\int_0^\infty r|V(r)|dr < \infty. \tag{23}$$

Then there exists a value of angular momentum l_m such that $(\mathfrak{M}_\pm)_{Q_l}$ are Abelian for all $l > l_m$.

(b) Suppose that $V(r)$ is everywhere attractive, $V(r) = -|V(r)|$, and that at least one of the following conditions is satisfied for given angular momentum L:

$$\begin{aligned} &-\int_0^R dr\, r\, V(r)\left(\frac{r}{R}\right)^{2L+1} - \int_R^\infty dr\, r\, V(r)\left(\frac{r}{R}\right)^{-(2L+1)} \\ &\geq 2L+1, \\ &-R\int_0^\infty dr\, V(r)\left[\left(\frac{r}{R}\right)^{-2L} - \left(\frac{r}{R}\right)^{-2L} R^2 V(r)\right]^{-1} > 1, \end{aligned} \tag{24}$$

where R denotes an arbitrary positive constant. Then $(\mathfrak{M}_\pm)_{Q_l}$ are non-Abelian for all $l \leq L$. Here we, of course, assume that each of the integrals in (24) exists and is finite.

Proof: (a) If (23) holds, then it is well known that the number of bound states is finite and that a value of angular momentum exists, say l_m, such that no bound states exist for $l > l_m$.[25] The result now follows from part (a) of Theorem 2.

(b) It was shown by Calogero[26] that each of the conditions (24) guarantees the existence of at least one bound state for each $l \leq L$. We again use part (a) of Theorem 2.

We now make some remarks concerning Theorem 2 and also derive some further conclusions for the assumed class of potentials. The relation (21) defines decompositions of $\mathfrak{B}_{F'}$ and $\mathfrak{B}_{F'}'$ as direct sums of factors,

$$\mathfrak{B}_{F'} = \sum_l{}' \mathfrak{B}_{Q_l}, \qquad \mathfrak{B}_{F'}' = \sum_l{}' \mathfrak{B}_{Q_l}', \tag{25}$$

the type of each of which is known from Theorem 1. However, the von Neumann algebras most directly associated with the totality of bound states are $\mathfrak{B}_F$ and $\mathfrak{B}_F'$. Since $F \in \mathfrak{B}$ we have that $\mathfrak{B}_F'$ is $*$-isomorphic to $\mathfrak{B}_{F'}'$. Combining this result with parts (a) and (b) of Theorem 2, one deduces that if bound states exist for each angular momentum $l = 0, 1, 2, \cdots$, then $\mathfrak{B}_F'$ is $*$-isomorphic to $\mathfrak{B}'$. The algebra $\mathfrak{B}_{F'}$ is infinite since $\mathfrak{B}$ is properly infinite (Corollary to Theorem 1). This is also clear from (25) since, from Theorem 1, each $\mathfrak{B}_{Q_l}$ is a factor of type I_∞. Moreover, using the method of proof of the above-noted corollary, one shows that $\mathfrak{B}_{F'}$ is properly infinite. We can also see that, contrary to the situation for their commutants, $\mathfrak{B}_F$ and $\mathfrak{B}_{F'}$ are in general *not* isomorphic because $\mathfrak{B}_F$ may be finite. An example of this situation is the case in which a finite number of s-wave bound states exist but no others. Then one finds $F' = Q_0$ so that $\mathfrak{B}_F'$ is a factor of type I_1 and, therefore, $\mathfrak{B}_F$ is a factor of type I_p, where p equals the number of bound states. In general, of course, where bound states of more than one value of l exist, $\mathfrak{B}_F$ is not a factor.

We can obtain an expression for $\mathfrak{B}_F$ somewhat (but not exactly) analogous to that for $\mathfrak{B}_{F'}$ in (25). From $F \leq F' = \sum_l' Q_l$, we obtain

$$\mathfrak{B}_F = \sum_l{}' \mathfrak{B}_{FQ_l}. \tag{26}$$

The $\mathfrak{B}_{FQ_l}$ are factors and their commutants are of type I_{2l+1} because of respective $*$-isomorphisms to $\mathfrak{B}'_{Q_l}$. We note, however, that the projections FQ_l in (26) are not central in $\mathfrak{B}$ as are the Q_l in (25).

It was seen above that, in the general case, $\mathfrak{B}_F$ and $\mathfrak{B}'_F$ are not factors. Our final result shows that, given certain hypotheses, the situation is different for $(\mathfrak{M}_\pm)_F$ and $(\mathfrak{M}'_\pm)_F$ so that, under these conditions, $\mathfrak{M}_\pm \neq \mathfrak{B}$. The assumptions needed for the proof of these assertions are weaker than those invoked to obtain the previous results (e.g., rotational invariance is not assumed) but, on the other hand, only a partial classification of the factors $(\mathfrak{M}_\pm)_F$ is obtained.

Proposition 2: Consider potentials $V(\mathbf{x})$ satisfying (5) such that the corresponding wave operators $\Omega_\pm$ exist and asymptotic completeness holds. Furthermore, assume that n bound states exist, where $n > 0$ may be either finite or infinite. If F denotes the projection onto the subspace spanned by the eigenfunctions of H, then $(\mathfrak{M}'_\pm)_F$ are factors of type I_n.

Proof: From asymptotic completeness and (6) we have $F \in \mathfrak{M}_\pm$. Let $F_\pm$ denote the respective central carriers of F relative to $\mathfrak{M}_\pm$. Then, since bound states exist and $G_\pm^\perp$ are minimal in $\mathfrak{M}_\pm \cap \mathfrak{M}'_\pm$, we have $F_\pm = G_\pm^\perp$, using the fact that $F \leq G_\pm^\perp$. Because $F \in \mathfrak{M}_\pm$ we have that $(\mathfrak{M}'_\pm)_F$ are $*$-isomorphic to $(\mathfrak{M}'_\pm)_{F_\pm} = (\mathfrak{M}'_\pm)_{G_\pm^\perp}$. It follows from Suzuki's work[16] and standard results concerning the tensor product of von Neumann algebras (Ref. 11, p. 24) that the factors $(\mathfrak{M}'_\pm)_{G_\pm^\perp}$ are respectively isomorphic to the factors $\mathfrak{L}(R_\pm^\perp)$, the algebras of all bounded operators on the orthocomplements of the respective ranges of $\Omega_\pm$. Now, using the hypothesis that n (negative-energy) bound states exist, asymptotic completeness, and the respective $*$-isomorphisms of $(\mathfrak{M}'_\pm)_F$ and $(\mathfrak{M}'_\pm)_{G_\pm^\perp}$, we find that $(\mathfrak{M}'_\pm)_F$ are factors of type I_n.

ACKNOWLEDGMENTS

I wish to thank Professor K. Hepp for correspondence concerning a previous version of this work. In addition, I am especially grateful to the referee for several beneficial suggestions including conjectures states in the paper as Proposition 1 and Theorem 1.

[1] J. M. Jauch, Helv. Phys. Acta **31**, 127 (1958).
[2] Reviews of this work are available: T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966), Chap. X; C. R. Putnam, *Commutation Properties of Hilbert Space Operators and Related Topics* (Springer, New York, 1967), Chap. V.
[3] J. D. Dollard, J. Math. Phys. **5**, 729 (1964).
[4] D. Mulherin and I. I. Zinnes, J. Math. Phys. **11**, 1402 (1970).
[5] E. B. Manoukian and E. Prugovečki, Can. J. Phys. **49**, 102 (1971).
[6] B. Simon, Commun. Math. Phys. **21**, 192 (1971).
[7] V. S. Buslaev and V. B. Matveev, Teor. Mat. Fiz. **2**, 367 (1970), [Theoret. Math. Phys. **2**, 266 (1970)].
[8] W. O. Amrein, Ph. A. Martin, and B. Misra, Helv. Phys. Acta **43**, 313 (1970).
[9] E. Prugovečki, Nuovo Cimento **4B**, 105 (1971).
[10] J. M. Jauch, B. Misra, and A. G. Gibson, Helv. Phys. Acta **41**, 513 (1968); J. M. Combes, "An Algebraic Approach to Quantum Scattering Theory," preprint, Centre de Physique Théorique, C.N.R.S., Marseille; J. V. Corbett, Phys. Rev. D **1**, 3331 (1970); R. B. Lavine, J. Functional Anal. **5**, 368 (1970).
[11] For information concerning von Neumann algebras we will refer to J. Dixmier, *Les algèbras d' opéraleurs dans l' espace Hilbertien (Algèbres de von Neumann)* (Gauthier-Villars, Paris, 1969), 2nd ed.
[12] For field theory see, for example, M. Guenin, in *Lectures in Theoretical Physics, Mathematical Methods of Theoretical Physics*, edited by W. E. Brittin, A. O. Barut, and M. Guenin (Gordon and Breach, New York, 1967), Vol. IX A, and H. Araki, in *Local Quantum Theory, International School of Physics "Enrico Fermi", Course XLV*, edited by R. Jost (Academic, New York, 1969); and for statistical mechanics, D. Ruelle, *Statistical Mechanics, Rigorous Results* (Benjamin, New York, 1969).
[13] T. Kato, Trans. Amer. Math. Soc. **70**, 195 (1951); also T. Kato, Ref. 2, Section V. 5. Kato's method can also be used to give a rigorous treatment of nonlocal separable potentials, but we will not consider this problem. See J. M. Combes, Nuovo Cimento **64A**, 111 (1969).
[14] R. B. Lavine, Commun. Math. Phys. **20**, 301 (1971).
[15] See, for example, J. M. Jauch and B. Misra, Helv. Phys. Acta **38**, 30 (1965).
[16] N. Suzuki, Proc. Japan Acad. **39**, 435 (1963). See also P. R. Halmos, *A Hilbert Space Problem Book* (Van Nostrand, Princeton, N. J., 1967), Problem 118.
[17] This assumption is not made in the case of Proposition 2 in Sec. IV.
[18] T. A. Green and O. E. Lanford III, J. Math. Phys. **1**, 139 (1960); S. T. Kuroda, *ibid.* **3**, 933 (1962).
[19] K. Kodaira, Amer. J. Math. **71**, 921 (1949). Kodaira's results also allow potentials with the more singular behavior as $r \to 0$ mentioned in the paragraph preceding (12).
[20] J. Weidmann, Math. Z. **98**, 268 (1967). See also Amrein *et al.*[8]
[21] See, for example, N. Dunford, Math. Ann. **162**, 294 (1966).
[22] See, for example, C. Pearcy and D. Topping, Mich. Math. J. **14**, 453 (1967).
[23] T. Kato, Ref. 2, Theorem V. 5. 7.
[24] See B. Simon, Ref. 6, and the review by T. Kato, Progr. Theoret. Phys. (Kyoto) Suppl. **40**, 3 (1967), for further references.
[25] V. Bargmann, Proc. Natl. Acad. Sci. (U.S.) **38**, 961 (1952). Bounds for l_m have been given by F. Calogero and G. Cosenza, Nuovo Cimento **45A**, 867 (1966), and by B. Simon, J. Math. Phys. **10**, 1123 (1969).
[26] F. Calogero, J. Math. Phys. **6**, 161, 1105 (1965). See also F. Calogero, *Variable Phase Approach to Potential Scattering* (Academic, New York, 1967), Chap. 23.

Asymptotic Behavior of Sturm-Liouville Systems

José Canosa
IBM Scientific Center, 2670 Hanover Street, Palo Alto, California 94304
(Received 29 August 1971; Revised Manuscript Received 4 January 1972)

It is shown that the bound-state solutions of regular and singular Sturm–Liouville systems have the same asymptotic behavior, in the limit of infinitely deep potential wells. This conclusion is illustrated explicitly by showing the profound similarity that exists between two seemingly unconnected problems: Mathieu's equation in a finite domain and Schrödinger's equation with Morse's potential.

1. INTRODUCTION

Sturm–Liouville systems occupy a very important place both in classical analysis and physical applications. Many one- and two-dimensional vibration problems lead to regular eigenvalue problems of the Sturm–Liouville type.[1] The Schrödinger equation in an infinite domain is a singular Sturm-Liouville system.[2]

The $\mathfrak{B}_{FQ_l}$ are factors and their commutants are of type I_{2l+1} because of respective $*$-isomorphisms to $\mathfrak{B}'_{Q_l}$. We note, however, that the projections FQ_l in (26) are not central in $\mathfrak{B}$ as are the Q_l in (25).

It was seen above that, in the general case, $\mathfrak{B}_F$ and $\mathfrak{B}'_F$ are not factors. Our final result shows that, given certain hypotheses, the situation is different for $(\mathfrak{M}_\pm)_F$ and $(\mathfrak{M}'_\pm)_F$ so that, under these conditions, $\mathfrak{M}_\pm \neq \mathfrak{B}$. The assumptions needed for the proof of these assertions are weaker than those invoked to obtain the previous results (e.g., rotational invariance is not assumed) but, on the other hand, only a partial classification of the factors $(\mathfrak{M}_\pm)_F$ is obtained.

Proposition 2: Consider potentials $V(\mathbf{x})$ satisfying (5) such that the corresponding wave operators $\Omega_\pm$ exist and asymptotic completeness holds. Furthermore, assume that n bound states exist, where $n > 0$ may be either finite or infinite. If F denotes the projection onto the subspace spanned by the eigenfunctions of H, then $(\mathfrak{M}'_\pm)_F$ are factors of type I_n.

Proof: From asymptotic completeness and (6) we have $F \in \mathfrak{M}_\pm$. Let $F_\pm$ denote the respective central carriers of F relative to $\mathfrak{M}_\pm$. Then, since bound states exist and $G_\pm^\perp$ are minimal in $\mathfrak{M}_\pm \cap \mathfrak{M}'_\pm$, we have $F_\pm = G_\pm^\perp$, using the fact that $F \leq G_\pm^\perp$. Because $F \in \mathfrak{M}_\pm$ we have that $(\mathfrak{M}'_\pm)_F$ are $*$-isomorphic to $(\mathfrak{M}'_\pm)_{F_\pm} = (\mathfrak{M}'_\pm)_{G_\pm^\perp}$. It follows from Suzuki's work[16] and standard results concerning the tensor product of von Neumann algebras (Ref. 11, p. 24) that the factors $(\mathfrak{M}'_\pm)_{G_\pm^\perp}$ are respectively isomorphic to the factors $\mathfrak{L}(R_\pm^\perp)$, the algebras of all bounded operators on the orthocomplements of the respective ranges of $\Omega_\pm$. Now, using the hypothesis that n (negative-energy) bound states exist, asymptotic completeness, and the respective $*$-isomorphisms of $(\mathfrak{M}'_\pm)_F$ and $(\mathfrak{M}'_\pm)_{G_\pm^\perp}$, we find that $(\mathfrak{M}'_\pm)_F$ are factors of type I_n.

ACKNOWLEDGMENTS

I wish to thank Professor K. Hepp for correspondence concerning a previous version of this work. In addition, I am especially grateful to the referee for several beneficial suggestions including conjectures states in the paper as Proposition 1 and Theorem 1.

[1] J. M. Jauch, Helv. Phys. Acta **31**, 127 (1958).
[2] Reviews of this work are available: T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966), Chap. X; C. R. Putnam, *Commutation Properties of Hilbert Space Operators and Related Topics* (Springer, New York, 1967), Chap. V.
[3] J. D. Dollard, J. Math. Phys. **5**, 729 (1964).
[4] D. Mulherin and I. I. Zinnes, J. Math. Phys. **11**, 1402 (1970).
[5] E. B. Manoukian and E. Prugovečki, Can. J. Phys. **49**, 102 (1971).
[6] B. Simon, Commun. Math. Phys. **21**, 192 (1971).
[7] V. S. Buslaev and V. B. Matveev, Teor. Mat. Fiz. **2**, 367 (1970), [Theoret. Math. Phys. **2**, 266 (1970)].
[8] W. O. Amrein, Ph. A. Martin, and B. Misra, Helv. Phys. Acta **43**, 313 (1970).
[9] E. Prugovečki, Nuovo Cimento **4B**, 105 (1971).
[10] J. M. Jauch, B. Misra, and A. G. Gibson, Helv. Phys. Acta **41**, 513 (1968); J. M. Combes, "An Algebraic Approach to Quantum Scattering Theory," preprint, Centre de Physique Théorique, C.N.R.S., Marseille; J. V. Corbett, Phys. Rev. D **1**, 3331 (1970); R. B. Lavine, J. Functional Anal. **5**, 368 (1970).
[11] For information concerning von Neumann algebras we will refer to J. Dixmier, *Les algèbras d' opéraleurs dans l' espace Hilbertien (Algèbres de von Neumann)* (Gauthier-Villars, Paris, 1969), 2nd ed.
[12] For field theory see, for example, M. Guenin, in *Lectures in Theoretical Physics, Mathematical Methods of Theoretical Physics*, edited by W. E. Brittin, A. O. Barut, and M. Guenin (Gordon and Breach, New York, 1967), Vol. IX A, and H. Araki, in *Local Quantum Theory, International School of Physics "Enrico Fermi", Course XLV*, edited by R. Jost (Academic, New York, 1969); and for statistical mechanics, D. Ruelle, *Statistical Mechanics, Rigorous Results* (Benjamin, New York, 1969).
[13] T. Kato, Trans. Amer. Math. Soc. **70**, 195 (1951); also T. Kato, Ref. 2, Section V. 5. Kato's method can also be used to give a rigorous treatment of nonlocal separable potentials, but we will not consider this problem. See J. M. Combes, Nuovo Cimento **64A**, 111 (1969).
[14] R. B. Lavine, Commun. Math. Phys. **20**, 301 (1971).
[15] See, for example, J. M. Jauch and B. Misra, Helv. Phys. Acta **38**, 30 (1965).
[16] N. Suzuki, Proc. Japan Acad. **39**, 435 (1963). See also P. R. Halmos, *A Hilbert Space Problem Book* (Van Nostrand, Princeton, N. J., 1967), Problem 118.
[17] This assumption is not made in the case of Proposition 2 in Sec. IV.
[18] T. A. Green and O. E. Lanford III, J. Math. Phys. **1**, 139 (1960); S. T. Kuroda, *ibid.* **3**, 933 (1962).
[19] K. Kodaira, Amer. J. Math. **71**, 921 (1949). Kodaira's results also allow potentials with the more singular behavior as $r \to 0$ mentioned in the paragraph preceding (12).
[20] J. Weidmann, Math. Z. **98**, 268 (1967). See also Amrein *et al.*[8]
[21] See, for example, N. Dunford, Math. Ann. **162**, 294 (1966).
[22] See, for example, C. Pearcy and D. Topping, Mich. Math. J. **14**, 453 (1967).
[23] T. Kato, Ref. 2, Theorem V. 5. 7.
[24] See B. Simon, Ref. 6, and the review by T. Kato, Progr. Theoret. Phys. (Kyoto) Suppl. **40**, 3 (1967), for further references.
[25] V. Bargmann, Proc. Natl. Acad. Sci. (U.S.) **38**, 961 (1952). Bounds for l_m have been given by F. Calogero and G. Cosenza, Nuovo Cimento **45A**, 867 (1966), and by B. Simon, J. Math. Phys. **10**, 1123 (1969).
[26] F. Calogero, J. Math. Phys. **6**, 161, 1105 (1965). See also F. Calogero, *Variable Phase Approach to Potential Scattering* (Academic, New York, 1967), Chap. 23.

Asymptotic Behavior of Sturm-Liouville Systems

José Canosa
IBM Scientific Center, 2670 Hanover Street, Palo Alto, California 94304
(Received 29 August 1971; Revised Manuscript Received 4 January 1972)

It is shown that the bound-state solutions of regular and singular Sturm–Liouville systems have the same asymptotic behavior, in the limit of infinitely deep potential wells. This conclusion is illustrated explicitly by showing the profound similarity that exists between two seemingly unconnected problems: Mathieu's equation in a finite domain and Schrödinger's equation with Morse's potential.

1. INTRODUCTION

Sturm–Liouville systems occupy a very important place both in classical analysis and physical applications. Many one- and two-dimensional vibration problems lead to regular eigenvalue problems of the Sturm–Liouville type.[1] The Schrödinger equation in an infinite domain is a singular Sturm-Liouville system.[2]

The aim of this paper is to discuss the asymptotic behavior of the Sturm–Liouville equation in Liouville normal form[2]

$$\frac{d^2y}{dx^2} + [\lambda - V(x)]y = 0, \tag{1}$$

when the continuous function $V(x)$ has the form of an extremely deep well.

As the one-dimensional Schrödinger equation is exactly in Liouville normal form, we will use the terminology and concepts of quantum mechanics to give physical content to our discussion right from the start. Equation (1) together with endpoint conditions satisfied at the boundaries of a finite domain, e.g.,

$$y(0) = y(\pi) = 0, \tag{2}$$

defines a regular Sturm–Liouville system, which in general has a finite number of negative eigenvalues and an infinite number of discrete positive eigenvalues, all of which give the bound states for the motion of a particle in a potential well $V(x)$ of length π and infinitely high walls.

In most quantum mechanics problems, the boundary conditions associated with (1) are

$$y(\pm\infty) \text{ finite}. \tag{3}$$

In this case, (1) and (3) define a singular Sturm–Liouville system, which in general has a discrete spectrum of negative eigenvalues giving the bound states, and a continuous spectrum of positive eigenvalues corresponding to the free particles.

The main result of this work is to show that the asymptotic behavior of the bound-state solutions is essentially the same for both the regular and singular Sturm–Liouville systems, provided that the potential well $V(x)$ is sufficiently deep. This behavior is independent of the type of homogeneous boundary conditions associated with the regular problem. Furthermore, the bound-state solutions can be obtained explicitly and fairly accurately by using standard perturbation theory.[3,4] It should be emphasized that the asymptotic problem considered cannot be treated by the classical WKB method because the potential is a very rapidly varying function of position; it is also one of the most difficult and expensive problems to solve with numerical methods because of the extremely rapid variation of the eigenfunctions in the region about the minimum of the potential.

As an illustration of the applicability of the method to a regular system, we have treated Mathieu's equation and have obtained exactly the first four terms of the asymptotic series for the eigenvalues due to Goldstein and Ince.[5–8] We also show the accuracy of the analytic approximations to Mathieu's functions by a comparison with the "exact" numerical results.[9] Although Sips[10,11] obtained higher order asymptotic results than ours, he did so by a special method developed specifically for Mathieu's equation, while our method is *equally* applicable to *any* equation of the form (1); furthermore, our results for Mathieu's functions have a much simpler and thus more useful form than Sips'.

Finally, we give the results obtained for the bound-state eigenvalues of a Schrödinger equation problem in an infinite domain. In this problem with Morse's potential,[12] the eigenvalues are obtained exactly. Regardless of accuracy considerations, the results obtained for the problems discussed show clearly that the asymptotic behavior of both the regular and singular Sturm–Liouville systems is essentially the same.

2. OUTLINE OF THE METHOD

The central idea of this work is to expand the potential $V(x)$ in a Taylor series about its minimum value, and solve the resulting approximate problem by standard perturbation theory. Physically, as the potential well is very deep, the classical turning points for the lowest bound states are very close to the minimum of the well; therefore the particle, even quantum-mechanically, can in the main "see" only the region of the potential near the minimum. We thus expect that the eigenfunctions obtained in this way will only be accurate in an asymptotically small neighborhood of the potential minimum; also, for the higher bound states, as the turning points intersect the potential well at higher points above the minimum, we expect that the accuracy of the approximation will decrease progressively relative to the lowest bound states.

This is indeed confirmed by the explicit results obtained. However, the region about the potential minimum is where the eigenfunctions have appreciable values; away from it, they are transcendentally small and in many applications it is adequate to know the eigenfunctions accurately only where they have non-negligible values. Furthermore, the eigenvalues obtained by the method are asymptotically accurate even if the eigenfunctions are not determined accurately in the classically forbidden region outside the turning points.

3. MATHIEU'S EQUATION

Mathieu's equation arises when solving by separation of variables the two-dimensional Helmholtz equation in a domain with an elliptic boundary (Ref. 1, p. 391). In standard form, Mathieu's equation reads

$$\frac{d^2y}{dx^2} + (\lambda - 2q\ \cos 2x)y = 0. \tag{4}$$

It is well known[8,13] that Eq. (4) has periodic solutions of period π and 2π when λ has a countably infinite set of characteristic values. The real parameter q determines the depth of the potential well $2q\ \cos 2x$; here we will discuss the asymptotic behavior of (4) in the limit of infinitely deep wells, i.e., when $q \to \infty$. Equation (4) with the associated boundary conditions

$$y'(0) = y'(\pi) = 0 \tag{5}$$

defines a regular Sturm–Liouville system[2] whose eigenfunctions are the even periodic Mathieu functions

$$\mathrm{ce}_r(x, q), \quad r = 0, 1, 2, \cdots. \tag{6}$$

The eigenfunctions of Eq. (4) with the boundary conditions

$$y(0) = y(\pi) = 0 \tag{7}$$

are the odd periodic Mathieu functions designated

$$\mathrm{se}_r(x, q), \quad r = 1, 2, 3, \cdots. \tag{8}$$

Here we follow the standard notation of Ref. 8. The reason for the notation ce_r (cosine elliptic) and se_r (sine elliptic) is that for $q = 0$ the eigenfunctions of systems (4) and (5) and (4) and (7) become, respectively, the trigometric cosines and sines.

As discussed in Sec. II, for very deep wells ($q \to \infty$) the lowest eigenvalues approach asymptotically the minimum value of the potential $V_{\min} = -2q$, i.e.,

$$\lambda = O(q). \tag{9}$$

Therefore, a simple division of Eq. (4) by the parameter q shows that for $q \to \infty$, we have a singular perturbation problem[14]

$$\begin{aligned} &\epsilon \frac{d^2y}{dx^2} + (\Lambda - 2\cos 2x)y = 0, \\ &\epsilon \equiv 1/q \to 0, \quad \Lambda \equiv \lambda/q = O(1). \end{aligned} \tag{10}$$

In general, the solution of these problems cannot be obtained as a single series that is uniformly valid in the whole domain.[14] A direct expansion of the solution of (10) in powers of the small parameter ϵ gives to all orders

$$y(x, \epsilon) = 0. \tag{11}$$

The meaning of the formal result (11) can be understood by considering that it is obtained by neglecting the term $\epsilon d^2y/dx^2$ in (10), and this is justified only in the region where the curvature of the solution is not large. As the curvature is greatest at the domain center $x = \frac{1}{2}\pi$, (11) suggests that the solution away from the center is transcendentally small, a conclusion consistent with the boundary conditions (5) and (7). To study the solution in the inner layer about the domain center $x = \frac{1}{2}\pi$, where it has nonnegligible values, we introduce the stretched variable

$$\bar{x} = (\pi/2 - x)/\epsilon^{1/2}, \tag{12}$$

and (10) becomes

$$\frac{d^2y}{d\bar{x}^2} + (\Lambda + 2\cos 2\sqrt{\epsilon}\,\bar{x})y = 0. \tag{13}$$

This transformation shows, as the small parameter ϵ has been eliminated from the highest order derivative term, that in the limit $\epsilon \to 0$ the thickness of the region about $\frac{1}{2}\pi$ where the solution is not negligibly small approaches zero as $\epsilon^{1/2}$.

The procedure is now straightforward, as described in Sec. II. One expands

$$\cos 2\sqrt{\epsilon}\,\bar{x} = 1 - 2\epsilon\bar{x}^2 + \tfrac{2}{3}\epsilon^2\bar{x}^4 - \tfrac{4}{45}\epsilon^3\bar{x}^6 + \cdots \tag{14}$$

so that (13) becomes

$$\frac{d^2y}{d\bar{x}^2} + (\Lambda + 2 - 4\epsilon\bar{x}^2 + \tfrac{4}{3}\epsilon^2\bar{x}^4 - \tfrac{8}{45}\epsilon^3\bar{x}^6)y = 0. \tag{15}$$

It is now convenient to introduce still a new independent variable

$$x^* = (4\epsilon)^{1/4}\bar{x}, \tag{16}$$

so that (15) takes the well-known form of the equation for an anharmonic oscillator

$$\frac{d^2y}{dx^{*2}} + (\rho - x^{*2} + \tfrac{1}{8}\sqrt{\epsilon}\,x^{*4} - \tfrac{1}{90}\epsilon x^{*6})y = 0, \tag{17}$$

where the eigenvalue ρ is defined as

$$\rho \equiv (\Lambda + 2)/2\epsilon^{1/2}. \tag{18}$$

The situation is now quite clear, as Mathieu's functions become transcendentally small except in an infinitesimally thin layer about $\frac{1}{2}\pi$, for all quantitative purposes the domain (whose length is π) can be considered to be of infinite length. Therefore, we will solve the equation for the anharmonic oscillator (17) subject to the usual boundary conditions

$$y(\pm\infty) \quad \text{finite}. \tag{19}$$

This means that in the limit $q \to \infty$, the even and odd periodic Mathieu functions ce_r and se_r approach each other asymptotically, a known analytic result[10,11,15] that has also been established numerically.[9] Heuristically, we make the hypothesis that for infinitely deep wells all regular Sturm–Liouville systems approach the same asymptotic limit, *independently* of the type of homogeneous boundary conditions. This is because, just as for problems in infinite domains, *both* the eigenfunctions and their derivatives vanish asymptotically at the finite domain boundaries, and this is indeed equivalent to any of the homogenous boundary conditions naturally associated with the Sturm–Liouville equation (Courant and Hilbert, p. 291).[1] This asymptotic limit is given by the solution of Schrödinger's equation for the anharmonic oscillator (17) and (19), which is now solved by standard perturbation theory.[3,4]

The unperturbed Hamiltonian is

$$H^0 = x^{*2}, \tag{20}$$

which gives the familiar equation for the harmonic oscillator, and the perturbed Hamiltonian is

$$H' = -\tfrac{1}{8}\sqrt{\epsilon}\,x^{*4} + \tfrac{1}{90}\epsilon x^{*6}. \tag{21}$$

The solution for the harmonic oscillator is

$$\rho_n^0 = 2n + 1, \tag{22}$$

$$y_n^0(x^*) = (2^n n!\sqrt{\pi})^{-1/2} e^{-x^{*2}/2} H_n(x^*), \quad n = 0, 1, 2, \ldots, \tag{23}$$

where H_n are the Hermite polynomials.

Second-order perturbation theory for the eigenvalues gives

$$\rho_n' = \rho_n^0 + (H')_{nn} + \sum_{n \neq k} [(H')_{kn}]^2/(\rho_n^0 - \rho_k^0), \tag{24}$$

while the first-order perturbation theory result for the eigenfunctions is

$$y'_n = y^0_n + \sum_{n \neq k} [(H')_{kn}/(\rho^0_n - \rho^0_k)]y^0_k. \tag{25}$$

In (24) and (25), $(H')_{kn}$ are the matrix elements

$$(H')_{kn} = (y^0_k, H' y^0_n) = \int_{-\infty}^{\infty} y^0_k H' y^0_n dx^*, \tag{26}$$

which are obtained in the usual way (Landau and Lifshitz, p. 136).[3] All the nonzero matrix elements necessary for our approximations (24) and (25) are

$$\begin{aligned}
(H')_{n,n} &= -\frac{3}{4}\frac{(2n+1)^2+1}{2}\frac{\sqrt{\epsilon}}{6} \\
&\quad + \frac{20n^3 + 30n^2 + 40n + 15}{8}\frac{\epsilon}{90}, \\
(H')_{n,n+2} &= (H')_{n+2,n} \\
&= -\frac{(4n+6)\sqrt{(n+2)(n+1)}}{4}\frac{\sqrt{\epsilon}}{6} + O(\epsilon), \\
(H')_{n,n+4} &= (H')_{n+4,n} \\
&= -\frac{\sqrt{(n+4)(n+3)(n+2)(n+1)}}{4}\frac{\sqrt{\epsilon}}{6} + O(\epsilon), \\
n &= 0, 1, 2, \cdots.
\end{aligned} \tag{27}$$

We now substitute the matrix elements (27) into the eigenvalue expression (24), which in this case is given explicitly by

$$\begin{aligned}
\rho'_n = \rho^0_n + (H')_{nn} &+ \frac{[(H')_{n-4,n}]^2}{\rho^0_n - \rho^0_{n-4}} + \frac{[(H')_{n-2,n}]^2}{\rho^0_n - \rho^0_{n-2}} \\
&+ \frac{[(H')_{n,n+2}]^2}{\rho^0_n - \rho^0_{n+2}} + \frac{[(H')_{n,n+4}]^2}{\rho^0_n - \rho^0_{n+4}}
\end{aligned} \tag{28}$$

as all the other matrix elements are zero. It should be noticed that the matrix elements $(H')_{n-4,n}$ are obtained from the last equation of (27) by substituting $n+4$ by n, $n+3$ by $n-1$, etc. The eigenvalues ρ are now transformed back into the original Mathieu eigenvalues λ using the definitions (18) and (10). After some elementary although delicate algebra we finally get

$$\begin{aligned}
\lambda_n &= -2q + 2(2n+1)q^{1/2} - \{[(2n+1)^2 + 1]/2^3\} \\
&\quad - \{[(2n+1)^3 + 3(2n+1)]/2^7\}q^{-1/2} + O(q^{-1}), \\
n &= 0, 1, 2, \ldots,
\end{aligned} \tag{29}$$

which gives exactly the first four terms of the classic result of Goldstein's and Ince's [Eq. (36), Ref. 5, also Eq. 20.2.30, p. 726, Ref. 8].

It is worthwhile to remark here that the first three terms of (29) are obtained simply by first order perturbation theory, keeping only for the perturbed Hamiltonian the first term in (21). It is clear that higher order terms in the asymptotic series (29) can be obtained by using higher order perturbation theory. However, the result (29) is sufficient for our purpose of illustrating the asymptotic behavior of regular Sturm–Liouville systems.

In Table I, we give the Mathieu eigenvalues obtained from the first three terms of Eq. (29), together with those obtained from the full four term asymptotic formula. It is seen that, as expected, the accuracy of the asymptotic treatment increases with the depth of the well, i.e., with the value of the parameter q. Thus

TABLE I. First five eigenvalues of Mathieu's equation.

	$-\lambda_r(q = 1\,600)$		
r	Three term asymp. form.	Four term asymp. form. (29)	Exact[a]
0	3120.250	3120.251	3120.251
1	2961.250	2961.257	2961.257
2	2803.250	2803.277	2803.278
3	2646.250	2646.321	2646.323
4	2490.250	2490.398	2490.403
	$-\lambda_r(q = 2\,500)$		
r	Three term asymp. form.	Four term asymp. form. (29)	Exact[a]
0	4900.250	4900.251	4900.251
1	4701.250	4701.256	4701.256
2	4503.250	4503.272	4503.272
3	4306.250	4306.307	4306.308
4	4110.250	4110.368	4110.372
	$-\lambda_r(q = 10\,000)$		
r	Three asymp. form.	Four term asymp. form. (29)	Exact[a]
0	19 800.250	19 800.250	19 800.250
1	19 401.250	19 401.253	19 401.253
2	19 003.250	19 003.261	19 003.261
3	18 606.250	18 606.278	18 606.279
4	18 210.250	18 210.309	18 210.310

[a]Determined numerically (Ref. 9), and also using higher order terms in Goldstein's and Ince's asymptotic formula.

for $q = 1\,600$, the four-term formula gives better than six significant figure accuracy, and seven and eight significant figure accuracy is obtained, respectively, for $q = 2\,500$ and 10 000. It should also be noticed that a simple first-order perturbation theory treatment, i.e., the three term asymptotic formula, gives better than four significant figure accuracy in all cases shown.

To conclude this section, we give some numerical results for the Mathieu functions computed from the first-order perturbation theory formula (25). For clarity we will write down the results explicitly. The fundamental eigenfunction is

$$\begin{aligned}
&\left.\begin{matrix}\mathrm{ce}_0(x^*, q) \\ \mathrm{se}_1(x^*, q)\end{matrix}\right\} \sim y'_0(x^*, q) \\
&= y^0_0 + \tfrac{1}{6}\left(\frac{3}{4\sqrt{2}}y^0_2 + \frac{\sqrt{3}}{8\sqrt{2}}y^0_4\right)\sqrt{\epsilon} + O(\epsilon),
\end{aligned} \tag{30}$$

where y^0_n are the orthonormal Hermite functions given in (23). In Table II we compare with the exact numerical values[9] the results obtained using the first term (harmonic oscillator) of (30) and the complete perturbation formula. As expected, the accuracy is only good near the minimum of the potential which is at the domain center $\frac{1}{2}\pi$. However, the accuracy is lost only where the function is already negligible compared with the values at the center. These results are in general agreement with those of Sips[10,11], who obtained asymptotic series for Mathieu's functions of higher order than (30). However, Sips' series are not as simple for hand computations because the space

TABLE II. Comparison of perturbation theory results for Mathieu's function $ce_0(x, q)$ or $se_1(x, q)$ with exact values.[a]

	$ce_0(90^0)/ce_0(85^0)$			$ce_0(90^0)/ce_0(80^0)$			$ce_0(90^0)/ce_0(70^0)$			$ce_0(90^0)/ce_0(60^0)$		
q	Harm. oscill.	First-order pert.	Exact	Harm. oscill.	First-order pert.	Exact	Harm. oscill.	First-order pert.	Exact	Harm. oscill.	First-order pert.	Exact
1 600	1.356	1.355	1.355	3.382	3.359	3.359	1.308 + 2	1.229 + 2	1.226 + 2	5.789 + 4	4.504 + 4	4.362 + 4
2 500	1.463	1.462	1.462	4.586	4.551	4.551	4.425 + 2	4.108 + 2	4.097 + 2	8.979 + 5	6.662 + 5	6.358 + 5
10 000	2.142	2.138	2.138	2.103 + 1	2.079 + 1	2.079 + 1	1.958 + 5	1.719 + 5	1.704 + 5	8.062 + 11	4.854 + 11	4.186 + 11

[a] The notation $1.5 + 2 = 1.5 \times 10^2$ is used. The exact values are from Ref. 9.

TABLE III. Zeroes of Mathieu's functions.[a]

	ce_2 and se_3			ce_4 and se_5					
q	Harm.osc.	First-order pert.	Exact	Harm.osc.	First-order pert.	Exact	Harm.osc.	First-order pert.	Exact
1 600	1.4917	1.4913	1.4914	1.3862	1.3856	1.3846	1.5121	1.5120	1.5117
2 500	1.5001	1.4998	1.4998	1.4057	1.4052	1.4046	1.5183	1.5182	1.5180
10 000	1.5208	1.5207	1.5207	1.4541	1.4539	1.4537	1.5337	1.5336	1.5336

[a] The exact values are from Ref. 9.

variable appears only as $\cos x^*$, so that one has transcendental polynomials in $\cos x^*$ [see Eq. (3), Ref. 11] while expressions as (30) are simple algebraic polynomials in x^*. It is clear that higher order perturbation theory would yield higher order terms in (30), but we do not think it to be a worthwhile exercise, because the analytic results are useful only if they are sufficiently simple for computation by elementary means.

As another illustration of the usefulness of these simple results, we have computed the zeros of the third and fifth eigenfunctions, namely

$$\left.\begin{matrix} ce_2(x^*, q) \\ se_3(x^*, q) \end{matrix}\right\} \sim y_2' = y_2^0 + \frac{(H')_{02}}{\rho_2^0 - \rho_0^0} y_0^0 + \frac{(H')_{24}}{\rho_2^0 - \rho_4^0} y_4^0 + \frac{(H')_{26}}{\rho_2^0 - \rho_6^0} y_6^0 \approx y_2^0 - \frac{1}{6}\left(\frac{3}{4\sqrt{2}} y_0^0 - \frac{7\sqrt{3}}{4} y_4^0\right)\sqrt{\epsilon}, \tag{31}$$

where for simplicity we have neglected the last term involving y_6^0; in this way the zeros of ce_2 and se_3 can be obtained analytically by solving a simple biquadratic equation. For this same reason, in the expression for ce_4 and se_5 we neglected the terms involving y_6^0 and y_8^0, so that we are left with

$$\left.\begin{matrix} ce_4(x^*, q) \\ se_5(x^*, q) \end{matrix}\right\} \sim y_4' \approx y_4^0 - \frac{1}{6}\left(\frac{\sqrt{3}}{8\sqrt{2}} y_0^0 + \frac{7\sqrt{3}}{4} y_2^0\right)\sqrt{\epsilon}, \tag{32}$$

so that again the zeros are given by solving the above biquadratic equation. The results obtained are given in Table III. The most striking fact in Table III is that the zeros of the harmonic oscillator eigenfunctions, i.e., of the Hermite polynomials, are accurate to four significant figures. The approximate first-order perturbation theory results have almost five figure accuracy. The reason for this accuracy is that, in the limit $q \to \infty$, all the zeros condense toward $\frac{1}{2}\pi$ and thus are located precisely in the inner layer about the center where the functions are obtained accurately.

4. SCHRÖDINGER'S EQUATION WITH MORSE'S POTENTIAL

In connection with the description of the bound states of diatomic molecules, Morse[12] defined the following Schrödinger equation problem:

$$\frac{d^2y}{dx^2} + [\lambda - q(e^{-2\alpha x} - 2e^{-\alpha x})]y = 0, \quad y(\pm\infty) \text{ finite.} \tag{33}$$

The problem (33) has an analytic solution for the bound states, which is given concisely in Landau and Lifshitz, p. 68.[3] The reason for having chosen this problem as an illustration of the method of Sec. II is that the potential well is strongly asymmetric, and thus it seemed a good test of a method where the zeroth order perturbation is the solution of a problem with a symmetric potential, i.e., the harmonic oscillator. Following Sec. II, we expand

$$q(e^{-2\alpha x} - 2e^{-\alpha x}) = -q + a^2qx^2 - a^3qx^3 + \tfrac{7}{12}a^4qx^4 + \cdots. \tag{34}$$

We now introduce a new independent variable

$$x^* = (a^2q)^{1/4}x \tag{35}$$

and the definitions

$$\rho \equiv \frac{\lambda + q}{(a^2q)^{1/2}}, \tag{36}$$

$$\epsilon \equiv a^{1/2}q^{-1/4}, \tag{37}$$

where the attractive strength of the well $1/\epsilon$ depends now on the two independent parameters q (the depth at $x = 0$) and a (a measure of the width of the well). In this way, problem (33) becomes

$$\frac{d^2y}{dx^{*2}} + (\rho - x^{*2} + \epsilon x^{*3} - \tfrac{7}{12}\epsilon^2 x^{*4})y = 0, \quad y(\pm\infty) \text{ finite,} \tag{38}$$

which is again the problem of an anharmonic oscilla-

TABLE IV. Bound-state eigenvalues of Schrödinger equation with Morse's potential.

	$-\lambda_n$		
n	Exact Eq.(43)	Numer. solution of Eqs.(33) and (47)	Numer. solution of Eqs.(33) and (48)
0	178.798	178.798	178.798
1	160.283	160.283	160.283
2	142.780	142.780	142.780
3	126.288	126.288	126.288
4	110.808	110.808	110.808
5	96.340	96.340	96.340
6	82.884	82.884	82.884
7	70.439	70.438	70.438
8	59.007	59.006	59.006
9	48.585	48.585	48.585
10	39.176	39.176	39.176
11	30.778	30.778	30.778
12	23.393	23.393	23.393
13	17.018	17.018	17.018
14	11.656	11.656	11.656
15	7.305	7.305	7.305
16	3.966	3.958	3.976
17	1.639	1.392	1.829
18	0.324	−1.232	0.011

tor. The unperturbed Hamiltonian (harmonic oscillator) is

$$H^0 = x^{*2}, \tag{39}$$

while the perturbed Hamiltonian is

$$H' = -\epsilon x^{*3} + \tfrac{7}{12}\epsilon^2 x^{*4}. \tag{40}$$

It should be noted that as the diagonal matrix elements of the first term of (40) $-\epsilon x^{*3}$ are zero, first-order perturbation theory of this part of the Hamiltonian does not give any correction $O(\epsilon)$ to the harmonic oscillator eigenvalues. To get any improvement, it is necessary to use second order perturbation theory. All the nonzero matrix elements of H' required are

$$\begin{aligned}(H')_{nn} &= \tfrac{7}{8}(n^2 + n + \tfrac{1}{2})\epsilon^2,\\ (H')_{n,n+1} &= (H')_{n+1,n} = -[\tfrac{9}{8}(n+1)^3]^{1/2}\epsilon,\\ (H')_{n,n+3} &= (H')_{n+3,n} = -[\tfrac{1}{8}(n+3)(n+2)(n+1)]^{1/2}\epsilon,\\ & n = 0, 1, 2, \cdots.\end{aligned} \tag{41}$$

Substituting the matrix elements (41) into the second-order perturbation theory formula for the eigenvalues (24), we get

$$\rho_n = (2n+1) - \tfrac{1}{4}(2n+1)^2\epsilon^2 + O(\epsilon^4), \quad n = 0, 1, 2, \cdots. \tag{42}$$

Transforming back to the original eigenvalues of (33) using the definition (36), one gets

$$\lambda_n = -q + (2n+1)aq^{1/2} - \tfrac{1}{4}(2n+1)^2a^2, \quad n = 0, 1, 2, \ldots, \tag{43}$$

which exceptionally is *identical* to the *exact* formula for the bound-state eigenvalues (Ref. 3, p. 69),

$$\lambda_n = -q[1 - (a/q^{1/2})(n + \tfrac{1}{2})]^2. \tag{44}$$

It is quite interesting to compare the result (43) for the eigenvalues of Schrödinger's equation with the first three terms of (29), the corresponding result for Mathieu's equation. It is seen that both asymptotic series are quite similar, and that the first three terms are of orders $q, q^{1/2}$, and 1 in the depth of the potential well. The results (29) and (43) more than anything else show quite clearly the strong similarity that exists between the asymptotic behavior of regular and singular Sturm–Liouville systems.

We now conclude with a numerical example which further illustrates the main conclusion of this work. Since the asymptotic behavior of regular and singular Sturm–Liouville problems is the same, we can approximate the solution of a regular problem by that of a singular problem or vice versa. Schrödinger's equation (33) for the values of the parameters

$$q = 188.4355, \quad a = 0.711\,248, \tag{45}$$

has 19 bound-state eigenvalues determined by the condition (Ref. 3, p. 68)

$$1/\epsilon^2 > n + \tfrac{1}{2}, \quad n = 0, 1, 2, \ldots, \tag{46}$$

where ϵ is defined in (37). We have approximated this problem by the regular Sturm–Liouville systems defined by Eq. (33) together with the boundary conditions

$$y(-1.9975) = y(8.0025) = 0 \tag{47}$$

and

$$y'(-1.9975) = y'(8.0025) = 0. \tag{48}$$

The 19 bound-state eigenvalues from the exact formula (43) and from the numerical solution[16] obtained using the boundary conditions (47) and (48) are given in Table IV. It is seen that the agreement between the first 16 eigenvalues is excellent, so that the regular Sturm–Liouville approximation to problem (33) with the boundary conditions (47) or (48) is valid for a surprisingly large number of eigenvalues and not only for the lowest bound states. The numerical results shown in Table IV for the regular Sturm–Liouville systems with boundary conditions (47) and (48) give further evidence that the asymptotic behavior of these problems is independent of the type of homogeneous boundary conditions. It should be noted that all of the eigenvalues obtained numerically are believed accurate to all figures given.

5. CONCLUSION

We have shown that in regular Sturm–Liouville problems with very deep wells, the eigenfunctions cluster in a small region about the minimum of the potential. As the eigenfunctions become transcendentally small at an infinitesimally small distance from the domain center (always chosen at the potential minimum), the problem, regardless of the homogeneous boundary conditions, can be treated as a problem for infinite domains with the boundary conditions (19). This indicates that both regular and singular Sturm–Liouville

systems have the same asymptotic behavior. This conclusion is illustrated by the remarkable similarity that exists between the asymptotic formulas for the eigenvalues of Mathieu's equation and of Schrödinger's equation with Morse's potential, two problems which at first sight seem totally unconnected. The solution is obtained by perturbation theory; the first order perturbation eigenvalues have already quite a remarkable accuracy, while the eigenfunctions are obtained accurately only in a thin layer about the domain center where they have their nonnegligible values.

[1] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Wiley-Interscience, New York, 1962), Vol. I, pp. 291 and 391.
[2] G. Birkhoff and G.-C. Rota, *Ordinary Differential Equations* (Ginn-Blaisdell, Waltham, Mass., 1962), 2nd ed., Chap. 10.
[3] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Nonrelativistic Theory* (Addison-Wesley, Reading, Mass., 1958), Chap. 6.
[4] R. H. Dicke and J. P. Wittke, *Introduction to Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1960), Chap. 14.
[5] S. Goldstein, Cambridge Phil. Soc. Trans. **23**, 303 (1927).
[6] E. L. Ince, Proc. Roy. Soc. (Edinburgh) **46**, 316 (1926).
[7] E. L. Ince, Proc. Roy. Soc. (Edinburgh) **47**, 294 (1927).
[8] M. Abramowitz and I. A. Stegun (Eds.), *Handbook of Mathematical Functions* (Dover, New York, 1965), Chap. 20.
[9] J. Canosa, J. Comput. Phys. **7**, 255 (1971).
[10] R. Sips, Trans. Amer. Math. Soc. **66**, 93 (1949).
[11] R. Sips, Trans. Amer. Math. Soc. **90**, 340 (1959).
[12] P. M. Morse, Phys. Rev. **34**, 57 (1929).
[13] N. W. McLachlan, *Theory and Application of Mathieu Functions* (Dover, New York, 1964).
[14] J. D. Cole, *Perturbation Methods in Applied Mathematics* (Ginn-Blaisdell, Waltham, Mass., 1968), p. 14.
[15] G. Blanch, Trans. Amer. Math. Soc. **97**, 357 (1960).
[16] J. Canosa and R. G. de Oliveira, J. Comput. Phys. **5**, 188 (1970).

On the Eigenvalues of the Invariant Operators of the Unitary Unimodular Group *SU(n)*

A. Partensky

Institut de Physique Nucléaire, Université Claude Bernard de Lyon (Institut National de Physique Nucléaire et de Physique des Particules), 69-Villeurbanne, France

(Received 18 June 1971; Revised Manuscript Received 29 November 1971)

By a purely infinitesimal method we derive the eigenvalues of the Biedenharn's invariant operators from I_2 to I_6 for the $SU(n)$ group. We show that the general formula, quoted from the Racah's work by Baird and Biedenharn, to obtain the eigenvalues of the invariant operators, is valid only for the two first I_2 and I_3 operators and not for the higher order invariants. We give the correct values for the first invariants till I_6.

1. INTRODUCTION

In a paper published in 1963, Biedenharn[1] has given an explicit construction of the independent invariants for the unimodular unitary group in n dimensions $SU(n)$. He proves that their invariants form a complete system of independent invariants suitable for uniquely labeling the irreducible inequivalent representations. In a second paper Baird and Biedenharn[2] pointed out that the evaluation of the invariant may be inferred from the form of the terms K_n in I_n involving the only h_i. If we denote the highest weight by vector **p** and by **R** the vector whose components are one half the sum of positive roots, we can define

$$J_n(\mathbf{p}) = K_n(\mathbf{p} + \mathbf{R}) - K_n(\mathbf{R}). \tag{1}$$

$K_n(\Lambda)$ is that part of the nth invariant I_n involving the h_i with $h_i \to \Lambda_i$. Baird and Biedenharn[2] demonstrated that the eigenvalue of the invariant operator I_3 for $SU(3)$ is J_3 and further said that: "Racah has asserted that the result we have just demonstrated for $SU(3)$ is true in general." So that, the eigenvalue of I_n should be simply J_n. The purpose of the present paper is, as suggested by Baird and Biedenharn,[2] to show by a purely infinitesimal proof that this result is only true for the I_2 and I_3 invariants of $SU(n)$ and breaks down for the higher order invariants. The correct formulas have been derived for I_4, I_5, and I_6.

The work of this paper is arranged in the following order. In Sec. 2, we summarize the results concerning the explicit construction of invariants in $SU(n)$ by setting the notations and giving all the expressions which are required for our purpose. In Sec. 3, we give an expression for the invariant I_k of the $SU(n)$ group derived in the fundamental representation and show that, in the particular case of the fundamental representation, the formula proposed by Baird and Biedenharn is not verified for the I_4 invariant. In Sec. 4, we develop the full calculation and give the formulas which actually have to be used to obtain the eigenvalues of the first invariants up to I_6.

2. SUMMARY

A. The Algebra of *SU(n)*

In order to properly parametrize the $SU(n)$ group, it is better to let the null trace condition naturally appear and not as a supplementary condition. In this view, Biedenharn[1] defines a new basis for the diagonal elements h_i, by inserting the real coefficients $\lambda_l^{(i)}$. We define here these coefficients, as a particular solution for the following equations:

$$\sum_{l=1}^{n} \lambda_l^{(i)} \lambda_l^{(j)} = \delta_{ij}, \qquad \sum_{i=0}^{n-1} \lambda_l^{(i)} \lambda_{l'}^{(i)} = \delta_{ll'}, \tag{2a}$$

$$\sum_{l=1}^{n} \lambda_l^{(i)} = 0 \text{ for } i \neq 0, \qquad \lambda_l^{(0)} = \frac{1}{\sqrt{n}} \text{ for any } l, \tag{2b}$$

$$\lambda_1^{(i)} \geq 0 \text{ for any } i. \tag{2c}$$

Condition (2c) is a subsidiary condition of interest only in making the first component of the weight vector positive.

The $n-1$ diagonal elements $h_i (i = 1, 2, \ldots, n-1)$ of $SU(n)$ of null trace are written

$$h_i \equiv (2n)^{-1/2} \sum_{l=1}^{n} \lambda_l^{(i)} e_{ll}. \tag{3a}$$

We define for convenience the element h_0 similarly,

$$h_0 \equiv (2n)^{-1/2} \sum_{l=1}^{n} \lambda_l^{(0)} e_{ll} = 2^{-1/2}\, n^{-1}\, I. \tag{3b}$$

systems have the same asymptotic behavior. This conclusion is illustrated by the remarkable similarity that exists between the asymptotic formulas for the eigenvalues of Mathieu's equation and of Schrödinger's equation with Morse's potential, two problems which at first sight seem totally unconnected. The solution is obtained by perturbation theory; the first order perturbation eigenvalues have already quite a remarkable accuracy, while the eigenfunctions are obtained accurately only in a thin layer about the domain center where they have their nonnegligible values.

[1] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Wiley-Interscience, New York, 1962), Vol. I, pp. 291 and 391.
[2] G. Birkhoff and G.-C. Rota, *Ordinary Differential Equations* (Ginn-Blaisdell, Waltham, Mass., 1962), 2nd ed., Chap. 10.
[3] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Nonrelativistic Theory* (Addison-Wesley, Reading, Mass., 1958), Chap. 6.
[4] R. H. Dicke and J. P. Wittke, *Introduction to Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1960), Chap. 14.
[5] S. Goldstein, Cambridge Phil. Soc. Trans. **23**, 303 (1927).
[6] E. L. Ince, Proc. Roy. Soc. (Edinburgh) **46**, 316 (1926).
[7] E. L. Ince, Proc. Roy. Soc. (Edinburgh) **47**, 294 (1927).
[8] M. Abramowitz and I. A. Stegun (Eds.), *Handbook of Mathematical Functions* (Dover, New York, 1965), Chap. 20.
[9] J. Canosa, J. Comput. Phys. **7**, 255 (1971).
[10] R. Sips, Trans. Amer. Math. Soc. **66**, 93 (1949).
[11] R. Sips, Trans. Amer. Math. Soc. **90**, 340 (1959).
[12] P. M. Morse, Phys. Rev. **34**, 57 (1929).
[13] N. W. McLachlan, *Theory and Application of Mathieu Functions* (Dover, New York, 1964).
[14] J. D. Cole, *Perturbation Methods in Applied Mathematics* (Ginn-Blaisdell, Waltham, Mass., 1968), p. 14.
[15] G. Blanch, Trans. Amer. Math. Soc. **97**, 357 (1960).
[16] J. Canosa and R. G. de Oliveira, J. Comput. Phys. **5**, 188 (1970).

On the Eigenvalues of the Invariant Operators of the Unitary Unimodular Group *SU(n)*

A. Partensky

Institut de Physique Nucléaire, Université Claude Bernard de Lyon (Institut National de Physique Nucléaire et de Physique des Particules), 69-Villeurbanne, France

(Received 18 June 1971; Revised Manuscript Received 29 November 1971)

By a purely infinitesimal method we derive the eigenvalues of the Biedenharn's invariant operators from I_2 to I_6 for the $SU(n)$ group. We show that the general formula, quoted from the Racah's work by Baird and Biedenharn, to obtain the eigenvalues of the invariant operators, is valid only for the two first I_2 and I_3 operators and not for the higher order invariants. We give the correct values for the first invariants till I_6.

1. INTRODUCTION

In a paper published in 1963, Biedenharn[1] has given an explicit construction of the independent invariants for the unimodular unitary group in n dimensions $SU(n)$. He proves that their invariants form a complete system of independent invariants suitable for uniquely labeling the irreducible inequivalent representations. In a second paper Baird and Biedenharn[2] pointed out that the evaluation of the invariant may be inferred from the form of the terms K_n in I_n involving the only h_i. If we denote the highest weight by vector **p** and by **R** the vector whose components are one half the sum of positive roots, we can define

$$J_n(\mathbf{p}) = K_n(\mathbf{p} + \mathbf{R}) - K_n(\mathbf{R}). \tag{1}$$

$K_n(\Lambda)$ is that part of the nth invariant I_n involving the h_i with $h_i \to \Lambda_i$. Baird and Biedenharn[2] demonstrated that the eigenvalue of the invariant operator I_3 for $SU(3)$ is J_3 and further said that: "Racah has asserted that the result we have just demonstrated for $SU(3)$ is true in general." So that, the eigenvalue of I_n should be simply J_n. The purpose of the present paper is, as suggested by Baird and Biedenharn,[2] to show by a purely infinitesimal proof that this result is only true for the I_2 and I_3 invariants of $SU(n)$ and breaks down for the higher order invariants. The correct formulas have been derived for I_4, I_5, and I_6.

The work of this paper is arranged in the following order. In Sec. 2, we summarize the results concerning the explicit construction of invariants in $SU(n)$ by setting the notations and giving all the expressions which are required for our purpose. In Sec. 3, we give an expression for the invariant I_k of the $SU(n)$ group derived in the fundamental representation and show that, in the particular case of the fundamental representation, the formula proposed by Baird and Biedenharn is not verified for the I_4 invariant. In Sec. 4, we develop the full calculation and give the formulas which actually have to be used to obtain the eigenvalues of the first invariants up to I_6.

2. SUMMARY

A. The Algebra of *SU(n)*

In order to properly parametrize the $SU(n)$ group, it is better to let the null trace condition naturally appear and not as a supplementary condition. In this view, Biedenharn[1] defines a new basis for the diagonal elements h_i, by inserting the real coefficients $\lambda_l^{(i)}$. We define here these coefficients, as a particular solution for the following equations:

$$\sum_{l=1}^{n} \lambda_l^{(i)} \lambda_l^{(j)} = \delta_{ij}, \qquad \sum_{i=0}^{n-1} \lambda_l^{(i)} \lambda_{l'}^{(i)} = \delta_{ll'}, \tag{2a}$$

$$\sum_{l=1}^{n} \lambda_l^{(i)} = 0 \text{ for } i \neq 0, \qquad \lambda_l^{(0)} = \frac{1}{\sqrt{n}} \text{ for any } l, \tag{2b}$$

$$\lambda_1^{(i)} \geq 0 \text{ for any } i. \tag{2c}$$

Condition (2c) is a subsidiary condition of interest only in making the first component of the weight vector positive.

The $n-1$ diagonal elements $h_i (i = 1, 2, \ldots, n-1)$ of $SU(n)$ of null trace are written

$$h_i \equiv (2n)^{-1/2} \sum_{l=1}^{n} \lambda_l^{(i)} e_{ll}. \tag{3a}$$

We define for convenience the element h_0 similarly,

$$h_0 \equiv (2n)^{-1/2} \sum_{l=1}^{n} \lambda_l^{(0)} e_{ll} = 2^{-1/2}\, n^{-1}\, I. \tag{3b}$$

h_0 is not a generator of $SU(n)$, but a multiple of unity and is the generator of the $U(1)$ group.

The $n(n-1)$ remaining operators keep Weyl's form, except for what concerns the normalization. They are called conventionally e_α and $e_{-\alpha}$, where α denotes the index pair (ij) and $-\alpha$ the indexing pair (ji) with $i<j$, respectively:

$$e_\alpha \equiv (2n)^{-1/2}\, e_{ij}, \qquad e_{-\alpha} \equiv (2n)^{-1/2}\, e_{ji}, \qquad i<j. \tag{4}$$

Biedenharn[1] defines $\chi_l^{(i)}$ as a particular solution of (2). which can be written

$$\chi_l^{(i)} = (-1)^{n-l+i} \times C\left(\frac{n-1}{2}, \frac{n-1}{2}, i; l-\frac{n+1}{2}, -l+\frac{n+1}{2}\right). \tag{5}$$

The phases $(-1)^{n-l}$ and $(-1)^i$ prove that Conditions (2b) and (2c) are, respectively, well fulfilled.

The definition of the $\chi_l^{(i)}$ with the help of Clebsch–Gordan's coefficients as used by Biedenharn, has the advantage of giving an explicit realization for the $\chi_l^{(i)}$ which allows one to define a canonical basis for the Lie algebra of $SU(n)$.

What is important, though, is not the explicit form of these coefficients, but Properties (2a) and (2b) they verify. The definition we take is less restraining, so that the expressions which will be valid for any particular realization of the $\chi_l^{(i)}$. This is an advantage because the Clebsch–Gordan coefficients do not always give the simplest and most interesting realization for a given $SU(n)$ group in any precise physical problem.

Let us write collectively x_A the n^2-1 generators of $SU(n)$. These generators x_A have been defined with the help of the $n\times n$ matrices which correspond to the fundamental representation. In this representation we can write the anticommutation relation

$$[x_A, x_B]_+ = \frac{\sqrt{2}}{n} h_0 \delta_A^{\bar{B}} + \sum_C [AB^C]\, x_C, \tag{6}$$

which defines the vectorial coupling coefficients $[AB^C]$.

Let us now denote by capital letters X_A (collectively for E_α and H_i), the generators in any representation, the small letters being reserved for the fundamental representation.

The metric tensor g_{AB} has the canonical form $g_{AB} = \delta_A^{\bar{B}}$ which is useful for raising or lowering the indices

$$\sum_M [AB^M]\, g_{MC} = [ABC],$$

For the $SU(n)$ group, the (AB^C) structure constants are given by the following equations which can be easily verified:

$$(ijk) = 0, \tag{7a}$$

$$(i\alpha-\alpha) = (2n)^{-1/2}(\chi_l^{(i)} - \chi_m^{(i)}), \qquad \alpha=(l,m), \tag{7b}$$

$$(\alpha\beta\gamma) = (l_1m_1, l_2m_2, l_3m_3) = (2n)^{-1/2}(\delta_{m_1l_2}\delta_{m_2l_3}\delta_{m_3l_1} - \delta_{l_1m_2}\delta_{l_2m_3}\delta_{l_3m_1}); \tag{7c}$$

and the vector coupling coefficients $[ABC]$ are given by the following:

$$[ijk] = \sqrt{\frac{2}{n}} \sum_{l=1}^{n} \chi_l^{(i)}\chi_l^{(j)}\chi_l^{(k)}, \tag{8a}$$

$$[i\alpha-\alpha] = (2n)^{-1/2}(\chi_l^{(i)} + \chi_m^{(i)}), \qquad \alpha=(l,m), \tag{8b}$$

$$[\alpha\beta\gamma] = [l_1m_1, l_2m_2, l_3m_3] = (2n)^{-1/2}[\delta_{m_1l_2}\delta_{m_2l_3}\delta_{m_3l_1} + \delta_{l_1m_2}\delta_{l_2m_3}\delta_{l_3m_1}]. \tag{8c}$$

(ABC) is totally antisymmetric, whereas $[ABC]$ is totally symmetric in all the indices.

We call root vector the vector $\alpha(\alpha_1, \alpha_2, \ldots, \alpha_{n-1})$ with components $\alpha_i = (i\alpha^\alpha)$ in the H_i space. As i is a vector, the components of which are all null (i = 0), we show easily that we have $(ABC) = 0$ and $[ABC] = 0$ except if $\mathbf{A} + \mathbf{B} + \mathbf{C} = 0$.

B. The $n-1$ Invariants of $SU(n)$

Biedenharn has shown that the set of operators

$$I_{k+1} = \sum_{AB} g_{AB}\, X_A\, T_B^{(k)}, \qquad k \text{ varying from 1 to } n-1 \tag{9}$$

with

$$T_A^{(1)} = X_A \quad \text{and} \quad T_C^{(k)} = \sum_{AB} [AB^C]\, X_A\, T_B^{(k-1)}$$

form the complete set of the $n-1$ invariants suitable for labeling the irreducible inequivalent representations of $SU(n)$.[3] Casimir's construction guaranties that I_{k+1} is an invariant if $T_B^{(k)}$ is a vector under $SU(n)$.

This is shown because of the relation

$$\sum_D [BC^D](AD^E) + [CA^D](BD^E) + [AB^D](CD^E) = 0, \tag{10}$$

which is obvious if we write

$$[x_A,[x_B,x_C]_+]_- + [x_B,[x_C,x_A]_+]_- + [x_C,[x_A,x_B]_+]_- \equiv 0. \tag{11}$$

We calculate the eigenvalues of these operators by application to the higher-weight vector denoted $|\bar{p}\rangle$. Applied to this vector, a raising operator gives zero

$$E_\alpha\,|\bar{p}\rangle = 0. \tag{12a}$$

The $T_A^{(k)}$ being also a vector, we have therefore in a similar way

$$T_\alpha^{(k)}\,|\bar{p}\rangle = 0. \tag{12b}$$

The diagonal operators (or operators of weight H_i) have p_i for eigenvalue

$$H_i\,|\bar{p}\rangle = p_i\,|\bar{p}\rangle. \tag{13}$$

In the case of the fundamental representation in which we note $|\text{F.R.}\rangle$, the vector of higher weight, we have

$$H_i\,|\text{F.R.}\rangle = (1/\sqrt{2n})\,\lambda_1^{(i)}\,|\text{F.R.}\rangle. \tag{14}$$

Let R be the vector equal to half the sum of the positive roots

$$R_i = \tfrac{1}{2}\sum_{\alpha^+} (\alpha - \alpha^i). \tag{15}$$

We can now calculate the eigenvalue of the Casimir's invariant

$$I_2 = \sum_{AB} g_{AB} X_A X_B = \sum_A X_A X_{-A}.$$

The calculation is very classic[4] but we give it again because it is a very clear illustration of the method used to calculate the eigenvalue of the invariants of higher order:

$$I_2 |\bar{p}\rangle = \left(\sum_i H_i^2 + \sum_{\alpha^+} E_\alpha E_{-\alpha}\right)|\bar{p}\rangle,$$

$$\sum E_\alpha E_{-\alpha}|\bar{p}\rangle = \left[\sum_{\alpha^+} E_{-\alpha}E_\alpha + \sum_{\alpha^+,i} (\alpha - \alpha i)H_i\right]|\bar{p}\rangle$$

$$= \sum_i 2R_i H_i |\bar{p}\rangle$$

or still with (1)

$$I_2 |\bar{p}\rangle = J_2 |\bar{p}\rangle. \tag{16}$$

3. EIGENVALUE OF THE INVARIANTS IN THE FUNDAMENTAL REPRESENTATION

In the fundamental representation we can do a direct evaluation of the invariant operators. The generators verify the anticommutation relation (6) which allows us to reduce the degree of the operators and finally express them as a multiple of the identity h_0.

Indeed, we have

$$x_A x_B = \tfrac{1}{2}\left(n^{-2}\delta_A^{-B} + \sum_D \{[AB^D] + (AB^D)\} x_D\right) \tag{17}$$

and the three following obvious relations

$$\sum_{AB} [AB^C]\,\delta_A^{-B} = 0,$$

$$\sum_{AB} [AB^C](AB^D) = 0,$$

$$\sum_{AB} [AB^C][AB^D] = [(n^2 - 4)/n^2]\,\delta_C^{-D}.$$

Let us consider the vector $t_A^{(k)}$. The above expressions give

$$t_C^{(2)} = \sum_{AB} [AB^C]\, x_A x_B = \tfrac{1}{2}[(n^2 - 4)/n^2]\, x_C,$$

which can be generalized as

$$t_C^{(k)} = [(n^2 - 4)/2n^2]^{k-1} x_C \tag{18}$$

and for the invariant of order $k + 1$ we have

$$I_{k+1} = \frac{n^2 - 1}{2n^2}\left(\frac{n^2 - 4}{2n^2}\right)^{k-1}. \tag{19}$$

This formula allows us to test the generalization of (16) by evaluating J_n in the fundamental representation. For that, we first establish a simple expression for the sum of the positive roots.

We consider the j vectors α^j, (j varying from 1 to $n - 1$) of components α_i^j (i varying from 1 to $n - 1$) defined by

$$\alpha_i^j = (2n)^{-1/2}(\lambda_j^{(i)} - \lambda_{j+1}^{(i)}). \tag{20}$$

We suppose that the i indices are ordered in such a way that the α_j roots are positive, and thus the first nonvanishing component of $\alpha^j(\alpha_1^j, \alpha_2^j, \ldots, \alpha_{n-1}^j)$ is positive. In the same way, we suppose that the j indices are chosen so that the α^j roots are in a decreasing order $\alpha^1 \geqslant \alpha^2 \geqslant \cdots \geqslant \alpha^{n-1}$.

After that, we can see that the α^j define the set of $n - 1$ simple roots of $SU(n)$ and R_i defined in (15) is written

$$R_i = \tfrac{1}{2}(2n)^{-1/2} \sum_{m>l} (\lambda_l^{(i)} - \lambda_m^{(i)}).$$

Some elementary manipulations on the summation transform it into

$$R_i = \tfrac{1}{2}(2n)^{-1/2} \sum_{k=1}^{n-1} k(n-k)(\lambda_k^{(i)} - \lambda_{k+1}^{(i)}). \tag{21a}$$

So we express one half the sum of the positive roots in terms of the simple roots of the group. By manipulating the summation, it can be still written

$$R_i = (2n)^{-1/2} \sum_{l=1}^{n} \gamma_l \lambda_l^{(i)} \tag{21b}$$

with

$$\gamma_l = \tfrac{1}{2}(n+1) - l. \tag{21c}$$

From now on, except where otherwise stated, we will always make the summation convention on the repeated indices, with the i indices (or $\lambda_i^{(i)}$ higher indices) varying from 1 to $n - 1$ and the l indices varying from 1 to n. The orthogonality relations (2a) are then written

$$\lambda_l^{(i)}\lambda_l^{(j)} = \delta_{ij}, \qquad \lambda_l^{(i)}\lambda_{l'}^{(i)} = \delta_{ll'} - 1/n. \tag{22}$$

The use of (15) and (21) gives

$$(h_i + R_i)|\text{F.R.}\rangle = (2n)^{-1/2}(\lambda_1^{(i)} + \gamma_\alpha \lambda_\alpha^i)|\text{F.R.}\rangle. \tag{23}$$

We see that the sum of the γ_l of odd powers vanishes,

$$\sum_{l=1}^{n} \gamma_l^q = [1 + (-1)^q] \sum_{\gamma=\epsilon}^{(n-1)/2} \gamma^q, \tag{24}$$

with $\epsilon = 1$ if n is odd and $\epsilon = \frac{1}{2}$ if n is even.

We can now derive the J_3 operator in the fundamental representation

$$J_3 = [ijk](h_i + R_i)(h_j + R_j)(h_k + R_k) - [ijk]R_i R_j R_k \tag{25}$$

and obtain

$$J_3 |\text{F.R.}\rangle = \left(\frac{n^2 - 1}{2n^2}\right)\left(\frac{n^2 - 4}{2n^2}\right)|\text{F.R.}\rangle, \tag{26}$$

which is in accordance with (19).

For J_4, the analogous calculation does not present any theoretical difficulty, but is long and rather tedious, and gives

$$J_4 |\text{F.R.}\rangle = \left(\frac{n^2 - 1}{2n^2}\right)\left[\left(\frac{n^2 - 4}{2n^2}\right)^2 + \frac{1}{6}\left(\frac{n^2 - 4}{2n^2}\right)\right]|\text{F.R.}\rangle. \tag{27}$$

This form is not in accordance with (19) and thus the generalization (16) is certainly unvalid for I_4.

4. EVALUATION OF THE INVARIANTS IN THE CASE OF AN IRREDUCIBLE NONEQUIVALENT REPRESENTATION OF HIGHER WEIGHT $|\bar{p}\rangle$

To make the notation easier to handle, we define

$$S_j^i \equiv \tfrac{1}{2} \sum_{\alpha^+} \{[\alpha - \alpha_i](\alpha - \alpha_j) + [\alpha - \alpha_j](\alpha - \alpha_i)\}, \tag{28a}$$

$$A_j^i \equiv \tfrac{1}{2}\sum_{\alpha^+}\{[\alpha-\alpha_i](\alpha-\alpha_j)-[\alpha-\alpha_j](\alpha-\alpha_i)\}, \quad (29a)$$

$$\Lambda_j^i \equiv [ijk](h_k + R_k).$$

S_j^i and A_j^i are expressed in terms of the $\lambda_l^{(i)}$,

$$S_j^i = \frac{1}{n}\sum_{l=1}^{n}\gamma_l \lambda_l^{(i)}\lambda_l^{(j)}, \quad (28b)$$

$$A_j^i = \frac{1}{2n}\sum_{q=1}^{n}\sum_{k=1}^{q}(\lambda_q^{(i)}\lambda_k^{(j)} - \lambda_k^{(i)}\lambda_q^{(j)}). \quad (29b)$$

There exist many equivalent ways in writing A_j^i and the two following relations, very useful in the calculation, are used here:

$$\sum_{q=1}^{n}\sum_{k=1}^{q}(\varphi_q\Psi_k + \varphi_k\Psi_q) = \sum_{q=1}^{n}\varphi_q\Psi_q + \sum_{q=1}^{n}\sum_{k=1}^{n}\varphi_q\Psi_k, \quad (30a)$$

$$\sum_{q=1}^{n}\sum_{k=1}^{q}\varphi_q\Psi_k = \sum_{q=1}^{n}\sum_{k=q}^{n}\varphi_k\Psi_q \quad (30b)$$

for any function φ_k and Ψ_k.

S_j^i and Λ_j^i are symmetric in the indices i and j exchanges, whereas A_j^i is antisymmetric.

We substitute in (10) $B = i$, $C = j$, $A = \alpha$, $E = \alpha$, and adding α^+ we obtain

$$S_j^i = [ijk]R_k. \quad (28c)$$

With these notations, we apply (9) to the higher weight $|\bar{p}\rangle$ to obtain

$$I_{n+1}|\bar{p}\rangle = (H_i + 2R_i)t_i^{(n)}|\bar{p}\rangle, \quad (31a)$$

$$T_i^{(k)}|\bar{p}\rangle = (\Lambda_j^i + A_j^i)T_j^{(k-1)}|\bar{p}\rangle \quad (31b)$$

that is,

$$I_{n+1}|\bar{p}\rangle = (H_i + 2R_i)(\Lambda_j^i + A_j^i)\cdots(\Lambda_l^k + A_l^k)H_l|\bar{p}\rangle \quad (32)$$

with $n-1$ factors $(\Lambda_l^j + A_l^j)$.

The symmetric part is designated by $[\mathfrak{S}]_l^i$ and the antisymmetric part by $[\mathfrak{A}]_l^i$ under the exchange of the indices i and l in the $(n-1)$ factors $(\Lambda_\beta^\alpha + A_\beta^\alpha)$:

$$(\Lambda_j^i + A_j^i)(\Lambda_\alpha^j + A_\alpha^j)\cdots(\Lambda_l^k + A_l^k) = [\mathfrak{S}]_l^i + [\mathfrak{A}]_l^i. \quad (33)$$

This allows us to write (33) in the following form:

$$I_{n+1}|\bar{p}\rangle = \{[(H_i + R_i)(H_l + R_l) - R_iR_l][\mathfrak{S}]_l^i + 2R_i(H_l + R_l) \times [\mathfrak{A}]_l^i\}|\bar{p}\rangle. \quad (34)$$

Now we must develop the quantities $[\mathfrak{S}]_l^i$ and $[\mathfrak{A}]_l^i$ for each invariant.

A. I_3 Invariant of $SU(n)$

We have

$$[\mathfrak{S}]_l^i = \Lambda_l^i \quad \text{and} \quad [\mathfrak{A}]_l^i = A_l^i,$$

which gives

$$I_3|\bar{p}\rangle = \{(H_i + R_i)\Lambda_l^i(H_l + R_l) - (2A_j^i + S_j^i) \times R_j(H_i + R_i)\}|\bar{p}\rangle. \quad (35)$$

with (21), (28), and (29) it can easily be shown that

$$(2A_j^i + S_j^i)R_j = 0. \quad (36a)$$

Remark: It must be emphasized again that in this formula, the repeated j index summation rules from 1 to $n-1$.

We also have

$$R_iS_j^iR_j = 0. \quad (36b)$$

With (36) we then write for I_3,

$$I_3|\bar{p}\rangle = \{(H_i + R_i)\Lambda_j^i(H_j + R_j) - R_iS_j^iR_j\}|\bar{p}\rangle$$

or still

$$I_3|\bar{p}\rangle = J_3|\bar{p}\rangle, \quad (37)$$

which generalizes to $SU(n)$ which Baird and Biedenharn[2] found in the particular case of $SU(3)$.

B. I_4 Invariant of $SU(n)$

We are going to develop the calculation for this invariant because it is typical and illustrates the process of calculation of any higher order invariant. We have

$$[\mathfrak{S}]_l^i = \Lambda_j^i\Lambda_l^j + A_j^iA_l^j,$$
$$[\mathfrak{A}]_l^i = \Lambda_j^iA_l^j + A_j^i\Lambda_l^j.$$

Bringing back these values in (34) and using (28) and (36), we obtain

$$I_4|\bar{p}\rangle = \{J_4 + (H_i + R_i)(H_l + R_l)[(A_j^i + S_j^i)(A_l^j - S_l^i) + R_\alpha S_j^\alpha[jil]] + \tfrac{5}{4}R_iS_j^iS_l^iR_l\}|\bar{p}\rangle. \quad (38)$$

We easily calculate

$$\tfrac{5}{4}R_iS_j^iS_l^jR_l = \frac{n^2-4}{12n^2}R_i^2 \quad \text{and} \quad R_i^2 = \frac{n^2}{12}\left(\frac{n^2-1}{2n^2}\right). \quad (39)$$

We shall now consider the following expression:

$$(A_j^i + S_j^i)(A_l^j - S_l^j) + R_\alpha S_j^\alpha[jil]. \quad (40)$$

With the help of (30) we can write for the term $A_j^iA_l^j$,

$$A_j^iA_l^j = \left(\frac{1}{2n}\delta_{ij} - \frac{1}{n}\sum_{q=1}^{n}\sum_{k=1}^{q}\lambda_k^{(i)}\lambda_q^{(j)}\right) \times \left(\frac{1}{2n}\delta_{jl} - \frac{1}{n}\sum_{q=1}^{n}\sum_{k=1}^{q}\lambda_k^{(j)}\lambda_q^{(l)}\right).$$

We now sum on j and use (30b) which can be written here as

$$\sum_{q=1}^{\alpha}\sum_{k=1}^{q}\lambda_k^{(i)} = \sum_{q=1}^{\alpha}(\alpha - q + 1)\lambda_q^{(i)}$$

and we obtain

$$A_j^iA_l^j = \frac{1}{4n^2}\delta_{il} - \frac{1}{2n^2}\sum_{q=1}^{n}\sum_{k=1}^{q}(\gamma_q - \gamma_k)(\lambda_q^{(i)}\lambda_k^{(l)} + \lambda_k^{(i)}\lambda_q^{(l)}) + \frac{2}{n^2}R_iR_l.$$

In a similar way we calculate the $A_j^iS_l^j$, $S_j^iA_l^j$, $S_j^iS_l^j$, and $R_\alpha S_j^\alpha[jil]$ terms and setting the results in (40), we find

$$(A_j^i + S_j^i)(A_l^j - S^j{}_l + R_\alpha S_j^\alpha[jil] = -\frac{n^2-4}{12n^2}\delta_l^i. \quad (41)$$

Bringing (39) and (41) in (38) we finally have

$$I_4|\bar{p}\rangle = \left(J_4 - \frac{n^2-4}{12n^2} J_2\right)|\bar{p}\rangle \tag{42}$$

which verifies (19). A corrective J_2 term in addition to J_4 appears.

C. I_5 invariant of $SU(n)$

The calculation is analogous to the previous one. We demonstrate the following formulas,

$$\{(A_j^k + S_j^k)[j\alpha\beta] - (A_j^\alpha + S_j^\alpha)[jk\beta]\}(A_k^\gamma + S_k^\gamma) + \{(A_j^k + S_j^k)[j\alpha\gamma] - (A_j^\gamma + S_j^\gamma)[jk\alpha]\}(A_k^\beta + S_k^\beta) + \{(A_j^k + S_j^k)[j\beta\gamma] - (A_j^\beta + S_j^\beta)[jk\gamma]\}(A_k^\alpha + S_k^\alpha) = \frac{1}{n^2}[\alpha\beta\gamma], \tag{43a}$$

$$R_i S_j^i[j\alpha k] = S_j^\alpha S_k^j + \frac{2}{n^2} R_\alpha R_k - \frac{n^2-1}{12n^2}\delta_{\alpha k}, \tag{43b}$$

$$(4A_j^i + S_j^i) S_l^j S_k^l R_k = \frac{n^2-13}{12n^2} S_k^i R_k, \tag{43c}$$

and obtain, after some calculation the eigenvalue of I_5:

$$I_5|\bar{p}\rangle = \{J_5 - [(n^2-8)/12n^2] J_3\}|\bar{p}\rangle. \tag{44}$$

D. Invariant I_6 of $SU(n)$

The calculation becomes very cumbersome and tedious for this invariant, so it will be omitted, but we indicate an indirect method which allows us to reach the result in an easier way. Indeed, I_6 expression, from (34), can be written

$$I_6|\bar{p}\rangle = \{J_6 + \mathcal{L}_1(\alpha\beta\mu\delta)(H_\alpha + R_\alpha)(H_\beta + R_\beta)(H_\mu + R_\mu) \times (H_\delta + R_\delta) + \mathcal{L}_2(\alpha\beta)(H_\alpha + R_\beta) + \mathcal{L}_3\}|\bar{p}\rangle. \tag{45a}$$

We will not give the long expressions of $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$. On the other hand we write *a priori* for I_6:

$$I_6|\bar{p}\rangle = \{J_6 + h(n)J_4 + g(n)J_2^2 + f(n)J_2\}|\bar{p}\rangle, \tag{45b}$$

where $h(n)$, $g(n)$, and $f(n)$ are three functions which we want to determine. By identifying [(45a) and (45b)] we have

$$\mathcal{L}_1(\alpha\beta\mu\delta) = h(n)[\alpha\beta j][j\mu\delta] + g(n)\delta_{\alpha\beta}\delta_{\mu\delta}, \tag{46a}$$

$$\mathcal{L}_2(\alpha\beta) = f(n)\delta_{\alpha\beta}. \tag{46b}$$

Multiplying the two members of (46a) by the following expression which brings in the symmetry the indices α, β, μ, and δ,

$$\tfrac{1}{4}(R_\alpha R_\beta R_\mu \delta_{\delta\gamma} + R_\delta R_\alpha R_\beta \delta_{\mu\gamma} + R_\mu R_\delta R_\alpha \delta_{\beta\gamma} + R_\beta R_\mu R_\delta \delta_{\alpha\gamma}). \tag{47}$$

We obtain, after all calculations,

$$-\frac{n^2-13}{12n^2} R_\alpha S_j^\alpha S_\gamma^j - \frac{1}{36}\left(\frac{n^2-1}{2n^2}\right)\left(\frac{n^2-4}{2n^2}\right) R_\gamma = h(n) R_\alpha S_j^\alpha S_\gamma^j + \frac{n^2-1}{24} g(n) R_\gamma$$

and

$$h(n) = -\frac{n^2-13}{12n^2}, \quad g(n) = -\frac{n^2-4}{6n^4}. \tag{48}$$

In the same way we calculate $f(n)$ and finally obtain for the I_6 eigenvalue

$$I_6|\bar{p}\rangle = \left\{J_6 - \frac{n^2-13}{12n^2} J_4 - \frac{n^2-4}{6n^4} J_2^2 - \frac{(n^2-4)(7n^2+8)}{720n^4} J_2\right\}|\bar{p}\rangle. \tag{49}$$

In conclusion, we must remark that we have not given the general eigenvalue of the I_k invariant of $SU(n)$ as function of the $K_{k'}$ $(k' \leq k)$. We have limited our work to the I_6 invariant which, in fact includes all the interesting physical problems. For higher invariant operators, it would be easier to express the Biedenharn's invariant in terms of the Gel'fand[5] invariants whose eigenvalues are known.[6]

---

1 L. C. Biedenharn, J. Math. Phys. **4**, 436 (1963).
2 G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
3 A. Klein, J. Math. Phys. **4**, 1283 (1963); U. Fano in Appendix A to G. E. Baird and L. C. Biedenharn, J. Math. Phys. **5**, 1723 (1964); B. Gruber and L. O'Raifeartaigh, J. Math. Phys. **5**, 1796 (1964); M. Umezawa, Nucl. Phys. **48**, 111 (1963); M. Umezawa, Nucl. Phys. **53**, 54 (1964); M. Micu, Nucl. Phys. **60**, 353 (1964); L. Egardt, Arkiv Fysik **27**, 193 (1964).
4 G. Racah, "Group Theory and Spectroscopy", Reprint of Lectures delivered at the Institute for Advanced Study, Princeton (1951).
5 I. M. Gel'fand, Mat. Sb **26**, 103 (1950)
6 J. D. Louck, Amer. J. Phys. **38**, 3 (1970); J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970) cf. Appendix B, p. 2401; A. M. Perelomov and V. S. Popov, Yadernaya Fizika **3**, 924 (1966) [Sov. J. Nucl. Phys. **3**, 676 (1966)]; *ibid.* **5**, 693 (1967) [*ibid.* **5**, 489 (1967)]; *ibid.* **7**, 460 (1968) [*ibid.* **7**, 290 (1968)].

The Spin Spectrum of an Unstable Particle

Gordon N. Fleming
Department of Physics, The Pennsylvania State University, University Park, Pennsylvania 16802
(Received 27 September 1971; Revised Manuscript Received 13 December 1971)

A state vector (vector-valued distribution) of an unstable particle can be a 3-momentum eigenvector in at most one inertial frame as a consequence of the necessary width of the energy spectrum. We investigate this frame dependence of momentum eigenvectors of unstable particles and demonstrate that:

(i) It is compatible with the principle of relativity.

(ii) It leads to a distinction between two ways of defining the spin of the particle which are equivalent for stable particles.

(iii) One definition, called kinematical spin, yields a precise value, and is determined by considerations of detailed balancing and other means of counting the degeneracy of the momentum eigenvectors.

(iv) The second definition, called dynamical spin, need not yield a sharp value, is determined by the dynamics of the decay process, and is measured by observing angular distributions of decay products.

It is shown that an unstable particle with vanishing kinematical spin (pion) may have a small admixture of non-isotropic angular distribution of unpolarized decay products in the rest frame of the decay products. The order of magnitude of the effect is estimated and the coupling between the mass and spin spectrum is studied in the case of local interactions without derivative coupling.

1. INTRODUCTION

In this paper I shall consider some of the consequences that follow, in a Lorentz invariant quantum theory, from the proposition that one can, in any inertial frame, identify detect or prepare an unstable system prior to decay at any instant of time. This is the basic premise of this study. At the risk of belaboring an obvious issue I will dwell for a moment on the meaning of the premise.

Consider a neutron. It is conventionally assumed, and we adhere to the position that one can talk meaningfully about the distinction between the probability for finding the neutron, a single particle to exist on one hand, and the probability for finding, not the neutron, but its decay products, a proton electron and antineutrino, to exist on the other. These probabilities change with time but at any time the distinction between the unstable parent particle and its decay products is clear cut. The premise then asserts that for any given instant in any given inertial frame there exist states of this system in which the probability for finding the unstable neutron (or more generally the parent particle) is momentarily unity and the probability for finding the decay products is momentarily zero.

Since a neutron lives so long and since we have no reason to regard a neutron as a composite system of its decay products prior to decay, the application of the premise here seems unobjectionable. The application of the premise becomes more doubtful as the lifetime decreases towards the point where we begin to doubt the reliability of all our spatio–temporal concepts and as the contribution of the decay products to the structure of the parent particle becomes more dominant.

Lurçat[1] has explicitly challenged the application of our premise to the strongly decaying hadronic resonances, denying to them a state vector description altogether and the possibility of ever being detected more directly than through the correlations among the decay products. He gave a quantitative form to this notion by introducing the concept of the *integrity* of an unstable system. If the unstable system has already decayed, its integrity is exactly zero. If it has not decayed, its integrity is exactly unity. For all other situations the integrity is not sharp but has an expectation value lying between zero and one. Lurçat suggests that the hadronic resonances always have integrity zero. This means that the concept of lifetime, as such, does not apply to hadronic resonances and that time dilation of the lifetime of the ultrahigh energy resonances could never render them detectable prior to decay. Such a restriction on the possible outcome of future experiments is very hypothetical and may have the character of raising a current technological limitation to the status of a fundamental principle.

In terms of the integrity our premise may be stated as declaring the existence, for any instant of time in any inertial frame, of states, for any unstable system, with integrity unity. The operator representing the integrity is just that projection operator which projects onto the unstable particle states and annihilates the decay products. Without embracing the premise in its entirety, the reader can regard this paper as a study pertaining to those metastable particles for which the premise is at least an excellent approximation.

The premise has been widely employed in the literature[2] (without the present degree of fanfare!) for discussions of the time dependence of the decay process (deviations from the exponential law, etc.) and the shape of the energy or mass spectrum. Such studies may be construed as a development of the consequences of time-translational invariance applied to the premise. In this paper we shall concentrate on the application of Lorentz invariance to the premise and the primary consequence of this will be the likelihood of a spin spectrum (as opposed to a sharp spin) for an unstable particle manifested in the angular distribution of the decay products. The order of magnitude of the width of the spin spectrum (barring selection rules) will be shown to be roughly given by

$$\Delta S \sim (v/C)\,(\Delta m/m),$$

where m is the expectation value of the rest mass of the unstable particle, Δm the width of the rest mass spectrum, and v the speed of the particle in that frame in which the system has unit integrity at some definite time. The frame dependence of the effect may be disquieting at first, but in fact is not in conflict with the equivalence of inertial frames or the invariance of the theory under the Lorentz group.[3]

In Table I I have listed the values of $\Delta m/m$ for a small selection of unstable systems. Clearly the last four entries, or others like them, would seem to

TABLE I. Ratio of width of mass spectrum to rest mass for several unstable systems.

Unstable system	$\Delta m/m$
neutron	10^{-27}
muon	10^{-17}
optical transitions in atoms	10^{-17}
lambda hyperon	10^{-14}
optical transitions in nuclei	10^{-12}
neutral pion	10^{-7}
neutral sigma hyperon	10^{-6}
excited nuclear levels	10^{-6}
hadron resonances	10^{-1}

provide the best hope of detecting the effect in question. However, the matter is not that simple. Spin assignment experiments involving hadronic resonances and excited nuclear levels are sufficiently dirty, by virtue of an abundance of competing processes, that they proceed by searching for a dominant harmonic contribution to the angular distribution of decay products near the peak of the mass distribution. The quantum number associated with the dominant harmonic is assigned as *the* spin of the system in question, and the residual angular distribution is attributed to the competing processes. Furthermore, theoretical considerations presented here will indicate that the spin spectrum is coupled to the mass spectrum in such a way as to suppress the wings of the spin spectrum at the peak of the mass spectrum. In the case of the neutral pion decay competing processes are not a problem but the Bose statistics of the two photon final state yield a selection rule so that the proper estimate of the width of the spin spectrum is $(v/c)^2\,(\Delta m/m)^2$.[4] The neutral sigma decay does look promising and will be studied in detail in a later paper.

The relevance of these considerations to the phenomena of the decay of the neutral K mesons and the mass spectrum of the A_2 meson will also be considered in subsequent papers.

Thus far the expressions unstable system and unstable particle have been used as more or less interchangeable. For the rest of the paper I would like to introduce a distinction. Henceforth, an *unstable system* is an enduring physical system which at some instant in some inertial frame consists of just one *unstable particle*. If the unstable system is left to itself, then as time passes the probability for the appearance of decay products increases and the probability for the presence of the unstable particle decreases. In the absence of measurements an unstable system can exist forever, but it consists of a single unstable particle for one instant at most. We will always work with Heisenberg picture state vectors which correspond to the entire evolutionary history of the unstable systems under consideration. The unstable system state vectors, however, will frequently be labeled by reference to the circumstances under which the system consists of one unstable particle.

2. MOMENTUM AND SPIN OF UNSTABLE SYSTEMS

Let $|\alpha, t_0\rangle$ be a Heisenberg picture state vector for an unstable system consisting of one unstable particle at time t_0 and in the instantaneous single-particle state denoted by α. Translation invariance asserts the existence of the spatially translated states $|\alpha_{\boldsymbol{\lambda}}, t_0\rangle$ (the physical system is translated by the displacement $\boldsymbol{\lambda}$), all of which also consist of one unstable particle at the time t_0. By superposition we can then construct the eigenvectors of the translation generators which we identify as the momentum eigenvectors. Neglecting internal degrees of freedom, we have

$$|\mathbf{p}, t_0\rangle \propto \int d^3\lambda\, e^{(i/\hbar)\mathbf{p}\cdot\boldsymbol{\lambda}}|\alpha_{\boldsymbol{\lambda}}\ t_0\rangle\,. \tag{2.1}$$

We cannot construct energy eigenvectors for unstable systems by superposing unstable system state vectors obtained by performing time displacements on a given one. Such a superposition could indeed yield an energy eigenvector, but it would not describe a system consisting of one unstable particle at any definite time. Unlike the case for single stable particles, time displacements have dynamical effects on unstable particles even when the latter are in momentum eigenstates. Consequently, unstable systems cannot exist in energy eigenstates.

In itself this is a familiar fact commonly referred to in terms of the energy or mass spectrum of the unstable system. When coupled to the existence of momentum eigenvectors in a Lorentz invariant theory, however, it leads to a result that is perhaps not so familiar and is universally ignored in practical calculations of decay amplitudes. *The Lorentz transform of an unstable system momentum eigenvector is not a momentum eigenvector.*

Under a pure Lorentz transformation the energy and momentum of a system change according to

$$\mathbf{p}'_{\parallel} = (1 - v^2/c^2)^{-1/2}\,(\mathbf{p}_{\parallel} + \mathbf{v}E/c^2), \tag{2.2a}$$

$$\mathbf{p}'_{\perp} = \mathbf{p}_{\perp}, \tag{2.2b}$$

$$E' = (1 - v^2/c^2)^{-1/2}\,(E + \mathbf{v}\cdot\mathbf{p}). \tag{2.2c}$$

If we initially have the uncertainties $\Delta\mathbf{p} = 0, \Delta E \neq 0$, then after the transformation we have

$$\Delta\mathbf{p}'_{\parallel} = (1 - v^2/c^2)^{-1/2}\,\mathbf{v}\Delta E/c^2, \qquad \Delta\mathbf{p}'_{\perp} = 0, \tag{2.3a}$$

$$\Delta E' = (1 - v^2/c^2)^{-1/2}\,\Delta E, \tag{2.3b}$$

and the system does not have sharp momentum in the new frame.

The equations indicate the induced spread in the momentum spectrum is proportional to the original spread in the energy spectrum, and for small Lorentz transformations, $v/c \ll 1$, performed on metastable particle states the induced spread in momentum would be extremely difficult to detect. Nevertheless, for purposes of theoretical analysis the concept of the frame dependence of momentum eigenvectors for unstable systems is important. In particular it implies the *nonexistence of a sharp rest frame for any unstable system with sharply defined nonvanishing momentum* in a given frame. This in turn renders the conventional technique of determining the spin of an unstable particle by transforming the observed angular distribution of decay products to the rest frame of the parent particle approximate in principle. In other words, even if the laboratory momentum of the unstable system is exactly known, the transformation to the rest frame is not sharply defined since it depends on the energy as well, which

cannot be sharp. More generally, two momentum eigenvectors for an unstable system with distinct momentum eigenvalues cannot be transformed into one another by a transformation in the Poincaré group. This is one sense in which the states of an unstable system do not form an irreducible representation space for the Poincaré group.[5]

In passing let us note that there should be no puzzlement over the fact that the transformation equations (2.2) indicate that the energy uncertainty always increases under a Lorentz transformation from the case $\Delta\mathbf{p} = 0$. At first this may seem to conflict with the dilation of the lifetime displayed by energetic metastable particles. Thus from $\Delta E \Delta t \gtrsim h$ we may expect that as the lifetime increases ΔE should decrease. But *lifetimes are well defined only for momentum eigenstates* and comparing two such eigenstates we have

$$E_1 = (\mathbf{p}_1^2 + m^2c^2)^{1/2}c, \qquad E_2 = (\mathbf{p}_2^2 + m^2c^2)^{1/2}c$$

and for the uncertainties

$$\Delta\mathbf{p}_1 = 0, \qquad \Delta E_1 \cong (mc^2/E_1)\,\Delta mc^2,$$
$$\Delta\mathbf{p}_2 = 0, \qquad \Delta E_2 \cong (mc^2/E_2)\,\Delta mc^2.$$

The energy spread of the more energetic momentum eigenstate is the smaller. This consideration indicates again the distinction, in the case of unstable systems between pairs of momentum eigenvectors on one hand and pairs of states connected by a Poincaré transformation on the other. Unlike the stable case the former pair is not a special case of the latter.

As the frame dependence of momentum eigenvectors is pursued further on in this paper, it will come to appear very likely that *in a certain sense* unstable particles do not possess a definite spin any more than they possess a definite mass. The *sense* in which this appears relates to the angular distribution of the decay products in *the rest frame of the decay products*. In fact it will become apparent that for unstable systems one can distinguish two ways of defining the spin of the system which in the limit of infinite lifetime become identical. I have chosen the names *kinematical* and *dynamical spin* for these two definitions since the former depends solely on the degeneracy of the momentum eigenvectors while the latter involves the behavior of the states of the system under the action of the Casimir invariant operators of the Poincaré group. It appears that spin determinations based on detailed balancing arguments and the like, involving as they do the counting of degenerate momentum eigenstates, determine the kinematical spin while the angular distribution of decay products is related to the dynamical spin. In a perturbative theory of the decay process the kinematical spin is identical with *the* spin of the particle in the stable limit while the spectrum of the dynamical spin is determined by the details of the interaction producing the decay. If the kinematical spin is nonzero, then the representation of the Poincaré group carried by the states of the unstable system is *not* multiplicity free.[1]

3. HYPERPLANE-DEPENDENT STATES OF UNSTABLE SYSTEMS

To carry out conveniently a discussion of the consequences of Lorentz invariance applied to the frame-dependent concepts of instantaneous or momentum eigenstates of unstable systems, we must find an invariant way of describing such states and a manifestly covariant way of labeling the state vectors that correspond to them. We do this by following the historical precedent set long ago by Tomonaga and Schwinger[6] and since adopted by many authors for various purposes. We note that an instant of time in a given inertial frame uniquely determines a three-dimensional section of space–time which in any inertial frame may be described as a spacelike hyperplane. The hyperplane is not instantaneous in those frames moving relative to the given one but is instead *tilted* with respect to the time axis. The invariant form of the assertion that an unstable system consists of an unstable particle at some given time in some given inertial frame is that the unstable system consists of an unstable particle on a particular spacelike hyperplane, viz., that hyperplane defined by the given instant in the given frame. The hyperplane itself is an invariant geometrical construct and the association of the unstable system with a particular hyperplane removes the need to employ any *particular* inertial frame in referring to it. The manifestly covariant labeling of the state vectors used in arbitrary inertial frames to describe unstable systems is achieved by employing the timelike unit vector η_μ, which defines the direction normal to the hyperplane in space-time, and the parameter τ, which measures the interval between the hyperplane and the space-time origin of the coordinate system along a line parallel to η_μ. The intersection of the hyperplane with the line in question is at the point $\tau\eta_\mu$, and all the space-time points lying in the hyperplane satisfy the equation of the hyperplane

$$\chi^\mu \eta_\mu = \tau. \tag{3.1}$$

Every spacelike hyperplane appears instantaneous in some inertial frame and in that frame we have $\eta_\mu = \eta_\mu^{(0)} \equiv (1, 0, 0, 0)$ and $\tau = \tau^{(0)} \equiv ct_0$, where t_0 is the time of the instantaneous hyperplane.[7]

Let $|\alpha, t_0\rangle$ denote the Heisenberg picture state vector for an unstable system consisting of one unstable particle at the time $t = t_0$ (the symbol α denotes the information needed to describe the instantaneous configuration of the unstable particle). In a frame related to the given one by the transformation,

$$\chi_\mu = \Lambda_\mu^\nu \chi_\nu^{(0)} + a_\mu, \tag{3.2}$$

the Heisenberg picture state vector for the *same system* is

$$|\alpha(\Lambda, a); \eta, \tau\rangle \equiv U(\Lambda, a)|\alpha, t_0\rangle, \tag{3.3}$$

where $\eta_\mu = \Lambda_\mu^\nu \eta_\nu^{(0)} = \Lambda_\mu^0$ and $\tau = \tau^{(0)} + a^\mu \eta_\mu = ct_0 + a^\mu \eta_\mu$ describe the hyperplane on which the unstable particle exists as viewed in the new frame. The unitary operator $U(\Lambda, a)$ inducing the Poincaré transformation on the Hilbert space is the familiar one,

$$U(\Lambda, a) = e^{iP_\mu a^\mu/\hbar}\, e^{-iM_{\mu\nu}\omega^{\mu\nu}(\Lambda)/2\hbar} \tag{3.4}$$

and the symbol $\alpha(\Lambda, a)$ denotes the modified description of the unstable particles configuration on the hyperplane. From Eq. (3.3), which may be construed

as a definition of the symbol on the left-hand side, and the group property of Poincaré transformations it follows that

$$|\alpha(\Lambda,a);\Lambda\eta,\tau+a\Lambda\eta\rangle = U(\Lambda,a)|\alpha;\eta,\tau\rangle. \tag{3.5}$$

Now consider the special case in which the original state is a momentum eigenstate, i.e., (neglecting internal degrees of freedom)

$$|\alpha,t_0\rangle = |\mathbf{p},t_0\rangle, \tag{3.6}$$

where

$$\mathbf{P}|\mathbf{p},t_0\rangle = \mathbf{p}|\mathbf{p},t_0\rangle. \tag{3.7a}$$

From

$$U^{-1}(\Lambda,a)P_\mu U(\Lambda,a) = \Lambda_\mu^\nu P_\nu \tag{3.8}$$

we find

$$K_\mu(\eta)|k;\eta,\tau\rangle = k_\mu|k;\eta,\tau\rangle, \tag{3.7b}$$

where

$$K_\mu(\eta) \equiv P_\mu - \eta_\mu\eta P, \qquad k_\mu \equiv \sum_{i=1}^{3}\Lambda_\mu^i p_i, \tag{3.9a}$$

$$\eta_\mu = \Lambda_\mu^0, \qquad \tau = ct_0 + a^\mu\eta_\mu \tag{3.9b}$$

and

$$|k;\eta,\tau\rangle \equiv U(\Lambda,a)|\mathbf{p},t_0\rangle\, e^{-ik_\mu a^\mu/\hbar}. \tag{3.10}$$

We see that under Poincaré transformation the instantaneous momentum eigenvector becomes a hyperplane dependent eigenvector of that part of the total 4-momentum that lies in the hyperplane, i.e., $K_\mu(\eta)$. Just as the original state does not have sharp energy, so the transformed state is not sharp in the part of P_μ normal to the hyperplane, i.e., $\eta_\mu\eta P$. I will sometimes refer to ηP as the hyperplane energy or the hyperplane Hamiltonian while $K_\mu(\eta)$ will be referred to as the hyperplane momentum. Note that the 4-vector eigenvalue k_μ has only three independent components since from its definition (3.9a) we have

$$k^\mu\eta_\mu = 0. \tag{3.11}$$

Finally, if we consider a general Poincaré transformation on the vector $|k;\eta,\tau\rangle$, we obtain

$$e^{ia\Lambda k/\hbar}|\Lambda k;\Lambda\eta,\tau+a\Lambda\eta\rangle = U(\Lambda,a)|k;\eta,\tau\rangle. \tag{3.12}$$

With these transformation equations for the hyperplane momentum eigenvectors under our belt, we can be more explicit about the meaning of the cryptic symbol $\alpha(\Lambda,a)$ in Eq. (3.3), (3.5). Again neglecting internal degrees of freedom, suppose

$$|\alpha,t_0\rangle \equiv \int d^3p\ \alpha(\mathbf{p})|\mathbf{p},t_0\rangle. \tag{3.13}$$

Then

$$\begin{aligned}|\alpha(\Lambda,a);\eta,\tau\rangle &= U(\Lambda,a)|\alpha,t_0\rangle\\ &= \int d^3p\ \alpha(\mathbf{p})\,U(\Lambda,a)|\mathbf{p},t_0\rangle\\ &= \int d^3p\ \alpha(\mathbf{p})\,e^{ika/\hbar}|k;\eta,\tau\rangle\\ &= \int d^4k\,\delta(\eta k)\,\alpha(\Lambda^{-1}k)\,e^{ika/\hbar}\,|k;\eta,\tau\rangle,\end{aligned} \tag{3.14}$$

and we see that

$$\alpha_{(\Lambda,a)}(k) \equiv \alpha(\Lambda^{-1}k)\,e^{ika/\hbar}. \tag{3.15}$$

Again from this case where we begin with the instantaneous state we have the general result

$$\begin{aligned}&|\alpha(\Lambda,a);\Lambda\eta,\tau+a\Lambda\eta\rangle\\ &\quad= \int d^4k\ \delta(k\Lambda\eta)\,\alpha_{(\Lambda,a)}(k)|k;\Lambda\eta,\tau+a\Lambda\eta\rangle,\end{aligned} \tag{3.16}$$

where Eq. (3.15) holds.

4. DECAY AMPLITUDES

We will consider the decay of an unstable particle with no internal degrees of freedom into two stable particles also devoid of internal degrees of freedom. In the case of the stable decay products the absence of internal degrees of freedom implies that the particles are spinless. For the unstable parent particle, however, such a conclusion would prejudge one of the major issues under consideration here. In fact we will find that the hyperplane dependence of the unstable system state vector enables the angular distribution of the decay products to deviate from isotropy in the rest frame of the decay products.

Let the initial state be $|k;\eta,\tau\rangle$ and the final state be $\langle p_1,p_2(+)|$, and eigenvector of the total 4-momentum,

$$\langle p_1,p_2(+)|P_\mu = \langle p_1,p_2(+)|(p_1+p_2)_\mu, \tag{4.1}$$

with the (+) signature indicating that in the infinite *future* the system consists of two stable particles with sharp 4-momenta p_1,p_2. The probability amplitude for the decay is

$$\langle p_1,p_2(+)|k;\eta,\tau\rangle.$$

Strictly speaking, of course, the initial state should be normalized by smearing in k about some central k_0 with a sharp distribution. The purpose of the present general investigation, however, will best be served if we concentrate on the idealized amplitude given above, the kernel of the realistic amplitude as it were.

Since both the initial and final states are eigenvectors of $K_\mu(\eta) \equiv P_\mu - \eta_\mu\eta P$, it follows that the amplitude must be proportional to the three-dimensional delta function of the hyperplane momentum. This is a covariant generalization of the Euclidean three-dimensional delta function and is given by

$$\delta_\eta^3(p_1+p_2-k) \equiv (2\pi\hbar)^{-3/2}\int d^4\lambda\ \delta(\eta\lambda)\,e^{i\lambda(p_1+p_2-k)/\hbar}. \tag{4.2}$$

It constrains k to the value

$$k_\mu = (p_1+p_2)_\mu - \eta_\mu(\eta p_1 + \eta p_2). \tag{4.3}$$

The τ dependence of the amplitude can be determined from (4.1) and

$$|k;\eta,\tau\rangle = e^{i\tau\eta P/\hbar}|k;\eta,0\rangle. \tag{4.4}$$

Thus

$$\langle p_1,p_2(+)|k;\eta,\tau\rangle = e^{i\tau(\eta p_1+\eta p_2)/\hbar}\,\langle p_1,p_2(+)|k;\eta,0\rangle. \tag{4.5}$$

Combining these results, we have

$$\begin{aligned}&\langle p_1,p_2(+)|k;\eta,\tau\rangle\\ &\quad= \delta_\eta^3(p_1+p_2-k)\,e^{i\tau(\eta p_1+\eta p_2)/\hbar}\,\mathfrak{F}(p_1,p_2;\eta),\end{aligned} \tag{4.6}$$

where k does not appear in $\mathfrak{F}$ because of (4.3). Now $\mathfrak{F}$ is a Lorentz invariant function of its vector variables and so can be written as

$$\mathfrak{F} = \mathfrak{F}(p_1p_2; \eta p_1, \eta p_2), \tag{4.7}$$

where the squares of the vectors are suppressed since they are not variable. In a careful calculation of the decay probability from a normalized initial state sharply peaked about some k_0 hyperplane momentum, the result would be proportional to the square of the absolute value of the form factor $\mathfrak{F}$. In first-order perturbation theory the η dependence in $\mathfrak{F}$ disappears and we have the conventional form factor for the $0 \to 0 + 0$ transition which yields an isotropic angular distribution in the rest frame of the decay products. Clearly the η dependence will allow deviations from isotropy. For example, consider an expansion of $\mathfrak{F}$ in powers of $(\eta p_1 + \eta p_2)$ and $(\eta p_1 - \eta p_2)$,

$$\mathfrak{F}(p_1p_2; \eta p_1, \eta p_2) = F_0(p_1p_2) + (\eta p_1 + \eta p_2)F_1(p_1p_2) + (\eta p_1 - \eta p_2)F_2(p_1p_2) + \cdots. \tag{4.8a}$$

In the rest frame with $\mathbf{p}_1 \equiv \mathbf{p} = -\mathbf{p}_2$, we have

$$\mathfrak{F} = F_0(p_{10}p_{20} + \mathbf{p}^2) + \eta_0(p_{10} + p_{20})F_1(p_{10}p_{20} + \mathbf{p}^2) + [\eta_0(p_{10} - p_{20}) - 2\boldsymbol{\eta}\cdot\mathbf{p}]F_2(p_{10}p_{20} + \mathbf{p}^2) + \cdots \tag{4.8b}$$

and the term proportional to $\boldsymbol{\eta}\cdot\mathbf{p}$ describes a P-wave angular distribution. If $F_2 \neq 0$, there is an $S = 1$ component in the spin spectrum of the unstable particle.

Notice that if the decay products are identical particles satisfying Bose statistics, then $\mathfrak{F}$ must be symmetric under the interchange $p_1 \leftrightarrow p_2$ and consequently odd powers of $(\eta p_1 - \eta p_2)$ cannot appear. In such a case the D wave is the simplest nonisotropic angular distribution one can have, and the nonvanishing components in the parent particle spin-spectrum would all be even.[8] This would very likely narrow the r.m.s. width of the spin spectrum since the strength of the higher spin terms in the expansion is likely to decrease rapidly. But more of that in a later section.

Consider now the case which conventionally would be described by saying the parent particle has spin one, i.e., the initial state is characterized not only by hyperplane parameters and a momentum eigenvalue but also by a polarization vector ϵ_μ satisfying $\eta\epsilon = 0$ and $\epsilon^2 = -1$. Also the state vector is linear in ϵ. Thus we have

$$\langle p_1p_2(+)|k, \epsilon; \eta, \tau\rangle = \delta_\eta^3(p_1 + p_2 - k)\, e^{i\tau(\eta p_1 + \eta p_2)/\hbar}\epsilon^\mu \mathfrak{F}_\mu(p_1, p_2; \eta), \tag{4.9}$$

where $\mathfrak{F}_\mu$ is now a vector function. The most general form for $\mathfrak{F}_\mu$ here is

$$\mathfrak{F}_\mu(p_1, p_2; \eta) = (p_1 + p_2)_\mu F_1(p_1p_2; \eta p_1, \eta p_2) + (p_1 - p_2)_\mu F_2(p_1p_2; \eta p_1, \eta p_2) + \epsilon_{\mu\alpha\beta\gamma}p_1^\alpha p_2^\beta \eta^\gamma F(p_1p_2; \eta p_1, \eta p_2). \tag{4.10}$$

In this case of the $1 \to 0 + 0$ transition the comparison with the results of the conventional treatment is more intricate than for the $0 \to 0 + 0$ transition. It is not sufficient merely to abolish all η dependence, for that still leaves us with two form factors F_1 and F_2. In fact, since the definition of the polarization vector ϵ employs the hyperplane vector η, one cannot consistently abolish η dependence. The correct limit is obtained by realizing that in the conventional treatment the parent particle is assumed to exist at a definite time in the rest frame of the decay products. In the present formalism such an hypothesis is characterized by giving η_μ the value

$$\eta_\mu = (p_1 + p_2)_\mu/[(p_1 + p_2)^2]^{1/2}. \tag{4.11}$$

In combination with the constraint on ϵ this renders F_2 the only form factor contributing to the decay amplitude as the conventional treatment demands.

If the reader will now reinspect the equations associated with the first example, i.e., the $0 \to 0 + 0$ transition, he will see that there too the conventional results are recaptured when (4.11) is satisfied. Since (4.11) is not an approximation as such, but rather a possible value for η, it follows that unstable systems with sharply defined rest frames [the physical condition determined by (4.11)] do not display the angular distribution effects we are considering. In other words we can never observe these effects from unstable systems that are known with great precision to decay at rest. For to know this is to know that $\mathbf{p} = 0$ precisely for the parent particle. But that can only be in an instantaneous state, i.e., one with $\boldsymbol{\eta} = 0$, η_0 1, and in such a state $\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p} = 0$ and (4.11) is satisfied. This consideration displays the source of the velocity dependence of the spin spectrum referred to in the introduction.

5. KINEMATICAL AND DYNAMICAL SPIN

There exist at least two distinct approaches to the concept of spin in relativistic quantum theory which yield equivalent results when applied to stable particles, but which are not equivalent when applied to unstable particles.

One approach employs the concept of the degeneracy of the momentum eigenvectors under the Euclidean group.[9] Consider such a vector, $|\mathbf{p}, \alpha\rangle$, in which α denotes information besides the momentum and independent of it which may be needed to determine the eigenvector uniquely (we suppress the time variable which would be required in the case of an unstable system). Now under a rotation R the momentum eigenvalue changes to $R\mathbf{p}$ and the momentum eigenvector to $U(R)|\mathbf{p}, \alpha\rangle$. The question of degeneracy is the question of the equality

$$U(R)|\mathbf{p}, \alpha\rangle \stackrel{?}{=} |R\mathbf{p}, \alpha\rangle \tag{5.1}$$

with the same α on both sides. If the equality always holds, we say the system has spin zero. If not, we look for the number of linearly independent eigenvectors $|\mathbf{p}, \alpha_n\rangle$ such that we always have

$$U(R)|\mathbf{p}, \alpha_n\rangle = \sum_{m=1}^{N} \mathfrak{D}_{mn}(R)|R\mathbf{p}, \alpha_m\rangle, \tag{5.2}$$

the $\mathfrak{D}$ being some R dependent coefficients. Writing the integer N as $2S + 1$, we say the system has spin S, and for the case $\mathbf{p} = 0$ we find

$$\mathbf{J}^2|0, \alpha_n\rangle = \hbar^2 S(S + 1)|0, \alpha_n\rangle. \tag{5.3}$$

By not involving the time or transformations to moving frames, the discussion is equally applicable to

Galilean or Lorentzian kinematics and to stable or unstable particles. It can, furthermore, be generalized to apply to unstable particles on arbitrary hyperplanes in a manner leaving the hyperplane fixed throughout the discussion. Consequently, to distinguish this approach from the following, I refer to the defined spin as *kinematical spin*.

The second approach to spin is concerned with the behavior of the states of the system under transformations which, interpreted actively, have dynamical content, i.e., time displacements and Lorentz transformations[10] or more generally those that change the hyperplane. The simplest expression of this spin concept is in terms of the Casimir invariants of the Poincaré group. For stable systems we have

$$P^2|p,\alpha\rangle = m^2c^2|p,\alpha\rangle, \tag{5.4a}$$

$$W^2|p,\alpha\rangle = -m^2c^2h^2S(S+1)|p,\alpha\rangle, \tag{5.4b}$$

where

$$W_\mu = \tfrac{1}{2}\epsilon_{\mu\alpha\beta\gamma}M^{\alpha\beta}P^\gamma \tag{5.5}$$

is the Pauli–Lubanski vector. For unstable particles which do not possess a precise mass we follow Lurçat[1] and write

$$\{W^2 + h^2S(S+1)P^2\}|k,\alpha;\eta,\tau\rangle = 0 \tag{5.6}$$

as the defining equation for the system to have *dynamical spin* S.

Now there always exists Lorentz transformations which will connect a given stable particle momentum eigenvector with a rest frame momentum eigenvector with $\mathbf{p} = 0$. But in the rest frame $W^2 = -P_0^2\mathbf{J}^2$ and the invariant equation (5.4b) is equivalent to (5.3). For stable particles the kinematical and dynamical spin are always equal.

As we saw earlier, however, the transformation of an unstable system momentum eigenvector to a sharp rest frame is not possible if the initial momentum was nonvanishing. Consequently, the equivalence of the kinematical and dynamical spin cannot be demonstrated for unstable systems.

For a clear example that the two spins are not equivalent consider a case in which the kinematical spin is given as zero. In the terms of the covariant formalism this means that the hyperplane momentum eigenvectors are nondegenerate and consequently they can be denoted by $|k;\eta,\tau\rangle$; no other symbols are necessary.

The spin operator associated with the kinematical spin can be constructed and seen to vanish in this case. We first define formally a position eigenvector for the unstable system on the (η,τ) hyperplane

$$|y;\eta,\tau\rangle \equiv (2\pi h)^{-3/2}\int d^4k\ \delta(\eta k)\, e^{iky/h}|k;\eta,\tau\rangle, \tag{5.7}$$

where $\eta y = 0$. For a stable system (or more precisely an elementary system in Wigner's sense) of vanishing spin this state is the hyperplane generalization of the Newton–Wigner position eigenvectors which were defined on instantaneous hyperplanes.[11] Since the transform defining these position eigenvectors on hyperplanes can be inverted, they form a complete basis in the same sense that the momentum eigenvectors do and they are orthogonal,

$$\langle y';\eta,\tau|y;\eta,\tau\rangle = \delta_\eta^3(y'-y) \tag{5.8a}$$

iff

$$\langle k';\eta,\tau|k;\eta,\tau\rangle = \delta_\eta^3(k'-k). \tag{5.8b}$$

We can, therefore, define, over the subspace of unstable system state vectors on the (η,τ) hyperplane, a position operator $Q_\mu(\eta,\tau)$ by

$$Q_\mu(\eta,\tau)|y;\eta,\tau\rangle = y_\mu|y;\eta,\tau\rangle. \tag{5.9}$$

Now consider a homogeneous transformation of the Poincaré group which leaves the hyperplane orientation η invariant. If the hyperplane were instantaneous, such a transformation would be a rotation, and so we may call this one a hyperplane rotation. The generators of such a three-parameter transformation are just the projections of the $M_{\mu\nu}$ orthogonal to η_λ, and they may be expressed in the convenient form

$$J_\mu(\eta) \equiv -\tfrac{1}{2}\epsilon_{\mu\alpha\beta\gamma}M^{\gamma\beta}\eta^\gamma, \tag{5.10}$$

which is clearly the hyperplane generalization of the total angular momentum operator. We can now express the kinematical spin operator defined in the same space as $Q_\mu(\eta,\tau)$ by the equation

$$J_\mu(\eta) - \epsilon_{\mu\alpha\beta\gamma}Q^\alpha(\eta,\tau)K^\beta(\eta)\eta^\gamma \equiv S_\mu(\eta,\tau). \tag{5.11}$$

The kinematical spin vanishes in the sense that

$$S_\mu(\eta,\tau)|y;\eta,\tau\rangle = 0 \tag{5.12}$$

as a direct consequence of

$$U(\Lambda,a)|y;\eta,\tau\rangle = |\Lambda y + a - \Lambda\eta(a\Lambda\eta);\Lambda\eta,\tau + a\Lambda\eta\rangle. \tag{5.13}$$

From completeness within the subspace we also clearly have

$$S_\mu(\eta,\tau)|k;\eta,\tau\rangle = 0. \tag{5.14}$$

If the dynamical spin were identical with the kinematical spin, then the Pauli–Lubanski vector

$$W_\mu \equiv \tfrac{1}{2}\epsilon_{\mu\alpha\beta\gamma}M^{\alpha\beta}P^\gamma \tag{5.15}$$

would yield zero also when applied to the hyperplane momentum eigenvector. But from

$$P^\gamma|k;\eta,\tau\rangle = \left(k^\gamma - ih\eta^\gamma\frac{\partial}{\partial\tau}\right)|k;\eta,\tau\rangle \tag{5.16}$$

and

$$M^{\alpha\beta}|k;\eta,\tau\rangle = ih\left(k^\alpha\frac{\partial}{\partial k_\beta} - k^\beta\frac{\partial}{\partial k_\alpha} + \eta^\alpha\frac{\partial}{\partial\eta_\beta} - \eta^\beta\frac{\partial}{\partial\eta_\alpha}\right)|k;\eta,\tau\rangle \tag{5.17}$$

we get

$$W_\mu|k;\eta,\tau\rangle = -ih\,\epsilon_{\mu\alpha\beta\gamma}\left(\eta^\alpha\frac{\partial}{\partial\eta_\beta}k^\gamma - ihk^\alpha\frac{\partial}{\partial k_\beta}\eta^\gamma\frac{\partial}{\partial\tau}\right)|k;\eta,\tau\rangle \stackrel{?}{=} 0. \tag{5.18}$$

The essential point is that the effect of the kinematical spin operator is determined solely by the behavior of the state vector under transformations that leave the hyperplane unchanged, while the Pauli–Lubanski vector and the Casimir invariants of the Poincaré group have an effect depending on the hyper-

plane dependence of the states. Furthermore, as we saw earlier, the hyperplane dependence of the state is precisely the source of the deviation from isotropy in the decay product rest frame angular distribution. For a stable system a momentum eigenvector on one hyperplane *is* a momentum eigenvector on any hyperplane, and this renders the hyperplane dependence trivial and forces equality between kinematical and dynamical spin.

We can pursue the spin content of the unstable system further by applying (5.18) to Eq. (4.6). Thus we have

$$\langle p_1 p_2(+)|W_\mu|k;\eta,\tau\rangle = -ih\tilde{\omega}_\mu(k,\eta)\langle p_1,p_2(+)|k;\eta,\tau\rangle \tag{5.19a}$$

and

$$\langle p_1,p_2(+)|W^2|k;\eta,\tau\rangle = -h^2\tilde{\omega}(k,\eta)^2\langle p_1,p_2(+)|k;\eta,\tau\rangle, \tag{5.19b}$$

where

$$\tilde{\omega}_\mu(k,\eta) \equiv \epsilon_{\mu\alpha\beta\gamma}\left(\eta^\alpha \frac{\partial}{\partial\eta_\beta} k^\gamma - ih k^\alpha \frac{\partial}{\partial k_\beta}\eta^\gamma \frac{\partial}{\partial\tau}\right). \tag{5.20}$$

If we now expand $\mathfrak{F}(p_1,p_2;\eta)$ in powers of the components of η orthogonal to p_1+p_2 [more precisely, in terms of basis functions of the little group of $(p_1+p_2)_\mu$ defined on the hyperboloid $\eta^2=1$], we will find that the differential operator $\tilde{\omega}^2$ just multiplies the $(s+1)$th term in the expansion by

$$s(s+1)(p_1+p_2)^2.$$

The basis functions in question are the traceless projections of the symmetric, transverse products

$$h_{\mu_1}\cdots h_{\mu_s},$$

where

$$h_\mu \equiv \eta_\mu - Q_\mu \eta Q/Q^2 \tag{5.21}$$

and

$$Q_\mu = (p_1+p_2)_\mu. \tag{5.22}$$

Hence the expansion of $\mathfrak{F}$ has the form

$$\mathfrak{F}(p_1,p_2;\eta) = \sum_{s=0}^{\infty} h_{\mu_1}\cdots h_{\mu_s} F^{\mu_1\cdots\mu_s}(q,Q,\eta Q). \tag{5.23}$$

where the F's are symmetric, transverse (orthogonal to Q), traceless tensor functions of $q_\mu = (p_1-p_2)_\mu$, Q_μ, and the scalar ηQ.

The $(s+1)$th term in the expansion of $\langle p_1,p_2(+)|k;\eta,\tau\rangle$ is

$$\delta_\eta^3(k-Q)\,e^{i\eta Q\tau/h} h_{\mu_1}\cdots h_{\mu_s} F^{\mu_1\cdots\mu_s}(Q,q;\eta Q),$$

and we have

$$\begin{aligned}
&\epsilon_{\mu\alpha\beta\gamma}\eta^\alpha\frac{\partial}{\partial\eta_\beta}k^\gamma\left[\delta_\eta^3(k-Q)^{i\eta Q\tau/h}h_{\mu_1}\cdots h_{\mu_s}F^{\mu_1\cdots\mu_s}(Q,q;\eta Q)\right]\\
&= \left[\epsilon_{\mu\alpha\beta\gamma}\eta^\alpha\frac{\partial}{\partial\eta_\beta}k^\gamma\delta_\eta^3(k-Q)e^{i\eta Q\tau/h}\right]h_{\mu_1}\cdots h_{\mu_s}F^{\mu_1\cdots\mu_s}\\
&\quad+\delta_\eta^3(k-Q)e^{i\eta Q\tau/h}\left[\epsilon_{\mu\alpha\beta\gamma}\eta^\alpha\frac{\partial}{\partial\eta_\beta}Q^\gamma h_{\mu_1}\cdots h_{\mu_s}\right]F^{\mu_1\cdots\mu_s}\\
&= \left[ih\,\epsilon_{\mu\alpha\beta\gamma}k^\alpha\frac{\partial}{\partial k_\beta}\eta^\gamma\frac{\partial}{\partial\tau}\delta_\eta^3(k-Q)e^{i\eta Q\tau/h}\right]h_{\mu_1}\cdots h_{\mu_s}F^{\mu_1\cdots\mu_s}\\
&\quad+\delta_\eta^3(k-Q)e^{i\eta Q\tau/h}\left[\epsilon_{\mu\alpha\beta\gamma}\eta^\alpha\frac{\partial}{\partial\eta_\beta}Q^\gamma h_{\mu_1}\cdots h_{\mu_s}\right]F^{\mu_1\cdots\mu_s},
\end{aligned} \tag{5.24}$$

where the replacement of k by Q inside the brackets in the second term of the second expression is permitted by the delta function, and the absence of a third term in the second expression is a consequence of the differential operator annihilating *any* function of ηQ, in the presence of the delta function, and, finally, the first term of the third expression follows from employing the Fourier representation of the product of the delta function and the exponential in the first term of the second expression. From (5.24) we have

$$\begin{aligned}
&\tilde{\omega}_\mu(k,\eta)\delta_\eta^3(k-Q)e^{i\eta Q\tau/h}h_{\mu_1}\cdots h_{\mu_s}F^{\mu_1\cdots\mu_s}\\
&\quad=\delta_\eta^3(k-Q)e^{i\eta Q\tau/h}[\omega_\mu(Q,\eta)h_{\mu_1}\cdots h_{\mu_s}]F^{\mu_1\cdots\mu_s},
\end{aligned}$$

where (5.25)

$$\omega_\mu(Q,\eta) \equiv \epsilon_{\mu\alpha\beta\gamma}\eta^\alpha\frac{\partial}{\partial\eta_\beta}Q^\gamma. \tag{5.26}$$

Similarly

$$\begin{aligned}
&\omega(k,\eta)^2\delta_\eta^3(k-Q)e^{i\eta Q\tau/h}h_{\mu_1}\cdots h_{\mu_s}F^{\mu_1\cdots\mu_s}\\
&=\delta_\eta^3(k-Q)e^{i\eta Q\tau/h}[\omega(Q,\eta)^2h_{\mu_1}\cdots h_{\mu_s}]F^{\mu_1\cdots\mu_s}.
\end{aligned} \tag{5.27}$$

But

$$\begin{aligned}
&[\omega(Q,\eta)^2h_{\mu_1}\cdots h_{\mu_s}]F^{\mu_1\cdots\mu_s}\\
&\quad=s(s+1)Q^2h_{\mu_1}\cdots h_{\mu_s}F^{\mu_1\cdots\mu_s},
\end{aligned} \tag{5.28}$$

which proves the assertion made just after Eq. (5.20).

The physical content of the preceeding derivation is embodied in the two equations

$$\begin{aligned}
&\langle p_1,p_2(+)|k;\eta,\tau\rangle = \delta_\eta^3(k-Q)\,e^{i\eta Q\tau/h}\\
&\qquad\times\sum_{s=0}^{\infty}h_{\mu_1}\cdots h_{\mu_s}F^{\mu_1\cdots\mu_s}(q,Q;\eta Q),
\end{aligned} \tag{5.29a}$$

$$\begin{aligned}
&\langle p_1,p_2(+)|W^2|k;\eta,\tau\rangle\\
&\quad=\delta_\eta^3(k-Q)e^{i\eta Q\tau/h}\sum_{s=0}^{\infty}[-h^2s(s+1)Q^2]\\
&\qquad\times h_{\mu_1}\cdots h_{\mu_s}F^{\mu_1\cdots\mu_s}(q,Q;\eta Q),
\end{aligned} \tag{5.29b}$$

which explicitly display the dynamical spin spectrum of the kinematically spinless unstable system.

Several possibilities are open with Eqs. (5.29). First, the unstable system may after all retain a definite dynamical spin (all but one of the F functions vanish), but it need not be the same as the kinematical spin (in the present case we may have $F=0$ and $F^\mu\neq 0$, for example). Second, the more likely arrangement is that there is no definite spin but that all or most of the F's are nonvanishing; in other words, there is a spin spectrum. Clearly the answer depends on the details of the dynamical interaction leading to instability. We would expect a large class of "weak" interactions to yield a dominant $F^{\mu_1\cdots\mu_s}$ for $s=$ kinematical spin and smaller F's surrounding the kinematical value. But how large this class is or how bizarre an interaction must be to deviate from this feature is an open matter requiring, for its resolution, the investigation of detailed dynamical models.

Although I shall not demonstrate the matter here, it should be clear that if the preceeding discussion had employed an unstable system with nonvanishing kinematical spin, then the analysis of the dynamical spin content of the decay amplitude would have involved

the addition of the kinematical spin to the "spin" coming from the hyperplane dependence alone. In particular a kinematical spin of s_0 and a "spin" s_1, associated with the $(s_1 + 1)$th term in the expansion of the hyperplane dependence will combine to contribute to all values of the dynamical spin from $|s_0 - s_1|$ to $s_0 + s_1$. In this way for $s_0 \neq 0$ a given dynamical spin value appears more than once in the analysis of the states of the unstable system, and therefore the representation of the Poincaré group carried by the unstable system is not multiplicity free.[1]

6. THE PERTURBATIVE APPROACH TO UNSTABLE SYSTEMS

In this section we will consider the relation between the ideas so far expressed here and the more traditional perturbative approach.[2] Thus we now regard the instantaneous 3-momentum eigenvector $|\mathbf{p}, t\rangle$ as an eigenvector of a time-dependent "unperturbed" Hamiltonian as well, i.e.,

$$P_0(t)|\mathbf{p}, t\rangle = (\mathbf{p}^2 + m_0^2 c^2)^{1/2}|\mathbf{p}, t\rangle \tag{6.1}$$

where m_0 is the unperturbed rest mass.

To generalize Eq. (6.1) to arbitrary hyperplanes, we must regard it as associated with the 3-momentum eigenvalue equation

$$\mathbf{P}|\mathbf{p}, t\rangle = \mathbf{p}|\mathbf{p}, t\rangle, \tag{6.2}$$

not withstanding the fact that the operator $\mathbf{P}$ is not time dependent. The generalization, which has been discussed before,[7,12] is effected by the defining equation

$$P_\mu^{(0)}(\eta, \tau) \equiv U(\Lambda, a)\{\Lambda_\mu^\nu P_\nu^{(0)}(t)\}U^{-1}(\Lambda, a), \tag{6.3}$$

where $\eta_\mu = \Lambda_\mu^0$, $\tau = ct + a^\mu \eta_\mu$, and $P_\nu^{(0)}(t) = (P_0(t), \mathbf{P})$. This definition yields the transformation rule

$$U^{-1}(\Lambda, a)P_\mu^{(0)}(\Lambda\eta, \tau + a\Lambda\eta)U(\Lambda, a) = \Lambda_\mu^\nu P^{(0)}(\eta, \tau) \tag{6.4}$$

and the relation

$$P_\mu - \eta_\mu \eta P = P_\mu^{(0)}(\eta, \tau) - \eta_\mu \eta P^{(0)}(\eta, \tau), \tag{6.5}$$

so that

$$P_\mu = P_\mu^{(0)}(\eta, \tau) + \eta_\mu V(\eta, \tau). \tag{6.6}$$

In other words the perturbing interaction V modified only that part of P_μ which is parallel to η_μ. The eigenvalue equations (6.1), (6.2) become

$$K_\mu(\eta)|k; \eta, \tau\rangle = k_\mu |k; \eta, \tau\rangle, \tag{6.7a}$$

$$\eta P^{(0)}(\eta, \tau)|k; \eta, \tau\rangle = \sqrt{m_0^2 c^2 - k^2}\,|k; \eta, \tau\rangle, \tag{6.7b}$$

and for some future purposes it will be convenient to introduce the unperturbed 4-momentum

$$p_\mu \equiv k_\mu + \eta_\mu \sqrt{m_0^2 c^2 - k^2}, \quad p^2 = m_0^2 c^2, \tag{6.8}$$

and the unperturbed 4-momentum eigenvector

$$|p; \eta, \tau\rangle \equiv e^{-i\eta p \tau/\hbar}|k; \eta, \tau\rangle, \tag{6.9}$$

satisfying

$$P_\mu^{(0)}(\eta, \tau)|p; \eta, \tau\rangle = p_\mu |p; \eta, \tau\rangle. \tag{6.10}$$

Besides the unperturbed generators of translations we can also introduce unperturbed generators of homogeneous Lorentz transformations. We already have

$$\mathbf{J}|\mathbf{p}, t\rangle = ih\,\mathbf{p} \times \frac{\partial}{\partial \mathbf{p}}|\mathbf{p}, t\rangle, \tag{6.11}$$

which yields

$$e^{(i/\hbar)\mathbf{J}\cdot\hat{n}\theta}|\mathbf{p}, t\rangle = |\mathbf{p}', t\rangle, \tag{6.12}$$

where

$$\mathbf{p}' = \hat{n}(\hat{n}\cdot\mathbf{p}) + \cos(\theta)[\mathbf{p} - \hat{n}(\hat{n}\cdot\mathbf{p})] + \sin(\theta)(\hat{n} \times \mathbf{p}). \tag{6.13}$$

The essential ingredient for completing the construction of the unperturbed generators of the Lorentz group is that the time parameter t of the eigenvector is not modified by the transformation, i.e., the transformation does not alter the hyperplane on which the unstable system is defined but rather treats the system as though it were stable ignoring the hyperplane dependence. Thus we introduce

$$M_{ij}^{(0)}(t) \equiv \epsilon_{ijk}J_k, \quad M_{i0}^{(0)}(t) = -M_{0i}^{(0)}(t), \tag{6.14}$$

such that

$$e^{-[iM_{\mu\nu}^{(0)}(t)\omega^{\mu\nu}(\Lambda)]/2\hbar}|p, t\rangle \equiv |\Lambda p, t\rangle, \tag{6.15}$$

where

$$|p, t\rangle \equiv e^{-ip_0 ct/\hbar}|p, t\rangle,$$

the instantaneous case of Eq. (6.9)

It follows from (6.15) that the $M_{\mu\nu}^{(0)}(t)$ satisfy the Lie algebra of the Lorentz group.

The generalization of the unperturbed homogeneous generators to arbitrary hyperplanes is effected by

$$M_{\mu\nu}^{(0)}(\eta, \tau) \equiv U(\Lambda, a)[\Lambda_\mu^\lambda \Lambda_\nu^\rho M_{\lambda\rho}^{(0)}(t) + a_\mu \Lambda_\nu^\rho P_\rho^{(0)}(t) - a_\nu \Lambda_\mu^\lambda P_\lambda^{(0)}(t)]U^{-1}(\Lambda, a), \tag{6.16}$$

where, again, $\eta_\mu = \Lambda_\mu^0$ and $\tau = ct + a\Lambda\eta$. As a result of this definition the $M_{\mu\nu}^{(0)}(\eta, \tau)$ transform in accordance with the rules,

$$U^{-1}(\Lambda, a)M_{\mu\nu}^{(0)}(\Lambda\eta, \tau + a\Lambda\eta)U(\Lambda, a) = \Lambda_\mu^\lambda \Lambda_\nu^\rho M_{\lambda\rho}^{(0)}(\eta, \tau) + a_\mu \Lambda_\nu^\rho P_\rho^{(0)}(\eta, \tau) - a_\nu \Lambda_\mu^\lambda P_\lambda^{(0)}(\eta, \tau). \tag{6.17}$$

Now for each hyperplane (η, τ), the $P_\mu^{(0)}(\eta, \tau)$ and the $M_{\mu\nu}^{(0)}(\eta, \tau)$ satisfy the Lie algebra of the Poincaré group, and so we may refer to the *unperturbed Poincaré group on the* (η, τ) *hyperplane*. The unitary operators for these groups are the

$$U^{(0)}(\Lambda, a; \eta, \tau) \equiv e^{iP_\mu^{(0)}(\eta, \tau)a^\mu/\hbar}e^{-[iM_{\mu\nu}^{(0)}(\eta, \tau)\omega^{\mu\nu}(\Lambda)]/2\hbar}, \tag{6.18}$$

and they satisfy

$$U^{(0)}(\Lambda, a; \eta, \tau)|p; \eta, \tau\rangle = e^{ia\Lambda p/\hbar}|\Lambda p; \eta, \tau\rangle, \tag{6.19a}$$

which is to be contrasted with

$$U(\Lambda, a)|p; \eta, \tau\rangle = e^{ia\Lambda p/\hbar}|\Lambda p; \Lambda\eta, \tau + a\Lambda\eta\rangle \tag{6.19b}$$

for the full Poincaré group operators.

Just as the translation generators P_μ and $P_\mu^{(0)}(\eta, \tau)$

differ by a term parallel to η_μ [see Eq. (6.6)] so the defining Eqs. (6.14)–(6.16) yield the form[12]

$$M_{\mu\nu} = M^{(0)}_{\mu\nu}(\eta,\tau) + \eta_\mu U_\nu(\eta,\tau) - \eta_\nu U_\mu(\eta,\tau), \quad (6.20)$$

where the U_μ can be chosen to satisfy $\eta U = 0$ without loss of generality.

From (5.16) and (6.9) we have

$$P_\mu |p;\eta,\tau\rangle = \left(p_\mu - ih\eta_\mu \frac{\partial}{\partial\tau}\right)|p;\eta,\tau\rangle, \quad (6.21)$$

which when compared with (6.6) and (6.10) yields

$$V(\eta,\tau)|p;\eta,\tau\rangle = ih\frac{\partial}{\partial\tau}|p;\eta,\tau\rangle. \quad (6.22)$$

Similarly, from (5.17) and (6.9), we have

$$M_{\mu\nu}|p;\eta,\tau\rangle = ih\left(p_\mu\frac{\partial}{\partial p^\nu} - p_\nu\frac{\partial}{\partial p^\mu} + \eta_\mu\frac{\partial}{\partial\eta^\nu} - \eta_\nu\frac{\partial}{\partial\eta^\mu}\right) \times |p;\eta,\tau\rangle, \quad (6.23)$$

which, when compared with

$$M^{(0)}_{\mu\nu}(\eta,\tau)|p;\eta,\tau\rangle = ih\left(p_\nu\frac{\partial}{\partial p^\nu} - p_\mu\frac{\partial}{\partial p^\mu}\right)|p;\eta,\tau\rangle|, \quad (6.24)$$

which follows from (6.19a), yields

$$U_\mu(\eta,\tau)|p;\eta\tau\rangle = ih\frac{\delta}{\delta\eta^\mu}|p;\eta,\tau\rangle, \quad (6.25)$$

where

$$\frac{\delta}{\delta\eta^\mu} \equiv \frac{\partial}{\partial\eta^\mu} - \eta_\mu\eta\frac{\partial}{\partial\eta} \quad (6.26)$$

is the differential operator that leaves η_μ on the unit hyperboloid. Thus we see that the dynamical effects of the interaction rendering the system unstable are embodied in the relativistic potentials $V(\eta,\tau)$ and $U_\mu(\eta,\tau)$.

It follows from the definition of the unperturbed Poincaré group that the kinematical spin of the unstable system is determined by the Casimir invariants of the unperturbed Poincaré group. Thus from

$$J_\mu(\eta)^2|p_\mu = \eta_\mu m_0 c;\eta,\tau\rangle = 0 \quad (6.27)$$

we obtain

$$W^{(0)}(\eta,\tau)^2|p;\eta\tau\rangle = 0, \quad (6.28)$$

where

$$W^{(0)}_\mu(\eta,\tau) \equiv -\tfrac{1}{2}\epsilon_{\mu\alpha\beta\gamma}M^{(0)\alpha\beta}(\eta,\tau)P^{(0)\gamma}(\eta,\tau) \quad (6.29)$$

because

$$|p;\eta,\tau\rangle = U^{(0)}(\Lambda,0;\eta,\tau)|p_\mu = \eta_\mu m_0 c;\eta,\tau\rangle, \quad (6.30)$$

where $p_\mu = \Lambda^\nu_\mu \eta_\nu m_0 c$. More generally from

$$J_\mu(\eta)^2|p_\mu = \eta_\mu m_0 c,\alpha;\eta,\tau\rangle = -h^2 s(s+1)|p_\mu = \eta_\mu m_0 c,\alpha;\eta,\tau\rangle \quad (6.31)$$

and

$$U^{(0)}(\Lambda,0;\eta,\tau)|p_\mu = \eta_\mu m_0 c,\alpha;\eta,\tau\rangle = \sum_{\beta=-s}^{s}\mathfrak{U}_{\beta\alpha}(p,\Lambda)|p,\beta;\eta,\tau\rangle, \quad (6.32)$$

we obtain

$$W^{(0)}_\mu(\eta,\tau)^2|p,\alpha;\eta,\tau\rangle = -m_0^2c^2h^2s(s+1)|p,\alpha;\eta,\tau\rangle. \quad (6.33)$$

But from (5.15) and (6.6), (6.20), (6.28) we have

$$W_\mu = W^{(0)}_\mu(\eta,\tau) - \tfrac{1}{2}\epsilon_{\mu\alpha\beta\gamma} \times [M^{(0)\alpha\beta}(\eta,\tau)\eta^\gamma V(\eta,\tau) + 2\eta^\alpha U^\beta(\eta,\tau)P^{(0)\gamma}(\eta,\tau)], \quad (6.34)$$

and we see again that outside of a particular dynamical model (specification of V and U_μ) we cannot say whether the dynamical spin determined by W_μ will be equal to the kinematical spin determined by $W^{(0)}_\mu(\eta,\tau)$ or will be sharp at all. It would seem to require a precise relation between V and U_μ to guarantee equality between the kinematical and dynamical spin. This relation will be investigated for a local field theory model in the next section.

I will close this section with a derivation of the action of the full Poincaré group operators on the unperturbed-hyperplane-dependent Poincaré group operators. In the case of the full Poincaré group we know that

$$U^{-1}(\Lambda,a)P_\mu U(\Lambda,a) = \Lambda^\nu_\mu P_\nu, \quad (6.35a)$$

$$U^{-1}(\Lambda,a)M_{\mu\nu}U(\Lambda,a) = \Lambda^\lambda_\mu\Lambda^\rho_\nu M_{\lambda\rho} + a_\mu\Lambda^\rho_\nu P_\rho - a_\nu\Lambda^\lambda_\mu P_\lambda \quad (6.35b)$$

and that

$$\begin{aligned} &U^{-1}(\Lambda_2,a_2)U(\Lambda_1,a_1)U(\Lambda_2,a_2) \\ &= U(\Lambda_2^{-1}\Lambda_1\Lambda_2, \Lambda_2^{-1}(\Lambda_1 a_2 + a_1 - a_2)) \\ &= e^{iU^{-1}(\Lambda_2,a_2)P_\mu U(\Lambda_2,a_2)a_1^\mu/h}\, e^{-iU^{-1}(\Lambda_2,a_2)M_{\mu\nu}U(\Lambda_2,a_2)\omega^{\mu\nu}(\Lambda_1)/2h}. \end{aligned} \quad (6.36)$$

By simple analogy, therefore, we conclude from (6.4), (6.17) that

$$\begin{aligned} &U^{-1}(\Lambda_2,a_2)U^{(0)}(\Lambda_1,a_1;\eta',\tau')U(\Lambda_2,a_2) \\ &= e^{iU^{-1}(\Lambda_2,a_2)P^{(0)}_\mu(\eta',\tau')U(\Lambda_2,a_2)a_1^\mu/h} \\ &\quad\times e^{-iU^{-1}(\Lambda_2,a_2)M^{(0)}_{\mu\nu}(\eta',\tau')U(\Lambda_2,a_2)\omega^{\mu\nu}(\Lambda_1)/2h} \\ &= U^{(0)}(\Lambda_2^{-1}\Lambda_1\Lambda_2, \Lambda_2^{-1}(\Lambda_1 a_2 + a_1 - a_2);\eta,\tau) \end{aligned} \quad (6.37)$$

where $\eta' = \Lambda_2\eta$, $\tau' = \tau + a_2\Lambda_2\eta$.

7. DECAY VIA LOCAL INTERACTIONS

In this last section I shall examine more closely the case of the decay of a kinematically spinless particle into two spinless stable particles. At a certain stage in the calculations the interactions responsible for the decay will be assumed *local* in a certain sense.

We begin with the decay amplitude

$$\langle p_1,p_2(+)|p;\eta,\tau\rangle = \delta^3_\eta(p_1+p_2-p)e^{i(\eta p_1+\eta p_2-\eta p)\tau/h}\mathfrak{F}(p_1p_2;\eta p_1,\eta p_2), \quad (7.1)$$

where $\langle p_1,p_2(+)|$ is the *outgoing* scattering eigenvector of the total 4-momentum with the two stable spinless particles possessing 4-momenta $p_{1\mu}$ and $p_{2\mu}$. If $V(\eta,\tau)$ is the potential responsible for the decay in the sense that the initial state $|p;\eta,\tau\rangle$ is an eigenvector of $P^{(0)}_\mu(\eta,\tau)$, where

$$P_\mu = P^{(0)}_\mu(\eta,\tau) + \eta_\mu V(\eta,\tau), \quad (7.2)$$

then the final state $\langle p_1,p_2(+)|$ is related to an eigenvector of $P^{(0)}_\mu(\eta,\tau)$ by

$$\langle p_1,p_2(+)| = \langle p_1,p_2(+);\eta,\tau| \times [I + V(\eta,\tau)(\eta p_1 + \eta p_2 - \eta P - io)^{-1}] \quad (7.3)$$

and the scattering eigenvector $\langle p_1,p_2(+);\eta,\tau|$ of $P_\mu^{(0)}(\eta,\tau)$ is assumed orthogonal to the initial state. Therefore, we have

$$\langle p_1,p_2(+)|p;\eta\tau\rangle = \langle p_1,p_2(+);\eta,\tau|V(\eta,\tau)(\eta p_1 + \eta p_2 - \eta P - io)^{-1} \times |p;\eta,\tau\rangle. \quad (7.4)$$

Next, to simplify the calculation, we retain the coupling through $V(\eta,\tau)$ of the two-particle scattering eigenvector of $P_\mu^{(0)}(\eta,\tau)$ to the unstable particle states and to these alone. In other words,

$$\langle p_1,p_2(+);\eta,\tau|V(\eta,\tau) \simeq \int d^4q\ \delta_+(q^2 - m_0^2c^2)\langle p_1,p_2(+);\eta,\tau| \times V(\eta,\tau)|q;\eta,\tau\rangle\langle q;\eta,\tau|. \quad (7.5)$$

Upon substituting this into (7.4) and writing[2]

$$\langle q;\eta,\tau|(\eta p_1 + \eta p_2 - \eta P - io)^{-1}|p;\eta\tau\rangle = 2\eta p\,\delta_\eta^3(q-p)[\eta p_1 + \eta p_2 - \eta P - R(\eta p;\eta p_1 + \eta p_2)]^{-1}, \quad (7.6)$$

where

$$2\eta p\,\delta_\eta^3(q-p)R(\eta p,\eta p_1 + \eta p_2) = \langle q;\eta\tau|V(\eta,\tau) + V(\eta,\tau) \times \Pi_\perp(\eta,\tau)(\eta p_1 + \eta p_2 - \eta P - io)^{-1} \times \Pi_\perp(\eta,\tau)V(\eta,\tau)|p;\eta\tau\rangle, \quad (7.7)$$

the Π's being projection operators onto the orthogonal complement of the subspace spanned by the unstable particle states $|p;\eta,\tau\rangle$, we obtain

$$\begin{aligned}\langle p_1,p_2(+)|p;\eta,\tau\rangle &= \int d^4q\ \delta_+(q^2 - m_0^2c^2) \times \langle p_1,p_2(+);\eta\tau|V(\eta,\tau)|q;\eta\tau\rangle 2\eta p\,\delta_\eta^3(q-p) \times [\eta p_1 + \eta p_2 - \eta P - R]^{-1} \\ &= \langle p_1,p_2(+);\eta\tau|V(\eta,\tau)|p;\eta,\tau\rangle \times [\eta p_1 + \eta p_2 - \eta P - R(\eta p,\eta p_1 + \eta p_2)]^{-1}. \end{aligned} \quad (7.8)$$

At this point we introduce the assumption of local interaction in the form

$$V(\eta,\tau) = -\int d^4x\ \delta(\eta x - \tau)\mathcal{L}_{\text{int}}(x), \quad (7.9)$$

where

$$\mathcal{L}_{\text{int}}(x) = e^{iPx/\hbar}\mathcal{L}_{\text{int}}(0)e^{-iPx/\hbar} = e^{iK(\eta)x/\hbar}\mathcal{L}_{\text{int}}(\eta\tau)e^{-iK(\eta)x/\hbar}. \quad (7.10)$$

This yields

$$\begin{aligned}\langle p_1,p_2(+);\eta,\tau|V(\eta,\tau)|p,\eta,\tau\rangle &= -\int d^4x\ \delta(\eta x - \tau)\,e^{i[(p_1+p_2-p)-\eta(\eta p_1+\eta p_2-\eta P)]x/\hbar} \times \langle p_1,p_2(+);\eta,\tau|\mathcal{L}_{\text{int}}(\eta,\tau)|p;\eta,\tau\rangle \\ &= -\int d^4y\ \delta(\eta y)\,e^{i(p_1+p_2-p)y/\hbar} \times \langle p_1,p_2(+);\eta,\tau|e^{i\eta P\tau/\hbar}\mathcal{L}_{\text{int}}(0)\,e^{-i\eta P\tau/\hbar}|p;\eta,\tau\rangle \\ &= -(2\pi\hbar)^3\,\delta_\eta^3(p_1 + p_2 - p) \times \langle p_1,p_2(+);\eta,0|e^{i\eta P^{(0)}(\eta,0)\tau/\hbar}\mathcal{L}_{\text{int}}(0) \times e^{-i\eta P^{(0)}(\eta,0)\tau/\hbar}|p;\eta,0\rangle \\ &= -(2\pi\hbar)^3\,\delta_\eta^3(p_1 + p_2 - p)e^{i(\eta p_1+\eta p_2-\eta p)\tau/\hbar} \times \langle p_1,p_2(+);\eta,0|\mathcal{L}_{\text{int}}(0)|p;\eta,0\rangle. \end{aligned} \quad (7.11)$$

The explicit η dependence left in the matrix element of $\mathcal{L}_{\text{int}}(0)$ can be eliminated if the local interaction generates a $U_\mu(\eta,\tau)$ given by

$$U_\mu(\eta,\tau) = -\int d^4x\ \delta(\eta x - \tau)(x_\mu - \eta_\mu\eta x)\mathcal{L}_{\text{int}}(x) \quad (7.12)$$

and is microcausal,

$$\delta(\eta x - \eta x')[\mathcal{L}_{\text{int}}(x),\mathcal{L}_{\text{int}}(x')] = 0, \quad (7.13)$$

for any timelike η_μ. These conditions lead to

$$[\mathcal{L}_{\text{int}}(0),U_\mu(\eta,0)] = 0, \quad (7.14)$$

and, since $\mathcal{L}_{\text{int}}(0)$ is a Lorentz scalar, i.e.,

$$[\mathcal{L}_{\text{int}}(0),M_{\mu\nu}] = 0, \quad (7.15a)$$

we also have

$$[\mathcal{L}_{\text{int}}(0),M_{\mu\nu}^{(0)}(\eta,0)] = 0. \quad (7.15b)$$

But

$$|p;\Lambda\eta,0\rangle = e^{-iM_{\mu\nu}\omega^{\mu\nu}(\Lambda)/2\hbar}\,e^{iM_{\mu\nu}^{(0)}(\eta,0)\omega^{\mu\nu}(\Lambda)/2\hbar}|p;\eta,0\rangle, \quad (7.16)$$

and similarly for $\langle p_1,p_2(+);\eta,0|$, so that

$$\langle p_1,p_2(+);\eta,0|\mathcal{L}_{\text{int}}(0)|p;\eta,0\rangle = \langle p_1,p_2(+);\ \Lambda\eta,0|\mathcal{L}_{\text{int}}(0)|p;\Lambda\eta,0\rangle. \quad (7.17)$$

The matrix element of the interaction Lagrangian density then is a Lorentz invariant function of the 4-momenta p_1,p_2, and p or

$$\langle p_1,p_2(+);\eta,0|\mathcal{L}_{\text{int}}(0)|p;\eta,0\rangle = F(p_1p_2;pp_1,pp_2), \quad (7.18)$$

where the dependence of F on the masses m_1,m_2, and m_0 is suppressed.

For the decay amplitude we now have

$$\langle p_1,p_2(+)|p;\eta,\tau\rangle = (2\pi\hbar)^3\,\delta_\eta^3(p_1 + p_2 - p)e^{i(\eta p_1+\eta p_2-\eta p)\tau/\hbar} \times F(p_1p_2;pp_1,pp_2)[\eta p_1 + \eta p_2 - \eta p - R \times(\eta p,\eta p_1 + \eta p_2)]^{-1}. \quad (7.19)$$

At first glance it may seem that there can be no deviation from spherical symmetry in the angular distribution since η_μ does not occur explicitly in F and its appearance in the "energy denominator" never involves $\eta p_1 - \eta p_2$. This is erroneous however because the three-dimensional delta function induces η dependence in F by forcing the replacement

$$p_\mu = (p_1 + p_2)_\mu - \eta_\mu(\eta p_1 + \eta p_2) + \eta_\mu\sqrt{m_0^2c^2 + (\eta p_1 + \eta p_2)^2 - (p_1 + p_2)^2}. \quad (7.20)$$

Thus, in the presence of the delta function,

$$F(p_1p_2;pp_1,pp_2) = F(p_1p_2;m_1^2c^2 + p_1p_2 - \eta p_1(\eta p_1 + \eta p_2) + \eta p_1\sqrt{(\eta p_1 + \eta p_2)^2 - \Delta}\; p_1p_2 + m_2^2c^2 - \eta p_2(\eta p_1 + \eta p_2) + \eta p_2\sqrt{(\eta p_1 + \eta p_2)^2 - \Delta}\cdots), \tag{7.21}$$

where $\Delta \equiv S - m_0^2c^2 = (p_1 + p_2)^2 - m_0^2c^2$. It follows from this that if the original F (prior to multiplication by the delta function) depends on $pp_1 - pp_2$ as well as $pp_1 + pp_2$, then the final F will depend on $\eta p_1 - \eta p_2$ as well as $\eta p_1 + \eta p_2$ and there will be an anisotropic angular distribution in the rest frame of the decay products. If the two decay products are not identical particles, then it is unlikely that F will be symmetric under interchange of p_1 with p_2. But in such a case (absence of Bose symmetry) F *must* depend on $pp_1 - pp_2$ and there *must* be deviation from spherical symmetry of the angular distribution in the decay product rest frame. Of course, if the decay products are identical, Bose symmetry will prevail, and a dependence of F on $pp_1 - pp_2$ can *only* occur through even powers of the difference. This becomes transformed into a similar dependence on $\eta p_1 - \eta p_2$ and, as indicated in Sec. 4, the result is the appearance of only even partial waves in the angular distribution and only even spin values in the spin spectrum of the parent particle.

On the other hand, inspection of (7.20) and (7.21) indicates that no deviation from spherical symmetry in the angular distribution can occur when $\Delta = 0$, i.e., when $(p_1 + p_2)^2 = p^2 \equiv m_0^2c^2$. At this value of Δ (usually the peak of the mass distribution) we have conservation of 4-momentum from (7.20), and (7.21) becomes

$$F(p_1p_2;pp_1,pp_2) = F[\tfrac{1}{2}(m_0^2 - m_1^2 - m_2^2)c^2; \tfrac{1}{2}(m_0^2 + m_1^2 - m_2^2)c^2; \tfrac{1}{2}(m_0^2 - m_1^2 + m_2^2)c^2]. \tag{7.22}$$

This result displays again the fact that in first-order perturbation theory the effect of interest can not be seen since one treats the parent particle state like a stable one (thereby eliminating explicit η dependence in the initial state) and also obtains conservation of total 4-momentum [thereby eliminating induced η dependence of the variety occurring in (7.21)]. For systems with a narrow width to their mass spectrum (so that the observed values of Δ are small), we may approximate (7.21) by a few terms in the Taylor series expansion of the Δ dependence. If we retain only first-order terms in Δ, we have

$$\begin{aligned} F &= F_{\Delta=0} + \frac{\Delta}{2}\frac{\partial F}{\partial p_1p_2}\bigg|_{\Delta=0} - \frac{\eta p_1}{\eta p_1 + \eta p_2}\frac{\partial F}{\partial pp_1}\bigg|_{\Delta=0} - \frac{\eta p_2}{\eta p_1 + \eta p_2}\frac{\partial F}{\partial pp_2}\bigg|_{\Delta=0} \\ &= F_{\Delta=0} + \frac{\Delta}{2}\left\{\frac{\partial F}{\partial p_1p_2} - \frac{1}{2}\left(\frac{\partial F}{\partial pp_1} + \frac{\partial F}{\partial pp_2}\right)\right\}_{\Delta=0} - \frac{\Delta}{4}\frac{\eta p_1 - \eta p_2}{\eta p_1 + \eta p_2}\left(\frac{\partial F}{\partial pp_1} - \frac{\partial F}{\partial pp_2}\right)_{\Delta=0}. \end{aligned} \tag{7.23}$$

The last term only yields a nonspherical angular distribution in the rest frame of decay products where the term becomes

$$-\frac{\Delta}{2}\frac{\eta\cdot\mathbf{p}_1}{\eta_0(p_{10} + p_{20})}\left(\frac{\partial F}{\partial pp_1} - \frac{\partial F}{\partial pp_2}\right)_{\Delta=0}. \tag{7.24}$$

The maximum value that $|\Delta|$ can have before the energy denominator in the decay amplitude becomes large enough to render the decay process negligible is of the order of magnitude

$$|\Delta|_{\max} \sim 2m_0\Gamma. \tag{7.25}$$

At the same time, if $\Gamma/m_0c^2 \ll 1$, then

$$p_{10} + p_{20} \sim m_0c$$

and

$$|\mathbf{p}_1| \sim [(m_0^2c^2 - (m_1 + m_2)^2c^2)(m_0^2c^2 - (m_1 - m_2)^2c^2)]^{1/2}/2m_0c. \tag{7.26}$$

Finally the value of $|\eta|/\eta_0$ is v/c where v is the velocity of the unstable particle in that frame in which the spatial 3-momentum is sharp. Therefore, the order of magnitude of (7.24) is

$$\frac{\Gamma}{m_0c^2}\frac{v}{c}\frac{[(m_0^2c^2 - (m_1 + m_2)^2c^2)(m_0^2c^2 - (m_1 - m_2)^2c^2)]^{1/2}}{2m_0^2c^2}\left|\, m_0^2c^2\left(\frac{\partial F}{\partial pp_1} - \frac{\partial F}{\partial pp_2}\right)\right|_{\Delta=0}\cos\theta, \tag{7.27}$$

where I have associated the quantity $m_0^2c^2$ with the derivatives of F so as to give that factor the same dimension as F itself. Since the F derivatives and the estimate of $|\mathbf{p}_1|/p_{10} + p_{20}$ varies with the dynamics and mode of decay, I take

$$(\Gamma/m_0c^2)(v/c) \tag{7.28}$$

as the natural dimensionless measure of the kinematical origins of the effect.

I will close this section with a demonstration that in the presence of local interactions without derivative coupling the eigenvalue of W_μ remains zero (if the kinematical spin has vanished) under just those conditions that we have seen to yield an isotropic angular distribution in the rest frame of the decay products.

We start from (6.34) into which we substitute (7.9) and (7.12) after noting from (6.20) that

$$J_\mu(\eta) = -\tfrac{1}{2}\epsilon_{\mu\alpha\beta\gamma}M^{(0)\alpha\beta}(\eta,\tau)\eta^\gamma, \tag{7.29}$$

and from (6.17) and (6.35b) that[12]

$$[K_\mu(\eta), U_\nu(\eta,\tau)] = ih(g_{\mu\nu} - \eta_\mu\eta_\nu)V(\eta,\tau). \tag{7.30}$$

we obtain

$$W_\mu - W_\mu^{(0)}(\eta,\tau) = J_\mu(\eta)V(\eta,\tau) - \tfrac{1}{2}\epsilon_{\mu\alpha\beta\gamma} \times [\eta^\alpha U^\beta(\eta,\tau) - \eta^\beta U^\alpha(\eta,\tau)]K^\gamma(\eta)$$

$$= J_\mu(\eta)V(\eta,\tau) - \tfrac{1}{2}\epsilon_{\mu\alpha\beta\gamma}K^\gamma(\eta)[\eta^\alpha U^\beta(\eta,\tau) - \eta^\beta U^\alpha(\eta,\tau)]$$
$$= \int d^4x\ \delta(\eta x - \tau)[J_\mu(\eta) - \tfrac{1}{2}\epsilon_{\mu\alpha\beta\gamma}K^\gamma(\eta)(\eta^\alpha x^\beta - \eta^\beta x^\alpha)]\mathcal{L}_{\rm int}(x)$$
$$= -\int d^4x\ \delta(\eta x - \tau)\{J_\mu(\eta) - \tfrac{1}{2}\epsilon_{\mu\alpha\beta\gamma}[x^\alpha K^\beta(\eta) - x^\beta K^\alpha(\eta)]\eta^\gamma\}\mathcal{L}_{\rm int}(x)$$
$$= -\int d^4x\ \delta(\eta x - \tau)e^{iK(\eta)x/\hbar}J_\mu(\eta)e^{-iK(\eta)x/\hbar}\mathcal{L}_{\rm int}(x)$$
$$= -\int d^4x\ \delta(\eta x - \tau)e^{iK(\eta)x/\hbar}J_\mu(\eta)\mathcal{L}_{\rm int}(\eta_\alpha\tau)e^{-iK(\eta)x/\hbar}. \quad (7.31)$$

Upon applying this operator to an unstable particle momentum eigenvector with vanishing kinematical spin, we have

$$W_\mu|p;\eta,\tau\rangle = (2\pi\hbar)^3\delta_\eta^3(P-p)J_\mu(\eta)\mathcal{L}_{\rm int}(\eta_\alpha\tau)|p;\eta,\tau\rangle. \quad (7.32)$$

Since the hyperplane rotation generator $J_\mu(\eta)$ commutes with $\mathcal{L}_{\rm int}(x)$ when $x_\mu = \eta_\mu\tau$ and

$$J_\mu(\eta)|p;\eta,\tau\rangle = \frac{ih}{2}\epsilon_{\mu\alpha\beta\gamma}p^\alpha\frac{\partial}{\partial p_\beta}\eta^\gamma|p;\eta,\tau\rangle, \quad (7.33)$$

we have

$$W_\mu|p;\eta,\tau\rangle = (2\pi\hbar)^3\delta_\eta^3(P-p)\mathcal{L}_{\rm int}(\eta_\mu\tau)\frac{ih}{2}\epsilon_{\mu\alpha\beta\gamma}p^\alpha\frac{\partial}{\partial p_\beta} \times|p;\eta,\tau\rangle\eta^\gamma. \quad (7.34)$$

Assuming only one decay mode, i.e., into the states $|p_1,p_2(+)\rangle$, we can write this as

$$W_\mu|p;\eta,\tau\rangle$$
$$= (2\pi\hbar)^3\int d^4p_1d^4p_2\ \delta_+(p_1^2 - m_1^2c^2)\delta_+(p_2^2 - m_2^2c^2) \times\delta_\eta^3(p_1+p_2-p)|p_1,p_2(+)\rangle\frac{ih}{2}\epsilon_{\mu\alpha\beta\gamma}p^\alpha\frac{\partial}{\partial p_\beta} \times\langle p_1,p_2(+)|\mathcal{L}_{\rm int}(\eta_\mu\tau)|p;\eta\tau\rangle\eta^\gamma$$
$$= (2\pi\hbar)^3\int d^4p_1d^4p_2\ \delta_+(p_1^2 - m_1^2c^2) \times\delta_+(p_2^2 - m_2^2c^2)\delta_\eta^3(p_1+p_2-p) \times e^{i(\eta p_1+\eta p_2-\eta p)\tau/\hbar}\frac{ih}{2}\epsilon_{\mu\alpha\beta\gamma}p^\alpha\frac{\partial}{\partial p_\beta} \times F(p_1p_2,pp_1,pp_2)\eta^\gamma\times|p_1,p_2(+)\rangle. \quad (7.35)$$

Now in the discussion following (7.21) we learned that an isotropic angular distribution in the rest frame of the decay products would occur if and only if F depended on pp_1 and pp_2 through the sum $pp_1 + pp_2$ and did not depend on $pp_1 - pp_2$. But this is just the condition that makes (7.35) vanish. Thus

$$\epsilon_{\mu\alpha\beta\gamma}p^\alpha\frac{\partial}{\partial p_\beta}\eta^\gamma F(p_1p_2,pp_1,pp_2) = \epsilon_{\mu\alpha\beta\gamma}p^\alpha\eta^\gamma\left(p_1^\beta\frac{\partial F}{\partial pp_1} + p_2^\beta\frac{\partial F}{\partial pp_2}\right) = \tfrac{1}{2}\epsilon_{\mu\alpha\beta\gamma}p^\alpha\eta^\gamma \times\left(\frac{(p_1+p_2)^\beta}{2}\frac{\partial F}{\partial(pp_1+pp_2)} + \frac{(p_1-p_2)^\beta}{2}\frac{\partial F}{\partial(pp_1-pp_2)}\right), \quad (7.36)$$

and the delta function $\delta_\eta^3(p_1+p_2-p)$ in (7.35) eliminates any contribution from the first term on the right of (7.36). The final result is

$$W_\mu|p;\eta,\tau\rangle = (2\pi\hbar)^3\int d^4p_1d^4p_2\,\delta_+(p_1^2 - m_1^2c^2)\delta_+(p_2^2 - m_2^2c^2) \times\delta_\eta^3(p_1+p_2-p)e^{i(\eta p_1+\eta p_2-\eta p)\tau/\hbar} \times\tfrac{1}{4}ih\,\epsilon_{\mu\alpha\beta\gamma}p^\alpha(p_1-p_2)^\beta\eta^\gamma \times\frac{\partial F(p_1p_2,pp_1+pp_2,pp_1-pp_2)}{\partial(pp_1-pp_2)} \times|p_1,p_2(+)\rangle, \quad (7.37)$$

and, if F depends only on $pp_1 + pp_2$, the dynamical spin vanishes.

---

1 F. Lurçat, Phys. Rev. **173**, 1461 (1968).
2 For a review of the conventional theory see M. Goldberger and K. Watson, *Collision Theory* (Wiley, New York, 1964), Chap. 8.
3 On the compatibility of frame dependent observables with Lorentz invariant theories, see R. L. Ingraham, Nuovo Cimento **39**, 131 (1965); G. N. Fleming, J. Math. Phys. **7**, 1959 (1967); S. McDonald, *ibid.* **10**, 1234 (1969).
4 A spin-one component in the spin spectrum of the neutral pion could, however, manifest itself in a rare three-photon decay mode. I am indebted to Professor H. Primakoff for this observation.
5 Besides Ref. 1 another recent conjecture on the group theoretic classification of unstable particle states has been put forward by L. S. Schulman, Ann. Phys. (N.Y.) **59**, 201 (1970).
6 S. Tomonaga, Progr. Theoret. Phys. (Kyoto) **1**, 27 (1946); J. Schwinger, Phys. Rev. **82**, 914 (1951).
7 A detailed discussion of these concepts is presented in the author's article mentioned in Ref. 3.
8 For the decay mode in question, see Ref. 4.
9 This approach is used in Ref. 2, pp. 36–40.
10 This approach is emphasized in S. Gasiovowicz, *Elementary Particle Physics* (Wiley, New York, 1966), Chap. 4.
11 T. Newton and E. Wigner, Rev. Mod. Phys. **21**, 400 (1949). See also G. N. Fleming, Phys. Rev. **137**, B188 (1965).
12 G. N. Fleming, J. Math. Phys. **9**, 193 (1968).

Probabilistic Interpretation of the Quantum Scattering Cross Section

Donald G. Babbitt*
Department of Mathematics, University of California, Los Angeles, California 90024

(Received 25 October 1971; Revised Manuscript Received 17 December 1971)

The probabilistic interpretation of the quantum scattering cross section in the case of potential scattering is discussed in terms of Poisson random measures on the impact parameter plane and the sphere of outgoing directions.

1. INTRODUCTION

In this paper we shall discuss the probabilistic interpretation of the scattering cross section for a quantum system described by a Hamiltonian $H = -\frac{1}{2}\sum_{j=1}^3(\partial/\partial x^j)^2 + V$ acting on $L^2(\mathbf{R}^3)$.[1] This is the quantum analog of the case we considered in an earlier paper.[2]

Almost all of the ingredients for a probabilistic description of a cross section in the quantum case already exist in the literature. The work of Hunziker,[3] where the quantum analog of the impact parameter

$$= J_\mu(\eta)V(\eta,\tau) - \tfrac{1}{2}\epsilon_{\mu\alpha\beta\gamma}K^\gamma(\eta)[\eta^\alpha U^\beta(\eta,\tau) - \eta^\beta U^\alpha(\eta,\tau)]$$
$$= \int d^4x\ \delta(\eta x - \tau)[J_\mu(\eta) - \tfrac{1}{2}\epsilon_{\mu\alpha\beta\gamma}K^\gamma(\eta)(\eta^\alpha x^\beta - \eta^\beta x^\alpha)]\mathcal{L}_{\rm int}(x)$$
$$= -\int d^4x\ \delta(\eta x - \tau)\{J_\mu(\eta) - \tfrac{1}{2}\epsilon_{\mu\alpha\beta\gamma}[x^\alpha K^\beta(\eta) - x^\beta K^\alpha(\eta)]\eta^\gamma\}\mathcal{L}_{\rm int}(x)$$
$$= -\int d^4x\ \delta(\eta x - \tau)e^{iK(\eta)x/\hbar}J_\mu(\eta)e^{-iK(\eta)x/\hbar}\mathcal{L}_{\rm int}(x)$$
$$= -\int d^4x\ \delta(\eta x - \tau)e^{iK(\eta)x/\hbar}J_\mu(\eta)\mathcal{L}_{\rm int}(\eta_\alpha\tau)e^{-iK(\eta)x/\hbar}. \quad (7.31)$$

Upon applying this operator to an unstable particle momentum eigenvector with vanishing kinematical spin, we have

$$W_\mu|p;\eta,\tau\rangle = (2\pi\hbar)^3\delta_\eta^3(P-p)J_\mu(\eta)\mathcal{L}_{\rm int}(\eta_\alpha\tau)|p;\eta,\tau\rangle. \quad (7.32)$$

Since the hyperplane rotation generator $J_\mu(\eta)$ commutes with $\mathcal{L}_{\rm int}(x)$ when $x_\mu = \eta_\mu\tau$ and

$$J_\mu(\eta)|p;\eta,\tau\rangle = \frac{ih}{2}\epsilon_{\mu\alpha\beta\gamma}p^\alpha\frac{\partial}{\partial p_\beta}\eta^\gamma|p;\eta,\tau\rangle, \quad (7.33)$$

we have

$$W_\mu|p;\eta,\tau\rangle = (2\pi\hbar)^3\delta_\eta^3(P-p)\mathcal{L}_{\rm int}(\eta_\mu\tau)\frac{ih}{2}\epsilon_{\mu\alpha\beta\gamma}p^\alpha\frac{\partial}{\partial p_\beta} \times|p;\eta,\tau\rangle\eta^\gamma. \quad (7.34)$$

Assuming only one decay mode, i.e., into the states $|p_1,p_2(+)\rangle$, we can write this as

$$W_\mu|p;\eta,\tau\rangle$$
$$= (2\pi\hbar)^3\int d^4p_1d^4p_2\ \delta_+(p_1^2 - m_1^2c^2)\delta_+(p_2^2 - m_2^2c^2) \times\delta_\eta^3(p_1+p_2-p)|p_1,p_2(+)\rangle\frac{ih}{2}\epsilon_{\mu\alpha\beta\gamma}p^\alpha\frac{\partial}{\partial p_\beta} \times\langle p_1,p_2(+)|\mathcal{L}_{\rm int}(\eta_\mu\tau)|p;\eta\tau\rangle\eta^\gamma$$
$$= (2\pi\hbar)^3\int d^4p_1d^4p_2\ \delta_+(p_1^2 - m_1^2c^2) \times\delta_+(p_2^2 - m_2^2c^2)\delta_\eta^3(p_1+p_2-p) \times e^{i(\eta p_1+\eta p_2-\eta p)\tau/\hbar}\frac{ih}{2}\epsilon_{\mu\alpha\beta\gamma}p^\alpha\frac{\partial}{\partial p_\beta} \times F(p_1p_2,pp_1,pp_2)\eta^\gamma\times|p_1,p_2(+)\rangle. \quad (7.35)$$

Now in the discussion following (7.21) we learned that an isotropic angular distribution in the rest frame of the decay products would occur if and only if F depended on pp_1 and pp_2 through the sum $pp_1 + pp_2$ and did not depend on $pp_1 - pp_2$. But this is just the condition that makes (7.35) vanish. Thus

$$\epsilon_{\mu\alpha\beta\gamma}p^\alpha\frac{\partial}{\partial p_\beta}\eta^\gamma F(p_1p_2,pp_1,pp_2) = \epsilon_{\mu\alpha\beta\gamma}p^\alpha\eta^\gamma\left(p_1^\beta\frac{\partial F}{\partial pp_1} + p_2^\beta\frac{\partial F}{\partial pp_2}\right) = \tfrac{1}{2}\epsilon_{\mu\alpha\beta\gamma}p^\alpha\eta^\gamma \times\left(\frac{(p_1+p_2)^\beta}{2}\frac{\partial F}{\partial(pp_1+pp_2)} + \frac{(p_1-p_2)^\beta}{2}\frac{\partial F}{\partial(pp_1-pp_2)}\right), \quad (7.36)$$

and the delta function $\delta_\eta^3(p_1+p_2-p)$ in (7.35) eliminates any contribution from the first term on the right of (7.36). The final result is

$$W_\mu|p;\eta,\tau\rangle = (2\pi\hbar)^3\int d^4p_1d^4p_2\,\delta_+(p_1^2 - m_1^2c^2)\delta_+(p_2^2 - m_2^2c^2) \times\delta_\eta^3(p_1+p_2-p)e^{i(\eta p_1+\eta p_2-\eta p)\tau/\hbar} \times\tfrac{1}{4}ih\,\epsilon_{\mu\alpha\beta\gamma}p^\alpha(p_1-p_2)^\beta\eta^\gamma \times\frac{\partial F(p_1p_2,pp_1+pp_2,pp_1-pp_2)}{\partial(pp_1-pp_2)} \times|p_1,p_2(+)\rangle, \quad (7.37)$$

and, if F depends only on $pp_1 + pp_2$, the dynamical spin vanishes.

---

1 F. Lurçat, Phys. Rev. **173**, 1461 (1968).
2 For a review of the conventional theory see M. Goldberger and K. Watson, *Collision Theory* (Wiley, New York, 1964), Chap. 8.
3 On the compatibility of frame dependent observables with Lorentz invariant theories, see R. L. Ingraham, Nuovo Cimento **39**, 131 (1965); G. N. Fleming, J. Math. Phys. **7**, 1959 (1967); S. McDonald, *ibid.* **10**, 1234 (1969).
4 A spin-one component in the spin spectrum of the neutral pion could, however, manifest itself in a rare three-photon decay mode. I am indebted to Professor H. Primakoff for this observation.
5 Besides Ref. 1 another recent conjecture on the group theoretic classification of unstable particle states has been put forward by L. S. Schulman, Ann. Phys. (N.Y.) **59**, 201 (1970).
6 S. Tomonaga, Progr. Theoret. Phys. (Kyoto) **1**, 27 (1946); J. Schwinger, Phys. Rev. **82**, 914 (1951).
7 A detailed discussion of these concepts is presented in the author's article mentioned in Ref. 3.
8 For the decay mode in question, see Ref. 4.
9 This approach is used in Ref. 2, pp. 36–40.
10 This approach is emphasized in S. Gasiovowicz, *Elementary Particle Physics* (Wiley, New York, 1966), Chap. 4.
11 T. Newton and E. Wigner, Rev. Mod. Phys. **21**, 400 (1949). See also G. N. Fleming, Phys. Rev. **137**, B188 (1965).
12 G. N. Fleming, J. Math. Phys. **9**, 193 (1968).

Probabilistic Interpretation of the Quantum Scattering Cross Section

Donald G. Babbitt*
Department of Mathematics, University of California, Los Angeles, California 90024

(Received 25 October 1971; Revised Manuscript Received 17 December 1971)

The probabilistic interpretation of the quantum scattering cross section in the case of potential scattering is discussed in terms of Poisson random measures on the impact parameter plane and the sphere of outgoing directions.

1. INTRODUCTION

In this paper we shall discuss the probabilistic interpretation of the scattering cross section for a quantum system described by a Hamiltonian $H = -\frac{1}{2}\sum_{j=1}^3(\partial/\partial x^j)^2 + V$ acting on $L^2(\mathbf{R}^3)$.[1] This is the quantum analog of the case we considered in an earlier paper.[2]

Almost all of the ingredients for a probabilistic description of a cross section in the quantum case already exist in the literature. The work of Hunziker,[3] where the quantum analog of the impact parameter

plane (among other things) has been defined, has been especially important for us. In fact our contribution is to make a slight refinement of Hunziker's treatment by using a Poisson random measure on the impact parameter plane to describe an infinite beam of incoming particles and a Poisson random measure on an appropriate subset of S^2 to describe the directions of the scattered beam. [$S^2 = \{(k_1, k_2, k_3) \in \mathbb{R}^3: k_1^2 + k_2^2 + k_3^2 = 1\}$.] Most of the paper will be devoted to embedding Hunziker's treatment of the quantum cross section into our setting. The only thing which will need proving (and the proofs are very easy) is that the limit of a finite beam yields an infinite beam in our sense and that the limit of the cross section of the finite beam, as the beam becomes infinite, equals the cross section of the infinite beam.

2. CROSS-SECTIONAL RANDOM MEASURES FOR FINITE BEAMS

As we mentioned in the introduction, we will be concerned with quantum scattering for a system described by a Hamiltonian $H = -\frac{1}{2}\Delta + V(x)$ acting on $L^2(\mathbb{R}^3)$. We will make several assumptions about H. The first, of course, is that H is self-adjoint in some sense, e.g., V is a Kato potential relative to $H_0 = -\frac{1}{2}\Delta$.[4] Let $U(t) = \exp(iHt)$. The second is that there exists a unitary propagator $U_0(t): -\infty < t < \infty$, associated with (H_0, H),[5] such that

$$\operatorname{s-lim}_{t \to +\infty} U(-t)U_0(t) = \Omega^{\pm} \tag{2.1}$$

exists with

$$\text{range } (\Omega^+) = \text{range } (\Omega^-); \tag{2.2}$$

i.e., we are *assuming asymptotic completeness*. Then $S = (\Omega^+)^{-1}\Omega^-$ will be the scattering matrix for the system (H_0, H).

Remark 1: If V is a "short range" potential, $U_0(t) = \exp(iH_0t)$. In fact, we could take as the definition of "short range" the existence of (2.1) with property (2.2) when $U_0(t) = \exp(iH_0t)$.

Let φ be a vector state with nonzero mean momentum, i.e., $\varphi \in L^2(\mathbb{R}^3)$, $\|\varphi\| = 1$, and $((\varphi, p_1\varphi), (\varphi, p_2\varphi), (\varphi, p_3\varphi)) \equiv \bar{p} \neq 0$, where $p_j = (i)^{-1}(\partial/\partial x^j)$, $j = 1, 2, 3$. An *incoming φ-beam of n particles with intensity ν*, denoted by $\mathcal{T}_{n,\nu,\varphi}$, is a set of n independent identically distributed random variables $\{X_1, \ldots, X_n\}$ with values in the *impact parameter plane*

$$I_{\bar{p}} = \{x \in \mathbb{R}^3 : x^1\bar{p}_1 + x^2\bar{p}_2 + x^3\bar{p}_3 \equiv x \cdot \bar{p} = 0\}.$$

We further assume that each X_j is uniformly distributed over the square A_n in $I_{\bar{p}}$ which is centered at 0 and with sides of length $(n/\nu)^{1/2}$.

The physical interpretation of the incoming φ beam is that it consists of n random incoming particles whose states in the distant past are asymptotic to $\{U_0(t)[\varphi^{X_1}], \ldots, U_0(t)[\varphi^{X_n}]\}$, where $\varphi^a(x) \equiv \varphi(x + a)$, $a \in I_{\bar{p}}$.

The *outgoing beam* will be the collection of random states $\{S[\varphi^{X_1}], \ldots, S[\varphi^{X_n}]\}$. The *physical interpretation of the outgoing beam* is that the outgoing states are asymptotically in the future like $\{U_0(t)S[\varphi^{X_1}], \ldots, U_0(t)S[\varphi^{X_n}]\}$.

