JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 4, NUMBER 4 APRIL 1963

Axiomatic Perturbation Theory. II. Partial Sum Formalism*

H. M. FRIED
Courant Institute of Mathematical Sciences, New York University, New York, N. Y.
(Received 19 November 1962)

A generalization of axiomatic perturbation theory is described in which propagator structure is introduced into the basic formalism, effectively resulting in a systematic rearrangement of the perturbation expressions representing a given process. The exact retarded functional of a simple Boson field theory is decomposed into a sum of approximate functionals, and all the n-point functions of the first few approximate functionals are determined from modified unitarity equations and a knowledge of the corresponding perturbation functions. The resulting amplitudes differ from the perturbation amplitudes by the replacement of the perturbative spectral functions with related spectral functions depending upon the latter and on the propagator structure. Sets of integral equations for the higher functionals are proposed and solved.

I. INTRODUCTION

IT is known that perturbation theory provides a general method for the construction of the n-point functions of relativistic field theory directly from the axioms, without the introduction of additional analyticity assumptions. By axiomatic perturbation theory, one means the non-Lagrangian methods[1-3] of approximating unitarity, and for this discussion, specifically those of I. The purpose of this note is to describe a variant of this procedure which, by introducing propagator self-energy structure into the formalism at an early stage, may be viewed as a rearrangement of the perturbation contributions representing a given process. This latter statement would be strictly correct if one could prove the convergence of both the conventional perturbation expressions and those obtained from the methods described here; since neither property is probably true, the partial sum formalism represents an independent, alternate method of passing from the axioms to arbitrary n-point functions of physical interest.

The essential modifications of the ordinary perturbation amplitudes, at least for the simplest cases which have been calculated, consist of the replacement of the perturbative spectral functions by related functions depending on the latter and on the propagator structure. One has the possibility of choosing a spectral function corresponding to the absorptive part of selected, proper self-energy propagator parts such that corresponding structure is inserted, in a systematic way, into all the other n-point functions. In principle, the choice of this function is open, the crudest selection corresponding to the replacement of "bare" lines by "dressed" lines representing the (daisy chain) sum of simple bubbles. In practice, a more natural choice is suggested by the possibility of removing all self-energy legs from the simplest set of nontrivial amputated n-point functions, with the consequence of inserting a more complicated structure into all lines. This latter choice will be made here, and a corresponding equation exhibited for the approximate propagator

* Supported in part by a Ford Foundation Grant.
[1] H. Lehmann, K. Symanzik, and W. Zimmerman, Nuovo Cimento 1, 205 (1955).
[2] K. Nishijima, Phys. Rev. 119, 485 (1960).
[3] H. M. Fried, J. Math. Phys. 3, 1107 (1962); hereafter denoted as I.

spectral function. Whether the solutions to this equation are physically interesting, i.e., whether the approximate one-(dressed)-particle-exchange scattering amplitude will contain resonances, is not as yet known. Even if this is the case, one must still deal with the ghosts which will appear in a more realistic theory. However, the possibility of passing from the axioms to n-point functions in a generalized manner is not without interest, and all that is contemplated here is the description of the formalism.

A second, and more practical reason for the choice of the spectral function just described, is that it permits the construction of a set of modified amplitudes directly from unitarity, making use of the known analyticity and symmetry properties of the ordinary perturbation functions. This allows the statement of the partial sum formalism in the weak sense of unitarity rather than in the stronger sense of integral unitarity.[4] The same techniques are not directly applicable to the higher approximations, but the assumed existence of the formalism in its strong sense leads to sets of crossing symmetric integral equations for all the higher amplitudes.

As in I, the discussion will center about the construction of the retarded functional $\langle R_x \rangle$ extensively described by Symanzik,[5] and, as in I, for the reason that the unitarity condition for the retarded functional is simpler than that of the time-ordered functional. (Essentially the same remark is that the retarded functions are all connected.) In Sec. II, a short review of the intermediate-state λ expansion defining the ordinary axiomatic perturbation theory is given, together with its generalization to the partial sum formalism in the weak sense. Construction of the simplest n-point functions and their "radiative corrections" is described in Secs. III and IV for the case of a self-interacting neutral scalar boson field with quanta of mass m, and similar calculations are sketched in Sec. V for a pseudoscalar field. The statement of the formalism in its strong sense and the resulting equations which may describe further corrections to the previous amplitudes are outlined in Sec. VI.

Before proceeding, it may be worthwhile to set down, following Symanzik, the basic formulas satisfied by the axiomatic retarded functional operator R_x. These are

(i) integral unitarity

$$R_{x,y} = i\theta(xy)[R_x, R_y], \tag{1}$$

which, for relativistic invariance, requires the space-like vanishing of the commutator of (1);

(ii) the asymptotic condition

$$R_x = :\exp\left(A_{IN}\frac{\delta}{\delta \mathbf{j}}\right):\langle R_x\rangle,$$

where the bracket means vacuum expectation value, the double dots denote a Wick product, and the symbol $A_{IN}\,\delta/\delta\mathbf{j}$ stands for the operator

$$\int du\, A_{IN}(u)K_u\frac{\delta}{\delta j(u)} = \int du\, A_{IN}(u)\frac{\delta}{\delta j(\mathbf{u})}.$$

Here, K_u represents the Klein–Gordon operator $m^2 - \Box_u$, and the operation $K_x\langle R_x\rangle = \langle R_x\rangle$ will be denoted by the phrase K–G amputation; when amputation in the sense of Symanzik[5] is intended, it will be denoted by the phrase S amputation. Combining (i) and (ii), one obtains the compact functional expression of integral unitarity,

$$\text{(iii)}\qquad \langle R_{x,y}\rangle = i\theta(xy)[\langle R_x\rangle, e^D, \langle R_y\rangle], \tag{2}$$

where

$$D = i\frac{\overleftarrow{\delta}}{\delta \mathbf{j}}\cdot\Delta^{(+)}\cdot\frac{\overrightarrow{\delta}}{\delta \mathbf{j}} = i\int du\, dv\,\frac{\overleftarrow{\delta}}{\delta j(\mathbf{u})}\,\Delta^{(+)}(u-v)\,\frac{\overrightarrow{\delta}}{\delta j(v)},$$

and the notation $[A, e^D, B] = Ae^DB - Be^DA$ has been used.

The conventional unitarity relations for the retarded n-point functions follow upon taking functional derivatives of the difference of (iii) and its permuted form.

$$\text{(iv)}\quad \langle R_{x,y_1y_2\cdots y_n}\rangle_0 - \langle R_{y_1,xy_2\cdots y_n}\rangle_0 = i\frac{\delta}{\delta j(y_2)}\cdots\frac{\delta}{\delta j(y_n)}[\langle R_x\rangle, e^D, \langle R_{y_i}\rangle]\Big|_{j\to 0},$$

where the subscript 0 on the left hand side of (iv) indicates that the source function, $j(z)$, has been set equal to zero. For the simple cases of spinless neutral boson fields treated here, all the retarded n-point functions are real, invariant, connected, and symmetric in their $n-1$ retarded coordinates.

II. THE λ EXPANSION AND A GENERALIZATION

The asymptotic condition (ii) explicitly exhibits dependence of the operator R_x on the operator A_{IN}, whose commutation properties then provide the structure of the unitarity condition (2). It is convenient to display the complete dependence of R_x on A_{IN}, as well as on the source j, by writing

$$R_x = R_x\{A_{IN}, j\} = :\exp\left(A_{IN}\frac{\delta}{\delta \mathbf{j}}\right):\langle R_x\{A_{IN}, j\}\rangle, \tag{3}$$

[4] As in I, integral unitarity denotes the axiomatic property of Eq. (1), while unitarity refers to the weaker statement, $R_{x,y} - R_{y,x} = i[R_x,R_y]$.

[5] K. Symanzik, J. Math. Phys. 1, 249 (1960).

where the c-number functional $\langle R_x\{A_{IN}, j\}\rangle$ then depends implicitly upon A_{IN} through the integral unitarity condition

$$\frac{\delta}{\delta j(y)} R_x\{A_{IN}, j\} = i\theta(xy)[R_x\{A_{IN}, j\}, R_y\{A_{IN}, j\}], \quad (4)$$

and its vacuum expectation value.

Consider now the scaling transformations

$$A_{IN} \to \lambda_1 A_{IN}, \qquad j \to \lambda_2 j,$$

where $\lambda_{1,2}$ are arbitrary, real numbers. The integral unitarity equation (4) becomes

$$\frac{1}{\lambda_2}\frac{\delta}{\delta j(y)} R_x\{\lambda_1 A_{IN}, \lambda_2 j\} = i\theta(xy)[R_x\{\lambda_1 A_{IN}, \lambda_2 j\}, R_y\{\lambda_1 A_{IN}, \lambda_2 j\}], \quad (5)$$

but in writing this the assumption of relativistic invariance has been made; namely, that the right-hand side of (5) shall vanish spacelike for arbitrary $\lambda_{1,2}$. This is quite plausible since j is an arbitrary source function and it is difficult to see how a change in the normalization of the in-field operator can destroy the spacelike commutivity of two operators which possess this property when $\lambda_1 = 1$. If this plausibility argument is accepted, then as indicated below, one has a simple reason for anticipating the relativistic invariance of all the amplitudes constructed in I. Further, invariance of the spacelike commutivity property under these scaling transformations suggests an immediate generalization which forms the conceptual basis of the partial sum formalism in the strong sense (Sec. VI).