In order to define cross-sectional random measure and the cross section of the beam, we must restrict our incoming states to those with a "reasonably sharp" momentum. We now make this more precise. Let φ be as above and let

$$C(\varphi) \equiv \{\lambda p : \lambda > 0, p \in \operatorname{supp}(\tilde{\varphi})\}$$

be the cone generated by the support of the Fourier transform $\tilde{\varphi}$ of φ. $C(\varphi)$ will be called *the cone of forward momenta* for $\mathcal{T}_{n,\nu,\varphi}$. We *require* from now on that $C(\varphi) \neq \mathbb{R}^3$.

We can now describe the cross sectional random measure for the beam $\mathcal{T}_{n,\nu,\varphi}$. First of all, it will be a random measure on $S^2(\varphi) \equiv S^2 \cap [C(\varphi)]'$, where $[\,]'$ denotes the complement of $[\,]$. $S^2(\varphi)$ is the *set of non-forward directions* for $\mathcal{T}_{n,\nu,\varphi}$. If Σ is a Borel subset of $S^2(\varphi)$, let $N_{n,\nu,\varphi}(\Sigma)$ be the (random) number of particles in the outgoing beam which asymptotically have their momentum directions in Σ. It is clear that this is a binomial random variable with parameter n (the maximum number of particles in the beam with outgoing directions in Σ) and with probability that the jth particle in the beam has outgoing momentum direction in Σ equal to

$$P_{n,\nu,\varphi} \equiv (\text{area } A_{n,\nu})^{-1} \int_{A_{n,\nu}} \int_{C(\Sigma)} |S[\varphi^a](p)|^2 dp da. \tag{2.3}$$

The random measure on $S^2(\varphi)$ defined by

$$\Phi_{n,\nu,\varphi}(\Sigma) \equiv N_{n,\nu,\varphi}(\Sigma)/\nu$$

is the *cross-sectional random measure for the beam* $\mathcal{T}_{n,\nu,\varphi}$. The measure on $S^2(\varphi)$ defined by

$$\sigma_{n,\nu}(\varphi, \Sigma) \equiv E(\Phi_{n,\nu,\varphi}(\Sigma)),$$

where $E(\cdot)$ denotes the expectation of $(\cdot)$, is the *cross section of the beam* $\mathcal{T}_{n,\nu,\varphi}$. From (2.3) and the fact that $N_{n,\nu,\varphi}(\Sigma)$ is binomial, it follows that

$$\sigma_n(\varphi, \Sigma) = \int_{A_{n,\nu}} \int_{C(\Sigma)} |S[\varphi^a](p)|^2 dp da, \tag{2.4}$$

where $C(\Sigma) \equiv \{\lambda x : \lambda > 0, x \in \Sigma\}$. Thus the cross section evaluated at Σ is the expected number of particles in the beam which asymptotically have outgoing momenta in Σ divided by the intensity. We shall see in the case of an infinite beam that the cross section is independent of the intensity of the incoming beam.

Remark 2: If we let $R = S - 1$, then

$$\sigma_{n,\nu}(\varphi, \Sigma) = \int_{A_{n,\nu}} \int_{C(\Sigma)} |R[\varphi^a](p)|^2 dp da, \tag{2.5}$$

since $\operatorname{supp}(\tilde{\varphi}) \cap C(\Sigma) = \{0\}$.

3. THE CROSS-SECTIONAL RANDOM MEASURE AND CROSS SECTION OF AN INFINITE BEAM

Infinite beams are somewhat more subtle to handle than finite beams. Intuitively an infinite φ beam should be given by an infinite collection of independent *uniformly distributed* random variables $\{X_1, X_2, \ldots,\}$ with values in $I_{\bar{p}}$. We would then reason essentially as we did in Sec. 2. The difficulty is that there is no such thing as a uniformly distributed

random variable with values in $I_{\bar{p}}$. Thus we must find an alternate way to describe an infinite beam. Fortunately a mathematical object exists which will allow us to do so. It is the notion of a Poisson process (or, if one prefers, random measure) based on a measure space which in our case will be either $I_{\bar{p}}$ with Lebesgue measure or S^2 with the canonical measure. See our earlier paper and references given there.[2]

We shall now give the definition of an infinite beam. φ is assumed to satisfy the same conditions as in Sec. 2. An infinite φ beam with intensity $\nu(>0)$ is a Poisson process $\mathcal{T}_{\nu,\varphi}$ on the impact parameter plane $I_{\bar{p}}$ with measure νda where da is Lebesgue measure on $I_{\bar{p}}$. The *physical interpretation* of $\mathcal{T}_{\nu,\varphi}$ is as follows. Let ω be a generic sample point and let $\{X_1(\omega), \cdots\} = \text{supp}(\mathcal{T}_{\nu,\varphi}(\omega))$. [Recall that $\mathcal{T}_{\nu,\varphi}(\omega)$ is a nonnegative integer-valued measure on $I_{\bar{p}}$. We may also assume the support is at most countable and $\mathcal{T}_{\nu,\varphi}(\omega)\{X_j(\omega)\} = 1$ for $j = 1, \cdots$.] The incoming beam will consist of particles which in the distant past are asymptotic to

$$\{U_0(t)\varphi^{X_1(\omega)}, \ldots, U_0(t)\varphi^{X_n(\omega)}, \ldots, t \to -\infty\}.$$

Remark 3: Finite φ beams may be treated in the same way as infinite φ beams. In fact, if A is a Borel subset of $I_{\bar{p}}$, let

$$\tilde{\mathcal{T}}^{(\omega)}_{n,\nu,\varphi}(A) = \text{the number of } \{X_1(\omega), \ldots, X_n(\omega)\} \text{ which lie in } A,$$

where $\{X_1, \ldots, X_n\} = \mathcal{T}_{n,\nu,\varphi}$ (see Sec. 2). $\tilde{\mathcal{T}}_{n,\nu,\varphi}$ will be a nonnegative integer-valued random measure on $I_{\bar{p}}$, supported by $A_{n,\nu}$. It is clear we can recover $\mathcal{T}_{n,\nu,\varphi}$ from $\tilde{\mathcal{T}}_{n,\nu,\varphi}$. The following proposition will show that the limit of finite φ beams is an infinite φ beam.

Proposition 1: $\{\tilde{\mathcal{T}}_{n,\nu,\varphi}(A)\}$ converges in distribution to $\mathcal{T}_{\nu,\varphi}(A)$ for each Borel subset A of $I_{\bar{p}}$.

Proof: $\tilde{\mathcal{T}}_{n,\nu,\varphi}(A)$ is a Bernoulli random variable with parameters n and $p_{n,\nu,\varphi} = [\text{Area } (A_{n,\nu})]^{-1} \int_{A\cap A_{n,\nu}} da$. Then, from the well-known convergence theorem of Bernoulli random variables to Poisson variables,[6] our proposition is proved.

The *cross-sectional random measure* of the beam $\mathcal{T}_{\nu,\varphi}$ is defined to be Poisson process $\Phi_{\nu,\varphi}$ on $S^2(\varphi)$ with measure

$$\sigma(\varphi, \Sigma) \equiv \int_{I_{\bar{p}}} \int_{C(\Sigma)} |\widetilde{R[\varphi^a]}(p)|^2 dp da. \tag{3.1}$$

[We allow $\sigma(\varphi, \Sigma)$ to be infinite.] The measure $\sigma(\varphi, \cdot)$ is called the *cross section* of the beam $\mathcal{T}_{\nu,\varphi}$. Note that it is independent of the intensity.

The *physical interpretation* of $\Phi_{\nu,\varphi}$ is analogous to the finite beam case. In particular, when Σ is a Borel subset of $S^2(\varphi)$, $\Phi_{\nu,\varphi}(\Sigma) = N_{\nu,\varphi}(\Sigma)/\nu$, where $N_{\nu,\varphi}(\Sigma)$ is the number (a random variable!) of particles in the incoming φ beam whose momenta have directions in Σ in the distant future. $\sigma(\varphi, \Sigma)$ is the *expected number* of such particles divided by the intensity ν.

The definitions of $\Phi_{n,\nu,\varphi}$ and $\sigma_{n,\nu}(\varphi, \cdot)$, the proposition above, and the following proposition justify this physical interpretation.

Proposition 2: The sequence $\{\Phi_{n,\nu,\varphi}(\Sigma)\}$ converges in distribution to $\Phi_{\nu,\varphi}(\Sigma)$ for each Borel subset Σ of $S^2(\varphi)$. Furthermore,

$$\lim_{n\to\infty} \sigma_{n,\nu}(\varphi, \Sigma) = \sigma(\varphi, \Sigma). \tag{3.2}$$

Proof: It suffices to prove (3.2) by the theorem on convergence of Bernoulli random variables to a Poisson random variable.[6] But (3.2) follows from (3.1), (2.5), and the monotone convergence theorem.

4. CONCLUDING REMARKS

Remark 4: In this paper we have only stressed the mathematical description and physical interpretation of certain scattering experimental parameters for a simple but important class of quantum mechanical systems. We would like to briefly describe how the hard analytic problems of scattering theory are related to our present discussion.

The first analytic problem has already been discussed, namely, the choice of an appropriate "free" propagator $\{U_0(t)\}$ and proof of (2.1) and (2.2). For example, see Amrein *et al.*,[5] Dollard,[5] and Hunziker[3] and references given there.

The other main analytic problem is to obtain a more tractable expression for the cross section (2.5). This has been described in a very lucid way by Hunziker.[3] We reproduce the main results of his discussion.

He shows that under reasonable conditions on V and φ, one has

$$\sigma(\varphi, \Sigma) = (2\pi)^2 \int_{C(\Sigma)} dp \int_{p'^2 = p^2} |T(p,p')|^2 |\tilde{\varphi}(p')|^2 d\Omega_E(p'), \tag{4.1}$$

where $d\Omega_E(p')$ is the canonical measure on the sphere $p'^2 = p^2$ and where $T(p,p')$ is obtained roughly as follows. One solves the Lippman–Schwinger equations

$$T(z) = V + V(H_0 - z)^{-1}T(z) \tag{4.2}$$

in momentum space for $\text{Im}z > 0$. Let $T(p,p';z)$ be the kernel for $T(z)$. Then

$$T(p,p') \equiv \lim_{\epsilon\to 0^+} T(p,p';p^2 + \tfrac{1}{2}i\epsilon).$$

It should be remarked that the study of (4.2) is also important for establishing (2.2).

Remark 5: The differential cross section enters the picture in the following way.[7] Let $\{\varphi_n\}$ be a sequence of states such that $|\tilde{\varphi}_n|^2$ converge to $\delta(\cdot - \bar{p})$. This corresponds to the incoming beam having a sharp momentum $\bar{p}$ which of course cannot be realized as a physical state. (Intuitively, in fact, $\varphi(x) = [\delta(0)^{-1/2} \exp(i\bar{p}\cdot x))]$. Then, under suitable conditions, (4.1) converges to

$$\sigma(\bar{p}, \Sigma) = (2\pi)^2 \int_\Sigma |T[(p,\bar{p})]|^2 \cdot [2E(\bar{p})]^{1/2} d\Omega(p),$$

where $E(\bar{p}) = \frac{1}{2}\bar{p}^2$, $p = (|\bar{p}|, \theta(p), \varphi(p))$ in spherical coordinates and $d\Omega(p) = \sin\theta(p) d\theta(p) d\varphi(p)$. The measure $\sigma(\bar{p}, \cdot)$ on $S^2 - \{\bar{p}/|\bar{p}|\}$ is called the cross section of the idealized incoming beam with sharp momentum $\bar{p}$. The density of this measure with regard to $d\Omega$;

$$\frac{d\sigma}{d\Omega}(\bar{p},\theta,\varphi) = [2E(\bar{p})]^{1/2}|T(p,\bar{p})|^2,$$

is the differential cross section of this idealized beam.

* This work was supported by NSF Grant GP-18127.

1 We have set $\hbar$ and m equal to 1 for simplicity.

2 D. G. Babbitt, J. Math. Phys. 12, 53 (1971).

3 The lecture notes of W. Hunziker in *Lectures in Theoretical Physics*, edited by A. O. Barut and W. E. Brittin (Gordon and Breach, New York, 1968), Vol. XA.

4 E. Nelson, J. Math. Phys. 5, 332 (1964), see Appendix.

5 W. O. Amrein, Ph. A. Martin, and B. Misra, "On the Asymptotic Condition of Scattering Theory," preprint from Institute de Physique Théorique, Genève (1969); J. D. Dollard, Rocky Mountain J. Math. 1, 5 (1971).

6 W. Feller, *An Introduction to Probability Theory and Its Application* (Wiley, New York, 1965), 2nd ed., p. 142.

7 We again follow Hunziker's treatment.

Multipole Radiation Solutions for a Class of Massive Particle Wave Equations*

Jack P. Hajj
Department of Mathematics, Suffolk University, Boston, Massachusetts
and
David L. Weaver
Department of Physics, Tufts University, Medford, Massachusetts

(Received 11 October 1971; Revised Manuscript Received 29 November 1971)

In this paper, the multipole radiation solutions are derived for the integer and half-integer spin massive particle wave equations of Weaver, Hammer, and Good. The rest system and massless limits of the solutions are found, the latter limit being shown to coincide with previous work by the present authors. Finally, as an application, the massive particle description is second quantized in terms of the multipole solutions and compared with some other formulations.

I. INTRODUCTION

It has been shown[1] that, for massless particles with integer or half-integer spin S, one may construct a Schrödinger type wave equation and solve this equation for the multipole radiation solutions, all in complete parallel to a treatment of the photon and Maxwell's equations. The purpose of this paper is to extend the construction of multipole solutions for any spin to include also particles with mass m, to examine the rest system and massless limits of the massive theory, and to secondquantize the massive theory in terms of angular momentum eigenstates.

The basic description to be used for massive particles should have a well-defined Hamiltonian operator, and a wavefunction that has simple Lorentz transformation properties and is not subject to auxiliary conditions. Such a description exists and has been investigated in detail by Weaver, Hammer, and Good.[2] Following Ref. 2, the wavefunction ψ, representing a particle of mass m and spin s, is $2(2s+1)$-dimensional and satisfies the wave equation

$$H\psi(\mathbf{x},t) = i\frac{\partial}{\partial t}\psi(\mathbf{x},t) \tag{1}$$

(the units are $\hbar = c = 1$). The Hamiltonian operator H is given by

$$H = E_0 S\beta S^{-1}, \tag{2}$$

where S is derived in Ref. 2 to have the form

$$S = \sum_{n=0}^{2s} d_n\left(\frac{\alpha\cdot\mathbf{P}}{|\mathbf{P}|}\right)^n\beta^n. \tag{3}$$

The coefficients d_n depend only on $|\mathbf{P}|$ the magnitude of the momentum operator $\mathbf{P}(=-i\nabla)$. The explicit forms of S and S^{-1} are tabulated in Ref. 2 for spins $0, \frac{1}{2}, 1$, and $\frac{3}{2}$, and general formulas from which one may derive the coefficients d_n for arbitrary spin are given in Ref. 3. Also, one has the definitions

$$E_0 \equiv \sqrt{|\mathbf{P}|^2+m^2},\quad \alpha \equiv \frac{1}{s}\begin{pmatrix}\mathbf{s} & 0\\ 0 & -\mathbf{s}\end{pmatrix},\quad \beta \equiv \begin{pmatrix}0 & I\\ I & 0\end{pmatrix}, \tag{4}$$

where I and $\mathbf{s}$ are the $(2s+1)$-dimensional unit and Hermitian spin matrices, respectively.

As shown in Ref. 2, Eq. (1) is completely equivalent to

$$E_0\hat{S}_t\beta\hat{S}_t^{-1}\psi = i\frac{\partial}{\partial t}\psi, \tag{5}$$

where the operator $\hat{S}_t$ is

$$\hat{S}_t = \exp[s\epsilon_0\tanh^{-1}(|\mathbf{P}|/E_0)\boldsymbol{\alpha}\cdot(\mathbf{P}/|\mathbf{P}|)]. \tag{6}$$

The operator

$$\epsilon_0 \equiv i\frac{\partial}{\partial t}\Big/\left|i\frac{\partial}{\partial t}\right|$$

is the sign of the energy operator. It differentiates between particle and antiparticle states.

Equations (1) and (2) imply that ψ satisfies the Klein-Gordon equation for mass m, and with respect to the homogeneous Lorentz group, ψ transforms as the direct sum of the $(s,0)$ and $(0,s)$ representations.

The outline of this paper is as follows: In Sec. II the multipole solutions of Eq. (1) are derived and discussed. The rest system and massless limits are then presented in Secs. III and IV with the latter limit being shown to coincide with previous work.[1] Finally, the massive description is secondquantized in terms of angular momentum states and compared with some other formulations in Sec. V.

II. MULTIPOLE SOLUTIONS

Making the substitution $\psi(\mathbf{x},t) = W(\mathbf{x})e^{-iEt}$ in Eq. (1), one obtains the eigenvalue problem

$$HW(\mathbf{x}) = EW(\mathbf{x}). \tag{7}$$

$$\frac{d\sigma}{d\Omega}(\bar{p},\theta,\varphi) = [2E(\bar{p})]^{1/2}|T(p,\bar{p})|^2,$$

is the differential cross section of this idealized beam.

* This work was supported by NSF Grant GP-18127.

1 We have set $\hbar$ and m equal to 1 for simplicity.

2 D. G. Babbitt, J. Math. Phys. 12, 53 (1971).

3 The lecture notes of W. Hunziker in *Lectures in Theoretical Physics*, edited by A. O. Barut and W. E. Brittin (Gordon and Breach, New York, 1968), Vol. XA.

4 E. Nelson, J. Math. Phys. 5, 332 (1964), see Appendix.

5 W. O. Amrein, Ph. A. Martin, and B. Misra, "On the Asymptotic Condition of Scattering Theory," preprint from Institute de Physique Théorique, Genève (1969); J. D. Dollard, Rocky Mountain J. Math. 1, 5 (1971).

6 W. Feller, *An Introduction to Probability Theory and Its Application* (Wiley, New York, 1965), 2nd ed., p. 142.

7 We again follow Hunziker's treatment.

Multipole Radiation Solutions for a Class of Massive Particle Wave Equations*

Jack P. Hajj
Department of Mathematics, Suffolk University, Boston, Massachusetts
and
David L. Weaver
Department of Physics, Tufts University, Medford, Massachusetts

(Received 11 October 1971; Revised Manuscript Received 29 November 1971)

In this paper, the multipole radiation solutions are derived for the integer and half-integer spin massive particle wave equations of Weaver, Hammer, and Good. The rest system and massless limits of the solutions are found, the latter limit being shown to coincide with previous work by the present authors. Finally, as an application, the massive particle description is second quantized in terms of the multipole solutions and compared with some other formulations.

I. INTRODUCTION

It has been shown[1] that, for massless particles with integer or half-integer spin S, one may construct a Schrödinger type wave equation and solve this equation for the multipole radiation solutions, all in complete parallel to a treatment of the photon and Maxwell's equations. The purpose of this paper is to extend the construction of multipole solutions for any spin to include also particles with mass m, to examine the rest system and massless limits of the massive theory, and to secondquantize the massive theory in terms of angular momentum eigenstates.

The basic description to be used for massive particles should have a well-defined Hamiltonian operator, and a wavefunction that has simple Lorentz transformation properties and is not subject to auxiliary conditions. Such a description exists and has been investigated in detail by Weaver, Hammer, and Good.[2] Following Ref. 2, the wavefunction ψ, representing a particle of mass m and spin s, is $2(2s+1)$-dimensional and satisfies the wave equation

$$H\psi(\mathbf{x},t) = i\frac{\partial}{\partial t}\psi(\mathbf{x},t) \tag{1}$$

(the units are $\hbar = c = 1$). The Hamiltonian operator H is given by

$$H = E_0 S\beta S^{-1}, \tag{2}$$

where S is derived in Ref. 2 to have the form

$$S = \sum_{n=0}^{2s} d_n\left(\frac{\alpha\cdot\mathbf{P}}{|\mathbf{P}|}\right)^n\beta^n. \tag{3}$$

The coefficients d_n depend only on $|\mathbf{P}|$ the magnitude of the momentum operator $\mathbf{P}(=-i\nabla)$. The explicit forms of S and S^{-1} are tabulated in Ref. 2 for spins $0, \frac{1}{2}, 1$, and $\frac{3}{2}$, and general formulas from which one may derive the coefficients d_n for arbitrary spin are given in Ref. 3. Also, one has the definitions

$$E_0 \equiv \sqrt{|\mathbf{P}|^2+m^2},\quad \alpha \equiv \frac{1}{s}\begin{pmatrix}\mathbf{s} & 0\\ 0 & -\mathbf{s}\end{pmatrix},\quad \beta \equiv \begin{pmatrix}0 & I\\ I & 0\end{pmatrix}, \tag{4}$$

where I and $\mathbf{s}$ are the $(2s+1)$-dimensional unit and Hermitian spin matrices, respectively.

As shown in Ref. 2, Eq. (1) is completely equivalent to

$$E_0\hat{S}_t\beta\hat{S}_t^{-1}\psi = i\frac{\partial}{\partial t}\psi, \tag{5}$$

where the operator $\hat{S}_t$ is

$$\hat{S}_t = \exp[s\epsilon_0\tanh^{-1}(|\mathbf{P}|/E_0)\boldsymbol{\alpha}\cdot(\mathbf{P}/|\mathbf{P}|)]. \tag{6}$$

The operator

$$\epsilon_0 \equiv i\frac{\partial}{\partial t}\Big/\left|i\frac{\partial}{\partial t}\right|$$

is the sign of the energy operator. It differentiates between particle and antiparticle states.

Equations (1) and (2) imply that ψ satisfies the Klein-Gordon equation for mass m, and with respect to the homogeneous Lorentz group, ψ transforms as the direct sum of the $(s,0)$ and $(0,s)$ representations.

The outline of this paper is as follows: In Sec. II the multipole solutions of Eq. (1) are derived and discussed. The rest system and massless limits are then presented in Secs. III and IV with the latter limit being shown to coincide with previous work.[1] Finally, the massive description is secondquantized in terms of angular momentum states and compared with some other formulations in Sec. V.

II. MULTIPOLE SOLUTIONS

Making the substitution $\psi(\mathbf{x},t) = W(\mathbf{x})e^{-iEt}$ in Eq. (1), one obtains the eigenvalue problem

$$HW(\mathbf{x}) = EW(\mathbf{x}). \tag{7}$$

Defining

$$\Sigma \equiv \begin{pmatrix} \mathbf{s} & 0 \\ 0 & \mathbf{s} \end{pmatrix}, \tag{8}$$

one has the commuting set of Hermitian operators H, J^2, J_3, Σ^2, and λ where

$$J_i = \epsilon_{ijk} x_j \left(-i\frac{\partial}{\partial x_k}\right) + \Sigma_i, \qquad \lambda = \Sigma \cdot \mathbf{P}/|\mathbf{P}|. \tag{9}$$

(Latin indices run from 1 to 3 and repeated indices are summed over.) The operator λ is the helicity operator with eigenvalues $\eta = -s, -s+1, \ldots, s$.

The problem is, therefore, to find physical solutions such that

$$HW = EW, \tag{10}$$

$$J^2 W = J(J+1)W, \tag{11}$$

$$J_3 W = MW, \quad -J \le M \le J, \tag{12}$$

$$\Sigma^2 W = s(s+1)W, \tag{13}$$

$$\lambda W = \eta W, \quad -s \le \eta \le s. \tag{14}$$

The simultaneous eigenfunctions of J^2, J_3, and Σ^2 are the spinor spherical harmonics $\Psi^M_{J,l,s}$ defined in Ref. 1, and the most general form for $W(\mathbf{x})$ is

$$W(\mathbf{x}) = \sum_{l=|J-s|}^{J+s} \left[h_l(r) \begin{pmatrix} \Psi^M_{J,l,s}(\theta,\varphi) \\ 0 \end{pmatrix} + k_l(r) \begin{pmatrix} 0 \\ \Psi^M_{J,l,s}(\theta,\varphi) \end{pmatrix} \right] \tag{15}$$

(r, θ, and φ are the spherical polar coordinates of $\mathbf{x}$), with the $2(2J+1)$ [or $2(2s+1)$ if $J < s$] radial functions (with the other indices suppressed) to be determined.

Defining $\epsilon \equiv E/|E|$, one can show that Eqs. (10) and (14) lead to the following equation for W:

$$\epsilon\beta W(\mathbf{x}) = \begin{bmatrix} \exp\left[-2\epsilon\eta \tanh^{-1}\left(\frac{\sqrt{E^2-m^2}}{|E|}\right)\right] & 0 \\ 0 & \exp\left[2\epsilon\eta \tanh^{-1}\left(\frac{\sqrt{E^2-m^2}}{|E|}\right)\right] \end{bmatrix} W(\mathbf{x}). \tag{16}$$

Substituting Eq. (15) into Eq. (16) one obtains

$$h_l(r) = \epsilon\delta^2(E,\eta) k_l(r), \qquad |J-s| \le l \le J+s, \tag{17}$$

where

$$\delta(E,\eta) = \exp\left[\epsilon\eta \tanh^{-1}\left(\frac{\sqrt{E^2-m^2}}{|E|}\right)\right] = \left(\frac{|E| + \sqrt{E^2-m^2}}{|E| - \sqrt{E^2-m^2}}\right)^{\epsilon\eta/2} \tag{18}$$

The helicity eigenvalue equation leads to the following equations for the k_l:

$$\mathbf{S}\cdot\mathbf{P}\left(\sum_{l=|J-s|}^{J+s} k_l \Psi^M_{J,l,s}\right) = \eta\sqrt{E^2-m^2}\left(\sum_{l=|J-s|}^{J+s} k_l \Psi^M_{J,l,s}\right). \tag{19}$$

This set of equations for the k_l has the same form as the eigenvalue problem for the massless Hamiltonian which was solved in Ref. 1. The radial functions k_l are, therefore, given by

$$k_l(r) = [C_l(\eta)/\rho^{1/2}] H_{l+1/2}(\rho),$$

where $\rho \equiv \sqrt{E^2-m^2}\, r$ and $H_{l+1/2}$ are Bessel or Hankel functions. The coefficients $C_l(\eta)$ are determined by the following set of homogeneous equations:

$$\begin{aligned} \tfrac{1}{2} i a^{J,s}_{|J-s|} C_{|J-s|+1}(\eta) &= \eta C_{|J-s|}(\eta), \\ -\tfrac{1}{2} i a^{J,s}_{J+s-1} C_{J+s-1}(\eta) &= \eta C_{J+s}(\eta), \\ -\tfrac{1}{2} i a^{J,s}_{l-1} C_{l-1}(\eta) + \tfrac{1}{2} i a^{J,s}_{l} C_{l+1}(\eta) &= \eta C_l(\eta), \\ |J-s|+1 \le l \le J+s-1, \end{aligned} \tag{20}$$

where the coefficients $a^{J,s}_l$ are defined, as in Ref. 1, to be

$$a^{J,s}_l \equiv \left(\frac{(l+s+J+2)(l+s-J+1)(l-s+J+1)(J+s-l)}{(2l+1)(2l+3)}\right)^{1/2}. \tag{21}$$

Equations (20) are similar to the set of equations found in Ref. 1 for the massless particle case with the important difference that now the eigenvalue η ranges from $-s$ to s rather than being restricted to the maximum and minimum values as in the massless theory. Thus, in the massive theory, the eigenvalues of total angular momentum J are permitted to have values less than s.

Equations (20) can be put in the matrix form

$$A_{J,s} C(\eta) = \eta C(\eta), \tag{22}$$

where $A_{J,s}$ is a $(2s+1)\times(2s+1)[(2J+1)\times(2J+1)]$ Hermitian matrix for $J \ge s (J < s)$, and $C(\eta)$ is the column vector

$$C(\eta) = \begin{bmatrix} C_{J+s}(\eta) \\ \cdot \\ \cdot \\ \cdot \\ C_{|J-s|}(\eta) \end{bmatrix}. \tag{23}$$

The eigenvalues of $A_{J,s}$ are found to be $-s, -s+1, \ldots, s$ for $J \ge s$, and $-J, -J+1, \ldots, J$ for $J < s$ so for $J < s$ not all the $2s+1$ helicity states are possible. However, since the eigenvalues of $A_{J,s}$ are possible eigenvalues of λ, the solutions of Eq. (22) exist. Furthermore, because of the Hermiticity of $A_{J,s}$, the coefficients $C_l(\eta)$ can be chosen to satisfy the following orthogonality and completeness relations

$$\sum_{l=|J-s|}^{J+s} C_l^*(\eta')C_l(\eta) = \delta_{\eta\eta'}, \quad \sum_{\eta} C_{l'}^*(\eta)C_l(\eta) = \delta_{ll'}, \tag{24}$$

(where * indicates complex conjugation). Also, since $A_{J,s}^* = -A_{J,s}$, the $C_l(\eta)$ can be chosen to satisfy

$$C_l^*(\eta) = C_l(-\eta). \tag{25}$$

Finally, the complete solution to the eigenvalue problem is

$$W_{E,\eta,J,M}^s(\mathbf{x}) = m^s \sum_{l=|J-s|}^{J+s} C_l(\eta)\rho^{-1/2}H_{l+1/2}(\rho) \times \begin{bmatrix} \delta(E,\eta)\epsilon\Psi_{J,l,s}^M(\theta,\varphi) \\ \delta^{-1}(E,\eta)\Psi_{J,l,s}^M(\theta,\varphi) \end{bmatrix} \tag{26}$$

(the factor m^s is included to facilitate the massless limit). For solutions $W_{E,\eta,J,M}^s(\mathbf{x})$ regular at the origin $\rho = 0$, one uses Bessel functions for $H_{l+1/2}$. To obtain outgoing spherical waves one uses, as in the massless case, Hankel functions of the first (second) kind,[4] $H^{(1)}$ ($H^{(2)}$), when $E > 0$ ($E > 0$).

In the case of outgoing spherical waves, the complete solution is

$$\Psi^{\pm}(J,M,s,\eta,|E|;\mathbf{x},t) = m^s \sum_{l=|J-s|}^{J+s} C_l(\eta)\rho^{-1/2}H_{l+1/2}^{(1,2)}(\rho) \times \begin{bmatrix} \delta(\pm|E|,\eta)\epsilon\Psi_{J,l,s}^M(\theta,\varphi) \\ \delta^{-1}(\pm|E|,\eta)\Psi_{J,l,s}^M(\theta,\varphi) \end{bmatrix} e^{\mp i|E|t}. \tag{27}$$

The upper (lower) signs are for positive (negative) E solutions and the first (second) kind of Hankel function is to be used with them.

Using Eq. (27) one may now construct the definite parity multipole solutions. From Ref. 2 the parity operator Π is defined by

$$\Pi \equiv i\beta P, \tag{28}$$

where $Pf(\mathbf{x}) = f(-\mathbf{x})$. Thus, for half-integer spin s the parity eigenstates are

$$\psi_{\text{mag}}^{\pm}(J,M,s,\eta,|E|;\mathbf{x},t) \equiv [1 + (-1)^{J-1/2}\Pi]\psi^{\pm}(J,M,s,\eta,|E|;\mathbf{x},t),$$
$$\psi_{\text{elec}}(J,M,s,\eta,|E|;\mathbf{x},t) \equiv [1 + (-1)^{J+1/2}\Pi]\psi^{\pm}(J,M,s,\eta,|E|;\mathbf{x},t), \tag{29}$$

the motivation being that

$$\Pi\psi_{\text{mag}} = (-1)^{J-1/2}\psi_{\text{mag}}, \quad \Pi\psi_{\text{elec}} = (-1)^{J+1/2}\psi_{\text{elec}}. \tag{30}$$

For integer s one defines

$$\psi_{\text{mag}}(J,M,s,\eta,|E|;\mathbf{x},t) \equiv [1 - i(-1)^J\Pi]\psi^{\pm}(J,M,s,\eta,|E|;\mathbf{x},t),$$
$$\psi_{\text{elec}}(J,M,s,\eta,|E|;\mathbf{x},t) \equiv [1 + i(-1)^J\Pi]\psi^{\pm}(J,M,s,\eta,|E|;\mathbf{x},t). \tag{31}$$

In this case the appropriate parity operator is $\beta P = -i\Pi$ and one finds

$$-i\Pi\psi_{\text{mag}} = (-1)^J\psi_{\text{mag}}, \quad -i\Pi\psi_{\text{elec}} = (-1)^{J+1}\psi_{\text{elec}}. \tag{32}$$

III. THE REST SYSTEM

Let v, the magnitude of the particle velocity be defined by $v \equiv \sqrt{E^2 - m^2}/|E|$. The rest system is defined by $v = 0$. To first order in v one has

$$\delta(E,\eta) = 1 + \epsilon\eta v, \quad \sqrt{E^2 - m^2} = mv \tag{33}$$

and therefore, for small v,

$$W_{E,\eta,J,M}^s(\mathbf{x}) = m^s \sum_{l=|J-s|}^{J+s} C_l(\eta)(mvr)^{-1/2}H_{l+1/2}(mvr)\begin{bmatrix} \epsilon & \Psi_{J,l,s}^M \\ & \Psi_{J,l,s}^M \end{bmatrix} + \epsilon\eta v m^s \sum_{l=|J-s|}^{J+s} C_l(\eta)(mvr)^{-1/2}H_{l+1/2}(mvr) \times \begin{bmatrix} \epsilon & \Psi_{J,l,s}^M \\ & -\Psi_{J,l,s}^M \end{bmatrix}. \tag{34}$$

In order to obtain a meaningful limit in Eq. (34) as $v \to 0$, one must use Bessel functions[4] $J_{l+1/2}(mvr)$ for $H_{l+1/2}$. Then, since

$$\lim_{x\to 0}(x)^{-1/2}J_{l+1/2}(x) = \begin{cases} 0, & l > 0 \\ \sqrt{2/\pi}, & l = 0, \end{cases} \tag{35}$$

one sees that $\lim W_{E,\eta,J,M}^s(\mathbf{x}) = 0$ as $v \to 0$ for all $J \neq s$, since for all such J the corresponding l is greater than zero. For $J = s$, the limit is nonvanishing since $l = 0$ occurs and one, then, has

$$\lim_{v\to 0} W_{E,\eta,J,M}^s(\mathbf{x}) = m^s \frac{C_0(\eta)}{\sqrt{2\pi}}\begin{bmatrix} \epsilon & |s,M\rangle \\ & |s,M\rangle \end{bmatrix}, \tag{36}$$

where $|s,M\rangle$ are the $(2s+1)$-dimensional eigenvectors of the operators s^2 and s_3. Thus, $\lim W_{E,\eta,J,M}^s$ as $v \to 0$ are the eigenvectors of Σ_3 and $m\beta$ belonging to the eigenvalues M and ϵm, respectively, where Σ_3 is the component of spin along the z direction and $m\beta$ is the rest system Hamiltonian.

IV. THE MASSLESS PARTICLE

Consider the behavior of $W_{E,\eta,J,M}^s(\mathbf{x})$ as $m \to 0$. One gets

$$\lim_{m\to 0} W_{E,\eta,J,M}^s(\mathbf{x}) = \sum_{l=|J-s|}^{J+s} C_l(\eta)\rho^{-1/2}H_{l+1/2}(\rho) \times \begin{bmatrix} \lim_{m\to 0} m^s\delta(E,\eta)\epsilon\Psi_{J,l,s}^M \\ \lim_{m\to 0} m^s\delta^{-1}(E,\eta)\Psi_{J,l,s}^M \end{bmatrix}, \tag{37}$$

where

$$\lim_{m\to 0} m^s\left(\frac{|E| + |\mathbf{P}|}{|E| - |\mathbf{P}|}\right)^{\pm\epsilon\eta/2} = (2|E|)^{\pm\epsilon\eta}\lim_{m\to 0} m^{(s\mp\epsilon\eta)}. \tag{38}$$

Since it is always true that $\epsilon\eta \leq s$, it follows that for $|\eta| \neq s$, $\lim m^{(s\mp\epsilon\eta)} = 0$ as $m \to 0$. Thus, for $J < s$, $\lim W_{E,\eta,J,M}^s(\mathbf{x}) = 0$ as $m \to 0$, since in this case $|\eta| \leq J < s$, while, for $J \geq s$,

$$\lim_{m\to 0} W_{E,\eta,J,M}^s(\mathbf{x}) = 0, \quad |\eta| \neq s,$$

$$= (2|E|)^s \sum_{l=|J-s|}^{J+s} C_l(\epsilon s)\rho^{-1/2}H_{l+1/2}(\rho)\begin{bmatrix}\epsilon & \Psi^M_{J,l,s}\\ & 0\end{bmatrix},$$
$$\eta = \epsilon s \qquad (39)$$
$$= (2|E|)^s \sum_{l=|J-s|}^{J+s} C_l(-\epsilon s)\rho^{-1/2}H_{l+1/2}(\rho)\begin{bmatrix}0 & \\ \Psi^M_{J,l,s} & \end{bmatrix},$$
$$\eta = -\epsilon s.$$

The second of Eqs. (39) is, except for normalization, in complete agreement with the massless solution found in Ref. 1.

V. SECOND QUANTIZATION

For the purposes of second quantization, it is convenient to rewrite the basic solutions, Eq. (26), as follows:

$$W^s_{E,\eta,J,M}(\mathbf{x}) = m^s\begin{bmatrix}\delta(E,\eta) & 0\\ 0 & \delta^{-1}(E,\eta)\end{bmatrix} \times \sum_{l=|J-s|}^{J+s} C_l(\eta)\rho^{-1/2}J_{l+1/2}(\rho)\begin{bmatrix}\epsilon & \Psi^M_{J,l,s}\\ & \Psi^M_{J,l,s}\end{bmatrix}, \qquad (40)$$

where the $(2s+1)$-dimensional unit matrix is suppressed in δ and δ^{-1}. In terms of the $(2s+1)$-dimensional spin matrices, Eq. (40) may be written

$$W^s_{E,\eta,J,M}(\mathbf{x}) = m^s\begin{bmatrix}\exp\left[\epsilon\tanh^{-1}\left(\frac{|\mathbf{P}|}{|E|}\right)\frac{\mathbf{s}\cdot\mathbf{P}}{|\mathbf{P}|}\right] & 0\\ 0 & \exp\left[-\epsilon\tanh^{-1}\frac{|\mathbf{P}|}{|E|}\frac{\mathbf{s}\cdot\mathbf{P}}{|\mathbf{P}|}\right]\end{bmatrix} \sum_{l=|J-s|}^{J+s} C_l(\eta)\rho^{-1/2}J_{l+1/2}(\rho)\begin{bmatrix}\epsilon & \Psi^M_{J,l,s}\\ & \Psi^M_{J,l,s}\end{bmatrix}. \qquad (41)$$

Then one defines a set of functions $\hat{W}^s_{E,\eta,J,M}(\mathbf{x})$ by

$$W^s_{E,\eta,J,M}(\mathbf{x}) = m^s \exp\left[\epsilon s\tanh^{-1}\left(\frac{|\mathbf{P}|}{|E|}\right)\frac{\boldsymbol{\alpha}\cdot\mathbf{P}}{|\mathbf{P}|}\right]\hat{W}^s_{E,\eta,J,M}(\mathbf{x}), \qquad (42)$$

so that

$$\hat{W}^s_{E,\eta,J,M}(\mathbf{x}) \equiv \sum_{l=|J-s|}^{J+s} C_l(\eta)\rho^{-1/2}J_{l+1/2}(\rho)\begin{bmatrix}\epsilon & \Psi^M_{J,l,s}\\ & \Psi^M_{J,l,s}\end{bmatrix}. \qquad (43)$$

The reason for this decomposition is that $\hat{W}$ satisfies

$$E_0\beta\hat{W}^s_{E,\eta,J,M}(\mathbf{x}) = E\hat{W}^s_{E,\eta,J,M}(\mathbf{x}), \qquad (44)$$

$$\lambda\hat{W}^s_{E,\eta,J,M}(\mathbf{x}) = \eta\hat{W}^s_{E,\eta,J,M}(\mathbf{x}) \qquad (45)$$

and, therefore, may be identified as the partial wave decomposition of the Foldy wavefunction.[5] Note that apart from an over-all constant $\hat{W}$ is identical to the first term in Eq. (34) and so is related to the rest system of the particle.

One may rewrite Eq. (44) as $\beta\hat{W}^s_{E,\eta,J,M} = \epsilon\hat{W}^s_{E,\eta,J,M}$. Using this form, Eq. (42) may be written

$$W^s_{E,\eta,J,M}(\mathbf{x}) = m^s S(\mathbf{P})\hat{W}^s_{E,\eta,J,M}(\mathbf{x}), \qquad (46)$$

where $S(\mathbf{P})$ is given by

$$S(\mathbf{P}) = \cosh(\gamma\lambda) - \gamma_5\beta\sinh(\gamma\lambda),$$
$$S^{-1}(\mathbf{P}) = \operatorname{sech}(2\gamma\lambda)S^{\dagger}(\mathbf{P}) \qquad (47)$$

with

$$\gamma \equiv \tanh^{-1}(|\mathbf{P}|/E_0)$$

and

$$\gamma_5 \equiv \begin{pmatrix}-I & 0\\ 0 & I\end{pmatrix}$$

(here $\dagger$ indicates the Hermitian conjugate).

The operator $S(\mathbf{P})$ and its properties for any spin have been discussed in detail by Williams, Draayer, and Weber.[3]

The most general state function that can be constructed from the set $W^s_{\epsilon E,\eta,J,M}(\mathbf{x})$ is given by

$$\psi(\mathbf{x},t) = \int_m^\infty dE \sum_{J,M,\eta,\epsilon} A_{J,M,\eta,\epsilon}(E)W^s_{\epsilon E,\eta,J,M}(\mathbf{x})e^{-i\epsilon Et} = m^s S(\mathbf{P})\Phi(\mathbf{x},t), \qquad (48)$$

where

$$\Phi(x,t) = \int_m^\infty dE \sum_{J,M,\eta,\epsilon} A_{J,M,\eta,\epsilon}(E)\hat{W}^s_{\epsilon E,\eta,J,M}(\mathbf{x})e^{-i\epsilon Et} \qquad (49)$$

is the Foldy time-dependent wavefunction satisfying

$$E_0\beta\Phi(\mathbf{x},t) = i\frac{\partial}{\partial t}\Phi(\mathbf{x},t). \qquad (50)$$

The $A_{J,M,\eta,\epsilon}(E)$ are expansion coefficients and will be replaced by operators in the second-quantized theory.

To quantize a field theory there are two possible approaches. One can construct a Lagrangian density, define from it generalized momenta, and impose the usual equal time commutation relations. To introduce the particle concept, one then expands the field operator on a complete set of functions, projects out the expansion coefficients (which are now operators), and derives their commutation relations. The second approach is to postulate the commutation relations of the expansion coefficients and to derive the relations among components of the field operators. The second method will be followed in this work.

Using Eq. (48) to make the connection between ψ and Φ, the Foldy wavefunction Φ will be quantized first. Thus, one rewrites Eq. (49) in operator form as follows:

$$\Phi(\mathbf{x},t) = \int_m^\infty dE\,\Gamma(E) \sum_{J,M,\eta} [a_{J,M,\eta}(E)\hat{W}^s_{E,\eta,J,M}(\mathbf{x})e^{-iEt} + b^\dagger_{J,M,\eta}(E)\hat{W}^s_{-E,\eta,J,M}(\mathbf{x})e^{iEt}], \qquad (51)$$

where

$$\Gamma(E) \equiv \sqrt{E/2}(E^2 - m^2)^{1/4}$$

and

$$a_{J,M,\eta}(E)/b^\dagger_{J,M,\eta}(E)$$

are the annihilation/creation operators for a particle/antiparticle with quantum numbers E, J, M, η. The operators satisfy the commutation/anticommutation relations for bosons/fermions with integral/half-integral spin

$$[a, a]_{\mp} = [b, b]_{\mp} = [a, b]_{\mp} = [a, b^{\dagger}]_{\mp} = 0 \tag{52}$$

and

$$[a_{J,M,\eta}(E), a^{\dagger}_{J',M',\eta'}(E')]_{\mp} = \delta_{JJ'}\delta_{MM'}\delta_{\eta\eta'}\delta(E-E'),$$
$$[b_{J,M,\eta}(E), b_{J',M',\eta'}(E')]_{\mp} = \delta_{JJ'}\delta_{MM'}\delta_{\eta\eta'}\delta(E-E'), \tag{53}$$

where $-/+$ means the commutator/anticommutator. From the above relations it follows that

$$[\Phi_\alpha(\mathbf{x}, t), \Phi_\beta(\mathbf{x}', t')]_{\mp} = 0,$$
$$[\Phi_\alpha(\mathbf{x}, t), \Phi^{\dagger}_\beta(\mathbf{x}', t')]_{\mp}$$
$$= \int_m^\infty dE\,\Gamma^2(E) \sum_{J,M,\eta} [(\hat{W}^s_{E,\eta,J,M}(\mathbf{x}))_\alpha (\hat{W}^s_{E,\eta,J,M}(\mathbf{x}))^*_\beta e^{-iE(t-t')}$$
$$\mp (\hat{W}^s_{-E,\eta,J,M}(\mathbf{x}))_\alpha (\hat{W}^s_{-E,\eta,J,M}(\mathbf{x}'))^*_\beta e^{iE(t-t')}], \tag{54}$$

where α, β range over the $2(2s+1)$ components of $\hat{W}^s_{E,\eta,J,M}$. At equal times one, therefore, must evaluate the integral

$$\int_m^\infty dE\,\Gamma^2(E) \sum_{J,M,\eta} \hat{W}^s_{E,\eta,J,M}(\mathbf{x})\hat{W}^{s\dagger}_{E,\eta,J,M}(\mathbf{x}')$$
$$= \int_m^\infty dE\Gamma^2(E) \sum_{J,M,\eta} \sum_{l,l'}, \frac{C_l(\eta)}{\sqrt{rr'}} \frac{C^*_{l'}(\eta)}{\sqrt{E^2-m^2}}$$
$$\times J_{l+1/2}(r\sqrt{E^2-m^2}) J_{l'+1/2}(r'\sqrt{E^2-m^2})$$
$$\times \Psi^M_{J,l,s}(\theta, \varphi)\Psi^{M\dagger}_{J,l,s}(\theta', \varphi') \otimes \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \tag{55}$$

where $\otimes$ denotes the direct product. The completeness relation for the $C_l(\eta)$, Eq. (24), enables one to perform the sum on η and then, on l'. The Bessel functions also form a complete set, the appropriate closure equation being

$$\int_0^\infty dx \times J_{l+1/2}(xr)J_{l+1/2}(xr') = 1/r\delta(r-r'). \tag{56}$$

Using this relation and the completeness of the spinor spherical harmonics one finds

$$\int_m^\infty dE\Gamma^2(E) \sum_{J,M,\eta} \hat{W}^s_{\pm E,\eta,J,M}(\mathbf{x})\hat{W}^{s\dagger}_{\pm E,\eta,J,M}(\mathbf{x}')$$
$$= \tfrac{1}{2}\delta(\mathbf{x}-\mathbf{x}')\begin{pmatrix} I & \pm I \\ \pm I & I \end{pmatrix}. \tag{57}$$

Then, for the two types of statistics the equal time relations among the components of the Foldy field operator are

$$[\Phi_\alpha(\mathbf{x}, t), \Phi^{\dagger}_\beta(\mathbf{x}', t)]_{+} = \delta_{\alpha\beta}\delta(\mathbf{x}-\mathbf{x}'), \tag{58}$$

$$[\Phi_\alpha(\mathbf{x}, t), \Phi^{\dagger}_\beta(\mathbf{x}', t)]_{-} = \beta_{\alpha\beta}\delta(\mathbf{x}-\mathbf{x}'). \tag{59}$$

From Eq. (58), $-i\Phi^{\dagger}(\mathbf{x}, t)$ is the operator canonically conjugate to $\Phi(\mathbf{x}, t)$ in the case of Fermi statistics, and for Bose statistics $-i\Phi^{\dagger}\beta$ is canonically conjugate to Φ. These results agree with those found by one of us (D.L.W.) using an expansion of Φ into plane waves.

Using Eq. (48), the relations for the wavefunction operator ψ are [α and β range from 1 to $2(2s+1)$]

$$[\psi_\alpha(\mathbf{x}, t), \psi_\beta(\mathbf{x}', t)]_{\mp} = 0 \tag{60}$$

and, for Fermi statistics

$$m^{-2s}[\psi_\alpha(\mathbf{x}, t), \psi^{\dagger}_\beta(\mathbf{x}', t)]_{+} = (S(\mathbf{P})S^{\dagger}(\mathbf{P}))_{\alpha\beta}\delta(\mathbf{x}-\mathbf{x}'), \tag{61}$$

which reduces, for spin 1/2, to

$$[\psi_\alpha(\mathbf{x}, t), \psi^{\dagger}_\beta(\mathbf{x}', t)]_{+} = E_0\delta_{\alpha\beta}\,\delta(\mathbf{x}-\mathbf{x}'), \tag{62}$$

the usual result for the second-quantized Dirac theory when one takes account of the normalization.

Similarly, for Bose statistics

$$m^{-2s}[\psi_\alpha(\mathbf{x}, t), \psi^{\dagger}_\beta(\mathbf{x}', t)]_{-}$$
$$= [H(\mathbf{P})/E_0\,S(\mathbf{P})S^{\dagger}(\mathbf{P})]_{\alpha\beta}\delta(\mathbf{x}-\mathbf{x}'). \tag{63}$$

These results are in agreement with the equal time commutators and anticommutators found by Nelson and Good[6] who quantized the plane wave expansion coefficients of $\psi(\mathbf{x}, t)$. It should be noted that for Bose statistics one should use $[(H(\mathbf{P})/E_0)\psi(\mathbf{x}, t)]^{\dagger}$ rather than $\psi^{\dagger}$ for the operator canonically conjugate to ψ or make some other adjustment (as discussed in Ref. 6) to obtain a casual commutator.

APPENDIX: PROPERTIES OF THE MATRIX $A_{J_1 s}$ AND THE COEFFICIENTS $C_l(\eta)$

Since $A_{J,s} = A_{s,J}$ one need only consider $A_{J,s}$ for $J \geq s$. Defining

$$a_r \equiv \tfrac{1}{2}a^{J,s}_{J+s-r}, \quad 1 \leq r \leq 2s, \tag{A1}$$

the eigenvalues λ of $A_{J,s}$ satisfy

$$\det(A_{J,s} - \lambda I)$$
$$= \lambda^{2s+1} - \lambda^{2s-1}\sum_{r=1}^{2s} a_r^2 + \lambda^{2s-3} \sum_{r_1 \leq r_2-2} a^2_{r_1}a^2_{r_2}$$
$$- \lambda^{2s-5} \sum_{\substack{r_1 \leq r_2-2 \\ r_1 \leq r_3-2}} a^2_{r_1}a^2_{r_2}a^2_{r_3} + \cdots = 0. \tag{A2}$$

The roots are found to be $\lambda = s, s-1, \ldots, -s$.

Now for $J < s$ one sees because of Eq. (A1) that the eigenvalues of $A_{J_1 s}$ are $\lambda = J, J-1, \ldots, -J$. This implies that for the massless particle, in which only the extreme values $\pm s$ of helicity are allowed, states with total angular momentum $J < s$ do not exist. This is confirmed by the fact that the massless limit of massive particle states with $J < s$ is zero.

In order to derive the general formula for the coefficients $C_l(\eta)$, consider a sequence of nonzero, real numbers $\{a_k\}$, and correspondingly, a class of Hermitian matrices A_k defined by

$$A_k \equiv \begin{bmatrix} 0 & -ia_1 & 0 & \cdots & & \\ ia_1 & 0 & -ia_2 & 0 & \cdots & \\ 0 & & & & & \\ \cdot & & & & & \\ \cdot & & & \cdots & ia_{k-1} & 0 & -ia_k \\ \cdot & & & & 0 & ia_k & 0 \end{bmatrix}. \tag{A3}$$

Let $A_k(\lambda)$ denote $\det(A_k - \lambda I)$. Then, the numbers $A_k(\lambda)$ satisfy

$$A_k(\lambda) + a_k^2 A_{k-2}(\lambda) = \lambda A_{k-1}(\lambda). \tag{A4}$$

If one defines $A_{-1}(\lambda) \equiv 1$ and $A_0(\lambda) \equiv \lambda$, then Eq. (A4) is satisfied for $k = 1, 2, \cdots$. Define now the numbers $X_k(\lambda)$ by

$$A_k(\lambda) \equiv (-i)^{k+1} a_1 a_2, \ldots, a_{k+1} X_k(\lambda), \tag{A5}$$

where $a_0 \equiv 1$. Then $X_k(\lambda)$ satisfy

$$-i a_{k+1} X_k(\lambda) + i a_k X_{k-2}(\lambda) = \lambda X_{k-1}(\lambda), \quad k = 1, 2, \cdots. \tag{A6}$$

If λ is such that $A_n(\lambda) = 0$, then the column vector $X(\lambda)$ defined by

$$X(\lambda) \equiv \begin{bmatrix} X_{-1}(\lambda) \\ X_0(\lambda) \\ \cdot \\ \cdot \\ \cdot \\ X_{n-1}(\lambda) \end{bmatrix} \tag{A7}$$

is an eigenvector of the matrix A_n belonging to the eigenvalue λ.

From the above discussion it follows that the coefficients $C_l(\eta)$, $J - s \le l \le J + s$ are given to within a multiplicative constant by

$$C_{J+s}(\eta) = 1, \quad C_{J+s-1}(\eta) = 2i\eta / a^{J,s}_{J+s-1},$$

$$C_{J+s-r}(\eta) = (2i)^r \Big/ \left(\prod_{k=1}^{r} a^{J,s}_{J+s-k} \right) \det \begin{vmatrix} \eta & -\frac{1}{2} i a^{J,s}_{J+s-1} & 0 & \cdots \\ \frac{1}{2} i a^{J,s}_{J+s-1} & \eta & -\frac{1}{2} i a^{J,s}_{J+s-2} & \cdots \\ \cdot & & & \\ \cdot & & & \\ \cdot & \cdots & \eta & -\frac{1}{2} i a^{J,s}_{J+s-r+1} \\ & & \frac{1}{2} i a^{J,s}_{J+s-r+1} & \eta \end{vmatrix}. \tag{A8}$$

The values of the coefficients $C_l(\eta)$ are listed below for some special cases. The normalization is that of Eqs. (24):

Spin $\frac{1}{2}$ (Dirac particle):

$$C_{J+1/2}(\eta) = -i\sqrt{2}\,\eta, \quad C_{J-1/2}(\eta) = 1/\sqrt{2}.$$

Spin 1 (Vector boson): For $J \ge s$, using the notation of Eq. (23),

$$C(1) = \frac{1}{\sqrt{2}} \begin{bmatrix} -i\sqrt{\dfrac{J}{2J+1}} \\ 1 \\ i\sqrt{\dfrac{J+1}{2J+1}} \end{bmatrix}, \quad C(0) = \begin{bmatrix} \sqrt{\dfrac{J+1}{2J+1}} \\ 0 \\ \sqrt{\dfrac{J}{2J+1}} \end{bmatrix},$$

$$C(-1) = C^*(1);$$

for $J < s$, i.e., $J = 0$, $C(0) = 1$.

Spin 3/2: For $J \ge s$,

$$C(\eta) = N_\eta \begin{bmatrix} 1 \\ 2i\eta\sqrt{\dfrac{J+1}{3J}} \\ \dfrac{3J - 4(J+1)\eta^2}{\sqrt{3(2J-1)(2J+3)}} \\ \frac{1}{2} i \left(\dfrac{J+1}{J(2J-1)(2J+3)} \right)^{1/2} \left(\dfrac{3J - 4(J+1)\eta^2}{\eta} \right) \end{bmatrix},$$

where $\eta = \frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}$, and N_η is the normalization constant. Explicitly,

$$C\left(\frac{3}{2}\right) = \frac{1}{4} \begin{bmatrix} \sqrt{\dfrac{2J-1}{J+1}} \\ i\sqrt{\dfrac{3(2J-1)}{J}} \\ -i\sqrt{\dfrac{3(2J+3)}{J+1}} \\ -i\sqrt{\dfrac{2J+3}{J}} \end{bmatrix}, \quad C\left(\frac{1}{2}\right) = \begin{bmatrix} \sqrt{\dfrac{3(2J+3)}{J+1}} \\ i\sqrt{\dfrac{2J+3}{J}} \\ \sqrt{\dfrac{2J-1}{J+1}} \\ i\sqrt{\dfrac{3(2J-1)}{J}} \end{bmatrix},$$

$$C(-\tfrac{3}{2}) = C^*(\tfrac{3}{2}), \; C(-\tfrac{1}{2}) = C(\tfrac{1}{2});$$

for $J < s$, i.e., $J = \frac{1}{2}$,

$$C(\eta) = \frac{1}{\sqrt{2}} \begin{bmatrix} -2i\eta \\ 1 \end{bmatrix}, \quad \eta = \pm \tfrac{1}{2}.$$

* Work supported in part by the U.S. Atomic Energy Commission.

1 J. P. Hajj and D. L. Weaver, J. Math. Phys. **10**, 1185 (1969).
2 D. L. Weaver, C. L. Hammer, and R. H. Good, Jr., Phys. Rev. **135**, B241 (1964).
3 S. A. Williams, J. P. Draayer, and T. A. Weber, Phys. Rev. **152**, 1207 (1966).
4 E. Jahnke and F. Emde, *Tables of Functions* (Dover, New York, 1945).
5 L. L. Foldy, Phys. Rev. **102**, 568 (1956).
6 T. J. Nelson and R. H. Good, Jr., Rev. Mod. Phys. **40**, 508 (1968).

Modified WKB Approximation Applied to the Solution of the Repulsive Singular Potential

Pao Lu and Sheldon S. Wald
Department of Physics, Arizona State University, Tempe, Arizona 85281
(Received 13 December 1971; Revised Manuscript Received 5 January 1972)

The conventional WKB method for phase shift calculations is known to fail for singular repulsive potentials at small incident energies. We show in this paper that the modified WKB method due to Miller and Good can be used instead and that it gives better agreement with the exact results of the scattering phase shifts at small incident energies and in the small coupling constant limit, as more terms in $\hbar^2$ are included.

1. INTRODUCTION

An excellent account of various approximation methods on singular potentials was recently published in a review article by Frank, Land, and Spector.[1] We see that the conventional WKB method for phase-shift calculations fails badly at low energies, while it gives very accurate results at high energies. In this paper we show that if use is made of the modified WKB method as given by Miller and Good[2] and later extended to include higher order terms in $\hbar^2$,[3,4] we can obtain a very interesting result, namely that the modified WKB method can be shown to be valid in regions where the conventional WKB method fails.

In Ref. 2 it is indicated that the conventional WKB method is a special case of the modified method. We will mention this in the beginning of the next section. We feel, therefore, sure that this modified method also applies wherever the conventional method is valid.

In what follows we choose to illustrate the problem by a repulsive r^{-4} potential and comparisons are made with the exact results[5] at the small energy and small coupling constant limit in order to verify the accuracy of the modified WKB approximation in the regions where the ordinary WKB approximation fails. We are not necessarily confining ourselves with such a potential, nor do we intend to produce better results than the existing ones as discussed in Ref. 1. The reason why we choose the r^{-4} potential is mainly due to the availability of the exact results to compare with. Here the modification on the WKB method is our main concern. If we want to obtain better results, we will have to go to higher order expansion terms than $\hbar^2$. This is not attempted here. Even to the order $\hbar^2$, we believe that the corresponding wavefunction which will be in closed form is more useful than some of the methods given in Ref. 1. However, we are not discussing any of these problems here. What we want to do here is to demonstrate the method and use the known exact r^{-4} potential scattering results as a guidance. We will see that the phase shifts of the modified WKB method are improved if the next high order terms in $\hbar^2$ are included.

In this investigation, we have focused our attention to the repulsive potential because it has one turning point. We choose our solved part with a single turning point also. This will appear again in the appropriate place. We are concentrating our efforts to the $L \neq 0$ cases. This is because the modified WKB method used is more accurate as L becomes large. Furthermore, the $L = 0$ case can be solved by other methods, like the variation method etc., where the solution will give a better approximation.

For reason of simplicity, we set $m = c = 1$. As we have been doing previously we will consider the eigenvalue of angular momentum $[l(l+1)]^{1/2}\hbar$, as the zeroth order in $\hbar$, for we think that it has a physical magnitude as energy or distance. Furthermore, we let $E \equiv \frac{1}{2}\hbar^2 K^2$ keeping in our mind that energy is of zeroth order in $\hbar$.

2. FORMULATION OF THE PROBLEM

The Schrödinger equation that we are going to solve is of the form

$$\frac{d^2\psi}{dr^2} + \left(\frac{p_1^2(r)}{\hbar^2}\right)\psi = 0, \tag{1}$$

where in our case, $p_1^2(r) = 2E - [l(l+1)/r^2]\hbar^2 - 2(g^2/r^4)$. Here we are only considering the repulsive potential, so that $g^2 > 0$. The Schrödinger equation that is used to approximate Eq. (1) is

$$\frac{d^2\phi}{dS^2} + \frac{p_2^2}{\hbar^2}\phi = 0 \tag{2}$$

with $p_2^2(S) = 2E - [l(l+1)/S^2]\hbar^2$. Here we remember that if $p_2^2(S) = 1$, we reduce to the conventional WKB method.[2] The reason why we are only able to consider the repulsive potential is due to the limitation of the method employed. In the case of $p_2^2 = 0$, there will be one turning point $S_1 = [l(l+1)]^{1/2}/K$ and in the corresponding case $p_1^2 = 0$, there will be one turning point r_1 also. This relation of one-to-one correspondence is important. As derived before[3] we have

$$\int_{r_1}^{r} p_1 dr + \frac{\hbar^2}{8}\oint\left(\frac{3p_1'^2}{p_2^3} - \frac{2p_1''}{p_2^2}\right)dr = \int_{S_1}^{S} p_2 dS + \frac{\hbar^2}{8}\oint\left(\frac{3p_2'^2}{p_2^3} - \frac{2p_2''}{p_2^2}\right)dS, \tag{3}$$

where as discussed previously, the contour integrals are each evaluated around the turning point S_1 or r_1 to the left and with infinity to the right.

If we successively use the integration by parts,

$$\int_{\infty - i\epsilon}^{\infty + i\epsilon} u\,dv = uv\Big|_{\infty - i\epsilon}^{\infty + i\epsilon} - \int_{\infty - i\epsilon}^{\infty + i\epsilon} v\,du, \tag{4}$$

we can get rid of the appearances of divergence in Eq. (3) which then yields the following formula:

$$\int_{r_1}^{r} p_1 dr + \frac{\hbar^2}{24}\left(\int_{r_1}^{r}\frac{t_1''^2}{t_1^{1/2}t_1'^2}dr - \int_{r_1}^{r}\frac{t_1'''}{t_1^{1/2}t_1'}dr\right)$$
$$\int_{S_1}^{S} p_2 dS + \frac{\hbar^2}{24}\left(\int_{S_1}^{S}\frac{t_2''^2}{t_2^{1/2}t_2'^2}dS - \int_{S_1}^{S}\frac{t_2'''}{t_2^{1/2}t_2'}dS\right) \tag{5}$$

with the following definitions for $t_1 = p_1^2$, $t_2 = p_2^2$ and the upper limit r and S which will eventually be set to be infinity later. Remember in Eq. (5) written above, we do not have any divergence in the second

order terms any more. Then we have, the phase shift formula,

$$\delta_r = \delta_S + \lim_{r\to\infty} K(S - r), \tag{6}$$

where $\delta_S = 0$ because the right-hand side of Eq. (3) is chosen to be that of a three-dimensional free particle.

3. CALCULATIONS TO THE ZEROTH ORDER OF $\hbar^2$

The formula that will give us the phase shifts needed to $\hbar^0$ is, simply the zeroth order of $\hbar^2$ in (5) which yields

$$\int_{r_1}^{r} p_1 dr = \int_{S_1}^{S} p_2 dS. \tag{7}$$

There is no difficulty in evaluating the right-hand side of Eq. (7) so that

$$\int_{S_1}^{S} p_2(S)dS = \int_{S_1}^{S} \frac{dS}{S}[\hbar^2K^2S^2 - l(l+1)\hbar^2]^{1/2} = \hbar K\{S - [l(l+1)]^{1/2}\pi/(2K)\}, \tag{8}$$

while the left-hand side is

$$\int_{r_1}^{r} p_1(r)dr = \int_{r_1}^{r} \frac{dr}{r^2}[\hbar^2K^2r^4 - l(l+1)\hbar^2r^2 - 2g^2]^{1/2} = \hbar K[r - 2(r_1^2 + r_2^2)^{1/2}E(\tfrac{1}{2}\pi, k) + (r_1^2 + r_2^2)^{1/2}F(\tfrac{1}{2}\pi, k)] \tag{9}$$

with

$$r_1^2 = (1/2K^2)(l(l+1) + \{[l(l+1)]^2 + 8g^2K^2/\hbar^2\}^{1/2}),$$
$$r_2^2 = (-1/2K^2)(l(l+1) - \{[l(l+1)]^2 + 8g^2K^2/\hbar^2\}^{1/2}),$$
$$k^2 = r_2^2/(r_1^2 + r_2^2),$$

and $F(\frac{1}{2}\pi, k), E(\frac{1}{2}\pi, k)$ or F, E being the complete elliptic integrals of the first and the second kind, respectively.

4. CALCULATIONS TO THE FIRST ORDER OF $\hbar^2$

It becomes a little more complicated to describe the higher order terms. Taking the first order of $\hbar^2$ in (5) for the S-side yields

$$\lim_{S\to\infty}\left\{\frac{\hbar^2}{24}\left[\int_{S_1}^{S}\left(\frac{t_2''^2}{(t_2^{1/2}t_2'^2)}\right)dS - \int_{S_1}^{S}\left(\frac{t_2'''}{(t_2^{1/2}t_2')}\right)dS\right]\right\} = \hbar[l(l+1)]^{1/2}\tfrac{1}{2}\pi + \hbar\{1/[l(l+1)]^{1/2}\}\tfrac{1}{16}\pi, \tag{10}$$

While, the r-side can be expressed as

$$\lim_{r\to\infty}\left\{\frac{\hbar^2}{24}\left[\int_{r_1}^{r}\left(\frac{t_1''^2}{(t_1^{1/2}t_1'^2)}\right)dr - \int_{r_1}^{r}\left(\frac{t_1'''}{(t_1^{1/2}t_1')}\right)dr\right]\right\} = -\hbar F/[8K(r_1^2 + r_2^2)^{1/2}] - \hbar r_3^2(F - J_1)/[4K(r_3^2 - r_2^2)(r_1^2 + r_2^2)^{1/2}] + 8g^4(F + J_2 - 2J_1)/\{3\hbar^3K[l(l+1)]^2 \times (r_1^2 + r_2^2)^{1/2}(r_3^2 - r_2^2)^2\}, \tag{11}$$

with the following definitions:

$$r_3^2 = 8g^2/[2l(l+1)\hbar^2], \tag{12a}$$

$$J_1 = \int_0^F \frac{du}{1 - \alpha^2\mathrm{sn}^2u}, \tag{12b}$$

$$J_2 = \int_0^F \frac{du}{(1 - \alpha^2\mathrm{sn}^2u)^2}, \tag{12c}$$

and

$$\alpha^2 = (r_2^2 - r_3^2)/(r_1^2 + r_2^2). \tag{12d}$$

Here we use the rotation and formulas from the handbook written by Byrd and Friedman.[6] The only deviation from this handbook is that here we have previously specified the symbol K as given by $K = (2mE)^{1/2}\hbar$, where energy is always in terms of K. So we prefer to use $F(\frac{1}{2}\pi, k^2)$ and $E(\frac{1}{2}\pi, k^2)$ or just F and E to denote the complete elliptic integrals of the first and the second kind, respectively. In so doing, we have

$$J_1 = \frac{k^2F}{k^2 - \alpha^2} - \frac{\pi\alpha^2\Lambda_0(\psi, k)}{2[\alpha^2(1-\alpha^2)(\alpha^2 - k^2)]^{1/2}}, \tag{13}$$

$$J_2 = \frac{1}{2(\alpha^2 - 1)(k^2 - \alpha^2)}\left(\alpha^2E + \frac{2k^4\alpha^2 - 2k^2 + \alpha^4k'^2}{k^2 - \alpha^2}F - \frac{\pi(2\alpha^2k^2 + 2\alpha^2 - \alpha^4 - 3k^2)\alpha^2\Lambda_0(\psi, k)}{2[\alpha^2(1-\alpha^2)(\alpha^2 - k^2)]^{1/2}}\right), \tag{14}$$

and

$$\psi = \sin^{-1}\left(\frac{\alpha^2}{\alpha^2 - k^2}\right)^{1/2}, \qquad k'^2 = 1 - k^2,$$

where $\Lambda_0(\psi, k)$ is called Heuman's lambda function and it is tabulated in Byrd and Friedman. As in our case $0 < -\alpha^2 < \infty$, we have

$$\Lambda_0(\psi, k') = \frac{2}{\pi}[E(\tfrac{1}{2}\pi, k)F(\psi, k') + F(\tfrac{1}{2}\pi, k)E(\psi, k') - F(\tfrac{1}{2}\pi, k)F(\psi, k')], \tag{15}$$

where $F(\psi, k')$ and $E(\psi, k')$ are elliptic integrals with argument ψ. And they are no longer complete integrals.

Equation (6) together with Eqs. (8) and (9) give us the phase shifts to the zeroth order of $\hbar^2$. With the addition of Eq. (10) to Eq. (8) and Eq. (11) to Eq. (9) we obtain the phase shifts to the first order of $\hbar^2$. Here we see that these phase shifts are true for all energies.

5. PHASE SHIFT AT THE SMALL q LIMIT

As a check on the formulas previously, we now consider the results that are given by exact solution, namely the Mathieus function. However, we see that the solutions to the Mathieu equation are only given in series form. Recently Coombs and Lin[4] gave the solution in the small energy and small coupling constant limit. By defining $q = Kq(2)^{1/2}$ we have, from their Eq. (3),

$$\sigma_S = \frac{q^2}{K}\left[(1-q)^2 + \pi^2q^2\left(\sum_{l=0}^{\infty}[(2l+3)(2l+1)]^{-1} \times P_l(\cos\theta)\right)^2 + \cdots\right], \tag{16}$$

where we have changed from the attractive case to the repulsive case by defining a correct q as was done above.

Comparing this with the following well-known equations

$$f(\theta) = \sum_{l=0}^{\infty}\frac{2l+1}{2iK}(e^{i2\delta l} - 1)P_l(\cos\theta), \tag{17a}$$

TABLE I. The exact phase shifts are $\delta_l = K^2g^2 2\pi S_1$, phase shifts to the zeroth order in $\hbar^2$ are $\delta_l^{(0)} = K^2g^2 2\pi S_2$, and phase shifts to the first order in $\hbar^2$ are $\delta_l^{(1)} = K^2g^2 2\pi S_3$.

l	$S_1 = \frac{1}{(2l+1)(2l-1)(2l+3)}$	$S_2 = \frac{1}{8[l(l+1)]^{3/2}}$	$S_3 = S_2\cdot\left(1+\frac{5}{8l(l+1)}\right)$
1	0.066 667	0.044 194	0.058 005
2	0.009 524	0.008 505	0.009 391
3	0.003 175	0.003 007	0.003 164
4	0.001 443	0.001 398	0.001 441
5	0.000 777	0.000 761	0.000 777

$$\sigma_S = |f(\theta)|^2, \tag{17b}$$

we obtain, for $l \geq 1$, the following expression for the phase shifts in the small q limit

$$\delta_l = 2\pi K^2 g^2/[(2l+1)(2l-1)(2l+3)], \tag{18}$$

where we keep only terms in q^2 by omitting the higher powers. By comparison with the exact result at its small energy limit, we can determine how good the WKB approximation is by simply expanding the WKB results to the small limit. So that, to the zeroth in $\hbar^2$, we have, for $l \geq 1$,

$$\delta_l^{(o)} = K(r - S) = \frac{K^2g^2}{[l(l+1)]^{3/2}}\frac{\pi}{4}, \tag{19}$$

which was obtained by keeping only terms to $q^2 = 2K^2g^2$, such that

$$F(\tfrac{1}{2}\pi, k^2) = \tfrac{1}{2}\pi(1 + \tfrac{1}{4}k^2 + \cdots), \tag{20}$$

$$E(\tfrac{1}{2}\pi, k^2) = \tfrac{1}{2}\pi(1 - \tfrac{1}{4}k^2 + \cdots), \tag{21}$$

and

$$k^2 = r_2^2/(r_1^2 + r_2^2) = 2g^2K^2/\hbar^2[l(l+1)]^2. \tag{22}$$

To the first order in $\hbar^2$, in addition to the same expansion formula as given in Eqs. (20), (21), and (22), we use the expansion formula, which is given in 906.05 of Ref. 6:

$$J_1 = \frac{\pi}{2}\sum_{m=0}^{\infty}\sum_{j=0}^{m}\frac{(2m)!(2j)!}{4^m 4^j (m!)^2 (j!)^2}k^{2j}(\alpha^2)^{m-j}.$$

Now from Eq. (22) and $\alpha^2 = -2g^2K^2/[l(l+1)]^2\hbar^2$ to the second order in q, we get

$$\delta_l^{(1)} = K(r - S) = \{K^2g^2\pi/4[l(l+1)]^{3/2}\}(1 + 5/8l(l+1)). \tag{23}$$

This comparison in the small q limit is shown in Table I. We see that it is agreeing in this limit. And the agreement improves as l becomes larger.

ACKNOWLEDGMENT

The authors want to express their sincere gratitude to the referee of this paper for his constructive comments.

1 W. M. Frank, D. J. Land, and R. M. Spector, Rev. Mod. Phys. **143**, 36 (1971).
2 S. C. Miller, Jr. and R. H. Good, Jr., Phys. Rev. **91**, 174 (1953).
3 P. Lu, Nuovo Cimento **58A**, 301 (1968).
4 P. Lu and E. M. Measure, Lett. Nuovo Cimento **2**, 1, 37 (1971).
5 J. A. Coombs and S. H. Lin, J. Chem. Phys. **54**, 2285 (1971).
6 R. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Physicists* (Springer-Verlag, Berlin, 1954).

Clebsch-Gordan Coefficients and Special Function Identities. I. The Harmonic Oscillator Group

Willard Miller, Jr.
Mathematics Department, University of Minnesota, Minneapolis, Minnesota 55455
(Received 29 January 1971)

It is shown that by constructing explicit realizations of the Clebsch–Gordan decomposition for tensor products of irreducible representations of a group G, one can derive a wide variety of special function identities with physical interest. In this paper, the representation theory of the harmonic oscillator group is used to give elegant derivations of identities involving Hermite, Laguerre, Bessel, and hypergeometric functions.

1. INTRODUCTION

In two recent papers Armstrong[1] and Cunningham[2] have employed Lie algebraic techniques to compute some integrals which are useful in the quantum mechanical treatment of the hydrogen atom. An advantage of such techniques is that they allow one to compute desired matrix elements for a quantum mechanical system directly from the symmetry properties of the system. There is no need to appeal to special function theory for an independent derivation. Moreover, the corresponding special function identities themselves can be more simply and elegantly derived on the basis of group theoretic considerations. The identities useful in quantum mechanics tend to be exactly those which are derivable from a study of the symmetry groups of quantum mechanical systems.

In this paper we extend the single example of Armstrong and Cunningham to a general method for the derivation of special function identities. The method is simple to describe. Let $\{\nu_j\}$ be a family of irreducible representations of the Lie algebra G and suppose the tensor product $\nu_k \otimes \nu_l$ can be decomposed into a direct sum of representations

$$\nu_k \otimes \nu_l \cong \sum_j \oplus\, n_j(k,l)\nu_j, \tag{1.1}$$

where the multiplicity $n_j(k,l)$ is either one or zero. Let $\{h_m^{(j)}\}$ be a suitably chosen basis (which we call *canonical*) for the representation space of ν_j. Then the vectors $\{h_m^{(k)} \otimes h_n^{(l)}\}$ form a basis for the representation space V of $\nu_k \otimes \nu_l$. On the other hand, from expression (1.1) we see that for each j such that $n_j(k,l) \neq 0$, we can find vectors $\{H_p^{(j)}\}$ which form a canonical basis for that subspace V_j of V which transforms irreducibly under ν_j. As is well known, the vectors $\{H_p^{(j)}\}$ can be expressed as linear combinations of the $\{h_m^{(k)} \otimes h_n^{(l)}\}$ via the Clebsch–Gordan (CG) coefficients

TABLE I. The exact phase shifts are $\delta_l = K^2g^2 2\pi S_1$, phase shifts to the zeroth order in $\hbar^2$ are $\delta_l^{(0)} = K^2g^2 2\pi S_2$, and phase shifts to the first order in $\hbar^2$ are $\delta_l^{(1)} = K^2g^2 2\pi S_3$.

l	$S_1 = \frac{1}{(2l+1)(2l-1)(2l+3)}$	$S_2 = \frac{1}{8[l(l+1)]^{3/2}}$	$S_3 = S_2\cdot\left(1+\frac{5}{8l(l+1)}\right)$
1	0.066 667	0.044 194	0.058 005
2	0.009 524	0.008 505	0.009 391
3	0.003 175	0.003 007	0.003 164
4	0.001 443	0.001 398	0.001 441
5	0.000 777	0.000 761	0.000 777

$$\sigma_S = |f(\theta)|^2, \tag{17b}$$

we obtain, for $l \geq 1$, the following expression for the phase shifts in the small q limit

$$\delta_l = 2\pi K^2 g^2/[(2l+1)(2l-1)(2l+3)], \tag{18}$$

where we keep only terms in q^2 by omitting the higher powers. By comparison with the exact result at its small energy limit, we can determine how good the WKB approximation is by simply expanding the WKB results to the small limit. So that, to the zeroth in $\hbar^2$, we have, for $l \geq 1$,

$$\delta_l^{(o)} = K(r - S) = \frac{K^2g^2}{[l(l+1)]^{3/2}}\frac{\pi}{4}, \tag{19}$$

which was obtained by keeping only terms to $q^2 = 2K^2g^2$, such that

$$F(\tfrac{1}{2}\pi, k^2) = \tfrac{1}{2}\pi(1 + \tfrac{1}{4}k^2 + \cdots), \tag{20}$$

$$E(\tfrac{1}{2}\pi, k^2) = \tfrac{1}{2}\pi(1 - \tfrac{1}{4}k^2 + \cdots), \tag{21}$$

and

$$k^2 = r_2^2/(r_1^2 + r_2^2) = 2g^2K^2/\hbar^2[l(l+1)]^2. \tag{22}$$

To the first order in $\hbar^2$, in addition to the same expansion formula as given in Eqs. (20), (21), and (22), we use the expansion formula, which is given in 906.05 of Ref. 6:

$$J_1 = \frac{\pi}{2}\sum_{m=0}^{\infty}\sum_{j=0}^{m}\frac{(2m)!(2j)!}{4^m 4^j (m!)^2 (j!)^2}k^{2j}(\alpha^2)^{m-j}.$$

Now from Eq. (22) and $\alpha^2 = -2g^2K^2/[l(l+1)]^2\hbar^2$ to the second order in q, we get

$$\delta_l^{(1)} = K(r - S) = \{K^2g^2\pi/4[l(l+1)]^{3/2}\}(1 + 5/8l(l+1)). \tag{23}$$

This comparison in the small q limit is shown in Table I. We see that it is agreeing in this limit. And the agreement improves as l becomes larger.

ACKNOWLEDGMENT

The authors want to express their sincere gratitude to the referee of this paper for his constructive comments.

1 W. M. Frank, D. J. Land, and R. M. Spector, Rev. Mod. Phys. **143**, 36 (1971).
2 S. C. Miller, Jr. and R. H. Good, Jr., Phys. Rev. **91**, 174 (1953).
3 P. Lu, Nuovo Cimento **58A**, 301 (1968).
4 P. Lu and E. M. Measure, Lett. Nuovo Cimento **2**, 1, 37 (1971).
5 J. A. Coombs and S. H. Lin, J. Chem. Phys. **54**, 2285 (1971).
6 R. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Physicists* (Springer-Verlag, Berlin, 1954).

Clebsch-Gordan Coefficients and Special Function Identities. I. The Harmonic Oscillator Group

Willard Miller, Jr.
Mathematics Department, University of Minnesota, Minneapolis, Minnesota 55455
(Received 29 January 1971)

It is shown that by constructing explicit realizations of the Clebsch–Gordan decomposition for tensor products of irreducible representations of a group G, one can derive a wide variety of special function identities with physical interest. In this paper, the representation theory of the harmonic oscillator group is used to give elegant derivations of identities involving Hermite, Laguerre, Bessel, and hypergeometric functions.

1. INTRODUCTION

In two recent papers Armstrong[1] and Cunningham[2] have employed Lie algebraic techniques to compute some integrals which are useful in the quantum mechanical treatment of the hydrogen atom. An advantage of such techniques is that they allow one to compute desired matrix elements for a quantum mechanical system directly from the symmetry properties of the system. There is no need to appeal to special function theory for an independent derivation. Moreover, the corresponding special function identities themselves can be more simply and elegantly derived on the basis of group theoretic considerations. The identities useful in quantum mechanics tend to be exactly those which are derivable from a study of the symmetry groups of quantum mechanical systems.

In this paper we extend the single example of Armstrong and Cunningham to a general method for the derivation of special function identities. The method is simple to describe. Let $\{\nu_j\}$ be a family of irreducible representations of the Lie algebra G and suppose the tensor product $\nu_k \otimes \nu_l$ can be decomposed into a direct sum of representations

$$\nu_k \otimes \nu_l \cong \sum_j \oplus\, n_j(k,l)\nu_j, \tag{1.1}$$

where the multiplicity $n_j(k,l)$ is either one or zero. Let $\{h_m^{(j)}\}$ be a suitably chosen basis (which we call *canonical*) for the representation space of ν_j. Then the vectors $\{h_m^{(k)} \otimes h_n^{(l)}\}$ form a basis for the representation space V of $\nu_k \otimes \nu_l$. On the other hand, from expression (1.1) we see that for each j such that $n_j(k,l) \neq 0$, we can find vectors $\{H_p^{(j)}\}$ which form a canonical basis for that subspace V_j of V which transforms irreducibly under ν_j. As is well known, the vectors $\{H_p^{(j)}\}$ can be expressed as linear combinations of the $\{h_m^{(k)} \otimes h_n^{(l)}\}$ via the Clebsch–Gordan (CG) coefficients

$$H_p^{(j)} = \sum_{m,n} C(k,m;l,n|j,p) h_m^{(k)} \otimes h_n^{(l)}. \tag{1.2}$$

Also, relations (1.2) can be inverted to express the $\{h_m^{(k)} \otimes h_n^{(l)}\}$ as linear combinations of the $\{H_p^{(j)}\}$. We suppose that the coefficients $C(\cdot)$ are known.

Consider a realization (model) of $\nu_k \otimes \nu_l$ such that V is a function space. Then the $h_m^{(k)} \otimes h_n^{(l)}$ are functions and (1.2) shows us how to construct the functions $H_p^{(j)}$. If, however, we can compute the functions $H_p^{(j)}$ directly from our model, we can view (1.2) as an identity relating two families of functions.

Armstrong and Cunningham considered an example where G was $sl(2,R)$, $\{\nu_j\}$ was the discrete series of representations, and V was a functional Hilbert space such that the basis vectors $h_m^{(k)} \otimes h_n^{(l)}, H_p^{(j)}$ were computable in terms of Laguerre polynomials. Substituting these results into (1.2) and using the known CG coefficients for the discrete series, they obtained an identity obeyed by Laguerre polynomials. (Actually these authors computed the matrix elements

$$\langle H_p^{(j)}, h_m^{(k)} \otimes h_n^{(l)} \rangle, \tag{1.3}$$

where $\langle \cdot, \cdot \rangle$ is the inner product on V but that is equivalent to a knowledge of (1.2). Rather than study (1.3) via the Wigner–Eckart theorem, we choose to examine the sums (1.2). This is because (1.2) makes sense in many cases where there is no convenient inner product space structure on V.)

The key to obtaining a variety of useful identities is in the construction of models of $\nu_k \otimes \nu_l$. Once a model is constructed the identity follows automatically. The author's works[3–6] contain a classification of these models for many of the symmetry groups of physics, in which the representation acts via differential and difference operators. Thus, choosing appropriate models from these papers we can substitute into (1.2) and obtain a wide variety of special function identities.

In this paper we consider the Lie algebra of the harmonic oscillator group S, a group which arises in the study of the harmonic oscillator problem in quantum mechanics. The irreducible representations and CG coefficients for S are computed in Ref. 3. In particular, some of the CG coefficients are expressible as hypergeometric functions and some as Laguerre polynomials. By choosing appropriate models we obtain identities involving Hermite, Laguerre, Bessel, and hypergeometric functions.

The identity (5.14) may be new. All results are obtained with a minimum of computation. We do not attempt to list all possible models but only a few which lead to especially interesting formulas.

In a subsequent paper we shall apply this method to the Lie algebras $su(2)$ and $sl(2,R)$, the latter related to the hydrogen atom problem. The CG coefficients and special function identities for these algebras are considerably more complicated than those presented here.

Unless otherwise stated, all variables appearing in this paper are real.

2. THE HARMONIC OSCILLATOR GROUP

We designate by S the real four-parameter group of matrices

$$g\{w,\alpha,\delta\} = \begin{pmatrix} 1 & \frac{1}{2}e^{-i\alpha}\overline{w} & i\delta - \frac{1}{8}w\overline{w} & \alpha \\ 0 & e^{-i\alpha} & -\frac{1}{2}w & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{2.1}$$

where $w = x + iy \in \mathbb{C}$ and α, δ are real. The group multiplication law is

$$g\{w,\alpha,\delta\}\cdot g\{w',\alpha',\delta'\} = g\{w + e^{-i\alpha}w',\ \alpha + \alpha',\ \delta + \delta' + \tfrac{1}{8}(\overline{w}w'e^{-i\alpha} - w\overline{w}'e^{i\alpha})\}. \tag{2.2}$$

In particular, $g\{0,0,0\}$ is the identity and the inverse of a group element is given by

$$g^{-1}\{w,\alpha,\delta\} = g\{-e^{i\alpha}w, -\alpha, \delta\}. \tag{2.3}$$

As a basis for the Lie algebra $\mathcal{S}$ of S we choose the matrices $\mathcal{J}_1, \mathcal{J}_2, \mathcal{J}_3, \mathcal{I}$ such that

$$\begin{aligned} g\{iy,0,0\} &= \exp y\,\mathcal{J}_1, \quad g\{x,0,0\} = \exp x\,\mathcal{J}_2, \\ g\{0,\alpha,0\} &= \exp\alpha\,\mathcal{J}_3, \quad g\{0,0,\delta\} = \exp\delta\,\mathcal{I}. \end{aligned} \tag{2.4}$$

It is easy to verify that these matrices satisfy the commutation relations

$$[\mathcal{J}_1,\mathcal{J}_2] = \tfrac{1}{2}\mathcal{I}, \quad [\mathcal{J}_3,\mathcal{J}_1] = \mathcal{J}_2, \quad [\mathcal{J}_3,\mathcal{J}_2] = -\mathcal{J}_1, \quad [\mathcal{J}_k,\mathcal{I}] = \mathcal{O}, \quad k = 1,2,3. \tag{2.5}$$

where $\mathcal{O}$ is the zero matrix. For many purposes a more convenient basis is provided by the matrices

$$\mathcal{J}^{\pm} = \mp\mathcal{J}_2 + i\mathcal{J}_1, \quad \mathcal{J}^3 = i\mathcal{J}_3, \quad \mathcal{E} = -i\mathcal{I}$$

in the complexification of $\mathcal{S}$. Here,

$$\begin{aligned} &[\mathcal{J}^3,\mathcal{J}^{\pm}] = \pm\,\mathcal{J}^{\pm}, \quad [\mathcal{J}^+,\mathcal{J}^-] = -\mathcal{E}, \\ &[\mathcal{E},\mathcal{J}^{\pm}] = [\mathcal{E},\mathcal{J}^3] = \mathcal{O}. \end{aligned} \tag{2.6}$$

The unitary irreducible representations of S were determined in Refs. 3 and 7. We list the results as given in Ref. 3. (In this reference, representations of the factor group S/D are computed where D is the cyclic group generated by $\exp 2\pi\mathcal{J}_3$. However, the modification of these results to compute representations of S is trivial.)

There are four classes of unitary irreducible representations. The first class consists of one-dimensional representations and is of no concern to us. The second class consists of representations (λ, l) where both λ and $l > 0$ are real numbers. Each (λ, l) can be defined on a Hilbert space $\mathcal{H}$ with ON basis $\{h_n : n = 0,1,2,\cdots\}$. Indeed, the defining relations are

$$\begin{aligned} &J^3 h_n = (n-\lambda)h_n, \quad Eh_n = lh_n \\ &J^+ h_n = [l(n+1)]^{1/2}h_{n+1}, \quad J^- h_n = (ln)^{1/2}h_{n-1}, \\ &\quad n = 0,1,2,\ldots, \end{aligned} \tag{2.7}$$

where $J^{\pm}, J^3, E$ are the linear operators on $\mathcal{H}$ corresponding to $\mathcal{J}^{\pm}, \mathcal{J}^3, E$ in the Lie algebra represen-

tation induced by (λ, l). The unitary operators $\mathbf{U}(g)$ which define the representation on $\mathcal{H}$ have matrix elements:

$$U_{n,m}(g) = \langle h_n, U(g)h_m\rangle = \exp[i\alpha(\lambda - m) + il\delta + i(n-m)\Theta]\exp\left(-\frac{lr^2}{2}\right)\left(\frac{n!}{m!}\right)^{1/2} \times (rl^{1/2})^{m-n}L_n^{(m-n)}(lr^2), \quad (2.8)$$

where $\langle\cdot,\cdot\rangle$ is the inner product on $\mathcal{H}$ and $L_n^{(m)}(x)$ is an associated Laguerre polynomial (See Ref. 8, Vol. II). We have introduced polar coordinates $re^{i\Theta} = \frac{1}{2}w$.

The third class consists of representations $(\lambda, -l)$ where again $l > 0$ and λ are real numbers. The representations are defined on the same Hilbert space $\mathcal{H}$, but the defining relations are now

$$J^3h_n = (-\lambda - n - 1)h_n, \quad Eh_n = -lh_n,$$
$$J^+h_n = (ln)^{1/2}h_{n-1}, \quad J^-h_n = [l(n+1)]^{1/2}h_{n+1}, \quad (2.9)$$
$$n = 0, 1, 2, \cdots.$$

The matrix elements are

$$V_{n,m}(g) = \langle h_n, \mathbf{V}(g)h_m\rangle = \exp[i\alpha(\lambda + m + 1) - il\delta + i(m-n)\Theta] \exp(-\tfrac{1}{2}lr^2) \times (n!/m!)^{1/2}(-l^{1/2}r)^{m-n}L_n^{(m-n)}(lr^2), \quad (2.10)$$

where $2re^{i\Theta} = w$.

The fourth class contains representations of the form $[\rho, s]$ where $\rho^2 > 0$ and s are real numbers with $0 \leq s < 1$. There is an equivalence $[\rho, s] \cong [-\rho, s]$, but all other pairs of representations are inequivalent. Each $[\rho, s]$ can be defined on a Hilbert space $\mathcal{H}$ with ON basis $\{k_m: m = 0, \pm 1, \pm 2, \cdots\}$. The defining relations are

$$J^3k_m = (m+s)k_m, \quad Ek_m = 0$$
$$J^+k_m = \rho k_{m+1}, \quad J^-k_m = \rho k_{m-1}, \quad (2.11)$$
$$m = 0, \pm 1, \pm 2, \ldots,$$

and the matrix elements are given by

$$W_{n,m}(g) = \langle k_n, \mathbf{W}(g)k_m\rangle = (-i)^{n-m}e^{i[(m-n)\theta + (m+s)\alpha]}J_{n-m}(\rho r), \quad (2.12)$$

where $g = g\{2re^{i\Theta}, \alpha, \delta\}$ and $J_n(x)$ is a Bessel function (Ref. 8, Vol. II).

Of special interest to us will be the Clebsch–Gordan series for the decomposition of a tensor product of two irreducible representations of S into a direct sum of such representations. Again we quote the results from Ref. 3. First we have the decomposition

$$(\lambda, l) \otimes (\lambda', l') \cong \sum_{a=0}^{\infty} \oplus (\lambda + \lambda' - a, l + l'). \quad (2.13)$$

A natural basis for the Hilbert space $\mathcal{H} \otimes \mathcal{H}'$ corresponding to the left-hand side of this expression is given by $\{h_{n,p} = h_n \otimes h'_p: n, p = 0, 1, 2, \cdots\}$, while a canonical basis for the subspace transforming according to $(\lambda + \lambda' - a, l + l')$ is denoted $\{h_m^{(\lambda+\lambda'-a, l+l')}: m = 0, 1, \cdots\}$. The CG coefficients relating these bases are

$$K[l, n; l', p \mid a, m] = \langle h_{n,p}, h_m^{(\lambda+\lambda'-a, l+l')}\rangle', \quad (2.14)$$

where $\langle\cdot,\cdot\rangle'$ is the inner product on $\mathcal{H} \otimes \mathcal{H}'$.

Explicitly,

$$\exp\left(\frac{l^{1/2}(zu + xv) + l'^{1/2}(wu - zv)}{(l + l')^{1/2}}\right) = \sum_{a,m,n,p=0}^{\infty} K[l, n; l', p \mid a, m]\frac{z^a w^m v^n u^p}{(a!m!n!p!)^{1/2}}. \quad (2.15)$$

It follows that these coefficients are zero unless $a + m = n + p$. Furthermore,

$$K[l, n; l', a + m - n \mid a, m] = (-1)^{m-n} \times \left(\frac{a!(l'/l)^{n-m}n!}{m!(a+m-n)!(1+l'/l)^{a+m}}\right)^{1/2} \times \frac{F(-m, n-a-m; n-m+1; -l'/l)}{\Gamma(n-m+1)}, \quad (2.16)$$

where $F(\alpha, \beta; \gamma; z)$ is the hypergeometric function and $\Gamma(z)$ is the gamma function (Ref. 8, Vol. I). In Ref. 3, several identities are derived for these coefficients based on relations (2.14) and (2.15).

The CG coefficients for the decomposition

$$(\lambda, -l) \otimes (\lambda', -l') \cong \sum_{a=0}^{\infty} \oplus (\lambda + \lambda' + a + 1, -l - l') \quad (2.17)$$

are given by

$$\langle h_{n,p}, h_m^{(\lambda+\lambda'+a+1, -l-l')}\rangle' = K[l, n; l', p \mid a, m], \quad (2.18)$$

identical with (2.14).

If $l > l' > 0$, we have

$$(\lambda, l) \otimes (\lambda', -l') \cong \sum_{a=0}^{\infty} \oplus (\lambda + \lambda' + a + 1, l - l') \quad (2.19)$$

with CG coefficients

$$\langle h_{n,j}, h_m^{(\lambda+\lambda'+a+1, l-l')}\rangle' = G[l, n; l', j \mid a, m]. \quad (2.20)$$

Here,

$$G[l, n; l', j \mid a, m] = (1 - l'/l)^{1/2}K[l - l', n; l', a \mid j, m]. \quad (2.21)$$

The representation $[\rho, s] \otimes (\lambda, l)$ can be defined on the Hilbert space $\mathcal{H} \otimes \mathcal{H}$. The Clebsch–Gordan series is

$$[\rho, s] \otimes (\lambda, l) \cong \sum_{a=-\infty}^{\infty} \oplus (\lambda - s + a, l) \quad (2.22)$$

and the CG coefficients are

$$\langle k_n \otimes h_j, h_m^{(\lambda-s+a, l)}\rangle' = E(n, j; a, m; \rho^2/l), \quad (2.23)$$

where $\{h_m^{(\lambda-s+a, l)}; m = 0, 1, 2, \cdots\}$ is a canonical basis for $(\lambda - s + a, l)$ and $\langle\cdot,\cdot\rangle'$ is the inner product on $\mathcal{H} \otimes \mathcal{H}$. These coefficients are zero unless $m - a = n + j$, in which case

$$E(n, j; a, m; \rho^2/l) = E(n + a, j; o, m; \rho^2/l) = E(m - j, j; o, m; \rho^2/l) \quad (2.24)$$
$$= (j!/m!)^{1/2}\exp(-\rho^2/2l)(\rho/l^{1/2})^{m-j}L_j^{(m-j)}(\rho^2/l).$$

The CG coefficients for the decomposition

$$[\rho,s]\otimes(\lambda,-l)\cong\sum_{a=-\infty}^{\infty}\oplus(\lambda-s+a,-l) \qquad (2.25)$$

are essentially identical to (2.23) so we omit them. Finally, the representations $(\lambda,l)\otimes(\lambda',-l)$ and $[\rho,s]\otimes[\rho',s']$ have direct integral rather than direct sum decompositions and we will not consider them here.

It follows immediately from their definitions that the CG coefficients satisfy unitarity relations. For example, from (2.14) we have

$$\sum_{a,m=0}^{\infty}K[l,n_1;l',p_1|a,m]K[l,n_2;l',p_2|a,m]=\delta_{n_1n_2}\delta_{p_1p_2}, \qquad (2.26)$$

$$\sum_{n,p=0}^{\infty}K[l,n;l',p|a_1,m_1]K[l,n;l',p|a_2,m_2]=\delta_{a_1a_2}\delta_{m_1m_2}.$$

(Note that these coefficients are all real.) Similar relations hold for the other CG coefficients.

3. IDENTITIES FOR THE MATRIX ELEMENTS OF S

As our first application of the preceding results we consider models of the representation $(\lambda,l)\otimes(\lambda',l')$ in terms of functions on the group S. Let $\mathfrak{F}$ be the space of all functions $f(g)$, $g\in S$, defined on S. The operators $\mathbf{P}(g)$,

$$[\mathbf{P}(g)f](g')=f(g'g), \quad g,g'\in S, \qquad (3.1)$$

determine a representation of S on $\mathfrak{F}$, the *left regular representation*. Let $U^{(\lambda,l)}_{b,n}(g)$ be the matrix element (2.8) corresponding to (λ,l) and (for fixed b) define functions $h_n(g)=U^{(\lambda,l)}_{b,n}(g)$, $n=0,1,2,\ldots$, in $\mathfrak{F}$. Then

$$[\mathbf{P}(g)h_n](g')=U^{(\lambda,l)}_{b,n}(g'g)=\sum_{j=0}^{\infty}U^{(\lambda,l)}_{j,n}(g)h_j(g') \qquad (3.2)$$

so that the $\{h_n(g')\}$ form an ON basis for a Hilbert subspace of $\mathfrak{F}$ which transforms according to the irreducible representation (λ,l). The last equality in (3.2) follows from the group multiplication property

$$U^{(\lambda,l)}_{b,n}(g'g)=\sum_{j=0}^{\infty}U^{(\lambda,l)}_{b,j}(g')\,U^{(\lambda,l)}_{j,n}(g), \quad g,g'\in S, \qquad (3.3)$$

of the matrix elements.

It follows that (for fixed b,c) the functions

$$h_{n,p}(g,g')=U^{(\lambda,l)}_{b,n}(g)U^{(\lambda',l')}_{c,p}(g'), \quad n,p=0,1,2,\ldots, \qquad (3.4)$$

on the group $S\times S$ form a natural basis for the representation $(\lambda,l)\otimes(\lambda',l')$ under the left regular representation. Using (2.13) and (2.14), we see that the functions

$$h^{(\lambda+\lambda'-a,l+l')}_m(g,g')=\sum_{n,p=0}^{\infty}K[l,n;l',p|a,m]\times U^{(\lambda,l)}_{b,n}(g)U^{(\lambda',l')}_{c,p}(g'), \quad m=0,1,2,\cdots, \qquad (3.5)$$

form a canonical basis for a model of $(\lambda+\lambda'-a,\ l+l')$. (Note that $K[\cdot]$ is zero unless $n+p=a+m$.)

We shall obtain an identity for the matrix elements by computing the functions $h^{(\lambda+\lambda'-a,l+l')}_m(g,g')\equiv h^a_m(g,g')$ in an alternate manner. This computation makes use of the obvious properties:

$$h^a_m(hh',kh')=\sum_{j=0}^{\infty}U^{(\lambda+\lambda'-a,l+l')}_{j,m}(h')h^a_j(h,k) \quad h,h',k\in S, \qquad (3.6)$$

and

$$U^{(\lambda,l)}_n(e)=\delta_{n,m}, \qquad (3.7)$$

where e is the identity element of S. Setting $g'=e$ in (3.5), we find

$$h^a_m(g,e)=K[l,a+m-c;l',c|a,m]U^{(\lambda,l)}_{b,a+m-c}(g).$$

Substituting this result in (3.6) with $k=e$, $h'=g'$, $h=g(g')^{-1}$, we obtain

$$h^a_m(g,g')=\sum_{j=0}^{\infty}K[l,a+j-c;l',c|a,j]\times U^{(\lambda+\lambda'-a,l+l')}_{j,m}(g')U^{(\lambda,l)}_{b,a+j-c}(g(g')^{-1}). \qquad (3.8)$$

The desired identity follows from a comparison of (3.5) and (3.8). In particular, if $g=g'$, we find the familiar identity:

$$\sum_{n,p}K[l,n;l',p|a,m]U^{(\lambda,l)}_{b,n}(g)U^{(\lambda',l')}_{c,p}(g)=K[l,b;l',c|a,b+c-a]U^{(\lambda+\lambda'-a,l+l')}_{b+c-a,m}(g). \qquad (3.9)$$

This identity can be written in several equivalent forms by making use of the unitarity of the K coefficients. Substitution of relations (2.8) and (2.16) into (3.9) leads to a special function identity. Similar identities can be derived in the same manner corresponding to each of the coefficients $G[\cdot]$ and $E[\cdot]$. Some of these are listed in Ref. 3.

4. IDENTITIES FOR HERMITE POLYNOMIALS

We now search for additional models of the representations $(\lambda,\pm l)$ and $[\rho,s]$. Many such models have been classified in [3]–[6] in terms of Lie algebras of differential and difference operators. We select a few of particular interest.

As shown in Ref. 3, the operators

$$J^+_x=-\frac{d}{dx}+lx, \quad J^-_x=\frac{d}{dx}, \quad E=l,$$
$$J^3_x=-l^{-1}\frac{d^2}{dx^2}+\frac{xd}{dx}-\lambda, \qquad (4.1)$$

and basis functions

$$h_n(x)=2^{-n/2}(n!)^{-1/2}H_n(x\sqrt{l/2}), \quad n=0,1,2,\ldots, \qquad (4.2)$$

determine a model of (λ,l), where $H_n(x)$ are Hermite polynomials (Ref. 8, Vol. II). Another model is given by the operators

$$J^+=e^{i\Theta}\left(-\frac{\partial}{\partial x}+\frac{lx}{2}\right), \quad J^-=e^{i\Theta}\left(\frac{\partial}{\partial x}+\frac{lx}{2}\right) \quad E=l,$$
$$J^3=-i\frac{\partial}{\partial\Theta} \qquad (4.3)$$

and basis functions

$$h_n(x,\Theta)=2^{-n/2}(n!)^{-1/2}\exp(-lx^2/4)H_n(x\sqrt{l/2})\times e^{i(n-\lambda)\Theta}. \qquad (4.4)$$

Suppose the operators $J^{\pm}_x,J^3_x,E_x$ and $J'^{\pm}_y,J'^3_y,E'_y$ are given by (4.1) and define models of the representations $(\lambda,l),(\lambda',l')$, respectively. Then the operators

$$J^{\pm}=J^{\pm}_x+J'^{\pm}_y, \quad J^3=J^3_x+J'^3_y, \quad E=E_x+E'_y \qquad (4.5)$$

and basis functions

$$h_{n,p}(x,y) = h_n(x)h_p(y) = 2^{-(n+p)/2}(n!p!)^{-1/2}H_n(x\sqrt{l/2})H_p(x\sqrt{l'/2}) \tag{4.6}$$

define a model of $(\lambda, l) \otimes (\lambda', l')$. We will use (2.13) and (4.5) to compute the basis functions $h_m^{(\lambda+\lambda'-a,l+l')} \equiv h_m^a(x,y)$ directly. It is easy to verify that the equations

$$J^- h_0^a = 0, \quad J^3 h_0^a = -(\lambda + \lambda')h_0^a \tag{4.7}$$

for $h_0^a(x,y)$ have unique solutions,

$$h_0^a(x,y) = c_a 2^{-a/2}(a!)^{-1/2}H_a[(x-y)\sqrt{\tfrac{1}{2}ll'/(l+l')}], \quad a = 0,1,2,\ldots, \tag{4.8}$$

where the c_a are constants. The remaining basis functions can be obtained from the recurrence relation

$$J^+ h_m^a = [(m+1)ll'/(l+l')]^{1/2}h_{m+1}^a, \quad m = 0,1,2,\cdots. \tag{4.9}$$

The solution is

$$h_m^a(x,y) = c_a(2ll')^{-m/2}(l+l')^m(m!)^{-1/2}H_a(u)H_m(v)$$

$$u = \sqrt{\frac{\frac{1}{2}ll'}{l+l'}}(x-y), \quad v = \frac{lx + l'y}{\sqrt{2(l+l')}}, \tag{4.10}$$

$$a, m = 0,1,2,\cdots.$$

To compute c_a we use the fact that

$$h_0^a = (a!)^{1/2}(l+l')^{-a/2}\sum_{k=0}^{a}\frac{l^{(a-k)/2}(l')^{k/2}(-1)^k}{\sqrt{k!\,(a-k)!}}h_{k,a-k}, \tag{4.11}$$

which follows from the explicit expression (2.16) for $K[l,k;l',a-k|a,0]$. Comparing the coefficient of x^a on both sides of this equation, we obtain

$$c_a = (-1)^a. \tag{4.12}$$

On the other hand, (2.14) yields the relation

$$h_m^a = \sum_{n,p} K[l,n;l',p|a,m]h_{n,p}. \tag{4.13}$$

Substitution of (2.16), (4.6), (4.10), and (4.12) into this relation yields the desired identity.

Another model of $(\lambda_1, l_1) \otimes (\lambda_2, l_2)$ is provided by the operators (4.3) and basis functions

$$h_{n,p}(x,\Theta) = 2^{-(n+p)/2}(n!p!)^{-1/2}\exp(-lx^2/4) \times H_n(x\sqrt{l_1/2})H_p(x\sqrt{l_2/2})e^{i(n+p-\lambda_1-\lambda_2)\Theta} = h_n(x,\Theta)h_p(x,\Theta), \tag{4.14}$$

where $l = l_1 + l_2$. Indeed,

$$J^+ h_{n,p} = h_n(x,\Theta)e^{i\Theta}\left(\frac{-\partial}{\partial x} + \frac{l_2 x}{2}\right)h_p(x,\Theta) + h_p(x,\Theta)e^{i\Theta}\left(\frac{-\partial}{\partial x} + \frac{l_1 x}{2}\right)h_n(x,\Theta) = \sqrt{l_2(p+1)}h_{n,p+1} + \sqrt{l_1(n+1)}h_{n+1,p} \tag{4.15}$$

with similar formulas for the other operators (4.3). On the other hand, from (2.13) it is obvious that the basis functions h_m^a for this model are given by

$$h_m^a(x,\Theta) = c_a 2^{-m/2}(m!)^{-1/2}\exp(-lx^2/4) \times H_m(x\sqrt{l/2})e^{i(m-\lambda_1-\lambda_2+a)\Theta}, \tag{4.16}$$

where c_a is a constant. We can use the identity (3.11) with $l = l_1$, $l' = l_2$ to compute c_a. Indeed, comparing coefficients of x^0, we find

$$c_a = \begin{cases}(-2)^{-a/2}(a!)^{1/2}/(a/2)! & \text{if } a \text{ is even} \\ 0 & \text{if } a \text{ is odd.}\end{cases} \tag{4.17}$$

(Note that (4.18) is actually the special case $x = y$ of the first identity derived in this section. However, the method of proof is much simpler.)

For our next model we observe that the operators

$$K^+ = -e^{i\Theta}\frac{\partial}{\partial x}, \quad K^- = e^{-i\Theta}\frac{\partial}{\partial x}, \quad K^3 = -i\frac{\partial}{\partial\Theta}, \quad E = 0 \tag{4.19}$$

and basis functions

$$k_n(x,\Theta) = (-i)^n e^{i(n+s)\Theta + i\rho x}, \quad n = 0, \pm 1, \pm 2, \cdots, \tag{4.20}$$

define the representation $[\rho, s]$. Therefore, the operators (4.3) and basis functions

$$k_n \otimes h_j(x,\Theta) = (-i)^n 2^{-j/2}(j!)^{-1/2} \times \exp(+i\rho x - lx^2/4)H_j(x\sqrt{l/2}) \times \exp[i(n+j+s-\lambda)\Theta], \quad j, \pm n = 0,1,2,\ldots, \tag{4.21}$$

determine a model of $[\rho, s] \otimes (\lambda, l)$. From the explicit form of the operators (4.3), we can directly compute the natural basis functions $h_m^{(\lambda-s+a,l)}(x,\Theta) \equiv h_m^a(x,\Theta)$ corresponding to the Clebsch–Gordan series. The results are clearly

$$h_m^a(x,\Theta) = c_a 2^{-m/2}(m!)^{-1/2}\exp(-lx^2/4)H_m(x\sqrt{l/2}) \times e^{i(m-\lambda+s-a)\Theta}, \tag{4.22}$$

where c_a is a constant. We compute the constant by evaluating the expression

$$h_m^a(x,\Theta) = \sum_{j=0}^{\infty} E(-j-a,j;a,0;\rho^2/l)k_{-j-a} \otimes h_j(x,\Theta) \tag{4.23}$$

at $x = 0$. the result is $c_a = i^a$, so the identity

$$h_m^a = \sum_{j,n} E(n,j;a,m;\rho^2/l)\,k_n \otimes h_j \tag{4.24}$$

becomes (after some simplification)

$$\exp(\rho^2 - 2i\rho x)H_m(x) = \sum_{j=0}^{\infty}(-2i\rho)^{m-j}L_j^{(m-j)}(2\rho^2)H_j(x). \tag{4.25}$$

A different group-theoretic interpretation of (4.25) is presented in Ref. 3, p. 106.

The operators

$$K^+ = \rho e^{i\Theta}, \quad K^- = \rho e^{-i\Theta}, \quad K^3 = -i\frac{\partial}{\partial\Theta}, \quad E = 0 \tag{4.26}$$

and basis functions

$$k_n(\Theta) = e^{i(n+s)\Theta}, \quad n = 0, \pm 1, \pm 2, \cdots, \tag{4.27}$$

define another model of $[\rho, s]$. It follows from this remark and expressions (4.3) and (4.4) that the operators

$$J^+ = e^{i\Theta}\left(\frac{-\partial}{\partial x} + \frac{lx}{2} + \rho\right), \quad J^- = e^{-i\Theta}\left(\frac{\partial}{\partial x} + \frac{lx}{2} + \rho\right)$$
$$E = l, \quad J^3 = -i\frac{\partial}{\partial\Theta} \tag{4.28}$$

and basis functions

$$k_n \otimes h_j(x,\Theta) = k_n(\Theta)h_j(x,\Theta)$$
$$= e^{i(n+s)\Theta}2^{-j/2}(j!)^{-1/2}\exp(-lx^2/4)$$
$$\times H_j(x\sqrt{l/2})\,e^{i(j-\lambda)\Theta},$$
$$j, \pm n = 0, 1, 2, \cdots, \tag{4.29}$$

define a model of $[\rho, s] \otimes (\lambda, l)$. In particular

$$J^+[k_n h_j] = (\rho e^{i\Theta}k_n)h_j + k_n e^{i\Theta}\left(-\frac{\partial}{\partial x} + \frac{lx}{2}\right)h_j,$$
$$J^3[k_n h_j] = \left(-i\frac{\partial}{\partial\Theta}k_n\right)h_j + k_n\left(-i\frac{\partial}{\partial\Theta}h_j\right), \tag{4.30}$$

with similar interpretations of the remaining operators. We can again compute the basis functions $h_m^{(\lambda-s+a,l)}(x,\Theta) \equiv h_m^a(x,\Theta)$ directly from (4.28) and (2.7):

$$h_m^a(x,\Theta) = c_a 2^{-m/2}(m!)^{-1/2}\exp[-l(x + 2\rho/l)^2/4]$$
$$\times H_m[(x - 2\rho/l)\sqrt{l/2}\,]\,e^{i(m-\lambda+s-a)\Theta}. \tag{4.31}$$

As usual, we compute c_a by evaluating (4.23) at $x = 0$. The result is $c_a = 1$, so our new identity becomes

$$\exp[-2x\rho - \rho^2]H_m(x + 2\rho)$$
$$= \sum_{j=0}^{\infty}(2\rho)^{m-j}L_j^{(m-j)}(2\rho^2)H_j(x). \tag{4.32}$$

A different group-theoretic derivation of this formula is given in Ref. 3, p. 106.

We omit the routine computation of the identities for Hermite polynomials obtained by decomposing $(\lambda, l) \otimes (\lambda', -l')$.

As a concluding remark we note that the identity

$$2^{-n/2}(n!)^{-1/2}H_n[(2l)^{-1/2}(J^+ + J^-)]h_0 = h_n,$$
$$n = 0, 1, 2, \ldots, \tag{4.33}$$

holds for the model of (λ, l) defined by (4.1), (4.2), since $J^+ + J^- = lx$. Therefore, (4.33) must hold for *all* models of (λ, l) as classified in Refs. 3–6. This identity is by no means obvious for the remaining models considered in this paper.

5. IDENTITIES FOR LAGUERRE FUNCTIONS

As shown in Ref. 3, p. 111, the operators

$$J^+ = e^{i\Theta}\left(\frac{\partial}{\partial x} - l\right), \quad J^- = e^{-i\Theta}\left(-x\frac{\partial}{\partial x} + \frac{i\partial}{\partial\Theta}\right),$$
$$J^3 = -i\frac{\partial}{\partial\Theta}, \quad E = l \tag{5.1}$$

and basis functions

$$h_n(x,\Theta) = (n!)^{1/2}l^{n/2}(lx)^{\lambda-n}L_n^{(\lambda-n)}(lx)\,e^{i(n-\lambda)\Theta},$$
$$n = 0, 1, 2, \cdots, \tag{5.2}$$

form a model of (λ, l). It follows that the operators (5.1) and basis functions

$$h_{n,p}(x,\Theta) = h_n \otimes h'_p = (n!p!)^{1/2}l_1^{\lambda_1-n/2}l_2^{\lambda_2-p/2}$$
$$\times x^{\lambda_1+\lambda_2-n-p}L_n^{(\lambda_1-n)}(l_1x)L_p^{(\lambda_2-p)}(l_2x)\,e^{i(n+p-\lambda_1-\lambda_2)\Theta}. \tag{5.3}$$

define a model of $(\lambda_1, l_1) \otimes (\lambda_2, l_2)$, where $l_1 + l_2 = l$. Here,

$$J^+(h_n h'_p) = h_n e^{i}\left(\frac{\partial}{\partial x} - l_2\right)h'_p + h_p e^{i\Theta}\left(\frac{\partial}{\partial x} - l_1\right)h_n \tag{5.4}$$

with a similar interpretation for the other operators. The basis functions $h_m^{(\lambda_1+\lambda_2-a,l)}(x,\Theta) \equiv h_m^a(x,\Theta)$ for this representation are easily obtained from (5.1) and (5.2):

$$h_m^a(x,\Theta) = c_a(m!)^{1/2}l^{m/2}(lx)^{\lambda_1+\lambda_2-a-m}L_m^{(\lambda_1+\lambda_2-a-m)}(lx)$$
$$\times e^{i(m-\lambda_1-\lambda_2+a)\Theta}, \quad m, a = 0, 1, 2, \cdots. \tag{5.5}$$

The value of c_a follows by equating coefficients of $x^{\lambda_1+\lambda_2-a}$ on both sides of expression (4.11) ($l = l_1$, $l' = l_2$):

$$c_a = \frac{\Gamma(\lambda_2 + 1)}{\Gamma(\lambda_2 - a + 1)}\,l_1^{\lambda_1+a/2}l_2^{\lambda_2-a/2}(l_1 + l_2)^{-\lambda_1-\lambda_2+a/2}$$
$$\times {}_1F_1(-\lambda_1 - a; \lambda_2 - a + 1; -l_2/l_1). \tag{5.6}$$

Thus,

$$\sum_{n,p} K[l_1, n; l_2, p \mid a, m](n!p!)^{1/2}l_1^{\lambda_1-n/2}$$
$$\times l_2^{\lambda_2-p/2}L_n^{(\lambda_1-n)}(l_1x)L_p^{(\lambda_2-p)}(l_2x)$$
$$= c_a(m!)^{1/2}(l_1 + l_2)^{\lambda_1+\lambda_2-a-m/2}L_m^{(\lambda_1+\lambda_2-n-p)}[(l_1 + l_2)x]. \tag{5.7}$$

In the special case where λ_1 and λ_2 are integers, this identity reduces to (3.9).

We can construct another model related to Laguerre polynomials by observing that the operators

$$K^+ = e^{i\Theta}\frac{\partial}{\partial x}, \quad K^- = e^{-i\Theta}\left(-x\frac{\partial}{\partial x} + \frac{i\partial}{\partial\Theta}\right), \quad K^3 = -i\frac{\partial}{\partial\Theta},$$
$$E = 0 \tag{5.8}$$

and basis functions

$$k_n(x,\Theta) = x^{-(n+s)/2}J_{-n-s}(2\rho\sqrt{x}\,)\,e^{i(n+s)\Theta},$$
$$n = 0, \pm 1, \pm 2, \ldots, \tag{5.9}$$

form a model of $[\rho, s]$. Hence, the operators (5.1) and basis functions

$$k_n \otimes h_j(x,\Theta) = (j!)^{1/2}l^{j/2}(lx)^{\lambda-j}x^{-(n+s)/2}$$
$$\times J_{-n-s}(2\rho\sqrt{x}\,)L_j^{(\lambda-j)}(lx)\,e^{i(n+s+j-\lambda)\Theta},$$
$$j, \pm n = 0, \pm 1, \pm 2, \ldots, \tag{5.10}$$

determine a model of $[\rho, s] \otimes (\lambda, l)$. The basis functions $h_m^{(\lambda-s+a,l)} \equiv h_m^a$ can be computed directly from (5.1) and (5.2):

$$h_m^a(x,\Theta) = c_a(m!)^{1/2}l^{m/2}(lx)^{(\lambda-s+a-m)}L_m^{(\lambda-s+a-m)}(lx)$$
$$\times e^{i(m-\lambda+s-a)\Theta}, \quad a, m = 0, 1, 2, \cdots. \tag{5.11}$$

We compute the constants c_a by comparing coefficients of $x^{\lambda-s+a}$ on both sides of (4.23). The result is

$$c_a = \frac{\exp(-\rho^2/2l)}{\Gamma(a-s+1)}\left(\frac{\rho}{l}\right)^{a-s} {}_1F_1\left(\frac{-\lambda; a-s+1; \rho^2}{l}\right)$$
$$= \exp\left(\frac{-\rho^2}{2l}\right)\frac{\Gamma(\lambda+1)}{\Gamma(\lambda+a-s+1)}\left(\frac{\rho}{l}\right)^{a-s} L_\lambda^{(a-s)}\left(\frac{\rho^2}{l}\right), \tag{5.12}$$

so that the identity

$$k_n \otimes h_j = \sum_m E(n,j;a,m;\rho^2/l)\, h_m^{(\lambda-s+a,l)} \tag{5.13}$$

reads

$$\exp(\rho^2)(\rho^2 x)^{(n+s)/2} J_{-n-s}(2\rho\sqrt{x})\, L_j^{(\lambda-j)}(x)$$
$$= \sum_{m=0}^{\infty} \frac{\Gamma(\lambda+1)}{\Gamma(\lambda+m-n-j-s+1)}(\rho^2)^{m-j} L_j^{(m-j)}(\rho^2)$$
$$\times L_\lambda^{(m-n-j-s)}(\rho^2) L_m^{(\lambda-s-n-j)}(x). \tag{5.14}$$

If $\lambda = j = 0$, $n+s = -a$, $l = 1$, this formula simplifies to the well-known expression

$$\exp(\rho^2)(\rho^2 x)^{-\alpha/2} J_\alpha(2\rho\sqrt{x}) = \sum_{m=0}^{\infty} \frac{\rho^{2m} L_m^{(\alpha)}(x)}{\Gamma(m+\alpha+1)}. \tag{5.15}$$

The special case of (5.14) with $j = 0$ was first derived by Erdelyi in 1937 (see Ref. 9, p. 141). However, the general formula with $j \neq 0$ may be new.

It is a routine computation to obtain models of the representation $(\lambda, l) \otimes (\lambda', -l')$, but this will be omitted.

6. DIFFERENCE OPERATOR MODELS

In this section, we construct Lie algebra models using difference operators. These models were classified in Ref. 4.

The operators

$$K^+ = e^{+i\Theta}(-L+1),$$
$$K^- = e^{-i\Theta}\left(-(x+1)R + x + 1 + \frac{i\partial}{\partial\Theta}\right), \tag{6.1}$$
$$K^3 = -i\frac{\partial}{\partial\Theta}, \qquad E = 0$$

and basis functions

$$k_n(x,\Theta) = \rho^{-n} L_x^{(-n-s)}(\rho^2)\, e^{i(n+s)\Theta}, \qquad n = 0, \pm 1, \pm 2, \ldots, \tag{6.2}$$

define a model of $[\rho, s]$. Here,

$$Rf(x,\Theta) = f(x+1,\Theta), \qquad Lf(x,\Theta) = f(x-1,\Theta). \tag{6.3}$$

On the other hand, the operators

$$H^+ = -le^{+i\Theta}, \qquad H^- = ie^{-i\Theta}\left(\frac{\partial}{\partial\Theta} + i\lambda\right), \qquad H^3 = -i\frac{\partial}{\partial\Theta}, \qquad E = l \tag{6.4}$$

and functions

$$h_j(\Theta) = (-1)^j l^{j/2}(j!)^{-1/2} e^{i(j-\lambda)\Theta}, \qquad j = 0, 1, 2, \ldots, \tag{6.5}$$

define a model of (λ, l). Thus, the operators

$$J^+ = e^{+i\Theta}(-L-l+1),$$
$$J^- = e^{-i\Theta}\left(-(x+1)R + x - \lambda + 1 + \frac{i\partial}{\partial\Theta}\right), \tag{6.6}$$
$$J^3 = -i\frac{\partial}{\partial\Theta}, \qquad E = l$$

and functions

$$k_n \otimes h_j(x,\Theta) = (-\sqrt{l}\,)^j (j!)^{-1/2} \rho^{-n} L_x^{(-n-s)}(\rho^2) e^{i(n+j+s-\lambda)\Theta}, \tag{6.7}$$

determine a model of $[\rho, s] \otimes (\lambda, l)$. We can compute the basis functions $h_m^{(\lambda-s+a,l)}(x,\Theta) \equiv h_m^a(x,\Theta)$ directly from (6.6), (2.7), and (2.22), with the result

$$h_m^a(x,\Theta) = c_a(-\sqrt{l}\,)^m (m!)^{-1/2} \frac{\Gamma(x-s+a+1)}{\Gamma(x+1)}$$
$$\times {}_2F_1(-m, s-a; s-a-x; l^{-1}) e^{i(m-\lambda+s-a)\Theta}. \tag{6.8}$$

(Here, c_a could be a periodic function of x with period one. However, it is easy to check that c_a is actually a constant.) The constant c_a can be evaluated from expression (4.23). Indeed, comparing (4.23) and (5.14) with $j = 0$, $x = 0$, we find

$$c_a = \rho^a \exp(\rho^2 - \rho^2/2l)/\Gamma(a-s+1). \tag{6.9}$$

Thus, the identity (5.13) becomes

$$\exp[\rho^2(1-l)] L_x^{(\alpha)}(l\rho^2)$$
$$= j! \sum_{m=0}^{\infty} \binom{x+\alpha+m-j}{x} \frac{(-l\rho^2)^{m-j}}{m!}$$
$$\times {}_2F_1(-m, j-\alpha-m; j-a-m-x; l^{-1})$$
$$\times L_j^{(m-j)}(\rho^2). \tag{6.10}$$

For our last example we consider the operators

$$J^+ = e^{i\Theta}(-L-l-l'+1),$$
$$J^- = e^{-i\Theta}\left(-(x+1)R + x - \lambda + 1 + \frac{i\partial}{\partial\Theta}\right), \tag{6.11}$$
$$J^3 = -i\frac{\partial}{\partial\Theta}, \qquad E = l + l'.$$

The functions

$$h_{n,p} = h_n \otimes h'_p(x,\Theta) = \frac{(-\sqrt{l}\,)^n}{\sqrt{n!}} \frac{(-\sqrt{l'}\,)^p}{\sqrt{p!}} \frac{\Gamma(x+\lambda'+1)}{\Gamma(x+1)}$$
$$\times {}_2F_1(-p, -\lambda'; -\lambda'-x; l'^{-1})$$
$$\times \exp(n+p-\lambda-\lambda')i\Theta, \qquad n, p = 0, 1, 2, \ldots, \tag{6.12}$$

and these operators define a model of $(\lambda, l) \otimes (\lambda', l')$. [In particular, the action of the operators (6.4) on the $\{h_n(\Theta)\}$ yields (λ, l).] Computing the basis functions $h_m^a \equiv h_m^{(\lambda+\lambda'-a,l+l')}(x,\Theta)$ directly, we find in analogy with (6.8):

$$h_m^a(x,\Theta) = c_a \frac{(-\sqrt{l+l'}\,)^m}{\sqrt{m!}} \frac{\Gamma(x+\lambda'-a+1)}{\Gamma(x+1)}$$
$$\times {}_2F_1(-m, a-\lambda'; a-\lambda'-x; (l+l')^{-1})$$
$$\times e^{i(m+a-\lambda-\lambda')\Theta}. \tag{6.13}$$

Using (4.11) to evaluate the constant, we obtain

$$c_a = \binom{\lambda'}{a} \sqrt{a!}\,(l+l')^{-a/2}(l/l')^{a/2}. \tag{6.14}$$

The resulting identity is

$$\binom{\lambda'}{a}\frac{\Gamma(x+\lambda'-a+1)}{\Gamma(x+\lambda'+1)}(-1)^{a+m}(l-l')^m l'^{-a} \times {}_2F_1(-m;a-\lambda',a-\lambda'-x;(l+l')^{-1}) = \sum_{n=0}^{a+m} \times \frac{(-1)^n l^{m-n}\,{}_2F_1(-m,n-a-m;-n-m+1;-l'/l)}{(a+m-n)!\,\Gamma(n-m+1)} \times {}_2F_1(n-a-m,-\lambda';-\lambda'-x;l'^{-1}). \quad (6.15)$$

1 L. Armstrong, Jr., J. Math. Phys. **12**, 953 (1971).
2 M. Cunningham, J. Math. Phys. **13**, 33 (1972).
3 W. Miller, *Lie Theory and Special Functions* (Academic, New York, 1968).
4 W. Miller, J. Math. Anal. Appl. **28**, 383 (1969).
5 W. Miller, SIAM J. Math. Anal. **1**, 246 (1970).
6 W. Miller, SIAM J. Math. Anal. 2, 307 (1971).
7 R. Streater, Commun. Math. Phys. **4**, 217 (1967).
8 A. Erdelyi *et al.*, *Higher Transcendental Functions*, Bateman Manuscript Project (McGraw-Hill, New York, 1953), Vols. I, II.
9 H. Buchholz, *The Confluent Hypergeometric Function* (Springer, New York, 1969).

Note on the Explicit Form of Invariant Operators for $O(n)$

Takayoshi Maekawa
Department of Physics, Kumamoto University, Kumamoto, Japan

(Received 20 September 1971)

A complete set of invariants of $O(n)$ is constructed explicitly and a method of deriving the corresponding invariants of $O(p,q)$ is briefly remarked.

It is assumed that the invariant operators for a Lie group are clarified implicitly by the researches of Killing, Cartan, and Weyl. It, however, is important to know the explicit form for the invariants in applications to physics. The subject is discussed by some authors[1-3] and an interesting form for the invariants is given for some special groups. But it seems that the explicit form for the invariants is not so simple as the Casimir operator. In this note, we give a complete system of independent invariants suitable for uniquely labeling the irreducible inequivalent representations of $O(n)$. A further discussion will be given in the near future together with some simple applications.[4]

The infinitesimal generators D_{jk}, $j,\ k = 1, 2, \ldots, n$, of $O(n)$ are defined as the quantities which satisfy the commutation relations

$$[D_{jk}, D_{lm}] = i(\delta_{jl}D_{km} + \delta_{km}D_{jl} - \delta_{jm}D_{kl} - \delta_{kl}D_{jm}), \quad (1)$$

where D_{jk} is antisymmetric ($D_{jk} = -D_{kj}$) and Hermitian. As is well known, the orthogonal group $O(n)$ has $[n/2]$ invariant operators, where $[n/2]$ is equal to $n/2$ or $(n-1)/2$ corresponding to even n or odd n. One of these invariants is the well-known Casimir operator

$$F^{(n)} = \tfrac{1}{2}D_{jk}D_{jk}, \quad (2)$$

where the superscript n of F denotes the dimension number. Unless stated otherwise, similar notation and the summation convention from 1 to n will be used.

We can give the result for the other invariant operators $G_p^{(n)}$ as follows:

$$G_p^{(n)} = \sum_{i_1<i_2<\cdots<i_{n-2p-2}} (C^{(n)}_{i_1 i_2 \cdots i_{n-2p-2}})^2, \quad (3)$$

$$G_p^{(n)} = C^{(n)} \quad \text{for an even } n \text{ and } p = (n-2)/2, \quad (4)$$

where p in (3) takes $1, 2, \ldots, (n-4)/2$ for an even n and $1, 2, \ldots, (n-3)/2$ for an odd n. It is straightforward to show that the $G_p^{(n)}$ in (3) and (4) are invariant. The sum on the right-hand side of (3) is taken over all satisfying the condition $i_1 < i_2 < \cdots < i_{n-2p-2}$. The C in (3) and (4) are given by virtue of D_{jk} as follows:

$$C^{(n)}_{i_1 i_2 \cdots i_{n-2p-2}} = \frac{1}{2^{p+1}(p+1)!}\mathcal{E}_{i_1 i_2 \cdots i_n} D_{i_{n-2p-1} i_{n-2p}} \times D_{i_{n-2p+1} i_{n-2p+2}} \cdots D_{i_{n-1} i_n},$$

$$\mathcal{E}_{i_1 i_2 \cdots i_n} = \begin{cases} +1, & \text{for an even permutation } (i_1 i_2 \cdots i_n) \text{ of } (12\cdots n), \\ -1, & \text{for an odd permutation } (i_1 i_2 \cdots i_n) \text{ of } (12\cdots n) \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

Thus together with (2) and (3) [and (4) for an even n], we have given the $[n/2]$ invariant operators for $O(n)$, whose explicit expressions can be easily given.

It can be seen that these invariant operators are independent and suitable for labeling the irreducible representations of $O(n)$. In order to see the situation, let us give an outline of the proof according to Biedenharn[1] and Micu[2]: When an invariant is evaluated in terms of the highest weight L and only the highest-order terms [only the terms containing the generators $H_j \equiv D_{2j-1\,2j} (j = 1, 2, \ldots, [n/2])$] in the invariant are considered, it becomes an invariant of the group S (the group of reflections on hyperplanes perpendicular to the roots). That is, the invariants $F^{(n)}$ and $G_p^{(n)}$ become

$$F^{(n)} \to \overline{F}^{(n)} = \sum^{[n/2]} L_j L_j, \quad (6)$$

$$G_p^{(n)} \to \overline{G}_p^{(n)} = \sum_{i_1<i_2<\cdots<i_{p+1}} (L_{i_1} L_{i_2} \cdots L_{i_{p+1}})^2, \quad (7)$$

$$G_p^{(n)} \to \overline{G}_p^{(n)} = L_1 L_2 \cdots L_{n/2} \quad \text{for an even } n \text{ and } p = (n-2)/2. \quad (8)$$

These invariants of S have the properties: Their Jacobian does not vanish identically and factorizes into $N[= n(n-2)/8$ for an even n and $(n-1)(n+1)/8$ for an odd $n]$ linear forms which, when equated to zero, give the reflecting hyperplanes that generate the

$$\binom{\lambda'}{a}\frac{\Gamma(x+\lambda'-a+1)}{\Gamma(x+\lambda'+1)}(-1)^{a+m}(l-l')^m l'^{-a} \times {}_2F_1(-m;a-\lambda',a-\lambda'-x;(l+l')^{-1}) = \sum_{n=0}^{a+m} \times \frac{(-1)^n l^{m-n}\,{}_2F_1(-m,n-a-m;-n-m+1;-l'/l)}{(a+m-n)!\,\Gamma(n-m+1)} \times {}_2F_1(n-a-m,-\lambda';-\lambda'-x;l'^{-1}). \quad (6.15)$$

1 L. Armstrong, Jr., J. Math. Phys. **12**, 953 (1971).
2 M. Cunningham, J. Math. Phys. **13**, 33 (1972).
3 W. Miller, *Lie Theory and Special Functions* (Academic, New York, 1968).
4 W. Miller, J. Math. Anal. Appl. **28**, 383 (1969).
5 W. Miller, SIAM J. Math. Anal. **1**, 246 (1970).
6 W. Miller, SIAM J. Math. Anal. 2, 307 (1971).
7 R. Streater, Commun. Math. Phys. **4**, 217 (1967).
8 A. Erdelyi *et al.*, *Higher Transcendental Functions*, Bateman Manuscript Project (McGraw-Hill, New York, 1953), Vols. I, II.
9 H. Buchholz, *The Confluent Hypergeometric Function* (Springer, New York, 1969).

Note on the Explicit Form of Invariant Operators for $O(n)$

Takayoshi Maekawa
Department of Physics, Kumamoto University, Kumamoto, Japan

(Received 20 September 1971)

A complete set of invariants of $O(n)$ is constructed explicitly and a method of deriving the corresponding invariants of $O(p,q)$ is briefly remarked.

It is assumed that the invariant operators for a Lie group are clarified implicitly by the researches of Killing, Cartan, and Weyl. It, however, is important to know the explicit form for the invariants in applications to physics. The subject is discussed by some authors[1-3] and an interesting form for the invariants is given for some special groups. But it seems that the explicit form for the invariants is not so simple as the Casimir operator. In this note, we give a complete system of independent invariants suitable for uniquely labeling the irreducible inequivalent representations of $O(n)$. A further discussion will be given in the near future together with some simple applications.[4]

The infinitesimal generators D_{jk}, $j,\ k = 1, 2, \ldots, n$, of $O(n)$ are defined as the quantities which satisfy the commutation relations

$$[D_{jk}, D_{lm}] = i(\delta_{jl}D_{km} + \delta_{km}D_{jl} - \delta_{jm}D_{kl} - \delta_{kl}D_{jm}), \quad (1)$$

where D_{jk} is antisymmetric ($D_{jk} = -D_{kj}$) and Hermitian. As is well known, the orthogonal group $O(n)$ has $[n/2]$ invariant operators, where $[n/2]$ is equal to $n/2$ or $(n-1)/2$ corresponding to even n or odd n. One of these invariants is the well-known Casimir operator

$$F^{(n)} = \tfrac{1}{2}D_{jk}D_{jk}, \quad (2)$$

where the superscript n of F denotes the dimension number. Unless stated otherwise, similar notation and the summation convention from 1 to n will be used.

We can give the result for the other invariant operators $G_p^{(n)}$ as follows:

$$G_p^{(n)} = \sum_{i_1<i_2<\cdots<i_{n-2p-2}} (C^{(n)}_{i_1 i_2 \cdots i_{n-2p-2}})^2, \quad (3)$$

$$G_p^{(n)} = C^{(n)} \quad \text{for an even } n \text{ and } p = (n-2)/2, \quad (4)$$

where p in (3) takes $1, 2, \ldots, (n-4)/2$ for an even n and $1, 2, \ldots, (n-3)/2$ for an odd n. It is straightforward to show that the $G_p^{(n)}$ in (3) and (4) are invariant. The sum on the right-hand side of (3) is taken over all satisfying the condition $i_1 < i_2 < \cdots < i_{n-2p-2}$. The C in (3) and (4) are given by virtue of D_{jk} as follows:

$$C^{(n)}_{i_1 i_2 \cdots i_{n-2p-2}} = \frac{1}{2^{p+1}(p+1)!}\mathcal{E}_{i_1 i_2 \cdots i_n} D_{i_{n-2p-1} i_{n-2p}} \times D_{i_{n-2p+1} i_{n-2p+2}} \cdots D_{i_{n-1} i_n},$$

$$\mathcal{E}_{i_1 i_2 \cdots i_n} = \begin{cases} +1, & \text{for an even permutation } (i_1 i_2 \cdots i_n) \text{ of } (12\cdots n), \\ -1, & \text{for an odd permutation } (i_1 i_2 \cdots i_n) \text{ of } (12\cdots n) \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

Thus together with (2) and (3) [and (4) for an even n], we have given the $[n/2]$ invariant operators for $O(n)$, whose explicit expressions can be easily given.

It can be seen that these invariant operators are independent and suitable for labeling the irreducible representations of $O(n)$. In order to see the situation, let us give an outline of the proof according to Biedenharn[1] and Micu[2]: When an invariant is evaluated in terms of the highest weight L and only the highest-order terms [only the terms containing the generators $H_j \equiv D_{2j-1\,2j} (j = 1, 2, \ldots, [n/2])$] in the invariant are considered, it becomes an invariant of the group S (the group of reflections on hyperplanes perpendicular to the roots). That is, the invariants $F^{(n)}$ and $G_p^{(n)}$ become

$$F^{(n)} \to \overline{F}^{(n)} = \sum^{[n/2]} L_j L_j, \quad (6)$$

$$G_p^{(n)} \to \overline{G}_p^{(n)} = \sum_{i_1<i_2<\cdots<i_{p+1}} (L_{i_1} L_{i_2} \cdots L_{i_{p+1}})^2, \quad (7)$$

$$G_p^{(n)} \to \overline{G}_p^{(n)} = L_1 L_2 \cdots L_{n/2} \quad \text{for an even } n \text{ and } p = (n-2)/2. \quad (8)$$

These invariants of S have the properties: Their Jacobian does not vanish identically and factorizes into $N[= n(n-2)/8$ for an even n and $(n-1)(n+1)/8$ for an odd $n]$ linear forms which, when equated to zero, give the reflecting hyperplanes that generate the

group S. In fact, the Jacobians become apart from a numerical factor

$$\frac{\partial(\overline{F}^{(n)}, \overline{G}_1^{(n)}, \ldots, \overline{G}_{(n-2)/2}^{(n)})}{\partial(L_1, L_2, \ldots, L_{n/2})} \simeq \prod_{i<j} (L_i - L_j)(L_i + L_j) \quad (9)$$

for an even n and

$$\frac{\partial(\overline{F}^{(n)}, \overline{G}_1^{(n)}, \ldots, \overline{G}_{(n-3)/2}^{(n)})}{\partial(L_1, L_2, \ldots, L_{(n-1)/2})} \simeq L_1 L_2 \cdots \times L_{(n-1)/2} \prod_{i<j} (L_i - L_j)(L_i + L_j) \quad (10)$$

for an odd n. Thus it demonstrates that the invariants $\overline{F}^{(n)}$ and $\overline{G}_p^{(n)}$ are a suitable basis for all the invariants of S, and then $F^{(n)}$ and $G_p^{(n)}$ are a suitable basis for all the invariants of $O(n)$.

Finally, it is noted that the invariant operators for the noncompact group $O(p, q)(p + q = n)$ are obtained explicitly from (2) and (3) [and (4) for an even n] by means of the substitution $D_{jk} \to iD_{jk}$ for $j \leq p, k \geq q$, or $j \geq q, k \leq p$ and retaining other D_{jk} in its original form.

ACKNOWLEDGMENT

The author would like to express his cordial thanks to Professor M. Ikeda for a helpful discussion.

[1] L. C. Biedenharn, J. Math. Phys. 4, 436 (1963). Detailed references are quoted in this paper.
[2] G. Racah, Lecture notes in "Group Theory and Spectroscopy," Princeton (1951); A. Klein, J. Math. Phys. 4, 1238 (1963); M. Umezawa, Nucl. Phys. 48, 111 (1963); 53, 54 (1964); 57, 65 (1964); M. Micu, *ibid.* 60, 353 (1964).
[3] I. M. Gel'fand and M. L. Zeitlin, Dokl. Akad. Nauk. SSSR, 71, 1017 (1950); H. S. Green, J. Math. Phys. 12, 2106 (1971). The author is indebted to the referee for drawing his attention to these articles.
[4] M. Ikeda and T. Maekawa (to be published).

Ray Theory of Diffraction by Open-Ended Waveguides. II. Applications

S. W. Lee
Department of Electrical Engineering, University of Illinois, Urbana, Illinois 61801
(Received 3 November 1971; Revised Manuscript Received 10 December 1971)

This series of papers presents an extension of Keller's diffraction ray method to problems involving two or more parallel plates by introducing a modified diffraction coefficient which takes care of coupling along a shadow boundary automatically. In Part I, a canonical problem was solved and the expression for the modified diffraction coefficient derived. In this part, we give a recipe for how to use this set of rays and illustrate it through several examples including (i) open-end parallel-plate waveguide, (ii) bifurcated waveguide, and (iii) an infinite array of parallel plates. The above three examples represent the only three types of problems in the edge diffraction theory that can be solved exactly by analytical techniques based on complex variables. In this paper it is demonstrated that all three exact solutions can be recovered by the present ray method. Moreover, in some problems where the analytical techniques cannot be conveniently applied, the ray method can often provide a useful approximate solution.

1. INTRODUCTION

Recently there have been several efforts to apply the geometrical ray method to diffraction problems involving waveguide discontinuities in open and closed regions. In attacking these problems, one of the most important steps is to describe the interaction between the upper edge and the lower plate for the structure sketched in Fig. 1. In the celebrated method developed by Yee, Felsen, and Keller (YKF method),[1] this interaction is accounted for by including the following two sets of rays in the field calculations:

(i) the rays that bounce back and forth between the two edges and

(ii) the rays that result from the multiple reflections and diffractions along the shadow boundary at $z = 0$.

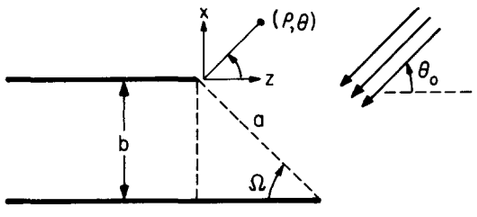


FIG. 1. Geometry of the canonical problem.

The rays in (i) can be easily evaluated. Furthermore, we usually have to consider the bounces no more than a few times since the ray amplitude decays as $(ka)^{-n/2}$ after the nth bounces ("a" is the distance between two edges; see Fig. 1). The calculation of rays in (ii) is much more difficult. In the YFK method, it is achieved by summing up the multiple scattering along the shadow boundary *based on a diffraction formula obtained from the classical half-plane problem.*[2] In other words, the solution in the YFK method is "built up" from that of a single half-plane under the assumption that $kb \gg 1$. Normally we would expect that the YFK method is valid only for reasonably large kb. However, numerical results show that it can give good accuracy for amazingly small kb (e.g., $b = 0.2\lambda$) in some cases.

A different approach to describe this interaction was taken in Paper I[3] (hereafter referred to as I), where we introduced a *modified diffraction coefficient* for the rays emerging from the upper edge. For the special case $kb \to \infty$, our modified diffraction coefficient is reduced to that of a half-plane, and, for finite kb, it includes the interaction between the upper edge and the lower plate along the shadow boundary at $z = 0$ *automatically*. Therefore, in applying our ray method to attack the problem sketched in Fig. 1, it is only necessary to consider the rays in (i) for the interaction part, but *not* the rays in (ii).

group S. In fact, the Jacobians become apart from a numerical factor

$$\frac{\partial(\overline{F}^{(n)}, \overline{G}_1^{(n)}, \ldots, \overline{G}_{(n-2)/2}^{(n)})}{\partial(L_1, L_2, \ldots, L_{n/2})} \simeq \prod_{i<j} (L_i - L_j)(L_i + L_j) \quad (9)$$

for an even n and

$$\frac{\partial(\overline{F}^{(n)}, \overline{G}_1^{(n)}, \ldots, \overline{G}_{(n-3)/2}^{(n)})}{\partial(L_1, L_2, \ldots, L_{(n-1)/2})} \simeq L_1 L_2 \cdots \times L_{(n-1)/2} \prod_{i<j} (L_i - L_j)(L_i + L_j) \quad (10)$$

for an odd n. Thus it demonstrates that the invariants $\overline{F}^{(n)}$ and $\overline{G}_p^{(n)}$ are a suitable basis for all the invariants of S, and then $F^{(n)}$ and $G_p^{(n)}$ are a suitable basis for all the invariants of $O(n)$.

Finally, it is noted that the invariant operators for the noncompact group $O(p, q)(p + q = n)$ are obtained explicitly from (2) and (3) [and (4) for an even n] by means of the substitution $D_{jk} \to iD_{jk}$ for $j \leq p, k \geq q$, or $j \geq q, k \leq p$ and retaining other D_{jk} in its original form.

ACKNOWLEDGMENT

The author would like to express his cordial thanks to Professor M. Ikeda for a helpful discussion.

[1] L. C. Biedenharn, J. Math. Phys. 4, 436 (1963). Detailed references are quoted in this paper.
[2] G. Racah, Lecture notes in "Group Theory and Spectroscopy," Princeton (1951); A. Klein, J. Math. Phys. 4, 1238 (1963); M. Umezawa, Nucl. Phys. 48, 111 (1963); 53, 54 (1964); 57, 65 (1964); M. Micu, *ibid.* 60, 353 (1964).
[3] I. M. Gel'fand and M. L. Zeitlin, Dokl. Akad. Nauk. SSSR, 71, 1017 (1950); H. S. Green, J. Math. Phys. 12, 2106 (1971). The author is indebted to the referee for drawing his attention to these articles.
[4] M. Ikeda and T. Maekawa (to be published).

Ray Theory of Diffraction by Open-Ended Waveguides. II. Applications

S. W. Lee
Department of Electrical Engineering, University of Illinois, Urbana, Illinois 61801
(Received 3 November 1971; Revised Manuscript Received 10 December 1971)

This series of papers presents an extension of Keller's diffraction ray method to problems involving two or more parallel plates by introducing a modified diffraction coefficient which takes care of coupling along a shadow boundary automatically. In Part I, a canonical problem was solved and the expression for the modified diffraction coefficient derived. In this part, we give a recipe for how to use this set of rays and illustrate it through several examples including (i) open-end parallel-plate waveguide, (ii) bifurcated waveguide, and (iii) an infinite array of parallel plates. The above three examples represent the only three types of problems in the edge diffraction theory that can be solved exactly by analytical techniques based on complex variables. In this paper it is demonstrated that all three exact solutions can be recovered by the present ray method. Moreover, in some problems where the analytical techniques cannot be conveniently applied, the ray method can often provide a useful approximate solution.

1. INTRODUCTION

Recently there have been several efforts to apply the geometrical ray method to diffraction problems involving waveguide discontinuities in open and closed regions. In attacking these problems, one of the most important steps is to describe the interaction between the upper edge and the lower plate for the structure sketched in Fig. 1. In the celebrated method developed by Yee, Felsen, and Keller (YKF method),[1] this interaction is accounted for by including the following two sets of rays in the field calculations:

(i) the rays that bounce back and forth between the two edges and

(ii) the rays that result from the multiple reflections and diffractions along the shadow boundary at $z = 0$.

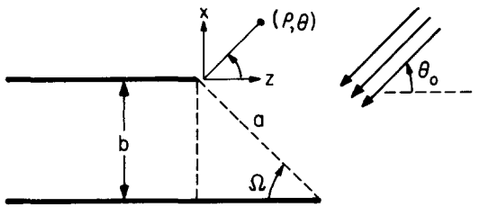


FIG. 1. Geometry of the canonical problem.

The rays in (i) can be easily evaluated. Furthermore, we usually have to consider the bounces no more than a few times since the ray amplitude decays as $(ka)^{-n/2}$ after the nth bounces ("a" is the distance between two edges; see Fig. 1). The calculation of rays in (ii) is much more difficult. In the YFK method, it is achieved by summing up the multiple scattering along the shadow boundary *based on a diffraction formula obtained from the classical half-plane problem.*[2] In other words, the solution in the YFK method is "built up" from that of a single half-plane under the assumption that $kb \gg 1$. Normally we would expect that the YFK method is valid only for reasonably large kb. However, numerical results show that it can give good accuracy for amazingly small kb (e.g., $b = 0.2\lambda$) in some cases.

A different approach to describe this interaction was taken in Paper I[3] (hereafter referred to as I), where we introduced a *modified diffraction coefficient* for the rays emerging from the upper edge. For the special case $kb \to \infty$, our modified diffraction coefficient is reduced to that of a half-plane, and, for finite kb, it includes the interaction between the upper edge and the lower plate along the shadow boundary at $z = 0$ *automatically*. Therefore, in applying our ray method to attack the problem sketched in Fig. 1, it is only necessary to consider the rays in (i) for the interaction part, but *not* the rays in (ii).

Furthermore, it is important to point out that our modified diffraction coefficient is derived from an asymptotic solution of a canonical problem for large ka, but *not* kb. Therefore, as long as the term $O(1/ka)$ can be safely ignored, our result is valid for any kb, no matter how small it is. This explains the fact that, in several problems with $ka \to \infty$, our ray method recovers their exact solutions.

The purposes for this part of the paper are the following three. First, we summarize the main results in I in a fashion such that without reading the quite involved derivations and manipulations in I the reader may gain a quick grasp of our ray method, and especially of the rules in applying it to attack some problems. This is done in Sec. 2 which contains essentially the same material as that in Sec. 6 in I. A major difference between the presentations in these two sections is that here we introduce only a single modification function [i.e. $f(\theta)$ in (2.3)] for both incoming and outgoing rays. By doing so, the reciprocity can be satisfied for this particular diffraction ray. A new contribution of the present Sec. 2 is that we give a recipe for the determination of the Green's function $G(\alpha)$ in a given problem, which is of vital importance in our later applications.

The second purpose for this part of the paper is to illustrate the application of our ray method by considering several examples involving edge diffraction in waveguides and periodic structures. The examples are (i) bifurcated waveguide (Sec. 3), (ii) step discontinuity in waveguide (Sec. 4), and (iii) infinite array of waveguides (Sec. 5). In the case of (i) and (iii), our ray method recovers the exact solutions as expected. For the problem in (ii), there is no known analytical solution, exact or approximate. Hence, there is no way to estimate the accuracy of the results obtained by our ray method.

The third purpose for this part of the paper is to make a quantitative comparison between the solution for the bifurcated waveguide obtained by the YFK method and that obtained by our ray method which is also the exact solution. Since the YFK solution is valid only for large waveguide dimensions, we need to extract the dominant asymptotic term from the exact solution for the purpose of fair comparison. This is accomplished by using asymptotic formulas developed by Weinstein,[4] which are summarized in Sec. 6 together with approximate formulas useful in simplifying our solutions for small waveguide dimensions. The comparison in Sec. 7 reveals that the YFK solution checks with the asymptotic expression of the exact solution in their first two dominant terms provided that the particular mode of interest is not close to cutoff. Otherwise, the YFK method does not seem able to provide a satisfactory solution.

2. THEORY

The canonical problem for our system of rays is the one shown in Fig. 1, namely, the diffraction of a plane wave by two staggered parallel plates. The incident plane wave can be either $TM(H_y, E_x, E_z)$ or $TE(E_y, H_x, H_z)$. This problem has been rigorously solved in I by the Wiener–Hopf technique, and its solution arranged in a form which admits a ray interpretation. The main conclusions are now summarized below.

A. Modified Diffraction Coefficient

To introduce our modification conveniently, consider first the special case $kb \to \infty$. In the absence of the lower plate, the field ψ (representing H_y for TM and E_y for TE) on the ray diffracted at the edge of the upper plate is given by the well-known expression [for $\exp(-i\omega t)$ time convention]

$$\psi(x, z) = (e^{i(k\rho - \pi/4)}/\sqrt{2\pi k\rho})D(\theta, \theta_0), \tag{2.1}$$

where θ_0 is the direction of the incident plane wave (with unit amplitude at $x = z = 0$) and (ρ, θ) is the observation point. The factor $D(\theta, \theta_0)$ is known as the diffraction coefficient (for a half-plane):

$$D(\theta, \theta_0) = \begin{cases} \dfrac{2i \sin\frac{1}{2}\theta_0 \sin\frac{1}{2}\theta}{\cos\theta_0 + \cos\theta}, & \text{for } TM, \\ \dfrac{-2i \cos\frac{1}{2}\theta_0 \cos\frac{1}{2}\theta}{\cos\theta_0 + \cos\theta}, & \text{for } TE, \end{cases} \tag{2.2}$$

where the angles θ and θ_0 take values between $-\pi$ and π. For finite kb, the diffraction coefficient is modified in order to take into account the interaction between the upper edge at $x = z = 0$ and the lower plate along the shadow boundary $z = 0$. Our result indicates that the modified diffraction coefficient $\bar{D}(\theta, \theta_0)$ takes the following form:

$$\bar{D}(\theta, \theta_0) = D(\theta, \theta_0) f(\theta) f(\theta_0), \tag{2.3a}$$

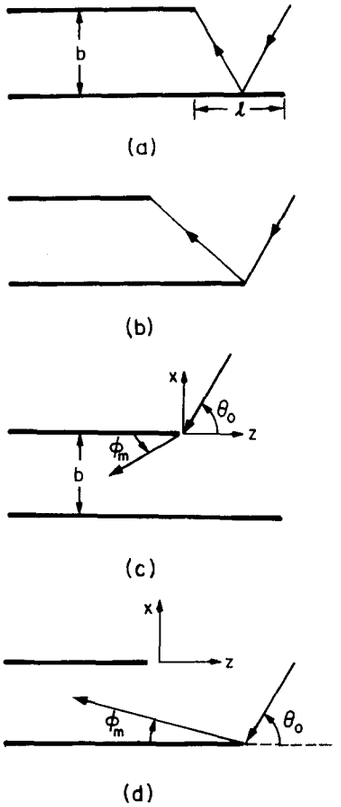


FIG. 2. Various rays in the canonical problem.

where

$$f(\theta) = \begin{cases} G_+(-k\cos\theta), & \pi/2 < |\theta_0| < \pi, \\ [G_+(k\cos\theta)]^{-1}, & |\theta_0| < \pi/2, \end{cases} \quad (2.3b)$$

where $G_+(\alpha)$ is the "plus factor"[5] of the normalized transformed Green's function, and will be detailed in subsection B below. The result in (2.3) is correct to the order of $(ka)^{-1/2}$. Note that $a = b \csc\Omega$. Thus, provided Ω is small enough, (2.3) is valid for any kb. It should be emphasized that the modification in (2.3) takes care of the multiple reflections and diffractions along the shadow boundary ($z = 0$) only, whereas the other interaction between the upper edge and the lower plate must be accounted for *separately*. Examples of latter types of interactions are the specularly reflected ray[8] shown in Fig. 2(a) and the ray bounced between two edges in Fig. 2(b), both of which can be easily calculated. For diffraction at the lower edge at ($x = -b$, $z = l$), the diffraction coefficient is given by $\bar{D}(\theta, \theta_0)$ in (2.2), not $\bar{D}(\theta, \theta_0)$ in (2.3). This is because of the fact that no shadow boundary can be found between the upper plate and the lower edge.

B. Green's Function

The function $G_+(\alpha)$ is related to the (normalized, transformed) Green's function $G(\alpha)$ by[6,7]

$$G_+(\alpha) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{ln\, G(\beta)}{\beta - \alpha}\, d\beta, \quad (2.4)$$

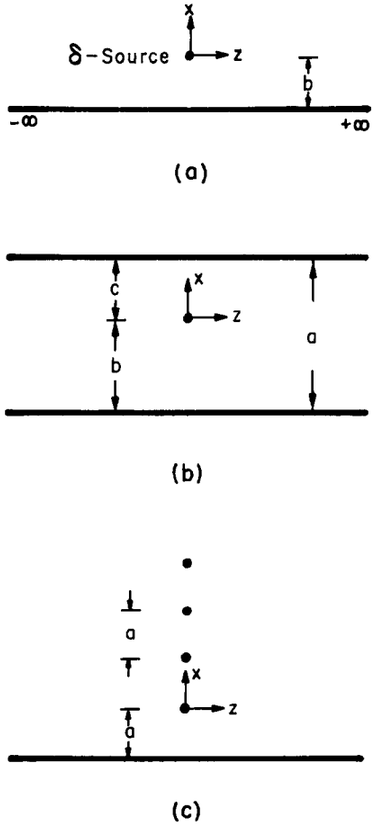


FIG. 3. Configurations for calculating the Green's function $G(\alpha)$.

which may be identified to the "plus part"[9] of $G(\alpha)$ in the Wiener–Hopf technique. The function $G(\alpha)$ is the normalized version of $\epsilon(x = 0, \alpha)$, which is the Fourier transformed Green's function for the tangential electric field (i.e., E_z for TM and E_y for TE) due to a point source at $x = z = 0$ in a configuration shown in Fig. 3(a). The normalization[10] is chosen such that

$$G(\alpha) = \epsilon(x = 0, \alpha)/[\lim_{b\to\infty} \epsilon(x = 0, \alpha)]. \quad (2.5)$$

To determine $\epsilon(x, \alpha)$ we may follow the following procedure.[7] The Green's function in the spatial domain satisfies the wave equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + k^2\right) E(x, z) = -\delta(x)\delta(y), \quad (2.6)$$

where E stands for E_z in the case of TM, and E_y of TE. Taking the Fourier transform of (2.6) defined by

$$\epsilon(x, \alpha) = \int_{-\infty}^{\infty} E(x, z) e^{i\alpha z} dz, \quad (2.7)$$

one has

$$\left(\frac{\partial^2}{\partial x^2} - \gamma^2\right) \epsilon(x, \alpha) = -\delta(x), \quad (2.8)$$

where $\gamma = \sqrt{\alpha^2 - k^2} = -i\sqrt{k^2 - \alpha^2}$. With the boundary condition that $\epsilon(x = -b, \alpha) = 0$, it is a simple matter to derive the result

$$\epsilon(x = 0, \alpha) = (1 - e^{-2\gamma b})/2\gamma. \quad (2.9)$$

It follows from (2.5) that

$$G(\alpha) = 1 - e^{-2\gamma b}. \quad (2.10)$$

With $G(\alpha)$ given in (2.10) the integral in (2.4) can be carried out explicitly with the result[11]

$$G_+(\alpha) = \sqrt{2kb}\sqrt{1 + \frac{\alpha}{k}}\, e^{-i\pi/4}\sqrt{\frac{\sin kb}{kb}} \times \exp\left[\frac{i\alpha b}{\pi}\left(1 - 0.57721 + \ln\frac{2\pi}{kb} + i\frac{\pi}{2}\right)\right] \times \exp\left(\frac{ib\gamma}{\pi}\ln\frac{\alpha - \gamma}{k}\right) \prod_{n=1}^{\infty}\left(1 + \frac{\alpha}{i\gamma_n}\right) e^{i\alpha b/n\pi}, \quad (2.11)$$

where $\gamma_n = [(n\pi/b)^2 - k^2]^{1/2} = -i[k^2 - (n\pi/b)^2]^{1/2}$. A short numerical table for $G_+(\alpha)$ was given in I. Further discussions on the computation of $G_+(\alpha)$ will be given in Sec. 6. A fact worth emphasizing is that in a given problem the function $G(\alpha)$ is identically the same for TM and TE waves.

C. Ray-to-Mode Conversion

Referring to Fig. 2(c), let us express the modal field solution in the waveguide as

$$\begin{Bmatrix} H_y \\ E_x \end{Bmatrix} = \sum_m \frac{1}{\epsilon_m} C_m \begin{Bmatrix} \cos(m\pi x/b) \\ \sin(m\pi x/b) \end{Bmatrix} e^{+\gamma_m z} \quad (2.12)$$

where $\epsilon_m = 2$, $m = 0$ and $\epsilon_m = 1$, $m \neq 0$. For propagating modes, the contribution to C_m due to a given ray is equal to

$$\psi^{(i)}(x=0, z=0)\bar{D}(\theta=\phi_m-\pi, \theta_0) \times [(2kb\cos\phi_m)^{-1}]N^{-1}, \quad (2.13)$$

where $\phi_m = \sin^{-1}(m\pi/kb)$ is the direction of propagation for the plane wave associated with the mth mode. The first factor in (2.13) is the amplitude of the incident ray evaluated at the edge, and the factor in [ ] in (2.13) is the *ray-to-mode conversion factor*. The constant N in (2.13) is a normalization factor and is equal to the ampltiude of the plane wave component in the direction of θ in modal field representation when evaluated at the edge. As an example, for the ray shown in Fig. 2(c) we have from (2.12) that

$$TE;\ N = -(1/2i)e^{-i(m\pi x/b)}e^{\gamma_m z}\Big|_{x=z=0} = \frac{i}{2}. \quad (2.14)$$

For the ray shown in Fig. 2(d), we have

$$TM;\ N = \tfrac{1}{2}e^{-i(m\pi x/b)}e^{\gamma_m z}\Big|_{\substack{x=-b\\ z=l}} = [(-1)^m/2]e^{\gamma_m l}. \quad (2.15)$$

With the main results summarized as above, we will now state the rules of using the set of new diffraction coefficients as follows:

(i) The new edge diffraction ray given in (2.1) and (2.3) can be used exactly in the same manner as the conventional Keller's edge diffraction ray given in (2.1) and (2.2). The only difference is that the new ray has already included the interaction between the edge and its neighboring plates along the shadow boundary at $z = 0$ (for all the configurations in Figures 1, 4, 5, and 7.) Therefore, in using the new rays, a separated account of such an interaction is no longer needed.

(ii) The Green's function appearing in the diffraction coefficient in (2.3) can be found by following the method outlined in Sec. 2B above [for the three common configurations shown in Fig. 3, their explicit forms are given in (2.11), (3.7), and (5.8), respectively].

(iii) In computing the field in a waveguide, the amplitude of a modal field in the waveguide (properly normalized) is equal to the product of the diffraction coefficient in (2.3) and a ray-to-mode conversion factor, as indicated in (2.13).

In the next three sections, we will illustrate how the new system of rays can be applied to solve some closed-region boundary value problems having edges of a half-plane type.

3. BIFURCATED WAVEGUIDE

As a first example, let us consider the problem of scattering by a bifurcated waveguide as shown in Fig. 4(a). The incident field is a TE_l mode from guide of width a,

$$E_y^{(i)} = \{-2ie^{il\pi b/a}\}\sin(l\pi/a)(x+b)e^{\gamma_{la}z}, \quad 0<x<a \text{ and } z>0, \quad (3.1)$$

where $\gamma_{nd} = \sqrt{(n\pi/d)^2 - k^2}$, $n = 1, 2, 3, \cdots$ and $d = a, b, c$. Note that the factor in $\{\ \}$ in (3.1) is the normalization factor so that the ray traveling in the direction $\theta = \phi_l = \sin^{-1}(l\pi/ka)$ has a *unit* amplitude at the edge of the bifurcated half-plane $x = z = 0$.

The problem at hand is to determine the scattered field in all three waveguides, namely

$$E_y = \begin{cases} \sum_{m=1}^{\infty} A_m\{2ie^{-im\pi b/a}\}\sin(m\pi/a)(x+b)e^{-\gamma_{ma}z}, \\ \qquad 0<x<a \text{ and } z>0 \\ \sum_{m=1}^{\infty} B_m\{-2ie^{im\pi b/b}\}\sin(m\pi/b)(x+b)e^{-\gamma_{mb}z}, \\ \qquad 0<x<b \text{ and } z<0, \\ \sum_{m=1}^{\infty} C_m\{2i\}\sin(m\pi/c)\,xe^{\gamma_{mc}z}, \\ \qquad b<x<a \text{ and } z<0. \end{cases} \quad (3.2)$$

It may be noted that the factors in $\{\ \}$ in (3.2) are again normalizations so that the ray amplitudes for the three rays shown in Fig. 4(a) are unity at $x = z = 0$ in the modal field. Such a normalization will make N defined in (2.13) unity in the later computation.

To apply our ray method, we need first to determine the transformed Green's function for the auxiliary problem, namely, the structure shown in Fig. 3(b), where the bifurcated half-plane has been *removed*. Explicitly we are looking for the solution to the transformed wave equation

$$\left(\frac{\partial^2}{\partial x^2} - \gamma^2\right)\epsilon(x, \alpha) = -\delta(x) \quad (3.3)$$

subject to the boundary conditions

$$\epsilon(x, \alpha) = 0, \quad \text{for } x = -b \text{ and } x = c. \quad (3.4)$$

A little manipulation leads to

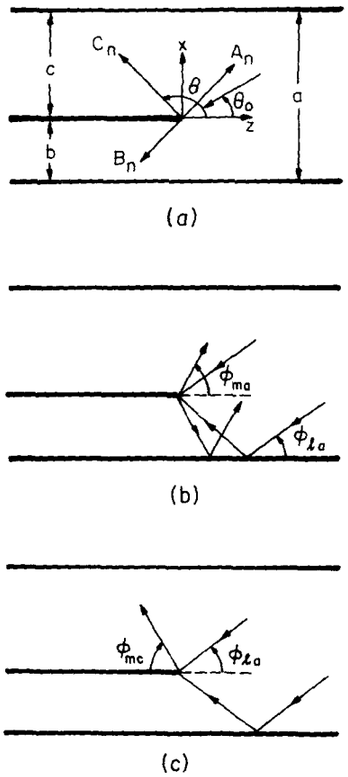


FIG. 4. Scattering in a bifurcated waveguide.

$$\epsilon(x=0,\alpha)=[(1-e^{-2\gamma b})(1-e^{-2\gamma c})]/[2\gamma(1-e^{-2\gamma a})] \tag{3.5}$$

It follows from (2.5) that the normalized transformed Green's function for the parallel plate waveguide is given by

$$G^{(1)}(\alpha)=[(1-e^{-2\gamma b})(1-e^{-2\gamma c})]/(1-e^{-2\gamma a}). \tag{3.6}$$

The superscript in $G^{(1)}(\alpha)$ is to differentiate it from the Green's function defined in (2.10). The "plus part" of $G^{(1)}(\alpha)$ as defined in (2.4) can be found explicitly, and is given by[7]

$$G_+^{(1)}(\alpha)=\sqrt{\frac{(1-e^{i2kb})(1-e^{i2kc})}{(1-e^{i2ka})}}\times\sqrt{1+\alpha/k}\,e^{i(\alpha/\pi)[b\ln(a/b)+c\ln(a/c)]}\times\prod_{n=1}^{\infty}\frac{(1+\alpha/i\gamma_{nb})(1+\alpha/i\gamma_{nc})}{(1+\alpha/i\gamma_{na})}. \tag{3.7}$$

We will discuss some computational aspects of $G_+^{(1)}(\alpha)$ in Secs. 6 and 7.

Once $G_+^{(1)}(\alpha)$ is found, we can immediately write down the solutions for modal coefficients by following the rules in Sec. 2, namely, for the field in guide of width a,

$$A_m=\left(\frac{-2i\cos\frac{1}{2}\phi_{la}\cos\frac{1}{2}\phi_{ma}}{\cos\phi_{la}+\cos\phi_{ma}}\frac{1}{G_+^{(1)}(k\cos\phi_{la})G_+^{(1)}(k\cos\phi_{ma})}\right)\times\left(\frac{1}{2ka\cos\phi_{ma}}\right)[(1-e^{i2kb\sin\phi_{la}})(1-e^{i2kb\sin\phi_{ma}})], \tag{3.8a}$$

$$\phi_{md}=\sin^{-1}(m\pi/kd),\qquad \text{for } d=a,b,c. \tag{3.8b}$$

Note that the two exponential terms in (3.8) correspond to the specularly reflected incoming and outgoing rays at the ground plane at $x=-b$, as sketched in Fig. 4(b). Similarly, we can write down the field in the two other guides, and the results are

$$B_m=\left(\frac{-2i\cos\frac{1}{2}\phi_{la}\cos[\frac{1}{2}(\phi_{mb}-\pi)]}{\cos\phi_{la}+\cos(\phi_{mb}-\pi)}\frac{G_+^{(1)}(-k\cos(\phi_{mb}-\pi))}{G_+^{(1)}(k\cos\phi_{la})}\right)\left(\frac{1}{2kb\cos\phi_{mb}}\right)(1-e^{i2kb\sin\phi_{la}}), \tag{3.9}$$

$$C_m=\left(\frac{-2i\cos\frac{1}{2}\phi_{la}\cos[\frac{1}{2}(\pi-\phi_{mc})]}{\cos\phi_{la}+\cos(\pi-\phi_{mc})}\frac{G_+^{(1)}(-k\cos(\pi-\phi_{mc}))}{G_+^{(1)}(k\cos\phi_{la})}\right)\left(\frac{1}{2kc\cos\phi_{mc}}\right)(1-e^{i2kb\sin\phi_{la}}). \tag{3.10}$$

It is interesting to note that, unlike the field in guide of width a, there is no specular reflection for the outgoing rays in the fields in the smaller guides [Fig. 4(c)].

The solutions given in (3.8)–(3.10) derived by our ray method may be compared with the exact solution of the bifurcated problem obtainable by either the residue calculus method[7,12] or Wiener-Hopf technique.[7] They are identical. This is not surprising in view of the fact that our ray theory summarized in Sec. 2 was based on an asymptotic solution for large $kb\csc\Omega$ (Fig. 1). In the bifurcation problem, $\Omega\to 0$, and hence our ray method recovers the exact solution.

Another point worth noting is that, even though our ray method is designed for computing propagating modes only, the solutions given (3.8)–(3.10) remain valid when $\{\phi_{la},\phi_{ma},\phi_{mb},\phi_{mc}\}$ become imaginary (for evanescent modes). However, such an extension of our ray method to cover evanescent modes has not been shown to be true in general.

4. STEP DISCONTINUITY IN WAVEGUIDE

In this section, we will consider a waveguide discontinuity problem which has no known analytical solution, exact or approximate. This is the step discontinuity shown in Fig. 5, with the wedge angle $\beta<\pi/2$.

The incident field from the larger waveguide is a TM_{l0} mode given by

$$H_y^i=(2/\epsilon_l)e^{il\pi b/a}\cos(l\pi/a)\,(x+b)e^{\gamma_l a z},\qquad 0<x<a \text{ and } z>0, \tag{4.1}$$

where $\epsilon_l=2$ if $l=0$ and $\epsilon_l=1$ if $l\neq 0$. The problem is to determine the scattered field,

$$H_y=\begin{cases}\sum_{m=0}^{\infty}A_m\frac{2}{\epsilon_m}e^{-im\pi b/a}\cos\left(\frac{m\pi}{a}\right)(x+b)e^{-\gamma_m a z}, & 0<x<a,\ z>0,\\ \sum_{m=0}^{\infty}C_m\frac{2}{\epsilon_m}\cos\left(\frac{m\pi}{c}\right)xe^{\gamma_m c z}, & b<x<c \text{ and } z<0.\end{cases} \tag{4.2}$$

The field in the triangular region between $x=0$ and $x=-a$ is quite complicated. Fortunately, the field there is generally not of primary interest.

Provided that β is much less than $(\pi/2)$, the interaction between the edge of a *wedge* and the waveguide walls is expected to be approximately the same as that between the edge of a *half-plane* and the waveguide walls. Under such an assumption, we can write down the solution to the present problem as below. First, we note that the diffraction coefficient $D(\theta,\theta_0)$ given in (2.2) is that for a *half-plane*, and therefore should be replaced by the corresponding expression for a wedge, which is given by the well-known expression

$$D_w(\theta,\theta_0)=\frac{\pi}{i(2\pi-\beta)}\sin\frac{\pi^2}{(2\pi-\beta)}\times\left[\left(\cos\frac{\pi(\theta-\theta_0)}{2\pi-\beta}-\cos\frac{\pi^2}{2\pi-\beta}\right)^{-1}+\left(\cos\frac{\pi(2\pi-\theta-\theta_0)}{2\pi-\beta}-\cos\frac{\pi^2}{2\pi-\beta}\right)^{-1}\right],\qquad \text{for } TM, \tag{4.3}$$

where β is the wedge angle. The diffraction coefficient in (4.3) no longer has the simple symmetry when θ (or θ_0) is changed to $-\theta$ (or $-\theta_0$) as we had in (2.2). Thus, we have to keep track of the five different rays shown in Fig. 5 separately. The result for the reflected field is

$$A_m = [D_w(\phi_{ma}, \phi_{la}) + D_w(-\phi_{ma}, \phi_{la})e^{i2kb\sin\phi_{ma}} + e^{i2kb\sin\phi_{la}} D_w(\phi_{ma}, -\phi_{la}) + e^{i2kb\sin\phi_{la}} \times D_w(-\phi_{ma}, -\phi_{la})e^{i2kb\sin\phi_{ma}}] \times \left(\frac{1}{G_+^{(1)}(k\cos\phi_{la})G_+^{(1)}(k\cos\phi_{ma})}\right) \times \left(\frac{1}{2ka\cos\phi_{ma}}\right), \quad (4.4)$$

where $G_+^{(1)}(\alpha)$ is given (3.7). The first, second, third, and fourth terms in $\{\}$ (4.4) correspond to the rays 1-3, 1-4, 2-3, and 2-4, respectively. A numerical example for A_0 (reflection coefficient of TEM) is presented in Fig. 6. We emphasize again that the result in (4.4) is approximate. Until there is a more rigorous way of attacking this problem, we cannot estimate its accuracy.

For the special case $\beta = 0$, the problem becomes that of the scattering of a TM wave in a bifurcated waveguide, and D_W in (4.3) is reduced to D in (2.2). Making use of the property of symmetry

$$D(\theta, \theta_0) = \text{sgn}(\theta, \theta_0)D(|\theta|, |\theta_0|), \quad \text{for } TM, \quad (4.5)$$

where sgn $x = +1$ if $x > 0$ and sgn $x = -1$ if $x < 0$, we obtain the reflected coefficients for a bifurcated waveguide (TM) from (4.4), namely,

$$A_m = \left(\frac{2i\sin\frac{1}{2}\phi_{la}\sin\frac{1}{2}\phi_{ma}}{\cos\phi_{la} + \cos\phi_{ma}} \frac{1}{G_+^{(1)}(k\cos\phi_{la})G_+^{(1)}(k\cos\phi_{ma})}\right)$$

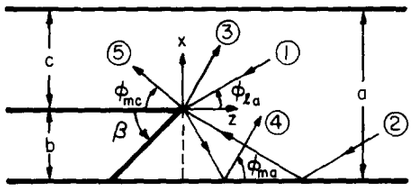


FIG. 5. Scattering by a step discontinuity in a waveguide.

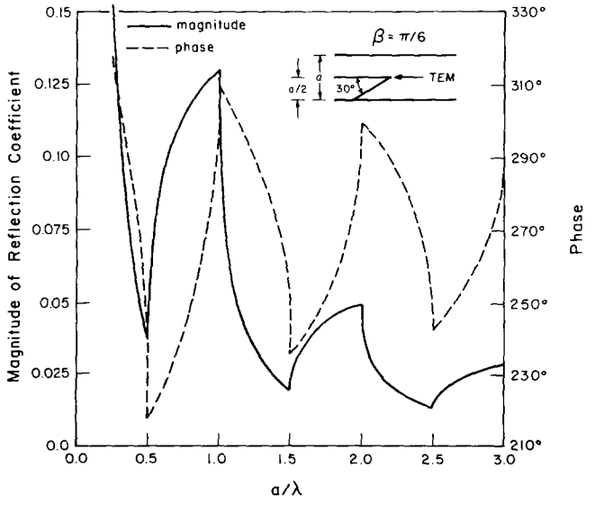


FIG. 6. Reflection coefficient for a TEM mode at the junction of a step discontinuity in a waveguide.

$$\times\left(\frac{1}{2ka\cos\phi_{ma}}\right)[(1 - e^{i2kb\sin\phi_{la}}) \times (1 - e^{i2kb\sin\phi_{ma}})], \quad \text{if } \beta = 0. \quad (4.6)$$

This result may be compared with (3.8) which gives the reflected coefficients for TE waves in the bifurcation problem. It is of particular interest to note that the factors in [ ] in (3.8) and (4.6) due to specular reflections are the same despite that the specular reflection coefficient for E_y in TE is (-1), and that for H_y in TM is $(+1)$. This is because of

$$D(\theta, -\theta_0) = \begin{cases} +D(\theta, +\theta_0), & \text{for } TE, \\ -D(\theta, +\theta_0), & \text{for } TM, \end{cases} \quad (4.7)$$

which contributes an extra minus sign for TM wave.

5. PERIODIC ARRAY OF WAVEGUIDES

The other class of closed-region boundary value problems are those involving periodic structures. Here we will consider one example of this type, namely, the scattering of a plane wave (TM) by an infinite periodic array of waveguides [Fig. 7(a)]. The incident field is

$$H_y^{(i)} = e^{-ik(x\sin\theta_0 + z\cos\theta_0)}, \quad z > 0. \quad (5.1)$$

The problem is to determine the scattered field

$$H_y = \begin{cases} \sum_{p=-\infty}^{\infty} A_p e^{-i\beta_p x}e^{-\Gamma_p z}, & z > 0, \\ \sum_{m=0}^{\infty} B_m \cos\left(\frac{m\pi}{b}\right) xe^{\gamma_m z}, & -b < x < 0 \text{ and } z < 0, \end{cases} \quad (5.2)$$

where $\beta_p = k\sin\theta_0 + (2p\pi/b)$, $\Gamma_p = (\beta_p^2 - k^2)^{1/2}$, and $\gamma_m = [(m\pi/b)^2 - k^2]^{1/2}$. In order to apply our ray method in a convenient manner, let us consider the equivalent problem shown in Fig. 7(b) with an incident field given by

$$H_y^i = [e^{-ikx\sin\theta_0} + e^{ik(x+2b)\sin\theta_0}]e^{-ikz\cos\theta}, \; z > 0. \quad (5.3)$$

From symmetry consideration, the scattered field for the problem in Fig. 7(b) may be shown to be

$$H_y = \sum_{p=-\infty}^{\infty} A_p(e^{-i\beta_p x} + e^{i2kb\sin\theta_0}e^{i\beta_p x})e^{-\Gamma_p z}, \; z > 0, \quad (5.4a)$$

$$H_y = \sum_{\substack{m=0 \\ \text{even}}}^{\infty} \frac{2}{\epsilon_m} B_m(1 + e^{ikb\sin\theta_0})\cos\left(\frac{m\pi}{b}\right) xe^{\gamma_m z} + \sum_{\substack{m=1 \\ \text{odd}}}^{\infty} 2B_m(1 - e^{ikb\sin\theta_0})\cos\left(\frac{m\pi}{b}\right) xe^{\gamma_m z}, \quad -b < x < 0 \text{ and } z < 0. \quad (5.4b)$$

Now we will solve this equivalent problem by our ray method.

First we need to determine the Green's function, which is the solution of the transformed wave equation

$$\left(\frac{\partial^2}{\partial x^2} - \gamma^2\right)\epsilon(x, \alpha) = -\sum_{n=0}^{\infty} \delta(x - nb)e^{-inkb\sin\theta_0}, \quad (5.5)$$

subject to the boundary condition $\epsilon(x = -b, \alpha) = 0$. A little algebra leads to the results

$$\epsilon(x,\alpha)=\frac{1-e^{-2\gamma b}}{2\gamma}\sum_{n=0}^{\infty}\frac{\sinh(n+1)\gamma a}{\sinh\gamma a}\,e^{-inka\sin\theta_0}$$
$$=\frac{1-e^{-2\gamma b}}{2\gamma(1-e^{b(\gamma-ik\sin\theta_0)})(1-e^{-b(\gamma+ik\sin\theta_0)})}. \quad (5.6)$$

It follows from (2.5) and (5.6) that

$$G^{(2)}(\alpha)=\frac{1-e^{-2\gamma b}}{(1-e^{-b(\gamma-ik\sin\theta)})(1-e^{-b(\gamma+ik\sin\theta_0)})}. \quad (5.7)$$

The factorization of $G^{(2)}(\alpha)$ can be found in Ref. 7, and is given by

$$G_+^{(2)}(\alpha)=\sqrt{\frac{1-e^{i2kb}}{(1-e^{ikb(1-\sin\theta_0)})(1-e^{ikb(1+\sin\theta_0)})}}$$
$$\times\sqrt{1+\frac{\alpha}{k}}\,\frac{e^{-i(\alpha b/\pi)\ln 2}}{1+\alpha/i\Gamma_0}$$
$$\times\prod_{n=1}^{\infty}\frac{1+\alpha/i\gamma_n}{(1+\alpha/i\Gamma_n)(1+\alpha/i\Gamma_{-n})}. \quad (5.8)$$

Once $G_+^{(2)}(\alpha)$ is found, the solution for the problem shown in Fig. 7(b) can be written down immediately. For the reflected field, we have

$$A_p=\left(\frac{2i\,\sin\frac{1}{2}\theta_0\sin\frac{1}{2}\psi_p}{\cos\theta_0+\cos\psi_p}\,\frac{1}{G_+^{(2)}(k\cos\theta_0)G_+^{(2)}(k\cos\psi_p)}\right)$$
$$\times\left(\frac{1}{kb\cos\psi_p}\right), \quad (5.9)$$

where $\psi_p=-\sin^{-1}(\beta_p/k)$, $0\le|\psi_p|<\pi/2$. In (5.9) we note the following:

(i) The specularly reflected rays are not included because of our representation of the fields in (5.3) and (5.4a), which automatically takes care of the specular reflection from the ground plane $x=-b$.

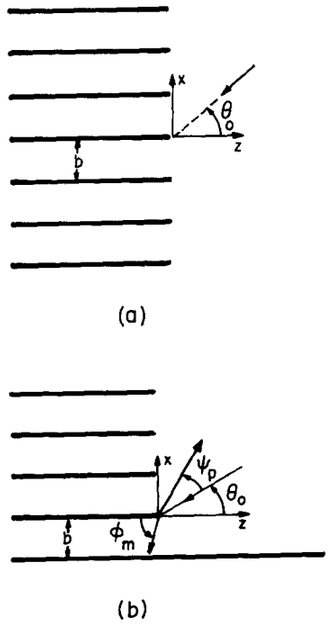


FIG. 7. Scattering by an infinite array of parallel plates.

(ii) Each waveguide mode (except for *TEM*) consists of two plane wave [e.g., see (2.12)]. Hence, the conversion factor for ray-to-space harmonics is simply twice that for ray-to-mode. This explains the absence of the factor 1/2 in the last factor in (5.9).

For the transmitted field, we have

$$B_m=\left(\frac{2i\,\sin\frac{1}{2}\theta_0\,\sin[\frac{1}{2}(\phi_m-\pi)]}{\cos\theta_0+\cos(\phi_m-\pi)}\,\frac{G_+(-k\cos(\phi_m-\pi))}{G_+(k\cos\theta_0)}\right)$$
$$\times\left(\frac{1}{2kb\cos\phi_m}\right)[(1-e^{i2kb\sin\theta_0})]N_m^{-1}, \quad (5.10)$$

where N_m is the normalization factor and may be found from (5.4b),

$$N_m=\begin{cases}(1+e^{ikb\sin\theta_0}), & m=0,2,4,\cdots,\\ (1-e^{ikb\sin\theta_0}), & m=1,3,5,\cdots.\end{cases} \quad (5.11)$$

The solutions given in (5.9)–(5.11) are again identical to the exact solutions obtainable through the Wiener–Hopf[7,13] or the residue calculus method.[7,14,15]

6. APPROXIMATE FORMULAS FOR THE GREEN'S FUNCTION

As illustrated in previous examples, the central step in calculating our modified ray amplitudes lies in the computation of the "plus part" of the normalized transformed Green's function $G_+(\alpha)$. In the present paper, we have given the explicit forms of $G_+(\alpha)$ for three commonly encountered configurations, namely (i) $G_+(\alpha)$ in (2.11) for a half-plane above a ground plane [cf. Fig. 4(a) and Fig. 1], (ii) $G_+(\alpha)$ in (3.7) for a half-plane in a parallel-plate waveguide [cf. Fig. 4(b) and Fig. 5], and (iii) $G_+^{(2)}(\alpha)$ in (5.8) for a periodic array of half-planes [cf. Fig. 4(c) and Fig. 7]. They are all expressed in terms of the form of *infinite* products. For small and moderate guide dimensions, the infinite products converge quite rapidly; however, for large guide dimensions, their computations can be very laborious. Take (2.11) as an example; it may be shown that the normalized truncation error for a N-term product is less than $(\alpha b/\pi)^2/N$ provided that $[(\alpha b/\pi)^2/N]\ll 1$. Thus for very large $(\alpha b/\pi)$, the convergence of the infinite product is slow and the computation of $G_+(\alpha)$ becomes laborious. In such an event, it is desirable to use an asymptotic formula given below.

For large ka, kb, and kc, all the three functions in (2.11), (3.7), and (5.8) can be approximately[16] expressed in terms of a universal function $U(s,p)$ defined by Weinstein.[4] The results are

$$G_+(k\cos\theta)\sim\exp\{U(s=\sqrt{2kb}\cos\theta,\ p=kb/\pi-|[kb/\pi]|)\}, \quad (6.1a)$$

$$G_+^{(1)}(k\cos\theta)\sim\exp\{U(s=\sqrt{2kb}\cos\theta,\ p=kb/\pi-|[kb/\pi]|)+U(s=\sqrt{2kc}\cos\theta,\ p=kc/\pi-|[kc/\pi]|)-U(s=2ka\cos\theta,\ p=ka/\pi-|[ka/\pi]|)\}, \quad (6.1b)$$

$$G_+^{(2)}(k\cos\theta)\sim\exp\{U(s=\sqrt{2kb}\cos\theta,p=kb/2\pi-|[kb/2\pi]|)-U(s=\sqrt{kb}\cos\theta,$$

$$p = [kb(1 + \sin\theta_0)/2\pi] - |[kb(1 + \sin\theta_0)/2\pi]|)$$
$$- U(s = \sqrt{kb}\cos\theta,\ p = [kb(1 - \sin\theta_0)/2\pi]$$
$$- |[kb(1 - \sin\theta_0)/2\pi]|)\}, \quad (6.1c)$$

for $ka, kb, kc \gg 1$,

where $|[d]|$ means the largest integer in d. The function $U(s, p)$ is defined as

$$U(s, p) = \frac{e^{-i\pi/4}}{\sqrt{2\pi}} \sum_{n=1}^{\infty} \frac{e^{in(2\pi p - s^2/2)}}{n} \int_{\sqrt{n}s}^{\infty} e^{it^2/2} dt \quad (6.2)$$

and may be approximated by

(i) $s \gg 1$,

$$U(s, p) = \frac{e^{-i3\pi/4}}{\sqrt{2\pi}s} \sum_{n=1}^{\infty} \frac{e^{i2\pi pn}}{n^{3/2}} + O\left(\frac{1}{s^3}\right), \quad (6.3a)$$

(ii) $s \ll 1$,

$$U(s, p) = \tfrac{1}{2}\ln(1 - e^{i2\pi p}) + \ln\left(1 + \frac{s}{\sqrt{4\pi p}}\right)$$
$$+ s\left(e^{-i\pi/4} \sum_{n=1}^{\infty} \frac{e^{i2\pi pn}}{\sqrt{2\pi n}} - \frac{1}{\sqrt{4\pi p}}\right) + O(s^2), \quad (6.3b)$$

(iii) $s \ll 1$ and $p \ll 1$,

$$U(s, p) = -\tfrac{1}{2}(\ln 2 + i\pi/2) + \ln(s + \sqrt{4\pi p})$$
$$- \tfrac{1}{2}(1 - i)0.824\sqrt{4\pi p} + O(s^2) + O(p) + O(s\sqrt{p}). \quad (6.3c)$$

An application of these asymptotic formulas will be given in Sec. 7.

For small ka, kb, and kc, the infinite products in (2.11), (3.7), and (5.8) converge very rapidly. As a matter of fact, if we drop the terms of $(ka/\pi)^2$, $(kb/\pi)^2$, $(kc/\pi)^2$ and higher, the following simplified results may be obtained:

$$G_+(k\cos\theta) \approx 2e^{-i\pi/4}\left|\cos\frac{\theta}{2}\right|\sqrt{\sin kb}\,\exp\{[(ikb\cos\theta)/\pi]$$
$$\times [1 - 0.57721 + \ln(2\pi/kb) + i\pi/2]\}$$
$$\times \exp[(ikb/\pi)|\theta|\,|\sin\theta|], \quad (6.4a)$$

$$G_+^{(1)}(k\cos\theta) \approx \sqrt{(1 - e^{i2kb})(1 - e^{i2kc})/(1 - e^{i2ka})}$$
$$\sqrt{2}\left|\cos\tfrac{1}{2}\theta\right| \times \exp\{[(ik\cos\theta)/\pi][b\ln(a/b)$$
$$+ c\ln(a/c)]\}, \quad (6.4b)$$

$$G_+^{(2)}(k\cos\theta)$$
$$\approx \sqrt{(1 - e^{i2kb})/[1 - e^{ikb(1+\sin\theta_0)}][1 - e^{ikb(1-\sin\theta_0)}]}$$
$$\times \frac{\sqrt{2}|\cos\theta/2|}{1 + (\cos\theta/|\cos\theta_0|)}$$
$$\times \exp\{-[(ikb\cos\theta)/\pi]\ln 2\},$$
$$\text{for } ka, kb, kc \ll 1. \quad (6.4c)$$

Through numerical computations it has been shown[17] that (6.4b) gives good accuracy for $(kb/\pi) < 1$ and $0 \le \theta_0 \le 60°$. We would expect the same accuracy from (6.4a) and (6.4c).

7. COMPARISON WITH YFK METHOD

The methodological difference between the YFK method and our ray method has been discussed in Sec. 1. Here we will make a quantitative comparison between their final solutions for a specific problem, namely, the bifurcated waveguide discussed in Sec. 3. Let us concentrate on the self-reflection coefficient of the TE_{l0} mode due to an incidence from the largest guide. The solution obtained by our ray method is given in (3.8) with $m = l$, which is also the exact solution. Under the assumption that

$$ka, kb, kc \gg 1, \quad (7.1)$$

we will compare it with the corresponding solution obtained by the YFK method for the following cases.

A. TE_{l0} Is Not Close to Cutoff

More precisely, we consider the case with

$$kd\cos\phi_{la} \gg 1, \quad \text{for } d = b, c. \quad (7.2)$$

Then we may use the formulas in (6.1b) and (6.3a) to expand $G_+^{(1)}(\alpha)$ appeared in (3.8), and the result is

$$A_l \sim \left(\frac{-2i\cos\tfrac{1}{2}\phi_{la}\cos\tfrac{1}{2}\phi_{la}}{\cos\phi_{la} + \cos\phi_{la}} \frac{(1 - e^{i2kb\sin\phi_{la}})(1 - e^{i2kb\sin\phi_{la}})}{2ka\cos\phi_{la}}\right)$$
$$\times \exp\left[\frac{e^{i\pi/4}}{\sqrt{2\pi}}\left(\frac{1}{\cos\phi_{la}} + \frac{1}{\cos\phi_{la}}\right)\right. \quad (7.3)$$
$$\left.\times \sum_{n=1}^{\infty}\left(\frac{e^{i2nkb}}{n\sqrt{2nkb}} + \frac{e^{i2nkc}}{n\sqrt{2nkc}} - \frac{e^{i2nka}}{n\sqrt{2nka}}\right) + O\left(\frac{1}{k}\right)\right],$$

where $O(1/k)$ means that terms of $(kd\cos\phi_{la})^{-1}$ with $d = b, c$ and higher have been ignored. If we further expand the exponential function in (7.3) and retain its leading term, we have

$$A_l \sim \left(\frac{-2i\cos\tfrac{1}{2}\phi_{la}\cos\tfrac{1}{2}\phi_{la}}{\cos\phi_{la} + \cos\phi_{la}} \frac{(1 - e^{i2kb\sin\phi_{la}})(1 - e^{i2kb\sin\phi_{la}})}{2ka\cos\phi_{la}}\right)$$
$$\times \{1 + M\}, \quad (7.4)$$

where M may be regarded as the contribution from the multiple scattering along the shadow boundary at $z = 0$ (Fig. 4) and is given by

$$M = \frac{e^{i\pi/4}}{\sqrt{2\pi}}\left(\frac{1}{\cos\phi_{la}} + \frac{1}{\cos\phi_{la}}\right)$$
$$\times \left[\sum_{n=1}^{\infty}\left(\frac{e^{i2nkb}}{n\sqrt{2nkb}} + \frac{e^{i2nkc}}{n\sqrt{2nkc}} - \frac{e^{i2nka}}{n\sqrt{2nka}}\right)\right] + O\left(\frac{1}{k}\right). \quad (7.5)$$

The corresponding formula obtained by Yee and Felsen[18] can be also expressed in the form of (7.4) except that M is different. Their M is given by[19]

$$M_{\mathrm{YFK}} = \frac{e^{i\pi/4}}{\sqrt{2\pi}}\left(\frac{1}{\cos\phi_{la}} + \frac{1}{\cos\phi_{la}}\right)\left(\frac{e^{i2kb}}{\sqrt{4\pi kb}} + \frac{e^{i2kc}}{\sqrt{4\pi kc}} + N\right), \quad (7.6)$$

where N is a doubly infinite summation, and is due to the contributions from the second and higher interactions along the shadow boundary at $z = 0$ (Fig. 4). Yee and Felsen[18] have shown numerically that N in most cases does not contribute significantly to M_{YFK}, and therefore its explicit form is not given in (7.6). Now let us compare (7.6) with (7.5). We note that they agree only for the two most dominant terms. A similar conclusion has been reached by Bowman[20] in the

comparison of the YFK solution and the exact solution for the problem of diffraction by an open-ended waveguide (Fig. 1 with $ka \to \infty$). At this point we should emphasize the fact that the result given in (7.4) and (7.5) is not the *rigorous* asymptotic expansion to the order of $(1/k)$ of the exact solution in (3.8). This is because the formula in (6.1b) itself is an approximate one, which was obtained in replacing $\gamma = \sqrt{\alpha^2 - k^2}$ by $\gamma \simeq -ik(1 - \alpha^2/2k^2)$ in $G^{(1)}(\alpha)$, and therefore the branch singularities at $\alpha = \pm k$ disappear. The absence of the branch singularities in $G^{(1)}(\alpha)$ accounts for the less accurate results in (7.4) and (7.5) when ka (or kb or kc) varies across the values of $q\pi$ for $q = 1, 2, 3, \ldots$. Therefore, the results given in (7.4) and (7.5) are not too accurate at the onset of a new propagating mode in any of the three guides in the bifurcated waveguide.

B. TE_{l0} Is Close to Cutoff

In such a case, it is convenient to introduce the parameter p_a through the definition

$$ka = \pi(l + p_a), \quad p_a \ll 1. \tag{7.7}$$

Then it follows that $\cos\phi_{la} = \sqrt{2\pi p_a/ka}$, and ϕ_{la} is close to $(\pi/2)$. The expansion of $G_+^{(1)}(k\cos\phi_{la})$ in (6.1b) will depend on the parameter

$$s = \sqrt{2kd}\cos\phi_{la} = 2\sqrt{p_a kd/l}, \quad \text{for } d = a, b, c;$$

when $s \gg 1$ we may use (6.3a), and when $s \ll 1$ we may use (6.3b) or (6.3c). When s assumes a fixed value, no simplified expression for $G_+^{(1)}(k\cos\phi_{la})$ is available and one has to use the exact one in (3.7). In the following let us concentrate on the most interesting case

$$p_a \ll 1, s \ll 1. \tag{7.8}$$

Making use of (6.3b) and (6.3c) in (6.16) gives the result

$$G_+(k\cos\phi_{la}) \sim e^{i(\pi/4)}(2\sqrt{2\pi p_a})^{-1}\exp[0.824(1-i) \times \sqrt{\pi p_a}]g(b)g(c), \tag{7.9}$$

where

$$g(d) = \sqrt{1 - e^{i2\pi kd}}\left(1 + \sqrt{\frac{p_a d}{p_d a}}\right)\exp\left[2\sqrt{\pi p_a}\sqrt{\frac{d}{a}} \times \left(e^{-i\pi/4}\sum_{n=1}^{\infty}\frac{e^{i2\pi kd}}{\sqrt{4\pi p_d}} - \frac{1}{\sqrt{4\pi p_d}}\right)\right],$$

$$p_d = \frac{kd}{\pi} - \left|\left[\frac{kd}{\pi}\right]\right|, \quad d = b, c.$$

Substitution of (7.9) into (3.8) leads to the following expression for the self-reflection coefficient for incident TE_{l0} mode:

$$A_l \sim (-1)\exp[-0.824(1-i)\sqrt{4\pi p_a}] \times [1 - e^{i2kb}\sin\phi_{la}/g(b)g(c)]^2,$$

which is valid under the conditions in (7.8) and (7.1). The YFK method does not seem able to give a satisfactory solution for this case.

8. CONCLUSION

In addition to the classical half-plane problem, there are only three types of problems in the edge diffraction theory that can be solved exactly. They are: (i) the open-end waveguide (Fig. 1 with $a = b$), (ii) the bifurcated waveguide (Fig. 4), and (iii) the periodic array of waveguides (Fig. 7). The known methods of attacking these problems are all based on the property of analytical functions in the complex plane (e.g., the Wiener–Hopf or the residue calculus techniques). They involve sophisticated mathematics and generally are not easy to apply. In the paper we have presented a recipe based on a ray-tracing procedure for attacking these problems. Following our recipe, we can write down the exact solutions almost *by inspection*. More important, in certain problems where the sophisticated analytical methods cannot be applied, our ray method often provides a useful approximate solution.

ACKNOWLEDGMENTS

The author appreciates many useful comments and encouragement from Professor G. A. Deschamps and Professor L. B. Felsen.

1 H. Y. Lee, L. B. Felsen, and J. B. Keller, SIAM J. Appl. Math. **16**, 268 (1968).
2 A more elaborate version than the one in (2.2) in order to be valid in the shadow boundary. It is also interesting to note that the ray amplitude so calculated decays as $(nkb)^{-1/2}$, instead of $(kb)^{-n/2}$ after the nth multiple scattering. This explains why the contribution from the rays in (ii) constitutes an integral part to the final result.
3 S. W. Lee, J. Math. Phys. **11**, 2830 (1970).
4 L. A. Weinstein, *Theory of Diffraction and the Factorization Methods* (Golem, Boulder, Colorado, 1969).
5 In the standard notation used in the Wiener-Hopf technique, [4, 6, 7] $G_+(-\alpha) = G_-(\alpha)$, which is the "minus factor" of $G(\alpha)$. Note that $G_+(\alpha)$ consists of all the spectrums for waves traveling in the $(+z)$ direction and $G_-(\alpha)$ for waves traveling in $(-z)$ direction [cf. the Fourier transform in (2.7)].
6 B. Noble, *Method Based on Weiner–Hopf Technique* (Macmillan, New York, 1958).
7 R. Mittra and S. W. Lee, *Analytical Methods in the Theory of Guided Waves* (Macmillan, New York, 1971).
8 This specularly reflected ray precisely compensates the discontinuity in the function $f(\theta)$ at $\theta = \pi/2$ because of the fact that $G_+(-k\cos\theta_0) = (1 - e^{i2kb\sin\theta})/[G_+(k\cos\theta_0)]$.
9 Implying the function $G_+(\alpha)$ is regular and free from zero in the upper half-plane of β defined by $\mathrm{Im}\beta > -\mathrm{Im}k$.
10 Precisely, this step is to factor out the part due to the contribution of the half-plane so that the remaining $G(\alpha)$ describe the interaction only. For simplicity, we call it "normalization."
11 This equation is identical to (3.7) in I. However, $(i\gamma_n)$ inside the product sign was misprinted as γ_n in (3.7) in I.
12 R. A. Hurd and H. Gruenberg, Can. J. Phys. **32**, 694 (1954).
13 J. F. Carlson and A. E. Heins, Quart. J. Appl. Math. **4**, 313 (1947).
14 F. Berz, Proc. IEE (London) **98**, Part II, 47 (1951).
15 E. A. M. Whitehead, Proc. IEE (London) **98**, Part III, 133 (1951).
16 The approximation comes in from the replacement of $\gamma = ik\sqrt{1 - (\alpha/k)^2}$ by $\gamma \simeq ik(1 - \alpha^2/2k^2)$ before the factorization formula in (2.4) is applied.
17 S. W. Lee and T. T. Fong, J. Appl. Phys. **43**, (1972) (to be published).
18 H. Y. Yee and L. B. Felsen, IEEE Trans. Microwave Theory Tech. **17**, 73 (1969).
19 It seems that Eq. (5.5) in Ref. 18 missed a multiplication factor (i). The result given in (7.6) above is based on (5.5) with the factor (i) added.
20 J. J. Bowman, SIAM J. Appl. Math. **18**, 818 (1970).

On the Structure of Relativity Groups

Luigi A. Lugiato*
Istituto di Scienze Fisiche dell'Università, Milano, Italy

and

Vittorio Gorini†‡
Institut für Theoretische Physik (I) der Universität, Marburg, Germany

and

Center for Particle Theory, The University of Texas, Austin, Texas 78712§
(Received 24 November 1971)

In this paper an attempt is made to define the class of equivalent reference frames in special relativity in terms of a purely kinematical characterization of the notion of free material point and using homogeneity of space and time, isotropy of space, and unidirectional time flow. The problem of finding the possible forms of the relativity group and the significance of a "space–time homogeneity axiom" which was used in previous papers are discussed.

I. INTRODUCTION

In some recent papers[1] the problem of the derivation of the Poincaré and Galilei groups from the basic principles of (i) relativity, (ii) homogeneity of space and time, and (iii) isotropy of space has been investigated extensively. In a sketchy way, the line of the argument is the following. First, by a suitable "formulation of the principle of space–time homogeneity"[1(a),1(e),2] one ensures that the space–time coordinate transformation between two equivalent frames of reference is affine. Secondly, one shows that the orthochronous Poincaré transformations, with the Galilei ones as the usual limiting case, are the only ones which are compatible with the principle of relativity, the isotropy of space, the synchronization of clocks in each frame at different space points, the existence of a unidirectional time flow, and the choice in all frames of common length and time standards.

Yet in the abovementioned papers, an operationally satisfactory, at least in principle, characterization of the class of equivalent frames is lacking. The usual characterization of the equivalent frames in special relativity is intrinsically a dynamical one, since it is based on the law of inertia, which is a statement about the properties of the motion of the center of mass of a free material body, being defined as one which is not acted upon by external forces.[3] Apart from the difficulties in principle which such a definition of a free body may encounter, we believe that in the search of the relativity groups one should try to rely on kinematical elements and on some generally recognized uniformity properties of space and time relative to the observer, rather than on dynamical (though very simple) laws, like the law of inertia. It is only after the possible relativity groups have been found following this line, that the request of invariance of the (known) dynamical laws should enable one to choose the correct transformation group among the various possibilities. In other words, the knowledge of the dynamics should enter only at the stage of selecting the correct group among the different solutions which have been found on a kinematical basis and on the ground of some minimal invariance properties which are generally accepted.

In this paper we stick to this line of approach by defining a class of frames (which we call *pre-inertial*) in terms of a *purely kinematical characterization of the notion of a free material point*, based on simple conceptual experiments of a type often used in the theory of relativity. At the same time, the principle of homogeneity of space and time will be formulated by stressing its significance as regards the space and time of the given observer, in contrast with the absolute nature of the space–time homogeneity axiom of Refs. 1(a) and 1(e) (to be labeled T from now on), which appears therefore to be an essentially independent and much stronger condition both from the physical as well as from the mathematical point of view.

Without the latter axiom, the problem of the linearity of the transformations and, consequently, of the Poincaré and the Galilei groups being the only solutions, is still open and requires further investigation.

In Sec. II we define the class of pre-inertial frames of reference in terms of our kinematical characterization of a free material point and requiring them to satisfy the conditions of space and time homogeneity, isotropy of space and unidirectional flow of time, and we show that this enables us to obtain a complete characterization of the transformations among frames which are relatively at rest.

In Sec. III we tentatively formulate a relativity principle for the class of pre-inertial frames and we state and briefly comment upon the (apparently difficult) mathematical problem that one has to solve in order to find the possible forms of the corresponding relativity group. Then we give a very concise but complete review of the way the Poincaré and Galilei groups can be univocally deduced if use is made of axiom T as an additional requirement. Finally, we briefly discuss the relation of axiom T to the usual characterization of a free material point and to the property of homogeneity of space and time.

In the following, greek indices $\mu, \nu, \cdots$ run from one to four, latin indices $i, j, \cdots$ from one to three. A linear transformation $x \to Mx$ (respectively, an affine transformation $x \to Mx + a$) of R^n will currently be identified to the corresponding matrix $M \in GL(n, R)$ [respectively, to the corresponding vector-matrix pair (a, M), $a \in R^n$, $M \in GL(n, R)$]. A matrix $G \in GL(n, R)$ will often be denoted by the family of its matrix elements $G = (G_{\alpha\beta})$ and an n-vector x by its components (x_α). The usual notation $\vec{x}$ is used for a 3-vector (x_i). We call a one-to-one map of a set onto itself a permutation of the set.

II. THE CLASS OF PRE-INERTIAL FRAMES REFERENCE

First of all, let us specify what we mean by the statement that the object A is a copy of the object B. An observer will say that A is a copy of B if, after reducing to rest both A and B in a properly smooth way, and putting A near B, after a series of comparisons he will judge that A and B are identical. In particular,

from a purely geometrical point of view, he will say that the two objects are congruent. We shall assume that the conclusion that A is (or is not) a copy of B is independent of the procedure by which A and B are put nearby, and of the observer. We suppose each observer to distinguish naturally between space and time and assume that he chooses a reference frame by which he is able to characterize each event by four real numbers, three space coordinates $x_i (i = 1, 2, 3)$, and a time coordinate t. Thus a reference frame (for a given observer) will consist essentially of a set of measuring rods or equivalent devices for the assignment of the location of events in space and of a set of clocks, each of which is a copy of the others, located at all space points, for the assignment of the times at which events happen. For definiteness, we shall assume without loss of generality that the coordinates are chosen in such a way that the domain of coordinate sets of all possible events is R^4 and that two distinct events have distinct coordinate sets.

We define a frame S' to be at rest relative to a given frame S if any object which is at rest in S is at rest in S' as well. This is equivalent to the statement that the coordinate transformation functions from S to $S'(x_4 = t, x_4' = t')$,

$$x_\mu' = f_\mu(x_1, x_2, x_3, x_4), \quad \mu = 1, 2, 3, 4, \tag{1}$$

must map a straight line which is parallel to the time axis onto a straight line which is parallel to the time axis and therefore are of the form

$$\begin{aligned} &\text{(a) } x_i' = f_i(x_1, x_2, x_3), \quad i = 1, 2, 3, \\ &\text{(b) } x_4' = f_4(x_1, x_2, x_3, x_4). \end{aligned} \tag{2}$$

Let

$$x_\mu = \phi_\mu(x_1', x_2', x_3', x_4'), \quad \mu = 1, 2, 3, 4, \tag{3}$$

be the inverse functions of (1) and suppose that there exists a 3-vector $\hat{\vec{x}} = (\hat{x}_1, \hat{x}_2, \hat{x}_3)$ such that $\vec{x}^{(1)} = \vec{\phi}(\hat{x}_1, \hat{x}_2, \hat{x}_3, x_4^{(1)}) \neq \vec{x}^{(2)} = \vec{\phi}(\hat{x}_1, \hat{x}_2, \hat{x}_3, x_4^{(2)})$ for some $x_4^{(1)}$ and $x_4^{(2)}$. Then transformation (1) would map the straight lines $\vec{x} = \vec{x}^{(1)}$ and $\vec{x} = \vec{x}^{(2)}$ onto the same straight line $\vec{x} = \hat{\vec{x}}$, which contradicts the fact that (1) is one to one. Hence the functions ϕ_i depend on x_1', x_2', and x_3' but not on x_4', so that S is at rest relative to S' and we can simply speak of "frames which are relatively at rest." Using similar arguments and observing that f is a permutation of R^4, one easily proves that $\vec{f}$ is a permutation of R^3 and that $\vec{\phi} = \overrightarrow{f^{-1}} = (\vec{f})^{-1}$.

We now postulate on experimental grounds the existence of a nonvoid class $\mathcal{O} = \{O\}$ of observers, with the class $\mathcal{J} = \{J\}$ of their corresponding frames $J \leftrightarrow O = O_J$, which we call *pre-inertial frames of reference*, such that

(A) with respect to each observer of the class $\mathcal{O}$, space appears to be Euclidean and the localization of events in space is given in terms of orthogonal, say, right-handed triads and

(B) the clocks of each frame $J \in \mathcal{J}$ can be synchronized in such a way that the following properties hold.

I: Let $J \in \mathcal{J}$ and let O_J be the observer which sets up J. Then O_J has copies of an apparatus (this might be thought for instance of being an ideal gun) which, being at rest in J, is able to start a material point such that

(a) the trajectory of the point is a straight line;

(b) if a second copy of the apparatus starts a second material point from the position in which the first point is at a certain instant of its motion, at the same instant and in the same direction, afterwards the two points proceed together;

(c) the law of motion $t \to \vec{x}(t)$ of the point is a continuous one-to-one function of t.[4]

II(a): If $J \in \mathcal{J}$, space and time are homogeneous according to O_J, meaning that

(i) for any given apparatus A_0 at rest in J and geometrically characterized by the set of spatial coordinates $\{\vec{x}\}$ and for any given 3-vector $\vec{a}$, O_J can in principle build a copy $A_{\vec{a}}$ of A_0, which is geometrically characterized by the set of coordinates $\{\vec{x} + \vec{a}\}$;

(ii) for any given experiment E_0 performed with A_0 and geometrically characterized by the set of space-time coordinates $\{\vec{x}, t\}$ and for any given 4-vector $\{\vec{a}, t_0\}$, O_J can in principle perform with $A_{\vec{a}}$ a corresponding copy experiment $E_{\vec{a},t_0}$ geometrically characterized by the set of coordinates $\{\vec{x} + \vec{a}, t + t_0\}$.

II(b): If $J \in \mathcal{J}$, space is isotropic according to O_J, meaning that

(i) for any given apparatus A_0 at rest in J and geometrically characterized by the set of spatial coordinates $\{\vec{x}\}$ and for any given proper rotation Q, O_J can in principle build a copy A_Q of A_0, which is geometrically characterized by the set of coordinates $\{Q\vec{x}\}$;

(ii) for any given E_0 performed with A_0 and geometrically characterized by the set of space-time coordinates $\{\vec{x}, t\}$, O_J can in principle perform with A_Q a corresponding copy experiment E_Q geometrically characterized by the set of coordinates $\{Q\vec{x}, t\}$.

III: Let J and J' be two frames of $\mathcal{J}$ which are supposed to be relatively at rest so that the transformation from J to J' has the form (2). Then, if $\vec{x}$ is any 3-vector and t_1 and t_2 any two distinct time instants, we assume that

$$\frac{f_4(\vec{x}, t_2) - f_4(\vec{x}, t_1)}{t_2 - t_1} > 0. \tag{4}$$

We shall now make several comments on points I–III. Point I gives the kinematical definition of a "free point." Indeed, with the help of II(a), one can show that

the motion of such points is uniform. In order to prove this, let us consider a "free point" fired at $t = 0$ and a second point fired at $t = t_1 \geq 0$, in the conditions described in I(b) relative to the first point. If $t \to \vec{x}(t)$, $t \geq 0$, and $t \to \vec{x}'(t)$, $t \geq t_1$, are the respective laws of motion of the two points, we have, by II(a),

$$\vec{x}'(t + t_1) = \vec{x}(t) + \vec{x}(t_1) - \vec{x}(0), \quad t_1 \geq t \geq 0. \tag{5}$$

If $t_2 \geq t_1$, by first setting $t = 0$ in (5) and then $t = t_2 - t_1$, we get

$$\vec{x}'(t_2) - \vec{x}'(t_1) = \vec{x}(t_2 - t_1) - \vec{x}(0), \quad t_2 \geq t_1 \geq 0. \tag{6}$$

On the other hand, by I(b),

$$\vec{x}'(t_2) - \vec{x}'(t_1) = \vec{x}(t_2) - \vec{x}(t_1), \quad t_2 \geq t_1 \geq 0. \tag{7}$$

Combining (6) and (7) we obtain

$$\vec{x}(t_2) - \vec{x}(t_1) = \vec{x}(t_2 - t_1) - \vec{x}(0), \quad t_2 \geq t_1 \geq 0. \tag{8}$$

The function $\vec{x}(t)$ has been defined for $t \geq 0$. However, it can be continued also for $-T \leq t < 0$, with the aid of a third material point fired at $t = -T$ in the direction of $\vec{x}(t_2) - \vec{x}(0)$, $t_2 > 0$, from a position such that at $t = 0$ it is in $\vec{x}(0)$. Since T is arbitrary, we can define a fictitious motion $\vec{x}(t)$ with $t \in (-\infty, +\infty)$, satisfying (8). Defining $\vec{g}(t) = \vec{x}(t) - \vec{x}(0)$, (8) gives

$$\vec{g}(t_2) - \vec{g}(t_1) = \vec{g}(t_2 - t_1), \quad t_2 \geq t_1 \geq 0. \tag{9}$$

On the other hand, by I(b) and II(a), $\vec{g}(T) = \vec{x}(T) - \vec{x}(0) = \vec{x}(0) - \vec{x}(-T) = -\vec{g}(-T)$, namely

$$\vec{g}(-t) = -\vec{g}(t), \quad t \in (-\infty, +\infty). \tag{10}$$

Combining (9) and (10) we get

$$\vec{g}(t) + \vec{g}(t') = \vec{g}(t + t'), \quad t, t' \in (-\infty, +\infty). \tag{11}$$

From (11) and I(c) one deduces in a standard way[5] that $\vec{g}(t)$ is a linear function of t, whereby

$$\vec{x}(t) = \vec{v}t + \vec{x}(0), \quad \vec{v} \neq 0. \tag{12}$$

One recognizes immediately that, due to II, the absolute value of the velocity $\vec{v}$ does not depend on its direction nor on the point in space and on the instant in time at which the "free point" was fired.

Now let P_1 and P_2 be two distinct points in space and let two "free points" be fired respectively at P_1 in the direction of P_2 at the time $t = 0$ as given by the clock in P_1 and at P_2 in the direction of P_1 at the time $t = 0$ as given by clock at P_2. The corresponding laws of motion are

$$\vec{x}(t) = P_1 + \alpha(P_2 - P_1)t \tag{13}$$

and

$$\vec{x}'(t) = P_1 + (1 - \alpha t)(P_2 - P_1). \tag{14}$$

Equating (13) to (14) gives

$$\vec{x}\left(\frac{1}{2\alpha}\right) = \vec{x}'\left(\frac{1}{2\alpha}\right) = \frac{P_2 + P_1}{2}, \tag{15}$$

showing that the two bullets collide at the midpoint of the straight line segment joining P_1 to P_2. We thus see that the synchronization of the clocks in a given frame, which is implied by conditions I and II is the standard one.[6]

We stress the fact that conditions I-III do not imply in general that the observer O_J, should see the "free points" fired by O_J as uniformly moving along straight lines. Indeed, the guns at rest in J are not in general at rest in J'.

Points II(ai) and II(bi), together with the definition of a copy, imply that objects do not deform when they are moved in space. This is in fact a translation of Assumption (a) that space appears to each observer to be Euclidean.

It follows from II(a) and II(b) that for every $J \in \mathcal{J}$ and for an arbitrary 4-vector $a = (\vec{a}, t_0)$ and an arbitrary proper rotation Q, one can build a space rotated, space–time translated frame $J_{(a,Q)}$ such that

$$\begin{aligned} \vec{x}_{(a,Q)} &= Q\vec{x} + \vec{a}, \\ t_{(a,Q)} &= t + t_0. \end{aligned} \tag{16}$$

Since $J_{(a,Q)}$ and J are relatively at rest, their corresponding observers can use the same guns and then it is immediately seen that $J_{(a,Q)}$ also belongs to $\mathcal{J}$. Actually, the family $\{J_{(a,Q)}\}_{a \in R^4, Q \in SO(3)}$ exhausts the set of frames of $\mathcal{J}$ which are at rest relative to J. In order to prove this, let J' be any frame of $\mathcal{J}$ which is at rest relative to J. Guns at rest in J are at rest in J' as well, which means that a material point which is free in J is seen to be free also in J'. Hence, since any straight line in space can be the trajectory of a free point due to II, the space part

$$x_i' = f_i(x_1, x_2, x_3), \quad i = 1, 2, 3, \tag{17}$$

of transformation (2) must map a straight line onto a straight line, its inverse having the same property. Then, using the fact that the only (algebraic!) automorphism of the real field is the identity (Appendix A), an almost straightforward application of the fundamental theorem of projective geometry[7] (Appendix B) shows that (17) is an affine permutation of R^3:

$$x_i' = \sum_{l=1}^{3} G_{il} x_l + a_i, \quad i = 1, 2, 3. \tag{18}$$

Now let A and B be two copies of an object, at rest in J and J'. Due to II(ai) and II(bi) they appear both to J and to J' as rotated and translated with respect to each other. In other words, transformation (18) must map figures which are relatively rotated and translated onto figures which are relatively rotated and translated. Therefore, by a trivial generalization of Lemma 2 of Ref. 1(d), we conclude that (18) must be of the form

$$x_i' = \sum_{j=1}^{3} Q_{ij}(kx_j) + a_i, \quad i = 1, 2, 3, \tag{19}$$

where $k > 0$ and $\{Q_{ij}\} \in SO(3)$ (due to the choice of right-handed triads). Clearly, if O_J and $O_{J'}$ both choose the same rule as unit of length, k has to be taken equal to one.

As to relation (2b), let P_1 and P_2 be any two distinct space points and let two material points be fired from P_1 to P_2 and, respectively, from P_2 to P_1, simultane-

ously at time t with respect to the clocks of J. Due to (15), the two bullets as seen by O_J will collide at the midpoint $\vec{x}_M = \frac{1}{2}(\vec{x}_{P_1} + \vec{x}_{P_2})$ of the straight line segment joining $\vec{x}_{P_1}$ to $\vec{x}_{P_2}$. But since

$$(\vec{x}_M)' = Q\vec{x}_M + \vec{a} = \tfrac{1}{2}\{Q\vec{x}_{P_1} + Q\vec{x}_{P_2} + \vec{a}\} = \tfrac{1}{2}(\vec{x}'_{P_1} + \vec{x}'_{P_2}) = \vec{x}'_M,$$

the collision will take place at the midpoint, as seen by $O_{J'}$ as well, who therefore will also judge that the two bullets were fired simultaneously. Therefore $f_4(\vec{x}_{P_1}, t) = t'_{P_1} = t'_{P_2} = f_4(x_{P_2}, t)$, which, by the arbitrariness of P_1, P_2, and t, implies that

$$t' = f_4(\vec{x}, t) = f_4(t). \tag{20}$$

Now let $t \to \vec{x}(t) = \vec{a} + t\vec{b}$ be the law of motion of a "free point" relative to J. We have $\vec{x}'(t') = Q\vec{x}(t) + \vec{c} = (Q\vec{a} + \vec{c}) + t(Q\vec{b}) = \vec{a}' + f_4^{-1}(t')(Q\vec{b})$ and due to the uniformity of the motion relative to J', $f_4^{-1}(t') = \alpha t' + \beta$, $\alpha \neq 0$. Then, if O_J and $O_{J'}$ choose the same time standard, α has to be chosen equal to one because of (4) and we have $t' = t + t_0$.

III. THE PROBLEM OF DERIVING THE RELATIVITY GROUP

Among the properties which define the class $\mathcal{J}$ of pre-inertial reference frames, II is clearly a statement of invariance of the description of physical phenomena within a maximal subclass of frames of $\mathcal{J}$ which are relatively at rest. We shall now tentatively formulate a *relativity principle* for the whole class $\mathcal{J}$ by assuming:

Axiom P: the frames of $\mathcal{J}$ are equivalent as regards the description of natural phenomena.

Generally speaking, this has to be intended as meaning that if two copies of a physical system have been given subjectively identical initial conditions in the respective laboratories of two frames of $\mathcal{J}$, the probabilities for equal outcomes of corresponding subjectively identical experiments performed at equal later times on the copies in the respective frames are the same.[8]

A consequence of Axiom P is that the set $\mathfrak{L}$ of transformations (1) relative to pairs of elements of $\mathcal{J}$ is a subgroup of the group $\mathcal{P}$ of permutations of R^4. Indeed, it is first of all evident that $\mathfrak{L}$ contains the identity map and that $f \in \mathfrak{L}$ implies $f^{-1} \in \mathfrak{L}$. Now, let g, $h \in \mathfrak{L}$ and suppose that, with obvious notations, $J' = g(J)$ and $\bar{J}' = h(\bar{J})$, with $J, J', \bar{J}, \bar{J}' \in \mathcal{J}$. Due to P, there exists in $\mathcal{J}$ a frame J'' which bears the same relation to J' as $\bar{J}'$ to $\bar{J}$: $J'' = h(J')$. Therefore, $J'' = h(g(J)) = hg(J)$ and the composite transformation also belongs to $\mathfrak{L}$.

$\mathfrak{L}$ might be termed the *space–time relativity group*.

We now define $\mathcal{K}$ to be the subset of $\mathcal{P}$ whose elements are the permutations of R^4 having the form (2). In other words,

$$\mathcal{K} = \{f \mid f \in \mathcal{P}; \text{f maps a straight line which is parallel to the } x_4 \text{ axis onto a straight line which is also parallel to the } x_4 \text{ axis}\}. \tag{21}$$

Using the same argument as the one following formula (2), one proves that $\mathcal{K}$ is a subgroup of $\mathcal{P}$. If $g \in \mathfrak{L} \cap \mathcal{K}$, it is a transformation connecting two frames of $\mathcal{J}$ which are relatively at rest and vice versa. On the other hand, we have seen in Sec. II that the group of transformations relative to pairs of frames belonging to a maximal subclass of frames of $\mathcal{J}$ which are relatively at rest is the semidirect product $\mathcal{R} = R^4 \otimes \mathcal{C}$ of the transformations of the form (16), where $a = (\vec{a}, t_0) \in R^4$ and

$$\mathcal{C} = \{C \mid C \in GL(4,R); C_{i4} = C_{4i} = 0, i = 1,2,3; C_{44} = 1; (C_{ij}) \in SO(3)\}. \tag{22}$$

Therefore, $\mathfrak{L}$ must satisfy

$$\mathfrak{L} \cap \mathcal{K} = \mathcal{R} \tag{23}$$

and the problem of finding the possible forms of the relativity group reduces to that of determining the family $\mathfrak{F} = \{\mathfrak{L}\}$ of all solutions of Eq. (23).

The task of finding $\mathfrak{F}$ seems to be a rather difficult one. We observe, however, that a particular subfamily of $\mathfrak{F}$ is $\{\mathcal{P}_\lambda\}_{\lambda \in [0,+\infty]}$, where

(i) if $0 < \lambda < +\infty$, $\mathcal{P}_\lambda$ is the proper orthochronous Poincaré group corresponding to an invariant velocity equal to $\lambda^{-1/2}$, namely

$$\mathcal{P}_\lambda = \{(a,L) \mid (a,L) \in \mathcal{P}; (a,L)x = Lx + a; a \in R^4; L \in GL(4,R); L^T g(\lambda) L = g(\lambda); g(\lambda) = \mathrm{diag}(-\lambda, -\lambda, -\lambda, 1); \det L = 1, L_{44} \geq 1\}, \tag{24}$$

(ii) $\mathcal{P}_0$ is the proper orthochronous inhomogeneous Galilei group, namely

$$\mathcal{P}_0 = \{(a,G) \mid (a,G) \in \mathcal{P}; (a,G)x = Gx + a; a \in R^4; G \in GL(4,R); G_{4i} = 0, i = 1,2,3; G_{44} = 1; (G_{ik}) \in SO(3)\}, \tag{25}$$

and

(iii) $$\mathcal{P}_{+\infty} = \mathcal{R}. \tag{26}$$

The following question is therefore at order. Is $\{\mathcal{P}_\lambda\}$ a proper subfamily of $\mathfrak{F}$? In other words, are there solutions of Eq. (23) which do not belong to the family $\{\mathcal{P}_\lambda\}$? Our feeling is that it might well be that the answer to this question is negative. On the other hand, if it will eventually turn out that other (perhaps not too pathological) solutions exist, it would be interesting to look whether they can be given a physical interpretation. Alternatively, one might look in this case for a further axiom to be imposed on the class $\mathcal{J}$, which should provide a necessary condition to rule out just these extra solutions, and try to interpret this axiom physically. Since we do not have at hand the solution of the problem, we shall confine ourselves to a review of the steps of the proof [which can be found in detail in Refs. 1(a) and 1(d)] that the "space–time homogeneity axiom" of Refs. 1(a) and 2 provides a sufficient condition for the family of possible relativity groups to be exactly $\{\mathcal{P}_\lambda\}$. The axiom can be stated as follows:

Axiom T: Let $J, J' \in \mathcal{J}$ and let $x \to x' = f(x)$ be the coordinate transformation from J to J'. If b is an

arbitrary 4-vector, let $J_{(b)}$ denote the frame space–time translated by b with respect to J:

$$x^{(b)} = x - b. \tag{27}$$

Then $J_{(b)}$ is seen also by J' as being space–time translated with respect to J, namely a 4-vector b' exists such that

$$f(x^{(b)}) = f(x) - b'. \tag{28}$$

We observe that for any $J \in \mathcal{J}$, the frame $J_{(a,Q)}$ defined by (16) satisfies T.

Using Axiom T and assuming f to be continuous, one can prove in a standard way[9] that f has to be an affine permutation of R^4:

$$f: x \to Mx + a. \tag{29}$$

Thus $\mathcal{L}$ is a subgroup of the affine group $R^4 \otimes GL(4,R)$ of R^4 and has the semidirect product structure $\mathcal{L} = R^4 \otimes \mathfrak{M}$, where $GL(4,R) \supseteq \mathfrak{M} \supseteq \mathcal{C}$ and $\mathfrak{M}$ is the group of transformations among the elements of a maximal class of frames of $\mathcal{J}$ having the same space–time origin. Within one such class, the group of transformations among frames which are relatively at rest is $\mathcal{C}$, so that Axiom T allows us to replace Eq. (23) by the equation

$$\mathfrak{M} \cap \mathfrak{K} = \mathcal{C}, \tag{30}$$

where $\mathfrak{K}$ is the subgroup of $GL(4,R)$ whose elements are the transformations which map any straight line parallel to the x_4 axis onto a straight line again parallel to the x_4 axis, namely

$$\mathfrak{K} = \{H \mid H \in GL(4,R); H_{i4} = 0, i = 1,2,3\}. \tag{31}$$

The entire family of solutions of Eq. (30) has been determined in Ref. 1(d) [see also Ref. 1(c)] and is given by the following:

Theorem: Let Φ denote the family of subgroups of $GL(4,R)$ defined by $\mathfrak{M} \in \Phi$ iff $\mathfrak{M} \cap \mathfrak{K} = \mathcal{C}$. Then $\Phi = \{\mathcal{G}_\lambda\}_{\lambda \in [0,+\infty]}$, where *a*) if $0 < \lambda < +\infty$, $\mathcal{G}_\lambda = \{L \mid L \in GL(4,R), (0,L) \in \mathcal{P}_\lambda\}$, *b*) $\mathcal{G}_0 = \{G \mid G \in GL(4,R); (0,G) \in \mathcal{P}_0\}$ and $\mathcal{G}_{+\infty} = \mathcal{C}$.

Since $\mathcal{P}_\lambda = R^4 \otimes \mathcal{G}_\lambda$, $\lambda \in [0,+\infty]$, this completes the proof that the addition of axiom T to conditions I–III of Sec. II is sufficient to restrict the choice of the possible relativity groups to the elements of the family $\{\mathcal{P}_\lambda\}$.

We conclude with some observations regarding the significance of Axiom T and its relation to the properties of free material points and to the homogeneity of space and time. Axiom T is essentially a statement that space–time translations have an absolute character for the frames of the class $\mathcal{J}$, and this can be seen to be equivalent to the statement that the notion of a free point is likewise absolute for these frames. Indeed, Axiom T implies that the elements of $\mathcal{L}$ are affine transformations, so that for any $J, J' \in \mathcal{J}$, observer $O_{J'}$ sees the "free points" fired by O_J as uniformly moving along straight lines. Conversely, suppose that the latter circumstance is true, and assume that the domain Δ of the possible velocity vectors $\vec{v}$ of free points in one frame of $\mathcal{J}$ contains a spherical neighborhood N of the origin (of course, free points whose velocities have different absolute values are fired by different guns). Then, the world lines of free material points going through a given space–time point x with respect to a given frame $J \in \mathcal{L}$ are straight lines whose union contains the inner part of a circular cone $C_x(N)$ with vertex in x and with axis the "rest" world line through x, and whose opening depends on N but can be taken to be x independent. Any straight line contained in one such cone is mapped by an element of $\mathcal{L}$ onto a straight line. Then it is easy to show that if $f \in \mathcal{L}$ and π is a plane such that one (and thus all) of its points x is the intersection of two straight lines lying in π and contained in $C_x(N)$, $f(\pi)$ is a plane.[10] Therefore, since any straight line can be obtained as the intersection of two such planes π and π', f (together with its inverse) maps a straight line onto a straight line; hence by the fundamental theorem of projective geometry (Appendix B),[7] it is an affine permutation of R^4. Axiom T is thus satisfied by the frames of $\mathcal{L}$.

The absolute character of translations and of the notion of free point was already guaranteed in the class of frames at rest relative to each other. Axiom T extends this property to the whole class $\mathcal{J}$. In a sense, it replaces the law of inertia.

As regards the relation of Axiom T to space–time homogeneity, it is clear that the former is not a statement of the homogeneity of space and time relative to a particular observer, but rather a postulate on the absolute character that space–time translations should have for the class of equivalent observers. It appears therefore to be an essentially independent and much stronger condition than the simple requirement of homogeneity of space and time relative to a given observer, as described in condition II(a).

ACKNOWLEDGMENTS

We are indebted to Professor V. Berzi for most illuminating discussions. Part of this work was done while one of the authors (V. G.) was visiting the Institut für Theoretische Physik (1) der Universität Marburg through an invitation of Professor H. D. Doebner which is herewith gratefully acknowledged. The same author is also indebted to Professor E. C. G. Sudarshan for some remarks and for inviting him to spend a period at the Center for Particle Theory of the University of Texas at Austin.

APPENDIX A

Let f be an automorphism of the field R of real numbers, namely f is a permutation of R satisfying

$$f(\alpha + \beta) = f(\alpha) + f(\beta) \tag{A1}$$

and

$$f(\alpha\beta) = f(\alpha) f(\beta), \tag{A2}$$

$\forall \alpha, \beta \in R$.

Taking $x = y = 0$ in (A1) we get

$$f(0) = 0 \tag{A3}$$

from which, setting $\beta = -\alpha$ in the same equation,

$$f(-\alpha) = -f(\alpha). \tag{A4}$$

If Q is the subfield of rational numbers, from (A1) we get in a standard way using induction[5] that

$$f(r\alpha) = rf(\alpha), \quad \forall \alpha \in R \text{ and } \forall r \in Q. \tag{A5}$$

Taking $\beta = 1$, (A2) gives $f(\alpha) = f(\alpha)f(1)$ and by (A3) and since f is a permutation, $f(1) = 1$ whence, setting $\alpha = 1$ in (A5),

$$f(r) = r, \quad \forall r \in Q. \tag{A6}$$

If $\alpha \geq \beta$ we have, using (A4), (A1), and (A2), $f(\alpha) - f(\beta) = f(\alpha - \beta) = f[(\alpha-\beta)^{1/2}(\alpha-\beta)^{1/2}] = \{f[(\alpha-\beta)^{1/2}]\}^2 \geq 0$,

$$\text{namely } \alpha > \beta \Rightarrow f(\alpha) > (\beta). \tag{A7}$$

Let $\alpha \in R$, $\alpha \in Q$, and suppose that $f(\alpha) > \alpha$. Then there exists $r \in Q$ such that $f(\alpha) > r > \alpha$. From $r > \alpha$, (A6), and (A7), we get $r > f(\alpha)$ which is a contradiction. Similarly, $f(\alpha) < \alpha$ also leads to a contradiction. Thus, combining with (A6), we get $f(\alpha) = \alpha, \forall \alpha \in R$.

APPENDIX B

Let V be a finite-dimensional vector space over a commutative field K.[11] We write $V^* = V - \{0\}$ and $K^* = K - \{0\}$, where 0 and **0** are the zeros (origins) of V and of K, respectively. The relation on V^*,

$$xRy \Leftrightarrow x = \lambda y \quad \text{for some } \lambda \in K^*, \tag{B1}$$

is an equivalence relation and the quotient set $P(V) = V^*/R$ is called the *projective space* associated to V. As a subset of V, an equivalence class is the difference set $\{\lambda a\}_{\lambda \in K^*}$, $a \in V^*$, of a straight line through the origin and the origin itself. As a point of $P(V)$, the equivalence class $\{\lambda a\}_{\lambda \in K^*}$ will be denoted by P_a. Three distinct points p_a, p_b, and p_c of $P(V)$ are said to be *collinear* if there exist λ, $\mu \in K^*$ such that $c = \lambda a + \mu b$. If V and V' are two vector spaces over the fields K and K', respectively, and I is an isomorphism of K onto K', a map ϕ of V into V' is said to be *semilinear* (relative to I) if

$$\begin{aligned} &\phi(\alpha x + \beta y) = I(\alpha)\phi(x) + I(\beta)\phi(y); \\ &\alpha, \beta \in K \text{ and } x, y \in V. \end{aligned} \tag{B2}$$

Theorem (the fundamental theorem of projective geometry[7,12]): Let V and V' be two finite-dimensional vector spaces of equal dimension $n \geq 3$ over the commutative fields K and K', respectively, and let $P(V)$ and $P(V')$ be the corresponding projective spaces. Let σ be a one-to-one map of $P(V)$ onto $P(V')$ such that, whenever $p_a, p_b, p_c \in P(V)$ are collinear, $\sigma(p_a)$, $\sigma(p_b), \sigma(p_c)$ are collinear. Then there exists a unique isomorphism I of K onto K' and a one-to-one semilinear map ϕ of V onto V' (relative to I) such that ϕ induces σ on $P(V)$ in the sense that $\sigma(p_a) = \{\phi(\lambda a)\}_{\lambda \in K^*}$, $a \in V^*$. If ψ is another similinear map of V onto V' having the same property, then $\psi(x) = \phi(\mu x)$, where μ is a given element of K^*, and by letting μ take all values in K^* we obtain all of these maps.

Now let f be a permutation of R^n ($n \geq 3$) such that f and f^{-1} map a straight line onto a straight line. Let π be a plane and let l_1 and l_2 be two distinct intersecting straight lines belonging to π. By hypothesis, $f(l_1)$ and $f(l_2)$ are two distinct intersecting straight lines which define a plane π'. The image $f(P)$ of a point P of π which belongs to $l_1 \cup l_2$ lies in π' by definition. Let a point Q belong to π and $Q \notin l_1 \cup l_2$. Draw a straight line l_Q through Q and intersecting both l_1 and l_2. Then, $f(l_Q)$ is a straight line intersecting $f(l_1)$ and $f(l_2)$ hence lying in π'. Therefore $f(Q) \in \pi'$ and $f(\pi) \subseteq \pi'$. Since f^{-1} also maps a straight line onto a straight line by hypothesis, we get similarly that $f(\pi) \supseteq \pi'$ so that $f(\pi) = \pi'$ and f maps a plane onto a plane. Then, if $g: R^n \to R^n$ is defined as

$$g(x) = f(x) - f(0), \quad x \in R^n, \tag{B3}$$

it induces a permutation σ of $P(R^n)$ which maps triplets of collinear points onto triplets of collinear points. Therefore, since R has only the identity automorphism (Appendix A), by the fundamental theorem of projective geometry, there exists a nonsingular linear transformation M of R^n onto itself inducing σ. The map

$$\hat{f} = M^{-1} \circ g \tag{B4}$$

is a permutation of R^n which maps a straight line onto a straight line and which maps a straight line through the origin onto itself. We prove that

$$\hat{f}: x \to \lambda x \tag{B5}$$

for some $\lambda \in R^*$. Indeed, let r be a straight line which does not go through the origin and let s_1 and s_2 be two distinct straight lines through the origin both of which intersect r (s_1 in P_1 and s_2 in P_2, say). Since $\hat{f}(s_1) = s_1$ and $\hat{f}(s_2) = s_2$, we have $\hat{f}(P_1) \in s_1$ and $\hat{f}(P_2) \in s_2$, hence the straight line $r' = \hat{f}(r)$ intersects s_1 in $\hat{f}(P_1)$ and s_2 in $\hat{f}(P_2)$, thus lying in the plane containing s_1, s_2, and r. We distinguish two cases.

Case 1: No matter how we choose r, $\hat{f}(r)$ is parallel to r. In this case, we prove (B5). Indeed, let y be an arbitrary point of R^{n*}. Then, since $\hat{f}$ maps a straight line through the origin onto itself, we have $\hat{f}(y) = \lambda(y)y$, $\lambda(y) \in R^*$. Let $z \neq y$ be any other point of R^{n*} and take r to be the straight line through z and y. Since r' is parallel to r we obviously have $\lambda(y) = \lambda(z)$, namely $\lambda(y)$ does not depend on y.

Case 2: There is a least one r such that r' is not parallel to r. Let s be the straight line through the origin which is parallel to r'. It intersects r in a point Q. Since s is mapped by $\hat{f}$ onto itself we must have $Q' = \hat{f}(Q) = r' \cap s$. But $r' \cap s = \emptyset$ and we have a contradiction. From (B3), (B4), and (B5) it follows that

$$f: x \to M(\lambda x) + f(0),$$

namely f is an affine permutation of R^n.

---

* Partially supported by a Consiglio Nazionale delle Ricerche Grant.

† On leave of absence from Istituto di Scienze Fisiche dell'Università, Milano.

‡ Supported in part by the U.S.A.E.C. under Contract ORO-(40-1) 3992 and by the A. V. Humboldt Stiftung with a grant. A travel grant under the Fulbright-Hays program is acknowledged.

§ Present address.

1 (a) V. Berzi and V. Gorini, J. Math. Phys. **10**, 1518 (1969); (b) V. Gorini and A. Zecca, J. Math. Phys. **11**, 2226 (1970); (c) V. Gorini, "Proceedings of the International Study Institute in Mathematical Physics," Istanbul, 1970; (d) V. Gorini, Commun. Math. Phys. **21**, 150 (1971); (e) V. Berzi and V. Gorini, "On Space-Time, Reference Frames and the Structure of Relativity Groups," to appear in Ann. Inst. Henri Poincaré. We refer to these papers, and in particular to (1a), for an extensive bibliography on the subject.

2 V. Lalan, Bull. Soc. Math France **65**, 83 (1937).

3 Compare e.g. V. Fock, *The Theory of Space, Time and Gravitation* (Pergamon-Macmillan, New York, 1964), Sec. 5.

4 This requirement could be replaced by the weaker condition of local boundedness.

5 Compare e.g. Ref. 1(a), Appendix A.

6 J. Aharoni, *The Special Theory of Relativity* (Oxford U.P., London, 1965), Chap. 1.

7 E. Artin, *Geometric Algebra* (Interscience, New York, 1957), Theorem 2.26.

8 For a thorough discussion of this point see R. M. F. Houtappel, H. Van Dam, and E. P. Wigner, Rev. Mod. Phys. **37**, 595 (1965).

9 See Footnotes 4 and 5.

10 Use the appropriate argument of Appendix B, by considering two distinct intersecting straight lines l_1 and l_2, lying in π and belonging to the cone through their intersection point. If $Q \in \pi$, $Q \notin l_1 \cup l_2$, the straight line l_Q of Appendix B through Q and intersecting l_1 and l_2 must be chosen belonging to $C_Q(N)$.

11 We require V to be commutative in order not to have to distinguish between left and right vector spaces over K. However, all considerations that we make here could be extended equally well to left and right vector spaces over noncommutative fields.

12 In Ref. 7 the theorem is stated and proved for the more general case of left vector spaces over not necessarily commutative fields.

Elementary Spinorial Excitations in a Model Universe

Antonio F. Rañada
Departamento de Física Teórica, Universidad de Madrid, Madrid, Spain.

and

Mario Soler
División de Física Teórica, Junta de Energía Nuclear, Madrid, Spain.

(Received 1 November 1971)

The generalized covariant Dirac equation for a certain five-dimensional universe is studied. If a torsion invariant is included in the free Lagrangian, it is shown that particlelike stable solutions exist having definite positive rest energy, spin, and corresponding antiparticles. The treatment is throughout classical.

I. INTRODUCTION

We show in this article the existence of stable particlelike classical elementary excitations in a model universe. The excitations are of the spinorial type and the results very similar to those of a model previously proposed by one of the authors[1] in a paper to be referred as I.

Our starting point combines two different considerations: on the one hand, if we are to find particlelike objects with a certain mass in some universe without introducing phenomenological parameters, we need to fix some scale or elementary length. There are two ways of doing this; one related to the microstructure of space-time, the other to its macrostructure. (Perhaps a third possibility would be to combine both.) We will assume the latter and consider a universe which might intuitively be called a thin three-dimensional sheet. That is, besides the ordinary three dimensions and time, a new dimension is postulated such that space is very narrow across. Its width will provide the necessary scale of length.

Our second basic consideration is the requirement of as much symmetry as can be provided for the particlelike object. The reason for this is simple. Ordinary symmetry under four-translations and Lorentz rotations is known to be a necessary but apparently not sufficient condition for a satisfactory classical theory of elementary particles. Invariance under generalized point dependent transformations is clearly stronger. It also provides in some cases a self-interaction which happens to be essential in order to have a normalizable theory. This is the kind of symmetry we will consider.

Coming to our first point, it is useful to analyze a massless boson field in the five-dimensional model universe, subject to the boundary conditions imposed by the two parallel planes $x^5 = \pm a$. (No space is supposed to exist outside these planes. Isotropy is assumed inside for the five-dimensional continuum, whenever boundaries are not reached.) (Similarly to what happens in any vibrating medium with boundaries, no flow of current can escape through any area of the boundary. This implies that the current component perpendicular to the boundary should vanish at each point of the boundary.)

It is clear that a mass term and a mass spectrum can be related to the existence of the bounded fifth dimension.

If the Klein-Gordon equation without a mass is written

$$\partial_\mu \partial^\mu \phi = 0 \tag{1}$$

with the metric $g_{\mu\nu}(1, -1, -1, -1, -1)$ and, in order to take into account the boundary conditions, ϕ is supposed to factorize as

$$\phi = A \begin{Bmatrix} \sin(n\pi/a)x^5 \\ \cos[(n + \frac{1}{2})\pi/a]x^5 \end{Bmatrix} \phi(\vec{x}, t), \tag{2}$$

where n is an integer, A a constant, and a the half-width of the universe observers unable to realize the existence of a very narrow a would tend to consider Eq. (1) an ordinary Klein-Gordon equation with a mass term

$$\Box\phi + m^2\phi = 0, \quad m^2 = n^2\pi^2/a^2, (n + \tfrac{1}{2})^2\pi^2/a^2. \tag{3}$$

As is seen one also finds a mass spectrum which is, of course, trivial. For the nonlinear equations which will be found later the mass spectrum cannot be calculated analytically.

A similar approach can be followed for spinorial excitations in the five-dimensional space. However,

‡ Supported in part by the U.S.A.E.C. under Contract ORO-(40-1) 3992 and by the A. V. Humboldt Stiftung with a grant. A travel grant under the Fulbright-Hays program is acknowledged.

§ Present address.

1 (a) V. Berzi and V. Gorini, J. Math. Phys. **10**, 1518 (1969); (b) V. Gorini and A. Zecca, J. Math. Phys. **11**, 2226 (1970); (c) V. Gorini, "Proceedings of the International Study Institute in Mathematical Physics," Istanbul, 1970; (d) V. Gorini, Commun. Math. Phys. **21**, 150 (1971); (e) V. Berzi and V. Gorini, "On Space-Time, Reference Frames and the Structure of Relativity Groups," to appear in Ann. Inst. Henri Poincaré. We refer to these papers, and in particular to (1a), for an extensive bibliography on the subject.

2 V. Lalan, Bull. Soc. Math France **65**, 83 (1937).

3 Compare e.g. V. Fock, *The Theory of Space, Time and Gravitation* (Pergamon-Macmillan, New York, 1964), Sec. 5.

4 This requirement could be replaced by the weaker condition of local boundedness.

5 Compare e.g. Ref. 1(a), Appendix A.

6 J. Aharoni, *The Special Theory of Relativity* (Oxford U.P., London, 1965), Chap. 1.

7 E. Artin, *Geometric Algebra* (Interscience, New York, 1957), Theorem 2.26.

8 For a thorough discussion of this point see R. M. F. Houtappel, H. Van Dam, and E. P. Wigner, Rev. Mod. Phys. **37**, 595 (1965).

9 See Footnotes 4 and 5.

10 Use the appropriate argument of Appendix B, by considering two distinct intersecting straight lines l_1 and l_2, lying in π and belonging to the cone through their intersection point. If $Q \in \pi$, $Q \notin l_1 \cup l_2$, the straight line l_Q of Appendix B through Q and intersecting l_1 and l_2 must be chosen belonging to $C_Q(N)$.

11 We require V to be commutative in order not to have to distinguish between left and right vector spaces over K. However, all considerations that we make here could be extended equally well to left and right vector spaces over noncommutative fields.

12 In Ref. 7 the theorem is stated and proved for the more general case of left vector spaces over not necessarily commutative fields.

Elementary Spinorial Excitations in a Model Universe

Antonio F. Rañada
Departamento de Física Teórica, Universidad de Madrid, Madrid, Spain.

and

Mario Soler
División de Física Teórica, Junta de Energía Nuclear, Madrid, Spain.

(Received 1 November 1971)

The generalized covariant Dirac equation for a certain five-dimensional universe is studied. If a torsion invariant is included in the free Lagrangian, it is shown that particlelike stable solutions exist having definite positive rest energy, spin, and corresponding antiparticles. The treatment is throughout classical.

I. INTRODUCTION

We show in this article the existence of stable particlelike classical elementary excitations in a model universe. The excitations are of the spinorial type and the results very similar to those of a model previously proposed by one of the authors[1] in a paper to be referred as I.

Our starting point combines two different considerations: on the one hand, if we are to find particlelike objects with a certain mass in some universe without introducing phenomenological parameters, we need to fix some scale or elementary length. There are two ways of doing this; one related to the microstructure of space-time, the other to its macrostructure. (Perhaps a third possibility would be to combine both.) We will assume the latter and consider a universe which might intuitively be called a thin three-dimensional sheet. That is, besides the ordinary three dimensions and time, a new dimension is postulated such that space is very narrow across. Its width will provide the necessary scale of length.

Our second basic consideration is the requirement of as much symmetry as can be provided for the particlelike object. The reason for this is simple. Ordinary symmetry under four-translations and Lorentz rotations is known to be a necessary but apparently not sufficient condition for a satisfactory classical theory of elementary particles. Invariance under generalized point dependent transformations is clearly stronger. It also provides in some cases a self-interaction which happens to be essential in order to have a normalizable theory. This is the kind of symmetry we will consider.

Coming to our first point, it is useful to analyze a massless boson field in the five-dimensional model universe, subject to the boundary conditions imposed by the two parallel planes $x^5 = \pm a$. (No space is supposed to exist outside these planes. Isotropy is assumed inside for the five-dimensional continuum, whenever boundaries are not reached.) (Similarly to what happens in any vibrating medium with boundaries, no flow of current can escape through any area of the boundary. This implies that the current component perpendicular to the boundary should vanish at each point of the boundary.)

It is clear that a mass term and a mass spectrum can be related to the existence of the bounded fifth dimension.

If the Klein-Gordon equation without a mass is written

$$\partial_\mu \partial^\mu \phi = 0 \tag{1}$$

with the metric $g_{\mu\nu}(1, -1, -1, -1, -1)$ and, in order to take into account the boundary conditions, ϕ is supposed to factorize as

$$\phi = A \begin{Bmatrix} \sin(n\pi/a)x^5 \\ \cos[(n + \frac{1}{2})\pi/a]x^5 \end{Bmatrix} \phi(\vec{x}, t), \tag{2}$$

where n is an integer, A a constant, and a the half-width of the universe observers unable to realize the existence of a very narrow a would tend to consider Eq. (1) an ordinary Klein-Gordon equation with a mass term

$$\Box\phi + m^2\phi = 0, \quad m^2 = n^2\pi^2/a^2, (n + \tfrac{1}{2})^2\pi^2/a^2. \tag{3}$$

As is seen one also finds a mass spectrum which is, of course, trivial. For the nonlinear equations which will be found later the mass spectrum cannot be calculated analytically.

A similar approach can be followed for spinorial excitations in the five-dimensional space. However,

both in the Klein-Gordon and in the Dirac cases, these linear Lorentz covariant equations have no satisfactory solutions from a classical point of view. We mean by satisfactory a theory having stable regular solutions with finite positive energy at rest.

Let us then introduce our stronger symmetry requirements and demand covariance under generalized (point dependent) Poincaré transformations. Our purpose is to study with this point of view the five-dimensional massless Dirac equation.

II. COVARIANT DIRAC EQUATION IN A NONSYMMETRIC FIVE-DIMENSIONAL SPACE-TIME

One can proceed writing the ordinary Lorentz covariant Lagrangian in 5-space

$$L_D = \bar{\psi} i\gamma^\mu \partial_\mu \psi - \partial_\mu \bar{\psi} i\gamma^\mu \psi, \tag{4}$$

where γ^μ stands for the usual Dirac matrices, supplemented with γ^5.

We use

$$\gamma^0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix},$$

$$\gamma^5 = -i\begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}, \quad \bar{\psi} = \psi^+\gamma^0.$$

Note that L_D is twice the usual Dirac Lagrangian. In general relativity this is the Lagrangian which couples with the gravitational field. See for instance Ref. 4.

In order to make the action integral invariant under the generalized transformations, a covariant derivative must be introduced in the form[2-6]

$$\partial_k \psi \equiv \psi_{,k} \to \mathfrak{D}_k \psi = h_k{}^\mu (\psi_{,\mu} + \tfrac{1}{2} A^{ij}{}_\mu S_{ij} \psi), \tag{5}$$

where the 25 fields $h_k{}^\mu$ are the contravariant components of a "fünfbein" system in five-dimensional space, the $A^{ij}{}_\mu$ are the local affine connections, and $S_{ij} = \frac{1}{4}[\gamma_i \gamma_j]$. (Except for obvious changes we use the same notation as Ref. 4.)

Since our goal is to have a well determined and consistent set of equations for the Dirac field ψ, we must add to L_D the free field Lagrangians corresponding to the new fields $h_k{}^\mu$ and $A^{ij}{}_\mu$ which have been introduced.

The free field Lagrangians must be scalars, and we need the covariant field strengths to construct them. These are deduced from the commutator of two covariant derivatives

$$[\mathfrak{D}_l, \mathfrak{D}_k] = \tfrac{1}{2} R^{ij}{}_{kl} S_{ij} - C^i{}_{kl} \mathfrak{D}_i, \tag{6}$$

where

$$R^{ij}{}_{kl} = h_k{}^\mu h_l{}^\nu (A^{ij}{}_{\mu,\nu} - A^{ij}{}_{\nu,\mu} - A^i{}_{k\mu} A^{kj}{}_\nu + A^i{}_{k\nu} A^{kj}{}_\mu), \tag{7}$$

$$C^i{}_{kl} = (h_k{}^\mu h_l{}^\nu - h_l{}^\mu h_k{}^\nu)(b^i{}_{\mu,\nu} + A^i{}_{g\nu} b^g{}_\mu), \tag{8}$$

and $b^i{}_\mu$ is the inverse of $h_i{}^\mu$ satisfying

$$b^i{}_\mu h_i{}^\nu = \delta_\mu{}^\nu, b^i{}_\mu h_j{}^\mu = \delta^i{}_j.$$

$R^{ij}{}_{kl}$ and $C^i{}_{kl}$ are, respectively, the curvature and torsion tensors expressed in the local system. Among the many invariants which can be constructed from these tensors, the contraction $R = R^{ij}{}_{ij}$ and the product $C = C^i{}_{kl} C_i{}^{kl}$ stand as the simplest, both being of second order in the $A^{ij}{}_\mu$ field variables. ($C^i{}_{kl}$ decomposes generally in irreducible components.[7,8] For the Dirac field it is completely antisymmetric.)

We thus naturally arrive at the Lagrangian

$$\mathfrak{L} = (R + \tfrac{1}{4}\chi C + k\mathfrak{L}_D)\mathfrak{K}, \tag{9}$$

where $\mathfrak{K} = [\det(h_k{}^\mu)]^{-1}$, χ and k are cosmological constants, and $\mathfrak{L}_D$ has to be expressed in terms of the covariant derivative (5), that is,

$$\mathfrak{L}_D = \bar{\psi} i\gamma^k \mathfrak{D}_k \psi - (\mathfrak{D}_k \bar{\psi}) i\gamma^k \psi. \tag{10}$$

$\mathfrak{L}$ is now invariant under generalized five-dimensional Poincaré transformations. The whole set of equations of motion implies variation with respect to the field variables $\bar{\psi}, h_k{}^\mu, A^{ij}{}_\mu$. These equations completely define a solution for the ψ and the other fields if adequate boundary conditions are provided.

Variation with respect to $\bar{\psi}$ gives the equation

$$i\gamma^k \mathfrak{D}_k \psi = 0. \tag{11}$$

Variation with respect to $A^{ij}{}_\mu$ gives

$$h_k{}^\mu C^k{}_{ij} - h_j{}^\mu C^k{}_{ik} - h_i{}^\mu C^k{}_{kj} + \chi h_k{}^\mu C_{ij}{}^k = k h_k{}^\mu S^k{}_{ij}. \tag{12}$$

One gets, from (12),

$$(1 + \chi) C_{ijk} = k S_{ijk}, \tag{13}$$

where $S^k{}_{ij}$ is the spin tensor of the five-dimensional Dirac field and has the usual value

$$S_{ijk} = -\tfrac{1}{2} i \bar{\psi} \gamma_i \gamma_j \gamma_k \psi \tag{14}$$

if i, j, k, are all different and is otherwise zero. Since only the spin density contributes to C_{ijk}, it is clear that torsion appears in the space as a consequence of the existence of matter. There are two contributions to (13) related to the terms R and C in the total Lagrangian. If these contributions happened to cancel each other, a peculiar situation arises, since the Dirac field must vanish and the other fields remain undetermined. How much space deviates from such situation is measured by the parameter $\lambda = -(1 + \chi)$. It will be seen later that a plus sign for λ is essential for the existence of particlelike elementary excitations with definite positive energy. Omission of the torsion term in (9), corresponding to $\lambda = -1$, would never make the self-interaction energy definite positive.

If we define $\sigma = k/\lambda$, (13) can be rewritten

$$C_{ijk} = -\sigma S_{ijk}. \tag{15}$$

From (12) the $A^{ij}{}_\mu$ fields may be solved in terms of the $h_k{}^\mu$ and ψ. It is convenient to write them in the form

$$A_{ij\mu} = {}^{(0)}A_{ij\mu} + {}^{(1)}A_{ij\mu},$$

where

$$\left.\begin{aligned} {}^{(0)}A_{ij\mu} &= \tfrac{1}{2}b^k{}_\mu\,(c_{kij} - c_{ijk} - c_{jki}) \\ c^k{}_{ij} &= (h_i{}^\mu h_j{}^\nu - h_j{}^\mu h_i{}^\nu)b^k{}_{\mu,\nu} \end{aligned}\right\},$$

and

$${}^{(1)}A_{ij\mu} = -\tfrac{1}{2}\sigma b^k{}_\mu S_{kij}. \tag{16}$$

The global affine connection has the form

$$\Gamma^\mu{}_{\lambda\nu} = -b^i{}_\lambda(h_i{}^\mu{}_{,\nu} - A^j{}_{i\nu}h_j{}^\mu). \tag{17}$$

From (16) and (17) it is easily seen that ${}^0A_{ij\mu}$ contributes only to the symmetric part of $\Gamma^\mu{}_{\lambda\nu}$ while ${}^{(1)}A_{ij\mu}$ contributes an unsymmetric part. It is possible however to eliminate the antisymmetric part of $\Gamma^\mu{}_{\lambda\mu}$ by means of Weyl's trick,[3] i.e., by expressing ${}^{(1)}A_{ij\mu}$ in terms of the matter and the $h_k{}^\mu$ fields. The remaining $\Gamma^\mu{}_{\lambda\nu}$ are then symmetric and we fall into an ordinary Riemannian space. The generalized Dirac equation (11) picks up then a nonlinear spin–spin interaction. This means that a theory which is linear in the ψ field when expressed in general affine space, turns out to be nonlinear in ordinary Riemannian space.

Microscopically there are no reasons to believe that curvature of space substantially affects the properties of elementary particles. We therefore introduce the simplifying assumption that once the antisymmetric part of the affine connection is eliminated at the expense of linearity, the Minkowsky limit is adequate.

The equation we arrive to is then

$$i\gamma^\mu\partial_\mu\psi + \tfrac{3}{4}\sigma\,(\bar{\psi}\psi)\psi = 0, \tag{18}$$

which corresponds to a Lagrangian

$$L = \tfrac{1}{2}L_D + \tfrac{1}{8}\sigma S^{ijk}S_{ijk}. \tag{19}$$

We are still left with the third set of equations of motion, coming from variation with respect to the $h_k{}^\mu$ fields. They may be written as

$$G^k{}_\mu + \tfrac{1}{4}\chi U^k{}_\mu = -kT^k{}_\mu, \tag{20}$$

where $G^k{}_\mu$ is Einstein's tensor, $T^k{}_\mu$ the energy momentum tensor, and $U^k{}_\mu$ is given as

$$\delta\int C\mathcal{K}d^5x = \int U^k{}_\mu\,\delta h_k{}^\mu\mathcal{K}d^5x. \tag{21}$$

As we see (20) differs by the term $U^k{}_\mu$ from the ordinary Einstein equations. A further difference is that $G^k{}_\mu$ and $T^k{}_\mu$ are not symmetric.[5]

However this should not be a question of concern, since Eq. (20) is written in a non-Riemannian space. The same equation can be expressed in an ordinary Riemannian space by eliminating the ${}^{(1)}A_{ij\mu}$ fields, after what the $U^k{}_\mu$ tensor no longer appears. It goes over to the $T^k{}_\mu$ tensor, which becomes symmetric, and (20) takes then the form of Einstein's equations.

III. LOWEST ENERGY SOLUTION OF THE MODEL

We now look for particlelike solutions of Eq. (18). In the Minkowsky limit our universe is a five-dimensional space subject to the condition $-a < x^5 < a$. We assume that nothing exists outside these boundaries. It is then natural to impose

$$j^5 = \bar{\psi}\gamma^5\psi = 0 \quad \text{for } x^5 = \pm a. \tag{22}$$

Since j^5 is the component of the matter current perpendicular to the boundaries, this implies that no current flows out of the Universe.

In order to solve Eq. (18) with condition (22), we start neglecting the nonlinear term. In order to factorize the solution as much as possible, it has been seen that the simplest prescription which fulfills condition (22) is as follows.

Let us take the spinors

$$\left.\begin{aligned} \psi_{a,n} &= \psi_{\mathrm{I},n}\cos[(2n+1)\mu\omega x^5] \\ \psi_{b,n} &= \psi_{\mathrm{II},n}\sin[(2n+1)\mu\omega x^5] \end{aligned}\right\}, \tag{23}$$

where $n = 0, 1, 2, \cdots, \mu$ is a parameter, and

$$\begin{aligned} \psi_{\mathrm{I},n} &= e^{-i\omega t}\begin{pmatrix} g_n\begin{pmatrix}1\\0\end{pmatrix} \\ if_n\begin{pmatrix}\cos\theta\\ e^{i\varphi}\sin\theta\end{pmatrix}\end{pmatrix}, \\ \psi_{\mathrm{II},n} &= e^{-i\omega t}(-1)^n\begin{pmatrix} -f_n\begin{pmatrix}\cos\theta\\ e^{i\varphi}\sin\theta\end{pmatrix} \\ ig_n\begin{pmatrix}1\\0\end{pmatrix}\end{pmatrix}. \end{aligned} \tag{24}$$

f_n and g_n are functions of $r = (x^2 + y^2 + z^2)^{1/2}$.

We now consider the combination

$$\psi_n = \psi_{a,n} + i\psi_{b,n} \tag{25}$$

and substitute ψ_n in (18) neglecting for the moment the nonlinear term. A short calculation leads to simple expressions for f_n and g_n in terms of Bessel functions. It is impossible, as should be expected, to obtain simultaneously good behavior at the origin and at infinity. Although the price that has to be paid is rather high as regards numerical computation, the introduction of nonlinearity provides solutions (characterized by the number of nodes) which are well behaved throughout and have a finite norm.

The rather special form (25) has been chosen so that the combination

$$\psi = \sum\psi_n \tag{26}$$

verifies automatically the natural boundary condition $j^5 = 0$ at $x^5 = \pm a$ if we take $\mu\omega a = \pi/4$. In the Fourier series (26), only odd arguments appear, each term in the series being a combination of odd and even harmonic functions of x^5. The x^5 derivative acts on each of these functions so as to provide a sort of mass term for the other.

When the nonlinear terms are considered, the picture seems to get obscured since (18) is no longer separable in the same simple "term by term" way. Fortunately, however, the complication amounts to a coupling of the successive radial functions and can be overcome. The x^5 dependence also factorizes.

The sensible thing to do is to try for the actual ψ the same expansion (26). It is now convenient to introduce

some changes in variables and functions in order to do away with unessential constants.

If we choose

$$g_n = (\omega/\tfrac{3}{4}\sigma)^{1/2}[2(1+\mu)]^{1/2}G_n,$$
$$f_n = (\omega/\tfrac{3}{4}\sigma)^{1/2}[2(1+\mu)]^{1/2}F_n, \qquad (27)$$
$$\rho = \omega(1+\mu)r, \quad \alpha = \omega x^5$$

and keep only the $n = 0$ term, the resulting equations are

$$F_0' + (2/\rho)F_0 + \nu G_0 + G_0(F_0^2 - G_0^2) = 0,$$
$$G_0' + F_0 + F_0(F_0^2 - G_0^2) = 0, \qquad (28)$$

where $\nu = (\mu - 1)/(\mu + 1)$ and prime means $(d/d\rho)$. Surprisingly enough, these are Eq. (12) of I.

In second order we find

$$F_0' + (2/\rho)F_0 + \nu G_0 + (G_0 + G_1)(F_0^2 - G_0^2 + 2F_0F_1 - 2G_0G_1) = 0,$$

$$G_0' + F_0 + (F_0 + F_1)(F_0^2 - G_0^2 + 2F_0F_1 - 2G_0G_1) = 0,$$

$$F_1' + (2/\rho)F_1 - (2+\nu)G_1 + G_0(F_0^2 - G_0^2 + 2F_0F_1 - 2G_0G_1) + G_1(F_1^2 - G_1^2) = 0,$$

$$G_1' - (1+2\nu)F_1 + F_0(F_0^2 - G_0^2 + 2F_0F_1 - 2G_0G_1) + F_1(F_1^2 - G_1^2) = 0. \qquad (29)$$

We have obtained a numerical solution (without nodes) for this system subject to the boundary condition $F_0, G_0, F_1, G_1, \to 0$ as $\rho \to \infty$. The solution is unique for each value of ν. The method is similar to the one used in I. It turns out that the physically relevant solutions (that is, those satisfying the above mentioned boundary conditions) tend very rapidly to exponentially decreasing solutions of the equation with no nonlinear term. Moreover, as should be expected, F_1 and G_1 decrease much faster than F_0, G_0.

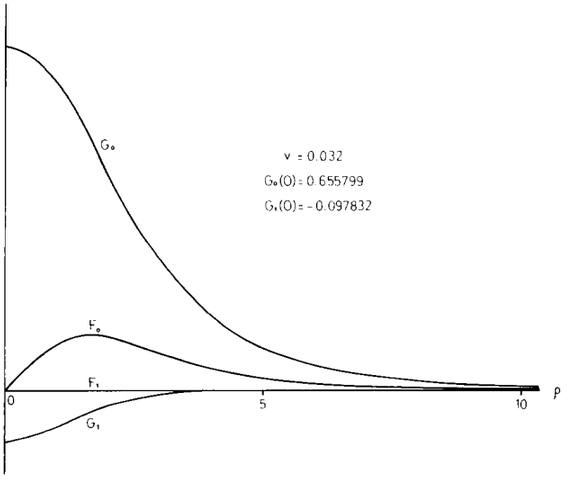


FIG. 1 Numerical solution of Eq. (18) including the first two terms of expansion (26) and satisfying the physical boundary conditions. The function F_1 is too small to appear in the drawing.

In Fig. 1 we show our results for the ν value corresponding to a minimum energy. The contribution of the second wave is quite small. It gets comparatively smaller for smaller values of ν and somewhat larger for larger values.

Of course a rigorous solution would entail evaluating the contribution of the rest of the series (26). This looks hopeless for such nonlinear equation. However, the smallness of the second wave gives strong evidence that for all practical purposes the first wave, which describes the model proposed in I, gives a satisfactory account of elementary spinorial excitations in our model universe. (The existence of bounded solutions can be proved by techniques derived from fixed point theorems. However, for the practical purpose of obtaining a spectrum of solutions corresponding to zero nodes, or eventually one or several nodes, we believe our method to be quite adequate.)

The energy can be expressed in first order as

$$E(\nu) = \tfrac{64}{3}(a^2/\sigma)(1+\nu)[(1-\nu)I_1 + I_2],$$

where

$$I_1 = \int_0^\infty (F_0^2 + G_0^2)\rho^2 d\rho,$$
$$I_2 = \int_0^\infty (F_0^2 - G_0^2)^2\rho^2 d\rho. \qquad (30)$$

The minimum value, which obtains for $\nu = 0.032$ is

$$E = \tfrac{64}{3}(a^2/\sigma)(13.72 + 1.26). \qquad (31)$$

The second wave gives a correction to the energy smaller than 1%. Figure 1 shows that most of the particle is concentrated in a sphere with radius $\rho = 5$, which by (27) corresponds to $r \approx 5/\omega(1+\mu) \approx 2.5/\omega$. On the other hand, the boundary conditions for the x^5 dependence provide the relation: $2a = \pi/2\omega\mu$, that is, $a \approx 0.7/\omega$. From these relations one might estimate the value of λ if the rest energy of the particles were experimentally known.

A model completely parallel to the one considered above can be constructed if the sign of the Lagrangian $\mathcal{L}_D$ for the matter field is reversed in (9). Since for this model the sign of the conserved current $j^\mu = -\bar{\psi}\gamma^\mu\psi$ is reversed, this suggests that the solutions be considered as antiparticles of the former ones. (See Ref. 9.) The equations are in this case

$$i\gamma^\mu\partial_\mu\psi - \tfrac{3}{4}\sigma(\bar{\psi}\psi)\psi = 0. \qquad (32)$$

If one makes in our solution to (18) the changes

$$\sin \leftrightarrow \cos,$$
$$\omega \leftrightarrow -\omega,$$
$$g_n \leftrightarrow (-1)^n g_n,$$
$$f_n \leftrightarrow (-1)^{n+1} f_n,$$

a solution to (32) is obtained having the same positive energy as the former.

IV. FINAL COMMENTS AND CONCLUSIONS

We would like to stress that the model which has been

explored above is remarkably simple and free of "ad hoc" assumptions.

It might look artificial to postulate a five-dimensional continuum, since space–time appears to be sufficiently described with four dimensions. The reasons for this initial postulate are not compelling but are relevant from the point of view of simplicity. In fact we have tried to eliminate the appearance of "ad hoc" mass terms in the equations describing elementary particles. These terms are disturbing, since it is not clear what dynamical mechanisms should be at their origin. No terms of this kind appear in the description of vibrations in solids which we took as a starting analogy. We accordingly tried to do away with mass terms and looked for more natural sources of the rest energy. The study of vibrations in thin films showed to us how a masslike behavior arises when a narrow transverse dimension exists. This we abstracted to be our x^5. Moreover, perhaps the remark can be made that a very simple model of closed expanding universe is the shell of a five-dimensional sphere. It is hard to understand that the width of the shell should be exactly zero.

Our second assumption of generalized symmetry is obviously much more general than invariance under ordinary Poincaré transformations.

It seems quite appropriate not to impose a rigid metric connection in space–time, independently of whether matter is or is not present. This provides the possibility that its structure can be twisted when the kind of matter which is present possesses spin. We have showed that if our particles were neutrinos and the width of the universe were known, the constant λ might be determined. Since this constant measures some sort of total twisting recovery strength of the space-time continuum, our model probes in this sense the geometrical properties of space–time. A necessary condition is that λ should be positive.

It has been known for quite a time[4,5] that spin can be defined as a dynamical property of classical fields. There is a very close analogy between charge and spin when one uses a gauge covariant formalism as has been done in this paper. The introduction of the fields $A^{ij}{}_{\mu}$ in order to define a covariant derivative exactly parallels the principle of "minimal electromagnetic interaction." Here, however, because of their different geometrical structure, not all fields possess the same "coupling constant", or, in other words, spin. Spinors have of course spin $\frac{1}{2}$, and this is the meaning of the $\frac{1}{2}$ factor which appears in Eq. (5).

Since we have a stable model which might describe the structure of an elementary particle, the dynamical definition of spin shows here all its relevance. Independently of the existence of Hermitian operators which in a nonlinear theory have no meaning, this particle has spin $\frac{1}{2}$ because of its geometrical nature which is of course completely classical.

We conclude that in spite of its unrealistic simplicity (no weak, electromagnetic, or strong interactions have been considered) the model shows that localized stable excitations can exist in a certain rather reasonable space–time. These classical particles would have a highly nontrivial rest energy spectrum (depending on the number of nodes of the wavefunctions), and other properties such as definite spin and corresponding antiparticles, usually considered typical quantum mechanical features.

ACKNOWLEDGMENT

We are indebted to Professor A. Galindo for his kind interest in this work as well as for a useful discussion.

[1] M. Soler, Phys. Rev. D **1**, 2766 (1970).
[2] H. Weyl, Z. Physik, **56**, 330 (1929).
[3] H. Weyl, Phys. Rev. **77**, 699 (1950).
[4] T. W. B. Kibble, J. Math. Phys. **2**, 212 (1961).
[5] D. W. Sciama, *Festschrift für Infeld* (Pergamon, New York, 1962).
[6] R. Utiyama, Phys. Rev. **101**, 1597 (1956).
[7] K. Hayashi and T. Nakano, Progr. Theoret. Phys. (Kyoto) **38**, 491 (1967).
[8] K. Hayashi, Progr. Theoret. Phys. (Kyoto) **39**, 494 (1968).
[9] M. Soler, Preprint JEN (unpublished) (1970).

Complete Solution of the Inverse Scattering Problem at Fixed Energy

Pierre C. Sabatier*†
Département de Physique Mathématique,‡ Université de Montpellier, Montpellier, France
and
Department of Physics, Indiana University, Bloomington, Indiana

(Received 19 April 1971)

Let $\mathfrak{D}$ be the class of functions which are bounded by $Cr^{-1+\epsilon}$ and $Cr^{-3-\epsilon}$, $\mathcal{S}$ the class of potentials $V(r)$ such that $V(r)$, $rV'(r)$, and $r^2V''(r)$ belong to $\mathfrak{D}$. $\mathcal{S}$ is dense in the class of potentials $\mathfrak{V}$ with finite norm $\int_0^\infty \rho\,|V(\rho)|\,d\rho$ in which almost all the results of potential scattering are derived. In this paper a *complete solution* of the inverse scattering problem at fixed energy is given in a class $\tilde{\mathcal{S}}$ of potentials which *contains* $\mathcal{S}$. This means that given any set of phase shifts bounded by $Cl^{-1-\epsilon}$, we construct *all* the potentials of $\tilde{\mathcal{S}}$ which fit this set of phase shifts. They depend on an arbitrary function. The fundamental tool in the solution is the "scattering structure function." The method is derived in such a way that an approximation theory and numerical computations are feasible. These, together with various studies of the solutions, are the object of forthcoming papers.

1. INTRODUCTION

We study the elastic scattering of a particle obeying the Schrödinger equation with a spherically symmetric potential, at an energy $E = \frac{1}{2}\hbar^2k^2/m$, m being the reduced mass and k the linear momentum. Our "inverse problem" is the construction of the potential from the phase shifts.

Being given a sequence $\delta(\delta = \{\delta_l\})$ of phase shifts, five questions are of interest. The two first ones are the *a priori* questions of the problem:

(i) Does a potential V exist, which generates δ, at the energy E, through the Schrödinger equation? Such a potential is called a solution of the inverse problem at the energy E.

explored above is remarkably simple and free of "ad hoc" assumptions.

It might look artificial to postulate a five-dimensional continuum, since space–time appears to be sufficiently described with four dimensions. The reasons for this initial postulate are not compelling but are relevant from the point of view of simplicity. In fact we have tried to eliminate the appearance of "ad hoc" mass terms in the equations describing elementary particles. These terms are disturbing, since it is not clear what dynamical mechanisms should be at their origin. No terms of this kind appear in the description of vibrations in solids which we took as a starting analogy. We accordingly tried to do away with mass terms and looked for more natural sources of the rest energy. The study of vibrations in thin films showed to us how a masslike behavior arises when a narrow transverse dimension exists. This we abstracted to be our x^5. Moreover, perhaps the remark can be made that a very simple model of closed expanding universe is the shell of a five-dimensional sphere. It is hard to understand that the width of the shell should be exactly zero.

Our second assumption of generalized symmetry is obviously much more general than invariance under ordinary Poincaré transformations.

It seems quite appropriate not to impose a rigid metric connection in space–time, independently of whether matter is or is not present. This provides the possibility that its structure can be twisted when the kind of matter which is present possesses spin. We have showed that if our particles were neutrinos and the width of the universe were known, the constant λ might be determined. Since this constant measures some sort of total twisting recovery strength of the space-time continuum, our model probes in this sense the geometrical properties of space–time. A necessary condition is that λ should be positive.

It has been known for quite a time[4,5] that spin can be defined as a dynamical property of classical fields. There is a very close analogy between charge and spin when one uses a gauge covariant formalism as has been done in this paper. The introduction of the fields $A^{ij}{}_{\mu}$ in order to define a covariant derivative exactly parallels the principle of "minimal electromagnetic interaction." Here, however, because of their different geometrical structure, not all fields possess the same "coupling constant", or, in other words, spin. Spinors have of course spin $\frac{1}{2}$, and this is the meaning of the $\frac{1}{2}$ factor which appears in Eq. (5).

Since we have a stable model which might describe the structure of an elementary particle, the dynamical definition of spin shows here all its relevance. Independently of the existence of Hermitian operators which in a nonlinear theory have no meaning, this particle has spin $\frac{1}{2}$ because of its geometrical nature which is of course completely classical.

We conclude that in spite of its unrealistic simplicity (no weak, electromagnetic, or strong interactions have been considered) the model shows that localized stable excitations can exist in a certain rather reasonable space–time. These classical particles would have a highly nontrivial rest energy spectrum (depending on the number of nodes of the wavefunctions), and other properties such as definite spin and corresponding antiparticles, usually considered typical quantum mechanical features.

ACKNOWLEDGMENT

We are indebted to Professor A. Galindo for his kind interest in this work as well as for a useful discussion.

[1] M. Soler, Phys. Rev. D **1**, 2766 (1970).
[2] H. Weyl, Z. Physik, **56**, 330 (1929).
[3] H. Weyl, Phys. Rev. **77**, 699 (1950).
[4] T. W. B. Kibble, J. Math. Phys. **2**, 212 (1961).
[5] D. W. Sciama, *Festschrift für Infeld* (Pergamon, New York, 1962).
[6] R. Utiyama, Phys. Rev. **101**, 1597 (1956).
[7] K. Hayashi and T. Nakano, Progr. Theoret. Phys. (Kyoto) **38**, 491 (1967).
[8] K. Hayashi, Progr. Theoret. Phys. (Kyoto) **39**, 494 (1968).
[9] M. Soler, Preprint JEN (unpublished) (1970).

Complete Solution of the Inverse Scattering Problem at Fixed Energy

Pierre C. Sabatier*†
Département de Physique Mathématique,‡ Université de Montpellier, Montpellier, France
and
Department of Physics, Indiana University, Bloomington, Indiana

(Received 19 April 1971)

Let $\mathfrak{D}$ be the class of functions which are bounded by $Cr^{-1+\epsilon}$ and $Cr^{-3-\epsilon}$, $\mathcal{S}$ the class of potentials $V(r)$ such that $V(r)$, $rV'(r)$, and $r^2V''(r)$ belong to $\mathfrak{D}$. $\mathcal{S}$ is dense in the class of potentials $\mathfrak{V}$ with finite norm $\int_0^\infty \rho\,|V(\rho)|\,d\rho$ in which almost all the results of potential scattering are derived. In this paper a *complete solution* of the inverse scattering problem at fixed energy is given in a class $\tilde{\mathcal{S}}$ of potentials which *contains* $\mathcal{S}$. This means that given any set of phase shifts bounded by $Cl^{-1-\epsilon}$, we construct *all* the potentials of $\tilde{\mathcal{S}}$ which fit this set of phase shifts. They depend on an arbitrary function. The fundamental tool in the solution is the "scattering structure function." The method is derived in such a way that an approximation theory and numerical computations are feasible. These, together with various studies of the solutions, are the object of forthcoming papers.

1. INTRODUCTION

We study the elastic scattering of a particle obeying the Schrödinger equation with a spherically symmetric potential, at an energy $E = \frac{1}{2}\hbar^2k^2/m$, m being the reduced mass and k the linear momentum. Our "inverse problem" is the construction of the potential from the phase shifts.

Being given a sequence $\delta(\delta = \{\delta_l\})$ of phase shifts, five questions are of interest. The two first ones are the *a priori* questions of the problem:

(i) Does a potential V exist, which generates δ, at the energy E, through the Schrödinger equation? Such a potential is called a solution of the inverse problem at the energy E.

(ii) Let us be given a class $\mathfrak{W}$ of functions containing a solution V of the inverse problem at the energy E. Is V the unique solution in $\mathfrak{W}$?

The answer to (i) is "in general" affirmative ("in general" meaning unless δ is an exceptional sequence, fulfilling special constraints). We therefore are led to the question:

(iii) Give a method for constructing a solution of the problem.

The answer to (ii) is definitely negative if $\mathfrak{W}$ is the class—say $\mathcal{P}$—of *all* the potentials leading to physically acceptable phase shifts. Only in very special subclasses of $\mathcal{P}$ can the answer of B be positive. This leads us to the two following questions, in which we call "equivalent" two potentials yielding the same sequence of phase shifts.

(iv) Give methods for constructing *all* the equivalent potentials in $\mathcal{P}$ or in a large enough, well defined, subclass of $\mathcal{P}$.

(v) Give an appraisal of the deviation from each other of all the equivalent potentials in $\mathcal{P}$ or in a large enough, well defined, subclass of $\mathcal{P}$.

In a more general framework, additional questions would be for instance: Let δ be given as a function of E, what conditions must be fulfilled for which at least one of the solutions is a static potential; and how do you get it? Our present interest is however strictly limited to the inverse problem at fixed energy. It is, therefore, convenient to use the following notations:

$$D_r^0 = r^2\left(\frac{\partial^2}{\partial r^2} + 1\right), \tag{1.1}$$

$$D_r^V = r^2\left(\frac{\partial^2}{\partial r^2} + 1 - V(r)\right), \tag{1.2}$$

so that the lth partial-wave equation reads

$$D_r^V \psi_l(r) = l(l+1)\psi_l(r). \tag{1.3}$$

The partial wave is the solution of (1.3) which behaves at the origin like $[r(l+\frac{1}{2})]^{-1}(\frac{1}{2}\pi)^{1/2} r^{l+1}$, so that for $V = 0$, we get

$$[\psi_l(r)]_{V=0} \equiv u_l(r) = (\tfrac{1}{2}\pi r)^{1/2} J_{l+1/2}(r). \tag{1.4}$$

The phase shifts are defined through the asymptotic behavior of $\psi_l(r)$:

$$\psi_l(r) = A_l \sin(r - l\pi/2 - \delta_l) + o(1), \quad r \to \infty. \tag{1.5}$$

A sufficient condition for the existence of the δ_l is that the following quantity is finite;

$$\|V\| = \int_0^a \rho^{1-\epsilon} |V(\rho)| d\rho + \int_a^\infty |V(\rho) d\rho, \tag{1.6}$$

where a is a fixed length.[1] This condition defines a set $\mathcal{U}$ of potential, in which $\|V\|$ is obviously a norm. However, the set $\mathcal{P}$ for which the δ_l are defined by (1.5) is much larger than $\mathcal{U}$: $\mathcal{P}$ contains, for instance, infinitely repulsive potentials.

Let us now summarize the approaches to the problem and our present knowledge of the question.

Roughly speaking, three kinds of formalisms have been used in the literature dealing with this problem. In the first one,[2-6] it is assumed *a priori* that the potential to be obtained is such that the *JWKB* approximation is valid for the phase shift. An additional assumption is more or less explicitly done, such that there is only one turning point. Hence, it is possible to define a function, say, $H(\lambda)$ from which $V(r)$ can straightforwardly be derived, at least for large enough r, and the phase shifts are given by

$$\delta(l) = \int_{l+1/2}^\infty \frac{\lambda}{\sqrt{\lambda^2 - (l-\frac{1}{2})^2}} H(\lambda) d\lambda. \tag{1.7}$$

Conversely, if δ is known as a differentiable function of l, the Abel transform (1.7) can be inversed and $H(\lambda)$ be obtained, then $V(r)$ from $H(\lambda)$. The method can be refined to include second order terms in *JWKB* approximation.[4,6] However, the only answer which can be done by this way to our five fundamental questions is—"Here is a potential $V(r)$ whose phase shifts are approximately the required ones." Although such an answer can be sufficient for certain physical purposes, we must observe that the inversion technique is itself not very well defined, since it includes an interpolation [from δ_l to $\delta(l)$], whose definition has been forgotten by most authors.[2,3,5,6] Yet, the Regge theorems, to be recalled below, show that the choice of an interpolation of $\delta(l)$ is the key of question (ii). Even for a rough application the method is questionable, since a small imaginary part in the potential leads to a dramatic situation as regards to the errors.

Due to Martin and Targonski,[7] the second way of approaching the problem gives a positive answer to (i) and (ii) for sets of phase shifts fulfilling some special constraints, and potentials of the Yukawa class. The method is not fitted to study questions (iii)–(v). It is not fitted either to make generalizations, except for those that may be very weak.[8]

The third approach, by far the richest one, can be considered as an extension to this problem of the Gel'fand–Levitan formalism. In such a formalism, being given two potentials V and W, one looks for a "transformation kernel" $K_V^W(r,r')$ which generates the wavefunction corresponding to W from the one corresponding to V through the formulas

$$\psi_l^W(r) = \psi_l^V(r) - \int_0^r K_V^W(r,\rho)\psi_l^V(\rho)\rho^{-2}d\rho, \tag{1.8}$$

$$K_V^W(r,r) = -\tfrac{1}{2}r\int_0^r \rho[W(\rho) - V(\rho)]d\rho. \tag{1.9}$$

All the scattering problem reduces therefore to the determination of $K_V^W(r,r')$. For this, it is convenient to introduce an auxiliary tool $f_V^W(r,r')$, which is a solution of the partial differential equation

$$\left.\begin{aligned}&[D_r^V - D_{r'}^V] f_V^W(r,r') = 0\\ &f_V^W(r,0) = f_V^W(0,r') = 0\end{aligned}\right\} \tag{1.10}$$

and which enable one to obtain $K_V^W(r,r')$ through what we think right to call the Regge–Newton equation

$$K_V^W(r,r') = f_V^W(r,r') - \int_0^r K_V^W(r,\rho) f_V^W(\rho,r)\rho^{-2}d\rho \tag{1.11}$$

obtained by analogy with the Gel'fand–Levitan equation. Two kinds of studies have used $f_V^W(r,r')$ as a fundamental concept. In the study developed by Regge,[9] and more thoroughly by Loeffel,[10] V is equal to one, and the class of potentials is $\mathcal{U}$. Questions (iii)–(v) are neglected, but (ii) is studied thoroughly. The function $f_1^W(r,r')$ is characterized by

a set of numbers, the "spectral data" $(\gamma, \{\nu_k\}, \{d_k\})$, which enable its construction through the expansion

$$f_1^W[x,y] = -(2\pi)^{-1}\int_{-\infty}^{+\infty}\gamma(\tau)(xy)^{i\tau+1/2}\,d\tau + \sum_{k=1}^{\infty} d_k (xy)^{\nu_k+1/2}. \quad (1.12)$$

The series converges uniformly for (x,y) in any compact contained in $R_+ \times R_+$. Using this expansion, it has been possible to prove the two following uniqueness theorems[10]:

Theorem I: Let V_1 and V_2 be in $\mathcal{U}$; let α_1 and α_2 be the corresponding Jost functions. If

$$\alpha_1(l) = \alpha_2(l)$$

for $\mathrm{Re}\, l \geq -\frac{1}{2}$, then

$$V_1(r) = V_2(r)$$

for almost all positive r.

Theorem II: Let $s_l = e^{2i\delta_l}$ and let $\sigma(l)$ be the interpolation of s_l obtained through the Jost functions (the socalled Regge interpolation). Let V_1 and V_2 now be in the class $\mathcal{U}$. If the corresponding Regge interpolations σ_1 and σ_2 satisfy

$$\sigma_1(l) = \sigma_2(l)$$

for all l with $\mathrm{Re}\, l > -\frac{1}{2}$, where both are holomorphic, then

$$V_1(r) = V_2(r)$$

for almost all positive r.

The problem of uniqueness reduces therefore to the step $s_l \to \sigma(l)$. At this point, there can be uniqueness only for particular classes of potentials, allowing the interpolation to be unique. This is the case, for instance, if Carlson's theorem applies (Yukawa classes $\cdots$) or if the Lagrange–Valiron theorem applies (several classes studied by the author[11]). If a subclass of $\mathcal{U}$ is chosen such that the solution is unique or if Regge interpolations are known, questions (i) and (iii) can be answered positively in certain cases.[10,12] However, attempts to answer questions (iv) and (v) through this formalism are lacking and the stability of a method using an interpolation process as an intermediate step is highly questionable.[13]

In the study initiated by Newton,[14] the function $f_0^W(r,r')$ is given by the series

$$f_0^W(r,r') = \sum_{l=0}^{\infty} c_l u_l(r) u_l(r'). \quad (1.13)$$

which is obviously a solution of (1.10). Hence $K_0^W(r,r')$ has the expansion

$$K_0^W(r,r') = \sum_{l=0}^{\infty} c_l \psi_l^W(r) u_l(r'). \quad (1.14)$$

Inserting this expansion in (1.11) and going to infinity yield a infinite system of linear equations relating the δ_l and the c_l, so that solving question (iii) reduces to inverting some infinite matrices. A formal solution of the inverse problem is therefore obtained in a very elegant way. It remains however, for solving (iii), to give explicitly the inverse matrices for solving (i) and (ii), to prove that physically acceptable potentials are obtained from a physically acceptable set of phase shifts. This has been done,[15] starting from a sequence $\delta = \{\delta_l\}$ of phase shifts such that

$$|\delta_l| < C(1+l)^{-3-\epsilon} \quad (1.15)$$

and obtaining, for almost every sequence of this class, a one-parameter family of potentials such that

$$\left.\begin{aligned} c_{2l} &= a_0 + O(l^{-2}) \\ c_{2l+1} &= a_1 + O(l^{-2}) \end{aligned}\right\} \quad l \to \infty. \quad (1.16)$$

Unless $a_0 = a_1$, these potentials behave asymptotically like $Cr^{-3/2}\sin(2r+\varphi)$, whereas for $a_0 = a_1$, they decrease faster than $Cr^{-2+\epsilon}$. This is the case for one and only one value of the arbitrary parameter in the family corresponding to a sequence δ. We can therefore assert the following: Let $\mathcal{C}_1^*$ be the class of potentials whose transformation kernel is of the form (1.14) and the c_l' fulfill (1.16), and $\mathcal{C}_1$ the subclass with $a_0 = a_1$. In $\mathcal{C}_1^*$, the answer to (i) and (iii) is affirmative. This holds for (ii) in $\mathcal{C}_1$ only. The uniqueness, or almost uniqueness, encountered in these cases, corresponds to very special properties of the potentials.[16] The method is easy to extend by allowing l in (1.13) to take noninteger values[17] and, possibly, complex values.[18] However, the limitations of such a method, for whatever generalization, comes from the convergence conditions when going to the asymptotic limit of the linear system, which is equivalent to (1.11). A necessary condition for this is that the $|c_\mu|$ be bounded by $C(\mathrm{Re}\mu)^{1/3}$. But this simple bound considerably limit the class—say $\mathcal{C}$—which can be attained by all these methods. A study of the Jost functions shows that expansions are then valid, comparable to certain dispersion formulas,[19] and in which the c_μ appears as interpolation coefficients.[11] In the series giving asymptotic quantities, the apparent rate of convergence is that of the series $|c_l|$. Modifications of the method[20] enable one to increase this rate of convergence. Actually, since the limit a_θ of c_l as $l \to \infty$ is proportional[21] to $\int_0^\infty \rho V(\rho)\,d\rho$, it is, in our opinion, more interesting to calculate exactly the contribution of this term and make manipulations on series including $(c_l - a_0)$.[22,23]

Extensions of the Gel'fand–Levitan formalism to include Coulombian potentials,[24] relativistic cases,[25] spin orbit potential[26] and tensor forces[27] are also available, but in a less advanced state.

In the following, we characterize the two ways of treating $f(r,r')$ as the spectral data approach and the interpolation coefficients approach. Facing the impressive list of results obtained through these methods, we are nevertheless disappointed for the following reasons:

(a) There is *no method* as yet available to answer questions (i)–(iii) in a class of potentials defined through simple properties of the potentials.

(b) There is *no* attempt made to solve question (iv).

(c) The main effect of the only preliminary attempt[28] to answer question (v) has been to convince the author that representations of $f(r,r')$ with interpolation coefficient are very particular, and a representation through the Fourier transform of $f(r,r)$ would be much more fruitful.

These remarks have led the author to undertake a

study which is now in course of publication, and whose two other parts have already been issued.[23,29,30] The remarks which have led us to our method are the following.

(d) $K_0^V(r,r')$ is a solution of the partial differential equation [we use in the following the simplified notation $K(r,r')$],

$$\left.\begin{aligned} &(D_r^V - D_{r'}^0)K(r,r') = 0 \\ &K(r,0) = K(0,r') = 0 \\ &K(r,r) = -\tfrac{1}{2}r\int_0^r \rho V(\rho)\,d\rho \end{aligned}\right\} \tag{1.17}$$

Conversely, straightforward techniques[14,31] show that the solution of (1.17) is the transformation kernel corresponding to $V(\rho)$, so that (1.8) holds. The only important operator is therefore $K(r,r')$, $f(r,r')$ being at most an auxiliary tool.

(e) In $K(r,r')$ the only important part for our problem is its asymptotic behavior for large r.

(f) If $f(r,r')$ is used as an auxiliary tool, the Fourier transform of $f(r,r)$ yields probably its most convenient characterization.

(g) Almost all the results in potential scattering being valid only in $\mathcal{U}$ (or in subclasses of $\mathcal{U}$), results valid in $\mathcal{U}$ or in a subclass of $\mathcal{U}$ large enough, and dense in $\mathcal{U}$ with respect to the norm (1.6), can be considered a sufficient generality. Following these remarks, we have obtained in a previous paper,[29] hereafter referred to as I, the following results:

Let $\mathcal{E}_l^k$ be the class of continuous functions $V(r)$ such that, for $x \geq 0$,

$$\left.\begin{aligned} &|r^j V(r)| \leq C(r/a)^{\epsilon'} \\ &|r^k V(r)| \leq C(a/r)^{\epsilon} \end{aligned}\right\} \tag{1.18}$$

and let $\bar{\mathcal{E}}$ be equal to $\mathcal{E}_{13}$, $\mathcal{E}$ be the subclass of $\bar{\mathcal{E}}$ such that $rV'(r)$ and $r^2V''(r)$ belong to $\bar{\mathcal{E}}$. Let V belong to $\mathcal{E}_{21}$; the differential equation (1.17) has a solution, thoroughly studied in I. For $V \in \mathcal{E}$ it can be put in the form

$$\left.\begin{aligned} &K(r,r') = \cos rk(r') + \sin r\bar{k}(r') + P^{(N)}(r,r') \\ &\frac{\partial}{\partial r}K(r,r') = -\sin rk(r') + \cos r\bar{k}(r') + Q^{(N)}(r,r') \end{aligned}\right\}, \tag{1.19}$$

where $P^{(N)}(r,r')$ and $Q^{(N)}(r,r')$ are "negligible functions" viz functions such that

$$\left.\begin{aligned} &\text{(a)}\ P^{(N)}(r,r') \to 0, \quad r \to \infty \\ &\text{(b)}\ \int_0^r |P^{(N)}(r,r')|r'^{-1}(1+r')^{-1} \to 0, \quad r \to \infty \end{aligned}\right\}. \tag{1.20}$$

Let us now insert (1.19) into (1.8). (Recall that, in the following, we definitely substitute the couple $(0,V)$ to the couple (V,W) used at the beginning of this section and drop the indices whenever they are not necessary.) By going to the asymptotic limit, we obtain the remarkable formula

$$\int_0^\infty \mathcal{K}(\rho)u_l(\rho)\rho^{-1}d\rho = \exp(-il\pi/2)[1 - A_l\exp(i\delta_l)], \tag{1.21}$$

where

$$\mathcal{K}(\rho) = \rho^{-1}[\bar{k}(\rho) + ik(\rho)]. \tag{1.22}$$

We have furthermore proved in I[29] that the data of $\mathcal{K}(\rho)$ enable one to construct $K(r,r')$ and therefore $V(r)$. $\mathcal{K}(\rho)$, $K(r,r')$, and $V(r)$ are therefore equivalent quantities from the information point of view, and the scattering problems are completely described by the formula (1.21). So as to label the importance of $\mathcal{K}(\rho)$ for the scattering problems, we have called it the "scattering structure function." The inverse problem formally reduces to the determination of the "scattering structure function" from the phase shifts. A direct determination, starting from (1.21), will be studied as a generalized moment problem in a forthcoming paper, of a more mathematical character. In the present one, we still use $f(r,r')$ as an auxiliary tool; but its importance is essentially that of a gadget for obtaining the s.s. functions.

In the present paper, we describe a method of solution of the inverse problem which answers the questions (i)–(v) in a class of potentials dense in $\mathcal{U}$. This method is described in Secs. 2 and 3 below. In Sec. 2, we first thoroughly study the properties of $K(r,r')$ and $f(r,r')$ for potentials of class $\mathcal{E}$, particularly their asymptotic behavior, which can be expressed for both functions by formulas similar to (1.19). A fundamental equation is given, which relates the s.s. function and $f(r,r')$. We then look for a characterization of $f(r,r')$. $f(r,r')$ can be characterized as a solution of (1.10); but it can also be characterized by an integral representation in which the key function is a Fourier transform of $f(r,r')$. We therefore study the properties of this Fourier transform for potential of $\mathcal{E}$ and, in view of obtaining the c_l, the properties of the function obtained by truncating the Fourier spectrum of $f(r,r)$. Section 2 contains finally the working program of a solution of the inverse problem. This working program is achieved in Sec. 3, where a complete solution of the problem is given and its answers to questions (i)–(iv) are clearly stated. As for question (v), it will be studied apart in a forthcoming paper.[32]

In a forthcoming paper, the properties of the solutions here obtained will be fully investigated. In particular, the trace method will be applied to the solutions, so as to obtain results corresponding to those obtained in the other inverse problem of quantum mechanics.[33] Alternate methods of constructing solutions will also be given. In one of them, like in the Newton's method and its generalizations, the main figures will be the interpolation coefficients. Although this method can be considered as the largest possible extension of studies of this kind, it presents the same defects, viz the potentials are defined by special properties of the c_l and not by simple mathematical properties.

2. PROPERTIES OF $K(r,r')$ AND $f(r,r')$

A. Properties of $K(r,r')$

The properties of $K(r,r')$ have been thoroughly studied in our previous paper (I) for potentials of class $\mathcal{E}$. Many of them, actually, hold for potentials of larger classes. We only give here a few results, necessary for understanding the following.[34] Let us first introduce the notations

$$K_1(r,r') = K(r,r') - K_0(r,r'), \tag{2.1}$$

where

$$K_0(r,r') = -\tfrac{1}{2}(rr')^{1/2}\int_0^{(rr')^{1/2}} \times J_0\left[(r-r')\left(1-\frac{\rho^2}{rr'}\right)^{1/2}\right]\rho V(\rho)\,d\rho \tag{2.2}$$

and

$$K_2(r,r') = K(r,r') - K_0^0(r,r'), \tag{2.3}$$

where

$$K_0^0(r,r') = (2\pi)^{-1} V_0 \int_0^1 G(r,r',u)u^{-2}du, \tag{2.4}$$

$$V_0 = \int_0^\infty \rho V(\rho)d\rho, \tag{2.5}$$

$$G(r,r',u) = \cos[(r-r')^2 + 4rr'u^2)^{1/2}] - \cos(r-r'). \tag{2.6}$$

$K_0^0(r,r')$ can also be given the remarkable form

$$K_0^0(r,r') = -(\pi)^{-1} V_0\Big[u_0(r)u_0(r') + 2\sum_{l=1}^{\infty} u_l(r)u_l(r')\Big]. \tag{2.7}$$

The interest of the separation (2.1) comes both from the fact that the symmetric function $K_0(r,r')$ is that part of $K(r,r')$ which is linear in V, and from the fact that remarkable bounds and integral expressions are valid for $K_1(r,r')$, for instance,

$$K_1(r,r') = (4r^2)^{-1}[K(r,r)]^2 \sin(r-r') + \int_r^{r'} \sin(r'-\rho)Q(r,\rho)\rho^{-2}d\rho, \tag{2.8}$$

$$K_1(r,r') = -(4r'^2)^{-1}[K(r',r')]^2 \sin(r'-r) + \int_{r'}^{r} \sin(r-\rho)S(\rho,r')\rho^{-2}d\rho, \tag{2.9}$$

where $Q(r,r')$ and $S(r,r')$ are continuous on $R^+ \times R^+$, and absolutely bounded by

$$C(rr')^{1/2}(rr')^{(1+\epsilon)/2}[1 + (rr')^{(1+\epsilon)/2}]^{-1}. \tag{2.10}$$

Bounds and integral expressions have also been obtained for

$$K_0^1(r,r') = K_0(r,r') - K_0^0(r,r'). \tag{2.11}$$

Although all these results are to be used in Appendix A, we do not state them here more completely, but only give some features of the asymptotic behavior of $K(r,r')$, using, for the various quantities, indices which refer to that part of $K(r,r')$ they refer to. Components of the s.s. function are proved to be given by the following formulas, where $\mathcal{K}(r)$ is equal to $[k(r)\tau + ik(r)]r^{-1}$:

$$\mathcal{K}_0(r) = -(2\pi r)^{-1/2} \times \int_0^\infty \exp[-i(r + \tfrac{1}{2}\rho^2/r) + i\pi/4]\rho V(\rho)d\rho, \tag{2.12}$$

$$\mathcal{K}_1(r) = r^{-1}\Big\{(4r^2)^{-1}[K(r,r)]^2 \exp(-ir) + \int_r^\infty \exp(-i\rho)S(\rho,r)\rho^{-2}d\rho\Big\}, \tag{2.13}$$

$$\mathcal{K}_1(r) = (2\pi r)^{-1/2}\int_0^1 u^{-3/2}du \int_0^\infty \times \exp[-\tfrac{1}{2}i(\rho^2 u/r + ur + r/u) + i\pi/4] \times V(\rho)K(\rho,\rho u)d\rho, \tag{2.14}$$

$$\mathcal{K}_0^0(r) = -\pi^{-1}V_0\int_0^\pi \exp[-ir\cos\theta]\cos(\theta/2)d\theta. \tag{2.15}$$

$\mathcal{K}_2(r)$ is[29] a bounded continuous function on R^+ going to zero like r^{-1} as r goes to ∞, whereas $\mathcal{K}_0^0(r)$ asymptotically behaves like

$$\mathcal{K}_0^0(r) = -(2\pi r)^{-1/2}V_0 \exp(-ir + i\pi/4) + 0(r^{-1}). \tag{2.16}$$

Besides, we have obtained in (I) several bounds for the remainders coming in when $K(r,r')$ approaches its asymptotic behavior. If we write again (1.19) as

$$\left(1 - i\frac{\partial}{\partial r}\right)K(r,r') = -ir'\mathcal{K}(r')\exp(ir) + \mathcal{R}(r,r'), \tag{2.17}$$

the components of $\mathcal{R}(r,r')$ are bounded by

$$|\mathcal{R}_0^0(r,r')| \le Cr'|r-r'|^{-1/2}, \tag{2.18}$$

$$|\mathcal{R}_0'(r,r')| \le \begin{cases} C \\ Cr'/r + Cr'^2/r + Cr'(r'/r)^{\epsilon/2} \end{cases} \quad \text{for } r' \le \tfrac{1}{2}r, \tag{2.19}$$

$$|\mathcal{R}_1(r,r')| \le Cr'[1 + (rr')^{1/2}]^{-1}. \tag{2.20}$$

Other bounds are given in the Appendix A. It is convenient to call the first term of an expansion like (2.17) the structural term, the other one, the remainder and the whole, a structural expansion.

In our previous paper (I), we did not give the asymptotic behaviour of $K(r,r')$ for fixed r, large r'. This however, is very easy to obtain, since $K_0(r,r')$ is symmetric, whereas $K_1(r,r')$ is given as well by (2.9) (which was used in I) and by (2.8), which can be used in the same way for the present purpose. The results are given in the following formulas, where we use the primes for denoting the quantities related to this type of asymptotic behavior:

$$\mathcal{K}_0'(r) = \mathcal{K}_0(r), \tag{2.21}$$

$$\mathcal{K}_1'(r) = -(4r^3)^{-1}[K(r,r)]^2 \exp(-ir) - ir^{-1}\int_r^\infty \exp(-i\rho)Q(r,\rho)\rho^{-2}d\rho. \tag{2.22}$$

Besides, the bounds of the various components of $|\mathcal{R}'(r,r')|$ are similar to those of $|\mathcal{R}(r',r)|$.

B. Properties of $f(r,r')$

We study $f(r,r')$ for a potential of class $\mathcal{E}$. Actually most of the formulas given below are valid for a potential of $\mathcal{E}_{21}$, in particular, formulas (2.23), (2.24), and (2.25). $f(r,r')$ can be related to $K(r,r')$ through the formula[35]

$$f(r,r') = K(r,r') + \int_0^r K(\tau,r)K(\tau,r')\tau^{-2}d\tau. \tag{2.23}$$

It is easy to obtain another formula for $f(r,r')$ by solving (1.10) through Riemann's method, following the work we have done in I for $K(r,r')$. The result[36] is

$$f(r,r') = f[(rr')^{1/2},(rr')^{1/2}] - (rr')^{1/2}\int_0^{(rr')^{1/2}} \times \rho^{-1}f(\rho,\rho)\frac{\partial}{\partial\rho}J_0\Big[(r-r')\Big(1 - \frac{\rho^2}{rr'}\Big)^{1/2}\Big]d\rho, \tag{2.24}$$

which can also be written[37] as

$$f(r,r') = -\tfrac{1}{2}(rr')^{1/2}\int_0^{(rr')^{1/2}} \times J_0[(r-r')(1-\rho^2/rr')^{1/2}]\rho\omega(\rho)d\rho, \tag{2.25}$$

where

$$\omega(r) = -2r^{-1}\frac{d}{dr}[r^{-1}f(r,r)]. \tag{2.26}$$

Comparison of (2.2) and (2.24) (as well as a direct

verification) shows that $K_0(r,r')$ is a solution of (1.10) and that $f(r,r')$ is related to $\omega(r)$ like $K_0(r,r')$ to $V(r)$. Formulas (2.24) and (2.25) yield the continuation of $f(r,r)$ in $R^+ \times R^+$ consistent with the partial differential equation (1.10). The continuation of $K(r,r)$ in this way is $K_0(r,r')$. We can also say that (2.24) gives a *characterization* of $f(r,r')$ by the function $f(r,r)$, this meaning that the data of $f(r,r)$ is sufficient for constructing the solution $f(r,r')$ of (1.10), which for $r = r'$ is equal to $f(r,r)$. Such a characterization is convenient for a complete study of the problem, and we shall give, here and there, some remarks which enable the interested reader to easily do it. However, it is (slightly) more convenient to characterize also $f(r,r')$ by a Fourier transform of $f(r,r)$.

C. Characterization of $f(r,r')$ by a Fourier Transform

Let us first study $K_0(r,r')$. From the results of I, we can write

$$K_0(r,r') = \int_0^\infty G(r,r',u)\Phi'(u)du, \tag{2.27}$$

where $G(r,r',u)$ is given by (2.6) and $\Phi'(u)$ is given by

$$\Phi'(u) = (2\pi)^{-1}V_0 u^{-2}\gamma(1-u) + \Phi_1'(u) \tag{2.28}$$

where γ is the step function, and $\Phi_1'(u)$ is locally integrable on R^+, going to zero like $Cu^{-3-\epsilon}$ as u goes to ∞ (this for a potential of class $\mathcal{E}$). The first term in (2.28) is responsible for the fact that $r^{-1}K_0(r,r)(= r^{-1}K(r,r)]$ goes to the constant $-\frac{1}{2}V_0$ as r goes to ∞. From (2.27) we have also derived the formula for the s.s. function:

$$\mathcal{K}_0(r) = ir^{-1}\exp(-ir)\int_0^\infty [\exp(2iru^2) - 1]\Phi'(u)du. \tag{2.29}$$

From (2.28), we easily derive the formula

$$K_0(r,r') = 4rr'\int_0^\infty w^{-1}\sin wu\Phi(u)du, \tag{2.30}$$

where

$$w = [(r-r')^2 + 4rr'u^2]^{1/2}, \tag{2.31}$$

$$\Phi(u) = -\int_u^\infty \Phi'(t)dt = -(2\pi)^{-1}V_0(u^{-1}-1) \times \gamma(t-u) + \Phi_1(u). \tag{2.32}$$

For $r = r'$, (2.30) reduces to

$$r^{-1}K_0(r,r) = 2\int_0^\infty \sin(2ru)\Phi(u)du. \tag{2.33}$$

We would like to extend these results to $f(r,r')$. Let us consider the function

$$\bar{f}(r,r) = f(r,r) - K(r,r). \tag{2.34}$$

From (2.23) it is easy to prove (Appendix A) that $r^{-1}\bar{f}(r,r)$ is a bounded continuous function, going to zero faster than $Cr^{-1}\log r$ as r goes to ∞, and therefore belongs to $L_2(0,\infty)$. We therefore can define in $L_2(0,\infty)$ a function $\Phi_2(u)$ such that this integral converges in the mean l.i.m.):

$$r^{-1}\bar{f}(r,r) = 2 \text{ l.i.m.} \int_0^\infty \sin(2ru)\Phi_2(u)du. \tag{2.35}$$

The continuation $\bar{f}(r,r')$ of $\bar{f}(r,r)$ which fits (1.10) clearly is $f(r,r') - K_0(r,r')$. Let us consider the formula (2.24) for it, and use the Parseval theorem and (2.35). We obtain, through well-known formulas[38] that

$$\bar{f}(r,r') = 4rr' \text{ l.i.m.} \int_0^\infty w^{-1}(\sin w)u\Phi_2(u)du. \tag{2.36}$$

Comparison with (2.30) shows that, if $r^{-1}f(r,r)$ is given by its Fourier transform

$$r^{-1}f(r,r) = 2\int_0^\infty \sin(2ru)F(u)du, \tag{2.37}$$

the following formula defines a continuation which is consistent with (1.10):

$$f(r,r') = 4rr'\int_0^\infty w^{-1}\sin w\, uF(u)du. \tag{2.38}$$

Besides, it is easy to prove on (2.38) that, provided convergence conditions are fulfilled, (2.38) is a solution of (1.10) reducing to (2.37) as r' goes to r. In those conditions, which certainly hold for potentials of classes larger than $\mathcal{E}$, (2.23) and (2.38) are therefore equivalent continuations of $f(r,r)$. This equivalence between the two characterizations of $f(r,r')$ either through $f(r,r)$ or through the Fourier transform, cannot be overemphasized. They are also equivalent (for equivalent classes of potentials) with the "Loeffel" characterization by spectral data. In contrast, the coefficients c_l can be used only, as we see below, when $f(r,r)$ has a truncated Fourier spectrum. Actually, this remark explains why, in a paper which want to be as comprehensive as possible, we like better the characterization by the Fourier transform: The coefficients c_l will appear there in a most natural way. The price to be paid is the supplementary assumptions of differentiability of the potentials—assumptions which can be suppressed for most results, as we have seen in I, Sec. 7.

D. Asymptotic Behavior of $f(r,r')$

We already know the asymptotic behavior of the first term in the right-hand side of (2.23). For a study of the second term, it is convenient to insert

$$K(\tau,r) = K_0^0(\tau,r) + K_0^1(\tau,r) + K_1(\tau,r) \tag{2.39}$$

into

$$\int_0^r [K(\tau,r)K(\tau,r')]\tau^{-2}d\tau. \tag{2.40}$$

Then we write down the asymptotic behavior of every term as given by formulas of the form (2.17). It is proved in Appendix A that all the remainders are "negligible" functions, so that we obtain in this way,

$$\left(1 - i\frac{\partial}{\partial r}\right)f(r,r') = -ir'\mathcal{F}(r')\exp(ir) + \mathcal{T}(r,r'), \tag{2.41}$$

where $\mathcal{T}(r,r')$ is a "negligible function"; $\mathcal{F}(r')$ and $\mathcal{T}(r,r')$ are related to $\mathcal{K}(r')$ and $K(r,r')$ by

$$\mathcal{F}(r') = \mathcal{K}(r') + \int_0^\infty \mathcal{K}'(\tau)\mathcal{K}(\tau,r')d\tau, \tag{2.42}$$

$$\mathcal{K}(r,r') = (rr')^{-1}K(r,r'), \tag{2.43}$$

$$\mathcal{T}(r,r') = \mathcal{R}(r,r') + \int_0^r \mathcal{R}'(\tau,r)K(\tau,r')\tau^{-2}d\tau + i\exp(ir)\int_r^\infty \mathcal{K}'(\tau)K(\tau,r')\tau^{-1}d\tau - ir^{-2}K(r,r)K(r,r'). \tag{2.44}$$

Furthermore, we prove in Appendix A that, for any positive ϵ,

$$|\mathcal{F}(\rho) - \mathcal{K}(\rho)| < C(1+\rho^{\epsilon})(1+\rho)^{-1}. \tag{2.45}$$

It is convenient to label two terms in $\mathcal{F}(r')$ like we did for $\mathcal{K}(r')$:

$$\mathcal{F}(r') = \mathcal{F}_0^0(r') + \mathcal{F}_2(r'), \tag{2.46}$$

where $\mathcal{F}_0^0(r')$ is nothing but $\mathcal{K}_0^0(r')$.

E. Asymptotic Fundamental Equation

We are now in position for deriving the asymptotic form of the Regge–Newton equation:

$$\mathcal{K}(r,r') = \mathcal{F}(r,r') - \int_0^r \mathcal{K}(r,\rho)\mathcal{F}(\rho,r')d\rho, \tag{2.47}$$

where we have used, for convenience, the notations (2.3) and

$$\mathcal{F}(r,r') = (rr')^{-1}f(r,r'). \tag{2.48}$$

Replacing $K(r,r')$ and $f(r,r')$ by their expansions (2.17) and (2.14), and letting r go to infinity in (2.47), we easily check that the contribution of the negligible functions vanishes and we obtain

$$\mathcal{K}(r') = \mathcal{F}(r') - \int_0^{\infty} \mathcal{K}(\rho)\mathcal{F}(\rho,r')d\rho, \tag{2.49}$$

which supersedes the Regge–Newton equation (2.47) in all the situations where only the behavior for large r comes in. This equation is therefore most convenient for scattering studies. One must notice that (2.49) contains exactly as much information as (2.47), since it follows from the analysis done in I (Sec. 7) that $\mathcal{K}(r,r')$ can be derived from $\mathcal{K}(r')$. Besides, the same device can be used for deriving $\mathcal{F}(r,r')$ from $\mathcal{F}(r)$. Actually, if we write

$$f_2(r,r') = f(r,r') - K_0^0(r,r'), \tag{2.50}$$

it is clear that the s.s. function corresponding to $f_2(r,r')$ is $\mathcal{F}_2(r')$.

Now, since $f_2(r,r')$ is a solution of the partial differential equation (1.10), just like $K_2^0(r,r')$ which we have studied in I, we can readily write, following [I, (1.26)],

$$f_2(r,r') = (2\pi)^{-1/2}(rr')^{1/2}\int_0^{\infty}[f_2(\rho)\sin(\varphi(\rho)) - \bar{f}_2(\rho) \times \cos(\varphi(\rho))]\rho^{-3/2}d\rho, \tag{2.51}$$

where

$$\varphi(\rho) = \tfrac{1}{2}\rho(rr'/\rho^2 + r/r' + r'/r) + \pi/4. \tag{2.52}$$

The formulas (2.42) and (2.49), together with the formulas relating $\mathcal{K}(r,r')$ to $\mathcal{K}(r')$ and $\mathcal{F}(r,r')$ to $\mathcal{F}(r')$ prove that the information contained in $\mathcal{F}(r)$ and that in $\mathcal{K}(r)$ are equivalent, and are also equivalent with that contained in $\mathcal{F}(r,r')$ or $\mathcal{K}(r,r')$. This holds provided that the integral equations there involved has a unique solution.

F. Further Characterization of $f(r,r')$

Formula (2.49) can be used to construct s. s. functions from the knowledge of $\mathcal{F}(r,r')$. $\mathcal{F}(r,r')$ can be constructed either from $F(u)$ through (2.38) or from $f(r,r)$ through (2.24). Actually, the knowledge of $\Phi(u)$ and $\bar{f}(r,r)$ or $\bar{\Phi}(u)$ and $\Phi_2(u)$ are sufficient for a characterization of $f(r,r')$. If we allow $r^{-1}\bar{f}(r,r)$ or $\Phi_2(u)$ to be *any* function of $L_2(0,\infty)$, we obtain a class of functions $\mathcal{F}(r,r')$, say, $\bar{\mathcal{F}}$, which contains the class associated with the class of potentials $\mathcal{E}$. However, $\bar{\mathcal{F}}$ is much too large for our purpose, and not convenient for our foregoing derivations. We therefore try to get more precise limitations on $\omega(r)$ and $\Phi_2(u)$ for a potential of class $\mathcal{E}$. From (2.23) and (2.26), we obtain

$$\omega(r) = V(r) + 2r^{-3}\int_0^r [K(\rho,r)]^2\rho^{-2}d\rho - 2r^{-4}[K(r,r)]^2 - 4r^{-2}\int_0^r K(\rho,r)\left(\frac{\partial}{\partial r}K(\rho,r)\right)\rho^{-2}d\rho. \tag{2.53}$$

For a potential in $\mathcal{E}$, we know that $|K(\rho,r)|$ is bounded by $C(\rho r)^{1+\epsilon/2}$, whereas $|(\partial/\partial r)K(\rho,r)|$ is bounded by $C\rho$. Hence, if we introduce $\omega_1(r)$ equal to $\omega(r) - V(r)$,

$$|\omega_1(r)| \le Cr^{\epsilon}(1+r^{\epsilon}). \tag{2.54}$$

The asymptotic behavior of $\omega(r)$ is derived in Appendix A, together with a careful appraisal of the remainder. The result is

$$|\omega_1(r) + 4r^{-2}[\alpha\cos(2r) + \beta\sin(2r)]| < Cr^{-3+\epsilon}, \tag{2.55}$$

where

$$\alpha = \int_0^{\infty}\bar{k}'(\rho)k'(\rho)\rho^{-2}d\rho, \tag{2.56}$$

$$\beta = \tfrac{1}{2}\int_0^{\infty}\{[\bar{k}'(\rho)]^2 - [k'(\rho)]^2\}\rho^{-2}d\rho. \tag{2.57}$$

From (2.54) and (2.55), we see that $r\omega(r)$ is a function of $L_2(0,\infty)$. Let us define its cosine Fourier transform $u\bar{\Phi}_2(u)$:

$$u\bar{\Phi}_2(u) = -(\pi)^{-1}\,\text{l.i.m.}\int_0^{\infty}\cos(2ru)r\omega_1(r)dr, \tag{2.58}$$

which is equivalent to

$$\frac{d}{dr}[r^{-1}\bar{f}(r,r)] = 4\,\text{l.i.m.}\int_0^{\infty}\cos(2ru)\bar{\Phi}_2(u)udu. \tag{2.59}$$

Let us now introduce

$$A(r) = 4r^{-1}[\alpha\cos 2r\gamma(r-a) + \beta\sin 2r\gamma(r-b)], \tag{2.60}$$

where γ is the Heaviside step function, a and b are conveniently chosen positive numbers. Clearly,

$$|r\omega_1(r) + A(r)| < Cr^{-2+\epsilon}. \tag{2.61}$$

Let $\bar{A}(u)$ the cosine Fourier transform of $A(r)$ be

$$\bar{A}(u) = \int_0^{\infty}\cos(2ru)A(r)dr. \tag{2.62}$$

After one integration by parts, we obtain for $\bar{A}(u)$ the following bound:

$$|\bar{A}(u)| < C(1+|1-u|)^{-1}(1+\log|1-u|). \tag{2.63}$$

From (2.58), it follows that $u\bar{\Phi}_2(u)$ is a function of $L_2(0\infty)$. From (2.55) and (2.63), it follows that this function is continuous on R^+, except for a logarithmic singularity for $u=1$. Let now $B(r)$ be the function

$$B(r) = \int_r^{\infty}A(\rho)d\rho. \tag{2.64}$$

$B(r)$ is a bounded function, which, for r going to infinity, goes to zero like

$$B(r) = 2r^{-1}(\beta \cos 2r - \alpha \sin 2r) + r^{-2}(\alpha \cos 2r - \beta \sin 2r) + 0(r^{-3}). \quad (2.65)$$

Let us now analyze (2.58). The value of the right-hand side is the limit, with respect to the norm in $L^2(0,\infty)$, of

$$\Phi_M = -\pi^{-1}\int_0^M \cos(2ru) r\omega_1(r)dr. \quad (2.66)$$

Let us integrate (2.66) by parts. Since $r^{-1}\bar{f}(r,r)$ goes to zero as r goes to zero and to ∞, we can write

$$r^{-1}\bar{f}(r,r) = -B(r) + \int_r^\infty [\rho\omega_1(\rho) + A(\rho)]d\rho \quad (2.67)$$

and, therefore,

$$\bar{\Phi}_M = -(\pi M)^{-1}\bar{f}(M,M) + 2\pi^{-1}u\int_0^M \sin(2ru) r^{-1}\bar{f}(r,r)dr. \quad (2.68)$$

It is easy to see that $\bar{\Phi}_M$ is a sequence convergent in $L^2(0,\infty)$ as M goes to ∞. Its limit is equal to

$$\bar{\Phi}_2(u) = 2\pi^{-1}\,\text{l.i.m.}\int_0^\infty \sin(2ru) r^{-1}\bar{f}(r,r)dr. \quad (2.69)$$

From (2.69) and (2.35), it follows that $\bar{\Phi}_2(u)$ and $\Phi_2(u)$ are equal almost everywhere. The function $\Phi_2(u)$ which characterizes $\bar{f}(r,r')$ through (2.35) has therefore, if the potential belongs to $\mathcal{E}$, the following properties:

(a) $\Phi_2(u)$ belongs to $L_2(0,\infty)$;

(b) $\Phi_2(u)$ is almost everywhere continuous;

(c) $u\bar{\Phi}_2(u)$ belongs to $L_2(0,\infty)$, and is continuous for any u but $u = 1$.

Since $\Phi_1(u)$ shows also those properties, so does

$$\psi(u) = \Phi_1(u) + \Phi_2(u) = F(u) + (2\pi)^{-1}V_0(u^{-1} - 1)\gamma(1-u) \quad (2.70)$$

The set $\mathfrak{F}$ of functions $f(r,r')$ characterized by giving a number V_0 and a function $\psi(u)$ contains the set of functions $f(r,r')$ corresponding to potentials of $\mathcal{E}$. $\mathfrak{F}$ is narrower than $\bar{\mathfrak{F}}$, and exhibits an interesting property: It is easy to study "truncated Fourier spectrum" approximations in $\mathfrak{F}$, viz. the functions obtained from a given element of $\mathfrak{F}$ by restricting $F(u)$ to a finite interval.

G. Projected Functions

Among the approximations described above, a particularly important one is obtained by restricting $F(u)$ to $[0,1]$. We call it the projection of $f(r,r')$. It yields two projected functions: the internal projection, whose spectrum is confined to $[0,1]$, and the external projection, whose spectrum is confined to $(1,\infty)$. Their sum is $f(r,r')$. Now, $f(r,r')$, according to (2.25), can be written as the sum of two terms, corresponding to two parts of $\omega(r)$ [given in (2.53)]. The first term $K_0(r,r')$ corresponds to $V(r)$. The second one, which we label $f_1(r,r')$, corresponds to the other terms in (2.53):

$$f(r,r') = K_0(r,r') + f_1(r,r'). \quad (2.71)$$

After a glance at the results of Paper I, recalled in Appendix A, we see that the interval projection $K_0^I(r,r')$ of $K_0(r,r')$ is readily obtained by substracting $\Phi_4(u)$ from $\Phi'(u)$ in (2.27), and adding to the function obtained in this way a quantity of the form $\gamma \sin r \sin r'$, so as to balance the integration by parts [for more details, see formula (2.92) below]. Therefore, $K_0^I(r,r')$ has a structural expansion and so does the external projection $K_0^E(r,r')$. Furthermore, the structure function $\mathcal{K}_0^E(\rho)$ is a continuous function on R_+, going to zero like ρ^{-1} as ρ goes to infinity.

Using now a well-known formula,[39] we obtain from (2.25) the following formula for $f_1(r,r')$:

$$f_1(r,r') = -2\pi^{-1}rr'\int_0^\infty \rho\omega_1(\rho)d\rho \times \int_0^\infty w^{-1}\sin w \cos(2\rho u)du, \quad (2.72)$$

so that the internal projection $f_1^I(r,r')$ is given by

$$f_1^I(r,r') = -2\pi^{-1}rr'\int_0^\infty \rho\omega_1(\rho)d\rho\int_0^1 w^{-1}\sin w\cos(2\rho u)du. \quad (2.73)$$

It follows from the definitions of $\omega_1(r)$ and (2.72) that $f_1(r,r)$ is equal to $\bar{f}(r,r)$. We prove in Appendix A that it is possible to define a and b in (2.60) in such a way that $B(0)$ be equal to zero. Let us do this choice, and set

$$\tilde{f}(r,r) = f_{10}(r,r) + f_{11}(r,r) \quad (2.74)$$

with

$$f_{10}(r,r) = -rB(r). \quad (2.75)$$

Hence, $(d/dr)r^{-1}f_{10}(r,r)$ is equal to $A(r)$, and therefore vanishes as r goes to zero. Let us now successively study the contributions of $f_{10}(r,r)$ and $f_1(r,r)$ to (2.73). For the first one, integrating by parts yields

$$4\pi^{-1}rr'\int_0^\infty \rho^{-1}f_{10}(\rho,\rho)d\rho\int_0^1 w^{-1}\sin w\sin(2\rho u)udu. \quad (2.76)$$

It is easy to see that a permutation of the integrations is possible. The function $\int_0^\infty B(\rho)\sin(2\rho u)d\rho$, which appears by the way, can be split into two parts:

$$\int_0^A B(\rho)\sin(2\rho u)d\rho, \quad (2.77)$$

which enables us a second integration by parts, obtaining

$$2\pi^{-1}\Big(\sin r \sin r'\int_0^\infty B(\rho)\sin(2\rho)d\rho + \int_0^1[\cos w - \cos(r-r')]du\int_0^A B(\rho)\rho\cos(2\rho u)d\rho\Big). \quad (2.78)$$

For the second one, we calculate directly $\int_A^\infty B(\rho)\sin 2\rho u d\rho$ and its first derivative with respect to ρ. It follows readily from (2.65) that

$$\int_A^\infty B(\rho)\sin 2\rho u d\rho = 2\int_A^\infty \rho^{-1}(\beta\cos 2\rho - \alpha\sin 2\rho)\sin(2\rho u)d\rho + \int_A^\infty \rho^{-2}(\alpha\cos 2\rho - \beta\sin 2\rho)\sin(2\rho u)du + \int_A^\infty \bar{B}(\rho)\sin(2\rho u)d\rho, \quad (2.79)$$

where

$$|\bar{B}(\rho)| < C\rho^{-3}. \quad (2.80)$$

The two first terms in (2.79) can easily be proved to be bounded functions on $[0,1]$, together with their first derivative, except for $u = 1$, where the first term has a logarithmic singularity, and its derivative exhibit the singularity $-\alpha(1-u)^{-1}$. As for the third term, it is obviously bounded together with its first derivative. The contribution of $\int_A^\infty B(\rho)\sin(2\rho u)d\rho$ to (2.76) can therefore be written, after one integration by parts,

$$\pi^{-1}\int_0^1 [\cos w - \cos(r+r')]\psi_1(u)du, \tag{2.81}$$

where

$$\psi_1(u) = \frac{d}{du}\left(\int_A^\infty B(\rho)\sin(2\rho u)d\rho\right). \tag{2.82}$$

The asymptotic behavior of (2.78) can be studied just like we did in Paper I and have partially recalled in A. The method essentially consists in replacing $[\cos w - \cos(r-r')]$ by $\{\cos[r + r'(2u^2-1)] - \cos(r-r')\}$, and proving that the remainders are negligible [the conditions fulfilled by the function of u in (2.78) are those fulfilled by $\Phi_3(u)$]. It follows that the contribution to the complex scattering structure function due to (2.78) is bounded in modulus by $C(1+r')^{-1}$.

The formula (2.81) can be studied in a very similar way, replacing again $\cos w$ by $\cos[r + r'(2u^2-1)]$, so as to obtain the asymptotic behavior of (2.81):

$$-2\pi^{-1}\cos r\int_0^1 \sin[r'(u^2-1)]\sin[r'u^2]\psi_1(u)du$$

$$+2\pi^{-1}\sin r\int_0^1 \sin[r'(u^2-1)]\cos[r'u^2]\psi_1(u)du. \tag{2.83}$$

As for the remainders, we know from I, Sec. 5, that they are bounded by

$$C\int_0^1 \sup\left[\left|\sin w\left(1-\frac{\dot{w}}{2w}\right)\right| + C\,|r-r'-w+2r'u^2|\,;C\right] \times |\psi(u)|\,du \tag{2.84}$$

and, therefore, if we refer to [I, (5.22)], by

$$\int_0^1 \sup[C, Cr'^2(r+r')^{-1}\,|1-u^2|]\,|\psi_1(u)|\,du. \tag{2.85}$$

It is easy to see that the remainders are therefore bounded by Cr'^2/r and $C(r'^2/r)^\epsilon$, and negligible. As for the s.s. function resulting from (2.83), it follows from the fact that $|\psi_1(u)|$ is the sum of an integrable function and of $-\alpha(1-u)^{-1}$ that the complex s.s. function has its modulus bounded by $C(1+r'^\epsilon)(1+r')^{-1}$. This can easily be proved for instance, on the real part, which according to (2.83), is bounded by

$$Cr'^{-1}\int_0^1 |\sin r'(u^2-1)|\,|u-1|^{-1}du < C \text{ and}$$

$$< Cr'^{-1}\left(\int_0^{1-\alpha}(1-u)^{-1}du + \int_{1-\alpha}^1 r'du\right) \tag{2.86}$$

by choosing α equal to $(r')^{-1+\epsilon}$ (for large r').

Therefore, the total contribution of $f_{10}(r,r)$ to the s.s. function of the internal projected function of $f_1(r,r')$ is bounded by $C(1+r'^\epsilon)(1+r')^{-1}$. Let us now study the contribution of $f_{11}(r,r)$. It follows from the definition of $B(r)$ and from (2.67) and (2.61) that $r^{-1}f_{11}(r,r)$ is a function of $L_2(0,\infty)$, going to zero as r goes to zero and as r goes to infinity. Furthermore, the function

$$\omega_2(r) = r^{-1}\frac{d}{dr}r^{-1}f_{11}(r,r) \tag{2.87}$$

is a bounded function of r, going to zero like $Cr^{-3+\epsilon}$ as r goes to infinity. With these conditions, it is proved in Appendix B that the function associated with the internal projected function of $f_{11}(r,r')$ exists, and is bounded by $C(1+r'^\epsilon)(1+r')^{-1}$.

Hence we have proved that the continuation of $\bar{f}(r,r)$ can be projected and that its internal projection has a structural expansion. Furthermore, the s.s. function $\mathfrak{F}_1^I(\rho)$ which is associated with this projection is bounded by

$$|\mathfrak{F}_1^I(\rho)| < C(1+\rho^\epsilon)(1+\rho)^{-1}. \tag{2.88}$$

If now we refer ourselves to (2.45), to (2.46), to the results obtained for $K_0^I(r,r')$, and to (2.71), we see that $\mathfrak{F}^I(\rho)$ can be written as the sum of two terms; the first one, which is nothing but $\mathfrak{F}_0^0(\rho)$ being associated with $K_0(r,r')$ only, whereas the second one, associated both with $K_0(r,r')$ and $f_1(r,r')$, is absolutely bounded like (2.88). Since the first one contains $\mathfrak{F}_0^0(\rho)$, it follows that the whole *external projection* of $f(r,r')$ is bounded like (2.88):

$$|\mathfrak{F}^E(\rho)| < C\,(1+\rho^\epsilon)(1+\rho)^{-1}. \tag{2.89}$$

This condition, which is necessarily fulfilled by a function $f(r,r')$ corresponding to a potential of class $\mathcal{E}$ but is not sufficient for guarantee that V is in $\mathcal{E}$, is *very important* for the analysis done below. We call $\mathfrak{F}^*$ the set of functions $f(r,r')$ whose external projection has a s.s. function fulfilling (2.89) and whose internal projection has a s.s. function sum of a term $\mathfrak{K}_0^0(\rho)$ and a term fulfilling (2.89). Any s.s. function of $\mathfrak{F}^*$ is therefore given by giving V_0 and a function fulfilling a bound of the form (2.89). Clearly the class $\mathfrak{T}_0$ corresponding to the potentials of class $\mathcal{E}$ is a subset on $\mathfrak{F}\cap\mathfrak{F}^*$, which is itself contained in $\mathfrak{F}$.

As a last remark on projected functions, let us notice that it is not completely equivalent to define the internal projection of $f(r,r')$ [given by (2.38)] through the formula

$$f^I(r,r') = 4rr'\int_0^1 w^{-1}\sin wuF(u)du, \tag{2.90}$$

which has been used in the definition above, or through the formula, which has generally no meaning

$$I = \int_0^1 [\cos w - \cos(r-r')]F'(u)du. \tag{2.91}$$

Actually,

$$I = f^I(r,r') - 2\sin r\sin r'F(1) \text{ if } F(1) \text{ is finite.} \tag{2.92}$$

The difference is obviously itself a structural expansion and its s.s. function fulfills the bound (2.89).

H. Bounds of $f^E(r,r')$

Let us now study the absolute bounds of the external projection $f^E(r,r')$ for a potential of class $\mathcal{E}$. From (2.71) it follows that we successively have to study $K_0^E(r,r')$ and $fw(r,r')$. $K_0^E(r,r')$ is equal to

$$K_0^E(r,r') = 4C_0rr'\int_1^\infty \omega^{-1}\sin wu\Phi(u)du, \tag{2.93}$$

where

$$\Phi(u) = -\int_u^\infty \Phi_4(t)dt, \tag{2.94}$$

the function $\Phi_4(t)$, given in Appendix A, is absolutely bounded by $Cu^{-3-\epsilon}$. Since $|w^{-1}\sin w|$ is bounded by $(1+w)^{-1}$, and therefore by $(1+r+r')^{-1}$, we get for $|K_0^E(r,r')|$ the bound $C(1+r+r')^{-1}rr'$.

Integration by parts of (2.93) yields a constant absolute bound.

Now, from (2.72) and (2.73) we obtain

$$f_1^E(r,r') = -2\pi^{-1}rr'\int_1^\infty w^{-1}\sin w du \times \int_0^\infty \rho\omega_1(\rho)\cos(2\rho u)du, \tag{2.95}$$

where $\rho\omega_1(\rho)$ is a bounded function, whose asymptotic behavior is given by (2.61). Let us successively study the contributions to (2.95) of $A(r)$ and of the remainder. For the first one we are led to terms of the form

$$rr'\int_1^\infty w^{-1}\sin w du\int_0^\infty \cos 2\rho\cos 2\rho u\rho^{-1}d\rho \tag{2.96}$$

and, therefore, after one integration by parts, to terms of the form

$$\int_1^\infty[\cos w - \cos(r+r')]\frac{d}{du}\left(u^{-1}\int_{a(u-1)}^\infty \rho^{-1}\cos 2\rho d\rho\right), \tag{2.97}$$

and the same term with $u+1$ instead of $(u-1)$. This leads one to a constant bound, except for the contribution of

$$\int_1^\infty[\cos w - \cos(r+r')][(u-1)u]^{-1}\cos[2a(u-1)]du, \tag{2.98}$$

which yields a bound $C(1+\log[1+rr'/(r+r')]$. On the other hand, since the ρ integral in (2.96) is easily shown to be absolutely bounded by $C\log(u-1)$ for any u, and by $C(u-1)^{-1}$ for large u, we get from (2.96) the bound $C(rr')^{1-\epsilon/2}(1+r+r')^{1-\epsilon}$

We have now to study the contribution of the remainder. We have shown in Appendix A that $(d/d\rho)(\rho\omega_1(\rho))$ is a function of $L^2(0,\infty)$. Let us denote by $\rho\omega_2(\rho)$ the remainder $[\rho\omega_1(\rho)+A(\rho)]$. Clearly $\rho^2\omega_2(\rho)$, $\rho\omega_2(\rho)$, and $(d/d\rho)\rho\omega_2(\rho)$ are functions of $L^2(0,\infty)$.

Using the Parseval theorem, it is easy to show that

$$\begin{aligned} rr'\int_1^\infty w^{-1}\sin w\, du\int_0^\infty \rho\omega_2(\rho)\cos(2\rho u)d\rho \\ = rr'\int_1^\infty \sin w du\int_0^T \rho\omega_2(\rho)\cos(2\rho u)d\rho + O(T^{-3/2+\epsilon}) \\ = -\tfrac{1}{2}rr'\int_1^\infty w^{-1}u^{-1}\sin w\, du\int_0^\infty \frac{d}{d\rho} \\ \times(\rho\omega_2(\rho))\sin(2\rho u)d\rho + O(T^{-1/2+\epsilon}). \end{aligned} \tag{2.99}$$

The ρ integral in the last expression is a function of u belonging to $L_2(0,\infty)$. The Schwarz inequality yields then readily the bound $Crr'(1+r+r')^{-1}$. With the same care, it is possible to integrate (2.99) by parts and obtain

$$\begin{aligned} -\tfrac{1}{2}\sin r\sin r'\int_0^\infty \rho\omega_2(\rho)\cos(2\rho)d\rho \\ -\tfrac{1}{4}\int_1^\infty[\cos w - \cos(r-r')]du \\ \times\int_0^\infty \rho\omega_2(\rho)\frac{d}{du}(u^{-1}\cos 2\rho u)d\rho \end{aligned} \tag{2.100}$$

from which the Schwarz inequality give a constant bound.

Gathering the bounds we have obtained, we can write

$$|\mathfrak{F}^E(r,r')| < \begin{cases} C(1+r+r')^{-1+\epsilon}(rr')^{-\epsilon/2} \\ C(rr')^{-1}\{1+\log[1+rr'/(r+r')]\} \end{cases}. \tag{2.101}$$

Taking care successively of the various cases defined by the positions of r, r', and rr' with respect to 1, we easily derive from (2.101) the more interesting bound

$$|\mathfrak{F}^E(r,r')| \le Cr^{-\epsilon/2}(1+r)^{-1+\epsilon}(r')^{-\epsilon/2}(1+r')^{-1+\epsilon}. \tag{2.102}$$

We denote by $\mathfrak{F}^{**}$ the class of functions $f(r,r')$ whose external projection fulfills (2.102). Two conditions are separately sufficient to guarantee that a function $f(r,r')$ belongs to $\mathfrak{F}^{**}$: either it corresponds to a potential of class $\mathcal{E}$; or $F(u)$ is locally integrable and goes to zero faster than $Cu^{-3-\epsilon}$ as u goes to infinity. This last condition is, actually, more restrictive than the first one, since we have seen that less restrictive asymptotic behavior of $F(u)$ come from $A(r)$.

I. The Five Requirements of a Solution

After the above "preparation," it is useful to define now precisely what we expect to obtain for being allowed to say the inverse problem is *completely* solved for the class $\mathcal{E}$ of potentials. Obviously we have such a complete solution if:

(1) Being given any sequence $\delta = \{\delta_l\}$ of phase shifts belonging to a certain sequences space Δ, we know a method $\mathfrak{M}$ for constructing a solution $f(r,r')$ of (1.10) such that (1.21) *and* (2.49) hold.

(2) The class $\mathfrak{F}_1$ of functions $f(r,r')$ obtained from the sequences δ_n of Δ contains the class $\mathfrak{F}_0$ of functions $f(r,r')$ corresponding to the potentials of $\mathcal{E}$.

(3) The space Δ contains the set Δ_0 of sequence of phase shifts corresponding to potentials of $\mathcal{E}$.

(4) Being given the sequence δ_0 of phase shifts corresponding to a potential V_0 of $\mathcal{E}$, V_0 does belong to the set of solutions obtained from δ_0 through $\mathfrak{M}$.

(5) A consequence of (2) and (4) is that the set $\mathfrak{V}_0$ of potentials obtained from all the sequences of Δ through $\mathfrak{M}$ contains $\mathcal{E}$. So as to obtain such a solution, first we construct the method $\mathfrak{M}$ for $\mathfrak{F}_1$ and Δ, then we check successively the five requirements. This program is now done in Sec. 3.

3. A COMPLETE SOLUTION OF THE INVERSE PROBLEM

So as to define the method $\mathfrak{M}$, we want to construct solutions $f(r,r')$ of (1.10) such that (1.21) and (2.49) hold. Now, a function $\mathfrak{F}(r,r')$ of $\mathfrak{F}$ is characterized either by $F(u)$ or by $\mathfrak{F}(r,r')$ and any of these characterizations, when used in (2.24) or (2.38), is sufficient to guarantee that $f(r,r')$ is a solution of (1.10), so that the whole Regge–Newton–Loeffel formalism applies. For obtaining a convenient method, we need to take (in the direct problem) or to obtain (in the inverse problem) $\mathfrak{F}(r,r')$ in a subclass $\mathfrak{F}_2$ of $\mathfrak{F}$ such that the following are true.

Assumption a: $\mathfrak{F}_2$ contains the subclass $\mathfrak{F}_0$ corresponding to the potentials of class $\mathcal{E}$.

Assumption b: Any function of $\mathfrak{F}_2$ has a structural expansion of the form (2.41), and its $s.s.$ function can be written in the form (2.46), with the bounds described at this point for its various components.

Assumption c: For any function of $\mathfrak{F}_2$, we can define structurable internal and external projections The external $s.s.$ function fulfills the bound (2.89). The internal $s.s.$ function fulfills a bound obtained by combining the bound of $\mathcal{K}_0(\rho)$ for a potential of $\mathcal{E}$ and (2.45) (or 2.88), and is therefore absolutely bounded by $C(1+\rho^{1/2})(1+\rho)^{-1}$, or $C(1+\rho)^{-1/2}$. The external projection fulfills the bound (2.102).

Assumption d: The representation (2.38) holds for any function of $\mathfrak{F}$, and $F(u)$ is the sum of $\Phi(u)$ defined by (2.32) and $\Phi_2(u)$ defined by (2.35). For any function of $\mathfrak{F}_2$, $\Phi_1(u)$ is continuous of R^+ and absolutely bounded by $Cu^{-2-\epsilon}$; $\Phi_2(u)$ and $u\Phi_2(u)$ belong to $L_2(0,\infty)$; $u\Phi_2(u)$ is continuous on R^+, except at $u=1$, for which it has a logarithmic singularity.

Assumption e. $\Phi_1(u)$ and $\Phi_2(u)$ are differentiable on $]0,1[$. $\Phi_1'(u)$ is continuous in $]0,1]$ and absolutely bounded by $Cu^{\epsilon-1}$. $\Phi_2'(u)$ is continuous on $]0,1[$. It is absolutely bounded by $Cu^{-\epsilon-1}$ as u goes to zero, diverges like $C_0(1-u)^{-1}$ as u goes to 1. Besides, $\Phi_1'(u)$ is continuous on $[1,\infty]$ and absolutely bounded by $Cu^{-3-\epsilon}$. $\Phi_2'(u)$ diverges like $C_0(u-1)^{-1}$ as u goes to 1 and $(u-1)\,\Phi_2'(u)$ is a function of $L_2(1,\infty)$.

Clearly Conditions (d) and (e) are weaker on Φ_2 than on Φ_1, so that the important conditions for $F(u)$ are those on Φ_2. It is only for making easier the reference to Sec. 2 that we split, here artificially, $\Phi_1(u)$ and $\Phi_2(u)$.

It follows from the results of Sec. 2, hereabove recalled, that Conditions (b), (c), and (d), are fulfilled by functions $f(r,r')$ associated with potentials of class $\mathcal{E}$. As for Condition (e), the part concerning $\Phi_1'(u)$ follows readily from the bound [I, (3.20)] derived in our first paper for a potential of class $\mathcal{E}$. The second part can be studied on (2.58), which we rewrite, using (2.60):

$$\Phi_2(u) = -(\pi u)^{-1}\int_0^\infty \cos(2ru)r\omega_2(r)dr + (\pi u)^{-1}\int_0^\infty \cos(2ru)A(r)dr. \qquad (3.1)$$

In (3.1) the first integral is, according to (2.55), absolutely convergent. The second integral is semi-convergent, so that we do not need the l.i.m. Denoting now the first part in the right-hand side of (3.1) by $\Phi_2^{(1)}(u)$, and using the facts that $r\omega_2(r)$ goes to zero as r goes to zero, that it is absolutely bounded by $Cr^{-2+\epsilon}$, and that its integral from 0 to ∞ vanishes, we easily obtain

$$\frac{d}{du}\Phi_2^{(2)}(u) = \text{l.i.m.}\,2(\pi u)^{-1}\int_0^\infty \sin(2ru)r^2\omega_2(r)dr + (\pi u^2)^{-1}\int_0^\infty [\cos(2ru)-1]r\omega_2(r)dr, \qquad (3.2)$$

which yields a contribution[1] absolutely bounded by $Cu^{-1}L_{1/c}(u)$. On the other hand, it is easy to obtain

$$\frac{d}{du}\Phi_2^{(2)}(u) = -2\pi^{-1}\frac{d}{du}\left(\int_0^A \sin(2ru)B(r)dr + \int_A^\infty \sin(2ru)B(r)dr\right). \qquad (3.3)$$

Clearly the first term in the right-hand side of (3.3) yields a bounded contribution. The second one has been studied under formula (2.82) and yields the divergence $-2\pi^{-1}\alpha(1-u)^{-1}$. It follows from (3.2) that $u(d/du)\Phi_2^{(1)}(u)$ is a function of $L_2(1,\infty)$. On the other hand, from (3.3), we obtain

$$\frac{d}{du}(\Phi_2^{(2)}(u)) = 4(\pi u)^{-1}\int_0^\infty \cos(2ru)A(r)dr. \qquad (3.4)$$

Standard manipulations of the right-hand side of (3.4) show that it is absolutely bounded by $Cu(u-1)^{-1}$. It is therefore *possible* to define a class $\mathfrak{F}_2$ fulfilling simultaneously the properties (a), (b), (c), (d), and (e). Besides, it is easy to find conditions which are *sufficient* to guarantee that a function $\mathfrak{F}(r,r')$ defined through (2.38) belongs to $\mathfrak{F}_2$. Actually, this goal is obviously attained if $F(u)$ fulfills the properties given for $\Phi(u)$, viz for $\Phi_1(u) - (2\pi)^{-1}V_0u^{-1}\gamma(1-u)$. It is also attained for any function constructed through (2.25) from a function $\omega(\rho)$ in $\mathcal{E}$. But the classes $\mathfrak{F}_2$ and $\mathfrak{F}_2^*$ thus defined do not contain $\mathfrak{F}_0$.

We now develop the method $\mathfrak{M}$ in both the direct and the reserve sense. Throughout the direct study, we explicitly assume that $\mathfrak{F}(r,r')$ belongs to $\mathfrak{F}_2$. Let $\mathfrak{F}^I(r,r')$ and $\mathfrak{F}^E(r,r')$ be the internal and the external projections of $\mathfrak{F}(r,r')$ and, $\mathfrak{F}^I(\rho)$ and $\mathfrak{F}^E(\rho)$ their structure functions. Let us *define* $\mathcal{K}^I(r)$ and $\mathcal{K}^E(r)$ by the formulas

$$\mathcal{K}^I(r) = \overline{\mathfrak{F}}^I(r) - \int_0^\infty \mathcal{K}(\rho)\overline{\mathfrak{F}}^I(\rho,r)d\rho, \qquad (3.5)$$

$$\mathcal{K}^E(r) = \overline{\mathfrak{F}}^E(r) - \int_0^\infty \mathcal{K}(\rho)\overline{\mathfrak{F}}^E(\rho,r)d\rho, \qquad (3.6)$$

where

$$\overline{\mathfrak{F}}^I(\rho,r) = \mathfrak{F}^I(\rho,r) + \delta\,\sin\rho\,\sin r(\rho r)^{-1}, \qquad (3.7)$$

$$\overline{\mathfrak{F}}^E(\rho,r) = \mathfrak{F}^E(\rho,r) - \delta\,\sin\rho\,\sin r(\rho r)^{-1}. \qquad (3.8)$$

The parameter δ is to be defined conveniently later. $\overline{\mathfrak{F}}^I(r)$ and $\overline{\mathfrak{F}}^E(r)$ are the $s.s.$ functions of $\overline{\mathfrak{F}}^I(r,r')$ and $\overline{\mathfrak{F}}^E(r,r')$. Their relations with $\overline{\mathfrak{F}}I(r)$ and $\mathfrak{F}^E(r)$ is obviously very simple.

Clearly, (2.49) is satisfied if and only if

$$\mathcal{K}^E(r) + \mathcal{K}^I(r) = \mathcal{K}(r). \qquad (3.9)$$

The idea of the method given here is to choose arbitrarily in $\mathfrak{F}_2$ the function $\mathfrak{F}_E(r,r')$. This can be done, for example, by choosing $F(u)$ for u larger than 1 and truncating (2.37), with convenient conditions on $F(u)$—a sufficient one being that $|F(u)|$ be smaller than $Cu^{-3-\epsilon}$. It can also be done by constructing $\mathfrak{F}^E(r,r')$ from an arbitrary function $\omega(r)$ belonging to $\mathcal{E}$. The result is then used to obtain $\mathcal{K}^E(r)$, and eventually to reduce (3.2) to an infinite system of linear equation which can be solved.

A. Formal Obtaining of $\mathcal{K}_E(r)$

Let us now study the resolvent equation of $\overline{\mathfrak{F}}^E(r,r')$,

$$\mathcal{J}^E(r,r') = \overline{\mathfrak{F}}^E(r,r') - \int_0^\infty \mathcal{J}^E(r,\rho)\overline{\mathfrak{F}}^E(\rho,r)d\rho. \qquad (3.10)$$

Introducing the function

$$\alpha(r) = r^{-\epsilon/2}(1+r)^{-1+\epsilon}, \tag{3.11}$$

where the positive number ϵ is arbitrarily small and, in particular, smaller than $\frac{1}{4}$, we see that $|\overline{\mathfrak{F}}^E(r,r')|$, according to (2.102), is smaller than $C\alpha(r)\alpha(r')$. Clearly the mapping

$$\rho \to \sigma = \int_0^\rho \alpha^2(x)dx \tag{3.12}$$

is a continuous bijection of R^+ into $[0,\delta]$, where

$$\delta = \int_0^\infty x^{-\epsilon}(1+x)^{-2+2\epsilon}dx = [\Gamma(1+\epsilon)]^2[\Gamma(2-2\epsilon)]^{-1}. \tag{3.13}$$

Let us now introduce the functions

$$\overline{\mathfrak{F}}^E(s,s') = [\alpha(r)\alpha(r')]^{-1}\overline{\mathfrak{F}}^E(r,r'), \tag{3.14}$$

$$\mathcal{J}^E(s,s') = [\alpha(r)\alpha(r')]^{-1}\mathcal{J}^E(r,r'), \tag{3.15}$$

where s and s' are the images of r and r' in the mapping (3.12).

Using (3.12), (3.14), and (3.15) in (3.10), we obtain

$$\mathcal{J}^E(s,s') = \overline{\mathfrak{F}}^E(s,s') - \int_0^\delta \mathcal{J}^E(s,\sigma)\overline{\mathfrak{F}}^E(\sigma,s')d\sigma. \tag{3.16}$$

(3.16) is a Fredholm equation with a continuous kernel on $[0,\delta]$. The Fredholm alternative holds. Unless $\mathfrak{F}^E(s,s')$ belongs to the special set $\{\mathfrak{F}_2\}$ of functions for which the homogeneous equation reduced from (3.16) has a nontrivial solution, (3.16) can be solved. Consider one of the possible characterizations of the functions $\mathfrak{F}^E(r,r')$ [for instance, since $F(u)$ necessarily belongs to $L_2(1,\infty)$, the distance derived from the quadratic norm]. Let $\{\mathfrak{F}_2\}$ be the set of functions $\mathfrak{F}(r,r')$ such that $\overline{\mathfrak{F}}^E(s,s')$ belongs to $\{\overline{\mathfrak{F}}_2\}$. It easily can be seen (for instance, on the quoted example), that there is at least a distance for which $\mathfrak{F}_2 - \{\mathfrak{F}_2\}$ is dense in $\mathfrak{F}_2$. In other words, we hardly restrict our choice by taking the "arbitrary" function $\mathfrak{F}^E(r,r')$ in $\mathfrak{F}_2 - \{\mathfrak{F}_2\}$. *This choice will always be assumed in the* following.[40] The solution of (3.16) can then be constructed by the Fredholm (N/D) formula, which yields a bounded function $\mathcal{J}^E(s.s')$. Let us then define $\mathcal{J}^E(r,r')$ by

$$\mathcal{J}^E(r,r') = \alpha(r)\alpha(r')\mathcal{J}^E(s,s'). \tag{3.17}$$

We see that $\mathcal{J}^E(r,r')$ is the solution of (3.10). This solution is unique with the assumptions above. Since (3.10) is the resolvent equation of $\overline{\mathfrak{F}}^E(r,r')$, $\mathcal{J}^E(r,r')$ can be used for transforming (3.6), using (3.9), into

$$\mathcal{K}^E(r) = \overline{\mathfrak{F}}^E(r) - \int_0^\infty \mathcal{K}^I(\rho)\overline{\mathfrak{F}}^E(\rho,r)d\rho - \int_0^\infty \overline{\mathfrak{F}}^E(\rho)\mathcal{J}^E(\rho,r)d\rho + \int_0^\infty d\rho \times \int_0^\infty \mathcal{K}^I(\sigma)\,\overline{\mathfrak{F}}^E(\sigma,\rho)\mathcal{J}^E(\rho,r)d\sigma. \tag{3.18}$$

Now, since $\overline{\mathfrak{F}}^E(r,r')$ is symmetric, so is $\mathcal{J}^E(r,r')$, and we can rewrite (3.19), using (3.10), as

$$\mathcal{K}^E(r) = \overline{\mathfrak{F}}^E(r) - \int_0^\infty \overline{\mathfrak{F}}^E(\rho)\mathcal{J}^E(\rho,r)d\rho - \int_0^\infty \mathcal{K}^I(\rho)\mathcal{J}^E(\rho,r)d\rho. \tag{3.19}$$

Hence, $\mathcal{K}^E(r)$ is given as a known functional of $\mathcal{K}^I(r)$, once $\mathfrak{F}^E(r,r')$ has been chosen in $\mathfrak{F}_2 - \{\mathfrak{F}_2\}$ so that $\overline{\mathfrak{F}}^E(r)$ and $\mathcal{J}^E(r,r')$ are known. Besides, since it readily follows from Assumption (c) that $|\mathcal{K}^E(\rho)|$ is bounded by $C\alpha(\rho)$, so is $|\mathcal{K}^E(\rho)|$, provided that

$$\int_0^\infty |\mathcal{K}^I(\rho)|\alpha(\rho)d\rho < \infty. \tag{3.20}$$

This is true, in particular, for a potential of $\mathcal{S}$, since $|\mathcal{K}(\rho)|$ is then bounded by $C(1+\rho)^{-1/2}$. We therefore take (3.20) as a *working* assumption in the following, to be checked in the reverse derivation of the interaction from the phase shifts.

B. The Coefficients c_l

Let us now study $\overline{\mathfrak{F}}^I(r,r')$. Using the assumptions which define $\mathfrak{F}_2$, we can write

$$\mathfrak{F}^I(r,r') = 4\int_0^1 w^{-1}\,\sin w\, F(u)udu. \tag{3.21}$$

Let us introduce the notations

$$f^I(r,r') = rr'\mathfrak{F}^I(r,r'), \tag{3.22}$$

$$\gamma = -(2\pi)^{-1}V_0. \tag{3.23}$$

Substituting in (3.21) a well-known formula,[41] we obtain

$$f^I(r,r') = \sum_0^\infty c_l^{(0)}u_l(r)u_l(r'), \tag{3.24}$$

where

$$c_l^{(0)} = 4(2l+1)\int_0^1 F(u)P_l(1-2u^2)udu. \tag{3.25}$$

Using now (2.70) and well-known formulas, we obtain

$$c_l^{(0)} - \gamma = 4(2l+1)\int_0^1 \psi(u)P_l(1-2u^2)udu. \tag{3.26}$$

Recall the formula

$$4(2l+1)uP_l(1-2u^2) = -\frac{d}{du}[P_{l+1}(1-2u^2) - P_{l-1}(1-2u^2)]. \tag{3.27}$$

From (3.26) and (3.27) we get

$$c_l^{(0)} - \gamma = 4(2l+1)\int_0^\alpha \psi(u)P_l(1-2u^2)udu + 4(2l+1)\int_\beta^1 \psi(u)P_l(1-2u^2)udu - [\psi(u)(P_l(1-2u^2) - P_l(1-2a^2))]_\alpha^\beta \tag{3.28}$$

It follows from Assumptions (d) and (e) that $|\psi(u)|$ is bounded by $C(u^{-\epsilon} + \log|1-u|)$ and $|\psi'(u)|$ by $C(u^{-1-\epsilon} + |1-u|^{-1})$. On the other hand, the following inequalities hold:

$$|P_{l+1}(1-2u^2) - P_{l-1}(1-2u^2)| < \begin{cases} c(l+\frac{1}{2})u^2, \\ c(l+\frac{1}{2})(1-u), \\ c(l+\frac{1}{2})^{-1/2}u^{-1/2}(1-u)^{-1/4}. \end{cases} \tag{3.29}$$

The two first inequalities come from (3.24) and from the well-known inequality $|P_l(x)| \le 1$. The last one comes from the asymptotic behavior of the P_l. Using

these inequalities in (3.28) with convenient α and β, we easily obtain

$$|c_l^{(0)} - \gamma| < C(l + \tfrac{1}{2})^{(-1+3\epsilon)/5}. \tag{3.30}$$

Let us now integrate (3.21) by parts. We obtain

$$\begin{aligned} f^I(r,r') &= 2\sin r \sin r' F(\tfrac{1}{2}) \\ &\quad + \int_0^{1/2} [\cos w - \cos(r-r')] F'(u)du \\ &\quad + \int_{1/2}^{1} [\cos w - \cos(r+r')] F'(u)du. \end{aligned} \tag{3.31}$$

It follows from the study done in Sec. 2 that we can obtain a structural expansion for $\mathfrak{F}^I(r,r')$ through replacing w by $[r + r'(2u^2 - 1)]$. The bounds of $F'(u)$ in the neighborhoods of zero and one are those encountered (respectively, in Paper I, Sec. 2) in the appraisals previously done for the remainders. They enable us to show that the remainder is a negligible function. We therefore can write

$$\begin{aligned} \mathfrak{F}^I(r') &= 2F(\tfrac{1}{2})(r')^{-1}\sin r' + i(r')^{-1} \\ &\quad \times \int_0^{1/2} \{\exp[ir'(2u^2-1)] - \exp(-ir')\} F'(u)du \\ &\quad + ir'^{-1}\int_{1/2}^{1} \{\exp[ir'(2u^2-1)] - \exp(ir')\} F'(u)du \end{aligned} \tag{3.32}$$

or, integrating by parts,

$$\mathfrak{F}^I(r') = 4\int_0^1 \exp[ir'(2u^2-1)]F(u)u\,du. \tag{3.33}$$

Now, from a well-known formula,

$$\exp[ir'(2u^2-1)] = \sum_{l=0}^{\infty} (2l+1)\, i^{-l}(r')^{-1}u_l(r')P_l(1-2u^2). \tag{3.34}$$

Inserting (3.34) into (3.33) and comparing (3.25), we obtain

$$\mathfrak{F}^I(r') = r'^{-1}\sum_0^{\infty} \exp(-il\pi/2)\, c_l^{(0)}\, u_l(r'). \tag{3.35}$$

Clearly, we would have obtained this formula readily from (3.24) by using the definition of $\mathfrak{F}^I(r')$ and the asymptotic form of $u_l(r')$. But the "negligibility" of the remainder would have been difficult to prove. Let us now introduce the coefficients

$$c_l = \begin{cases} c_l^{(0)} & \text{for } l \neq 0 \\ c_0^{(0)} + \delta & \text{for } l = 0. \end{cases} \tag{3.36}$$

From (3.6), (3.24), and (3.35), we obtain the formulas

$$\overline{\mathfrak{F}}^I(r,r') = (rr')^{-1}\sum_{l=0}^{\infty} c_l u_l(r)u_l(r'), \tag{3.37}$$

$$\overline{\mathfrak{F}}^I(r') = (r')^{-1}\sum_{l=0}^{\infty} \exp(\tfrac{1}{2}il\pi)c_l u_l(r'). \tag{3.38}$$

Inserting now (3.35) and (3.24) into (3.5), we obtain for $\mathfrak{K}^I(r)$ the following expansion:

$$\begin{aligned} \mathfrak{K}^I(r) &= \sum_{l=0}^{\infty} c_l r^{-1} u_l(r)[\exp(-il\pi/2)] \\ &\quad - \int_0^{\infty} \mathfrak{K}(\rho)u_l(\rho)\rho^{-1}d\rho. \end{aligned} \tag{3.39}$$

Then via (1.21), (3.39) yields

$$\mathfrak{K}^I(r) = \sum_{l=0}^{\infty} c_l A_l \exp[i(\delta_l - l\pi/2)]r^{-1}u_l(r). \tag{3.40}$$

It is quite remarkable that the formula (3.40) is completely independent of the choice of $\mathfrak{F}^E(r,r')$. It follows from the bound $C(1+\rho)^{-1/2}$ of $|\mathfrak{K}(\rho)|$ that

$$\left| \int_0^{\infty} \mathfrak{K}(\rho)u_l(\rho)\rho^{-1}d\rho \right| < C(1+l)^{-1/2}. \tag{3.41}$$

Comparing (3.41) and (1.21), we see that A_l goes to 1, and δ_l goes to zero, as l goes to infinity. On the other hand, it follows from (3.27) that the coefficients c_l are equibounded. Hence $\mathfrak{K}^I(r)$, as given by (3.40), is an entire function of r. According to a previous paper,[42] the expansion of $\mathfrak{K}^I(r)$ in a series of Bessel functions is unique, and the $c_l A_l$ are completely determined by giving $\mathfrak{K}^I(r)$.

C. Algebraic Equations

Let us now introduce the functions or numbers:

$$\overline{\mathfrak{K}}^E(r) = \overline{\mathfrak{F}}^E(r) - \int_0^{\infty} \overline{\mathfrak{F}}^E(\rho)\mathfrak{J}^E(\rho,r)d\rho, \tag{3.42}$$

$$B_l \exp[i(\sigma_l - l\pi/2)] = \int_0^{\infty} \overline{\mathfrak{K}}^E(\rho)u_l(\rho)\rho^{-1}d\rho, \tag{3.43}$$

$$g_{ll'} = \int_0^{\infty}\int_0^{\infty} \mathfrak{J}^E(x,y)u_l(x)u_{l'}(y)x^{-1}y^{-1}\,dxdy. \tag{3.44}$$

Clearly, once $\mathfrak{F}^E(r,r')$ has been chosen, all these are known quantities.

Now, inserting (3.40) into (1.21), taking into account (3.9), yields the relation

$$\begin{aligned} A_l \exp[i(\delta_l - l\pi/2)] &= \exp(-il\pi/2) - \int_0^{\infty} \mathfrak{K}^E(\rho)u_l(\rho)\rho^{-1}d\rho \\ &\quad - \sum_{l'} c_{l'} L_{ll'} A_{l'} \exp[i(\delta_{l'} - l'\pi/2)], \end{aligned} \tag{3.45}$$

where

$$L_{ll'} = \int_0^{\infty} u_l(\rho)u_{l'}(\rho)\rho^{-2}d\rho. \tag{3.46}$$

From (3.19), (3.42), and (3.40), we get

$$\begin{aligned} \mathfrak{K}^E(r) &= \overline{\mathfrak{K}}^E(r) - \sum_{l'=0}^{\infty} c_{l'}A_{l'}\exp(i\delta_{l'} - l'\pi/2) \\ &\quad \times \int_0^{\infty} \rho^{-1}u_{l'}(\rho)\mathfrak{J}^E(\rho,r)d\rho. \end{aligned} \tag{3.47}$$

Inserting (3.47) into (3.45) gives the infinite system of linear equations,

$$\begin{aligned} A_l \exp[i(\delta_l - l\pi/2)] + B_l \exp[i(\sigma_l - l\pi/2)] &= \exp(-il\pi/2) - \sum_{l'} c_{l'}L_{ll'}A_{l'}\exp[i(\delta_{l'} - l'\pi/2)] \\ &\quad + \sum_{l'} c_{l'}g_{ll'}A_{l'}\exp[i(\delta_{l'} - l'\pi/2)]. \end{aligned} \tag{3.48}$$

When $\mathfrak{F}^E(r,r')$ and δ are given, $\mathfrak{J}^E(r,r')$, $\overline{\mathfrak{F}}^E(r)$, $\overline{\mathfrak{K}}^E(r)$, B_l, σ_l, $g_{ll'}$, can be considered as known quantities. The problem is to obtain consistently the c_l and the A_l from (3.42). When this is done, it will be possible to construct $\overline{\mathfrak{F}}^I(r,r')$ through (3.37), $\overline{\mathfrak{F}}^I(r')$ through (3.38), and $\mathfrak{K}^I(r)$ through (3.40). Inserting $\mathfrak{K}^I(\rho)$ in (3.19) yields $\mathfrak{K}^E(r)$. Hence, the whole structure of the scattering problem is known. If the potential is wanted, it is easy to construct it from $\mathfrak{F}(r,r')$ through the Regge–Newton equation (1.11) and formula (1.17). The infinite system (3.42) is therefore the fundamental system of the problem.

D. Reduction of the Fundamental System

Let us write (3.48) in the equivalent form,

$$A_l[1 + c_l(L_{ll} - g_{ll})] = \exp(-i\delta_l) - B_l \times \exp[i(\sigma_l - \delta_l)] - \sum_{l' \neq l} c_{l'}A_{l'} \times \exp[i(\delta_{l'} - \delta_l) + i(l - l')\pi/2](L_{ll'} - g_{ll'}). \quad (3.49)$$

The real and the imaginary part of (3.49) give two systems of equations:

$$A_l[1 + c_l(L_{ll} - g_{ll})] = \cos\delta_l - B_l \cos(\sigma_l - \delta_l) - \sum_{l' \neq l} M_{ll'}c_{l'}A_{l'} \sin(\delta_{l'} - \delta_l) + \sum_{l' \neq l} g_{ll'}c_{l'}A_{l'} \times \cos[\delta_{l'} - \delta_l + (l - l')\pi/2], \quad (3.50)$$

where

$$M_{ll'} = i \exp[i(l - l')\pi/2]L_{ll'} \quad (3.51)$$

or

$$M_{ll'} = \begin{cases} 0 \text{ for } l - l' \text{ even} \\ [(l' + \tfrac{1}{2})^2 - (l + \tfrac{1}{2})^2]^{-1} \text{ for } l - l' \text{ odd} \end{cases} \quad (3.52)$$

and

$$\sin\delta_l + B_l \sin(\sigma_l - \delta_l) = \sum_{l'} M_{ll'}c_{l'}A_{l'} \cos(\delta_{l'} - \delta_l) + \sum_{l'} g_{ll'}c_{l'}A_{l'} \sin[\delta_{l'} - \delta_l + (l - l')\pi/2]. \quad (3.53)$$

Let us now set, for convenience,

$$a_l = c_l A_l \cos\delta_l, \quad (3.54)$$

$$\eta_l = (\cos\delta_l)^{-1}B_l \sin(\sigma_l - \delta_l) = B_l(\sin\sigma_l - \cos\sigma_l \tan\delta_l), \quad (3.55)$$

$$\bar{h}_{ll} = \begin{cases} (-1)^{(l-l')/2}g_{ll'} & \text{for } l - l' \text{ even and} \neq 0 \\ 0 \text{ for } (l - l') \text{ odd or zero} \end{cases} \quad (3.56)$$

$$h_{ll} = \begin{cases} (-1)^{(l-l'-1)/2}g_{ll'} & \text{for } (l - l') \text{ odd} \\ 0 \text{ for } (l - l') \text{ even} \end{cases} \quad (3.57)$$

We get from (3.53),

$$\tan\delta_l + \eta_l = \sum_{l'} M_{ll'}a_{l'}(1 + \tan\delta_l\tan\delta_{l'}) + \sum_{l'} a_{l'}[h_{ll'}(1 + \tan\delta_l \tan\delta_{l'}) + \bar{h}_{ll'}(\tan\delta_{l'} - \tan\delta_l)] \quad (3.58)$$

or, in matrix notations, **e** being the unit vector, **a** the vector with components a_l,

$$(\eta + \tan\Delta)\mathbf{e} = M(1 + S)\mathbf{a}, \quad (3.59)$$

where

$$S = M^{-1}\tan\Delta M \tan\Delta + M^{-1}\tan\Delta h \tan\Delta + M^{-1}h + M^{-1}\bar{h}\tan\Delta - M^{-1}\tan\Delta\,\bar{h}. \quad (3.60)$$

The matrix M^{-1} has been constructed by the author in a previous paper. Hence, if δ is given, S can be constructed, and solving (3.53) reduces to inverting $(1 + S)$. The solution of (3.53) from the known δ_l yields to a_l. From the a_l, a simple division yields the quantity c_lA_l. Inserting these values in (3.50) yields the A_l and therefore the c_l, from which we construct all the unknown quantities of the scattering problem. So as to give a more precise meaning to this formalism, let us more thoroughly study $\mathcal{J}^E(r,r')$ and the g_l.