With Eq. (3) modified to read

$$R_x\{\lambda_1 A_{IN}, \lambda_2 j\} = :\exp\left[\left(\frac{\lambda_1}{\lambda_2}\right) A_{IN}\frac{\delta}{\delta \mathbf{j}}\right]: \langle R_x\{\lambda_1 A_{IN}, \lambda_2 j\}\rangle, \quad (6)$$

Eq. (5) may be rewritten in the form

$$\frac{\delta}{\delta j(y)} R_x\{\lambda_1 A_{IN}, \lambda_2 j\} = i\lambda_2\theta(xy) :\exp\left[\left(\frac{\lambda_1}{\lambda_2}\right) A_{IN}\frac{\delta}{\delta \mathbf{j}}\right]: \times\left[\langle R_x\{\lambda_1 A_{IN}, \lambda_2 j\}\rangle, \times \exp\left[\left(\frac{\lambda_1}{\lambda_2}\right)^2 D\right], \langle R_y\{\lambda_1 A_{IN}, \lambda_2 j\}\rangle\right]. \quad (7)$$

For the special choice of the parameters $\lambda_{1,2}$,

$$\lambda_1 = \lambda^{-\frac{1}{2}}, \qquad \lambda_2 = \lambda^{-1},$$

and with the definition

$$R_x^\lambda\{A_{IN}, j\} = R_x\{\lambda^{-\frac{1}{2}} A_{IN}, \lambda^{-1} j\},$$

Eq. (6) and the vacuum expectation value of (7) become

$$R_x^\lambda\{A_{IN}, j\} = :\exp\left(\lambda^{\frac{1}{2}} A_{IN}\frac{\delta}{\delta \mathbf{j}}\right): \langle R_x^\lambda\{A_{IN}, j\}\rangle, \quad (8)$$

$$\frac{\delta}{\delta j(y)}\langle R_x^\lambda\rangle = \langle R_{x,y}^\lambda\rangle = \frac{i}{\lambda}\theta(xy)[\langle R_x^\lambda\rangle, e^{\lambda D}, \langle R_y^\lambda\rangle]. \quad (9)$$

Equations (8) and (9) represent the basic relations of I, which serve to define the successive, crossing symmetric, approximate functionals of perturbation theory; with

$$\langle R_x^\lambda\rangle = \sum_{i=0}^{\infty} \lambda^i \langle R_x^{(i)}\rangle,$$

the expansion of (9) in powers of λ produces integral unitarity equations for the Born functional $\langle R_x^{(0)}\rangle$ and functionals of higher index, $\langle R_x^{(i)}\rangle$, $j \geq 1$, whose n-point functions contain all the renormalized radiative corrections to the n-point Born functions. If, by the previous plausibility argument, the right-hand side of (9) is to vanish spacelike for arbitrary λ, then when (9) is expanded in powers of λ each term in the functional expansion must vanish spacelike, and the relativistic invariance of all amplitudes so constructed is assured.

The essential part of this scheme is the replacement of Eq. (2) by (9), and the identification of the exact functional with the sum of all the approximate functionals,

$$\langle R_x\rangle = \sum_{i=0}^{\infty} \langle R_x^{(i)}\rangle,$$

since Eqs. (2) and (9) become identical when $\lambda = 1$. The partial sum formalism in its weak sense may now be defined by rewriting the difference of (2) and its permuted form in the manifestly identical form

$$\langle R_{x,y}\rangle - \langle R_{y,x}\rangle = i[\langle R_x\rangle, e^{D+D_B-D_B}, \langle R_y\rangle], \quad (10)$$

where D_B is an arbitrary functional differential operator, here chosen as[6]

$$D_B = i\frac{\overleftarrow{\delta}}{\delta \mathbf{j}}\cdot\Delta_B^{(+)}\cdot\frac{\overrightarrow{\delta}}{\delta \mathbf{j}} = i\int du\, dv \frac{\overleftarrow{\delta}}{\delta j(\mathbf{u})}\Delta_B^{(+)}(u-v)\frac{\overrightarrow{\delta}}{\delta j(\mathbf{v})},$$

with

$$\Delta_B^{(+)}(x) = \int_{M^2}^{\infty} d\kappa^2\, C_B(\kappa^2)\Delta^{(+)}(x, \kappa^2), \quad M^2 \geq 4m^2.$$

[6] The Klein–Gordon operator entering into the definition of D_B is to refer to the mass m.

If the same λ-expansion procedure is applied to (10), where the latter is rewritten in the form

$$\langle R^{\lambda}_{x,y}\rangle - \langle R^{\lambda}_{y,x}\rangle = \frac{i}{\lambda}[\langle R^{\lambda}_{x}\rangle, \exp(\lambda D + \lambda^{\alpha} D_B - \lambda^{\beta} D_B), \langle R^{\lambda}_{y}\rangle], \quad (11)$$

with α and β integers, $\beta > \alpha > 0$, then for $\lambda = 1$, Eqs. (11), (10), and the permuted form of (2) become identical. However, the expansion of (11) in powers of λ generates unitarity equations for new, approximate functionals which differ from those of the ordinary perturbation theory, although their sum (assuming convergence) is just the exact functional; this then corresponds to a "rearrangement" of the perturbation series for each n-point function. For all 2-point functions, where there is no danger in using the formalism in its stronger sense; the unitarity relations obtained from (11) will be replaced by corresponding integral unitarity equations.

Although this procedure provides the basis for a generalized formalism, it is intuitively clear that the rearrangements so obtained bear a close resemblance to the ordinary perturbation forms. The measure of this similarity is an indication of how much effort must be expended before one obtains results which differ significantly from, and illustrate the essential modifications of the ordinary perturbation forms. Roughly speaking, for certain choices of the parameters α, β one must calculate at least the first few approximate propagators (which will be identical to the renormalized perturbation functions) before seeing the effect of the D_B terms; this occurs for $\beta > \alpha \geq 2$. The simplest and perhaps most easily interpretable results follow from the choice $\beta > \alpha = 1$, and the discussion of the scalar Boson field functions to follow will be based upon the simplest of these possibilities, $\alpha = 1$, $\beta = 2$ (and $M^2 = 4m^2$).

With $\langle R^{\lambda}_{x}\rangle = \sum_{j=0}^{\infty} \lambda^j \langle \bar{R}^{(j)}_{x}\rangle$, the barred notation serving to distinguish the functionals of index j from their perturbation counterparts of I, the expansion in powers of λ of

$$\langle R^{\lambda}_{x,y}\rangle - \langle R^{\lambda}_{y,x}\rangle = \frac{i}{\lambda}[\langle R^{\lambda}_{x}\rangle, \exp[\lambda(D + D_B) - \lambda^2 D_B], \langle R^{\lambda}_{y}\rangle] \quad (12)$$

defines an infinite set of approximate unitarity equations,

$$\langle \bar{R}^{(0)}_{x,y}\rangle - \langle \bar{R}^{(0)}_{y,x}\rangle = i[\langle \bar{R}^{(0)}_{x}\rangle, D + D_B, \langle \bar{R}^{(0)}_{y}\rangle], \quad (13)$$

$$\langle \bar{R}^{(1)}_{x,y}\rangle - \langle \bar{R}^{(1)}_{y,x}\rangle = i\Big\{\Big[\langle \bar{R}^{(0)}_{x}\rangle, \frac{(D + D_B)^2}{2!} - D_B, \langle \bar{R}^{(0)}_{y}\rangle\Big] + [\langle \bar{R}^{(1)}_{x}\rangle, D + D_B, \langle \bar{R}^{(0)}_{y}\rangle] + [\langle \bar{R}^{(0)}_{x}\rangle, D + D_B, \langle \bar{R}^{(1)}_{y}\rangle]\Big\}, \text{ etc.} \quad (14)$$

As indicated in Sec. VI, there is reason to believe that each of these relations may be replaced by the stronger integral unitarity statement; but, with the exception of the 2-point functions, only these simpler unitarity conditions are needed to determine the functionals of lower index.

III. FUNCTIONS OF INDEX ZERO

A. The 2-Point Function

In the limit of zero source, Eq. (13) becomes

$$\langle \bar{R}^{(0)}_{x,y}\rangle_0 - \langle \bar{R}^{(0)}_{y,x}\rangle_0 = i^2 \int \Delta'(u - v)\langle \bar{R}^{(0)}_{x,u}\rangle_0 \langle \bar{R}^{(0)}_{y,v}\rangle_0, \quad (15)$$

where

$$\Delta'(x) = \Delta(x, m^2) + \Delta_B(x),$$

$$\Delta_B(x) = \Delta_B^{(+)}(x) + \Delta_B^{(-)}(x) = \int_{4m^2}^{\infty} d\kappa^2\, C_B(\kappa^2)\Delta(x, \kappa^2).$$

The existence of a stable particle of mass m implies that the Fourier transform of the exact 2-point function has a pole of unit residue on the mass shell. If the corresponding residue of the zero-index Fourier transform is denoted by Z, it follows from Eq. (15) that $Z = 0$ or 1. The choice $Z = 1$ then implies that the Fourier transforms of all the remaining 2-point functions are not to have a mass shell pole,

$$\int dv\, \Delta^{(\pm)}(u - v, m^2)\langle \bar{R}^{(j)}_{x,v}\rangle_0 = 0, \qquad j \geq 1, \quad (16)$$

since (15) may be rewritten in the form

$$\langle \bar{R}^{(0)}_{x,y}\rangle_0 - \langle \bar{R}^{(0)}_{y,x}\rangle_0 = -\Delta(x - y, m^2) + i^2 \int \Delta_B(u - v)\langle \bar{R}^{(0)}_{x,u}\rangle_0 \langle \bar{R}^{(0)}_{y,v}\rangle_0. \quad (17)$$

Because it is known that the exact momentum space propagator satisfies a simple dispersion relation, each of the approximate 2-point functions may be taken to obey the same relation; in the language used here, this means that Eq. (17) and all similar unitarity statements for the higher-index propagators may be replaced by the corresponding integral unitarity conditions,

$$\langle \bar{R}^{(0)}_{x,y}\rangle_0 = \Delta_R(x-y) + i^2\theta(xy) \times \int \Delta_B(u-v)\langle \bar{R}^{(0)}_{x,u}\rangle_0\langle \bar{R}^{(0)}_{y,v}\rangle_0, \quad (18)$$

since equations of the form of (18) will always be relativistic statements, and there is no danger in using the formalism in its strong sense here. If the momentum space propagator is defined by

$$\langle R_{x,y}\rangle_0 = (2\pi)^{-4}\int dp\, r(p)\, \exp\,[ip\cdot(x-y)],$$

the Fourier transform of (18) becomes

$$\bar{r}^{(0)}(s) = \frac{1}{m^2 - s - i\epsilon} + \int_{4m^2}^{\infty} \frac{ds'\,\Pi(s')\,|\bar{r}^{(0)}(s')|^2}{s' - s - i\epsilon}, \quad (19)$$

where we have taken $s = -p^2$, $p_0 > 0$, and set $\Pi(s) = (s-m^2)^2 C_B(s) > 0$.

The reason for the phrase "partial sum formalism" is now evident, since Eq. (19) has exactly the same structure as the axiomatic dispersion relation for the exact propagator.[5] In the latter case, $\Pi(s)$ is given by the absorptive part of the proper self-energy diagrams; in this case, $\Pi(s)$ is arbitrary, with special choices corresponding to the selection of special self-energy terms. For example, if $\Pi(s)$ is replaced by the lowest-order perturbation bubble contribution (two-particle phase space term in a theory with nonvanishing vertex of coupling constant g),

$$\Pi(s) \to \frac{g^2}{32\pi^2}\,\theta(s-4m^2)\left[1 - \frac{4m^2}{s}\right]^{\frac{1}{2}}, \quad (20)$$

then the solution to (19) is just the daisy chain representing the sum of an infinite string of such bubbles, and this structure will automatically be introduced into every other n-point function.

The appropriate solution to (19), omitting CDD zeros, is given by

$$\bar{r}^{(0)}(s) = [m^2 - s - i\epsilon]^{-1} \times \left[1 - (m^2 - s)\int_{4m^2}^{\infty} \frac{ds'\,\Pi(s')}{(s' - s - i\epsilon)(m^2 - s')^2}\right]^{-1}, \quad (21)$$

and it will be convenient, following the discussion in the introduction, to leave open the choice of $\Pi(s)$ until the next section. It may be worthwhile to remark that the inclusion of CDD zeroes in (21) would give results quite different from those obtained by summing the conventional perturbation graphs.

B. The 3-Point Function

A single functional differentiation of Eq. (13) yields, in the limit of zero source,

$$\langle \bar{R}^{(0)}_{x,yz}\rangle_0 - \langle \bar{R}^{(0)}_{y,xz}\rangle_0 = i\{[\langle \bar{R}^{(0)}_{x,z}\rangle,\, D + D_B,\, \langle \bar{R}^{(0)}_{y}\rangle] + [\langle \bar{R}^{(0)}_{x}\rangle,\, D + D_B,\, \langle \bar{R}^{(0)}_{y,z}\rangle]\}. \quad (22)$$

A solution of (22) is suggested by the form of the exact n-point amplitude which appears as a general S-amputated function adjoined to "dressed" propagator legs.[5] With a judicious choice of legs the remaining structure of the amputated amplitude may be reduced to a form sufficiently simple such that it may be read off from the corresponding perturbation construction of I. The simplest illustration of the procedure occurs for this 3-point function of index zero, and proceeds as follows:

If the S-amputated amplitude[7] $\langle \hat{R}^{(0)}_{\mathbf{x},yz}\rangle_0$ is defined by the relation

$$\langle \bar{R}^{(0)}_{x,yz}\rangle_0 = \int \langle \bar{R}^{(0)}_{x,x'}\rangle_0\langle \bar{R}^{(0)}_{y',y}\rangle_0\langle \bar{R}^{(0)}_{z',z}\rangle_0\langle \hat{R}^{(0)}_{\mathbf{x}',\mathbf{y}'\mathbf{z}'}\rangle_0, \quad (23)$$

the substitution of (23) into (22), together with the use of (15), leads to the unitarity statement for the amputated 3-point amplitude,

$$\int \langle \bar{R}^{(0)}_{x,x'}\rangle_0\langle \bar{R}^{(0)}_{y,y'}\rangle_0\langle \bar{R}^{(0)}_{z',z}\rangle_0[\langle \hat{R}^{(0)}_{\mathbf{x}',\mathbf{y}'\mathbf{z}'}\rangle_0 - \langle \hat{R}^{(0)}_{\mathbf{y}',\mathbf{y}'\mathbf{z}'}\rangle_0] = 0.$$

This is satisfied by[8]

$$\langle \hat{R}^{(0)}_{\mathbf{x},yz}\rangle_0 = g\,\delta(x-y)\,\delta(y-z), \quad (24)$$

where g is an arbitrary constant (with dimensions of mass) which, as in I, is to represent the renormalized charge of the theory. Equation (24) is just the lowest-order K–G amputated vertex of perturbation theory; the only difference between the complete zero-index 3-point functions in this formalism and in the conventional perturbation theory is that here we have legs given by $\langle \bar{R}^{(0)}_{x,y}\rangle_0$ rather than by $\Delta_R(x-y)$.

C. The 4-Point Function

A second functional differentiation of Eq. (15) yields, in the limit of zero source,

$$\begin{aligned}\langle \bar{R}^{(0)}_{x,y_1y_2y_3}\rangle_0 - \langle \bar{R}^{(0)}_{y_1,xy_2y_3}\rangle_0 &= i\{[\langle \bar{R}^{(0)}_{x,y_2y_3}\rangle,\, D + D_B,\, \langle \bar{R}^{(0)}_{y_1}\rangle] \\ &+ [\langle \bar{R}^{(0)}_{x}\rangle,\, D + D_B,\, \langle \bar{R}^{(0)}_{y_1,y_2y_3}\rangle] \\ &+ [\langle \bar{R}^{(0)}_{x,y_2}\rangle,\, D + D_B,\, \langle \bar{R}^{(0)}_{y_1,y_3}\rangle] \\ &+ [\langle \bar{R}^{(0)}_{x,y_3}\rangle,\, D + D_B,\, \langle \bar{R}^{(0)}_{y_1,y_2}\rangle]\}_{j\to 0}, \end{aligned} \quad (25)$$

and this can be treated in the same manner as Eq. (22). Defining the zero-index S-amputated ampli-

[7] S amputation will be used for all the amputated $\langle \hat{R}^{(j)}_{\mathbf{x},y_1\cdots y_n}\rangle_0$ functions.

[8] This choice satisfies the required properties of connectedness, reality, symmetry, and Condition B of I.

tude $\langle \hat{R}^{(0)}_{x,y_1,y_2y_3}\rangle_0$,

$$\langle \bar{R}^{(0)}_{x,y_1y_2y_3}\rangle_0 = \int \langle \bar{R}^{(0)}_{x,x'}\rangle_0 \langle \bar{R}^{(0)}_{y_1',y_1}\rangle_0 \times \langle \bar{R}^{(0)}_{y_2',y_2}\rangle_0 \langle \bar{R}^{(0)}_{y_3',y_3}\rangle_0 \langle \hat{R}^{(0)}_{x',y_1'y_2'y_3'}\rangle_0,$$

substitution into (25), with the use of (24) and (15), provides a unitarity condition analogous to that of the vertex function

$$\hat{R}^{(0)}_{x,y_1y_2y_3}\rangle_0 - \langle \hat{R}^{(0)}_{y_1,xy_2y_3}\rangle_0 = g^2\delta(x-y_2)[\langle \bar{R}^{(0)}_{y_2,y_3}\rangle_0 - \langle \bar{R}^{(0)}_{y_3,y_2}\rangle_0]\delta(y_3-y_1) + (y_2 \leftrightarrow y_3), \quad (26)$$

where both sides of (26) are understood to be integrated over the zero-index propagators. This is satisfied by

$$\langle \hat{R}^{(0)}_{x,y_1y_2y_3}\rangle_0 = g^2 \sum_{p=1}^{3} \delta(x-y_1)\langle \bar{R}^{(0)}_{y_1,y_2}\rangle_0 \delta(y_2-y_3), \quad (27)$$

where $\sum_{p=1}^{3}$ denotes the sum of the three circular permutations of the coordinates $y_{1,2,3}$. Equations (26) and (27) have the same form as the corresponding K–G amputated perturbation expressions, with $\langle \bar{R}^{(0)}_{x,y}\rangle_0$ replacing $\Delta_R(x-y)$.

It is evident from these examples that the construction of all the remaining n-point functions of zero index can be carried through in a similar manner, resulting in S-amputated amplitudes which differ from the corresponding K–G amputated Born functions by the replacement of internal $\Delta_R(x-y)$ lines by $\langle \bar{R}^{(0)}_{x,y}\rangle_0$ lines; the complete zero-index functions are then obtained by adjoining $\langle \bar{R}^{(0)}_{x,y}\rangle_0$ legs to each S-amputated amplitude.