E. Properties of $\mathcal{J}^E(r,r')$ and $g_{ll'}$

From (A67), (A85), (A100), we know that, for a potential of class $\mathcal{E}$, $\rho\omega(\rho)$ can be absolutely bounded by $C\rho^\epsilon$, and its derivative by $C\rho^{\epsilon-1}$. From (2.25), it follows that $f(r,r')$ and its first and second order derivatives are continuous functions for any finite r and r', that $f(r,r')$ is a solution of (1.10), and that the following bounds hold:

$$|f(r,r')| < C(rr')^{1+\epsilon/2}, \quad (3.61)$$

$$\left|\frac{\partial}{\partial r'}f(r,r')\right| < C(1 + rr' + r'^2)r^{1+\epsilon/2}r'^{\epsilon/2}, \quad (3.62)$$

$$\left|\frac{\partial^2}{\partial r'^2}f(r,r')\right| < C(1 + rr' + r'^2)r^{1+\epsilon/2}r'^{-1+\epsilon/2}. \quad (3.63)$$

On the other hand, the series

$$\sum_0^\infty \bar{c}_l u_l(r)u_l(r'), \quad (3.64)$$

where $\{\bar{c}_l\}$ is any sequence of bounded numbers, can easily be proved[43] to be absolutely bounded by Crr' if $\bar{c}_0 \neq 0$, by $C(rr')^2$ if $\bar{c}_0 = 0$, and its partial derivatives by the ratios of these bounds by r, r^2, r', r'^2, rr' according to the case. Hence, for a potential of class $\mathcal{E}$, the parameter δ can be chosen in such a way that $f^E(r,r')$ satisfies (3.61), (3.62), and (3.63). It can also be said that for *a potential of class* $\mathcal{E}$, if $f^E(r,r')$ satisfies (3.61), then $c_0 = 0$. In the following, we complete the Assumptions (a) – (e) defining $\mathfrak{F}_2$ by the following one, which defines the subclass $\mathfrak{F}_3$ of $\mathfrak{F}_2$.

Assumption f: In $\mathfrak{F}_3$, $f^E(r,r')$ satisfies the bounds (3.61)–(3.63). That the special set $\{\mathfrak{F}_3\}$ for which the homogeneous form of (3.14) has a nontrivial solution is rare in $\mathfrak{F}_3$ can be proved like for $\{\mathfrak{F}_2\}$ in $\mathfrak{F}_2$.

Now, from (2.90) and (3.8), we can write

$$f^E(r,r') = \int_1^\infty [\cos w - \cos(r + r')]F'(u)du - \delta\sin r \sin r'. \quad (3.65)$$

If $|F'(u)|$ was bounded, like $|\Phi_1'(u)|$, by $Cu^{-3-\epsilon}$, we would have been able, after choosing conveniently δ, to obtain the bound

$$Cr^{\epsilon+1}(1 + r^{\epsilon+1})^{-1}r'^{\epsilon+1}(1 + r'^{\epsilon+1})^{-1} \text{ for } |\bar{f}^E(r,r')|,$$

similar bounds for the derivatives, and this would have made much easier the discussion below. With Assumptions (d) and (e), it is proved in Appendix C that $|f^E(r,r')|$ is bounded by $C[1 + |\log(r^{-1} + r'^{-1})|]$. Taking this into account, with (3.61), yields the improved bound (valid in $\mathfrak{F}_3$),

$$|\bar{f}^E(r,r')| < C\beta(r)\beta(r') \quad (3.66)$$

with

$$\beta(r) = r^{1+\epsilon/2}(1 + r^{1+\epsilon/2})^{-1}(1 + |\log r|), \quad (3.67)$$

$\beta(r)$ is a better bound than $\alpha(r)$. Using it like we did for $\alpha(r)$, we readily prove that

$$|\mathcal{J}^E(r,r')| < C(rr')^{-1}\beta(r)\beta(r'). \quad (3.68)$$

Assumptions (d) and (e) are unfortunately only sufficient to prove that the differentiation on the right-hand side of (3.10) with respect to r' is possible once (Appendix C). When this is done, we let r' go to ∞ and

replace $f(\rho,r')$ and $(\partial/\partial r')f(\rho,r')f$ by their structural expansion. It is easy to see that the remainders are negligible. The function $\mathcal{J}^E(r,r')$ has, therefore, a structural expansion and its s.s. function is given by

$$\mathcal{J}^E(r) = \bar{\mathcal{F}}^E(r) - \int_0^\infty \mathcal{J}^E(r,\rho)\bar{\mathcal{F}}^E(\rho)d\rho. \tag{3.69}$$

(3.69) can also be considered as an integral equation for $\bar{\mathcal{F}}^E(r)$. It follows from our assumptions that its solution is unique and is given by

$$\mathcal{J}^E(r) = \bar{\mathcal{F}}^E(r) - \int_0^\infty \bar{\mathcal{F}}^E(r,\rho)\,\mathcal{J}^E(\rho)d\rho. \tag{3.70}$$

We can also compare (3.69) with (3.49). Since $\mathcal{J}^E(r,r')$ is symmetric and since the solution is unique, we derive the formula

$$\bar{\mathcal{K}}^E(r) = \mathcal{J}^E(r). \tag{3.71}$$

Let us now study the equation

$$\mathcal{J}_A^E(r,r') = \bar{\mathcal{F}}^E(r,r') - \int_0^A \mathcal{J}_A^E(r,\rho)\bar{\mathcal{F}}^E(\rho,r')d\rho. \tag{3.72}$$

With some algebra, we can derive the equivalent equation

$$\mathcal{J}_A^E(r,r') = \mathcal{J}^E(r,r') + \int_A^\infty \mathcal{J}_A^E(r,\rho)\mathcal{J}^E(\rho,r')d\rho. \tag{3.73}$$

It is easy to see from (3.68) that we can choose A in such a way that (3.73) can be solved by iteration. This choice will be assumed in the following. The sum of the Neumann iterated series give then a function $\mathcal{J}_A^E(r,r')$ which is bounded like (3.68). Inserting this bound in (3.72) and using successively (3.61), (3.62), and (3.63), we easily prove that $\mathcal{J}_A^E(r,r')$ is, for any finite A, twice differentiable, and that its derivatives satisfies bounds like (3.61), (3.62), and (3.63). Needless to say, these bounds depend on A and it may be that the derivatives have no limit as A goes to ∞. Actually, we have been able to prove that they converge almost everywhere if $|F'(u)|$ is bounded by $Cu^{-3-\epsilon}$; but we have not been able to extend the proof with only the Assumptions (d) and (e). Fortunately, this is not necessary for our purpose. Let us define from (3.72) the s.s. function $\mathcal{J}_A^E(r)$ corresponding to $\mathcal{J}_A^E(r,r')$. It is easy to see that $\mathcal{J}_A^E(r)$ converges to $\mathcal{J}^E(r)$ as A goes to ∞. Let $\bar{g}_A(r)$ and $g_A(r)$ be the real and the imaginary part of $r\mathcal{J}_A^E(r)$, $\bar{g}(r)$ and $g(r)$ the real and the imaginary part of $r\mathcal{J}^E(r)$, and $\bar{f}(r)$ and $f(r)$ the real and the imaginary part of $r\bar{\mathcal{F}}^E(r)$. It is easy to see that Eqs. (3.69) and (3.70) hold when $(0,\infty)$ is replaced by $(0,A)$ and the quantities $\mathcal{J}_A^E(r,\rho)$ and $\mathcal{J}^E(r)$ are replaced by $\mathcal{J}_A^E(r,\rho)$ and $\mathcal{J}_A^E(r)$. In the same way, we define the A restriction of B_l and σ_l by

$$B_l^A \exp[i(\sigma_l^A - l\pi/2)] = \int_0^A \mathcal{J}_A^E(\rho)\,u_l(\rho)\rho^{-1}d\rho. \tag{3.74}$$

Let us now study the function

$$h_A(r,r') = (D_r^{(0)} - D_{r'}^{(0)})g_A^E(r,r'), \tag{3.75}$$

where $D_r^{(0)}$ has been defined by (1.1), and $g_A^E(r,r')$ is defined from $\mathcal{J}_A^E(r,r')$ like $f(r,r')$ from $\mathcal{F}(r,r')$.

From (3.72) we easily obtain

$$h_A(r,r') = h_A^{(0)}(r,r') - \int_0^A h_A(r,\rho)\bar{f}^E(\rho,r')\rho^{-2}d\rho, \tag{3.76}$$

where

$$h_A^{(0)}(r,r') = g_A^E(r,A)\left(\frac{\partial}{\partial\rho}\bar{f}^E(\rho,r')\right)_{\rho=A} - \bar{f}^E(A,r')\left(\frac{\partial}{\partial\rho}g_A^E(r,\rho)\right)_{\rho=A}. \tag{3.77}$$

It is easy to see that $h_A^{(0)}(r,r')$ is equal to

$$h_A^{(0)}(r,r') = g_A(r)\bar{f}(r') - \bar{g}(r)f(r') + H_N(r')g_N(A,r) + H_N(r)f_N(A,r'), \tag{3.78}$$

where g_N and f_N are "negligible functions" and $|H_N(r)|$ is bounded by $Cr\alpha(r)$. Comparing now (3.76) and (3.72) and taking into account (3.68) and the A form of (3.70), we get

$$h_A(r,r') = g_A(r)\bar{g}_A(r') - \bar{g}_A(r)g_A(r') + \mathcal{K}(r,r',A), \tag{3.79}$$

where $\mathcal{K}(r,r',A)$ goes to zero as A goes to ∞, for any fixed r and r' in such a way that $\int_0^\infty\int_0^\infty |\mathcal{K}(r,r',A)|\,[r(1+r)]^{-1}[r'(1+r')]^{-1}$ goes to zero. An analysis of the asymptotic behavior of (3.73) as r' goes to ∞ shows that $|g(r) - g_A(r)|$ and $|\bar{g}(r) - \bar{g}_A(r)|$ are bounded by $r\,\alpha(r)\int_A^\infty \alpha(\rho)(1+\rho)^{-1+\epsilon}\,d\rho$. The functions $g(r)$ and $\bar{g}(r)$ can therefore be substituted to $g_A(r)$ and $\bar{g}_A(r)$ without changing the properties of the remainder $\mathcal{K}$. Incidently, this proves that $h_A(r,r')$ has a limit even in the conditions in which this cannot be ascertained for the second derivatives of $g^E(r,r')$. Let us now write, using (3.44),

$$g_{ll'} = \lim_{A\to\infty} g_{ll}^A, \tag{3.80}$$

where

$$g_{ll'}^A = \int_0^A\int_0^A \mathcal{J}_A^E(x,y)u_l(x)u_{l'}(y)\,x^{-1}y^{-1}dxdy \tag{3.81}$$

Clearly,

$$\begin{aligned}[l(l+1) - l'(l'+1)]g_{ll'} &= \int_0^A\int_0^A g_A^E(r,r')[u_{l'}(r')D_r^0 u_l(r) \\ &\quad - u_l(r)D_r^0 u_{l'}(r')]r^{-2}dr\,r'^{-2}dr' \\ &= B_l^A \sin[\sigma_l^A + \tfrac{1}{2}\pi(l-l')] \\ &\quad - B_l^A \sin[\sigma_l^A + \tfrac{1}{2}\pi(l'-l)] \\ &\quad + \int_0^A\int_0^A u_l(r)u_{l'}(r')h_A(r,r')r^{-2}r'^{-2}dr\,dr' \\ &\quad + o(1),\end{aligned} \tag{3.82}$$

where the remainder $o(1)$ comes from the fact that we have replaced the Bessel function by their asymptotic behavior. Using then (3.79) and going to the limit $A\to\infty$ yields finally the remarkable formula

$$\begin{aligned}g_{ll'} = [l(l+1) - l'(l'+1)]^{-1}\{&B_l \sin[\sigma_l + \tfrac{1}{2}\pi(l-l')] \\ &- B_l \sin[\sigma_l + \tfrac{1}{2}\pi(l'-l)] \\ &+ B_l B_{l'} \sin[\sigma_l - \sigma_{l'} - \tfrac{1}{2}\pi(l-l')]\}.\end{aligned} \tag{3.83}$$

F. Bounds for B_l

In the following, we assume that $|(d/du)\Phi_2^{(1)}(u)|$ is a function of $L_{1+\epsilon^{-1}}(0,\infty)$, or, in other word, that

Assumption G: $(d/du)\Phi_2(u)$ is the sum of $C(u-1)^{-1}$ and of a function of $L_{1+\epsilon^{-1}}(0,\infty)$.

Using now (3.43) and (3.70), we obtain

$$B_l \exp[i(\sigma_l - l\pi/2)] = \int_0^\infty \bar{\mathfrak{F}}^E(r)u_l(r)r^{-1}dr - \int_0^\infty \mathfrak{J}^E(\rho)\rho^{-1}d\rho \int_0^\infty \bar{\mathfrak{F}}^E(\rho,r)u_l(r)r^{-1}dr. \quad (3.84)$$

We know that the asymptotic behavior of $\bar{\mathfrak{F}}^E(\rho,r)$ and $\mathfrak{F}^E(\rho,r)$ can be obtained from (3.65) in the form

$$f^E(r,r') \sim \int_0^\infty \{\cos[r + r'(2u^2-1)] - \cos(r+r')\} \times F'(u)du - \delta \sin r \sin r' \quad (3.85)$$

$$\mathfrak{F}^E(r') = ir'^{-1}\exp(ir')\int_1^\infty [\exp(2ir')(u^2-1) - 1] \times F'(u)du + \delta(\sin r')/r'. \quad (3.86)$$

Substituting (3.86) into the first term in the right-hand side of (3.84), we are led to integrals of the form

$$I = \int_r^\infty \exp(i\alpha r')u_l(r')r'^{-2}dr'. \quad (3.87)$$

Using well-known formulas,[44] it is easy to show that $|I|$ is smaller than Cl^{-2} as l goes to ∞. In the same way, we easily prove, by substituting (3.85) in the second term of the right-hand side of (3.84), that (3.85) yields there a contribution smaller than Cl^{-2} as l goes to infinity. Unfortunately, Assumptions (a) $\leftrightarrow$ (g) are not sufficient to yield such a beautiful result when the whole function $f^E(r,r')$ is used. The remainders lead there to terms of the form

$$\int_1^\infty F'(u)du \int_0^\infty \mathfrak{J}^E(\rho)\rho^{-1}d\rho \int_0^\infty R(r,\rho,u)u_l(r)r^{-2}dr, \quad (3.88)$$

where, according to results of I (Sec. 5),

$$|R(r,\rho,u)| < C\rho^2(u^2-1)[r+\rho^2(u^2-1)]^{-1} + C\rho^2u^2(u^2-1)\{r+\rho^2u^2(u^2-1)\}. \quad (3.89)$$

When (3.89) is used in (3.88), replacing all the terms by an absolute upper bound, and noticing that the contribution of $u_l(r)$ for r smaller than $\frac{1}{2}l$ goes to zero faster than any power of l, we can prove that (3.88) goes to zero faster than $l^{(-5/4)+\epsilon}$, and therefore faster than $l^{-1-\epsilon_0}$, where ϵ_0 is a positive number.

We therefore get the bound

$$|B_l| < Cl^{-1-\epsilon_0}. \quad (3.90)$$

G. Solutions of the Fundamental System

From a previous study,[45] we know the matrix M^{-1},

$$M^{-1} = -\mu M \mu, \quad (3.91)$$

where μ is the diagonal matrix, with elements

$$\left.\begin{aligned}\mu_{2n} &= 4\pi^{-1}[\Gamma(n+\tfrac{1}{2})]^2[\Gamma(n+1)]^{-2}(2n+\tfrac{1}{2})\\ \mu_{2n+1} &= 4\pi^{-1}[\Gamma(n+\tfrac{3}{2})]^2[\Gamma(n+1)]^{-2}(2n+\tfrac{3}{2})\end{aligned}\right\}. \quad (3.92)$$

The inverse of M is not unique. However, since we are interested in a solution $\{a_l\}$ uniformly bounded, we can write

$$\alpha\mathbf{v} + M^{-1}(\eta + \tan\Delta)\mathbf{e} = (1+S)\mathbf{a}, \quad (3.93)$$

where $\mathbf{v}$ is a vector such that

$$M\mathbf{v} = 0 \quad (3.94)$$

and α is an arbitrary parameter. It is possible to show (we do not in this paper) that the the unique solution of (3.94) with bounded components is the vector with even order components equal to μ_{2n}, odd order components equal to zero. Let us now study $(1+S)$. From (3.57) and (3.83) we easily obtain

$$h = MD + DM + DMD + EME, \quad (3.95)$$

where D and E are the diagonal matrices with elements

$$D_{ll} = B_l \cos\sigma_l \quad (3.96)$$

and

$$E_{ll} = B_l \sin\sigma_l, \quad (3.97)$$

$$\bar{h} = \bar{M}E + E\bar{M} + E\bar{M}D - D\bar{M}E, \quad (3.98)$$

and $\bar{M}$ is the matrix

$$\bar{M}_{ll'} = \begin{cases}[l(l+1) - l'(l+1)]^{-1} & \text{for } l-l' \text{ even} \neq 0\\ 0 \quad \text{for } l-l' \text{ odd or } l=l'.\end{cases} \quad (3.99)$$

Let us now recall the following inequality[46] valid for any $\beta > -1$,

$$\sum_{l' \neq l} |l^2 - l'^2|^{-1} l^{-\beta} < C(l^{-\beta-1} + l^{-2}). \quad (3.100)$$

Let us assume that the phase shifts δ_l go to zero like $Cl^{-1-\epsilon_1}$ as l goes to ∞. It is easy to see from (3.60), (3.90), (3.95), (3.98), and (3.100) that, $L^{\pm\bar{\epsilon}}$ being the diagonal matrix with elements $(l+\frac{1}{2})^{\pm\bar{\epsilon}}$,

$$Q = -L^{-\bar{\epsilon}}\mu^{1/2}\, S\, \mu^{1/2}L^{\bar{\epsilon}} \quad (3.101)$$

maps any vector with components $< Cl^{-\bar{\epsilon}-\epsilon}$ into a vector with components going to zero (faster than $Cl^{-1-\bar{\epsilon}}$) as l goes to ∞. Besides, it is easy to see that Q is completely continuous. The spectrum of Q is discrete and it is possible to use the Fredholm method for inverting $(1-Q)$, provided that one is not an eigenvalue of Q. Q depends continuously on the phase shifts δ and on the function $f^E(r,r')$ chosen above. When that function is equal to zero or bounded in such a way that the B_l are all very small, and when, in the same problem, all the phase shifts are small, it is possible to construct $(1-Q)^{-1}$ by iteration. This can be proved readily. It follows that one can be an eigenvalue of Q only for "rare" sets of phase shifts and other parameters. The Fredholm method can therefore be used "in general." It yields

$$(1-Q)^{-1} = 1 + N/D, \quad (3.102)$$

where

$$D = 1 + \sum_1^\infty \frac{(-1)^n}{n!} \sum_{l_1,l_2,\ldots,l_n} \begin{bmatrix} Q^{l_1}_{l_1} & Q^{l_2}_{l_1} & \cdots & Q^{l_n}_{l_1} \\ Q^{l_1}_{l_2} & & & \\ \cdot & & & \\ \cdot & & & \\ Q^{l_1}_{l_n} & & \cdots & Q^{l_n}_{l_n} \end{bmatrix}, \quad (3.103)$$

$$N_l^{l'} = \sum_0^\infty \frac{(-1)^n}{n!} \sum_{l_1,\ldots,l_n} \begin{bmatrix} Q_l^{l'} & Q_l^{l_1} & \cdots & Q_l^{l_n} \\ Q_{l_1}^{l'} & Q_{l_1}^{l_1} & \cdots & Q_{l_1}^{l_n} \\ \cdot & & & \\ \cdot & & & \\ Q_{l_n}^{l'} & & \cdots & Q_{l_n}^{l_n} \end{bmatrix} \tag{3.104}$$

Let us now expand $N_l^{l'}$ with respect to the first column in each determinant. We obtain in this way,

$$N_l^{l'} = \sum_0^\infty \frac{(-1)^n}{n!} \sum_{l_1,\ldots,l_n} [Q_l^{l'} \mathfrak{A}_l(l, l_1, \ldots, l_n) + Q_{l_1}^{l'} \mathfrak{A}_{l_1}(l, l, \ldots, l_n) + \ldots], \tag{3.105}$$

where $\mathfrak{A}_{l_p}$ is the cofactor of $Q_{l_p}^{l'}$ in the determinant in (3.104). The application of N to a vector w with components bounded by $Cl^{-\bar{\epsilon}-\epsilon}$ yields

$$\sum N_l^{l'} w_{l'} = \sum_0^\infty \frac{(-1)^n}{n!} \sum_{l_1,\ldots,l_n} [\bar{w}_l \mathfrak{A}_l + \bar{w}_{l_1} \mathfrak{A}_{l_1} + \cdots], \tag{3.106}$$

where

$$\bar{w}_l = \Sigma\, Q_l^{l'} w_{l'}. \tag{3.107}$$

On the other hand it is easy to see that the following bound holds for any positive $\bar{\epsilon}$:

$$|Q_{l_p}^{l_n}| < C(1 + l_p)^{-1-\bar{\epsilon}}. \tag{3.108}$$

Using Hadamard's theorem, we obtain from (3.108),

$$|\mathfrak{A}_{l_p}(l, l_1, \ldots, l_n)| < C^n n^{n/2} (1 + l)^{-1-\bar{\epsilon}} (1 + l_1)^{-1-\bar{\epsilon}} \cdots \times (1 + l_{p-1})^{-1-\epsilon} (1 + l_{p+1})^{-1-\bar{\epsilon}} \cdots (1 + l_n)^{-1-\bar{\epsilon}}. \tag{3.109}$$

Clearly, (3.109) implies that the series in (3.105) converge and yield a number $N_l^{l'}$ absolutely bounded by $C(1 + l)^{-1-\bar{\epsilon}}$. The formula (3.102) yields therefore the inverse of $(1 - Q)$. It is easy to see that a modification of $\bar{\epsilon}$ does not modify D. The multiplication of Q by $\mu^{1/2}$ on the left and $\mu^{-1/2}$ on the right does not either. In other words, the use of Q was only a trick for showing the convergence of (3.103) and (3.104) by Hadamard's theorem and we could have done this by using only S if we had introduced more refined bounds. This proves that $(1 + S)^{-1}$ can be put in the form

$$(1 + S)^{-1} = 1 + \bar{N}/D, \tag{3.110}$$

where

$$N_l^{l'} = \mu_l^{1/2} (1 + l)^{\bar{\epsilon}} N_l^{l'} \mu_{l'}^{-1/2} (1 + l')^{-\bar{\epsilon}}. \tag{3.111}$$

Clearly $\bar{N}_l^{l'}$ is bounded uniformly. Besides, using (3.106) and (3.11), we can write

$$\sum \bar{N}_l^{l'} \omega_{l'} = \sum_0^\infty \frac{(-1)^n}{n!} \sum_{l_1,\ldots,l_n} [\bar{\omega}_l \mathfrak{B}_l + \bar{\omega}_{l_1} \mathfrak{B}_{l_1} + \ldots], \tag{3.112}$$

where

$$\bar{\omega}_{l_p} = \sum_{l'} S_{l_p}^{l'} \omega_{l'}, \tag{3.113}$$

$$\mathfrak{B}_{l_p} = \mu_l^{1/2} (1 + l)^{\bar{\epsilon}} \mu_{l_p}^{-1/2} (1 + l_p)^{-\bar{\epsilon}}\ \mathfrak{A}_{l_p}. \tag{3.114}$$

This proves that $(1 + S)^{-1}$ can be applied to any vector to which S *can be applied and yields*, through (3.113), a bounded number. This is the case with the left-hand side of (3.93), according to the assumptions done on the phase shifts and the bounds of the B_l. We therefore obtain

$$\mathbf{a} = (1 + S)^{-1}[\alpha \mathbf{v} + M^{-1}(\eta + \tan\Delta)\mathbf{e}]. \tag{3.115}$$

It is clear from the discussion above that a_l is bounded as l goes to ∞. The coefficients $c_l A_l$ are therefore bounded and it follows from (3.50) that the A_l go to one as l goes to ∞. The c_l are therefore bounded as l goes to ∞. We can even go further by expanding the $\mathfrak{B}_l$ (or the $\mathfrak{A}_l$, with respect to the first line), so as to put in evidence S_l^{li} (or Q_l^{li}). By this way it is not difficult to show that

$$a_l \sim \begin{cases} C_0 + 0(l^{-\epsilon}) & \text{for } l \text{ even} \to \infty \\ C_1 + 0(l^{-\epsilon}) & \text{for } l \text{ odd} \to \infty \end{cases}, \tag{3.116}$$

where $\epsilon = \mathrm{Inf}(\epsilon_0, \epsilon_1)$. Inserting this result in (3.40), so as to take apart the contribution of the constants, which can be obtained[47] exactly, in (3.40), and roughly bounding the contribution of the remainder, it is easy to see that condition (3.20) holds, so that the model is self-consistent.

Unfortunately, (3.116) (or the cleverness of the author) is not sufficient to guarantee that all the potentials constructed in this way have a physical meaning. This, however, would be possible if the δ_l and the B_l were bounded by $Cl^{-3-\epsilon}$, since we could use a method we have already introduced.[47]

H. Conclusion

We have to check that our solution satisfies the five requirements of Sec. 2I.

(1) Let Δ be the space of sequences $s = \{s_l\}$ such that $|\tan s_l| < C(1 + l)^{-\epsilon_1 - 1}$, the method $\mathfrak{M}$ described above enables one to construct *all* the solution $f(r, r')$ of (1.10) belonging to the class $\mathfrak{F}_1$ defined below and such that (1.21) *and* (2.49) hold.

(2) The class $\mathfrak{F}_1$ contains all the functions $f(r, r')$ whose external projection, as defined in Sec. 2, fulfills Assumptions a $\leftrightarrow$ f described in Sec. 3, and whose internal projection is characterized by coefficients c_l behaving asymptotically like (3.116). This class contains $\mathfrak{F}_0$ according to (3.30) and to the choice of Assumptions a $\leftrightarrow$ f. To obtain the solution, the external projection is chosen arbitrarily, and then the c_l are determined from the δ_l and from parameters related to the external projection.

(3) We have to check that, for a potential of class $\mathcal{E}$, the set $|\tan\delta_l|$ belongs to Δ. Now, according to a result Martin,[48]

$$|\sin\delta_l| < \{1 - [\pi/2l + 1)]^{1/2} \int_0^\infty \rho\, |V(\rho)|\, d\rho\}^{-1} \times \int_0^\infty [u_l(\rho)]^2 |V(\rho) d\rho. \tag{3.117}$$

For a potential of class $\mathcal{E}$, this bound goes to zero like $l^{-2-\epsilon}$ as l goes to ∞. Therefore the set $|\tan\delta_l|$ belongs to Δ if (and only if) no phase shifts is equal to $\frac{1}{2}\pi$, viz. provided there is no resonance.

(4) If we mean by method $\mathfrak{M}$ the set of all solutions of the fundamental system, the fourth requirement is

certainly fulfilled. If we mean the set obtained by Fredholm method, it may be, after all, that the potential V_0 corresponds to the spectrum. The solution must then be obtained by methods available to solve inhomogeneous equations in the case of an eigenvalue (for special free terms).

(5) The fifth requirement is therefore fulfilled for almost every element of $\mathcal{V}_0$.

We can now give a definite answer to questions (1)–(v). The answer to (i) and (iii) is positive and the answer to (ii) is negative. The method $\mathfrak{M}$ is a positive answer to (iv). The way by which the method $\mathfrak{M}$ is undertaken enables an easy answer to question (v), since an appraisal of the deviations between functions whose Fourier spectrum is different outside a finite interval is a classical problem. Practical answers are given in a previous paper.[23] More general theoretical answers will be given in a forthcoming paper.[32]

ACKNOWLEDGMENTS

The beginning of this paper has been written in the warm hospitality of Indiana University Physics Department. The end of this paper has been written in the Mathematics Department of Montpellier University. We thank our colleagues of both departments for their interest in this work.

APPENDIX A

We derive the asymptotic behavior of $f(r,r') - K(r,r')$, $f(r,r) - K(r,r)$, and their derivatives. All our results are given for potentials of $\mathcal{E}$, i.e., $\mathcal{E}^3_{13}$.

Improvement of Certain Bounds

So as to obtain certain results, we need to improve the bounds previously[29] obtained for the remainders when $K(r,r')$ and $(\partial/\partial r)K(r,r')$ are replaced by their asymptotic behavior. We are interested in the case

$$r' \leq \tfrac{1}{2}, \quad r \geq 1. \tag{A1}$$

It follows from (2.17), (2.27), and (2.29) that

$$\mathrm{Re}\mathcal{R}_0(r,r') = \int_0^\infty \{\cos w - \cos[r + r'(2u^2 - 1)]\} \times \Phi'(u)du, \tag{A2}$$

where for $u \leq 1$

$$\Phi'_1(u) = \Phi_3(u) = -(4\pi)^{-1}u^{-2}\int_0^\infty [u^{-1}\sin(2us) - 2s] \times \frac{d}{ds}s^{-1}\frac{d}{ds}(s^3V(s))ds \tag{A3}$$

and, for $u > 1$,

$$\Phi'(u) = \Phi_4(u) = -(4\pi)^{-1}u^{-2}\int_0^\infty u^{-1}\sin(2us) \times \frac{d}{ds}s^{-1}\frac{d}{ds}(s^3V(s))ds. \tag{A4}$$

Let us now study the contribution to (A2) coming from the values of u smaller than one. Let us introduce $\alpha\,(0 < \alpha < 1)$ and integrate by parts from α to one. We obtain

$$\begin{aligned}&\int_0^\alpha \{\cos w - \cos[r + r'(2u^2 - 1)]\}\,\Phi'(u)\,du \\ &\quad + (4r)^{-1}\sin(r + r')\,\Phi(1) - [r^{-1}w(\alpha) - 1](4r'\alpha)^{-1} \\ &\quad \times \sin(w(\alpha))\,\Phi(\alpha) - (4r'\alpha)^{-1}[\sin w(\alpha) \\ &\quad - \sin(r - r' + 2r'\alpha^2)]\,\Phi(\alpha) \\ &\quad + \int_\alpha^1 [\sin(r - r' + 2r'u^2) - \sin w] \\ &\quad \times \frac{d}{du}[(4r'u)^{-1}\Phi(u)]du \\ &\quad + \int_\alpha^1 \sin w(1 - r^{-1}w)\frac{d}{du}[(4r'u)^{-1}\Phi(u)]du \\ &\quad - \int_\alpha^1 w^{-1}\sin w\,\Phi(u)du.\end{aligned} \tag{A5}$$

From (2.31), it is easy to see that, for $u < 1$,

$$|1 - r^{-1}w| < Cr'/r \tag{A6}$$

and according to [I, (5.6)]

$$|r - r' + 2r'u^2 - w| < Cr^{-1}r'^2u^2. \tag{A7}$$

From (A3) it is easy to show that $|\Phi_3(u)|$ is bounded by $Cu^{\epsilon-1}$, and $|(d/du)u^{-1}\Phi_3(u)|$ by $Cu^{\epsilon-3}$, for a potential in $\mathcal{E}$. Using these bounds in (A5), with

$$\alpha = \mathrm{Inf}(1, r'^{-1/2}), \tag{A8}$$

we see that the contribution of $\Phi_3(u)$ is absolutely bounded by $Cr^{-1}(1 + r')$, whereas the one of $(2\pi)^{-1}V_0u^{-2}$ is absolutely bounded by $Cr^{-1}(1 + r'^{3/2})$.

Let us now study the values of u larger than one and introduce for this $A > 1$. Integrating by parts (A2) from one to A, we get

$$\begin{aligned}&-(4r)^{-1}\sin(r + r')\,\Phi(1) + [r^{-1}w(A) - 1](4r'A)^{-1} \\ &\quad \times \sin[w(A)]\,\Phi(A) + (4r'A)^{-1}[\sin w(A) \\ &\quad - \sin(r - r' + 2r'A^2)]\,\Phi(A) + \int_1^A [\sin(r - r' + 2r'u^2) \\ &\quad - \sin w]\frac{d}{du}[(4r'u)^{-1}\Phi(u)]du \\ &\quad + \int_1^A \sin w(wr^{-1} - 1)\frac{d}{du}[(4r'u)^{-1}\Phi(u)]du \\ &\quad - \int_1^A w^{-1}\sin w\;\Phi(u)\,du \\ &\quad + \int_A^\infty [\cos w - \cos(r - r' + 2r'u^2)]\,\Phi(u)du.\end{aligned} \tag{A9}$$

From (A4), it is easy to show that $|\Phi_4(u)|$ is bounded by $Cu^{-3-\epsilon}$, and $|(d/du)u^{-1}\Phi_4(u)|$ by Cu^{-5}, whereas $|1 - r^{-1}w|$ is bounded by Cu^2r'/r, and, according to [I, (5.24)],

$$|\sin w - \sin(r - r' + 2r'u^2)| \leq Cu^2[(r'^2/r) + r'^2u^2(r + 2r'u^2)^{-1}]. \tag{A10}$$

Using these bound in (A9), with

$$A = r^{1/\epsilon}, \tag{A11}$$

we are led for the contribution of $\Phi(u)$ to the bound

$$Cr^{-1}[1 + r'(1 + \log r)].$$

Collecting these results and using notations referring to the parts of $K_0(r,r')$ which are concerned, we obtain, in the conditions (A1),

$$|\mathrm{Re}\mathcal{R}^0_0(r,r')| \leq Cr^{-1}(1 + r'^{3/2}), \tag{A12}$$

$$|\mathrm{Re}\mathcal{R}^1_0(r,r')| \leq Cr^{-1}[1 + r'(1 + \mathrm{Log}r)]. \tag{A13}$$

We can study in the same way

$$\mathrm{Im}\mathcal{R}_0(r,r') = \int_0^\infty \{\sin[r + r'(2u^2-1)] - \sin w\dot{w}/(2w)\}\Phi'(u)du. \quad \text{(A14)}$$

Since $|1 - \dot{w}/(2w)|$ is smaller than $Cr'u^2/r$ either for u smaller than one or for u larger than one, the replacement of $\dot{w}/(2w)$ by one affords only a bound Cr'/r, and the remaining term can be studied exacted like (A2). The bounds (A12) and (A13) therefore hold for $\mathrm{Im}\mathcal{R}_0(r,r')$. Replacing r by r' and r' by r lead in the same way to the bounds of $\mathcal{R}_0(r,r')$. Besides, it is easy to derive from (2.8) and (2.9) the values of $\mathcal{R}_1(r,r')$ and $\mathcal{R}_1'(r,r')$:

$$\mathcal{R}_1(r,r') = i\int_r^\infty \exp[i(r-\rho)]S(\rho,r')\rho^{-2}d\rho(r\to\infty), \quad \text{(A15)}$$

$$\mathcal{R}_1'(r,r') = i\int_{r'}^\infty \exp[i(r'-\rho)]Q(r,\rho)\rho^{-2}d\rho(r'\to\infty), \quad \text{(A16)}$$

from which the bounds of $|Q|$ and $|S|$ yield[49] (2.20). Let us now study the asymptotic behavior of (2.23). The asymptotic form is given by (2.42). For getting the bounds, we first need bounds for $|K_0^0(r,r')|$, $|K_0^1(r,r')|$, and $|K_1(r,r')|$. From [I, (3.1)] and [I, (3.4)], we know that

$$|K_0^0(r,r')| < \begin{cases} C(rr')^{1/2}(rr')^{1/2+\epsilon/2}[1+(rr')^{1/2+\epsilon/2}]^{-1} \\ C(rr')^{1/2}[1+|r-r'|]^{-1/2}. \end{cases} \quad \text{(A17)}$$

Besides, a careful analysis of (2.30), with $\Phi(u) = u^{-1}$, shows that

$$|K_0^0(r,r')| < \begin{cases} Cr\ \log(1+r) \\ Cr'\ \log(1+r') \end{cases}. \quad \text{(A18)}$$

Now, inserting $\Phi_3(u)$ and $\Phi_4(u)$ into (2.30), we easily show that [compare I, (3.39)]

$$|K_0^1(r,r')| < \begin{cases} C \\ Crr'[1+|r-r'|]^{-1} \end{cases}. \quad \text{(A19)}$$

From [I. (4.72)], [I, (4.85)–(4.87)], we get

$$|K_1(r,r')| < \begin{cases} C(rr')^{1/2}(rr')^{1/2+\epsilon/2}[1+(rr')^{(1+\epsilon)/2}]^{-1} \\ C \\ Cr \\ Cr' \end{cases}. \quad \text{(A20)}$$

It is interesting to combine these bounds so as to obtain more convenient ones for example, combining the three first bounds of $|K_1(r,r')|$, we can obtain

$$|K_1(r,r')| < C\frac{r}{1+r}r'^{\eta} \quad \text{(A21)}$$

and, from the ones of $K_0^0(r,r')$,

$$|K_0^0(r,r')| < \begin{cases} \{Cr\tau^{\eta}[1+\log(1+r)]\} \\ Cr^{1/2+\epsilon}r'^{1/2-\epsilon}[1+|\log(1+r)|]^{1/2} \end{cases}, \quad \text{(A22)}$$

where η is an arbitrary number of $[0,1)$.

On the other hand, we know from I that

$$|\mathcal{K}_0^0(\rho)| < C(1+\rho)^{-1/2}, \quad \text{(A23)}$$

$$|\mathcal{K}_2(\rho)| < C(1+\rho)^{-1}. \quad \text{(A24)}$$

From [I, (5.39)] and from the combination of (A18) and (A23), we know the following bounds for $|\mathcal{R}_0^0(r,r')|$:

$$|\mathcal{R}_0^0(r,r')| \le \begin{cases} Cr'|r-r'|^{-1/2} \\ C(rr')^{1/2} \end{cases}. \quad \text{(A25)}$$

From [I, (5.18)] and [I, (5.28)], we get

$$|\mathcal{R}_0^1(r,r')| \le \begin{cases} C \\ C[r'/r + r'^2/r + r'(r'/r)^{\epsilon/2}] \text{for } r' \le \frac{1}{2}r \end{cases} \quad \text{(A26)}$$

and from (A15) and [I, (4.82)]

$$|\mathcal{R}_1(r,r')| \le Cr'[1+(rr')^{1/2}]^{-1}. \quad \text{(A27)}$$

Similar bounds hold for the "primed" terms corresponding to the asymptotic behaviors when, for fixed r,r' goes to infinity. From the symmetry of $K_0(r,r')$, it follows that

$$\mathcal{R}_0'(r,r') = \mathcal{R}_0(r',r) \quad \text{(A28)}$$

(either with the subindex zero or the subindex one). From (2.9), we have

$$|\mathcal{R}_1'(r,r')| \le Cr[1+(rr')^{1/2}]^{-1}. \quad \text{(A29)}$$

Asymptotic Behavior of (2.40)

So as to study the asymptotic behavior of (2.40), we replace $K(\tau,r)$ by (2.39), and use its asymptotic behavior (2.17). The structural terms (2.42) are obtained in this way as candidates for the asymptotic form of $f(r,r')$. Differentiating (2.23) with respect to r leads one to

$$\frac{\partial}{\partial r}f(r,r') = \frac{\partial}{\partial r}K(r,r') + \int_0^r \frac{\partial}{\partial r}K(\tau,r)\,K(\tau,r')\tau^{-2}d\tau + r^{-2}K(r,r)K(r,r'). \quad \text{(A30)}$$

To show that the extra term $r^{-2}K(r,r)K(r,r')$ is a negligible function is trivial. Inserting the asymptotic form of $(\partial/\partial r)K(\tau,r)$ into (A30) leads one to a candidate for the asymptotic form of $f(r,r')$ which is consistent with the one obtained above, and written in (2.42). So as to prove that these "candidates" are really the asymptotic form, we have only to prove that both the contributions of the "remainders" and the contributions of the path (r,∞) in the integrations are negligible functions. This is easy to do with the help of the bounds (A19)–(A28). We give the details only for one of the cross products (actually the most difficult to handle):

$$I(r,r') = \int_0^r K_0^0(\tau,r)K_0^0(\tau,r')\tau^{-2}d\tau,$$
$$\bar{I}(r,r') = \int_0^r \frac{\partial}{\partial r}K_0^0(\tau,r)K_0^0(\tau,r')\tau^{-2}d\tau. \quad \text{(A31)}$$

A candidate for the asymptotic form of $I(r,r')$, whose

derivative is a candidate of the asymptotic form of $\bar{I}(r,r')$, is

$$\mathcal{J}(r,r') = \cos r \int_0^\infty k_0^0(\tau)K_0^0(\tau,r')\tau^{-2}d\tau + \sin r \int_0^\infty \bar{k}_0^0(\tau)K_0^0(\tau,r')\tau^{-2}d\tau \quad \text{(A32)}$$

Using (A23) and (A22a) we see that $|\mathcal{J}(r,r')|$ is bounded by Cr'. Using (A23) and (A17b), we obtain for $r' > 1$

$$|\mathcal{J}(r,r')| < Cr^{1/2}\int_0^{r'-1}\frac{u^{-1/2}du}{|1-u|^{1/2}} + C\int_{r'-1}^\infty \frac{du}{u|1-u|^{1/2}} \quad \text{(A33)}$$

so that, for large r', $|\mathcal{J}(r,r')|$ is bounded by $C\log r'$. Gathering the two bounds here obtained give the contribution of $|\mathcal{J}(r,r')|$. We have now to evaluate the remainders. Clearly we can limit our study to the following terms, $I_1(r,r')$ and $I_2(r,r')$,

$$I_1(r,r') = \int_r^\infty |k_0^0(\tau)|\,|K_0^0(\tau,r')|\,\tau^{-2}d\tau. \quad \text{(A34)}$$

Since $I(r)$ converges, $I_1(r,r')$ goes to zero as r goes to infinity. The evaluation of $\int_0^r \rho^{-1}(1+\rho)^{-1}I_1(r,\rho)d\rho$ leads to the evaluation of

$$\int_r^\infty |k_0(\tau)|\tau^{-2}d\tau \int_0^r \rho^{-1}(1+\rho)^{-1}|K_0^0(\tau,\rho)|\,d\tau, \quad \text{(A35)}$$

which, using (A22b), is easily proved to behave like $\int_r^\infty |k_0(\tau)|\tau^{-(3/2)-\epsilon}d\tau$, and therefore goes to zero as r goes to ∞. $I_1(r,r')$ is therefore a negligible function of r.

On the other hand, we pick the term

$$I_2(r,r') = \int_0^r |\mathcal{R}'^0_0(\tau,r')|K_0^0(\tau,r')|\tau^{-2}\,d\tau. \quad \text{(A36)}$$

Using (A22b) and (A25a), together with (A28) we get for $|I_2|$ the bound $Cr^{-(1/2)+\epsilon}r'^{1-\epsilon}$, which goes to zero as r goes to ∞. Using (A22b) and (A25a), we are able to obtain for $|\int_0^r \rho^{-1}(1+\rho)^{-1}|I_2(r,\rho)|d\rho$ the successive upper bounds $|\int_0^r |\mathcal{R}_0^0(r,\tau)|\tau^{-3/2-\epsilon}d\tau|$ and $Cr^{-\epsilon}$, which achieve to prove that $|I_2(r,r')|$ is a negligible function. It is clear that all the other terms encountered throughout the appraisal of the remainders for the asymptotic evaluation of (A31) can be studied either like $I_1(r,r')$ or like $I_2(r,r')$. The proofs for the other terms of $\int_0^r K(\tau,r)K(\tau,r')\tau^{-2}d\tau$ are similar and somewhat simpler.

Further Improvements of Bounds

The bounds we have obtained up to now are sufficient for deriving the asymptotic behavior of $\omega(r)$ and $\omega'(r)$, but not to get a fine appraisal of the remainders. So as to be able to use the results of Appendix B, we need better bounds for $\mathcal{R}'_1(r,r')$ and we need to introduce a special trick for the evaluation of $(\partial/\partial r)\int_0^r [K_0(\tau,r)]^2\tau^{-2}d\tau$ and similar quantities. Let us therefore study these tools by now.

A Fine Bound for $\mathcal{R}'_1(r,r')$

We first derive a fine bound for $M(r,r')$, using the results of I. Using the first of the bounds [I, (3.6)] for $|M_0(r,r')|$, and using the bound [I, (4.33)] for $|n(r,r\theta,\rho,\rho u)|$, we obtain for $|m_1^-(r,r\theta)|$, the following inequalities, which we use in place of [I, (4.57)] (for $\theta \geq 1$):

$$|m_1^-(r,r\theta)| \leq \int_1^\theta du \int_0^{(r^2\theta/u)^{1/2}} |n(r,r\theta,\rho,\rho u)| \times |m_0^-(\rho,\rho u|\rho d\rho \quad \text{(A37)}$$

$$\leq C\int_1^\theta u^{-1}(\theta-u)^{-1/4}[1-(\theta u)^{-1}]^{-1/4}du \times \int_0^{(r^2\theta/u)^{1/2}} (r^2\theta-\rho^2u)^{-1/4}|V(\rho)|\rho d\rho \quad \text{(A38)}$$

splitting the integration domain into two parts, and using for $|V(\rho)|$ the bound $C\rho^{-2}$ on the upper interval, this yields for the ρ integral the bound

$$C(r^2\theta)^{-1/4}\int_0^{(1/2)(r^2\theta/u)^{1/2}} |V(\rho)|\rho d\rho + C\int_{(1/2)(r^2\theta/u)^{1/2}}^{(r^2\theta/u)^{1/2}} (r^2\theta-\rho^2u)^{-1/4}\rho^{-1}d\rho \quad \text{(A39)}$$

and therefore, for $|m_1^-(r,r\theta)|$, the bound

$$C(r^2\theta)^{-1/4}\int_1^\theta u^{-1}(\theta-u)^{-1/4}[1-(\theta u)^{-1}]^{-1/4}du. \quad \text{(A40)}$$

This yields the bound

$$C(r^2\theta)^{-1/4}[1-\theta^{-1}]^{-1/4}(\theta-1)^{-1/4} \times \int_0^1 (1-s)^{-1/4}(\theta-1)[1+s(\theta-1)]^{-1}ds \quad \text{(A41)}$$

or

$$|m_1^-(r,r\theta)| < C(r'-r)^{-1/2}[1-\log(r'/r)] \quad \text{for } r' \geq r. \quad \text{(A42)}$$

The same device yields for $\theta \leq 1$

$$|m_1^-(r,r\theta)| \leq C(r^2\theta)^{-1/4}\int_\theta^1 (\theta^{-1}-u^{-1})^{-1/4} \times (1-\theta u)^{-1/4}u^{-1}du. \quad \text{(A43)}$$

The mapping $(\theta = \eta^{-1},\ u = t^{-1})$ yields an integral similar to the one in (A40), and we obtain for $|m_1(r,r\theta)|$ the following bound, valid for any positive number θ:

$$|m_1^-(r,r\theta)| \leq C\,|r'-r|^{-1/2}[1+|\log(r'/r)|]. \quad \text{(A44)}$$

For $|m_1^+(r,r\theta)|$, using the first and the third bound [I, (3.8)] in place of $|m_0^+(r,r\theta)|$, we obtain in the case $\theta \leq 1$,

$$|m_1^+(r,r\theta)| \leq C\int_\theta^1 (\theta^{-1}-u^{-1})^{-1/4}(1-\theta u)^{-1/4}u^{-1}du \times \int_0^{(r^2\theta/u)^{1/2}} (r^2\theta-\rho^2u)^{-1/4}\rho\,|V(\rho)|\rho^2u \times [1+\rho^{3/2}(1-u)^{3/2}]^{-1}d\rho. \quad \text{(A45)}$$

Let us now use in the ρ integral the inequalities

$$|V(\rho)| < C\rho^{-3-\epsilon}, \quad \text{(A46)}$$

$$1+X \geq X^\alpha \quad \text{for } X \geq 0,\ 0 < \alpha \leq 1. \quad \text{(A47)}$$

They enable us to bound the ρ integral by

$$C(r^2\theta)^{-1/4}u(1-u)^{-1+\epsilon}\Big[\int_0^{(1/2)(r^2\theta/u)^{1/2}} \rho^{2+\epsilon}|V(\rho)|d\rho + \int_{(1/2)(r^2\theta/u)^{1/2}}^{(r^2\theta/u)^{1/2}} \rho^{-1}d\rho\Big] \quad \text{(A48)}$$

or

$$C(r^2\theta)^{-1/4}u(1-u)^{-1+\epsilon}. \tag{A49}$$

Inserting (A49) into (A45), we easily obtain, since $(1-\theta)^\epsilon < 1$,

$$|m_1^+(r,r\theta)| \le C[r(1-\theta)]^{-1/2}, \quad \theta < 1. \tag{A50}$$

For θ larger than one, and provided that ϵ be smaller than $\frac{1}{4}$, it is easy to use the same token and obtain, for any value of r and r' in R^+,

$$|m_1^+(r,r')| \le C|r-r'|^{-1/2}. \tag{A51}$$

From (A44) and (A51) and, on the other hand, [I, (4.66)] and [I, (4.68)], we see that $M_1(r,r')$ is bounded as follows:

$$|M_1(r,r')| < \begin{cases} C(rr')^{1/2} \\ C|r-r'|^{-1/2}[1+|\log(r/r')|](rr')^{1/2}. \end{cases} \tag{A52}$$

If now we refer to the analysis done in I, and especially expressed, for instance, by [I, (4.50)], for appraising $|M(r,r') - M_0(r,r') - M_1(r,r')|$, we see that using the first bound (A52) leads one to a majoration like the second bound for this remainder. We can therefore write

$$|M(r,r') - M_0(r,r')| < \begin{cases} C(rr')^{1/2} \\ C|r-r'|^{-1/2}[1+|\log(r/r')|](rr')^{1/2}. \end{cases} \tag{A53}$$

and it is easy to see that the same bound (A53) holds for $|Q(r,r')|$ and $|S(r,r')|$. Now, since it follows from (2.8) that

$$|\mathcal{R}_1'(r,r')| < \int_{r'}^\infty |Q(r,\rho)|\rho^{-2}d\rho, \tag{A54}$$

we get, for $r' \ge r$,

$$|\mathcal{R}_1'(r,r')| \le [1+\log(r'/r)]r^{1/2}r'^{-1} \tag{A55}$$

Transformation of $\frac{\partial}{\partial r}K_0(r,r')$

We now derive a formula which enables us to transform all the integrals containing a derivative of $K_0(r,r')$. Starting from (2.27), applying $[(\partial/\partial r) + (\partial/\partial r')]$, and integrating by parts, we obtain

$$\left(\frac{\partial}{\partial r} + \frac{\partial}{\partial r'}\right)K_0(r,r') = \frac{1}{2}\frac{r+r'}{rr'}L_0(r,r'), \tag{A56}$$

where

$$L_0(r,r') = -\int_0^\infty [\cos w - \cos(r-r')]\frac{d}{du}(u\Phi'(u))du. \tag{A57}$$

Comparing (A57) and (2.27), we see that $L_0(r,r')$ is a function similar with $K_0(r,r')$. Clearly,

$$-\frac{d}{du}(u\Phi'(u)) = (2\pi)^{-1}V_0u^{-2} - \frac{d}{du}(u\Phi_3(u)) \quad \text{for } u \le 1.$$
$$= -\frac{d}{du}(u\Phi_4(u)) \quad \text{for } u \ge 1 \tag{A58}$$

Referring ourselves to (A3) and (A4), we can easily prove that $|(d/du)u\Phi_3(u)|$ is bounded by $Cu^{\epsilon-1}$ and $|(d/du)u\Phi_4(u)|$ is bounded by Cu^{-3}. It is possible to go further by noticing that $(d/du)(u^4\Phi_4(u))$ is the Fourier transform of a function belonging to $L_{1+\epsilon}(0,\infty)$. From the well-known theory[50] of the Fourier transform of L_p functions $(1 < p \le 2)$, we therefore obtain the bound

$$\left|\frac{d}{du}u\Phi_4(u)\right| < Cu^{-3}L(u), \tag{A59}$$

where $L(u)$ is a function of $L_{(1+\epsilon)/\epsilon}(0,\infty)$. With these bounds it is possible to obtain for $L_0(r,r')$ an asymptotic "structure" similar with that of $K_0(r,r')$, with a slight modification concerning the remainder "$|\mathcal{R}_0'(r,r')|$," in which, now, $Cr(r'/r)^{\epsilon/2}$ has to be replaced by $Cr'(r'/r)^{\epsilon/2(1+\epsilon)}$ which is obtained through Hölder's inequality.

The asymptotic form of (A56) strongly suggests the following equalities:

$$\bar{k}_0'(\rho) + \frac{\partial k_0'(\rho)}{\partial\rho} = \tfrac{1}{2}\rho^{-1}l_0'(\rho), \tag{A60}$$

$$-k_0'(\rho) + \frac{\partial \bar{k}_0'(\rho)}{\partial\rho} = \tfrac{1}{2}\rho^{-1}\bar{l}_0'(\rho). \tag{A61}$$

Both are easily proved, with the help of elementary algebra and integrations by parts, from the formulas

$$k_0'(\rho) = 2\int_0^\infty \sin(\rho u^2)\sin[\rho(1-u^2)]\Phi'(u)du, \tag{A62}$$

$$\bar{k}_0'(\rho) = -2\int_0^\infty \sin(\rho u^2)\cos[\rho(1-u^2)]\Phi'(u)du, \tag{A63}$$

$$l_0'(\rho) = -2\int_0^\infty \sin(\rho\mu^2)\sin[\rho(1-\mu^2)]\frac{d}{du}(u\Phi'(u))du, \tag{A64}$$

$$\bar{l}_0'(\rho) = 2\int_0^\infty \sin(\rho u^2)\cos[\rho(1-u^2)]\frac{d}{du}(u\Phi'(u))du. \tag{A65}$$

It is interesting to notice that, for a potential of class $\mathcal{E}$, just like $K_0(r,r')$ behaves like $(rr')^{1+\epsilon}$ as rr' goes to zero, the s.s. functions go to zero like $\rho^{1+\epsilon/2}$ as ρ goes to zero. This is easy to prove. In relation to that, we clearly notice also that from (A3) and (A4) we have, for a potential of class $\mathcal{E}$,

$$\int_0^\infty u^2\Phi'(u)du = 0. \tag{A66}$$

Asymptotic Behavior of $\omega(r)$

From (2.23) and (2.26), we obtain

$$\omega(r) = V(r) + 2r^{-3}\int_0^r [K(\rho,r)]^2\rho^{-2}d\rho - 2r^{-4}[K(r,r)]^2 - 4r^{-2}\int_0^r K(\rho,r)\left(\frac{\partial}{\partial r}K(\rho,r)\right)\rho^{-2}d\rho. \tag{A67}$$

We are interested in the behavior of $\omega(r)$ for r much larger than one. From (A17), (A19), and (A20), it is easy to see that the second term of (A67) is absolutely bounded by $Cr^{-3+\epsilon}$, a bound which obviously holds for the first and the third term. Let us now study the last term and, for this, the following integrals:

$$A = \int_0^r K_0(\rho,r)\left(\frac{\partial}{\partial r}K_0(\rho,r)\right)\rho^{-2}d\rho, \tag{A68}$$

$$B = \int_0^r K_1(\rho,r)\left(\frac{\partial}{\partial r}K_0(\rho,r)\right)\rho^{-2}d\rho, \tag{A69}$$

$$C = \int_0^r K_0(\rho,r)\left(\frac{\partial}{\partial r}K_1(\rho,r)\right)\rho^{-2}d\rho, \tag{A70}$$

$$D = \int_0^r K_1(\rho,r)\left(\frac{\partial}{\partial r}K_1(\rho,r)\right)\rho^{-2}d\rho. \tag{A71}$$

From (A56) we easily obtain

$$A = \tfrac{1}{2}\int_0^r K_0(\rho,r)L_0(\rho,r)\rho^{-3}d\rho - \int_0^r [K_0(\rho,r)]^2\rho^{-3}d\rho + \tfrac{1}{2}r^{-1}\int_0^r K_0(\rho,r)L_0(\rho,r)\rho^{-2}d\rho - \tfrac{1}{2}r^{-2}[K_0(r,r)]^2. \quad (A72)$$

The third and the fourth term in (A72) are bounded by $Cr^{-1+\epsilon}$ as r goes to infinity. Inside the first and the second one, it is possible to replace K_0 and L_0 by their asymptotic behavior and increase the domain of integration up to ∞. Using (A17), (A19), and (A20), we see that the missing parts of the integrals (from r to ∞), together with the parts between $\frac{1}{2}r$ and r, are bounded by

$$C\int_{r/2}^{\infty}\rho^{-3}d\rho + C\int_{r/2}^{\infty} r\rho[1+|\rho-r|]^{-1}\rho^{-3}d\rho \quad (A73)$$

and therefore by $Cr^{-1}\log r$ or $Cr^{-1+\epsilon}$. Besides, the contribution of the remainders $|\mathcal{R}_0'(\tau,r)|$ or similar terms is bounded, using (A12), by

$$Cr^{-1}\int_0^{r/2}(1+\rho^{1/2})^{-1}\rho^2(r-\rho)^{-1/2}(r\rho)^{1/2}\rho^{-3}d\rho \quad (A74)$$

or $Cr^{-1+\epsilon}$. The replacement of K_0 and L_0 in (A72) by their asymptotic behavior introduces therefore errors of the order of $r^{-1+\epsilon}$. The leading term is easily seen to be

$$A = (\cos^2 r)\left(\tfrac{1}{2}\int_0^{\infty} k_0'(\rho)l_0'(\rho)\rho^{-3}d\rho - \int_0^{\infty}[k_0'(\rho)]^2\rho^{-3}d\rho\right) + (\sin^2 r)\left(\tfrac{1}{2}\int_0^{\infty}\bar{k}_0'(\rho)\bar{l}_0'(\rho)\rho^{-3}d\rho - \int_0^{\infty}[\bar{k}_0'(\rho)]^2\rho^{-3}d\rho\right) + \sin r\cos r\left(\tfrac{1}{2}\int_0^{\infty}[k_0'(\rho)\bar{l}_0'(\rho) + \bar{k}_0'(\rho)l_0'(\rho)]\rho^{-3}d\rho - 2\int_0^{\infty}k_0'(\rho)\bar{k}_0'(\rho)\rho^{-3}d\rho\right) + O(r^{-1+\epsilon}). \quad (A75)$$

Using (A60) and (A61), it is easy to reduce (A75) to:

$$A = (\cos^2 r - \sin^2 r)\int_0^{\infty}\bar{k}_0'(\rho)k_0'(\rho)\rho^{-2}d\rho + \sin r\cos r\left(\int_0^{\infty}\rho^{-2}\{[\bar{k}_0'(\rho)]^2 - [k_0'(\rho)]^2\}d\rho\right) + O(r^{-1+\epsilon}), \quad (A76)$$

which is precisely the formula we could obtain by replacing readily K_0 by its asymptotic behavior (but we could not appraise the remainder in such a way). Furthermore, it is clear from (2.16) that the integrals in (A76) are only semiconvergent.

Let us now study B and C, and, for this, recall that

$$|K_1(\rho,r) - [k(\rho,\rho)]^2\sin(\rho-r) - \int_{\rho}^{\infty}\sin(r-\tau)Q(\rho,\tau)\tau^{-2}d\tau| \le \mathcal{R}_1'(\rho,r), \quad (A77)$$

$$\left|\frac{\partial}{\partial r}K_1(\rho,r) + [k(\rho,\rho)]^2\cos(\rho-r) + \int_{\rho}^{\infty}\cos(r-\tau)Q(\rho,\tau)\tau^{-2}d\tau\right| \le \mathcal{R}_1'(\rho,r). \quad (A78)$$

Using in (A69) and (A70) the bounds (A19), (A23), (A25), and (A26) for the terms indexed 0, it is easy to see that replacing $K_1(\rho,r)$ and $(\partial/\partial r)K_1(\rho,r)$ by their asymptotic behavior introduces only errors of the order of $Cr^{-1+\epsilon}$. Let us now study

$$B_1 = -\int_0^r\left(\frac{\partial}{\partial r}K_0(\rho,r)\right)\left(\int_{\rho}^{\infty}\sin(r-\tau)Q(\rho,\tau)\tau^{-2}d\tau\right)\times\rho^{-2}d\rho, \quad (A79)$$

$$C_1 = \int_0^r K_0(\rho,r)\left(\int_{\rho}^{\infty}\cos(r-\tau)Q(\rho,\tau)\tau^{-2}d\tau\right)\rho^{-2}d\rho. \quad (A80)$$

The integrals containing Q can be absolutely bounded by $C\rho^{-1/2}$. Using this and again (A19), (A23), (A25), and (A26), it is easy to prove that the terms indexed 0 in (A79) and (A80) can be replaced by their asymptotic behavior, and the $(0,r)$ interval of integration by $(0,\infty)$ with an error smaller than $Cr^{-3+\epsilon}$. Hence we are led to study

$$B_2 = \int_0^r [k(\rho,\rho)]^2\sin(\rho-r)\left(\frac{\partial}{\partial r}K_0(\rho,r)\right)\rho^{-2}d\rho, \quad (A81)$$

$$C_2 = -\int_0^r [k(\rho,\rho)]^2\cos(\rho-r)K_0(\rho,r)\rho^{-2}d\rho. \quad (A82)$$

Let us now use (A56) and integrate by parts. We obtain

$$B_2 + C_2 = \int_0^r K_0(\rho,r)\sin(\rho-r)\frac{d}{d\rho}[k(\rho,\rho)]^2\rho^{-2}d\rho + (2r)^{-1}\int_0^r [k(\rho,\rho)]^2L_0(\rho,r)\sin(\rho-r)\rho^{-2}d\rho + \tfrac{1}{2}\int_0^r [k(\rho,\rho)]^2\sin(\rho-r)L_0(\rho,r)\rho^{-3}d\rho - 2\int_0^r K_0(\rho,r)[k(\rho,\rho)]^2\sin(\rho-r)\rho^{-3}d\rho. \quad (A83)$$

In (A83) the remainders coming from the replacement of the integration interval $(0,r)$ by $(0,\infty)$, and the various remainders coming from the replacement of K_0 and L_0 by their asymptotic behavior can easily be appraised using (A23) and (A25), together with the second term in (A83). All these are bounded by $Cr^{-1+\epsilon}$ for large r. The asymptotic form of (A83) can again be transformed using (A60) and (A61). As expected, the result is the one we could have obtained readily by substituting in B and C the asymptotic behavior of K_0 and K_1.

Such a direct substitution can be done in D, and the use of (A55) readily yields remainders of order $Cr^{-1+\epsilon}$. From the above analysis we therefore conclude that

$$(A+B+C+D) = (\cos^2 r - \sin^2 r)\int_0^{\infty}\bar{k}'(\rho)k'(\rho)\rho^{-2}d\rho + \sin r\cos r\int_0^{\infty}\{[\bar{k}'(\rho)]^2 - [k'(\rho)]^2\}\rho^{-2}d\rho + O(r^{-1+\epsilon}) \quad (A84)$$

and therefore, as $r \to \infty$,

$$\omega(r) = -4r^{-2}[\cos 2r\int_0^{\infty}\bar{k}'(\rho)k'(\rho)\rho^{-2}d\rho + \tfrac{1}{2}\sin 2r\int_0^{\infty}\{[\bar{k}'(\rho)]^2 - [k'(\rho)]^2\}\rho^{-2}d\rho] + O(r^{-3+\epsilon}). \quad (A85)$$

Study of $B(r)$

From (2.60) and (2.64) we want to prove that a and b can be chosen in such a way that $B(r)$ is equal to zero for $r < \mathrm{Inf}(a,b)$. Let us prove this for a. We have to prove that a can be chosen in such a way that $\int_a^{\infty}\cos\rho\,\rho^{-1}d\rho$ be equal to zero. Let a be smaller than $\pi/2$.

Integrating once by parts, it is easy to see that

$$\left| \int_{\pi}^{\infty} \cos\rho\, \rho^{-1} d\rho \right| \leq \pi^{-1}. \tag{A86}$$

Now $\int_{\pi/2}^{\pi} \cos\rho\, \rho^{-1} d\rho$ is obviously negative, and smaller than $-\pi^{-1} \int_{\pi/2}^{\pi} |\cos\rho| d\rho$ or $-\pi^{-1}$. Hence,

$$\int_{\pi/2}^{\infty} \cos\rho\, \rho^{-1} d\rho \leq 0. \tag{A87}$$

It is easy to see that for any positive a, $\int_a^{\infty} \cos\rho\, \rho^{-1} d\rho$ is a continuous function. But $\int_{\epsilon}^{\pi/2} \cos\rho\, \rho^{-1} d\rho$ is larger than $\int_{\epsilon}^{\pi/4} \cos\rho\, \rho^{-1} d\rho$, and therefore larger than $\frac{1}{2}\sqrt{2} \log(\frac{1}{4}\pi\epsilon)$. Choosing ϵ small enough, we see that we succeed to prove that $\int_{\epsilon}^{\infty} \cos\rho\, \rho^{-1} d\rho$ can be made ≥ 0. Since it is a continuous function, it has therefore a zero between ϵ and $\pi/2$. We choose a equal to this zero. The proof for b is somewhat simpler.

Further Bounds of $K_0^0(r,r')$ and $\frac{\partial}{\partial r} K_0(r,r')$

So as to get a better bound for $K_0^0(r,r')$, we start from the formula

$$K_0^0(r,r') = C_0 rr' \int_0^1 w^{-1} \sin(w) du, \tag{A88}$$

which follows from [I, (3.13)], and we successively study the cases $(r \leq 2, r' \leq 10)$, $(r \leq 2, r' > 10)$, $(r \geq 2, r' \leq 10)$, $(r \geq 2, r' \geq 10, |r - r'| \geq 0.1r')$, $(r \geq 2, r' \geq 10$, and $|r - r'| < 0.1r')$.

We easily obtain

$$|K_0^0(r,r')| \leq Cr \log(1 + r). \tag{A89}$$

From (A88), we derive the following bounds:

$$\left| \frac{\partial}{\partial r} K_0^0(r,r') \right|$$

$$\leq r' \left| \int_0^1 w^{-1} \sin w du \right| + rr' \int_0^1 |\cos w - w^{-1} \sin w| \times |r - r' + 2r'u^2| w^{-2} du \tag{A90}$$

$$\leq Cr' + Crr' \int_0^1 |r - r' + 2r'u^2| (1 + w^2)^{-1} du \tag{A91}$$

$$\leq Cr' \quad \text{for } r' \leq r. \tag{A92}$$

Besides, the first term in the rhs of (A90) is bounded by $Cr'(r - r')^{-(1-\epsilon)/2} (rr')^{-(1+\epsilon)/4}$. The second one (before taking the absolute value), can be integrated by parts, after having splitted the integration interval. We obtain in this way as an upper bound,

$$C|rr' \int_0^{\alpha} w^{-2} |r - r' + 2r'u^2| du + r' \int_{\alpha}^1 w^{-1} du + 1 + \alpha^{-1}|. \tag{A93}$$

w^2 in the first term of (A93) can be replaced by $(r - r')^{2-\epsilon} (rr')^{\epsilon/2} u^{\epsilon}$.

Taking $\alpha = (rr')^{-(1/2)+1/4\epsilon} (r - r')^{+(1/2)-(\epsilon/2)}$, we obtain by this way the bound

$$\left| \frac{\partial}{\partial r} K_0^0(r,r') \right| \leq C[1 + (rr')^{(1/2)-1/4\epsilon} (r - r')^{-(1/2)+\epsilon/2} + r'(r - r')^{-(1-\epsilon)/2} (rr')^{-(1+\epsilon)/4}. \tag{A94}$$

It is also useful to extend the domain of validity of the bounds obtained in I for $|(\partial/\partial r) K_0'(r,r')|$. This can be done easily if we notice on formulas [I, (3.28)] and [I, (3.29)] that

$$\left| \frac{\partial}{\partial r} C(r,r',u) \right| < C|r - r' - w + 2r'u^2| \tag{A95}$$

$$< Cr'u^2 |rr'(1 - u^2)| \, |(r + w)^2 - r'^2|^{-1} \tag{A96}$$

$$< Cr' |1 - u^2| \quad \text{for } r' \leq r. \tag{A97}$$

Hence, following from the properties of Φ_3 and Φ_4 recalled at the beginning of the present appendix, we get

$$\left| \frac{\partial}{\partial r} K_0'(r,r') \right| \leq Cr' \quad \text{for } r' \leq r. \tag{A98}$$

Bounds of $r\omega_1'(r)$

From (A67), we derive the value of $\omega_1'(r)$:

$$\omega_1'(r) = 8r^{-5}[K(r,r)]^2 - 2r^{-2}V(r)K(r,r) - 6r^{-4} \int_0^r [K(\rho,r)]^2 \rho^{-2} d\rho + 12r^{-3} \int_0^r K(\rho,r) \frac{\partial}{\partial r} K(\rho,r) \rho^{-2} d\rho - 4r^{-2} \int_0^r \left(\frac{\partial}{\partial r} K(\rho,r) \right)^2 \rho^{-2} d\rho - 4r^{-2} \int_0^r K(\rho,r) \frac{\partial^2}{\partial r^2} K(\rho,r) \rho^{-2} d\rho. \tag{A99}$$

Using alternately for $|K(r,r')|$ the bounds recalled in (A17), (A18), (A19), and (A20), for $(\partial/\partial r) K_0^0(r,r')|$ the bounds (A92) and (A94), for $|(\partial/\partial r)K(r,r') - (\partial/\partial r)K_0^0(r,r')|$ the bounds Cr' and C which both hold if $r' \leq r$ {see (A98), [I, (3.40)], and [I, (4.90)]}, we easily obtain

$$r|\omega_1'(r)| \leq \begin{cases} C \\ Cr^{-1+\epsilon}. \end{cases} \tag{A100}$$

APPENDIX B: "PROJECTED FUNCTIONS"

Let us be given a function $\sigma(\rho)$ such that

(1) $\rho\sigma(\rho)$ is a continuous function of ρ, which goes to zero faster than $C\rho^{-2+\epsilon}$ as ρ goes to ∞;

(2) $$\int_0^{\infty} \rho\sigma(\rho) = 0; \tag{B1}$$

(3) The function $\sum(r,r')$, defined by

$$\sum(r,r') = -\tfrac{1}{2}(rr')^{1/2} \int_0^{(rr')^{1/2}} \times J_0[(r - r')(1 - \rho^2/rr')^{1/2}] \rho\sigma(\rho) d\rho \tag{B2}$$

can be expanded as

$$\sum(r,r') = \cos r \sum(r') + \sin r \, \overline{\sum}(r') + \sum_N(r,r') \tag{B3}$$

$$\frac{\partial}{\partial r} \sum(r,r') = -\sin r \sum(r') + \cos r \, \overline{\sum}(r') + \overline{\sum}_N(r,r'), \tag{B4}$$

where $\sum_N$ and $\overline{\sum}_N$ are negligible functions. Using well-known formulas,[51] we can write (B2) in the form

$$\sum(r,r') = -2\pi^{-1}rr' \int_0^\infty \rho\sigma(\rho)d\rho \int_0^\infty w^{-1} \times \sin w \cos(2\rho u)du, \quad \text{(B5)}$$

where w is defined by (2.31). The projected function $\sum_p(r,r')$ is equal to

$$\sum_p(r,r') = -2\pi^{-1}rr' \int_0^\infty \rho\sigma(\rho)d\rho \int_0^1 w^{-1} \times \sin w \cos(2\rho u)du. \quad \text{(B6)}$$

Now, let us set

$$P(r,r',\rho) = rr' \int_0^1 w^{-1} \sin w \cos(2\rho u)du = \tfrac{1}{2}\sin r \sin r' \cos 2\rho - \tfrac{1}{4}\int_0^1 G(r,r',u)\frac{d}{du} \times [(1-\cos 2\rho u)u^{-1}]du + \tfrac{1}{4}\int_0^1 G(r,r',u)u^{-2}du. \quad \text{(B7)}$$

Inserting the left-hand side of (B7) into (B6), we notice that, due to (B1), the contribution of the last term of (B7) vanishes. We therefore obtain

$$\sum_p(r,r') = -\pi^{-1}\sin r \sin r' \int_0^\infty \rho\sigma(\rho)\cos(2\rho)d\rho + (2\pi)^{-1}\int_0^1 G(r,r',u)u^{-1}\Omega(u)du, \quad \text{(B8)}$$

where

$$\Omega(u) = u\int_0^\infty \rho\sigma(\rho)\frac{d}{du}\left(\frac{1-\cos 2\rho u}{u}\right)d\rho. \quad \text{(B9)}$$

Clearly,

$$\Omega(u) = 2\int_0^\infty \rho^2\sigma(\rho)\sin(2\rho u)d\rho - \int_0^\infty \rho^2\sigma(\rho) \times [(\rho u)^{-1}(1-\cos 2\rho u)]du. \quad \text{(B10)}$$

Since $\rho^2\sigma(\rho)$ is a function of $L_2(0,\infty)$, the first term in the right-hand side of (B10) is a function of $L_2(0,\infty)$. Noticing now that $\rho^{5/2-\epsilon}\sigma(\rho)$ is a function of $L_2(0,\infty)$, and with the help of Schwarz's inequality, we are able to get for the second term the bound $Cu^{-\epsilon}$. We therefore can write

$$\int_0^1 \Omega^2(u)du < C. \quad \text{(B11)}$$

We now intend to show that $\sum_p(r,r')$ can be expanded like $\sum(r,r')$ is in (B3) and (B4). For this, we use a technique studied in Paper I, and available for integral expressions like (B8). According to [I, (5.3)] and [I, (5.4)], the terms remaining when $G(r,r',u)$ has been replaced by its asymptotic behavior in (B8) are bounded for $r' \le r$ by

$$C\int_0^1 |1 - \dot{w}/(2w)|u^{-1}|\Omega(u)|du + C\int_0^1 u^{-1}|\Omega(u)|\mathrm{Sup}[C, |r-r'-w+2r'u^2|]|du. \quad \text{(B12)}$$

Using [I, (3.31)] and [I, (5.6)], we readily obtain for the remainders the bound

$$R \le Cr'/r + Cr'^2/r \quad \text{(B13)}$$

valid for $r' \le \frac{1}{2}r$.

For r' between $\frac{1}{2}r$ and r, it follows from (B12) the bound

$$C\int_0^1 r'^2u[(r-r')^2 + 4rr'u^2|^{-1}|\Omega(u)|du + C\int_\alpha^1 u^{-1}|\Omega(u)|du + C\int_0^\alpha rr'u[(r-r')^2 + 4rr'u^2|^{-1/2}|\Omega(u)|du. \quad \text{(B14)}$$

Using the Schwarz's inequality and taking $\alpha = (rr')^{-1/3}$ yields, after some elementary inequalities, the bound

$$R \le Cr'^2(rr')^{-3/4}(r-r')^{-1/2}\left(\log\frac{r+r'}{r-r'}\right)^{1/2} + C(rr')^{1/6} + C(rr')^{5/12}(r-r')^{-1/2} \quad \text{(B15)}$$

valid for $r' \le r$.

It is easy to see that the bounds (B13) and (B15) for the remainders prove that these are negligible function. Indeed, the fact the remainders go to zero as r goes to infinity is obvious in (B13). The fact that the integral

$$\int_0^r |R(r,r')|r'^{-1}(1+r')^{-1}dr' \quad \text{(B16)}$$

goes to zero as r goes to infinity is readily seen by using (B13) between 0 and $r^{1/2}$, (B15) between $r^{1/2}$ and r.

Inserting in (B8) the asymptotic behavior of $G(r,r',u)$ as we did in I, we obtain for $\sum_p(r,r')$ expansions (B3) and (B4), with the substitution

$$\sum \to \sum_P, \quad \text{(B17)}$$

where

$$\sum_P(r') = 2\int_0^1 \sin(r'u^2)\sin[r'(1-u^2)]u^{-1}\Omega(u)du \quad \text{(B18)}$$

$$\bar{\sum}_P(r') = -\pi^{-1}\sin r' \int_0^\infty \rho\sigma(\rho)\cos(2\rho)d\rho - 2\int_0^1 \sin(r'u^2)\cos[r'(1-u^2)]u^{-1}\Omega(u)du. \quad \text{(B19)}$$

So as to obtain bounds for $\sum_P(r')$, let us insert (B9) in (B18):

$$\sum_P(r') = \int_0^\infty \rho\sigma(\rho)G(r',\rho)d\rho, \quad \text{(B20)}$$

where

$$G(r',\rho) = 2\int_0^1 \sin(r'u^2)\sin[r'(1-u^2)]\frac{d}{du} \times [u^{-1}(1-\cos 2\rho u)]du \quad \text{(B21)}$$

$$= 4r'\int_0^1 (1-\cos 2\rho u)\sin[r'(2u^2-1)]du. \quad \text{(B22)}$$

Clearly $|G(r',\rho)|$ is bounded by Cr'. Using now for $|u^{-1}\sin(2\rho u)|$ and $|u^{-2}(1-\cos 2\rho u)|$ the bound $u^{-1}(\rho u)^\epsilon$, it is easy to show from (B21) that $|G(r',\rho)|$ is also bounded by $C\rho^{1+\epsilon}$. Inserting these bounds in (B20) we get for $\sum_P(r')$ the bounds

$$|\sum_P(r')| < \begin{matrix} Cr' \\ Cr'^\epsilon \end{matrix}. \quad \text{(B23)}$$

Similar remarks lead to

$$|\bar{\sum}_P(r')| < \begin{matrix} Cr' \\ Cr'^\epsilon \end{matrix}. \quad \text{(B24)}$$

It must be noticed that, throughout the derivation of these properties of $\sum_P(r,r')$, we never have used Assumption (3) on $\sum(r,r')$, as expressed by (B3) and (B4). Only, if we are interested in the function

$$\sum_Q(r,r') = -2\pi^{-1}rr' \int_0^\infty \rho V(\rho)d\rho \int_1^\infty w^{-1} \times \sin w \cos(2\rho u)du, \quad \text{(B25)}$$

we can guarantee for this function the existence of an expansion like (B3) and (B4) if such an expansion

exists for $\sum(r,r')$, and bounds like (B23) and (B24) if they exist for $\sum(r')$ and $\overline{\sum}(r')$. All these results are obvious since

$$\sum_Q(r,r') = \sum(r,r') - \sum_P(r,r'). \tag{B26}$$

Let us notice also that the expansions (B3) and (B4) certainly hold if $\sigma(\rho)$ goes to zero faster than $Cr^{-3-\epsilon}$, according to a study done in the Sec. 7 of I.

APPENDIX C

We start from (3.65), and split the integration interval by the bound two. Since $|F'(u)|$ is integrable on $(2,\infty)$, and behaves like $C(u-1)^{-1}$ as u goes to one, it is easy to check that the contribution of the upper interval is absolutely bounded by C, the other by $C\{1+\log[(r+r')/rr']\}$. From (3.65) we then obtain

$$-\frac{\partial f(r,r')}{\partial r} = \int_1^\infty \left(\sin w\,\frac{\dot w}{2w} - \sin(r+r')\right)F'(u)du + \delta\cos r\sin r'. \tag{C1}$$

Again, we split the integration interval. On $(1,2)$, we are led to integrals of the form

$$\int_1^2 \left|\frac{\dot w}{(2w)} - 1\right|(u-1)^{-1}du, \tag{C2}$$

which is bounded by C, and

$$\int_1^2 \mathrm{Sup}[|r+r'-w|, C](u-1)^{-1}du, \tag{C3}$$

which is again bounded by $C(1+\log(r^{-1}+r'^{-1}))$. We are therefore led to study

$$\int_2^\infty \sin w\; G(u)\,\frac{\dot w}{(2w)}\,u^{-1}du, \tag{C4}$$

where $G(u)\in L_2(2,\infty)$. Let us now introduce the notations

$$\left.\begin{aligned}\beta &= \tfrac{1}{2}(rr')^{-1/2}|r-r'|\\ x &= (u^2+\beta^2)^{1/2}\end{aligned}\right\}, \tag{C5}$$

we can put (C4) into the form

$$\int_{(4+\beta^2)^{1/2}}^\infty \sin[2(rr')^{1/2}x]\,[\beta(x^2-\beta^2)^{-1} + (r'/r)^{1/2}]\times G(\sqrt{x^2-\beta^2})dx. \tag{C6}$$

Using the Parseval theorem, we therefore obtain

$$\left|\frac{\partial f(r,r')}{\partial r}\right| < C[1+\log(r^{-1}+r'^{-1})] + \mathrm{Sup}\{C,(rr')^{-1/4}|r-r'|^{1/2}L_2[(rr')^{1/2}]\} + C(r'/r)^{1/2}L_2[(rr')^{1/2}]. \tag{C7}$$

Together with the bounds for small r', this is sufficient to guarantee that it is possible to differentiate almost everywhere an integral of the form $\int f(r,\rho)g(\rho,r')\rho^{-2}d\rho$.

---

* Supported in part by the U.S. Army Research Office, Durham, North Carolina.

† Postal Address: Département de Physique Mathématique, Faculté des Sciences, 34 - Montpellier, France.

‡ Physique Mathématique et Théorique, Equipe de Recherche associée au CNRS.

1 In the following, C is a general constant, **C** a general positive constant, and ϵ an arbitrarily small positive number. $L_p(x)$ is a general function belonging to L_p, either on R^+ or, when stated otherwise, in a given part of R^+. These quantities are not intended to have the same value every time they are used. A review of the main results given in the present paper is available in a lecture of the author, reviewing the inverse scattering problem at fixed energy, and to be published in the Proceedings of the Workshop on the Mathematics of Profile Inversion, held at Ames Research Center, Moffet Field, California in July 1971.

2 Q. B. Firsov, Zh. Eksp. Teor. Fiz. **24**, 279 (1953).

3 J. A. Wheeler, Phys. Rev. **99**, 630 (1955).

4 P. C. Sabatier, "Le problème inverse à énergie donnée en mécanique quantique," Thèse Faculté des Sciences Orsay (1966), pp. 145–54.

5 W. H. Miller, J. Chem. Phys. **51**, 3631 (1969).

6 G. Vollmer, Z. Physik, **226**, 423 (1969).

7 A. Martin and G. Targonski, Nuovo Cimento **20**, 1182 (1961).

8 A. Martin (private communication).

9 T. Regge, Nuovo Cimento **14**, 951 (1949).

10 J. J. Loeffel, Ann. Inst. Henri Poincaré **8**, 339 (1968).

11 P. C. Sabatier, J. Math. Phys. **8**, 1957 (1967).

12 G. Burdet, M. Giffon, and E. Predazzi, Nuovo Cimento **36**, 1337 (1965).

13 G. A. Viano, Nuovo Cimento **63**, 581 (1969).

14 R. G. Newton, J. Math. Phys. **3**, 75 (1962).

15 P. C. Sabatier, J. Math. Phys. **7**, 1515 (1966).

16 P. C. Sabatier, J. Math. Phys. **7**, 2079 (1966).

17 P. C. Sabatier, J. Math. Phys. **8**, 905 (1967).

18 P. C. Sabatier, J. Math. Phys. **9**, 1241 (1968), Sec. 2A.

19 R. G. Newton, J. Math. Phys. **8**, 1556 (1967).

20 J. R. Cox and K. W. Thompson, J. Math. Phys. **11**, 804, 815 (1970).

21 See P. C. Sabatier, Ref. (15), Eqs. (3.66) and (3.67); Ref. 17, Eqs. (3.8) and (3.16) for $\alpha = \beta$.

22 Quyen-Van Phu, Thesis (unpublished).

23 P. C. Sabatier and Quyen Van Phu, Phys. Rev. D **4**, 127 (1971).

24 C. Coudray and M. Coz, Ann. Phys. (N.Y.) **61**, 488 (1970).

25 C. Coudray and M. Coz, J. Math. Phys. **12**, 1166 (1971).

26 P. C. Sabatier, J. Math. Phys. **9**, 1241 (1968).

27 M. Hooshyar, J. Math. Phys. **12**, 2243 (1971).

28 P. C. Sabatier, Compt. Rend. **265**, 5 (1967).

29 P. C. Sabatier, J. Math. Phys. **12**, 1393 (1971).

30 In the logical order, Ref. 29 can be labelled as I, the present paper as II, Ref. 23 as IV, other parts being in preparation. In the following, we sometimes refer to formulas of I by their number following the label I.

31 The details of these tedious calculations are given, for example, in Ref. 4, p. 195.

32 P. C. Sabatier (to be published).

33 R. G. Newton, Phys. Rev. **101**, 1588 (1956); V. S. Buslaev, Dokl. Akad. Nauk SSSR **5**, 1067 (1962) [Sov. Phys. Dokl. **7**, 295 (1962)]; F. Calogero and A. Degasperis, J. Math. Phys. **9**, 90 (1968).

34 To compare these formulas with those written in I. one should keep in mind that $k = 1$ in the present paper.

35 See for example Ref. 11, formula (1.36b) for $z = z_1$.

36 (2.24) is similar to [I, (2.21)] taken for $V = 0$, which hence reduces to [I, (1.22)]. This is not surprising, since $K_0(r,r')$ is a solution of (1.10).

37 The derivation is formal. The justification is easy, using the properties of $\omega(r)$ derived below.

38 A. Erdélyi *et al.*, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Formula 7.7.(5).

39 A. Erdélyi *et al.*, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Formula 1.13(47).

40 It is easy to see that the simple reductions of the class $\mathfrak{F}_2$ we do later can be done consistently with trivial modifications of the exceptional set.

41 Reference 38, Formula 7.15(30).

42 Reference 17, Appendix A.

43 Use the fact, proved in Ref. 15, that $|u_l(r)|$, for any r, is bounded by $Cl^{1/3}$, and that the maximum, for large l, is for $r \sim l$, so that it can easily be seen that $|r^{-1}u_l(r)|$ is bounded by $Cl^{-2/3}$, etc.

44 Reference 38, Formula 7.7(29).

45 Reference 15, Sec. 2

46 Reference 16, Appendix III.

47 Reference 15, Sec. 3 and Appendix.

48 A. Martin, "Progress in Elementary Particle and Cosmic Ray Physics. VIII," Formula 136.

49 These bounds are sharper than those obtained in I.

50 E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, (Oxford U.P., New York, 1937).

51 Reference 39, Formula 1.7(30).

On the Stability of Periodic Orbits for Nonlinear Oscillator Systems in Regions Exhibiting Stochastic Behavior*

Gary H. Lunsford and Joseph Ford
School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332
(Received 5 January 1972)

A computer has been used to determine the stability character of periodic orbits for the Hamiltonian oscillator system $H = \frac{1}{2}(p_1^2 + p_2^2 + q_1^2 + q_2^2) + q_1^2 q_2 - \frac{1}{3}q_2^3$. Using procedures developed by Greene [J. Math. Phys. 9, 760 (1968)], empirical evidence has been obtained indicating that this system has a dense or near dense set of unstable periodic orbits throughout its stochastic (unstable) regions of phase space. The extent to which such stochastic regions exhibit C-system behavior, i.e., ergodicity and mixing, is discussed. Finally, the above Hamiltonian system is shown to be intimately related to the Fermi-Pasta-Ulam system as well as to the Toda lattice.

I. INTRODUCTION

When the system energy is sufficiently large, Henon and Heiles[1] have demonstrated by computer experiment that sizeable regions of phase space for the classical, two-oscillator Hamiltonian system

$$H = \tfrac{1}{2}(p_1^2 + p_2^2 + q_1^2 + q_2^2) + q_1^2 q_2 - \tfrac{1}{3}q_2^3 \qquad (1)$$

are filled with unstable trajectories so wild and erratic that the system motion can be regarded as stochastic, at least in the sense that the system trajectories wander freely over part or most of the energy surface. Indeed in these stochastic regions, computer calculations show that two close initial states evolve in time such that the phase space distance between the states increases exponentially. Such computer experiments indicate (but do not prove) that the stochastic regions exhibit C-system[2] behavior, i.e., ergodicity and mixing. Contopoulos[3] on the other hand has carried out computer experiments on a system similar to Hamiltonian (1) and has determined that distinct trajectories lying in a stochastic region do not cover this region as uniformly as one would expect under the assumption of ergodicity. It appears that, unlike C-systems, Hamiltonian (1) may possess stable periodic orbits and small regions of stability lying inside its macroscopic stochastic regions.

It is therefore of interest to investigate the stability character of periodic orbits for Hamiltonian (1). In this paper using techniques previously described by Greene,[4] we develop empirical evidence indicating that the stochastic regions for Hamiltonian (1) are characterized by a dense or near dense set of unstable periodic orbits. Our results, like those of Greene, are empirical and do not constitute a mathematical proof; moreover they do not preclude the existence of stable periodic orbits in the stochastic regions. However, they do make it clear that such periodic orbits, if they exist, are likely to affect only regions of very small measure in phase space. As a consequence in the macroscopic view of a physicist, the physical properties of stochastic regions like those of Hamiltonian (1) are likely to be indistinguishable from those of a C-system.

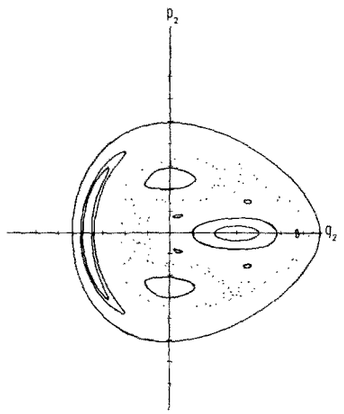


FIG. 1. The mapping generated by Hamiltonian (1) at energy $E = \frac{1}{8}$. Central fixed points lie on the q_2 and p_2 axes at the center of each region of ovals. Each oval in the stable regions was generated by a single, quasi-periodic trajectory of Hamiltonian (1). The set of scattered dots shown in the unstable or stochastic region was also generated by a single trajectory. (Here 1 tick = 0.1.)

II. MATHEMATICAL PRELIMINARIES

In order to define our notation and concepts and in order to make this paper as self-contained as possible, this section briefly reviews the highlights of relevant earlier work. Of necessity, much detail has been omitted; if confusion arises, the reader is urged to peruse the extremely clear Refs. 1 and 4 as well as Chap. 4 of Ref. 2.

Henon and Heiles survey the allowed motion for Hamiltonian (1) by, in effect, reducing the motion to a plane area-preserving mapping. For fixed energy E, it is possible to generate a plane area-preserving mapping in the (q_2, p_2) plane using Hamiltonian (1). Starting with a given (q_2, p_2) point in the plane and setting $q_1 = 0$, we may solve Eq. (1) for $p_1 \geq 0$. This provides us with an initial mapping point (q_2, p_2) and with a set of initial conditions (q_1, p_1, q_2, p_2) for integrating a trajectory of Hamiltonian (1). The initial (q_2, p_2) point then maps into the (q_2, p_2) coordinates of the next point on this system trajectory for which $q_1 = 0$ and $p_1 \geq 0$. Clearly each trajectory generates a unique set of points in the (q_2, p_2) plane, and we may use Poincaré's theorem on integral invariants to prove that the mapping is area-preserving.

Figure 1 shows the mapping generated by Hamiltonian (1) at the energy $E = \frac{1}{8}$. On the q_2 or p_2 axis at the center of each region of ovals is a fixed point of the mapping (hereafter called a central fixed point) generated by a periodic orbit of the differential equation system. If as usual we denote the mapping by the equation

$$(q_2^{(1)}, p_2^{(1)}) = T(q_2^{(0)}, p_2^{(0)}), \qquad (2)$$

then these central fixed points satisfy $(q_2, p_2) = T(q_2, p_2)$. Each oval about a central fixed point is—to computer accuracy—an invariant curve generated by a quasi-periodic orbit of Hamiltonian (1), and sequential mapping iterates of an initial (q_2, p_2) point lying on an oval rotate about the central fixed point. The average angle of rotation (divided by 2π) is called the rotation number ω, and the ω associated with each invariant curve varies smoothly as one progresses

out from a central fixed point. Moser[5] has shown that mathematically exact invariant curves surround the central fixed points provided, among other things, that the associated rotation numbers are irrationals satisfying the inequality

$$|\omega - l/k| > \mathcal{E}/k^{5/2} \tag{3}$$

for all integers l and k, where $\mathcal{E}$ is a constant independent of k.

Between these exact Moser curves, one might expect to find a dense set of invariant curves with rational rotation numbers $\omega = P/Q$ (P and Q relative prime) composed purely of fixed points of Q iterations of the mapping T^Q. However a theorem due to Birkhoff[6] assures us that in general only remnants of such ω-rational invariant curves appear in the form of $2Q$, interleaved fixed points, half being stable elliptic fixed points of T^Q and half being unstable hyperbolic fixed points.[7] Indeed according to Birkhoff, extremely accurate computer calculation of the Fig. 1 mapping in a neighborhood of the central elliptic fixed point on the positive q_2 axis would be expected to reveal a picture similar to that sketched in Fig. 2. An indication of this structure for the Henon–Heiles system at energy $E = \frac{1}{8}$ appears in Fig. 3, which shows only a few of the hundreds of fixed points found for this system. Returning to Fig. 2, we note that the Moser curves are very closely spaced near the central fixed point. Farther out, the interleaved Birkhoff fixed points become large enough to be seen on a macroscopic scale. Moreover at the edges of the figure, the intersecting "hash" of separatrices from the hyperbolic fixed points becomes apparent. Slightly beyond the outer curves in Fig. 2, the Moser curves vanish[4] and the elliptic members of the Birkhoff fixed point families themselves become hyperbolic.[4] Consequently as we shall show later, beyond the family of five elliptic fixed points in Fig. 1 belonging to T^5, many if not all fixed points are hyperbolic, leading to the erratic splatter of points seen in Fig. 1 which were generated by a single trajectory of the Henon–Heiles system.

In order to demonstrate that the stochastic region of Fig. 1 contains a dense or near dense set of hyperbolic fixed points, corresponding to unstable periodic trajectories for the Henon–Heiles system, we apply the calculative techniques of Greene[4] to the Henon–Heiles mapping. Suppose in the (q_2, p_2) plane we have found a fixed point (q, p) of T^Q satisfying

$$(q, p) = T^Q(q, p); \tag{4}$$

let us then follow Greene and write Eq. (4) as a linearized mapping, valid in a small neighborhood of this fixed point, in the vector-matrix form

$$\begin{bmatrix} (q^{(1)} - q) \\ (p^{(1)} - p) \end{bmatrix} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{bmatrix} (q^{(0)} - q) \\ (p^{(0)} - p) \end{bmatrix}, \tag{5}$$

where the $\{M_{ij}\}$ denote the matrix elements of the linearized transformation. We may now use a computer to calculate the matrix elements of M. First we choose a point $(q^{(0)}, p^{(0)})$ near the fixed point (q, p) such that $(p^{(0)} - p) = 0$ and $(q^{(0)} - q) \neq 0$. We then solve the Henon–Heiles differential equation system obtaining $(q^{(1)} - q)$ and $(p^{(1)} - p)$. Thence we have

$$M_{11} = (q^{(1)} - q)/(q^{(0)} - q) \quad \text{and}$$
$$M_{21} = (p^{(1)} - p)/(q^{(0)} - q). \tag{6}$$

Similarly picking an initial point such that $(q^{(0)} - q) = 0$ and $(p^{(0)} - p) \neq 0$ and then calculating the new $(q^{(1)} - q)$ and $(p^{(1)} - p)$, we find

$$M_{12} = (q^{(1)} - q)/(p^{(0)} - p) \quad \text{and}$$
$$M_{22} = (p^{(1)} - p)/(p^{(0)} - p). \tag{7}$$

The conditions that the determinant of M should equal unity and that (due to symmetry of Fig. 1 about the q_2 axis) the diagonal elements of M should be equal can be used to insure accuracy in the calculations. Now introduce the quantity R, called the residue of the fixed point and defined as

FIG. 2. Using increased computer integration accuracy, the mapping in a neighborhood of the central fixed point on the q_2 axis in Fig. 1 would be expected to appear as shown in this sketch. The circles represent the exact invariant curves predicted by the Moser theorem, and the alternating elliptic-hyperbolic fixed point families are of the type predicted by Birkhoff.

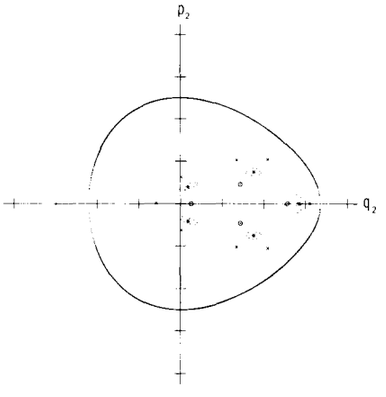


FIG. 3. Selected fixed points of the mapping generated by Hamiltonian (1) at energy $E = \frac{1}{8}$ illustrating that some of the structure sketched in Fig. 2 can actually be observed. The innermost set of four fixed points located at the dots inside small circles are elliptic members of the T^4 family. The fixed points located at the asterisks are elliptic members of the same T^5 family which can be seen in Fig. 1. Surrounding each member of this T^5 family is a set of eleven hyperbolic fixed points belonging to T^{55}. Finally, the eight fixed points located at the x symbols are hyperbolic members of T^8. With the exception of the T^8 family, each family of fixed points shown represents half of an alternating elliptic-hyperbolic set. However, all 16 members of the two T^8 families, which lie in the stochastic region of Fig. 1, are hyperbolic. (Here 1 tick = 0.2)

$$R \equiv \tfrac{1}{2} - \tfrac{1}{4}\,\mathrm{Tr}(M), \tag{8}$$

from which we may determine the stability character of the fixed point. Greene[4] shows that the fixed point is elliptic when $0 < R < 1$ and hyperbolic when $R < 0$ or $R > 1$. As mentioned earlier, the $2Q$ fixed points of T^Q form two families of Q members. Each member of one family has the same positive residue while each member of the other family has the same negative residue. Empirically it is found that the absolute value of the two residues are approximately equal. Consequently when $|R| < 1$, T^Q has alternating elliptic-hyperbolic members; on the other hand when $|R| > 1$, all $2Q$ fixed points of T^Q are hyperbolic, half being hyperbolic with $R < 1$ and half being hyperbolic with $R > 1$.

Let us now briefly return to Fig. 2 which shows a few of the Moser invariant curves and a few of the interleaved elliptic-hyperbolic fixed point families. All members of this set lie on approximate circles (or ellipses) about the central fixed point, and we shall refer to them as members of the first order hierarchy. Since the Moser and Birkhoff theorems are also valid in some neighborhood of each first order, elliptic fixed point, Fig. 2 is reproduced on a microscopic scale about each first order, elliptic fixed point yielding the second order of the hierarchy. One similarly proceeds to define higher orders in the hierarchy, boxes within boxes ad infinitum. The fixed points of first order have rotation numbers $\omega = P/Q$ about the central fixed point with P and Q relative prime. The higher order fixed points yield $\omega = P/Q$ with P and Q having a common integer factor. Here Q is the number of fixed points in the family corresponding to the integer exponent of T^Q, and P is the integral number of 2π rotations around the central fixed point when following from a fixed point to its image through all Q members of the family. For example, each member of the first order, elliptic family of T^5 in Fig. 3 has $\omega = \frac{1}{5}$, while each member of the second order, hyperbolic family of T^{55} has $\omega = \frac{11}{55}$. Greene chooses to fix his attention on the dense or near dense set of first order fixed points in the hierarchy, and for these fixed points he observes that the residue of a fixed point of T^Q may be related to Q by the equation

$$|R| = \alpha[f(P/Q)]^{Q/2}, \tag{9}$$

where (P/Q) is the rotation number of the fixed point, where the absolute value sign on R permits use of negative residue fixed points, and where $f(P/Q)$ is an empirically determined function. If we can now empirically determine a smooth $f(P/Q)$ function, then in those (q_2, p_2) plane regions where $f(P/Q) > 1$ we have that $|R| > 1$, which implies that all or almost all first order fixed points in that region are hyperbolic. In the following section, we present f curves for the Henon–Heiles system.

III. RESULTS

In seeking data from which to plot f curves for the Henon–Heiles system, we chose to search the positive q_2 axis for first order fixed points lying about the central fixed point on the q_2 axis. Due to the (q_2, p_2) plane symmetry about the q_2 axis, at least one fixed point of T^Q for odd Q must lie on axis. This axis would thus be especially rich in fixed points. As a preliminary calculation, we located about 200 first order fixed points lying on the q_2 axis for Henon–Heiles system energy $E = \frac{1}{12}$. These fixed points were quite well distributed over almost the complete interval between the central fixed point on the q_2 axis and the maximum positive q_2 value allowed by $E = \frac{1}{12}$. Inverse rotation numbers began at about $\omega^{-1} = 5.0$ near the central fixed point and monotonically increased to about $\omega^{-1} = 8.5$ near the maximum allowed q_2 value.

For each fixed point we computed R and thence f, using a preliminary value of unity for α. A plot of $f = |R|^{2/Q}$ versus inverse rotation number $\omega^{-1} = Q/P$ provided us with a graph looking like a bad case of the measles. This somewhat random splatter of points however had, just as Greene predicts, a lower envelope representing a minimum f curve. We then found that this minimum f curve could be determined quite easily by plotting f values for sequences of fixed points having inverse rotation numbers given by $\omega^{-1} = k \pm (1/n)$, where $k = 5, 6, 7$, or 8 and $n > 1$ is an

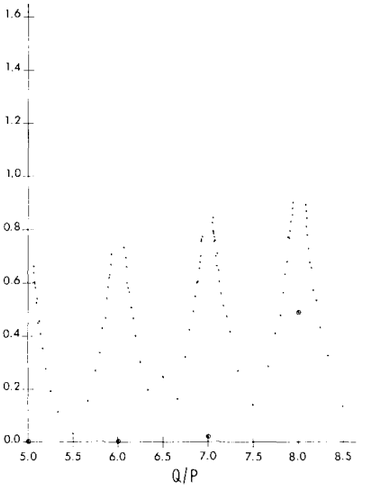


FIG. 4. The minimum f curve for Henon–Heiles system energy $E = \frac{1}{12}$.

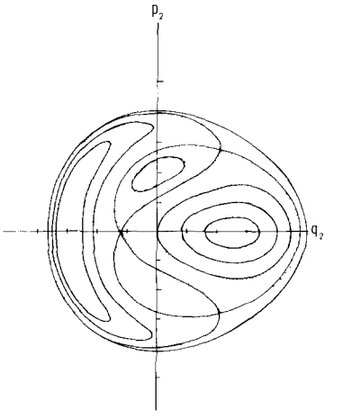


FIG. 5. The mapping generated by Hamiltonian (1) at energy $E = \frac{1}{12}$. As anticipated from Fig. 4, no macroscopic stochastic regions appear. Figure 4 nonetheless indicates that microscopic stochastic regions would be observed with sufficient computer accuracy. (Here 1 tick = 0.1)

integer. The data for a typical sequence is presented in Table I. The minimum f curve itself is presented in Fig. 4. In this figure, α equals unity. We recomputed the graph first using $\alpha = 10$ and then $\alpha = 0.1$. The net result was a slight shift upward or downward of the Fig. 4 curve of no consequence for our purposes; we thus used $\alpha = 1$ in all our calculations.

Although the minimum f curve in Fig. 4 has narrow spikes which rise above unity and although there is a general increase in f values with increasing $\omega^{-1} = Q/P$, the most significant feature of Fig. 4 is that the f curve does not rise and remain above unity. As a consequence, we would not expect to find any macroscopic stochastic regions in a (q_2, p_2) mapping for the Henon–Heiles system at energy $E = \frac{1}{12}$, and this expectation is verified in Fig. 5. The spikes in Fig. 4 which rise above unity (Greene points out that such spikes exist in every Q/P region) imply that there are many microscopic stochastic zones in Fig. 5, each containing a dense or near dense set of first order, hyperbolic fixed points. Despite the fact that these microscopic regions are much too narrow to be seen in Fig. 5, their existence is expected, though not proved, from heuristic arguments[4,8] based on the Moser and Birkhoff theorems; moreover, their appearance can be empirically demonstrated by extremely accurate computer calculations.

Finally in Fig. 6 we present the minimum f curve for the Henon–Heiles system at the energy $E = \frac{1}{8}$ used in Figs. 1 and 3. This minimum curve is again based on data for sequences of fixed points having inverse rotation numbers given by $\omega^{-1} = k \pm (1/n)$, where here $k = 3, 4,$ or 5. In Fig. 1, $\omega^{-1} = Q/P$ varies monotonically from about 3.0 near the central fixed point on the q_2 axis to about 5.5 slightly beyond the ovals around the elliptic fixed point of T^5. The f curve in Fig. 6, as opposed to its cousin in Fig. 4, rises quite rapidly with Q/P. Excepting the fixed points with $Q/P = 5$ which have an f value and residue slightly less than unity, the minimum f curve lies above unity for all $Q/P \geq 4.7$, implying that a dense or near dense set of hyperbolic first order fixed points exists in the stochastic region beyond the T^5 elliptic points in Fig. 1. Indeed one would anticipate that this stochastic

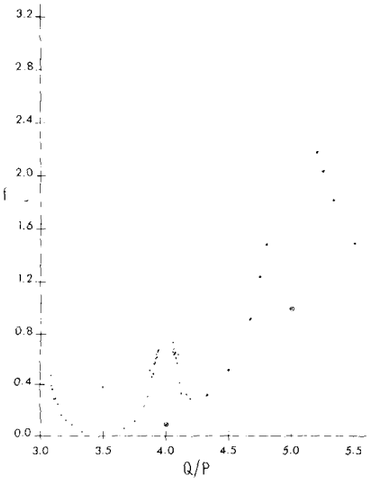


FIG. 6. The minimum f curve for Henon–Heiles system energy $E = \frac{1}{8}$.

TABLE I. The quantity $f = |R|^{2/Q}$ as a function of inverse rotation number $\omega^{-1} = Q/P$ for selected fixed points in the sequence having $\omega^{-1} = 7 + (1/n)$, where n is an integer, and for Henon–Heiles system energy $E = \frac{1}{12}$.

n	Q	P	Q/P	Residue	$f = \lvert R\rvert^{2/Q}$
2	15	2	7.50	4.15×10^{-7}	1.41×10^{-1}
3	22	3	7.33	5.65×10^{-7}	2.70×10^{-1}
4	29	4	7.25	3.22×10^{-6}	4.18×10^{-1}
5	36	5	7.20	1.56×10^{-6}	4.76×10^{-1}
6	43	6	7.17	6.25×10^{-7}	5.15×10^{-1}
7	50	7	7.14	-5.20×10^{-7}	5.61×10^{-1}
8	57	8	7.13	5.98×10^{-7}	6.05×10^{-1}
9	64	9	7.11	8.47×10^{-7}	6.46×10^{-1}
10	71	10	7.10	5.20×10^{-6}	7.10×10^{-1}
15	106	15	7.07	-3.07×10^{-7}	7.54×10^{-1}
20	141	20	7.05	1.30×10^{-6}	8.25×10^{-1}

region completely surrounds the ovals shown in Fig. 1 which encircle the elliptic fixed points of T^5, a feature we shall return to later.

The above argument for a dense or near dense set of hyperbolic fixed points throughout the stochastic region would be considerably strengthened had we been able to continue the f curve of Fig. 6 to larger values of $\omega^{-1} = Q/P$. Unfortunately, searching the q_2 axis for fixed points lying in the highly unstable stochastic region required more computer time and accuracy than was available to us. Nonetheless, the following evidence supporting our central conclusion was obtained. We tracked a number of individual Q/P families as a function of energy from their first appearance near the central fixed point as alternating elliptic-hyperbolic families with $|R| < 1$, through their transition to hyperbolic-hyperbolic families with $|R| > 1$, and through their motion well into the stochastic sea with $|R| \gg 1$. As the energy was increased, the residues for all the individual families studied first oscillated a bit between zero and unity and then rose monotonically to values greatly exceeding unity.

IV. DISCUSSION

Recently, Sinai[9] rigorously proved that the hard sphere gas is ergodic and mixing by, in essence, showing it to be a C-system[2] (strictly speaking, a K-system[2].) One is thus led to inquire[10] whether or not systems with attractive as well as repulsive interactions can exhibit C-system behavior. Since a definitive, mathematically rigorous answer to this question appears to lie far in the future, computer-based, empirical studies of simple systems with attractive interactions such as Hamiltonian (1) can be particularly illuminating. Now an essential property of C-systems[2] is that points on initially close orbits separate exponentially with time. Therefore, all periodic orbits for C-systems are unstable. As a consequence, the elliptic fixed points of Fig. 1, generated by the stable periodic orbits of the Henon–Heiles system, graphically demonstrate that Hamiltonian (1) is not a C-system. Nonetheless Fig. 1 reveals macroscopic unstable regions which are suggestive of emerging C-system behavior. In particular, Henon and Heiles have shown empirically that points on initially close orbits in such regions do exponentiate apart. Moreover, they show that the macroscopic unstable region includes most of phase space for sufficiently large energy.

In this paper we have presented additional evidence indicating C-system behavior in the unstable regions of the Henon–Heiles system by showing that such regions contain a dense or near dense set of unstable periodic orbits. Unfortunately our empirical evidence does not preclude the appearance of microscopic stable regions embedded in the macroscopic unstable region. First, our calculations say nothing about the periodic orbits corresponding to higher order fixed points in the hierarchy; and, second, they say nothing about the maverick periodic orbits which apparently fit nowhere in the hierarchy scheme. Despite these drawbacks, one anticipates that at worst the dense set of first order, unstable periodic trajectories would allow the stable orbits, if present, to influence only microscopic regions of phase space. In support of this conjecture, one observes using a computer that distinct orbits started deep within the unstable region of Fig. 1 each generate an essentially indistinguishable, random-looking set of points.

An exception to this behavior does occur however where, as in Fig. 1, the stochastic sea touches or encloses a macroscopic stable region. For example, the boundary between the stable regions associated with the T^5 elliptic fixed points in Fig. 1 and the surrounding stochastic sea is not sharp. A narrow transition region exists, and orbits in this region may linger there for protracted time intervals before finally escaping into the stochastic sea. The present authors believe that this exceptional behavior is responsible for Contopoulos' results mentioned in Sec. I. In Fig. 1 the boundary layer between the stable regions and the stochastic sea is sizeable and consequently the measure of the trajectories influenced by it is not small. However as the energy in Hamiltonian (1) increases, these macroscopic stable regions and their troublesome boundaries shrink to negligible size. Thus for sufficiently large energy, Hamiltonian (1) would be expected to yield C-system behavior over most of its phase space.

In conclusion, let us emphasize that the Henon–Heiles system does provide insight into the generic behavior of systems with attractive forces. First it is worth noting, as shown in Appendix A, that any three-particle system with a Hamiltonian of the form

$$H = (1/2m)(P_1^2 + P_2^2 + P_3^2) + V(Q_1 - Q_3) + V(Q_2 - Q_1) + V(Q_3 - Q_2) \quad (10)$$

can be reduced to Hamiltonian (1) in the cubic approximation, provided that $V(r)$ has a nonzero cubic term in its power series expansion. Thus the three-particle Fermi–Pasta–Ulam[11] system and the three-particle Toda lattice[12] are intimately related to Hamiltonian (1). In addition numerous computer studies[8,13] of various Hamiltonian oscillator systems reveal behavior differing little from that of the Henon–Heiles system. Finally although Hamiltonian (1) yields predominantly C-system behavior only for large energy, Ford and Lunsford[14] demonstrate that the same type stochastic behavior can occur for arbitrarily small energy when the number of oscillators is three or greater.

ACKNOWLEDGMENT

The authors particularly wish to thank Dr. John M. Greene of the Plasma Physics Laboratory, Princeton University, whose clear and illuminating correspondence lighted a path through our fog of misconceptions. Errors which remain are the sole responsibility of the authors.

APPENDIX A

Let us expand the potentials V in Hamiltonian (10) as a power series in their arguments, retaining terms only through cubic order. We may then write Hamiltonian (10) as

$$H = \tfrac{1}{2}[P_1^2 + P_2^2 + P_3^2 + (Q_1 - Q_3)^2 + (Q_2 - Q_1)^2 + (Q_3 - Q_2)^2] - (\alpha/3)[(Q_1 - Q_3)^3 + (Q_2 - Q_1)^3 + (Q_3 - Q_2)^3], \quad (A1)$$

where without loss of generality we have set the mass m and the "spring constant" k of the quadratic terms equal to unity. Into Hamiltonian (A1) let us introduce the canonical transformation

$$Q_i = \sum_{j=1}^{3} A_{ij}\zeta_j, \qquad P_i = \sum_{j=1}^{3} A_{ij}\eta_j, \quad (A2)$$

where the matrix A is given by

$$A = \begin{pmatrix} 3^{-1/2} & 2^{-1/2} & 6^{-1/2} \\ 3^{-1/2} & 0 & -(2/3)^{1/2} \\ 3^{-1/2} & -2^{-1/2} & 6^{-1/2} \end{pmatrix}. \quad (A3)$$

we find

$$H = \tfrac{1}{2}[\eta_1^2 + \eta_2^2 + \eta_3^2 + 3\zeta_2^2 + 3\zeta_3^2] + (3\alpha/2^{1/2})(\zeta_2\zeta_3^2 - \tfrac{1}{3}\zeta_2^3). \quad (A4)$$

Since ζ_1 does not appear in Hamiltonian (A4), η_1 is a constant of the motion. We may thus drop the term η_1^2 from Hamiltonian (A4), obtaining the reduced, two degrees of freedom Hamiltonian

$$H = \tfrac{1}{2}(\eta_2^2 + \eta_3^2 + 3\zeta_2^2 + 3\zeta_3^2) + (3\alpha/2^{1/2})(\zeta_2\zeta_3^2 - \tfrac{1}{3}\zeta_2^3). \quad (A5)$$

The equations of motion for Hamiltonian (A5) are

$$\begin{aligned} \ddot{\zeta}_2 &= -3\zeta_2 - (3\alpha/2^{1/2})(\zeta_3^2 - \zeta_2^2), \\ \ddot{\zeta}_3 &= -3\zeta_3 - (3\alpha/2^{1/2})(2\zeta_2\zeta_3). \end{aligned} \quad (A6)$$

Now change the time scale via $t = \tau/3^{1/2}$, retaining the dot notation for $(d/d\tau)$, and let $\zeta_2 = (2^{1/2}/\alpha)q_2$ and $\zeta_3 = (2^{1/2}/\alpha)q_1$; then Eq. (A6) may be written

$$\ddot{q}_1 = -q_1 - 2q_1q_2, \qquad \ddot{q}_2 = -q_2 - q_1^2 + q_2^2. \quad (A7)$$

The Hamiltonian for Eq. (A7) is

$$H = \tfrac{1}{2}(p_1^2 + p_2^2 + q_1^2 + q_2^2) + q_1^2q_2 - \tfrac{1}{3}q_2^3, \quad (A8)$$

i.e., the Henon–Heiles Hamiltonian (1). Thus in the cubic approximation, Hamiltonian (10) is equivalent to Hamiltonian (1).

---

* This work supported in part by the National Science Foundation.

[1] M. Henon and C. Heiles, Astron. J. **69**, 73 (1964).

[2] V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968), Chaps. 3 and 4.

[3] G. Contopoulos, Astron. J. **76**, 147 (1971).
[4] J. M. Greene, J. Math. Phys. **9**, 760 (1968).
[5] J. Moser, Nachr. Adad. Wiss. Goettingen, II. Math. Physik. Kl. No. 1 (1962).
[6] G. D. Birkhoff, *Collected Mathematical Papers* (American Mathematical Society, Providence, R.I., 1950), Vol. I, p. 673, Vol. II, p. 252.
[7] A fixed point is said to be elliptic if invariant curves in its immediate neighborhood are ellipses; it is called hyperbolic if nearby invariant curves are hyperbolic.
[8] G. H. Walker and J. Ford, Phys. Rev. **188**, 416 (1969).
[9] Ja. G. Sinai in *Statistical Mechanics, Proceedings of the IUPAP Meeting, Copenhagen, 1966*, edited by T. A. Bak (Benjamin, New York, 1967), p. 559; Ja. G. Sinai, Russian Math. Surveys **25**, No. 2, 137 (1970).
[10] A. S. Wightman in *Statistical Mechanics at the Turn of the Decade*, edited by E. G. D. Cohen (Dekker, New York, 1971), p. 1.
[11] See *Enrico Fermi: Collected Papers, Vol. II* (U. of Chicago Press, Chicago, 1965), p. 978.
[12] M. Toda, J. Phys. Soc. Japan **23**, 501 (1967).
[13] N. Saito, N. Ooyama, Y. Aizawa, and H. Hirooka, Progr. Theoret. Phys. (Kyoto) Suppl. **45**, 209 (1970).
[14] J. Ford and G. H. Lunsford, Phys. Rev. A 1, 59 (1970).

Solution of Faddeev Equations for a One-Dimensional System

Chanchal K. Majumdar
Tata Institute of Fundamental Research, Bombay, India

(Received 27 January 1971)

Exact solutions are derived for the Faddeev equations for a one-dimensional system of three distinguishable particles of identical mass, interacting pairwise through a delta function potential. The method here employed uses scattering boundary conditions directly on the system of equations.

Dodd[1] has shown that solutions can be found to the Faddeev equations[2] for a one-dimensional system of three particles of identical mass interacting pairwise with a delta function potential. This system was discussed by McGuire[3] and Yang.[4] Dodd constructs the various scattering amplitudes by utilizing the known form of McGuire's configuration space solutions. He then demonstrates that these satisfy the Faddeev equations for the system. His method can therefore be characterized as indirect. Obviously a more direct method of solution, without using the knowledge of the configuration space solution, is of some interest, particularly, with a view to handling problems where the wavefunctions are not known *a priori*.

We shall present such a direct solution of the problem. The main point is to incorporate the boundary conditions in Faddeev equations. In some recent work on atomic problems, Chen, Chung, and Kramer[5] have shown how to do this; an explanation of their procedure is given by Shastry and Rajagopal.[6] We adopt this method to solve the equations.

1. FADDEEV EQUATIONS

The mass of each particle is taken to be unity. Let x_1, x_2, x_3 be their coordinates. The Hamiltonian of our problem is

$$H = \tfrac{1}{2}(k_1^2 + k_2^2 + k_3^2) - g\delta(x_2 - x_3) - g\delta(x_3 - x_1) - g\delta(x_1 - x_2). \quad (1.1)$$

We define the momenta

$$\begin{aligned} K &= (6)^{-1/2}(k_1 + k_2 + k_3), \\ q_1 &= \tfrac{1}{2}(3)^{-1/2}(2k_1 - k_2 - k_3), \quad & p_1 &= \tfrac{1}{2}(k_2 - k_3), \\ q_2 &= \tfrac{1}{2}(3)^{-1/2}(2k_2 - k_3 - k_1), \quad & p_2 &= \tfrac{1}{2}(k_3 - k_1), \\ q_3 &= \tfrac{1}{2}(3)^{-1/2}(2k_3 - k_1 - k_2), \quad & p_3 &= \tfrac{1}{2}(k_1 - k_2). \end{aligned} \quad (1.2)$$

We can put $K = 0$. Then the kinetic part of (1.1) becomes

$$H_0 = p_1^2 + q_1^2 = p_2^2 + q_2^2 = p_3^2 + q_3^2. \quad (1.3)$$

The sets (p_1, q_1), (p_2, q_2), and (p_3, q_3) are related; for instance,

$$\begin{aligned} q_2 &= -\tfrac{1}{2}q_1 + \tfrac{1}{2}\sqrt{3}p_1, \quad & p_2 &= \tfrac{1}{2}\sqrt{3}q_1 - \tfrac{1}{2}p_1, \\ q_3 &= -\tfrac{1}{2}q_1 - \tfrac{1}{2}\sqrt{3}p_1, \quad & p_3 &= \tfrac{1}{2}\sqrt{3}q_1 - \tfrac{1}{2}p_1. \end{aligned} \quad (1.4)$$

We denote the state vector by

$$|k_1 k_2 k_3\rangle \equiv |p_1 q_1\rangle_1 = |p_2 q_2\rangle_2 = |p_3 q_3\rangle_3. \quad (1.5)$$

The subscript denotes which particular set of (1.2) we are using. The states (1.5) are complete and normalized by

$${}_i\langle p'q'|pq\rangle_i = \delta(p' - p)\delta(q' - q). \quad (1.6)$$

The Faddeev equations are written as

$$T^1 = T_1 + T_1 G_0(T^2 + T^3), \quad T^2 = T_2 + T_2 G_0(T^3 + T^1),$$
$$T^3 = T_3 + T_3 G_0(T^1 + T^2), \quad G_0 = (E - H_0)^{-1}. \quad (1.7)$$

The operators T_i are obtained from the off-energy-shell extension of the t-matrices of the two particle subsystems:

$${}_i\langle p_i q_i | T_i(E) | p_i' q_i'\rangle_i = -\frac{g}{2\pi} \cdot \frac{\delta(q_i - q_i')(E - q_i^2)^{1/2}}{(E - q_i^2)^{1/2} - \tfrac{1}{2}ig}, \quad i = 1, 2, 3. \quad (1.8)$$

For any state $|y\rangle$, we take matrix elements of (1.7) and write

$$\begin{aligned} {}_1\langle p_1 q_1 | T^1 | y\rangle &= {}_1\langle p_1 q_1 | T_1 | y\rangle - (g/2\pi)\Psi_1(q_1; y), \\ {}_2\langle p_2 q_2 | T^2 | y\rangle &= {}_2\langle p_2 q_2 | T_2 | y\rangle - (g/2\pi)\Psi_2(q_2; y), \\ {}_3\langle p_3 q_3 | T^3 | y\rangle &= {}_3\langle p_3 q_3 | T_3 | y\rangle - (g/2\pi)\Psi_3(q_3; y). \end{aligned} \quad (1.9)$$

Then from (1.7), (1.8), and (1.4), we get

$$\Psi_1(q_1; y) = \chi_1(q_1; y) + \frac{\sqrt{3}g}{4\pi} \frac{(E - q_1^2)^{1/2}}{(E - q_1^2)^{1/2} - \tfrac{1}{2}ig} \int_{-\infty}^{\infty} \frac{\Psi_2(q''; y)\, dq''}{q''^2 + q''q_1 + q_1^2 - \tfrac{3}{4}E} + \frac{\sqrt{3}g}{4\pi} \frac{(E - q_1^2)^{1/2}}{(E - q_1^2)^{1/2} - \tfrac{1}{2}ig} \int_{-\infty}^{\infty} \frac{\Psi_3(q''; y)\, dq''}{q''^2 + q''q_1 + q_1^2 - \tfrac{3}{4}E}. \quad (1.10)$$

Here χ_1 is given by

[3] G. Contopoulos, Astron. J. **76**, 147 (1971).
[4] J. M. Greene, J. Math. Phys. **9**, 760 (1968).
[5] J. Moser, Nachr. Adad. Wiss. Goettingen, II. Math. Physik. Kl. No. 1 (1962).
[6] G. D. Birkhoff, *Collected Mathematical Papers* (American Mathematical Society, Providence, R.I., 1950), Vol. I, p. 673, Vol. II, p. 252.
[7] A fixed point is said to be elliptic if invariant curves in its immediate neighborhood are ellipses; it is called hyperbolic if nearby invariant curves are hyperbolic.
[8] G. H. Walker and J. Ford, Phys. Rev. **188**, 416 (1969).
[9] Ja. G. Sinai in *Statistical Mechanics, Proceedings of the IUPAP Meeting, Copenhagen, 1966*, edited by T. A. Bak (Benjamin, New York, 1967), p. 559; Ja. G. Sinai, Russian Math. Surveys **25**, No. 2, 137 (1970).
[10] A. S. Wightman in *Statistical Mechanics at the Turn of the Decade*, edited by E. G. D. Cohen (Dekker, New York, 1971), p. 1.
[11] See *Enrico Fermi: Collected Papers, Vol. II* (U. of Chicago Press, Chicago, 1965), p. 978.
[12] M. Toda, J. Phys. Soc. Japan **23**, 501 (1967).
[13] N. Saito, N. Ooyama, Y. Aizawa, and H. Hirooka, Progr. Theoret. Phys. (Kyoto) Suppl. **45**, 209 (1970).
[14] J. Ford and G. H. Lunsford, Phys. Rev. A 1, 59 (1970).

Solution of Faddeev Equations for a One-Dimensional System

Chanchal K. Majumdar
Tata Institute of Fundamental Research, Bombay, India

(Received 27 January 1971)

Exact solutions are derived for the Faddeev equations for a one-dimensional system of three distinguishable particles of identical mass, interacting pairwise through a delta function potential. The method here employed uses scattering boundary conditions directly on the system of equations.

Dodd[1] has shown that solutions can be found to the Faddeev equations[2] for a one-dimensional system of three particles of identical mass interacting pairwise with a delta function potential. This system was discussed by McGuire[3] and Yang.[4] Dodd constructs the various scattering amplitudes by utilizing the known form of McGuire's configuration space solutions. He then demonstrates that these satisfy the Faddeev equations for the system. His method can therefore be characterized as indirect. Obviously a more direct method of solution, without using the knowledge of the configuration space solution, is of some interest, particularly, with a view to handling problems where the wavefunctions are not known *a priori*.

We shall present such a direct solution of the problem. The main point is to incorporate the boundary conditions in Faddeev equations. In some recent work on atomic problems, Chen, Chung, and Kramer[5] have shown how to do this; an explanation of their procedure is given by Shastry and Rajagopal.[6] We adopt this method to solve the equations.

1. FADDEEV EQUATIONS

The mass of each particle is taken to be unity. Let x_1, x_2, x_3 be their coordinates. The Hamiltonian of our problem is

$$H = \tfrac{1}{2}(k_1^2 + k_2^2 + k_3^2) - g\delta(x_2 - x_3) - g\delta(x_3 - x_1) - g\delta(x_1 - x_2). \quad (1.1)$$

We define the momenta

$$\begin{aligned} K &= (6)^{-1/2}(k_1 + k_2 + k_3), \\ q_1 &= \tfrac{1}{2}(3)^{-1/2}(2k_1 - k_2 - k_3), \quad & p_1 &= \tfrac{1}{2}(k_2 - k_3), \\ q_2 &= \tfrac{1}{2}(3)^{-1/2}(2k_2 - k_3 - k_1), \quad & p_2 &= \tfrac{1}{2}(k_3 - k_1), \\ q_3 &= \tfrac{1}{2}(3)^{-1/2}(2k_3 - k_1 - k_2), \quad & p_3 &= \tfrac{1}{2}(k_1 - k_2). \end{aligned} \quad (1.2)$$

We can put $K = 0$. Then the kinetic part of (1.1) becomes

$$H_0 = p_1^2 + q_1^2 = p_2^2 + q_2^2 = p_3^2 + q_3^2. \quad (1.3)$$

The sets (p_1, q_1), (p_2, q_2), and (p_3, q_3) are related; for instance,

$$\begin{aligned} q_2 &= -\tfrac{1}{2}q_1 + \tfrac{1}{2}\sqrt{3}p_1, \quad & p_2 &= \tfrac{1}{2}\sqrt{3}q_1 - \tfrac{1}{2}p_1, \\ q_3 &= -\tfrac{1}{2}q_1 - \tfrac{1}{2}\sqrt{3}p_1, \quad & p_3 &= \tfrac{1}{2}\sqrt{3}q_1 - \tfrac{1}{2}p_1. \end{aligned} \quad (1.4)$$

We denote the state vector by

$$|k_1 k_2 k_3\rangle \equiv |p_1 q_1\rangle_1 = |p_2 q_2\rangle_2 = |p_3 q_3\rangle_3. \quad (1.5)$$

The subscript denotes which particular set of (1.2) we are using. The states (1.5) are complete and normalized by

$${}_i\langle p'q'|pq\rangle_i = \delta(p' - p)\delta(q' - q). \quad (1.6)$$

The Faddeev equations are written as

$$T^1 = T_1 + T_1 G_0(T^2 + T^3), \quad T^2 = T_2 + T_2 G_0(T^3 + T^1),$$
$$T^3 = T_3 + T_3 G_0(T^1 + T^2), \quad G_0 = (E - H_0)^{-1}. \quad (1.7)$$

The operators T_i are obtained from the off-energy-shell extension of the t-matrices of the two particle subsystems:

$${}_i\langle p_i q_i | T_i(E) | p_i' q_i'\rangle_i = -\frac{g}{2\pi} \cdot \frac{\delta(q_i - q_i')(E - q_i^2)^{1/2}}{(E - q_i^2)^{1/2} - \tfrac{1}{2}ig}, \quad i = 1, 2, 3. \quad (1.8)$$

For any state $|y\rangle$, we take matrix elements of (1.7) and write

$$\begin{aligned} {}_1\langle p_1 q_1 | T^1 | y\rangle &= {}_1\langle p_1 q_1 | T_1 | y\rangle - (g/2\pi)\Psi_1(q_1; y), \\ {}_2\langle p_2 q_2 | T^2 | y\rangle &= {}_2\langle p_2 q_2 | T_2 | y\rangle - (g/2\pi)\Psi_2(q_2; y), \\ {}_3\langle p_3 q_3 | T^3 | y\rangle &= {}_3\langle p_3 q_3 | T_3 | y\rangle - (g/2\pi)\Psi_3(q_3; y). \end{aligned} \quad (1.9)$$

Then from (1.7), (1.8), and (1.4), we get

$$\Psi_1(q_1; y) = \chi_1(q_1; y) + \frac{\sqrt{3}g}{4\pi} \frac{(E - q_1^2)^{1/2}}{(E - q_1^2)^{1/2} - \tfrac{1}{2}ig} \int_{-\infty}^{\infty} \frac{\Psi_2(q''; y)\, dq''}{q''^2 + q''q_1 + q_1^2 - \tfrac{3}{4}E} + \frac{\sqrt{3}g}{4\pi} \frac{(E - q_1^2)^{1/2}}{(E - q_1^2)^{1/2} - \tfrac{1}{2}ig} \int_{-\infty}^{\infty} \frac{\Psi_3(q''; y)\, dq''}{q''^2 + q''q_1 + q_1^2 - \tfrac{3}{4}E}. \quad (1.10)$$

Here χ_1 is given by

$$-\frac{g}{2\pi}\chi_1(q_1;y) = \int_{-\infty}^{\infty} \frac{{}_1\langle p_1q_1|T_1|p''q''\rangle_2 dp''dq''\,{}_2\langle p''q''|T_2|y\rangle}{E-p''^2-q''^2} + \int_{-\infty}^{\infty} \frac{{}_1\langle p_1q_1|T_1|p''q''\rangle_3 dp''dq''\,{}_3\langle p''q''|T_3|y\rangle}{E-p''^2-q''^2} \tag{1.11}$$

The equations for $\Psi_2(q_2;y)$ and $\Psi_3(q_3;y)$ can be easily written down by looking at (1.10) and (1.11).

2. SOLUTION

The bound state solution is obtained by ignoring the inhomogeneous term χ in (1.10). The equations for Ψ_1, Ψ_2, and Ψ_3 become identical; hence Ψ is independent of the labels 1, 2, or 3. One has then to solve a single integral equation. It is easy to show that $\Psi(q) = (q^2 + \frac{3}{4}g^2)^{-1}$ is a solution and the three-particle bound state has the binding energy $E_B = -g^2$.

Consider now the scattering solutions. Starting with the pair (23) bound and particle one incident on it, we have three possible processes below the break-up threshold:

$$\begin{aligned} 1 + (23) &\to 1 + (23), \\ &\to 2 + (31), \\ &\to 3 + (12). \end{aligned} \tag{2.1}$$

The last two represent rearrangement collisions.

According to Chen, Chung, and Kramer, we have to put the inhomogeneous term in (1.10) on to the energy shell and factor out properly normalized bound state function, if present. It is seen from (1.8) that the two particle bound state energy is $-\frac{1}{4}g^2$, so that the initial energy is, on the energy shell,

$$E = q_0^2 - \tfrac{1}{4}g^2, \tag{2.2}$$

where q_0 is measured in set $(q_1p_1)_1$. As the configuration (23) is bound, T_1 must have a pole at the appropriate energy, but not T_2 and T_3. Also note that the two particle bound state function $G_0|p\rangle = (p^2 + \frac{1}{4}g^2)^{-1}|p\rangle$ has the normalization

$$N^2 = \left(\int_{-\infty}^{\infty} \frac{dp}{(p^2+\frac{1}{4}g^2)^2}\right)^{-1} = \frac{g^3}{4\pi}. \tag{2.3}$$

For the initial state $|y\rangle$, we let q_0^2 approach $E + \frac{1}{4}g^2$; then divide out the normalization (2.3), that is

$$\lim_{q_0^2 \to E+\frac{1}{4}g^2} \frac{E-q_0^2+\frac{1}{4}g^2}{(g^3/4\pi)} \cdot {}_1\langle p_1q_1|T_1|q_0, -\tfrac{1}{4}g^2\rangle_1 = \delta(q_1-q_0). \tag{2.4}$$

Also,

$$\lim_{q_0^2 \to E+\frac{1}{4}g^2} \frac{E-q_0^2+\frac{1}{4}g^2}{(g^3/4\pi)} \cdot {}_2\langle p_2q_2|T_2|q_0, -\tfrac{1}{4}g^2\rangle_1 = 0,$$

$$\lim_{q_0^2 \to E+\frac{1}{4}g^2} \frac{E-q_0^2+\frac{1}{4}g^2}{(g^3/4\pi)} \cdot {}_3\langle p_3q_3|T_3|q_0, -\tfrac{1}{4}g^2\rangle_1 = 0. \tag{2.5}$$

Thus

$$\chi_1(q_1;q_0) = 0,$$

$$\chi_2(q_2;q_0) = -\frac{\sqrt{3}}{2} \cdot \frac{(E-q_2^2)^{1/2}}{(E-q_2^2)^{1/2} - \frac{1}{2}ig} \times \frac{1}{(q_2-x)(q_2-\bar{x})},$$

$$\chi_3(q_3;q_0) = -\frac{\sqrt{3}}{2} \cdot \frac{(E-q_3^2)^{1/2}}{(E-q_3^2)^{1/2} - \frac{1}{2}ig} \times \frac{1}{(q_3-x)(q_3-\bar{x})}, \tag{2.6}$$

where

$$x = -\tfrac{1}{2}q_0 + i\tfrac{1}{4}\sqrt{3}g, \quad \bar{x} = -\tfrac{1}{2}q_0 - i\tfrac{1}{4}\sqrt{3}g. \tag{2.7}$$

It is also convenient to use only half-shell extensions, so that henceforth we put

$$E = q_0^2 - \tfrac{1}{4}g^2. \tag{2.8}$$

Define

$$t = (\tfrac{1}{4}g^2 - q_0^2 + q^2)^{1/2}. \tag{2.9}$$

Then we have

$$\Psi_1(q) = \frac{\sqrt{3}g}{4\pi} \frac{it}{it-\frac{1}{2}ig} \times \left(\int_{-\infty}^{\infty} \frac{\Psi_2(q'')dq''}{q''^2+q''q+q^2-\frac{3}{4}(q_0^2-\frac{1}{4}g^2)} + \int_{-\infty}^{\infty} \frac{\Psi_3(q'')dq''}{q''^2+q''q+q^2-\frac{3}{4}(q_0^2-\frac{1}{4}g^2)}\right), \tag{2.10}$$

$$\Psi_2(q) = -\frac{\sqrt{3}}{2} \frac{it}{it-\frac{1}{2}ig} \cdot \frac{1}{(q-x)(q-\bar{x})} + \frac{\sqrt{3}g}{4\pi} \cdot \frac{it}{it-\frac{1}{2}ig} \times \left(\int_{-\infty}^{\infty} \frac{\Psi_3(q'')dq''}{q''^2+q''q+q^2-\frac{3}{4}(q_0^2-\frac{1}{4}g^2)} + \int_{-\infty}^{\infty} \frac{\Psi_1(q'')dq''}{q''^2+q''q+q^2-\frac{3}{4}(q_0^2-\frac{1}{4}g^2)}\right). \tag{2.11}$$

The equation for Ψ_3 is obtained from (2.11) by replacing two by three. As a result of the identity of mass, we have

$$\Psi_2(q) = \Psi_3(q), \tag{2.12}$$

and can restrict ourselves to only two unknown functions. Write

$$\Psi_1(q) = \frac{it}{it-\frac{1}{2}ig}\varphi_1(q), \quad \Psi_2(q) = \frac{it}{it-\frac{1}{2}ig}\varphi_2(q). \tag{2.13}$$

The equations to be solved are [compare Ref. 1, Eq. (15)]

$$\varphi_1(q) = \frac{\sqrt{3}g}{4\pi} \int_{-\infty}^{\infty} \frac{dq''}{(q''-y)(q''-\bar{y})} \times \frac{(q_0^2-\frac{1}{4}g^2-q''^2)^{1/2}}{(q_0^2-\frac{1}{4}g^2-q''^2)^{1/2}-\frac{1}{2}ig} \cdot \varphi_2(q''), \tag{2.14}$$

$$\varphi_2(q) = -\frac{\sqrt{3}}{2}\frac{1}{(q-x)(q-\bar{x})} + \frac{\sqrt{3}g}{4\pi}\int_{-\infty}^{\infty}\frac{dq''}{(q''-y)(q''-\bar{y})} \times \frac{(q_0^2 - \frac{1}{4}g^2 - q''^2)^{1/2}}{(q_0^2 - \frac{1}{4}g^2 - q''^2)^{1/2} - \frac{1}{2}ig}\cdot[\varphi_1(q'') + \varphi_2(q'')], \tag{2.15}$$

where

$$y = -\tfrac{1}{2}q + i\tfrac{1}{2}\sqrt{3}t, \qquad \bar{y} = -\tfrac{1}{2}q - i\tfrac{1}{2}\sqrt{3}t. \tag{2.16}$$

As solution, we try the ansatz

$$\varphi_1(q) = \frac{it - \frac{1}{2}ig}{it}\left(\frac{A_1}{q - q_0 - i\epsilon} + \frac{B_1}{q-x} + \frac{C_1}{q-\bar{x}}\right), \tag{2.17}$$

$$\varphi_2(q) = \frac{it - \frac{1}{2}ig}{it}\left(\frac{A_2}{q - q_0 - i\epsilon} + \frac{B_2}{q-x} + \frac{C_2}{q-\bar{x}}\right). \tag{2.18}$$

A_1, B_1, C_1 and A_2, B_2, C_2 are independent of q, but depend on q_0 and g. The last two terms are suggested by the inhomogeneous term, while the existence of the pole at $q = q_0$ is required by the fact that on the energy shell we have finite scattering [see Eqs. (2.21) and (2.22) below)].

We have to determine six unknowns—A_1, B_1, C_1 and A_2, B_2, C_2. Substituting (2.17) and (2.18) into (2.14) and (2.15), and carrying out the elementary complex integrations, we get six linear equations for them, and two additional equations which can be shown to be identically satisfied. The solutions are

$$A_1 = -\frac{2}{\sqrt{3}(q_0 - \frac{1}{2}\sqrt{3}\,ig)},$$

$$B_1 = 0, \qquad C_1 = \frac{2}{\sqrt{3}(q_0 - \frac{1}{2}\sqrt{3}\,ig)} \tag{2.19}$$

and

$$A_2 = -\frac{\sqrt{3}\,q_0 + \frac{1}{2}ig}{\sqrt{3}(q_0 - \frac{1}{2}\sqrt{3}\,ig)(\sqrt{3}q_0 - \frac{1}{2}ig)},$$

$$B_2 = -\frac{\sqrt{3}q_0 + \frac{1}{2}ig}{ig(\sqrt{3}q_0 - \frac{1}{2}ig)}, \quad C_2 = \frac{\sqrt{3}q_0 + \frac{1}{2}ig}{\sqrt{3}ig(q_0 - \frac{1}{2}\sqrt{3}\,ig)}. \tag{2.20}$$

The pole at $q_0 = \frac{1}{2}\sqrt{3}\,ig$ corresponds to the three particle bound state of energy $E_B = q_0^2 - \frac{1}{4}g^2 = -g^2$. Now to get the scattering amplitudes, we must put (1.9) back on to the energy shell[5]; the physical collision amplitudes are then determined to within a constant factor c. We have

$$\lim_{q^2 \to E + \frac{1}{4}g^2}(E - q^2 + \tfrac{1}{4}g^2)\left(-\frac{g}{2\pi}\Psi_1(q)\right) = \lim_{q^2 \to q_0^2}(q_0^2 - q^2)\left(-\frac{g}{2\pi}\Psi_1(q)\right) = -\frac{g}{2\pi}\cdot\frac{4q_0}{\sqrt{3}(q_0 - \frac{1}{2}\sqrt{3}ig)} \tag{2.21}$$

$$\lim_{q^2 \to E + \frac{1}{4}g^2}(E - q^2 + \tfrac{1}{4}g^2)\left(-\frac{g}{2\pi}\cdot\Psi_2(q)\right) = \lim_{q^2 \to q_0^2}(q_0^2 - q^2)\left(-\frac{g}{2\pi}\Psi_2(q)\right) = -\frac{g}{2\pi}\,\frac{2q_0(\sqrt{3}q_0 + \frac{1}{2}ig)}{\sqrt{3}(q_0 - \frac{1}{2}\sqrt{3}\,ig)(\sqrt{3}q_0 - \frac{1}{2}ig)}. \tag{2.22}$$

As the density of states is $(dE/dq)^{-1}_{q=q_0} = \frac{1}{2}q_0^{-1}$, the S-matrix for the bound state scattering, including the undermined constant c, is

$$S = \begin{pmatrix} 1 + \dfrac{2icg}{\sqrt{3}(q_0 - \frac{1}{2}\sqrt{3}ig)} & \dfrac{icg(\sqrt{3}q_0 + \frac{1}{2}ig)}{\sqrt{3}(q_0 - \frac{1}{2}\sqrt{3}ig)(\sqrt{3}q_0 - \frac{1}{2}ig)} & \dfrac{icg(\sqrt{3}q_0 + \frac{1}{2}ig)}{\sqrt{3}(q_0 - \frac{1}{2}ig)(\sqrt{3}q_0 - \frac{1}{2}ig)} \\ \dfrac{icg(\sqrt{3}q_0 + \frac{1}{2}ig)}{\sqrt{3}(q_0 - \frac{1}{2}\sqrt{3}ig)(\sqrt{3}q_0 - \frac{1}{2}ig)} & 1 + \dfrac{2icg}{\sqrt{3}(q_0 - \frac{1}{2}\sqrt{3}ig)} & \dfrac{icg(\sqrt{3}q_0 + \frac{1}{2}ig)}{\sqrt{3}(q_0 - \frac{1}{2}\sqrt{3}ig)(\sqrt{3}q_0 - \frac{1}{2}ig)} \\ \dfrac{icg(\sqrt{3}q_0 + \frac{1}{2}ig)}{\sqrt{3}(q_0 - \frac{1}{2}\sqrt{3}ig)(\sqrt{3}q_0 - \frac{1}{2}ig)} & \dfrac{icg(\sqrt{3}q_0 + \frac{1}{2}ig)}{\sqrt{3}(q_0 - \sqrt{3}ig)(\sqrt{3}q_0 - \frac{1}{2}ig)} & 1 + \dfrac{2icg}{\sqrt{3}(q_0 - \frac{1}{2}\sqrt{3}ig)} \end{pmatrix}. \tag{2.23}$$

Demanding that S be unitary, we find $c = 1$ ($c = 0$ is obviously an unphysical root). This completes the solution.

3. CONCLUSION

Dodd[1] has discussed general features of the scattering amplitudes. He remarks that this problem is soluble because of the cancellation of the two-body branch cut by a compensating factor in the exact amplitudes. In some recent work, it was shown[7] that the energies of the three spin wave bound states of the linear Heisenberg chain of spin-$\frac{1}{2}$ particles can be found by solving the corresponding Faddeev equations. An examination shows great mathematical similarity—especially in the feature of the cancellation of the two-body branch cuts in the two problems. However, the spin problem has more complicated mathematical structure, and explicit analytical solutions have not been obtained yet.

[1] L. R. Dodd, J. Math. Phys. 11, 207 (1970). Please note the printing mistake in Eq. (4); 6 should be 3.

[2] L. D. Faddeev, Zh. Eksp. Teor. Fiz. 39, 1459 (1960) [Sov. Phys. JEPT 12, 1014 (1961)].

[3] J. B. McGuire, J. Math. Phys. 5, 622 (1964).
[4] C. N. Yang. Phys. Rev. 168, 1920 (1968).
[5] J. C. Y. Chen, K. T. Chung, and P. J. Kramer, Phys. Rev. 184, 64 (1969).
[6] C. S. Shastry and A. K. Rajagopal, Phys. Rev. A 2, 781 (1970).
[7] C. K. Majumdar, Phys. Rev. B 1, 287 (1970); C. K. Majumdar and G. Mukhopadhyay, Phys. Letters 31A, 321 (1970).

An Approximate Solution for the Static, Spherically Symmetric Metric Due to a Point Charged Mass in Brans-Dicke Theory

M. N. Mahanta and D. R. K. Reddy
Mathematics Department, Indian Institute of Technology, New Delhi-29, India
(Received 11 January 1971)

An approximate solution of the field equations of Brans–Dicke theory is obtained for a static, spherically symmetric metric due to a charged point mass which can be considered to be an analog of the Reissner–Nordstrom metric in Einstein's theory.

1. INTRODUCTION

In a previous paper,[1] we have obtained an approximate solution of the vacuum static case of spherical symmetry in Brans–Dicke theory, starting from the variational principle[2]

$$\delta \int (\psi R - \omega \psi_{,i}\psi^{,i}/\psi)\sqrt{-g}\, d^4x = 0, \tag{1}$$

where R is the scalar curvature, ω the dimensionless constant, and ψ the scalar playing the role of G^{-1}, following a technique used first by Weyl[3] and then by Pauli.[4] A similar solution for the static, spherically symmetric metric due to a point charged mass is well worth consideration.

In this paper we obtain an approximate solution of the field equations of Brans–Dicke theory for a static, spherically symmetric metric due to a point charged mass following the same technique. The solution is then compared with the Reissner–Nordstrom solution in Einstein's theory.

2. FIELD EQUATIONS

We consider the line element for the static, spherically symmetric metric case in the form

$$ds^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2 + l(x^1dx^1 + x^2dx^2 + x^3dx^3)^2 + g_{44}(dx^4)^2. \tag{2}$$

The scalar curvature R has been calculated to be[1,4]

$$R = -\frac{1}{r^2\Delta}\frac{d}{dr}\left(\frac{r^2 g'_{44}}{\Delta}\right) + \frac{2}{r}\frac{\Delta'}{\Delta^3}g_{44} - \frac{2}{r^2\Delta}\frac{d}{dr}\left(\frac{r g_{44}}{\Delta}\right) - \frac{2}{r^2}, \quad \text{at } x^1 = r,\ x^2 = x^3 \doteq 0, \tag{3}$$

where dashes denote differentiation with respect to r and

$$g_{11} = h^2 = 1 + lr^2,\ \Delta = \sqrt{-g} = h\sqrt{-g_{44}}. \tag{4}$$

A charged point mass besides giving rise to a radially symmetrical gravitational (tensor–scalar field) also gives rise to a similar electrostatic field. Since both fields influence one another mutually, they may be determined cojointly and simultaneously.

For the electromagnetic case, the variational principle (1) can be written as[5]

$$\delta \int \left(\psi R + \frac{8\pi}{c^4}F_{ij}F^{ij} - \frac{\omega\psi_{,i}\psi^{,i}}{\psi}\right)\sqrt{-g}\, d^4x = 0, \tag{1'}$$

where

$F_{14}F^{41} = g^{11}g^{44}F_{14}F_{14} = -\Phi'^2/\Delta^2$ (for the electrostatic field) and hence in the region devoid of masses and charges, Eg. (1') can be written, using CGS units, as[3]

$$\delta \int \left(\psi R - \frac{8\pi}{c^4}r^2\frac{\Phi'^2}{\Delta} + \omega g_{44}\frac{\psi'^2 r^2}{\psi\Delta}\right)dr = 0, \tag{5}$$

where

$d^4x = d^4x^4\, d\Omega\, dr\, r^2$ ($d\Omega$ is an element of solid angle at the origin), R is given by (3), and Φ denotes the electrostatic potential.

Now variation with respect to ψ, Δ, g_{44}, and Φ in (5), respectively, leads to the following field equations:

$$-\frac{d}{dr}\left(\frac{r^2 g'_{44}}{\Delta}\right) + \frac{2r g_{44}\Delta'}{\Delta^2} - \frac{2d}{dr}\left(\frac{r g_{44}}{\Delta}\right) - 2\Delta = \frac{\omega r^2 g_{44}\psi'^2}{\Delta\psi^2} + 2\omega\frac{d}{dr}\left(\frac{r^2 g_{44}}{\Delta}\frac{\psi'}{\psi}\right), \tag{6}$$

$$-\frac{2rg'_{44}}{\Delta^2} - \frac{2g_{44}}{\Delta^2} - 2 = \frac{\omega r^2 g_{44}}{\Delta^2}\frac{\psi'^2}{\psi^2} - \frac{kr^2}{\psi}\frac{\Phi'^2}{\Delta^2}, \tag{7}$$

$$\frac{2r\Delta'}{\Delta^2} = -\frac{\omega r^2\psi'^2}{\Delta\psi^2}, \tag{8}$$

$$\frac{d}{dr}\left(\frac{r^2\Phi'}{\Delta}\right) = 0, \tag{9}$$

where $k = 8\pi/c^4$.

3. SOLUTION OF THE FIELD EQUATIONS

We try the following solution:

$$\frac{d}{dr}\left(\frac{r^2 g_{44}}{\Delta}\frac{\psi'}{\psi}\right) = 0. \tag{10}$$

And so from Eq. (6),

$$-\frac{d}{dr}\left(\frac{r^2 g'_{44}}{\Delta}\right) + \frac{2r g_{44}\Delta'}{\Delta^2} - \frac{2d}{dr}\left(\frac{r g_{44}}{\Delta}\right) - 2\Delta = \frac{\omega r^2 g_{44}}{\Delta}\frac{\psi'^2}{\psi^2}. \tag{6'}$$

Also we have from (7) and (8),

$$-\frac{2rg'_{44}}{\Delta^2} - \frac{2g_{44}}{\Delta^2} - 2 = \frac{\omega r^2 g_{44}}{\Delta^2}\frac{\psi'^2}{\psi^2} - \frac{kr^2\Phi'^2}{\psi\Delta^2}, \tag{7'}$$

[3] J. B. McGuire, J. Math. Phys. 5, 622 (1964).
[4] C. N. Yang. Phys. Rev. 168, 1920 (1968).
[5] J. C. Y. Chen, K. T. Chung, and P. J. Kramer, Phys. Rev. 184, 64 (1969).
[6] C. S. Shastry and A. K. Rajagopal, Phys. Rev. A 2, 781 (1970).
[7] C. K. Majumdar, Phys. Rev. B 1, 287 (1970); C. K. Majumdar and G. Mukhopadhyay, Phys. Letters 31A, 321 (1970).

An Approximate Solution for the Static, Spherically Symmetric Metric Due to a Point Charged Mass in Brans-Dicke Theory

M. N. Mahanta and D. R. K. Reddy
Mathematics Department, Indian Institute of Technology, New Delhi-29, India
(Received 11 January 1971)

An approximate solution of the field equations of Brans–Dicke theory is obtained for a static, spherically symmetric metric due to a charged point mass which can be considered to be an analog of the Reissner–Nordstrom metric in Einstein's theory.

1. INTRODUCTION

In a previous paper,[1] we have obtained an approximate solution of the vacuum static case of spherical symmetry in Brans–Dicke theory, starting from the variational principle[2]

$$\delta \int (\psi R - \omega \psi_{,i}\psi^{,i}/\psi)\sqrt{-g}\, d^4x = 0, \tag{1}$$

where R is the scalar curvature, ω the dimensionless constant, and ψ the scalar playing the role of G^{-1}, following a technique used first by Weyl[3] and then by Pauli.[4] A similar solution for the static, spherically symmetric metric due to a point charged mass is well worth consideration.

In this paper we obtain an approximate solution of the field equations of Brans–Dicke theory for a static, spherically symmetric metric due to a point charged mass following the same technique. The solution is then compared with the Reissner–Nordstrom solution in Einstein's theory.

2. FIELD EQUATIONS

We consider the line element for the static, spherically symmetric metric case in the form

$$ds^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2 + l(x^1dx^1 + x^2dx^2 + x^3dx^3)^2 + g_{44}(dx^4)^2. \tag{2}$$

The scalar curvature R has been calculated to be[1,4]

$$R = -\frac{1}{r^2\Delta}\frac{d}{dr}\left(\frac{r^2 g'_{44}}{\Delta}\right) + \frac{2}{r}\frac{\Delta'}{\Delta^3}g_{44} - \frac{2}{r^2\Delta}\frac{d}{dr}\left(\frac{r g_{44}}{\Delta}\right) - \frac{2}{r^2}, \quad \text{at } x^1 = r,\ x^2 = x^3 \doteq 0, \tag{3}$$

where dashes denote differentiation with respect to r and

$$g_{11} = h^2 = 1 + lr^2,\ \Delta = \sqrt{-g} = h\sqrt{-g_{44}}. \tag{4}$$

A charged point mass besides giving rise to a radially symmetrical gravitational (tensor–scalar field) also gives rise to a similar electrostatic field. Since both fields influence one another mutually, they may be determined cojointly and simultaneously.

For the electromagnetic case, the variational principle (1) can be written as[5]

$$\delta \int \left(\psi R + \frac{8\pi}{c^4}F_{ij}F^{ij} - \frac{\omega\psi_{,i}\psi^{,i}}{\psi}\right)\sqrt{-g}\, d^4x = 0, \tag{1'}$$

where

$F_{14}F^{41} = g^{11}g^{44}F_{14}F_{14} = -\Phi'^2/\Delta^2$ (for the electrostatic field) and hence in the region devoid of masses and charges, Eg. (1') can be written, using CGS units, as[3]

$$\delta \int \left(\psi R - \frac{8\pi}{c^4}r^2\frac{\Phi'^2}{\Delta} + \omega g_{44}\frac{\psi'^2 r^2}{\psi\Delta}\right)dr = 0, \tag{5}$$

where

$d^4x = d^4x^4\, d\Omega\, dr\, r^2$ ($d\Omega$ is an element of solid angle at the origin), R is given by (3), and Φ denotes the electrostatic potential.

Now variation with respect to ψ, Δ, g_{44}, and Φ in (5), respectively, leads to the following field equations:

$$-\frac{d}{dr}\left(\frac{r^2 g'_{44}}{\Delta}\right) + \frac{2r g_{44}\Delta'}{\Delta^2} - \frac{2d}{dr}\left(\frac{r g_{44}}{\Delta}\right) - 2\Delta = \frac{\omega r^2 g_{44}\psi'^2}{\Delta\psi^2} + 2\omega\frac{d}{dr}\left(\frac{r^2 g_{44}}{\Delta}\frac{\psi'}{\psi}\right), \tag{6}$$

$$-\frac{2rg'_{44}}{\Delta^2} - \frac{2g_{44}}{\Delta^2} - 2 = \frac{\omega r^2 g_{44}}{\Delta^2}\frac{\psi'^2}{\psi^2} - \frac{kr^2}{\psi}\frac{\Phi'^2}{\Delta^2}, \tag{7}$$

$$\frac{2r\Delta'}{\Delta^2} = -\frac{\omega r^2\psi'^2}{\Delta\psi^2}, \tag{8}$$

$$\frac{d}{dr}\left(\frac{r^2\Phi'}{\Delta}\right) = 0, \tag{9}$$

where $k = 8\pi/c^4$.

3. SOLUTION OF THE FIELD EQUATIONS

We try the following solution:

$$\frac{d}{dr}\left(\frac{r^2 g_{44}}{\Delta}\frac{\psi'}{\psi}\right) = 0. \tag{10}$$

And so from Eq. (6),

$$-\frac{d}{dr}\left(\frac{r^2 g'_{44}}{\Delta}\right) + \frac{2r g_{44}\Delta'}{\Delta^2} - \frac{2d}{dr}\left(\frac{r g_{44}}{\Delta}\right) - 2\Delta = \frac{\omega r^2 g_{44}}{\Delta}\frac{\psi'^2}{\psi^2}. \tag{6'}$$

Also we have from (7) and (8),

$$-\frac{2rg'_{44}}{\Delta^2} - \frac{2g_{44}}{\Delta^2} - 2 = \frac{\omega r^2 g_{44}}{\Delta^2}\frac{\psi'^2}{\psi^2} - \frac{kr^2\Phi'^2}{\psi\Delta^2}, \tag{7'}$$

$$2rg_{44}\frac{\Delta'}{\Delta^2} = -\omega r^2\frac{g_{44}}{\Delta}\frac{\psi'^2}{\psi^2}. \tag{8'}$$

Equations (6'), (7'), and (8') constitute only two independent equations which can be taken as

$$\left.\begin{aligned}\frac{d}{dr}\left(\frac{rg_{44}}{\Delta}\right) &= -\Delta\left[1-\frac{kr^2\Phi'^2}{2\psi\Delta^2}\right],\\ \frac{d}{dr}\left(\frac{r^2g'_{44}}{\Delta}\right) &= \Delta\left[\frac{4rg_{44}\Delta'}{\Delta^3}-\frac{kr^2\Phi'^2}{\psi\Delta^2}\right].\end{aligned}\right\} \tag{11}$$

Also from (9) we get

$$\frac{r^2\Phi'}{\Delta} = \text{const} = e \tag{12}$$

(which is identified with the charge e of the point mass). In view of the difficulty in finding an exact solution of the Eqs. (11) and (12), we find an approximate solution correct up to the second order in $1/r$ by using the method of successive approximation.

Let us consider the field equations (6), (7), (8), and (9). When $\psi = \text{const} = \psi_0$ (which is to be calculated to the second order in ψ_0^{-1}), we get the Reissner–Nordstrom solution

$$\left.\begin{aligned}\Delta &= 1,\\ g_{44} &= -1+2m/r-ke^2/r^2,\\ \Phi &= e/r.\end{aligned}\right\} \tag{13}$$

Now substituting the values from (12) and (13) in the right-hand side of Eqs. (11) and integrating with respect to r, we get

$$g_{44}/\Delta = -(1 + C/2r^2 - \lambda/r) \tag{14}$$

and

$$g'_{44}/\Delta = \mu/r^2 + C/r^3, \tag{15}$$

where λ, μ, C are constants. Here C is given by

$$C = 8\pi e^2/\psi_0 c^4. \tag{16}$$

Eliminating Δ between (14) and (15), integrating with respect to r, and keeping the terms up to the second order in $1/r$ in the expansion of the resulting expression for g_{44}, we get

$$g_{44} \approx -[1 + \mu/r + (C + \lambda\mu)/2r^2]. \tag{17}$$

In (17) and hereafter in the calculation of g_{11}, Φ, ψ and Δ, we neglect the terms yielding ψ_0^{-2} or G_0^2 which are very small, after identifying the constants later with the Brans–Dicke's weak-field approximations. So (17) becomes now

$$g_{44} \approx -(1 + \mu/r + C/2r^2). \tag{18}$$

Now using (18) in (14), we get

$$\Delta \approx [1 + (\lambda + \mu)/r]. \tag{19}$$

Now in view of Eqs. (4), (18), and (19) we get

$$h^2 = g_{11} \approx [1 + (2\lambda + \mu)/r - C/2r^2]. \tag{20}$$

Again using (14) in (10) and integrating, we get the value of ψ as

$$\psi \approx P[1 + A/r + (1/2r^2)(A\lambda + A^2)], \quad P = \text{const}, \quad A = \text{const}. \tag{21}$$

Now equating the coefficients of $1/r$ in (18), (20), and (21) to Brans–Dicke's[2] weak-field approximations, we get (since to the first order these values should be the same)

$$\left.\begin{aligned}-\mu &= 2MG_0/c^2,\\ 2\lambda + \mu &= (2MG_0/c^2)[(1+\omega)/(2+\omega)],\end{aligned}\right.$$

so that

$$\left.\begin{aligned}\lambda &= (2MG_0/c^2)/(3+2\omega)/(4+2\omega),\\ P &= G_0^{-1}(4+2\omega)(3+2\omega)^{-1},\\ A &= 2MG_0/(4+2\omega)c^2r,\end{aligned}\right\} \tag{22}$$

where

$$G_0 = \psi_0^{-1}(4+2\omega)(3+2\omega)^{-1}. \tag{23}$$

Also from (12) and (19), using (22), we get

$$\Phi = e[(MG_0/r^2(4+2\omega)c^2) + 1/r]. \tag{24}$$

Now considering (16), (18), (20), (22), and (24), we can write down the approximate solution for the metric (2) due to a point charged mass in Brans–Dicke's theory as

$$\left.\begin{aligned}g_{44} &\approx -1 + \frac{2MG_0}{c^2r} - \frac{4\pi e^2G_0}{c^4r^2}\left(\frac{3+2\omega}{4+2\omega}\right),\\ g_{11} &\approx 1 + \frac{2MG_0}{c^2r}\left(\frac{1+\omega}{2+\omega}\right) - \frac{4\pi e^2G_0}{c^4r^2}\left(\frac{3+2\omega}{4+2\omega}\right),\\ \Phi &\approx e\left(\frac{MG_0}{c^2r^2(4+2\omega)} + \frac{1}{r}\right),\\ \psi &\approx \psi_0 + \frac{2M}{(3+2\omega)c^2r} + \frac{4M^2\psi_0^{-1}}{r^2c^4}\left(\frac{2+\omega}{(3+2\omega)^2}\right).\end{aligned}\right\} \tag{25}$$

Thus, we see that the solution (25) is an analog of the Reissner–Nordstrom solution in Einstein's theory. It is interesting to note that the coefficient of $1/r^2$ in g_{44} and g_{11} are multiplied by the factor $(3 + 2\omega)/(4 + 2\omega)$. Also an additional term occurs in the electrostatic potential Φ.

ACKNOWLEDGMENT

We are thankful to The Council of Scientific and Industrial Research, India for the financial assistance.

[1] M. N. Mahanta and D. R. K. Reddy J. Math Phys. **12**, 929 (1971).
[2] C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).
[3] H. Weyl, *Space–Time Matter* (Dover, New York, 1922).
[4] W. Pauli, *Theory of Relativity* (Pergamon, Oxford, London, 1958).
[5] P. Jordan, *Schwerkraft und Weltall* (Vieweg, Braunschweig, 1955).

About the Solutions of Dirac's Equations in the Presence of New Magnetic Fields

André Hautot
University of Liège, Sart Tilman, Belgium
(Received 2 December 1970)

Dirac's equations are completely solved for four entirely new configurations of the external magnetic field. The use of the quaternionic formalism simplifies the problem of separation of variables. For three of those new magnetic fields the radial equations just obtained are not classical. Section II studies them and gives their solutions in term of usual transcendental functions. When energy is quantized, energy levels are given.

The problem of finding all electrical or magnetical fields which allow a complete integration of Dirac's equations has always interested physicists. Lam[1] and Stanciu[2] have established a quasicomplete bibliography about that problem. However they omitted an important contribution of Harish-Chandra.[3] We bring our contribution by solving Dirac's equation for four entirely new configurations of the external magnetic field. The mathematical problem involved is rather complicated, and may easily be solved only if we use the following two-step procedure:

(I) separation of variables by a quaternionic method we have previously indicated,[4]

(II) resolution of the separated differential equations which, as we shall see, are not classical.

I. SEPARATION OF VARIABLES

(A) Let us first consider Dirac's equations in cylindrical coordinates:

$$ie^{-k\varphi}\frac{\partial u}{\partial r}+\frac{1}{r}\,je^{-k\varphi}\frac{\partial u}{\partial \varphi}+k\frac{\partial u}{\partial z}=-u\left(\frac{m_0c}{\hbar}\,j+\frac{\sqrt{-1}E}{c\hbar}i\right)-\frac{\epsilon\sqrt{-1}}{\hbar}\,(\mathrm{vect}P)u, \qquad (1)$$

where i, j, k are the symbols of quaternions, $\sqrt{-1}$ the symbol of complex number, $\mathrm{vect}P = iA_x + jA_y + kA_z$ the magnetic potential, ϵ the charge of the electron.

The solution of (1) may be found under the separated form ($m = \cdots, -2, -1, 0, 1, 2, \cdots$)

$$u = e^{k\varphi/2}\;e^{(\sqrt{-1}/\hbar)p_z z}\;e^{\sqrt{-1}(m+1/2)\varphi}\times(R_1+iR_2+jR_3+kR_4) \qquad (2)$$

if and only if

$$\mathrm{vect}P = e^{k\varphi/2}\,(jS_3+kS_4)e^{-k\varphi/2},$$

where S_3 and S_4 are arbitrary functions of r.

The corresponding magnetic field may be deduced; in vectorial familiar notation it is written ($r^2 = x^2+y^2$).

$$\mathbf{B} = \mathrm{curl}\mathbf{A} = \left(\frac{y}{r}\frac{dS_4}{dr}, -\frac{x}{r}\frac{dS_4}{dr}, \frac{dS_3}{dr}+\frac{S_3}{r}\right). \qquad (3)$$

Introducing (2) into (1), we deduce the radial equation for the quaternion $R = R_1 + iR_2 + jR_3 + kR_4$:

$$ir\frac{dR}{dr}+\tfrac{1}{2}\,iR+j\sqrt{-1}\,(m+\tfrac{1}{2})R+\frac{\sqrt{-1}}{\hbar}\,p_z rkR+rR\left(\frac{m_0c}{\hbar}\,j+\frac{\sqrt{-1}E}{c\hbar}\,i\right)+\frac{\epsilon\sqrt{-1}}{\hbar}\,r\,(jS_3+kS_4)R=0. \qquad (4)$$

This quaternionic equation is equivalent to a system of four coupled differential equations. The extreme complexity of that system can be avoided by using the quaternionic notation. Multiplying (4) at both sides by the constant quaternion $Q = (m_0c/\hbar)\,j + (\sqrt{-1}E/c\hbar)\,i$, we see that the equation remains unchanged if we consider the new unknown function RQ. Therefore, we have

$$RQ = \lambda R, \qquad \lambda = \text{scalar}.$$

That equation is equivalent to an algebraic homogenous system of four equations; it is compatible only if

$$\lambda^2 = a^2 - b^2 \quad \text{where} \quad a = E/c\hbar \quad \text{and} \quad b = m_0c/\hbar.$$

Equation (4) is then equivalent to a coupled system of two differential equations plus two algebraic relations which give R_1 and R_4, for example, when R_2 and R_3 are known. Eliminating R_1 and R_4 and putting

$$T_2 = R_2 + R_3\sqrt{-1}, \qquad T_3 = R_3 + R_2\sqrt{-1},$$

there only remains the following differential system:

$$\begin{cases} r\dfrac{dT_2}{dr}+(m+1)T_2+(a+b)\dfrac{p_z+\lambda\hbar+\epsilon S_4}{\lambda\hbar}\,rT_3 \\ \quad +\dfrac{\epsilon}{\hbar}\,S_3rT_2=0, \\ r\dfrac{dT_3}{dr}-mT_3+(a-b)\dfrac{p_z-\lambda\hbar+\epsilon S_4}{\lambda\hbar}\,rT_2 \\ \quad -\dfrac{\epsilon}{\hbar}\,S_3rT_3=0. \end{cases} \qquad (5)$$

That system is exactly soluble in four cases. Let us study them. We first point out that $S_4 = \text{const}$ and $S_3 = \alpha/r$ are to be rejected because they lead to $\mathbf{B} = 0$.

1. $S_3 = (B/2)r, \quad S_4 = 0.$

Thus $\mathbf{B} = (0, 0, B) = \text{constant magnetic field}$. This problem is classical.[5] The following are new.

2. $S_3 = \alpha \quad (= \text{const}), \quad S_4 = 0.$

Thus $\mathbf{B} = (\alpha/\sqrt{x^2+y^2})(0, 0, 1)$.

Decoupling system (5), we obtain second-order differential equations for T_2 and T_3 which are confluent hypergeometric equations (we deal with the case $m \geq 0$; analogous calculations hold when $m < 0$):

$$\begin{cases} T_2 = \mathrm{const}\,r^{m+1}e^{-\sqrt{N}r}F(1-n;2m+3;2\sqrt{N}r), \\ T_3 = \mathrm{const}\,r^{m}\,e^{-\sqrt{N}r}F(-n;2m+1;2\sqrt{N}r), \end{cases}$$

where

$$N = AD + \left(\frac{\epsilon\alpha}{\hbar}\right)^2 \text{ and } A = (a+b)\frac{p_z+\lambda\hbar}{\lambda\hbar}, \quad D = (a-b)\frac{p_z-\lambda\hbar}{\lambda\hbar},$$

$n = 0, 1, 2, 3, \cdots$, when $n = 0$, $F(1; 2m + 3; 2\sqrt{N}r)$ is not a polynomial, but in this case $T_2 = 0$ because the constant it contains is zero on account of (5).

The energy levels are given by the polynomial condition:

$$E^2 = m_0^2c^4 + c^2p_z^2 + (\epsilon\alpha c)^2 \frac{4n(n + 2m + 1)}{(2n + 2m + 1)^2}$$

3. $S_3 = 0, \quad S_4 = \gamma/r.$

Thus $\mathbf{B} = [\gamma/(x^2 + y^2)^{3/2}](-y, x, 0).$

Decoupling system (5), we obtain nonclassical second-order differential equations for T_2 and T_3:

$$r^2[d(a + b) - Ar]\frac{d^2T_2}{dr^2} + \{Ar^2 + 2[d(a + b) - Ar]r\}\frac{dT_2}{dr} + \{(m + 1)Ar - m(m + 1)[d(a + b) - Ar] - [d(a - b) - Dr][d(a + b) - Ar]^2\}\ T_2 = 0 \qquad (6)$$

(with $d = -\epsilon\gamma/\lambda\hbar$) (idem for T_3).

Problem 4 leads to the same conclusion.

4. $S_3 = \alpha, \quad S_4 = \gamma/r.$

$$\mathbf{B} = \frac{1}{\sqrt{x^2 + y^2}}\left(\frac{-\gamma y}{x^2 + y^2}, \frac{\gamma x}{x^2 + y^2}, \alpha\right)$$

is in fact the superposition of the two former fields. The theorems contained in Sec. II are necessary to solve completely problems 3 and 4.

(B) Before enunciating these theorems, we shall solve the same problem in spherical coordinates. We start with Dirac's equations in the quaternionic formalism:

$$ke^{k\varphi/2}e^{-j\theta}e^{-k\varphi/2}\frac{\partial u}{\partial r} + ie^{-k\varphi/2}e^{-j\theta}e^{-k\varphi/2}\frac{1}{r}\frac{\partial u}{\partial\theta} + \frac{je^{-k\varphi}}{r\sin\theta}\frac{\partial u}{\partial\varphi} = -uQ - \frac{\epsilon\sqrt{-1}}{\hbar}(\text{vect}\,P)u \qquad (7)$$

The solution of (7) may be found under the separated form ($m = \cdots, -2, -1, 0, 1, 2, \cdots$)

$$u = e^{k\varphi/2}e^{j\theta/2}e^{\sqrt{-1}(m+1/2)\varphi}\,\textcircled{u}\,(\theta)R(r),$$

where $\textcircled{u} = \textcircled{u}_1 + k\,\textcircled{u}_4$ and $R = (R_1 + iR_2)(1 - j)$ if $\text{vect}\,P = (1/r)e^{k\varphi/2}e^{j\theta/2}\,jg\,(\theta)e^{-j\theta/2}e^{-k\varphi/2}$.

The corresponding magnetic field is

$$\mathbf{B} = (g\cot\theta + g')\mathbf{r}/r^3.$$

g is an arbitrary scalar function of θ (except $1/\sin\theta$ because $\mathbf{B} = 0$ in that case). Both $\textcircled{u}$ and R satisfy a quaternionic equation. The R equation is analogous to the radial equation when there is no magnetic field. The $\textcircled{u}$ equation is identical to the corresponding $\textcircled{u}$ equation in the relativistic hydrogen problem apart from a factor depending upon g:

$$j\frac{d\textcircled{u}}{d\theta} - \frac{\sqrt{-1}(m + \frac{1}{2})}{\sin\theta}\,i\,\textcircled{u} + \tfrac{1}{2}j\cot\theta\,\textcircled{u} - \frac{\epsilon\sqrt{-1}}{\hbar}gi\textcircled{u} = a\,\textcircled{u}\,i.$$

That equation which becomes real if we put

$$\begin{cases} 2\textcircled{u}_1 = T_1 + \sqrt{-1}T_4 \\ -2\textcircled{u}_4 = T_4 + \sqrt{-1}T_1 \end{cases}$$

is generally insoluble except in the two following cases.

1. $g = \mu\cot\theta.$

We deduce $\mathbf{B} = \mu\mathbf{r}/r^3.$

Harish-Chandra[3] has solved that problem when there is no electrical field. Yet equations are also soluble when Coulomb field is simultaneously considered: It might be shown that the $\textcircled{u}$ equations remains unchanged while the radial equation is analogous to the hydrogen radial equation except that $(j + \frac{1}{2})^2$ must be replaced by $(j + \frac{1}{2})^2 - (\epsilon\mu/\hbar)^2$. The energy levels are therefore given by

$$E^2 = m_0^2c^4\left(1 + \frac{Z^2\alpha^2}{[n + \sqrt{(j + \frac{1}{2})^2 - (\epsilon\mu/\hbar)^2 - Z^2\alpha^2}]^2}\right)^{-1}$$

2. $g = \nu\tan\theta$

We deduce $\mathbf{B} = \nu(1 + r^2/z^2)\mathbf{r}/r^3.$

If $m \geqslant 0$, the $\textcircled{u}$ equation when decoupled leads to $[u = \cos^2(\theta/2);\ T_1 = u^{m/2}(1 - u)^{(m+1)/2}(1 - 2u)^{\epsilon\nu/\hbar}\,V_1]$

$$u(1 - u)(1 - 2u)V_1'' + [(1 - 2u)(m + 1 - 2mu - 3u) - (4\epsilon\nu/\hbar)u(1 - u)]V_1' + (A + Bu)V_1 = 0, \qquad (8)$$

where A and B are constants. The theorems of Sec. II (idem for V_4) are absolutely necessary to solve completely that problem.

II. THREE NEW DIFFERENTIAL EQUATIONS

We have studied the three following second-order differential equations:

$$DP'' + (az^2 + bz + c)P' + (d + ez + fz^2)P = 0,$$

where $D \equiv z, z(z - 1)$ or $z(1 - z)(\alpha - z).$

They give rise to the following theorems.

Theorem 1: $zP_n'' + (az^2 + bz + c)P_n' + (d + ez + fz^2)P_n = 0$ admits polynomial solutions if

$$\left\{\begin{array}{l} \left.\begin{array}{l} e = -an \\ f = 0 \end{array}\right\} \text{necessary conditions} \quad (n = 0, 1, 2, \cdots) \\ c = -j \quad (= 0, -1, -2, \cdots \text{fixed}) \\ \left|\begin{array}{cccc} S_0 & T_0 & & \\ R_1 & S_1 & T_1 & \\ \cdot & \cdot & \cdot & \\ & \cdot & \cdot & \cdot \\ & & R_{j-1} & S_{j-1} & T_{j-1} \\ & & & R_j & S_j \end{array}\right| = 0, \end{array}\right.$$

where $\begin{cases} R_k = a(k - 1 - n), \\ S_k = d + bk, \\ T_k = (k + 1)(k + c). \end{cases}$

The polynomials are written as a linear combination of $j+1$ Hermite polynomials:

$$\boxed{P_n = \sum_{k=0}^{j} A_k H_{n-k}\left(\frac{az+b}{\sqrt{-2a}}\right),} \qquad H_0 = 1, H_1 = 2x, \text{ etc.}$$

The A_k are solutions of the recurrent system

$$\mathcal{R}_k A_{k-1} + \mathcal{S}_k A_k + \mathcal{T}_k A_{k+1} = 0$$

with
$$\begin{cases} \mathcal{R}_k = -\sqrt{-2a}(n-k+1)(k-j-1), \\ \mathcal{S}_k = (d+bk), \\ \mathcal{T}_k = -\tfrac{1}{2}\sqrt{-2a}\,(k+1). \end{cases}$$

Theorem 2: $z(z-1)P_n'' + (az^2+bz+c)P_n' + (d+ez+fz^2)P_n = 0$ admits polynomial solutions if

$$\begin{cases} e = -an \\ f = 0 \end{cases} \text{ necessary conditions } (n = 0, 1, 2, \cdots)$$

$$c = j \quad (= 0, 1, 2, \ldots \text{ fixed})$$

$$\begin{vmatrix} S_0 & T_0 & & & \\ R_1 & S_1 & T_1 & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot \\ & & & R_{j-1} & S_{j-1} & T_{j-1} \\ & & & & R_j & S_j \end{vmatrix} = 0,$$

where
$$\begin{cases} R_k = a(k-1-n), \\ S_k = d + k(b+k-1), \\ T_k = (k+1)(c-k). \end{cases}$$

The polynomials are written as a linear combination of $j+1$ Laguerre polynomials:

$$\boxed{P_n = \sum_{k=0}^{j} A_k F[k-n, a+b+c; a(1-z)]}$$

$$F(a, b, y) = 1 + \frac{a}{b}\frac{y}{1!} + \cdots .$$

The A_k are solutions of the recurrent system

$$\mathcal{R}_k A_{k-1} + \mathcal{S}_k A_k + \mathcal{T}_k A_{k+1} = 0$$

with
$$\begin{cases} \mathcal{R}_k = (k-1-c)(k-1-n), \\ \mathcal{S}_k = d - cn + k(b+2c-2k+2n), \\ \mathcal{T}_k = (k+1)(k+1-n-a-b-c). \end{cases}$$

Remark 1: Another solution deduces from the former if we simultaneously replace

z by $1-z$, $\quad c$ by $-(a+b+c)$,
a by $-a$, $\quad d$ by $d-an$,
b by $2a+b$, $\quad n$ by n.

Theorem 3: $z(1-z)(\alpha-z)P_n'' + (az^2+bz+c)P_n' + (d+ez+fz^2)P_n = 0$ admits polynomial solutions if

$$\begin{cases} e = -n(n+a-1) \\ f = 0 \end{cases} \text{ necessary conditions } (n = 0, 1, 2, \cdots)$$

$$c = -\alpha j \quad (= 0, -\alpha, -2\alpha, \cdots \text{ fixed})$$

$$\begin{vmatrix} S_0 & T_0 & & & \\ R_1 & S_1 & T_1 & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot \\ & & & R_{j-1} & S_{j-1} & T_{j-1} \\ & & & & R_j & S_j \end{vmatrix} = 0,$$

where
$$\begin{cases} R_k = (k-1)(a+k-2) - n(n+a-1) \\ S_k = d + bk - (\alpha+1)k(k-1), \\ T_k = (k+1)(c+\alpha k). \end{cases}$$

The polynomials are written as a linear combination of $j+1$ Jacobi polynomials:

$$\boxed{P_n = \sum_{k=0}^{j} A_k F\left(k-n, n-k+a+j-1; \frac{a+b+c}{1-\alpha}; \frac{1-z}{1-\alpha}\right).}$$

The A_k are solutions of the recurrent system

$$\mathcal{R}_k A_{k-1} + \mathcal{S}_k A_k + \mathcal{T}_k A_{k+1} = 0$$

with
$$\begin{cases} \mathcal{R}_k = \dfrac{(\alpha-1)(k-1-n)[(a+b+c)/(1-\alpha) - n + k - a - j](2n-k+a+j)(k-j-1)}{(2n-2k+a+j+1)(2n-2k+a+j)}, \\ \mathcal{S}_k = d - \alpha Bk - \alpha n(k-j) - \dfrac{1-\alpha}{(B-A-1)(B-A+1)} \\ \qquad \times \{jAB(2C-B-A-1) - [Bk + n(k-j)](1-B^2-A^2-C+CB+CA)\}, \\ \mathcal{T}_k = \dfrac{(\alpha-1)(n-k+a+j-2)[(a+b+c)/(1-\alpha) - k - 1 + n](k+1)(2n+a-k-2)}{(2n-2k+a+j-3)(2n-2k+a+j-2)}, \end{cases}$$

with $A = k - n$, $B = n - k + a + j - 1$,
and $C = (a + b + c)/(1 - \alpha)$.

Remark 2: As in Theorem 2, other solutions occur when we replace in the starting equation

(a) z by $1 - z$, (c) z by $(\alpha - 1)z + 1$,
(b) z by $\alpha(1 - z)$, (d) z by $(1 - \alpha)z + \alpha$,
(e) z by αz.

The demonstrations are similiar for the three theorems. Here we prove only Theorem 3. We start from the equation

$$z(1 - z)(\alpha - z)P_n'' + (az^2 + bz + c)P_n' + (d + ez + fz^2)P_n = 0. \quad (9)$$

Let us introduce into (9) the polynomial form $P_n = \sum_{k=0}^{n} \lambda_k z^k$; it follows that $f = 0$ and $e = -n(n + a - 1)$ are necessary and that

$$[(k - 1)(a + k - 2) - n(n + a - 1)]\lambda_{k-1} + [d + bk - (\alpha + 1)k(k - 1)]\lambda_k + (k + 1)(c + \alpha k)\lambda_{k+1} = R_k\lambda_{k-1} + S_k\lambda_k + T_k\lambda_{k+1} = 0.$$

That condition is satisfied if

$$\begin{vmatrix} S_0 & T_0 & & & \\ R_1 & S_1 & T_1 & & \\ \cdot & \cdot & \cdot & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot \\ & & R_{n-1} & S_{n-1} & T_{n-1} \\ & & & R_n & S_n \end{vmatrix} = 0. \quad (10)$$

It seems very difficult to deduce the general polynomial solution because the order of the determinant increases with the order n of the polynomial. However, if we suppose $c = -\alpha j$ (j *fixed* integer) $T_j = 0$ and condition (10) may be replaced by

$$\begin{vmatrix} S_0 & T_0 & & & \\ R_1 & S_1 & T_1 & & \\ \cdot & \cdot & \cdot & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot \\ & & R_{j-1} & S_{j-1} & T_{j-1} \\ & & & R_j & S_j \end{vmatrix} = 0$$ whose order does not depend upon n.