IV. FUNCTIONS OF INDEX ONE

A. Corrections to the Zero-Index Propagator

All the corrections to the zero-index 2-point function have the same general form and it is convenient to discuss them all at the same time, reserving the explicit calculations for the propagator of index one. From the integral unitarity form of (14), and similar relations of higher index, one obtains

$$\langle \bar{R}^{(j)}_{x,y}\rangle_0 = \theta(xy)Q^{(j)}(x-y) + i^2\theta(xy) \times \int \Delta'(u-v)\{\langle \bar{R}^{(j)}_{x,u}\rangle_0\langle \bar{R}^{(0)}_{y,v}\rangle_0 + \langle \bar{R}^{(0)}_{x,u}\rangle_0\langle \bar{R}^{(j)}_{y,v}\rangle_0\}, \quad (28)$$

where the inhomogeneous $Q^{(j)}$ terms are constructed from n-point functions of index less than j. Each $\theta\cdot Q^{(j)}$ will have the spectral representation

$$\theta(x)Q^{(j)}(x) = \int_{4m^2} d\kappa^2\, \Pi^{(j)}(\kappa^2)\Delta_R(x,\kappa^2),$$

where $\Pi^{(j)}(\kappa^2)$ is real and vanishes for $\kappa^2 < 4m^2$; this is compatible with (16), and (28) may therefore be rewritten in the form

$$\langle \bar{R}^{(j)}_{x,y}\rangle_0 = \theta(xy)Q^{(j)}(x-y) + i^2\theta(xy) \times \int \Delta_B(u-v)\{\langle \bar{R}^{(j)}_{x,u}\rangle_0\langle \bar{R}^{(0)}_{y,v}\rangle_0 + \langle \bar{R}^{(0)}_{x,u}\rangle_0\langle \bar{R}^{(j)}_{y,v}\rangle_0\}. \quad (29)$$

In momentum space, with $s = -p^2$, $p_0 > 0$, Eq. (29) corresponds to an Omnès equation[9];

$$\bar{r}^{(j)}(s) = \bar{Q}^{(j)}(s) + \int_{4m^2} \frac{ds'\,\Pi(s')}{s'-s-i\epsilon} \times \{\bar{r}^{(j)}(s')\bar{r}^{(0)*}(s') + \bar{r}^{(j)*}(s')\bar{r}^{(0)}(s')\},$$

$$\bar{Q}^{(j)}(s) = \int_{4m^2} \frac{ds'\,\Pi^{(j)}(s')}{s'-s-i\epsilon},$$

with solution

$$\bar{r}^{(j)}(s) = e^{u(s)} \int_{4m^2} \frac{ds'}{s'-s-i\epsilon} \times \exp[-\Delta(s')]\Pi^{(j)}(s'), \quad (30)$$

$$u(s) = \frac{1}{\pi}\int_{4m^2} \frac{ds'\,\delta(s')}{s'-s-i\epsilon} = \Delta(s) + i\delta(s),$$

where, with the aid of (19),

$$\exp[i\delta(s)] = 1 + 2\pi i\Pi(s)\bar{r}^{(0)}(s),$$

and

$$\delta(s) = \tan^{-1}\left[\frac{2\pi\Pi(s)\ \mathrm{Re}\,\bar{r}^{(0)}(s)}{1-2\pi\Pi(s)\ \mathrm{Im}\,\bar{r}^{(0)}(s)}\right], \qquad \delta(s) = 0, \quad s < 4m^2.$$

An arbitrary, additive solution to the homogeneous problem has been omitted.

For the simplest case, $j = 1$, $Q^{(1)}$ may be read off from the integral unitarity form of (14),

$$\theta(xy)Q^{(1)}(x-y) = -\tfrac{1}{2}\theta(xy) \times \int [\Delta'^{(+)}(u_1-v_1)\Delta'^{(+)}(u_2-v_2) - \Delta'^{(-)}\Delta'^{(-)}] \times \langle \bar{R}^{(0)}_{x,u_1u_2}\rangle_0\langle \bar{R}^{(0)}_{y,v_1v_2}\rangle_0 + \Delta_R(x-y) - \langle \bar{R}^{(0)}_{x,y}\rangle_0. \quad (31)$$

With the aid of (24), the integral of (31) may be written in the form

$$-\tfrac{1}{2}g^2\theta(xy)\int \langle \bar{R}^{(0)}_{x,u}\rangle_0[\bar{\Delta}^{(+)}(u-v)\bar{\Delta}^{(+)}(u-v) - \bar{\Delta}^{(-)}\bar{\Delta}^{(-)}]\langle \bar{R}^{(0)}_{y,v}\rangle_0,$$

where

$$\bar{\Delta}^{(\pm)}(x-y) \equiv \int \langle \bar{R}^{(0)}_{x,u}\rangle_0\Delta'^{(\pm)}(u-v)\langle \bar{R}^{(0)}_{y,v}\rangle_0. \quad (32)$$

In momentum space, (31) then becomes

$$\bar{Q}^{(1)}(s) = [m^2 - s - i\epsilon]^{-1} - \bar{r}^{(0)}(s) + \int_{4m^2} \frac{ds'\,q^{(1)}(s')}{s'-s-i\epsilon}, \quad (33)$$

[9] R. Omnès, Nuovo Cimento **8**, 316 (1958).

where

$$q^{(1)}(s) = \tfrac{1}{2}g^2\,|\bar{r}^{(0)}(s)|^2 \int \frac{ds_1}{\pi}\,\mathrm{Im}\,\bar{r}^{(0)}(s_1) \times \int \frac{ds_2}{\pi}\,\mathrm{Im}\,\bar{r}^{(0)}(s_2)\cdot\Sigma^{(2)}(s, s_1s_2), \qquad (34)$$

and $\Sigma^{(2)}$ is the two-particle phase space integral defined by

$$\theta(p)\Sigma^{(2)}(-p^2, s_1s_2) = (2\pi)^{-3}\int dk_1\,dk_2\;\delta(k_1^2 + s_1)\delta(k_2^2 + s_2) \times \theta(k_1)\theta(k_2)\delta(k_1 + k_2 - p),$$

$$\Sigma^{(2)}(s, s_1s_2) = \frac{1}{16\pi^2}\,\theta(s + s_1 - s_2) \times \theta(s + s_2 - s_1)\theta(s^2 - 2s[s_1 + s_2] + [s_1 - s_2]^2)\left[1 - 2\left(\frac{s_1 + s_2}{s}\right) + \left(\frac{s_1 - s_2}{s}\right)^2\right]^{\frac{1}{2}}.$$

It is clear that if $\Pi(s)$ is set equal to zero, (33), (34), and (30) just reproduce the renormalized bubble contribution of order g^2. It is not difficult to see that $q^{(1)}(s)$ vanishes for $s < 4m^2$; this is true because the lowest mass of $\mathrm{Im}\,\bar{r}^{(0)}(s)$ lies at m^2 and not because $\Pi(s)$ has been chosen to vanish for $s < 4m^2$. To this order of approximation, the retarded propagator is then given by

$$r(s) \sim \bar{r}^{(0)}(s) + \exp\,[u(s)]\int_{4m^2}^{\infty}\frac{ds'\,e^{-\Delta(s')}}{s' - s - i\epsilon} \times [q^{(1)}(s') - \Pi(s')\,|\bar{r}^{(0)}(s')|^2]. \qquad (35)$$

B. The 3-Point Function

The vertex function of index one is the first amplitude having a nontrivial structure for its S-amputated part. From (14), one sees that the function satisfies the unitarity equation

$$\langle\bar{R}^{(1)}_{x,yz}\rangle_0 - \langle\bar{R}^{(1)}_{y,xz}\rangle_0 = i^2\int\Delta'(u - v)\{\langle\bar{R}^{(1)}_{x,z\mathrm{u}}\rangle_0\langle\bar{R}^{(0)}_{y,\mathrm{v}}\rangle_0 + \langle\bar{R}^{(0)}_{x,\mathrm{u}}\rangle_0\langle\bar{R}^{(1)}_{y,z\mathrm{v}}\rangle_0\} + i^2\int\Delta_B(u - v)\{\langle\bar{R}^{(1)}_{x,\mathrm{u}}\rangle_0\langle\bar{R}^{(0)}_{y,z\mathrm{v}}\rangle_0 + \langle\bar{R}^{(0)}_{x,z\mathrm{u}}\rangle_0\langle\bar{R}^{(1)}_{y,\mathrm{v}}\rangle_0\} - i^2\int\Delta_B(u - v)\{\langle\bar{R}^{(0)}_{x,\mathrm{u}}\rangle_0\langle\bar{R}^{(0)}_{y,z\mathrm{v}}\rangle_0 + \langle\bar{R}^{(0)}_{x,z\mathrm{u}}\rangle_0\langle\bar{R}^{(0)}_{y,\mathrm{v}}\rangle_0\} - \frac{i}{2}\int[\Delta'^{(+)}(u_1 - v_1)\Delta'^{(+)}(u_2 - v_2) - \Delta'^{(-)}\Delta'^{(-)}] \times\{\langle\bar{R}^{(0)}_{x,z\mathrm{u}_1\mathrm{u}_2}\rangle_0\langle\bar{R}^{(0)}_{y,\mathrm{v}_1\mathrm{v}_2}\rangle_0 + \langle\bar{R}^{(0)}_{x,\mathrm{u}_1\mathrm{u}_2}\rangle_0\langle\bar{R}^{(0)}_{y,z\mathrm{v}_1\mathrm{v}_2}\rangle_0\}. \qquad (36)$$