Now let us introduce $P_n = \sum_{k=0}^{j} A_k F(k - n, n - k + a + j - 1; (a + b + c)/(1 - \alpha); (1 - z)/(1 - \alpha))$ into equation (9) taking into account the differential equation satisfied by the F functions, the following remains to be proved:

$$\sum_{k=0}^{j} A_k \{ j(1 - z)(\alpha - z)\frac{dF(k - n)}{dz} + [(2kn - k^2 + ak - k + jk - nj)z - d] \times F(k - n) \} = 0.$$

Remembering the recurrence relations existing between $(1 - z)(\alpha - z)dF(k - n)/dz$ and $zF(k - n)$ on one side, $F(k - 1 - n)$ and $F(k + 1 - n)$ and $F(k - n)$ on another side, it follows that

$$\sum_{k=0}^{j} F(k - n)\{\mathcal{R}_k A_{k-1} + \mathcal{S}_k A_k + \mathcal{T}_k A_{k+1}\} = 0,$$

where $\mathcal{R}_k$, $\mathcal{S}_k$, and $\mathcal{T}_k$ are given above. Therefore, our theorem is proved and the A_k are solutions of $\mathcal{R}_k A_{k-1} + \mathcal{S}_k A_k + \mathcal{T}_k A_{k+1} = 0$. Moreover, for our three theorems the reader will verify the curious equality correct for all j-values:

$$\begin{vmatrix} S_0 & T_0 & & & \\ R_1 & S_1 & T_1 & & \\ \cdot & \cdot & \cdot & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot \\ & & R_{j-1} & S_{j-1} & T_{j-1} \\ & & & R_j & S_j \end{vmatrix} = \begin{vmatrix} \mathcal{S}_0 & \mathcal{T}_0 & & & \\ \mathcal{R}_1 & \mathcal{S}_1 & \mathcal{T}_1 & & \\ \cdot & \cdot & \cdot & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot \\ & & \mathcal{R}_{j-1} & \mathcal{S}_{j-1} & \mathcal{T}_{j-1} \\ & & & \mathcal{R}_j & \mathcal{S}_j \end{vmatrix}$$

Remarks:

(a) Those theorems remain entirely valid when n is not an integer. In that case of course the solutions are not polynomial.
(b) In the third theorem we deduce from Remark 2 that condition $c = -\alpha j$ can be replaced by $a + b + c = -j(1 - \alpha)$ or by

$$a\alpha^2 + b\alpha + c = -j\alpha(\alpha - 1) \quad (11)$$

following the case.

III. APPLICATION TO THE PROBLEMS OF SEC. I

A. Dirac's Equations in the Magnetic Field,

$$\mathbf{B} = \frac{1}{\sqrt{x^2 + y^2}}\left(\frac{-\gamma y}{x^2 + y^2}, \frac{\gamma x}{x^2 + y^2}, \alpha\right)$$

The radial equations for T_2 and T_3 may be solved by using Theorem 2 of Sec. II. It is easily seen that $j = 1$. T_2 and T_3 appear as linear combinations of two confluent hypergeometric functions while energy levels are given by the polynomial condition

$$E^2 = m_0^2c^4 + c^2p_z^2 + (\epsilon\alpha c)^2 - \left(\frac{\epsilon\alpha c(2m + 1) + 2cp_z\epsilon\gamma/\hbar}{2n + \sqrt{(2m + 1)^2 + 4(\epsilon\gamma/\hbar)^2}}\right)^2.$$

We pass over the explicit writing of T_2 and T_3 for the sake of brevity.

B. Dirac's Equations in the Magnetic Field,

$$\mathbf{B} = \nu(1 + r^2/z^2)\,\mathbf{r}/r^3$$

The θ functions T_1 and T_4 may be found by Theorem 3 of Sec. II only if one among the three relations (11) may be satisfied. Only the third may be, provided that

$$\nu = -j\hbar/2\epsilon \quad \text{(remember } j \text{ is an integer).} \quad (12)$$

Remembering that ν has the dimensions of a magnetic pole, we see that relation (12) is absolutely identical with Dirac's rule of quantization concerning magnetic poles,[6] rule he has established in a very different way. Therefore, we may conclude with this very curious remark: Dirac's equation are exactly soluble only when the parameter ν is quantized following the rules of quantum mechanics.

We omit again the explicit writing of the functions T_1 and T_4. They appear as linear combinations of $j+1$ hypergeometric functions. Energy would be quantized if we added Coulomb electrical field.

[1] L. Lam, Phys. Letters **31A**, 406 (1970).
[2] G. N. Stanciu, Phys. Letters **23A**, 232 (1966).
[3] Harish-Chandra, Phys. Rev. **74**, 883 (1948).
[4] André Hautot, Physica **48**, 609 (1970).
[5] I. I. Rabi, Z. Phys. **49**, 507 (1928).
[6] P. A. M. Dirac, Phys. Rev. **74**, 817 (1948).

Two-Time Spin-Pair Correlation Function of the Heisenberg Magnet at Infinite Temperature. II

Tohru Morita*
Department of Physics, Ohio University, Athens, Ohio 45701

(Received 22 November 1971)

The two-time spin-pair correlation function $\sigma(R,t)$ is expressed as a product of a Gaussian distribution function and a power series with respect to time. The width of the Gaussian distribution function is determined by the second derivative with respect to time of the autocorrelation function $\sigma(0,t)$. The coefficients of the power series are determined up to terms of order t^8 for the square and s.c. isotropic Heisenberg and XY magnets of spin $\frac{1}{2}$ at infinite temperature, and up to terms of t^{10} for the linear magnets. The expressions obtained by truncating the power series to the exactly known terms give a very good fit at short times to the exact expression for the linear XY model, and to the results of the computer simulation due to Windsor for the s.c. Heisenberg ferromagnet. The Fourier-time transforms of the expressions thus obtained are shown to be the Gram–Charlier expansions. Thus we conclude that the Gram–Charlier expansion of the Fourier-time transform of $\sigma(R,t)$, $\tilde{\sigma}(R,\omega)$, and that of its Fourier-space transform $I(k,t)$, $S(k,\omega)$, is especially useful in the cases when the short-time behavior of $\sigma(R,t)$ or $I(k,t)$ is considered important; generally speaking, this is the case at large k or at ω which is not very small. When the width of the Gaussian distribution function is determined by the second derivative of $I(k,t)$, a convergent result is not obtained for small values of k. For larger k, the convergence is as good as for the case when the width is determined by the autocorrelation function.

1. INTRODUCTION

The two-time spin-pair correlation function is the quantity of primary importance in the theory of neutron scattering from magnetic materials.[1,2] In a number of papers,[3-8] calculations are focused on the function at infinite temperature, for simplicity of the treatment. For example, Windsor[4] performed a computer simulation calculation for the Heisenberg magnet of classical spins at infinite temperature. Carboni and Richards[5] gave an exact calculation, for a finite chain, of the Heisenberg magnet of spin $\frac{1}{2}$ at infinite temperature.

In the preceding paper,[9] the author gave the numerical values of the coefficients of the expansion in powers of time of the two-time spin-pair correlation function $\langle s_i^z(t)s_f^z(0)\rangle\rangle$ of spin $\frac{1}{2}$ at infinite temperature. In the present paper, we express $\langle s_i^z(t)s_f^z(0)\rangle\rangle$ and its Fourier-space transform $\langle S_k^z(t)S_{-k}^z(0)\rangle\rangle$ as a product of a Gaussian distribution function and a power series.

It is noted that the Fourier-time transform of this expression gives the Gram–Charlier expansion, which was proposed by Collins and Marshall[7] for the analysis of the Fourier-time transform of $\langle S_k^z(t)S_{-k}^z(0)\rangle\rangle$. The convergence of those expansions is discussed for the Heisenberg magnet and the XY magnet. For the one-dimensional XY magnet, the results are compared with the exact solution.

In the remaining part of this introduction, definitions adopted in this paper are given. The Hamiltonian of the system is

$$H = -\sum_f \sum_g [J_\perp(f,g)s_f^- s_g^+ + J_\parallel(f,g)s_f^z s_g^z], \quad (1.1)$$

where $J_\perp(f,g)$ and $J_\parallel(f,g)$ are equal to $J_\perp$ and $J_\parallel$, respectively, when f and g are nearest neighbors of each other and zero otherwise. For the Heisenberg magnet, $J_\perp = J_\parallel \equiv J$. If $J_\parallel = 0$ and the system is one-dimensional, the system is the XY model. We shall call the case of $J_\parallel = 0$ the XY magnet in general. The two-time spin-pair correlation function $\sigma(R_{if},t)$ is defined by

$$\sigma(R_{if},t) = \langle s_i^z(t)s_f^z(0)\rangle\rangle - \langle s_i^z\rangle\langle s_f^z\rangle, \quad (1.2)$$

where

$$s_i^z(t) = e^{iHt}s_i^z e^{-iHt}.$$

We shall introduce the Fourier-space transform of $s_i^z(t)$ by

$$S_k^z(t) = \sum_f s_f^z(t)\, e^{ik\cdot R_f}. \quad (1.3)$$

The Fourier-space transform of $\sigma(R_{if},t)$ is the so-called "intermediate scattering function." It will be denoted by $I(k,t)$:

$$I(k,t) = \sum_f \sigma(R_{if},t)\, e^{ik\cdot R_{if}}. \quad (1.4)$$

It is the correlation function of $S_k^z(t)$:

$$I(k,t) = N^{-1}\langle S_k^z(t)S_{-k}^z(0)\rangle\rangle, \qquad k \neq 0, \quad (1.5)$$

where N is the total number of spins in the system. We shall focus our attention mainly on $\sigma(R_{if},t)$ in Secs. 2 and 3, and on $I(k,t)$ in Sec. 4.

The θ functions T_1 and T_4 may be found by Theorem 3 of Sec. II only if one among the three relations (11) may be satisfied. Only the third may be, provided that

$$\nu = -j\hbar/2\epsilon \quad \text{(remember } j \text{ is an integer).} \quad (12)$$

Remembering that ν has the dimensions of a magnetic pole, we see that relation (12) is absolutely identical with Dirac's rule of quantization concerning magnetic poles,[6] rule he has established in a very different way. Therefore, we may conclude with this very curious remark: Dirac's equation are exactly soluble only when the parameter ν is quantized following the rules of quantum mechanics.

We omit again the explicit writing of the functions T_1 and T_4. They appear as linear combinations of $j+1$ hypergeometric functions. Energy would be quantized if we added Coulomb electrical field.

[1] L. Lam, Phys. Letters **31A**, 406 (1970).
[2] G. N. Stanciu, Phys. Letters **23A**, 232 (1966).
[3] Harish-Chandra, Phys. Rev. **74**, 883 (1948).
[4] André Hautot, Physica **48**, 609 (1970).
[5] I. I. Rabi, Z. Phys. **49**, 507 (1928).
[6] P. A. M. Dirac, Phys. Rev. **74**, 817 (1948).

Two-Time Spin-Pair Correlation Function of the Heisenberg Magnet at Infinite Temperature. II

Tohru Morita*
Department of Physics, Ohio University, Athens, Ohio 45701

(Received 22 November 1971)

The two-time spin-pair correlation function $\sigma(R,t)$ is expressed as a product of a Gaussian distribution function and a power series with respect to time. The width of the Gaussian distribution function is determined by the second derivative with respect to time of the autocorrelation function $\sigma(0,t)$. The coefficients of the power series are determined up to terms of order t^8 for the square and s.c. isotropic Heisenberg and XY magnets of spin $\frac{1}{2}$ at infinite temperature, and up to terms of t^{10} for the linear magnets. The expressions obtained by truncating the power series to the exactly known terms give a very good fit at short times to the exact expression for the linear XY model, and to the results of the computer simulation due to Windsor for the s.c. Heisenberg ferromagnet. The Fourier-time transforms of the expressions thus obtained are shown to be the Gram–Charlier expansions. Thus we conclude that the Gram–Charlier expansion of the Fourier-time transform of $\sigma(R,t)$, $\tilde{\sigma}(R,\omega)$, and that of its Fourier-space transform $I(k,t)$, $S(k,\omega)$, is especially useful in the cases when the short-time behavior of $\sigma(R,t)$ or $I(k,t)$ is considered important; generally speaking, this is the case at large k or at ω which is not very small. When the width of the Gaussian distribution function is determined by the second derivative of $I(k,t)$, a convergent result is not obtained for small values of k. For larger k, the convergence is as good as for the case when the width is determined by the autocorrelation function.

1. INTRODUCTION

The two-time spin-pair correlation function is the quantity of primary importance in the theory of neutron scattering from magnetic materials.[1,2] In a number of papers,[3-8] calculations are focused on the function at infinite temperature, for simplicity of the treatment. For example, Windsor[4] performed a computer simulation calculation for the Heisenberg magnet of classical spins at infinite temperature. Carboni and Richards[5] gave an exact calculation, for a finite chain, of the Heisenberg magnet of spin $\frac{1}{2}$ at infinite temperature.

In the preceding paper,[9] the author gave the numerical values of the coefficients of the expansion in powers of time of the two-time spin-pair correlation function $\langle s_i^z(t)s_f^z(0)\rangle\rangle$ of spin $\frac{1}{2}$ at infinite temperature. In the present paper, we express $\langle s_i^z(t)s_f^z(0)\rangle\rangle$ and its Fourier-space transform $\langle S_k^z(t)S_{-k}^z(0)\rangle\rangle$ as a product of a Gaussian distribution function and a power series.

It is noted that the Fourier-time transform of this expression gives the Gram–Charlier expansion, which was proposed by Collins and Marshall[7] for the analysis of the Fourier-time transform of $\langle S_k^z(t)S_{-k}^z(0)\rangle\rangle$. The convergence of those expansions is discussed for the Heisenberg magnet and the XY magnet. For the one-dimensional XY magnet, the results are compared with the exact solution.

In the remaining part of this introduction, definitions adopted in this paper are given. The Hamiltonian of the system is

$$H = -\sum_f \sum_g [J_\perp(f,g)s_f^- s_g^+ + J_\parallel(f,g)s_f^z s_g^z], \quad (1.1)$$

where $J_\perp(f,g)$ and $J_\parallel(f,g)$ are equal to $J_\perp$ and $J_\parallel$, respectively, when f and g are nearest neighbors of each other and zero otherwise. For the Heisenberg magnet, $J_\perp = J_\parallel \equiv J$. If $J_\parallel = 0$ and the system is one-dimensional, the system is the XY model. We shall call the case of $J_\parallel = 0$ the XY magnet in general. The two-time spin-pair correlation function $\sigma(R_{if},t)$ is defined by

$$\sigma(R_{if},t) = \langle s_i^z(t)s_f^z(0)\rangle\rangle - \langle s_i^z\rangle\langle s_f^z\rangle, \quad (1.2)$$

where

$$s_i^z(t) = e^{iHt}s_i^z e^{-iHt}.$$

We shall introduce the Fourier-space transform of $s_i^z(t)$ by

$$S_k^z(t) = \sum_f s_f^z(t)\, e^{ik\cdot R_f}. \quad (1.3)$$

The Fourier-space transform of $\sigma(R_{if},t)$ is the so-called "intermediate scattering function." It will be denoted by $I(k,t)$:

$$I(k,t) = \sum_f \sigma(R_{if},t)\, e^{ik\cdot R_{if}}. \quad (1.4)$$

It is the correlation function of $S_k^z(t)$:

$$I(k,t) = N^{-1}\langle S_k^z(t)S_{-k}^z(0)\rangle\rangle, \qquad k \neq 0, \quad (1.5)$$

where N is the total number of spins in the system. We shall focus our attention mainly on $\sigma(R_{if},t)$ in Secs. 2 and 3, and on $I(k,t)$ in Sec. 4.

2. SHORT-TIME EXPANSION OF $\sigma(R_{if}, t)$

The short-time expansion of the two-time spin-pair correlation function $\sigma(R_{if}, t)$ has been given in the following form:

$$\sigma(R_{if}, t) = \sigma_\alpha^{(0)}(R_{if}) + \sum_{n=1}^{\infty} \frac{(-1)^n}{(2n)!}\, \sigma_\alpha^{(2n)}(R_{if})(Jt)^{2n}, \tag{2.1}$$

where $\sigma_\alpha^{(2n)}(R_{if})$ is $\sigma_t^{(2n)}(R_{if})$ or $\sigma_0^{(2n)}(R_{if})$ according as the system is the isotropic Heisenberg magnet or the XY magnet. For the case of the XY magnet, J must be replaced by $J_\perp$. The numerical values of $\sigma_t^{(2n)}(R_{if})$ and $\sigma_0^{(2n)}(R_{if})$ are listed in Tables IV–VI of Ref. 9.

When $R_{if} = 0$, $\sigma(0, t)$ is equal to $\sigma_\alpha^{(0)}(0)$ at $t = 0$ and decays with finite second derivative $-\sigma_\alpha^{(2)}(0) J^2$. It is natural to approximate this behavior by the Gaussian distribution function

$$\sigma(0, t) \approx \sigma_\alpha^{(0)}(0) e^{-\tau^2/2}, \tag{2.2}$$

where

$$\tau^2 = (Jt)^2\, \sigma_\alpha^{(2)}(0)/\sigma_\alpha^{(0)}(0). \tag{2.3}$$

For spin $\frac{1}{2}$ at infinite temperature, $\sigma_\alpha^{(0)}(0) = \frac{1}{4}$ and $\sigma_\alpha^{(2)}(0) = \frac{1}{2} z$, and hence

$$\tau^2 = 2z(Jt)^2, \tag{2.4}$$

where z is the coordination number of the lattice.

Assuming this general decay of $\sigma(R_{if}, t)$, we express $\sigma(R_{if}, t)$ as a product of $\exp(-\tau^2/2)$ and a power series in t:

$$\sigma(R_{if}, t) = \tfrac{1}{4} e^{-\tau^2/2}\left(\eta_\alpha^{(0)}(R_{if}) + \sum_{n=1}^{\infty} \frac{\eta_\alpha^{(2n)}(R_{if})}{(2n)!\,(2z)^n}\, \tau^{2n}\right). \tag{2.5}$$

The factor $\frac{1}{4}$ on the right-hand side is for $\sigma_\alpha^{(0)}(0)$. By comparing (2.1) and (2.5), we can determine the exact coefficients $\eta_\alpha^{(2n)}(R_{if})$ for $0 \le n \le N$ when $\sigma_\alpha^{(2n)}(R_{if})$ are known for $0 \le n \le N$. The coefficients $\eta_\alpha^{(2n)}(R_{if})$ for the Heisenberg magnet ($\alpha = t$) and the XY magnet ($\alpha = 0$), respectively, are given in Table I for $0 \le 2n \le 10$ for the linear chain. The corresponding tables for $0 \le 2n \le 8$ for the square and simple cubic lattices are Tables II and III.

The $\sigma(R_{if}, l)$ are calculated as a function of t by truncating the sum on the right-hand side of (2.5) at the 2nd, 3rd, $\cdots$ terms. When one needs m steps along the axis to go from the ith site to the fth site, the first nonzero coefficient $\eta_\alpha^{(2n)}(R_{if})$ occurs for $n = m$. The approximation where n coefficients $\eta_\alpha^{(2m)}(R_{if})$ through $\eta_\alpha^{(2m+2n-2)}(R_{if})$ are used and the higher terms are ignored in the sum of (2.5) will be called the "n-term approximation." Typical curves are shown in Figs. 1–5. For the origin, the Gaussian approximation (2.2) is the two-term approximation. For convenience of comparison with previous work, $4\sigma(R_{if}, t)$ is plotted as a function of $\tau = (2z)^{1/2} Jt$. The curves are labeled as n term or simply as n ($n = 1, 2, 3, 4$, or 5) for the n-term approximations.

When we are interested in neutron diffraction, the quantity of interest is the space-time Fourier trans-

TABLE I. Coefficients $\eta_\alpha^{(2n)}(R_{if})$ for the Linear Chain

I.a. Isotropic Heisenberg Magnet of Spin $\frac{1}{2}$.

R_{if}	$\eta_t^{(0)}$	$\eta_t^{(2)}$	$\eta_t^{(4)}$	$\eta_t^{(6)}$	$\eta_t^{(8)}$	$\eta_t^{(10)}$
(0)	1	0	− 4	68	5812	279184
(1)	0	2	20	216	2288	24784
(2)	0	0	6	210	6664	219120
(3)	0	0	0	20	1512	90852
(4)	0	0	0	0	70	9240
(5)	0	0	0	0	0	252

I.b. XY Model.

R_{if}	$\eta_0^{(0)}$	$\eta_0^{(2)}$	$\eta_0^{(4)}$	$\eta_0^{(6)}$	$\eta_0^{(8)}$	$\eta_0^{(10)}$
(0)	1	0	− 12	− 160	420	173376
(1)	0	2	24	300	2800	− 47880
(2)	0	0	6	240	8680	322560
(3)	0	0	0	20	1680	112140
(4)	0	0	0	0	70	10080
(5)	0	0	0	0	0	252

TABLE II. Coefficients $\eta_\alpha^{(2n)}(R_{if})$ for the Square Lattice

II.a. Isotropic Heisenberg Magnet.

R_{if}	$\eta_t^{(0)}$	$\eta_t^{(2)}$	$\eta_t^{(4)}$	$\eta_t^{(6)}$	$\eta_t^{(8)}$
(0, 0)	1	0	8	440	44120
(1, 0)	0	2	28	680	22080
(2, 0)	0	0	0	310	17192
(3, 0)	0	0	0	20	2296
(4, 0)	0	0	0	0	70
(1, 1)	0	0	12	680	39424
(2, 1)	0	0	0	60	7224
(3, 1)	0	0	0	0	280
(2, 2)	0	0	0	0	420

II.b. XY Magnet.

R_{if}	$\eta_0^{(0)}$	$\eta_0^{(2)}$	$\eta_0^{(4)}$	$\eta_0^{(6)}$	$\eta_0^{(8)}$
(0, 0)	1	0	− 40	− 1472	− 46536
(1, 0)	0	2	40	948	21168
(2, 0)	0	0	6	400	25928
(3, 0)	0	0	0	20	2800
(4, 0)	0	0	0	0	70
(1, 1)	0	0	12	800	51408
(2, 1)	0	0	0	60	8400
(3, 1)	0	0	0	0	280
(2, 2)	0	0	0	0	420

TABLE III. Coefficients $\eta_\alpha^{(2n)}(R_{if})$ for the s.c. Lattice

III.a. Isotropic Heisenberg Magnet.

R_{if}	$\eta_t^{(0)}$	$\eta_t^{(2)}$	$\eta_t^{(4)}$	$\eta_t^{(6)}$	$\eta_t^{(8)}$
(0, 0, 0)	1	0	36	444	206892
(1, 0, 0)	0	2	36	1656	63952
(2, 0, 0)	0	0	6	410	36232
(3, 0, 0)	0	0	0	20	3080
(4, 0, 0)	0	0	0	0	70
(1, 1, 0)	0	0	12	880	77728
(2, 1, 0)	0	0	0	60	9576
(3, 1, 0)	0	0	0	0	280
(2, 2, 0)	0	0	0	0	420
(1, 1, 1)	0	0	0	120	19488
(2, 1, 1)	0	0	0	0	840

III.b. XY Magnet.

R_{if}	$\eta_0^{(0)}$	$\eta_0^{(2)}$	$\eta_0^{(4)}$	$\eta_0^{(6)}$	$\eta_0^{(8)}$
(0, 0, 0)	1	0	− 84	− 5472	− 324996
(1, 0, 0)	0	2	56	2012	62832
(2, 0, 0)	0	0	6	560	53480
(3, 0, 0)	0	0	0	20	3920
(4, 0, 0)	0	0	0	0	70
(1, 1, 0)	0	0	12	1120	106512
(2, 1, 0)	0	0	0	60	11760
(3, 1, 0)	0	0	0	0	280
(2, 2, 0)	0	0	0	0	420
(1, 1, 1)	0	0	0	120	23520
(2, 1, 1)	0	0	0	0	840

form of $\sigma(R_{if}, t)$. The Fourier transform with respect to time is taken with the aid of the following formula[11]:

$$\int_0^\infty \tau^{2n} e^{-\tau^2/2a^2} \cos u\tau \, d\tau = (\tfrac{1}{2}\pi)^{1/2}(-1)^n a^{2n+1} e^{-a^2u^2/2} He_{2n}(au), \quad (2.6)$$

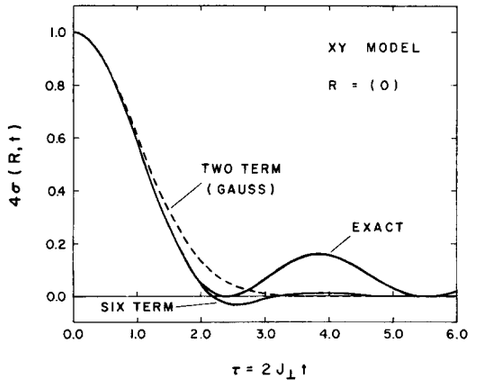


FIG. 1. Time dependence of the autocorrelation function for the one-dimensional XY model. The two-term approximation is the Gaussian distribution function. The six-term approximation is given by Eq. (2.5) and Table I.b. The exact curve is due to Katsura *et al.*[10]

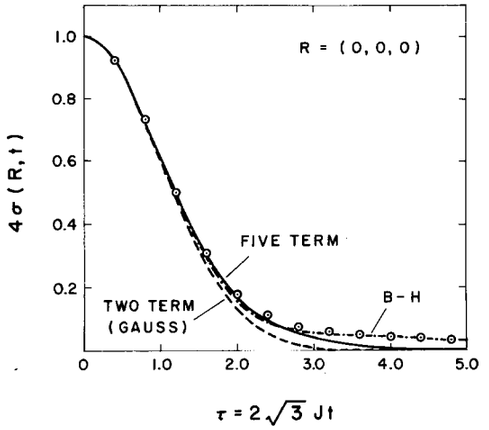


FIG. 2. Time dependence of the autocorrelation function for the s.c. Heisenberg magnet. The two-term approximation is the Gaussian distribution function. The five-term approximation is given by Eq. (2.5) and Table III.a. Circles with center dot are due to Windsor.[4] Dashed curve (B–H) is the result of Blume and Hubbard.[6]

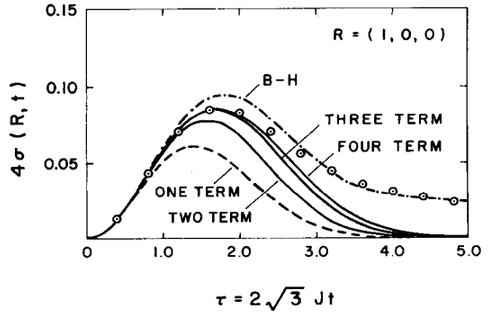


FIG. 3. Time dependence of the correlation function of nearest neighbor spins for the s.c. Heisenberg magnet. The one- through four-term approximations are given. The circles with a center dot are due to Windsor and the dashed curve (B–H) to Blume and Hubbard.

where $He_n(x)$ is the Hermite polynomial defined by

$$He_n(x) = n! \sum_{m=0}^{[n/2]} \frac{(-1)^m x^{n-2m}}{m!\, 2^m (n-2m)!} \quad (2.7)$$

Parameter a is included in (2.6) for later use.

We define the Fourier-time transform of $\sigma(R_{if}, t)$ by

$$\tilde{\sigma}(R_{if}, \omega) = \int_0^\infty \sigma(R_{if}, t) \cos\omega t \, dt. \quad (2.8)$$

Then we have

$$\tilde{\sigma}(R_{if}, \omega) = \left(\frac{\pi}{2}\right)^{1/2} \frac{1}{4(2z)^{1/2}J} e^{-u^2/2}\left(\eta_\alpha^{(0)}(R_{if}) + \sum_{n=1}^{\infty} \frac{(-1)^n \eta_\alpha^{(2n)}(R_{if})}{(2n)!\,(2z)^n} He_{2n}(u)\right), \quad (2.9)$$

where

$$u^2 = \omega^2/2zJ^2. \quad (2.10)$$

The curves of $\tilde{\sigma}(R_{if}, \omega)$ are shown in Figs. 6–8.

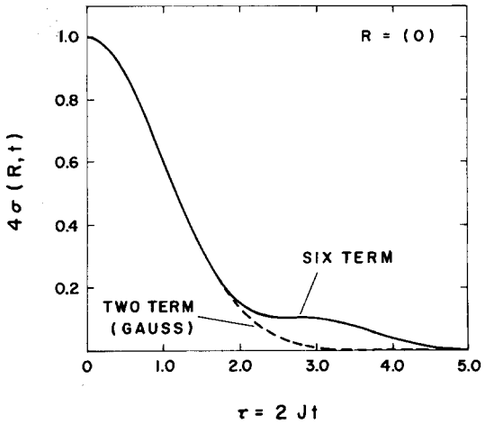


FIG. 4. Time dependence of the autocorrelation function for the linear Heisenberg magnet. The two- and six-term approximations are shown.

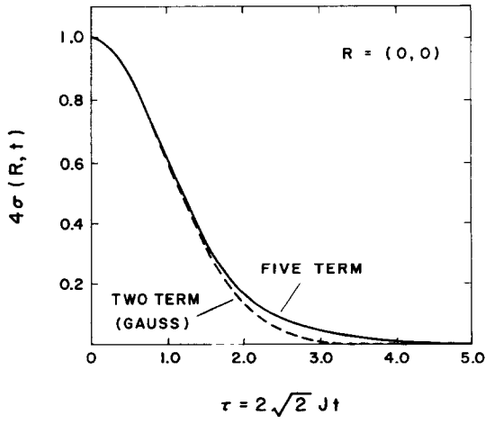


FIG. 5. Time dependence of the autocorrelation function for the square Heisenberg magnet. The two- and five-term approximations are shown.

The inverse relation to (2.8) is

$$\sigma(R_{if}, t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \tilde{\sigma}(R_{if}, \omega) \cos\omega t \, d\omega. \qquad (2.11)$$

Expanding both sides in powers of t, we confirm that the moments of ω calculated by (2.9) are the same as the ones given by the coefficients of (2.1). Hence the expansion (2.9) represents the Gram–Charlier expansion introduced by Collins and Marshall[7] for the neutron scattering problem. Thus Tables I–III give the coefficients for the Gram–Charlier expansion.

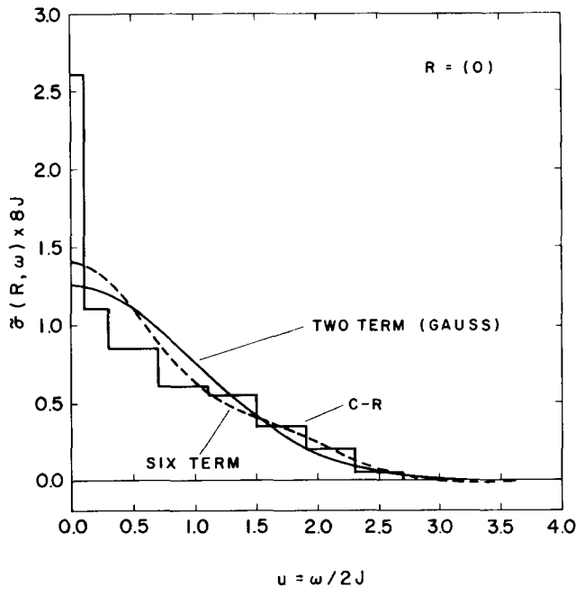


FIG. 6. Fourier-time transform of the autocorrelation function for the linear Heisenberg magnet; cf. Fig. 4. The six-term approximation and Carboni and Richards' result (C–R) are compared.

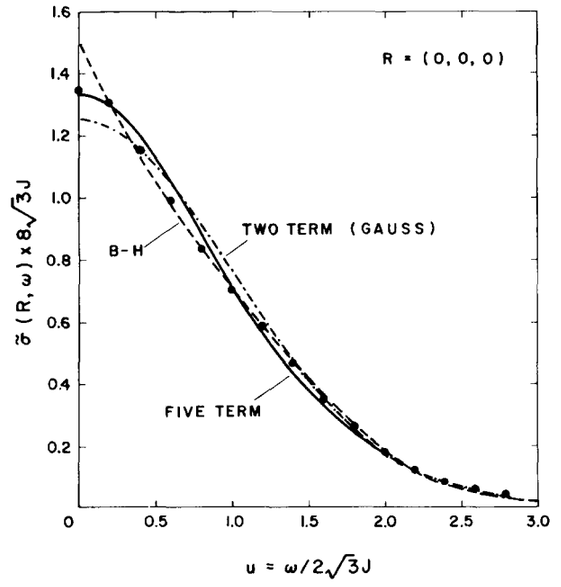


FIG. 7. Fourier-time transform of the auto-correlation function for the s.c. Heisenberg magnet. The five-term approximation is compared with the two-term (Gaussian) approximation, Windsor's result (black circles), and Blume and Hubbard's result (dashed curve); cf. Fig. 2.

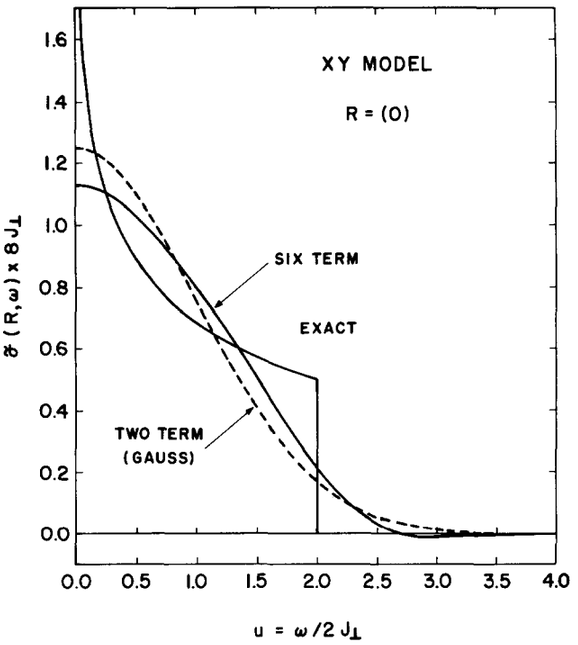


FIG. 8. Fourier-time transform of the autocorrelation function of the one-dimensional XY model. The six-term approximation is compared with the two-term (Gaussian) approximation and the exact curve; cf. Fig. 1.

3. DISCUSSION OF THE RESULTS FOR $\sigma(R_{if}, l)$ and $\tilde{\sigma}(R_{if}, \omega)$

Figure 1 shows the autocorrelation function for the one-dimensional XY model. The six-term approximation represents the curve which Eq. (2.5) gives when $\eta_\alpha^{(2n)}(R_{if})$, given for $R_{if} = 0$ in Table I.b, are used up to $2n = 10$ and higher terms are ignored. It is compared with the exact result obtained by Katsura *et al.*[10] We notice that the six-term approximation is a great improvement over the two-term approximation (Gaussian distribution function) for $0 \leq \tau \lesssim 2.0$. But the approximation is not good for $\tau > 2.0$.

Figures 2 and 3 show the autocorrelation function and the correlation function of spins on the nearest neighbors, for the s.c. lattice. The curves are obtained by Eq. (2.5) with Table IIIa. The curves in the five-term and four-term approximations, respectively, show very good agreement with Windsor's results (scaled properly) for $0 \leq \tau \lesssim 2.0$. The agreement with Blume and Hubbard's[6] results is not so complete in this range. For $2.0 < \tau$, the present result deviates from Windsor's result. This deviation is due to the inaccuracy of the present result. This situation is clear in Fig. 3, where the curves obtained in the two- and three-term approximations are also given. The convergence is good for $0 \leq \tau \lesssim 2.0$ at the four-term approximation. But, the values for $2 < \tau$ are still changing towards the curve due to Windsor as the higher terms are included.

From these observations, we conclude that Eq. (2.5) with Tables I–III gives a very good estimate of $\sigma(R_{if}, t)$ for $0 \leq \tau \lesssim 2.0$, if more than two nonzero values of $\eta_\alpha^{(2n)}(R_{if})$ are available. All these figures show that the curves decay to zero too fast at $\tau > 2.0$.

Figure 4 shows the time dependence of the autocorrelation function of the one-dimensional Heisenberg magnet. In the six-term approximation, we have a tail extending up to $\tau \sim 4.0$. But in Windsor's result, we see that the tail is extended to still larger values of τ. This situation can be seen more clearly in Fig. 6, where the Fourier-time transform is compared. At $\omega = 0$, $\tilde{\sigma}(R, \omega)$ in the six-term approximation is a little higher than that of the Gaussian curve. This is due to the tail around $\tau \sim 3$ of Fig. 4. But the result of Carboni and Richards[5] suggests a divergence of $\tilde{\sigma}(R_{if}, \omega)$ at $\omega = 0$, which is connected with a very long tail in $\sigma(R_{if}, t)$. Except for the divergence of $\tilde{\sigma}(R, \omega)$ at $\omega = 0$, the curve obtained by the six-term approximation fits well to the Carboni and Richards result.

Figure 5 shows the autocorrelation function for the square lattice. Compared with Windsor's result, the tail is again too short-ranged.

Figure 7 shows the Fourier-time transform of the curves given in Fig. 2 for the autocorrelation function of the Heisenberg magnet for the s.c. lattice. The five-term approximation is slightly nearer to Windsor's curve, compared with the two-term approximation.

Figure 8 shows the Fourier-time transforms of the curves shown in Fig. 1 of the autocorrelation function for the one-dimensional XY model. The good fit of the six-term approximation at short time in Fig. 1 is now indicated only by a tendency toward improving from the two-term approximation to the exact one. At the end of the last section, we noticed that the present approximations used in evaluating $\sigma(R_{if}, t)$ gives the Gram–Charlier expansion for $\tilde{\sigma}(R_{if}, \omega)$. The above analysis shows that improvements in $\tilde{\sigma}(R_{if}, \omega)$ are attained by including more terms if the values of $\sigma(R_{if}, t)$ for $0 < \tau \lesssim 2.0$ play a main role in its calculation for the given value of ω. Thus we cannot expect very good results for $\tilde{\sigma}(R_{if}, \omega)$ at small ω which is most affected by the tail of $\sigma(R_{if}, t)$.

4. $I(k, t)$ and $S(k, \omega)$

In this section, we investigate the Fourier-space transforms of the quantities discussed in the preceding section. On the one hand, we simply take the Fourier-space transforms of the results (2.5) and (2.9) obtained in the preceding section:

$$I(k,t) = \tfrac{1}{4} e^{-\tau^2/2} \left(\zeta_\alpha^{(0)}(k) + \sum_{n=1}^{\infty} \frac{\zeta_\alpha^{(2n)}(k)}{(2n)!\,(2z)^n} \tau^{2n} \right), \quad (4.1)$$

and

$$S(k,\omega) = \left(\frac{\pi}{2}\right)^{1/2} \frac{1}{4(2z)^{1/2} J} e^{-u^2/2} \left(\zeta_\alpha^{(0)}(k) + \sum_{n=1}^{\infty} \frac{(-1)^n \zeta_\alpha^{(2n)}(k)}{(2n)!\,(2z)^n} He_{2n}(u) \right), \quad (4.2)$$

where

$$\zeta_\alpha^{(2n)}(k) = \sum_f \eta_\alpha^{(2n)}(R_{if})\, e^{ik\cdot R_{if}}. \quad (4.3)$$

Examples of curves of $I(k, t)$ and $S(k, \omega)$ obtained by (4.1) and (4.2) are shown by dashed lines in Figs. 9–12 and 13–15, respectively. Fig. 9 shows the comparison of $I(k, t)$ for the one-dimensional XY model with the exact result.[10] Agreement is good at $0 < \tau \lesssim 2.0$. $I(k, t)$ obtained by truncating the expansion (4.1) decays too fast and cannot reproduce the oscillatory behavior of the exact $J_0(2\tau)$. The Fourier-time transform $S(k, \omega)$ is compared in Fig. 13. We notice now that the approximate expression cannot reproduce the singular behavior of $S(k, \omega)$. The curves for $I(k, t)$ obtained by (4.1) are considered to be very good for $0 < \tau \lesssim 2.0$.

In another approach, we substitute (2.1) into (1.4) and obtain the short time expansion of $I(k, t)$ as

$$I(k,t) = I_\alpha^{(0)}(k) + \sum_{n=1}^{\infty} \frac{(-1)^n I_\alpha^{(2n)}(k)}{(2n)!\,(2z)^n} \tau^{2n}, \quad (4.4)$$

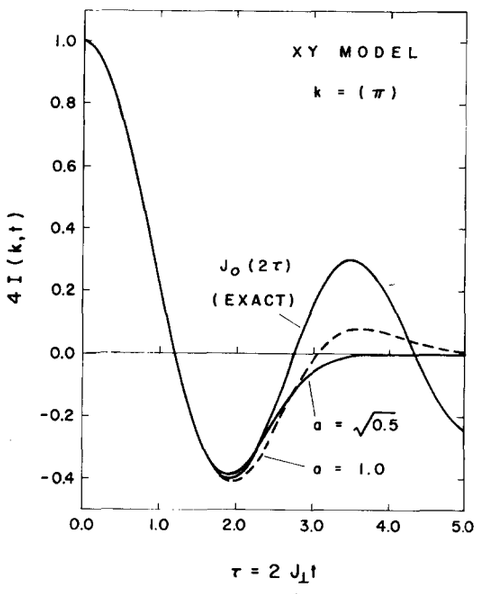


FIG. 9. Time dependence of $I(k, t)$ for the one-dimensional XY model, where $k = (\pi)$. The six-term approximations are compared with the exact result.

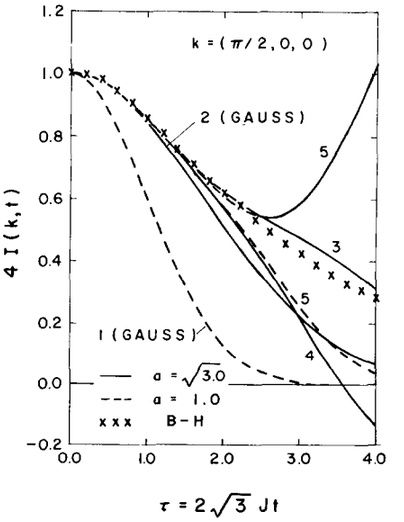


FIG. 10. Time dependence of $I(k, t)$ for the s.c. Heisenberg magnet, where $k = (\pi/2, 0, 0)$. The solid lines show the one- through five-term approximations, the Gaussian distribution function is determined by the coefficients of $O(t^0)$ and $O(t^2)$. The expansion does not converge. The dashed lines are for the case where a is determined by the expansion of the autocorrelation function. They are compared with Blume and Hubbard's curved (B–H).

$$I_\alpha^{(2n)}(k) = \sum_f \sigma_\alpha^{(2n)}(R_{if})\, e^{ik\cdot R_{if}}. \tag{4.5}$$

In analogy with the case of $\sigma(R_{if}, t)$ for $R_{if} = 0$ in Sec. 2, we introduce the Gaussian distribution function by using the first two terms of (4.4):

$$I_\alpha^{(0)}(k)\, e^{-\tau^2/2a^2}, \tag{4.6}$$

where

$$a^2 = 2zI_\alpha^{(0)}(k)/I_\alpha^{(2)}(k). \tag{4.7}$$

$I(k,t)$ is expressed as a product of (4.6) and a power series of τ^2:

$$I(k,t) = I_\alpha^{(0)}(k)\, e^{-\tau^2/2a^2}\left(1 + \sum_{n=1}^{\infty} \frac{\xi^{(2n)}(k)}{(2n)!\,(2z)^n}\,\tau^{2n}\right). \tag{4.8}$$

With the aid of formula (2.7), its Fourier-time transform is

$$S(k,\omega) = I_\alpha^{(0)}(k)\left(\frac{\pi}{2}\right)^{1/2} \frac{a e^{-a^2u^2/2}}{(2z)^{1/2}J} \times \left(1 + \sum_{n=1}^{\infty} (-1)^n \frac{\xi^{(2n)}(k)\, a^{2n}}{(2n)!\,(2z)^n}\, He_{2n}(au)\right). \tag{4.9}$$

This is the Gram–Charlier expansion of $S(k,\omega)$ proposed by Collins and Marshall.[7]

Curves obtained for $I(k,t)$ and $S(k,\omega)$ by (4.8) and (4.9) are shown by solid lines in Figs. 9–12 and 13–15, respectively. For small values of k, parameter a^2 is larger than unity and expansion (4.8) for $I(k,t)$ does not converge; such an example is shown in Fig. 10. This feature is stronger for larger dimen-

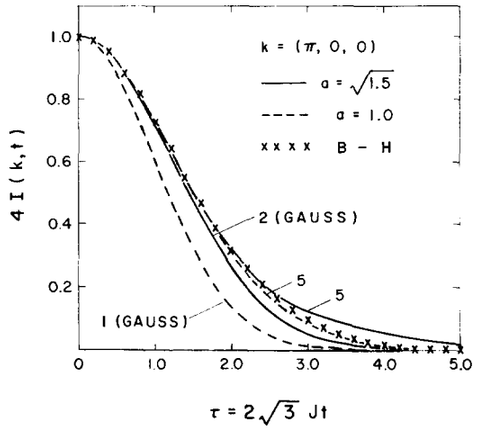


FIG. 11. Time dependence of $I(k,t)$ for the s.c. Heisenberg magnet, where $k = (\pi, 0, 0)$. Convergence is good for both values of $a^2 = 1.5$ and $a^2 = 1.0$. The curves are compared with Blume and Hubbard's result (B–H).

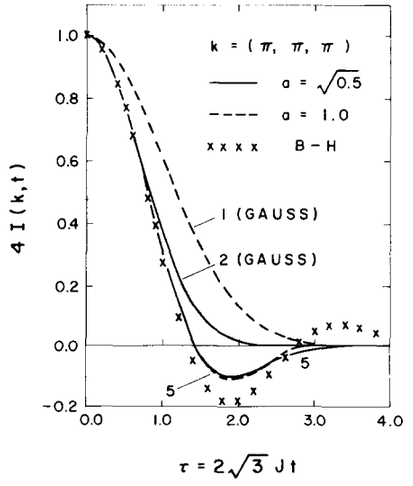


FIG. 12. Time dependence of $I(k,t)$ for the s.c. Heisenberg magnet, where $k = (\pi, \pi, \pi)$.

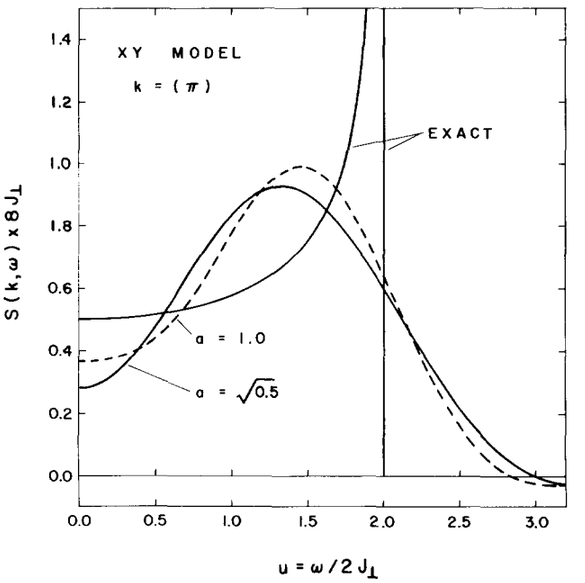


FIG. 13. Frequency-dependence of $S(k,\omega)$ for the one-dimensional XY model, where $k = (\pi)$. The six-term approximations are compared with the exact result.

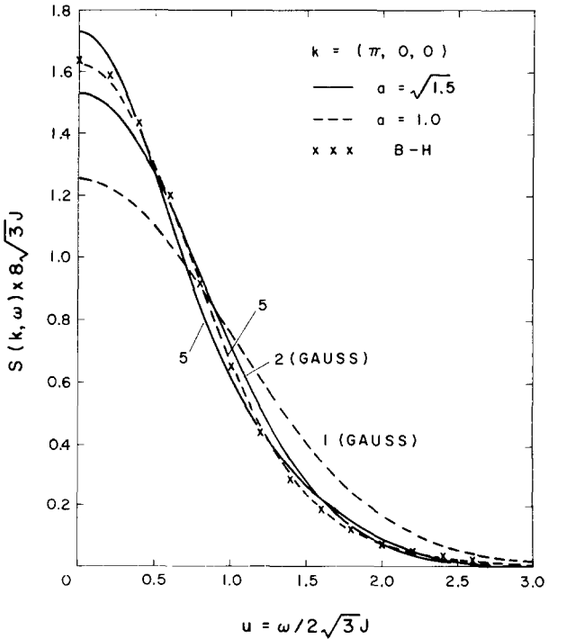


FIG. 14. Frequency-dependence of $S(k,\omega)$ for the s.c. Heisenberg magnet, where $k = (\pi, 0, 0)$; cf. Fig. 11.

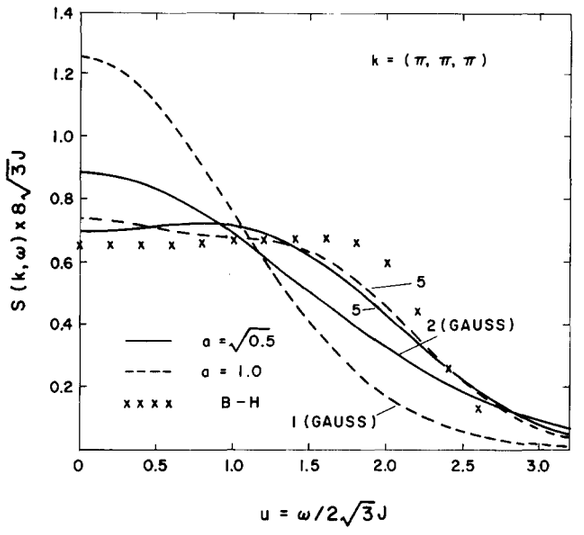


FIG. 15. Frequency-dependence of $S(k,\omega)$ for the s.c. Heisenberg magnet, where $k = (\pi,\pi,\pi)$; cf. Fig. 12.

sion. For large values of k, a^2 is smaller than unity and $I(k,t)$ and $S(k,\omega)$ converge fairly well. We notice that the curves obtained by (4.1) and (4.8) agree well with each other at $0 \leq \tau \lesssim 2.0$. From a comparison of Fig. 10 with Figs. 11 and 12, we see that $I(k,t)$ for small k decays slower than for larger k. This means that the present result gives a better fit for larger k. Discrepancy of the curves in the five-term approximation from Blume and Hubbard's curve near $\tau \sim 2.0$ in Fig. 12 will be due to a similar discrepancy of $\sigma(R_{if},t)$ for nearest neighbors shown in Fig. 3.

5. CONCLUSION

The short-time expansions of the two-time spin-pair correlation function $\sigma(R_{if},t)$ and its Fourier transform $I(k,t)$ are expressed as a product of a Gaussian distribution function and a power series. When the Gaussian distribution function is chosen to fit the first two terms of the expansion of the autocorrelation function, the results for $\sigma(R_{if},t)$ and $I(k,t)$ are found very good for $0 < \tau \lesssim 2.0$. At $\tau > 2.0$, the quantities damps too fast. In case when the values at $\tau < 2.0$ of $\sigma(R_{if},t)$ or $I(k,t)$ play an important role, these expansions must be useful. Generally speaking, this is the case for $I(k,t)$ and $S(k,\omega)$ at large k and for $\tilde{\sigma}(R_{if},\omega)$ at ω which is not very small.

In the Gram–Charlier expansion proposed by Collins and Marshall,[7] the Gaussian distribution function is determined by the first two terms of the expansion of $I(k,t)$. It is found that it gives a divergent result for small values of k.

ACKNOWLEDGMENTS

The author is grateful to Dr. T. Horiguchi for valuable discussions. The numerical computations were made with the aid of computer IBM 360/44 of Computer Center of Ohio University.

* On leave of absence from Department of Applied Science, Faculty of Engineering, Tohoku University, Sendai, Japan.

1 L. Van Hove, Phys. Rev. **95**, 1374 (1954).
2 P. G. de Gennes, in *Magnetism*, edited by G. T. Rado and H. Suhl (Academic, New York, 1963), Vol. III, p. 115.
3 P. G. de Gennes, J. Phys. Chem. Solids **4**, 223 (1967).
4 C. G. Windsor, in *Neutron Inelastic Scattering* (IAEA, Vienna, 1968), Vol. II, p. 83.
5 F. Carboni and P. M. Richards, Phys. Rev. **177**, 889 (1969).
6 M. Blume and J. Hubbard, Phys. Rev. B **1**, 3815 (1970).
7 M. F. Collins and W. Marshall, Proc. Phys. Soc. **92**, 390 (1967).
8 R. A. Tahir-Kheli and D. G. McFadden, Phys. Rev. **182**, 604 (1969).
9 T. Morita, J. Math. Phys. **12**, 2062 (1971).
10 S. Katsura, T. Horiguchi, and M. Suzuki, Physica **46**, 69 (1970).
11 W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, New York, 1966), p. 254.

Eigenvalue Problem for Lagrangian Systems. VI

E. M. Barston
Courant Institute of Mathematical Sciences, New York University, New York, New York 10012
(Received 22 November 1971)

The gyroscopic Lagrangian system $\ddot{\xi} + A\dot{\xi} + H\xi(t) = 0$ on the finite-dimensional complex Hilbert space E is shown to be stable if and only if there exist Hermitian operators $i\alpha$ and $P > 0$ such that $A = P\alpha + \alpha P$ and $H = P(\frac{1}{4} + \alpha^2)P$. Structural stability is shown to be equivalent to the uniqueness of P and α, and to the existence of a Liapunov operator on $E \times E$ of the form $Lp(T)$, where $L = \begin{pmatrix} iA & i \\ i & 0 \end{pmatrix}$, $T = \begin{pmatrix} 0 & iI \\ iA & i \end{pmatrix}$, and $p(x)$ is a real polynomial of degree not exceeding the least of the quantities $2 \dim E - 1$ and $4N+1$, where N denotes the number of negative eigenvalues of H.

I. INTRODUCTION

This paper continues the discussion presented earlier[1–5] of the linear gyroscopic Lagrangian system

$$\ddot{\xi} + A\dot{\xi} + H\xi(t) = 0, \qquad t \geq 0. \tag{1}$$

Here the operators H and iA are assumed to be time-independent linear Hermitian operators on and into the complex n-dimensional Hilbert space E, with $\xi(t) \in E$ for each $t \geq 0$. We obtain some necessary and sufficient conditions for stability (Sec. II); in particular, it is shown that the system (1) is stable if and only if there exist Hermitian operators $i\alpha$ and $P > 0$ such that $A = P\alpha + \alpha P$ and $H = P(\frac{1}{4} + \alpha^2)P$. This result provides a simple means of generating stable systems. In Sec. III we consider the problem of the construction of Liapunov functionals. It was shown in Ref. 1 that for any real polynomial $p(x)$, the operator $Lp(T)$ is Hermitian and $\langle\zeta, Lp(T)\zeta\rangle$ is a constant of the motion of the equivalent system $\dot{\zeta} = W\zeta$ in $E \times E$, where

$$W = \begin{pmatrix} 0 & I \\ -H & -A \end{pmatrix}, \quad L = \begin{pmatrix} -iA & -i \\ i & 0 \end{pmatrix}, \quad \text{and} \quad T = -iW.$$

Thus if $Lp(T) > 0$, $\langle\zeta, Lp(T)\zeta\rangle$ is a Liapunov functional

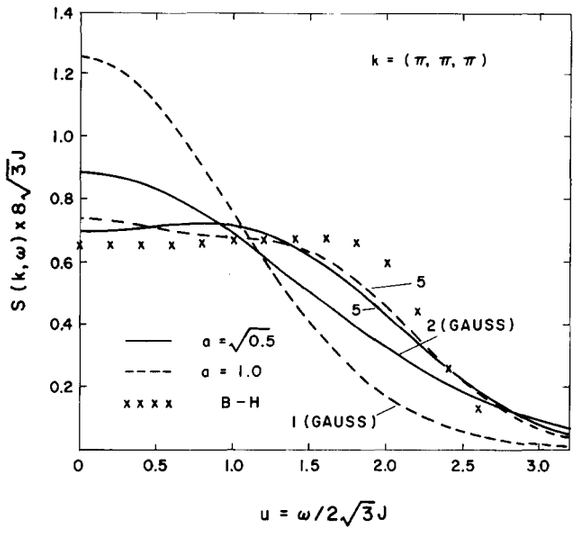


FIG. 15. Frequency-dependence of $S(k,\omega)$ for the s.c. Heisenberg magnet, where $k = (\pi,\pi,\pi)$; cf. Fig. 12.

sion. For large values of k, a^2 is smaller than unity and $I(k,t)$ and $S(k,\omega)$ converge fairly well. We notice that the curves obtained by (4.1) and (4.8) agree well with each other at $0 \leq \tau \lesssim 2.0$. From a comparison of Fig. 10 with Figs. 11 and 12, we see that $I(k,t)$ for small k decays slower than for larger k. This means that the present result gives a better fit for larger k. Discrepancy of the curves in the five-term approximation from Blume and Hubbard's curve near $\tau \sim 2.0$ in Fig. 12 will be due to a similar discrepancy of $\sigma(R_{if},t)$ for nearest neighbors shown in Fig. 3.

5. CONCLUSION

The short-time expansions of the two-time spin-pair correlation function $\sigma(R_{if},t)$ and its Fourier transform $I(k,t)$ are expressed as a product of a Gaussian distribution function and a power series. When the Gaussian distribution function is chosen to fit the first two terms of the expansion of the autocorrelation function, the results for $\sigma(R_{if},t)$ and $I(k,t)$ are found very good for $0 < \tau \lesssim 2.0$. At $\tau > 2.0$, the quantities damps too fast. In case when the values at $\tau < 2.0$ of $\sigma(R_{if},t)$ or $I(k,t)$ play an important role, these expansions must be useful. Generally speaking, this is the case for $I(k,t)$ and $S(k,\omega)$ at large k and for $\tilde{\sigma}(R_{if},\omega)$ at ω which is not very small.

In the Gram–Charlier expansion proposed by Collins and Marshall,[7] the Gaussian distribution function is determined by the first two terms of the expansion of $I(k,t)$. It is found that it gives a divergent result for small values of k.

ACKNOWLEDGMENTS

The author is grateful to Dr. T. Horiguchi for valuable discussions. The numerical computations were made with the aid of computer IBM 360/44 of Computer Center of Ohio University.

* On leave of absence from Department of Applied Science, Faculty of Engineering, Tohoku University, Sendai, Japan.

1 L. Van Hove, Phys. Rev. **95**, 1374 (1954).
2 P. G. de Gennes, in *Magnetism*, edited by G. T. Rado and H. Suhl (Academic, New York, 1963), Vol. III, p. 115.
3 P. G. de Gennes, J. Phys. Chem. Solids **4**, 223 (1967).
4 C. G. Windsor, in *Neutron Inelastic Scattering* (IAEA, Vienna, 1968), Vol. II, p. 83.
5 F. Carboni and P. M. Richards, Phys. Rev. **177**, 889 (1969).
6 M. Blume and J. Hubbard, Phys. Rev. B **1**, 3815 (1970).
7 M. F. Collins and W. Marshall, Proc. Phys. Soc. **92**, 390 (1967).
8 R. A. Tahir-Kheli and D. G. McFadden, Phys. Rev. **182**, 604 (1969).
9 T. Morita, J. Math. Phys. **12**, 2062 (1971).
10 S. Katsura, T. Horiguchi, and M. Suzuki, Physica **46**, 69 (1970).
11 W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, New York, 1966), p. 254.

Eigenvalue Problem for Lagrangian Systems. VI

E. M. Barston
Courant Institute of Mathematical Sciences, New York University, New York, New York 10012
(Received 22 November 1971)

The gyroscopic Lagrangian system $\ddot{\xi} + A\dot{\xi} + H\xi(t) = 0$ on the finite-dimensional complex Hilbert space E is shown to be stable if and only if there exist Hermitian operators $i\alpha$ and $P > 0$ such that $A = P\alpha + \alpha P$ and $H = P(\frac{1}{4} + \alpha^2)P$. Structural stability is shown to be equivalent to the uniqueness of P and α, and to the existence of a Liapunov operator on $E \times E$ of the form $Lp(T)$, where $L = \begin{pmatrix} iA & i \\ i & 0 \end{pmatrix}$, $T = \begin{pmatrix} 0 & iI \\ iA & i \end{pmatrix}$, and $p(x)$ is a real polynomial of degree not exceeding the least of the quantities $2 \dim E - 1$ and $4N+1$, where N denotes the number of negative eigenvalues of H.

I. INTRODUCTION

This paper continues the discussion presented earlier[1–5] of the linear gyroscopic Lagrangian system

$$\ddot{\xi} + A\dot{\xi} + H\xi(t) = 0, \qquad t \geq 0. \tag{1}$$

Here the operators H and iA are assumed to be time-independent linear Hermitian operators on and into the complex n-dimensional Hilbert space E, with $\xi(t) \in E$ for each $t \geq 0$. We obtain some necessary and sufficient conditions for stability (Sec. II); in particular, it is shown that the system (1) is stable if and only if there exist Hermitian operators $i\alpha$ and $P > 0$ such that $A = P\alpha + \alpha P$ and $H = P(\frac{1}{4} + \alpha^2)P$. This result provides a simple means of generating stable systems. In Sec. III we consider the problem of the construction of Liapunov functionals. It was shown in Ref. 1 that for any real polynomial $p(x)$, the operator $Lp(T)$ is Hermitian and $\langle\zeta, Lp(T)\zeta\rangle$ is a constant of the motion of the equivalent system $\dot{\zeta} = W\zeta$ in $E \times E$, where

$$W = \begin{pmatrix} 0 & I \\ -H & -A \end{pmatrix}, \quad L = \begin{pmatrix} -iA & -i \\ i & 0 \end{pmatrix}, \quad \text{and} \quad T = -iW.$$

Thus if $Lp(T) > 0$, $\langle\zeta, Lp(T)\zeta\rangle$ is a Liapunov functional

for the system $\dot{\zeta} = W\zeta$. Furthermore, $Lp(T)$ then symmetrizes T and hence reduces (1) to self-adjoint form. Thus the construction of such operators (called Liapunov operators) is of some interest. We show that an operator $Lp(T) > 0$ exists if and only if the system (1) is structurally stable (i.e., if and only if the system is stable and remains stable for all sufficiently small Hermitian perturbations of H and anti-Hermitian perturbations of A), and that if an operator $Lp(T) > 0$ exists at all, then one exists for a polynomial $p(x)$ of degree not exceeding the least of the numbers $2\dim E - 1$ and $4N + 1$, where N denotes the number of negative eigenvalues of H counted according to their multiplicity.

In the sequel, we make extensive use of the results and notation of Ref. 1; in particular, the latter is used freely without further explanation. The reader should consult Ref. 1 for definitions.

II. SOME NECESSARY AND SUFFICIENT CONDITIONS FOR STABILITY

Theorem 1: The system (1) is stable if and only if there exists a linear operator D_1 from E into E such that

$$H = (D_1 - iA)D_1 \tag{2}$$

and

$$P \equiv D_1 + D_1^\dagger - iA > 0. \tag{3}$$

The existence of D_1 implies the existence of a complete L-canonical set of eigenvectors

$$\zeta_k = \begin{pmatrix} \xi_k \\ i\omega_k\xi_k \end{pmatrix}, \qquad k = 1, 2, \ldots, 2n,$$

such that Eqs. (28)–(36) of Ref. 1 hold.

Proof: Suppose the system (1) is stable. The existence of an operator D_1 satisfying Eqs. (2) and (3) is an immediate consequence of Theorems 2(C), 5(B), and 7 of Ref. 1.

Conversely, assume that Eqs. (2) and (3) hold. Then

$$PD_1 = (D_1 - iA)D_1 + D_1^\dagger D_1 = H + D_1^\dagger D_1. \tag{4}$$

Since $P > 0$ and $H + D_1^\dagger D_1$ is Hermitian, there exist n linearly independent vectors ξ_k and real numbers $\omega_k, k = 1, 2, \ldots, n$, satisfying the equations

$$\omega_k P\xi_k = (H + D_1^\dagger D_1)\xi_k, \qquad k = 1, 2, \ldots, n, \tag{5}$$

$$(\xi_k, P\xi_l) = \delta_{kl}, \qquad k = 1, 2, \ldots, n. \tag{6}$$

Thus by Eq. (4), $\omega_k P\xi_k = PD_1\xi_k, k = 1, 2, \ldots, n$, so that $D_1\xi_k = \omega_k\xi_k, k = 1, \ldots, n$. Now

$$H_\omega \equiv \omega^2 - \omega iA - H = \omega^2 - \omega iA - (D_1 - iA)D_1 = (\omega + D_1 - iA)(\omega - D_1), \tag{7}$$

so that $H_{\omega_k}\xi_k = 0, k = 1, \ldots, n$. Let $D_2 \equiv iA - D_1^\dagger$. Then $P = D_1 - D_2 = -(D_2 + D_2^\dagger - iA)$, so that $PD_2 = -(D_2 - iA)D_2 - D_2^\dagger D_2 = -H - D_2^\dagger D_2$. Since $P > 0$ and $H + D_2^\dagger D_2$ is Hermitian, there exist n linearly independent vectors ξ_k and n real numbers $\omega_k, k = n + 1, \ldots, 2n$, satisfying the equations

$$\omega_k P\xi_k = -(H + D_2^\dagger D_2)\xi_k, \qquad k = n + 1, \ldots, 2n, \tag{8}$$

$$(\xi_k, P\xi_l) = \delta_{kl}, \qquad k, l = n + 1, \ldots, 2n. \tag{9}$$

Then $PD_2\xi_k = \omega_k P\xi_k, k = n + 1, \ldots, 2n$, so that $D_2\xi_k = \omega_k\xi_k, k = n + 1, \ldots, 2n$. For real ω, the adjoint of Eq. (7) gives $H_\omega = (\omega - D_1^\dagger)(\omega + D_1^\dagger - iA) = (\omega - D_1^\dagger) \times (\omega - D_2)$, so that $H_{\omega_k}\xi_k = 0, k = n + 1, \ldots, 2n$, Therefore the ξ_k are, for all $k = 1, \ldots, 2n$, eigenvectors of Eq. (1) with eigenvalues ω_k, and therefore the vectors

$$\zeta_k \equiv \begin{pmatrix} \xi_k \\ i\omega_k\xi_k \end{pmatrix}, \quad k = 1, \ldots, 2n,$$

are eigenvectors of the operator T with corresponding real eigenvalues ω_k. Now

$$\langle\zeta_k, L\zeta_l\rangle = (\omega_k + \omega_l)(\xi_k, \xi_l) - (\xi_k, iA\xi_l) = (\omega_k\xi_k, \xi_l) + (\xi_k, \omega_l\xi_l) - (\xi_k, iA\xi_l),$$

so that

$$\langle\zeta_k, L\zeta_l\rangle = (\xi_k, [D_1^\dagger + D_1 - iA]\xi_l) = (\xi_k, P\xi_l) = \delta_{kl}, \qquad 1 \le k, l \le n,$$

$$\langle\zeta_k, L\zeta_l\rangle = (\xi_k, [D_2^\dagger + D_2 - iA]\xi_l) = -(\xi_k, P\xi_l) = -\delta_{kl}, \qquad n + 1 \le k, l \le 2n,$$

$$\langle\zeta_k, L\zeta_l\rangle = (\xi_k, [D_1^\dagger + D_2 - iA]\xi_l) = 0, \qquad 1 \le k \le n, \quad n + 1 \le l \le 2n,$$

and therefore

$$\langle\zeta_k, L\zeta_l\rangle = \gamma_k\delta_{kl},$$

where

$$\gamma_k = \begin{cases} 1, & k = 1, \ldots, n, \\ -1, & k = n + 1, \ldots, 2n. \end{cases} \tag{10}$$

This completes the verification of Eqs. (28)–(36) of Ref. 1. It follows at once from Eq. (10) that the vectors $\zeta_k, k = 1, \ldots, 2n$, form a basis for E^2 and thus constitute a complete L-canonical set of eigenvectors with real eigenvalues ω_k; the stability of the system (1) now follows from Theorem 5 of Ref. 1.

Corollary 1: The system (1) is stable if and only if there exists a positive definite Hermitian operator P and an anti-Hermitian operator a such that

$$H = [\tfrac{1}{2}(P - iA) + a][\tfrac{1}{2}(P + iA) + a]. \tag{11}$$

If Eq. (11) holds for some $P > 0$ and an anti-Hermitian operator a, then a is the unique solution of the equation $Pa + aP = \frac{1}{2}i(AP - PA)$.

Proof: Assume that the system is stable. Then by Theorem 1, there exists D_1 satisfying Eqs. (2) and (3). Let $a \equiv D_1 - \frac{1}{2}(P + iA)$, where $P \equiv D_1 + D_1^\dagger - iA > 0$. Then $a^\dagger = D_1^\dagger - \frac{1}{2}(P + iA)$ and $a + a^\dagger = D_1 + D_1^\dagger - (P + iA) = 0$, so that a is anti-Hermitian. Therefore $D_1 = \frac{1}{2}(P + iA) + a$ and $H = (D_1 - iA)D_1 = [\frac{1}{2}(P - iA) + a][\frac{1}{2}(P + iA) + a]$. Conversely, suppose there exists $P > 0$ and an anti-Hermitian operator a such that Eq. (11) holds. Define $D_1 \equiv \frac{1}{2}(P + iA) + a$. Then by Eq. (11), $H = (D_1 - iA)D_1$ and $D_1 + D_1^\dagger - iA = \frac{1}{2}(P + iA) + a + \frac{1}{2}(P + iA) - a - iA = P > 0$; thus Theorem 1 implies that the system is stable. Since H is Hermitian, we infer from Eq. (11) that $[\frac{1}{2}(P - iA)$

$+ a][\frac{1}{2}(P + iA) + a] = [\frac{1}{2}(P + iA) - a][\frac{1}{2}(P - iA) - a]$, which holds if and only if $Pa + aP = \frac{1}{2}i(AP - PA)$. Since $P > 0$, this equation has a unique solution for a.

Corollary 2: The system (1) is stable if and only if there exists a positive-definite Hermitian operator P and an anti-Hermitian operator α such that

$$A = P\alpha + \alpha P \tag{12}$$

and

$$H = P(\tfrac{1}{4} + \alpha^2)P. \tag{13}$$

Proof: Suppose the system is stable. Then by Corollary 1, Eq. (11) holds for $P > 0$ and a anti-Hermitian, with a the unique solution of $Pa + aP = \frac{1}{2}i(AP - PA)$. We define α to be the unique (since $P > 0$) solution of the equation $P\alpha + \alpha P = A$, which implies that α is anti-Hermitian. Now $P[\frac{1}{2}i(\alpha P - P\alpha)] + [\frac{1}{2}i(\alpha P - P\alpha)]P = \frac{1}{2}i(AP - PA)$, so that $a = \frac{1}{2}i(\alpha P - P\alpha)$. The substitution of $P\alpha + \alpha P$ for A and $\frac{1}{2}i(\alpha P - P\alpha)$ for a into Eq. (11) yields Eq. (13). Conversely, let Eqs. (12) and (13) hold with $P > 0$. Equation (12) implies that α is anti-Hermitian, so that $a \equiv \frac{1}{2}i(\alpha P - P\alpha)$ is anti-Hermitian. We have

$$\begin{aligned}[\tfrac{1}{2}(P - iA) + a][\tfrac{1}{2}(P + iA) + a] &= \tfrac{1}{4}(P - iP\alpha - i\alpha P + i\alpha P - iP\alpha) \\ &\quad \times (P + iP\alpha + i\alpha P + i\alpha P - iP\alpha) \\ &= P(\tfrac{1}{4} + \alpha^2)P = H,\end{aligned}$$

and the system is stable by Corollary 1.

Given a stable system (1) and operators α (anti-Hermitian) and P (> 0) satisfying Eqs. (12) and (13), it is not difficult to construct from α and P a positive-definite operator $\Pi \in V$ (a positive $\Pi \in V$ must exist by Theorem 5 of Ref. 1). Indeed,

$$\begin{aligned}\Pi &\equiv \begin{pmatrix} \frac{1}{4}P - P\alpha P^{-1}\alpha P & -P\alpha P^{-1} \\ P^{-1}\alpha P & P^{-1} \end{pmatrix} \\ &= \begin{pmatrix} P^{1/2} & 0 \\ 0 & P^{-1/2} \end{pmatrix} \begin{pmatrix} \frac{1}{4} + \beta\beta^\dagger & -\beta \\ -\beta^\dagger & 1 \end{pmatrix} \begin{pmatrix} P^{1/2} & 0 \\ 0 & P^{-1/2} \end{pmatrix},\end{aligned} \tag{14}$$

where $\beta \equiv P^{1/2}\alpha P^{-1/2}$, is clearly positive definite, and

$$\Pi W = \begin{pmatrix} P(\frac{1}{4}\alpha + \alpha^3)P & P(\frac{1}{4} + \alpha^2) \\ -(\frac{1}{4} + \alpha^2)P & -\alpha \end{pmatrix}$$

is anti-Hermitian, so that $\Pi \in V$ by Theorem 2(B) of Ref. 1. Note that if r is any invertible operator from E^2 into E^2 that commutes with W, then $r^\dagger \Pi r$ is also a positive-definite operator in V. This follows from Theorem 2(B) of Ref. 1, since $r^\dagger \Pi r$ is Hermitian and $\tilde{a} \equiv \Pi W$ is anti-Hermitian, so that $r^\dagger \Pi r W = r^\dagger \Pi W r = r^\dagger \tilde{a} r$ is anti-Hermitian.

The following theorem gives necessary and sufficient conditions guaranteeing the uniqueness of the operators D_1 of Eqs. (2) and (3) and P and α of Eqs. (12) and (13).

Theorem 2: Suppose the system (1) is stable. Then the operator D_1 satisfying Eqs. (2) and (3) and the operators P (> 0) and α (anti-Hermitian) satisfying Eqs. (12) and (13) are uniquely determined if and only if, for each eigenspace S_ω, $\langle \zeta, L\zeta \rangle$ is of one sign for all nonzero $\zeta \in S_\omega$.

Proof: Since the system is stable, all the eigenvalues ω are real and $\sum_\omega \dim S_\omega = 2n$ (Theorem 5 of Ref. 1). Choose a basis for each eigenspace S_ω according to Theorem 3(E) of Ref. 1, with $G = L$; the union of all these basis sets, enumerated as $\{\zeta_1, \ldots, \zeta_{2n}\}$, forms a complete L-canonical set of eigenvectors satisfying

$$\langle \zeta_k, L\zeta_l \rangle = \gamma_k \delta_{kl}, \tag{15}$$

where $\gamma_k = \pm 1$, $k = 1, \ldots, 2n$. By Theorem 7 of Ref. 1, we may assume, without loss of generality, that

$$\gamma_k = \begin{cases} 1, & k = 1, \ldots, n, \\ -1, & k = n+1, \ldots, 2n. \end{cases} \tag{16}$$

Each

$$\zeta_k = \begin{pmatrix} \xi_k \\ i\omega_k \xi_k \end{pmatrix},$$

where $\xi_k \in E$ and ω_k is the eigenvalue corresponding to the eigenvector ζ_k. Suppose that for some eigenvalue Ω, $\langle \zeta, L\zeta \rangle$ is not of one sign on S_Ω. The nonempty set $\{\zeta_k \mid \zeta_k \in S_\Omega, 1 \le k \le 2n\}$ is a basis for S_Ω, so that if the γ_k in the set $\{\gamma_k \mid \zeta_k \in S_\Omega\}$ were either all positive or all negative, then by Eq. (15), $\langle \zeta, L\zeta \rangle$ would be strictly positive or negative, respectively, for all nonzero $\zeta \in S_\Omega$. Therefore $\{\gamma_k \mid \zeta_k \in S_\Omega\}$ contains γ_k's of both signs, and $m \equiv \dim S_\Omega \ge 2$. Without loss of generality, we may assume that $\zeta_k \in S_\Omega$ for $1 \le k \le r$ and $n + 1 \le k \le n + m - r$, where $1 \le r < m$, i.e., $\{\zeta_k \mid 1 \le k \le r, n + 1 \le k \le n + m - r\}$ is the basis of eigenvectors for S_Ω. Then the set $\{\xi_k \mid \zeta_k \in S_\Omega\}$ is linearly independent. Let z be any complex number. We define

$$\tilde{\zeta}_k(z) \equiv \begin{cases} (|z|^2 + 1)^{1/2}\zeta_1 + z\zeta_{n+1}, & k = 1, \\ \bar{z}\zeta_1 + (|z|^2 + 1)^{1/2}\zeta_{n+1}, & k = n + 1, \\ \zeta_k, & 2 \le k \le n, n + 2 \le k \le 2n. \end{cases}$$

Then $\langle \tilde{\zeta}_k(z), L\tilde{\zeta}_l(z) \rangle = \gamma_k \delta_{kl}$, $k, l = 1, \ldots, 2n$, where the γ_k are given by Eq. (16), and the $\zeta_k(z)$, $k = 1, \ldots, 2n$, form a complete L-canonical set of eigenvectors. Let $\tilde{\xi}_k(z)$ denote the first component of $\tilde{\zeta}_k(z)$, so that

$$\tilde{\zeta}_k(z) = \begin{pmatrix} \tilde{\xi}_k(z) \\ i\omega_k \tilde{\xi}_k(z) \end{pmatrix}, \qquad k = 1, \ldots, 2n.$$

By Theorem 6 of Ref. 1, $\{\tilde{\xi}_1(z), \ldots, \tilde{\xi}_n(z)\}$ is a basis for E. For each complex number z, we define the operator $D_1(z)$ from E into E by $D_1(z)\xi = \sum_1^n \alpha_k \omega_k \tilde{\xi}_k(z)$ where $\xi = \sum_1^n \alpha_k \tilde{\xi}_k(z)$. Since the set $\{\xi_k \mid 1 \le k \le r, k = n + 1\}$ is linearly independent, it follows that $D_1(z_1) \ne D_1(z_2)$ if $z_1 \ne z_2$. But $D_1(z)$ satisfies Eqs. (2) and (3) for every complex z, as demonstrated in the proof of Theorem 7 of Ref. 1; thus we conclude that if $\langle \zeta, L\zeta \rangle$ is not of one sign on each S_ω, then there are infinitely many operators D_1 satisfying Eqs. (2) and (3). Suppose now that for each S_ω, $\langle \zeta, L\zeta \rangle$ is of one sign for all nonzero $\zeta \in S_\omega$. Let $\Omega_1, \Omega_2, \ldots, \Omega_q$ denote the distinct elements of the set of eigenvalues $\{\omega \mid \langle \zeta, L\zeta \rangle > 0 \text{ on } S_\omega \text{ for } \zeta \ne 0\}$. By Theorem 7 of Ref. 1, every complete L-canonical set of eigenvectors has n positive γ_k and n negative γ_k; hence

$\sum_1^q \dim S_{\Omega_k} = n$, so that E is the direct sum of the subspaces

$$S_{\Omega_k} \equiv \{\xi \mid \begin{pmatrix} \xi \\ i\Omega_k \xi \end{pmatrix} \in S_{\Omega_k}\}, \qquad k = 1, \ldots, q.$$

Then for every $\eta \in E$, η admits the unique representation $\eta = \sum_1^q \beta_k \eta_k$, where $\eta_k \in S_{\Omega_k}$, $k = 1, \ldots, q$, and we define the operator D from E into E by $D\eta \equiv \sum_1^q \beta_k \Omega_k \eta_k$. Let D_1 satisfy Eqs. (2) and (3). Then by the construction of Theorem 1, there exist n vectors $\xi_k \in E$ satisfying Eqs. (5) and (6); $\{\xi_k\}_1^n$ is a basis for E; $D_1 \xi_k = \omega_k \xi_k$, $k = 1, \ldots, n$; and

$$\zeta_k \equiv \begin{pmatrix} \xi_k \\ i\omega_k \xi_k \end{pmatrix}, \qquad k = 1, \ldots, n,$$

are eigenvectors of T with eigenvalues ω_k satisfying $\langle \zeta_k, L\zeta_l \rangle = \delta_{kl}$, $k, l = 1, \ldots, n$. Then for each k, $1 \le k \le n$, $\omega_k = \Omega_l$ for some l, $1 \le l \le q$, and $\xi_k \in S_{\Omega_l}$; hence $D\xi_k = \omega_k \xi_k$, $k = 1, \ldots, n$. Since $\{\xi_k\}_1^n$ is a basis for E, we conclude that $D_1 = D$, which proves that D_1 is unique. We have now demonstrated the theorem for the operator D_1. We complete the proof by showing that D_1 is unique if and only if P and α are unique. Suppose then that D_1 is unique, and let $A = P_l \alpha_l + \alpha_l P_l$, $H = P_l(\frac{1}{4} + \alpha_l^2)P_l$, $l = 1, 2$, where $P_l > 0$ and α_l is anti-Hermitian. Define $D_{1l} \equiv \frac{1}{2}(P_l + iA) + a_l$, $a_l \equiv \frac{1}{2} i(\alpha_l P_l - P_l \alpha_l)$, $l = 1, 2$. Then D_{1l} satisfies $H = (D_{1l} - iA)D_{1l}$ and $P_l = D_{1l} + D_{1l}^\dagger - iA > 0$ for $l = 1, 2$, so that $D_{11} = D_{12}$. Therefore $P_1 + iA = D_{11} + D_{11}^\dagger = D_{12} + D_{12}^\dagger = P_2 + iA$, so that $P_1 = P_2$ and hence $\alpha_1 = \alpha_2$. Finally, suppose that P and α are unique. Let D_{11} and D_{12} satisfy Eqs. (2) and (3), i.e.,

$$H = (D_{1l} - iA)D_{1l}, \qquad l = 1, 2, \tag{2'}$$

$$D_{1l} + D_{1l}^\dagger - iA > 0, \qquad l = 1, 2. \tag{3'}$$

Then as shown in the proofs of Corollaries 1 and 2, $P_l \equiv D_{1l} + D_{1l}^\dagger - iA$ and α_l, defined as the unique solution of $P_l \alpha_l + \alpha_l P_l = A$, satisfy $A = P_l \alpha_l + \alpha_l P_l$ and $H = P_l(\frac{1}{4} + \alpha_l^2)P_l$. By the uniqueness of P, we must have $P_1 = P_2 = P$, i.e., $D_{1l} + D_{1l}^\dagger - iA = P$ for $l = 1, 2$. Hence $D_{1l} = \frac{1}{2}(P + iA) + a_l$, $l = 1, 2$, where a_l is anti-Hermitian. Thus by Eq. (2'), $H = [\frac{1}{2}(P - iA) + a_l][\frac{1}{2}(P + iA) + a_l]$, $l = 1, 2$, and we conclude from Corollary 1 that a_l must be the unique solution a of the equation $Pa + aP = \frac{1}{2} i(AP - PA)$, i.e., $a_1 = a_2$, and therefore $D_{11} = D_{12}$. This completes the proof of the theorem.

Corollary 2 provides a simple characterization of the class of stable systems, and is particularly useful for synthesizing operator pairs A and H for which the system (1) will be stable—one merely chooses a positive definite operator P and an anti-Hermitian operator α and computes A and H from Eqs. (12) and (13). Given a particular operator A, the entire class of operators H for which the system (1) will be stable can be constructed by letting P range through the set of positive operators; for each such P the unique solution α of Eq. (12) is computed, and then the corresponding H (now a function of P) is obtained from Eq. (13). The problem of determining an operator A, given the operator H, so that (1) will be stable, is also readily solved: The operator $P > 0$ is chosen so that $P^{-1}HP^{-1} - \frac{1}{4} \le 0$ (i.e., so that $H - \frac{1}{4}P^2$ is negative definite or negative semi-definite) and α is taken to by any one of the anti-Hermitian square roots of $P^{-1}HP^{-1} - \frac{1}{4}$; A is then computed from Eq. (12). Observe that if the operators P (> 0) and α (anti-Hermitian) satisfying Eqs. (12) and (13) are known, then $D_1 = (\frac{1}{2} + i\alpha)P$ satisfies Eqs. (2) and (3), so that by Eqs. (5), (6), (8), and (9), the $2n$ eigenvectors ξ_k of Eq. (1) with eigenvalues ω_k satisfy the self-adjoint (with weight operator P) eigenvalue problems

$$\omega_k P\xi_k = P(i\alpha + \tfrac{1}{2})P\xi_k, \qquad k = 1, \ldots, n,$$

$$\omega_k P\xi_k = P(i\alpha - \tfrac{1}{2})P\xi_k, \qquad k = n+1, \ldots, 2n,$$

where $(\xi_k, P\xi_l) = \delta_{kl}$ for $k, l = 1, \ldots, n$ and $k, l = n+1, \ldots, 2n$, so that all the usual variational principles apply.

If H and A commute, then it is easy to see that the system (1) is stable if and only if $4H - A^2 > 0$; in this case, we can take $P = (4H - A^2)^{1/2}$ and $\alpha = \frac{1}{2}P^{-1}A$, and P, α, H, and A all commute. Indeed, if H and A commute and $4H - A^2 > 0$, set $P \equiv (4H - A^2)^{1/2} > 0$ and $\alpha \equiv \frac{1}{2}P^{-1}A$. Then P commutes with H, A, and α, and it is readily verified that Eqs. (12) and (13) hold, so that the system is stable. Conversely, suppose that H and A commute and the system is stable. Then there exists an orthonormal basis of simultaneous eigenvectors $\{\eta_k\}_1^n$ of A and H such that $H\eta_k = \lambda_k \eta_k$, $iA\eta_k = \mu_k \eta_k$, where λ_k and μ_k are real, $k = 1, \ldots, n$. Then $H_\omega \eta_k = 0$ provided $\omega = \frac{1}{2}[\mu_k \pm (\mu_k^2 + 4\lambda_k)^{1/2}]$. Clearly we must have $\mu_k^2 + 4\lambda_k \ge 0$, $k = 1, \ldots, n$, or the system would be exponentially unstable. In fact we must have $\mu_k^2 + 4\lambda_k > 0$ for every $k = 1, \ldots, n$, for if $\mu_k^2 + 4\lambda_k = 0$ for some k, then $\xi(t) \equiv t\eta_k \exp(\frac{1}{2} i\mu_k t)$ satisfies Eq. (1) and the system would be unstable. Thus $\mu_k^2 + 4\lambda_k > 0$ for all $k = 1, 2, \ldots, n$, which implies that $4H - A^2 > 0$.

III. STRUCTURAL STABILITY AND THE CONSTRUCTION OF LIAPUNOV OPERATORS

It was shown in Theorem 5 of Ref. 1 that the system (1) is stable if and only if there exists a positive-definite operator $\Pi \in V$. We call such an operator a Liapunov operator. Given the real polynomial $p(x)$, the operator $Lp(T)$ is in V by Theorem 2 of Ref. 1; we now proceed to determine the subclass of stable systems possessing Liapunov operators of the form $Lp(T)$. Suppose that the system (1) is stable. Then there exists a complete L-canonical set of eigenvectors $\{\zeta_k\}_1^{2n}$ with real eigenvalues ω_k satisfying

$$\omega_k \zeta_k = T\zeta_k, \qquad k = 1, \ldots, 2n \tag{17}$$

and Eqs. (15) and (16). Let $\zeta \in E^2$. Then $\zeta = \sum_1^{2n} \beta_k \zeta_k$, so that

$$\begin{aligned} \langle \zeta, Lp(T)\zeta \rangle &= \sum_{k,l=1}^{2n} \bar{\beta}_k \beta_l \langle \zeta_k, Lp(T)\zeta_l \rangle \\ &= \sum_{k,l=1}^{2n} \bar{\beta}_k \beta_l p(\omega_l) \langle \zeta_k, L\zeta_l \rangle \\ &= \sum_1^{2n} |\beta_k|^2 p(\omega_k)\gamma_k. \end{aligned} \tag{18}$$

Clearly $Lp(T)$ can be positive if and only if

$$p(\omega_k)\gamma_k > 0, \quad k = 1, \ldots, 2n. \tag{19}$$

Suppose $\omega_k = \omega_l$ for $k \neq l$. Then if Eq. (19) holds, $p(\omega_k)\gamma_k > 0$ and $p(\omega_k)\gamma_l = p(\omega_l)\gamma_l > 0$, so that γ_k and γ_l must be of the same sign; thus for each eigenspace S_ω, $\langle \eta, L\eta\rangle$ must be of one sign for all nonzero $\eta \in S_\omega$. Conversely, suppose that $\langle \eta, L\eta\rangle$ ($\eta \neq 0$) is of one sign on each S_ω, and let $\Omega_1, \ldots, \Omega_m$ denote the distinct eigenvalues of T. Then $m \leq 2n$, and we can construct a real polynomial $p(x)$ of degree not exceeding $m - 1$ such that for each $k = 1, \ldots, m$,

$$p(\Omega_k) = \begin{cases} \Delta_k > 0 & \text{if } \langle \eta, L\eta\rangle > 0 \text{ on } S_{\Omega_k},\ \eta \neq 0, \\ \Delta_k < 0 & \text{if } \langle \eta, L\eta\rangle < 0 \text{ on } S_{\Omega_k},\ \eta \neq 0. \end{cases} \tag{20}$$

Then Eq. (19) holds, and $Lp(T)$ is a Liapunov operator. Hence a Liapunov operator of the form $Lp(T)$ with $p(x)$ a real polynomial of degree $\leq 2n - 1$ exists if and only if the system (1) is stable and $\langle \eta, L\eta\rangle$ is of one sign on each S_ω ($\eta \neq 0$).

The existence of a Liapunov operator of the form $Lp(T)$ implies that the system (1) remains stable for all sufficently small perturbations of the operators A and H which keep these operators anti-Hermitian and Hermitian, respectively. A system having this property is said to be structurally stable; more precisely, we say that the system (1) is structurally stable if and only if there exists $\epsilon > 0$ such that for all Hermitian operators $i\tilde{A}$ and $\tilde{H}$ satisfying $\|\tilde{A}\| \leq \epsilon$ and $\|\tilde{H}\| \leq \epsilon$, the system

$$\ddot{\xi} + (A + \tilde{A})\dot{\xi} + (H + \tilde{H})\xi(t) = 0 \tag{21}$$

is stable. The structural stability of a system possessing a Liapunov operator of the form $Lp(T)$ follows from the fact that L and T are continuous in A and H (indeed, if

$$T' = \begin{pmatrix} 0 & -i \\ i(H + \tilde{H}) & i(A + \tilde{A}) \end{pmatrix} \text{ and } L' = \begin{pmatrix} -i(A + \tilde{A}) & -i \\ i & 0 \end{pmatrix},$$

then $\|T' - T\| \leq \sqrt{2}\epsilon$ and $\|L' - L\| \leq \epsilon$ provided $\|\tilde{A}\| \leq \epsilon$ and $\|H\| \leq \epsilon$) so that $Lp(T)$ is continuous in A and H. The application of Theorem 2 of Ref. 1 to the system (21) shows that $L'p(T')$ is Hermitian and will be a Liapunov operator for the system (21) if $L'p(T') > 0$; but if $Lp(T) > 0$, the continuity of $Lp(T)$ in A and H and the fact that $L'p(T')$ is Hermitian implies the existence of $\epsilon > 0$ such that $L'p(T') > 0$ provided $\|\tilde{A}\| \leq \epsilon$ and $\|\tilde{H}\| \leq \epsilon$.

Suppose, on the other hand, that the system (1) does not admit a Liapunov operator of the form $Lp(T)$. Then either the system is unstable or the system is stable but $\langle \zeta, L\zeta\rangle$ is not of one sign on each eigenspace S_ω ($\zeta \neq 0$). Suppose the latter is true; then for some eigenvalue ω (real) there exist eigenvectors $\zeta_1, \zeta_2 \in S_\omega$ such that $\langle \zeta_k, L\zeta_l\rangle = \gamma_k \delta_{kl}$, $k, l = 1, 2$, with $\gamma_1 = 1$ and $\gamma_2 = -1$. Let $\zeta \equiv \zeta_1 + \zeta_2$. Then $\zeta \in S_\omega$, $\zeta \neq 0$, so that

$$\zeta = \begin{pmatrix} \eta \\ i\omega\eta \end{pmatrix},$$

with $\eta \neq 0, H_\omega \eta = 0$, and

$$2\omega(\eta, \eta) - (\eta, iA\eta) = \langle \zeta, L\zeta\rangle = \gamma_1 + \gamma_2 = 0. \tag{22}$$

Let h be any Hermitian operator from E into E satisfying $h\eta = -(2i\omega + A)\eta$; the necessary and sufficient condition for the existence of such an h is that $(\eta, [2i\omega + A]\eta) = 0$, which is just Eq. (22). Then it is readily verified that for every $\epsilon > 0$, $\xi(t) \equiv \eta \exp(i\omega + \epsilon)t$ satisfies the equation

$$\ddot{\xi} + A\dot{\xi} + (H + \epsilon h - \epsilon^2 I)\,\xi(t) = 0,$$

so that the system (1) is not structurally stable.

We summarize these results together with the results of Theorem 2 in the next theorem.

Theorem 3: The following statements are equivalent:

(A) The system (1) is structurally stable.

(B) A Liapunov operator of the form $Lp(T)$ exists where $p(x)$ is a real polynomial of degree not exceeding $2n - 1$.

(C) There exist unique operators P (> 0) and α (anti-Hermitian) satisfying Eqs. (12) and (13).

(D) There exists a unique operator D_1 satisfying Eqs. (2) and (3).

(E) The system (1) is stable, and for each eigenspace S_ω, $\langle \zeta, L\zeta\rangle$ is of one sign for all nonzero $\zeta \in S_\omega$.

Now that we have seen that structural stability is equivalent to the existence of Liapunov operators of the form $Lp(T)$, we consider the problem of constructing such operators. Although it is not necessary to consider polynomials $p(x)$ of degree exceeding $2n - 1$, this is not terribly comforting if n is not small. Fortunately, however, the number N of negative eigenvalues (counted according to their multiplicities) of H is a primary factor in determining the maximum degree of the polynomial $p(x)$ required, and can often be used to advantage.

Theorem 4: Let H have J positive eigenvalues (counted according to their multiplicities) and let M be the dimension of the nullspace of H. Let the system (1) be stable, so that there exists a complete L-canonical set of eigenvectors $\{\zeta_k\}_1^{2n}$ satisfying Eqs. (15) and (16), with real eigenvalues ω_k enumerated so that $\omega_1 \geq \omega_2 \geq \cdots \geq \omega_n$ and $\omega_{n+1} \geq \omega_{n+2} \geq \cdots \geq \omega_{2n}$. Then $\omega_J > 0, \omega_{J+M} \geq 0, \omega_{2n-J+1} < 0$, and $\omega_{2n-J-M+1} \leq 0$, i.e., at least J of the $\omega_1, \ldots, \omega_n$ are positive and at least $J + M$ are nonnegative, while at least J of the $\omega_{n+1}, \ldots, \omega_{2n}$ are negative and at least $J + M$ are nonpositive.

Proof: We write

$$\zeta_k = \begin{pmatrix} \xi_k \\ i\omega_k \xi_k \end{pmatrix}, \quad k = 1, \ldots, 2n,$$

and by Theorem 7 of Ref. 1, there exist linear operators D_1, D_2, and $P > 0$ such that Eqs. (33), (35), and (36) of Ref. 1 hold, i.e., $\eta_k \equiv P^{1/2}\xi_k$ and ω_k are eigenvectors and eigenvalues, respectively, of the Hermitian eigenvalue problems

$$\omega_k \eta_k = P^{-1/2}(H + D_1^\dagger D_1)P^{-1/2}\eta_k, \quad k = 1, \ldots, n,$$
$$\omega_k \eta_k = -P^{-1/2}(H + D_2^\dagger D_2)P^{-1/2}\eta_k, k = n + 1, \ldots, 2n. \tag{23}$$

Since $D_1^\dagger D_1 \geq 0$, it follows immediately from the minimax principle for the eigenvalues of a Hermitian operator that $H + D_1^\dagger D_1$ has at least as many positive eigenvalues as does H and at least as many nonnegative eigenvalues as H; since $P^{-1/2}$ is Hermitian positive definite, the number of positive eigenvalues as well as the number of zero eigenvalues of the operators $H + D_1^\dagger D_1$ and $P^{-1/2}(H + D_1^\dagger D_1)P^{-1/2}$ are identical. Therefore $P^{-1/2}(H + D_1^\dagger D_1)P^{-1/2}$ has at least J positive eigenvalues and at least $J + M$ nonnegative eigenvalues. Similarly, since $D_2^\dagger D_2 \geq 0$, $-P^{-1/2}(H + D_2^\dagger D_2)P^{-1/2}$ has at least J negative eigenvalues and at least $J + M$ nonpositive eigenvalues.

Theorem 5: Let N ($< n \equiv \dim E$) be the number of negative eigenvalues of H counted according to their multiplicity. Suppose the system (1) admits a Liapunov operator of the form $Lp(T)$, where $p(x)$ is a real polynomial in x (i.e., suppose the system is structurally stable). Then the degree of $p(x)$ need not exceed $4N + 1$, i.e., there exists a real polynomial $f(x)$ of degree less than or equal to $4N + 1$ such that $Lf(T)$ is a Liapunov operator.

Proof: Let N ($< n$) be the number of negative eigenvalues of H and let $\{\xi_k\}_1^{2n}$ be a complete L-canonical set of eigenvectors satisfying Eqs. (15) and (16) with real eigenvalues ω_k enumerated so that $\omega_1 \geq \omega_2 \geq \cdots \geq \omega_n$ and $\omega_{n+1} \geq \omega_{n+2} \geq \cdots \geq \omega_{2n}$. If 0 is not an eigenvalue of H, we conclude from Theorem 4, with $J = n - N$, that $\omega_{n+N+1} < 0 < \omega_{n-N}$. If 0 is an eigenvalue of H, then 0 is an eigenvalue of T [$H\eta = 0$ implies $T\begin{pmatrix}\eta\\0\end{pmatrix} = 0$]. By assumption, the system is structurally stable, so that the sets $Q_+ \equiv \{\omega_1, \ldots, \omega_n\}$ and $Q_- \equiv \{\omega_{n+1}, \ldots, \omega_{2n}\}$ are disjoint, and therefore either $0 \in Q_+$ or $0 \in Q_-$, but 0 cannot be in both. If $0 \in Q_+$, then we conclude from Theorem 4 (with $J + M = n - N$) that $\omega_{n+N+1} < 0 \leq \omega_{n-N}$, while if $0 \in Q_-$, Theorem 4 implies that $\omega_{n+N+1} \leq 0 < \omega_{n-N}$. Thus, in any case, $\omega_{n+N+1} < \omega_{n-N}$. Let r be the greatest integer $\leq n$ such that $\omega_r > \omega_{n+N+1}$, and let u be the least integer $\geq n + 1$ such that $\omega_u < \omega_r$. Then $n - N \leq r \leq n$, $n + 1 \leq u \leq n + N + 1$, $\omega_k < \omega_u$ for $r + 1 \leq k \leq n$, and $\omega_k > \omega_r$ for $n + 1 \leq k \leq u - 1$. Let $\omega_u < \rho < \omega_r$, and define $p_1(x) \equiv x - \rho$. Then $p_1(\omega_k)\gamma_k > 0$ for $1 \leq k \leq r$ and $u \leq k \leq 2n$. Let $\Omega_1^+, \ldots, \Omega_q^+$ denote the distinct elements of the set $\{\omega_k | r + 1 \leq k \leq n\}$ and let $\Omega_1^-, \ldots, \Omega_y^-$ denote the distinct elements of the set $\{\omega_k | n + 1 \leq k \leq u - 1\}$, so that $q \leq n - r \leq N$ and $y \leq u - 1 - n \leq N$. Since $\Omega_k^+ \in Q_+$ and $Q_+ \cap Q_- = \phi$, there exists $\epsilon > 0$ such that the q closed intervals $I_k^+ \equiv [\Omega_k^+ - \epsilon, \Omega_k^+ + \epsilon]$ are all disjoint and satisfy $I_k^+ \cap Q_- = \phi$, $k = 1, \ldots, q$. Similarly, there exists $\delta > 0$ such that the y closed intervals $I_k^- \equiv [\Omega_k^- - \delta, \Omega_k^- + \delta]$ are all disjoint and satisfy $I_k^- \cap Q_+ = \phi$, $k = 1, \ldots, y$. We define

$$p(x) \equiv p_1(x)\left[\prod_1^q (x - \Omega_k^+ + \epsilon)(x - \Omega_k^+ - \epsilon)\right] \times \left[\prod_1^y (x - \Omega_k^- + \delta)(x - \Omega_k^- - \delta)\right]. \quad (24)$$

Then by construction, Eq. (19) holds, so that $Lp(T) > 0$ and is therefore a Liapunov operator. The degree of $p(x)$ is given by $1 + 2q + 2y \leq 1 + 2N + 2N = 1 + 4N$, and the proof is complete.

ACKNOWLEDGMENT

The work presented here was sponsored by the Air Force Office of Scientific Research, under Research Grant No. AFOSR 71-2053.

[1] E. M. Barston, J. Math. Phys. 12, 1116 (1971).
[2] E. M. Barston, J. Math. Phys. 8, 523 (1967).
[3] E. M. Barston, J. Math. Phys. 8, 1886 (1967).
[4] E. M. Barston, J. Math. Phys. 9, 2069 (1968).
[5] E. M. Barston, J. Math. Phys. 12, 1867 (1971).

Exchange Interaction Model of Ferromagnetism*

H. H. Chen† and R. K. Joseph
Department of Electrical Engineering, The Johns Hopkins University, Baltimore, Maryland 21218
(Received 4 January 1971)

The Schrödinger exchange operator for arbitrary spin has been used to form an interaction Hamiltonian for a nearest-neighbor model of ferromagnetism. Through use of the cluster expansion method and new group theoretic results in conjunction with the diagrammatic method, eight terms in the high temperature series for the zero-field partition function and the low-field susceptibility are obtained for arbitrary spin and general crystal lattice. Critical parameters are estimated from these series by means of various ratio tests and Padé approximants. For the cubic lattices the Curie temperature T_c and the critical index γ are given by

$$k_B T_C/J = 0.547(q - 1.6)(Y^{-1} + 0.21)$$

and

$$\gamma = 0.48 + 2.16\,Y^{-1} \quad \text{for } S > \tfrac{1}{2};$$
$$= 1.41 \pm 0.02 \quad \text{for } S = \tfrac{1}{2},$$

respectively, where $Y = 2S + 1$. Comparison of these results with those appropriate to the Heisenberg model as well as to experimental values is made. The concept of multipolar ordering is also discussed. It is shown that for the present model all of the $2S$ "independent" multipolar phase transitions are exactly degenerate with the usual dipolar transition.

I. INTRODUCTION

For the Heisenberg model of ferromagnetism, high temperature series for various thermodynamic quantities have been extensively studied and used to investigate critical properties of ferromagnetic systems. If we consider a ferromagnetic system contain-

Since $D_1^\dagger D_1 \geq 0$, it follows immediately from the minimax principle for the eigenvalues of a Hermitian operator that $H + D_1^\dagger D_1$ has at least as many positive eigenvalues as does H and at least as many nonnegative eigenvalues as H; since $P^{-1/2}$ is Hermitian positive definite, the number of positive eigenvalues as well as the number of zero eigenvalues of the operators $H + D_1^\dagger D_1$ and $P^{-1/2}(H + D_1^\dagger D_1)P^{-1/2}$ are identical. Therefore $P^{-1/2}(H + D_1^\dagger D_1)P^{-1/2}$ has at least J positive eigenvalues and at least $J + M$ nonnegative eigenvalues. Similarly, since $D_2^\dagger D_2 \geq 0$, $-P^{-1/2}(H + D_2^\dagger D_2)P^{-1/2}$ has at least J negative eigenvalues and at least $J + M$ nonpositive eigenvalues.

Theorem 5: Let N ($< n \equiv \dim E$) be the number of negative eigenvalues of H counted according to their multiplicity. Suppose the system (1) admits a Liapunov operator of the form $Lp(T)$, where $p(x)$ is a real polynomial in x (i.e., suppose the system is structurally stable). Then the degree of $p(x)$ need not exceed $4N + 1$, i.e., there exists a real polynomial $f(x)$ of degree less than or equal to $4N + 1$ such that $Lf(T)$ is a Liapunov operator.

Proof: Let N ($< n$) be the number of negative eigenvalues of H and let $\{\xi_k\}_1^{2n}$ be a complete L-canonical set of eigenvectors satisfying Eqs. (15) and (16) with real eigenvalues ω_k enumerated so that $\omega_1 \geq \omega_2 \geq \cdots \geq \omega_n$ and $\omega_{n+1} \geq \omega_{n+2} \geq \cdots \geq \omega_{2n}$. If 0 is not an eigenvalue of H, we conclude from Theorem 4, with $J = n - N$, that $\omega_{n+N+1} < 0 < \omega_{n-N}$. If 0 is an eigenvalue of H, then 0 is an eigenvalue of T [$H\eta = 0$ implies $T\begin{pmatrix}\eta\\0\end{pmatrix} = 0$]. By assumption, the system is structurally stable, so that the sets $Q_+ \equiv \{\omega_1, \ldots, \omega_n\}$ and $Q_- \equiv \{\omega_{n+1}, \ldots, \omega_{2n}\}$ are disjoint, and therefore either $0 \in Q_+$ or $0 \in Q_-$, but 0 cannot be in both. If $0 \in Q_+$, then we conclude from Theorem 4 (with $J + M = n - N$) that $\omega_{n+N+1} < 0 \leq \omega_{n-N}$, while if $0 \in Q_-$, Theorem 4 implies that $\omega_{n+N+1} \leq 0 < \omega_{n-N}$. Thus, in any case, $\omega_{n+N+1} < \omega_{n-N}$. Let r be the greatest integer $\leq n$ such that $\omega_r > \omega_{n+N+1}$, and let u be the least integer $\geq n + 1$ such that $\omega_u < \omega_r$. Then $n - N \leq r \leq n$, $n + 1 \leq u \leq n + N + 1$, $\omega_k < \omega_u$ for $r + 1 \leq k \leq n$, and $\omega_k > \omega_r$ for $n + 1 \leq k \leq u - 1$. Let $\omega_u < \rho < \omega_r$, and define $p_1(x) \equiv x - \rho$. Then $p_1(\omega_k)\gamma_k > 0$ for $1 \leq k \leq r$ and $u \leq k \leq 2n$. Let $\Omega_1^+, \ldots, \Omega_q^+$ denote the distinct elements of the set $\{\omega_k | r + 1 \leq k \leq n\}$ and let $\Omega_1^-, \ldots, \Omega_y^-$ denote the distinct elements of the set $\{\omega_k | n + 1 \leq k \leq u - 1\}$, so that $q \leq n - r \leq N$ and $y \leq u - 1 - n \leq N$. Since $\Omega_k^+ \in Q_+$ and $Q_+ \cap Q_- = \phi$, there exists $\epsilon > 0$ such that the q closed intervals $I_k^+ \equiv [\Omega_k^+ - \epsilon, \Omega_k^+ + \epsilon]$ are all disjoint and satisfy $I_k^+ \cap Q_- = \phi$, $k = 1, \ldots, q$. Similarly, there exists $\delta > 0$ such that the y closed intervals $I_k^- \equiv [\Omega_k^- - \delta, \Omega_k^- + \delta]$ are all disjoint and satisfy $I_k^- \cap Q_+ = \phi$, $k = 1, \ldots, y$. We define

$$p(x) \equiv p_1(x)\left[\prod_1^q (x - \Omega_k^+ + \epsilon)(x - \Omega_k^+ - \epsilon)\right] \times \left[\prod_1^y (x - \Omega_k^- + \delta)(x - \Omega_k^- - \delta)\right]. \quad (24)$$

Then by construction, Eq. (19) holds, so that $Lp(T) > 0$ and is therefore a Liapunov operator. The degree of $p(x)$ is given by $1 + 2q + 2y \leq 1 + 2N + 2N = 1 + 4N$, and the proof is complete.

ACKNOWLEDGMENT

The work presented here was sponsored by the Air Force Office of Scientific Research, under Research Grant No. AFOSR 71-2053.

[1] E. M. Barston, J. Math. Phys. 12, 1116 (1971).
[2] E. M. Barston, J. Math. Phys. 8, 523 (1967).
[3] E. M. Barston, J. Math. Phys. 8, 1886 (1967).
[4] E. M. Barston, J. Math. Phys. 9, 2069 (1968).
[5] E. M. Barston, J. Math. Phys. 12, 1867 (1971).

Exchange Interaction Model of Ferromagnetism*

H. H. Chen† and R. K. Joseph
Department of Electrical Engineering, The Johns Hopkins University, Baltimore, Maryland 21218
(Received 4 January 1971)

The Schrödinger exchange operator for arbitrary spin has been used to form an interaction Hamiltonian for a nearest-neighbor model of ferromagnetism. Through use of the cluster expansion method and new group theoretic results in conjunction with the diagrammatic method, eight terms in the high temperature series for the zero-field partition function and the low-field susceptibility are obtained for arbitrary spin and general crystal lattice. Critical parameters are estimated from these series by means of various ratio tests and Padé approximants. For the cubic lattices the Curie temperature T_c and the critical index γ are given by

$$k_B T_C/J = 0.547(q - 1.6)(Y^{-1} + 0.21)$$

and

$$\gamma = 0.48 + 2.16\,Y^{-1} \quad \text{for } S > \tfrac{1}{2};$$
$$= 1.41 \pm 0.02 \quad \text{for } S = \tfrac{1}{2},$$

respectively, where $Y = 2S + 1$. Comparison of these results with those appropriate to the Heisenberg model as well as to experimental values is made. The concept of multipolar ordering is also discussed. It is shown that for the present model all of the $2S$ "independent" multipolar phase transitions are exactly degenerate with the usual dipolar transition.

I. INTRODUCTION

For the Heisenberg model of ferromagnetism, high temperature series for various thermodynamic quantities have been extensively studied and used to investigate critical properties of ferromagnetic systems. If we consider a ferromagnetic system contain-

ing N particles of spin S with isotropic nearest-neighbor exchange interactions, the Heisenberg Hamiltonian is given by

$$\mathcal{H} = -2J\sum_{<ij>} \mathbf{S}_i\cdot\mathbf{S}_j - g\mu H\sum_{i=1}^{N} S_{zi}, \tag{1}$$

where J is the nearest-neighbor exchange constant, $\mathbf{S}_i$ is the spin operator of an atom located at the lattice site labeled i, g is the gyromagnetic ratio, μ the Bohr magneton, H the z-directed external magnetic field, and S_{zi} is the z component of $\mathbf{S}_i$. The first term in the Hamiltonian represents the exchange energy and the summation is taken over all nearest-neighbor pairs of atoms, denoted $\langle ij\rangle$; the sum in the second term is over all atoms and is the Zeeman energy of the system.

The Heisenberg Hamiltonian linear in $\mathbf{S}_i\cdot\mathbf{S}_j$ which arises from a consideration of the Coulomb interaction together with the Pauli principle is in fact only the lowest-order significant term in a perturbation expansion which when carried further leads to terms nonlinear in $\mathbf{S}_i\cdot\mathbf{S}_j$.[1]

In order to study the properties of systems containing nonlinear terms $J_n(\mathbf{S}_i\cdot\mathbf{S}_j)^n$, in the Hamiltonian, the Schrödinger exchange operator has been used to form an interaction Hamiltonian for a nearest-neighbor model of ferromagnetism,[2,3]

$$\mathcal{H} = -J\sum_{<ij>} P_{ij} - g\mu H\sum_{i=1}^{N} S_{zi}. \tag{2}$$

Here P_{ij} is the Schrödinger exchange operator and is a polynomial of degree $2S$ in $\mathbf{S}_i\cdot\mathbf{S}_j$:

$$P_{ij} = \sum_{n=0}^{2S} A_n(S)(\mathbf{S}_i\cdot\mathbf{S}_j)^n, \qquad i \neq j. \tag{3}$$

The coefficients A_n are determined from the property that P_{ij} exchanges, or permutes, the spin coordinates of two atoms labeled i and j:

$$P_{ij}\,|\,m\rangle_i\,|\,m'\rangle_j = |\,m'\rangle_i\,|\,m\rangle_j \tag{4}$$

or

$$P_{ij}O(i,j) = O(j,i)P_{ij}, \tag{5}$$

where $O(i,j)$ is any operator which contains the spin operators of atoms i and j and $|\,m\rangle_i$ and $|\,m\rangle_j$ are eigenstates of S_{zi} and S_{zj}, respectively. Schrödinger[4] has explicitly shown that

$$P_{ij} = (-1)^{2S}\left(1 + \sum_{p=1}^{2S}\frac{(-1)^p}{(p!)^2}\prod_{q=1}^{p}[M - q(q-1)]\right), \tag{6}$$

where $M \doteq 2\,[S(S+1) + \mathbf{S}_i\cdot\mathbf{S}_j]$. The coefficients A_n typically have the values

$$\begin{aligned}
&S = \tfrac{1}{2}\colon\ A_0 = \tfrac{1}{2},\quad A_1 = 2,\\
&S = 1\colon\ A_0 = -1,\quad A_1 = 1,\quad A_2 = 1,\\
&S = \tfrac{3}{2}\colon\ A_0 = -\tfrac{67}{32},\quad A_1 = -\tfrac{9}{8},\quad A_2 = \tfrac{11}{18},\quad A_3 = \tfrac{2}{9},\\
&S = 2\colon\ A_0 = -1,\quad A_1 = -\tfrac{5}{2},\quad A_2 = -\tfrac{13}{16},\quad A_3 = \tfrac{1}{6},\quad A_4 = \tfrac{1}{36},
\end{aligned} \tag{7}$$

etc.

Inclusion of these special combinations of nonlinear terms in the Hamiltonian may not be realized in nature. However, a study of this model gives detailed information about what effect such nonlinear terms should have on the critical properties of the system.

Due to the permutation property of the Schrödinger exchange operator (5), the high temperature series can be extended further with less effort for the present Hamiltonian than for the Heisenberg Hamiltonian. For the case $S = \frac{1}{2}$, the present Hamiltonian is identical to that of the Heisenberg model; Baker *et al.*[5] obtained terms through T^{-9} for the close-packed lattices and T^{-10} for the loose-packed lattices for the zero-field partition function series and the low-field susceptibility series. For $S = 1$, Allan and Betts[2] obtained eight terms in these series for the face-centered cubic lattice. Such a large number of terms was obtained through the use of the cluster expansion method in conjunction with a technique making use of "Branching diagrams"; this is practicable only for the case of S equal to $\frac{1}{2}$ or 1. We have developed a new method which can be applied to the case of arbitrary spin directly and hence obtained eight terms in the high temperature series for general crystal lattices.

High temperature series expansions and the cluster expansion method are discussed in Secs. II and III, respectively; these ideas can be applied to any of the spin Hamiltonians usually studied. Sections IV and V contain group theoretical considerations and the diagrammatic method required to calculate the series coefficients for the present model. Details of the calculations are given in Sec. VI and the explicit series results are presented in Sec. VII. Several checking procedures on the results are considered in Sec. VIII. In Sec. IX these high temperature series are used to estimate various critical properties by means of ratio tests and the method of Padé approximants. The significance of the results as well as the concept of multipolar ordering is found in Sec. X.

II. HIGH TEMPERATURE SERIES EXPANSIONS

A. Preliminary Remarks

For any spin Hamiltonian $\mathcal{H}$ the high temperature series expansion method introduced by Kramers[6] and Opechowski[7] makes use of a result of the form

$$\begin{aligned}
Z &= \mathrm{tr}\,e^{-\beta\mathcal{H}}\\
&= \mathrm{tr}\mathbf{I}[1 - \beta\langle\mathcal{H}\rangle + (\beta^2/2!)\langle\mathcal{H}^2\rangle\\
&\quad - (\beta^3/3!)\langle\mathcal{H}^3\rangle + \cdots],
\end{aligned} \tag{8}$$

where Z is the partition function, $\beta = (k_BT)^{-1}$, k_B is Boltzmann's constant, $\langle\mathcal{H}^n\rangle = \mathrm{tr}\mathcal{H}^n/\mathrm{tr}\mathbf{I}$, and $\mathbf{I}$ is the unit matrix. Related thermodynamic functions can then be expressed as ascending series in powers of $1/T$ by evaluating the leading coefficients in the series for various crystal lattices. The first few terms of these series provide a good approximation to each thermodynamic quantity at high temperatures. Furthermore, extrapolations from such truncated expansion series are considered to be the most powerful theoretical approach yet developed for obtaining estimates of the various critical parameters.

B. Zero-Field Partition Function and Related Thermodynamic Functions

Since the various thermodynamic functions are re-

lated to the partition function by $\ln Z$, it is convenient to express the partition function in the form $\ln Z$. Moreover, since $\ln Z$ is an extensive quantity, considerable simplification can be made in the derivation of high temperature series by means of the cluster expansion method, as will be discussed in the next section. We introduce the following convenient notations: $Y = 2S + 1$, $X = S(S + 1)$, $\alpha = g\mu H/J$, $K = J/k_BT$, and

$$Q = \sum_{i=1}^{N} S_{zi}, \tag{9}$$

$$\mathcal{P} = \sum_{<ij>} P_{ij}, \quad \text{for the exchange interaction model,} \tag{10}$$

$$= 2\sum_{<ij>} \mathbf{S}_i \cdot \mathbf{S}_j, \quad \text{for the Heisenberg model,}$$

$$= 2\sum_{<ij>} S_{zi} S_{zj}, \quad \text{for the Ising model.}$$

Then,

$$\mathcal{H} = -J(\mathcal{P} + \alpha Q) = \mathcal{H}_0 - J\alpha Q. \tag{11}$$

For a system which consists of N particles of spin S, $\mathrm{tr} I = Y^N$. In zero external field,

$$Z = Y^N \left(1 + \sum_{n=1}^{\infty} \frac{K^n}{n!} \langle \mathcal{P}^n \rangle\right), \tag{12}$$

and

$$\ln Z = N \ln Y + \ln \left(1 + \sum_{n=1}^{\infty} \frac{K^n}{n!} \langle \mathcal{P}^n \rangle\right), \tag{13}$$

from which it follows on expanding the logarithm

$$\ln Z = N \ln Y + \sum_{n=1}^{\infty} e_n \frac{K^n}{n!}, \tag{14}$$

with

$$e_n = \sum \frac{(-1)^{\alpha_1+\alpha_2+\cdots+\alpha_p-1}(\alpha_1 + \alpha_2 + \cdots + \alpha_p - 1)!}{\alpha_1!\alpha_2!\ldots\alpha_p!(a_1!)^{\alpha_1}(a_2!)^{\alpha_2}\ldots(a_p!)^{\alpha_p}} \times \langle \mathcal{P}^{a_1} \rangle^{\alpha_1} \langle \mathcal{P}^{a_2} \rangle^{\alpha_2} \ldots \langle \mathcal{P}^{a_p} \rangle^{\alpha_p}. \tag{15}$$

The summation is taken over all partitions of the integer n, namely, all sets of positive integers $(a_1, a_2, \ldots, a_p; \alpha_1, \alpha_2, \ldots, \alpha_p)$ which satisfy the conditions

$$a_1\alpha_1 + a_2\alpha_2 + \cdots + a_p\alpha_p = n \tag{16}$$

and

$$a_1 < a_2 < \cdots < a_p. \tag{17}$$

Each thermodynamic function series can now be directly obtained from Eq. (14):

internal energy:
$$E = k_BT^2 \frac{\partial}{\partial T} (\ln Z)$$
$$= -J \frac{\partial}{\partial K} (\ln Z), \tag{18}$$

entropy:
$$S = k_B \frac{\partial}{\partial T} (T \ln Z)$$
$$= -k_B K^2 \frac{\partial}{\partial K} \frac{(\ln Z)}{K}, \tag{19}$$

specific heat:
$$C_v = k_B \frac{\partial}{\partial T} \left(T^2 \frac{\partial}{\partial T} \ln Z\right)$$
$$= k_B K^2 \frac{\partial^2}{\partial K^2} (\ln Z). \tag{20}$$

C. Low-Field Susceptibility

The low-field susceptibility is defined as

$$\chi = \lim_{H \to 0} k_B T \frac{\partial^2}{\partial H^2} \ln Z = \lim_{\alpha \to 0} k_B T \left(\frac{g\mu}{J}\right)^2 \frac{\partial^2}{\partial \alpha^2} \ln Z. \tag{21}$$

Since $\mathcal{P}$ and Q commute, $(\partial/\partial\alpha)(\mathcal{P} + \alpha Q)^n = n(\mathcal{P} + \alpha Q)^{n-1} Q$. It is then straightforward to show that

$$\chi = [(g\mu)^2/k_BT]\Delta(Q), \tag{22}$$

where $\Delta(Q) = \langle Q^2 \rangle_\beta - \langle Q \rangle_\beta^2$ is the zero-field thermal fluctuation of Q, and, for any operator A,

$$\langle A \rangle_\beta = \frac{\mathrm{tr} A e^{-\beta \mathcal{H}_0}}{\mathrm{tr} e^{-\beta \mathcal{H}_0}} = \sum_{n=0}^{\infty} \frac{K^n}{n!} \langle \mathcal{P}^n A \rangle \Big/ \sum_{n=0}^{\infty} \frac{K^n}{n!} \langle \mathcal{P}^n \rangle. \tag{23}$$

As a consequence of the fact that $\ln Z$ is an even function of the external field, $\mathrm{tr}\, \mathcal{P}^n Q = 0$ for all n; hence, $\langle Q \rangle_\beta = 0$, so that we can rewrite χ in the form

$$\chi = \frac{(g\mu)^2 X}{3k_BT} \left(N + \sum_{n=1}^{\infty} a_n \frac{K^n}{n!}\right). \tag{24}$$

It is easy to show that the coefficients a_n satisfy the recursion relation

$$a_n = \frac{3}{X} \langle \mathcal{P}^n Q^2 \rangle - \sum_{k=0}^{n-1} \frac{n!}{k!(n-k)!} a_k \langle \mathcal{P}^{n-k} \rangle, \tag{25}$$

with $a_0 = N$. In order to obtain the terms e_k and a_k, we must calculate the quantities $\langle \mathcal{P}^n \rangle$ and $\langle \mathcal{P}^n Q^2 \rangle$ for $n \leq k$. Here the matrices $\mathcal{P}$ and Q are of order $(2S + 1)^N$. For real crystals, $N \to \infty$, and direct computation of the required traces is impossible. However, there are two alternative methods of handling the calculation. In the next section we discuss the cluster expansion method, while the diagrammatic method is considered in Sec. V.

III. CLUSTER EXPANSION METHOD

A. Preliminary Remarks

The use of the cluster expansion method in deriving series expansions for magnetic systems was first suggested by Domb.[8] It was pointed out that high temperature series for extensive quantities for infinite lattices ($N \to \infty$) can be obtained simply by calculating the corresponding series for clusters of finite sites. One advantage of this method as compared to the diagrammatic method is that the number of configurations that one has to consider are considerably smaller than that required in the diagrammatic method. More important, however, is the fact that most of the calculations called for in the cluster expansion method can easily be done on a fast computer.

The general method of the cluster expansions has been developed by many authors and derived in a number of different ways. In this section we introduce a new proof of an essential theorem of this method (Theorem 1), which can be applied directly to the calculation of the high temperature series for any spin Hamiltonian.

B. General Concepts

A (linear) graph is a collection of points with lines joining certain pairs of points. If a subset of points

are joined successively by lines, the assembly of these lines is called a path connecting the initial and the final points. A graph is said to be connected if any two points in the graph are connected by a path. Otherwise, the graph is said to be disconnected. Clearly, any graph consists of connected graphs, and each connected graph is called a component. If the initial point and the final point of a path coincide, we speak of a cycle. A set of different cycles is said to be independent if none of the cycles can be made up of parts of other cycles. The maximum number of independent cycles in a graph is called the cyclomatic number of the graph. It is well known that for a connected graph

$$c = \ell - p + 1, \tag{26}$$

where c is the cyclomatic number, ℓ the number of lines, and p the number of points in a graph. In general, if we denote the number of connected components in a graph by n, then

$$c = l - p + n. \tag{27}$$

A connected graph is said to be closed if any point in the graph has at least two lines connected to it. Otherwise, it is said to be open.

If g_r and g_s are two graphs having no points in common, the union of these two graphs, denoted $g_r \cup g_s$, is the collection of all points and lines of g_r and g_s. A graph g is a subgraph of G if any point in g is a point in G and any line in g is a line in G. A graph G' is said to be isomorphic with G if there is a one-to-one correspondence between their points such that pairs of points are joined by lines in G' if and only if the corresponding pairs of points are joined in G. The lattice constant[9] of a graph g on a graph G is the number of subgraphs of G isomorphic with g, denoted $(g; G)$. $(g; G)$ is sometimes abbreviated as $[g]$ if G is not specified. Lattice constants of disconnected graphs can be expressed in terms of lattice constants of connected graphs. For example, consider the equation shown in Fig. 1. The first term on the right-hand side is of second order while the others are of first order in the lattice constants of connected graphs. In general, lattice constants of disconnected graphs having n connected components will consist of terms from the first order to the nth order in lattice constants of connected graphs.

C. New Derivation of the Cluster Expansions

Let $\phi(g)$ be any quantity associated with the graph g. ϕ is said to be extensive if, for any graphs g_r and g_s having no points in common,

$$\phi(g_r \cup g_s) = \phi(g_r) + \phi(g_s); \tag{28}$$

that is, the quantity ϕ of two graphs considered together is the sum of the quantities of the two graphs considered separately. The number of lines and the number of points in a graph are obviously extensive. $(g; G)$ is also an extensive quantity of G, i.e.,

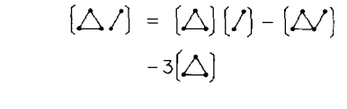


FIG. 1. Example of expressing the lattice constant of a disconnected graph in terms of lattice constants of connected graphs.

$$(g; g_r \cup g_s) = (g; g_r) + (g; g_s). \tag{29}$$

Suppose that a graph G consists of the number π_i of connected graphs g_i, $i = 1, 2, 3, \cdots$. Let $\phi(G; t)$ be an extensive quantity of G and t be a set of parameters independent of graphs. By the extensive property of ϕ,

$$\phi(G; t) = \sum_i \phi(g_i; t)\pi_i. \tag{30}$$

Using the extensive property of the lattice constant as expressed by Eq. (29), set $g = g_j$ for $j = 1, 2, 3, \cdots$. We then have a set of linear equations

$$(g_j; G) = \sum_i (g_j; g_i)\,\pi_i, \quad j = 1, 2, 3, \cdots. \tag{31}$$

If the graphs are labeled in the graph dictionary order such that

$$l_i \leq l_j, \quad \text{for } i < j, \tag{32}$$

where l_i and l_j are the numbers of lines of graphs g_i and g_j, respectively, it is then obvious that

$$\begin{aligned} (g_j; g_i) &= 0, \quad \text{for } j > i, \\ &= 1, \quad \text{for } j = i. \end{aligned} \tag{33}$$

Define a matrix A with elements $A_{mn} = (g_m; g_n)$. Equation (33) then means that A is a triangular matrix with the lower triangular elements equal to zero. Furthermore, all the diagonal elements are unity. Therefore, A is nonsingular and its inverse exists. From Eq. (31), we get

$$\pi_i = \sum_j (A^{-1})_{ij}(g_j; G), \quad i = 1, 2, 3, \ldots, \tag{34}$$

where $(A^{-1})_{ij}$ are elements of the inverse matrix of A. Substituting Eq. (34) into Eq. (30) yields

$$\begin{aligned} \phi(G; t) &= \sum_i \sum_j \phi(g_i; t)(A^{-1})_{ij}(g_j; G) \\ &= \sum_j F_j(t)(g_j; G), \end{aligned} \tag{35}$$

where $F_j(t) = \sum_i (g_i; t)(A^{-1})_{ij}$ is independent of G. Since G may be any graph, let $G = g_i$. Equation (35) yields

$$\phi(g_i; t) = \sum_j F_j(t)(g_j; g_i). \tag{36}$$

On substituting Eq. (33) into Eq. (36) and rearranging terms, we obtain an important theorem formulated by Sykes *et al.*[10]

Theorem 1: If $\phi(G; t)$ satisfies the extensive property, then ϕ can be expressed by Eq. (35) in which $F_j(t)$ are given by the recursion formula

$$F_j(t) = \phi(g_j; t) - \sum_{i=1}^{j-1} (g_i; g_j)F_i(t) \tag{37}$$

and

$$F_1(t) = \phi(g_1; t). \tag{38}$$

D. Application to Magnetic Systems

In a nearest-neighbor model of any of the spin Hamiltonians, if spin sites are represented by points and interactions between nearest-neighbor sites are repre-

sented by lines joining the corresponding pairs of points, then systems represented by graphs isomorphic to each other will have the same physical properties and systems which consist of a number of independent subsystems will be represented by disconnected graphs. Let $\phi(G)$ be any quantity of a magnetic system represented by a graph G. It is clear then that $\ln Z(G)$ and $\chi(G)$ satisfy Eq. (28) since $\ln Z$ and χ are extensive thermodynamic quantities. We can then use Theorem 1 to calculate $\ln Z(G; t)$ and $\chi(G; t)$ for any cluster or crystal lattice G and for a set of parameters t, such as temperature T, spin value S, external magnetic field H, exchange constant J, etc. In high temperature series expansions, we express F_j [in Eq. (35)] in powers of $K(= J/k_B T)$

$$F_j = \sum_{n\ 0}^{\infty} f_n(g_j) K^n. \tag{39}$$

Theorem 1 is useful because in most cases $f_n(g_j) = 0$ for $n < l_j$. This will be explicitly shown in Sec. V for the present Hamiltonian when $\phi = \ln Z$ or χ. Therefore, if we want to obtain series up to the kth power in K for $\phi(G)$, we need only calculate $\phi(g_j)$ and hence F_j for connected graphs having up to k lines. Thus, instead of considering infinite lattices, we have reduced the problem to clusters of finite size.

IV. APPLICATION OF GROUP THEORY

A. Symmetric Group

A rearrangement of the order of N symbols is called a permutation. There are $N!$ number of possible permutations on N symbols. The set of all these permutations form a group called the symmetric group of degree N, denoted S_N. Each elements in S_N can be written as a product of independent cycles. For example, the permutation by which $\{a, b, c, d, e, f, g\}$ is replaced by $\{b, f, c, e, d, a, g\}$ can be written as $(abf)(c)(de)(g)$. Each ( ) is called a cycle, and the number of symbols in ( ) is the order of the cycle. In writing a permutation as the product of cycles, cycles of order 1 need not be mentioned. Furthermore, the sequence of appearance of the cycles as well as the first symbol in each cycle is arbitrary.

Cycles having no symbols in common are said to be independent. A cycle of order 2 is called an interchange. Any cycle of higher order can be expressed as a product of interchanges (having symbols in common). For example,

$$(abc\cdots de) = (ae)(ad)\cdots(ac)(ab). \tag{40}$$

Permutations which are products of an even number of interchanges are called even permutations. Otherwise, they are called odd permutations. It is straightforward to show that

$$(fg)(fa\cdots bgc\cdots\cdots d) = (fa\cdots b)(gc\cdots d) \tag{41}$$

and

$$(fg)(fa\cdots b)(gc\cdots d) = (fa\cdots\cdots bgc\cdots\cdots d). \tag{42}$$

Equations (41) and (42) imply that for any group element, say P, when multiplied by an interchange (fg), $(fg)P$ has one more cycle than P if f, g belong to the same cycle in P and has one less cycle than P if f, g belong to different independent cycles in P.

Elements in S_N are divided into classes. Elements which have the same cycle structure are transforms of one another by elements in S_N and belong to the same class. Elements which belong to the class $(a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_p^{\alpha_p})$ have α_1 cycles of order a_1, α_2 cycles of order $a_2, \ldots$, and α_p cycles of order a_p. The number of classes in S_N is equal to the number of partitions of the integer N, and each class is labeled by a partition of N. The number of group elements in a class is called the order of the class. For the class $k = (a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_p^{\alpha_p})$, its order h_k is

$$h_k = N!/\alpha_1!\alpha_2! \cdots \alpha_p! a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_p^{\alpha_p}. \tag{43}$$

Associated with S_N there is a finite number of inequivalent representations, called irreducible representations. Each irreducible representation is also labeled by a partition of N. Since elements of a class are transforms of one another, their matrices in any representation, say ν, have the same trace. This value is referred to as the character of the class k in the representation ν, denoted $\chi_k^{(\nu)}$.

Below we summarize some group theoretic results which will be useful in deriving high temperature series for the present Hamiltonian.

For the irreducible representations ν and ν', we have the orthogonality relations

$$\sum_k \frac{h_k}{h} \chi_k^{(\nu)} \chi_k^{(\nu')} = \delta_{\nu\nu'}, \tag{44}$$

and

$$\sum_\nu \frac{h_k}{h} \chi_k^{(\nu)} \chi_{k'}^{(\nu)} = \delta_{kk'}, \tag{45}$$

where h_k is the order of the class k, $h = N!$ is the order of S_N, and the summations $\sum_k$ and $\sum_\nu$ are taken over all classes k and all irreducible representations ν, respectively.

Let $P^{(\Gamma)}$ be any matrix representation of an element P in S_N. $P^{(\Gamma)}$ can be resolved into a direct sum of n_ν number of ν irreducible representations, i.e., the matrix $P^{(\Gamma)}$ now takes the form of a series of blocks, the irreducible representations, placed along the principal diagonal, which can be written as

$$P^{(\Gamma)} = \sum_\nu n_\nu P^{(\nu)}. \tag{46}$$

From Eq. (44), we get

$$n_\nu = \sum_k \frac{h_k}{h} \chi_k^{(\nu)} \chi_k^{(\Gamma)}. \tag{47}$$

For two irreducible representations labeled by a pair of conjugate partitions ν and $\tilde{\nu}$, we have

$$\chi_k^{(\nu)} = \pm \chi_k^{(\tilde{\nu})}, \tag{48}$$

with the plus sign applying for even classes and the minus sign for odd classes of permutations.

If we sum the matrices of an irreducible representation ν for all elements of a class k, we obtain a multiple of the unit matrix:

$$\sum_{P \in k} P^{(\nu)} = [(h_k \chi_k^{(\nu)})/\chi_{1^N}^{(\nu)}]\mathrm{I}, \tag{49}$$

where $\chi_{1^N}^{(\nu)}$ is the matrix dimension of the irreducible representation ν.

Using these equations, we now prove several useful new theorems.

B. New Theorems

Theorem 2: If R is any matrix which commutes with all elements of S_N in a matrix representation Γ, then

$$\mathrm{tr}[(\mathcal{P}^{(\Gamma)})^n R] = \sum_\nu \sum_k \mathrm{tr}(\mathcal{P}^{(\nu)})^n \frac{h_k}{h} \chi_k^{(\nu)} \mathrm{tr}[P_k^{(\Gamma)} R], \tag{50}$$

where $\mathcal{P}^{(\Gamma)}$ and $\mathcal{P}^{(\nu)}$ are sums of elements in S_N in the Γ and the ν representations, respectively, P_k is any element in the class k, and the summations are taken over all irreducible representations ν and all classes k.

Proof: The condition that R commutes with all elements of S_N in a matrix representation Γ and the fact that elements of a class are transforms of one another imply that $\mathrm{tr}P^{(\Gamma)}R$ have the same value for all elements P which belong to the same class. Since products of elements in S_N are also elements in S_N, $\mathcal{P}^n$ as well as $\mathcal{P}$ is a sum of elements in S_N. It is then sufficient to show that for any element, say P, in S_N

$$\mathrm{tr}P^{(\Gamma)}R = \sum_\nu \sum_k \mathrm{tr}P^{(\nu)} \frac{h_k}{h} \chi_k^{(\nu)} \mathrm{tr}P_k^{(\Gamma)} R. \tag{51}$$

Let P belong to the class k'; $\mathrm{tr}P^{(\nu)} = \chi_{k'}^{(\nu)}$. By Eq. (45)

$$\begin{aligned} &\sum_\nu \sum_k \chi_{k'}^{(\nu)} \frac{h_k}{h} \chi_k^{(\nu)} \mathrm{tr}P_k^{(\Gamma)} R \\ &\quad = \sum_k \delta_{kk'} \mathrm{tr}P_k^{(\Gamma)} R \\ &\quad = \mathrm{tr}P_{k'}^{(\Gamma)} R. \end{aligned} \tag{52}$$

Equation (51) and hence Eq. (50) then follow.

Theorem 3: If ν and $\tilde{\nu}$ are two irreducible representations labeled by a pair of conjugate partitions of N and $\mathcal{P}$ is a sum of odd permutations in S_N, then

$$\mathrm{tr}(\mathcal{P}^{(\nu)})^n = (-1)^n \mathrm{tr}(\mathcal{P}^{(\tilde{\nu})})^n. \tag{53}$$

Proof: Since a product of n odd permutations is an even permutation if n is an even number and is an odd permutation if n is odd, Eq. (53) follows from Eq. (48).

C. The $(2S+1)^N$-Dimensional Representation of S_N

For a system containing N particles of spin S, it is clear that matrices of the Schrödinger exchange operators P_{ij} and their products form a $(2S+1)^N$-dimensional representation of S_N. It is convenient to choose the basis of the $(2S+1)^N$-dimensional vector space as the eigenstates of the z component of the spins of the N particles, $|m_1\rangle |m_2\rangle \cdots |m_N\rangle$. In this section we restrict our attention to this representation. Unless otherwise specified, matrix representation of operators will be in this representation. We first prove the following results:

$$\begin{aligned} \mathrm{tr}(ijk\cdots l) &= \sum_{m_i} \sum_{m_j} \cdots \sum_{m_k} \sum_{m_l} (\delta_{m_i,m_j} \delta_{m_i,m_k} \cdots \delta_{m_i,m_l}) \\ &= \sum_{m_i} (1) = Y. \end{aligned} \tag{54}$$

All sums in this equation, and those following directly after, range from $-S$ to $+S$. Similarly,

$$\begin{aligned} \mathrm{tr}S_{zi}^n (ijk\cdots l) &= \sum_{m_i} \sum_{m_j} \cdots \sum_{m_k} \sum_{m_l} (m_i^n \delta_{m_i,m_j} \delta_{m_i,m_k} \cdots \delta_{m_i,m_l}) \\ &= \sum_{m_i} (m_i^n) = YW_n, \end{aligned} \tag{55}$$

where

$$W_n = Y^{-1} \sum_m (m^n), \tag{56}$$

and

$$\begin{aligned} &\mathrm{tr}S_{zi}^n S_{zj}^n (ik\cdots jl\cdots) \\ &\quad = \sum_{m_i} \sum_{m_k} \cdots \sum_{m_j} \sum_{m_l} \cdots (m_i^n m_j^n \delta_{m_i,m_k} \cdots \delta_{m_i,m_j} \delta_{m_i,m_l} \cdots) \\ &\quad = \sum_{m_i} (m_i^{2n}) = YW_{2n}. \end{aligned} \tag{57}$$

Let us express any element in S_N, say P, as a product of independent cycles:

$$P = (abc\cdots d)(ijk\cdots l)\cdots(xy\cdots z). \tag{58}$$

P may be considered as a direct product of each cycle,

$$P = (abc\cdots d) \times (ijk\cdots l) \times \cdots \times (xy\cdots z). \tag{59}$$

By the trace property of the direct product,

$$\mathrm{tr}(A \times B) = (\mathrm{tr}A)(\mathrm{tr}B), \tag{60}$$

it follows from Eqs. (54), (55), and (57) that

$$\mathrm{tr}S_{zi}^n P = W_n \mathrm{tr}P, \tag{61}$$

and

$$\begin{aligned} \mathrm{tr}S_{zi}^n S_{zj}^n P &= W_{2n} \mathrm{tr}P, && \text{if } i = j \text{ or } i, j \text{ belong to the same cycle in } P, \\ &= W_n^2 \mathrm{tr}P, && \text{if } i, j \text{ belong to different independent cycles in } P. \end{aligned} \tag{62}$$

For the case $n = 1$, $W_1 = 0$ and $W_2 = X/3$. If the element P belongs to the class $k = (a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_p^{\alpha_p})$, then, from Eqs. (54), (60), and (62),

$$\mathrm{tr}P_k = Y^{\alpha_1+\alpha_2+\cdots+\alpha_p} \tag{63}$$

and

$$\begin{aligned} \mathrm{tr}P_k Q^2 &= \sum_i \sum_j \mathrm{tr}S_{zi} S_{zj} P_k \\ &= (X/3)(\alpha_1 a_1^2 + \alpha_2 a_2^2 + \cdots + \alpha_p a_p^2) Y^{\alpha_1+\alpha_2+\cdots+\alpha_p}, \end{aligned} \tag{64}$$

where the summations are from 1 to N.

It also follows from Eqs. (41), (42) and Eqs. (54), (60) that

$$\begin{aligned} \mathrm{tr}P_{ij}P &= Y\mathrm{tr}P, && \text{if } i, j \text{ belong to the same cycle in } P, \\ &= Y^{-1}\mathrm{tr}P, && \text{if } i, j \text{ belong to different independent cycles in } P. \end{aligned} \tag{65}$$

V. DIAGRAMMATIC METHOD

A. Preliminary Remarks

The diagrammatic method has been extensively used in deriving high temperature series for various spin Hamiltonians. In this section, the procedure previously used for the Heisenberg Hamiltonian by Rushbrooke and Wood[11] will be modified for the exchange interaction Hamiltonian. The labor of evaluating series

coefficients by this method is much greater than the labor involved in the previously described cluster expansion method. However, there are a number of important results that can be directly proved by the diagrammatic technique which are not at all obvious from a consideration of the cluster expansion method.

B. Zero-Field Partition Function

For the exchange interaction model $\mathcal{P}^n$ in Eq. (12) is a sum of products ΠP_{ij} and each product contains n factors P_{ij}. There is a correspondence between products in $\mathcal{P}^n$ and diagrams of n lines on the lattice. For each of the n factors P_{ij} in the product, when we draw a straight line connecting lattice sites i and j, we obtain a diagram of n lines. The diagrams may be connected or disconnected, and may have more than one line joining a pair of points.

Following Rushbrooke and Wood,[11] we can write

$$\langle \mathcal{P}^n \rangle = \sum_{i=n} [D_i]\langle D_i \rangle, \tag{66}$$

where $\sum_{i=n}$ sums over all diagrams D_i of n lines. $[D_i]$ is the number of times that the diagram D_i will occur on the lattice. $\langle D_i \rangle$ is the weighting factor, or the contribution of the diagram D_i to $\langle \mathcal{P}^n \rangle$. For a diagram D having n lines between p points, following Rushbrooke and Wood,

$$\langle D \rangle = Y^{-p} \sum_{\text{perm}} \text{tr}(\,)(\,)\cdots(\,), \tag{67}$$

where each bracket is a Schrödinger exchange operator P_{ij} which corresponds to the line joining points i and j in the diagram D and $\sum_{\text{perm}}$ sums over all different permutations in the order of appearence of the brackets.

The occurrence factor of a diagram on a lattice has the same meaning as the lattice constant of a graph. The only difference is that for graphs we speak of diagrams which have at most one line connecting a pair of points. Some typical examples of the relation between the occurrence factors of diagrams and lattice constants of graphs are shown in Fig. 2. As mentioned in Sec. III, lattice constants of disconnected graphs can be expressed in terms of lattice constants of connected graphs. Let $\Lambda_i\Phi$ signify that part of Φ which is of ith order in the lattice constants of connected graphs. We can then write

$$\langle \mathcal{P}^n \rangle = \Lambda_1\langle \mathcal{P}^n \rangle + \Lambda_2\langle \mathcal{P}^n \rangle + \cdots . \tag{68}$$

As shown in the section on the cluster expansion method, Eq. (35), any extensive quantity, such as $\ln Z$ or χ is of first order in lattice constants of connected graphs. Hence by Eq. (12),

$$\ln Z = N \ln Y + \sum_{n=1}^{\infty} \frac{K^n}{n!} \Lambda_1\langle \mathcal{P}^n \rangle. \tag{69}$$

Those parts of higher order in lattice constants of connected graphs will cancel exactly when transformed from Z to $\ln Z$. Comparing Eqs. (14) and (69), we have

$$e_n = \Lambda_1\langle \mathcal{P}^n \rangle = \sum_{i=n} \Lambda_1[D_i]\langle D_i \rangle. \tag{70}$$

In the limit $N \to \infty$, lattice constants of connected graphs will be directly proportional to N. Λ_1 then means nothing more than "the part proportional to N."

C. Low-Field Susceptibility

When a finite external magnetic field is applied, the partition function can be written in the form

$$\ln Z = N \ln Y + \sum_{n=1}^{\infty} \frac{K^n}{n!} \Lambda_1\langle (\mathcal{P} + \alpha Q)^n \rangle. \tag{71}$$

Since $\mathcal{P}$ and Q commute and $\langle \mathcal{P}^n Q \rangle = 0$, it is easy to show that

$$\chi = \lim_{H\to 0} k_B T \frac{\partial^2}{\partial H^2} \ln Z = \frac{g^2\mu^2}{k_B T} \sum_{n=0}^{\infty} \frac{K^n}{n!} \Lambda_1\langle \mathcal{P}^n Q^2 \rangle. \tag{72}$$

For $n = 0$, $\langle Q^2 \rangle = NX/3$. Comparing Eqs. (24) and (72) yields

$$a_n = (3/X)\Lambda_1\langle \mathcal{P}^n Q^2 \rangle. \tag{73}$$

The situation in calculating $\langle \mathcal{P}^n Q^2 \rangle$ is similar to that involved in $\langle \mathcal{P}^n \rangle$. We again have a correspondence between terms ΠP_{ij} produced on expanding $\mathcal{P}^n$ and diagrams of n lines on the lattice, but now, besides the n lines, a diagram will contain two crosses. These are denoted +, and are placed on those sites from which we have extracted terms like $S_{zi}S_{zj}$ from Q^2. When two crosses coincide, we speak of a double cross, denoted *.

There are five kinds of ways in which two crosses may be added to a diagram D:

(a) D^*, a double cross superposed on D,

(b) D^{++}, two crosses superposed on different sites of D,

(c) D^{+} +, one cross superposed on D, another not on D,

(d) D *, a double cross not superposed on D,

(e) D + +, two separated crosses not superposed on D.

Since $\langle S_{zi} \rangle = 0$ and $\langle S_{zi}^2 \rangle = X/3$, by the trace property of the direct product, we immediately find that diagrams (c) and (e) have zero contributions to $\langle \mathcal{P}^n Q^2 \rangle$ and that diagram (d) contributes an amount $(X/3)\langle D \rangle$ to $\langle \mathcal{P}^n Q^2 \rangle$. Let p be the number of points contained in the diagram D. Then we have that the occurrence factor, for diagram (d), $[D\,*] = [D][\,\cdot\,] - p[D]$. Therefore, as far as terms of first order in lattice constants of connected graphs are concerned, it is equivalent to say that the diagram (d) has occurrence factor $[D]$ associated with the weighting factor $-(pX/3)\langle D \rangle$.

For diagram (a) the double cross can be superposed on any of the p points. For diagram (b) there are $p(p-1)$ ways to superpose the two crosses on D. If

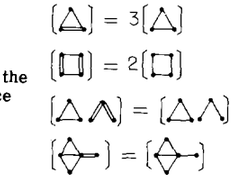


FIG. 2. Examples of the relation between the occurrence factors of diagrams and lattice constants of graphs.

we sum up these p^2 diagrams, with occurrence factor $[D]$, their contributions to $\langle \mathcal{O}^n Q^2\rangle$, denoted $\langle \overline{DQ}^2\rangle$, will be

$$\langle \overline{DQ}^2\rangle = Y^{-p} \sum_{\text{perm}} \text{tr}(\ )(\ )\cdots(\ )(\sum_{i=1}^{p} S_{zi})^2. \tag{74}$$

Therefore, including all five kinds of diagrams, with occurrence factor $[D]$, the weighting factor of the diagram D to $\Lambda_1\langle \mathcal{O}^n Q^2\rangle$, denoted $\langle DQ^2\rangle$, will be

$$\langle DQ^2\rangle = \langle \overline{DQ}^2\rangle - (pX/3)\langle D\rangle. \tag{75}$$

Hence

$$a_n = (3/X)\sum_{l_i=n} \Lambda_1[D_i][\langle \overline{D_iQ}^2\rangle - p_i(X/3)\langle D_i\rangle]. \tag{76}$$

where p_i is the number of points in D_i. By rewriting the products ΠP_{ij} as products of independent cycles, the traces in Eqs. (67) and (74) can be obtained from Eqs. (63) and (64). $\langle D_i\rangle$ and $\langle D_iQ^2\rangle$ are then determined. Note that for the present Hamiltonian these five kinds of diagrams can be considered together and the calculation of the coefficients a_n is greatly simplified. For the Heisenberg Hamiltonian, weighting factors of these five kinds of diagrams have to be considered separately.

D. Further Theorems

If we express $\Lambda_1[D_i]$ in terms of lattice constants of connected graphs, denoted $[g_i]$, we can rewrite e_n and a_n as follows:

$$e_n = \sum_i [g_i] e_n(g_i) \tag{77}$$

and

$$a_n = \sum_i [g_i] a_n(g_i), \tag{78}$$

where $\sum_i$ sums over all connected graphs g_i. We then prove the following:

Theorem 4:

$$e_n(g_i) = a_n(g_i) = 0, \quad \text{if } n < \ell_i. \tag{79}$$

Proof: This is obvious since those diagrams with occurrence factor containing $[g_i]$ must have ℓ_i lines or more, and they contribute to e_n and a_n for $n \geq \ell_i$.

Theorem 5: If we define cycles of diagrams similar to those for graphs presented in Sec. III, then for diagrams containing n lines and c independent cycles their contributions to e_n and a_n contain terms Y^{-n+2k}, where k ranges from 0 to c, but $2k \geq n$.

Proof: As mentioned in Sec. IV, for any group element P, $(fg)P$ has one more cycle than P if f, g be long to the same cycle in P and has one less than P if f, g belong to different independent cycles in P. Assume that the diagram contains p points. The identity element will have p different independent cycles, namely, (1) (2) $\cdots$ (p). Consider one of the products

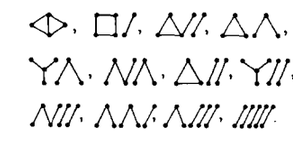


FIG. 3. Let the first graph be g_i. All diagrams of five lines with occurrence factors containing $[g_i]$ are shown. This illustrates the fact that diagrams of ℓ_i lines with occurrence factors containing $[g_i]$ consist of subgraphs of g_i.

of n factors (ij) occurring in Eqs. (67) and (74), and multiply successively to the identity element, first, the nth factor, then the $(n-1)$th factor, etc., and finally the first factor. Since each multiplication either decrease or increase the number of different independent cycles by 1, the resultant product will contain $p - n + 2k$ different independent cycles. Here k is the number of times that i, j happen to occur in the same cycle in the product which is to be multiplied by (ij). This can happen only when i, j are joined by paths other than the line ij, or ij must be a line of a cycle. Therefore, $k \leq c$. Also, the number of independent cycles may not be greater than p, $2k \leq n$. Equation (63) says that for any permutation containing t different independent cycles its trace in the $(2S+1)^p$-dimensional matrix representation is Y^t. Theorem 5 then follows from a consideration of Eqs. (63), (67), (70) and Eqs. (64), (74), and (76).

A consequence of this theorem is that $e_n(g_i)$ and $a_n(g_i)$ contain terms $Y^{-n}, Y^{-n+2}, \cdots$, which we can write as

$$e_n(g_i) = e_n^{(n)}(g_i)Y^{-n} + e_n^{(n-2)}(g_i)Y^{-n+2} + \cdots + e_n^{(1)}(g_i)Y^{-1} \quad [\text{or } e_n^{(0)}(g_i)] \tag{80}$$

and

$$a_n(g_i) = a_n^{(n)}(g_i)Y^{-n} + a_n^{(n-2)}(g_i)Y^{-n+2} + \cdots + a_n^{(1)}(g_i)Y^{-1} \quad [\text{or } a_n^{(0)}(g_i)], \tag{81}$$

Theorem 6:

$$e_n^{(-n+2k)}(g_i) = a_n^{(-n+2k)}(g_i) = 0, \quad \text{for } k > c_i + n - \ell_i, \tag{82}$$

where c_i is the number of independent cycles and ℓ_i is the number of lines in the connected graph g_i.

Proof: Consider first the case that $n = \ell_i$. Diagrams of ℓ_i lines which have an occurrence factor containing $[g_i]$ are those which consist of subgraphs of g_i, and hence cannot have more cycles than g_i. For example, let g_i be the connected graph in Fig. 3. Other diagrams of five lines with occurrence factor containing $[g_i]$ are those disconnected graphs shown in the figure. Hence from Theorem 5, Theorem 6 is proved for the case $n = \ell_i$. When $n > \ell_i$, we can superpose the additional $n - \ell_i$ lines on g_i. Each line superposed on g_i is equivalent to forming an additional cycle in the resultant diagram, and the maximum number of independent cycles will be $c_i + n - \ell_i$ in some of the n diagrams which have occurrence factors containing $[g_i]$. This then completes the proof of Theorem 6.

Theorem 7:

$$a_n^{(0)}(g_i) = 0 \quad \text{for all } g_i. \tag{83}$$

Proof: Consider a diagram D of p points and n lines which has occurrence factor containing $[g_i]$. Those products ΠP_{ij} in Eqs. (67) and (74) which contribute to $a_n^{(0)}$ must be equal to the identity operator. Since $\text{tr}\mathbf{I}Q^2 = pX/3$, Eq. (83) follows from Eq. (75).

Theorem 8:

$$e_n^{(1)}(g_i) = \tfrac{1}{2} a_n^{(1)}(g_i) \quad \text{for all } g_i. \tag{84}$$

Proof: Those products in Eqs. (67) and (74) which contribute to $e_n^{(1)}$ and $a_n^{(1)}$ must contain $p-1$ cycles, or belong to the class $1^{p-2}2$. From Eqs. (63) and (64), for permutations P belonging to the class $1^{p-2}2$,

$$\mathrm{tr}P\left(\sum_{i=1}^{p} S_{zi}\right)^2 = \frac{X}{3}(p+2)Y^{p-1} \tag{85}$$

and

$$\mathrm{tr}P = Y^{p-1}. \tag{86}$$

From Eqs. (67), (74), and (75), the contributions of these products to $\langle D\rangle$ and $(3/X)\langle DQ^2\rangle$ will be Y^{-1} and $2Y^{-1}$, respectively, Theorem 8 then follows.

VI. CALCULATIONS OF SERIES COEFFICIENTS

A. Clusters With Up to Seven Sites

From the discussion in Sec. III we know that in order to obtain the high temperature series for general lattices up to the seventh power in $K(=J/k_BT)$, we must calculate the corresponding series for clusters having up to seven lines. For computational convenience we group these clusters into two categories, those having up to seven points and those containing eight points. Consider first clusters with up to seven points and seven lines. They are labeled in the graph dictionary order such that Eq. (33) is satisfied.[12]

To calculate the coefficients in the zero-field partition function and the low-field susceptibility series for finite clusters we make use of Eqs. (15) and (25), in which the quantities $\mathrm{tr}\mathcal{P}^n$ and $\mathrm{tr}\mathcal{P}^nQ^2$ for $n \leq 7$ are obtained from Theorem 2 by setting R equal to the unit matrix and Q^2, respectively. Now $\mathcal{P}^{(\Gamma)}$ in Eq. (50) is a sum of Schrödinger exchange operators which correspond to the lines in each cluster. h_k and $\chi_k^{(\nu)}$ are available in a number of texts which deal specifically with the symmetric group.[13] The values of $\mathrm{tr}P_k$ and $\mathrm{tr}P_kQ^2$ can be calculated by Eqs. (63) and (64). Corresponding to $\mathcal{P}^{(\Gamma)}, \mathcal{P}^{(\nu)}$ is a sum of interchanges in the irreducible representation ν. The explicit form of the permutation matrices in any irreducible representation can be obtained by the technique introduced by Yamanouchi.[14]

For two conjugate representations $\mathrm{tr}\mathcal{P}^n$ are related to each other by Eq. (53) and hence only one of the traces need by calculated. The quantities $\mathrm{tr}(\mathcal{P}^{(\nu)})^n$ were calculated on a computer for all required clusters with up to seven points. The size of the greatest matrix involved in the calculation is of dimension 35×35.

B. Clusters Containing Eight Sites

For clusters with seven lines and eight sites, we can still evaluate the high temperature series following the procedure just described. However, for the symmetric group of degree eight some of the irreducible representations are of dimension 70×70 and the trace calculation on a computer would have been too expensive for us. From Eq. (26) we have that clusters with seven lines and eight points contain no cycles. It then follows immediately from Theorem 4 that for these clusters $e_n(g_i) = a_n(g_i) = 0$ for $n \leq 6$ and from Theorem 6 that $e_7(g_i)$ and $a_7(g_i)$ can be written in the form

$$e_7(g_i) = e_7^{(7)}(g_i)Y^{-7} \tag{87}$$

and

$$a_7(g_i) = a_7^{(7)}(g_i)Y^{-7}. \tag{88}$$

For $S = \frac{1}{2}$ the exchange interaction Hamiltonian is identical to the Heisenberg Hamiltonian, and the high temperature series for these two models are the same. It is known that for the Heisenberg model graphs with n lines containing no cycles will not contribute to e_n and the only graph with n lines containing no cycles which contributes to a_n is the simple chain.[15] Therefore, $e_7(g_i) = a_7(g_i) = 0$ for all clusters of eight points with the exception that for the simple chain $a_7(g_i) = 10\,080Y^{-7}$. This value was obtained from Eq. (88) together with the results of Domb and Wood[15]; for the case of spin $\frac{1}{2}$ and for the simple chain of eight points they obtained $a_7(g_i) = 10\,080 \times 2^{-7}$.

VII. SERIES RESULTS

The simplest way to represent the series coefficients e_n and a_n [see Eqs. (14) and (24)] for any finite cluster or crystal lattice is in the form of Eqs. (77) and (78) together with Eqs. (80) and (81). The number of graphs which contribute to $e_1 - e_7$ are 1, 1, 2, 4, 7, 15, and 29, respectively, while the numbers of graphs that contribute to $a_1 - a_7$ are 1, 2, 4, 8, 16, 35, 82, respectively. The values of $e_n^{(m)}(g_i)$ and $a_n^{(m)}(g_i)$ for these graphs are given in Appendix D of Ref. 12.

For regular lattices, e.g., body-centered cubic and face-centered cubic lattices with $N \to \infty$, the lattice constants of connected graphs are proportional to N, and lattice constants of open graphs can be expressed in terms of the lattice constants of closed graphs and the coordination number of the lattice. We rewrite Eqs. (14) and (24) in the form

$$\ln Z/N = \ln Y + \sum_{n=2}^{\infty} e_n K^n, \tag{89}$$

and

$$\chi = \frac{C}{T}\left(1 + \sum_{n=1}^{\infty} a_n K^n\right), \tag{90}$$

where $C = N(g\mu)^2X/3k_B$. Note that the numbers $n!$ do not appear in these expressions and that e_1 has been set equal to zero by adjusting the zero of energy such that the internal energy is equal to zero at infinite temperature, that is, the constant Y^{-1} is subtracted from P_{ij} to make $\mathcal{H}$ traceless. The coefficients e_n and a_n will be written in the form

$$e_n = e_n^{(n)}Y^{-n} + e_n^{(n-2)}Y^{-n+2} + \cdots + e_n^{(1)}Y^{-1} \quad (\text{or } e_n^{(0)}), \tag{91}$$

$$a_n = a_n^{(n)}Y^{-n} + a_n^{(n-2)}Y^{-n+2} + \cdots + a_n^{(2)}Y^{-2} \quad (\text{or } a_n^{(1)}Y^{-1}). \tag{92}$$

The various quantities $e_n^{(m)}$ and $a_n^{(m)}$ are then given below:

$$e_2^{(0)} = \tfrac{1}{4}(\sigma+1), \quad e_2^{(2)} = -\tfrac{1}{4}(\sigma+1),$$

$$e_3^{(1)} = -\tfrac{1}{6}[(\sigma+1) - 6p_3],$$

$$e_3^{(3)} = \tfrac{1}{6}[(\sigma+1) - 6p_3], \quad e_4^{(0)} = -\tfrac{1}{24}[(\sigma+1)^2 - 12p_3],$$

$$e_4^{(2)} = \tfrac{1}{24}[(\sigma+1)(\sigma+4) - 84p_3 + 24p_4],$$

$$e_4^{(4)} = -\tfrac{1}{8}[(\sigma+1) - 24p_3 + 8p_4],$$

$$e_5^{(1)} = \tfrac{1}{60}[(\sigma+1)(5\sigma+4) - 15p_3(2\sigma+7) + 40p_4 + 30p_{5a}],$$

$$e_5^{(3)} = -\tfrac{1}{12}[(\sigma+1)(\sigma+2) - 3p_3(2\sigma+31) + 56p_4 - 12p_5 + 30p_{5a}],$$

$$e_5^{(5)} = \tfrac{1}{10}[(\sigma+1) - 60p_3 + 40p_4 - 10p_5 + 20p_{5a}],$$

$$e_6^{(0)} = \tfrac{1}{720}[(\sigma+1)(11\sigma^2+20\sigma+8) - 36p_3(7\sigma+4) + 72p_4 + 132p_{5a} + 96p_{6d}],$$

$$e_6^{(2)} = -\tfrac{1}{720}[(\sigma+1)(11\sigma^2+110\sigma+68) - 108p_3 \times (19\sigma+23) + 24p_4(20\sigma+93) - 600p_5 + 3132p_{5a} - 360p_{6a} - 360p_{6b} + 240p_{6c} + 1536p_{6d}],$$

$$e_6^{(4)} = \tfrac{1}{24}[(\sigma+1)(3\sigma+4) - 6p_3(10\sigma+51) + 8p_4 \times (2\sigma+39) - 140p_5 + 364p_{5a} + 24p_6 - 60p_{6a} - 60p_{6b} + 8p_{6c} + 192p_{6d}],$$

$$e_6^{(6)} = -\tfrac{1}{12}[(\sigma+1) - 114p_3 + 120p_4 - 60p_5 + 132p_{5a} + 12p_6 - 24p_{6a} - 24p_{6b} + 72p_{6d}],$$

$$e_7^{(1)} = -\tfrac{1}{5040}[(\sigma+1)(245\sigma^2+385\sigma+136) - 42p_3 \times (37\sigma^2+288\sigma+18) + 560p_4(5\sigma+6) - 910p_5 + 28p_{5a}(77\sigma+312) - 1260p_{6a} - 1456p_{6b} + 2352p_{6c} + 3864p_{6d} - 798p_{7c} - 840p_{7g} - 644p_{7h}],$$

$$e_7^{(3)} = \tfrac{1}{720}[(\sigma+1)(35\sigma^2+175\sigma+88) - 6p_3 \times (37\sigma^2+1128\sigma+348) + 80p_4(41\sigma+78) - 10p_5(60\sigma+343) + 52p_{5a}(29\sigma+234) + 720p_6 - 3060p_{6a} - 3808p_{6b} + 2256p_{6c} + 9192p_{6d} + 360p_{7a} + 360p_{7b} - 1314p_{7c} - 240p_{7e} + 360p_{7f} - 3000p_{7g} - 1532p_{7h}],$$

$$e_7^{(5)} = -\tfrac{1}{12}[2(\sigma+1)^2 - 3p_3(30\sigma+59) + 48p_4(\sigma+7) - 5p_5(2\sigma+47) + 10p_{5a}(2\sigma+59) + 84p_6 - 192p_{6a} - 216p_{6b} + 32p_{6c} + 576p_{6d} - 12p_7 + 30p_{7a} + 30p_{7b} - 68p_{7c} - 4p_{7e} + 30p_{7f} - 120p_{7g} - 96p_{7h}],$$

$$e_7^{(7)} = \tfrac{1}{14}[(\sigma+1) - 168p_3 + 280p_4 - 210p_5 + 476p_{5a} + 84p_6 - 168p_{6a} - 182p_{6b} + 504p_{6d} - 14p_7 + 28p_{7a} + 28p_{7b} - 56p_{7c} + 28p_{7f} - 84p_{7g} - 84p_{7h}], \tag{93}$$

and

$$a_1^{(1)} = (\sigma+1), \quad a_2^{(2)} = (\sigma+1)(\sigma-1),$$

$$a_3^{(1)} = -\tfrac{1}{3}[(\sigma+1) - 6p_3],$$

$$a_3^{(3)} = [(\sigma+1)(\sigma-1)^2 - 18p_3], \quad a_4^{(2)} = -\tfrac{1}{12} \times [(\sigma+1)(3\sigma-8) - 48p_3(\sigma-5) - 56p_4],$$

$$a_4^{(4)} = [(\sigma+1)(\sigma-1)^3 - 18p_3(2\sigma-5) - 32p_4],$$

$$a_5^{(1)} = \tfrac{1}{30}[(\sigma+1)(5\sigma+4) - 15p_3(2\sigma+7) + 40p_4 + 30p_{5a}],$$

$$a_5^{(3)} = -\tfrac{1}{6}[(\sigma+1)(\sigma^2-4\sigma+6) - 6p_3(6\sigma^2-34\sigma+85) - 8p_4(7\sigma-36) - 50p_5 + 204p_{5a}],$$

$$a_5^{(5)} = [(\sigma+1)(\sigma-1)^4 - 54p_3(\sigma^2-4\sigma+5) - 64p_4 \times (\sigma-3) - 50p_5 + 132p_{5a}],$$

$$a_6^{(2)} = \tfrac{1}{360}[(\sigma+1)(53\sigma^2-169\sigma-136) - 18p_3 \times (40\sigma^2-182\sigma-347) - 48p_4(3\sigma+115) + 1380p_5 + 24p_{5a}(30\sigma-347) + 756p_{6a} + 828p_{6b} - 216p_{6c} - 4944p_{6d}],$$

$$a_6^{(4)} = -\tfrac{1}{12}[(\sigma+1)(\sigma^3-6\sigma^2+14\sigma-16) - 12p_3 \times (8\sigma^3-52\sigma^2+138\sigma-219) - 8p_4 \times (21\sigma^2-124\sigma+344) - 20p_5(10\sigma-57) + 8p_{5a}(102\sigma-515) - 156p_6 + 636p_{6a} + 636p_{6b} + 64p_{6c} - 2544p_{6d}],$$

$$a_6^{(6)} = [(\sigma+1)(\sigma-1)^5 - 18p_3(4\sigma^3-21\sigma^2+42\sigma-34) - 32p_4(3\sigma^2-14\sigma+21) - 50p_5(2\sigma-7) + 24p_{5a}(11\sigma-40) - 72p_6 + 192p_{6a} + 188p_{6b} + 72p_{6c} - 624p_{6d}],$$

$$a_7^{(1)} = -\tfrac{1}{2520}[(\sigma+1)(245\sigma^2+385\sigma+136) - 42p_3 \times (37\sigma^2+228\sigma+18) + 560p_4(5\sigma+6) - 910p_5 + 28p_{5a}(77\sigma+312) - 1260p_{6a} - 1456p_{6b} + 2352p_{6c} + 3864p_{6d} - 798p_{7c} - 840p_{7g} - 644p_{7h}],$$

$$a_7^{(3)} = \tfrac{1}{360}[(\sigma+1)(46\sigma^3-205\sigma^2+347\sigma+264) - 6p_3(180\sigma^3-1029\sigma^2+3411\sigma+1292) - 4p_4(192\sigma^2-1076\sigma-7012) + 10p_5 \times (28\sigma-1495) + 4p_{5a}(270\sigma^2-1694\sigma+15589) + 2976p_6 + 12p_{6a}(126\sigma-1243) + 4p_{6b} \times (414\sigma-4531) - 16p_{6c}(27\sigma-406) - 48p_{6d} \times (206\sigma-1135) + 1388p_{7a} + 1454p_{7b} - 7820p_{7c} + 320p_{7d} - 348p_{7e} + 1292p_{7f} - 12540p_{7g} - 8272p_{7h}],$$

$$a_7^{(5)} = \tfrac{1}{6}[(\sigma+1)(\sigma^3-5\sigma^2+10\sigma-10) + 6p_3 \times (10\sigma^4-74\sigma^3+218\sigma^2-327\sigma+339) + 8p_4 \times (14\sigma^3-95\sigma^2+279\sigma-554) + 10p_5 \times (15\sigma^2-98\sigma+307) - 4p_{5a} \times (153\sigma^2-962\sigma+2399) + 12p_6(13\sigma-83) - 12p_{6a}(53\sigma-267) - 12p_{6b}(53\sigma-295) - 8p_{6c} \times (8\sigma-51) + 48p_{6d}(53\sigma-228) + 112p_7 - 456p_{7a} - 456p_{7b} + 1392p_{7c} - 96p_{7d} - 52p_{7e} - 456p_{7f} + 1992p_{7g} + 1796p_{7h}],$$

$$a_7^{(7)} = [(\sigma+1)(\sigma-1)^6 - 18p_3(5\sigma^4 - 32\sigma^3 + 84\sigma^2 - 110\sigma + 62) - 128p_4(\sigma^3 - 6\sigma^2 + 14\sigma - 14) - 50p_5 \times (3\sigma^2 - 16\sigma + 28) + 4p_{5a}(99\sigma^2 - 546\sigma + 971) - 144p_6(\sigma - 4) + 384p_{6a}(\sigma - 4) + 2p_{6b} \times (188\sigma - 777) + 144p_{6c}(\sigma - 4) - 1248p_{6d} \times (\sigma - 4) - 98p_7 + 256p_{7a} + 252p_{7b} - 580p_{7c} + 72p_{7d} + 96p_{7e} + 260p_{7f} - 588p_{7g} - 844p_{7h}]. \quad (94)$$

Here the p_{nx} are lattice constants for closed graphs per lattice site and $q = \sigma + 1$ is the coordination number of the lattice. The values p_{nx} and q for various crystal lattices are well known.[8] For convenience, the coefficients e_n for the cubic lattices and for several spin values are shown in Appendix A, and values of a_n for the cubic lattices and a number of two dimensional lattices are given in Appendix B. It is generally found that, similarly to the Heisenberg model,[11] the coefficients in these series expansions increase in smoothness as q increase. However, the series coefficients for the present model are much more irregular than those of the Heisenberg model, especially for large values of spin.

VIII. CHECKING PROCEDURES

Since there are numerous possibilities for errors to be made in the computations, it is important to be able to check the general expressions for the coefficients $e_n^{(m)}$ and $a_n^{(m)}$ before using them to estimate critical parameters. We have considered a large number of finite clusters of eight lines. For these clusters we can calculate the zero-field partition function and the low-field susceptibility series from Eqs. (93) and (94) by substituting the appropriate lattice constants. Comparing the results with those obtained from a direct machine calculation as described in Section VI yields full agreement up to e_7 and a_7 in all cases.

By observing the general expressions for $e_n^{(m)}$ and $a_n^{(m)}$ that we obtained [Eqs. (93) and (94)], we see that

$$\sum_m e_n^{(m)} = 0, \quad m = n, n-2, n-4, \cdots, 1 \text{ (or } 0), \quad (95)$$

and

$$e_n^{(1)} = \tfrac{1}{2} a_n^{(1)}. \quad (96)$$

These two equations hold for all n and provide an additional check on our results. Equation (96) has been proved in Theorem 8, while Eq. (95) is a necessary consequence of the fact that, for $S = 0$, $\ln Z = 0$.

As a final check, we see that when we set S equal to $\frac{1}{2}$ and 1, respectively, our general results for e_n and a_n reduce exactly to those obtained previously by Domb and Wood[15] and by Allan and Betts,[16] respectively.

IX. ANALYSIS OF SERIES

A. Estimates of Curie Temperatures and Critical Indices

The Curie temperature T_C and the critical index γ for the susceptibility series are defined by

$$\chi \sim (T - T_C)^{-\gamma}, \quad \text{for } T \to T_C^+, \quad (97)$$

or

$$\sim (K_C - K)^{-\gamma}, \quad \text{for } K \to K_C^-, \quad (98)$$

where $K_C = J/k_B T_C$. For the face-centered cubic lattice T_C and γ were first estimated by means of the ratio method. From Eq. (98) the coefficients a_n in Eq. (90) have the property that for large n

$$\frac{a_n}{a_{n-1}} = \left(\frac{a_n}{a_{n-2}}\right)^{1/2} \to \frac{1}{K_C}\left(1 + \frac{\gamma - 1}{n}\right). \quad (99)$$

If we plot the two sets of values a_n/a_{n-1} and $(a_n/a_{n-2})^{1/2}$ vs $1/n$, each of the plots tends to a straight line as n increases and intersects with $1/n = 0$ at K_C^{-1} with slope $(\gamma - 1)K_C^{-1}$. If we plot $(a_n)^{1/n}$ vs $1/n$, the plot also intersects with $1/n = 0$ at K_C^{-1} but does not approach the intersection in a simple linear fashion for large n. Figure 4 illustrates these plots for the face-centered cubic lattice and for $S = 1$. The Curie temperature and the critical index can also be estimated by the method of Padé approximants.[17] The $[M, N]$ Padé approximant to a polynomial $f(z)$ is given by $P(z)/Q(z)$, where $P(z)$ and $Q(z)$ are polynomials of degree N and M, respectively, such that $Q(0) = 1$ and $f(z)$ agrees with the expansion of $P(z)/Q(z)$ for the first $M + N + 1$ terms.

If the divergence of χ at K_C is in the form shown in Eq. (98), then

$$(\chi)^{1/\gamma} \sim (K_C - K)^{-1}, \quad (100)$$

$$\frac{\chi'}{\chi} = \frac{d}{dK}\ln\chi = \frac{\gamma}{K_C - K}, \quad (101)$$

and

$$\frac{\chi''\chi}{(\chi')^2} = \frac{d}{dK}\left(\ln\frac{d\chi}{dK}\right) \Big/ \frac{d}{dK}(\ln\chi) = 1 + \frac{1}{\gamma}. \quad (102)$$

These equations suggest four methods of determining K and/or γ from χ:

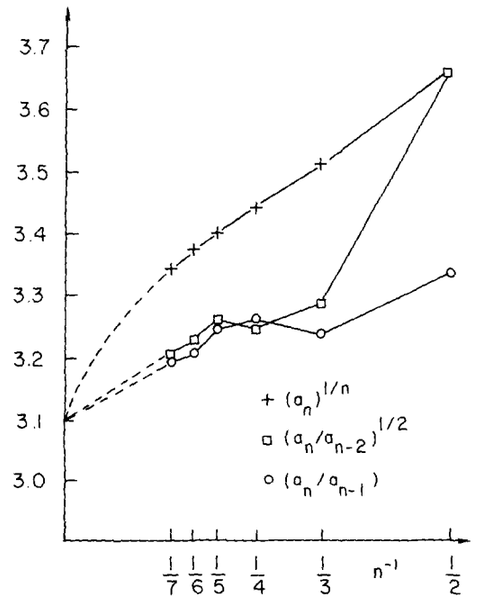


FIG. 4. Plots of $(a_n)^{1/n}$, $(a_n/a_{n-2})^{1/2}$, and (a_n/a_{n-1}) n^{-1} for the face-centered cubic lattice, $S = 1$.

(a) Choosing γ, K_C can be presented by appropriate poles of the Padé approximants to $(\chi)^{1/\gamma}$.

(b) Choosing K_C, γ can be obtained by evaluating Padé approximants to $(K_C - K)\chi'/\chi$ at $K = K_C$.

(c) For a Padé approximant to χ'/χ the appropriate pole gives K_C and the residue at this pole gives $-\gamma$.

(d) Evaluating Padé approximants to $\chi''\chi/(\chi')^2$ at $K = K_C$ gives $1 + 1/\gamma$.

For the face-centered cubic lattice and for $S = \frac{1}{2} - 3$ we have estimated K_C and γ by all the four methods. In method (a) instead of tabulating the various approximants in Padé table for several values of γ we have plotted the various approximants of K_C as functions of γ in the $K_C - \gamma$ plane. The proper values of γ and K_C are easily obtained from these curves in the region in which the various approximants coalesce. Figure 5 shows several of these curves for the face-centered cubic lattice and for $S = 1$. In method (b) the approximants of γ as functions of K_C are also plotted in the $K_C - \gamma$ plane. These curves are roughly parallel to and close to curves obtained from method (a) in the region of interest. The various approximants in methods (c) and (d) are rather irregular. However, they are not inconsistent with what methods (a) and (b) yield. The estimates of K_C and γ from Padé analysis are also in agreement with estimates based on the various ratio tests. The final estimates of γ and T_C are given in Figures 6 and 7, respectively. These results can be simply described by the equations

$$k_B T_C/J = 1.19 + 5.70Y^{-1} \tag{103}$$

and

$$\gamma = 0.48 + 2.16Y^{-1}. \tag{104}$$

Equations (103) and (104) hold for all S with the exception that $\gamma = 1.41 \pm 0.02$ for $S = \frac{1}{2}$.

For the body-centered cubic lattice and the simple cubic lattice the series coefficients are in general too irregular to estimate both K_C and γ either by ratio tests or by the Padé approximant method. However, within our precision, γ seems to be the same for all of the cubic lattices for each S. If the values of γ for the body-centered cubic and the simple cubic lattices are chosen to be the same as those of the face-centered cubic lattice, then an estimate of K_C from Padé approximants to $(\chi)^{1/\gamma}$ suggests that for all of the cubic lattices T_C can be described to within a few percent by

$$k_B T_C/J = 0.547(q - 1.6)(Y^{-1} + 0.21). \tag{105}$$

The estimates of T_C for the body-centered cubic and

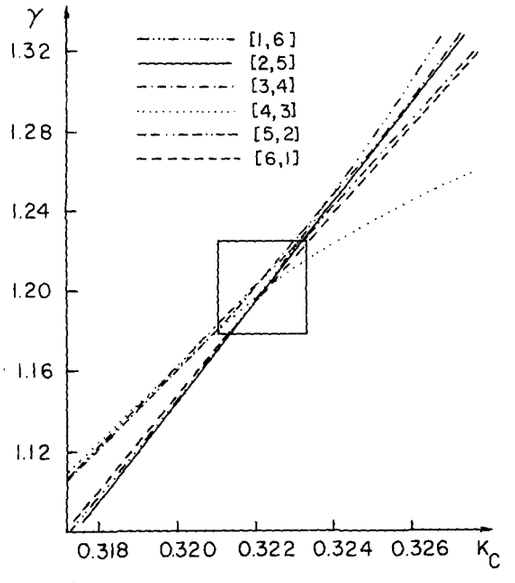


FIG. 5. Plots of several Padé approximants of K_C as functions of γ for the fcc lattice, $S = 1$.

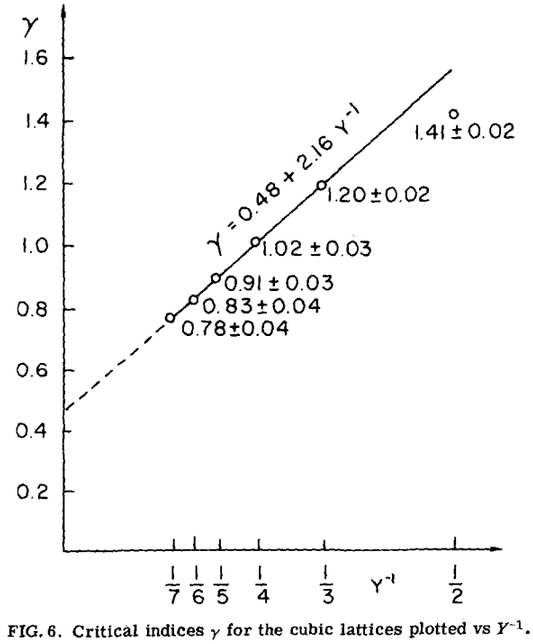


FIG. 6. Critical indices γ for the cubic lattices plotted vs Y^{-1}.

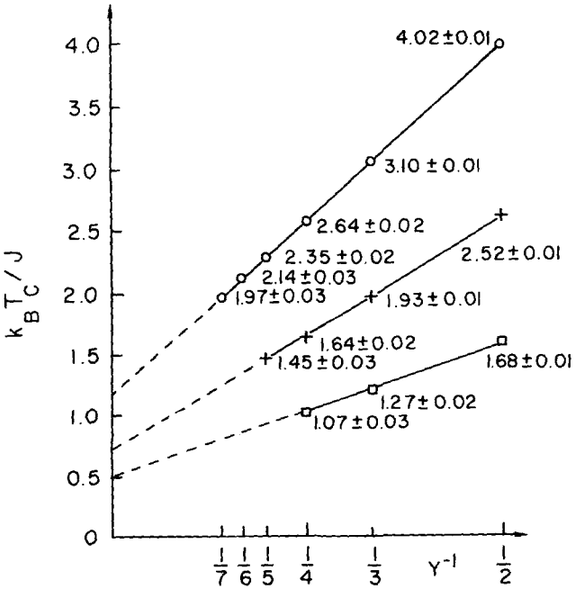


FIG. 7. Curie temperatures, $k_B T_C/J$, for the fcc lattice (denoted ○), b.c.c. lattice (denoted +), and sc lattice (denoted □) plotted vs Y^{-1}.

the simple cubic lattices are also contained in Figure 7.

For the two-dimensional lattices no consistent results could be obtained since the series are erratic and consist of positive and negative terms.

We have also investigated the specific heat series. The scatter in the various Padé approximants is too large to draw any conclusions concerning the critical temperature and the nature of the divergence of the specific heat at T_C. Considerably longer series would be needed for this purpose.

B. Critical Energy and Critical Entropy

Making use of the values of K_C in Eq. (105) the entropy series and the internal energy series obtained from Eq. (18) and (19) were also analyzed by the Padé approximant method. For the face-centered cubic lattice we found that

$$(S_\infty - S_C)/S_\infty = 0.494 - 0.353Y^{-1} \tag{106}$$

and

$$(E_\infty - E_C)/JN = 3.98 - 5.04Y^{-1}, \tag{107}$$

for all S with the exception that, for $S = \frac{1}{2}$, $(E_\infty - E_C)/JN = 1.60$. Here $S_\infty = Nk_B \ln Y$ and $E_\infty = 0$ are the entropy and the internal energy, respectively, at infinite temperature. For the body-centered cubic and the simple cubic lattices the convergence of the Padé approximants is in general fairly poor. However, the critical change of the entropy for the body-centered cubic and the simple cubic lattices seem also to vary linearly in Y^{-1} and for each spin value it is higher than that for the face-centered cubic lattice by an amount 0.025 and 0.055, respectively.

X. DISCUSSION AND CONCLUSIONS

Since there are $2S$ linearly independent isotropic interactions $(\mathbf{S}_i \cdot \mathbf{S}_j)^n [1 \leq n \leq 2S]$ for spin S particles, there are other kinds of order parameters[18,19] besides $\sum_i S_{zi}$, with which phase transitions might be associated. In order to study the possibility of such transitions we shall consider first the following modified Hamiltonian:

$$\mathcal{H}_n = \mathcal{H}_0 - \zeta Q_n, \tag{108}$$

where

$$\mathcal{H}_0 = -J \sum_{\langle ij \rangle} P_{ij} = -J\mathcal{P}, \tag{109}$$

$$Q_n = \sum_{i=1}^{N} S_{zi}^n, \quad n = 1, 2, \ldots 2S, \tag{110}$$

and ζ is some (fictitious) external magnetic field. We then define a generalized susceptibility χ_n by

$$\chi_n = \lim_{\zeta \to 0} \beta^{-1} \frac{\partial^2}{\partial \zeta^2} \ln \operatorname{tr} e^{-\beta \mathcal{H}_n}. \tag{111}$$

By definition $\chi_1 \sim \chi$. Since $\mathcal{P}$ and Q_n commute, as did $\mathcal{P}$ and Q, we have

$$\chi_n = \beta \Delta(Q_n). \tag{112}$$

In Ref. 20 the following theorem is proved.

Theorem 9:

$$\chi_n/\chi_1 = D_n(S) = (3/X)(W_{2n} - W_n^2). \tag{113}$$

Here $D_n(S)$ is a quantity which is independent of both *lattice* and *temperature* and W_n is given by Eq. (56). This means that for a given lattices, all of the $\chi_n \to \infty$ at the same T_C in exactly the same way.

Since the (dipolar) susceptibility (21) can also be written in the form

$$\chi = \tfrac{1}{3} g^2 \mu^2 \beta \sum_{i,j} \langle \mathbf{S}_i \cdot \mathbf{S}_j \rangle_\beta, \tag{114}$$

it is really necessary to investigate quantities of the form $\sum_{i,j} \langle (\mathbf{S}_i \cdot \mathbf{S}_j)^n \rangle_\beta$ for $1 \leq n \leq 2S$. However, for both computational and theoretical reasons, we have restricted our attention to the single quantity $\sum'_{i,j} \langle P_{ij} \rangle_\beta$. Here the prime in the double sum over i and j means that terms for which $i = j$ are to be excluded. The following theorem is also proved in Ref. 20.

Theorem 10:

$$\sum'_{i,j} \langle P_{ij} - (1/Y)\mathrm{I} \rangle_\beta = (4/Y) \sum'_{i,j} \langle \mathbf{S}_i \cdot \mathbf{S}_j \rangle_\beta. \tag{115}$$

Since $\langle \mathbf{S}_i^2 \rangle_\beta = X$, it immediately follows from this result and Eq. (114) that the left-hand side of Eq. (115) diverges at the same T_C and in exactly the same way as χ. Hence we must conclude that for the exchange interaction model the $2S$ "independent" multipolar phase transitions are in fact exactly degenerate with the dipolar transition. This means that the values of T_C and γ predicted on the basis of analysis of the high temperature susceptibility series are completely characteristic of this model.

Theorem 10 also suggests another approach for the derivation of the susceptibility series. Since $\sum'_{i,j} \langle P_{ij} \rangle_\beta$ is equal to χ except for a constant, χ can be obtained by calculating the quantities $\operatorname{tr} \sum'_{i,j} P_{ij} \mathcal{P}^n$ for each cluster. As was shown in Sec. IV, each Schrödinger exchange operator can be resolved into a direct sum of irreducible representations of the symmetric group. For a cluster of N sites it follows from Eqs. (46) and (49) that

$$\operatorname{tr} \sum_{i,j}' P_{ij} \mathcal{P}^n = \sum_\nu n_\nu [N(N-1)\chi^{(\nu)}_{1^{N-2}2} / \chi^{(\nu)}_{1^N}] \operatorname{tr} (\mathcal{P}^{(\nu)})^n, \tag{116}$$

where n_ν is given by Eq. (47) and $\chi_k^{(\nu)} = \operatorname{tr} P_k$ is given by Eq. (63). The labor required to derive χ by this approach is about the same as the previous approach which made use of Theorem 2.

The numerical results for the critical properties of the exchange interaction model have been shown in the previous section. Comparison of these results to those appropriate to the Heisenberg model[21,22] then shows that for $S > \frac{1}{2}$:

(1) Both T_C and γ for the present model are lower than those of the Heisenberg model.

(2) For the present model T_C is a decreasing function of spin, while T_C is an increasing function of spin for the Heisenberg model.

(3) For the exchange interaction model γ depends strongly on S and even becomes less than unity for large enough spin. On the other hand, for the Heisenberg model γ is a weak function of spin and is greater than 1.33 for all spin. Experimental values[23] of γ appear to lie in between the estimated values for these two models. Fisher[24] has recently suggested that the

observed value of the critical index γ_{obs} is related to the theoretical value by

$$\gamma_{obs} = \gamma/(1-\alpha), \tag{117}$$

where α is the theoretical value of the critical exponent of the specific heat series defined by

$$C_v \sim (T - T_C)^{-\alpha}, \quad \text{for } T \to T_C^+. \tag{118}$$

For $0 < \alpha < 1, \gamma_{obs} > \gamma$. This then shows that the theoretical estimates of γ when renormalized by the factor $(1-\alpha)^{-1}$ will be closer to the experimental values for the present model, while for the Heisenberg model the renormalized values of γ are in even further disagreement with the experimental values. This suggests that the present model may have more physical significance than originally thought.

(4) For the present model the fraction of the total entropy change occurring above T_C is higher than that of the Heisenberg model.

Hence we see that the inclusion of nonlinear terms in the Hamiltonian significantly affects the theoretical estimates of the critical parameters of magnetic systems.

The fact that T_C for the present model is lower than that of the Heisenberg model should be useful in a study of the possibility of a phase transition for the two-dimensional Heisenberg ferromagnet. If we can show that $T_C > 0$ for the two-dimensional exchange interaction model then there will be phase transitions for the two-dimensional Heisenberg model. However, as was previously mentioned, for large spin and/or for small coordination number, the series coefficients for the low-field susceptibility series are irregular and estimates of critical parameters from the high temperature series become quite inaccurate. It is then necessary that even more terms in the high temperature series be obtained.

APPENDIX A: VALUES OF e_n FOR SEVERAL CRYSTAL LATTICES

	$S=\frac{1}{2}$	$S=1$	$S=\frac{3}{2}$	$S=2$	$S=\frac{5}{2}$
Face-centered cubic lattice					
e_2	2.25	2.666 67	2.8125	2.88	2.916 67
e_3	2.25	1.777 78	1.406 25	1.152	0.972 22
e_4	0.468 75	−0.740 74	−1.259 76	−1.5168	−1.660 88
e_5	−1.3125	−1.481 48	−1.294 92	−1.107 45	−0.956 02
e_6	3.515 62	4.348 42	4.251 71	4.123 01	4.029 17
e_7	22.578 12	13.333 33	8.656 61	6.282 34	4.910 30
Body-centered cubic lattice					
e_2	1.5	1.777 78	1.875	1.92	1.944 44
e_3	−0.5	−0.395 06	−0.3125	−0.256	−0.216 05
e_4	0.4375	−1.086 42	−1.738 28	−2.0608	−2.241 52
e_5	0.375	2.304 53	2.378 91	2.159 62	1.919 24
e_6	0.531 25	1.285 78	3.688 96	5.259 01	6.241 93
e_7	−0.460 42	−6.285 14	−9.534 58	−9.961 30	−9.486 78
Simple cubic lattice					
e_2	1.125	1.333 33	1.406 25	1.44	1.458 33
e_3	−0.375	−0.296 29	−0.234 37	−0.192	−0.16204
e_4	−0.421 87	−0.962 96	−1.186 52	−1.296	−1.357 06
e_5	0.656 25	1.037 04	0.963 87	0.844 03	0.737 27
e_6	0.720 31	1.383 54	2.103 39	2.539 84	2.806 17
e_7	−1.579 69	−2.815 36	−3.148 86	−2.999 27	−2.739 34

APPENDIX B: VALUES OF a_n FOR SEVERAL CRYSTAL LATTICES

	$S=\frac{1}{2}$	$S=1$	$S=\frac{3}{2}$	$S=2$	$S=\frac{5}{2}$
Face-centered cubic lattice					
a_1	6.0	4.0	3.0	2.4	2.0
a_2	30.0	13.333 33	7.5	4.8	3.333 33
a_3	138.0	43.111 11	19.5	10.848	6.888 89
a_4	611.25	140.555 55	53.25	26.4336	15.472 22
a_5	2658.55	456.039 51	142.4	60.490 24	30.751 23
a_6	11 432.5125	1461.636 21	372.440 62	134.670 33	59.845 01
a_7	48 726.7262	4664.542 43	988.930 48	317.953 49	132.515 04
Body-centered cubic lattice					
a_1	4.0	2.666 67	2.0	1.6	1.333 33
a_2	12.0	5.333 33	3.0	1.92	1.333 33
a_3	34.666 67	9.777 78	3.833 33	1.770 66	0.888 89
a_4	95.833 33	21.851 85	8.208 33	4.043 73	2.351 85
a_5	262.7	45.195 06	16.4125	8.990 72	6.134 56
a_6	708.041 66	83.481 48	16.807 29	3.551 53	0.066 36
a_7	1893.289 68	168.439 35	24.685 73	−3.132 39	−10.696 58
Simple cubic lattice					
a_1	3.0	2.0	1.5	1.2	1.0
a_2	6.0	2.666 67	1.5	0.96	0.666 67
a_3	11.0	2.888 89	1.0	0.368	0.111 11
a_4	20.6250	4.722 22	1.781 25	0.880 80	0.513 89
a_5	39.025	7.451 85	3.340 63	2.208 64	1.712 04
a_6	68.777 08	7.190 53	0.618 09	−0.487 37	−0.647 19
a_7	119.429 76	8.197 00	−1.840 97	−3.868 18	−4.266 62

APPENDIX B: VALUES OF a_n FOR SEVERAL CRYSTAL LATTICES (Continued)

	$S=\frac{1}{2}$	$S=1$	$S=\frac{3}{2}$	$S=2$	$S=\frac{5}{2}$
Plane triangular					
a_1	3.0	2.0	1.5	1.2	1.0
a_2	6.0	2.666 66	1.5	0.96	0.666 67
a_3	8.5	2.888 89	1.4375	0.88	0.611 11
a_4	9.375	2.5	1.078 13	0.5928	0.375
a_5	11.025	1.624 69	0.098 44	−0.282 24	−0.386 73
a_6	16.964 58	2.069 14	0.209 89	−0.134 44	−0.192 21
a_7	21.152 68	4.766 14	2.472 38	1.771 23	1.428 05
Plane square					
a_1	2.0	1.333 33	1.0	0.8	0.666 67
a_2	2.0	0.888 89	0.5	0.32	0.222 22
a_3	1.333 33	0.148 15	−0.083 33	−0.138 67	−0.148 15
a_4	1.083 33	0.481 48	0.270 83	0.173 33	0.120 37
a_5	1.183 33	0.737 45	0.685 42	0.617 81	0.550 82
a_6	0.509 72	−0.677 02	−0.677 26	−0.053 45	−0.411 81
a_7	−4.821 83	−0.963 41	−1.125 63	−1.213 00	−1.190 45

* Part of a dissertation submitted by H. H. Chen to the Electrical Engineering Department of the Johns Hopkins University for the degree of Doctor of Philosophy, September 1970.

† Present Address: Department of Physics, New York University, University Heights, New York, New York 10453.

1 P. W. Anderson, Solid State Phys. **14**, 99 (1963); C. Herring, Magnetism IIB, 1 (1966).

2 G. A. T. Allan and D. D. Betts, Proc. Phys. Soc. (London) **91**, 341 (1967).

3 R. I. Joseph, Phys. Rev. **163**, 523 (1967).

4 E. Schrödinger, Proc. Roy. Irish Acad. **47**, 39 (1941).

5 G. A. Baker, H. E. Gilbert, J. Eve, and G. S. Rushbrooke, Phys. Letters **20**, 146 (1966).

6 H. A. Kramers, Commun. Kamerlingh Onnes Lab. Leiden, Suppl. No. 83 (1936).

7 W. Opechowski, Physica **4**, 181 (1937); **6**, 1112 (1939).

8 C. Domb, Advan. Phys. **9**, 149 (1960).

9 The lattice constants defined here are the so-called high temperature lattice constant. See, for example, Ref. 8.

10 M. F. Sykes, J. W. Essam, B. R. Heap, and B. J. Hiley, J. Math. Phys. **7**, 1557 (1966).

11 G. S. Rushbrooke and P. J. Wood, Mol. Phys. **1**, 257 (1958).

12 H. H. Chen, thesis (The Johns Hopkins University, 1970). See Appendix A.

13 D. E. Littlewood, *The Theory of Group Characters* (Oxford at the Clarendon Press, Oxford, 1940).

14 T. Yamanouchi, Proc. Phys.-Math. Soc. Japan **19**, 436 (1937).

15 C. Domb and D. W. Wood, Proc. Phys. Soc. (London) **86**, 1 (1965).

16 For $S = 1$ and the fcc lattice we get $a_7 = 4664.5424$ while the results of Ref. 2 gives $a_7 = 4658.4738$. In private communication with Dr. Allan the mistake leading to the latter erroneous value has been discovered.

17 G. A. Baker, Advan. Theoret. Phys. **1**, 1 (1965).

18 M. Suzuki, Progr. Theoret. Phys. (Kyoto) **42**, 1086 (1969).

19 M. Blume and Y. Y. Hsieh, J. Appl. Phys. **40**, 1249 (1969) (abstract only).

20 H. H. Chen and R. I. Joseph, Phys. Rev. B **2**, 2706 (1970).

21 H. E. Stanley and T. A. Kaplan, J. Appl. Phys. **38**, 977 (1967).

22 M. E. Fisher, Rept. Progr. Phys. **30**, 615 (1967).

23 M. Vicentini-Missoni, R. I. Joseph, M. S. Green, and J. M. H. Levelt Sengers, Phys. Rev. B **1**, 2312 (1970).

24 M. E. Fisher, Phys. Rev. **176**, 257 (1968).

Canonical Dynamics of Spinning Particles in Gravitational and Electromagnetic Fields*

H. P. Künzle

Department of Mathematics, University of Alberta, Edmonton 7, Canada

(Received 28 October 1971)

In terms of the canonical form and the connection form on the bundle of Lorentz frames P over a space-time manifold V, a presymplectic form ω is defined on P, which induces a Poisson bracket on the set of real valued functions on the phase space of the system representing a spinning particle in an exterior gravitational and electromagnetic field. This structure coincides with the unique Poincaré invariant one for the free particle. Moreover, the projections into V of the integral manifolds of the kernel of ω on P yield precisely the world lines of a spinning particle as obtained for the dipole approximation of Dixon's equations of motion for extended test bodies in general relativity.

1. INTRODUCTION

It is well known that if a classical system can be described by a Lagrangian L, then not only the equations of motion are uniquely determined, but also a symplectic structure $\tilde{\omega}$ on the set of all motions M (equivalent to the phase space) of the system. The 2-form $\tilde{\omega}$ in turn induces a Lie algebra structure, the Poisson bracket, on the set of real valued functions on M (the "observables"). For several reasons it can be argued[1] that, apart from the equations of motion, this symplectic structure is all that is needed (for the comparison with quantized systems) and it seems considerably more fundamental than the Lagrangian itself. For example, the Lagrangian is not quite unique for given equations of motion and given $\tilde{\omega}$, and there are no Galilei or Poincaré invariant Lagrangians, but there are symplectic manifolds on which these groups act transitively. Moreover, to find a Lagrangian formulation it is necessary to separate phase space variables into position and momentum variables, a distinction which sometimes—as in the case of a particle with spin—looks slightly artificial.

One purpose of this paper is to illustrate, in the case of a test particle with spin in a curved space-time (a system described for fixed rest mass m and spin magnitude s at a given time t by initial data consisting of three position coordinates x^i, the 4-momentum p^α, and the 4-spin vector s^α, subject to the constraints $p^\alpha p_\alpha = -m^2$, $s^\alpha s_\alpha = s^2$, and $p^\alpha s_\alpha = 0$), that it may be easier to guess a suitable form for the symplectic

APPENDIX B: VALUES OF a_n FOR SEVERAL CRYSTAL LATTICES (Continued)

	$S=\frac{1}{2}$	$S=1$	$S=\frac{3}{2}$	$S=2$	$S=\frac{5}{2}$
Plane triangular					
a_1	3.0	2.0	1.5	1.2	1.0
a_2	6.0	2.666 66	1.5	0.96	0.666 67
a_3	8.5	2.888 89	1.4375	0.88	0.611 11
a_4	9.375	2.5	1.078 13	0.5928	0.375
a_5	11.025	1.624 69	0.098 44	−0.282 24	−0.386 73
a_6	16.964 58	2.069 14	0.209 89	−0.134 44	−0.192 21
a_7	21.152 68	4.766 14	2.472 38	1.771 23	1.428 05
Plane square					
a_1	2.0	1.333 33	1.0	0.8	0.666 67
a_2	2.0	0.888 89	0.5	0.32	0.222 22
a_3	1.333 33	0.148 15	−0.083 33	−0.138 67	−0.148 15
a_4	1.083 33	0.481 48	0.270 83	0.173 33	0.120 37
a_5	1.183 33	0.737 45	0.685 42	0.617 81	0.550 82
a_6	0.509 72	−0.677 02	−0.677 26	−0.053 45	−0.411 81
a_7	−4.821 83	−0.963 41	−1.125 63	−1.213 00	−1.190 45

* Part of a dissertation submitted by H. H. Chen to the Electrical Engineering Department of the Johns Hopkins University for the degree of Doctor of Philosophy, September 1970.

† Present Address: Department of Physics, New York University, University Heights, New York, New York 10453.

1 P. W. Anderson, Solid State Phys. **14**, 99 (1963); C. Herring, Magnetism IIB, 1 (1966).

2 G. A. T. Allan and D. D. Betts, Proc. Phys. Soc. (London) **91**, 341 (1967).

3 R. I. Joseph, Phys. Rev. **163**, 523 (1967).

4 E. Schrödinger, Proc. Roy. Irish Acad. **47**, 39 (1941).

5 G. A. Baker, H. E. Gilbert, J. Eve, and G. S. Rushbrooke, Phys. Letters **20**, 146 (1966).

6 H. A. Kramers, Commun. Kamerlingh Onnes Lab. Leiden, Suppl. No. 83 (1936).

7 W. Opechowski, Physica **4**, 181 (1937); **6**, 1112 (1939).

8 C. Domb, Advan. Phys. **9**, 149 (1960).

9 The lattice constants defined here are the so-called high temperature lattice constant. See, for example, Ref. 8.

10 M. F. Sykes, J. W. Essam, B. R. Heap, and B. J. Hiley, J. Math. Phys. **7**, 1557 (1966).

11 G. S. Rushbrooke and P. J. Wood, Mol. Phys. **1**, 257 (1958).

12 H. H. Chen, thesis (The Johns Hopkins University, 1970). See Appendix A.

13 D. E. Littlewood, *The Theory of Group Characters* (Oxford at the Clarendon Press, Oxford, 1940).

14 T. Yamanouchi, Proc. Phys.-Math. Soc. Japan **19**, 436 (1937).

15 C. Domb and D. W. Wood, Proc. Phys. Soc. (London) **86**, 1 (1965).

16 For $S = 1$ and the fcc lattice we get $a_7 = 4664.5424$ while the results of Ref. 2 gives $a_7 = 4658.4738$. In private communication with Dr. Allan the mistake leading to the latter erroneous value has been discovered.

17 G. A. Baker, Advan. Theoret. Phys. **1**, 1 (1965).

18 M. Suzuki, Progr. Theoret. Phys. (Kyoto) **42**, 1086 (1969).

19 M. Blume and Y. Y. Hsieh, J. Appl. Phys. **40**, 1249 (1969) (abstract only).

20 H. H. Chen and R. I. Joseph, Phys. Rev. B **2**, 2706 (1970).

21 H. E. Stanley and T. A. Kaplan, J. Appl. Phys. **38**, 977 (1967).

22 M. E. Fisher, Rept. Progr. Phys. **30**, 615 (1967).

23 M. Vicentini-Missoni, R. I. Joseph, M. S. Green, and J. M. H. Levelt Sengers, Phys. Rev. B **1**, 2312 (1970).

24 M. E. Fisher, Phys. Rev. **176**, 257 (1968).

Canonical Dynamics of Spinning Particles in Gravitational and Electromagnetic Fields*

H. P. Künzle

Department of Mathematics, University of Alberta, Edmonton 7, Canada

(Received 28 October 1971)

In terms of the canonical form and the connection form on the bundle of Lorentz frames P over a space-time manifold V, a presymplectic form ω is defined on P, which induces a Poisson bracket on the set of real valued functions on the phase space of the system representing a spinning particle in an exterior gravitational and electromagnetic field. This structure coincides with the unique Poincaré invariant one for the free particle. Moreover, the projections into V of the integral manifolds of the kernel of ω on P yield precisely the world lines of a spinning particle as obtained for the dipole approximation of Dixon's equations of motion for extended test bodies in general relativity.

1. INTRODUCTION

It is well known that if a classical system can be described by a Lagrangian L, then not only the equations of motion are uniquely determined, but also a symplectic structure $\tilde{\omega}$ on the set of all motions M (equivalent to the phase space) of the system. The 2-form $\tilde{\omega}$ in turn induces a Lie algebra structure, the Poisson bracket, on the set of real valued functions on M (the "observables"). For several reasons it can be argued[1] that, apart from the equations of motion, this symplectic structure is all that is needed (for the comparison with quantized systems) and it seems considerably more fundamental than the Lagrangian itself. For example, the Lagrangian is not quite unique for given equations of motion and given $\tilde{\omega}$, and there are no Galilei or Poincaré invariant Lagrangians, but there are symplectic manifolds on which these groups act transitively. Moreover, to find a Lagrangian formulation it is necessary to separate phase space variables into position and momentum variables, a distinction which sometimes—as in the case of a particle with spin—looks slightly artificial.

One purpose of this paper is to illustrate, in the case of a test particle with spin in a curved space-time (a system described for fixed rest mass m and spin magnitude s at a given time t by initial data consisting of three position coordinates x^i, the 4-momentum p^α, and the 4-spin vector s^α, subject to the constraints $p^\alpha p_\alpha = -m^2$, $s^\alpha s_\alpha = s^2$, and $p^\alpha s_\alpha = 0$), that it may be easier to guess a suitable form for the symplectic

structure from geometrical characteristics of the system than to find first a Lagrangian.

In Sec. 2 the basic notions of evolution space P and space of motion M introduced by Souriau[1,2] are reviewed and slightly generalized. In the literature, P corresponds to $TQ \times \mathbb{R}$ if Q is the configuration space, TQ its tangent bundle, and $\mathbb{R}$ the time continuum, or to $T(Q \times \mathbb{R})$ if the Lagrangian is expressed in homogeneous form. The manifold P is then equipped with a presymplectic form ω (a contact structure in the nonhomogeneous case) and the leaves of the foliation defined by ker ω are in one-to-one correspondence with the possible motions of the system (or with the points of the phase space T^*Q considered as initial data of a motion at a fixed time). Therefore, M defined as the set leaves $P/\ker\,\omega$ is called the space of motions and carries a naturally induced symplectic structure $\tilde{\omega}$. We will simply drop the assumption that rank $\omega = \dim P - 1$ or $\dim P - 2$.

In Sec. 3 we choose for P the principal bundle of Lorentz frames (for short, Lorentz bundle) over space–time V and show that a presymplectic potential θ (a 1-form such that $\omega = d\theta$ is a presymplectic form) can be chosen that is expressed in a simple way in terms of the given geometrical structure of P and that can heuristically be justified at least as well as the choice of any particular Lagrangian. For the flat space case this presymplectic manifold (P, ω) is then easily seen to be equivalent as evolution space to the one Souriau[1] chooses to describe what he calls a Poincaré invariant classical elementary system with mass and spin. For the latter, Souriau shows by group theoretical methods that there exists a unique symplectic space of motions $(M, \tilde{\omega})$, while he chooses his evolution space from the consideration that the time development of such a free particle should be trivial, i.e., a timelike geodesic in Minkowski space–time. In our treatment, working downwards from the given structure on the evolution space, however, we get not only the symplectic structure on $M = P/\ker\,\omega$ (in principle at least), but also explicitly the equations of motion for the particle in a gravitational field. This is so because the worldlines $x^\alpha(s)$ of the particle in space–time are simply the projections into V of the leaves of ker ω in P. Expressed in terms of $x^\alpha(s), p^\alpha(s)$, and $S^{\alpha\beta}(s)$, these world lines are obtained by integrating

$$\dot{p}^\alpha = -\tfrac{1}{2}R^\alpha{}_{\beta\gamma\delta}v^\beta S^{\gamma\delta} \tag{1.1}$$

and

$$\dot{S}^{\alpha\beta} = 2p^{[\alpha}v^{\beta]} \tag{1.2}$$

where $\dot{} := v^\alpha\nabla_\alpha$ and $v^\alpha := dx^\alpha/ds$ and $S^{\alpha\beta}$ is the antisymmetric spin tensor related to s^α by $S^{\alpha\beta} := \epsilon^{\alpha\beta\gamma\delta}u_\gamma s_\delta$, $s^\alpha = -\tfrac{1}{2}\epsilon^{\alpha\beta\gamma\delta}u_\beta S_{\gamma\delta}$, where u^α is the unit vector in the direction of p^α. These are precisely the equations Dixon[3,4] gets for extended test bodies in general relativity provided higher moments than monopole and dipole moments are neglected, his starting point being the conservation law for the stress–energy tensor $T_{\alpha\beta}$ of the field theory.

The same agreement with Dixon's equations is obtained in Sec. 4 where the motion of a charged particle with spin and magnetic dipole moment proportional to the spin is studied in an exterior gravitational and electromagnetic field. In addition to showing the existence of a canonical form of Dixon's fairly complicated equations, this approach seems to confirm the correctness of the latter since the derivation is from a completely different point of view. (We do, however, agree with Dixon in the *a priori* assumptions that the 4-momentum and spin be orthogonal,

$$s^\alpha p_\alpha = 0 \quad \text{or} \quad S^{\alpha\beta}p_\beta = 0, \tag{1.3}$$

and that the 4-momentum be not necessarily parallel to the 4-velocity.) These assumptions seem to be very natural in view of Souriau's group theoretical study in the flat space case, but they were not always made in the earlier literature[5].

As an illustration of the form Noether's theorem takes in this non-Lagrangian canonical formalism, we compute in Sec. 5 explicitly the integrals of motion of the spinning particle obtained from space–time symmetries.

2. CANONICAL DYNAMICS

This section contains a review of the general framework of canonical dynamics for classical systems with a finite number of degrees of freedom.[6] We follow most closely Souriau[1], but much of the formalism is also introduced in Abraham and Marsden[7] and Godbillon.[8] Since our applications are to a relativistic system, we only work with the extended configuration and phase space, in other words, always think of the time as included among the position variables.

A *canonical dynamical system* is given by any presymplectic manifold (P, ω) (i.e., ω is a closed 2-form on P with constant rank less than the dimension of P), where the manifold P is called an *evolution* space of the system.[9] The kernel of a closed form ω always defines an involutive differential system ker ω (for $q \in P$, $\ker_q\,\omega := \{Z \in T_qP/Z \lrcorner \omega = 0\}$[10]) and hence a foliation. The set of leaves of this foliation, $M = P/\ker\,\omega$, then has a natural induced differentiable structure provided there exists a neighborhood of every point in P that intersects every leaf in a connected (or void) set.[11] This we assume to be the case (by possibly restricting consideration to some open submanifold of P). The manifold M then also has a unique *symplectic* structure $\tilde{\omega}$ such that $p^*\tilde{\omega} = \omega$, where p is the canonical projection $p : P \to M$. Interpreting the leaves of ker ω as physical motions, one calls M the *space of motions*. It is symplectomorphic to the (nonextended) phase space in the physics literature.

To describe by (P, ω) not only the canonical structure on the phase space but also the time development of the system we assume that P is some fibre bundle $\pi : P \to V$, where V is space–time (or a product of space–time with itself for many-particle systems, or some other manifold describing the *external* configuration of the system. Note that V need *not* represent *all* position variables one needs in the Lagrangian formulation.) The possible trajectories of the particle(s) in space–time are then obtained by projecting the leaves of ker ω into V. The result will normally be a (nonparametrized) curve in V, but also higher-dimensional surfaces may occur as, for example, null hypersurfaces representing wavefronts in the case of zero-rest-mass particles with spin in Souriau's model.[12]

A transformation group $\psi : G \times P \to P$ leaving ω invariant (i.e., satisfying $\psi_a^* \omega = \omega$ for all $a \in G$) is called a *dynamical group* of the canonical dynamical system. Noether's theorem, which states that to every parameter of G there corresponds an integral of motion can then be stated in this formalism as follows.

Since G leaves ω invariant and therefore also the leaves of $\ker \omega$ there is an induced action $\tilde{\psi} : G \times M \to M$ on the space of motions. Let $\bar{A}$ and $\tilde{A}$ be the Killing vector fields[13] on P and M, respectively, corresponding to an element A of the Lie algebra $\mathfrak{g}$ of G. Then $p_* \bar{A} = \tilde{A}$ and $\pounds_{\bar{A}} \omega = \pounds_{\tilde{A}} \tilde{\omega} = 0$ or, equivalently, $d(\bar{A} \lrcorner \omega) = 0$ and $d(\tilde{A} \lrcorner \tilde{\omega}) = 0$.[14] Thus, there exist functions $\tilde{f}_A$ and f_A defined (locally) up to a constant by $\tilde{A} \lrcorner \tilde{\omega} = -d\tilde{f}_A$ and

$$f_A = \pi^* \tilde{f}_A \tag{2.1}$$

satisfying $\bar{A} \lrcorner \omega = -df_A$. In view of (2.1), f_A is clearly a "constant of motion."

The map $\mu : P \to \mathfrak{g}^*$ (the dual of the Lie algebra $\mathfrak{g}$) defined by $f_A(q) = \langle A / \mu(q) \rangle$ for all $q \in P$, where $\langle / \rangle$ is the scalar product of $\mathfrak{g} \otimes \mathfrak{g}^*$, is called a *moment* of the dynamical group G. Whether μ exists globally is a nontrivial topological question depending on G, ψ and (P, ω) and discussed by Souriau[1] for some special cases.

The functions f_A are particularly easy to find if there exists a presymplectic potential θ of ω that is itself invariant under G. Then one has simply $f_A = \bar{A} \lrcorner \theta$.

If P is a fibre bundle over V and there is a unique lift of (at least some class of) diffeomorphisms $\phi : V \to V$ into bundle automorphisms $\tilde{\phi} : P \to P$, there often exists a particular dynamical group, consisting of all bundle automorphisms $\tilde{\phi}$ of P that leave ω invariant. They are easy to interpret physically and can be called *obvious symmetries*.[15]

3. THE LORENTZ BUNDLE AS EVOLUTION SPACE

In this section let P always denote the Lorentz bundle over space-time (V, g). A fibre $\pi^{-1}(x)$ for $x \in V$ consists of all tetrads $\{e_a \mid a = 0, 1, 2, 3\}$ at x satisfying $g(e_a, e_b) = \eta_{ab}$.[16] The Lorentz group G operates simply transitively on each fibre of P in the sense of operating separately on each vector of any frame. If (x^α) is a local coordinate system in V and $e_a = e_a^\alpha \partial_\alpha$, then (x^α, e_a^β) is the corresponding fibre coordinate system in P. Let $\{\theta^a\}$, where $\theta^a = \theta_\alpha^a dx^\alpha$, be the basis of $T_x^* V$ dual to $\{e_a\}$ when the latter is considered as a basis of $T_x V$. Then

$$\theta_\alpha^a e_a^\beta = \delta_\alpha^\beta \quad \text{and} \quad \theta_\alpha^a e_b^\alpha = \delta_b^a \tag{3.1}$$

and $(\theta^0, \theta^1, \theta^2, \theta^3)$ is the $\mathbb{R}^4$-valued canonical form of P.[13] We represent the Lie algebra $\mathfrak{g}$ of the Lorentz group by the set of 6×6 matrices A satisfying $A \cdot \eta + \eta \cdot A^T = 0$. A basis of $\mathfrak{g}$ is given by the ten matrices E_{ab}, $a < b$, $E_{ab} + E_{ba} = 0$, whose components are $(E_{ab})_k^l = \eta_{ak} \delta_b^l - \delta_a^l \eta_{bk}$. The $\mathfrak{g}$-valued connection form $\underset{\sim}{\omega} = \omega^a{}_b E_a{}^b$ of the metric connection on P is then in local coordinates given by[17]

$$\omega^a{}_b = \theta_\alpha^a (de_b^\alpha + \Gamma_{\beta\gamma}^\alpha e_b^\gamma dx^\beta) \tag{3.2}$$

and satisfies

$$\omega_{ab} + \omega_{ba} = 0. \tag{3.3}$$

Note that the ten real valued 1-forms θ^a and $\omega^a{}_b$ form a basis of each cotangent space $T_q^* P$.

To consider P as evolution space for a spinning particle, we interpret $x \in V$ as the event corresponding to the instantaneous position of the particle, e_0 as the direction of its 4-momentum $p^\alpha \partial_\alpha$, and e_1 as the direction of its spin 4-vector $s^\alpha \partial_\alpha$ (or, equivalently, $e_2 \wedge e_3$ as the plane element defined by the tensorial spin $\frac{1}{2} S^{\alpha\beta} \partial_\alpha \wedge \partial_\beta$). The frame vectors e_2 and e_3 are not interpreted separately because P has one dimension more than necessary for an evolution space of a system with four degrees of freedom.

For a system with fixed nonzero rest mass m and fixed spin magnitude s, we choose as presymplectic potential on P

$$\theta = m\theta^0 + s\omega_{23}. \tag{3.4}$$

An *a priori heuristic* justification might be as follows: The Lorentz group operates on the fibres of P. Let $\bar{A}$ be the Killing vector field corresponding to $A = A^a{}_b E_a{}^b \in \mathfrak{g}$. Then $\bar{A}$ is a vertical vector field on P and the $\bar{E}_a{}^b$ span the vertical subspace of $T_q P$; therefore, $\bar{A} \lrcorner \theta = s\bar{A} \lrcorner \omega_{23} = sA_{23}$, or $f_{E_2{}^3} = s$, while $f_{E_a{}^b} = 0$ for all other $E_a{}^b \in \mathfrak{g}$. But s is the only "internal" nontrivial constant of motion we associate with the system (internal, because it does not arise from space-time symmetries). The rest mass m, on the other hand, is associated with translations of space-time in the direction of the 4-momentum. Indeed, suppose ξ is a vector field on V coinciding with $e_0^\alpha \partial_\alpha$ at a given event x, and let $\tilde{\xi}$ be its horizontal lift onto P. Then $\tilde{\xi} \lrcorner \omega_b^a = 0$ and $\tilde{\xi} \lrcorner \theta^a = \xi^a$, the tetrad components of ξ with respect to e_a. Therefore, $\tilde{\xi} \lrcorner \theta = m\xi^0 = m$ if $\xi(x) = e_0(x)$, i.e., if ξ corresponds to a space-time translation in the direction of $p^\alpha \partial_\alpha$ by unit proper time.

The verification that $\omega := d\theta$ is a presymplectic form on P, the computation of $\ker \omega$ and the trajectories in V, and the proof that the canonical structure on $M = P/\ker \omega$ coincides with the one of Souriau in the flat space case are quite straightforward.

To compute ω, we use the structure equations of Cartan for a connection without torsion,

$$d\theta^a = -\omega^a{}_r \wedge \theta^r$$

and

$$d\omega^a{}_b = -\omega^a{}_r \wedge \omega^r{}_b + \tfrac{1}{2} R^a{}_{brs} \theta^r \wedge \theta^s$$ [18]

and find from (3.4) using also (3.3)

$$\omega = \tfrac{1}{2} s R_{23rs} \theta^r \wedge \theta^s - m\theta^r \wedge \omega_{0r} - s\omega_{2r} \wedge \omega^r{}_3. \tag{3.5}$$

At this stage already we can see that (P, ω) with ω given by (3.5) defines the same symplectic manifold $P/\ker \omega$ as the unique one obtained by group theoretical methods by Souriau.[19,20,1] Indeed, for a flat connection we have from (3.5) and (3.2) (in Cartesian coordinates)

$$\begin{aligned} \omega &= m\theta^r \wedge \omega_{r0} - s\omega_{20} \wedge \omega_{30} + s\omega_{21} \wedge \omega_{31} \\ &= m\theta_\alpha^r \eta_{rs} \theta_\beta^s dx^\alpha \wedge de_0^\beta - s\eta_{2r}\eta_{3s} \theta_{[\alpha}^r \theta_{\beta]}^s \\ &\quad \times (de_0^\alpha \wedge de_0^\beta - de_1^\alpha \wedge de_1^\beta), \end{aligned}$$

which can be written as

$$\omega = m\eta_{\alpha\beta}dx^\alpha \wedge de_0^\beta - \tfrac{1}{2}\, s\epsilon_{\mu\nu\alpha\beta}e_0^\mu e_1^\nu \times (de_0^\alpha \wedge de_0^\beta - de_1^\alpha \wedge de_1^\beta). \quad (3.6)$$

This is of the same form (up to some conventions and notations) as the presymplectic form Souriau[21] chooses on the manifold

$$W = \{(x, e_0, e_1) \in \mathbb{R}^4 \times \mathbb{R}^4 \times \mathbb{R}^4 \mid (-\eta_{\alpha\beta}e_0^\alpha e_0^\beta) = \eta_{\alpha\beta}e_1^\alpha e_1^\beta = 1, \eta_{\alpha\beta}e_0^\alpha e_1^\beta = 0\}.$$

The form ω in (3.6) can be considered as a pull back to the Lorentz bundle of Souriau's ω on W under the obvious projection. This process increases only the kernel by the one more dimension P has than W. Thus we have clearly a symplectomorphism from $P/\ker\,\omega_P$ to $W/\ker\,\omega_W$.

To compute now ker ω, note first that a vector on P is determined by its components

$$v^a := Z \lrcorner\, \theta^a \quad (3.7)$$

and

$$\dot{e}_b^a := Z \lrcorner\, \omega_b^a, \quad (3.8)$$

where by (3.3)

$$\eta_{ac}\dot{e}_b^c + \eta_{bc}\dot{e}_a^c = 0. \quad (3.9)$$

It follows from (3.5)–(3.9) that

$$Z \lrcorner\, \omega = (sv^sR_{sr23} - m\dot{e}_r^0)\theta^r - mv^r\omega_{0r} + s(\dot{e}_2^r\omega_{r3} - \dot{e}_3^r\omega_{r2}).$$

This 1-form is equal to zero if

$$m\dot{e}_a^0 = -sR_{ar23}v^r, \quad (3.10)$$

$$mv^1 = 0, \quad mv^2 = -s\dot{e}_3^0, \quad mv^3 = s\dot{e}_2^0 \quad (3.11)$$

and $s\dot{e}_2^1 = s\dot{e}_3^1 = 0$. Using (3.11) and (3.9) in (3.10), we find

$$v^2 = -s^2m^{-2}\Delta^{-1}R_{0323}v^0, \quad v^3 = s^2m^{-2}\Delta^{-1}R_{0223}v^0 \quad (3.12)$$

provided $\Delta := 1 + s^2m^{-2}R_{2323} \neq 0$. Substituting (3.12) again into (3.10), we get all $\dot{e}_a^0 = \dot{e}_0^a$. It follows that all v^a and $\dot{e}_b^a$ are either zero or well-determined multiples of v^0, except $\dot{e}_3^2 = -\dot{e}_2^3$, which remain completely arbitrary. This proves that dim ker $\omega = 2$ (generically, for $m \neq 0 \neq s$). If $s = 0 \neq m$, on the other hand, $v^a = 0$ for $a \neq 0$ and $\dot{e}_a^0 = 0$ for all a, and $\dot{e}_2^1 = -\dot{e}_1^2$, $\dot{e}_3^1 = -\dot{e}_1^3$, $\dot{e}_3^2 = -\dot{e}_2^3$, and v^0 remain arbitrary, showing that dim ker $\omega = 4$. Thus, the dimension of M turns out to be eight or six, as it should, for the particle with nonzero or zero spin, respectively.

The equations of the world line of the particle are best obtained if everything is expressed in terms of local fibre coordinates. The projection map has then the simple form $\pi: (x^\alpha, e_a^\alpha) \to x^\alpha$. Suppose $Z = v^\alpha\partial/\partial x^\alpha + E_a^\alpha\partial/\partial e_a^\alpha$ is a vector field on P everywhere in ker ω. Then an integral curve of Z satisfies

$$\frac{dx^\alpha}{d\lambda} = v^\alpha \quad \text{and} \quad \frac{de_a^\alpha}{d\lambda} = E_a^\alpha.$$

But by (3.1) and (3.2) $v^a = Z \lrcorner\, \theta^a = v^\alpha\theta_\alpha^a$ and $\dot{e}_b^a = Z \lrcorner\, \omega_b^a = \theta_\alpha^a(E_b^\alpha + \Gamma_{\beta\gamma}^\alpha e_b^\gamma v^\beta) = \theta_\alpha^a(de_b^\alpha/d\lambda + \Gamma_{\beta\gamma}^\alpha e_b^\gamma dx^\beta/d\lambda) = \theta_\alpha^a v^\beta\nabla_\beta e_b^\alpha$, i.e., v^a and $\dot{e}_b^a$ are the tetrad components of the tangent vector to the world line (which can be arbitrarily parametrized) and of the covariant derivative of e_a along the world line, respectively.

It is not hard to translate the tetrad components of v and e_a obtained in (3.10)–(3.12) into components with respect to a local coordinate system. If $p^\alpha := me_0^\alpha$ and $s^\alpha := se_1^\alpha$ or $S^{\alpha\beta} = 2se_2^{[\alpha}e_3^{\beta]}$, one finds

$$mv^\alpha = v^0(p^\alpha + \tfrac{1}{2}m^{-2}\Delta^{-1}S^{\alpha\beta}R_{\beta\gamma\mu\nu}p^\gamma S^{\mu\nu}), \quad (3.13)$$

$$\dot{p}^\alpha = -\tfrac{1}{2}R^\alpha_{\lambda\mu\nu}v^\lambda S^{\mu\nu}, \quad (3.14)$$

and

$$(\delta_\beta^\alpha + e_0^\alpha e_{0\beta})\dot{s}^\beta = 0, \quad (3.15)$$

with $\Delta = 1 + \tfrac{1}{4}m^{-2}R_{\alpha\beta\gamma\delta}S^{\alpha\beta}S^{\gamma\delta}$. Equation (3.14) is the same as (1.1), and, since (1.3) holds, (3.15) is seen to be equivalent to (1.2) while (3.13) can be derived from (1.1)–(1.3). The factor v^0 is still arbitrary. It can be fixed by requiring v^α to be a unit vector, i.e., by choosing proper time as the curve parameter. We have thus recovered Dixon's equations of motion by our choice (3.4) of the presymplectic potential θ.

Remark: The foregoing construction breaks down wherever $\Delta = 0$. The kernel of ω may then have a different dimension. Moreover, the 4-velocity v^α in (3.13) may turn out to be null or spacelike. These shortcomings are not due to this particular derivation, but they are inherent in the equations of motion (1.1)–(1.3). Dixon indeed derived them under the assumption that space–time curvature and spin are not too large. We see that a sufficient condition for the consistency of the equations of motion is of the form (spin)$^2\cdot$(curvature) $<$ (mass)^2. (This is satisfied for most realistic cases, for in the extreme case of an electron near the horizon of a Schwarzschild black hole of mass M the equations remain valid provided $M > 10^{19}$ g or equivalently $r_{\text{grav}} > 10^{-9}$ cm). (spin)$^2\cdot$(curvature) $<$ (mass)^2. (This is satisfied for most realistic cases, for in the extreme case of an electron near the horizon of a Schwarzschild black hole of mass M the equations remain valid provided $M > 10^{19}$ g or equivalently $r_{\text{grav}} > 10^{-9}$ cm).

4. MOTION IN AN ELECTROMAGNETIC FIELD

We indicate in this section how some modifications in the symplectic potential θ of (3.4) lead to a system describing a particle with fixed "rest" mass m, spin magnitude s, charge e, and magnetic dipole moment μs in a combined gravitational and electromagnetic field. These modifications are formally the same as in the special relativistic case treated by Souriau[1] and the result will again be exactly the equations of motion for Dixon's[4] extended charged bodies if all other moments are neglected.

Assume that the particle has a magnetic dipole moment parallel and proportional to the spin by a constant factor μ ($\mu = ge/2m$ in our units, where g is the Landé factor which is very close to 2 for the electron, for example). Then the symplectic potential θ in (3.4) must be replaced by

$$\theta = M\theta^0 + s\omega_{23} + eA, \quad (4.1)$$

where $M := m + \frac{1}{2}\mu S^{\alpha\beta}F_{\alpha\beta} = m + \mu s F_{23}$ (in the frame $\{e_a\}$ we have $\frac{1}{2}\mu S^{\alpha\beta}F_{\alpha\beta} = \frac{1}{2}\mu \mathbf{s}\cdot\mathbf{B}$ which is the familiar spin contribution to the rest energy in a magnetic field) and $A = A_\alpha dx^\alpha = A_r\theta^r$ is a vector potential of the electromagnetic field $F = dA$.[22] Since

$$dM = \mu s(\nabla_1 F_{23}\theta^r + F_{2r}\omega^r{}_3 + F_{r3}\omega^r{}_2)\ ^{23},$$

we have for the presymplectic form explicitly

$$\omega = (\tfrac{1}{2}sR_{23rs} + \tfrac{1}{2}eF_{rs} - \mu s\delta^0_r\nabla_s F_{23})\theta^r\wedge\theta^s - M\theta^r\wedge\omega_{0r} - \mu s\theta^0\wedge(F_{2r}\omega^r{}_3 + F_{r3}\omega^r{}_2) - s\omega_{2r}\wedge\omega^r{}_3, \quad (4.2)$$

and, if v^a and $\dot{e}^a_b$ are as in (3.7) and (3.8), respectively,

$$Z\lrcorner\,\omega = (sv^sR_{sr23} + ev^sF_{sr} - \mu s v^0\nabla_r F_{23} - M\dot{e}^0_r)\theta^r + \mu s(F_{2r}\dot{e}^r_3 + F_{r3}\dot{e}^r_2 + v^r\nabla_r F_{23})\theta^0 - Mv^r\omega_{0r} + s(\dot{e}^r_2 - \mu v^0F_2{}^r)\omega_{r3} - s(\dot{e}^r_3 + \mu v^0F^r{}_3)\omega_{r2}.$$

There follows for $Z\in\ker\omega$

$$\begin{aligned} v^1 &= 0,\\ v^2 &= -sM^{-2}\Delta^{-1}v^0(sR_{0323} + (e-\mu M)F_{03} - \mu s\nabla_3F_{23}),\\ v^3 &= sM^{-2}\Delta^{-1}v^0(sR_{0223} + (e-\mu M)F_{02} - \mu s\nabla_2F_{23}), \end{aligned} \quad (4.3)$$

and

$$\begin{aligned} M\dot{e}^2_0 &= v^0\Delta^{-1}\{sR_{0223} + [e-\mu M(1-\Delta)]F_{02} - \mu s\nabla_2F_{23}\},\\ M\dot{e}^3_0 &= v^0\Delta^{-1}\{sR_{0323} + [e-\mu M(1-\Delta)]F_{03} - \mu s\nabla_3F_{23}\},\\ M\dot{e}^1_0 &= v^0(sR_{0123} + eF_{01} - \mu s\nabla_1F_{23}) - v^2(sR_{1223} + eF_{12}) - v^3(sR_{1323} + eF_{13}),\\ \dot{e}^2_1 &= \mu F_{12}v^0, \qquad \dot{e}^3_1 = \mu F_{13}v^0, \end{aligned} \quad (4.4)$$

where it is assumed that $\Delta := 1 + s^2M^{-2}R_{2323} + esM^{-2}F_{23} \neq 0 \neq M$. Again dim ker $\omega = 2$ in the generic case for $m \neq 0 \neq s$ and dim ker $\omega = 4$ for $s = 0 \neq m$; thus, qualitatively the symplectic structure on the space of motions is not changed by the presence of a weak electromagnetic field.

The equations of motion reduce to those of Souriau[24] and Bargmann, Michel, and Telegdi[25] for flat space and *weak* electromagnetic fields, as can be seen directly from (4.3) and (4.4). (Note that putting the Landé factor g equal to 2 only really simplifies the equations of motion if only terms linear in F are considered.) If we transform (4.3) and (4.4) into tensor equations, we find

$$Mv^\alpha = v^0\{p^\alpha + M^{-2}\Delta^{-1}S^{\alpha\beta}[\tfrac{1}{2}(R_{\beta\lambda\mu\nu}p^\lambda + \mu M\nabla_\beta F_{\mu\nu})S^{\mu\nu} + (e-\mu M)F_{\beta\lambda}p^\lambda]\}, \quad (4.5)$$

$$\dot{p}^\alpha = -\tfrac{1}{2}R^\alpha{}_{\lambda\mu\nu}v^\lambda S^{\mu\nu} - eF^\alpha{}_\lambda v^\lambda - \tfrac{1}{2}\mu v^0\nabla^\alpha F_{\mu\nu}S^{\mu\nu}, \quad (4.6)$$

and

$$(\delta^\alpha_\beta + e^\alpha_0 e_{0\beta})(\dot{s}^\beta + \mu v^0F^\beta{}_\gamma s^\gamma) = 0, \quad (4.7)$$

where $p^\alpha = Me^\alpha_0$ and the parametrization is still arbitrary. Equations (4.6) and (4.7) are equivalent to Dixon's equations of motion for the charged spinning particle with magnetic dipole moment.

5. SYMMETRIES AND CONSERVED QUANTITIES

We apply the remarks of Sec. 2 about dynamical groups to the case of a spinning particle.

The first group action to consider is the action of the Lorentz group on the fibres of the principal bundle P. This action is such that if $B = B^a_bE^b_a \in \mathfrak{g}$ then the corresponding Killing vector field $\overline{B}$ on P satisfies $B\lrcorner\,\theta^a = 0$ and $\overline{B}\lrcorner\,\omega^a_b = B^a_b$. A straightforward calculation of $\pounds_{\overline{B}}\omega$ using these relations shows that only the 1-parameter group $\exp(t)\ E_{23}$ leaves ω invariant (if $m \neq 0 \neq s$) and that the corresponding conserved quantity is $\overline{E}_{23}\lrcorner\,\theta = s$. If $s = 0$ (but $m \neq 0$), the whole $SO(3)$ generated by E_{12}, E_{13}, and E_{23} operates as a dynamical group, but the corresponding integrals of motion $E_{12}\lrcorner\,\theta$, $E_{13}\lrcorner\,\theta$, and $E_{23}\lrcorner\,\theta$ all vanish.

Since P is a principal bundle with a connection, there exists a unique horizontal lift $\hat{\xi}$ of every vector field ξ on V. A computation, however, shows that in order to leave θ invariant the vector field ξ must be covariantly constant on V, i.e., $\nabla_\alpha\xi_\beta = 0$, and satisfy $e\xi\lrcorner\,F = 0$, a situation which can only occur in very special gravitational and electromagnetic fields. The corresponding quantity $\hat{\xi}\lrcorner\,\theta = M\xi^0 = -\xi^\alpha p_\alpha$ will therefore be conserved only in these rare cases.

The following third possibility is the only physically interesting one. If $\phi: V\to V$ is any diffeomorphism, there is a well-defined lift to the tangent bundle $\phi_*: TV\to TV$ and therefore a lift $\tilde{\phi}$ to the bundle of linear frames. In fibre coordinates, if ξ is the vector field on V generating a one-parameter group of diffeomorphisms $\{\phi_t\}$, the generator $\tilde{\xi}$ of $\{\tilde{\phi}_t\}$ on the bundle of linear frames has the form

$$\tilde{\xi} = \xi^\alpha\frac{\partial}{\partial x^\alpha} + e^\beta_a\frac{\partial\xi^\alpha}{\partial x^\beta}\frac{\partial}{\partial e^\alpha_a}$$

whence it follows that $\tilde{\xi}\lrcorner\,\theta^a = \xi^a$ and $\tilde{\xi}\lrcorner\,\omega^a{}_b = \nabla_b\xi^a$. By (3.3) the condition that $\tilde{\xi}$ be tangent to the Lorentz bundle P on P is $\eta_{ar}\nabla_b\xi^r + \eta_{br}\nabla_a\xi^r = 0$; in other words, ξ must be a Killing vector field of the metric g in order to lift to P. Computing $\pounds_{\tilde{\xi}}\theta$ for this case, we find

$$\pounds_{\tilde{\xi}}\theta = e\pounds_\xi A + \mu s(\pounds_\xi F)_{23}\theta^0,$$

where $\pounds_\xi A$ and $\pounds_\xi F$ stand for the respective pull backs to P of the Lie derivatives of A and F in V.

Solving $\pounds_{\tilde{\xi}}\omega = d\pounds_{\tilde{\xi}}\theta = 0$ for ξ gives $\pounds_\xi F = 0$ whether s is zero or not. Thus, we recover the result that the "obvious" symmetries ξ for the charged particle with or without spin are those that leave the gravitational field and the electromagnetic field invariant, i.e., satisfy $\pounds_\xi g = \pounds_\xi F = 0$.

The corresponding conserved quantity $f_{\tilde{\xi}}$ is given locally up to a constant by

$$df_{\tilde{\xi}} = -\tilde{\xi}\lrcorner\,\omega = d(M\xi^0 + s\nabla_3\xi_2) - e\xi\lrcorner\,F.$$

If one introduces a vector potential A satisfying $\pounds_\xi A = 0$ and $dA = F$ (which can always be done locally), $f_{\tilde{\xi}}$ can be written

$$f_{\tilde{\xi}} = M\xi^0 + s\nabla_3\xi_2 + e\xi^rA_r.$$

(Note that except for the arbitrary constant, $f_{\tilde{\xi}}$ is not gauge dependent.)

In the case where ξ is a distinguished timelike Killing vector field of a stationary space–time also

satisfying $\pounds_\xi F = 0$, one would call the corresponding f_ξ the *total energy* E of the particle, in tensor form,

$$E = -\xi^\alpha p_\alpha - \tfrac{1}{2} S^{\alpha\beta}\nabla_\alpha \xi_\beta + e\xi^\alpha A_\alpha.$$

6. CONCLUSION

This derivation of the equations of motion for particles with spin in general relativity does not, of course, replace Dixon's derivation from the conservation laws of the field theory. It does, however, confirm his conclusion that they seem to be the best available equations for a classical spinning particle. Whether or not one finds it desirable to equip all classical dynamical systems (of not purely phenomenological type) with a natural canonical structure, this approach shows that replacing the Lagrangian as the starting point by the presymplectic potential θ or the presymplectic 2-form ω gives more flexibility in the choice of the evolution space and avoids the introduction of somewhat artificial position variables (like angles whose canonical momenta are the spin components). Nevertheless, the choice of a symplectic potential as in (3.4) may not be more difficult or more arbitrary than the choice of a Lagrangian. (In both cases one is guided by the flat space version of the theory.)

In his Minkowski space model of the free spinning particle, Souriau[1] is also able to describe particles with zero rest mass and nonzero spin s whose trajectories in space–time turn out to be null hypersurfaces or wavefronts. It is not hard to set up the corresponding model in our framework (representing the Lorentz bundle P by null tetrads instead of quasiorthonormal tetrads). It turns out, however, that the difficulties encountered in the computation of ker ω on P (in Sec. 3) for large curvatures and spin present themselves in the zero-rest-mass case for arbitrarily small nonzero curvature (or electromagnetic fields) already. That is, the dimension of ker ω will change for different positions, spins, and momenta, thus not defining even locally a proper quotient manifold. Moreover, the trajectory in space–time, even if well defined, does not, as one might have hoped, become a null line or null hypersurface, but can be a spacelike curve or even a spacelike surface, making a physical interpretation almost impossible. One might therefore conclude that Souriau's classical zero-rest-mass particle with spin is extremely unstable in interactions with exterior gravitational and electromagnetic fields.

Finally, it is quite possible that more complex systems (particles with more internal degrees of freedom) can be treated in a similar way with some fibre bundle over space–time as evolution space. The Lorentz bundle P, however, having dimension ten only cannot be used for a system with more than four degrees of freedom like the spinning particle just described.

ACKNOWLEDGMENT

The author would like to thank W. Kundt for several discussions and for reading the manuscript.

---

* Supported in part by the National Research Council of Canada.

1 J.-M. Souriau, *Structure des systèmes dynamiques* (Dunod, Paris, 1970).

2 J.-M. Souriau, Commun. Math. Phys. **1**, 374 (1966).

3 W. G. Dixon, Nuovo Cimento **34**, 317 (1964).

4 W. G. Dixon, Proc. Roy. Soc. London **A314**, 499 (1970).

5 Cf. Refs. 3 and 4 for references to the earlier literature on the spinning particle in electromagnetic and gravitational fields.

6 The general definitions hold also for field theories if one allows infinite-dimensional manifolds. But in the following the term "classical system" will always refer to one with finitely many degrees of freedom.

7 R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin, New York, 1967).

8 C. Godbillon, *Géométrie differentielle et mécanique analytique* (Hermann, Paris, 1969).

9 This definition is slightly more general than the one proposed by J. Sniatycki and W. M. Tulczyjew, Ann. Inst. Henri Poincaré **15**, 177 (1971), which would not apply to our description of the spinning particle evolution space.

10 $Z \lrcorner \omega$ denotes the contraction of the tangent vector Z with the p-form ω, in coordinates, if $Z = Z^\alpha \partial_\alpha$ and $\omega = (1/p!) \times \omega_{\alpha_1 \cdots \alpha_p} dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_p}$, then $Z \lrcorner \omega = [1/(p-1)!] \times Z^{\alpha_1} \omega_{\alpha_1 \alpha_2 \cdots \alpha_p} dx^{\alpha_2} \wedge \cdots \wedge dx^{\alpha_p}$.

11 See, for example, J. Dieudonné, *Éléments d'analyse* (Gauthier-Villars, Paris, 1970), Vol. 3, p. 58, or R. S. Palais, Mem. Amer. Math. Soc. **22**, (1957).

12 Ref. 1, p. 189.

13 See, for example, S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry. I* (Interscience, New York, 1963).

14 $\pounds_Z K$ denotes the Lie derivative along the vector field Z of a tensor field K.

15 R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966).

16 η_{ab} = diag (−1, 1, 1, 1). Latin indices will denote tetrad components, Greek indices components with a respect to a local coordinate system (x^α) in V. They will be raised and lowered by η_{ab} and $g_{\alpha\beta}$, respectively.

17 See Ref. 13, p. 142.

18 $R^a{}_{bcd} = R^\alpha{}_{\beta\gamma\delta} \theta^a_\alpha e^\beta_b e^\gamma_c e^\delta_d$ are the tetrad components of the curvature tensor.

19 J.-M. Souriau, Compt. Rend. **263**, 1191 (1966).

20 J.-M. Souriau, Compt. Rend. **265**, 165 (1967).

21 Ref. 1, p. 182.

22 If it is desirable to avoid the gauge arbitrariness in A, one can directly define the presymplectic form $\omega = d(M\theta^0 + s\omega_{23}) + eF$.

23 $\nabla_a F_{bc}$ are the tetrad components of the tensor $\nabla_\alpha F_{\beta\gamma} : \nabla_a F_{bc} := \nabla_\alpha F_{\beta\gamma} e^\alpha_a e^\beta_b e^\gamma_c$.

24 Ref. 1, p. 206.

25 V. Bargmann, L. Michel, and V. L. Telegdi, Phys. Rev. Letters **2**, 435 (1959).

Spin-Dependent Operators in the Occupation-Branching Number Representation

J. F. Gouyet
Laboratoire d'Optique et Physique Moléculaire, Ecole Polytechnique, Paris, France
(Received 4 January 1971)

In two preceding papers, a general method was given of building up open-shell wavefunctions which are eigenfunctions of the spin operators S^2. This method leads to a second quantizationtype formalism, where degenerate multiplets are described in what we call the occupation-branching number representation. In the paper referred to as Paper II, equivalent monoelectronic and bielectronic operators were defined and a diagrammatic technique was developed. In the present paper, the equivalent spin-dependent operators in the occupation-branching number representation are built up, in the cases of monoelectronic and bielectronic operators. The spin–orbit and the spin–spin coupling are the most important operators of this kind. The associated diagrammatic representation is given.

I. INTRODUCTION

Diagrammatic techniques using the perturbation theory[1] have been applied in nuclear structure[2] and, more recently, in atomic and molecular problems[3] in cases of nondegenerate unperturbed states. The general extension of the linked-diagram expansion for the degenerate states of a system of fermions has been given by Kato,[4] Bloch,[5] and Bloch and Horowitz,[6] and with use of degenerate Brillouin–Wigner perturbation theory by Brandow.[7] Treatments of configuration interaction in open-shell atoms are given by Sandars[8] and Tolmachev.[9]

An extension of the linked diagram expansion for degenerate states including spin symmetry has been given in two previous works.[10,11] The unperturbed states of a given configuration have a definite spin S and a supplementary quantum number γ connected with the branching path, which distinguishes the different degenerate states. The building up principle is closely related to the second quantization. The representation was called[11] occupation-branching number representation. Equivalent mono and bielectronic operators were defined but only for spin dependent operators. In this paper, we build up the equivalent operators in the spin-dependent case. Diagrams associated with the matrix elements between excited configurations, which play the most important part in the perturbation theory, are given for these operators.

II. THE OCCUPATION–BRANCHING NUMBER REPRESENTATION[11]

A molecular, or more generally p-body, unperturbed eigenstate, is defined by the occupation number N_i of a couple $(\xi_i \Sigma_i)$ given in a standard order. ξ_i is the orbital quantum number and $\Sigma_i = \pm \frac{1}{2}$ gives the way in which the electron of the ith occupied orbital is coupled to the foregoing electrons with respect to the spin.

Let

$$|\{N\}\rangle = |\cdots N_{2k}(\xi_k - \tfrac{1}{2}) N_{2k-1}(\xi_k \tfrac{1}{2}) \cdots N_2(\xi_1 - \tfrac{1}{2}) \times N_1(\xi_1 \tfrac{1}{2}))\rangle, \quad N_i = 0, 1 \tag{1}$$

be a state with p electrons and with total spin

$$S = \tfrac{1}{2}(N_1 - N_2) + \cdots + \tfrac{1}{2}(N_{2k-1} - N_{2k}) + \cdots \equiv \Sigma_1 + \Sigma_2 + \cdots + \Sigma_p, \tag{2}$$

where $\Sigma_1, \ldots, \Sigma_p$ are the Σ_i of the occupied $(\xi_i \Sigma_i)$ (that is to say, $N_i = 1$).

Such a state, with ordered $(\xi_i \Sigma_i)$, is called a standard state.

The different manners in which the Σ's are distributed are associated with a path γ of the branching diagram

$$\gamma = (\Sigma_p, \ldots, \Sigma_2, \Sigma_1). \tag{3}$$

For example, the two doublets of a three-open-shell system have the two following γ:

$$\gamma_1 = (\tfrac{1}{2}, -\tfrac{1}{2}, \tfrac{1}{2}; \underbrace{-\tfrac{1}{2}, \tfrac{1}{2}, \ldots, -\tfrac{1}{2}, \tfrac{1}{2}}_{\text{closed shells}}),$$

$$\gamma_2 = (-\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}; \underbrace{-\tfrac{1}{2}, \tfrac{1}{2}, \ldots, -\tfrac{1}{2}, \tfrac{1}{2}}_{\text{closed shells}}).$$

A representation of the two γ, corresponding to the two doublets is given in Fig. 1.

A scalar product can be defined and we have shown[11] that

$$\langle\langle\{N\} \mid \{N'\}\rangle\rangle = \delta_{NN'}. \tag{4}$$

Note that on paper II,[11] an ambiguity remains in the normalization of $|\{N\}\rangle\rangle$ in the case of complete shells following shells of which the total spin S_c is different from zero. In this case, the coefficient of normalization N_c of $|\{N\}\rangle\rangle$ must be

$$N_c = \prod_c \left(\frac{2S_c + 1}{2S_c + 2}\right)^{1/2}.$$

Naturally when all $S_c = 0$, it reduces to $N_c = 2^{-c/2}$. More details can be found in Appendix C.

The creation operator $\mathfrak{B}^{\xi\Sigma\dagger}$ creates an electron on the orbital ξ coupled, to give the total spin $S' = S + \Sigma$:

$$\mathfrak{B}^{\xi\Sigma\dagger} |\{N\}\rangle\rangle = |1_{\xi\Sigma}\{N\}\rangle\rangle.$$

This state is, in general, not standard and not normalized to unity.

An operator $\mathfrak{B}^{\xi\Sigma\dagger}$ can be written in the form

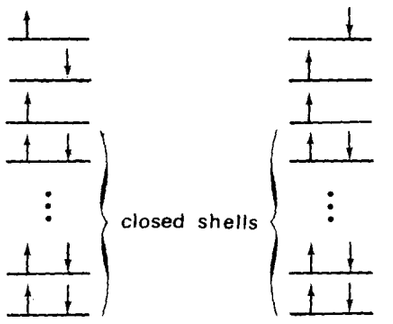


FIG. 1. First doublet with $\gamma = \gamma_1$; second doublet with $\gamma = \gamma_2$.

$$\mathfrak{B}^{\xi\Sigma\dagger} = w\mathfrak{A}^{\xi\Sigma\dagger} = wa^{\xi\sigma\dagger}S_{\sigma}^{\Sigma\dagger} \tag{5}$$

with a summation on $\sigma = \pm\frac{1}{2}$; $a^{\xi\sigma\dagger}$ is the usual creation operator of a spin orbital $(\xi\sigma)$; $S_{\sigma}^{\Sigma\dagger}$ an operator[12] giving the coupling scheme and $W = [S_{op}]^{-1/2}$ is a normalization operator (inverse square root of the spin multiplicity).

The annihilator is the adjoint of $\mathfrak{B}^{\xi\Sigma\dagger}$:

$$\mathfrak{B}^{\xi\Sigma} = \mathfrak{A}^{\xi\Sigma}\cdot w = a^{\xi\sigma}S_{\sigma}^{\Sigma}\cdot w. \tag{6}$$

III. REDUCED MATRIX ELEMENTS

Let us consider two states,

$$|{}^{[S]}N_{\gamma}\rangle\rangle = N_c\,\mathfrak{B}^{\xi_n\Sigma_n\dagger}\cdots\mathfrak{B}^{\xi_1\Sigma_1\dagger}|0\rangle\rangle, \tag{7}$$

$$|{}^{[S']}N'_{\gamma'},\rangle\rangle = N_{c'}\,\mathfrak{B}^{\xi'_n\Sigma'_n\dagger}\cdots\mathfrak{B}^{\xi'_1\Sigma'_1\dagger}|0\rangle\rangle. \tag{8}$$

The $(\xi\Sigma)$ are in a standard order and $N_c, N_{c'}$ are the normalization coefficients. Let us consider a spin-dependent operator ${}^{[s_1]}\hat{O}$ of which components are $O_{S_1}^{M_1}$. Then the reduced matrix element can be written [see Paper I, Eq. (31)[10]]

$$\begin{aligned}\langle N'_{S'\gamma'}\|O_{S_1}\|N_{S\gamma}\rangle &\equiv \langle\langle {}^{[S']}N'_{\gamma'}|{}^{[S_1]}\hat{O}|{}^{[S]}N_{\gamma}\rangle\rangle\\ &= \langle 0|\,a^{\xi'_1\sigma'_1}\cdots a^{\xi'_n\sigma'_n}O_{S_1}^{M_1}a^{\xi_n\sigma_n\dagger}\cdots a^{\xi_1\sigma_1\dagger}\,|0\rangle\\ &\quad\times(z^0|\,S_{\sigma'_1}^{\Sigma'_1}w\cdots S_{\sigma'_n}^{\Sigma'_n}ww_{M_1}^{S_1}wS_{\sigma_n}^{\Sigma_n\dagger}\cdots wS_{\sigma_1}^{\Sigma_1\dagger}\,|z^0),\end{aligned} \tag{9}$$

where $w_{M_1}^{S_1}$ has been defined as the components of a unit operator W^{S_1}.

IV. MONOELECTRONIC OPERATORS

The operator ${}^{[S_1]}F$ can be written

$${}^{[S_1]}\hat{F} = \sum_i {}^{[S_1]}\hat{f}(i), \tag{10}$$

i designating the coordinates of the ith particle.

The components are $F_{S_1}^{M_1}$:

$${}^{[S_1]}\hat{F} = F_{S_1}^{M_1}w_{M_1}^{S_1}, \qquad F_{S_1}^{M_1} = \sum_i f_{S_1}^{M_1}(i). \tag{11}$$

Then the reduced matrix element takes the following form;

$$\begin{aligned}&\langle\langle {}^{[S']}\Psi'_{\gamma'}|{}^{[S_1]}\hat{F}|{}^{[S]}\Psi_{\gamma}\rangle\rangle\\ &\quad= N_{c'}N_c\sum_p(-1)^p\sum_i\prod_{j\neq i}\delta^{\xi'_j,\xi_{pj}}\delta^{\sigma'_j,\sigma_{pj}}\\ &\qquad\times\langle\xi'_i\sigma'_i|\,f_{S_1}^{M_1}\,|\xi_i\sigma_i\rangle\\ &\qquad\times(z^0|\,S_{\sigma'_1}^{\Sigma'_1}w\cdots S_{\sigma'_n}^{\Sigma'_n}ww_{M_1}^{S_1}wS_{\sigma_n}^{\Sigma_n\dagger}\cdots wS_{\sigma_1}^{\Sigma_1\dagger}|z^0),\end{aligned}$$

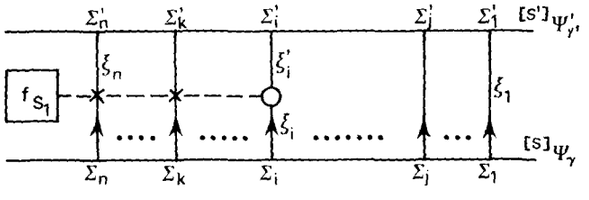


FIG. 2.

where p is a permutation of the spin-orbitals (ξ_j, σ_j).

Using the Wigner–Eckart theorem

$$\langle\xi'_i\sigma'_i|\,f_{S_1}^{M_1}\,|\xi_i\sigma_i\rangle = -\begin{pmatrix}\frac{1}{2} & M_1 & \sigma_i\\ \sigma'_i & S_1 & \frac{1}{2}\end{pmatrix}\langle\xi'_i\|f_{S_1}\|\xi_i\rangle$$

and writing

$${}^{[S_1]}T^{\sigma'_i,\sigma_i} = -\begin{pmatrix}\frac{1}{2} & M_1 & \sigma_i\\ \sigma'_i & S_1 & \frac{1}{2}\end{pmatrix}w_{M_1}^{S_1},$$

we have

$$\begin{aligned}&\langle\langle {}^{[S']}\Psi'_{\gamma'}|{}^{[S_1]}\hat{F}|{}^{[S]}\Psi_{\gamma}\rangle\rangle\\ &\quad= N_{c'}N_c\sum_p(-1)^p\sum_i\prod_{j\neq i}\delta^{\xi'_j,\xi_{pj}}\delta_j^{\sigma'_j\sigma_{pj}}\\ &\qquad\times(z^0|\,S_{\sigma'_1}^{\Sigma'_1}w\cdots S_{\sigma'_n}^{\Sigma'_n}w^{[S_1]}T^{\sigma'_i,\sigma_i}wS_{\sigma_n}^{\Sigma_n\dagger}\cdots wS_{\sigma_1}^{\Sigma_1\dagger}|z^0).\end{aligned} \tag{12}$$

First, this formula enables us to define an equivalent operator, and secondly, to interpret graphically the reduced matrix element.

To seek the equivalent operator we use an identification method. We have shown in Sec. II that a monoelectronic scalar equivalent operator can be written

$$\begin{aligned}F_{\text{equ}} = w^{-1}\,{}^{[0]}\hat{F} &= \sum_{\xi\xi'\Sigma}\mathfrak{B}^{\xi'\Sigma\dagger}\langle\xi'|f|\xi\rangle\mathfrak{B}^{\xi\Sigma}\\ &= \sum_{\xi\xi'\Sigma}\langle\xi'|f|\xi\rangle\,a^{\xi'\sigma\dagger}a^{\xi\sigma}wS_{\sigma'}^{\Sigma\dagger}S_{\sigma}^{\Sigma}w.\end{aligned} \tag{13}$$

As the reduced matrix element $\langle\xi'\|f\|\xi\rangle$ is equal to $\sqrt{2}\,\langle\xi'|f|\xi\rangle$, then, if we try to find ${}^{[S_1]}F_{\text{equ}}$, we obtain

$$\begin{aligned}w\,{}^{[S_1]}F_{\text{equ}} &= {}^{[S_1]}\hat{F}\\ &= \sum_{\substack{\xi'\xi\\ \Sigma'\Sigma}}\langle\xi'\|f_{S_1}\|\xi\rangle\,a^{\xi'\sigma'\dagger}a^{\xi\sigma}wS_{\sigma'}^{\Sigma'\dagger}C_{\Sigma'\Sigma}^{\text{op}}S_{\sigma}^{\Sigma}w,\end{aligned}$$

where $C_{\Sigma'\Sigma}^{\text{op}}$ is an operator to be determined.

We have to work out the identification

$$\delta^{\sigma'_i\sigma'}\delta^{\sigma_i\sigma}wS_{\sigma'}^{\Sigma'\dagger}C_{\Sigma'\Sigma}^{\text{op}}S_{\sigma}^{\Sigma}w \equiv {}^{[S_1]}T^{\sigma'_i;\sigma_i}.$$

The values $C_{\Sigma'\Sigma}$ of $C_{\Sigma'\Sigma}^{\text{op}}$ are obtained by the identification

$$(z_{M'}^{S'}|\,{}^{[S_1]}T^{\sigma';\sigma}\,|z_M^S) \equiv \sum_{\Sigma'\Sigma}C_{\Sigma'\Sigma}(z_{M'}^{S'}|wS_{\sigma'}^{\Sigma'\dagger}S_{\sigma}^{\Sigma}w|z_M^S)$$

for all S, M, S', M' of the basis spin vectors $|z_M^S\rangle$.

It is shown in Appendix A that

$$C_{\Sigma'\Sigma} = (-1)^{1/2+S+S_0+S_1}\begin{Bmatrix}\frac{1}{2} & \frac{1}{2} & S_1\\ S' & S & S_0\end{Bmatrix} \tag{14}$$

with $S_0 = S - \Sigma = S' - \Sigma'$.

Then

$$C_{\Sigma'\Sigma}^{\text{op}} = (-1)^{S_1}(-1)^{1/2-\Sigma'}(-1)^{[S_{\text{op}}]}\times\begin{Bmatrix}\frac{1}{2} & \frac{1}{2} & S_1\\ S_{\text{op}}+\Sigma' & S_{\text{op}}+\Sigma & S_{\text{op}}\end{Bmatrix} \tag{15}$$

so that

$$C_{\Sigma'\Sigma}^{\text{op}}|\,z_{M_0}^{S_0}) = C_{\Sigma'\Sigma}|\,z_{M_0}^{S_0} = C_{\Sigma'\Sigma}|\,z_{M_0}^{S_0}).$$

The coefficient $C_{\Sigma'\Sigma}$ has an important asymptotic property[15]

$$C_{\Sigma'\Sigma} \underset{S_0\to\infty}{\sim} \frac{(-1)^{S_1-\Lambda_1}}{[S_0]^{\frac{1}{2}}} \begin{pmatrix} \frac{1}{2} & \Sigma & \Lambda_1 \\ \Sigma' & \frac{1}{2} & S_1 \end{pmatrix}, \quad \text{where } \Lambda_1 = \Sigma' - \Sigma. \tag{16}$$

The equivalent operator can now be written as

$$^{[S_1]}\hat{F} = \sum_{\substack{\xi'\xi \\ \Sigma'\Sigma}} \langle \xi' \| {}^{[S_1]} f \| \xi \rangle \, \mathfrak{B}^{\xi'\Sigma'\dagger} \, C^{\mathrm{op}}_{\Sigma'\Sigma} \mathfrak{B}^{\xi\Sigma}. \tag{17}$$

We then use the generalized definition

$$\langle (\xi'\Sigma' | {}^{[S_1]} f | \xi\Sigma) \rangle = \langle \xi' \| f_{S_1} \| \xi \rangle \, C^{\mathrm{op}}_{\Sigma'\Sigma}, \tag{18}$$

which gives for $^{[S_1]}F$:

$$^{[S_1]}\hat{F} = \sum_{\substack{\xi'\xi \\ \Sigma'\Sigma}} \mathfrak{B}^{\xi'\Sigma'\dagger} \langle (\xi'\Sigma' | {}^{[S_1]} f | \xi\Sigma) \rangle \, \mathfrak{B}^{\xi\Sigma}. \tag{19}$$

Now graphically we can use formula (12) and suppose the sets $\{\xi'_j\}$ and $\{\xi_j\}$ in a standard order. For instance, if the two sets differ only by the orbitals ξ'_i and ξ_j, respectively,

$$\begin{aligned} \langle (^{[S']}\Psi'_{\gamma'} | {}^{[S_1]}\hat{F} | {}^{[S]}\Psi_\gamma) \rangle &= \langle \xi'_i \| f_{S_1} \| \xi_i \rangle \prod_{j\neq i} \delta^{\sigma'_j,\sigma_j} \\ &\times (z^0 | S^{\Sigma'_1}_{\sigma'_1} w \cdots S^{\Sigma'_i}_{\sigma'_i} w \cdots S^{\Sigma'_n}_{\sigma'_n} w \\ &\times {}^{[S_1]}T^{\sigma_i;\sigma_i} w S^{\Sigma_n\dagger}_{\sigma_n} \cdots w S^{\Sigma_i\dagger}_{\sigma_i} \cdots w S^{\Sigma_1\dagger}_{\sigma_1} | z^0 \end{aligned}$$

Using the results of Appendix B, it is possible to calculate the expression

$$\begin{aligned} &\prod_{j\neq i} \delta^{\sigma'_j\sigma_j} (z^0 | S^{\Sigma'_1}_{\sigma'_1} w \cdots S^{\Sigma'_i}_{\sigma'_i} w \cdots S^{\Sigma'_n}_{\sigma'_n} \\ &\quad \times {}^{[S_1]}T^{\sigma'_i\sigma_i} w S^{\Sigma_n\dagger}_{\sigma_n} \cdots w S^{\Sigma_i\dagger}_{\sigma_i} \cdots w S^{\Sigma_1\dagger}_{\sigma_1} | z^0) \\ &= \prod_{j=1}^{i-1} \delta^{\Sigma'_j\Sigma_j} O^{\Sigma'_i\Sigma_i}_{S_1}(s_{i-1}) \prod_{k=i+1}^{n} X^{\Sigma'_k\Sigma_k}_{S_1}(s_{k-1}, s'_{k-1}) \end{aligned} \tag{20}$$

with $s_k = \Sigma_1 + \Sigma_2 + \cdots \Sigma_k$, $s'_k = \Sigma'_1 + \Sigma'_2 + \cdots \Sigma'_k$, and

$$O^{\Sigma'_1\Sigma}(s) = (-1)^{1/2-\Sigma} \times \left(\frac{(1+\Sigma'-\Sigma)!(1+\Sigma-\Sigma')!(2s+1+2\Sigma+2\Sigma')}{3![s](2s+1+\Sigma+\Sigma')} \right)^{1/2} \times C_{\Sigma\Sigma'},$$

$(s+\Sigma)^{1/2}(s'+\Sigma')^{1/2} X^{\Sigma'\Sigma}_{S_1}(s,s')$ being given by Table I.

This can be graphically represented by Fig. (8). The cross on the line $(\Sigma'_k \Sigma_k)$, $k \geq i$, recalls the existence of a coupling coefficient $X^{\Sigma'_k\Sigma_k}$. The circle on the line $(\Sigma'_i \Sigma_i)$ recalls the action of the spin operator on the propagator of particle i: The coupling operator is $\langle (\xi\Sigma | {}^{[1]}f | \xi'\Sigma') \rangle = \langle \xi \| {}^{[1]}f \| \xi' \rangle \cdot O^{\Sigma'\Sigma}(s_{i-1})$ for a given s_{i-1}.

The graphical representation is shown in Fig. 2.

V. BIELECTRONIC OPERATORS

The operator $^{[S_1]}G$ can be written

$$^{[S_1]}\hat{G} = \tfrac{1}{2} \sum_{i\neq j} {}^{[S_1]}\hat{g}(i,j). \tag{21}$$

The components of $^{[S_1]}g$ are $g^{M_1}_{S_1}$ and we consider the following form for $g^{M_1}_{S_1}$:

$$g^{M_1}_{S_1}(i,j) \equiv g(i,j) \{ u^{M_a}_{S_a}(i) \, v^{M_b}_{S_b}(j) \}^{M_1}_{S_1} = g \times \begin{pmatrix} M_1 & S_a & S_b \\ S_1 & M_a & M_b \end{pmatrix} u^{M_a}_{S_a} v^{M_b}_{S_b}, \tag{22}$$

where $g^{(i,j)}$ is the space-bielectronic part and $^{[S_a]}u$, $^{[S_b]}v$ purely spin operators coupled in S_1.

Then the reduced matrix element is

$$\begin{aligned} &\langle (^{[S']}\Psi'\gamma' | {}^{[S_1]}\hat{G} | {}^{[S]}\Psi\gamma) \rangle \\ &= \tfrac{1}{2} N_{c'} N_c \sum_p (-1)^p \sum_{i\neq j} \prod_{k\neq i,j} \delta^{\xi'_k,\xi_{pk}} \delta^{\sigma'_k,\sigma_{pk}} \\ &\quad \times \langle \xi'_i \sigma'_i \xi'_j \sigma'_j | g^{M_1}_{S_1} | \xi_i \sigma_i \xi_j \sigma_j \rangle \langle z^0 | S^{\Sigma'_1}_{\sigma'_1} w \cdots \\ &\quad \times S^{\Sigma'_n}_{\sigma'_n} w w^{S_1}_{M_1} w S^{\Sigma_n\dagger}_{\sigma_n} \cdots w S^{\Sigma_1\dagger}_{\sigma_1} | z^0 \rangle. \end{aligned} \tag{23}$$

We define an operator T' by the relation

$$\begin{aligned} &\langle \xi'_i \sigma'_i \xi'_j \sigma'_j | g^{M_1}_{S_1} | \xi_i \sigma_i \xi_j \sigma_j \rangle w^{S_1}_{M_1} \\ &= \langle \xi'_i \xi'_j | g | \xi_i \xi_j \rangle \langle \tfrac{1}{2} \| u_{S_a} \| \tfrac{1}{2} \rangle \langle \tfrac{1}{2} \| v_{S_b} \| \tfrac{1}{2} \rangle \\ &\quad \times {}^{(S_a S_b) S_1} T^{\sigma'_i\sigma'_j;\sigma_j\sigma_i} \end{aligned}$$

with

$$\begin{aligned} {}^{(S_a S_b) S_1} T^{\sigma'_i\sigma'_j;\sigma_j\sigma_i} &= w^{S_1}_{M_1} \begin{pmatrix} M_1 & S_a & S_b \\ S_1 & M_a & M_b \end{pmatrix} \begin{pmatrix} \frac{1}{2} & M_a & \sigma_i \\ \sigma' & S_a & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & M_b & \sigma'_j \\ \sigma'_j & S_b & \frac{1}{2} \end{pmatrix}. \end{aligned}$$

The problem is now to find an operator $C^{\mathrm{op}}_{\Sigma'_a\Sigma'_b\Sigma_b\Sigma_a}$ so that

$$\begin{aligned} &{}^{(S_a S_b) S_1} T^{\sigma'_a\sigma'_b;\sigma_b\sigma_a} \\ &\equiv w S^{\Sigma'_a\dagger}_{\sigma'_a} w S^{\Sigma'_b\dagger}_{\sigma'_b} C^{\mathrm{op}}_{\Sigma'_a\Sigma'_b\Sigma_b\Sigma_a} S^{\Sigma_b}_{\sigma_b} w S^{\Sigma_a}_{\sigma_a} w. \end{aligned}$$

$C^{\mathrm{op}}_{\Sigma'_a\Sigma'_b\Sigma_b\Sigma_a}$ is determined by an identification similar to the one of the monoelectronic case and the demonstration can be found in Appendix D:

$$\begin{aligned} C^{\mathrm{op}}_{\Sigma'_a\Sigma'_b\Sigma_b\Sigma_a} &= (-1)^{(1/2)\Sigma'_b + 2(\Sigma_a+\Sigma_b)+S_a+S_1} \\ &\quad \times [S_1]^{1/2} [S_{\mathrm{op}} + \Sigma_b]^{1/2} [S_{\mathrm{op}} + \Sigma'_b]^{1/2} \\ &\quad \times \begin{Bmatrix} \frac{1}{2} & \frac{1}{2} & S_b \\ S_{\mathrm{op}} + \Sigma'_b & S_{\mathrm{op}} + \Sigma_b & S_{\mathrm{op}} \end{Bmatrix} \\ &\quad \times \begin{Bmatrix} S_a & S_b & S_1 \\ \frac{1}{2} & S_{\mathrm{op}} + \Sigma_b & S_{\mathrm{op}} + \Sigma_a + \Sigma_b \\ \frac{1}{2} & S_o + \Sigma'_b & S_{\mathrm{op}} + \Sigma'_a + \Sigma'_b \end{Bmatrix}. \end{aligned} \tag{24}$$

In the particular case where $S_a = S_b = S_1 = 0$, we find

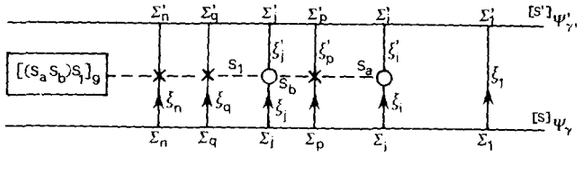


FIG. 3.

$$C_{\Sigma'_a\Sigma'_b\Sigma_b\Sigma_a} = \left(\frac{1}{\sqrt{2}}\right)^2 \frac{\delta^{\Sigma'_a\Sigma_a}\delta^{\Sigma'_b\Sigma_b}}{[S]^{1/2}},$$

which corresponds to the scalar bielectronic operator.

The generalized definition

$$\langle(\xi'_1\Sigma'_1\xi'_2\Sigma'_2|^{[(S_aS_b)S_1]}\hat{g}|\xi_1\Sigma_1\xi_2\Sigma_2)\rangle \equiv \langle\xi'_1\tfrac{1}{2}\xi'_2\tfrac{1}{2}\|g_{(S_aS_b)S_1}\|\xi_1\tfrac{1}{2}\xi_2\tfrac{1}{2}\rangle C^{op}_{\Sigma'_1\Sigma'_2\Sigma_2\Sigma_1} \quad (25)$$

gives

$$^{[S_1]}\hat{G} = \sum_{(\xi,\Sigma)} \mathcal{B}^{\xi'_1\Sigma'_1\dagger}\mathcal{B}^{\xi'_2\Sigma'_2\dagger}\langle(\xi'_1\Sigma'_1\xi_2\Sigma'_2|^{[S_1]}\hat{g}|\xi_1\Sigma_1\xi_2\Sigma_2)\rangle \times \mathcal{B}^{\xi_2\Sigma_2}\mathcal{B}^{\xi_1\Sigma_1}. \quad (26)$$

Graphically we use expression (23) in a similar manner to the monoelectronic case [formula (20)] as it is seen in Appendix E.

The spin-recoupling part gives then

$$M = \prod_{k=1}^{i-1}\delta^{\Sigma'_k\Sigma_k}O^{\Sigma'_i\Sigma_i}_{S_a}(s_i - 1) \times \prod_{k=i+1}^{j-1}X^{\Sigma'_k\Sigma_k}_{S_a}(s_{k-1}, s'_{k-1})O^{\Sigma'_j\Sigma_j}_{S_b(S_aS_1)}(s_{j-1}s'_{j-1}) \times \prod_{k=j+1}^{n}X^{\Sigma'_k\Sigma_k}_{S_1}(s_{k-1}s'_{k-1}), \quad (27)$$

where

$$O^{\Sigma'\Sigma}_{S_b(S_aS_1)}(s,s') = (-1)^{s'-s+2\Sigma-S_b-S_1}[s]^{1/2}[s']^{1/2}[S_1]^{1/2} \times \begin{Bmatrix} S_a & S_b & S_1 \\ s & \frac{1}{2} & s+\Sigma \\ s' & \frac{1}{2} & s'+\Sigma' \end{Bmatrix}.$$

This is graphically shown on Fig. 3.

VI. CONCLUSION

This paper gives in the branching occupation number representation, the expressions for the equivalent monoelectronic and bielectronic spin-dependent operators which can be then inserted in a more general perturbation treatment of excited states. For clarity and simplicity, it has been restricted to particle representation. In the particle-hole representation, hole-hole and particle-hole interactions introduce little modifications in the foregoing derivations.

ACKNOWLEDGMENTS

The author would like to thank Professor R. Lefebvre and Professor F. Teissier du Cros for their interest in this work and Dr. M. Boiteux for many valuable discussions and suggestions.

APPENDIX A: DETERMINATION OF THE $C^{op}_{\Sigma'\Sigma}$ IN $(\xi'\Sigma'|^{[S_1]}f|\xi\Sigma)\rangle$

The expression of the $C_{\Sigma'\Sigma}$ is obtained by the following identification for all S, M, S', M':

$$(z^{S'}_{M'}|^{[S_1]}T^{\sigma'\sigma}|z^{S}_{M}) = \sum_{\Sigma'\Sigma} C_{\Sigma'\Sigma}(z^{S'}_{M'}|wS^{\Sigma'\dagger}_{\sigma'}S^{\Sigma}_{\sigma}w|z^{S}_{M}).$$

Using the expression

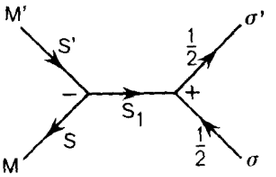
$$(z^{S'}_{M'}|w^{S_1}_{M_1}|z^{S}_{M}) = \begin{pmatrix} M' & S_1 & S \\ S' & M_1 & M \end{pmatrix},$$

$$S^{\Sigma}_{\sigma}w|z^{S}_{M}) = [S-\Sigma]^{1/2}(-1)^{S-M}(-1)^{1/2-\Sigma} \times \begin{pmatrix} \frac{1}{2} & S & S-\Sigma \\ -\sigma & M & -M+\sigma \end{pmatrix}|z^{S-\Sigma}_{M-\sigma}),$$

we obtain, after some manipulations,

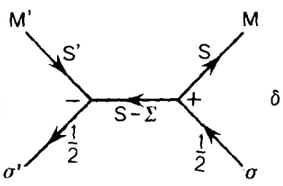
$$\begin{pmatrix} M' & S_1 & S \\ S' & M_1 & M \end{pmatrix}\begin{pmatrix} \frac{1}{2} & M_1 & \sigma \\ \sigma' & S_1 & \frac{1}{2} \end{pmatrix} = \sum_{\Sigma'\Sigma} C_{\Sigma'\Sigma}[S-\Sigma] \times \begin{pmatrix} \frac{1}{2} & M' & -M'+\sigma \\ \sigma' & S' & S'-\Sigma \end{pmatrix}\begin{pmatrix} \sigma & S & S-\Sigma \\ \frac{1}{2} & M & -M+\sigma \end{pmatrix}.$$

This can be graphically represented, by using Yutsis–Levinson and Vanagas angular diagrams:[16]



Part A

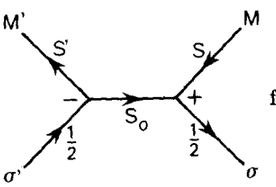
$= \sum_{\Sigma'\Sigma} C_{\Sigma'\Sigma}[S-\Sigma]$



$\delta_{S'-\Sigma',S-\Sigma}$.

Part B
Eq. I

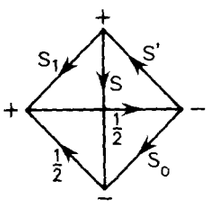
Considering the following angular diagram



for a given S_0,

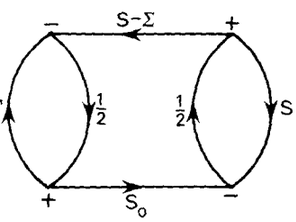
Eq. II

we can join the corresponding free lines of the identity (I) and obtain



Part A

$= C_{\Sigma'\Sigma}[S-\Sigma]$



$\delta_{S-\Sigma,S'-\Sigma'}$.

Part B
Eq. III

Then,

$$C_{\Sigma'\Sigma} = (-1)^{1/2+S+S_0+S_1} \begin{Bmatrix} \frac{1}{2} & \frac{1}{2} & S_1 \\ S' & S & S_0 \end{Bmatrix}$$

with $S_0 = S - \Sigma = S' - \Sigma'$.

Two cases are possible:

a. $S_1 = 0$, then

$$C_{\Sigma'\Sigma} = \frac{\delta_{\Sigma'\Sigma}}{\sqrt{2}[S]^{1/2}} \quad \text{and}$$

$$^{[0]}T^{\sigma'\sigma} = \frac{1}{\sqrt{2}}\left(\sum_{\Sigma} w S^{\Sigma\dagger}_{\sigma'} S^{\Sigma}_{\sigma} w\right)[S_{\text{op}}]^{-1/2}$$

b. $S_1 = 1$, then

$$C^{\text{op}}_{\Sigma'\Sigma} = \frac{(-1)^{(1/2)-\Sigma'}}{[S_{\text{op}}]^{1/2}} \left(\frac{(1+\Sigma'-\Sigma)!(1+\Sigma-\Sigma')!}{3!}\right)^{1/2} \times \left(\frac{[S_{\text{op}}]+2(\Sigma+\Sigma')}{[S_{\text{op}}]+(\Sigma+\Sigma')}\right)^{1/2}.$$

APPENDIX B: DEMONSTRATION OF EQ. 20

To calculate the expression

$$M = \prod_{j\neq i} \delta^{\sigma'_j\sigma_j} (z^0 | S^{\Sigma'_1}_{\sigma'_1} w \cdots S^{\Sigma'_i}_{\sigma'_i} w \cdots S^{\Sigma'_n}_{\sigma'_n} w \times {}^{[S_1]}T^{\sigma'_i\sigma_i} w S^{\Sigma_n\dagger}_{\sigma_n} \cdots w S^{\Sigma_i\dagger}_{\sigma_i} \cdots w S^{\Sigma_1\dagger}_{\sigma_1} | z^0),$$

we use a recursion formula. Then writing $S = \bar{S} + \Sigma$, $M = M + \sigma$ and $S' = \bar{S}' + \Sigma', \bar{M}' = M' + \sigma'$, we first determine

$$K = \delta^{\sigma'\sigma}(z^{\bar{S}'}_{\bar{M}'} | S^{\Sigma'}_{\sigma'} w w^{S_1}_{M_1} w S^{\Sigma\dagger}_{\sigma} | z^{\bar{S}}_{\bar{M}}).$$

Using the definitions of $W^{S_1}_{M_1}$ and $S^{\Sigma\dagger}_{\sigma}$ acting on $|z^{\bar{S}}_{\bar{M}}\rangle$, we obtain

$$K = (-1)^{(1/2)+S'+\bar{S}+S_1}[\bar{S}]^{1/2}[\bar{S}']^{1/2} \begin{Bmatrix} S & \bar{S} & \frac{1}{2} \\ \bar{S}' & S' & S_1 \end{Bmatrix} \times (z^{\bar{S}'}_{\bar{M}'} | w^{S_1}_{M_1} | z^{\bar{S}}_{\bar{M}}) = (z^{\bar{S}'}_{\bar{M}'} | w^{S_1}_{M_1} | z^{\bar{S}}_{\bar{M}}) \times X.$$

Note that for $S_1 = 0$ we simply have the case of scalar operators

$$X = \delta^{\bar{S}\bar{S}'}\delta^{\Sigma'\Sigma}[S]^{-1/2}.$$

The other possible case is $S_1 = 1$, that is to say, the case of spin-dependent operators.

The table giving the possible values of $[S]^{1/2}[S']^{1/2} X(\bar{S}, \bar{S}')$ is given in Table I.

The last expression to be determined is

$$-\begin{pmatrix} \frac{1}{2} & M_1 & \sigma \\ \sigma' & S_1 & \frac{1}{2} \end{pmatrix} (z^{\bar{S}'}_{\bar{M}'} | S^{\Sigma'}_{\sigma'} w w^{S_1}_{M_1} w S^{\Sigma\dagger}_{\sigma} | z^{\bar{S}}_{\bar{M}}),$$

which is equal to

$$(-1)^{(1/2)+S'+\bar{S}+S_1} \begin{Bmatrix} \frac{1}{2} & \frac{1}{2} & S_1 \\ S' & S & \bar{S} \end{Bmatrix} \delta^{\bar{S}\bar{S}'}.$$

Then the element M is

TABLE I

$\Sigma' \backslash \Sigma$	$\frac{1}{2}$	$-\frac{1}{2}$
$\frac{1}{2}$	$\sqrt{(\bar{S}+\bar{S}'+3)(\bar{S}+\bar{S}')}$	$-\sqrt{(\bar{S}-\bar{S}'+1)(\bar{S}'-\bar{S}+2)}$
$\frac{1}{2}$	$\sqrt{(\bar{S}'-\bar{S}+1)(\bar{S}-\bar{S}'+2)}$	$\sqrt{(\bar{S}+\bar{S}'-1)(\bar{S}+\bar{S}'+2)}$

$$M = \prod_{j=1}^{i=1} \delta^{\Sigma'_j\Sigma_j} \times (-1)^{(1/2)+s'_{i-1}+S_1} \begin{Bmatrix} \frac{1}{2} & \frac{1}{2} & S_1 \\ s'_i & s_i & s_{i-1} \end{Bmatrix} \times \prod_{k=1}^{n-1} (-1)^{1/2+s'_{k+1}+s_k+S_1} \times [s_k]^{1/2}[s'_k]^{1/2} \begin{Bmatrix} s_{k+1} & s'_{k-1} & S_1 \\ s'_k & s_k & \frac{1}{2} \end{Bmatrix},$$

where $s_k = \Sigma_1 + \Sigma_2 + \cdots + \Sigma_k$ and $s'_k = \Sigma'_1 + \Sigma'_2 + \cdots + \Sigma'_k$.

APPENDIX C: ABOUT NORMALIZATION OF THE STATES

Normalization must be stated precisely as follows. In Paper I we have put[1]

$$\langle\langle {}^{[S']}\Psi_{\gamma'} | {}^{[S]}\Psi_{\gamma}\rangle\rangle = \langle \Psi_{S'\gamma'} \| \Psi_{S\gamma}\rangle \equiv \delta_{\gamma'\gamma}[S]^{1/2}.$$

But in the occupation branching representation, it is easier to normalize to unity. This is done in Paper II and with the foregoing definitions of $|{}^{[S']}N_{\gamma'}\rangle\rangle$ and $|{}^{[S]}N_{\gamma}\rangle\rangle$ (formulas 7 and 8).

It is then, more convenient to introduce the operator $^{[0]}1 \equiv w$, so that

$$\langle\langle {}^{[S]}\Psi_{\gamma'} | {}^{[0]}1 | {}^{[S]}\Psi_{\gamma}\rangle\rangle \equiv \langle \Psi_{S'\gamma'} \| \Psi_{S\gamma}\rangle \equiv \delta_{\gamma\gamma'}[S]^{1/2},$$

while

$$\langle\langle {}^{[S']}\Psi_{\gamma'} | {}^{[S]}\Psi_{\gamma}\rangle\rangle \equiv \langle \Psi_{S'\gamma'} \| w^{-1} \| \Psi_{S\gamma}\rangle = \delta_{\gamma\gamma'}[S].$$

The $^{[S]}\psi_{\gamma}$ are given by formula (25) of Paper I[1] with the normalization constant $C_{\gamma} = (n!/d_{\gamma}(n))^{1/2}[S]^{1/2}$ [formula (38)[1]]. Then,

$$|{}^{[S']}\Psi_{\gamma'}\rangle\rangle \equiv [S']^{1/2}|{}^{[S']}N_{\gamma'}\rangle\rangle \text{ and } |{}^{[S]}\Psi_{\gamma}\rangle\rangle \equiv [S]^{1/2}|{}^{[S]}N_{\gamma}\rangle\rangle$$

so that $\langle\langle {}^{[S']}N_{\gamma'} | {}^{[S]}N_{\gamma}\rangle\rangle = \delta_{\gamma\gamma'}$.

The exact determination of N_c remains, in a general excited state which possesses doubly occupied orbitals. The normalization constant N_c is shown to be

$$N_c = \prod_c \left(\frac{2S_c+1}{2S_c+2}\right)^{1/2},$$

using the fact that the standard ordered $\mathfrak{B}^{-1/2\dagger}$ $\mathfrak{B}^{\xi 1/2\dagger}$ and $\mathfrak{B}^{\xi 1/2}\mathfrak{B}^{\xi-1/2}$ satisfy

$$\mathfrak{B}^{\xi-1/2\dagger}\,\mathfrak{B}^{\xi 1/2\dagger} \equiv \frac{[S_{\text{op}}+\frac{1}{2}]^{1/2}}{[S_{\text{op}}]^{1/2}}(\xi^2)^{0\dagger},$$

$$\mathfrak{B}^{\xi 1/2}\,\mathfrak{B}^{\xi-1/2} \equiv \frac{[S_{\text{op}}+\frac{1}{2}]^{1/2}}{[S_{\text{op}}]^{1/2}}(\xi^2)^{0},$$

$(\xi^2)^{0\dagger}$ and $(\xi^2)^0$ being, respectively, the creation and the annihilation operator of a singlet pair on the orbital ξ as is seen in Paper II,[2] formula 28.

The demonstration uses the fact that

$$\sqrt{\frac{S_{op}+1}{[S_{op}]}}\, wS_{\sigma'}^{-1/2\dagger} wS_{\sigma}^{1/2\dagger} - \sqrt{\frac{S_{op}}{[S_{op}]}}\, wS_{\sigma'}^{1/2\dagger}$$

$$wS_{\sigma}^{-1/2\dagger} \equiv \frac{\epsilon_{\sigma'\sigma}}{\sqrt{2}}$$

is antisymmetrical, while

$$\sqrt{\frac{S_{op}}{[S_{op}]}}\, wS_{\sigma'}^{-1/2\dagger} wS_{\sigma}^{1/2\dagger} + \sqrt{\frac{S_{op}+1}{[S_{op}]}}\, wS_{\sigma'}^{1/2\dagger} wS_{\sigma}^{-1/2\dagger}$$

is symmetrical and gives the relation

$$\sqrt{\frac{S_{op}}{[S_{op}]}}\, \mathfrak{B}^{\xi -1/2\dagger}\, \mathfrak{B}^{\xi 1/2\dagger} + \sqrt{\frac{S_{op}+1}{[S_{op}]}}\, \mathfrak{B}^{\xi 1/2\dagger}\, \mathfrak{B}^{\xi -1/2\dagger} \equiv 0.$$

One can also use a direct demonstration in doing simply that scalar product of two standard states.

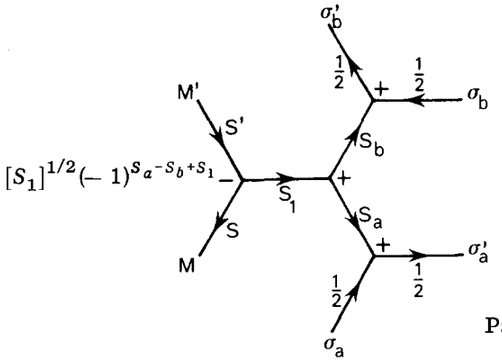
APPENDIX D: DETERMINATION OF THE $C_{\Sigma'_a\Sigma'_b\Sigma_b\Sigma_a}$

The identity for all S, M, S', M':

$$(z_{M'}^{S'} |\, {}^{(S_a S_b) S_1} T^{\sigma'_a \sigma'_b \sigma_b} |\, z_M^S)$$
$$\equiv \sum_{(\Sigma)} (z_{M'}^{S'} |\, wS_{\sigma'_a}^{\Sigma'_a\dagger} wS_{\sigma'_b}^{\Sigma'_b\dagger} S_{\sigma_b}^{\Sigma_b} wS_{\sigma_a}^{\Sigma_a} w\, |\, z_M^S)\, C_{\Sigma'_a\Sigma'_b\Sigma_a\Sigma_b}$$

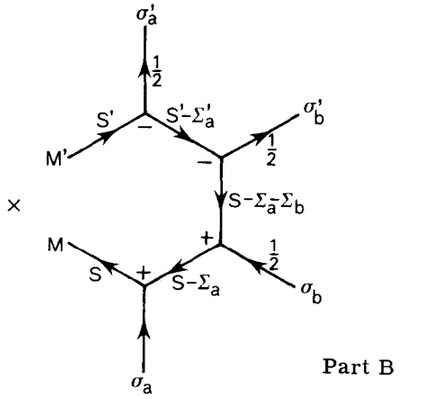
enables us to determine the values of $C_{\Sigma'_a\ldots}$ of $C^{op}_{\Sigma'_a\ldots}$.

This can be easily done graphically using Yutsis diagrams:



Part A

$$\equiv -\sum_{(\Sigma)} C_{\Sigma'_a\Sigma'_b\Sigma_b\Sigma_a}^{[S-\Sigma_a]^{1/2}\,[S'-\Sigma'_a]^{1/2}\,[S-\Sigma_a-\Sigma_b]}$$



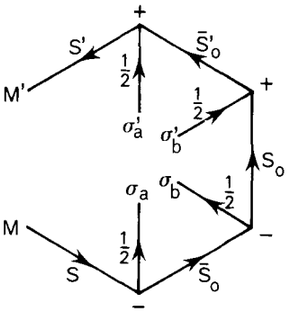
Part B

with

$$S - \Sigma_a - \Sigma_b = S' - \Sigma'_a - \Sigma'_b.$$

Eq. IV

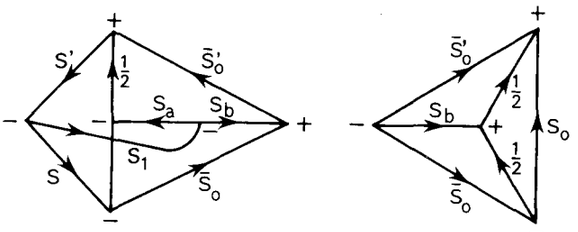
Considering the following angular diagram,



for given $\bar{S}_0, \bar{S}_0, \bar{S}'_0$,

Eq. V

we can join the corresponding free lines and obtain



Part A Part B

$$\times (-1)^{S_a - S_b + S_1} [S_1]^{1/2}$$
$$\equiv -\sum_{(\Sigma)} C_{\Sigma'_a\Sigma'_b\Sigma_b\Sigma_a}\, \delta^{\bar{S}_0, S-\Sigma_a}\, \delta^{S_0, S-\Sigma_a-\Sigma_b}\, \delta^{\bar{S}'_0, S'-\Sigma'_a}$$
$$\times [S-\Sigma_a]^{1/2}\, [S'-\Sigma_a]^{1/2},$$

Eq. VI

that is to say

$$C_{\Sigma'_a\Sigma'_b\Sigma_b\Sigma_a} = (-1)^{(1/2)-\Sigma'_b+2(\Sigma_a+\Sigma_b)+S_a+S_1}$$
$$\times [S_1]^{1/2} [S_0]^{1/2} [\bar{S}'_0]^{1/2} \begin{Bmatrix} \frac{1}{2} & S_b & S_1 \\ \bar{S}'_0 & \bar{S}_0 & S_0 \end{Bmatrix}$$
$$\times \begin{Bmatrix} S_a & S_b & S_1 \\ \frac{1}{2} & \bar{S}_0 & S \\ \frac{1}{2} & \bar{S}'_0 & S' \end{Bmatrix}$$

with

$$\bar{S}_0 = S_0 + \Sigma_b, \qquad \bar{S}'_0 = S_0 + \Sigma'_b,$$
$$S = S_0 + \Sigma_a + \Sigma_b, \qquad S' = S_0 + \Sigma'_a + \Sigma'_b.$$

APPENDIX E: DETERMINATION OF M FOR THE BIELECTRONIC OPERATOR [FORMULA (27)]

The only new problem is here to demonstrate the identity

$$-\begin{pmatrix} \frac{1}{2} & M_b & \sigma \\ \sigma' & S_b & \frac{1}{2} \end{pmatrix} \begin{pmatrix} M_1 & S_a & S_b \\ S_1 & M_a & M_b \end{pmatrix} (z_{M'}^{S'} |\, S_{\sigma}^{\Sigma'} ww_{M_1}^{S_1} S_{\sigma}^{\Sigma\dagger} |\, z_M^S)$$
$$\equiv (z_{M'}^{S'} |\, w_{M_a}^{S_a} |\, z_M^S) \times O_{S_b(S_a S_1)}^{\Sigma'\Sigma}(S, S').$$

Using again Yutsis angular diagram, we easily obtain

$$O^{\Sigma'\Sigma}_{S_b(S_aS_1)}(S,S') = (-1)^{S'-S+2\Sigma-S_b-S_1}[S]^{1/2}[S']^{1/2} \times \begin{Bmatrix} S_a & S_b & S_1 \\ S & \frac{1}{2} & S+\Sigma \\ S' & \frac{1}{2} & S'+\Sigma' \end{Bmatrix}.$$

There remains then only

$$-\begin{pmatrix} \frac{1}{2} & M_a & \sigma_i \\ \sigma_i' & S_a & \frac{1}{2} \end{pmatrix} W^{S_a}_{M_a},$$

which is the same kind of relation as in the mono-electronic case.

1. K. A. Brueckner, Phys. Rev. **97**, 1353 (1955); **100**, 36 (1955); *The Many-Body Problem* (Wiley, New York, 1959). J. Goldstone, Proc. Roy. Soc. (London) **A239**, 267 (1957). N. M. Hugenholtz, Physica **23**, 481 (1957).
2. K. A. Brueckner and J. L. Gammel, Phys. Rev. **109**, 1023 (1958); K. A. Brueckner and K. S. Masterton, Jr., *ibid.* **128**, 2267 (1962).
3. H. P. Kelly, Phys. Rev. **131**, 684 (1963); **136**, B 896, (1964). J. Čižek, J. Chem. Phys. **45**, 11, 4256 (1966). V. V. Tolmachev, "The Field Form of the Perturbation Theory Applied to Many-Electron Problems of Atoms and Molecules," Rotaprint, University of Tarta, 1963 (in Russian) (unpublished). P. G. H. Sandars, in Colloques Internationaux du CNRS (Paris, 1966) (unpublished). J. R. Malrieu, J. Chem. Phys. **47**, 4555 (1967).
4. T. Kato, Progr. Theoret. Phys. (Kyoto) **4**, 154 (1949).
5. C. Bloch, Nucl. Phys. **6**, 329 (1958).
6. C. Bloch and J. Horowitz, Nucl. Phys. **8**, 91 (1958).
7. B. H. Brandow, *Many Body Description of Nuclear Structure and Reaction*, edited by C. Bloch (Academic, New York, 1966).
8. P. G. H. Sandars, Advan. Chem. Phys. **14**, 365 (1969), a linked diagram treatment of configuration interaction in open shell atoms.
9. V. V. Tolmachev, Advan. Chem. Phys. 14, 421 (1969).
10. J. F. Gouyet, Phys. Rev. A **2**, 139 (1970).
11. J. F. Gouyet, Phys. Rev. A **2**, 1286 (1970).
12. The notation $S^{\Sigma\dagger}_\sigma$ includes spin creation and annihilation operators. We have the following connections with Schwinger (Ref. 13) and Bargmann (Ref. 14) operators:
$$S^{1/2\dagger}_\sigma = a^\dagger_\sigma = z_\sigma;$$
$$S^{-1/2\dagger}_\sigma = \tilde{a}_\sigma = (-1)^{1/2-\sigma}a_{-\sigma} = (-1)^{1/2-\sigma}\frac{\partial}{\partial z_{-\sigma}}.$$
13. J. Schwinger, in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. Van Dam (Academic, New York, 1965), p. 229.
14. V. Bargmann, Rev. Mod. Phys. **34**, 829 (1962).
15. K. Alder, A. Bohr, T. Huns, B. Mottelson, and A. Winther, Rev. Mod. Phys. **28**, 432 (1956).
16. A. P. Yutsis, I. B. Levinson, and W. Vanagas, *Mathematical Apparatus of the Theory of Angular Momentum* (Israel Program for Scientific Translations, 1962).

Analytic Unimodular Matrix Groups Containing the Image D^j of the Rotation Group: A Simple Derivation

L. Abellanas* and A. Galindo†

IHES, Bures-sur-Yvette, France

(Received 9 November 1971; Revised Manuscript Received 4 January 1972)

The existence of closed connected subgroups $\mathfrak{X}$ of $\mathfrak{SU}(2j+1)$ such that $D^j[\mathfrak{SU}(2)] \lneq \mathfrak{X} \lneq \mathfrak{SU}(2j+1)$ is investigated in terms of Lie algebras. We look for a more general solution to this problem. Besides the trivial (orthogonal or symplectic) candidates, Racah pointed out that $\mathfrak{X} = \mathcal{G}_2$ is another relevant example, for the case $j = 3$. This augmented chain made possible some very interesting theoretical calculations in the framework of atomic spectroscopy. Actually these are all possible solutions. We give a quite simple proof of this fact, using some well-known properties of the 6-j symbols.

1. INTRODUCTION

In the theoretical analysis of some physical problems (atomic spectroscopy, etc.)[1] the question often arises of finding closed connected subgroups $\mathfrak{X} < \mathfrak{SU}(2j+1)$ (j integer or half-integer) such that

$$(\mathfrak{SU}(2))_j \lneq \mathfrak{X} \lneq \mathfrak{SU}(2j+1) \tag{1}$$

where $(\mathfrak{SU}(2))_j$ is the image of $\mathfrak{SU}(2)$ under the irrep D^j. (Given two groups $\mathcal{G}, \mathcal{G}'$, the symbol $\mathcal{G}' < \mathcal{G}$ stands for "$\mathcal{G}'$ is a subgroup of $\mathcal{G}$"; and $\mathcal{G}' \lneq \mathcal{G}$ stands for "$\mathcal{G}'$ is a proper subgroup of $\mathcal{G}$." Similar notation will be used for Lie algebras. The symbol $\subset$ is only used to denote set-theoretic inclusion.) The classical groups $\mathfrak{SO}(2j+1)$ (j integer > 1), $\mathfrak{Sp}(2j+1)$ (j half-integer $> \frac{1}{2}$) are well-known examples of such $\mathfrak{X}$. As to further candidates, Racah[2] realized that for $j = 3$ the exceptional Cartan group $\mathcal{G}_2$ provides a refined chain

$$(\mathfrak{SU}(2))_3 < \mathcal{G}_2 < \mathfrak{SO}(7) < \mathfrak{SU}(7), \tag{2}$$

which has proved quite useful in analyzing complex atomic spectra.

Dynkin[3] has completely classified all possible inclusion types of irreducible groups of unimodular matrices. Dynkin's analysis is general and exhaustive and makes full use of the theory of semisimple, classical, and exceptional, Lie algebras. A look at Dynkin's tables shows that, with the only exception (2), just the above mentioned possibilities $\mathfrak{SO}(2j+1)$ or $\mathfrak{Sp}(2j+1)$ for $\mathfrak{X}$ remain. That this result is proven does not seem to be generally known by physicists although the fact is "apparently" true.

The aim of this short note is to present a simple, direct analysis of problem (1), using some elementary properties of the 6-j symbols. The arithmetical nature of this analysis should be noted.

2. THE CLASSICAL SOLUTION

It proves convenient to reformulate our problem in terms of complex Lie algebras. Namely, we wish to find for a given $j > \frac{1}{2}$ the more general Lie algebra X such that

$$(A_1)_j \lneq X \lneq A_{2j}. \tag{3}$$

(A_n is Cartan's notation for the complexified Lie algebra of $\mathfrak{SU}(n+1)$ and $(A_1)_j$ stands for the image of A_1 under D^j.) From (3) it follows that A_{2j}, X are naturally endowed with A_1-module structures.

It is known[4] that the A_1-module A_{2j} decomposes into a multiplicity-free direct sum of irreducible modules

$$A_{2j} = \bigoplus_{L \in I} [A_1]_{j,L}, \qquad I \equiv \{1, 2, \ldots, 2j\} \tag{4}$$

$$O^{\Sigma'\Sigma}_{S_b(S_aS_1)}(S,S') = (-1)^{S'-S+2\Sigma-S_b-S_1}[S]^{1/2}[S']^{1/2} \times \begin{Bmatrix} S_a & S_b & S_1 \\ S & \frac{1}{2} & S+\Sigma \\ S' & \frac{1}{2} & S'+\Sigma' \end{Bmatrix}.$$

There remains then only

$$-\begin{pmatrix} \frac{1}{2} & M_a & \sigma_i \\ \sigma_i' & S_a & \frac{1}{2} \end{pmatrix} W^{S_a}_{M_a},$$

which is the same kind of relation as in the mono-electronic case.

1. K. A. Brueckner, Phys. Rev. **97**, 1353 (1955); **100**, 36 (1955); *The Many-Body Problem* (Wiley, New York, 1959). J. Goldstone, Proc. Roy. Soc. (London) **A239**, 267 (1957). N. M. Hugenholtz, Physica **23**, 481 (1957).
2. K. A. Brueckner and J. L. Gammel, Phys. Rev. **109**, 1023 (1958); K. A. Brueckner and K. S. Masterton, Jr., *ibid.* **128**, 2267 (1962).
3. H. P. Kelly, Phys. Rev. **131**, 684 (1963); **136**, B 896, (1964). J. Čižek, J. Chem. Phys. **45**, 11, 4256 (1966). V. V. Tolmachev, "The Field Form of the Perturbation Theory Applied to Many-Electron Problems of Atoms and Molecules," Rotaprint, University of Tarta, 1963 (in Russian) (unpublished). P. G. H. Sandars, in Colloques Internationaux du CNRS (Paris, 1966) (unpublished). J. R. Malrieu, J. Chem. Phys. **47**, 4555 (1967).
4. T. Kato, Progr. Theoret. Phys. (Kyoto) **4**, 154 (1949).
5. C. Bloch, Nucl. Phys. **6**, 329 (1958).
6. C. Bloch and J. Horowitz, Nucl. Phys. **8**, 91 (1958).
7. B. H. Brandow, *Many Body Description of Nuclear Structure and Reaction*, edited by C. Bloch (Academic, New York, 1966).
8. P. G. H. Sandars, Advan. Chem. Phys. **14**, 365 (1969), a linked diagram treatment of configuration interaction in open shell atoms.
9. V. V. Tolmachev, Advan. Chem. Phys. 14, 421 (1969).
10. J. F. Gouyet, Phys. Rev. A **2**, 139 (1970).
11. J. F. Gouyet, Phys. Rev. A **2**, 1286 (1970).
12. The notation $S^{\Sigma\dagger}_\sigma$ includes spin creation and annihilation operators. We have the following connections with Schwinger (Ref. 13) and Bargmann (Ref. 14) operators:
$$S^{1/2\dagger}_\sigma = a^\dagger_\sigma = z_\sigma;$$
$$S^{-1/2\dagger}_\sigma = \tilde{a}_\sigma = (-1)^{1/2-\sigma}a_{-\sigma} = (-1)^{1/2-\sigma}\frac{\partial}{\partial z_{-\sigma}}.$$
13. J. Schwinger, in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. Van Dam (Academic, New York, 1965), p. 229.
14. V. Bargmann, Rev. Mod. Phys. **34**, 829 (1962).
15. K. Alder, A. Bohr, T. Huns, B. Mottelson, and A. Winther, Rev. Mod. Phys. **28**, 432 (1956).
16. A. P. Yutsis, I. B. Levinson, and W. Vanagas, *Mathematical Apparatus of the Theory of Angular Momentum* (Israel Program for Scientific Translations, 1962).

Analytic Unimodular Matrix Groups Containing the Image D^j of the Rotation Group: A Simple Derivation

L. Abellanas* and A. Galindo†

IHES, Bures-sur-Yvette, France

(Received 9 November 1971; Revised Manuscript Received 4 January 1972)

The existence of closed connected subgroups $\mathfrak{X}$ of $\mathfrak{SU}(2j+1)$ such that $D^j[\mathfrak{SU}(2)] \lneq \mathfrak{X} \lneq \mathfrak{SU}(2j+1)$ is investigated in terms of Lie algebras. We look for a more general solution to this problem. Besides the trivial (orthogonal or symplectic) candidates, Racah pointed out that $\mathfrak{X} = \mathcal{G}_2$ is another relevant example, for the case $j = 3$. This augmented chain made possible some very interesting theoretical calculations in the framework of atomic spectroscopy. Actually these are all possible solutions. We give a quite simple proof of this fact, using some well-known properties of the 6-j symbols.

1. INTRODUCTION

In the theoretical analysis of some physical problems (atomic spectroscopy, etc.)[1] the question often arises of finding closed connected subgroups $\mathfrak{X} < \mathfrak{SU}(2j+1)$ (j integer or half-integer) such that

$$(\mathfrak{SU}(2))_j \lneq \mathfrak{X} \lneq \mathfrak{SU}(2j+1) \tag{1}$$

where $(\mathfrak{SU}(2))_j$ is the image of $\mathfrak{SU}(2)$ under the irrep D^j. (Given two groups $\mathcal{G}, \mathcal{G}'$, the symbol $\mathcal{G}' < \mathcal{G}$ stands for "$\mathcal{G}'$ is a subgroup of $\mathcal{G}$"; and $\mathcal{G}' \lneq \mathcal{G}$ stands for "$\mathcal{G}'$ is a proper subgroup of $\mathcal{G}$." Similar notation will be used for Lie algebras. The symbol $\subset$ is only used to denote set-theoretic inclusion.) The classical groups $\mathfrak{SO}(2j+1)$ (j integer > 1), $\mathfrak{Sp}(2j+1)$ (j half-integer $> \frac{1}{2}$) are well-known examples of such $\mathfrak{X}$. As to further candidates, Racah[2] realized that for $j = 3$ the exceptional Cartan group $\mathcal{G}_2$ provides a refined chain

$$(\mathfrak{SU}(2))_3 < \mathcal{G}_2 < \mathfrak{SO}(7) < \mathfrak{SU}(7), \tag{2}$$

which has proved quite useful in analyzing complex atomic spectra.

Dynkin[3] has completely classified all possible inclusion types of irreducible groups of unimodular matrices. Dynkin's analysis is general and exhaustive and makes full use of the theory of semisimple, classical, and exceptional, Lie algebras. A look at Dynkin's tables shows that, with the only exception (2), just the above mentioned possibilities $\mathfrak{SO}(2j+1)$ or $\mathfrak{Sp}(2j+1)$ for $\mathfrak{X}$ remain. That this result is proven does not seem to be generally known by physicists although the fact is "apparently" true.

The aim of this short note is to present a simple, direct analysis of problem (1), using some elementary properties of the 6-j symbols. The arithmetical nature of this analysis should be noted.

2. THE CLASSICAL SOLUTION

It proves convenient to reformulate our problem in terms of complex Lie algebras. Namely, we wish to find for a given $j > \frac{1}{2}$ the more general Lie algebra X such that

$$(A_1)_j \lneq X \lneq A_{2j}. \tag{3}$$

(A_n is Cartan's notation for the complexified Lie algebra of $\mathfrak{SU}(n+1)$ and $(A_1)_j$ stands for the image of A_1 under D^j.) From (3) it follows that A_{2j}, X are naturally endowed with A_1-module structures.

It is known[4] that the A_1-module A_{2j} decomposes into a multiplicity-free direct sum of irreducible modules

$$A_{2j} = \bigoplus_{L \in I} [A_1]_{j,L}, \qquad I \equiv \{1, 2, \ldots, 2j\} \tag{4}$$

and that a basis in $[A_1]_{j,L}$ consists of the $(2j+1)\times(2j+1)$ matrices

$$(T_M^L)^m{}_n = \sqrt{2j+1}\begin{pmatrix} m & L & j \\ j & M & n \end{pmatrix}, \quad M = -L, -L+1, \ldots, L. \tag{5}$$

Therefore X will be of the form

$$X = \bigoplus_{L\in I(X)} [A_1]_{j,L}, \tag{6}$$

where $I(X) \subset I$. As $[A_1]_{j,1} = (A_1)_j$, necessarily $1 \in I(X)$.

Since

$$\left[[A_1]_{j,L_1}, [A_1]_{j,L_2}\right]$$

is again an A_1-module, we have

$$\left[[A_1]_{j,L_1}, [A_1]_{j,L_2}\right] = \bigoplus_{L_3\in I} c_j(L_1,L_2,L_3)[A_1]_{j,L_3}, \tag{7}$$

where $c_j(L_1,L_2,L_3) = 0,1$.

Consequently, the following statement holds:

Lemma 1: X will be a Lie algebra satisfying (3) if and only if X is of the form (6), $\{1\} \neq I(X) \neq I$, and $L_1, L_2 \in I(X)$ implies $L_3 \in I(X)$ whenever $c_j(L_1,L_2,L_3) = 1$.

Now from (5) one easily deduces

$$T^{L_1}_{M_1} T^{L_2}_{M_2} = (-)^{L_1+L_2+2j}\sqrt{2j+1} \sum_{L_3,M_3} (2L_3+1) \times \begin{Bmatrix} L_1 & L_2 & L_3 \\ j & j & j \end{Bmatrix} \begin{pmatrix} L_1 & L_2 & M_3 \\ M_1 & M_2 & L_3 \end{pmatrix} T^{L_3}_{M_3}, \tag{8}$$

whence

$$[T^{L_1}_{M_1}, T^{L_2}_{M_2}] = (-1)^{L_1+L_2+2j}\sqrt{2j+1} \times \sum_{L_3M_3} [1-(-)^{L_1+L_2+L_3}](2L_3+1) \times \begin{Bmatrix} L_1 & L_2 & L_3 \\ j & j & j \end{Bmatrix} \begin{pmatrix} L_1 & L_2 & M_3 \\ M_1 & M_2 & L_3 \end{pmatrix} T^{L_3}_{M_3}. \tag{9}$$

Therefore

Lemma 2: $c_j(L_1,L_2,L_3) = 1$ if and only if

(i) $L_1+L_2+L_3$ is odd,

(ii) $\begin{Bmatrix} L_1 & L_2 & L_3 \\ j & j & j \end{Bmatrix} \neq 0$.

Two direct consequences of Lemma 2 are worth mentioning: First, $c_j(L_1,L_2,L_3)$ is totally symmetric in L_1, L_2, L_3; and secondly,

$$\bigoplus_{L\in I_-} [A_1]_{j,L}, \text{ with } I_- \equiv \{L \in I : L \text{ odd}\},$$

is a Lie algebra. If $\Gamma^j = D^j(i\sigma_y)$ denotes the element of $(\mathfrak{SU}(2))_j$ representing a rotation of angle $-\pi$ around the Oy axis, it is known[4] that the matrix Γ^j, in the standard basis, is symmetric for integer j, antisymmetric for half-integer j. Moreover

$$\Gamma^j T^L_M (\Gamma^j)^{-1} = (-)^L (T^L_M)^T \tag{10}$$

and therefore

$$\bigoplus_{L\in I_-} [A_1]_{j,L} = B_j, \quad j \text{ integer} \qquad = C_{j+\frac{1}{2}}, \quad j \text{ half-integer}. \tag{11}$$

Expressions (11) exhibit the A_1-module structure of the classical solutions of (3).

3. THE MORE GENERAL SOLUTION

To proceed further, a closer look at $c_j(L_1,L_2,L_3)$ is necessary. First, it is plain that

(3.1) $c_j(1,L_2,L_3) = 1$ iff $L_2 = L_3$.

On the other hand, using Sato's condensed expressions[5] for the 6-j symbols, Lemma 2 can be replaced by the equivalent

Lemma 3: $c_j(L_1,L_2,L_3) = 1$ if and only if

(i) $L_1+L_2+L_3$ is odd,

(ii) L_1, L_2, L_3 satisfy the triangular condition ($|L_1-L_2| \leq L_3 \leq L_1+L_2$),

(iii) $[L_1L_2(2j-L_3) - L_1L_2 \times (2j+1+L_3)^{(-1)}]^{(L_1+L_2-L_3)} \neq 0$.

[provided one assumes $L_3 \geq \min(L_1,L_2)$; otherwise, just permute appropriately the indices 1, 2, 3 in the inequality under (iii)].

We use Sato's shorthand notation

$$x^{(\sigma)} \equiv x(x-1)\cdots(x-\sigma+1)$$
$$x^{(-\sigma)} \equiv (x+1)\cdots(x+\sigma), \quad \sigma \in \mathbb{Z}_+$$

and the binomial expansion of $\{x_1y_1z_1 - x_2y_2z_2^{(-1)}\}^{(\sigma)}$ is symbolically understood.

From Lemma 3 the following set of conclusions can simply be drawn:

Provided that (i) and (ii) of Lemma 3 are satisfied, then

(3.2) $c_j(L_1,L_2,2j) = 1$,

(3.3) $c_j(L_1,L_2,2j-1) = 1$ iff

$$(L_1-\Delta_3+1)(L_2-\Delta_3+1)(4j+\Delta_3) \neq L_1L_2\Delta_3 \tag{12}$$

with $\Delta_3 \equiv L_1+L_2-L_3 = L_1+L_2-(2j-1)$. In particular,

$$c_j(L,L,2j-1) = 1 \quad \text{iff } j(2j+1) \neq L(L+1), \tag{13}$$

(3.4) $c_j(L_1,L_2,L_1+L_2-1) = 1$,

(3.5) $c_j(3,L,L) = 1$ iff

$$5(2L+1)^2 - 12(2j+1)^2 \neq 17. \tag{14}$$

An appendix contains a brief discussion of the diophantine equation

$$5(2L+1)^2 - 12(2j+1)^2 = 17 \tag{15}$$

related to (14).

We are now in a position to find the more general Lie algebra X satisfying (3). Such X will have the form given by (6) with $\{1\} \subsetneq I(X) \subsetneq I$ (therefore $j > 1$).

Suppose that $L \in I(X)$, $L \neq 1$. Three mutually excluding possibilities may arise:

$$\begin{aligned} &\text{(a) } c_j(L,L,3) = 1; \\ &\text{(b) } c_j(L,L,3) = 0, \quad L \neq 2j-1; \\ &\text{(c) } c_j(L,L,3) = 0, \quad L = 2j-1. \end{aligned} \tag{16}$$

Case (a): By Lemma 1, $3 \in I(X)$, and repeated application of (3.4) leads to $I \subset I(X)$, $\{L + 2n : n \in \mathbb{Z}\} \cap I \subset I(X)$. Consequently, if L is even, the Lie algebra X as given by (6) will exhaust A_{2j}[i.e., $I(X) = I$]. For the inclusion to be proper, L must be odd and therefore $X = B_j$ (j integer > 1) or $X = C_{j+\frac{1}{2}}$ (j half-integer $> \frac{1}{2}$).

Case (b): The couple L, j must satisfy (15). Consequently $L > j$, j is integer (see Appendix) and (13) holds, since Eqs. (15) and $j(2j+1) = L(L+1)$ have no simultaneous (positive integer) solution. We conclude thus by (3.3) that $2j - 1 \in I(X)$. Being $2j - 1 \neq L$, the couple $2j-1, j$ satisfies (14), and hence (3.5) implies $3 \in I(X)$. The argument given for Case (a) can now be repeated, and the same conclusions thereof follow.

Case (c): By (3.5) the specific numerical values $j = 3$, $L = 5$ are required. Thus $X = [A_1]_{3,1} \oplus [A_1]_{3,5}$ is a 14-dimensional Lie subalgebra of A_6. Such Lie algebra X is simple [any proper ideal of X should be an A_1-module and therefore necessarily $[A_1]_{3,1}$ or $[A_1]_{3,5}$; but none of these modules is an ideal of X, as $c_3(1,5,5) = 1$ by (3.1)]. The only simple Lie algebra of dimension 14 is the exceptional G_2, and so the chain (2) emerges as

$$(A_1)_3 < G_2 < B_3 < A_6. \tag{17}$$

In conclusion, the following statement has therefore been proved.

Proposition: Let X be any Lie algebra satisfying $(A_1)_j \subsetneq X \subsetneq A_{2j}$.

Then either

$$\begin{aligned} &X = B_j, \quad j \text{ integer} > 1 \\ &\quad = C_{j+\frac{1}{2}}, \quad j \text{ half-integer} > \tfrac{1}{2} \end{aligned}$$

or

$$X = G_2, \quad j = 3. \qquad \text{(QED)}$$

ACKNOWLEDGMENTS

The authors wish to thank warmly Professor L. Motchane for the most kind hospitality extended to them at the IHES, and one of the authors (A.G.) for the financial support as well. They also gratefully acknowledge Professor Michel for raising the problem, for his constant encouragement, and for many extremely valuable discussions.

APPENDIX

In the preceding discussion we succeeded in showing that, unless $j = 3$ and $L = 5$, $[A_1]_{j,L} \subset X$, $L > 1$, $j > 1$ imply $[A_1]_{j,3} \subset X$. This we did either directly [Case (a)] or indirectly [Case (b)], depending on whether $c_j(L,L,3) = 1$ or 0. And once it was shown that $3 \in I(X)$, automatically (3.4) took care of the rest. The particular value $L = 3$ seems thus to be so relevant that it may have some interest knowing which L's are directly connected to it for a given j, i.e., $c_j(L,L,3) = 1$. According to (3.5) this will be the case iff $L > 1$, $j > 1$ and (14) holds. Alternatively, $c_j(L,L,3) = 0$ iff either $L = 1$ or the integers $x = 2L + 1$, $y = 2j + 1$ satisfy the Diophantine equation

$$5x^2 - 12y^2 = 17. \tag{A1}$$

All (integer) solutions x, y of (A1) consist of odd numbers: x even would imply $17 \equiv 0 \pmod 4$; y even would require $x^2 \equiv 5 \pmod 8$, but always $x^2 \equiv 0, 1, 4 \pmod 8$.

Therefore, $c_j(L,L,3) = 1$ whenever $L > 1$, j half-integer $> \frac{1}{2}$.

The general solution of (A1) is of the form

$$\begin{aligned} x &= 5\alpha - 36\beta, \\ \pm y &= 3\alpha - 25\beta, \end{aligned} \tag{A2}$$

where α, β are (integer) solutions of the Diophantine Pell's equation[6]

$$\alpha^2 - 60\beta^2 = 1 \tag{A3}$$

associated to (A1). In fact, x, y as given by (A2) satisfy (A1). Conversely, if x, y are solutions of (A1), then $x^2 + y^2 \equiv 0 \pmod{17}$, whence $y \equiv \pm 4x \pmod{17}$.

Therefore, as from (A2)

$$\begin{aligned} 17\alpha &= 25x \mp 36y, \\ 17\beta &= 3x \mp 5y, \end{aligned} \tag{A4}$$

both rhs of (A4) will be multiples of 17 and α, β will be integer solutions of (A3).

All solutions of Pell's equation (A3) are given[6] by

$$|\alpha_n| + |\beta_n|\sqrt{60} = (31 + 4\sqrt{60})^n, \quad n \geq 0 \tag{A5}$$

and therefore $c_j(L,L,3) = 0$ iff either $L = 1$ or $L = 2$, $j = 1$, or

$$\begin{aligned} 2L + 1 &= 5|\alpha_n| - 36\beta_n, \\ 2j + 1 &= |25\beta_n - 3|\alpha_n||, \end{aligned} \tag{A6}$$

with $n \geq 1$. And thus, for instance, (A6) gives

$$\begin{aligned} n = 1, &\quad L = 5, \ j = 3 \\ &\quad L = 149, \ j = 96 \\ n = 2, &\quad L = 338, \ j = 218 \\ &\quad L = 9266, \ j = 5981 \end{aligned}$$

etc.

---

* Supported by G.I.F.T. (Spain).

† On leave of absence from the University of Madrid.

1 See, for instance, the beautiful review by L. Michel, "Application of Group Theory to Quantum Physics: Algebraic Aspects," in *Lecture Notes in Physics*, 6 (Batelle Seattle 1969 Rencontres) (Springer-Verlag, Berlin, 1970).

2 G. Racah, Phys. Rev. **76**, 1352 (1949).

3 E. B. Dynkin, in Amer. Math. Soc. Transl., Ser. 2, 6, 245 (1957).

4 See, for instance, M. G. Doncel, L. Michel, and P. Minnaert, "Polarization Density Matrix," Preprint PTB-37 (Bordeaux, 1970).

5 M. Sato, Progr. Theoret. Phys. **13**, 405 (1955).

6 L. J. Mordell, *Diophantine Equations* (Academic, New York, 1969).

A Functional Calculus Approach to the Ursell-Mayer Functions

C. Y. Shen
Department of Mathematics, Simon Fraser University, Burnaby 2, British Columbia, Canada
(Received 16 September 1971; Revised Manuscript Received 11 November 1971)

A more complete treatment of the Ursell-Mayer functions is given here by way of a functional calculus defined on the algebra of complex-valued functions on the exponential of a set. Through measure-theoretical considerations, we are able to establish relationships between these functions and certain linear operators which relate probability densities and correlation functions in statistical mechanics. Also we point out that these functions provide a way of representing states of an infinite system.

1. INTRODUCTION

In statistical mechanics, the equilibrium state is usually determined by specifying the zero-density function $\psi_n(x_1, \ldots, x_n) = e^{-\beta H_n(x_1, \cdots, x_n) - \tau n}$, where $H_n(x_1, \ldots, x_n)$ is the system Hamiltonian and β, τ are constants. For the study of imperfect gases, it is sometime convenient to work with the Ursell-Mayer functions $\varphi_n(x_1, \ldots, x_n)$, which are defined in terms of $\psi_n(x_1, \ldots, x_n)$ by the scheme

$$\begin{aligned}
\psi_1(x_1) &= \varphi_1(x_1), \\
\psi_2(x_1, x_2) &= \varphi_2(x_1, x_2) + \varphi_1(x_1)\varphi_1(x_2), \\
\psi_3(x_1, x_2, x_3) &= \varphi_3(x_1, x_2, x_3) + \varphi_2(x_2, x_3)\varphi_1(x_1) \\
&\quad + \varphi_2(x_1, x_3)\varphi_1(x_2) + \varphi_2(x_1, x_2)\varphi_1(x_3) \\
&\quad + \varphi_1(x_1)\varphi_1(x_2)\varphi_1(x_3), \\
&\vdots
\end{aligned}$$

These functions can also be introduced elegantly through a functional calculus. The germ of this idea can be found in Refs. 1 and 2. Since the mechanical systems under consideration here often consist of finitely but variably many identical particles, it is more natural to regard the configuration space of such systems as the set of all finite unordered sequences of points in Euclidean 3-space. Carter and Prenter[3] pointed out the technical advantages of this viewpoint, and part of their work, namely the exponential construction of a set and of a measure space, is briefly described in Sec. 2. Our first object is to give a rigorous and more complete treatment of the Ursell-Mayer functions within the framework of exponential spaces and via a functional calculus. By doing so, we are able to prove some interesting properties about the Ursell-Mayer functions.

We begin our discussion by introducing an algebra $\mathfrak{A}(X_e)$ of complex-valued functions defined on the exponential of a set. Multiplication in this algebra corresponds to the star product used by Schwartz and Ruelle. Through power series expansion, we establish in Sec. 3 the homomorphism between the algebra of locally analytic functions and $\mathfrak{A}(X_e)$. The Ursell-Mayer function φ of a function $\psi \in \mathfrak{A}(X_e)$ such that $\psi(0) > 0$ is defined to be

$$\varphi = \operatorname{Log}\psi, \quad \psi = \exp\varphi,$$

where Log is the principal branch of the logarithm. Our main result, stated in Sec. 4, shows that the exp and Log mappings commute with certain linear operators T, U which connect probability densities and correlation functions in statistical mechanics.[4,5] Also, we point out that the Ursell-Mayer functions can be used to represent states of infinite systems.

2. THE ALGEBRA $\mathfrak{A}(X_e)$

Given a nonempty set X, we denote the collection of all unordered sequences of finite length in X by X_e. As pointed out in Ref. 3, an unordered sequence x, with length $l(x) = n$, may be written as a formal product

$$x = x_1 \cdots x_n,$$

where the order of factors is irrelevant. When $n = 0$, we write $x = 0$. By collecting equal factors, the product can also be written in the form

$$x = t_1^{r_1} \cdots t_m^{r_m},$$

where $t_1, \ldots, t_m$ are the distinct factors of x and $r_1, \ldots, r_m$ are the corresponding multiplicities. We adopt the convention that, for all $t \in X$, $t^r = 0$ if $r = 0$. Also, we denote the multiplicity of each factor t of x by $r(x, t)$, and define $r(x, t) = 0$, if t is not a factor of x.

There is a natural binary operation on X_e, called *concatenation*, defined as follows: If $x = x_1 \cdots x_m$ and $y = y_1 \cdots y_n$, then

$$xy = x_1 \cdots x_m y_1 \cdots y_n,$$

with 0 as the identity element. Notice that $r(xy, t) = r(x, t) + r(y, t)$ for all $x, y \in X_e$. Let $x \in X_e$ and let n be a positive integer. An *n-partition* of x is an n-tuple of unordered sequences $(x_1, \ldots, x_n)$ such that $x = x_1 \cdots x_n$. The set of all n-partitions of x is denoted by $P_n(x)$. The *index of an unordered sequence* $x = t_1^{r_1} \cdots t_m^{r_m}$ is $I(x) = r_1! \cdots r_m!$ and the *index of an n-partition* $(x_1, \ldots, x_n)$ of x is

$$I(x_1, \ldots, x_n) = I(x)/I(x_1) \cdots I(x_n).$$

Now let $\mathfrak{A}(X_e)$ denote the family of all complex-valued functions defined on X_e. If $\varphi, \psi \in \mathfrak{A}(X_e)$, their star product $\varphi * \psi$ is defined on X_e by the equation

$$\varphi * \psi(x) = \sum \{I(x_1, x_2)\varphi(x_1)\psi(x_2) : (x_1, x_2) \in P_2(x)\}. \tag{2.1}$$

The notation for star products is simplified by introducing a special symbol $\sum^*$. If S is a set of partitions of an unordered sequence x, then

$$\sum{}^* \{f(x_1, \ldots, x_n) : (x_1, \ldots, x_n) \in S\}$$

will denote the corresponding sum in which each term is multiplied by the index of the partition $(x_1, \ldots, x_n)$. In this notation we have

$$\varphi * \psi(x) = \sum{}^* \{\varphi(x_1)\psi(x_2) : (x_1, x_2) \in P_2(x)\}.$$

With this product, $\mathfrak{A}(X_e)$ becomes a commutative algebra with identity 1^* given by

$$1^*(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

By a simple inductive argument, we obtain

Theorem 2.1: Let $(\varphi_1, \ldots, \varphi_k)$, $k \geq 1$, be a finite sequence of functions in $\mathfrak{A}(X_e)$. Then

$$(\varphi_1 * \cdots * \varphi_k)(x) = \sum^*\{\varphi_1(x_1) \cdots \varphi_k(x_k): (x_1, \ldots, x_k) \in P_k(x)\}.$$

The nth power of an element φ under the star operation will be denoted simply by φ^n. For $n = 0$, we define $\varphi^0 = 1^*$. A function $\varphi \in \mathfrak{A}(X_e)$ has a multiplicative inverse if and only if $\varphi \in \mathfrak{A}_1 = \{\varphi \in \mathfrak{A}(X_e)$ $\varphi(0) \neq 0\}$. An exact formula for the inverse is given by Eq. (3.3) below. The complement of $\mathfrak{A}_1$,

$$\mathfrak{A}_0 = \{\varphi \in \mathfrak{A}(X_e): \varphi(0) = 0\},$$

is a maximal ideal in $\mathfrak{A}(X_e)$. It also possesses another interesting property given by

Theorem 2.2: Let $\varphi \in \mathfrak{A}_0$. If $n > l(x)$, then $\varphi^n(x) = 0$.

The proof follows trivially from Theorem 2.1 and the fact that if $(x_1, \ldots, x_n) \in P_n(x)$ and $n > l(x)$, then $x_i = 0$ for some $i = 1, 2, \ldots, n$. As a consequence of Theorem 2.2, arbitrary infinite power series $\sum_{n=0}^{\infty} c_n\varphi^n$ are easily defined for members of $\mathfrak{A}_0$ by the equation

$$\left(\sum_{n=0}^{\infty} c_n\varphi^n\right)(x) = \sum_{n=0}^{l(x)} c_n\varphi^n(x). \tag{2.2}$$

This important observation furnishes the basis of our discussion in the next section.

Another useful way of writing Eq. (2.2) can be given as follows: An n-partition $(x_1, \ldots, x_n)$ of an unordered sequence x is *proper* if $x_i \neq 0$ for all $i = 1, \ldots, n$.

The class of all proper n-partitions of x will be denoted by $Q_n(x)$. Furthermore, $Q(x)$ will denote the class of proper n-partitions of x for all $n = 1, 2, \cdots$. That is,

$$Q(x) = \cup\{Q_n(x): 0 < n \leq l(x)\}.$$

By Theorem 2.1 and the above definition, Eq. (2.2) is equivalent to

$$\begin{aligned} &\left(\sum_{n=0}^{\infty} c_n\varphi^n\right)(0) = c_0, \\ &\left(\sum_{n=0}^{\infty} c_n\varphi^n\right)(x) \\ &\quad = \sum^*\{c_k\varphi(x_1) \cdots \varphi(x_k): (x_1, \ldots, x_k) \in Q(x)\} \\ &\quad \text{if } x \neq 0. \end{aligned} \tag{2.3}$$

For each $x \in X_e$, we define a linear operator D_x: $\mathfrak{A}(X_e) \to \mathfrak{A}(X_e)$ by the equation

$$D_x\varphi(y) = \varphi(xy).$$

This corresponds to a similar operator used by Ruelle.[1] Notice that $D_xD_y = D_{xy}$ for all $x, y \in X_e$. When $x = 0$, D_x is the identity. When $l(x) = 1$, D_x is a *derivation* in the sense that

$$D_x(\varphi * \psi) = \varphi * D_x\psi + D_x\varphi * \psi.$$

Theorem 2.3 below shows that D_x behaves like a differential operator of order $l(x)$.

Lemma 2.1: Let f be a function on $X_e \times X_e$ into a linear space. Then, for each $x, y \in X_e$ such that $l(x) = 1$,

$$\sum^*\{f(xu, v) + f(u, xv): (u, v) \in P_2(y)\} = \sum^*\{f(u, v): (u, v) \in P_2(xy)\}.$$

Proof: For each $y \in X_e$, we have

$$\begin{aligned} &\sum^*\{f(xu, v): (u, v) \in P_2(y)\} \\ &\quad = \sum\left\{\frac{I(y)r(xy, x)r(xu, x)}{I(u)I(v)r(xy, x)r(xu, x)} f(xu, v): (u, v) \in P_2(y)\right\} \\ &\quad = \sum\left\{\frac{I(xy)r(u', x)}{I(u')I(v')r(xy, x)} f(u', v'): (u', v') \in P_2(xy)\right. \\ &\quad \left.\text{and } x \text{ is a factor of } u'\right\}. \end{aligned}$$

Since $r(u', x) = 0$ if x is not a factor of u', the above sum is equal to

$$\sum\left\{\frac{I(xy)r(u', x)}{I(u')I(v')r(xy, x)} f(u', v'): (u', v') \in P_2(xy)\right\}$$

By a similar argument, we obtain

$$\begin{aligned} &\sum^*\{f(u, xv): (u, v) \in P_2(y)\} \\ &\quad = \sum\left\{\frac{I(xy)r(v', x)}{I(u')I(v')r(xy, x)} f(u', v'): (u', v') \in P_2(xy)\right\} \end{aligned}$$

The lemma now follows by combining the above sums and using the identity $r(u, x) + r(v, x) = r(uv, x)$.

Theorem 2.3: Let $\varphi, \psi \in \mathfrak{A}(X_e)$ and x be a member of X_e. Then

$$D_x(\varphi * \psi) = \sum^*\{D_u\varphi * D_v\psi: (u, v) \in P_2(x)\}.$$

Proof: The proof is by induction on $l(x)$. Clearly the result holds for $x = 0$. When $l(x) = 1$, we have

$$(D_x\varphi * \psi + \varphi * D_x\psi)(y) = \sum^*\{\varphi(xu)\psi(v) + \varphi(u)\psi(xv): (u, v) \in P_2(y)\}.$$

Applying Lemma 2.1 with $f(u, v) = \varphi(u)\psi(v)$, we see the right-hand sum is equal to $D_x(\varphi * \psi)(y)$. Now suppose the result holds for $l(x) = k \geq 1$, we write $x = yz$, where $l(y) = 1$ and $l(z) = k$. Then

$$\begin{aligned} D_x(\varphi * \psi) &= D_y(D_z(\varphi * \psi)) \\ &= \sum^*\{D_{yu}\varphi * D_v\psi + D_u\varphi * D_{yv}\psi: (u, v) \in P_2(z)\}. \end{aligned}$$

Again using Lemma 2.1 with $f(u, v) = D_u\varphi * D_v\psi$, we obtain the desired result.

By analogy with the theory of ordinary differential equations, we have the following simple fact:

Theorem 2.4: Let $\psi \in \mathfrak{A}(X_e)$. Then for all $x \in X$, the pair of equations

$$\begin{aligned} D_x\varphi &= \varphi * \psi \\ \varphi(0) &= 0 \end{aligned}$$

has the unique solution $\varphi = 0$.

Now suppose that the set X is endowed with a σ-finite measure structure. Carter and Prenter[3] have shown that it induces a natural σ-finite measure structure on X_e as follows: Let $\mathfrak{X}$ be a σ-algebra of subsets of X and let ξ be a σ-finite measure defined on $\mathfrak{X}$. For each nonnegative integer n, $(X^{\cdot n}, \mathfrak{X}^{\cdot n}, \xi^{\cdot n})$ denotes the nth product space. When $n = 0, X^{\cdot 0} = \{0\}$ and $\xi^{\cdot 0}(\{0\}) = 1$. Let $X_e^{\cdot} = \cup_{n=0}^{\infty} X^{\cdot n}$. Then

$$\mathfrak{X}_e^{\cdot} = \left\{ \bigcup_{n=0}^{\infty} A_n : A_n \in \mathfrak{X}^{\cdot n} \text{ for each } n \right\}$$

is a σ-algebra of subsets of $X_e^{\cdot}$, and the set function $\xi_e^{\cdot}$ defined on $\mathfrak{X}_e^{\cdot}$ by

$$\xi_e^{\cdot}(E) = \sum_{n=0}^{\infty} \frac{1}{n!} \xi^{\cdot n}(E \cap X^{\cdot n})$$

is a σ-finite measure. The quotient space of $(X_e^{\cdot}, \mathfrak{X}_e^{\cdot}, \xi_e^{\cdot})$ under the natural projection $p: X_e^{\cdot} \to X_e$ (here members of X_e are regarded as equivalence classes of elements in $X_e^{\cdot}$ under the equivalence relation of rearrangement) is called the *exponential space* of $(X, \mathfrak{X}, \xi)$ and is denoted by $(X_e, \mathfrak{X}_e, \xi_e)$.

Let $\mathcal{L}_1(X_e)$ be the set of all ξ_e-integrable functions, and let $L_1(X_e)$ be the corresponding set of equivalence classes. In view of the following known theorem, $L_1(X_e)$ is a commutative Banach algebra with identity 1^*.

Theorem 2.5: If $\varphi_1, \varphi_2 \in \mathcal{L}_1(X_e)$, then so is $\varphi_1 * \varphi_2 \in \mathcal{L}_1(X_e)$.

Moreover,

(i) $\int \varphi_1 * \varphi_2 \, d\xi_e = (\int \varphi_1 d\xi_e)(\int \varphi_2 d\xi_e)$

(ii) $\int |\varphi_1 * \varphi_2| \, d\xi_e \le (\int |\varphi_1| \, d\xi_e)(\int |\varphi_2| \, d\xi_e)$.

3. THE OPERATIONAL CALCULUS

In this section we will set up an operational calculus for $\mathfrak{A}(X_e)$, which is suggested by the standard operational calculus for Banach spaces. The spectrum of an element $\varphi \in \mathfrak{A}(X_e)$ is defined to be the set of all complex numbers z such that $\varphi - z1^*$ has no inverse. Since $\varphi - z1^*$ has no inverse if and only if it belongs to $\mathfrak{A}_0$, the spectrum of φ consists of the single point $\varphi(0)$. If F is a locally analytic function such that $\varphi(0) \in \text{dom}F$, we define $F(\varphi)$ by using the Taylor's series expansion of F about $\varphi(0)$. That is,

$$F(\varphi) = \sum_{n=0}^{\infty} a_n [\varphi - \varphi(0)1^*]^n \tag{3.1}$$

where $a_n = [F^{(n)}(\varphi(0))]/n!$. Since $\varphi - \varphi(0)1^*$ belongs to $\mathfrak{A}_0$, this infinite series is defined in the sense of Eq. (2.2). Also, by Eq. (2.3), we have

$$\begin{aligned} &F(\varphi)(0) = F(\varphi(0)), \\ &F(\varphi)(x) = \sum{}^*\{a_n \varphi(x_1) \cdots \varphi(x_n) : (x_1, \ldots, x_n) \in Q(x)\} \\ &\qquad \text{if } x \neq 0. \end{aligned} \tag{3.2}$$

Consider the class of all locally analytic functions whose domains contain $\varphi(0)$. Two such functions F, G are *equivalent relative* to φ if $F(z) = G(z)$ on some neighborhood of $\varphi(0)$. The set of equivalence classes of such functions forms a commutative algebra with multiplicative identity, where the algebraic operations are defined in the obvious way by using representative functions. We shall denote this algebra by $\mathfrak{F}_\varphi$.

Theorem 3.1: For each $\varphi \in \mathfrak{A}(X_e)$, the mapping $F \to F(\varphi)$ given by Eq. (3.1) determines a homomorphism of the algebra $\mathfrak{F}_\varphi$ into $\mathfrak{A}(X_e)$. Moreover, it maps the constant function $F(z) = 1$ into 1^* and the identity function $F(z) = z$ into φ.

Proof: Clearly, the mapping $F \to F(\varphi)$ is linear. Let $F(z), G(z)$ be two members of $\mathfrak{F}_\varphi$ with expansions

$$F(z) = \sum_{m=0}^{\infty} a_m [z - \varphi(0)]^m,$$

$$G(z) = \sum_{n=0}^{\infty} b_n [z - \varphi(0)]^n.$$

For $x = 0$, $(F(\varphi) * G(\varphi))(0) = (FG(\varphi))(0) = a_0 b_0$. When $l(x) > 0$, we have

$$(F(\varphi) * G(\varphi))(x) = \sum \left\{ \frac{I(x)}{I(u)I(v)} F(\varphi)(u) G(\varphi)(v) : (u, v) \in P_2(x) \right\}$$

Separating out the terms with $u = 0$ and $v = 0$ and expanding $F(\varphi)(u), G(\varphi)(v)$, we obtain for this sum

$$\begin{aligned} &a_0 \sum \left\{ \frac{I(x)}{I(y_1) \cdots I(y_l)} b_l \varphi(y_1) \cdots \varphi(y_l) : (y_1, \ldots, y_l) \in Q(x) \right\} \\ &\quad + b_0 \sum \left\{ \frac{I(x)}{I(y_1) \cdots I(y_l)} a_l \varphi(y_1) \cdots \varphi(y_l) : (y_1, \ldots, y_l) \in Q(x) \right\} \\ &\quad + \sum \left\{ \frac{I(x)}{I(u)I(v)} \left(\sum \left\{ \frac{I(u)}{I(u_1) \cdots I(u_m)} a_m \varphi(u_1) \cdots \varphi(u_m) : (u_1, \ldots, u_m) \in Q(u) \right\} \right) \right. \\ &\quad \times \left(\sum \left\{ \frac{I(v)}{I(v_1) \cdots I(v_n)} b_n \varphi(v_1) \cdots \varphi(v_n) : (v_1, \ldots, v_n) \in Q(v) \right\} \right): \\ &\quad \left. \times (u, v) \in Q(x) \right\}. \end{aligned}$$

The last term can be rewritten as

$$\sum \left\{ \frac{I(x)}{I(u_1) \cdots I(u_m) I(v_1) \cdots I(v_n)} \times a_m b_n \varphi(u_1) \cdots \varphi(u_m) \varphi(v_1) \cdots \varphi(v_n) : \right.$$

$$\left. (u_1, \ldots, u_m) \in Q(u), (v_1, \ldots, v_n) \in Q(v) \right.$$
$$\left. \text{for some } u, v \in X_e \text{ such that } (u, v) \in Q_2(x) \right\}.$$

Collecting terms corresponding to the same partition in $Q(x)$ and combining it with the first two terms above, we have

$$\sum^*\left\{\left(\sum_{k=0}^{l} a_l b_{l-k}\right)\varphi(y_1)\cdots\varphi(y_l):(y_1,\ldots,y_l)\in Q(x)\right\},$$

which is equal to $((FG)(\varphi))(x)$. The rest of the theorem follows immediately from the definition of $F(\varphi)$.

As a consequence of Theorem 3.1 and Eq. (3.2), φ^{-1} is equal to $F(\varphi)$, where $F(z)=1/z$. That is,

$$\varphi^{-1}(0)=\frac{1}{\varphi(0)},$$
$$\varphi^{-1}(x)=\sum^*\left\{\frac{(-1)^k}{\varphi(0)^{k+1}}\varphi(x_1)\ldots\varphi(x_k):(x_1,\ldots,x_k)\in Q(x)\right\}$$
$$\text{if } x\neq 0. \tag{3.3}$$

In the foregoing discussion, we focus our attention on a particular function $\varphi\in\mathfrak{A}(X_e)$. Now we reverse this point of view and consider the mapping $\varphi\to F(\varphi)$ for a fixed locally analytic function F. More precisely, for each such function F, let $\mathfrak{D}_F$ be the class of all members $\varphi\in\mathfrak{A}(X_e)$ such that $\varphi(0)\in\mathrm{dom}F$. Then the mapping $\varphi\to F(\varphi)$ is a function on $\mathfrak{D}_F$ into $\mathfrak{A}(X_e)$. For purposes of the following theorem, we shall denote this function by F^*.

Theorem 3.2: The mapping $F\to F^*$ defined above preserves functional composition in the sense that for all locally analytic functions F,G,

$$(F\circ G)^*=F^*\circ G^*$$

That is, $(F\circ G)(\varphi)$ is defined if and only if $F(G(\varphi))$ is defined, in which case $(F\circ G)(\varphi)=F(G(\varphi))$.

Proof: It is easy to show that $(F\circ G)(\varphi)$ is defined if and only if $F(G(\varphi))$ is defined. Now suppose $\varphi(0)\in\mathrm{dom}(F\circ G)$ and $G(z),F(z)$ have the following expansions about $\varphi(0),G(\varphi(0))$, respectively:

$$G(z)=G(\varphi(0))+\sum_{n=1}^{\infty}d_n[z-\varphi(0)]^n,$$
$$F(z)=\sum_{n=0}^{\infty}c_n[z-G(\varphi(0))]^n.$$

Then the expansion of $(F\circ G)(z)$ about $\varphi(0)$ is given by

$$(F\circ G)(z)=\sum_{n=0}^{\infty}q_n[z-\varphi(0)]^n,$$

where $q_0=c_0$ and for $n>0$

$$q_n=\sum_{k=1}^{n}c_k\left(\sum\left\{d_{j_1}\cdots d_{j_k}:j_1+\ldots+j_k=n,\ \text{all } j\text{'s}\geq 1\right\}\right)$$

Clearly, $F(G(\varphi))(0)=((F\circ G)(\varphi))(0)=c_0$. Suppose $x\neq 0$.

We have

$$F(G(\varphi))(x)=\sum\left\{\frac{I(x)}{I(u_1)\cdots I(u_k)}c_kG(\varphi)(u_1)\cdots G(\varphi)(u_k):(u_1,\ldots,u_k)\in Q(x)\right\}$$
$$=\sum\left\{\frac{I(x)c_k}{I(u_{11})\cdots I(u_{1j_1})\cdots I(u_{k1})\cdots I(u_{kj_k})}d_{j_1}\cdots d_{j_k}\varphi(u_{11})\ldots\varphi(u_{1j_1})\ldots\varphi(u_{kj_k}):\right.$$
$$(u_{11},\ldots,u_{1j_1})\in Q(u_1),\ldots,(u_{k1},\ldots,u_{kj_k})\in Q(u_k)\text{ for some }u_1,\ldots,u_k\in X_e\text{ such that}$$
$$\left.(u_1,\ldots,u_k)\in Q(x)\right\}.$$

Collecting terms corresponding to the same partition $(u_{11},\ldots,u_{1j_1},\ldots,u_{k1},\ldots,u_{kj_k})$ in $Q(x)$, we obtain

$$\sum\left\{\frac{I(x)}{I(x_1)\cdots I(x_n)}\sum_{k=1}^{n}c_k\left(\sum\left\{d_{i_1}\cdots d_{j_k}:j_1+\cdots+j_k=n,\right.\right.\right.$$
$$\left.\left.\left.\text{all } j\text{'s}\geq 1\right\}\right)\varphi(x_1)\cdots\varphi(x_n):(x_1,\ldots,x_n)\in Q(x)\right\},$$

which is equal to $(F\circ G)(\varphi)(x)$.

The operator D_x of Sec. 2 acts as a differential operator on $F(\varphi)$ in the sense of

Theorem 3.3: Let x be a member of X_e such that $l(x)=1$, and let F be a locally analytic function with derivative F'. Then, for each $\varphi\in\mathfrak{D}_F$

$$D_xF(\varphi)=F'(\varphi)*D_x\varphi$$

Proof: Let F have the expansion

$$F(z)=\sum_{n=0}^{\infty}b_n[z-\varphi(0)]^n.$$

It is easy to show that $(F'(\varphi)*D_x\varphi)(0)=(D_xF(\varphi))(0)=b_1\varphi(x)$. When $y\neq 0$, we have

$$(F'(\varphi)*D_x\varphi)(y)=F'(\varphi)(0)\varphi(xy)+\sum\left\{\frac{I(y)}{I(u)I(v)}\right.$$
$$\times\left(\sum\left\{\frac{I(u)}{I(u_1)\ldots I(u_k)}(k+1)b_{k+1}\varphi(u_1)\cdots\varphi(u_k):\right.\right.$$
$$\left.\left.(u_1,\ldots,u_k)\in Q(u)\right\}\right)D_x\varphi(v):(u,v)\in P_2(y),$$
$$\left.u\neq 0\right\}.$$

Let $u_{k+1}=xv$. Since $r(u_{k+1},x)=0$ if x is not a factor of u_{k+1}, the above sum is equal to

$$F'(\varphi)(0)\varphi(xy)+\sum\left\{\frac{I(xy)}{I(u_1)\cdots I(u_{k+1})}\frac{r(u_k,x)}{r(xy,x)}\right.$$
$$\times(k+1)b_{k+1}\varphi(u_1)\cdots\varphi(u_{k+1}):(u_1,\ldots,u_{k+1})$$
$$\left.\in Q(xy)\text{ and }k\geq 2\right\}.$$

By the symmetry of this sum, it can be expressed as

$$F'(\varphi)(0)\varphi(xy)+\sum\left\{\frac{I(xy)}{I(u_1)\ldots I(u_{k+1})}\frac{r(u_j,x)}{r(xy,x)}\right.$$
$$\times b_{k+1}\varphi(u_1)\cdots\varphi(u_{k+1}):(u_1,\ldots,u_j,\ldots,u_k)$$
$$\left.\in Q(xy),\quad k\geq 2,\quad j=1,2,\ldots,k+1\right\}.$$

Using the identity $r(u_1,x)+\cdots+r(u_{k+1},x)=r(xy,x)$ and the fact that $F'(\varphi)(0)=b_1$, we finally have

$$b_1\varphi(xy)+\sum\left\{\frac{I(xy)}{I(u_1)\cdots I(u_{k+1})}b_{k+1}\varphi(u_1)\cdots\varphi(u_{k+1}):\right.$$
$$\left.(u_1,\ldots,u_{k+1})\in Q(xy),\quad k\geq 2\right\}=(D_xF(\varphi))(y).$$

Finally, we have

Theorem 3.4: Let $\varphi\in\mathfrak{L}_1(X_e)$ and let $F(z)$ be a locally analytic function such that $\varphi(0)\in\mathrm{dom}F$. Then $F(\varphi)\in\mathfrak{L}_1(X_e)$ if the series

$$\sum_{n=1}^{\infty}\frac{|F^{(n)}(\varphi(0))|}{n!}\left(\int|\varphi|\,d\xi_e-|\varphi(0)|\right)^n$$

converges. In this case, $\int F(\varphi)d\xi_e=F(\int\varphi\,d\xi_e)$.

The proof of this theorem involves simple applications of Lebesgue convergence theorems.

4. URSELL–MAYER FUNCTIONS

For a systematic treatment of Ursell–Mayer functions, we first introduce the exponential mapping $\varphi \to \exp\varphi$. This mapping is defined on $\mathfrak{A}(X_e)$ and has range $\mathfrak{A}_1$. Using Eq. (3.2), we have

$$(\exp\varphi)(0) = e^{\varphi(0)},$$

$$(\exp\varphi)(x) = \sum{}^{*}\left\{\frac{e^{\varphi(0)}}{k!}\varphi(x_1)\cdots\varphi(x_k) : (x_1,\ldots,x_k) \in Q(x)\right\} \quad \text{if } x \neq 0. \tag{4.1}$$

Alternatively, $\exp\varphi$ may be defined as the unique solution ψ of the pair of equations

$$D_x\psi = \psi * D_x\varphi,$$
$$\psi(0) = e^{\varphi(0)},$$

where $l(x) = 1$. The addition formula

$$\exp(\varphi_1 + \varphi_2) = \exp\varphi_1 * \exp\varphi_2 \tag{4.2}$$

is easily established using this characterization of $\exp\varphi$.

A generalization of

$$D_x(\exp\varphi) = \exp\varphi * D_x\varphi$$

is given by

Theorem 4.1: Let x be a member of X_e such that $l(x) \geq 1$. Then

$$D_x(\exp\varphi) = \sum{}^{*}\{(1/k!)\exp\varphi * D_{x_1}\varphi * \cdots * D_{x_k}\varphi : (x_1,\ldots,x_k) \in Q(x)\}.$$

Proof: We have seen that this holds for $l(x) = 1$. Suppose it holds for $l(x) = k \geq 1$. When $l(x) = k + 1$, we write $x = yz$, where $l(y) = 1$ and $l(z) = k$. By Theorem 3.3, we have

$$D_x(\exp\varphi) = D_z(\exp\varphi * D_y\varphi) = \sum{}^{*}\{D_u(\exp\varphi) * D_{vy}\varphi : (u,v) \in P_2(z)\}.$$

Separating out the term with $u = 0$ and using the induction hypothesis, we obtain

$$\exp\varphi * D_x\varphi + \sum\left\{\frac{I(z)}{I(v)I(u_1)\cdots I(u_k)}\frac{1}{k!}\exp\varphi * D_{u_1}\varphi * \cdots * D_{u_k}\varphi * D_{vy}\varphi : (u_1,\ldots,u_k) \in Q(u), (u,v) \in P_2(z), u \neq 0\right\}.$$

Setting $u_{k+1} = vy$, writing $I(z)/I(v) = [I(x)/I(u_{k+1})] \times [r(u_{k+1},y)/r(x,y)]$ and noting that $r(u_{k+1},y) = 0$ if y is not a factor of u_{k+1}, we obtain for the above sum

$$\sum\left\{\frac{I(x)}{I(u_1)\cdots I(u_{k+1})}\frac{r(u_{k+1},y)}{r(x,y)}\frac{1}{k!}\exp\varphi * D_{u_1}\varphi * \cdots * D_{u_{k+1}}\varphi : (u_1,\ldots,u_{k+1}) \in Q(x),\ k \geq 1\right\}.$$

By the symmetry of this sum and the identity $r(u_1,y) + \cdots + r(u_k,y) = r(x,y)$, this becomes

$$\sum{}^{*}\left\{\frac{1}{(k+1)!}\exp\varphi * D_{u_1}\varphi * \cdots * D_{u_{k+1}}\varphi : (u_1,\ldots,u_{k+1}) \in Q(x),\ k \geq 1\right\}.$$

The induction step is completed by combining this with the term $\exp\varphi * D_x\varphi$.

Let $\log z$ be an arbitrary branch of the logarithm.

The function $\varphi \to \log\varphi$ is defined for all $\varphi \in \mathfrak{A}_1$ such that $\varphi(0)$ does not lie on the branch cut. By using Eq. (3.2), we have

$$(\log\varphi)(0) = \log\varphi(0)$$

$$(\log\varphi)(x) = \sum{}^{*}\left\{\frac{(-1)^{k-1}}{k\varphi(0)^k}\varphi(x_1)\cdots\varphi(x_k) : (x_1,\ldots,x_k) \in Q(x)\right\} \quad \text{if } x \neq 0. \tag{4.3}$$

Notice that the value of $(\log\varphi)(x)$ is independent of the branch of $\log z$ if $x \neq 0$. It follows from Theorem 3.2 that

$$\exp(\log\varphi) = \varphi.$$

Moreover, for each $\varphi \in \mathfrak{A}(X_e)$ there is a branch of $\log z$ such that

$$\log(\exp\varphi) = \varphi.$$

In particular, this holds if $\varphi(0)$ is positive and $\log z$ is the principal branch. This, together with Eq. (4.2), yields the addition formula

$$\mathrm{Log}(\varphi_1 * \varphi_2) = \mathrm{Log}\varphi_1 + \mathrm{Log}\varphi_2, \tag{4.4}$$

where $\mathrm{Log}z$ is the principal branch and $\varphi_1(0), \varphi_2(0)$ are both positive. Finally, from Theorem 3.3, we have

$$D_x\log\varphi = \varphi^{-1} * D_x\varphi \quad \text{if } l(x) = 1.$$

Definition 4.1: Let $\mathfrak{A}_+$ denote the class of all $\psi \in \mathfrak{A}(X_e)$ such that $\psi(0)$ is real and strictly positive. For each $\psi \in \mathfrak{A}_+$, the corresponding Ursell–Mayer function φ is given by

$$\varphi = \mathrm{Log}\psi.$$

An important property of Ursell–Mayer functions is that φ vanishes when ψ factors in the following sense: Let Y, Z be a disjoint pair of subsets of X, and let $\psi \in \mathfrak{A}_+$ be such that $\psi(yz) = \psi(y)\psi(z)$ for all $y \in Y_e$, $z \in Z_e$. Then $\mathrm{Log}\psi(yz) = 0$ for all $y \in Y_e, z \in Z_e$ such that $l(y), l(z) \geq 1$ (see Ref. 4, Theorem 3.4.2, and Ref. 2, p. 60). Regarding integrability, we have, by Theorem 3.4,

Theorem 4.2: Suppose $\psi \in \mathfrak{A}_+$ is integrable. Then $\varphi = \mathrm{Log}\,\psi$ is integrable if $\int |\psi|\, d\xi_e < 2\psi(0)$.

In fact, the following example shows that $\mathrm{Log}\psi$ need not be integrable if $\int |\psi|\, d\xi_e \geq 2\psi(0)$: Let X be a singleton set $\{a\}$ with total measure 1, α be a positive number, and

$$\psi(x) = \begin{cases} 1 & \text{if } x = 0 \\ \alpha & \text{if } x = a \\ 0 & \text{otherwise} \end{cases}.$$

Then

$$\int \psi d\xi_e = 1 + \alpha,$$

$$(\mathrm{Log}\psi)(x) = \begin{cases} 0 & \text{if } x = 0 \\ (n-1)!(-1)^{n-1}\alpha^n & \text{if } x = a^n, \quad n = 1, 2, \ldots \end{cases}$$

and

$$\int |\mathrm{Log}\psi| d\xi_e = \sum_{n=1}^{\infty} \frac{\alpha^n}{n},$$

which converges if and only if $\alpha < 1$.

We conclude the discussion by showing that the exp and Log mappings commute with certain linear operators T, U which arise from the study of probability densities and correlation functions in statistical mechanics.[4,5] To this end, we introduce the following subspace $\mathfrak{N}$ of $\mathfrak{L}_1(X_e)$: Let $z^l\varphi$ denote the pointwise product $(z^l\varphi)(x) = z^{l(x)}\varphi(x)$. Then $\mathfrak{N}$ is the set of all ξ_e-measurable functions such that, for each complex number z and all $x \in X_e$, the function $z^l D_x\varphi \in \mathfrak{L}_1(X_e)$. We define two linear operators T, U: $\mathfrak{N} \to \mathfrak{N}$ by the equations

$$T\varphi(x) = \int D_x\varphi(y) d\xi_e(y), \tag{4.5}$$

$$U\varphi(x) = \int (-1)^{l(y)} D_x\varphi(y) d\xi_e(y). \tag{4.6}$$

It is known that T maps $\mathfrak{N}$ one-to-one onto $\mathfrak{N}$ and with U as its inverse (see Ref. 5, Theorem 3.1).

Theorem 4.3: If $\varphi \in \mathfrak{N}$, then $\exp\varphi \in \mathfrak{N}$ and

$$T \exp\varphi = \exp(T\varphi), \qquad U \exp\varphi = \exp(U\varphi).$$

Proof: It follows from Theorem 4.1 and the identities $z^l 1^* = 1^*$, $z^l F(\varphi) = F(z^l\varphi)$, that

$$z^l D_x(\exp\varphi) = \sum{}^*\{(1/k!) \exp(z^l\varphi) * z^l D_{x_1}\varphi * \cdots * z^l D_{x_k}\varphi : (x_1, \ldots, x_k) \in Q(x)\}.$$

Therefore,

$$\int |z^l D_x(\exp\varphi)| d\xi_e \le \int \sum{}^*\{(1/k!) |\exp(z^l\varphi)| * |z^l D_{x_1}\varphi| * \cdots * |z^l D_{x_k}\varphi| : (x_1, \ldots, x_k) \in Q(x)\} d\xi_e$$

Now $\exp(z^l\varphi)$ is integrable by Theorem 3.4. Hence, in view of Theorem 2.5, this expansion is equal to

$$\sum{}^*\{(1/k!) \int |\exp(z^l\varphi)| d\xi_e \int |\exp(D_{x_1}\varphi)| d\xi_e \cdots \times \int |z^l D_{x_k}\varphi| d\xi_e : (x_1, \ldots, x_k) \in Q(x)\}$$

so that $z^l D_x(\exp\varphi)$ is integrable. By a similar calculation for the case when $z = \pm 1$, together with Eq. (4.1) and the identities

$$\int \exp(\varphi) d\xi_e = \exp[T\varphi(0)], \quad \int \exp[(-1)^l\varphi] d\xi_e = \exp[U\varphi(0)],$$

one obtains the remaining results.

Corollary 4.1:

(i) If $\varphi(0), T\varphi(0)$ are both positive and $\mathrm{Log}\varphi \in \mathfrak{N}$, then $\varphi \in \mathfrak{N}$ and $T\,\mathrm{Log}\varphi = \mathrm{Log}(T\varphi)$.

(ii) If $\varphi(0), U\varphi(0)$ are both positive and $\mathrm{Log}\varphi \in \mathfrak{N}$, then $\varphi \in \mathfrak{N}$ and $U\,\mathrm{Log}\varphi = \mathrm{Log}(U\varphi)$.

In Ref. 5, we have shown that the state of an infinite system can be represented by a family of compatible probability densities $\{w_n\}$ or by the correlation function ρ. If we consider the Ursell–Mayer functions $u_n = \mathrm{Log}w_n$, $\zeta_n = \mathrm{Log}\rho^{(n)}$, the above results show that if either u_n or ζ_n belongs to $\mathfrak{N}$, then $w_n, \rho^{(n)}, u_n, \zeta_n$ belong to $\mathfrak{N}$ and $\zeta_n = Tu_n$, $u_n = U\zeta_n$. In other words, the following "commutative diagram" holds.

$$\begin{array}{ccc} \zeta_n & \underset{T}{\overset{U}{\rightleftarrows}} & u_n \\ \mathrm{Log}\uparrow\downarrow\exp & & \exp\uparrow\downarrow\mathrm{Log} \\ \rho^{(n)} & \underset{U}{\overset{T}{\rightleftarrows}} & w_n \end{array}$$

Thus, in particular, the Ursell–Mayer functions u_n, ζ_n can be used to represent the state of an infinite system.

---

1 D. Ruelle, Rev. Mod. Phys. **36**, 580 (1964).
2 J. Schwartz, "Statistical Mechanics", lecture notes (Courant Institute of Mathematical Sciences, New York University, 1957).
3 D. S. Carter and P. M. Prenter, Z. Wahrscheinlichkeitstheorie Geb. **21**, 1 (1972).
4 C. Y. Shen, Technical Report No. 35 (Department of Mathematics, Oregon State University, Corvallis, 1967).
5 C. Y. Shen and D. S. Carter, J. Math. Phys. **12**, 1263 (1971).

Pulse Propagation in a Dispersive Medium with Moving Density Profile*

G. M. Whitman†
Newark College of Engineering, Newark, N.J.

and

L. B. Felsen
Polytechnic Institute of Brooklyn, Farmingdale, N.Y.

(Received 4 January 1971)

Space–time ray methods are developed and employed to obtain asymptotic solutions for pulse propagation in media with slow spatial and temporal inhomogeneities. Emphasis is placed on cold isotropic plasmas whose plasma frequency ω_p has a space–time dependence of the form $(z\text{-}Vt)$, thereby simulating ordered profile motion. Asymptotic results are constructed directly by the ray method, are related to those in a spatially inhomogeneous but temporally invariant plasma by the Lorentz transformation, and are also compared with an exact closed form solution for the special case of exponential variation of $\omega_p(z - Vt)$. The presentation stresses physical attributes of the propagation process through extensive use of space–time ray diagrams and plane wave dispersion surfaces.

I. INTRODUCTION

Ray methods have provided a highly useful and physically appealing procedure for analyzing propagation and scattering of high-frequency time–harmonic fields in a relatively arbitrary environment. More recently, such techniques have been applied also to time-dependent fields in dispersive media, with space–time rays generalizing the conventional monochromatic rays of geometrical optics. Previous studies have dealt with the space–time ray theory for dispersive media with temporally invariant, spatially inhomogeneous[1,2] and spatially invariant, temporally inhomogeneous[3] properties. In the present paper, these earlier results are generalized to dispersive media exhibiting "slow" temporal (t) and spatial ($\mathbf{r}$) variation, with emphasis on a functional dependence in the form $(\mathbf{r} - \mathbf{V}t)$ expressive of ordered motion with uniform velocity $\mathbf{V}$. Slow variation implies essentially applicability of WKB-type considerations. Although not treated here, abrupt changes can also be accommodated. Examined in detail are Green's functions for cold isotropic plasmas with plasma frequency $\omega_p(\mathbf{r}, t) = \omega_p(\mathbf{r} - \mathbf{V}t)$, which may simulate the motion of an ionization profile.

For various source configurations, electromagnetic fields in isotropic plasmas which vary sufficiently slowly in both space and time can be expressed approximately in terms of Green's functions $G(\mathbf{r}, t)$ satisfying the scalar wave equation

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\omega_p^2(\mathbf{r}, t)}{c^2}\right)G(\mathbf{r}, t) = -\delta(\mathbf{r})\delta(t) \qquad (1)$$

subject to the causality condition

$$G \equiv 0, \quad t < 0. \qquad (1')$$

∇ is the gradient operator in real space and c is the speed of light in vacuum. We begin in Sec. II by applying the direct ray method[1,3] in order to determine the solution asymptotically at large distances and at long observation times. By this procedure, the field is found along curves in space–time called "rays," each ray defining a path in $(\mathbf{r}, t)$ space traversed by a wavepacket. As noted above, attention is confined essentially to media which vary in space–time according to $(\mathbf{r} - \mathbf{V}t)$; in particular, a one-dimensional spatial model is treated so that $\omega_p(\mathbf{r} - \mathbf{V}t) \to \omega_p(z - Vt)$. Under these circumstances, it is found that rays are characterized by the constancy of $a = (\omega - kV)$, where ω is the instantaneous frequency and k the instantaneous wavenumber of the corresponding wavepacket. This conservation condition facilitates solution (analytically or graphically) and interpretation of the propagation problem, and it generalizes the analogous conditions ω = constant and k = constant for temporally and spatially invariant media, respectively. The asymptotic solution is constructed and a number of interesting physical properties are discussed in detail.

In Sec. III, the one-dimensional analog of (1′) with $\omega_p = \omega_p(z - Vt)$ is treated by an alternative procedure involving (Lorentz) transformation to a new reference frame (z', t') that moves with velocity V relative to the original frame (z, t). When this is done, the field in the (z', t') space is governed by a scalar wave equation identical with the original one except that $\omega_p(z, t)$ is replaced by $\omega_p'(z')$ descriptive of a dispersive, *time-invariant,* inhomogeneous, isotropic medium. The problem thus simplifies and can be treated by integral (Fourier) transform methods as an alternative to the direct ray procedure. For illustration, $\omega_p'(z')$ is taken as an exponentially varying function for which an exact solution is available.[4] The asymptotic form of this solution exhibits certain peculiarities which are not easily interpreted per se but become clarified when viewed in terms of space–time rays, since it is then possible to separate effects of the medium from those due to the excitation. This aspect of the discussion has implications of greater generality than demonstrated by this special example, and it highlights peculiarities attributable to an unduly idealized model of the medium (see Sec. V).

Exact solutions, albeit for special problems, provide an important adjunct to the asymptotic theory since their availability permits assessment of the validity of the asymptotic results, and also provides means for constructing transition functions in parameter regimes where simple space–time ray considerations fail. Section IV contains a certain class of exact solutions (derived in the Appendix), a special case of which is the previously mentioned exponential dependence of $\omega_p(z - Vt)$.

II. DIRECT RAY METHOD

A. Construction of the Solution

Since ω_p in (1) depends on both space and time, a reduction of (1) by integral transforms is generally not possible. For asymptotic solutions, one can, however, apply the "direct ray method."[1,3] To do this, we introduce into (1) a large parameter λ via the transformation $t \to \lambda t$, $\mathbf{r} \to \lambda\mathbf{r}$, to get the equation

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \lambda^2 \frac{\omega_p^2(\mathbf{r},t)}{c^2}\right) G(\mathbf{r},t) = -\delta(\mathbf{r})\delta(t), \quad (2)$$

subject to the causality condition

$$G \equiv 0, \quad t < 0. \quad (2')$$

It is understood that $G(\mathbf{r},t)$ and $\omega_p(\mathbf{r},t)$ in (2) are not the same functions that appear in (1) although the same notation is used for simplicity. The transformation from (1) to (2) implies the regime of long distances and observation times, and of slowly varying ω_p.