With the aid of (24), (27), and (32), the last line of (36) may be written in the form

$$-\frac{i}{2}g^3\int\langle\bar{R}^{(0)}_{x,x'}\rangle_0\langle\bar{R}^{(0)}_{z',z}\rangle_0\langle\bar{R}^{(0)}_{y,v}\rangle_0 \times[\bar{\Delta}^{(+)}(u - v)\bar{\Delta}^{(+)}(u - v) - \bar{\Delta}^{(-)}\bar{\Delta}^{(-)}] \times\langle\bar{R}^{(0)}_{x',u}\rangle_0\cdot\delta(x' - z') - (x\leftrightarrow y) - ig^3\int\langle\bar{R}^{(0)}_{x,x'}\rangle_0\langle\bar{R}^{(0)}_{z',z}\rangle_0\langle\bar{R}^{(0)}_{y,v}\rangle_0 \times[\bar{\Delta}^{(+)}(x' - v)\bar{\Delta}^{(+)}(z' - v) - \bar{\Delta}^{(-)}\bar{\Delta}^{(-)}] \times\langle\bar{R}^{(0)}_{x',z'}\rangle_0 - (x\leftrightarrow y), \qquad (37)$$

and, using the unitarity relations of the previous 2-point functions, it is easy to see that the first line of (37) may be replaced by the combination

$$g\int\langle\bar{R}^{(0)}_{x,x'}\rangle_0\langle\bar{R}^{(0)}_{z',z}\rangle_0\delta(x' - z')\{\langle\bar{R}^{(1)}_{x',v}\rangle_0 - \langle\bar{R}^{(1)}_{y,x'}\rangle_0 + i^2\int\Delta_B(u - v)\langle\bar{R}^{(0)}_{x',\mathrm{u}}\rangle_0\langle\bar{R}^{(0)}_{y,\mathrm{v}}\rangle_0 - i^2\int\Delta_B(u - v)[\langle\bar{R}^{(1)}_{x',\mathrm{u}}\rangle_0\langle\bar{R}^{(0)}_{y,\mathrm{v}}\rangle_0 + \langle\bar{R}^{(0)}_{x',\mathrm{u}}\rangle_0\langle\bar{R}^{(1)}_{y,\mathrm{v}}\rangle_0]\} - (x\leftrightarrow y). \qquad (38)$$

When the terms of Eq. (38) are combined with the second and third lines of (36), one obtains

$$g\int\langle\bar{R}^{(0)}_{x,x'}\rangle_0\langle\bar{R}^{(0)}_{z',z}\rangle_0\delta(x' - z')\{\langle\bar{R}^{(1)}_{x',v}\rangle_0 - \langle\bar{R}^{(1)}_{y,x'}\rangle_0 + \int\Delta_B(u - v)\langle\bar{R}^{(1)}_{x',\mathrm{u}}\rangle_0\langle\bar{R}^{(0)}_{y,\mathrm{v}}\rangle_0\} - (x\leftrightarrow y). \qquad (39)$$

The right-hand side of the unitarity condition for this vertex function is then given by (39), the second term of (37), and the first term of (36). The significance of the combination appearing in (39) lies in the presence of the $\delta(x' - z')$ factor, which indicates that these terms (which would be called contact terms in the ordinary perturbation theory) contribute a self-energy part to the legs of the *S-amputated* vertex. This presupposes the use of the zero-index propagator for the external legs of the entire function, which is effectively required by the form of the first line of (36). The appearance of such terms in the legs of all higher-index amplitudes is just a reflection of the fact that the zero-index propagator does not, in general, give an adequate description of the legs of every n-point function, but its use is nevertheless forced by the formalism's close relation to the conventional perturbation theory. The terms appearing in the curly bracket of (39) are sufficiently complicated (they depend upon y rather than y') to rule out guessing the

corresponding part of the vertex from the form of the ordinary perturbation terms, and in general the best one can do is to rely upon a different calculational procedure (e.g., based upon the equations of Sec. VI). However, there is still the possibility of choosing the spectral function $\Pi(s)$ such that this and all the n-point functions of index one may be obtained directly from their unitarity relations and a knowledge of the corresponding perturbation terms; that is, it is possible to require the terms of (39) to vanish, which will then lead to a condition on $\Pi(s)$ and which will, incidentally, permit the calculational method of this section to go through.

If the curly bracket of (39) is set equal to zero, in momentum space one obtains the condition

$$\bar{r}^{(1)}(s) - \bar{r}^{(1)*}(s) = 2\pi i \cdot \Pi(s)\bar{r}^{(1)}(s)\bar{r}^{(0)*}(s) \tag{40}$$

or

$$\bar{r}^{(1)*}(s) = \exp\,[-i\delta(s)]\bar{r}^{(1)}(s). \tag{41}$$

Equation (41) implies that

$$\bar{r}^{(1)}(s) = \xi \exp\,[\tfrac{1}{2}u(s)], \tag{42a}$$

where ξ is a real constant; this was the reason for omitting a solution to the homogeneous problem, since the latter would have had the phase $\delta(s)$ rather than $\frac{1}{2}\,\delta(s)$ [which, incidentally, is the phase of $\bar{r}^{(0)}(s)$]. Equations (41) and (42a) and the unitarity form of (29) lead to the further restrictions

$$\Pi^{(1)}(s) + \tan\frac{\delta(s)}{2}\cdot\exp\,[\Delta(s)]\,\frac{P}{\pi} \times \int_{4m^2}^{\infty}\frac{ds'}{s'-s}\exp\,[-\Delta(s')]\Pi^{(1)}(s') = 0, \tag{42b}$$

$$\Pi^{(1)}(s) + \frac{\xi}{\pi}\exp\,[\tfrac{1}{2}\Delta(s)]\cdot\sin\frac{\delta(s)}{2} = 0, \tag{42c}$$

which are implicit statements about the spectral function $\Pi(s)$. Certainly, the simplest (if not the only) solution to Eqs. (42a, b, c) is obtained by choosing $\Pi^{(1)}(s) = \xi = 0$; in this case $\langle\bar{R}^{(1)}_{x,y}\rangle_0 = 0$, and one obtains from the absorptive part of (33), or from (35), the relation

$$\Pi(s) = \frac{g^2}{2}\int\frac{ds_1}{\pi}\,\mathrm{Im}\,\bar{r}^{(0)}(s_1) \cdot\int\frac{ds_2}{\pi}\,\mathrm{Im}\,\bar{r}^{(0)}(s_2)\cdot\Sigma^{(2)}(s, s_1 s_2), \tag{43}$$

which, with (21), represents a nonlinear integral equation for this spectral function. Equation (43) may be rewritten in the form

$$\Pi(s) = \frac{g^2 s^2}{32\pi^4}\int_0^1 d\xi\int_0^{\frac{1}{2}(1+\xi^2)} d\eta\cdot\mathrm{Im}\,\bar{r}^{(0)}\left[\frac{s}{2}(\eta+\xi)\right] \times \mathrm{Im}\,\bar{r}^{(0)}\left[\frac{s}{2}(\eta-\xi)\right]\cdot(1+\xi^2-2\eta)^{\frac{1}{2}}, \tag{44}$$

and it is clear that the crudest approximation to the right-hand side of (43) or (44) leads to the bubble contribution of (20). Together, Eqs. (43) and (21) correspond to introducing into the two-particle phase space spectral function, the dressed propagator lines defined in terms of the same function.

To summarize this argument: the original, arbitrary spectral function $\Pi(s)$ may be chosen such that the S-amputated vertex of index one contains no self-energy legs; the choice represented by (43) is the simplest, and corresponds to setting the propagator of index one equal to zero. In effect, this forces the zero-index propagator to be correct to this order of approximation, and guarantees that the prescription of using unitarity and a comparison with the ordinary perturbation forms can be carried through.

Defining the S-amputated vertex of index one by the relation

$$\langle\bar{R}^{(1)}_{x,yz}\rangle_0 = \int\langle\bar{R}^{(0)}_{x,x'}\rangle_0\langle\bar{R}^{(0)}_{y',y}\rangle_0\langle\bar{R}^{(0)}_{z',z}\rangle_0\langle\hat{R}^{(1)}_{\mathbf{x}',\mathbf{y}'\mathbf{z}'}\rangle_0,$$

with $\langle\bar{R}^{(1)}_{x,y}\rangle_0 = 0$, and with the aid of (15), one obtains from (36) and (37), the unitarity condition

$$\int\langle\bar{R}^{(0)}_{x,x'}\rangle_0\langle\bar{R}^{(0)}_{y,y'}\rangle_0\langle\bar{R}^{(0)}_{z',z}\rangle_0[\langle\hat{R}^{(1)}_{\mathbf{x}',\mathbf{y}'\mathbf{z}'}\rangle_0 - \langle\hat{R}^{(1)}_{\mathbf{y}',\mathbf{x}'\mathbf{z}'}\rangle_0] = -ig^3\int\langle\bar{R}^{(0)}_{x,x'}\rangle_0\langle\bar{R}^{(0)}_{y,y'}\rangle_0\langle\bar{R}^{(0)}_{z',z}\rangle_0 \times[\bar{\Delta}^{(+)}(x'-y')\bar{\Delta}^{(+)}(z'-y') - \bar{\Delta}^{(-)}\bar{\Delta}^{(-)}]\langle\bar{R}^{(0)}_{z',z'}\rangle_0 - (x\leftrightarrow y),$$

or

$$\langle\hat{R}^{(1)}_{\mathbf{x},\mathbf{yz}}\rangle_0 - \langle\hat{R}^{(1)}_{\mathbf{y},\mathbf{xz}}\rangle_0 = -ig^3\langle\bar{R}^{(0)}_{x,z}\rangle_0[\bar{\Delta}^{(+)}(x-y)\bar{\Delta}^{(+)}(z-y) - \bar{\Delta}^{(-)}\bar{\Delta}^{(-)}] - (x\leftrightarrow y). \tag{45}$$

This resembles the corresponding perturbation equation except that the latter's self-energy legs (due to simple bubbles which were necessarily present because K–G amputation was used) are missing, and the $\bar{\Delta}^{(\pm)}$ functions replace the simple $\Delta^{(\pm)}$ functions. On the basis of the perturbation vertex of index one, one may write

$$\langle\hat{R}^{(1)}_{\mathbf{x},\mathbf{yz}}\rangle_0 = -ig^3\theta(xy) \times\{\langle\bar{R}^{(0)}_{x,z}\rangle_0[\bar{\Delta}^{(+)}(x-y)\bar{\Delta}^{(+)}(z-y) - \bar{\Delta}^{(-)}\bar{\Delta}^{(-)}] - (x\leftrightarrow y)\}, \tag{46}$$

or

$$\langle \bar{R}^{(1)}_{x,yz}\rangle_0 = -ig^3\{\langle \bar{R}^{(0)}_{x,z}\rangle_0\theta(zy) \times [\bar{\Delta}^{(+)}(x-y)\bar{\Delta}^{(+)}(z-y) - \bar{\Delta}^{(-)}\bar{\Delta}^{(-)}] + \langle \bar{R}^{(0)}_{x,y}\rangle_0\theta(yz) \times [\bar{\Delta}^{(+)}(x-z)\bar{\Delta}^{(+)}(y-z) - \bar{\Delta}^{(-)}\bar{\Delta}^{(-)}]\}, \quad (47)$$

representing a relativistic, retarded solution to the unitarity equation (45), which is symmetric in its retarded coordinates. The property of symmetry, made explicit by the equality of (46) and (47), follows from (32) and the statement

$$\langle \bar{R}^{(0)}_{x,y}\rangle_0 = i^2\theta(xy)[\bar{\Delta}^{(+)}(x-y) + \bar{\Delta}^{(-)}] = i^2\theta(xy)\bar{\Delta}(x-y).$$

As in I, a pure contact term, $Z_{(1)}\,\delta(x-y)\,\delta(y-z)$, may be added to the right-hand side of (46) or (47), with the finite constant $Z_{(1)}$ chosen such that the Fourier transform of these equations vanishes on the mass shell of all invariants; the constant g is then renormalized to this approximation. The form of the S-amputated vertex of index one constructed in this manner corresponds to the simple triangle diagram of perturbation theory, with all internal $\Delta^{(*)}$ lines replaced by $\bar{\Delta}^{(*)}$ lines.

C. The 4-Point Function

With the propagator of index one again set equal to zero, all self-energy legs will disappear from the amputated scattering amplitude. In addition, there will occur a cancellation of spurious internal self-energy structure corresponding to the bubble term of the fourth-order perturbation graph; such a term has already been included in the zero-index 4-point function, [Eq. (27)] and the necessary cancellation is neatly provided by the formalism.

These cancellations may easily be seen without explicitly calculating the entire amplitude; the unitarity condition here is

$$\begin{aligned}\langle \bar{R}^{(1)}_{x,y_1y_2y_3}\rangle_0 &- \langle \bar{R}^{(1)}_{y_1,xy_2y_3}\rangle_0 \\ = i\Big\{&\Big[\langle \bar{R}^{(0)}_{x,y_2y_3}\rangle, \frac{(D+D_B)^2}{2!} - D_B, \langle \bar{R}^{(0)}_{y_1}\rangle\Big] \\ &+ \Big[\langle \bar{R}^{(0)}_{x}\rangle, \frac{(D+D_B)^2}{2!} - D_B, \langle \bar{R}^{(0)}_{y_1,y_2y_3}\rangle\Big] \\ &+ \Big[\langle \bar{R}^{(0)}_{x,y_2}\rangle, \frac{(D+D_B)^2}{2!} - D_B, \langle \bar{R}^{(0)}_{y_1,y_3}\rangle\Big] \\ &+ \Big[\langle \bar{R}^{(0)}_{x,y_3}\rangle, \frac{(D+D_B)^2}{2!} - D_B, \langle \bar{R}^{(0)}_{y_1,y_2}\rangle\Big] \\ &+ [\langle \bar{R}^{(1)}_{x,y_2}\rangle, D+D_B, \langle \bar{R}^{(0)}_{y_1,y_3}\rangle] \\ &+ [\langle \bar{R}^{(1)}_{x,y_3}\rangle, D+D_B, \langle \bar{R}^{(0)}_{y_1,y_2}\rangle] \\ &+ [\langle \bar{R}^{(0)}_{x,y_2}\rangle, D+D_B, \langle \bar{R}^{(1)}_{y_1,y_3}\rangle] \\ &+ [\langle \bar{R}^{(0)}_{x,y_3}\rangle, D+D_B, \langle \bar{R}^{(1)}_{y_1,y_2}\rangle] \\ &+ [\langle \bar{R}^{(1)}_{x,y_2y_3}\rangle, D+D_B, \langle \bar{R}^{(0)}_{y_1}\rangle] \\ &+ [\langle \bar{R}^{(0)}_{x}\rangle, D+D_B, \langle \bar{R}^{(1)}_{y_1,y_2y_3}\rangle]\Big\}_{j\to 0}, \end{aligned} \quad (48)$$

and the removal of the unwanted self-energy terms occurs in the first through the fourth lines of (48). Since the S-amputated vertex of index one has no self-energy legs, these can only occur in the first two lines of (48), appearing there because of the structure of the S-amputated zero-index 5-point function. The significant terms of the latter will be those three (out of a total of 15) which have both internal coordinates equal [factors proportional to $\delta(u_1-u_2)$]; when substituted into the first two lines of (48) these terms produce three distinct self-energy legs which are exactly cancelled by the counterterms

$$-i\{[\langle \bar{R}^{(0)}_{x,y_2y_3}\rangle, D_B, \langle \bar{R}^{(0)}_{y_1}\rangle] + [\langle \bar{R}^{(0)}_{x}\rangle, D_B, \langle \bar{R}^{(0)}_{y_1,y_2y_3}\rangle]\} \quad (49)$$

of the first two lines of (48). The three distinct terms of (49) correspond to the three permutations of (27).

The cancellation obtained in the third and fourth lines of (48) corresponds to removing that part of the fourth-order scattering amplitude in which a simple bubble is inserted into the single internal line representing a one-meson exchange. Using (24), (27), and (43), the arithmetic of the cancellation is easily seen.

Because of the very close similarity of (48) to the corresponding integral unitarity equation in the perturbation theory, the S-amputated 4-point function of index one can be read off directly from the K–G amplitude of index one, by dropping all internal and external self-energy parts of the latter (simple bubbles) and replacing each $\Delta^{(*)}$ line by a corresponding $\bar{\Delta}^{(*)}$ line. It is intuitively clear that the same prescription is valid for every n-point function of index one.

The functionals of indices zero and one are thus completely determined, and by themselves may be viewed as an "axiomatic model" of a simple field theory. How realistic such a model may be depends on the solutions to (43), and is an entirely different question. It should be noted that the scattering amplitude here will have nontrivial dependence on the invariant variables s, t, and u, while possessing crossing symmetry.

V. ZERO-VERTEX THEORY

In this section, the partial sum formalism is outlined for the case of a self-interacting neutral

pseudoscalar boson field, and the derivation of the simplest n-point functions will be sketched briefly. It will be convenient to change the value of the integer β of Eq. (11) to $\beta = 3$ (and $M^2 = 9m^2$), thereby avoiding the introduction of spurious complications. One then considers the λ-expansion of

$$\langle R^{\lambda}_{x,y}\rangle - \langle R^{\lambda}_{y,x}\rangle = \frac{i}{\lambda}[\langle R^{\lambda}_{x}\rangle, \exp[\lambda(D + D_B) - \lambda^3 D_B], \langle R^{\lambda}_{y}\rangle], \quad (50)$$

where, as in the scalar field case, all the 2-point unitarity statements derived from (50) are understood to hold in the stronger sense of integral unitarity.

A. Functions of Index Zero

The propagator of index zero is again given by Eqs. (15), (18), and (21), with $\Pi(s)$ as yet unspecified. The zero-index vertex is taken to vanish identically, and the zero-index 4-point function then satisfies the unitarity condition (25), where the vertex contributions to the latter are missing. Defining the S-amputated 4-point amplitude in terms of zero-index propagator legs, one obtains the simple unitarity statement

$$\langle \hat{R}^{(0)}_{\mathbf{x},y_1,y_2,y_3}\rangle_0 - \langle \hat{R}^{(0)}_{y_1,\mathbf{x},y_2,y_3}\rangle_0 = 0,$$

with solution[8]

$$\langle \hat{R}^{(0)}_{\mathbf{x},y_1,y_2,y_3}\rangle_0 = G\delta(x - y_1)\delta(y_1 - y_2)\delta(y_2 - y_3), \quad (51)$$

where G is a dimensionless number denoting, as in I, the renormalized coupling constant. Equation (51) resembles the corresponding perturbation term in exactly the same manner as (24) resembles the lowest-order K–G amputated perturbation vertex. It is clear that the entire zero-index functional is composed of n-point functions ($n > 2$) for which

(i) all odd-n-point functions vanish;

(ii) all even-n-point functions consist of zero-index propagator legs adjoined to S-amputated amplitudes resembling the K–G amputated Born functions of perturbation theory, except that each $\Delta_R(x - y)$ factor of the latter is replaced by a corresponding $\langle \bar{R}^{(0)}_{x,y}\rangle_0$ factor.

B. Functions of Index One

The propagator of index one satisfies the integral unitarity equation

$$\langle \bar{R}^{(1)}_{x,y}\rangle_0 = i\theta(xy)\{[\langle \bar{R}^{(1)}_{x}\rangle, D + D_B, \langle \bar{R}^{(0)}_{y}\rangle] + [\langle \bar{R}^{(0)}_{x}\rangle, D + D_B, \langle \bar{R}^{(1)}_{y}\rangle]\}_0, \quad (52)$$

which resembles the form of (29) except that the inhomogeneous $Q^{(j)}$ term is absent. The proper solution to (52) is therefore $\langle \bar{R}^{(1)}_{x,y}\rangle_0 = 0$; this is not surprising since the corresponding perturbation function also vanishes.