To solve (2) outside the source region, we assume an asymptotic solution of the form

$$G(\mathbf{r},t) \sim A_0(\mathbf{r},t)e^{i\lambda S(\mathbf{r},t)} + \sum_{m=1}^{\infty} \frac{1}{(i\lambda)^m} A_m(\mathbf{r},t)e^{i\lambda S(\mathbf{r},t)}, \quad (3)$$

where $S(\mathbf{r},t)$ and $A_m(\mathbf{r},t)$, $m = 0, 1, 2, \ldots$, are independent of λ. Substituting (3) into (2) for $(\mathbf{r},t) \neq (0,0)$, interchanging orders of differentiation and summation, and equating coefficients of different powers of λ to zero gives the dispersion equation

$$c^2k^2 - \omega^2 + \omega_p^2(\mathbf{r},t) = 0, \quad (4)$$

and the transport equation for the zeroth order amplitude function $A_0(\mathbf{r},t)$,

$$\left(\nabla\cdot\mathbf{k} + \frac{1}{c^2}\frac{\partial\omega}{\partial t} + 2\mathbf{k}\cdot\nabla + \frac{2}{c^2}\omega\frac{\partial}{\partial t}\right)A_0 = 0, \quad (5)$$

as well as transport equations for $A_m(\mathbf{r},t)$, $m \geq 1$. We restrict ourselves in the following to the lowest order approximation of G given by the first term on the right-hand side of (3). The wave vector $\mathbf{k}$ and frequency ω in (4) and (5) are defined, respectively, as

$$\mathbf{k} = \nabla S, \quad \omega = -\frac{\partial S}{\partial t}. \quad (5')$$

To solve (4), one introduces space–time ray trajectories via the characteristic (or ray) equations[1,5]

$$\frac{d\mathbf{r}}{dt} = \nabla_k\omega(\mathbf{k};\mathbf{r},t), \quad \frac{d\mathbf{k}}{dt} = -\nabla\omega(\mathbf{k};\mathbf{r},t), \quad \frac{d\omega}{dt} = \frac{\partial\omega}{\partial t}. \quad (6)$$

$\nabla_k\omega = \mathbf{v}_g$ defines the group velocity and ∇_k is the gradient operator in $\mathbf{k}$-space. The last term in (6), which follows from the first two equations and the derivative relation $d\omega/dt = \nabla_k\omega\cdot d\mathbf{k}/dt + \nabla\omega\cdot d\mathbf{r}/dt + \partial\omega/\partial t$, shows that ω is not constant along a ray if the medium properties vary with time, while the second equation in (6) shows that the wave vector $\mathbf{k}$ is not constant along a ray if the medium parameters vary with position. If the medium varies arbitrarily in space and time, it is not possible, in general, to solve (6) explicitly. However, explicit solutions can be found for special cases.

To find the phase of a wave along a given ray, one integrates the derivative relation $dS/dt = \nabla S\cdot d\mathbf{r}/dt + \partial S/\partial t$ from one space–time point on the ray to another. Via (5'), the result is

$$S(\mathbf{r},t) - S(\mathbf{r}_1,t_1) = \int_{\mathbf{r}_1,t_1}^{\mathbf{r},t} (\mathbf{k}\cdot\mathbf{v}_g - \omega)dt. \quad (7)$$

To determine the wave amplitude A_0, it is convenient to write (5) in the form

$$\Box\cdot\left(A_0^2(\mathbf{r},t)\frac{\omega(\mathbf{r},t)}{c^2}\mathbf{V}_g(\mathbf{r},t)\right) = 0, \quad \Box = \nabla + \alpha_0\frac{\partial}{c\partial t}, \quad (8)$$

where $\Box$ is the gradient operator in the four-dimensional $(\mathbf{r},ct)$ space, α_0 is a unit vector along the t coordinate, and $\mathbf{V}_g = \mathbf{v}_g + \alpha_0 c$ is the group velocity 4-vector tangent to the space–time ray. Integrating (8) over a volume contained in a space–time ray tube bounded by hyperplanes t = constant as shown in Fig. 1 and using the divergence theorem with the recognition that $\mathbf{V}_g\cdot\alpha_0 = c$, one finds

$$A_0(\mathbf{r},t) = A_0(\mathbf{r}_1,t_1)\,[\omega(\mathbf{r}_1,t_1)\Delta\mathbf{r}_1/\omega(\mathbf{r},t)\Delta\mathbf{r}]^{1/2} \quad (8')$$

along a ray, where $\Delta\mathbf{r}$ is the ray tube cross-section in 4-space (i.e., the spatial volume occupied by the wavepacket in 3-space) in the hyperplane t = constant. Determination of the initial field values at $(\mathbf{r}_1,t_1)$ requires separate consideration. The results in (7) and (8') are valid for $\mathbf{r}$ in 3-space dimensions and arbitrary (though slowly varying) $\omega_p(\mathbf{r},t)$; for propagation in one space dimension and ω_p defined as in (9), these equations reduce to (15) and (17), respectively.

Considerations in this paper are restricted primarily to propagation in one space dimension z and to spatial and temporal dependence of ω_p in the form of a moving electron density profile,

$$\omega_p = \omega_p(q), \quad (9)$$

where

$$q = z - \beta ct, \quad \beta = V/c. \quad (9')$$

In this case, one finds from the last two equations in (6) that $(d/dt)(\omega - kV) = 0$ or

$$\omega - \beta ck = a, \quad (10)$$

where a is constant along a ray. Combining (10) with the dispersion relation (4) gives the frequency (assumed positive) and wavenumber of the wavepacket in terms of $\omega_p(q)$ and the ray parameter a as [see (14) for justification of the inequality]

$$\omega = \gamma[a\gamma \pm \beta\sqrt{a^2\gamma^2 - \omega_p^2(q)}], \quad v_g \gtrless V \quad (11)$$

and

$$ck = \gamma[\beta a\gamma \pm \sqrt{a^2\gamma^2 - \omega_p^2(q)}], \quad v_g \gtrless V, \quad (12)$$

where $\gamma = (1-\beta^2)^{-1/2}$. It then follows from the first equation of (6), with $d\omega/dk = c^2k/\omega$ from (4), that the ray trajectories are defined by

$$\frac{dz}{cdt} = \frac{v_g}{c} = \frac{\beta a\gamma \pm \sqrt{a^2\gamma^2 - \omega_p^2(q)}}{a\gamma \pm \beta\sqrt{a^2\gamma^2 - \omega_p^2(q)}}, \quad v_g \gtrless V. \quad (13)$$

Since $dz/cdt = dq/cdt + \beta$ from (9'), it follows that

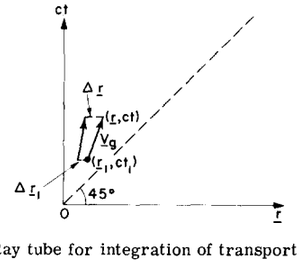


Fig. 1. Ray tube for integration of transport equation.

$$\frac{v_g}{c} - \frac{V}{c} = \frac{dq}{cdt} = \pm f_{\pm}(q), \qquad v_g \gtrless V, \tag{14}$$

where

$$f_{\pm}(q) = \sqrt{a^2\gamma^2 - \omega_p^2(q)}\{\gamma^2[a\gamma \pm \beta\sqrt{a^2\gamma^2 - \omega_p^2(q)}]\}^{-1} \tag{14'}$$

Thus, the ray equation has two separate forms depending on whether or not the group velocity of the wavepacket described by a exceeds the profile speed V.

From (7), (9), (9′), and (10), on a ray segment in (q, ct) space between source and turning point (if any; see Sec. IIB), the phase $S = S(z, t)$ of the signal at (z, t) on a ray can be expressed in terms of an integral over q as

$$S - S_1 = \int_{q_1}^{q} k(\xi)d\xi - a(t - t_1), \tag{15}$$

where $k(q)$ is given by (12) and $q_1 = z_1 - \beta ct_1$, $S_1 = S(z_1, t_1)$. Equation (15) reduces to the correct result when $\beta \to 0$.[5] Alternatively, it can be shown from (4), (14), and the result $v_g = c^2k/\omega$, that one may write instead of (15) [note that $dS/dt = -\omega + kv_g = -\omega_p^2(q)/\omega(q)$ and $dS/dq = (dS/dt)\cdot(dq/dt)^{-1}$]:

$$S - S_1 = \mp \frac{\gamma}{c}\int_{q_1}^{q} \frac{\omega_p^2(\xi)d\xi}{\sqrt{a^2\gamma^2 - \omega_p^2(\xi)}}, \qquad v_g \gtrless V. \tag{15'}$$

The amplitude A_0 of the signal is found conveniently by expressing (8′) in terms of q:

$$A_0(z, t) = \bar{A}_0(z_1, t_1)[\omega(a, q)|\Delta q/\Delta a|]^{-1/2}, \tag{16}$$

where

$$\bar{A}_0(z_1, t_1) = A_0(z_1, t_1)[\omega(a, q_1)|\Delta q_1/\Delta a|]^{1/2} \tag{16'}$$

involves the amplitude at the initial point (z_1, t_1) on the ray. Use has been made of the fact that at a fixed instant of time, the ray tube cross section $\Delta z = \Delta q$ and that the frequency of the signal can be expressed as a function of a and q as seen from (11). $(\Delta q/\Delta a)$ can be found by writing the ray equation (14) in the integral form $t = t(q)$ and differentiating implicitly. $\bar{A}_0(z_1, t_1)$ and $S(z_1, t_1)$ are determined from the asymptotic approximation of the known exact solution with $\omega_p = \omega_p(0) = \text{constant}$,[1,6] since the plasma frequency is assumed to remain essentially constant over the interval of time $0 \le t \le t_1$, with t_1 large enough to validate use of the asymptotic form (local homogeneity assumption; rays in this region of space-time are essentially straight). It is then found that $S_1 \to 0$ when $(z_1, t_1) \to (0, 0)$ and that along a ray path in (q, ct) space between the source and a turning point (if any),

$$A_0(z, t) = \frac{c^{3/2}e^{i\pi/4}}{2\sqrt{2\pi\gamma\lambda}[a^2\gamma^2 - \omega_p^2(q_1)]^{1/4}[a^2\gamma^2 - \omega_p^2(q)]^{1/4}\left|\int_{q_1}^{q} d\xi\ \omega_p^2(\xi)/[a^2\gamma^2 - \omega_p^2(\xi)]^{3/2}\right|^{1/2}}. \tag{17}$$

Similar expressions for S and A_0 are obtained on ray segments beyond a turning point, by using the appropriate equation for the turned ray.

B. Ray Diagrams, Dispersion Surfaces, and Physical Interpretation

While the ray equation (14) must generally be solved numerically, a graphical scheme can be employed to visualize the space-time ray trajectories. It is recalled that space-time rays are normal to the dispersion surface descriptive at given $(\mathbf{r}, t)$ of the real solutions of (4) in $(\mathbf{k}, \omega)$ space.[5] For the one-dimensional case, one selects a value $a = a_1$ in (10) and locates the point of intersection P_0 of the line $\omega - \beta ck = a_1$ with the dispersion curve corresponding to the starting point $z = z_0$, $t = t_0$ of the ray. The initial ray direction is normal to the curve at P_0, the sense of the normal being such that time increases along the ray. After a time interval Δt, the ray reaches the point $z_1 = z_0 + \Delta z$, $t_1 = t_0 + \Delta t$. The medium now is characterized by $\omega_p(z_1 - \beta ct_1)$, which value of plasma frequency gives rise to a new dispersion curve. The ray direction at z_1, t_1 is then found from the normal at the intersection point of the new dispersion curve and the line $a = a_1$. This process is continued for subsequent space-time intervals to yield the entire ray trajectory (see Fig. 2). The procedure also forms the basis for numerical integration of the ray equation. After the ray paths have been determined,

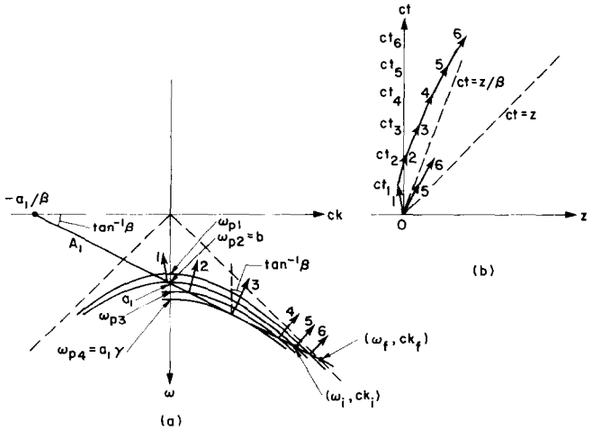


Fig. 2. Graphical construction of ray paths in an isotopic electron plasma whose plasma frequency decreases monotonically as a function of $(z - \beta ct)$. (a) Dispersion surfaces $\omega(z, t; k) = [c^2k^2 + \omega_p^2(z - \beta ct)]^{1/2}$ and tracking condition $(\omega - \beta ck) = a_1$ for a typical space-time ray. (b) Space-time ray trajectories.

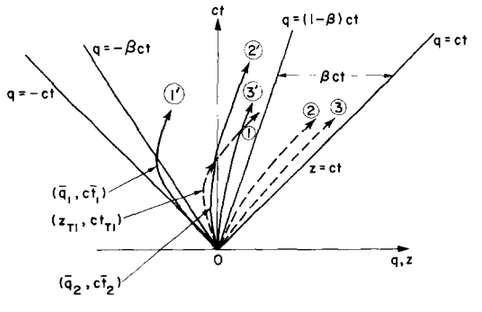


Fig. 3. Ray trajectories in (q, ct) space and (z, ct) space for an isotropic plasma whose plasma frequency decreases monotonically with $q = z - \beta ct$. Solid arrows (→): (q, ct) space. Dashed arrows ((– – →): (z, ct) space. 1 and 1′, 2 and 2′, and 3 and 3′ identify rays with initial values $v_g < 0$, $0 < v_g < V$, and $v_g > V$, respectively.

the phase and amplitude of the signal can be calculated in a manner analogous to that discussed in Ref. 3.

For a monotonically decreasing profile as in (18), it follows from (14) (see Fig. 3) that if a wavepacket initially has a speed $v_g < V, q$ becomes increasingly more negative along the ray, reaches a minimum value $\bar{q}$ [the turning point in (q, ct) space where $dq/cdt = 0$], increases to zero, becomes positive, and finally approaches $+\infty$ as $t \to \infty$; correspondingly, dq/cdt, which is initially negative when $q = 0$, becomes less negative as $q \to \bar{q}$, is zero at $\bar{q}$, thenceforth becomes increasingly more positive and approaches $(1 - \beta)$ as $q \to \infty$. Since $z = q + \beta ct$, a ray trajectory in (z, ct) space (rays 1, 2, 3 in Fig. 3) is similar to a ray trajectory in (q, ct) space (rays 1′, 2′, 3′ in Fig. 3) and differs only in that each point in (q, ct) space along a ray is shifted to the right by an amount (βct), with associated slopes related by $v_g/c = dq/cdt + \beta$, where $v_g = dz/dt$. The distance (βct) is given graphically by the horizontal segment contained between the lines $q = (1-\beta)ct$ and $q = ct$ shown in Fig. 3.

If $v_g > 0$ initially, i.e., $dq/dt > -\beta$ (rays 2′ and 3′ in Fig. 3), then the wavepacket in (z, ct) space (ray 2 in Fig. 3) will continue to propagate in the $+z$ direction with increasing speed, with the result that $v_g \to c$, as $t \to \infty$ (for monotonically decreasing $\omega_p(q)$, one has $\omega_p \to 0$ as $q \to \infty$). However, if $v_g < 0$ initially, i.e., $dq/cdt < -\beta$ (ray 1′ in Fig. 3), the wavepacket will continue to propagate in the negative z direction with decreasing speed; when $dq/cdt \to -\beta$, the packet will slow down and reverse direction [$dz/dt = 0$ at (z_T, ct_T)] and thenceforth continue to travel in the positive z direction with ever increasing speed approaching c as $t \to \infty$ (ray 1 in Fig. 3). Because of the above variation in q (see Fig. 3, rays 1′ or 2′) and since the plasma frequency is assumed to be a monotonically decreasing function of q, $\omega_p(q)$ increases initially along the ray, reaches a maximum equal to $(a\gamma)$ at $(\bar{q}, c\bar{t})$, and then decreases to zero as $q \to \infty$. The ray turns at $(\bar{q}, c\bar{t})$ because $\omega_p(\bar{q}) = a\gamma$ is the maximum allowable value of plasma frequency for which the frequency and wavenumber of the associated wavepacket are real [see (11) and (12)]. Note that $\omega_p(\bar{q}) = a\gamma$ corresponds to $v_g = V$, so that a wavepacket sees the largest plasma frequency along its trajectory when its group speed equals the profile speed V. Moreover, $k \neq 0$ when $\omega_p(\bar{q}) = a\gamma$ but $k = 0$ when $\omega_p = a$, which latter value corresponds to $v_g = 0$. Thus in (z, ct) space, the ray turns at a time $t_T < \bar{t}$ and at a plasma frequency $\omega_p = a$ which is smaller than the maximum allowable value $\omega_p(\bar{q}) = a\gamma$. The maximum plasma frequency encountered by a given ray can be determined graphically since it corresponds to the dispersion curve which is tangent to the line $a =$ constant. In Fig. 2(a), ray 3 is normal to the line $a = a_1$ which is tangent to the dispersion curve identified by $\omega_{p4} = a_1\gamma$. Thus the wavepacket identified by the parameter a_1 reaches the group speed $v_g = V$, when $q = \bar{z} - \beta c\bar{t}_2 = \bar{q}$ as given in Fig. 2(b) and $\omega_p(\bar{q}) = \omega_{p4}$. Figure 2(b) shows a typical ray trajectory, and one observes from Fig. 2(a), as noted above, that the ray turns when $\omega_p = a_1$, which value is smaller than $\omega_p(\bar{q}) = a_1\gamma$ but larger than $\omega_p = b$.

When the initial group speed of a wavepacket is greater than V, then from (14), q can be shown to increase along a ray (see Fig. 3, ray 3′) and dq/cdt, initially positive, becomes more positive as $t \to \infty$. Therefore, v_g increases continually and approaches c as $t \to \infty$ (Fig. 3, ray 3). Since $\omega_p(q)$ decreases along the ray, the normals to the corresponding dispersion curves proceed along the line $a =$ constant to larger and larger values of both frequency and wavenumbers [see normals 5 and 6 in Fig. 2(a)]. The graphical construction makes it possible to predict the range over which the central frequency and wavenumber of a given wavepacket vary. As an example, consider the wavepacket identified by a_1, propagating in space–time along the initial direction labeled 5 in Fig. 2(a), with the initial value of v_g greater than V. The initial (ω, ck) values are denoted by (ω_i, ck_i), whereas the final values are (ω_f, ck_f) found at the intersection of the lines $a = a_1$ and $\omega = ck$.

In the limit $\beta \to 0$, the turning point in (z, ct) space occurs at the maximum value of $\omega_p(z)$ seen by a ray.[5] This same observation follows from the above discussion since in this limit, the line $a = a_1$ in Fig. 2(a) becomes horizontal (i.e., $\omega = \omega_1$). The maximum plasma frequency encountered by the wavepacket then corresponds to the dispersion curve which is tangent to this horizontal line; tangency occurs at $\omega = \omega_1 = \omega_p(z)$, $k = 0$.

C. Exponential Profile

For certain profiles of $\omega_p(q)$, the ray equation (14) can be integrated to yield explicit forms for $t = t(q)$. As an illustration, we assume an exponentially varying plasma frequency,

$$\omega_p = be^{-q/h}, \quad q = z - \beta ct, \tag{18}$$

where b and h are positive constants. Thus, at any fixed position z, the plasma frequency grows exponentially with t, whereas at any instant of time, the plasma frequency decays to zero as $z \to \infty$. A study of (14) or application of the graphical procedure as in Fig. 2 reveals that all of the rays eventually travel in the positive z direction. For a ray which either does not turn ($v_g > V$ initially) or has not turned as yet ($v_g < V$ initially and $t < \bar{t}$, where $\bar{t}$ is defined as that instant of time when the wavepacket group speed is V), the ray trajectory is found explicitly by integrating (14) from $q = 0$ to some value q. The resulting equation describing the ray trajectory is [the integral becomes elementary on introducing the change of variable $x = (b/a\gamma)\exp(-q/h)$]

$$\frac{1-u}{1+u}\cdot\frac{1+u_0}{1-u_0} = e^{\mp 2(c\beta\gamma^{-2}-\beta q)/h} = e^{\mp 2(ct-\beta z)/h}, \quad v_g \gtrless V, \tag{19}$$

where

$$u = u(q) = (1 - b^2e^{-2q/h}/a^2\gamma^2)^{1/2}, \qquad u_0 = u(0) = (1 - b^2/a^2\gamma^2)^{1/2}. \tag{19′}$$

However, if $v_g < V$ initially and $t > \bar{t}$, then the ray trajectory is found by integrating $dq/cdt = -f_-(q)$ from $q = 0$ to $q = \bar{q}$ defined as that value of q for which $v_g = V$, and combining this result with the integration of $dq/cdt = f_+(q)$ from $\bar{q}$ to q since v_g is now greater than V. The ray trajectory for $t > \bar{t}$ (i.e., after the turning point) is now described by the relation

$$\frac{1-u}{1+u}\cdot\frac{1-u_0}{1+u_0} = e^{-2(c\beta\gamma^{-2}-\beta q)/h} = e^{-2(ct-\beta z)/h}. \tag{20}$$

To express the ray parameter a explicitly in terms of z and t, it is convenient to form the ratio

$$(1-u^2)/(1-u_0^2) = e^{-2(z-\beta ct)/h} \tag{21}$$

and to introduce new coordinates (η, ξ) defined in (37). Multiplying and dividing (21) and (19) yields the two relations $(1 \pm u) = D_{\pm}(1 \pm u_0)$, where D_+ and D_- equal $e^{2\eta'}$ and $e^{-2\xi'}$, respectively, when $v_g > V$; but equal $e^{-2\xi'}$ and $e^{2\eta'}$, respectively, when $v_g < V$ and $t < \bar{t}$, with $\xi' = \bar{a}_1\xi, \eta' = \bar{a}_2\eta$, $\bar{a}_{1,2} = 1 \mp \beta$. Solving for u_0, and using (19′) and a few trigonometric identities, one finds the ray path parametrically in terms of $a = \omega - \beta ck$ as

$$\sqrt{2}(\omega - \beta ck)\gamma = \frac{\pm\, b e^{-(z-\beta ct)/2h} \sinh[(ct-\beta z)/h]}{\{\cosh[(ct-\beta z)/h] - \cosh[(z-\beta ct)/h]\}^{1/2}}, \tag{22}$$

for $a \gtrless 0$. One may show that (22) remains valid for $t > \bar{t}$. In the limit of vanishing profile speed $\beta \to 0$, i.e., for the stationary inhomogeneity $\omega_p \to b e^{-z/h}$, $dq/cdt = f_+(q)$ may be shown to reduce to the correct form of the ray equation when v_g is initially positive; the solution (22) also produces the appropriate ray family. Similar conclusions follow when $V = 0$, $v_g < 0$. It may also be shown that the ray tube cross section Δq or Δz corresponding to (22) is nonzero at all points along a ray so that the ray configuration does not form a caustic [see (46)].

For $\omega_p(q)$ given by (18), it is also possible to determine explicit expressions for the phase and amplitude of the signal. With $(z_1, t_1) = (0, 0)$, which implies $S_1 = 0$, and before turning points (if any), (15′) can be evaluated to yield

$$S = \mp\, (\gamma^2 ah/c)(u - u_0), \qquad v_g \gtrless V, \qquad t < \bar{t}, \tag{23}$$

where u and u_0 are defined in (19′). Following the procedure used in deriving (22), one may reduce (23) to

$$S = -\, b\zeta^{1/2}, \tag{24}$$

where

$$\zeta = 2\left(\frac{\gamma h}{c}\right)^2 e^{-(z-\beta ct)/h}\left[\cosh\left(\frac{ct-\beta z}{h}\right) - \cosh\left(\frac{z-\beta ct}{h}\right)\right]. \tag{24′}$$

Similarly, for $\omega_p(q)$ given by (18) and with $q_1 \to 0$, (17) reduces to

$$A_0(z,t) = \frac{c^{3/2} e^{i\pi/4}}{2\gamma[2\pi ah\lambda(u-u_0)]^{1/2}}, \tag{25}$$

which yields on use of (23) and (24),

$$A_0(z,t) = \frac{c e^{i\pi/4}}{2(2\pi b\lambda)^{1/2}\zeta^{1/4}}, \tag{26}$$

where ζ is given by (24′). Thus from (3), (24), and (26), with λ set equal to unity [i.e., for G as defined in (1)], the leading term in the asymptotic approximation of the scalar wavefield is [note that from (4), the solution contains contributions from (k, ω) and $(-k, -\omega)$]

$$G(z,t) \sim \frac{c}{(2\pi b)^{1/2}\zeta^{1/4}} \cos(b\zeta^{1/2} - \pi/4), \tag{27}$$

which expression is valid at distant observation points sufficiently long after the arrival of the first response. The above result checks with the asymptotic approximation of the exact solution in Sec. III, obtained on replacing the Bessel function in (35) by its large argument form. Equation (27) also reduces to the correct result when $\beta \to 0$ (i.e., $\omega_p \to b e^{-z/h}$); when $h \to \infty$ with β/h finite (i.e., $\omega_p \to b e^{\beta ct/h}$, which is studied in Ref. 3); and when $h \to \infty$ (i.e., $\omega_p \to b$). From (24) or from (27) and the defining relation (5′), the time-dependent local oscillation frequency (in radians) of the signal is seen to be $|b(\partial/\partial t)\zeta^{1/2}|$ and the local spatial wavenumber $|b(\partial/\partial z)\zeta^{1/2}|$.

D. Field Behavior near the Wavefront

As discussed in Refs. 3, 5, the wavepacket description of propagation in dispersive media is not applicable everywhere in space–time but fails in so-called transition regions. One such transition region, relevant for the present problem, surrounds the wavefronts $ct \sim |z|$. The procedure in Refs. 2, 5 for determining the fields in this parameter regime can be used here and yields for the problem described by the spatially one-dimensional analog of (1) with arbitrary $\omega_p(q)$:

$$G(z,t) \sim \tfrac{1}{2} c\, J_0(2\sqrt{\nu\tau})\, U(\tau), \tag{28}$$

where J_0 is the zeroth order Bessel function, $U(\tau)$ is the Heaviside unit function, and

$$\tau = \gamma(1 \pm \beta)(t \mp z/c), \qquad \nu = \pm(\gamma/2c)\int_0^q \omega_p^2(\tilde{q})\, d\tilde{q}, \quad q \gtrless 0, \tag{28′a}$$

with $q = z - \beta ct$. Equation (28) is obtained conveniently by first solving the problem in (z', t') space via the Lorentz transformation (see Sec. III), applying the procedure in Refs. 2, 5, and transforming to (z, t) space. For the special case $\omega_p = b e^{-q/h}$, it follows from (28′) that

$$\nu = (\pm\, b^2\gamma h/4c)(1 - e^{-2q/h}), \qquad q \gtrless 0. \tag{29}$$

It can be shown that the early time response (28), with $|\nu| \gg 1$ given by (29), connects with the asymptotic result (27) when $2\sqrt{\nu\tau}$ is sufficiently large; this implies that the Bessel function can be replaced by its large argument form although $(ct - |z|)$ remains small.

III. THE LORENTZ TRANSFORMATION

Instead of looking for solutions of the one-dimensional analog of (1) [with $\omega_p(z,t) = \omega_p(z - Vt)$] in (z, t) space, it is advantageous to introduce new coordinates (z', t') via the Lorentz transformation

$$z' = \gamma(z - \beta ct), \qquad ct' = \gamma(ct - \beta z), \tag{30}$$

thereby reducing the problem to one in a time-independent medium:

$$\left(\frac{\partial^2}{\partial z'^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t'^2} - \frac{\omega_p'^2(z')}{c^2}\right) G'(z', t') = -\,\delta(z')\delta(t'), \tag{31}$$

subject to the causality condition

$$G' \equiv 0, \qquad t' < 0. \tag{31′}$$

As is well known, the transformation (30) leaves the operator $(\partial^2/\partial z^2 - \partial^2/c^2\partial t^2)$ and the source term $\delta(z)\delta(t)$ invariant. In (31), $G(z,t) = G'(z', t')$ and

$\omega_p(z,t) = \omega_p(z'/\gamma) = \omega_p'(z')$. The asymptotic analysis of (31) by the direct ray method is available[1] and has been further elaborated by graphical methods utilizing the plane wave dispersion surfaces.[5] The solution to (31) can be formulated rigorously in terms of temporal Laplace transforms suitable for subsequent asymptotic evaluation by the saddle point method[1]; the asymptotic evaluation of the exact solution may be used for comparison to validate the ray solution. An analogous situation occurs when ω_p depends only on time so that the roles of z and t are essentially interchanged.[3] When $\omega_p = b\ \exp[-z'/\gamma h]$, an exact closed form solution for G' can be found from Ref. 4 while Ref. 3 deals with its counterpart for $\omega_p = b\ \exp[ct'/\gamma h]$. Once a solution in the moving (z', t') frame is obtained, be it exact or approximate, one may find the solution in the stationary frame simply by transforming back to the original (z, t) space. If in the limit of large $V, \omega_p(z - Vt)$ reduces to a function of time only, then solutions in the stationary frame reduce to solutions in a temporally varying spatially homogeneous environment. Thus, solutions for time varying media can be generated from those for spatially varying media by using the Lorentz transformation as described above and then passing to the mathematical limit $V \to \infty$.

For illustration, let us again consider the cold electron plasma characterized by the exponentially varying plasma frequency (18). Applying the Lorentz transformation, one has

$$\omega_p(z,t) = be^{-(z-\beta ct)/h} = be^{-z'/\gamma h} \equiv \omega_p'(z'). \tag{32}$$

Using the ray method or evaluating the exact integral representation for $G'(z', t')$ asymptotically,[1] one can show that

$$G'(z',t') \sim c[/(2\pi b)^{1/2}(\zeta')^{1/4}]^{-1} \cos(b\sqrt{\zeta'} - \tfrac{1}{4}\pi), \tag{33}$$

where

$$\zeta' = 2\left(\frac{\gamma h}{c}\right)^2 e^{-z'/\gamma h}\left(\cosh\frac{ct'}{\gamma h} - \cosh\frac{z'}{\gamma h}\right). \tag{33'}$$

This result checks with the asymptotic approximation of the exact solution

$$G'(z',t') = \tfrac{1}{2}c\ J_0(b\sqrt{\zeta'})U(ct' - |z'|), \tag{34}$$

when the zeroth order Bessel function J_0 is replaced by its large argument form; $U(x) = 0$ or 1 for $x < 0$ and $x > 0$, respectively. The exact solution in (34) can be readily inferred from Eq. (3.18) of Ref. 4. Transforming (33) and (34) to the stationary frame via (30), one obtains, respectively, (27) and its associated exact solution

$$G(z,t) = \tfrac{1}{2}c\ J_0(b\sqrt{\zeta})U(ct - |z|), \tag{35}$$

where ζ is given by (24').

For $\omega_p'(z')$ given by (32) with $h < 0$, the ray construction associated with Fig. 2 of Ref. 5 is applicable. Once the rays are drawn, the Lorentz transformation can then be used to map these rays from (z', t') onto (z, t) space. A ray configuration results which is essentially the same as that depicted in Fig. 2.

In the moving frame wherein the medium appears only spatially inhomogeneous, ω' remains constant along a ray.[5] Since the asymptotic wave solution as found above behaves locally like a plane wave, its phase can be shown to remain invariant under the Lorentz transformation.[7] As a consequence, the frequency and wavenumber of a wavepacket viewed from both the moving (primed) and stationary (unprimed) frames are related by the equations

$$\omega' = \gamma(\omega - \beta ck), \qquad ck' = \gamma(ck - \beta\omega). \tag{36}$$

Thus, constancy of ω' along a ray in (z', t') space implies the constancy of $(\omega - \beta ck)$ along the same ray viewed from the stationary frame, thereby confirming the result in (10).

If $\beta \to 0$, the straight line A_1 in Fig. 2(a) becomes horizontal. The construction in Fig. 2 therefore reduces to that in Ref. 5. If, on the other hand, $V, h \to \infty$ such that (V/h) remains finite, then line A_1 becomes vertical and the construction of the ray configuration due to localized sources in a temporally varying medium follows.[3] In this same limiting case, all of the results in this section reduce to those in Sec. IIC of Ref. 3.

IV. EXACT CLOSED FORM SOLUTIONS

As mentioned earlier, the exponential form of $\omega_p(z, t)$ in (18) has been chosen for illustration because an exact closed form solution can then be found. In fact, it is possible to accommodate a more general class of $\omega_p(z, t)$ profiles. If (η, ξ) are normal coordinates defined by the relations

$$2h\eta = ct - z, \qquad 2h\xi = ct + z, \tag{37}$$

then as shown in the Appendix, closed form solutions may be obtained for plasma frequencies with product separability of the form

$$\omega_p(z,t) = bN(\eta)Z(\xi), \qquad N(0) = 1, \tag{38}$$

where b is a constant. In particular, the exact solution of the scalar wave equation

$$\left(\frac{\partial^2}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\omega_p^2(z,t)}{c^2}\right)\overline{G}(z,t) = -\delta(z)\frac{\partial}{\partial t}\delta(t), \tag{39}$$

subject to the causality condition $\overline{G} \equiv 0$, $t < 0$, and with ω_p as in (38) is given by

$$\overline{G} = \frac{c^2}{2}\left[\delta(ct - |z|) - W_0(\eta,\xi)U(ct - |z|)\right], \tag{40}$$

where

$$W_0(\eta,\xi) = b\left(\frac{v^2 - w^2}{2}\ \frac{J_1(b\sqrt{\zeta})}{\sqrt{\zeta}}\right) \tag{40'a}$$

with

$$\zeta = \left(4h\int_0^\eta N^2(\hat{\eta})d\hat{\eta}\right)\left(\frac{h}{c^2}\int_0^\xi Z^2(\hat{\xi})d\hat{\xi} + B\right), \tag{40'b}$$

$$\frac{v^2 - w^2}{2} = \frac{h}{c^2}\int_0^\xi Z^2(\hat{\xi})d\hat{\xi} - \frac{4h}{A^2}\int_0^\eta N^2(\hat{\eta})d\hat{\eta} + B. \tag{40'c}$$

A and B are constants, $U(x)$ is the Heaviside unit function, and J_1 is the Bessel function of first order.

For the exponential variation in (18), N and Z in (40b) and (40c) have the form

$$N(\eta) = e^{\bar{\alpha}_2\eta}, \qquad Z(\xi) = e^{-\bar{\alpha}_1\xi}, \tag{41}$$

where $\bar{a}_{1,2} = 1 \mp \beta$. Since $N(\eta) \to 1$ and $Z(\xi) \to 1$, and therefore

$$\int_0^{\eta} N^2(\hat{\eta})d\hat{\eta} \approx \eta, \qquad \int_0^{\xi} Z^2(\hat{\xi})d\hat{\xi} \approx \xi, \tag{42}$$

when $\eta \to 0$ and $\xi \to 0$, respectively, ζ and $(v^2 - w^2)$ given by (40b) and (40c) simplify, and the constants A^2 and B can be found by comparison with the corresponding terms of the time differentiated exact solution (with $\omega_p = b$) given in Ref. 6.

The result is

$$A^2 = -4c^2, \qquad B = 0. \tag{43}$$

Substituting (41) and (43) into (40b) and (40c), and evaluating the integrals, one obtains (24′) for ζ, and

$$\frac{v^2 - w^2}{2} = \frac{h}{c^2}\left(\frac{e^{-(1-\beta)(ct+z)/2h}\sinh[(1-\beta)(ct+z)/2h]}{(1-\beta)} + \frac{e^{(1+\beta)(ct-z)/2h}\sinh[(1+\beta)(ct-z)/2h]}{(1+\beta)}\right); \tag{44}$$

when combined with (40), these relations furnish the exact solution of (39), with ω_p given by (18). The result may also be derived from (35) by noting that $\bar{G}$ is the negative of the derivative $\partial G/\partial t'$, evaluated at $t' = 0$ provided that the temporal source function is taken as $\delta(t - t')$ instead of $\delta(t)$. As noted previously, the limiting cases $\beta \to 0$, and $\beta, h \to \infty$ with β/h finite, reduce properly to results given in Refs. 4 and 3, respectively.

V. DISCUSSION

In this paper, the space–time ray theory developed previously for media with spatial or temporal variation has been extended to the case where medium parameters may vary (slowly) in space and time. For the particularly interesting case of space–time dependence in the form $(\mathbf{r} - \mathbf{V}t)$, descriptive of ordered motion of a disturbance in the medium, it has been shown that the ray trajectories are described by constancy of $(\omega - \mathbf{k} \cdot \mathbf{V})$, the physical consequences of which have been discussed. As in previous work on simpler problem categories, the ray method provides a physical interpretation of wave phenomena that is not as apparent from other treatments. The discussion has sought to clarify relevant propagation processes, in particular those relating to wavepackets with initial speeds greater or less than the profile speed V.

For further illustration of the role of space–time rays in the interpretation of transient solutions in dispersive media, we return to certain peculiarities exhibited by the exact solution for impulsive $[d\delta(t')/dt']$ excitation of a medium described in the moving (z', t') frame by $\omega_p' = b \exp(-z'/\gamma h)$ as in (32), with b, $h > 0$. The exact solution given in Ref. 4 [see also $\partial G'/\partial t'$ in (34)] diverges for observation times $t' \to \infty$. To understand this behavior properly, it is necessary to separate the effects of the medium from those due to the excitation. We seek an interpretation by examining first the space–time ray trajectories. Since $\omega_p' \to \infty$ as $z' \to -\infty$, rays proceeding initially to the left are eventually turned back; however, wavepackets with higher frequency penetrate deeper into the medium before being reflected and therefore emerge at later observation times (see Fig. 4). It can be shown that if adjacent rays are distinguished by a constant frequency increment $\Delta\omega'$ (it is recalled that $\omega' =$ constant on a ray), then the cross section of the reflected ray tubes decreases as t' increases, with z' fixed; in fact, as $t' \to \infty$, one has $\Delta z'/\Delta\omega' \to c\,\Delta t'/\Delta\omega' \to 0$ [see (47)]. From (8′), the field amplitude $A_0(z', t')$ varies as $\bar{A}_0(z_1', t_1')(\Delta z'/\Delta\omega')^{-1/2}$, where $\bar{A}_0(z_1', t_1') = A_0(z_1', t_1')(\Delta z_1'/\Delta\omega')^{1/2}$ incorporates the amplitude at a reference (initial) point (z_1', t_1') on the ray. Evidently, because of the shrinking ray tube cross section, $A_0(z', t') \to \infty$ at long observation times t' unless the spectral content of the source function as contained in $\bar{A}_0(z_1', t_1')$ overcomes this growth.

Details of the calculation illustrate the validity of these remarks. From (30) and (36), the ray trajectory in (z', t') space, given parametrically in terms of $\omega' > 0$, follows from (22) as

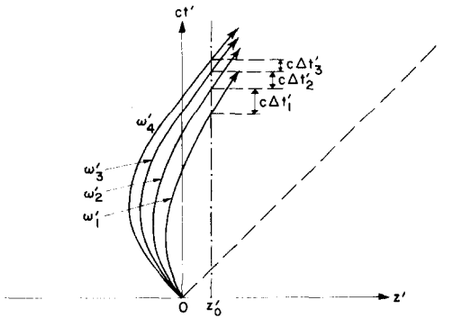


Fig. 4. Ray tubes in the moving reference frame. Cross sections in the plane $z' = z_0'$ are $c\Delta t_j'$, $j = 1, 2, 3$. The frequency increment $\Delta\omega'$ between adjacent rays is constant and $\omega_{i+1}' = \omega_i' + \Delta\omega' > \omega_i'$, $i = 1, 2, 3, 4$. $\omega_p(z')$ decreases monotonically with z'.

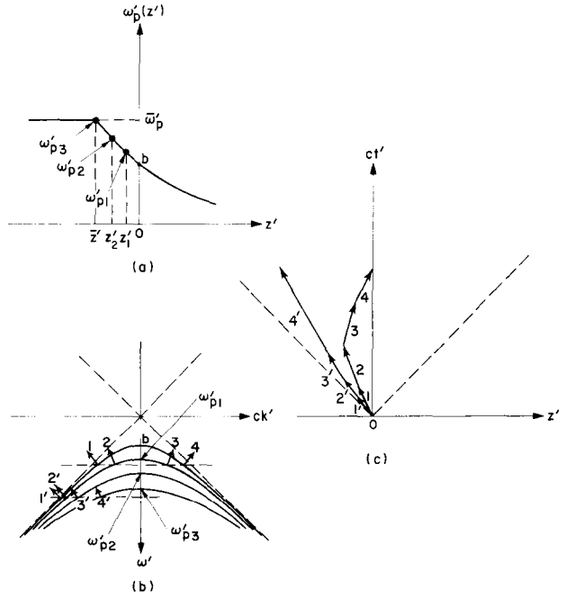


Fig. 5. Ray paths in a spatially inhomogeneous isotropic plasma whose plasma frequency ω_p' is a constant $\bar{\omega}_p' > 0$ in the region $z \leq \bar{z}'$ but decreases monotonically in the region $z' \geq \bar{z}'$, with $\omega_p' \to 0$ as $z' \to \infty$. (a) Profile of $\omega_p'(z')$. (b) Dispersion surfaces $\omega'(z', k') = [c^2k'^2 + \omega_p'^2(z')]^{1/2}$, with $(\omega_p')_{\max} = \bar{\omega}_p'$. (c) Space–time ray trajectories in piecewise constant approximation.

$$\sqrt{2}\,\omega' = \frac{b\,e^{-z/2\gamma h}\sinh(ct'/\gamma h)}{[\cosh(ct'/\gamma h) - \cosh(z'/\gamma h)]^{1/2}}. \tag{45}$$

Differentiating (45) with respect to z' for fixed t', and inverting the result, yields

$$\frac{dz'}{d\omega'} = \frac{2\gamma h}{\omega'}\left(\frac{\sinh(z'/\gamma h)}{[\cosh(ct'/\gamma h) - \cosh(z'/\gamma h)]} - 1\right)^{-1} \tag{46}$$

Evidently, $\omega' \to \infty$ on rays reaching a fixed observation point z' at very long observation times, and $|\Delta z'/\Delta\omega'| \to 0$ for such rays. From previous results dealing with pulsed plane wave propagation in a homogeneous plasma,[1] it is known that for a source function $\delta(t')$,

$$\overline{A}_0(z_1', t_1') = c^{3/2}e^{i\pi/4}\{2\sqrt{2\pi\omega'}[\omega'^2 - \omega_p'^2(z_1')]^{1/4}\}^{-1} \tag{47}$$

while for a source function $(d/dt')\delta(t')$,

$$\overline{A}_0(z_1', t_1') = (c^2/2\sqrt{2\pi})e^{-i\pi/4}. \tag{48}$$

The phase in these expressions is taken so as to correspond to the known result when $\omega_p \to b$. As noted previously, z_1' and t_1' are chosen sufficiently near the source to render the medium locally homogeneous ($\omega_p' \approx b$), but nevertheless large enough to validate use of the asymptotic expressions for the fields; both requirements can be met for sufficiently weak medium inhomogeneities. One may now observe that for the source associated with (47), the high-frequency components are excited so weakly as to nullify the growth of $|\Delta z'/\Delta\omega'|^{-1/2}$ in (46), but that this is not the case for excitation by a doublet as in (48). Thus, separate features of the transient field behavior, attributable to the source spectrum and the medium profile, respectively, are well clarified by the space–time ray approach.

The space–time ray analysis also makes evident that the above-described amplitude divergence due to ray tube contraction cannot occur in a physical medium whose plasma frequency grows up to a *finite* value $\omega_p'(\bar{z}')$ and is constant for $z' < \bar{z}'$. As noted from Fig. 5, not all rays are now turned back; wavepackets with sufficiently high frequency penetrate the medium to arbitrarily large distances z'. Although the cross sections of the reflected ray tubes decrease as t' increases, the family of reflected rays ceases after a certain value of t' and therefore does not exhibit the $\Delta z' \to 0$ limit noted earlier.

APPENDIX: EXACT SOLUTION OF THE SCALAR WAVE EQUATION (39) WITH $\omega_p(z, t)$ GIVEN BY (38).

Assume that

$$\overline{\overline{G}} = \frac{c^2}{2}W(\eta, \xi), \tag{A1}$$

where (η, ξ) are defined in (37). The left-hand side of (39) can then be expressed in terms of (η, ξ) as

$$\left[\frac{\partial^2}{\partial\eta\,\partial\xi} + \left(\frac{hb}{c}N(\eta)Z(\xi)\right)^2\right]W(\eta, \xi) = \frac{2h^2}{c^2}\delta(z)\frac{\partial}{\partial t}\delta(t). \tag{A2}$$

Moreover, we assume a causal solution of the form

$$W = \delta(ct - |z|) - W_0(w, v)U(ct - |z|), \tag{A3}$$

where

$$W_0(w, v) = \tfrac{1}{2}b(v^2 - w^2)J_1(b\sqrt{\zeta})/\sqrt{\zeta} \tag{A3'a}$$

with

$$\zeta = A^2(\tfrac{1}{2}w^2 + \hat{a})(\tfrac{1}{2}v^2 + \hat{a}), \quad w = w(\eta), \quad v = v(\xi), \tag{A3'b}$$

$A^2, \hat{a}$ being constants. With $\omega_p(z, t)$ given by (38), we seek solutions of (39) that reduce to the time derivative of the known Green's function G when ω_p is constant.[6] The form (A3) is chosen because of its similarity, after appropriate changes of variables, with known exact solutions to (39) when $\omega_p(z, t)$ is a constant, an exponential in space,[4] or an exponential in time.[3]

Substitution of (A3) into (A2) gives, for $z > 0$,

$$\alpha_1(\eta, \xi)U(\eta) + \alpha_2(\eta, \xi)\delta(\eta) = -\frac{2h^2}{c^2}\delta(z)\frac{\partial}{\partial t}\delta(t), \tag{A4}$$

where

$$\alpha_1 = \frac{\partial^2 W_0}{\partial\xi\,\partial\eta} + K^2W_0, \quad \alpha_2 = \frac{\partial W_0}{\partial\xi} - \frac{K^2}{2h}, \quad K^2 = \left(\frac{hbN(\eta)Z(\xi)}{c}\right)^2. \tag{A4'a}$$

From (A3a) and (A3b) and the identities

$$\frac{\partial^2\zeta}{\partial w\,\partial v} = \frac{1}{\zeta}\frac{\partial\zeta}{\partial w}\frac{\partial\zeta}{\partial v}, \quad v\frac{\partial\zeta}{\partial w} - w\frac{\partial\zeta}{\partial v} = \frac{v^2 - w^2}{2\zeta}\frac{\partial\zeta}{\partial w}\frac{\partial\zeta}{\partial v}, \tag{A4'b}$$

as well as $dJ_1(x)/dx = J_0(x) - x^{-1}J_1(x)$, one may write (A4a) as

$$\alpha_1 = \frac{b^3}{\sqrt{\zeta}}\frac{v^2 - w^2}{2}\left[-\frac{A^2}{4}\frac{d}{d\eta}\left(\frac{w^2}{2}\right)\frac{d}{d\xi}\left(\frac{v^2}{2}\right) + \left(\frac{hNZ}{c}\right)^2\right] \times J_1(b\sqrt{\zeta}) \tag{A4'c}$$

and

$$\alpha_2 = \frac{b^2}{2}\left\{\frac{1}{\sqrt{\zeta}}\frac{dv}{d\xi}\left[\frac{2v}{b}J_1(b\sqrt{\zeta}) + \frac{v^2 - w^2}{2\sqrt{\zeta}}\frac{\partial\zeta}{\partial v}\left(J_0(b\sqrt{\zeta}) - \frac{2J_1(b\sqrt{\zeta})}{b\sqrt{\zeta}}\right)\right] - h\left(\frac{NZ}{c}\right)^2\right\}. \tag{A4'd}$$

Equation (A4) stipulates that $\alpha_1 \equiv 0$ and that $\alpha_2 \to 0$ as $\eta \to 0$ when z or t is nonzero. The former condition implies that the bracketed term in (A4c) is zero, i.e.,

$$\frac{d}{d\eta}\left(\frac{w^2}{2}\right)\frac{d}{d\xi}\left(\frac{v^2}{2}\right) = \left(\frac{2hNZ}{cA}\right)^2. \tag{A5}$$

The latter condition, we assume, implies that $\zeta \to 0$ as $\eta \to 0$. Applying this condition to (A4d), recalling from (38) that $N(0) = 1$, and replacing Bessel functions by their small argument forms, we find the relation

$$\frac{d}{d\xi}\left(\frac{v^2}{2}\right) = h\left(\frac{Z}{c}\right)^2 \tag{A6}$$

in the limit as $\eta \to 0$. The assumption $\zeta \to 0$ as $\eta \to 0$ can be satisfied by [see (A3b)]

$$\tfrac{1}{2}w^2 + \hat{a} = 0 \tag{A7}$$

in the limit as $\eta \to 0$. Substitution of (A6) into (A5) yields

$$\frac{d}{d\eta}\left(\frac{w^2}{2}\right) = h\left(\frac{2N}{A}\right)^2. \tag{A8}$$

Integrating (A6) and (A8) with use of (A7), one obtains w^2 and v^2 as integrals over η and ξ, respectively, from which (40b) and (40c) follow with $B = v^2(0)/2 + \hat{a}$. If $\omega_p \to b$ in some limit then A^2, B are given as in (43).

The above scheme can also be used to find solutions of the three-dimensional scalar wave equation

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\omega_p^2(r,t)}{c^2}\right)G(\mathbf{r},t) = -\delta(\mathbf{r})\delta(t), \quad \text{(A9)}$$

where r is the radial coordinate from the origin, and $\omega_p(r,t)$ is given by (38), with (η, ξ) defined in (37) provided that z is replaced by r. The causal solution of (A9) can be shown to be

$$G(\mathbf{r},t) = (c/4\pi r)W(\eta,\xi), \quad \text{(A10)}$$

where $W(\eta,\xi)$ is given by (A3) with ζ and $(v^2 - w^2)$ expressed in terms of $N(\eta)$ and $Z(\xi)$ as in (40b) and (40c). The constants A^2 and B are $4c^2$ and zero, respectively, when $\omega_p \to b$ in some limit. Thus, for example, we can find the exact closed form solution for a radially symmetric exponentially varying plasma frequency profile with exponential time-dependence of the form

$$\omega_p(r,t) = b e^{\sigma_1(r-Vt)+\sigma_2(c^2t^2+r^2)}, \quad \text{(A11)}$$

where $\sigma_{1,2}$ and V are constants. When $\sigma_1 = 0$, $\omega_p(r,t)$ given by (A11) is Gaussian in both space and time.

* The research reported herein was sponsored by the Air Force Cambridge Research Laboratories, Office of Aerospace Research, under Contract F 19628-68-C-0072, but the paper does not necessarily reflect endorsement by the sponsor.

† Work initiated while author was a postdoctoral fellow with the Electrophysics Department, Polytechnic Institute of Brooklyn Farmingdale, New York.

1 R. M. Lewis, *Asymptotic Solutions of Differential Equations and Their Applications*, University of Wisconsin Symposium Proceedings (Wiley, New York, 1964); R. M. Lewis and N. Bleistein, "Asymptotic Solution of a Dispersive Hyperbolic Equation with Variable Coefficients," N.Y.U. Courant Institute of Mathematical Sciences, Research Report No. EMp-206 (1965).

2 L. B. Felsen, IEEE Trans. Antennas Propagation **17**, 191, (1969).

3 L. B. Felsen and G. M. Whitman, IEEE Trans. Antennas Propagation **18**, 242 (1970).

4 E. Ott and J. Shmoys, Can. J. Phys. **47**, 1435 (1969).

5 L. B. Felsen, SIAM Rev. **12**, 424 (1970).

6 B. Van der Pol and H. Bremmer, *Operational Calculus* (Cambridge U. P., Cambridge, 1955), Sec. 16. 3, p. 349.

7 C. M. Papas, *Theory of Electromagnetic Wave Propagation* (McGraw-Hill, New York, 1965), Sec. 7. 2.

A Bubble-Time Canonical Formalism for Geometrodynamics*

Karel Kuchař
Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540
(Received 19 October 1971; Revised Manuscript Received 21 January 1972)

A functional differential version of the ADM canonical formalism is proposed. The existence of a canonical transformation separating canonical variables into internal coordinates, energy-momentum densities, and two pairs of true dynamical variables is assumed. The evolution of dynamical variables is governed by functional differential Hamilton's equations. They satisfy certain integrability conditions ensuring the internal path independence of dynamical evolution. The change of dynamical variables along any spacelike hypersurface is given by their Lie derivatives. This allows an elimination of $3\infty^3$ components of the Hamilton equation, leading to a functional differential Hamilton equation based on a single bubble time. The Hamilton–Jacobi theory is built along the same lines. The formalism is illustrated in the mini-phase-space of the cylindrical Einstein–Rosen wave.

1. INTRODUCTION

The remarkable dynamical structure of Einstein's gravitational law has been recognized since the very beginning of the general theory of relativity. Later on, its connection with the canonical formalism was understood.[1] We know that Einstein's equations can be split into two sets, the dynamical equations and the initial value equations. The dynamical equations tell us how the canonically conjugate variables g_{ik} and π^{ik} (representing the intrinsic geometry of a spacelike hypersurface and its extrinsic curvature) change if we push the hypersurface through space–time. However, the canonical variables g_{ik} and π^{ik} cannot be freely specified on a given hypersurface, because they are limited by the initial value equations.

Both sets of equations are intimately interconnected. Firstly, if the initial value equations are satisfied on an initial hypersurface, and if we evolve the canonical variables in accordance with the dynamical equations, then the changed canonical variables satisfy the initial value equations on the changed hypersurface. Secondly, if the canonical variables satisfy the initial value equations on *any* spacelike hypersurface, then their change from one spacelike hypersurface to another is necessarily governed by the dynamical equations.[2] This indicates that the initial value equations contain all the dynamics of the gravitational field.

Einstein's equations are thus highly repetitive. They spell out the same dynamical evolution in two different ways; once by means of the dynamical equations, the second time through the initial value equations. The redundancy of equations is brought in by the redundancy of the canonical variables. The initial value equations imply that the canonical variables g_{ik} and π^{ik} are not independent, but that four of them can be expressed by means of the remaining ones. Moreover, four of these remaining variables can be frozen by picking up a definite slicing of space–time and a definite labeling of the slices, i.e., by imposing four coordinate conditions. This leaves us with only two pairs of true dynamical variables. The extra variables are the price we must pay for the freedom to choose the space–time coordinates at will.

This opens two alternative routes for the further development of the canonical formalism. The first one was followed by Dirac[3] and DeWitt,[4] who worked with the original redundant variables g_{ik}, π^{ik} subject to the

Integrating (A6) and (A8) with use of (A7), one obtains w^2 and v^2 as integrals over η and ξ, respectively, from which (40b) and (40c) follow with $B = v^2(0)/2 + \hat{a}$. If $\omega_p \to b$ in some limit then A^2, B are given as in (43).

The above scheme can also be used to find solutions of the three-dimensional scalar wave equation

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\omega_p^2(r,t)}{c^2}\right)G(\mathbf{r},t) = -\delta(\mathbf{r})\delta(t), \quad \text{(A9)}$$

where r is the radial coordinate from the origin, and $\omega_p(r,t)$ is given by (38), with (η, ξ) defined in (37) provided that z is replaced by r. The causal solution of (A9) can be shown to be

$$G(\mathbf{r},t) = (c/4\pi r)W(\eta,\xi), \quad \text{(A10)}$$

where $W(\eta,\xi)$ is given by (A3) with ζ and $(v^2 - w^2)$ expressed in terms of $N(\eta)$ and $Z(\xi)$ as in (40b) and (40c). The constants A^2 and B are $4c^2$ and zero, respectively, when $\omega_p \to b$ in some limit. Thus, for example, we can find the exact closed form solution for a radially symmetric exponentially varying plasma frequency profile with exponential time-dependence of the form

$$\omega_p(r,t) = b e^{\sigma_1(r-Vt)+\sigma_2(c^2t^2+r^2)}, \quad \text{(A11)}$$

where $\sigma_{1,2}$ and V are constants. When $\sigma_1 = 0$, $\omega_p(r,t)$ given by (A11) is Gaussian in both space and time.

* The research reported herein was sponsored by the Air Force Cambridge Research Laboratories, Office of Aerospace Research, under Contract F 19628-68-C-0072, but the paper does not necessarily reflect endorsement by the sponsor.

† Work initiated while author was a postdoctoral fellow with the Electrophysics Department, Polytechnic Institute of Brooklyn Farmingdale, New York.

1 R. M. Lewis, *Asymptotic Solutions of Differential Equations and Their Applications*, University of Wisconsin Symposium Proceedings (Wiley, New York, 1964); R. M. Lewis and N. Bleistein, "Asymptotic Solution of a Dispersive Hyperbolic Equation with Variable Coefficients," N.Y.U. Courant Institute of Mathematical Sciences, Research Report No. EMp-206 (1965).

2 L. B. Felsen, IEEE Trans. Antennas Propagation **17**, 191, (1969).

3 L. B. Felsen and G. M. Whitman, IEEE Trans. Antennas Propagation **18**, 242 (1970).

4 E. Ott and J. Shmoys, Can. J. Phys. **47**, 1435 (1969).

5 L. B. Felsen, SIAM Rev. **12**, 424 (1970).

6 B. Van der Pol and H. Bremmer, *Operational Calculus* (Cambridge U. P., Cambridge, 1955), Sec. 16. 3, p. 349.

7 C. M. Papas, *Theory of Electromagnetic Wave Propagation* (McGraw-Hill, New York, 1965), Sec. 7. 2.

A Bubble-Time Canonical Formalism for Geometrodynamics*

Karel Kuchař
Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540
(Received 19 October 1971; Revised Manuscript Received 21 January 1972)

A functional differential version of the ADM canonical formalism is proposed. The existence of a canonical transformation separating canonical variables into internal coordinates, energy-momentum densities, and two pairs of true dynamical variables is assumed. The evolution of dynamical variables is governed by functional differential Hamilton's equations. They satisfy certain integrability conditions ensuring the internal path independence of dynamical evolution. The change of dynamical variables along any spacelike hypersurface is given by their Lie derivatives. This allows an elimination of $3\infty^3$ components of the Hamilton equation, leading to a functional differential Hamilton equation based on a single bubble time. The Hamilton–Jacobi theory is built along the same lines. The formalism is illustrated in the mini-phase-space of the cylindrical Einstein–Rosen wave.

1. INTRODUCTION

The remarkable dynamical structure of Einstein's gravitational law has been recognized since the very beginning of the general theory of relativity. Later on, its connection with the canonical formalism was understood.[1] We know that Einstein's equations can be split into two sets, the dynamical equations and the initial value equations. The dynamical equations tell us how the canonically conjugate variables g_{ik} and π^{ik} (representing the intrinsic geometry of a spacelike hypersurface and its extrinsic curvature) change if we push the hypersurface through space–time. However, the canonical variables g_{ik} and π^{ik} cannot be freely specified on a given hypersurface, because they are limited by the initial value equations.

Both sets of equations are intimately interconnected. Firstly, if the initial value equations are satisfied on an initial hypersurface, and if we evolve the canonical variables in accordance with the dynamical equations, then the changed canonical variables satisfy the initial value equations on the changed hypersurface. Secondly, if the canonical variables satisfy the initial value equations on *any* spacelike hypersurface, then their change from one spacelike hypersurface to another is necessarily governed by the dynamical equations.[2] This indicates that the initial value equations contain all the dynamics of the gravitational field.

Einstein's equations are thus highly repetitive. They spell out the same dynamical evolution in two different ways; once by means of the dynamical equations, the second time through the initial value equations. The redundancy of equations is brought in by the redundancy of the canonical variables. The initial value equations imply that the canonical variables g_{ik} and π^{ik} are not independent, but that four of them can be expressed by means of the remaining ones. Moreover, four of these remaining variables can be frozen by picking up a definite slicing of space–time and a definite labeling of the slices, i.e., by imposing four coordinate conditions. This leaves us with only two pairs of true dynamical variables. The extra variables are the price we must pay for the freedom to choose the space–time coordinates at will.

This opens two alternative routes for the further development of the canonical formalism. The first one was followed by Dirac[3] and DeWitt,[4] who worked with the original redundant variables g_{ik}, π^{ik} subject to the

initial value constraints. The second one was pursued by Arnowitt, Deser, and Misner (ADM),[1] who tried to eliminate all superfluous variables and identify the independent dynamical degrees of freedom of the gravitational field. Either of these approaches has certain disadvantages. In the Dirac formalism, true dynamical variables are mixed with variables specifying the hypersurface and its coordinatization. This obscures the physical interpretation of the formalism, especially in its quantum version. In the ADM method, one system of coordinates is picked out by coordinate conditions, which obliterate the arbitrariness of slicing.

To avoid these difficulties, we shall build an intermediate formalism lying somewhere between the Dirac formalism and the ADM formalism. The new formalism resembles the ADM approach in so far as it identifies the true dynamical variables and solves the constraint equations; but it shares a common feature with the Dirac approach that it imposes no coordinate conditions. In fact, it can be viewed in two different ways, either as the Dirac formalism carried into a new representation by a canonical transformation, or as an infinite number of ADM formalisms, one for each choice of a coordinate condition, stacked together in one compact notation. Its main advantages over the Dirac and the ADM formalisms are, respectively, that a clear separation of true dynamical variables from the superfluous degrees of freedom is maintained, while the slicing of space–time is kept completely arbitrary. Many results and ideas fitting into this approach are already contained or anticipated in the classical series of papers by ADM. It is more an emphasis than a content which is being added to their penetrating analysis. We want to stress that the dynamical structure of Einstein's law is best revealed when the slicing of space–time is left arbitrary and all equations are written down as functional differential equations with respect to a corresponding bubble time. The proposed formalism provides a natural language for the discussion of questions belonging properly to the no man's land between the Dirac and ADM formalisms. Here are a few typical examples: Does the ADM formalism depend on choice of coordinate conditions? Does the evolution of state depend on path? How many equations comprised in the functional differential Einstein–Hamilton–Jacobi equation are actually independent? What is the physical meaning of the integrability conditions of this equation? The answers to such questions will emerge as we proceed.

The first step in building the new formalism is the separation of the twelve canonical variables g_{ik}, π^{ik} into three groups, which we shall denote by the letters X, Π, and Y. We have already seen that not all canonical variables g_{ik}, π^{ik} are true dynamical variables, but that four of them are actually functions $X^\iota(\mathbf{x})$ fixing the position of the hypersurface and endowing it with a special system of spatial coordinates (Fig. 1a).[5] We shall call these functions *internal coordinates*, because they are constructed from the geometrical properties g_{ik}, π^{ik} of the hypersurface, not externally assigned to its points, like the coordinate labels t and $\mathbf{x}$. The four functions X^ι of three labels give us the parameter equations of the hypersurface

$$X^\iota = X^\iota(\mathbf{x}). \tag{1.1}$$

We can change the labels $\mathbf{x}$ arbitrarily without changing the hypersurface, the functions (1.1) and the new functions

$$X^\iota = X^\iota(\mathbf{x}(\bar{\mathbf{x}})) \tag{1.2}$$

representing the same hypersurface. We can eliminate the labels altogether and write the equation of the hypersurface directly as

$$X^0 = X^0(\mathbf{x}(\mathbf{X})) \equiv T(\mathbf{X}). \tag{1.3}$$

In other words, Eq. (1.1) goes over into Eq. (1.3) if we use the three internal coordinates X^k themselves as privileged labels. If the functions (1.1) depend smoothly on an additional parameter t,

$$X^\iota = X^\iota(t, \mathbf{x}), \tag{1.4}$$

they define an internal trajectory, i.e., a continuous family of spacelike hypersurfaces ordered by a time parameter t, drawn in the space of internal coordinates.

The function (1.3) fixes the position of a spacelike hypersurface in space–time, playing the role of Old Father Time of classical physics. The "instantaneous" initial value equations contain this time function hidden among other canonical variables as Achilles amidst the daughters of King Lykamedes. It is then no wonder that these equations carry a dynamical meaning.

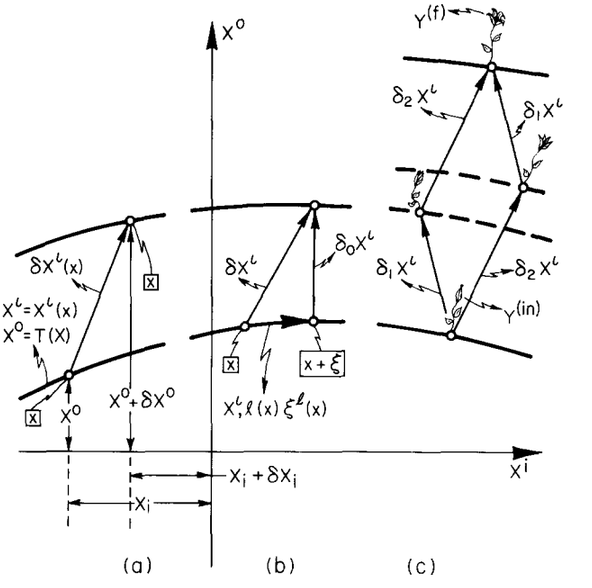


Fig. 1. (a) *Displacement of a hypersurface.* Internal coordinates X^ι are plotted as cartesian coordinates. Arbitrary labels $\mathbf{x}$ are attached to the points of the hypersurface. The hypersurface can be characterized either by the parameter equations $X^\iota = X^\iota(\mathbf{x})$ or internally as $X^0 = T(\mathbf{X})$. The displacement $\delta X^\iota(\mathbf{x})$ connects points with the same label $\mathbf{x}$ on two nearby hypersurfaces. (b) *Reducibility of the Hamilton equations.* An arbitrary displacement δX^ι can be decomposed into a displacement $X^\iota{}_{,\ell}(\mathbf{x})\xi^\ell(\mathbf{x})$ along the hypersurface, and the displacement $\delta_0 X^\iota$ in the direction of the X^0 coordinate lines. The change of the dynamical variables under the displacement $X^\iota{}_{,\ell}(\mathbf{x})\xi^\ell(\mathbf{x})$ can be generated by a Lie derivative. (c) *Path independence of dynamical evolution.* Starting from an initial hypersurface, we pass to the final hypersurface along two different paths $\delta_1 X^\iota, \delta_2 X^\iota$ and $\delta_2 X^\iota, \delta_1 X^\iota$. The initial values $Y^{(\mathrm{in})}$ of the dynamical variables (symbolized by flowers) evolve into the same final values $Y^{(f)}$, irrespective of the path taken.

The split of twelve canonical coordinates g_{ik}, π^{ik} into three groups of four functions was originally introduced by ADM who stressed that the functions X^ι are not dynamically determined and may be used as independent variables in place of coordinates.[6] Developing their formalism, however, ADM chose an alternative route and froze X^ι into preferred time and space coordinates by coordinate conditions. What follows may be thought of as an elaboration of the original ADM idea. We impose no coordinate conditions on the internal coordinates, leaving the slicing of space–time completely arbitrary. The time variable is allowed to change by one amount at one point and by another amount at another point, creating a bubble deformation of the original hypersurface. This distinguishes the new formalism from the latter development of the ADM formalism, in which a one-parameter family of slices is picked out by coordinate conditions and the permissible variations go only from one slice of the family to another. In the Dirac formalism, all spacelike hypersurfaces are equally permissible; however, we do not know which variable plays the role of time. This explains the name we shall use to describe the new formalism: the bubble-time canonical formalism.

The second group of canonical variables Π_ι contains the variables canonically conjugate to internal coordinates X^ι. The variable $-\Pi_0(\mathbf{x})$ can be interpreted as a gravitational energy density and the three variables $-\Pi_i(\mathbf{x})$ as gravitational momentum densities. Finally, the third group of canonical variables $Y_\iota(\mathbf{x})$ consists of two pairs of canonically conjugate variables.

$$Y_\iota(\mathbf{x}) = \{g^A(\mathbf{x}), \pi_A(\mathbf{x})\}, \quad A = 1, 2,$$

representing the true dynamical degrees of freedom of the gravitational field.

Unfortunately, no general agreement has been reached as to how to separate the internal coordinates X^ι and the conjugate energy-momentum densities Π_ι from the true dynamical variables Y_ι. It is difficult to find an Ulyssean trick luring Achilles out of his disguise. The first attempt was made by Arnowitt, Deser, and Misner[1] who decomposed the canonical variables g_{ik} and π^{ik} into transversal traceless TT, transversal T, and longitudinal L parts with respect to a fictitious flat space background, and identified the individual variables as

$$\begin{aligned} &X^0 = -\tfrac{1}{2}\Delta^{-1}\pi^T, \quad X^i = g^L{}_i, \\ &\Pi_0 = \Delta g^T, \quad \Pi_i = 2(\Delta\pi^{Li} + \pi^{Lk}{}_{,ki}), \\ &g^A \Leftrightarrow g^{TT}{}_{ik}, \quad \pi_A \leftrightarrow \pi^{TTik}. \end{aligned} \tag{1.5}$$

However, this decomposition clearly depends on the choice of the background and does not constitute an invariant procedure. Some time ago, a covariant decomposition was tried by Deser.[7] Quite recently, York attempted to identify two of the true dynamical degrees of freedom with the independent components of the conformal curvature tensor of the spacelike hypersurface which is an identically TT symmetric tensor in the covariant sense.[8] However, it is difficult to write down a canonical transformation leading to Deser's and York's variables, and it is therefore not yet certain how convenient they are from the point of view of the canonical formalism.

A fresh insight into the nature of internal coordinates was gained by studying the dynamics of truncated models with most of the gravitational degrees of freedom frozen by imposed symmetries.[9-11] It would be premature to make any definite conclusion about the construction of the internal coordinates on the basis of a few simple models studied so far. In fact, different constructions have been attempted. In homogeneous cosmological models,[9] the internal spatial coordinates are frozen and it seems advantageous to identify time with a simple function of the metric (the logarithm of the volume of the universe). For cylindrical gravitational waves, on the other hand, an internal radial coordinate is a function of the metric, and a natural time variable is a certain functional of the momentum.[10] Virtually the same prescription for the internal radial coordinate and time works also in the case of the Schwarzschild field.[11] In the last two examples, as well as in the original ADM procedure,[1] an internal time is constructed from the momenta π^{ik}, i.e., from the extrinsic curvature of the hypersurface. The term "extrinsic time" seems therefore appropriate for it. On the other hand, we have an "intrinsic time" for cosmological models, constructed out of the intrinsic metric of the hypersurface.

So far it seems that quite different choices of internal time and space coordinates simplify the Hamilton equations for different particular models. It might even happen that different separations of internal coordinates from dynamical variables work equally well in a single model. This would mean that it is merely a matter of taste which degrees of freedom are taken to provide the space and time structure with respect to which the change of the remaining degrees of freedom is registered. Passing from the truncated models to general geometrodynamics, our belief in a unique and natural separation of internal coordinates and energy-momentum densities from true dynamical variables seems even more open to doubt. It may be that there is only a multitude of various separations dictated by expediency when solving particular problems.

Under these circumstances, it may seem impossible to develop a bubble-time formalism. How can we proceed, if we do not know how to construct the internal time and other canonical variables we need? This essential weakness, which the new formalism shares with the ADM method, turns into a strength when viewed from a different angle. Let us assume that we are presented with *some* set of new canonical variables X, Π, Y (subject to very mild restrictions) constructed from the old canonical variables g_{ik}, π^{ik} by a canonical transformation. We can then build a bubble-time formalism and show it has a number of interesting features, without ever knowing the actual form of the canonical transformation. Our results are thus completely independent of the actual construction of the canonical variables X, Π, Y.

We develop the bubble-time formalism in several steps. We start in Sec. 2, by reviewing the ADM canonical formalism and introducing a compact notation. In Sec. 3, we separate the canonical variables into groups and write down the Hamilton equations in the new representation. In Section 4, we assume that the initial value equations can be solved with respect to the quantities $\Pi_\iota(\mathbf{x})$. This gives us the energy-momentum densities as functions of the true dynamical

variables $Y_\iota(\mathbf{x})$ depending on the internal space and time coordinates $X^\iota(\mathbf{x})$. These functionals yield the bubble-time Hamilton equations which govern the *internal dynamics* of the gravitational field, i.e., how the dynamical variables $Y_\iota(\mathbf{x})$ evolve as functionals of the internal coordinates $X^\iota(\mathbf{x})$.

This dynamics is to be distinguished from an *external dynamics* which people usually have in mind when thinking about the time evolution of the gravitational field. Here, a one-parameter family of spacelike hypersurfaces labeled by an arbitrary parameter t is picked out. The external stacking of the hypersurfaces is given by prescribing arbitrarily the lapse and shift functions $N(t, \mathbf{x}), N_i(t, \mathbf{x})$. The original canonical variables $g_{ik}(\mathbf{x}), \pi^{ik}(\mathbf{x})$ are chosen at a fixed moment $t^{(\mathrm{in})}$ in such a way that they satisfy the initial value equations, and their change in t is determined. No split into internal coordinates X^ι, energy–momentum densities Π_ι, and dynamical variables Y_ι is ever attempted. However, we prove in Sec. 4 that the internal dynamics induces an external dynamics of the field. In particular, to every internal path (1.4) there corresponds a unique "external path" $N(t, \mathbf{x}), N_i(t, \mathbf{x})$. This is a new version of the sandwich conjecture within the bubble-time formalism.

The bubble-time Hamilton equations must satisfy two consistency conditions which are discussed in Secs. 5 and 6. The first condition tells us that the evolution of the dynamical variables must not depend on the internal path between the fixed initial and final hypersurfaces. This condition is satisfied by virtue of the completeness relations among the initial value constraints. The second condition requires that the change of the dynamical variables predicted by the bubble-time Hamilton equations along a spacelike hypersurface should coincide with a Lie derivative within this hypersurface. Due to this condition, the original Hamilton equation which has $4\infty^3$ components can be reduced to a single bubble-time Hamilton equation which has only ∞^3 components. This is done in Sec. 7. The Hamilton–Jacobi theory in the bubble-time canonical formalism is discussed in Sec. 8. The consistency conditions again allow a reduction of the Hamilton–Jacobi equation to a single functional differential equation for a principal Hamilton functional depending on a single bubble-time variable and on two dynamical variables. In Sec. 9, we show how the new formalism works by applying it to a simple mini-phase-space model, for which the canonical transformation separating internal coordinates from true dynamical variables is explicitly known. This model is given by the Einstein–Rosen cylindrical gravitational waves.

2. STANDARD CANONICAL FORMALISM

The bubble-time canonical formalism can be developed from the standard action functional of the gravitational field brought into a canonical form by Arnowitt, Deser, and Misner[1]:

$$S = \int dt \int d^3x \, (\pi^{ik}(\mathbf{x}) g_{ik}(\mathbf{x}) - N_\iota(\mathbf{x}) \mathcal{H}^\iota(\mathbf{x})). \tag{2.1}$$

All variables entering into this action functional have a simple geometrical meaning.[1,2]

The canonical coordinates g_{ik} give us the metric of a spacelike hypersurface $t = \text{const}$. The momenta π^{ik} are related to the extrinsic curvature K^{ik} of the hypersurface by the formula

$$\pi^{ik} = -g^{1/2}(K^{ik} - K g^{ik}), \tag{2.2}$$

and the coefficients $N_\iota = \{N, N_i\}$ tell us how the initial hypersurface of a constant t is related to a displaced hypersurface of a constant $t + \delta t$. The lapse function $N(\mathbf{x})$ gives the proper time separation $\delta\tau(\mathbf{x})$ between these two hypersurfaces measured in the normal direction to the first hypersurface

$$\delta\tau(\mathbf{x}) = N(\mathbf{x}) \delta t.$$

The shift functions $N_i(\mathbf{x})$ determine how the spatial system of coordinates on the hypersurface of a constant $t + \delta t$ is displaced with respect to the spatial system of coordinates on the hypersurface of a constant t. If the normal to the first hypersurface drawn at a point with the coordinates $\mathbf{x}$ intersects the second hypersurface at a point with the coordinates $\mathbf{x} + \delta\mathbf{x}$, then

$$\delta x^i(\mathbf{x}) = -N^i(\mathbf{x}) \delta t, \qquad \text{where } N^i = g^{ik} N_k.$$

The components of $\mathcal{H}^\iota = \{\mathcal{H}, \mathcal{H}^i\}$ are constructed from the canonical variables g_{ik}, π^{ik},

$$\mathcal{H}(\mathbf{x}) = g^{-1/2}(\mathbf{x}) \{ \pi_{ik}(\mathbf{x}) \pi^{ik}(\mathbf{x}) - \tfrac{1}{2} (\pi^i_i(\mathbf{x}))^2 \} - g^{1/2}(\mathbf{x}) R(\mathbf{x}), \tag{2.3}$$

$$\mathcal{H}^i(\mathbf{x}) = -2\pi^{ik}{}_{|k}(\mathbf{x}), \tag{2.4}$$

and they are called super-Hamiltonian and supermomentum of the gravitational field. The Hamiltonian

$$H = \int d^3x \, N_\iota(\mathbf{x}) \mathcal{H}^\iota(\mathbf{x}) \tag{2.5}$$

is their linear combination, with lapse and shift functions as undetermined coefficients.

If we vary the action function (2.1) with respect to the canonical variables, we get the dynamical equations

$$\dot{g}_{ik}(\mathbf{x}) = \frac{\delta H}{\delta \pi^{ik}(\mathbf{x})}, \qquad \dot{\pi}^{ik}(\mathbf{x}) = -\frac{\delta H}{\delta g_{ik}(\mathbf{x})}. \tag{2.6}$$

If we vary it with respect to the variables N_ι which play the role of the Lagrange multipliers, we obtain the initial value equations

$$\mathcal{H}^\iota = 0. \tag{2.7}$$

To facilitate the further handling of complicated expressions, it is advisable to introduce a condensed notation. We shall use all types of indices (Greek indices, lower case and capital Latin indices) to represent at the same time the coordinate labels $\mathbf{x}$, different indices, like $\iota, \kappa, \ldots$, representing different labels, like $\mathbf{x}, \mathbf{x}', \ldots$, and the Einstein summation convention over a repeated index implying the integration over the continuous variables $\mathbf{x}$. With this understanding, the Hamiltonian (2.5) can be written down simply as $H = N_\iota \mathcal{H}^\iota$. Sometimes we wish to overrule this convention, in which case we write down the dependence on $\mathbf{x}$ explicitly, as in Eqs. (2.3) and (2.4), where no integration over $\mathbf{x}$ is assumed.

The initial value equations (2.7) are the only constraints on the canonical variables g_{ik}, π^{ik} on a spacelike hypersurface. The Poisson brackets between

these constraints do not lead to new constraints, because they are expressible by the original constraints themselves:

$$[\mathcal{H}^\iota, \mathcal{H}^\kappa] = C^{\iota\kappa}{}_\lambda \mathcal{H}^\lambda. \tag{2.8}$$

Here, the compact notation is used for the first time. The coefficients $C^{\iota\kappa}{}_\lambda$ are certain combinations of the Kronecker deltas, the delta-functions $\delta(\mathbf{x}, \mathbf{x}')$, and their first partial derivatives.[13] The completeness relations (2.8) are essential for the consistency of the bubble-time formalism—they ensure the path independence of dynamical evolution.

In addition to the Poisson brackets (2.8), we need to know the Poisson brackets of the basic canonical variables g_{ik}, π^{ik} with the supermomentum $\mathcal{H}^i$. Because $\mathcal{H}^i\xi_i$ generates infinitesimal coordinate transformations $x^i \to \bar{x}^i = x^i + \xi^i$, its Poisson brackets with g_{ik}, π^{ik} yield the Lie derivatives of these variables:[4]

$$\begin{aligned}
[g_{ik}(\mathbf{x}), \mathcal{H}^l\xi_l] &= \pounds_\xi\, g_{ik}(\mathbf{x}) \\
&\equiv g_{ik,l}(\mathbf{x})\xi^l(\mathbf{x}) + g_{kl}(\mathbf{x})\xi^l{}_{,i}(\mathbf{x}) + g_{il}(\mathbf{x})\xi^l{}_{,k}(\mathbf{x}), \\
[\pi^{ik}(\mathbf{x}), \mathcal{H}^l\xi_l] &= \pounds_\xi\, \pi^{ik}(\mathbf{x}) \\
&\equiv (\pi^{ik}(\mathbf{x})\xi^l(\mathbf{x}))_{,l} - \pi^{il}(\mathbf{x})\xi^k{}_{,l}(\mathbf{x}) - \pi^{lk}(\mathbf{x})\xi^i{}_{,l}(\mathbf{x}).
\end{aligned} \tag{2.9}$$

The relations (2.9) hold for an arbitrary vector ξ^l.

3. SEPARATION OF VARIABLES

Following the general ideas introduced in Sec. 1, we assume that there exists a canonical transformation

$$g_{ik}, \pi^{ik} \to X^\iota, \Pi_\iota, Y_\iota \equiv \{g^A, \pi_A\} \tag{3.1}$$

leading from the old canonical variables g_{ik}, π^{ik}, which are treated all on the same footing, to the new canonical variables X^ι, Π_ι, Y_ι, to which completely different roles, namely those of internal coordinates, energy-momentum densities, and true dynamical variables are ascribed.

The new canonical variables are functionals of the old canonical variables, depending on coordinate labels $\mathbf{x}$ as parameters. These functionals are in general nonlocal: to know (for example) $X^0(\mathbf{x})$ at the point $\mathbf{x}$, we need to know $g_{ik}(\mathbf{x})$, $\pi^{ik}(\mathbf{x})$ for all values of $\mathbf{x}$, i.e., on the whole hypersurface. The canonical transformation (3.1) may mix superspace with the geometrodynamical phase space, so that the new configuration space X^ι, g^A is no longer identical with superspace. The canonical transformation (1.5) given by ADM neatly illustrates both these points; it is nonlocal, and it performs the mixing.

The canonical transformation (3.1) should be given by an invariant prescription independent of the labeling $\mathbf{x}$ of the hypersurface. The functionals $\bar{X}^\iota(\bar{\mathbf{x}})[\bar{g}_{ik}, \bar{\pi}^{ik}]$ should be therefore constructed from the transformed functions $\bar{g}_{ik}(\bar{\mathbf{x}})$, $\bar{\pi}^{ik}(\bar{\mathbf{x}})$ in the same way as the functionals $X^\iota(\mathbf{x})[g_{ik}, \pi^{ik}]$ are constructed from the original functions $g_{ik}(\mathbf{x})$, $\pi^{ik}(\mathbf{x})$, the same statement being applicable to the variables Π_ι and Y_ι.[14] Moreover, to maintain the interpretation of the quantities $X^\iota(\mathbf{x})$ as geometrically privileged time and space coordinates of a fixed point on the hypersurface, we must require that not only their *form*, but also their actual *values* do not change under the relabeling. They should therefore be scalars with respect to spatial transformations. Similarly, to maintain the interpretation of the quantities $\Pi_\iota(\mathbf{x})$ as energy and momenta densities per unit coordinate cell, they must transform as scalar densities under the spatial transformations. Continuing in the same spirit, we require that the dynamical coordinates $g^A(\mathbf{x})$ be scalars and the conjugate momenta $\pi_A(\mathbf{x})$ be scalar densities.

The transformation (1.5) fails to meet this double requirement of a form and value invariance. Start in a system of coordinates $\mathbf{x}$ and decompose the variables $g_{ik}(\mathbf{x})$, $\pi^{ik}(\mathbf{x})$ into TT, T, and L parts according to the ADM flat-space-background procedure. Define the canonical variables X^ι, Π_ι, Y_ι by Eqs. (1.5). Pass to the new system of coordinates $\bar{\mathbf{x}}$ by a coordinate transformation and decompose the transformed variables $\bar{g}_{ik}(\bar{\mathbf{x}})$ and $\bar{\pi}^{ik}(\bar{\mathbf{x}})$ according to the same flat-space-background procedure. Define the new canonical variables $\bar{X}^\iota$, $\bar{\Pi}_\iota$, $\bar{Y}_\iota$ from the TT, T, and L parts of $\bar{g}_{ik}$ and $\bar{\pi}^{ik}$ by the same equation (1.5). The new variables $\bar{X}^\iota$, $\bar{\Pi}_\iota$, $\bar{Y}_\iota$ depend on the old variables X^ι, Π_ι, Y_ι nonlocally in a complicated way, different classes of old variables being mixed together to yield the new variables. The variables X^ι, Π_ι, Y_ι are thus far away from being scalars, scalar densities, or other simple geometrical objects. ADM themselves did not run into this trouble, because they assumed that the TT, T, L decomposition should not be performed on an arbitrary slice and in an arbitrary system of coordinates $\mathbf{x}$, but only on such slices and in such systems of coordinates which are picked out by coordinate conditions

$$\pi^{ii}{}_{,jj} - \pi^{ij}{}_{,ij} = 0, \qquad g_{ij,j} = 0$$

equivalent to choosing X^0 and X^i themselves as preferred t and x^i labels:

$$t = -\tfrac{1}{2}\Delta^{-1}\pi^T, \qquad x^i = g^L{}_i.$$

They were thus not constrained by the requirement that the new variables X^ι, Π_ι, Y_ι be given by form-invariant functionals of g_{ik}, π^{ik} and of nothing else on any slice in any coordinate system $\mathbf{x}$.

A good example of canonical transformation leading to variables with correct transformation properties is provided by the canonical transformation in mini-phase-space of the cylindrical gravitational wave, leading to the Einstein–Rosen time T and cylindrical radius R as internal coordinates X^ι, to the C-energy density and C-momentum density as the variables Π_ι, and to the wave amplitude ψ and the conjugate momentum π_ψ as dynamical variables. We shall investigate this transformation in the final section of this paper.

The canonical transformation (3.1) is thus limited by two requirements: it must be form-invariant, and the new canonical variables must behave as scalars or pseudoscalars with respect to the relabeling of the hypersurface. To see what it implies, let us fix our attention on the functionals Y_ι. The requirement of their form-invariance means that

$$Y_\iota(\mathbf{x})[g_{ik}, \pi^{ik}] = Y_\iota(\mathbf{x})[g_{ik}, \pi^{ik}]. \tag{3.2}$$

On the other hand, the transformation properties of

the dynamical variables Y_ι can be expressed in terms of their Lie derivatives

$$\bar{Y}_\iota(\mathbf{x})[\bar{g}_{ik},\bar{\pi}^{ik}] = Y_\iota(\mathbf{x})[g_{ik},\pi^{ik}] + \pounds_\xi Y_\iota(\mathbf{x})[g_{ik},\pi^{ik}]. \quad (3.3)$$

However, $\bar{g}_{ik}$ differs from g_{ik} and $\bar{\pi}_{ik}$ differs from π^{ik} also by an appropriate Lie derivative

$$g_{ik}(\mathbf{x}) = \bar{g}_{ik}(\mathbf{x}) - \pounds_\xi g_{ik}(\mathbf{x}), \qquad \pi^{ik}(\mathbf{x}) = \bar{\pi}^{ik}(\mathbf{x}) - \pounds_\xi \pi^{ik}(\mathbf{x}). \quad (3.4)$$

Putting Eqs. (3.2)–(3.4) together, we get the fundamental relation

$$\pounds_\xi Y_\iota = \frac{\delta Y_\iota}{\delta g^{ik}} \pounds_\xi g_{ik} + \frac{\delta Y_\iota}{\delta \pi^{ik}} \pounds_\xi \pi^{ik}. \quad (3.5)$$

Similar relations hold for the variables X^ι and Π_ι. In general, any functional $Z(\mathbf{x})[z_a]$ of any number of function variables $z_a(\mathbf{x})$ which is defined by an invariant prescription independent of the labeling $\mathbf{x}$ satisfies the equation

$$\pounds_\xi Z = \frac{\delta Z}{\delta z_a} \pounds_\xi z_a. \quad (3.6)$$

The canonical transformation (3.1) brings the action functional (2.1) into the form

$$S = \int dt\, (\Pi_\iota \dot{X}^\iota + \pi_A \dot{g}^A - N_\iota \mathcal{H}^\iota[X^\kappa, \Pi_\kappa, Y_\kappa]). \quad (3.7)$$

The super-Hamiltonian and supermomentum $\mathcal{H}^\iota$ become functionals of the new canonical variables, depending on the label $\mathbf{x}$ as a parameter. In special cases it may happen that $\mathcal{H}^\iota$ are local functionals of the new canonical variables, i.e., they are constructed out of the variables $X^\kappa(\mathbf{x})$, $\Pi_\kappa(\mathbf{x})$, $Y_\kappa(\mathbf{x})$ and a finite number of their derivatives taken in the same point $\mathbf{x}$, even if the new canonical variables themselves are nonlocal functionals of the old canonical variables. In general, however, the functionals $\mathcal{H}^\iota$ are nonlocal functionals of the new canonical variables.

If we vary the action functional (3.7) with respect to the lapse and shift functions N_ι, we get the initial value equations (2.7), and if we vary it with respect to the new canonical variables, we get the dynamical equations (2.6) in the new representation,

$$\dot{X}^\iota = N_\kappa \frac{\delta \mathcal{H}^\kappa}{\delta \Pi_\iota}, \quad (3.8)$$

$$\dot{\Pi}_\iota = -N_\kappa \frac{\delta \mathcal{H}^\kappa}{\delta X^\iota}, \quad (3.9)$$

$$\dot{g}^A = N_\kappa \frac{\delta \mathcal{H}^\kappa}{\delta \pi_A}, \quad (3.10)$$

$$\dot{\pi}_A = -N_\kappa \frac{\delta \mathcal{H}^\kappa}{\delta g^A}. \quad (3.11)$$

In the Hamilton equations (3.8)–(3.11), the internal coordinates are already separated from the rest of the canonical variables; but the path is still specified externally by the lapse and shift functions. The internal path is obtained from the external path by solving the Hamilton equation (3.8). The Hamilton equations (3.8), (3.9), and (3.11) were written down in this functional differential form for the first time by ADM.[15]

4. BUBBLE-TIME HAMILTON EQUATION

Let us now assume that the initial value equations (2.7) can be solved with respect to the energy-momentum variables Π_κ. We write the solution in the form

$$\Pi_\kappa(\mathbf{x}) = -P_\kappa(\mathbf{x})[X^\lambda, Y_\lambda], \quad (4.1)$$

where P_κ are functionals of the function variables $X^\lambda(\mathbf{x})$, $Y_\lambda(\mathbf{x})$, depending on $\mathbf{x}$ as parameters. Again, these functionals are in general nonlocal, except for special cases like the cylindrical wave. Because (4.1) is a solution of the equations $\mathcal{H}^\iota(\mathbf{x})[X^\kappa, \Pi_\kappa, Y_\kappa] = 0$, the composite functionals

$$\mathcal{H}^\iota(\mathbf{x})[X^\kappa, -P_\kappa[X^\lambda, Y_\lambda], Y_\kappa] \quad (4.2)$$

must be identically equal to zero for arbitrary function variables $X^\kappa(\mathbf{x})$, $Y_\kappa(\mathbf{x})$.

The variables $P_\kappa(\mathbf{x})[X^\lambda, Y_\lambda]$ will be interpreted as gravitational energy-momentum densities constructed from the dynamical variables $Y_\lambda(\mathbf{x})$ on the hypersurface $X^\lambda = X^\lambda(\mathbf{x})$. We may change these densities by a "constant" amount $\delta\Phi[X^\lambda]/\delta X^\kappa(\mathbf{x})$ (independent of the dynamical variables $Y_\lambda(\mathbf{x})$) when we subject the original set $X^\lambda, \Pi_\lambda, Y_\lambda$ to a canonical transformation

$$\tilde{X}^\lambda = X^\lambda, \quad \tilde{\Pi}_\lambda = \Pi_\lambda - \delta\Phi[X^\mu]/\delta X^\lambda, \quad \tilde{Y}_\lambda = Y_\lambda$$

generated by an arbitrary functional $\Phi[X^\mu]$ of internal coordinates. The solution of the initial value equations written in terms of the tilded variables yields the shifted energy-momentum densities

$$-\tilde{\Pi}_\kappa = \tilde{P}_\kappa[\tilde{X}^\lambda, \tilde{Y}_\lambda] = P_\kappa[X^\lambda, Y_\lambda] + \frac{\delta\Phi[X^\lambda]}{\delta X^\kappa}. \quad (4.3)$$

The transformation (4.3) corresponds to a change of the zero point of energy in particle mechanics.

Provided the usual rules of the algebraic game are to be trusted in the functional game we play, the initial value equations can be solved with respect to the variables Π_κ, if the "matrix" $\mathcal{H}^{\iota\kappa} \equiv \delta\mathcal{H}^\iota/\delta\Pi_\kappa$ has an inverse $\mathcal{H}^{-1}{}_{\kappa\lambda}$,

$$\mathcal{H}^{\iota\kappa}\mathcal{H}^{-1}{}_{\kappa\lambda} = \delta^\iota_\lambda. \quad (4.4)$$

The symbol δ^ι_λ on the right-hand side of Eq. (4.4) represents the product of the Kronecker delta and the three-dimensional delta-function $\delta(\mathbf{x},\mathbf{x}')$, which is a bidensity of zero weight at the first argument and of unit weight at the second.

Because the composite functional (4.2) identically vanishes, we get by its variation with respect to the variables X^κ and Y_κ a couple of equations

$$\frac{\delta\mathcal{H}^\iota}{\delta X^\kappa} - \mathcal{H}^{\iota\lambda}\frac{\delta P_\lambda}{\delta X^\kappa} = 0,$$
$$\frac{\delta\mathcal{H}^\iota}{\delta Y_\kappa} - \mathcal{H}^{\iota\lambda}\frac{\delta P_\lambda}{\delta Y_\kappa} = 0. \quad (4.5)$$

Using Eq. (4.4), we solve Eqs. (4.5) with respect to the variational derivatives $\delta P_\lambda/\delta X^\kappa$ and $\delta P_\lambda/\delta Y_\kappa$,

$$\frac{\delta P_\lambda}{\delta X^\kappa} = \mathcal{H}^{-1}{}_{\lambda\iota}\frac{\delta\mathcal{H}^\iota}{\delta X^\kappa},$$

$$\frac{\delta P_\lambda}{\delta Y_\kappa} = \mathcal{K}^{-1}{}_{\lambda\iota}\frac{\delta \mathcal{K}^\iota}{\delta Y_\kappa}. \tag{4.6}$$

We can now "deparametrize" the action functional (3.7) by substituting into it the solution (4.1) of the initial value equations

$$S = \int dt\,(\pi_A \dot{g}^A - P_\iota \dot{X}^\iota). \tag{4.7}$$

The quantities X^ι are no longer to be varied, but are thought of as some prescribed functions of t and $\mathbf{x}$. The Hamilton equations are those corresponding to the Hamiltonian $P_\iota \dot{X}^\iota$,

$$\dot{Y}_\iota = [Y_\iota, P_\kappa]\dot{X}^\kappa. \tag{4.8}$$

The Poisson brackets [ ] are evaluated using g^A, π_A as a complete set of canonical variables, whereas the Poisson brackets **[ ]** were evaluated using a larger set X^ι, Π_ι, g^A, π_A. In the deparametrized formalism, Π_ι is expressed as a functional of X^ι, g^A, and π_A, and X^ι are no longer canonical variables.

Equations (4.8) can also be obtained directly from the Hamilton equations (3.8)–(3.11), without referring to the variational principle. Using Eq. (4.4), we solve Eq. (3.8) with respect to the quantities N_κ,

$$N_\kappa = \dot{X}^\iota \mathcal{K}^{-1}{}_{\iota\kappa}. \tag{4.9}$$

Substituting this expression for N_κ into Eqs. (3.10) and (3.11) and taking account of Eqs. (4.6), we get once more Eqs. (4.8). The remaining Hamilton equation (3.5) turns out to be, after using Eqs. (4.1), (4.6), and (4.8) in succession, equivalent to the relation

$$X^\kappa\left(\frac{\delta P_{[\iota}}{\delta X^{\kappa]}} + [P_\iota, P_\kappa]\right) = 0. \tag{4.10}$$

We shall return to its meaning in Sec. 5. The Hamilton equations (4.8) were derived in essentially the same way in Ref. 15.

The transition from the Hamilton equations (3.8)–(3.11) to the Hamilton equations (4.8) through the elimination of the lapse and shift functions represents a switch from the external path approach to the internal path approach. Indeed, the deparametrized Hamilton equations (4.8) allow us to follow the change in the dynamical variables Y_ι as we go along the internal path (1.4). For this purpose, we never need to know the spacing between the neighboring hypersurfaces and the shift in their labeling $\mathbf{x}$. However, we can recover the lapse and shift functions corresponding to the internal path (1.4) by going back to Eq. (4.9). The "matrix" $\mathcal{K}^{-1}{}_{\iota\kappa}$ in it is a functional of X^λ and Y_λ, which we obtain by solving Eq. (4.4) and then substituting for Π_κ its value (4.2). Equation (4.9) provides another version of the much discussed "thin-sandwich" conjecture. In the standard formulation of this conjecture, we specify freely the spatial metric g_{ik} (six quantities per space point) and its rate of change $\dot{g}_{ik}$ (another six quantities per space point), and we hope that the lapse and shift functions are determined uniquely under the appropriate boundary conditions. Here, we specify freely the internal coordinates X^ι (four quantities per space point), their rate of change $\dot{X}^\iota$ (another four quantities per space point), and the true dynamical variables Y_ι (still another four quantities per space point). The lapse and shift functions are then given by Eq. (4.9). The boundary conditions are involved when finding out the "inverse matrix" $\mathcal{K}^{-1}{}_{\iota\kappa}$ from Eq. (4.4). Let us note that the freely specifiable data X^ι, $\dot{X}^\iota$, and Y_ι do not necessarily correspond to the freely specifiable data g_{ik}, $\dot{g}_{ik}$, because the general canonical transformation (3.2) mixes superspace with momentum space.

Having reconstructed the external trajectory $N_\iota(t, \mathbf{x})$ from the internal trajectory $X^\iota(t, \mathbf{x})$, we return to the evolution problem. Let us prescribe the dynamical variables $Y^{(in)}_\iota(\mathbf{x})$ on an initial hypersurface $X^\iota = X^{(in)\iota}(\mathbf{x})$ and try to determine their values $Y^{(f)}_\iota(\mathbf{x})$ on a final hypersurface $X^\iota = X^{(f)\iota}(\mathbf{x})$. We connect the initial hypersurface with the final hypersurface by an internal path (1.4), with

$$\begin{aligned} X^\iota(t^{(in)}, \mathbf{x}) &= X^{(in)\iota}(\mathbf{x}), \\ X^\iota(t^{(f)}, \mathbf{x}) &= X^{(f)\iota}(\mathbf{x}), \end{aligned} \tag{4.11}$$

and solve the Hamilton equations (4.8) under the initial conditions

$$Y_\iota(t^{(in)}, \mathbf{x}) = Y^{(in)}_\iota(\mathbf{x}).$$

We thus get $Y^{(f)}_\iota(\mathbf{x})$ as $Y_\iota(t^{(f)}, \mathbf{x})$. However, the initial hypersurface may be connected with the final hypersurface by an infinite number of internal paths (1.4) and (4.11). We expect that the final values $Y^{(f)}_\iota(\mathbf{x})$ of the dynamical variables remain the same whatever path we choose. Because any hypersurface may be picked out as a final hypersurface, the dynamical variables should depend only on the hypersurface $X^\iota = X^\iota(\mathbf{x})$ itself, not on the way in which the hypersurface fits into some internal path. Rather than simple functions of t, the dynamical variables should be functionals of $X^\iota(\mathbf{x})$. Due to the arbitrariness of the direction $\dot{X}^\iota$ in which the internal path leaves the hypersurface X^ι, the Hamilton equations (4.8) imply that the functionals $Y_\iota[X^\kappa]$ satisfy the functional differential equation

$$\frac{\delta Y_\iota}{\delta X^\kappa} = [Y_\iota, P_\kappa], \tag{4.12}$$

under the initial conditions

$$Y_\iota[X^{(in)\kappa}] = Y^{(in)}_\iota.$$

Equation (4.16) is the bubble-time Hamilton equation which is the central point of the new formalism.

5. PATH INDEPENDENCE OF DYNAMICAL EVOLUTION

If we prescribe the internal path and go along it, we recover the differential equation (4.8) out of the functional differential equation (4.12), because

$$\dot{Y}_\iota[X^\kappa(t)] = \frac{\delta Y_\iota}{\delta X^\kappa}\Big|_{X^\lambda = X^\lambda(t)}\dot{X}^\kappa. \tag{5.1}$$

On the other hand, the set of all partial differential equations (4.8) taken along different internal paths was summarized by one functional differential equation (4.12) only under our assumption that the evolution of dynamical variables is path independent.

Let us prove that the dynamical evolution predicted by the functional differential equation (4.13) is really path independent. This amounts to the integrability condition

$$\frac{\delta}{\delta X^{[\lambda}}\,[Y_\iota, P_\kappa]] = 0, \tag{5.2}$$

which the right-hand side of Eq. (4.12) must satisfy because the second variational derivatives of Y_ι with respect to X^κ commute,

$$\delta^2 Y_\iota / \delta X^{[\kappa} \delta X^{\lambda]} = 0 \quad \text{(see Fig. 1c)}.$$

Condition (5.2) can be cast into a simpler form. Using the original equation (4.12), we get

$$\frac{\delta}{\delta X^{[\lambda}}\,[Y_\iota, P_{\kappa]}] = -\,[[Y_\iota, P_{[\kappa}], P_{\lambda]}] - \left[Y_\iota, \frac{\delta P_{[\kappa}}{\delta X^{\lambda]}}\right].$$

On the other hand, by virtue of the Jacobi identity and the antisymmetry of the Poisson brackets,

$$[[Y_\iota, P_{[\kappa}], P_{\lambda]}] = [Y_\iota, [P_\kappa, P_\lambda]].$$

The integrability condition (5.2) therefore reduces to the condition

$$\frac{\delta P_{[\kappa}}{\delta X^{\lambda]}} + [P_\kappa, P_\lambda] = 0. \tag{5.3}$$

Equation (5.3) is equivalent to Eq. (4.10) which was obtained as a consequence of the parametrized Hamilton equations (3.8)–(3.11). However, we can check that it is actually an identity valid independently of the dynamical Hamilton equations (4.8), by virtue of the way the energy-momentum densities are constructed out of the variables X^ι, Y_ι. In fact, it is a consequence of the completeness conditions (2.8) of the initial value equations.

The Poisson brackets in Eqs. (2.8) are evaluated using g^A, π_A, and X^ι, Π_ι as independent canonical variables. After that, we can substitute into Eqs. (2.8) the solution (4.1) of the initial value equations and get

$$\frac{\delta \mathcal{H}^{[\mu}}{\delta g^A}\frac{\delta \mathcal{H}^{\nu]}}{\delta \pi_A} + \frac{\delta \mathcal{H}^{[\mu}}{\delta X^\alpha}\frac{\delta \mathcal{H}^{\nu]}}{\delta \Pi_\alpha} = 0.$$

If we multiply the last equation by ${\mathcal{H}^{-1}}_{\kappa\mu}{\mathcal{H}^{-1}}_{\lambda\nu}$ and use Eqs. (4.4) and (4.5), we obtain exactly the integrability condition (5.3). This proves the consistency of the bubble-time Hamilton equation (4.12).

Another way of seeing that Eq. (5.3) ought to be identically satisfied is to follow the argument of ADM.[6] They point out that Eq. (5.3) guarantees the preservation of the initial value equations $\mathcal{H}^\iota = 0$ in time and that it turns into the Bianchi identity once the equations of motion are used to eliminate the time derivatives.

6. DISPLACEMENT ALONG SPACELIKE HYPERSURFACE

The functional differential equation (4.12) describes how the dynamical variables at a point with the label $\mathbf{x}$ change if we create a little bubble δX^κ around another point with the label $\mathbf{x}'$ [remember that the indices in Eq. (4.12) represent at the same time the coordinate labels]. Going from the old hypersurface $X^\kappa = X^\kappa(\mathbf{x})$ to a new hypersurface $X^\kappa = X^\kappa(\mathbf{x}) + \delta X^\kappa(\mathbf{x})$, the change in the dynamical variables $Y_\iota(\mathbf{x})$ predicted by Eq. (4.12) amounts to

$$\delta Y_\iota = \frac{\delta Y_\iota}{\delta X^\kappa}\,\delta X^\kappa = [Y_\iota, P_\kappa]\delta X^\kappa. \tag{6.1}$$

Note that the displacement $\delta X^\kappa(\mathbf{x})$ connects the points with the same labels on the old and new hypersurfaces (see Fig. 1a).

In one case, however, the change in the dynamical variables can be written down directly, without using the Hamilton equation (4.12) at all—namely, if the displacements $\delta X^\iota(\mathbf{x})$ are all tangential to the hypersurface $X^\iota = X^\iota(\mathbf{x})$,

$$\delta X^\iota(\mathbf{x}) = {X^\iota}_{,l}(\mathbf{x})\xi^l(\mathbf{x}), \tag{6.2}$$

connecting the points which originally had the labels $\mathbf{x}$ and $\mathbf{x} + \boldsymbol{\xi}(\mathbf{x})$. The change δY_ι then consists of two parts. The first part is due to the step from the old point $\mathbf{x}$ to the new point $\mathbf{x} + \xi$, and is simply equal to $Y_{\iota,l}(\mathbf{x})\xi^l(\mathbf{x})$. However, because the displacement $\delta X^\iota(\mathbf{x})$ should connect the points with the same label, we must yet relabel the hypersurface in such a way that the point with the original label $\mathbf{x} + \xi$ acquires the new label $\mathbf{x}$. This induces the second part of the change of the dynamical variables, depending on their transformation properties. For the scalars g^A it vanishes, and for the scalar densities π_A it is equal to $\pi_A(\mathbf{x})\xi^i{}_{,i}(\mathbf{x})$. Putting both parts of the total change δY_ι together, we get just the Lie derivative $\pounds_\xi$ of the dynamical variable Y_ι,

$$\delta Y_\iota(\mathbf{x}) = \pounds_\xi\, Y_\iota(\mathbf{x}) = \begin{cases} g_{A,\iota}(\mathbf{x})\xi^l(\mathbf{x}) \\ (\pi_A(\mathbf{x})\xi^l(\mathbf{x}))_{,l} \end{cases} \tag{6.3}$$

Because the direct approach should lead to the same answer for δY_ι as the Hamilton equation, we can compare Eqs. (6.1)–(6.3) and conclude that

$$\pounds_\xi\, Y_\iota(\mathbf{x}) = \int d^3x'\, [Y_\iota(\mathbf{x}), P_\kappa(\mathbf{x}')]{X^\kappa}_{,l}(\mathbf{x}')\xi^l(\mathbf{x}'). \tag{6.4}$$

Equation (6.4) is equivalent to several other statements. Substituting for the Lie derivatives of the dynamical variables the actual expressions (6.3) and taking account of the arbitrariness of ξ^l, we can replace Eq. (6.4) by two equations:

$$\begin{aligned} g^A{}_{,l}(\mathbf{x})\delta(\mathbf{x}, \mathbf{x}') &= \frac{\delta}{\delta \pi_A(\mathbf{x})}\,P_\kappa(\mathbf{x}'){X^\kappa}_{,l}(\mathbf{x}'), \\ \pi_A(\mathbf{x})\delta_{,l}(\mathbf{x}, \mathbf{x}') &= -\frac{\delta}{\delta g^A(\mathbf{x})}\,P_\kappa(\mathbf{x}'){X^\kappa}_{,l}(\mathbf{x}'). \end{aligned} \tag{6.5}$$

Treating them as variational equations for the functional $P_\kappa(\mathbf{x}'){X^\kappa}_{,l}(\mathbf{x}')$ of variables X^ι and Y_ι, we easily guess their solution,

$$P_\kappa(\mathbf{x}){X^\kappa}_{,l}(\mathbf{x}) = \pi_A(\mathbf{x})g^A{}_{,l}(\mathbf{x}). \tag{6.6}$$

In fact, an arbitrary invariant functional $F_l(\mathbf{x})[X^\mu]$ of internal coordinates X^μ transforming as a vector density could be added to the right-hand side of Eq. (6.6) as an integration constant. However, we shall prove at the end of this section that this functional must be put equal to zero.

Equation (6.6) represents $3^{\infty 3}$ equations for $4^{\infty 3}$ unknowns $P_\kappa(\mathbf{x})$. We can therefore express $3^{\infty 3}$ of these

unknowns, say $P_k(\mathbf{x})$, in terms of the remaining ∞^3 ones, say $P_0(\mathbf{x})$:

$$P_k(\mathbf{x}) = \{\pi_A(\mathbf{x}) g^A{}_{,l}(\mathbf{x}) - P_0(\mathbf{x}) X^0{}_{,l}(\mathbf{x})\} \frac{\partial x^l}{\partial X^k}. \tag{6.7}$$

The gravitational momentum density P_k on the hypersurface $X^\iota = X^\iota(\mathbf{x})$ is thereby expressed explicitly by means of the gravitational energy density P_0 and the dynamical variables g^A, π_A.[16]

Equation (6.4), or any of the equivalent equations (6.5)–(6.7), imposes another consistency requirement on the formalism. If it were not satisfied identically, by virtue of the structure of the functionals $P_\kappa(\mathbf{x})[X^\lambda, Y_\lambda]$, it would represent a constraint on the initial data $Y_\iota(\mathbf{x})$ which could not be therefore prescribed arbitrarily. However, we show now that Eq. (6.4) is an identity, due to the fact that the supermomentum $\mathcal{H}^i$ generates the coordinate transformations, and the dynamical variables are defined in an invariant way.

The invariant character of the canonical transformation (3.1) allowed us to express the Lie derivatives of the dynamical variables Y_ι by means of the Lie derivatives of the old canonical variables g_{ik}, π^{ik} in Eq. (3.5). However, the Lie derivatives of the canonical variables g_{ik}, π^{ik} are generated by the supermomentum $\mathcal{H}^i$ according to Eqs. (2.9). We can therefore rewrite Eq. (3.5) in the form

$$\pounds_\xi Y_\iota = [Y_\iota, \mathcal{H}_k \xi^k]. \tag{6.8}$$

Because the Poisson bracket is an invariant of canonical transformations, we can evaluate it after introducing Y_ι as canonical variables.

We get

$$[g^A, \mathcal{H}_k \xi^k] = \frac{\delta \mathcal{H}_k \xi^k}{\delta \pi_A},$$

$$[\pi_A, \mathcal{H}_k \xi^k] = -\frac{\delta \mathcal{H}_k \xi^k}{\delta g^A}. \tag{6.9}$$

At this stage, we substitute the functionals $P_\kappa[X^\lambda, Y_\lambda]$ for Π_κ to the right-hand sides of Eq. (6.9), and use Eq. (4.5). This brings us to the formula

$$\pounds_\xi Y_\iota = \frac{\delta \mathcal{H}_k \xi^k}{\delta \Pi_\lambda} [Y_\iota, P_\lambda] \tag{6.10}$$

for the Lie derivative of dynamical variables. We can now go through the same sequence of steps to calculate the Lie derivative of internal coordinates X^λ,

$$\pounds_\xi X^\lambda = [X^\lambda, \mathcal{H}_k \xi^k] = \frac{\delta \mathcal{H}_k \xi^k}{\delta \Pi_\lambda}. \tag{6.11}$$

Combining Eq. (6.11) with Eq. (6.10), we conclude that the consistency condition (6.4) is identically satisfied.

We obtained an integral form (6.6) of the consistency condition (6.4), picking up an integration functional $F_l(\mathbf{x})[X^\mu]$ which should have been added to the right-hand side of Eq. (6.6). We show now that this integration functional is actually equal to zero. We multiply the integrability condition (5.3) by $X^\kappa{}_{,l}(\mathbf{x})$ and sum over the index κ (no integration is implied over the corresponding label $\mathbf{x}$). Using Eq. (6.6) with $F_l(\mathbf{x})[X^\mu]$ on its right-hand side, we get

$$\frac{\delta P_\kappa(\mathbf{x})}{\delta X^\lambda(\mathbf{x}')} X^\kappa{}_{,l}(\mathbf{x}) - \frac{\delta P_\lambda(\mathbf{x}')}{\delta X^\kappa(\mathbf{x})} X^\kappa{}_{,l}(\mathbf{x}) + [\pi_A(\mathbf{x}) g^A(\mathbf{x})_{,l}, P_\lambda(\mathbf{x}')] = 0. \tag{6.12}$$

The first tem in Eq. (6.12) may be rewritten as

$$\frac{\delta P_\kappa(\mathbf{x})}{\delta X^\lambda(\mathbf{x}')} X^\kappa{}_{,l}(\mathbf{x}) = \frac{\delta}{\delta X^\lambda(\mathbf{x}')} (P_\kappa(\mathbf{x}) X^\kappa{}_{,l}(\mathbf{x})) - P_\lambda(\mathbf{x}) \delta_{,l}(\mathbf{x}, \mathbf{x}'). \tag{6.13}$$

The second term in Eq. (6.12) can be obtained from the relation

$$\pounds_\xi P_\lambda = \frac{\delta P_\lambda}{\delta X^\kappa} \pounds_\xi X^\kappa + \frac{\delta P_\lambda}{\delta Y_\kappa} \pounds_\xi Y_\kappa$$

expressing the fact that P_λ is an invariant functional of X^κ and Y_κ. Writing down the Lie derivatives of scalars X^κ and scalar densities Y_κ, P_λ explicitly, taking account of the arbitrariness of $\xi^l(\mathbf{x})$ and grouping the terms with $\delta P_\lambda / \delta Y_\kappa$ into a Poisson bracket, we get

$$-P_\lambda(\mathbf{x}) \delta_{,l}(\mathbf{x}, \mathbf{x}') = \frac{\delta P_\lambda(\mathbf{x}')}{\delta X^\kappa(\mathbf{x})} X^\kappa{}_{,l}(\mathbf{x}) - [\pi_A(\mathbf{x}) g^A{}_{,l}(\mathbf{x}), P_\lambda(\mathbf{x}')]. \tag{6.14}$$

Substituting now from Eqs. (6.13) and (6.14) into Eq. (6.12), all terms except one cancel and we get

$$\frac{\delta}{\delta X^\lambda(\mathbf{x}')} (P_\kappa(\mathbf{x}) X^\kappa{}_{,l}(\mathbf{x})) = 0.$$

This implies that the functional $F_l(\mathbf{x})[X^\mu]$ in Eq. (6.6) does not depend on internal coordinates at all and ought to be discarded. Note that the dynamical equations (4.12) were never used in this proof.

7. REDUCED FORMALISM

The identity (6.4) allows us to reduce the number of independent components of the functional differential Hamilton equation (4.12). Recall that this equation has $4\infty^3 \times 4\infty^3$ components, each index representing at the same time a coordinate label $\mathbf{x}$. Assume that $Y_\iota(\mathbf{x})[X^\kappa]$ are some functionals of X^κ independent of the labeling of the hypersurface, but otherwise completely arbitrary. According to Eq. (3.6),

$$\pounds_\xi Y_\iota(\mathbf{x}) = \int d^3x' \frac{\delta Y_\iota(\mathbf{x})}{\delta X^\kappa(\mathbf{x}')} X^\kappa{}_{,l}(\mathbf{x}') \xi^l(\mathbf{x}'). \tag{7.1}$$

Using the identity (6.4) and taking account of the arbitrariness of ξ^l, we conclude that the functionals Y_ι must satisfy the equations

$$\left(\frac{\delta Y_\iota}{\delta X^\kappa(\mathbf{x})} - [Y_\iota, P_\kappa(\mathbf{x})]\right) X^\kappa{}_{,l}(\mathbf{x}) = 0 \tag{7.2}$$

(summation over κ, but no integration over $\mathbf{x}$!). However, these equations are just $4\infty^3 \times 3\infty^3$ linear combinations of the Hamilton equations (4.12). So, if $Y_\iota(\mathbf{x})[X^\kappa]$ are invariantly defined functionals of X^κ and if they satisfy the $4\infty^3 \times \infty^3$ equations obtained by putting $\kappa = 0$ in Eq. (4.12), they automatically satisfy the remaining $4\infty^3 \times 3\infty^3$ equations obtained by putting

$\kappa = k$. This is also intuitively clear, because an arbitrary displacement $\delta X^\iota(\mathbf{x})$ can be decomposed into a displacement $\delta X^\iota(\mathbf{x}) = X^\iota{}_{,\lambda}(\mathbf{x})\,\xi^\lambda(\mathbf{x})$ lying completely in the hypersurface, and a displacement $\delta X^\kappa(\mathbf{x}) = \delta_0 X^0(\mathbf{x})\delta^{0\kappa}$ in the direction of the X^0 lines (see Fig. 1b). The change of Y_ι under the displacement in the hypersurface being trivially generated by the Lie derivative, we need to know only the change of Y_ι under the displacement along the X^0 lines to determine the change of Y_ι under an arbitrary displacement $\delta X^\iota(\mathbf{x})$.

The reduction in the number of independent equations for Y_ι is best displayed when adopting the internal coordinates X^k as privileged labels x^k. The hypersurfaces are then defined by deparametrized equations (1.3). The dynamical variables in privileged labeling reduce to functionals $Y_\iota(\mathbf{X})[T]$ of a single function variable $T(\mathbf{X})$,

$$Y_\iota(\mathbf{X})[T(\mathbf{X})] \equiv Y_\iota(\mathbf{x} = \mathbf{X})[X^0 = X^0(\mathbf{x}), X^k = x^k]. \tag{7.3}$$

Similarly, the energy-momentum functionals P_ι reduce to functionals of five function variables $T(\mathbf{X})$, $Y_\iota(\mathbf{X})$. If we change the hypersurface but preserve the privileged labeling, the dynamical variables observe the functional differential equation

$$\frac{\delta Y_\iota(\mathbf{X})}{\delta T(\mathbf{X}')} = [Y_\iota(\mathbf{X}), P_0(\mathbf{X}')] \tag{7.4}$$

with a single bubble-time variable $T(\mathbf{X})$.

Suppose we know a solution $Y_\iota(\mathbf{X})[T]$ of this functional differential equation. We can then easily construct a corresponding solution $Y_\iota(\mathbf{x})[X^\kappa]$ of the full set of functional differential equations (4.12). For a given set of functions $X^\kappa(\mathbf{x})$, the inverse functions $\mathbf{x} = \mathbf{x}(\mathbf{X})$ specify the transformation from the internal coordinates $\mathbf{X}$ to arbitrary coordinates $\mathbf{x}$, and the composite function $X^0 = X^0(\mathbf{x}(\mathbf{X})) \equiv T(\mathbf{X})$ gives the deparametrized form of the equation of the hypersurface. We *define* the functionals $Y_\iota(\mathbf{x})[X^\kappa]$ by taking the known functionals $Y_\iota(\mathbf{X})[T]$ and transforming them by the coordinate transformation $\mathbf{X} \to \mathbf{x} = \mathbf{x}(\mathbf{X})$. By the very method of their definition, the variables $Y_\iota(\mathbf{x})[X^\kappa]$ as functionals of $X^\kappa(\mathbf{x})$ do not depend on the labeling of the hypersurface, and therefore satisfy Eq. (7.2). It is also easy to check that they satisfy the $\kappa = 0$ component of Eq. (4.12), due to the fact that the functionals $Y_\iota(\mathbf{X})[T]$ satisfy Eq. (7.4). They therefore satisfy all components of Eq. (4.12). In this sense, the entire dynamics of the gravitational field is reduced to a single bubble-time functional differential equation (7.4).

8. EINSTEIN–HAMILTON–JACOBI EQUATION

It is now a standard procedure to put the canonical version of Einstein's theory into the Hamilton–Jacobi form. Peres,[17] and after him many other investigators,[18] used what may be called the metric representation. In it, the momenta $\pi^{ik}(\mathbf{x})$ are generated from a single functional $S[g_{ik}]$ of the metric,

$$\pi^{ik}(\mathbf{x}) = \frac{\delta S}{\delta g_{ik}(\mathbf{x})}. \tag{8.1}$$

The equations for this functional are obtained when substituting the momenta (8.1) into the initial value equations (2.3), (2.4), and (2.7). In this way we get

$$g^{-1/2}(\mathbf{x})\Big(g_{il}(\mathbf{x})g_{km}(\mathbf{x}) - \tfrac{1}{2}g_{ik}(\mathbf{x})g_{lm}(\mathbf{x})\Big)\frac{\delta S}{\delta g_{ik}(\mathbf{x})}\frac{\delta S}{\delta g_{lm}(\mathbf{x})} - g^{1/2}(\mathbf{x})R(\mathbf{x}) = 0, \tag{8.2}$$

$$\left(\frac{\delta S}{\delta g_{ik}(\mathbf{x})}\right)_{|k} = 0 \tag{8.3}$$

It is well known that Eq. (8.3) is equivalent to the statement that Hamilton's principal functional S does not depend on the labeling of the hypersurface. Rather than being a functional of $6\infty^3$ components of the metric tensor $g_{ik}(\mathbf{x})$, it is a functional of $3\infty^3$ equivalence classes of such metrics which are connected with each other by three-dimensional transformations. In other words, it is a functional of a 3-geometry.

From the $3\infty^3$ arguments of the Hamilton principal functional, $2\infty^3$ arguments should represent the dynamical degrees of freedom of the gravitational field, and the remaining ∞^3 arguments a bubble time. Equation (8.2) is called the Einstein–Hamilton–Jacobi equation. It governs the change of Hamilton's principal functional when passing from one 3-geometry to another or, in our new language, the evolution of Hamilton's principal functional of the dynamical variables in the time variable.

In the metric representation, the time variable is never separated from the dynamical variables. It is then easy to forget that time is actually one of the arguments of Hamilton's principal functional and wonder how the time evolution of the gravitational field can possibly be described by a Hamilton–Jacobi equation which seemingly refers only to a single spatial hypersurface.

The interpretation of the Hamilton–Jacobi formalism becomes more transparent in the new representation we are advocating in this paper. Instead of taking S to be a functional of $6\infty^3$ components of the metric tensor, we take it to be a functional of $4\infty^3 + 2\infty^3$ variables $X^\kappa(\mathbf{x}), g^A(\mathbf{x})$. We generate the conjugate momenta $\Pi_\kappa(\mathbf{x}), \pi_A(\mathbf{x})$ by variational derivatives

$$\Pi_\kappa(\mathbf{x}) = \frac{\delta S}{\delta X^\kappa(\mathbf{x})}, \qquad \pi_A(\mathbf{x}) = \frac{\delta S}{\delta g^A(\mathbf{x})}, \tag{8.4}$$

and turn the initial value equations (2.7) into equations for the Hamilton principal functional S,

$$\mathcal{H}^\iota(\mathbf{x})\left[X^\kappa, \frac{\delta S}{\delta X^\kappa}, g^A, \frac{\delta S}{\delta g^A}\right] = 0. \tag{8.5}$$

We assumed that the initial value equations could be solved with respect to the energy-momentum variables, the solution being given by Eq. (4.1). The Hamilton–Jacobi equations (8.5) can be therefore solved with respect to the variational derivatives $\delta S/\delta X^\kappa$,

$$\frac{\delta S}{\delta X^\kappa} + P_\kappa\left[X^\lambda, \frac{\delta S}{\delta g^A}\right] = 0. \tag{8.6}$$

The bubble-time Hamilton–Jacobi equation (8.6) confronts us with the same consistency problems as the bubble-time Hamilton equations (4.12). Firstly, we must require that the evolution of Hamilton's principal functional from an initial to a final hypersurface does not depend on the path. Secondly, the evolution of Hamilton's principal functional in the direction tangential to a spatial hypersurface should be trivial, provi-

ded Hamilton's principal functional is not affected by the relabeling of the hypersurface.

The first requirement is equivalent to an integrability condition of Eq. (8.6), following from the fact that the variational derivatives of S commute: $\delta^2 S/\delta X^{[\kappa}\delta X^{\lambda]} = 0$. This integrability condition coincides with the integrability condition (5.3) of the Hamilton equation (4.12), which we have proved to be an identity.

Passing to the second consistency requirement, we write down condition (3.6) that Hamilton's principal functional does not depend on the labeling of the hypersurface,

$$\frac{\delta S}{\delta X^{\kappa}(\mathbf{x})} X^{\kappa}{}_{,l}(\mathbf{x}) + \frac{\delta S}{\delta g^{A}(\mathbf{x})} g^{A}{}_{,l}(\mathbf{x}) = 0. \tag{8.7}$$

Expressing $\delta S/\delta X^{\kappa}$ from the Hamilton–Jacobi equation, we get

$$-\frac{\delta S}{\delta g^{A}(\mathbf{x})} g^{A}{}_{,l}(\mathbf{x}) + P_{\kappa}(\mathbf{x})\left[X^{\lambda}, g^{A}, \frac{\delta S}{\delta g^{A}}\right] X^{\kappa}{}_{,l}(\mathbf{x}) = 0. \tag{8.8}$$

This relation is actually an identity, valid for any functional $S[X^{\lambda}, g^{A}]$ because of the structure of the energy-momentum densities P_{κ} as functionals of X^{λ} and Y_{λ}. In fact, Eq. (8.8) is the same equation as the identity (6.6).

We can now reverse the argument and say that the $3\infty^3$ linear combinations

$$\left(\frac{\delta S}{\delta X^{\kappa}(\mathbf{x})} + P_{\kappa}(\mathbf{x})\left[X^{\lambda}, g^{A}, \frac{\delta S}{\delta g^{A}}\right]\right) X^{\kappa}{}_{,l}(\mathbf{x}) = 0$$

of the $4\infty^3$ Hamilton–Jacobi equations (8.6) are automatically satisfied, if S is any functional which does not depend on the labeling of the hypersurface. Therefor, if S is such a functional and if it in addition satisfies the $\kappa = 0$ component of the Hamilton–Jacobi equation (8.6), it also satisfies the remaining $\kappa = k$ components of this equation. This allows us to reduce the $4\infty^3$ Hamilton–Jacobi equations (8.6) to ∞^3 equations. As in Sec. 7, we pick up the special labeling $\mathbf{X} = \mathbf{x}$ of the hypersurface and write down a single functional differential equation

$$\frac{\delta S}{\delta T(\mathbf{X})} + P_0(\mathbf{X})\left[T, g^{A}, \frac{\delta S}{\delta g^{A}}\right] = 0 \tag{8.9}$$

for a functional $S[T, g^{A}]$ of *three* functions $T(\mathbf{x}), g^{A}(\mathbf{X})$ of $\mathbf{X}$. Assuming we know a solution $S[T, g^{A}]$ of this equation, we can define a new functional $S[X^{\kappa}, g^{A}]$ of *six* functions $X^{\kappa}(\mathbf{x}), g^{A}(\mathbf{x})$ by putting

$$S[X^{\kappa}(\mathbf{x}), g^{A}(\mathbf{x})] = S[T(\mathbf{X}) = X^{0}(\mathbf{x}(\mathbf{X})), g^{A}(\mathbf{x}(\mathbf{X}))]. \tag{8.10}$$

The new functional $S[X^{\kappa}, g^{A}]$ does not depend on the labeling of the hypersurface by the very method of its construction and it satisfies the $4\infty^3$ Hamilton–Jacobi equations (8.6). The whole content of Einstein's gravitational law is thus hidden in a single variational Hamilton–Jacobi equation (8.5) for a Hamilton's principal functional S of two functions $g^{A}(\mathbf{X})$ representing the dynamical degrees of freedom, and one function $T(\mathbf{X})$ representing a bubble time. This equation determines $S[T^{(f)}, g^{A}]$ on a final hypersurface $T = T^{(f)}(\mathbf{X})$, if we know $S[T^{(\mathrm{in})}, g^{A}]$ on an initial hypersurface $T = T^{(\mathrm{in})}(\mathbf{x})$.

What is even better, we can replace this variational equation in a bubble time by a partial differential equation in a coordinate time, due to the path independence of the dynamical evolution. We simply connect the initial hypersurface with the final hypersurface by an arbitrary internal trajectory $T = T(t, \mathbf{X})$, and ask how Hamilton's principal functional changes if we pass from one hypersurface of the internal trajectory to the next. In the same manner as the functional differential Hamilton equation (4.12) reduces to the ordinary differential equation (4.8), the functional differential Hamilton–Jacobi equation (8.9) reduces to one ordinary differential equation

$$\dot{S}(t) + \int d^3X\, P_0(\mathbf{x})\left[T(t), g^{A}, \frac{\delta S(t)}{\delta g^{A}}\right] \dot{T}(t, \mathbf{X}) = 0. \tag{8.11}$$

This is as far as any conceivable reduction of the evolution problem can go. Let us note that $P_0(\mathbf{X})$ is the same functional as that which enters the single bubble-time Hamilton equation (7.4), and that

$$S(t)[g^{A}(\mathbf{X})] \equiv S[T(t, \mathbf{X}), g_{A}(\mathbf{X})].$$

If we know a complete solution $S[T, g^{A}; \alpha^{A}]$ of the Hamilton–Jacobi equation (8.9) depending on two arbitrary functions $\alpha^{A}(\mathbf{X})$ (none of which is solely additive), we can follow the well-known procedure and find out the evolution of the dynamical variables $Y_{\iota}(\mathbf{X})[T]$ starting from their initial values $Y^{(\mathrm{in})}{}_{\iota}(\mathbf{X})$ on an initial hypersurface $T = T^{(\mathrm{in})}(\mathbf{X})$. We write down the equations

$$\pi_{B} = \frac{\delta}{\delta g^{B}} S[T, g^{A}; \alpha^{A}],$$

$$\beta_{B} = \frac{\delta}{\delta \alpha^{B}} S[T, g^{A}; \alpha^{A}],$$

containing two new functions $\beta_{B}(\mathbf{X})$. If we substitute into these equations the initial values $g^{(\mathrm{in})A}(\mathbf{X})$ and $\pi^{(\mathrm{in})}{}_{A}(\mathbf{X})$ of the dynamical variables on the hypersurface $T = T^{(\mathrm{in})}(\mathbf{X})$, we can treat them as equations determining the unknown functions $\alpha^{A}(\mathbf{X}), \beta_{A}(\mathbf{X})$ in terms of the initial data. After that, we pass to an arbitrary hypersurface $T = T(\mathbf{X})$ and solve the same equations with respect to the dynamical variables g^{A}, π_{A}, obtaining the dynamical trajectory

$$Y_{\iota}(\mathbf{X})[T; Y^{(\mathrm{in})}{}_{\kappa}].$$

Once we know the dynamical trajectory, we can repeat the steps explained in the preceding sections and reconstruct the solution $g_{ik}(t, \mathbf{x}), N_{\iota}(t, \mathbf{x})$ of Einstein's gravitational law.

9. AN EXAMPLE: CYLINDRICAL WAVES

It is nice to have a model for which the explicit form of the canonical transformation (3.1) is known and all steps of the general formalism can be worked out. The Einstein–Rosen cylindrical gravitational waves are ideal for this purpose. They represent the simplest model with an infinite number of degrees of freedom allowing for a bubble time. When constructing the model, we apply the canonical formalism to a part of the original phase space by restricting the metric g_{ik}, the momentum π^{ik}, and the lapse and shift functions N, N_{i} by an intransitive group of motions with spacelike Killing vectors. The restriction on the

lapse and shift functions means that only those slices which respect the space–time symmetries are permissible. The region of the phase space in which the restricted dynamics takes place is called mini-phase-space. The general schema of the reduction and the canonical formalism for cylindrical waves was presented in an earlier paper.[10] The reader is referred to it for the details.

Here we start by noting that a system of coordinates r, φ, z exists in which the cylindrically symmetric canonical variables g_{ik}, π^{ik} and the lapse and shift functions on any permissible slice assume the form

$$g_{11} = e^{\gamma-\psi}, \quad g_{22} = R^2 e^{-\psi}, \quad g_{33} = e^{\psi}, \quad g_{12} = g_{13} = g_{23} = 0, \tag{9.1}$$

$$\pi^{11} = \pi_\gamma e^{\psi-\gamma}, \quad \pi^{22} = \tfrac{1}{2} R \pi_R e^{\psi}, \quad \pi^{33} = (\pi_\gamma + \tfrac{1}{2} R \pi_R + \pi_\psi) e^{-\psi}, \pi^{12} = \pi^{13} = \pi^{23} = 0, \tag{9.2}$$

$$N = N(r), \quad N_1 = N_1(r), \quad N_2 = N_3 = 0. \tag{9.3}$$

The variables γ, R, ψ and $\pi_\gamma, \pi_R, \pi_\psi$ are functions of the radial coordinate r only. The coordinates φ and z are fixed by the symmetry requirements up to the trivial transformations

$$\varphi \to \bar{\varphi} = \pm \varphi + \varphi_0, \quad z \to \bar{z} = \alpha z + z_0$$

with constant coefficients φ_0, z_0, α. On the other hand, the radial coordinate r can be changed arbitrarily,

$$r \to \bar{r} = f^{-1}(r). \tag{9.4}$$

Under the transformation (9.4), the momenta $\pi_\gamma, \pi_R, \pi_\psi$ behave as scalar densities

$$\pi_\gamma(r) \to \bar{\pi}_\gamma(\bar{r}) = f'(\bar{r}) \pi_\gamma(f(\bar{r})), \text{ etc.},$$

and the functions R and Ψ as scalars

$$R(r) \to \bar{R}(\bar{r}) = R(f(\bar{r})), \text{ etc.},$$

whereas the function γ has an anomalous transformation property

$$\gamma(r) \to \bar{\gamma}(\bar{r}) = \gamma(f(\bar{r})) + 2 \ln f'(\bar{r}).$$

(Prime denotes the differentiation with respect to the radial coordinate r.) The functions γ, R, ψ and $\pi_\gamma, \pi_R, \pi_\psi$ can be treated as canonically conjugate variables, because the action functional (2.1) (the integration over z being limited by the two "planes" $z = z_0$ and $z = z_0 + 1$, to avoid an infinite action) maintains the canonical form

$$S = 2\pi \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dr (\pi_\gamma \dot{\gamma} + \pi_R \dot{R} + \pi_\psi \dot{\psi} - N\mathcal{H} - N_1 \mathcal{H}^1),$$

when the substitutions (9.1)–(9.3) are performed. The super-Hamiltonian and supermomentum can be calculated directly through their definitions (2.3), (2.4), and they are

$$\begin{aligned} \mathcal{H} &= e^{(1/2)(\psi-\gamma)} (-\pi_\gamma \pi_R + \tfrac{1}{2} R^{-1} \pi_\psi^2 + 2R'' - \gamma' R' + \tfrac{1}{2} R \psi'^2), \\ \mathcal{H}^1 &= e^{\psi-\gamma} (-2\pi_\gamma' + \gamma' \pi_\gamma + R' \pi_R + \psi' \pi_\psi), \\ \mathcal{H}^2 &= \mathcal{H}^3 = 0. \end{aligned} \tag{9.5}$$

The structure of the super-Hamiltonian and supermomentum in the canonical variables γ, R, ψ and $\pi_\gamma, \pi_R, \pi_\psi$ is complicated. A remarkable simplification is achieved by introducing new variables $T, R; \Pi_T, \Pi_R; \psi, \pi_\psi$ by the canonical transformation

$$\begin{aligned} T(r) &= T(\infty) + \int_\infty^r -\pi_\gamma(r) dr, \\ \Pi_T(r) &= -\gamma'(r) + \{\ln(R'^2(r) - \pi_\gamma^2(r))\}', \\ \Pi_R(r) &= \Pi_R(r) + (\ln\{(R'(r) - \pi_\gamma(r))(R'(r) + \pi_\gamma(r))^{-1}\})', \\ R(r) &= R(r), \quad \psi(r) = \psi(r), \quad \pi_\psi(r) = \pi_\psi(r). \end{aligned} \tag{9.6}$$

The transformation (9.6) is an example of the canonical transformation (3.1) in mini-phase-space, with

$$X^0 \leftrightarrow T, \; X^1 \leftrightarrow R, \quad \Pi_0 \leftrightarrow \Pi_T, \; \Pi_1 \leftrightarrow \Pi_R, \quad g^1 \leftrightarrow \psi, \; \pi_1 \leftrightarrow \pi_\psi.$$

The φ and z components of internal coordinates and momentum densities, as well as the second polarization of the dynamical variables, are suppressed by the cylindrical symmetry of the model. The new variables have the desired transformation properties under the relabeling (9.4) of the radial coordinate: T, R, ψ are scalars, and Π_T, Π_R, Π_ψ are scalar densities. The canonical variables T, R and Π_T, Π_R can be identified with familiar quantities; T and R are the Einstein–Rosen coordinates in which the Einstein equations for cylindrical waves are usually treated, and the momenta $-\Pi_T$ and $-\Pi_R$ coincide with the C-energy density and C-flux introduced by Thorne.[19]

The permissible slices and their radial labeling are further restricted by the boundary conditions

$$\begin{aligned} &T' \to 0, \; (\gamma \to 0), \; R \to r \quad \text{for } r \to 0, \\ &T \to t, \; R \to r \quad \text{for } r \to \infty \text{ or for } t \to \pm\infty. \end{aligned}$$

The dynamical variables must also satisfy certain boundary conditions, but we need not use them in this paper.

The super-Hamiltonian and the radial supermomentum can be expressed by means of the new variables, thus we get

$$\begin{aligned} \mathcal{H} &= e^{(1/2)(\psi-\gamma)} (R' \Pi_T + T' \Pi_R + \tfrac{1}{2} R^{-1} \pi_\psi^2 + \tfrac{1}{2} R \psi'^2), \\ \mathcal{H}^1 &= e^{\psi-\gamma} (T' \Pi_T + R' \Pi_R + \psi' \pi_\psi). \end{aligned} \tag{9.7}$$

The old variable γ in Eqs. (9.7) is thought to be expressed through the new canonical variables by inverting the canonical transformation (9.6):

$$\gamma(r) = \int_0^r -\Pi_T(r) dr + \ln(R'^2(r) - T'^2(r)). \tag{9.8}$$

It is now straightforward to write down the Hamilton equations (3.8)–(3.11). The variational derivatives of $\mathcal{H}$ and $\mathcal{H}^1$ with respect to the new canonical variables are easily calculated, especially if we take into account that the initial value equations are satisfied so that all terms proportional to $\mathcal{H}$ and $\mathcal{H}^1$ can be put equal to zero. Because of this, we need not vary the exponentials in expressions (9.7). We write here only equation (3.8):

$$\begin{aligned} \dot{T} &= N R' e^{(1/2)(\psi-\gamma)} + N_1 T' e^{\psi-\gamma}, \\ \dot{R} &= N T' e^{(1/2)(\psi-\gamma)} + N_1 R' e^{\psi-\gamma}. \end{aligned} \tag{9.9}$$

Equations (9.9) are easily inverted to give the lapse and shift functions

$$N = (R'^2 - T'^2)^{-1}(\dot{T}R' - \dot{R}T')e^{(1/2)(\gamma-\psi)}$$
$$N_1 = (R'^2 - T'^2)^{-1}(\dot{R}R' - \dot{T}T')e^{\gamma-\psi}. \quad (9.10)$$

We can now return to the expressions (9.7) for super-Hamiltonian and supermomentum and solve the initial value equations (2.7) with respect to the C-energy density and C-flux,

$$-\Pi_T = P_0(r)[T,R;\psi,\pi_\psi]$$
$$= (R'^2 - T'^2)^{-1}(R'\tfrac{1}{2}R^{-1}\pi_\psi^2 + \tfrac{1}{2}R\psi'^2) - T'\psi'\pi_\psi),$$
$$-\Pi_R = P_1(r)[T,R;\psi,\pi_\psi]$$
$$= (R'^2 - T'^2)^{-1}(-T'(\tfrac{1}{2}R^{-1}\pi_\psi^2 + \tfrac{1}{2}R\psi'^2) + R'\psi'\pi_\psi). \quad (9.11)$$

At this point we are able to prove our version of the thin-sandwich conjecture for the cylindrical waves. If we specify freely the internal coordinates T,R, their rate of change $\dot{T},\dot{R}$, and the true dynamical variables ψ,π_ψ, we can calculate the lapse and shift functions from Eqs. (9.11), (9.8), and (9.10). In particular, if we use the Einstein–Rosen system of coordinates T,R, i.e., if we prescribe the internal trajectory by the equations

$$T = t, \quad R = r,$$

we get from Eqs. (9.10) the lapse and shift functions

$$N = e^{(1/2)(\gamma-\psi)}, \quad N_1 = 0.$$

In this way we have recovered the well-known space-time metric of the cylindrical waves from the internal path and the true dynamical variables.

Knowing the energy-momentum densities (9.11) as functions of dynamical variables, we can write down the bubble-time Hamilton equation (4.12) as

$$\frac{\delta\psi(r)}{\delta T(r^*)} = (R'^2 - T'^2)^{-1}(T'\pi_\psi - RR'\psi')\delta'(r,r^*),$$
$$\frac{\delta\psi(r)}{\delta R(r^*)} = -(R'^2 - T'^2)^{-1}(R'\pi_\psi - RT'\psi')\delta'(r,r^*),$$
$$\frac{\delta\pi_\psi(r)}{\delta T(r^*)} = (R'^2 - T'^2)^{-1}(R^{-1}R'\pi_\psi - T'\psi')\delta(r,r^*),$$
$$\frac{\delta\pi_\psi(r)}{\delta R(r^*)} = -(R'^2 - T'^2)^{-1}(R^{-1}T'\pi_\psi - R'\psi')\delta(r,r^*). \quad (9.12)$$

Due to the presence of the delta functions, either r or r^* can be used as arguments of the variables on the right-hand side of Eqs. (9.12). We see that Eqs. (9.12) are local; the change of the hypersurface around the point r^* does not affect the dynamical variables at a distant point $r \neq r^*$.

It is easy to check directly that the integrability conditions (5.3) are satisfied by the energy-momentum densities (9.11). It is also easy to check the second consistency condition, preferably in the form given by Eq. (6.6). We immediately see that the appropriate combination of the energy-momentum densities (9.11) gives

$$P_0T' + P_1R' = \pi_\psi\psi'.$$

Choosing now R as a special radial labeling of the hypersurfaces, we reduce the Hamilton equations (9.12) to equations containing a single bubble-time $T(R)$:

$$\frac{\delta\psi(R)}{\delta T(R^*)} = (1 - T_{,R}^2)^{-1}(T_{,R}\pi_\psi(R) - R\psi_{,R})\delta_{,R}(R,R^*),$$
$$\frac{\delta\pi_\psi(R)}{\delta T(R^*)} = (1 - T_{,R}^2)^{-1}(R^{-1}\pi_\psi - T_{,R}\psi_{,R})\delta(R,R^*). \quad (9.13)$$

It is equally straightforward to write down the Hamilton–Jacobi equations (8.6) or (8.9). Let us confine our attention to the reduced equation (8.9) for the Hamilton principal functional $S[T,\psi]$ of two function variables $T(R),\psi(R)$:

$$\frac{\delta S}{\delta T(R)} + \tfrac{1}{2}(1 - T_{,R}^2)^{-1}\left(R^{-1/2}\frac{\delta S}{\delta\psi(R)} - R^{1/2}T_{,R}\psi_{,R}\right)^2 + \tfrac{1}{2}R\psi_{,R}^2 = 0. \quad (9.14)$$

The Hamilton-Jacobi equation (9.14) can be obtained by a WKB approximation from a functional differential Schrödinger equation for the state functional $\Psi[T,\psi]$. The quantization of cylindrical gravitational waves starting from this Schrödinger equation has been extensively studied in Ref. 10.

ACKNOWLEDGMENTS

Helpful comments on a preliminary draft of this paper by Stanley Deser, Jiří Langer, John A. Wheeler, and especially James W. York are gratefully acknowledged. Thanks are also due to the referee of this paper who pointed out several connections with the earlier work of ADM.

---

* Work supported in part by the National Science Foundation Grant No. 30799X to Princeton University.

1 R. Arnowitt, S. Deser, and C. W. Misner, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), Chap. 7 and the series of earlier papers quoted here.
2 See, e.g., the forthcoming book *Gravitation* by C. W. Misner, K. S. Thorne, and J. A. Wheeler.
3 P. A. M. Dirac, Can. J. Math. **2**, 129 (1950); Proc. Roy. Soc. (London) **A246**, 326 (1958); **A246**, 333 (1958); Phys. Rev. **114**, 924 (1959).
4 B. S. DeWitt, Phys. Rev. **160**, 1113 (1967).
5 We introduce the following convention about the range of our indices: The Greek indices run from 0 to 3, the lower case Latin indices from 1 to 3, and the capital Latin indices from 1 to 2.
6 R. Arnowitt, S. Deser, and C. W. Misner, Phys. Rev. **117**, 1595 (1960), Sec. 2.
7 S. Deser, Ann. Inst. Henri Poincaré **7**, 149 (1967).
8 J. W. York, Phys. Rev. Letters **26**, 1656 (1971).
9 C. W. Misner, Phys. Rev. Letters **22**, 1071 (1969); Phys. Rev. **186**, 1319 (1969); **186**, 1328 (1969).
10 K. Kuchař, Phys. Rev. D **4**, 955 (1971).
11 Y. Nutku, in Proceedings of the Pittsburgh Conference on Relativity, July 13-17, 1970, to be published.
12 Standard notation is used in this paper: Comma denotes the partial derivatives and verticle stroke the covariant derivatives with respect to the spatial metric. Two indices enclosed by square brackets indicate antisymmetrization, $A_{\ldots[ik\ldots l]\ldots} = A_{\ldots ik\ldots l\ldots} - A_{\ldots[k\ldots i\ldots]}$. Lower case Latin indices are raised and lowered by the spatial metric tensor. The determinant of g_{ik} is denoted by g, the scalar curvature of a spacelike hypersurface by R.

[13] The actual form of $C^{ik}{}_{\lambda}$ is not important for the purpose of this paper. The Poisson brackets (2.8) are evaluated in Ref. 4.

[14] The notation $X^{\iota}(\mathbf{x})[g_{ik}, \pi^{ik}]$ indicates that X^{ι} is a function of $\mathbf{x}$ and simultaneously a functional of $g_{ik}(\mathbf{x})$ and $\pi^{ik}(\mathbf{x})$. E.g., the scalar potential Φ as a function of the charge density $\rho(\mathbf{x})$ would be written in this notation as $\Phi(\mathbf{x})[\rho] = \int d^3x'\, \rho(\mathbf{x}')/|\mathbf{x} - \mathbf{x}'|$.

[15] R. Arnowitt, S. Deser, and C. W. Misner, J. Math. Phys. 1, 434 (1960).

[16] When $\pi^{TT\,ik}$ and $g^{TT\,ik}$ are used for π_A and g^A, and the internal coordinates are frozen into preferred coordinates by imposing the ADM coordinate conditions, Eq. (6.7) goes over into Eq. (A4) of Ref. 15.

[17] A. Peres, Nuovo Cimento 26, 53 (1962).

[18] For a recent and thorough discussion of the Einstein–Hamilton–Jacobi equation, see U. Gerlach, Phys. Rev. 177, 1929 (1969).

[19] K. S. Thorne, Phys. Rev. 138, 251 (1965).

A Property of Zeros of the Partition Function for Ising Spin Systems

Elliott H. Lieb*
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
and
David Ruelle
Institut des Hautes Etudes Scientifiques, 91 Bures-sur-Yvette, France
(Received 22 November 1971)

Given an Ising antiferromagnet on a lattice with an AB substructure (bipartite lattice), one can consider the associated ferromagnet in which all the exchange constants are negated. Suppose the ferromagnet is above its critical temperature in the sense that there is an arc $(-\theta, \theta)$ of the unit circle on which the partition function has no zeros in $z = \exp(2\beta H)$. We prove that the original antiferromagnet partition function will have no zeros in z in the disc orthogonal to the unit disc and passing through the two end points of the arc. In other words, the antiferromagnet free energy is analytic in the magnetic field for small fields.

1. STATEMENT OF RESULTS

Let $\mathbb{Z}^\nu$ be the lattice of points with ν integer coordinates. We assume that a function ϵ on $\mathbb{Z}^\nu$ with values ± 1 is given such that

(a) ϵ is not identically $+1$,

(b) $\epsilon(x)\epsilon(y) = \epsilon(x+y)$.

We can then say that the point x is *even*, resp. *odd*, if $\epsilon(x) = +1$, resp. -1.

A spin variable with two possible values $\sigma_x = \pm 1$ is associated with each lattice site $x \in \mathbb{Z}^\nu$ and we introduce a translation-invariant pair interaction J such that $\sum_{x \in \mathbb{Z}^\nu} |J(x)| < +\infty$. If $\Lambda = \{x_1, \ldots, x_m\}$ is a finite subset of $\mathbb{Z}^\nu$, the energy of the spin configurtion $\sigma = (\sigma_{x_1}, \ldots, \sigma_{x_m})$ is

$$E(\sigma) = \tfrac{1}{2} \sum_{x \in \Lambda} \sum_{y \in \Lambda} J(x-y)(1 - \sigma_x\sigma_y) - H \sum_{x \in \Lambda} \sigma_x,$$

where H is the magnetic field. The free energy per site at temperature β^{-1} is then

$$\psi_J(\beta, H) = -\lim_{\Lambda \to \infty} \beta^{-1} |\Lambda|^{-1} \log \sum_{\sigma} \exp[-\beta E(\sigma)],$$

where $|\Lambda|$ is the number of points in Λ and $\Lambda \to \infty$ may be taken to mean that Λ is a rectangular box with all sides tending to infinity.

Theorem 1: (a) Let J be ferromagnetic, i.e., $J \geq 0$, and let β^{-1} be above the critical temperature in the sense that some neighborhood of $H = 0$ is free of zeros of the partition function $\sum_\sigma \exp[-\beta E(\sigma)]$. Then some neighborhood of $H = 0$ is also free of zeros of the partition function for the interaction ϵJ. In particular $\psi_{\epsilon J}(\beta, H)$ is analytic around $H = 0$ [this can be applied to antiferromagnetic nearest neighbor (Ising) interactions because they are of the form ϵJ].

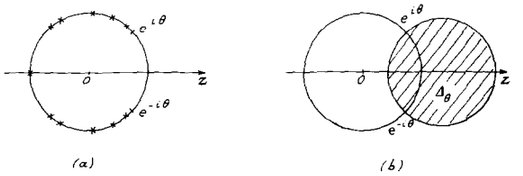


FIG. 1. (a) Interaction J: the crosses are zeros of the partition function. (b) Interaction ϵJ: the cross-hatched region Δ_θ is free of zeros of the partition function.

(b) More precisely, let $z = e^{2\beta H}$. We shall show that if the partition function for the interaction J does not vanish when $|z| = 1$, $|\arg z| < \theta$, then the partition function for the interaction ϵJ does not vanish when $z \in \Delta_\theta$. Here Δ_θ is the open region containing the point 1, and bounded by the circle orthogonal to $\{z : |z| = 1\}$ going through $e^{\pm i\theta}$ (see Fig. 1). In particular, $\psi_{\epsilon J}$ is analytic in Δ_θ.

The proof uses the theory of analytic functions of several complex variables, and is accomplished in two steps. The first step (Sec. 2) is an application of Borchers' double cone theorem.[1] The second step (Sec. 3) is the computation of a holomorphy envelope (a limiting case of the holomorphy envelope of two polydiscs; see Ref. 2).

At the end of Sec. 3, some extensions of the above theorem are indicated.

2. PROOF OF PART (a) OF THEOREM 1

Let us fix $\beta > 0$ and write

$$F_J^\Lambda(H_1, H_2) = -\beta^{-1} |\Lambda|^{-1} \log \sum_{\sigma} \exp[-\beta E^*(\sigma)],$$

$$E^*(\sigma) = \tfrac{1}{2} \sum_{x \in \Lambda} \sum_{y \in \Lambda} J(x-y)(1 - \sigma_x\sigma_y) - \sum_{x \in \Lambda} \{\tfrac{1}{2}[1 + \epsilon(x)]H_1\sigma_x + \tfrac{1}{2}[1 - \epsilon(x)]H_2\sigma_x\}.$$

It is easily seen that

$$\psi_J(\beta, H) = \lim_{\Lambda \to \infty} F_J^\Lambda(H, H),$$

$$\psi_{\epsilon J}(\beta, H) = -\beta^{-1} \sum_{x \in \mathbb{Z}^\nu} \tfrac{1}{2}[\epsilon(x) - 1)]\, J(x) + \lim_{\Lambda \to \infty} F_J^\Lambda(H; -H).$$

We shall now make use of a theorem of Borchers.[1,3]

[13] The actual form of $C^{ik}{}_{\lambda}$ is not important for the purpose of this paper. The Poisson brackets (2.8) are evaluated in Ref. 4.

[14] The notation $X^{\iota}(\mathbf{x})[g_{ik}, \pi^{ik}]$ indicates that X^{ι} is a function of $\mathbf{x}$ and simultaneously a functional of $g_{ik}(\mathbf{x})$ and $\pi^{ik}(\mathbf{x})$. E.g., the scalar potential Φ as a function of the charge density $\rho(\mathbf{x})$ would be written in this notation as $\Phi(\mathbf{x})[\rho] = \int d^3x'\, \rho(\mathbf{x}')/|\mathbf{x} - \mathbf{x}'|$.

[15] R. Arnowitt, S. Deser, and C. W. Misner, J. Math. Phys. 1, 434 (1960).

[16] When $\pi^{TT\,ik}$ and $g^{TT\,ik}$ are used for π_A and g^A, and the internal coordinates are frozen into preferred coordinates by imposing the ADM coordinate conditions, Eq. (6.7) goes over into Eq. (A4) of Ref. 15.

[17] A. Peres, Nuovo Cimento 26, 53 (1962).

[18] For a recent and thorough discussion of the Einstein–Hamilton–Jacobi equation, see U. Gerlach, Phys. Rev. 177, 1929 (1969).

[19] K. S. Thorne, Phys. Rev. 138, 251 (1965).

A Property of Zeros of the Partition Function for Ising Spin Systems

Elliott H. Lieb*
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
and
David Ruelle
Institut des Hautes Etudes Scientifiques, 91 Bures-sur-Yvette, France
(Received 22 November 1971)

Given an Ising antiferromagnet on a lattice with an AB substructure (bipartite lattice), one can consider the associated ferromagnet in which all the exchange constants are negated. Suppose the ferromagnet is above its critical temperature in the sense that there is an arc $(-\theta, \theta)$ of the unit circle on which the partition function has no zeros in $z = \exp(2\beta H)$. We prove that the original antiferromagnet partition function will have no zeros in z in the disc orthogonal to the unit disc and passing through the two end points of the arc. In other words, the antiferromagnet free energy is analytic in the magnetic field for small fields.

1. STATEMENT OF RESULTS

Let $\mathbb{Z}^\nu$ be the lattice of points with ν integer coordinates. We assume that a function ϵ on $\mathbb{Z}^\nu$ with values ± 1 is given such that

(a) ϵ is not identically $+1$,

(b) $\epsilon(x)\epsilon(y) = \epsilon(x+y)$.

We can then say that the point x is *even*, resp. *odd*, if $\epsilon(x) = +1$, resp. -1.

A spin variable with two possible values $\sigma_x = \pm 1$ is associated with each lattice site $x \in \mathbb{Z}^\nu$ and we introduce a translation-invariant pair interaction J such that $\sum_{x \in \mathbb{Z}^\nu} |J(x)| < +\infty$. If $\Lambda = \{x_1, \ldots, x_m\}$ is a finite subset of $\mathbb{Z}^\nu$, the energy of the spin configurtion $\sigma = (\sigma_{x_1}, \ldots, \sigma_{x_m})$ is

$$E(\sigma) = \tfrac{1}{2} \sum_{x \in \Lambda} \sum_{y \in \Lambda} J(x-y)(1 - \sigma_x\sigma_y) - H \sum_{x \in \Lambda} \sigma_x,$$

where H is the magnetic field. The free energy per site at temperature β^{-1} is then

$$\psi_J(\beta, H) = -\lim_{\Lambda \to \infty} \beta^{-1} |\Lambda|^{-1} \log \sum_{\sigma} \exp[-\beta E(\sigma)],$$

where $|\Lambda|$ is the number of points in Λ and $\Lambda \to \infty$ may be taken to mean that Λ is a rectangular box with all sides tending to infinity.

Theorem 1: (a) Let J be ferromagnetic, i.e., $J \geq 0$, and let β^{-1} be above the critical temperature in the sense that some neighborhood of $H = 0$ is free of zeros of the partition function $\sum_\sigma \exp[-\beta E(\sigma)]$. Then some neighborhood of $H = 0$ is also free of zeros of the partition function for the interaction ϵJ. In particular $\psi_{\epsilon J}(\beta, H)$ is analytic around $H = 0$ [this can be applied to antiferromagnetic nearest neighbor (Ising) interactions because they are of the form ϵJ].

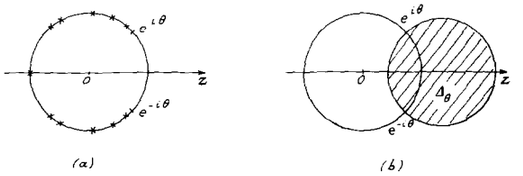


FIG. 1. (a) Interaction J: the crosses are zeros of the partition function. (b) Interaction ϵJ: the cross-hatched region Δ_θ is free of zeros of the partition function.

(b) More precisely, let $z = e^{2\beta H}$. We shall show that if the partition function for the interaction J does not vanish when $|z| = 1$, $|\arg z| < \theta$, then the partition function for the interaction ϵJ does not vanish when $z \in \Delta_\theta$. Here Δ_θ is the open region containing the point 1, and bounded by the circle orthogonal to $\{z : |z| = 1\}$ going through $e^{\pm i\theta}$ (see Fig. 1). In particular, $\psi_{\epsilon J}$ is analytic in Δ_θ.

The proof uses the theory of analytic functions of several complex variables, and is accomplished in two steps. The first step (Sec. 2) is an application of Borchers' double cone theorem.[1] The second step (Sec. 3) is the computation of a holomorphy envelope (a limiting case of the holomorphy envelope of two polydiscs; see Ref. 2).

At the end of Sec. 3, some extensions of the above theorem are indicated.

2. PROOF OF PART (a) OF THEOREM 1

Let us fix $\beta > 0$ and write

$$F_J^\Lambda(H_1, H_2) = -\beta^{-1} |\Lambda|^{-1} \log \sum_{\sigma} \exp[-\beta E^*(\sigma)],$$

$$E^*(\sigma) = \tfrac{1}{2} \sum_{x \in \Lambda} \sum_{y \in \Lambda} J(x-y)(1 - \sigma_x\sigma_y) - \sum_{x \in \Lambda} \{\tfrac{1}{2}[1 + \epsilon(x)]H_1\sigma_x + \tfrac{1}{2}[1 - \epsilon(x)]H_2\sigma_x\}.$$

It is easily seen that

$$\psi_J(\beta, H) = \lim_{\Lambda \to \infty} F_J^\Lambda(H, H),$$

$$\psi_{\epsilon J}(\beta, H) = -\beta^{-1} \sum_{x \in \mathbb{Z}^\nu} \tfrac{1}{2}[\epsilon(x) - 1)]\, J(x) + \lim_{\Lambda \to \infty} F_J^\Lambda(H; -H).$$

We shall now make use of a theorem of Borchers.[1,3]

Theorem 2 (double cone theorem): Let C be an open convex cone in $\mathbb{R}^n$, with apex at the origin. Let the function F be defined and analytic in a domain $D \subset \mathbb{C}^n = \mathbb{R}^n + i\mathbb{R}^n$ such that

(i) $D \supset (C + i\mathbb{R}^n) \cup (-C + i\mathbb{R}^n)$,

(ii) D contains the points $i[\alpha x + (1-\alpha)y]$, where x, $y \in \mathbb{R}^n$, $x - y \in C$ and α runs between 0 and 1.

Then F is analytic in a fixed complex neighborhood of the set $i[(x - C) \cap (y + C)]$

We apply this theorem to the function $F_J^\Lambda(H_1, H_2)$. Here $n = 2$ and C consists of the points with coordinates > 0 in $\mathbb{R}^2$. Thus (i) means that F_J^Λ is analytic for $\mathrm{Re}\, H_1 > 0$ and $\mathrm{Re}\, H_2 > 0$, or $\mathrm{Re}\, H_1 < 0$ and $\mathrm{Re}\, H_2 < 0$. Condition (i) is satisfied by F_J^Λ as a consequence of the Lee–Yang circle theorem.[4] By assumption, (ii) is satisfied with

$$x = (K, K), \quad y = (-K, -K)$$

where $2\beta K = \theta$ and θ is defined in part (b) of Theorem 1. Therefore, F_Λ^J is analytic in a fixed complex neighborhood of the set

$$\{(H_1, H_2) \in i\mathbb{R}^2 : |H_1| < K, \ |H_2| < K\}.$$

In particular, when we take $H_1 = -H_2 = H$ and let $\Lambda \to \infty$, we obtain the result announced in part (a) of Theorem 1.

3. CALCULATION OF THE DOMAIN OF HOLOMORPHY

Consider a finite Ising spin system composed of N sites and a collection of $\binom{N}{2}$ interaction constants $\{J_{ij}\}$ which are nonnegative. (Previously we assumed the system was translation invariant, but for this section there is no need to do so.) To each site we assign a magnetic field H_i, $i = 1, \ldots, N$. As was explained before, by letting $H_i = H$ for i an even site and $H_i = -H$ for i an odd site the system is, in fact, equivalent to an Ising antiferromagnet. In this section we shall allow the $\{H_i\}$ to be arbitrary. The advantage of doing so is that we can thereby consider a more general system in which the spin magnetic moment varies from site to site.

It is convenient to work with the "activity" variables

$$z_i = \exp(2\beta H_i), \quad i = 1, \ldots, N, \tag{3.1}$$

so that the partitions function can be expressed as

$$Z(H_1, \ldots, H_N) = \exp\left(-\beta \sum_{i=1}^{N} H_i\right) P(z), \tag{3.2}$$

where P is a polynomial and $z = (z_1, \ldots, z_N)$ is regarded as a point in $\mathbb{C}^N$.

Notation: (a) D denotes the open unit disc in $\mathbb{C}$:

$$D = \{z \in \mathbb{C} : |z| < 1\}.$$

(b) D^N denotes the symmetric unit polydisc in $\mathbb{C}^N$:

$$D^N = \{z \in \mathbb{C}^N : |z_i| < 1, \ i = 1, \ldots, N\}.$$

(c) E^N denotes $[\mathrm{Int}\,(\sim D)]^N$:

$$E^N = \{z \in \mathbb{C}^N : |z_i| > 1, \ i = 1, \ldots, N\}$$

(d) For $0 < \theta < \pi$, $A_\theta \subset \mathbb{C}$ is the open arc:

$$A_\theta = \{z \in \mathbb{C} : z = e^{i\psi}, \ -\theta < \psi < \theta\}$$

(e) $B_\theta = \partial D \backslash A_\theta$ is the (closed) arc complementary to A_θ.

(f) $A_\theta^N \subset \mathbb{C}^N$ is

$$A_\theta^N = \{z \in \mathbb{C}^N : z_i \in A_\theta, \ i = 1, \ldots, N\}.$$

Given that $P(z)$ has no zeros in $D^N \cup E^N$ (the Lee-Yang theorem) and given that, for $z \in \mathbb{C}$, $P(z, z, \ldots, z)$ has no zeros in A_θ, we have established previously that there exists a complex neighborhood O of A_θ^N in which $P(z)$ is free of zeros. In this section we address ourselves to obtaining a minimal estimate of O by calculating the envelope of holomorphy of $D_N \cup E_N \cup O$. As will be seen, this envelope is, in fact, a domain in $\mathbb{C}^N$. In this domain, $f(z) = 1/P(z)$ is holomorphic and hence $P(z)$ has no zeros.

For $z \in \mathbb{C}$ we change variables $z \to \psi$ as follows:

$$\exp(\psi + i\theta) = -(z - e^{i\theta})/(z - e^{-i\theta}) \tag{3.3}$$

so that

$$\psi = i(\pi - \theta) + \ln[(z - e^{i\theta})/(z - e^{-i\theta})] \tag{3.4}$$

with the logarithm defined to be holomorphic in $\mathbb{C}\backslash B_\theta$ and $\to 1$ as $z \to \infty$.

Alternatively,

$$z = \cosh \tfrac{1}{2}(\psi - i\theta)/\cosh \tfrac{1}{2}(\psi + i\theta) \tag{3.5}$$

so that $\psi \to -\psi$ is equivalent to $z \to 1/z$.

The mapping (3.3) maps $\mathbb{C}\backslash B_\theta$ conformally onto

$$K_\theta = \{\psi = s + it : -\infty < s < \infty, \ -\pi < t < \pi\}\backslash\{i(\pi - \theta)\}. \tag{3.6}$$

To describe the mapping, we refer to Figs. 2 and 3. If we fix $t \neq \pi - \theta$, then as s goes from $-\infty$ to ∞ a curve is traced out in the z plane. This curve is an arc of a circle passing through the points $e^{i\theta}$ (corresponding to $s = -\infty$) and $e^{-i\theta}$ (corresponding to $s = \infty$). The line segment $L = [e^{i\theta}, e^{-i\theta}]$ corresponds to $t = -\theta$. Arcs to the right of this segment correspond to $-\theta < t < \pi - \theta$. In particular, $t = 0$ corresponds to A_θ. If $-\pi < t < -\theta$, the arcs are to the left of L and inside D with the boundary of B_θ being reached as $t \downarrow -\pi$. If $\pi - \theta < t < \pi$, the arcs are to the left of L and outside D with B_θ being reached as $t \uparrow \pi$. The exceptional case is $t = \pi - \theta$, where $s \in \{-\infty, 0\}$ corresponds to the vertical line $\{e^{i\theta}, e^{i\theta} + i\infty\}$ while $s \in \{0, \infty\}$ corresponds to the vertical line $\{e^{-i\theta} - i\infty, e^{-i\theta}\}$. A complete circle through the points $e^{i\theta}$ and $e^{-i\theta}$ is always composed of two arcs whose t values, t and t', satisfy $|t - t'| = \pi$. If a value of t (and hence an arc) is fixed, then for every point on the arc the angle subtended by the line segment L is $|\pi - |t + \theta||$. Among the circles through $\{e^{i\theta}, e^{-i\theta}\}$ there is a particularly important one to which we shall return later, namely the circle orthogonal to ∂D. It has the property that it is invariant under inversion: $z \to 1/z$. The arc of this circle inside D corresponds to $t = -\pi/2$ while the arc outside D corresponds to $t = \pi/2$.

Except when $\theta = \pi/2$, this circle has a radius $|\tan\theta|$ and its center is at $z = (1/\cos\theta, 0)$. When $\theta = \pi/2$ the "circle" is the imaginary axis.

Returning to $\mathbb{C}^N$, we apply the conformal map (3.3) to each z_i, i.e., $z_i \to \psi_i$, $i = 1, \ldots, N$, so that $1/P(z) = f(z) \to h(\psi)$ with $\psi = (\psi_1, \ldots, \psi_N)$. D^N is mapped onto

$$\delta^N = \{\psi : \psi_j = s_j + it_j,\ -\infty < s_j < \infty,\ -\pi < t_j < 0,\ j = 1, \ldots, N\}$$

while E^N is mapped onto $\epsilon^N = \mu^N \backslash \gamma^N$, where

$$\mu^N = \{\psi : \psi_j = s_j + it_j,\ -\infty < s_j < \infty,\ 0 < t_j < \pi,\ j = 1, \ldots, N\}$$

and

$$\gamma^N = \{\psi : \psi_j = s_j + it_j,\ -\infty < s_j < \infty,\ 0 < t_j < \pi,\ j = 1, \ldots, N$$

and $\psi_j = i(\pi - \theta)$ for some $j\}$.

The set A_θ^N is mapped onto

$$\alpha^N = \{\psi : \psi_j = s_j + it_j,\ -\infty < s_j < \infty,\ t_j = 0,\ j = 1, \ldots, N\}.$$

Since each z_i is a holomorphic function of ψ_i in K_θ, $h(\psi)$ is holomorphic in $\delta^N \cup \epsilon^N \cup \Omega^N$, where Ω^N is some complex neighborhood of α_θ^N. This is a consequence of the double cone theorem stated in Sec. 2.

The next step is to establish analyticity of $h(\psi)$ in some neighborhood of γ^N so that we can replace ϵ^N by μ^N. Let G^N be the inverse image of γ^N in $\hat{\mathbb{C}}^N$ so that $z \in G^N$ means that some z_j is infinite. The mere fact that P is a polynomial does not guarantee that $1/P$ is holomorphic in a neighborhood of G^N.

[Remark: The fact that P comes from an Ising ferromagnetic spin system does guarantee this analyticity, however, because the coefficient Q of $\prod_{j=1}^n z_i$ (for example) in P, when considered as a polynomial in $z_{n+1}, \ldots, z_N$, is nonzero in E^{N-n}. This is so because Q is the polynomial of an Ising ferromagnetic spin system of $N - n$ spins which interact with n other spins fixed in the + 1 direction. Hence $1/P = 0$ on G^N.]

To remedy the lack of analyticity in the general case, we can replace $f(z)$ by $F(z) = f(z)\,\gamma(z)$, where $\gamma(z) = \prod_{j=1}^N \rho(z_j)$ and $\rho(z) = \sin^M\theta[z^2 + 1 - 2z\cos\theta]^{-M/2}$ for a sufficiently large positive integer M. Clearly $\rho(z)$ is holomorphic away from the cut B_θ *and is never zero* except at $z = \infty$, so that F has the same analyticity properties as f. However, $F = 0$ on G^N and is analytic in some neighborhood of G^N. In terms of ψ, the equivalent substitution is

$$h(\psi) \to H(\psi) = h(\psi) \prod_{j=1}^N \cosh\tfrac{1}{2}(\psi + i\theta)^M.$$

Hence, H is holomorphic on some neighborhood of the tube $T = \delta^N \cup \mu^N \cup \alpha^N$. Since any neighborhood of T contains a connected component which contains $\delta^N \cup \mu^N$, we can use the tube theorem[5] to assert that H is also holomorphic on the convex hull of T which we shall call T'. It is easy to see that

$$T' = \bigcup_{0 \le a \le \pi} T_a,$$

where T_a is the tube with imaginary base $\{(t_1, \ldots, t_N) : -\pi + a < t_j < a, j = 1, \ldots, N\}$. I.e., T' is the union of translates of δ^N (or μ^N). The inverse image of T_a is a symmetric domain $D_a^N \subset \mathbb{C}^N$ where $\partial D_a \subset \mathbb{C}$ is a circle through the points $e^{i\theta}$ and $e^{-i\theta}$. As a goes from 0 to π, all circles are reached; however, D_a is either the disc interior to the circle or else the exterior of the circle and is chosen such that D_a contains A_θ.

A more geometric interpretation of our conclusion is this:

Theorem 3: Let $P(z_1, \ldots, z_N)$ be the polynomial of a ferromagnetic Ising spin system of N spins and suppose that $P(z, \ldots, z) \neq 0$ when $z \in A_\theta = \{e^{i\varphi} : -\theta < \varphi < \theta\}$ for some $\theta \in (0, \pi)$. Let C be any circle in $\mathbb{C}$ that includes the points $e^{i\theta}$ and $e^{-i\theta}$ and let Δ be the interior or exterior or C chosen such that $\Delta \supset A_\theta$. Then $P(z_1, \ldots, z_N) \neq 0$ if $z_i \in \Delta$ for $i = 1, \ldots, N$.

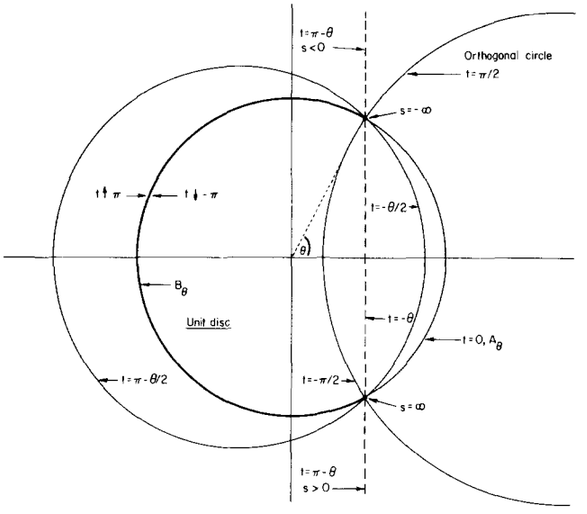


FIG. 2. The unit disc and several circles through the points $e^{i\theta}$ and $e^{-i\theta}$ are shown for the case $\theta < \pi/2$. If $\psi = s + it$, each arc corresponds to some fixed t value (as shown) and $-\infty < s < \infty$. The Ising antiferromagnetic partition function has no zeros in the *interior* of the orthogonal circle.

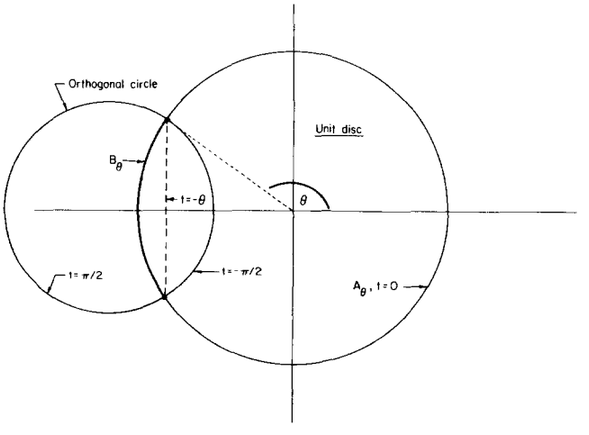


FIG. 3. The same as Fig. 2 except that $\theta > \pi/2$. In this case the antiferromagnetic partition function has no zeros in the exterior of the orthogonal circle.

To apply Theorem 3 to the antiferromagnet considered in the previous section, we must set $z_i = z = \exp(2\beta H)$ for i an even site and $z_i = 1/z$ for i an odd site. Let A_θ be the arc in which the ferromagnetic partition function has no zeros and let C_θ be the circle through the points $e^{i\theta}$ and $e^{-i\theta}$ that is *orthogonal* to the unit circle. Let $\triangle_\theta$ be the *interior* of C_θ when $0 < \theta < \pi/2$ and the *exterior* of C_θ when $\pi/2 < \theta < \pi$. When $\theta = \pi/2$, $\triangle_{\pi/2}$ is the right-hand complex plane, $\triangle_{\pi/2} = \{z : \mathrm{Re}(z) > 0\}$. Then the antiferromagnetic partition function is nonzero when $z \in \triangle_\theta$.

Remarks: (i) Theorem 3 can be applied to any system that has the Lee–Yang property such as the Heisenberg ferromagnet. It is only for the Ising model, however, that the replacement of z by $1/z$ on odd sites is equivalent to the replacement of the ferromagnet by an antiferromagnet. In particular, Theorem 3 can be applied to an Ising model of arbitrary spin, not just spin $\frac{1}{2}$.

(ii) Instead of using the variable $z = \exp(2\beta H)$ one could have used $\xi = \exp(\beta H)$ so that the angle θ would be replaced by $\theta/2$. The new domain $\triangle'$ would be smaller in the sense that $(\triangle')^2 \subset \triangle$. In other words, the fact that the polynomial is a function of $z = \xi^2$ is a stronger hypothesis and naturally leads to a larger domain in which there are no zeros.

4. EXTENSIONS

Consider an assembly of spins with ferromagnetic pair interaction J in a magnetic field H, and assume that the grand partition function has no zeros for $\mathrm{Re}\, H = 0$, $|\mathrm{Im}\, H| < K$. Suppose now that different magnetic fields $H_1, \ldots, H_n$ act on different classes of spins, and let $F(H_1, \ldots, H_n)$ be the corresponding free energy. We deduce as in Sec. 2 that F is analytic in a complex neighborhood of the set

$$\{(H_1, \ldots, H_n) \in i\mathbb{R}^n : |H_1| < K, \ldots, |H_n| < K\}.$$

Let $z_i = e^{2\beta H_i}$ and $f(z_1, \ldots, z_n) = F(H_1, \ldots, H_n)$. The argument in Sec. 3 shows that f is analytic when all z_i are in the same D_a, D_a being any open region containing the point 1 and bounded by a circle through the points $e^{\pm 2i\beta K}$.

The free energy for various (not all!) spin systems with interactions of antiferromagnetic or other type and magnetic moment varying from site to site may be written (up to constants) as

$$F(\epsilon_1 H, \ldots, \epsilon_n H) = f(z^{\epsilon_1}, \ldots, z^{\epsilon_n}),$$

where $\epsilon_1, \ldots, \epsilon_n$ are some real numbers. The above results prove analyticity in a neighborhood of $z = 1$, more precisely for all z such that

$$1, z^{\epsilon_1}, \ldots, z^{\epsilon_n}$$

are all inside (or outside) the same circle through $e^{\pm 2i\beta K}$. However, if the region of analyticity for z contains a piece of the negative real axis, it will in general be necessary to exclude it (introduce a cut). This is so because, if ϵ_i is not a positive integer, the functions z^{ϵ_i} are not holomorphic at 0 and ∞. Even if all ϵ_i are integers, it is still necessary to introduce a cut from -1 to $-\infty$.

ACKNOWLEDGMENTS

This paper was written while the authors participated in the 1971 Battelle "Recontre" at the Battelle Seattle Research Center. We are grateful for the hospitality shown us. E. L. would like to thank R. Stora for introducing him to the mapping (3.3) and its relevance to the tube theorem. The same mapping has been used[2] to find the envelope of holomorphy of the union of two polydiscs.

---

* Work supported by National Science Foundation Grant GP 26526.

1 H. J. Borchers, Nuovo Cimento **19**, 787 (1961).
2 V. Glaser and J. Bros, "L'envelope d'holomorphie de l'union de deux polycercles," CERN preprint (1961).
3 H. Epstein, thesis (to appear).
4 T. D. Lee and C. N. Yang, Phys. Rev. **87**, 410 (1952).
5 S. Bochner and W. T. Martin, *Several Complex Variables* (Princeton U.P., Princeton, N. J., 1948).

Path Integrals, Asymptotics, and Singular Perturbations*

David W. McLaughlin
Department of Mathematics, University Heights Campus,
and
Courant Institute of Mathematical Sciences, New York University, New York, New York 10012
(Received 6 August 1971; Revised Manuscript Received 17 January 1972)

In this paper we evaluate several Feynman path integrals asymptotically with respect to various parameters in order to gain mathematical insight into the asymptotic evaluation of function space integrals with oscillatory integrands, a type of integral which is beginning to appear in areas of physics other than quantum mechanics. In each integral studied, the integrand factors into a product of two functionals, one of which is dominant in the limit under consideration. By systematically exploiting this feature, we obtain the asymptotic behavior of path integrals for the physical situations of (1) weakly complex potentials, (2) high energy and complex potentials, (3) weak real potentials, and (4) strong real potentials. In the complex cases, the techniques indicate a means to handle (complex valued) turning points. In the sections treating strong and weak potentials, we relate the relative ease with which one may exploit the factorization of the integrand to the theories of regular and singular perturbations. In the singular case, several examples are presented, one of which is a high energy evaluation of the path integral associated with the "Langer transformed" radial equation. Finally, using more conventional techniques, we construct the complete asymptotic series for each case, thus formally establishing that we have obtained the leading term in an asymptotic expansion of the path integral.

INTRODUCTION

When Feynman introduced the path integral in 1948,[1] he did so primarily for calculational purposes. He sought a representation of the propagator of quantum mechanics from which one could obtain quantum properties in much the same way as one can derive from

To apply Theorem 3 to the antiferromagnet considered in the previous section, we must set $z_i = z = \exp(2\beta H)$ for i an even site and $z_i = 1/z$ for i an odd site. Let A_θ be the arc in which the ferromagnetic partition function has no zeros and let C_θ be the circle through the points $e^{i\theta}$ and $e^{-i\theta}$ that is *orthogonal* to the unit circle. Let $\triangle_\theta$ be the *interior* of C_θ when $0 < \theta < \pi/2$ and the *exterior* of C_θ when $\pi/2 < \theta < \pi$. When $\theta = \pi/2$, $\triangle_{\pi/2}$ is the right-hand complex plane, $\triangle_{\pi/2} = \{z : \mathrm{Re}(z) > 0\}$. Then the antiferromagnetic partition function is nonzero when $z \in \triangle_\theta$.

Remarks: (i) Theorem 3 can be applied to any system that has the Lee–Yang property such as the Heisenberg ferromagnet. It is only for the Ising model, however, that the replacement of z by $1/z$ on odd sites is equivalent to the replacement of the ferromagnet by an antiferromagnet. In particular, Theorem 3 can be applied to an Ising model of arbitrary spin, not just spin $\frac{1}{2}$.

(ii) Instead of using the variable $z = \exp(2\beta H)$ one could have used $\xi = \exp(\beta H)$ so that the angle θ would be replaced by $\theta/2$. The new domain $\triangle'$ would be smaller in the sense that $(\triangle')^2 \subset \triangle$. In other words, the fact that the polynomial is a function of $z = \xi^2$ is a stronger hypothesis and naturally leads to a larger domain in which there are no zeros.

4. EXTENSIONS

Consider an assembly of spins with ferromagnetic pair interaction J in a magnetic field H, and assume that the grand partition function has no zeros for $\mathrm{Re}\, H = 0$, $|\mathrm{Im}\, H| < K$. Suppose now that different magnetic fields $H_1, \ldots, H_n$ act on different classes of spins, and let $F(H_1, \ldots, H_n)$ be the corresponding free energy. We deduce as in Sec. 2 that F is analytic in a complex neighborhood of the set

$$\{(H_1, \ldots, H_n) \in i\mathbb{R}^n : |H_1| < K, \ldots, |H_n| < K\}.$$

Let $z_i = e^{2\beta H_i}$ and $f(z_1, \ldots, z_n) = F(H_1, \ldots, H_n)$. The argument in Sec. 3 shows that f is analytic when all z_i are in the same D_a, D_a being any open region containing the point 1 and bounded by a circle through the points $e^{\pm 2i\beta K}$.

The free energy for various (not all!) spin systems with interactions of antiferromagnetic or other type and magnetic moment varying from site to site may be written (up to constants) as

$$F(\epsilon_1 H, \ldots, \epsilon_n H) = f(z^{\epsilon_1}, \ldots, z^{\epsilon_n}),$$

where $\epsilon_1, \ldots, \epsilon_n$ are some real numbers. The above results prove analyticity in a neighborhood of $z = 1$, more precisely for all z such that

$$1, z^{\epsilon_1}, \ldots, z^{\epsilon_n}$$

are all inside (or outside) the same circle through $e^{\pm 2i\beta K}$. However, if the region of analyticity for z contains a piece of the negative real axis, it will in general be necessary to exclude it (introduce a cut). This is so because, if ϵ_i is not a positive integer, the functions z^{ϵ_i} are not holomorphic at 0 and ∞. Even if all ϵ_i are integers, it is still necessary to introduce a cut from -1 to $-\infty$.

ACKNOWLEDGMENTS

This paper was written while the authors participated in the 1971 Battelle "Recontre" at the Battelle Seattle Research Center. We are grateful for the hospitality shown us. E. L. would like to thank R. Stora for introducing him to the mapping (3.3) and its relevance to the tube theorem. The same mapping has been used[2] to find the envelope of holomorphy of the union of two polydiscs.

---

* Work supported by National Science Foundation Grant GP 26526.

1 H. J. Borchers, Nuovo Cimento **19**, 787 (1961).
2 V. Glaser and J. Bros, "L'envelope d'holomorphie de l'union de deux polycercles," CERN preprint (1961).
3 H. Epstein, thesis (to appear).
4 T. D. Lee and C. N. Yang, Phys. Rev. **87**, 410 (1952).
5 S. Bochner and W. T. Martin, *Several Complex Variables* (Princeton U.P., Princeton, N. J., 1948).