The vertex function vanishes, and the 4-point amplitude then satisfies the unitarity equation

$$\begin{aligned}\langle \bar{R}^{(1)}_{x,y_1y_2y_3}\rangle_0 - \langle \bar{R}^{(1)}_{y_1,xy_2y_3}\rangle_0 &= i\Big\{\Big[[\langle \bar{R}^{(0)}_{x,y_2}\rangle, \frac{(D + D_B)^2}{2!}, \langle \bar{R}^{(0)}_{y_1,y_3}\rangle\Big] \\ &+ \Big[\langle \bar{R}^{(0)}_{x,y_3}\rangle, \frac{(D + D_B)^2}{2!}, \langle \bar{R}^{(0)}_{y_1,y_2}\rangle\Big] \\ &+ [\langle \bar{R}^{(1)}_{x,y_2y_3}\rangle, D + D_B, \langle \bar{R}^{(0)}_{y_1}\rangle] \\ &+ [\langle \bar{R}^{(0)}_{x}\rangle, D + D_B, \langle \bar{R}^{(1)}_{y_1,y_2y_3}\rangle]\Big\}_0. \end{aligned} \quad (53)$$

Using Eq. (51) and the zero-index propagator for the legs, one finds the unitarity condition for the S-amputated amplitude:

$$\begin{aligned}\langle \hat{R}^{(1)}_{\mathbf{x},y_1y_2y_3}\rangle_0 - \langle \hat{R}^{(1)}_{y_1,\mathbf{x}y_2y_3}\rangle_0 &= -\frac{i}{2}G^2\delta(x - y_2) \\ &\times [\bar{\Delta}^{(+)}(y_2 - y_3)\bar{\Delta}^{(+)}(y_2 - y_3) \\ &- \bar{\Delta}^{(-)}\bar{\Delta}^{(-)}]\delta(y_3 - y_1) + (y_2 \leftrightarrow y_3). \end{aligned} \quad (54)$$

There are no self-energy legs in (54), and there are none in the corresponding perturbation amplitude.[10] On the basis of the latter, one may write the relativistic, retarded, crossing symmetric solution to (54):

$$\begin{aligned}\langle \hat{R}^{(1)}_{\mathbf{x},y_1y_2y_3}\rangle_0 &= -\frac{i}{2}G^2\sum_{p=1}^{3}\delta(x - y_1)\theta(y_1y_2) \\ &\times [\bar{\Delta}^{(+)}(y_1 - y_2)\bar{\Delta}^{(+)}(y_1 - y_2) - \bar{\Delta}^{(-)}\bar{\Delta}^{(-)}]\delta(y_2 - y_3) \\ &+ Z'_{(1)}\delta(x - y_1)\delta(y_1 - y_2)\delta(y_2 - y_3), \end{aligned} \quad (55)$$

where $Z'_{(1)}$ is a renormalization (or subtraction) constant multiplying a pure connected contact term permitted by (54). In the ordinary perturbation theory, $Z'_{(1)}$ is logarithmically divergent, corresponding to the single subtraction which must be made in that amplitude (and which then permits the identification of G with the renormalized coupling constant to order G^2). If the S-amputated momentum-space amplitude is defined by

$$\hat{\Gamma}^{(1)}(q; p_1p_2p_3) = (2\pi)^{-4}\int dx\, dy_1\, dy_2\, dy_3 \times \exp[i(q\cdot x + p_1\cdot y_1 + p_2\cdot y_2 + p_3\cdot y_3)]\langle \hat{R}^{(1)}_{\mathbf{x},y_1y_2y_3}\rangle_0,$$

[10] This is, in fact, the reason for choosing $\beta = 3$, since the choice $\beta = 2$ would have introduced spurious self-energy parts which are not even present in the ordinary perturbation terms.

the Fourier transform of (55) leads to

$$\hat{r}^{(1)}(q; p_1p_2p_3) = \frac{G^2}{2}\,\delta(q + p_1 + p_2 + p_3)$$
$$\times \sum_{p=1}^{3} \int_{4m^2}^{\infty} d\kappa^2[\kappa^2 + (p_2 + p_3)^2 - i\epsilon(p_2 + p_3)]^{-1}$$
$$\times \int \frac{ds_1}{\pi}\,\mathrm{Im}\,\bar{r}^{(0)}(s_1)\cdot\int \frac{ds_2}{\pi}\,\mathrm{Im}\,\bar{r}^{(0)}(s_2)\cdot\Sigma^{(2)}(\kappa^2, s_1s_2)$$
$$+ Z'_{(1)}\,\delta(q + p_1 + p_2 + p_3), \tag{56}$$

where $\sum_{p=1}^{3}$ now refers to the permutation of the three momenta $p_{1,2,3}$ in the s, t, and u combinations $p_2 + p_3$, $p_1 + p_2$, $p_1 + p_3$. Since (56) contains the G^2 perturbation contribution, unless some remarkable cancellations occur, $Z'_{(1)}$ will again be logarithmically divergent (but no worse); as in the perturbation case, this will accomplish the renormalization of G to this "order," and the constant $Z'_{(1)}$ will disappear from the final, finite result.

All the n-point functions of index one may be read off from their perturbation counterparts following the rules (i) of A and the slight modification of (ii),

(ii′) all even-n-point functions consist of zero-index propagator legs adjoined to S amplitudes resembling the K–G amputated index one functions except that each $\Delta^{(*)}$ factor of the latter is replaced by a corresponding $\bar{\Delta}^{(*)}$ factor.

C. Functions of Index Two

Expansion of the integral unitarity form of (50) yields the equation defining the propagator of index two:

$$\langle\bar{R}^{(2)}_{x,y}\rangle_0 = i\theta(xy)\Big\{\Big[\langle\bar{R}^{(0)}_x\rangle, \frac{(D + D_B)^3}{3!} - D_B, \langle\bar{R}^{(0)}_y\rangle\Big]$$
$$+ [\langle\bar{R}^{(2)}_x\rangle, D + D_B, \langle\bar{R}^{(0)}_y\rangle]$$
$$+ [\langle\bar{R}^{(0)}_x\rangle, D + D_B, \langle\bar{R}^{(2)}_y\rangle]\Big\}_0, \tag{57}$$

where the index-one propagator and the vertex functions of indices zero and one have been set equal to zero. The solution to (57) will be of the same general form as (30). In momentum space this reads

$$\bar{r}^{(2)}(s) = e^{u(s)}\int_{4m^2}^{\infty} \frac{ds'\,e^{-\Delta(2')}}{s' - s - i\epsilon}\,\Pi^{(2)}(s'),$$

where $\Pi^{(2)}(s)$ denotes the absorptive part of the Fourier transform of the inhomogeneous terms of (57). Anticipating the result that the vanishing of $\langle\bar{R}^{(2)}_{x,y}\rangle_0$ will again remove all self-energy legs from the remaining S-amputated functions of index two, this condition then provides an equation for the spectral function $\Pi(s)$. Since

$$\Pi^{(2)}(s) = q^{(2)}(s) - \Pi(s)\,|\bar{r}^{(0)}(s)|^2,$$

where

$$q^{(2)}(s) = \frac{G^2}{3!}\,|\bar{r}^{(0)}(s)|^2$$
$$\times \int \frac{ds_1}{\pi}\,\mathrm{Im}\,\bar{r}^{(0)}(s_1)\cdot\int \frac{ds_2}{\pi}\,\mathrm{Im}\,\bar{r}^{(0)}(s_2)$$
$$\times \int \frac{ds_3}{\pi}\,\mathrm{Im}\,\bar{r}^{(0)}(s_3)\cdot\Sigma^{(3)}(s, s_1s_2s_3), \tag{58}$$

and $\Sigma^{(3)}$ is the three-particle phase space integral defined by

$$\theta(p)\Sigma^{(3)}(-p^2, s_1s_2s_3) = (2\pi)^{-6}\int dk_1\,dk_2\,dk_3$$
$$\times\,\delta(k_1^2 + s_1)\,\delta(k_2^2 + s_2)\,\delta(k_3^2 + s_3)$$
$$\times\,\theta(k_1)\,\theta(k_2)\,\theta(k_3)\,\delta(k_1 + k_2 + k_3 - p),$$

if $\Pi^{(2)}(s) = 0$ a comparison of (58) and (19) yields the nonlinear equation

$$\Pi(s) = \frac{G^2}{3!}\int \frac{ds_1}{\pi}\,\mathrm{Im}\,\bar{r}^{(0)}(s_1)\cdot\int \frac{ds_2}{\pi}\,\mathrm{Im}\,\bar{r}^{(0)}(s_2)$$
$$\times \int \frac{ds_3}{\pi}\,\mathrm{Im}\,\bar{r}^{(0)}(s_3)\cdot\Sigma^{(3)}(s, s_1s_2s_3),$$

which has the same form and interpretation as (43).

Aside from questions of physical usefulness, as indicated previously, this is a useful procedure for it now permits all the remaining n-point functions of index two to be read off, in modified form, directly from the corresponding perturbation amplitudes. The prescription is defined by the rules (i) of A and the slight modification of (ii′) of B,

(ii″) all even-n-point functions consist of zero-index propagator legs adjoined to S-amplitudes resembling the K–G index-two functions of perturbation theory, except that each $\Delta^{(*)}$ factor of the latter is replaced by a corresponding $\bar{\Delta}^{(*)}$ factor, and the self-energy legs appearing in the K–G amputated functions are to be discarded.

Taken by themselves, the functionals of indices zero, one, and two, may be viewed as defining an "axiomatic model" of a pseudoscalar field theory.