Path Integrals, Asymptotics, and Singular Perturbations*

David W. McLaughlin
Department of Mathematics, University Heights Campus,
and
Courant Institute of Mathematical Sciences, New York University, New York, New York 10012
(Received 6 August 1971; Revised Manuscript Received 17 January 1972)

In this paper we evaluate several Feynman path integrals asymptotically with respect to various parameters in order to gain mathematical insight into the asymptotic evaluation of function space integrals with oscillatory integrands, a type of integral which is beginning to appear in areas of physics other than quantum mechanics. In each integral studied, the integrand factors into a product of two functionals, one of which is dominant in the limit under consideration. By systematically exploiting this feature, we obtain the asymptotic behavior of path integrals for the physical situations of (1) weakly complex potentials, (2) high energy and complex potentials, (3) weak real potentials, and (4) strong real potentials. In the complex cases, the techniques indicate a means to handle (complex valued) turning points. In the sections treating strong and weak potentials, we relate the relative ease with which one may exploit the factorization of the integrand to the theories of regular and singular perturbations. In the singular case, several examples are presented, one of which is a high energy evaluation of the path integral associated with the "Langer transformed" radial equation. Finally, using more conventional techniques, we construct the complete asymptotic series for each case, thus formally establishing that we have obtained the leading term in an asymptotic expansion of the path integral.

INTRODUCTION

When Feynman introduced the path integral in 1948,[1] he did so primarily for calculational purposes. He sought a representation of the propagator of quantum mechanics from which one could obtain quantum properties in much the same way as one can derive from

the various integral representations of the special functions many of their properties. He successfully applied his representation in various perturbation[2] and variational[3] calculations. Since then, various authors (Morette,[4] Gutzwiller,[5] Pechukas,[6] Gel'fand and Yaglom,[7] Buslaev,[8] to mention a few) have realized that the path integral is particularly suited for small $\hbar$ or semiclassical approximations. To date, these approaches have been disappointing in that few unknown facts about quantum mechanics have been learned from them. However, there are other areas of physics in which function space integrals with oscillatory integrands occur (for example, in the study of scattering from a random medium[9]). In these areas the meaning of various approximations of the underlying partial differential equations is not as well understood as is the Schrödinger equation, and it is hoped that the function space integral will offer some insight into the various approximations. Such considerations lead us to believe that once methods to approximate these integrals over function space are better understood, new physical results will follow. Therefore, at this point it seems mathematically appropriate to perform asymptotic evaluations of several Feynman path integrals which have not been treated previously, but where the results still can be compared with those of more conventional techniques.

While asymptotic evaluations of path integrals are necessarily formal, when properly understood and carefully presented, they are very convincing. To be so, however, the particular limit under consideration must be precisely stated and the calculations systematically performed. In this study, the primary purpose is to enhance the understanding of the mathematical nature of this function space integral through systematically performing several asymptotic evaluations. We have attempted to clearly present all necessary formal arguments. In particular, we asymptotically evaluate the path integral for several classes of integrands, each class representing a different physical situation (complex potentials with small imaginary parts, complex potentials at high energies, weak potentials, and strong potentials). Mathematically, each class is characterized by a different large parameter. Thus, in contrast with most previous work, we are forced to evaluate the integral asymptotically with respect to parameters other than $\hbar$.

All of the integrals studied are related since the integrands all factor into the product of two functionals, one of which is dominant. This factorization is really the key feature which enables us to obtain each approximation clearly. Although natural, apparently these factorizations have not been exploited in previous asymptotic calculations of path integrals.

Perhaps one of the more interesting results of this work is the relationship between factorizations which may be utilized and regular and singular perturbation theory. This relationship is discussed in some detail. The first sections study integrands with complex potentials. These sections seem particularly relevant to random scattering since there the integrands are similar.

In Sec. 1, we develop the path integral solution of the time-dependent Schrödinger equation with a complex potential and discuss the difficulties in obtaining from this representation an asymptotic expansion in the parameter $\hbar$. In Sec. 2 we circumvent these difficulties by seeking an asymptotic solution of the path integral when the complex part of the potential is weak. The notion of the integrand as a product of two factors first occurs in this section. In Sec. 3, we consider a problem of more practical interest, namely, the time–energy Fourier transform of the path integral (with complex potential). There we derive high energy asymptotic expansions of this "time-path" integral. Although our results are valid in n dimensions, when specialized to one dimension they indicate a means to handle (complex) turning points. In Sec. 4 we discuss the limits of weak and strong potentials, obtaining approximations in each case. For weak potentials one immediately obtains the eikonal approximation from the path integral. In the case of strong potentials, the approximation techniques are more delicate, the approximation being singular. Since this singular nature seems particularly interesting, we present several concrete examples. One of these is the high energy evaluation of the radial Schrödinger equation complete with the correct $(l + \frac{1}{2})^2$ factor.

To date no one has estimated the accuracy of stationary phase calculations of the path integral directly from the integral itself. In lieu of such estimates, we have been forced to compare our results with those of more conventional techniques. These formal calculations are given in the appendices and, in each case, they agree with the path integral results. We remark that, in the body of path integral literature, verifications that the approximation is asymptotic to the solution usually proceed in a manner similar to that used by Pauli[10] [that is, by applying the Schrödinger operator to the approximation, one notices that the result is, say, $O(\hbar^2)$]. The procedure we adopt, due to Keller and collaborators,[11,12] has the advantage of systematically constructing the asymptotic expansion to all orders in the small parameter, while at the same time giving some rationale for the specification of certain underdetermined functions which are inherent in such expansions. Since Keller's technique has not been previously applied to the time-dependent Schrödinger equation (although its applicability was clear), we present these asymptotic constructions in some detail in several appendices. We believe that Keller's constructions are the most systematic available to obtain the complete asymptotic series, although it appears that the path integral eventually will offer a more systematic approach since one needs only to evaluate an "integral" asymptotically.

1. HEURISTIC CONSTRUCTION FOR A COMPLEX POTENTIAL

We seek the fundamental solution of the Schrödinger equation with a complex potential; $V(\mathbf{x}) = U(\mathbf{x}) - igW(\mathbf{x}), gW \geq 0$; i.e., the solution of the following initial value problem:

$$i\hbar \frac{\partial}{\partial t} K(\mathbf{x}, t|\mathbf{y}, 0) = \left(-\frac{\hbar^2}{2m}\nabla_{\mathbf{x}}^2 + V(\mathbf{x})\right) K(\mathbf{x}, t|\mathbf{y}, 0), \quad t > 0, \tag{1.1}$$

$$K(\mathbf{x}, t|\mathbf{y}, 0) = 0, \quad t < 0, \tag{1.2}$$

and

$$\lim_{t\to 0^+} K(\mathbf{x}, t|\mathbf{y}, 0) = \delta^n(\mathbf{x} - \mathbf{y}). \tag{1.3}$$

Here $\mathbf{x}$ and $\mathbf{y}$ are vectors in R^n. K may be represented as a path integral,

$$K(\mathbf{x}, t|\mathbf{y}, 0) = \int_{P(\mathbf{x},t|\mathbf{y},0)} D\mathbf{z}\, \exp\{(i/\hbar)S[\mathbf{z}(\cdot), t]\} \times \exp\{-(g/\hbar)\omega[\mathbf{z}(\cdot), t]\}, \quad (1.4)$$

where

$$S[\mathbf{z}(\cdot), t] \equiv \int_0^t \left[\tfrac{1}{2}m\left(\frac{d\mathbf{z}}{d\sigma}\right)^2 - U(\mathbf{z}(\sigma))\right] d\sigma, \quad (1.5)$$

$$\omega[\mathbf{z}(\cdot), t] \equiv \int_0^t W(\mathbf{x}(\sigma))\, d\sigma, \quad (1.6)$$

and $P(\mathbf{x}, t|\mathbf{y}, 0)$ is that collection of real valued paths connecting the space–time points $(\mathbf{y}, 0)$ and $(\mathbf{x}, t)$.

[Throughout the paper, P will be defined by

$$\{P(\mathbf{x}, t|\mathbf{y}, 0) \equiv \{\text{continuous, real, vector valued functions } \mathbf{z}(\sigma) \text{ of the time } \sigma \text{ such that } \mathbf{z}(0) = \mathbf{y},\ \mathbf{z}(t) = \mathbf{x}\}\} \quad (1.7)$$

However, functionals appearing in the integrand (such as S and ω) will be defined differently in each section of the paper.]

Explicitly the path integral in (1.4) is defined by

$$\int_{P(\mathbf{x},t|\mathbf{y},0)} D\mathbf{z}\, \exp\left(\frac{i}{\hbar}S[\mathbf{z}(\cdot), t]\right) \exp\left(-\frac{g}{\hbar}\omega[\mathbf{z}(\cdot), t]\right) \equiv \lim_{N\to\infty}\left(\frac{m}{2\pi i\hbar\Delta t}\right)^{nN/2} \int_{-\infty}^{\infty} d^n z_1 \cdots \int_{-\infty}^{\infty} d^n z_{N-1} \times \exp\left\{\frac{i}{\hbar}\sum_{j=1}^{N}\left[\left(\frac{m}{2}\frac{(\mathbf{z}_j - \mathbf{z}_{j-1})^2}{\Delta t} - [U(\mathbf{z}_j) - igW(\mathbf{z}_j)]\Delta t\right)\right]\right\}, \quad (1.8)$$

where $\Delta t \equiv t/N$, $\mathbf{z}_0 \equiv \mathbf{y}$, and $\mathbf{z}_N \equiv \mathbf{x}$. As long as this $(N-1)$–fold integral exists, the path integral may be seen to solve Eqs. (1.1) and (1.3) by replacing the real potential in Feynman's original "proof"[1] with a complex one. We remark that $W(\mathbf{x})$ is restricted to be positive just to insure the existence of the integral.

Since the potential is complex, one is motivated to consider extending the class of paths P which are being summed over to the collection of complex valued paths connecting the space–time points $(\mathbf{y}, 0)$ and $(\mathbf{x}, t)$, $CP(\mathbf{x}, t|\mathbf{y}, 0)$. Without questioning its meaning, we consider this formal extension,

$$K \equiv \int_{CP(\mathbf{x},t|\mathbf{y},0)} D\mathbf{z}\, \exp\{(i/\hbar)S[\mathbf{z}(\cdot), t]\} \times \exp\{-(g/\hbar)\omega[\mathbf{z}(\cdot), t]\}, \quad (1.9)$$

and seek to evaluate it asymptotically as $\hbar$ vanishes. For small $\hbar$ one would expect this extension to be dominated by those paths which make the functional $S[\mathbf{z}(\cdot), t] + i\omega[\mathbf{z}(\cdot), t]$ stationary with respect to neighboring paths in the class CP. Thus, we are led to seek (complex valued) solutions of the variational problem,

$$\delta(S + i\omega)|_{\mathbf{z}(\cdot)} = 0, \quad \mathbf{z}(0) = \mathbf{y}, \quad \mathbf{z}(t) = \mathbf{x}, \quad (1.10)$$

which, assuming sufficient analyticity, are just the solutions of

$$m\frac{d^2\mathbf{z}(\tau)}{d\tau^2} = -\nabla V(\mathbf{z}), \quad \mathbf{z}(0) = \mathbf{y}, \quad \mathbf{z}(t) = \mathbf{x}. \quad (1.11)$$

For simplicity we assume (1.11) admits only one solution, $\boldsymbol{\xi}(\cdot)$. (If it admitted several, we would sum the final formula over the candidates.) Treating the formal expression (1.9) as a path integral, we expand the integrand about $\boldsymbol{\xi}(\cdot)$, truncate this expansion after the quadratic term, and evaluate the remaining Gaussian path integral. The result is

$$K \sim A\, \exp((i/\hbar)\{S[\boldsymbol{\xi}(\cdot), t] + ig\omega[\boldsymbol{\xi}(\cdot), t]\}), \quad \text{as } \hbar \to 0, \quad (1.12)$$

where

$$A = (2\pi i\hbar)^{-n/2}\left[\det\left(\frac{\partial^2(S[\boldsymbol{\xi}(\cdot), t] + ig\omega[\boldsymbol{\xi}(\cdot), t])}{\partial\mathbf{x}_i\partial\mathbf{y}_j}\right)\right]^{1/2}. \quad (1.13)$$

It is not at all clear what such formal calculations mean. Nevertheless, in Appendix A we apply the techniques of Keller,[11,12] and construct a series which is asymptotic to the solution of the initial value problem (1.1)–(1.3). The leading term of this series agrees with (1.13); in particular, it contains the path $\boldsymbol{\xi}(\cdot)$. Thus, the dominant path in the asymptotic solution is the complex valued path $\boldsymbol{\xi}(\cdot)$, and, if one is to obtain it from a path integral, the path $\boldsymbol{\xi}(\cdot)$ must be a member of the collection being summed over. That is, one must give the proper interpretation to an expression such as (1.9) in order that it be well defined and equal to (1.4). We do not know how to handle this problem primarily because such expressions diverge due to terms such as $\exp\{-[m\,\mathrm{Re}(\Delta)\mathrm{Im}(\Delta)/\Delta t]\}$ which appear in the integrand. [Here, $\Delta \equiv (\mathbf{z}_i - \mathbf{z}_{i-1})$, $i = 1, 2, \ldots, N$.]

However, these problems of extension can be avoided if real parameters dominate. In these cases, it will be unnecessary to introduce representations such as (1.9) and will be sufficient to use representation (1.4) to generate the asymptotic behavior. In the following two sections, we consider two situations of this type, namely, complex potentials with weak imaginary parts and complex potentials at high energies.

2. WEAKLY COMPLEX POTENTIALS

Again we consider the initial value problem (1.1)–(1.3) and the path integral representation of its solution, Eq. (1.4). The integrand of this path integral is seen to be a product of two functionals, $\exp\{(i/\hbar)S[\mathbf{z}(\cdot), t]\}$ and $\exp\{-(g/\hbar)\omega[\mathbf{z}(\cdot), t]\}$. For small $\hbar$ the first is a rapidly oscillating functional of the path $\mathbf{z}(\cdot)$, while the second is a positive functional. In addition the latter depends critically upon the ratio $g/\hbar$. When this ratio is large, it is a rapidly decreasing exponential; when small, it is slowly varying. Thus, when g is $O(\hbar)$ or smaller, the second factor is a slowly varying, positive functional, and the path integral should be dominated by that path(s) in $P(\mathbf{x}, t|\mathbf{y}, 0)$ which makes the argument of the oscillatory exponential stationary with respect to neighboring paths (in P). Such a path must be a solution of the variational problem

$$\delta S|_{\mathbf{z}(\cdot)} = 0, \quad \mathbf{z}(0) = \mathbf{y}, \quad \mathbf{z}(t) = \mathbf{x}, \quad (2.1)$$

or equivalently the two-point boundary value problem

$$m\frac{d^2\mathbf{z}}{d\sigma^2}(\sigma) = -\nabla U, \quad \mathbf{z}(0) = \mathbf{y}, \quad \mathbf{z}(t) = \mathbf{x}. \quad (2.2)$$

For simplicity we assume this problem has a unique

solution, $\xi(\cdot)$. We now replace the smooth functional with its value at the path $\xi(\cdot)$, remove the resulting constant from the integrand, expand the functional $S[\mathbf{z}(\cdot), t]$ about $\xi(\cdot)$ neglecting terms of higher than second order, and obtain

$$K(\mathbf{x}, t|\mathbf{y}, 0) \sim \tilde{K}(\mathbf{x}, t|\mathbf{y}, 0), \quad \text{for sufficiently small } \hbar \text{ and } g, \tag{2.3}$$

where

$$\tilde{K}(\mathbf{x}, t|\mathbf{y}, 0) \equiv \exp\{(i/\hbar)S[\xi(\cdot), t]\} \exp\{-(g/\hbar)\omega[\xi(\cdot), t]\} \times \int_{P(\mathbf{0},t|0,0)} D\eta \exp\{(i/\hbar) Q[\eta(\cdot), t]\} \tag{2.4}$$

and where the quadratic functional Q is calculated to be

$$Q[\eta(\cdot), t] = \int_0^t \left[\frac{m}{2}\left(\frac{d\eta}{d\sigma}\right)^2 - \tfrac{1}{2}(\nabla^2 U(x))_{\mathbf{x}=\xi(\sigma)}\eta^2(\sigma)\right] d\sigma. \tag{2.5}$$

Being Gaussian, the remaining path integral may be explicitly calculated,[5] yielding

$$\tilde{K}(\mathbf{x}, t|\mathbf{y}, 0) = B \exp\{(i/\hbar)S[\xi(\cdot), t]\} \exp\{-(g/\hbar)\omega[\xi(\cdot), t]\}, \tag{2.6}$$

where

$$B \equiv (-1)^n (2\pi i\hbar)^{-n/2} \left[\left|\det\left(\frac{\partial^2 S[\xi(\cdot), t]}{\partial \mathbf{x}_i \partial \mathbf{y}_j}\right)\right|\right]^{1/2} e^{-i(\pi/2)\beta(\xi(\cdot))} \tag{2.7}$$

and $\beta(\xi(\cdot))$ is the total reduction in rank of the determinant along the path $\xi(\cdot)$.

There are several points to notice about this result. Firstly, the dominant path is real and determined by the real part of the potential. From Sec. 1 we know that for this to be so, g must not be large. Secondly, the approximation seems natural; that is, when V is weakly complex, one might expect to be able to use only its real part to define the classical mechanics, to construct the approximate kernel from rays associated with this classical mechanics in a typical WKB fashion, and finally to take the complex part of the potential into account with an exponential sink of probability, a sink depending upon the time integral of the complex part of the potential evaluated along the "classical path $\xi(\cdot)$." Thirdly, the approximation is clearly invalid if $\mathbf{x}$ (or $\mathbf{y}$) is located at a zero of $\{\det[(\partial/\partial \mathbf{x}_i)(\partial/\partial \mathbf{x}_j)S(\xi(\cdot), t)]\}^{-1}$. These occur at caustics associated with the classical rays of the real potential U. Even if $\mathbf{x}$ and $\mathbf{y}$ are not located at a caustic, $\xi(\cdot)$ may pass through several "on its way between $\mathbf{x}$ and $\mathbf{y}$." $\beta(\xi(\cdot))$ counts the number of times this occurs, and the path integral quickly indicates the phase associated with each occurrence. Certainly the "natural argument" presented above does not indicate how one should account for this phenomena.

Finally, we have checked the order of accuracy of $\tilde{K}$ [Eq. (2.6)] in two ways. One method is to take the asymptotic expansion as $\hbar$ vanishes given in Appendix A [Eq. (A25)], solve Eq. (A26) for the rays by a power series in g, and use only the first term of this expansion to define $S[\xi(\cdot), t]$. More directly, one may seek an asymptotic expansion of K of which (2.7) is the leading term. This we do in Appendix B. Both techniques confirm that $\tilde{K}$ is the correct leading term in a formal asymptotic expansion of K.

3. HIGH ENERGY AND COMPLEX POTENTIALS

While in the time dependent picture, the calculation is extremely direct, for practical purposes it is often desirable to work in the time independent framework. Consider $G(\mathbf{x}, \mathbf{y}, E)$, the Fourier transform of $K(\mathbf{x}, t|\mathbf{y}, 0)$:

$$G(\mathbf{x}, \mathbf{y}, E) \equiv \frac{1}{i\hbar}\int_0^\infty dt \exp\left(\frac{i}{\hbar}Et\right) K(\mathbf{x}, t|\mathbf{y}, 0)$$
$$= \frac{1}{i\hbar}\int_0^\infty dt \int_{P(\mathbf{x},t|\mathbf{y},0)} D\mathbf{z} \times \exp\left(\frac{i}{\hbar}\{Et + S[\mathbf{z}(\cdot), t]\}\right)\exp\left(-\frac{g}{\hbar}\omega[\mathbf{z}(\cdot), t]\right). \tag{3.1}$$

Here E is given a small positive imaginary part ϵ, and (3.1) is interpreted as the limit as ϵ vanishes.

In this section we interpret the Fourier transform together with the path integral as a "double integral" which we seek to evaluate asymptotically for large E. We begin by placing the integral in an equivalent form:

$$G(\mathbf{x}, \mathbf{y}, E)$$
$$= \frac{1}{i\hbar\sqrt{E}}\int_0^\infty d\tau \lim_{N\to\infty}\left(\frac{m\sqrt{E}N}{2\pi i\hbar\tau}\right)^{nN/2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} d^n\mathbf{z}_1\cdots d^n\mathbf{z}_{N-1}$$
$$\times \exp\left\{\frac{i\sqrt{E}}{\hbar}\left[\tau + \sum_{j=1}^{N}\left(\frac{m}{2}\frac{N}{\tau}\left|\mathbf{z}_j - \mathbf{z}_{j-1}\right|^2 - V(\mathbf{z}_j)\frac{\tau}{NE}\right)\right]\right\} \tag{3.2}$$
$$\equiv \frac{1}{i\hbar\sqrt{E}}\int_0^\infty d\tau \int_{P(\mathbf{x},1|\mathbf{y},0)} D\mathbf{z} \exp\left(\frac{i}{\hbar}\frac{i\sqrt{E}}{\hbar}\phi[\mathbf{z}(\cdot), \tau, E]\right), \tag{3.3}$$

where

$$\phi[\mathbf{z}(\cdot), \tau, E] \equiv \tau + \int_0^1\left(\frac{m}{2\tau}\left(\frac{d\mathbf{z}}{d\sigma}\right)^2 - \frac{\tau}{E}V(\mathbf{z}(\sigma))\right)d\sigma. \tag{3.4}$$

When V is bounded, whether complex or not, a very simple approximation is applicable. The integrand factors rather naturally into the product of two exponentials,

$$\exp\{(i\sqrt{E}/\hbar)\phi\} = \exp\{(i\sqrt{E}/\hbar)T[\mathbf{z}(\cdot), \tau]\} \times \exp\{-(i/\hbar\sqrt{E})\mu[\mathbf{z}(\cdot), \tau]\}, \tag{3.5}$$

where

$$T[\mathbf{z}(\cdot), \tau] = \tau + \int_0^1 \frac{m}{2\tau}\left(\frac{d\mathbf{z}(\sigma)}{d\sigma}\right)^2 d\sigma \tag{3.6}$$

and

$$\mu[\mathbf{z}(\cdot), \tau] = \tau \int_0^1 V(\mathbf{z}(\sigma))\, d\sigma. \tag{3.7}$$

The first factor becomes highly oscillatory as E becomes large, while the second becomes more and more slowly varying as a function of the "path" $(\mathbf{z}(\cdot), \tau)$. For large E the "time-path" integral will be dominated by those paths $(\xi(\cdot), \tau_0)$ about which the functional T is stationary with respect to neighboring paths in P and nearby times τ. But these are just the solutions of the variational problem

$$\delta T|_{\xi(\cdot)} = 0, \quad \mathbf{z}(0) = \mathbf{y}, \quad \mathbf{z}(1) = \mathbf{x} \tag{3.8}$$

with τ_0 determined by the constraint

$$1 = \frac{m}{2\tau_0^2}\int_0^1\left(\frac{d\xi}{d\sigma}\right)^2 d\sigma. \tag{3.9}$$

The Euler equation associated with (3.8) is

$$\frac{m}{\tau}\frac{d^2\xi(\sigma)}{d\sigma^2}=0, \quad \xi(0)=\mathbf{y}, \quad \xi(1)=\mathbf{x}. \tag{3.10}$$

Thus, $\xi(\sigma)$ is given by

$$\xi(\sigma)=\mathbf{y}+(\mathbf{x}-\mathbf{y})\sigma, \tag{3.11}$$

and τ_0 is calculated to be

$$\tau_0=+[\tfrac{1}{2}m(\mathbf{x}-\mathbf{y})^2]^{1/2}. \tag{3.12}$$

As in Sec. 2, we evaluate the functional $\exp\{-(i/\hbar\sqrt{E})\mu[\mathbf{z}(\cdot),\tau]\}$ along the path $(\xi(\cdot),\tau_0)$ and remove this constant from the integral. The remaining time-path integral is merely that of a free particle and may be calculated explicitly,

$$G(\mathbf{x},\mathbf{y},E)\sim\tilde{G}(\mathbf{x},\mathbf{y},E) \quad \text{as } E\to\infty, V \text{ bounded}, \tag{3.13}$$

where

$$\tilde{G}(\mathbf{x},\mathbf{y},E)\equiv D\exp\left(\frac{i}{\hbar}[2mE]^{1/2}(\mathbf{x}-\mathbf{y})\right)\times\exp\left[-\frac{i\tau_0}{\hbar\sqrt{E}}\int_0^1 V(\mathbf{y}+(\mathbf{x}-\mathbf{y})\sigma)d\sigma\right] \tag{3.14}$$

and where

$$D\equiv-(i/\hbar)(m/2E)^{1/2}. \tag{3.15}$$

Thus, for large E and bounded V the eikonal approximation is immediately obtained from the path integral even when V is complex. In this approximation the potential is accounted for by integration along the "free particle rays."

Clearly in many actual cases when E is finite, (3.14) is inadequate. For in such cases, there may exist "forbidden regions" of space in which the character of G is not oscillatory. Equation (3.14) shows no such regions. Also, in such cases there will exist rays which reflect from the forbidden areas. These are not present in Eq. (3.14) which contains only a direct ray.

To include such effects, we reconsider the factorization of the integrand in Eq. (3.3) by writing ϕ as the sum of its real and imaginary parts. Viewed in this manner, it seems natural to factor the integrand as

$$\exp[(i\sqrt{E}/\hbar)\phi]=\exp\{(i\sqrt{E}/\hbar)F[\mathbf{z}(\cdot),\tau,E]\}\times\exp\{-(g/\hbar\sqrt{E})\omega[\mathbf{z}(\cdot),\tau,E]\} \tag{3.16}$$

where

$$F[\mathbf{z}(\cdot),\tau,E]\equiv\tau+\int_0^1\left[\frac{m}{2\tau}\left(\frac{d\mathbf{z}}{d\sigma}\right)^2-\frac{\tau}{E}U(\mathbf{z})\right]d\sigma \tag{3.17}$$

and

$$\omega[\mathbf{z}(\cdot),\tau,E]\equiv\int_0^1\tau W(\mathbf{z})d\sigma. \tag{3.18}$$

The first factor is wildly oscillatory for large E. The paths $(\mathbf{z}(\cdot),\tau)$ which make F stationary with respect to neighboring paths will dominate the integral provided the complex part of the potential is weak relative to the real part, for example, if W is bounded and g is sufficiently small. In this case, $(\xi(\cdot))$ will be the solution of the variational problem for F,

$$(\delta F)_{\xi(\cdot)}=0, \quad \xi(0)=\mathbf{y}, \quad \xi(1)=\mathbf{x}, \tag{3.19}$$

with the critical time τ_0 determined by the constraint

$$\frac{1}{2}\frac{m}{\tau_0^2}\left(\frac{d\mathbf{z}}{d\sigma}\right)^2+\frac{U}{E}=1. \tag{3.20}$$

The Euler equations of this problem are

$$\frac{m}{\tau^2}\frac{d^2\mathbf{z}(\sigma)}{d\sigma^2}=-\frac{1}{E}\nabla_z U, \quad \mathbf{z}(0)=\mathbf{y}, \quad \mathbf{z}(1)=\mathbf{x}. \tag{3.21}$$

We assume that $\mathbf{x}$ and $\mathbf{y}$ are in the same allowed region of the potential U. As we have discussed, Eqs. (3.20) and (3.21) may admit several solutions which we index by a set J, $(\xi_j(\cdot),\tau_j)$, $j\in J$. Since only the immediate vicinities of these paths will contribute significantly to the "time-path" integral, we write it as a sum over J, each term of which is integrated only over the immediate vicinity of the path $(\xi_j(\cdot),\tau_j)$,

$$G(\mathbf{x},\mathbf{y},E)\sim\frac{1}{i\hbar\sqrt{E}}\sum_{j\in J}\int_{R_j}d\tau\int_{P_j}D\mathbf{z}\exp\left(\frac{i\sqrt{E}}{\hbar}\phi[\mathbf{z}(\cdot),\tau,E]\right), \quad \text{as } \sqrt{E}\to\infty, \tag{3.22}$$

where $(P_j\times R_j)$ are small regions of $(\mathbf{z}(\cdot),\tau)$ space containing $(\xi_j(\cdot),\tau_j)$.

Again we evaluate each slowly varying functional at the path $(\xi_j(\cdot),\tau_j)$, remove it from the integrand, and evaluate the remaining "time-path" integrals by stationary phase. This is most easily accomplished by first evaluating the path integral asymptotically, and then taking the Fourier transform of the result.[5] Expanding F about ξ_j through second order in $\eta_j(\cdot)\equiv(\mathbf{z}(\cdot)-\xi_j(\cdot))$ and performing the resulting path integral, we obtain

$$\int_{P_j}D\mathbf{z}\exp\left(\frac{i\sqrt{E}}{\hbar}F[\mathbf{z}(\cdot),\tau,E]\right)\sim\exp\left(\frac{i\sqrt{E}}{\hbar}F[\xi_j(\cdot),\tau,E]\right)\int_{P(0,1\,|\,0,0)}D\mathbf{z}\exp\left\{\frac{i\sqrt{E}}{\hbar}\int_0^1\left[\frac{m}{2\tau}\left(\frac{d\eta}{d\sigma}\right)^2-\frac{\tau}{2E}(\nabla^2U)_{\mathbf{z}=\xi_j}\right]d\sigma\right\} \tag{3.23}$$

$$\sim\exp\left(\frac{i\sqrt{E}}{\hbar}F[\xi_j(\cdot),\tau,E]\right)H_j(\mathbf{x},\tau\,|\,\mathbf{y},0)\equiv\tilde{K}_j(\mathbf{x},\tau\,|\,\mathbf{y},0), \tag{3.24}$$

where

$$H_j\equiv\left(\frac{\sqrt{E}}{2\pi i\hbar}\right)^{+n/2}\left[\det\left(\frac{\partial^2F[\xi_j(\cdot),\tau,E]}{\partial x_k\partial y_l}\right)\right]^{1/2}. \tag{3.25}$$

We now asymptotically evaluate the Fourier transform of (3.24) by expanding it about τ_j and evaluating the Gaussian integral. The result is

$$\int_{R_j}d\tau\exp\left(\frac{i\sqrt{E}}{\hbar}\tau\right)\tilde{K}_j(\mathbf{x},\tau\,|\,\mathbf{y},0)\sim H_j(\mathbf{x},\tau_j\,|\,\mathbf{y},0)\exp\left(\frac{i\sqrt{E}}{\hbar}[\tau_j+F[\xi_j(\cdot),\tau_j,E]\right)\int_{-\infty}^{\infty}d\sigma\exp\left(\frac{i\sqrt{E}}{2\hbar}\left(\frac{d^2F}{d\tau^2}\right)_{\tau=\tau_j}\sigma^2\right) \tag{3.26}$$

$$=L_j\exp\left(\frac{i\sqrt{E}}{\hbar}[\tau_j+F[\xi_j(\cdot),\tau_j,E]\right), \tag{3.27}$$

where

$$L_j \equiv \left[\frac{2\pi i\hbar}{-\sqrt{E}}\left(\frac{d^2F}{d\tau^2}\right)^{-1}\right]^{1/2} H_j. \tag{3.28}$$

Equation (3.27) may be considerably simplified if the classical equations, (3.20) and (3.21), are used to eliminate τ_j and to express the result directly in terms of $(\mathbf{x}, \mathbf{y}, E)$. The final expression together with the explicit specifications of the branches of the square roots is

$$G(\mathbf{x}, \mathbf{y}, E) \sim \tilde{G}(\mathbf{x}, \mathbf{y}, E), \quad \text{as } E \to \infty, \tag{3.29}$$

where

$$\tilde{G}(\mathbf{x}, \mathbf{y}, E) \equiv [(-1)^n/2\pi\hbar^{n-1}] \sum_{j\subset J} I_j(\mathbf{x}, \mathbf{y}, E) \times \exp[(i/\hbar)R_j(\mathbf{x}, \mathbf{y}, E)] \exp[-(g/\hbar\sqrt{E})\omega_j], \tag{3.30}$$

$$R_j \equiv \int_{\mathbf{y}}^{\mathbf{x}} + [P(\mathbf{x}')]^{1/2} d\mathbf{x}' \text{ (evaluated along the classical path } j \text{ connecting } \mathbf{x} \text{ and } \mathbf{y} \text{ at energy } E), \tag{3.31}$$

$$I_j \equiv \left[\left|\det\begin{pmatrix}\partial_{xy}R_j & \partial_{xE}R_j \\ \partial_{yE}R_j & \partial_{EE}R_j\end{pmatrix}\right|\right]^{1/2} e^{-i(\pi/2)\beta_j}, \tag{3.32}$$

β_j is the total reduction in rank of I_j along the classical path in question, and ω_j is the functional ω evaluated along the classical path.

Many of the comments at the end of Sec. 2 for the time-dependent case apply here as well. The time-path integral tells us that in this approximation we need only use the real part of the potential to define the appropriate classical mechanics with the complex part entering only through an exponential sink of probability, a sink depending upon the time integral of this complex part evaluated along the classical path.

The phase differences of $(-i\beta_j\pi/2)$ between the various terms arise from treating the caustics associated with the classical rays of the real potential $U(\cdot)$. The interpretation of these phases is particularly clear in the one-dimensional case. For example, let $U(\cdot)$ be a continuous, monotonic, increasing function such that $U(x_0) = E$. If $x < y < x_0$, Eqs. (3.20) and (3.21) will possess two solutions $(\xi_d(\cdot), \tau_d)$ and $(\xi_r(\cdot), \tau_r)$ corresponding to the direct and reflected classical rays. In this special example, Eq. (3.30) reduces to

$$\tilde{G}(x, y, E) = \frac{m}{\hbar}[P(x)P(y)]^{-1/2}\left\{\exp\left[-\frac{i}{\hbar}\int_y^x \left(P(z) + img\frac{W(z)}{P(z)}\right)dz\right] + e^{-i\pi/2}\exp\left[+\frac{i}{\hbar}\int_y^{x_0}\left(P(z) + img\frac{W(z)}{P(z)}\right)dz - \frac{i}{\hbar}\int_{x_0}^{x}\left(P(z) + img\frac{W(z)}{P(z)}\right)dz\right]\right\}, \tag{3.33}$$

where $P(z) = +2m[E - U(z)]^{1/2}$. We emphasize that the phase difference of $e^{-i\pi/2}$ arises directly from the evaluation of the time-path integral and depends only upon the real part of the potential. Had we proceeded with the more conventional WKB approaches, it would have arisen from "connection formulas for complex valued turning points."

Each approximation derived in this section is the leading term in an asymptotic expansion of $G(\mathbf{x}, \mathbf{y}, E)$, as may be verified by methods similar to those applied to the time dependent case. We omit these calculations since the formulas are so closely related to the well-known high-energy WKB and eikonal approximations and turn to a strong potential limit.

4. WEAK AND STRONG POTENTIALS—REGULAR AND SINGULAR PERTURBATIONS

In this section we consider only real potentials and fix all constants with the exception of the potential strength g. The path integral to be studied is

$$K(\mathbf{x}, t|\mathbf{y}, 0) = \int_{P(\mathbf{x},t|\mathbf{y},0)} D\mathbf{z} \exp\{(i/\hbar)T[\mathbf{z}(\cdot), t]\} \times \exp\{-(i/\hbar)g\omega[\mathbf{z}(\cdot), t]\}, \tag{4.1}$$

where

$$T[\mathbf{z}(\cdot), t] \equiv \int_0^t \tfrac{1}{2}m\left(\frac{d\mathbf{z}}{d\sigma}\right)^2 d\sigma, \tag{4.2}$$

and

$$\omega[\mathbf{z}(\cdot), t] \equiv \int_0^t V(\mathbf{z}(\sigma))d\sigma. \tag{4.3}$$

When g is very small, it is clear that we may proceed as in the first part of Sec. 3, that is, to remove the second factor from the integrand evaluated along the "free particle path" and to evaluate the free propagator:

$$K(\mathbf{x}, t|\mathbf{y}, 0) \sim (m/2\pi i\hbar t)^{n/2} \exp\{(i/\hbar)T[\xi(\cdot), t]\} \times \exp\{-(i/\hbar)g\omega[\xi(\cdot), t]\}, \tag{4.4}$$

where

$$\xi(\sigma) = \mathbf{y} + (\mathbf{x} - \mathbf{y})\sigma/t. \tag{4.5}$$

This is just the time-dependent eikonal approximation, valid for weak potentials.

A far more intriguing limit is that of large g. In this case, the second factor in (4.1) makes the integrand wildly oscillatory for large values of g. In analogy with all preceding material, we would expect the path integral to be dominated by those paths $\xi(\cdot)$ which make the functional $\omega[\mathbf{z}(\cdot), t]$ stationary. However, the following calculation,

$$\delta\omega[\mathbf{z}(\cdot), t] = \delta\int_0^t V(\mathbf{z}(\sigma))d\sigma = \int_0^t (\nabla V \cdot \delta\mathbf{z})d\sigma = 0 \;\forall \delta\mathbf{z}_j \Longrightarrow \frac{\partial}{\partial z_i}V(\mathbf{z}) = 0, \quad i = 1, 2, 3, \ldots, n, \tag{4.6}$$

shows that the only possible solutions are constants associated with critical points of the function V. However, these will not be members of the class $P(\mathbf{x}, t|\mathbf{y}, 0)$ since the boundary conditions will not be satisfied. This simple analysis is just too naive.

The difficulty is overcome by invoking a change of variable. We replace $T - g\omega$ by the equivalent functional

$$S[\mathbf{z}(\cdot), g^{1/2}t] \equiv g^{1/2}\int_0^{g^{1/2}t}\left[\tfrac{1}{2}m\left(\frac{d\mathbf{z}}{d\sigma}\right)^2 - V(\mathbf{z}(\sigma))\right]d\sigma. \tag{4.7}$$

Under this change, K is written as

$$K(\mathbf{x}, t|\mathbf{y}, 0) = \int_{P(\mathbf{x},\sqrt{g}\,t|\mathbf{y},0)} D\mathbf{z} \exp\{(i/\hbar)g^{1/2}S[\mathbf{z}(\cdot), g^{1/2}t]\}. \tag{4.8}$$

Now as $\sqrt{g}$ increases, $S[\mathbf{z}(\cdot), g^{1/2}t]$ becomes large (this is particularly so if $g^{1/2}t$ is fixed or if $V(\mathbf{z}) \leq 0$) and the complete integrand, $\exp[(ig^{1/2}/\hbar)S]$, becomes a rapidly oscillating functional of the path $\mathbf{z}(\cdot)$. This oscillatory behavior forces the path integral to be dominated by those paths $\xi(\cdot)$ which make

the entire functional S stationary with respect to neighboring paths in the class $P(\mathbf{x}, \sqrt{g}\,t\,|\,\mathbf{y}, 0)$. The Euler (or ray) equations associated with this variational principle are

$$m\,\frac{d^2\boldsymbol{\xi}}{d\sigma^2} = -\nabla_{\xi} V, \qquad \boldsymbol{\xi}(0) = \mathbf{y}, \quad \boldsymbol{\xi}(\sqrt{g}\,t) = \mathbf{x}. \tag{4.9}$$

As above, we expand the integrand through second order about the function $\boldsymbol{\xi}(\cdot)$ and perform the resulting Gaussian path integral obtaining

$$K_\tau(\mathbf{x}, \tau\,|\,\mathbf{y}, 0) \underset{g\to\infty}{\sim} \left(\frac{2\pi i\hbar}{g^{1/2}}\right)^{-n/2} \left[\det\left(\frac{\partial^2 S(\boldsymbol{\xi}(\cdot), \tau)}{\partial x_i \partial y_j}\right)\right]^{1/2} \exp\left(\frac{i\sqrt{g}}{\hbar}\, S(\boldsymbol{\xi}(\cdot), \tau)\right), \tag{4.10}$$

where $\tau \equiv \sqrt{g}\,t$, and $K_\tau \equiv K(\mathbf{x}, \tau/\sqrt{g}\,|\,\mathbf{y}, 0)$. That (4.10) is the leading term in an asymptotic approximation is verified in Appendix C by more conventional techniques.

Thus, in the variables $(\mathbf{x}, \tau)$ the problem associated with large g reduces to the same calculations involved in the usual small $\hbar$, fixed g approximation. There is one important difference, however. We would actually like to solve the problem in the "physical" variables $(\mathbf{x}, t)$. In this case, as long as (4.10) is valid through $\tau = O(\sqrt{g})$, we need only an approximate solution of the ray Eqs. (4.9) valid uniformly in $\sigma \in [0, t]$, t fixed > 0. Thus, it may be possible to use the large parameter $\sqrt{g}$ which appears in (4.9) to obtain an explicit representation of $\boldsymbol{\xi}(\cdot)$. In the analogous small $\hbar$ case, when the solution is sought in the "physical" $(\mathbf{x}, t)$ variables, there is no $\hbar$ dependence in the ray equation to be exploited.

To see how one could use this additional g dependence of the ray equations, we change variables to remove the parameter g from the boundary conditions,

$$m\,\frac{d^2\boldsymbol{\xi}}{d\sigma'^2} = -g\,\nabla V, \qquad \boldsymbol{\xi}(0) = \mathbf{y}, \qquad \boldsymbol{\xi}(t) = \mathbf{x}. \tag{4.11}$$

We seek an asymptotic approximation of (4.11) valid in the limit of large g. The source of the difficulty in the "naive method" described above is now apparent, that is, the problem of finding these approximate rays is a singular perturbation problem. In attempting to use only the potential energy functional ω to define the rays, we lost the boundary conditions. This loss of boundary conditions is a reflection of the singular nature of the problem. On the other hand, the problem of a weak potential was successfully treated by using only the kinetic energy functional T to define the rays. In that case, had we followed the procedure of "stretching the time variable," the ray equations would have been the same,

$$m\,\frac{d^2\boldsymbol{\xi}}{d\sigma^2} = -g\nabla V, \qquad \boldsymbol{\xi}(0) = \mathbf{y}, \qquad \boldsymbol{\xi}(t) = \mathbf{x}, \tag{4.12}$$

but the solution is desired for small g. Solving (4.12) in the limit of vanishing g is merely a regular perturbation problem which involves no loss of boundary conditions. This singular-regular nature is the underlying reason that the problem of dominant ω is more delicate than that of dominant T.

A one-dimensional parabolic barrier with smooth variable coefficients,

$$V(\xi, \sigma) = -\tfrac{1}{2} f(\epsilon\sigma)\xi^2 - h(\sigma)\xi - h^2(\sigma)/2f(\epsilon\sigma), \qquad f(\cdot) > 0, \epsilon^2 = 1/g, \tag{4.13}$$

provides a simple example which illustrates one method by which the g dependence of the ray equations may be utilized. In this case, $V(\xi, \sigma)$ is negative and Eq. (4.10) certainly is applicable. Rather than calculate the ray ξ exactly, we seek a uniformly valid approximation through the methods of singular perturbation theory.[13]

In this case with m set equal to 1, Eq. (4.12) becomes

$$\frac{\epsilon^2 d^2\xi(\sigma)}{d\sigma^2} = f(\epsilon\sigma)\xi(\sigma) + h(\sigma), \qquad \xi(0) = y, \qquad \xi(t) = x. \tag{4.14}$$

Here y, x and $h(\sigma)/f(\epsilon\sigma)$ are assumed to satisfy the inequality $y \le -h(\sigma)/f(\epsilon\sigma) \le x$ so that the barrier lies between the initial and final points. The auxiliary equation,

$$\epsilon^2 D^2 - f(\epsilon\sigma) = 0, \tag{4.15}$$

has two roots,

$$D = \pm\,(1/\epsilon)[f(\epsilon\sigma)]^{1/2}, \tag{4.16}$$

both of which are singular as ϵ vanishes. This indicates that the solution will possess both a left boundary layer near $\sigma = 0$ and a right boundary layer near $\sigma = t$, and leads us to seek a solution of the form

$$\xi(\sigma) = \xi_L(\sigma_L; \epsilon) + \xi_I(\sigma; \epsilon) + \xi_R(\sigma_R; \epsilon), \tag{4.17}$$

where $\sigma_L \equiv \sigma/\epsilon$, $\sigma_R \equiv (t - \sigma)/\epsilon$. Working first in the interior region away from both boundaries, we seek an asymptotic expansion to $\xi_I(\sigma; \epsilon)$ of the form

$$\xi_I(\sigma; \epsilon) \sim \sum_{j=0}^{\infty} \xi_{I_j}(\sigma)\epsilon^j. \tag{4.18}$$

Placing expansion (4.18) into the ray equation (4.14), we find that the ξ_{I_j} satisfy

$$f(\epsilon\sigma)\xi_{I_j}(\sigma) + h(\sigma) = \frac{d^2\xi_{I_{j-2}}}{d\sigma^2}, \qquad j \in (0, 1, 2, \cdots), \qquad \xi_{I_{-1}} = \xi_{I_{-2}} = 0. \tag{4.19}$$

Thus, the leading behavior of $\xi_I(\sigma)$ is seen to be

$$\xi_I(\sigma) \sim -h(\sigma)/f(\epsilon\sigma). \tag{4.20}$$

In the left boundary layer, $\xi_L(\sigma_L) \equiv \xi(\epsilon\sigma_L) - \xi_I(\epsilon\sigma_L)$ satisfies

$$\frac{d^2\xi_L}{d\sigma_L^2} = f(\epsilon^2\sigma_L)\xi_L. \tag{4.21}$$

Writing ξ_L as $(A_L(\sigma_L; \epsilon)\exp\{-\int_0^{\sigma_L}[f(\epsilon^2\sigma)]^{1/2}d\sigma\})$, we find A_L satisfies

$$\frac{d^2A_L}{d\sigma_L^2} - 2[f(\epsilon^2\sigma_L)]^{1/2}\frac{dA_L}{d\sigma_L} = \frac{\epsilon^2}{2}\frac{f'(\epsilon^2\sigma_L)A_L}{[f(\epsilon^2\sigma_L)]^{1/2}}. \tag{4.22}$$

Seeking A_L as $\sum_{j=0}^{\infty} A_{L_j}\epsilon^j$, we find

$$\frac{d^2A_{L_j}}{d\sigma_L^2} - 2[f(\epsilon^2\sigma_L)]^{1/2}\frac{dA_{L_j}}{d\sigma_L} = \frac{f'(\epsilon^2\sigma_L)}{2[f(\epsilon^2\sigma_L)]^{1/2}}A_{L_{j-2}},$$

$$j \in (0, 1, 2, \cdots), \quad A_{-1} = A_{-2} = 0. \tag{4.23}$$

Demanding only bounded solutions as $\sigma_L \to +\infty$, we find the leading behavior of $\xi_L(\sigma_L)$ to be

$$\xi_L(\sigma_L) \sim a_L e^{-F_L(\sigma_L)}, \tag{4.24}$$

where

$$F_L(\sigma_L) \equiv \int_0^{\sigma_L} [f(\epsilon^2\sigma)]^{1/2}\, d\sigma, \tag{4.25}$$

and a_L is a constant to be determined. Similarly, in the right boundary layer we find

$$\xi_R(\sigma_R) \sim a_R e^{-F_R(\sigma_R)}, \tag{4.26}$$

where

$$F_R(\sigma_R) \equiv \int_0^{\sigma_R} [f(\epsilon t - \epsilon^2\sigma)]^{1/2}\, d\sigma, \tag{4.27}$$

and a_R another constant. Hence the leading behavior of the ray $\xi(\sigma)$ is given as

$$\xi(\sigma) \sim a_L e^{-F_L(\sigma/\epsilon)} - [h(\sigma)/f(\epsilon\sigma)] + a_R e^{-F_R[(t-\sigma)/\epsilon]}. \tag{4.28}$$

The boundary conditions $\xi(0) = y$, $\xi(t) = x$ specify a_L and a_R to be

$$a_L = [y + h(0)/f(0)] - [x + h(t)/f(\epsilon t)]e^{-F_L(t/\epsilon)} + O(e^{-2F_L(t/\epsilon)}) \tag{4.29}$$

and

$$a_R = [x + h(t)/f(\epsilon t)] - [y + h(0)/f(0)]e^{-F_L(t/\epsilon)} + O(e^{-2F_L(t/\epsilon)}). \tag{4.30}$$

The physical interpretation of this approximation is clear. In the limit of a steep barrier, the particle, in a time of $O(\epsilon)$, proceeds exponentially from the initial point y to the top of the barrier, "rides" with the barrier until a time σ such that $(t - \sigma)$ is $O(\epsilon)$, and finally proceeds exponentially to the final point x.

Using Eq. (4.25) for the ray, we calculate the leading asymptotic behavior of K to be

$$K \sim K_0 \equiv (2\pi i\hbar)^{-1/2}(\tilde{D})^{1/2} \exp[(i/\hbar)\tilde{S}], \quad \text{as } \epsilon \to 0, \tag{4.31}$$

where

$$\tilde{S} \equiv \int_0^t d\sigma \left\{ \frac{1}{2}\left(-\frac{h'(\sigma)}{f(\epsilon\sigma)} + \frac{\epsilon h(\sigma)f'(\epsilon\sigma)}{f^2(\epsilon\sigma)}\right)^2 + \frac{f(\epsilon\sigma)}{\epsilon^2}\left(a_L^2 e^{-2F_L(\sigma_L)} + a_R^2 e^{-2F_R(\sigma_R)}\right) + \frac{[f(\epsilon\sigma)]^{1/2}}{\epsilon}\left[\left(\frac{h'(\sigma)}{f(\epsilon\sigma)} - \frac{\epsilon h(\sigma)f'(\epsilon\sigma)}{f^2(\epsilon\sigma)}\right) \times (a_L e^{-F_L(\sigma_L)} - a_R e^{-F_R(\sigma_R)})\right]\right\}, \tag{4.32}$$

and

$$\tilde{D} \equiv -\frac{2}{\epsilon^2}\int_0^t d\sigma f(\epsilon\sigma)(e^{-F_R(t/\epsilon)}e^{-2F_L(\sigma_L)} + e^{-F_L(t/\epsilon)}e^{-2F_R(\sigma_R)}). \tag{4.33}$$

In order to be certain that K_0 is a valid approximation, we applied the Schrödinger operator to it. This rather tedious calculation establishes that

$$\left[i\hbar\frac{\partial}{\partial t} + \frac{\hbar^2}{2}\frac{\partial^2}{\partial x^2} + \frac{1}{2\epsilon^2}\left(f(\epsilon t)x^2 + 2h(t)x + \frac{h^2(t)}{f(\epsilon t)}\right)\right]K_0 = [O(\epsilon)]K_0 \tag{4.34}$$

as long as t is bounded away from zero. (Clearly here we had to assume f and h were sufficiently smooth.)

While (4.26) seems complicated, it is explicit in the sense that the integrand is explicitly given and not just "the solution of the ray equations." It is this feature which makes the technique interesting in that one could apply it to problems whose ray equations are not explicitly solvable.

In the problem at hand, it is instructive to compare the approximation with the exact solution when $h \equiv 0$ and $f \equiv 1$. In that case, K_0 is calculated to be

$$K_0 \sim (\pi i\epsilon\hbar)^{-1/2}e^{-t/2\epsilon}\exp[(i/2\hbar\epsilon)(x^2 + y^2 - 4xy\, e^{-t/\epsilon})], \tag{4.35}$$

while the exact answer is

$$K = (2\pi i\hbar\epsilon)^{-1/2}[\sinh(t/\epsilon)]^{-1/2} \times \exp\{(i/2\hbar\epsilon)[(x^2 + y^2)\coth(t/\epsilon) - 2xy\ \operatorname{csch}(t/\epsilon)]\}. \tag{4.36}$$

These agree for small ϵ as long as t is bounded away from zero, although K_0 is considerably the simpler.

We close this section by applying these path integral techniques to the radial Schrödinger equation,

$$\left(-\frac{1}{2}\frac{d^2}{dr^2} + V(r) + \frac{l(l+1)}{r^2} - E\right)G(r, r'; E, l) = -\delta(r - r'). \tag{4.37}$$

As is well known, the singularity at the origin of the angular momentum barrier renders this equation ill-suited for high energy approximation. Following Langer,[14] we remove this singularity with the change of variable $r = e^x$,

$$\left(-\frac{1}{2}\frac{d^2}{dx^2} + W(x) + L\right)g(x, x'; E, L) = -\delta(x - x'), \tag{4.38}$$

where

$$L \equiv (l + \tfrac{1}{2})^2, \quad W \equiv (V - E)e^{2x}, \quad g \equiv (r'r)^{-1/2}G(r, r'; E, l), \quad \text{and} \quad e^{x'} \equiv r'. \tag{4.39}$$

Treating L as a continuous variable, we write g as

$$g = i\int_0^\infty d\lambda\ e^{-i\lambda L}\hat{g}(x, x'; E, \lambda), \tag{4.40}$$

where $\hat{g}$ is the solution of

$$\left(-\frac{\partial^2}{2\partial x^2} + W(x) - i\frac{\partial}{\partial\lambda}\right)\hat{g}(x, x'; E, \lambda) = 0, \quad \lim_{\lambda\to 0}\hat{g} = \delta(x - x'). \tag{4.41}$$

The solution of (4.41), $\hat{g}$, may be represented as a path integral,

$$\hat{g}(x, x'; E, \lambda) = \int_{P(x,\lambda|x',0)} Dz \exp\{iS[z(\cdot), E, \lambda]\}, \tag{4.42}$$

where

$$S[z(\cdot), E, \lambda] \equiv \int_0^\lambda \left[\frac{1}{2}\left(\frac{dz}{d\sigma}\right)^2 + (E - V)e^{2z}\right]d\sigma. \tag{4.43}$$

For large E, bounded V, the potential Ee^{2z} dominates, and the methods of this section apply. Replacing λ with $\lambda'/\sqrt{E}$ in the definition of the path integral, we obtain

$$\hat{g}(x,x';E,\lambda'/\sqrt{E}) = \int_{P(x,\lambda'|x',0)} Dz\, \exp\{i\sqrt{E}\, S_0[z(\cdot),E,\lambda']\}, \quad (4.44)$$

where

$$S_0[z(\cdot),E,\lambda'] \equiv \int_0^{\lambda'} \left[\frac{1}{2}\left(\frac{dz}{d\sigma}\right)^2 + \left(1-\frac{V(z)}{E}\right)e^{2z}\right] d\sigma. \quad (4.45)$$

For large E, fixed λ', this last path integral is dominated by that path(s) which makes S_0 stationary. The Euler equations of this variational principle are

$$\frac{d^2z}{d\sigma^2} = +\left(2-\frac{2V(z)}{E}-\frac{V'}{E}\right)e^{2z}, \quad z(0)=x', \quad z(\lambda')=x. \quad (4.46)$$

For every λ', if V is sufficiently smooth and E sufficiently large, these will possess a unique solution $\xi(\sigma,E,\lambda')$. Hence, for large E, $\hat{g}$ will be approximated by

$$\hat{g}\left(x,x';E,\frac{\lambda'}{\sqrt{E}}\right) \sim \left(\frac{\sqrt{E}}{2\pi i}\right)^{1/2}\left(\left|\frac{\partial^2 S_0[\xi(\cdot),E,\lambda']}{\partial x\partial x'}\right|\right)^{1/2} \times \exp\{i\sqrt{E}\, S_0[\xi(\cdot),\lambda',E]\}. \quad (4.47)$$

We now change variables from λ to $\lambda'/\sqrt{E}$ in the Fourier transform to obtain

$$g(x;x';E,L) = \frac{i}{\sqrt{E}}\int_0^\infty d\lambda'\, e^{-iL\lambda'/\sqrt{E}}\, g\left(x,x';E,\frac{\lambda'}{\sqrt{E}}\right) \sim \frac{i}{\sqrt{E}}\int_0^\infty d\lambda'\, e^{-iL\lambda'/\sqrt{E}}\left[\left(\frac{\sqrt{E}}{2\pi i}\right)^{1/2}\left(\left|\frac{\partial^2 S_0}{\partial x\partial x'}\right|\right)^{1/2} \times \exp(i\sqrt{E}\, S_0)\right]. \quad (4.48)$$

Due to the oscillatory nature of $\exp(i\sqrt{E}\, S_0)$, for large E this Fourier transform is dominated by neighborhoods of the critical points λ_0' of the integrand as defined by

$$-\frac{L}{E}+\left(\frac{\partial S_0}{\partial \lambda'}\right)_{\lambda'=\lambda_0'} = 0. \quad (4.49)$$

This equation is easily seen to possess two solutions, λ_d' and λ_r', corresponding to the direct path and the one that reflects from the angular momentum barrier, and g is approximated by

$$g(x,x';E,L) \sim \frac{i}{\sqrt{E}}\left[\left(\frac{\sqrt{E}}{2\pi i}\right)^{1/2}\left(\left|\frac{\partial^2 S}{\partial x\partial x'}\right|\right)^{1/2}_{\lambda'=\lambda} \exp\left(-\frac{iL\lambda_d'}{\sqrt{E}}+i\sqrt{E}\, S_0[\xi_d(\cdot),E,\lambda_d']\right) \times \int_{-\infty}^{\infty} d\eta\, \exp\left(i\,\frac{\sqrt{E}}{2}\, S_0''(\lambda_d')\eta^2\right) + \left(\left|\frac{\partial^2 S}{\partial x\partial x'}\right|\right)^{1/2}_{\lambda'=\lambda_r'} \exp\left(-\frac{iL\lambda_r'}{\sqrt{E}}+i\sqrt{E}\, S_0[\xi_r(\cdot),E,\lambda_r']\right)\int_{-\infty}^{\infty} d\eta\, \exp\left(\frac{i\sqrt{E}}{2}\, S_0''(\lambda_r')\eta^2\right)\Bigg], \quad (4.50)$$

where

$$S_0''(\lambda_{d,r}') \equiv \left[\frac{\partial^2}{\partial\lambda^2}\, S(\xi(\cdot),E,\lambda)\right]_{\lambda=\lambda_{d,r}'} \gtrless 0. \quad (4.51)$$

Evaluating these Gaussian integrals while using (4.49) to eliminate $\lambda_{d,r}'$, we obtain

$$G(x,x';E,l) \sim i[P(r)P(r')]^{-1/2} \times [\exp(i\int_{r'}^{r} P(r)dr) + e^{-i\pi/2}\exp(-i\int_{r'}^{r_0} P(r)dr + i\int_{r_0}^{r} P(r)dr)], \quad (4.52)$$

where

$$P(r) = \left[2\left(E-U(r)-\frac{(l+\frac{1}{2})^2}{r^2}\right)\right]^{1/2}, \quad (4.53)$$

and r_0 is defined by $P(r_0)=0$. The phase difference of $e^{-i\pi/2}$ between the two terms in (4.52) arises from the difference in sign of $S_0''(\lambda_{d,r})$, Eq. (4.51), forcing a different rotation of contour for each Gaussian integral. Thus we see that the potential dominance which occurs after the "Langer transform" of the radial equation is rather naturally treated by these techniques. The final formula contains the correct $(l+\frac{1}{2})^2$ factor in the local wave number.

5. STRONG SCALAR AND VECTOR POTENTIALS

In this section we consider the following more general path integral in the limit of large g:

$$K(\mathbf{r}_f,t|\mathbf{r}_I,0) \equiv \int_{P(\mathbf{r}_f,t|\mathbf{r}_I,0)} D\mathbf{z}\, \exp(i/\hbar)S[\mathbf{z}(\cdot),t,g], \quad (5.1)$$

where

$$S[\mathbf{z}(\cdot),t,g] \equiv \int_0^t \left[\tfrac{1}{2}m\left(\frac{d\mathbf{z}}{d\sigma}\right)^2 - g^2V(\sigma,\mathbf{z}(\sigma)) + ga_j(\sigma,\mathbf{z}(\sigma))\frac{dz_j}{d\sigma}\right], \quad (5.2)$$

and where V and a_i are sufficiently smooth functions of σ and $\mathbf{z}$. In the limit of large g, this is a singular problem. As in Sec. 4, we change variables to obtain

$$K(\mathbf{r}_f,\tau/g\,|\,\mathbf{r}_I,0) = \int_{P(\mathbf{r}_f,\tau|\mathbf{r}_I,0)} D\mathbf{z} \times \exp\{(ig/\hbar)S_0[\mathbf{z}(\cdot),\tau,g]\}, \quad \tau = gt \quad (5.3)$$

where

$$S_0[\mathbf{z}(\cdot),\tau,g] \equiv \int_0^\tau \left[\frac{m}{2}\left(\frac{d\mathbf{z}}{d\sigma}\right)^2 - V\left(\frac{\sigma}{g},\mathbf{z}(\sigma)\right) + a_j\left(\frac{\sigma}{g},\mathbf{z}(\sigma)\right)\frac{dz_j}{d\sigma}\right]d\sigma. \quad (5.4)$$

By the same procedures discussed above, we obtain, for fixed τ, large g,

$$K \sim \tilde{K} \equiv \left(\frac{g}{2\pi i\hbar}\right)^{1/2}\left[\left|\det\left(\frac{\partial^2 S_0(\xi)}{\partial\mathbf{r}_I\partial\mathbf{r}_f}\right)\right|\right]^{1/2} \times e^{-i\beta(\xi)\pi/2}\exp\left(\frac{ig}{\hbar}\, S_0[\xi(\cdot),\tau,g]\right), \quad (5.5)$$

where $\xi(\cdot)$ is the solution (assumed unique) of the two-point boundary value problem

$$m\frac{d^2\xi_j}{d\sigma^2} = -\frac{\partial V}{\partial\xi_j}-\frac{\partial a_j}{\partial\sigma}+\left(\frac{\partial a_k}{\partial\xi_j}-\frac{\partial a_j}{\partial\xi_k}\right)\frac{d\xi_k}{d\sigma},$$

$$\xi(0) = \mathbf{r}_I, \xi(\tau) = \mathbf{r}_f, \quad (5.6)$$

and where $\beta(\xi)$ is the total reduction in rank of the determinant along $\xi(\cdot)$.

Once again, assuming (5.5) valid through τ of $O(g)$, it may be advantageous to express the ray equations explicitly as functions of t,

$$m\epsilon^2 \frac{d^2\xi_k}{d\sigma^2} = -\frac{\partial V}{\partial \xi_k} + \epsilon\left(\frac{\partial a_j}{\partial \xi_k} - \frac{\partial a_k}{\partial \xi_j}\right)\frac{d\xi_j}{d\sigma} - \epsilon\frac{\partial a_k}{\partial \sigma},$$
$$\xi(0) = \mathbf{r}_I, \ \xi(t) = \mathbf{r}_f, \ \epsilon = 1/g. \quad (5.7)$$

One can then seek the rays explicitly for small values of the parameter ϵ by the techniques of singular perturbation theory.

In order to illustrate the calculation of the rays in several dimensions with the techniques of singular perturbation theory, we pick the simple example of motion across a parabolic barrier in the presence of a uniform B field, specifically,

$$V(\mathbf{r}) = -\tfrac{1}{2}Ex^2, \quad \mathbf{a}(\mathbf{r}) = (B/2)(-y\hat{x} + x\hat{y}), \quad (5.8)$$

where E and B are constants satisfying the inequality $E - B^2 > 0$. For these potentials the ray equations may be solved explicitly, offering a check on the singular perturbation calculations. For this special case, the ray equations (5.7) become

$$\epsilon^2 \frac{d^2x}{d\sigma^2} = Ex + \epsilon B\frac{dy}{d\sigma}, \quad (5.9)$$

$$\epsilon^2 \frac{d^2y}{d\sigma^2} = -\epsilon B\frac{dx}{d\sigma}, \quad (5.9')$$

together with the boundary conditions

$$x(0) = x_I, \ x(t) = x_f, \quad y(0) = y_I, \ y(t) = y_f. \quad (5.10)$$

Here we have set $m = 1$ and ignored the "z motion" since it is trivial.

The characteristic polynomial associated with the ray equations (5.9) is

$$\omega(\lambda) \equiv \det\begin{vmatrix} (\epsilon^2\lambda^2 - E) & -\epsilon\lambda B \\ \epsilon\lambda B & \epsilon^2\lambda^2 \end{vmatrix} = \epsilon^2\lambda^2(B^2 + \epsilon^2\lambda^2 - E), \quad (5.11)$$

the four roots of which are $0, 0, \pm[(E - B^2)/\epsilon^2]^{1/2}$. Two of these are singular as ϵ goes to zero, indicating that the solution will possess two boundary layers, a "left" layer near $\sigma = 0$ and a "right" one near $\sigma = t$.

In the "inner region" away from both boundary layers, we seek a solution of the form

$$\mathbf{R}(\sigma;\epsilon) \sim \sum_{j=0}^{\infty} \mathbf{R}_j\epsilon^j, \quad \mathbf{R} \equiv (X, Y). \quad (5.12)$$

Inserting this into the ray equations, we find that the $\mathbf{R}_j$ satisfy

$$EX_j + B\dot{Y}_{j-1} = \ddot{X}_{j-2}, \ B\dot{X}_j + \ddot{Y}_{j-1} = 0, \quad \mathbf{R}_{-1} = \mathbf{R}_{-2} \equiv 0,$$
$$j = 0, 1, 2, \ldots, \quad (5.13)$$

where dots denote derivatives with respect to σ. These possess the solutions

$$X_j = c_j, \quad Y_j = -(E/B)c_{j+1}\sigma + d_j, \quad j = 0, 1, 2, \ldots, \quad (5.14)$$

where c_0 equals 0 and the remaining c_j and d_j are constants to be determined.

Turning now to the left boundary layer, we assume a solution of the form

$$\mathbf{r}_L(\sigma_L;\epsilon) \equiv \mathbf{r}(\epsilon\sigma_L;\epsilon) - \mathbf{R}(\epsilon\sigma_L;\epsilon) \sim \sum_{j=0}^{\infty} \mathbf{r}_{Lj}\epsilon^j, \quad \sigma_L \equiv \sigma/\epsilon. \quad (5.15)$$

Inserting this ansatz into the ray equations, we discover $\mathbf{r}_{Lj} \equiv (x_{Lj}; y_{Lj})$ satisfy

$$\frac{d^2x_{Lj}}{d\sigma_L^2} = B\frac{dy_{Lj}}{d\sigma_L} + Ex_{Lj}$$
$$\frac{d^2y_{Lj}}{d\sigma_L^2} = -B\frac{dx_{Lj}}{d\sigma_L}, \ j = 0, 1, 2, \cdots. \quad (5.16)$$

The solutions of (5.16) which vanish at infinity are

$$x_{Lj} = a_{Lj}e^{-\omega\sigma_L}, \quad y_{Lj} = +(B/\omega)a_{Lj}e^{-\omega\sigma_L},$$
$$\omega = (E - B^2)^{1/2}, \quad j = 0, 1, 2, \cdots, \quad (5.17)$$

where the a_{Lj} are constants to be determined. Similarly, in the right boundary layer, the solution,

$$\mathbf{r}_R(\sigma_R) \equiv \mathbf{r}(t - \epsilon\sigma_R) - \mathbf{R}(t - \epsilon\sigma_R) \sim \sum_{j=0}^{\infty} \mathbf{r}_{Rj}\epsilon^j,$$

is found to be

$$x_{Rj} = a_{Rj}e^{-\omega\sigma_R}, \quad y_{Rj} = -(B/\omega)a_{Rj}e^{-\omega\sigma_R},$$
$$j = 0, 1, 2, \cdots, \quad (5.18)$$

where the a_{Rj} are constants to be determined.

Thus, the asymptotic behavior of the solution is given by

$$x(\sigma) \sim \sum_{j=0}^{\infty} \epsilon^j(a_{Lj}e^{-\omega\sigma/\epsilon} + c_j + a_{Rj}e^{-\omega[(t-\sigma)/\epsilon]})$$
$$y(\sigma) \sim \sum_{j=0}^{\infty} \epsilon_j\left(\frac{B}{\omega}a_{Lj}e^{-\omega\sigma/\epsilon} - \frac{E}{B}c_{j+1}\sigma + d_j - \frac{B}{\omega}a_{Rj}e^{-\omega(t-\sigma)/\epsilon}\right),$$
$$\text{as } \epsilon \to 0, \quad (5.19)$$

where $\omega \equiv +[E - B^2]^{1/2}$, $c_0 = 0$, and a_{Lj}, a_{Rj}, c_j, and d_j are constants to be determined by the boundary condition

$$\begin{pmatrix} A_j \\ B_j \\ D_j \\ E_j \end{pmatrix} = \begin{pmatrix} 1 & e^{-\omega t/\epsilon} & 0 & 0 \\ e^{-\omega t/\epsilon} & 1 & 0 & 0 \\ B/\omega & -(B/\omega)e^{-\omega t/\epsilon} & 1 & 0 \\ (B/\omega)e^{-\omega t/\epsilon} & -B/\omega & 1 & -(E/B)t \end{pmatrix}\begin{pmatrix} a_{Lj} \\ a_{Rj} \\ d_j \\ c_{j+1} \end{pmatrix}, \quad (5.20)$$

where

$$(A_j, B_j, D_j, E_j) \equiv (X_I, X_f, Y_I, Y_f), \quad j = 0,$$
$$= (-c_j, -c_j, 0, 0), \quad j = 1, 2, \cdots. \quad (5.21)$$

Inverting this system, we find the constants to be given by

$$a_{Lo} = W^{-1}(X_I - X_f e^{-\omega t/\epsilon}),$$
$$a_{Ro} = W^{-1}(X_f - X_I e^{-\omega t/\epsilon}),$$
$$d_0 = y_I - (B/\omega)W^{-1}[x_i(1 + e^{-2\omega t/\epsilon}) - 2X_f e^{-\omega t/\epsilon}],$$
$$c_1 = (B/Et)[(y_I - y_f) - (B/\omega)W^{-1}(X_I + X_f)(1 - e^{-\omega t/\epsilon})^2], \quad (5.22)$$

and

$$a_{Rj} = a_{Lj} = -W^{-1}c_j(1 - e^{-\omega t/\epsilon})$$
$$d_j = W^{-1}c_j(B/\omega)(1 - e^{-\omega t/\epsilon})^2 \quad (5.23)$$
$$c_{j+1} = 2W^{-1}c_j(B^2/\omega Et)(1 - e^{-\omega t/\epsilon})^2, \quad j = 1, 2, \cdots,$$

where $W \equiv (1 - e^{-2\omega t/\epsilon})$.

As mentioned above, the rays for this problem may be calculated explicitly, and Eqs. (5.19) are found to agree with their asymptotic expansions. We remark that for this simple case the equations valid in the boundary layers (5.16) are the complete ray equations and the only actual simplification is in the boundary conditions. This is not so in general. For example, we have calculated the rays for motion across a potential barrier in the presence of a slowly varying magnetic field (and its corresponding electric field) by these techniques. There a simplification was obtained at each step. However, since the algebra was more involved, the problem did not seem as appropriate an illustrative example as the simple problem discussed above.

It is not difficult to think of other mathematical potentials where techniques such as these could be applicable. However, we hope they could be of use in problems of direct physical interest. In addition to random scattering, one possibility is the class of problems where the potential is the sum of several, one of which is strong relative to the others. Finally we remark that, while in this paper we have treated all parameters as if they were dimensionless, in physical applications it would be necessary to convert to dimensionless expansion parameters.

ACKNOWLEDGMENTS

It is our pleasure to acknowledge many helpful conversations with Professor J. B. Keller, Professor Robert E. O'Malley, and Professor L. S. Schulman during the course of this work.

APPENDIX A

In this appendix using techniques of Keller, Lewis and Cohen[11,12] as applied to the heat equation, we obtain the asymptotic expansion of K, the solution of the initial value problem (1.1)–(1.3), through all orders in $\hbar$. The construction is formal in the sense that no mathematical proof is presented to establish that the sequence of partial sums is indeed asymptotic to K. For convenience we modify the initial condition slightly to read

$$\lim_{t \to 0^+} K(\mathbf{x}, t | \mathbf{y}, 0) = (2\pi i\hbar)^{n/2}\delta^n(\mathbf{x} - \mathbf{y}), \quad (A1)$$

and seek an asymptotic expansion of K of the form

$$K(\mathbf{x}, t | \mathbf{y}, 0) \sim \exp\left(\frac{i}{\hbar} S(\mathbf{x}, t | \mathbf{y}, 0)\right) \sum_{j=0}^{\infty} (i\hbar)^j b_j(\mathbf{x}, t | \mathbf{y}, 0). \quad (A2)$$

Placing (A2) into the Schrödinger equation and, assuming sufficient analyticity, we obtain the following set of equations for S and b_j:

$$\frac{1}{2m}(\nabla_x S)^2 + V(\mathbf{x}) + \frac{\partial S}{\partial t} = 0 \quad (A3)$$

and

$$\frac{1}{2m}\nabla_x^2 b_{j-2} + \frac{\nabla_x S \cdot \nabla_x b_{j-1}}{m} + \frac{\nabla_x^2 S}{2m} b_{j-1} + \frac{\partial}{\partial t} b_{j-1} = 0,$$
$$j = 1, 2, \cdots, \; b_{-1} \equiv 0. \quad (A4)$$

Equations (A3), (A4) may be viewed as first-order differential equations which may be solved recursively by the method of characteristics.[15] The only obvious condition imposed upon the functions S and b_j is that (A2) must satisfy (A1). The set (A3), (A4) is thus underdetermined. We shall use the exactly solvable "canonical problem" $V(\mathbf{x}) = V(\mathbf{x})|_{\mathbf{x}=\mathbf{y}} \equiv V(\mathbf{y})$ as a rationale for any additional specifications needed for uniqueness. This canonical problem has the explicit solution

$$K_c = \left(\frac{m}{t}\right)^{n/2} \exp\left[\frac{i}{\hbar}\left(\frac{m}{2}\frac{|\mathbf{x} - \mathbf{y}|^2}{t} - V(\mathbf{y})t\right)\right]. \quad (A5)$$

We solve Eq. (A3) by the method of characteristics. Define E and K by

$$E \equiv -\frac{\partial S}{\partial t} \quad \text{and} \quad \mathbf{K} \equiv \nabla_x S. \quad (A6)$$

In terms of $\mathbf{K}$ and E, Eq. (A3) becomes

$$E = (\mathbf{K}^2/2m) + V. \quad (A7)$$

The characteristic equations are then calculated to yield the ray equations,

$$\frac{dt}{d\sigma} = 1, \quad \frac{d\mathbf{x}}{d\sigma} = \frac{\mathbf{K}}{m}, \quad \frac{dE}{d\sigma} = 0, \quad \frac{d\mathbf{K}}{d\sigma} = -\nabla_x V, \quad (A8)$$

and the equation for $S(\sigma)$,

$$\frac{dS}{d\sigma} = \frac{\mathbf{K}^2}{m} - E. \quad (A9)$$

These equations constitute a system of five ordinary differential equations with complex coefficients which will enable us to solve for S along a given ray, a solution of Eqs. (A8) which satisfies the consistency condition (A7). In all, we need to specify five (complex) constants, two of which are vectors.

If we wish to identify the parameter σ with the time, one of these is eliminated. Solving the analogous ray equations for the canonical problem, and comparing the results with its exact solution (A5), indicates a convenient specification of another of these constants,

$$\mathbf{x}(\sigma = 0) = \mathbf{y}. \quad (A10)$$

Once $\mathbf{K}$ is known, the consistency condition determines the constant E. The rays are then uniquely indexed by a parameter Γ,

$$\mathbf{K}(\sigma = 0) = \boldsymbol{\Gamma}. \tag{A11}$$

Then, along the ray indexed by $\boldsymbol{\Gamma}$,

$$S_\Gamma(\sigma) = S_\Gamma(\sigma = 0) + \int_0^\sigma \left(\frac{\mathbf{K}_\Gamma^2}{2m} - V(\mathbf{x}_\Gamma)\right) d\sigma'. \tag{A12}$$

For convenience we transform to ray coordinates (σ, Γ). The Jacobian of this transformation,

$$J(\sigma, \boldsymbol{\Gamma}) \equiv \left|\frac{\partial(t, \mathbf{x})}{\partial(\sigma, \boldsymbol{\Gamma})}\right|, \tag{A13}$$

satisfies

$$\frac{dJ}{d\sigma} = J\left(\nabla_x \cdot \frac{d\mathbf{x}}{d\sigma} + \frac{\partial}{\partial t}\frac{dt}{d\sigma}\right) = J\left(\nabla_x \cdot \frac{\mathbf{K}}{m}\right). \tag{A14}$$

Now Eqs. (A4) may be written along the rays as

$$\frac{d}{d\sigma} b_{j-1} + \left(\frac{\nabla_x \cdot \mathbf{K}}{2m}\right) b_{j-1} = -\frac{1}{2m} \nabla_x^2 b_{j-2}, \qquad j = 1, 2, \cdots, \; b_{-1} = 0, \tag{A15}$$

which, when combined with Eq. (A14), yields

$$\frac{d}{d\sigma}\left[J^{1/2} b_{j-1}\right] = \frac{-J^{1/2}\nabla_x^2}{2m} b_{j-2}, \qquad j = 1, 2, \cdots, \; b_{-1} = 0. \tag{A16}$$

Equations (A16) may be integrated along the rays,

$$J^{1/2} b_{j-1}^{1/2} = -\frac{1}{2m} \int_0^\sigma (J^{1/2}\nabla_x^2 b_{j-2})\, d\sigma' + d_{j-1}, \tag{A17}$$

where $\{d_j\}$ are constants. In order to complete the construction of the asymptotic expansion, we need specify $S(\sigma = 0), \Gamma, d_j$. To do this, we again compare with the canonical problem. For it the corresponding quantities, labeled with a superscript c, are

$$S_\Gamma^c(\sigma) = \tfrac{1}{2} m\{[\mathbf{x}_\Gamma(\sigma) - \mathbf{y}]^2/\sigma\} - V(\mathbf{y})\sigma + S_\Gamma^c(\sigma = 0), \tag{A18}$$

$$[J_\Gamma^c(\sigma)\, b_0^c(\sigma)]^{1/2} = (\sigma/m)^{1/2} b_0^c(\sigma) = d_0^c, \tag{A19}$$

$$[J_\Gamma^c(\sigma)\, b_j^c(\sigma)]^{1/2} = d_j^c, \qquad i = 1, 2, \cdots. \tag{A20}$$

Comparing this with the exact solution of the canonical problem (A5), we see that a consistent specification is

$$\mathbf{x}_\Gamma(\sigma = t) = \mathbf{x} \quad \text{(which determines } \Gamma\text{)}, \tag{A21}$$

$$S_c(\sigma = 0) = 0, \tag{A22}$$

$$d_0^c = 1, \tag{A23}$$

$$d_j^c = 0, \qquad j = 1, 2, \cdots. \tag{A24}$$

Following Keller, we assume that these same specifications apply to the problem at hand. Thus, the formal asymptotic expansion of K is given by

$$K(\mathbf{x}, t|\mathbf{y}, 0) \sim e^{i/\hbar S(\mathbf{x}, t|\mathbf{y}, 0)} \sum_{j=0}^\infty (i\hbar)^j b_j(\mathbf{x}, t|\mathbf{y}, 0), \tag{A25}$$

where

$$S(\mathbf{x}, t|\mathbf{y}, 0) = \int_0^t \left[\frac{m}{2}\left(\frac{d\mathbf{z}}{d\sigma}\right)^2 - V(\mathbf{z}(\sigma))\right] d\sigma$$

and where $\mathbf{z}(\sigma)$ is the solution (assumed unique) of

$$m\frac{d^2\mathbf{z}(\sigma)}{d\sigma^2} = -\nabla_x V, \qquad \mathbf{z}(0) = \mathbf{y}, \; \mathbf{z}(t) = \mathbf{x}. \tag{A26}$$

The b_j are given in terms of the ray $\mathbf{z}(\sigma)$ by

$$b_0 = (J^{-1/2})_{\sigma=t} = \left[\det\left(\frac{\partial \mathbf{K}(\sigma = 0)}{\partial \mathbf{x}(\sigma = 0)}\right)\right]^{1/2} = \left[\det\left(\frac{\partial S(\mathbf{x}, t|\mathbf{y}, 0)}{\partial x_i \partial y_j}\right)\right]^{1/2} \tag{A27}$$

and

$$b_j = -J^{-1/2} \int_0^t \left(\frac{J^{1/2}\nabla_x^2 b_{j-1}}{2m}\right) d\sigma, \tag{A28}$$

where the integration is performed along the ray $\mathbf{z}(\sigma)$. We remark that if the potential is complex, the ray will be complex valued and that the leading term of (A25) is in agreement with Eq. (1.12). Also, this construction will only be valid until the first caustic is reached. One would then be forced to perform a local analysis in the vicinity of the caustic.

APPENDIX B

In this appendix we consider the Schrödinger equation in the form

$$i\hbar \frac{\partial}{\partial t} K(\mathbf{x}, t|\mathbf{y}, 0) = -\frac{\hbar^2}{2m}\nabla_x^2 K(\mathbf{x}, t|\mathbf{y}, 0) + [V(\mathbf{x}) + \hbar\tilde{V}(\mathbf{x})]K(\mathbf{x}, t|\mathbf{y}, 0), \tag{B1}$$

together with the initial condition

$$\lim_{t\to 0^+} K(\mathbf{x}, t|\mathbf{y}, 0) = (2\pi i\hbar)^{n/2}\delta^n(\mathbf{x} - \mathbf{y}). \tag{B2}$$

Here $V(\mathbf{x})$ and $\tilde{V}(\mathbf{x})$ are (possibly) complex functions of $\mathbf{x}$. We seek a solution of the form

$$K(\mathbf{x}, t|\mathbf{y}, 0) = \exp[(i/\hbar)S(\mathbf{x}, t|\mathbf{y}, 0)], \tag{B3}$$

where S is a complex function which may be expressed asymptotically as a power series in $\hbar$,

$$S(\mathbf{x}, t|\mathbf{y}, 0) \sim \sum_{j=0}^\infty \hbar^j S_j(\mathbf{x}, t|\mathbf{y}, 0). \tag{B4}$$

Placing (B4) and (B3) into (B1) and equating coefficients of like powers of $\hbar$ results in the following system of equations:

$$-\frac{\partial}{\partial t} S_0 = \frac{1}{2m}(\nabla_x S_0)^2 + V, \tag{B5}$$

$$-\frac{\partial}{\partial t} S_1 = -\frac{i}{2m}\nabla_x^2 S_0 + \frac{1}{m}\nabla_x S_0 \cdot \nabla_x S_1 + \tilde{V}, \tag{B6}$$

$$-\frac{\partial}{\partial t} S_j = -\frac{i}{2m}\nabla_x^2 S_{j-1} + \frac{1}{m}\nabla_x S_0 \cdot \nabla_x S_j + \frac{1}{2m}\sum_{k=1}^{j-1} (\nabla_x S_k)\cdot(\nabla_x S_{j-k}), \qquad j \in (2, 3, \cdots). \tag{B7}$$

Equation (B5) is identical with (A3), so that equations (A6)–(A9) again apply. Notice that only V and not $\tilde{V}$ enters these equations. Using the same canonical problem as in Appendix A, we find that the same rays, indexed by Γ, once again apply. Integrating (B5) along these rays, we obtain

$$S_0^\Gamma(\sigma) = \int_0^\sigma \left(\frac{\mathbf{K}_\Gamma^2(\sigma')}{2m} - V(\mathbf{x}_\Gamma)\right) d\sigma' + S_0^\Gamma(\sigma = 0). \tag{B8}$$

To integrate (B6) along a ray, we realize that it is equivalent to

$$\frac{d}{d\sigma} S_1(\mathbf{x}_\Gamma(\sigma), \sigma | \mathbf{y}, 0) = \frac{i}{2m} \nabla^2 S_0 - \tilde{V}. \tag{B9}$$

If we define $A_1(\mathbf{x}_\Gamma(\sigma), \sigma | \mathbf{y}, 0)$ by

$$-i \ln A_1 = S_1 + \int_0^\sigma \tilde{V}(\mathbf{x}_\Gamma(\sigma'))\, d\sigma', \tag{B10}$$

then it satisfies

$$\frac{dA_1}{d\sigma} = -\frac{1}{2m} (\nabla_x^2 S_0) A_1. \tag{B11}$$

But (B11) is the same equation as (A4), and thus its solution is

$$J^{1/2} A_1^{1/2} = d_0. \tag{B12}$$

Similarly Eqs. (B7) may be integrated along the rays,

$$S_j(\mathbf{x}_\Gamma(\sigma), \sigma | \mathbf{y}, 0) = \int_0^\sigma f_j(\sigma')\, d\sigma' + S_j(\sigma = 0), \qquad j = 2, 3, \cdots, \tag{B13}$$

where

$$f_j(\sigma) \equiv \frac{i}{2m} \nabla_x^2 S_{j-1} - \frac{1}{2m} \sum_{k=1}^{j-1} (\nabla_x S_k) \cdot (\nabla_x S_{j-k}). \tag{B14}$$

Once again the various constants are specified by the initial condition and the canonical problem. Picking that ray ξ which passes through $\mathbf{x}$ at $\sigma = t$, we find the leading term to be

$$K(\mathbf{x}, t | \mathbf{y}, 0) \underset{\hbar \to 0}{\sim} \left[\det\left(\frac{\partial S(\mathbf{x}, t | \mathbf{y}, 0)}{\partial x_i \partial y_j} \right) \right]^{1/2} \times \exp\left\{ \frac{i}{\hbar} \int_0^t \left[\frac{m}{2} \left(\frac{d\xi}{d\sigma} \right)^2 - V(\xi) \right] d\sigma \right\} \times \exp\left(-i \int_0^t \tilde{V}(\xi(\sigma))\, d\sigma \right). \tag{B15}$$

We remark that the rays are determined only by the dominant part of the potential V. If V is real and $\tilde{V}$ is pure imaginary, this appendix applies to Sec. 2. If V and $\tilde{V}$ are both real, it applies to Sec. 4. Again (B15) is valid only until the first caustic is reached. Since the path integral is a convenient tool to treat caustics, we will not perform here the local analysis needed to pass through one. Finally we selected $\tilde{V}$ linear in $\hbar$ only for convenience. We could have been considering a double asymptotic expansion in both $\hbar$ and the strength of $\tilde{V}$, say g.

APPENDIX C

In this appendix we show that the large g approximation of Sec. 4 is the first term of an asymptotic expansion in the parameter g^{-1}. Consider the Schrödinger equation with delta function data,

$$\left(-\tfrac{1}{2} \nabla_x^2 + g^2 V(\mathbf{x}) - i \frac{\partial}{\partial t} \right) K(\mathbf{x}, t | \mathbf{y}, 0) = 0, \qquad t > 0, \tag{C1}$$

$$\lim_{t \to 0^+} K(\mathbf{x}, t | \mathbf{y}, 0) = (2\pi i)^{n/2} \delta^n(\mathbf{x} - \mathbf{y}). \tag{C2}$$

Dividing Eq. (C1) by g^2 and making the change of variables $\tau = gt$, we find

$$\left(-\frac{1}{2g^2} \nabla_x^2 + V(x) - \frac{i}{g} \frac{\partial}{\partial \tau} \right) K(\mathbf{x}, \tau | \mathbf{y}, 0) = 0. \tag{C3}$$

For fixed τ, large g, Eq. (C3) is of the form studied in Appendix A; hence, $K(\mathbf{x}, \tau | \mathbf{y}, 0)$ possesses an expansion of the form

$$K(\mathbf{x}, \tau | \mathbf{y}, 0) \sim \exp(igS(\mathbf{x}, \tau | \mathbf{y}, 0)) \sum_{j=0}^{\infty} g^{-j} A_j(\mathbf{x}, \tau | \mathbf{y}, 0), \qquad g \to \infty. \tag{C4}$$

The calculations of Appendix A show that the first term of (C4) is in agreement with Eq. (4.10).

* Work supported in part by a National Science Foundation predoctoral fellowship which the author held at Indiana University and in part by the Air Force Office of Scientific Research, under Grant No. AFOSR-71-2013.

1 R. Feynman, Rev. Mod. Phys. **20**, 367 (1948).
2 R. Feynman, Phys. Rev. **80**, 440 (1950).
3 R. Feynman, Phys. Rev. **97**, 3, 660 (1955).
4 C. Morette, Phys. Rev. **81**, 848 (1951).
5 M. Gutzwiller, J. Math. Phys. **8**, 10, 1979 (1967).
6 P. Pechukas, Phys. Rev. **181**, 1, 166 (1969).
7 I. M. Gel'fand and A. M. Yaglom, J. Math. Phys. **1**, 48 (1960).
8 V. S. Buslaev, "Continuum Integrals and the Asymptotic Behavior of the Solutions of Parabolic Equations as $t \to 0$. Application to Diffraction," in *Topics in Mathematical Physics*, edited by Sh. Berman, English transl. (Consultants Bureau, New York, 1967).
9 P. L. Chow, "Function Space Integration and Random Scattering," to be published.
10 W. Pauli, "Feldquantisierung," lecture notes Zurich (1950-1951), Appendix.
11 J. Cohen and R. M. Lewis, J. Inst. Math. Appl. **3**, 266 (1967).
12 J. Cohen, F. Hagen, and J. B. Keller, J. Math. Anal. Appl. (to be published).
13 R. E. O'Malley, Advan. Math. **2**, 4, 365 (1968).
14 R. Langer, Phys. Rev. **51**, 669 (1937).
15 R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol. II.

Generalized Mechanics

Herbert Tesser
Department of Physics, Pratt Institute, Brooklyn, New York
(Received 29 April 1970; Revised Manuscript Received 10 December 1970)

The quantization scheme for generalized Hamiltonians recently proposed by Hayes is shown to be inconsistent. The theory of constraints developed by Dirac is used and leads to a consistent theory.

1. INTRODUCTION

In recent years several attempts[1] have been made to extend the mechanics of systems described by Lagrangians containing second and higher order derivatives. In one of these works, Hayes[2] has concentrated on the quantization of such systems. We will show that the techniques used by Hayes lead to a theory in which the Hamiltonian is no longer the generator of time translation for the system, and for which canonical quantization cannot be performed. The constraint

To integrate (B6) along a ray, we realize that it is equivalent to

$$\frac{d}{d\sigma} S_1(\mathbf{x}_\Gamma(\sigma), \sigma | \mathbf{y}, 0) = \frac{i}{2m} \nabla^2 S_0 - \tilde{V}. \tag{B9}$$

If we define $A_1(\mathbf{x}_\Gamma(\sigma), \sigma | \mathbf{y}, 0)$ by

$$-i \ln A_1 = S_1 + \int_0^\sigma \tilde{V}(\mathbf{x}_\Gamma(\sigma'))\, d\sigma', \tag{B10}$$

then it satisfies

$$\frac{dA_1}{d\sigma} = -\frac{1}{2m} (\nabla_x^2 S_0) A_1. \tag{B11}$$

But (B11) is the same equation as (A4), and thus its solution is

$$J^{1/2} A_1^{1/2} = d_0. \tag{B12}$$

Similarly Eqs. (B7) may be integrated along the rays,

$$S_j(\mathbf{x}_\Gamma(\sigma), \sigma | \mathbf{y}, 0) = \int_0^\sigma f_j(\sigma')\, d\sigma' + S_j(\sigma = 0), \qquad j = 2, 3, \cdots, \tag{B13}$$

where

$$f_j(\sigma) \equiv \frac{i}{2m} \nabla_x^2 S_{j-1} - \frac{1}{2m} \sum_{k=1}^{j-1} (\nabla_x S_k) \cdot (\nabla_x S_{j-k}). \tag{B14}$$

Once again the various constants are specified by the initial condition and the canonical problem. Picking that ray ξ which passes through $\mathbf{x}$ at $\sigma = t$, we find the leading term to be

$$K(\mathbf{x}, t | \mathbf{y}, 0) \underset{\hbar \to 0}{\sim} \left[\det\left(\frac{\partial S(\mathbf{x}, t | \mathbf{y}, 0)}{\partial x_i \partial y_j} \right) \right]^{1/2} \times \exp\left\{ \frac{i}{\hbar} \int_0^t \left[\frac{m}{2} \left(\frac{d\xi}{d\sigma} \right)^2 - V(\xi) \right] d\sigma \right\} \times \exp\left(-i \int_0^t \tilde{V}(\xi(\sigma))\, d\sigma \right). \tag{B15}$$

We remark that the rays are determined only by the dominant part of the potential V. If V is real and $\tilde{V}$ is pure imaginary, this appendix applies to Sec. 2. If V and $\tilde{V}$ are both real, it applies to Sec. 4. Again (B15) is valid only until the first caustic is reached. Since the path integral is a convenient tool to treat caustics, we will not perform here the local analysis needed to pass through one. Finally we selected $\tilde{V}$ linear in $\hbar$ only for convenience. We could have been considering a double asymptotic expansion in both $\hbar$ and the strength of $\tilde{V}$, say g.

APPENDIX C

In this appendix we show that the large g approximation of Sec. 4 is the first term of an asymptotic expansion in the parameter g^{-1}. Consider the Schrödinger equation with delta function data,

$$\left(-\tfrac{1}{2} \nabla_x^2 + g^2 V(\mathbf{x}) - i \frac{\partial}{\partial t} \right) K(\mathbf{x}, t | \mathbf{y}, 0) = 0, \qquad t > 0, \tag{C1}$$

$$\lim_{t \to 0^+} K(\mathbf{x}, t | \mathbf{y}, 0) = (2\pi i)^{n/2} \delta^n(\mathbf{x} - \mathbf{y}). \tag{C2}$$

Dividing Eq. (C1) by g^2 and making the change of variables $\tau = gt$, we find

$$\left(-\frac{1}{2g^2} \nabla_x^2 + V(x) - \frac{i}{g} \frac{\partial}{\partial \tau} \right) K(\mathbf{x}, \tau | \mathbf{y}, 0) = 0. \tag{C3}$$

For fixed τ, large g, Eq. (C3) is of the form studied in Appendix A; hence, $K(\mathbf{x}, \tau | \mathbf{y}, 0)$ possesses an expansion of the form

$$K(\mathbf{x}, \tau | \mathbf{y}, 0) \sim \exp(igS(\mathbf{x}, \tau | \mathbf{y}, 0)) \sum_{j=0}^{\infty} g^{-j} A_j(\mathbf{x}, \tau | \mathbf{y}, 0), \qquad g \to \infty. \tag{C4}$$

The calculations of Appendix A show that the first term of (C4) is in agreement with Eq. (4.10).

* Work supported in part by a National Science Foundation predoctoral fellowship which the author held at Indiana University and in part by the Air Force Office of Scientific Research, under Grant No. AFOSR-71-2013.

1 R. Feynman, Rev. Mod. Phys. **20**, 367 (1948).
2 R. Feynman, Phys. Rev. **80**, 440 (1950).
3 R. Feynman, Phys. Rev. **97**, 3, 660 (1955).
4 C. Morette, Phys. Rev. **81**, 848 (1951).
5 M. Gutzwiller, J. Math. Phys. **8**, 10, 1979 (1967).
6 P. Pechukas, Phys. Rev. **181**, 1, 166 (1969).
7 I. M. Gel'fand and A. M. Yaglom, J. Math. Phys. **1**, 48 (1960).
8 V. S. Buslaev, "Continuum Integrals and the Asymptotic Behavior of the Solutions of Parabolic Equations as $t \to 0$. Application to Diffraction," in *Topics in Mathematical Physics*, edited by Sh. Berman, English transl. (Consultants Bureau, New York, 1967).
9 P. L. Chow, "Function Space Integration and Random Scattering," to be published.
10 W. Pauli, "Feldquantisierung," lecture notes Zurich (1950-1951), Appendix.
11 J. Cohen and R. M. Lewis, J. Inst. Math. Appl. **3**, 266 (1967).
12 J. Cohen, F. Hagen, and J. B. Keller, J. Math. Anal. Appl. (to be published).
13 R. E. O'Malley, Advan. Math. **2**, 4, 365 (1968).
14 R. Langer, Phys. Rev. **51**, 669 (1937).
15 R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol. II.

Generalized Mechanics

Herbert Tesser
Department of Physics, Pratt Institute, Brooklyn, New York
(Received 29 April 1970; Revised Manuscript Received 10 December 1970)

The quantization scheme for generalized Hamiltonians recently proposed by Hayes is shown to be inconsistent. The theory of constraints developed by Dirac is used and leads to a consistent theory.

1. INTRODUCTION

In recent years several attempts[1] have been made to extend the mechanics of systems described by Lagrangians containing second and higher order derivatives. In one of these works, Hayes[2] has concentrated on the quantization of such systems. We will show that the techniques used by Hayes lead to a theory in which the Hamiltonian is no longer the generator of time translation for the system, and for which canonical quantization cannot be performed. The constraint

theory developed by Dirac[3] will be used to illustrate how a consistent theory is constructed.

2. LAGRANGIAN THEORY

For a Lagrangian depending upon variables q_j and its derivatives $q_j^m = d^m q_j/dt^m$, Ostrogradsky[4] formulated a technique which yields the equations of motion, the generalized moments, and the Hamiltonian of the system. We consider the action

$$S = \int_{t_1}^{t_2} L\left(q_j, \dot{q}_j, \ldots, \frac{d^N q_j}{dt^N}\right) dt, \tag{2.1}$$

where $j = 1, 2, \ldots, R$. The independent variables are the q_j and the derivatives q_j^m. Variation of the action, holding the endpoints fixed, leads to the equations of motion

$$\sum_{n=0}^{N} (-1)^n \frac{d^n}{dt^n}\left(\frac{\partial L}{\partial q_j^n}\right) = 0. \tag{2.2}$$

The appearance of $\ddot{q}_j$ and higher order derivatives forces us to use the method of Ostrogradsky in constructing the Hamiltonian. The Ostrogradsky method treats the variables

$$q_j^l \Leftrightarrow P_{j,l} = \sum_{i=0}^{N-l-1} (-1)^i \frac{d^i}{dt^i}\left(\frac{\partial L}{\partial q_j^{i+l+1}}\right), \quad l \leq N-1, \tag{2.3}$$

as independent coordinates and their canonical momenta. The Hamiltonian is defined by

$$H = \sum_{j=1}^{R} \sum_{l=0}^{N-1} (P_{j,l} q_j^{l+1}) - L. \tag{2.4}$$

Hayes claims that the Hamiltonian is a function of $2N + 1$ variables, N momenta, the generalized coordinate, and N derivatives of the generalized coordinate. However, direct calculation shows that the Hamiltonian is a function only of the $N - 1$ derivatives of the generalized coordinates. We have from Eq. (2.4),

$$\frac{\partial H}{\partial q_j^N} = P_{j,N-1} - \frac{\partial L}{\partial q_j^N}. \tag{2.5}$$

From Eq. (2.3) we see that the definition of $P_{j,N-1}$ is

$$P_{j,N-1} = \frac{\partial L}{\partial q_j^N}.$$

Thus

$$\frac{\partial H}{\partial q_j^N} = 0.$$

As Hayes points out, there is no guarantee that Eq. (2.3) can be used to eliminate the highest derivative of the q_j from the Hamiltonian. If q_j^N cannot be explicitly eliminated from the Hamiltonian, then it must appear as the multiplier of a constraint. We shall return to the question of the constraint problem in Sec. 4.

Ostrogradsky's definition of the momenta, coordinates, and Hamiltonian ensures that the usual relations

$$\frac{\partial H}{\partial q_j^m} = -\dot{P}_{j,m}, \tag{2.6a}$$

$$\frac{\partial H}{\partial P_{j,m}} = \dot{q}_j^m \tag{2.6b}$$

are valid for the canonical pairs. Equations (2.6) follow from the relation

$$\dot{P}_{j,m} + P_{j,m-1} = \frac{\partial L}{\partial q_j^m}. \tag{2.7}$$

3. GENERALIZED COMMUTATORS

Hayes proposes that the commutation relations among the canonical variables be generalized. He postulates that

$$[P_{j,n}, q_k^m] = -i\hbar C(n,N)\delta_{jk}\delta_{nm}, \tag{3.1}$$

where $C(n,N)$ is a c number and is to be chosen "such that the usual uncertainty relation holds."

This change in the commutation relations necessitates a change in the form of the Schrödinger equation. Consider the expectation value of an operator $\hat{O}$

$$\langle \hat{O} \rangle = \int \prod_{j=1}^{R} \prod_{m=0}^{N-1} dq_j^m [\psi^* \hat{O} \psi]. \tag{3.2}$$

We have

$$\frac{d\langle \hat{O} \rangle}{dt} = \int \prod_{j=1}^{R} \prod_{m=0}^{N-1} dq_j^m \left[\dot{\psi}^* \hat{O} \psi + \psi^* \hat{O} \dot{\psi} + \psi^* \frac{\partial O}{\partial t} \psi\right]. \tag{3.3}$$

Take $\hat{O}$ to be a linear combination of the $P_{j,m}$ and q_j^m. It is clear that the Schrödinger equation

$$H\psi = i\hbar \frac{\partial \psi}{\partial t} \tag{3.4}$$

or a simple generalization involving the $C(n,N)$, will not properly describe the time-dependence of the wavefunction. If all the $C(n,N)$ are equal, we could obtain a proper theory by replacing $\hbar$ by $\hbar C(n,N)$ in (3.4), but this amounts to merely rescaling $\hbar$. A simple example where rescaling does not work is described by the Lagrangian for two noninteracting particles:

$$L = \sum_i \left[\left(-\frac{m_1}{2}\right)\ddot{x}_i x_i + \frac{m_2}{2}\dot{y}_i^2\right]. \tag{3.5}$$

We now investigate the consequences of the commutation relations proposed by Eq. (3.1). Let us suppose that there are no constraints (or that they have been removed by the method suggested in Sec. 4). The Hayes and Ostrogradsky variables may be related by the linear transformation

$$\hat{P}_{Hn} = \hat{P}_{On}, \qquad \hat{Q}_{Hn} = C(n,N)\hat{Q}_{On}, \tag{3.6}$$

where ${}_H\hat{Q}$ and ${}_H\hat{P}$ (${}_O\hat{Q}$ and ${}_O\hat{P}$) are the Hayes (Ostrogradsky) position and momentum operators and other subscripts have been suppressed. See Hayes, Ref. 2, Eqs. (19) and (20).

Hayes commutation relations are then satisfied by the variables ${}_HQ_n$ and ${}_HP_n$. Since the $C(n,N)$ are c, numbers the rules of the algebra are unaffected by the transformation. Furthermore, the vectors of the Hilbert space are unaffected by the transformation $|{}_H Q\rangle = |{}_O Q\rangle$. To complete the study of the Hayes variables, we need only construct the algebra table for the new variables.

Suppose the vectors ${}_O\hat{Q}$ are eigenvectors of the ${}_OQ_n$.

Trivially, then, the elements of the Hayes matrices are

$$\langle Q_{n'}|\underset{H}{\hat{Q}_n}|Q_n\rangle = \underset{H}{Q}\,\delta_{nn'}, \tag{3.7a}$$

$$\langle Q_{n'}|\underset{H}{\hat{P}_n}|Q_n\rangle = -i\hbar C(n,N)\frac{\partial}{\partial Q_n}\delta(Q_n - Q_{n'}). \tag{3.7b}$$

The multiplication and addition tables follow immediately.

4. CONSISTENCY OF THE THEORY

If a Lagrangian contains high order derivatives and Ostrogradsky's method is used, we find that constraint relations exist among the canonical variables. Furthermore, some of these constraints do not commute. The constraints act to reduce the size of the Hilbert space of state vectors describing the system. Dirac argues that noncommuting (second class) constraints must be eliminated from the theory, since such constraints can generate unphysical vectors in the Hilbert space.

The extra variables are eliminated by transforming one constraint to a canonical variable γ_2. The other constraint then also transforms. This constraint is solved for π_2, the variable conjugate to γ_2. The pair of constraints are eliminated by substituting for π_2 and $\gamma_2 = 0$ in the new Hamiltonian.

Suppose we have two constraints G_1 and G_2,

$$G_1 = G_1[P_n^j - U(Q_{m,j}, P_r^j)] = 0, \qquad r, m \neq n \tag{4.1a}$$

and

$$G_2 = G_2[Q_{n,j} - V(Q_{m,j}, P_r^j)] = 0, \quad r, m \neq n,$$
$$[G_1, G_2] \neq 0. \tag{4.1b}$$

Here U and V are functions of the coordinates and moments other than P_n^j and $Q_{n,j}$.

Make the transformation

$$W = \pi_2 G_2 + W',$$

where W' is not a function of π_2 but completes the description of the transformation.

Then

$$\gamma_2 = \frac{\partial W}{\partial \pi_2} = G_2 = 0. \tag{4.2}$$

In general π_2 will not be zero. However, the constraint G_1 will also transform under W,

$$G_1(P_n^j, P_m^j, \ldots, Q_{n,j}, Q_{m,j} \cdots)$$
$$\Rightarrow G_1(F(\pi_{n,j}, \pi_{m,j}, \ldots, \gamma_{n,j}, \gamma_{m,j} \cdots)) = 0. \tag{4.3}$$

π_2 is removed from the theory by solving Eq. (4.3) for π_2.

The substitutions of $\gamma_2 = 0, \pi_2 = f(\pi_m, \gamma^n)$ are made in the transformed Hamiltonian reducing the number of canonical pairs by one.

As a specific example, we consider the Lagrangian introduced by Hayes,

$$L = -\tfrac{1}{2}\sum_{j=1}^{3} m x_j \ddot{x}_j - V(x_1, x_2, x_3), \tag{4.4}$$

which generates the familiar equation of motion

$$m\ddot{x}_j = -\frac{\partial V}{\partial x_j}. \tag{4.5}$$

The momenta are

$$P_{j,0} = \tfrac{1}{2} m \dot{x}_j \tag{4.6a}$$

and

$$P_{j,1} = -\tfrac{1}{2} m \dot{x}_j] \tag{4.6b}$$

Since x_j and $\dot{x}_j$ are considered to be independent variables, Eqs. (4.6) are constraints. Let

$$G_1 = [P_{j,0} - \tfrac{1}{2} m \dot{x}_j] = 0 \tag{4.7a}$$

and

$$G_2 = [P_{j,1} + \tfrac{1}{2} m x_j] = 0. \tag{4.7b}$$

For the system described by the Lagrangian (4.4), Hayes evaluates $C(0,2)$ and $C(1,2)$ and finds

$$C(0,2) = C(1,2) = \tfrac{1}{2}.$$

Thus, we have

$$[G_1, G_2] \neq 0. \tag{4.8}$$

If the system is to be quantized, we see that there can exist no state vector for which (4.7a), (4.7b), and (4.8) are valid operator equations. To eliminate the extra variables, we will transform the constraints using the method outlined above. The Lagrangian (4.4) gives rise to the Hamiltonian

$$H = \sum_j [P_{j,0}\dot{x}_j + \ddot{x}_j(P_{j,1} + \tfrac{1}{2} m x_j)] + V(x_1, x_2, x_3). \tag{4.9}$$

We first perform the transformation $(\dot{x}_j \to P_{j,1}) \to (Q_{j,1}, \pi_{j,1})$.

As the generating function, we choose

$$W_1 = \sum_j \frac{(P'_{j,1} P_{j,1})}{m}. \tag{4.10}$$

The transformation equations

$$Q'_{j,1} = \frac{\partial W_1}{\partial P'_{j,1}}, \qquad \dot{x}_j = -\frac{\partial W}{\partial P_{j,1}} \tag{4.11}$$

yield

$$P'_{j,1} = -m\dot{x}_j, \tag{4.12a}$$

$$Q'_{j,1} = P_{j,1}/m. \tag{4.12b}$$

The second transformation from variables $(P'_{j,1}, Q'_{j,1}, P_{j,0}, x_j)$ to variables $(\pi''_j, \gamma''_j, \pi'_j, \gamma'_j)$ is generated by the transformation

$$W_2 = \pi'_j(Q'_{j,1} + \tfrac{1}{2} x_j) + \pi''_j(Q'_{j,1} - \tfrac{1}{2} x_j). \tag{4.13}$$

Upon solving the transformation equations, we have

$$\gamma'_j = -(Q'_{j,1} + \tfrac{1}{2} x_j), \tag{4.14a}$$

$$\gamma''_j = -(Q'_{j,1} - \tfrac{1}{2} x_j), \tag{4.14b}$$

$$P'_{j,1} = \pi'_j + \pi''_j, \tag{4.14c}$$

and

$$P_{j,0} = \tfrac{1}{2}(\pi'_j - \pi''_j). \tag{4.14d}$$

Solving Eqs. (4.12c) and (4.12d) for π'_j and π''_j, we obtain

$$\pi_j'' = \tfrac{1}{2}P_{j,1}' - P_{j,0} \tag{4.15a}$$

$$\pi_j' = \tfrac{1}{2}P_{j,1}' + P_{j,0} \tag{4.15b}$$

From the constraint relations, we see that $\gamma_j' = 0$ and $\pi_j' = 0$. Thus, using Eqs. (4.10) and (4.13), we have transformed the constraints into a conjugate pair of variables. The theory is made complete when γ_j' and π_j' are set to zero in the new Hamiltonian,

$$H' = \sum_j [-\tfrac{1}{2}(\pi_j' - \pi_j'')(\pi_j' + \pi_j'') - \ddot{x}_j\gamma_j'] + V[(\gamma_j'' - \gamma_j')].$$

We obtain the usual Hamiltonian without constraints, [5]

$$H' = \sum_j \frac{P_j''^2}{2m} + V(\gamma_j''). \tag{4.16}$$

5. CONCLUSION

We have shown that the formalism developed by Hayes to treat generalized Hamiltonians is not consistent. The Dirac theory of constraints must be used to eliminate superfluous variables and to generate a consistent theory

Since the variables P_j'' and γ_j'' obey the usual commutation relations, we can retain the usual Schrödinger equation with Eq. (4.16) defining the Hamiltonian operator.

---

[1] See, for example, M. Borneas Phys. Rev. **186**, 1299 (1969), and references contained therein.
[2] C. F. Hayes, J. Math. Phys. **10**, 1555 (1969).
[3] P. A. M. Dirac, Can. J. Math. **2**, 129 (1950); **3**, 1 (1951).
[4] E. T. Whittaker, *Analytical Dynamics* (Cambridge U. P., Cambridge, 1937), p. 265.
[5] This form of the Hamiltonian is unique. The ambiguities demonstrated by C. F. Hayes and J. M. Jankowski [Nuovo Cimento **58**, 494 (1968)] arise because they make transformations on the Hamiltonian which are not canonical.

Noncommutative Probability on von Neumann Algebras

S. Gudder and J.-P. Marchand*
Department of Mathematics, University of Denver, Denver, Colorado 80210
(Received 11 November 1971; Revised Manuscript Received 21 January 1972)

We generalize ordinary probability theory to those von Neumann algebras $\mathbb{A}$, for which Dye's generalized version of the Radon–Nikodym theorem holds. This includes the classical case in which $\mathbb{A}$ is an Abelian von Neumann algebra generated by an observable or complete set of commuting observables. Via Gleason's theorem, this also includes the case of ordinary quantum mechanics, in which $\mathbb{A} = \mathbb{B}(H)$ is the von Neumann algebra of all bounded operators on a separable Hilbert space H. Particular consideration is given to the concepts of conditioning, sufficient statistics, coarse-graining, and filtering.

1. INTRODUCTION

In two fundamental papers Dye[1] and Segal[2] have introduced models of noncommutative probability and integration theories. In both models a pair $(\mathbb{A}, w)$ consisting of a measure algebra $\mathbb{A}$ and a measure w on $\mathbb{A}$ have been specified. But while Segal places restrictions on the "gage" w, requiring it to be unitarily invariant, Dye requires the algebra $\mathbb{A}$ to be a σ-finite and finite von Neumann algebra. Umegaki[3] and his collaborators have carried these theories to a considerable extent. However, they continue to impose invariance and finiteness restrictions.

Our purpose here is to investigate the non-Boolean probability theory of ordinary quantum mechanics in a separable Hilbert space H. $\mathbb{A}$ is therefore (as least in the absence of superselection rules) the von Neumann algebra $\mathbb{B}(H)$ of all bounded linear operators on H, and the measures are the countably additive positive linear functionals w on $\mathbb{A}$. In this context, neither Dye's or Segal's frameworks appear to be suitable. In Dye's case $\mathbb{B}(H)$ is not finite (unless H is finite dimensional). In Segal's case there are no nontrivial bounded gages (again, unless H is finite dimensional) and the only unbounded gages on $\mathbb{B}(H)$ are proportional to the trace functional τ, all one-dimensional projections in $\mathbb{B}(H)$ being unitarily equivalent. In this latter case τ cannot be normalized to obtain a probability measure (state) so that a probability theory is out of the question. On the other hand, although $\mathbb{B}(H)$ is not finite if H is infinite dimensional, Dye's main result, namely the generalized Radon–Nikodym theorem, remains true for $\mathbb{B}(H)$. As we shall show, this follows using a result of Gleason.[4]

We shall derive our results as far as possible for an arbitrary von Neumann algebra $\mathbb{A}$ and arbitrary measures w, while including in each section their applications to the "extreme" cases, $\mathbb{A}$ Abelian (classical case) and $\mathbb{A} = \mathbb{B}(H)$ (pure quantum case). These results briefly are as follows.

In Sec. 2 we combine the theorems of Dye and Gleason which leads to a relation between the density operators of quantum mechanics and the Radon–Nikodym derivative in Dye's sense. Some of the results of this section are not new[5]; their derivation is, however, simple enough to justify a restatement of the proofs.

In Sec. 3 we generalize the notion of conditioning with respect to a von Neumann subalgebra $\mathbb{B}$. For reasons of physical relevance and mathematical simplicity, particular attention is given to the case of an Abelian $\mathbb{B}$. This then leads to severe restrictions on the structure of $\mathbb{B}$ and w. It is shown that conditioning with respect to Abelian $\mathbb{B}$ exists if and only if $\mathbb{B}$ contains a discrete part and the density operator W of w commutes with $\mathbb{B}$. In this case the conditional expectation of w given $\mathbb{B}$ appears as a Radon–Nikodym derivative very much like in classical probability.

In Sec. 4 we consider the concepts of equivalence of states relative to a subalgebra $\mathbb{B}$ and of sufficient statistics. Physically, $\mathbb{B}$ may be interpreted as a partial measurement (for instance, measurement of the macroscopic observables in statistical mechanics). $\mathbb{B}$-equivalence then leads to the concept of macrocells, and $\mathbb{B}$ is a sufficient statistic for those states which have equal $\mathbb{B}$-expectations.[6] We show that $\mathbb{B}$-equivalence and $\mathbb{B}$-sufficiency are comple-

$$\pi_j'' = \tfrac{1}{2}P_{j,1}' - P_{j,0} \tag{4.15a}$$

$$\pi_j' = \tfrac{1}{2}P_{j,1}' + P_{j,0} \tag{4.15b}$$

From the constraint relations, we see that $\gamma_j' = 0$ and $\pi_j' = 0$. Thus, using Eqs. (4.10) and (4.13), we have transformed the constraints into a conjugate pair of variables. The theory is made complete when γ_j' and π_j' are set to zero in the new Hamiltonian,

$$H' = \sum_j [-\tfrac{1}{2}(\pi_j' - \pi_j'')(\pi_j' + \pi_j'') - \ddot{x}_j\gamma_j'] + V[(\gamma_j'' - \gamma_j')].$$

We obtain the usual Hamiltonian without constraints, [5]

$$H' = \sum_j \frac{P_j''^2}{2m} + V(\gamma_j''). \tag{4.16}$$

5. CONCLUSION

We have shown that the formalism developed by Hayes to treat generalized Hamiltonians is not consistent. The Dirac theory of constraints must be used to eliminate superfluous variables and to generate a consistent theory

Since the variables P_j'' and γ_j'' obey the usual commutation relations, we can retain the usual Schrödinger equation with Eq. (4.16) defining the Hamiltonian operator.

---

[1] See, for example, M. Borneas Phys. Rev. **186**, 1299 (1969), and references contained therein.
[2] C. F. Hayes, J. Math. Phys. **10**, 1555 (1969).
[3] P. A. M. Dirac, Can. J. Math. **2**, 129 (1950); **3**, 1 (1951).
[4] E. T. Whittaker, *Analytical Dynamics* (Cambridge U. P., Cambridge, 1937), p. 265.
[5] This form of the Hamiltonian is unique. The ambiguities demonstrated by C. F. Hayes and J. M. Jankowski [Nuovo Cimento **58**, 494 (1968)] arise because they make transformations on the Hamiltonian which are not canonical.

Noncommutative Probability on von Neumann Algebras

S. Gudder and J.-P. Marchand*
Department of Mathematics, University of Denver, Denver, Colorado 80210
(Received 11 November 1971; Revised Manuscript Received 21 January 1972)

We generalize ordinary probability theory to those von Neumann algebras $\mathbb{A}$, for which Dye's generalized version of the Radon–Nikodym theorem holds. This includes the classical case in which $\mathbb{A}$ is an Abelian von Neumann algebra generated by an observable or complete set of commuting observables. Via Gleason's theorem, this also includes the case of ordinary quantum mechanics, in which $\mathbb{A} = \mathbb{B}(H)$ is the von Neumann algebra of all bounded operators on a separable Hilbert space H. Particular consideration is given to the concepts of conditioning, sufficient statistics, coarse-graining, and filtering.

1. INTRODUCTION

In two fundamental papers Dye[1] and Segal[2] have introduced models of noncommutative probability and integration theories. In both models a pair $(\mathbb{A}, w)$ consisting of a measure algebra $\mathbb{A}$ and a measure w on $\mathbb{A}$ have been specified. But while Segal places restrictions on the "gage" w, requiring it to be unitarily invariant, Dye requires the algebra $\mathbb{A}$ to be a σ-finite and finite von Neumann algebra. Umegaki[3] and his collaborators have carried these theories to a considerable extent. However, they continue to impose invariance and finiteness restrictions.

Our purpose here is to investigate the non-Boolean probability theory of ordinary quantum mechanics in a separable Hilbert space H. $\mathbb{A}$ is therefore (as least in the absence of superselection rules) the von Neumann algebra $\mathbb{B}(H)$ of all bounded linear operators on H, and the measures are the countably additive positive linear functionals w on $\mathbb{A}$. In this context, neither Dye's or Segal's frameworks appear to be suitable. In Dye's case $\mathbb{B}(H)$ is not finite (unless H is finite dimensional). In Segal's case there are no nontrivial bounded gages (again, unless H is finite dimensional) and the only unbounded gages on $\mathbb{B}(H)$ are proportional to the trace functional τ, all one-dimensional projections in $\mathbb{B}(H)$ being unitarily equivalent. In this latter case τ cannot be normalized to obtain a probability measure (state) so that a probability theory is out of the question. On the other hand, although $\mathbb{B}(H)$ is not finite if H is infinite dimensional, Dye's main result, namely the generalized Radon–Nikodym theorem, remains true for $\mathbb{B}(H)$. As we shall show, this follows using a result of Gleason.[4]

We shall derive our results as far as possible for an arbitrary von Neumann algebra $\mathbb{A}$ and arbitrary measures w, while including in each section their applications to the "extreme" cases, $\mathbb{A}$ Abelian (classical case) and $\mathbb{A} = \mathbb{B}(H)$ (pure quantum case). These results briefly are as follows.

In Sec. 2 we combine the theorems of Dye and Gleason which leads to a relation between the density operators of quantum mechanics and the Radon–Nikodym derivative in Dye's sense. Some of the results of this section are not new[5]; their derivation is, however, simple enough to justify a restatement of the proofs.

In Sec. 3 we generalize the notion of conditioning with respect to a von Neumann subalgebra $\mathbb{B}$. For reasons of physical relevance and mathematical simplicity, particular attention is given to the case of an Abelian $\mathbb{B}$. This then leads to severe restrictions on the structure of $\mathbb{B}$ and w. It is shown that conditioning with respect to Abelian $\mathbb{B}$ exists if and only if $\mathbb{B}$ contains a discrete part and the density operator W of w commutes with $\mathbb{B}$. In this case the conditional expectation of w given $\mathbb{B}$ appears as a Radon–Nikodym derivative very much like in classical probability.

In Sec. 4 we consider the concepts of equivalence of states relative to a subalgebra $\mathbb{B}$ and of sufficient statistics. Physically, $\mathbb{B}$ may be interpreted as a partial measurement (for instance, measurement of the macroscopic observables in statistical mechanics). $\mathbb{B}$-equivalence then leads to the concept of macrocells, and $\mathbb{B}$ is a sufficient statistic for those states which have equal $\mathbb{B}$-expectations.[6] We show that $\mathbb{B}$-equivalence and $\mathbb{B}$-sufficiency are comple-

mentary concepts. The notion of $\mathbb{B}$-equivalence has been used for some time now for the purpose of (i) measurement theory in quantum mechanics, (ii) nonequilibrium statistical mechanics, (iii) ergodic theory, and (iv) equilibrium statistical mechanics and phase transitions. It has been explicitly pointed out previously[7] that this notion plays a central role in the formulation of coarse-graining which we generalize in Sec. 5.

In Sec. 5 the physical notion of coarse-graining[7] is generalized to suitable von Neumann algebras. Coarse-graining of a state w on $\mathbb{A}$ with respect to a subalgebra $\mathbb{B}$ roughly means choosing in the $\mathbb{B}$-equality class of w the state with maximal entropy. Physically this is related to the notion of minimal phenomenology by answering the question: What is the most probable state of the system if only a partial measurement $\mathbb{B}$ has been carried out? It turns out that the concept of coarse-graining is meaningful only if a dominating state (in physics, for instance an equilibrium state) is specified beforehand, relative to which w admits a density. In the Abelian case, coarse-graining can be expressed analytically as a Radon–Nikodym derivative and geometrically as a projection onto $\mathbb{B}$. Since in quantum mechanics the trace τ dominates all states, it is natural to coarse-grain with respect to τ. This is, however, possible only in the restricted case where $\mathbb{B}$ contains projections with finite-dimensional range, as already pointed out by Emch.[7]

In Sec. 6 we consider the familiar notion of filtering.[8] The difference between filtering and coarse-graining is as follows: While coarse-graining asks for the most probable state of a system whose *a priori* probabilities relative to $\mathbb{B}$ are known, filtering asks the question: Given a system in state w, what state is it in after a measurement of $\mathbb{B}$? In other words, while both filtering and coarse-graining transform w into $\mathbb{B}$-equal states, filtering conserves the information of w within the subspaces invariant under $\mathbb{B}$ maximally while coarse-graining destroys it. Geometrically, filtering corresponds to projecting onto the commutant $\mathbb{B}'$ of $\mathbb{B}$ ($\mathbb{B} \subseteq \mathbb{B}'$).

A last general remark. In the noncommutative case the Radon–Nikodym theorem of Dye is much more complicated than in the classical theory. In fact let w, v be measures on $\mathbb{A}$ with w absolutely continuous with respect to v. Then by Dye's theorem we have $w(A) = v(T^*AT)$ for every $A \in \mathbb{A}$ for some square-integrable operator T. This reduces to the usual form $w(A) = v((dw/dv)A)$ only if $v(T^*AT) = v(TT^*A)$, where $dw/dv = TT^*$, which in turn holds only if either $\mathbb{A}$ is Abelian or v is cyclic. This problem lurks in the background of all our subsequent work, and it is therefore helpful to call it to the reader's attention from the beginning.

2. DYE'S THEOREM

Let H be a Hilbert space, $\mathbb{A}$ a von Neumann algebra of operators on H, $P_{\mathbb{A}}$ the set of all (self-adjoint) projections in $\mathbb{A}$. $\mathbb{A}$ is *σ-finite* if any collection of mutually orthogonal projections is at most countable. $\mathbb{A}$ is *finite* if there exist no partial isometries Ω in $\mathbb{A}$ such that $\Omega P \Omega^* = I$ for $P \neq I$. If H is separable, then the von Neumann algebra $\mathbb{A} = \mathbb{B}(H)$ of all bounded operators in H is σ-finite but not finite if dim $H = \infty$. If $\mathbb{A}$ is Abelian and H separable or if $\mathbb{A}$ is of types $\mathrm{I}_n, \mathrm{II}_1$, then it is σ-finite and finite.

A measure w on $P_{\mathbb{A}}$ is a nonnegative mapping $w: P_{\mathbb{A}} \to \mathrm{R}^+$ such that (1) $w(0) = 0$, (2) $w(\sum A_i) = \sum w(A_i)$ for any countable set of mutually orthogonal projections in $P_{\mathbb{A}}$. An *integral* w on $\mathbb{A}$ is a positive linear functional $w: \mathbb{A} \to C$ satisfying (1) and (2). An integral w on $\mathbb{A}$ is a *state* on $\mathbb{A}$ if $w(I) = 1$. Let w, v be measures or integrals. Then w is *absolutely continuous* with respect to v ($w \prec v$) if $v(P) = 0$ implies $w(P) = 0$, $P \in P_{\mathbb{A}}$. If $w \prec v$ and $v \prec w$, then w and v are *equivalent* ($w \sim v$). Note that $\sim$ is indeed an equivalence relation.

Let w be an integral on $\mathbb{A}$. $L_2(\mathbb{A}, w)$ is the Hilbert space of operators generated by the operators in $\mathbb{A}$ with respect to the inner product $(A_1, A_2) = w(A_1{}^*A_2)$. $L_1(\mathbb{A}, w)$ is the Banach space of operators generated by the operators in $\mathbb{A}$ with respect to the L_1-norm $\|A\|_1 = w(|A|) = w((A^*A)^{1/2}) = \||A|^{1/2}\|_2^2$. A is called *$\mathbb{A}$-integrable, $\mathbb{A}$-square-integrable* if $A \in L_1(\mathbb{A}, w)$, $A \in L_2(\mathbb{A}, w)$, respectively.

Theorem 2.1: (Dye) Let $\mathbb{A}$ be a σ-finite, finite von Neumann algebra and $w \prec v$ integrals on $\mathbb{A}$. Then there exists an operator $T \in L_2(\mathbb{A}, v)$ such that $w(A) = v(T^*AT)$ for all $A \in \mathbb{A}$.

TT^* is called the (generalized) *Radon–Nikodym derivative* of w with respect to v ($TT^* \equiv dw/dv$). Notice $TT^* \in L_1(\mathbb{A}, v)$. The theorem of Dye admits various interpretations for various choices of the algebra $\mathbb{A}$. We consider, in particular, the two extreme cases: (1) $\mathbb{A}$ Abelian (classical case), (2) $\mathbb{A} = \mathbb{B}(H)$(pure quantum case).

(1) The Classical Case

Let (Ω, Σ, μ) be a finite measure space, $\mathbb{A}$ the algebra of bounded Σ-measurable functions A on Ω. $\mathbb{A}$ can be thought of as a von Neumann algebra in the Hilbert space $L_2(\Omega, \Sigma, \mu)$ if we define $A_f g(\omega) = f(\omega)g(\omega)$, $f \in \mathbb{A}$, $g \in L_2(\Omega, \Sigma, \mu)$. Then the expectation $w(A) = \int A d\mu$ is an integral on $\mathbb{A}$. $(\mathbb{A}, w)$ is called a *measure algebra*. Since $\mathbb{A}$ is Abelian, Dye's theorem applies: If $w \prec v$, there exists a function $T \in L_2(\mathbb{A}, v)$ such that $w(A) = v(T^*AT) = v(T^*TA) = v((dw/dv)A)$, for all $A \in \mathbb{A}$, where T^* is now the complex conjugate of T and $dw/dv = TT^* = |T|^2 \in L_1(\mathbb{A}, v)$ is the ordinary Radon–Nikodym derivative. Conversely, it is well known that if $\mathbb{A}$ is an Abelian von Neumann algebra and w a measure on $\mathbb{A}$, then $(\mathbb{A}, w)$ is isomorphic to a measure algebra $(\hat{\mathbb{A}}, \hat{w})$ on a measure space $(\Omega, \Sigma, \hat{w})$ so Dye's theorem reduces to the Radon–Nikodym theorem.

(2) The Pure Quantum Case

Let H be a separable Hilbert space and $\mathbb{A} = \mathbb{B}(H)$. If dim $H = \infty$, $\mathbb{A}$ is σ-finite but not finite so that Dye's theorem does not directly apply. But in this special case we have the following theorem.

Theorem 2.2: (Gleason) If dim $H > 2$ and w is a measure on $P_{\mathbb{A}}$, then there is a unique positive trace class operator W such that $w(A) = \tau(WA)$ for all $A \in P_{\mathbb{A}}$.

Note that linearity of w is not an assumption but

rather is a consequence of this theorem. Gleason's theorem can be extended to obtain the following.

Theorem 2.3: If $\dim H > 2$ and w is an integral on $\mathbb{A}$, then there exists a unique positive trace class operator W such that $w(A) = \tau(WA)$ for all $A \in \mathbb{A}$. (We call W the *density operator* of w.)

Proof: Since w restricted to $P_{\mathbb{A}}$ is a measure, Gleason's theorem being applied, there is a positive trace class operator W such that $w(A) = \tau(WA)$ for all $A \in P_{\mathbb{A}}$. If $A = \int \lambda P^A(d\lambda)$ is self-adjoint, we have $w(A) = \int \lambda w[P^A(d\lambda)] = \int \lambda \tau[WP^A(d\lambda)] = \tau[W \int \lambda P^A(d\lambda)] = \tau(WA)$. If $A \in \mathbb{A}$ is arbitrary, $w(A) = \frac{1}{2}w(A + A^*) + (1/2i)w[i(A - A^*)] = \tau(WA)$. It is clear that W is unique.

We can now prove Dye's theorem for type I_∞ algebras.

Theorem 2.4: Let $\mathbb{A} = \mathbb{B}(H)$ be a type I_∞ algebra and $w \prec v$ integrals on $\mathbb{A}$. Then there exists a unique operator $dw/dv \in L_1(\mathbb{A}, v)$ such that $w(A) = v((dw/dv)A)$ for all $A \in \mathbb{A}$.

Proof: By Theorem 2.3 there exist unique operators W, V such that $w(A) = \tau(WA)$, $v(A) = \tau(VA)$ for all $A \in \mathbb{A}$. Let E_V be the support of V (i.e., the orthogonal complement of the projection onto the null space of V). Since $w \prec v$, it follows that $E_V W = WE_V = W$. Since V is one-to-one on $E_V H$, the operator $V^{-1}E_V$ is well defined. Define $dw/dv = V^{-1}E_V W$. Then $w(A) = \tau(WA) = \tau(V(dw/dv)A) = v((dw/dv)A)$ for all $A \in \mathbb{A}$.

Finally we relate Gleason's theorem to Dye's theorem.

Theorem 2.5: If $2 < \dim H < \infty$, then $W = dw/d\tau$.

Proof: The trace τ is an integral on $\mathbb{A}$ which majorizes all other integrals ($w \prec \tau$ for all integrals w). Therefore, by the theorems of Gleason and Dye, $w(A) = \tau(WA) = \tau(T^*AT) = \tau(TT^*A) = \tau((dw/d\tau)A)$ for all $A \in \mathbb{A}$, and, since W is unique, $W = dw/dv = TT^*$.

3. CONDITIONAL EXPECTATION

We now want to generalize the classical notion of conditioning to our noncommutative case. Classically, if $(\Omega, \mathbb{A}, \mu)$ is a probability space and $\mathbb{B}$ a sub σ-algebra of $\mathbb{A}$, then for $A \in \mathbb{A}$ the conditional probability of A given $\mathbb{B}$ is a function $p_\mu(A \mid \mathbb{B})$ measurable relative to $\mathbb{B}$ and satisfying $\int_B p_\mu(A \mid \mathbb{B})d\mu = \mu(A \cap B) \equiv \mu_A(B)$ for all $B \in \mathbb{B}$. Let us first find the quantum analog of $\mu_A(B)$. Now by definition $\mu_A(B) = \mu(A \mid B) \times \mu(B)$, where $\mu(A \mid B)$ is the conditional probability of A given B, so that it suffices to find the quantum analog of $\mu(A \mid B)$.

Let w be a state on $\mathbb{A} = \mathbb{B}(H)$. According to the usual quantum formalism, after the *a priori* measurement of $B \in P_{\mathbb{A}}$ the initial state W becomes $BWB/\tau(WB)$, and hence the conditional probability of $A \in P_{\mathbb{A}}$ in this state is

$$w(A/B) = \tau(BWBA)/\tau(WB) = \tau(WBAB)/\tau(WB) = w(BAB)/w(B).$$

Therefore, $w_A(B) \equiv w(BAB)$ should be considered as the quantum analog of μ_A.

Definition 3.1: Let $\mathbb{A}$ be an arbitrary von Neumann algebra on H and $\mathbb{B} \subseteq \mathbb{A}$ a von Neumann subalgebra of $\mathbb{A}$. Let w be an integral on $\mathbf{A}$. Then the $\mathbb{B}$-*expectation of* $A \in \mathbb{A}$ *with respect to* w is an operator $E_w(A \mid \mathbb{B}) \in \mathbb{B}$ satisfying $w[BE_w(A \mid \mathbb{B})B] = w_A(B) \equiv w(BAB)$ for all $B \in P_{\mathbb{B}}$.

As we shall see the $\mathbb{B}$-expectation need not, in general, exist. Thus $A \to E_w(A \mid \mathbb{B})$ must be thought of as a map whose domain is a subset of $\mathbb{A}$. Also $E_w(A \mid \mathbb{B})$ is unique only to a projection in $\mathbb{B}$ of w measure zero. One can ensure uniqueness by requiring $E_w(A \mid \mathbb{B})$ to have the same support as w on $\mathbb{B}$.

Theorem 3.1: The $\mathbb{B}$-expectation has the following properties:

(1) It maps a subset of $\mathbb{A}$ onto $\mathbb{B}$.

(2) Linearity: $E_w(\alpha_1 A_1 + \alpha_2 A_2 \mid \mathbb{B}) = \alpha_1 E_w(A_1 \mid \mathbb{B}) + \alpha_2 E_w(A_2 \mid \mathbb{B})$.

(3) Involution: $E_w(A^* \mid \mathbb{B}) = E_w(A \mid \mathbb{B})^*$.

(4) Conservation of identity: $E_w(I \mid \mathbb{B}) = I$.

(5) Idempotence: $E_w(E_w(A \mid \mathbb{B}) \mid \mathbb{B}) = E_w(A \mid \mathbb{B})$.

Proof: (1) Let $B' \in \mathbb{B}$. Then the defining property $w[BE_w(B' \mid \mathbb{B})B] = w(BB'B)$ for all $B \in P_{\mathbb{B}}$ implies $E_w(B' \mid \mathbb{B}) = B'$. Hence E_w is onto. (2) $w[BE_w(\alpha_1 A_1 + \alpha_2 A_2) \mid \mathbb{B})B] = w[B(\alpha_1 A_1 + \alpha_2 A_2)B] = \alpha_1 w(BA_1B) + \alpha_2 w(BA_2B) = \alpha_1 w[BE_w(A_1 \mid \mathbb{B})B] + \alpha_2 w[BE_w(A_2 \mid \mathbb{B})B] = w[B(\alpha_1 E_w(A_1 \mid \mathbb{B}) + \alpha_2 E_w(A_2 \mid \mathbb{B}))B]$ for all $B \in \mathbb{B}$. (3) $w[BE_w(A^* \mid \mathbb{B})B] = w(BA^*B) = w[(BAB)^*] = [w(BAB)]^* = (w[BE_w(A \mid \mathbb{B})B])^* = w[(BE_w(A \mid \mathbb{B})B)^*] = w[BE_w(A \mid \mathbb{B})^*B]$ for all $B \in P_{\mathbb{B}}$. (4) and (5) follow from the proof of (1).

(1) Classical Case

Let (Ω, Σ, μ) be a measure space, $(\mathbb{A}, w)$ the corresponding measure algebra, and $\mathbb{B} \subseteq \mathbb{A}$ a subalgebra. The projections $P_{\mathbb{B}}$ are the characteristic functions χ_B on Ω with B a set in the sub σ-algebra of Σ corresponding to the subalgebra $\mathbb{B}$ of $\mathbb{A}$. The above definition now reads for $A = \chi_A$, $w(\chi_B E_w(\chi_A \mid \mathbb{B})\chi_B) = w\chi_A(\chi_B) = w(\chi_B \chi_A \chi_B)$ for all $\chi_B \in P_{\mathbb{B}}$, or, in terms of Σ and μ, $\int_B p_\mu(A \mid \mathbb{B})d\mu = \mu(A \cap B)$. But this is the defining property of the ordinary conditional expectation of A given $\mathbb{B}$.

(2) Pure Quantum Case

Let H be a separable Hilbert space, $\mathbb{A} = \mathbb{B}(H)$ and $\mathbb{B} \subseteq \mathbb{A}$ a von Neumann subalgebra. In the theorem that follows we first prove some general results concerning the restriction $w_A^{\mathbb{B}}$ of w_A to $\mathbb{B}$.

Theorem 3.2: (1) $w_A^{\mathbb{B}}$ is a measure for all $A \in \mathbb{A}$ if and only if $W \in \mathbb{B}'$. (2) $w_A^{\mathbb{B}}$ is a measure for all states w on $\mathbb{A}$ if and only if $A \in \mathbb{B}'$. (3) If $w_A^{\mathbb{B}}$ is a measure, then $w_A^{\mathbb{B}} \prec w^{\mathbb{B}}$. (4) $w_{A^*}^{\mathbb{B}}(B) = w_A^{\mathbb{B}}(B)^*$.

Proof: (1) Suppose $W \in \mathbb{B}'$ and B_i is a sequence of mutually orthogonal projections in $\mathbb{B}$. Then $w_A(\sum B_i) = w(\sum B_i A \sum B_j) = \sum_{i,j} w(B_i A B_j)$

$$= \sum_{i,j} \tau(WB_i A B_j) = \sum_{i,j} \tau(B_j WB_i A) = \sum_{i,j} \tau(B_j B_i WA)$$

$$= \sum \tau(B_i WA) = \sum \tau(WB_i AB_i) = \sum w(B_i A B_i)$$

$$= \sum w_A(B_i).$$

Conversely, suppose w_A is additive on $\mathbb{B}$ for every $A \in \mathbb{A}$ and $B \in P_{\mathbb{B}}$. Then $w_A(B + B^\perp) = w_A(I) = w(A) = w_A(B) + w_A(B^\perp) = w(BAB) + w(B^\perp AB^\perp) = 2w(BAB) - w(BA) - w(AB) + w(A)$. Hence $2w(BAB) = w(BA) + w(AB)$ or $2\tau(BWBA) = \tau((WB + BW)A)$ for every $A \in \mathbb{A}$. Let $A = P_\phi$, the one-dimensional projection on the vector ϕ, to obtain $2\langle BWB\phi, \phi\rangle = \langle (WB + BW)\phi, \phi\rangle$ for every $\phi \in H$. Since these operators are self-adjoint, $2BWB = WB + BW$. Multiplying this last equation by B on the left and right, we obtain $BW = WB = BWB$. The proof of (2) is similar. (3) If $w(B) = 0$, we have $\tau(BWB) = \tau(WB) = 0$. Since BWB is positive, we conclude that $BWB = 0$. Hence $w_A(B) = w(BAB) = \tau(BWBA) = 0$. (4) Finally $w_{A*}(B) = \tau(WBA^*B) = \tau[(BABW)^*] = [\tau(BABW)]^* = [w(BAB)]^* = w_A(B)^*$.

We now consider the case in which $\mathbb{B}$ is Abelian. The study of conditioning with respect to non-Abelian subalgebras is more difficult and its operational meaning in physics less transparent. In the Abelian case, the existence of the $\mathbb{B}$-expectation for all $A \in \mathbb{A}$ imposes rather heavy limitations on the algebra $\mathbb{B}$ and the states.

Theorem 3.3: Let $\mathbb{B}$ be an Abelian subalgebra of $\mathbb{A}$. The following statements are equivalent:

(1) $E_w(A \mid \mathbb{B})$ exists for all $A \in \mathbb{A}$,

(2) $w_A^{\mathbb{B}}$ is a measure on $P_{\mathbb{B}}$ for all $A \in \mathbb{A}$,

(3) $W \in \mathbb{B}'$.

If $E_w(A \mid \mathbb{B})$ exists, then it is unique and is given by $E_w(A \mid \mathbb{B}) = \int (dw_A'/dw')P(d\lambda)$, where $\mathbb{B} = \{P(\Delta): \Delta \in B(R)\}''$, $w'(\Delta) = w(P(\Delta))$, $w_A'(\Delta) = w(P(\Delta)AP(\Delta))$.

Proof: It is clear that $w_A^{\mathbb{B}}$ is a measure if $E_w(A \mid \mathbb{B})$ exists. Conversely, suppose $w_A^{\mathbb{B}}$ is a measure. Then, via Theorem 3.2 (3), $w_A' \prec w'$ so that dw_A'/dw' exists. Now for $B = P(\Delta_0) \in \mathbb{B}$, $w[\int (dw_A'/dw')P(d\lambda)B] = \int (dw_A'/dw')w[P(d\lambda \cap \Delta_0)] = \int_{\Delta_0} (dw_A'/dw')dw' = w_A'(\Delta_0) = w(BAB) = w_A(B)$. Hence $\int (dw_A'/dw')P(d\lambda)$ is a $\mathbb{B}$-expectation for A. Thus (1) and (2) are equivalent. The equivalence of (2) and (3) follows from Theorem 3.2 (1). To show uniqueness, suppose $E_w(A \mid \mathbb{B})$ is a $\mathbb{B}$-expectation for A. Then, since $E_w(A \mid \mathbb{B}) \in \mathbb{B}$, we have $E_w(A \mid \mathbb{B}) = \int f(\lambda)P(d\lambda)$ for some Borel measurable function f. If $\Delta_0 \in B(R)$, we obtain

$$\begin{aligned}\int_{\Delta_0} f(\lambda)dw' &= \int_{\Delta_0} f(\lambda)w(P(d\lambda)) = \int f(\lambda)w[P(d\lambda \cap \Delta_0)\\ &= \int f(\lambda)w[P(d\lambda)P(\Delta_0)] = w[\int f(\lambda)P(d\lambda)P(\Delta_0)]\\ &= w[E_w(A \mid \mathbb{B})P(\Delta_0)] = w_A[P(\Delta_0)]\\ &= w[P(\Delta_0)AP(\Delta_0)].\end{aligned}$$

Also $\int_{\Delta_0} (dw_A'/dw')dw' = w_A'(\Delta_0) = w[P(\Delta_0)AP(\Delta_0)]$. Hence $f = dw_A'/dw'$ a.e. $[w']$.

Next we express the conditional expectation as a Radon–Nikodym derivative, in analogy to the classical case.

Theorem 3.4: If $\mathbb{B}$ is Abelian, $E_w(A \mid \mathbb{B})$ exists for a positive A, and $W \in \mathbb{B}'$, then $E_w(A \mid \mathbb{B}) = dw_A^{\mathbb{B}}/dw^{\mathbb{B}}$ in the sense of Dye.

Proof: Suppose $E_w(A \mid \mathbb{B})$ exists and A is positive. Thus $w_A^{\mathbb{B}}$ is a measure. We now show that $w_A^{\mathbb{B}}$ has a unique extension to an integral on $\mathbb{B}$. If $T = \int f(\lambda)P(d\lambda) \in \mathbb{B}$ is self-adjoint, define $w_A^{\mathbb{B}}(T) = \int f(\lambda)w_A^{\mathbb{B}}[P(d\lambda)]$ and, if $T \in \mathbb{B}$, define $w_A^{\mathbb{B}}(T) = \frac{1}{2}w_A(T + T^*) + (1/2i)w_A^{\mathbb{B}}[i(T - T^*)]$. Since A is positive, it follows that $w_A^{\mathbb{B}}$ is a positive measure so that $w_A^{\mathbb{B}}$ becomes a positive functional on $\mathbb{B}$. To show $w_A^{\mathbb{B}}$ is linear on $\mathbb{B}$, it suffices to show that $w_A^{\mathbb{B}}$ is additive on self-adjoint operators in $\mathbb{B}$. Suppose $T_1 = \int f_1(\lambda)P(d\lambda)$, $T_2 = \int f_2(\lambda)P(d\lambda) \in \mathbb{B}$ are self-adjoint. Then $w_A^{\mathbb{B}}(T_1 + T_2) = w_A^{\mathbb{B}}[\int (f_1 + f_2)(\lambda)P(d\lambda)] = \int (f_1 + f_2)(\lambda)w_A^{\mathbb{B}}[P(d\lambda)] = \int f_1(\lambda)w_A^{\mathbb{B}}[P(d\lambda)] + \int f_2(\lambda)w_A^{\mathbb{B}}[P(d\lambda)] = w_A^{\mathbb{B}}(T_1) + w_A^{\mathbb{B}}(T_2)$.

It is easily seen that this positive linear extension is unique. Since $\mathbb{B}$ is a finite von Neumann algebra and $w_A^{\mathbb{B}} < w^{\mathbb{B}}$, the Dye derivative $dw_A^{\mathbb{B}}/dw^{\mathbb{B}}$ exists. By Dye's theorem there is a $T \in L_2(\mathbb{B}, w)$ such that for every $B \in \mathbb{B}$, $w_A^{\mathbb{B}}(B) = w^{\mathbb{B}}(T^*BT) = w^{\mathbb{B}}(T^*TB) = w^{\mathbb{B}}((dw_A^{\mathbb{B}}/dw^{\mathbb{B}})B)$. On the other hand, $w_A^{\mathbb{B}}(B) = w^{\mathbb{B}}(E_w(A \mid \mathbb{B})B)$ for every $B \in \mathbb{B}$ whence the conclusion.

The following lemma shows the type of restriction imposed on $\mathbb{B}$ by the requirement that $W \in \mathbb{B}'$.

Lemma 3.1: If $\mathbb{B} \subseteq \mathbb{B}'$, $W \in \mathbb{B}'$, then $\mathbb{B}$ contains a discrete part $P_d\mathbb{B}$ and the support E_W of W is contained in P_d.

Proof: Since $\mathbb{B}$ is Abelian, there is a bounded self-adjoint operator $A = \int \lambda P^A(d\lambda)$ such that $\mathbb{B} = \{P^A(\Delta): \Delta \in B(R)\}''$. Let P_d and P_c be the discrete and continuous parts of H relative to A. Now P_dH is spanned by the eigenvectors of A. On the other hand, let $W = \sum w_iP_i$, $w_i > 0$, be the spectral form of W. If E_W is the support of W, we have $E_W = \sum P_i$, and, since W is compact, the projections P_i have finite-dimensional range. Now $W \in \mathbb{B}'$ implies $P_i \in \mathbb{B}'$. Therefore, the P_i reduce A, and we have $A = \sum P_iAP_i + P_0AP_0$, where P_0 is the projection onto the null space of W. Noticing that all the operators P_iAP_i are self-adjoint, we see that their eigenvectors ϕ_j^i span all the subspaces P_iH. On the other hand, the ϕ_j^i are eigenvectors of A and hence $\phi_j^i \in P_dH$. So we conclude $E_WH = \sum P_iH \subseteq P_dH$.

We therefore obtain the following corollary to Theorem 3.4.

Corollary 3.1: Suppose $\mathbb{B}$ is Abelian. (1) $E_w(A \mid \mathbb{B})$ exists for all $A \in \mathbb{A}$ if and only if $\mathbb{B}$ has a discrete part $\{B_i\}''$ and $W \in \mathbb{B}'$. (2) If $E_w(A \mid \mathbb{B})$ exists for all $A \in \mathbb{A}$, then it is unique and

$$E_w(A \mid \mathbb{B}) = \int (dw_A'/dw')P(d\lambda) = \sum_{w(B_i)\neq 0}[w_A(B_i)/w(B_i)]B_i.$$

We mention here, for completeness, Umegaki's characterization theorem for $\mathbb{B}$-expectations in his framework (which as we mentioned in the introduction is different from ours) and then verify his conditions in our case when $\mathbb{B}$ is Abelian.

Theorem 3.5: (Umegaki[3]) $E_w(A \mid \mathbb{B})$ is a $\mathbb{B}$-expectation if and only if it is (1) a linear map from $\mathbb{A}$ onto $\mathbb{B}$, (2) involutary, (3) conserves the identity, (4) satisfies $E_w[AE_w(A' \mid \mathbb{B}) \mid \mathbb{B}] = E_w[E_w(A \mid \mathbb{B})A' \mid \mathbb{B}] = E_w(A \mid \mathbb{B})E_w(A' \mid \mathbb{B})$, (5) contractive: $w(\mid E(A \mid \mathbb{B})\mid) \leq w(\mid A \mid)$, $\mid A \mid = (A^*A)^{1/2}$.

Theorem 3.6: If $\mathbb{B} \subseteq \mathbb{B}'$ and $E_w(A \mid \mathbb{B})$ exists for all $A \in \mathbb{A}$, then it satisfies Umegaki's conditions (1)–(5).

Proof: Conditions (1)–(3) have been demonstrated in Theorem 3.1. (4) Applying Corollary 3.1 (2), we have

$$\begin{aligned}\mathrm{E}_w(AE_w(A'\mid\mathbb{B})\mid\mathbb{B}) &= \sum_i \frac{w_{AE_w(A'\mid\mathbb{B})}(B_i)}{w(B_i)}B_i \\ &= \sum_i \frac{w(B_iAE_w(A'\mid\mathbb{B})B_i)}{w(B_i)}B_i \\ &= \sum_i \frac{w\{B_iA\sum_j[w_{A'}(B_j)/w(B_j)]B_j\}}{w(B_i)}B_i \\ &= \sum_{i,j}\frac{w_{A'}(B_j)}{w(B_i)}\frac{w(B_iAB_j)}{w(B_i)}B_i = \sum_i \frac{w_{A'}(B_i)}{w(B_i)}\frac{w_A(B_i)}{w(B_i)}B_i.\end{aligned}$$

The other two expressions lead to the same result. (5) Applying Theorem 3.1 (3) and Corollary 3.1 (2), we have

$$\begin{aligned}w(|E_w(A\mid\mathbb{B})|) &= w\{[E_w(A^*\mid\mathbb{B})E_w(A\mid\mathbb{B})]^{1/2}\} \\ &= w\left[\left(\sum\frac{w_{A^*}(B_i)}{w(B_i)}\frac{w_A(B_i)}{w(B_i)}B_i\right)^{1/2}\right] \\ &= w\left[\left(\sum\left|\frac{w_A(B_i)}{w(B_i)}\right|^2 B_i\right)^{1/2}\right] = w\left[\sum\left|\frac{w_A(B_i)}{w(B_i)}\right|B_i\right] \\ &= \sum\left|\frac{w_A(B_i)}{w(B_i)}\right|w(B_i) = \sum|w(AB_i)|.\end{aligned}$$

By the polar decomposition we obtain $A = T|A|$, where T is a partial isometry. Hence

$$\sum|w(AB_i)| = \sum|w(T|A|B_i)| \le \sum w(|A|B_i) = w(|A|).$$

4. $\mathbb{B}$-EQUALITY AND $\mathbb{B}$-SUFFICIENCY

We now generalize the classical concepts of *a priori* probabilities on a subalgebra and sufficient statistics. Let $\mathbb{A}$ and $\mathbb{B} \subseteq \mathbb{A}$ be von Neumann algebras. Two states w, v on $\mathbb{A}$ are $\mathbb{B}$-*equal* (or $\mathbb{B}$-*equivalent*) if their restrictions $w^{\mathbb{B}}, v$ to $\mathbb{B}$ coincide. We then use the notation $w \overset{\mathbb{B}}{=} v$. Let $\{w_i: i \in I\}$ be a (possibly uncountable) collection of states on $\mathbb{A}$. We say that $\mathbb{B}$ is *sufficient* for $\{w_i: i \in I\}$ if for any $A \in \mathbb{A}$ there is an $E_A \in \mathbb{B}$ such that whenever $E_{w_i}(A\mid\mathbb{B})$ exists we have $E_{w_i}(A\mid\mathbb{B}) = E_A$ a.e. $[w_i]$ for all $i \in I$. Two states w, v on $\mathbb{A}$ are $\mathbb{B}$-*sufficient* (denoted $w \overset{\mathbb{B}}{\sim} v$) if $w \sim v$ and $\mathbb{B}$ is sufficient for $\{w, v\}$. Notice $w \overset{\mathbb{B}}{\sim} v$ if and only if $w \sim v$ and $E_w(A\mid\mathbb{B}) = E_v(A\mid\mathbb{B})$ for all $A \in \mathbb{A}$ when these expressions exist. The relations $\overset{\mathbb{B}}{=}$ and $\overset{\mathbb{B}}{\sim}$ are equivalence relations. We call the equivalence classes $\mathbb{B}$-*cells* and $\mathbb{B}$-*suffs*, respectively, and use the notation $S_w^{\mathbb{B}} = \{v: v \overset{\mathbb{B}}{=} w\}$, $\mathbf{S}_w^{\mathbb{B}} = \{v: v \overset{\mathbb{B}}{\sim} w\}$.

The $\mathbb{B}$-cells are uniquely determined by a measurement of $\mathbb{B}$ alone; on the other hand, two states of a $\mathbb{B}$-cell cannot be distinguished by the measurement of $\mathbb{B}$ alone. The $\mathbb{B}$-cells thus represent exactly the information on the states contained in a measurement of $\mathbb{B}$. In quantum statistics this concept is basic. Let, for instance, $\mathbb{B}$ represent the (commutative) algebra of macroscopic observables. The $\mathbb{B}$-cells then appear as the so-called macrocells, containing the macroscopically equivalent states. The $\mathbb{B}$-suffs, on the other hand, yield no information whatsoever on the $\mathbb{B}$-expectation of w except for the support of $w^{\mathbb{B}}$. For example, $E_w(B\mid\mathbb{B}) = B$ is independent of w and hence, for $\mathbb{A} = \mathbb{B}$, all states (with equal supports) are mutually $\mathbb{B}$-sufficient. The concept of sufficient statistics has so far mainly been used in the classical theory; see, however, Umegaki.[6]

The two concepts are, in a certain sense, complementary to each other. This is indicated by the following extreme cases. Let $\mathbb{B} = \{0, I\}$ be the trivial subalgebra. Then $S_w^{\mathbb{B}}$ contains all states on A, while $\mathbf{S}_w^{\mathbb{B}}$ contains only w, since $E_w(A\mid\mathbb{B}) = w(A)I$. Now let $\mathbb{B} = \mathbb{A}$. Then $S_w^{\mathbb{A}}$ contains only the state w, while $\mathbf{S}_w^{\mathbb{A}}$ contains all states equivalent to w, since $E_w(A\mid\mathbb{A}) = A$ independently of w.

Theorem 4.1: (1) If $w \in S_{w_1}^{\mathbb{B}} \cap \mathbf{S}_{w_2}^{\mathbb{B}}$, then w is unique and $w(A) = w_1[E_{w_2}(A\mid\mathbb{B})]$ for all $A \in \mathbb{A}$. (2) $\{w\} = S_w^{\mathbb{B}} \cap \mathbf{S}_w^{\mathbb{B}}$. (3) If $\mathbb{B}$ is Abelian and $\mathbb{A} = \mathbb{B}(H)$, then $\{w\} = S_{w_1}^{\mathbb{B}} \cap \mathbf{S}_{w_2}^{\mathbb{B}}$, where w is given in (1).

Proof: (1) $w(A) = w_A(I) = w[E_w(A\mid\mathbb{B})] = w_1[E_{w_2}(A\mid\mathbb{B})]$. (2) follows from (1). (3) Let w be defined as in (1). Then, for $B \in \mathbb{B}$, $w(B) = w_1[E_{w_2}(B\mid\mathbb{B})] = w_1(B)$ so that $w \in S_{w_1}^{\mathbb{B}}$. Furthermore, for $B \in \mathbb{B}$, $A \in \mathbb{A}$ we have, by Theorem 3.5 (4), $E_{w_2}(AB\mid\mathbb{B}) = E_{w_2}(AE_w(B\mid\mathbb{B})\mid\mathbb{B}) = E_{w_2}(A\mid\mathbb{B})E_w(B\mid\mathbb{B}) = E_{w_2}(A\mid\mathbb{B})B$. We then obtain, for $B \in P_{\mathbb{B}}$, $A \in \mathbb{A}$, the following: $W[E_w(A\mid\mathbb{B})B] = w_A(B) = w(AB) = w_1[E_{w_2}(AB\mid\mathbb{B})] = w_1[E_{w_2}(A\mid\mathbb{B})B] = w[E_{w_2}(A\mid\mathbb{B})B]$. Hence $E_w(A\mid\mathbb{B}) = E_{w_2}(A\mid\mathbb{B})$ a.e. $[w]$. It is also easy to show that $w \sim w_2$ so that $w \in \mathbf{S}_{w_2}^{\mathbb{B}}$.

Theorem 4.1 (2) can be interpreted by saying that if the *a priori* probability $w^{\mathbb{B}}$ of B and the $\mathbb{B}$-expectations of all $A \in \mathbb{A}$ with respect to w are known, then the state w is uniquely defined by $w(A) = w^{\mathbb{B}}[E_w(A\mid\mathbb{B})]$.

We now give a characterization of $\mathbb{B}$-sufficiency which differs from the one of Halmos and Savage.[6]

Theorem 4.2: Suppose $\mathbb{B} \subseteq \mathbb{B}'$ and $\mathbb{A} = \mathbb{B}(H)$. Then $\mathbb{B}$ is sufficient for a set of states $M \subseteq \mathbb{B}'$ on $\mathbb{A}$ if and only if there exists a state v on $\mathbb{A}$ such that $M^{\mathbb{B}} \sim v^{\mathbb{B}}$ and $dw_A^{\mathbb{B}}/dv_A^{\mathbb{B}} = dw^{\mathbb{B}}/dv^{\mathbb{B}}$ a.e. $[v_A]$ for every $A \in \mathbb{A}$, $w \in M$.

Proof: In order to see that the condition is sufficient, we claim $E_w(A\mid\mathbb{B}) = E_v(A\mid\mathbb{B})$ a.e. $[w]$ for all $A \in \mathbb{A}$. Indeed, $w^{\mathbb{B}}[E_v(A\mid\mathbb{B})B] = v^{\mathbb{B}}[(dw^{\mathbb{B}}/dv^{\mathbb{B}})E_v(A\mid\mathbb{B})B] = v^{\mathbb{B}}[E_v(A\mid\mathbb{B})Bdw^{\mathbb{B}}/dv^{\mathbb{B}}] = v_A^{\mathbb{B}}(B(dw^{\mathbb{B}}/dv^{\mathbb{B}})) = v_A^{\mathbb{B}}(B(dw_A^{\mathbb{B}}/dv^{\mathbb{B}})) = w_A^{\mathbb{B}}(B) = w^{\mathbb{B}}[E_w(A\mid\mathbb{B})B]$ for all $B \in P^{\mathbb{B}}$. For necessity, via a proof similar to Lemma 7 of Halmos and Savage, $M^{\mathbb{B}}$ has an equivalent countable subset $\{w_i^{\mathbb{B}}\}$. Let $v(A) = \sum\alpha_i w_i(A)$, $A \in \mathbb{A}$, $\alpha_i > 0$, $\sum\alpha_i = 1$. Then $M^{\mathbb{B}} \sim v^{\mathbb{B}}$ and $v \in \mathbb{B}'$. Let $E_w(A\mid\mathbb{B}) = E(A\mid\mathbb{B})$ a.e. $[w]$, for every $w \in M$. For fixed $w \in M$ we have $v_A^{\mathbb{B}}((dw^{\mathbb{B}}/dv^{\mathbb{B}})B) = \sum\alpha_i w_{i_A}^{\mathbb{B}}((dw^{\mathbb{B}}/dv^{\mathbb{B}})B) = \sum\alpha_i w_i^{\mathbb{B}}[E_{w_i}(A\mid\mathbb{B})(dw^{\mathbb{B}}/dv^{\mathbb{B}})B] = \sum\alpha_i w_i^{\mathbb{B}}[E(A\mid\mathbb{B})(dw^{\mathbb{B}}/dv^{\mathbb{B}})B] = v^{\mathbb{B}}[E(A\mid\mathbb{B})(dw^{\mathbb{B}}/dv\ )B] = w^{\mathbb{B}}[E(A\mid\mathbb{B})B] = w^{\mathbb{B}}[E_w(A\mid\mathbb{B})B] = w_A^{\mathbb{B}}(B) = v_A^{\mathbb{B}}((dw_A^{\mathbb{B}}/dv_A^{\mathbb{B}})B)$. Hence $dw^{\mathbb{B}}/dv^{\mathbb{B}} = dw_A/dv_A^{\mathbb{B}}$ a.e. $[v_A]$.

Classical Case

Let (Ω, Σ, μ) be a measure space, $\{\Delta_i \in \Sigma\}$ a countable partition of Ω and $\Delta \in \Sigma$. We use the notation $w(f) =$

$\int f d\mu$. If $\mathbb{B}$ is the sub σ-algebra generated by $\{\Delta_i\}$ we have $E_w(\chi_\Delta \mid \mathbb{B}) = \sum[w_\Delta(\chi_{\Delta_i})/w(\chi_{\Delta_i})]\chi_{\Delta_i}$ and, applying Theorem 4.1 (3), $\mu(\Delta) = w[E_w(\chi_\Delta \mid \mathbb{B})] = \sum[\mu(\Delta \cap \Delta_i)/\mu(\Delta_i)]\,\mu(\Delta_i) = \sum\mu(\Delta \mid \Delta_i)\mu(\Delta_i)$, which is a familar elementary formula. Similarly, if $\mu \in S^{\mathbb{B}}_{\mu_1} \cap \mathbf{S}^{\mathbb{B}}_{\mu_2}$, we obtain $\mu(\Delta) = w_1[E_{w_2}(\chi_\Delta \mid \mathbb{B})] = \sum[\mu_2(\Delta \cap \Delta_i)/\mu_2(\Delta_i)]\,\mu_1(\Delta_i)$.

5. COARSE-GRAINING

Let $\mathbb{A}$ be an arbitrary von Neumann algebra and let w, v be integrals on $\mathbb{A}$ satisfying $w < v$. We say w *admits a v-density dw/dv* if there exists an operator $dw/dv \in L_1(\mathbb{A}, v)$ such that $w(A) = v((dw/dv)A)$ for all $A \in \mathbb{A}$. We denote the set of such integrals by $D(v) = \{w\colon w < v, w$ admits a v-density$\}$. There are many important special cases in which w admits a v-density when $w < v$. For example, in the classical case $(\mathbb{A} \subseteq \mathbb{A}')$, in the pure quantum case $[\mathbb{A} = \mathbb{B}(H)]$, and when $\mathbb{A}$ is finite, σ-finite and v is cyclic, we know this is true. In these special cases, the ordinary Radon–Nikodym derivative, the density operator W, and Dye's derivative, respectively, are the densities.

Definition 5.1: Let $\mathbb{B} \subseteq \mathbb{A}$ be von Neumann algebras and $w \in D(v)$ states on $\mathbb{A}$. The $(\mathbb{B}, v)$ coarse-graining of w is a state $C^{\mathbb{B}}_v w \in D(v) \cap S^{\mathbb{B}}_w$ satisfying $dC^{\mathbb{B}}_v w/dv \in L_1(\mathbb{B}, v)$.

This definition is sufficiently general to encompass the usual cases of coarse-graining in the classical and pure quantum cases. Usually the $\mathbb{B}$ coarse-graining of w is defined as a state in $S^{\mathbb{B}}_w$ with maximal entropy; however, our definition is more general since the maximal entropy may be infinite in which case there may not exist a unique coarse-graining state.

Lemma 5.1: (i) A state w' on $\mathbb{A}$ is a $(\mathbb{B}, v)$ coarse-graining of w if and only if (1) $w' \in D(v), dw'/dv \in L_1(\mathbb{B}, v)$ and (2) $v((dw'/dv)B) = w(B)$ for all $B \in \mathbb{B}$. (ii) If $C^{\mathbb{B}}_v w$ exists, it is unique.

Proof: (i) Since $w'(B) = v((dw'/dv)B)$ for all $B \in \mathbb{B}, w' \in S^{\mathbb{B}}_w$ if and only if $w(B) = v((dw'/dv)B)$ for all $B \in \mathbb{B}$. (ii) If w' is a $(\mathbb{B}, v)$ coarse-graining of w, by (2) $dw'/dv = dC^{\mathbb{B}}_v w/dv$ a.e. $[v]$ so that $w' = C^{\mathbb{B}}_v w$.

This last lemma permits us to verify the special cases easily. For example, if $w \in D(v)$, then $dC^{\mathbb{A}}_v w/dv = dw/dv$ and $dC^{\mathbb{B}}_w w/dw = I$ since $v((dw/dv)A) = w(A)$ for all $A \in \mathbb{A}$ and $w((dC^{\mathbb{B}}_w w/dw)B) = w(B)$ for all $B \in \mathbb{B}$.

Theorem 5.1: If $w^{\mathbb{B}} \in D(v^{\mathbb{B}})$, then $C^{\mathbb{B}}_v w$ exists, is unique a.e. $[v^{\mathbb{B}}]$ and $dC^{\mathbb{B}}_v w/dv = dw^{\mathbb{B}}/dv^{\mathbb{B}}$ a.e. $[v^{\mathbb{B}}]$.

Proof: Since $w^{\mathbb{B}} \in D(v^{\mathbb{B}})$, we have $dw^{\mathbb{B}}/dv^{\mathbb{B}} \in L_1(\mathbb{B}, v)$ and $v[(dw^{\mathbb{B}}/dv^{\mathbb{B}})B] = v^{\mathbb{B}}[(dw^{\mathbb{B}}/dv^{\mathbb{B}})B] = w^{\mathbb{B}}(B)$ $w(B)$ for all $B \in \mathbb{B}$. Therefore, the integral $w'(A) = v[(dw^{\mathbb{B}}/dv^{\mathbb{B}})A]$ satisfies (1) and (2) and so is a $(\mathbb{B}, v)$ coarse-graining of w. This proves existence. Now let $C^{\mathbb{B}}_v w$ be a $(\mathbb{B}, v)$ coarse-graining of w. Then (1) and (2) imply $v^{\mathbb{B}}[(dC^{\mathbb{B}}_v/dv)B] = v[(dC^{\mathbb{B}}_v w/dv)B] = w(B)$ $= w^{\mathbb{B}}(B) = v^{\mathbb{B}}[(dw^{\mathbb{B}}/dv^{\mathbb{B}})B]$ for every $B \in \mathbb{B}$. It follows that $dC^{\mathbb{B}}_v w/dv = dw^{\mathbb{B}}/dv^{\mathbb{B}}$ a.e. $[v^{\mathbb{B}}]$.

Let w, v be states on $\mathbb{A}$ satisfying $w \in D(v)$ and $dw/dv \in L_2(\mathbb{A}, v)$. Then the *information $I_v(w)$ of w with respect to v* is defined by $I_v(w) = \|dw/dv\|$, where $\|\cdot\|$ is the norm in $L_2(\mathbb{A}, v)$.

Theorem 5.2: If there is a state in $S^{\mathbb{B}}_w$ with finite information, then $C^{\mathbb{B}}_v w$ exists and is the unique state in $S^{\mathbb{B}}_w$ with minimal information.

Proof: Let $w_0 \in S^{\mathbb{B}}_w \cap D(v)$, $dw_0/dv \in L_2(\mathbb{A}, v)$. Now $L_2(\mathbb{B}, v)$ is a closed subspace of the Hilbert space $L_2(\mathbb{A}, v)$. Therefore, there exist unique elements A_1, A_2 such that $dw_0/dv = A_1 + A_2$, $A_1 \in L_2(\mathbb{B}, v)$, $A_2 \in L_2(\mathbb{B}, v)^{\perp}$. Now $v(A_2 B) = (A_2^*, B) = 0$ for every $B \in L_2(\mathbb{B}, v)$, where $(\cdot\,, \cdot)$ is the scalar product in $L_2(\mathbb{A}, v)$. Hence $v(A_1 B) = v[(A_1 + A_2)B] = v((dw_0/dv)B) = w_0(B) = w(B)$ for every $B \in \mathbb{B}$. If w_1 is the state $w_1(A) = v(A_1 A)$, it follows that $w_1 = C^{\mathbb{B}}_v w$. This implies $A_1 = dC^{\mathbb{B}}_v w/dv$. Hence $I_v(w_0) = \|dw_0/dv\| = (\|dC^{\mathbb{B}}_v w/dv\|^2 + \|A_2\|^2)^{1/2} \geq \|dC^{\mathbb{B}}_v w/dv\| = I_v(C^{\mathbb{B}}_v w)$. Thus $C^{\mathbb{B}}_v w$ has minimal information in $S^{\mathbb{B}}_w$. To show $C^{\mathbb{B}}_v w$ is the unique state in $S^{\mathbb{B}}_w$ with minimal information, suppose $w' \in S_w$ has minimal information. Then, if we write $dw'/dv = A_1' + A_2'$, $A_1' \in L_2(\mathbb{B}, v)$, $A_2' \in L_2(\mathbb{B}, v)^{\perp}$, we obtain, as above, $(\|dC^{\mathbb{B}}_v w/dv\|^2 + \|A_2'\|^2)^{1/2} = I_v(w') = I_v(C^{\mathbb{B}}_v w) = \|dC^{\mathbb{B}}_v w/dv\|$. Hence $A_2' = 0$ and $dw'/dv \in L_2(\mathbb{B}, v) \cap L_1(\mathbb{A}, v)$. Thus $dw'/dv \in L_1(\mathbb{B}, v)$ and hence w' is a $(\mathbb{B}, v)$ coarse-graining of w. It follows from Lemma 5.1 (ii) that $w' = C^{\mathbb{B}}_v w$.

Theorem 5.3: Let $P^{\mathbb{B}}$ be the projection of $L_2(\mathbb{A}, v)$ onto $L_2(\mathbb{B}, v)$. If $dw/dv \in L_2(\mathbb{A}, v)$ then $C^{\mathbb{B}}_v w$ exists, $dC^{\mathbb{B}}_v w/dv \in L_2(\mathbb{B}, v)$ and $dC^{\mathbb{B}}_v w/dv = P^{\mathbb{B}} dw/dv$.

Proof: From the proof on Theorem 5.2, it follows that $dw/dv = A_1 + A_2 = (dC^{\mathbb{B}}_v w/dv) + A_2$, $dC_v w/dv \in L_2(\mathbb{B}, v)$, $A_2 \in L_2(\mathbb{B}, v)^{\perp}$. Hence $P^{\mathbb{B}} A_2 = 0$ and $P^{\mathbb{B}} dw/dv = dC^{\mathbb{B}}_v w/dv$.

It follows from Theorem 5.3 that if $dw/dv \in L_2(\mathbb{A}, v)$ then $C^{\mathbb{B}}_v w$ is the element of $L_2(\mathbb{B}, v)$ such that $\|dC^{\mathbb{B}}_v w/dv - dw/dv\|$ is a minimum.

(1) Classical Case

If $\mathbb{A}$ is Abelian, $w < v$ states on $\mathbb{A}$, then dw/dv exists uniquely a.e. $[v]$ and is a density. It follows from Theorem 5.1 that $C^{\mathbb{B}}_v w(A) = v((dw^{\mathbb{B}}/dv^{\mathbb{B}})A)$ exists and is unique. The above theory can easily be generalized to integrals and even integrals with infinite values. The following example shows that $C^{\mathbb{B}}_v w$ may exist even if v has infinite values. However, this imposes some restrictions on $\mathbb{B}$ and w. Let λ be Lebesque measure and $\mu < \lambda$ on R. Let $\mathbb{A}$ be the Borel sets on R, $\{\Delta_i \in \mathbb{A}\}$ a countable partition of R and $\mathbb{B}$ the sub σ-algebra generated by $\{\Delta_i\}$. Then according to Theorem 5.1 we would expect $dC^{\mathbb{B}}_\lambda \mu/d\lambda = d\mu^{\mathbb{B}}/d\lambda^{\mathbb{B}} = \sum_{\lambda(\Delta_i) \neq 0}[\mu(\Delta_i)/\lambda(\Delta_i)]\chi_{\Delta_i}$. However, this formula is only valid if $\lambda(\Delta_i) < \infty$ whenever $\mu(\Delta_i) \neq 0$. In the other case the $(\mathbb{B}, \lambda)$ coarse-graining would not exist.

(2) Pure Quantum Case

Let $\mathbb{A} = \mathbb{B}(H)$. In this case there are many states v such that $w \in D(v)$ for every state w. Indeed, if ϕ_i is a complete orthonormal set for H, $\alpha_i > 0$, $\sum\alpha_i = 1$, let $V = \sum\alpha_i P_{\phi_i}$ be the density operator for v. In particular, if $\dim H = n$, then τ/n with density operator I/n is such a state. Again the above theory can be generalized to a certain extent to infinite value integrals. For example $w \in D(\tau)$ for every integral w and $dw/d\tau = W$ the density operator of w.

In this case it is natural to introduce the so-called Liouville space $L_2(\mathbb{A}, \tau)$ and its subspace $L_2(\mathbb{B}, \tau)$ and to consider the $(\mathbb{B}, \tau)$ coarse-graining. If $\mathbb{B}$ is Abelian, then one would expect $C_\tau^{\mathbb{B}} w(A) = \tau((dw^{\mathbb{B}}/d\tau^{\mathbb{B}})A)$. Hence, in simplified notation, if $C_\tau^{\mathbb{B}} w$ has the density operator $C^{\mathbb{B}}W$, then we would have $C^{\mathbb{B}}W = dw^{\mathbb{B}}/d\tau^{\mathbb{B}}$. As in the classical case $C_\tau^{\mathbb{B}} w$ exists if $\mathbb{B}$ and w are suitably restricted.

As an example, let $\mathbb{B} = \{B_i\}''$ be a countable partition. Then one would expect $dw^{\mathbb{B}}/d\tau^{\mathbb{B}} = \sum_{\tau(B_i)\neq 0} [w(B_i)/\tau(B_i)]B_i = dC_\tau^{\mathbb{B}} w/d\tau = C^{\mathbb{B}}W$. This formula is valid if and only if the projections B_i not annihilated by w are finite, $\tau(B_i) < \infty$. In the other case, $dC_\tau^{\mathbb{B}} w/d\tau$ does not exist. In this example Theorem 5.3 is particularly transparent geometrically, since when it exists $C^{\mathbb{B}}W = \sum_{\tau(B_i)\neq 0}[w(B_i)/\tau(B_i)]B_i = \sum_{\tau(B_i)\neq 0}(W, B_i/\sqrt{\tau(B_i)})B_i/\sqrt{\tau(B_i)} = P^{\mathbb{B}}W$ and we have:

Lemma 5.2: $C^{\mathbb{B}}W = E_\tau(W \mid \mathbb{B})$.

Proof:

$$E_\tau(W \mid \mathbb{B}) = \sum_{\tau(B_i)\neq 0} \frac{\tau_w(B_i)}{\tau(B_i)} B_i = \sum_{\tau(B_i)\neq 0} \frac{\tau(B_i W B_i)}{\tau(B_i)} B_i$$

$$= \sum_{\tau(B_i)\neq 0} \frac{w(B_i)}{\tau(B_i)} B_i = C\ W.$$

For the information we obtain, as in Theorem 5.2, $I_\tau(w_0) = \|dw_0/d\tau\| = \|W_0\| = [\tau(W_0^2)]^{1/2} \geq \|dC_\tau^{\mathbb{B}} w_0/d\tau\| = \|C^{\mathbb{B}}W\| = I_\tau(C^{\mathbb{B}}W)$ for any $w_0 \in S_w^{\mathbb{B}}$. If we define the *entropy* of W by $H(W) = -\tau(W \log W)$, we notice that $-H(W)$ has the same extrema as $I_\tau(W) = [\tau(W^2)]^{1/2}$. Hence Theorem 5.2 may be stated as follows:

Theorem 5.4: If $C^{\mathbb{B}}W$ exists and $H(C^{\mathbb{B}}W) < \infty$, then $C^{\mathbb{B}}W$ is the unique state in $S_w^{\mathbb{B}}$ with maximal entropy.

Remark: As we have seen, coarse-graining with respect to τ is rather restricted, since it exists only if $\mathbb{B}$ has a finite discrete part. Furthermore, the classical case does not admit a universally dominating measure like τ in the quantum case (for example, the Lebesgue measure λ on R dominates only the absolutely continuous measures on R). These observations were part of our motivation for defining a relative coarse-graining of w with respect to another state v which may not be τ or λ. Our concept may, however, be still too restricted, since we assumed the existence of densities. It may be possible to define the coarse-graining in terms of Dye's operators T rather than TT^*.

6. FILTERING

In this section we relate the ordinary concepts of filtering (or pinching, or reduction) of states with respect to a discrete observable to our previous theory and to coarse-graining in particular. Our results represent essentially a generalization to infinite algebras of similar results obtained previously by Davis.[8] In his paper, Davis also gives a rather detailed and critical discussion of the concept of conditioning, the main concern of Sec. 3.

Let $\mathbb{A} = \mathbb{B}(H)$ be a type I von Neumann algebra, $\mathbb{B} = \{B_i\}''$ a discrete Abelian subalgebra generated by the partition $\{B_i\}$ and w a state on $\mathbb{A}$ with density operator W. The $\mathbb{B}$-*filtering* is a map $F^{\mathbb{B}}$ from states into states defined by $F^{\mathbb{B}}W = \sum B_i W B_i$. [Notice that $F^{\mathbb{B}}W$ is in fact a state, since it is positive and $\tau(F^{\mathbb{B}}W) = \sum \tau(WB_i) = \tau(W) = 1$.]

Theorem 6.1: Let $P^{\mathbb{B}'}$ be the projection of the Liouville space $L_2(\mathbb{A}, \tau)$ onto $L_2(\mathbb{B}', \tau)$. Then $F^{\mathbb{B}}W = P^{\mathbb{B}'}W$.

Proof: $F^{\mathbb{B}}W \in \mathbb{B}'$ since for any $B_j \in \{B_i\}$ we have $(F^{\mathbb{B}}W)B_j = \sum_i B_i W B_i B_j = B_j W B_j = B_j \sum_i B_i W B_i = B_j(F^{\mathbb{B}}W)$. It follows that $F^{\mathbb{B}}W \in L_2(\mathbb{B}', \tau)$. Since $W = F^{\mathbb{B}}W + (W - F^{\mathbb{B}}W)$, it remains to show that $W - F^{\mathbb{B}}W \in L_2(\mathbb{B}', \tau)^\perp$. It suffices to show that $(W - F^{\mathbb{B}}W) \perp A$ for all $A \in \mathbb{B}'$. If $A \in \mathbb{B}'$, we have $(W, A) = (W, \sum B_i A) = \sum (W, B_i A) = \sum (W, B_i A B_i) = \sum \tau(WB_i A B_i) = \sum \tau(B_i W B_i A) = \tau(\sum B_i W B_i A) = \tau(F^{\mathbb{B}}WA) = (F^{\mathbb{B}}W, A)$. Hence $(W - F^{\mathbb{B}}W, A) = 0$.

Theorem 6.2: $F^{\mathbb{B}}W$ is the state in $S_w^{\mathbb{B}} \cap \mathbb{B}'$ with maximal information.

Proof: We already showed that $F^{\mathbb{B}}W \in \mathbb{B}'$. Now $F^{\mathbb{B}}W \in S_w^{\mathbb{B}}$ since $\tau(F^{\mathbb{B}}WB_j) = \tau(\sum_i B_i W B_i B_j) = \tau(WB_j) = w(B_j)$ for every $B_j \in \{B_i\}$. Since $\|F^{\mathbb{B}}W - W\|^2$ is minimal among all elements of $L_2(\mathbb{B}', \tau)$ and $\|F^{\mathbb{B}}W - W\|^2 = \|W\|^2 - \|F^{\mathbb{B}}W\|^2$, we must have $I_\tau(F^{\mathbb{B}}W) = \|F^{\mathbb{B}}W\|$ maximal.

Theorem 6.3: Coarse-graining and filtering are idempotent operators satisfying: (1) $C^{\mathbb{B}}C^{\mathbb{B}} = C^{\mathbb{B}}$, $F^{\mathbb{B}}F^{\mathbb{B}} = F^{\mathbb{B}}$, $C^{\mathbb{B}}F^{\mathbb{B}} = F^{\mathbb{B}}C^{\mathbb{B}} = C^{\mathbb{B}}$, (2) $F^{\mathbb{B}} = C^{\mathbb{B}}$ if and only if $\mathbb{B}$ is maximal Abelian.

Proof: (1) follows from the fact that $C^{\mathbb{B}}$ and $F^{\mathbb{B}}$ are projections onto $L_2(\mathbb{B}, \tau)$ and $L_2(\mathbb{B}', \tau)$, respectively, and that $L_2(\mathbb{B}, \tau)$ is a closed subspace of $L_2(\mathbb{B}', \tau)$ since $\mathbb{B} \subseteq \mathbb{B}'$. The same reasoning proves (2) since $F^{\mathbb{B}} = P^{\mathbb{B}'} = P^{\mathbb{B}} = C^{\mathbb{B}}$ if and only if $\mathbb{B} = \mathbb{B}'$ is maximal Abelian.

A simple example helps to visualize the difference between filtering and coarse-graining. Let $\dim H = 3$, $\mathbb{A} = \mathbb{B}(H)$, $\mathbb{B} = \{B, B^\perp\}''$, where $B = \mathrm{diag}(1, 1, 0)$, $B^\perp = \mathrm{diag}(0, 0, 1)$. Then

$$S_w^{\mathbb{B}} = \{V: V_{11} + V_{22} = W_{11} + W_{22}, V_{33} = W_{33}\};$$

$$F^{\mathbb{B}}W = \begin{pmatrix} W_{11} & W_{12} & 0 \\ W_{21} & W_{22} & 0 \\ 0 & 0 & W_{33} \end{pmatrix};$$

$$C^{\mathbb{B}}W = \mathrm{diag}(\tfrac{1}{2}(W_{11} + W_{22}), \tfrac{1}{2}(W_{11} + W_{22}), W_{33}).$$

As seen in this example, both filtering and coarse-graining are reduced by the subspaces $B_i \in \mathbb{B}$. But while filtering conserves the entire information within the "boxes," coarse-graining destroys it completely.

---

* The research for this work was conducted under N.S.F. Grant GP-23304.

[1] H. A. Dye, Trans. Amer. Math. Soc. **72**, 243 (1952).
[2] I. E. Segal, Ann. Math. **57**, 401 (1953).

[3] H. Umegaki, Tohoku Math. J. **6**, 177 (1954); Kodai Math. Sem. Rep. **14**, 59 (1962); Proc. Japan Acad. **37**, 459 (1961); M. Nakamura and T. Turumaru, Tohoku Math. J. **6**, 182 (1954).
[4] A. M. Gleason, J. Ratl. Mech. Anal. **6**, 885 (1957).
[5] See, for example, J. Schwartz, *W*-Algebras* (Gordon and Breach, New York, 1967).
[6] P. R. Halmos and L. J. Savage, Ann. Math. Stat. **20**, 225 (1949); S. Kullback and R. A. Leibler, *ibid.* **22**, 79 (1959); H. Umegaki, Kodai Math. Sem. Rep. **11**, 51 (1959).
[7] G. Emch, Helv. Phys. Acta 38, 164 (1965); G. Emch and C. Favre, "Coarse-Graining in Liouville Space and Ergodicity," preprint, Geneva (1965).
[8] C. Davis, Ill. J. Math. **3**, 538 (1959); E. B. Davies and J. T. Lewis, Commun. Math. Phys. **17**, 239 (1970); C. M. Edwards, *ibid.* **16**, 207 (1970).