VI. STRONG PARTIAL SUM FORMALISM

It is not clear whether the procedure of calculating S-amputated functions on the basis of the corresponding perturbation theory K–G amputated functions can be continued for general amplitudes of higher index. It is clear that these S-amputated

functions will themselves contain new self-energy legs corresponding to nonzero higher-index propagators, unless it is also possible to introduce, in a consistent way, new spectral functions $\Pi(s)$ (each, e.g., corresponding to the onset of a new cut) such that only the zero-index propagator does not vanish. In lieu of answers to these questions, it is desirable to have, at least in principle, equations for calculating the higher-index functions, and these are provided by the formalism in the stronger sense of integral unitarity. It should be emphasized that the remarks to follow are based upon a plausibility argument of relativistic invariance similar to that of Sec. II; it is therefore possible that the discussion of this section it is in part incorrect, although this does not appear to be likely.

The relativistic invariance of the λ-expansion defining the perturbation amplitudes of I is based upon the spacelike vanishing of the commutator of (5) for arbitrary, real scaling parameters $\lambda_{1,2}$. The generalization to the partial sum formalism in its stronger sense is suggested by considering those additional transformations on A_{IN} which leave unchanged the spacelike commutivity property; that such transformations will, in general, also change the content of the unitarity condition is not immediately pertinent to this discussion, since such unwanted modifications will be removed later on. In particular, if there exists an operator B_{IN} with the properties

(i) $[A_{IN}(x), B_{IN}(y)] \equiv 0$,

(ii) $[B_{IN}(x), B_{IN}(y)] = c$ number vanishing spacelike,

then it is reasonable to conjecture that the operator $R_x\{A_{IN} + B_{IN}, j\}$, obtained from $R_x\{A_{IN}, j\}$ by the replacement

$$A_{IN}(x) \rightarrow A_{IN}(x) + B_{IN}(x),$$

will have a commutator which vanishes spacelike:

$$[R_x\{A_{IN} + B_{IN}, j\}, R_y\{A_{IN} + B_{IN}, j\}] = 0 \quad \text{spacelike.} \tag{59}$$

This can be made plausible by an examination of the simplest terms which enter into the commutator of (59), but it would be most desirable to have a proof of this property.

An operator B_{IN} fulfilling requirements (i) and (ii) may be constructed by enlarging the Hilbert space to include "particles" or, in general, continuous mass distributions which do not contain the physical mass m of the field A_{IN}. The commutator of two B_{IN} fields will then have the representation

$$[B_{IN}(x), B_{IN}(y)] = i\Delta_B(x - y) = i\int_{M^2}^{\infty} d\kappa^2\, C_B(\kappa^2)\Delta(x - y, \kappa^2),$$

where $M > m$. It should be emphasized that the introduction of extraneous particles or mass distributions into the existing axiomatic description of a single particle of mass m is *not* contemplated here; the only aim of this discussion is to represent an operator B_{IN} such that (59) is made plausible. With the asymptotic condition, Eq. (3), modified to read[11]

$$R_x\{A_{IN} + B_{IN}, j\} = :\exp\left[(A_{IN} + B_{IN})\frac{\delta}{\delta j}\right]: \times \langle R_x\{A_{IN} + B_{IN}, j\}\rangle, \tag{60}$$

the result of this argument is that the vacuum expectation value of the commutator of (59) shall be expected to vanish spacelike:

$$\Big[\langle R_x\{\lambda_1 A_{IN} + \lambda_3 B_{IN}, \lambda_2 j\}\rangle, \times \exp\left[\left(\frac{\lambda_1}{\lambda_2}\right)^2 D + \left(\frac{\lambda_3}{\lambda_2}\right)^2 D_B\right], \times \langle R_y\{\lambda_1 A_{IN} + \lambda_3 B_{IN}, \lambda_2 j\}\rangle\Big] = 0 \quad \text{spacelike,} \tag{61}$$

where the arbitrary, real scaling parameters $\lambda_{1,2,3}$ have been introduced.

If Eq. (61) is true, then when (11) is rewritten in the integral unitarity form

$$\langle R^{\lambda}_{x,y}\rangle = \frac{i}{\lambda}\,\theta(xy) \times [\langle R^{\lambda}_x\rangle, \exp(\lambda D + \lambda^{\alpha} D_B - \lambda^{\beta} D_B), \langle R^{\lambda}_y\rangle], \tag{62}$$

with

$$\lambda_2 = \lambda^{-1}, \qquad \lambda_1 = \lambda^{-\frac{1}{2}}, \qquad \lambda_3 = \lambda^{-1}[\lambda^{\alpha} - \lambda^{\beta}]^{\frac{1}{2}},$$

and

$$\langle R^{\lambda}_x\rangle = \langle R_x\{\lambda^{-\frac{1}{2}} A_{IN} + \lambda^{-1}[\lambda^{\alpha} - \lambda^{\beta}]^{\frac{1}{2}} B_{IN}, \lambda^{-1} j\}\rangle,$$

each term in the expansion of (62) in powers of λ will be relativistically invariant. This will certainly be the case for all 2-point functions, which property has been used in the previous sections. Assuming that this is true in general, means that one obtains for the general $(n + 1)$-point function of index j, a relativistic integral equation of form similar to that of (28);

[11] It is essential that the K–G operator appearing in the exponent of (60) refer to the mass m.

$$\langle \bar{R}^{(j)}_{x,y_1\cdots y_n}\rangle_0 = \theta(xy_1)Q^{(j)}(x, y_1 \cdots y_n)$$
$$+ i^2\theta(xy_1)\int \Delta'(u-v)$$
$$\times \{\langle \bar{R}^{(j)}_{x,uy_2\cdots y_n}\rangle_0\langle \bar{R}^{(0)}_{y_1,v}\rangle_0 + \langle \bar{R}^{(0)}_{x,u}\rangle_0\langle \bar{R}^{(j)}_{y_1,vy_2\cdots y_n}\rangle_0\}, \quad (63)$$

where the inhomogeneous $Q^{(j)}$ term is composed of functions with indices less than or equal to j; the only j-index amplitudes appearing in $Q^{(j)}$ will be $(m+1)$-point functions with $m < n$. The existence of solutions to (63) means that one then has, as in the perturbation theory, a meaningful expansion procedure, since higher-index functions are obtained in terms of lower-index functions. The only solutions to (63) which are acceptable, are those satisfying the requirements of connectedness, reality, retardedness, symmetry of the retarded coordinates, and condition B of I.

As in I, one expects to be able to construct crossing symmetric solutions because of the explicit operator nature of the transformation

$$R_x \to R^\lambda_x$$
$$= :\exp\left[(\lambda^{\frac{1}{2}}A_{IN} + [\lambda^\alpha - \lambda^\beta]^{\frac{1}{2}}B_{IN})\frac{\delta}{\delta \mathfrak{j}}\right]: \langle R^\lambda_x\rangle. \quad (64)$$

That is, the operator statement

$$R^\lambda_{\cdot,y} = \frac{i}{\lambda}\theta(xy)[R^\lambda_x, R^\lambda_y], \quad (65)$$

together with (64), just produces (62) upon taking the vacuum expectation value of (65). Hence all subsequent functional derivatives of (62) are equally well specified by calculating the vacuum-expectation value of the derivatives of (65). But any functional derivatives of (65) may be rearranged in the manifestly symmetric form

$$R^\lambda_{x,y_1\cdots y_n} = \frac{i^n}{\lambda^n}\sum_P \theta(xy_1)\cdots\theta(y_{n-1}y_n)$$
$$\times [\cdots [R^\lambda_x, R^\lambda_{y_1}], R^\lambda_{y_2}]\cdots], R^\lambda_{y_n}],$$

or

$$\langle R^\lambda_{x,y_1\cdots y_n}\rangle = \frac{i^n}{\lambda^n}\sum_P \theta(xy_1)\cdots\theta(y_{n-1}y_n)$$
$$\times [\cdots [\langle R^\lambda_x\rangle, \exp(\lambda D + \lambda^\alpha D_B - \lambda^\beta D_B), \langle R^\lambda_{y_1}\rangle], \cdots],$$
$$\times \exp(\lambda D + \lambda^\alpha D_B - \lambda^\beta D_B), \langle R^\lambda_{y_n}\rangle], \quad (66)$$

where the symbol $\sum_P$ denotes a sum on the permutation of the coordinates $y_1,\ldots,_n$. Equation (63) represents the integral equations obtained by the expansion of (62) in powers of λ, together with $n-1$ functional differentiations; the identical equations, but written in a crossing symmetric manner, are obtained by the corresponding λ-expansion of (66).

Solutions to these equations may be constructed using a simple generalization of the ordinary perturbation procedure given in I. One first notes that $Q^{(j)}$ is antisymmetric in x and y_1, and that the unitarity form of (63) may be rewritten, using (15) and the zero-index propagators as legs, in the form of the previous sections;

$$\int \langle \bar{R}^{(0)}_{x,x'}\rangle_0\langle \bar{R}^{(0)}_{y_1,y_1'}\rangle_0\langle \bar{R}^{(0)}_{y_2',y_2}\rangle_0\cdots\langle \bar{R}^{(0)}_{y_n',y_n}\rangle_0$$
$$\times [\langle \hat{R}^{(j)}_{x',y_1'\cdots y_n'}\rangle_0 - \langle \hat{R}^{(j)}_{y_1',x'\cdots y_n'}\rangle_0]$$
$$= Q^{(j)}(x, y_1\cdots y_n). \quad (67)$$

In momentum space, the Fourier transform of the absorptive part of the index-j, S-amputated amplitudes is then given by the transform of $Q^{(j)}$ divided by the factors $\bar{\tau}^{(0)}(-q)\bar{\tau}^{(0)}(-p_1)\bar{\tau}^{(0)}(p_2)\cdots\bar{\tau}^{(0)}(p_n)$, where $q, p_1, \cdots, p_n$ are the momentum variables conjugate to $x, y_1, \cdots, y_n$. Since $Q^{(j)}$ is assumed to be composed of known functions, the difference of $\langle \hat{R}^{(j)}_{x,y_1\cdots y_n}\rangle_0$ and its permuted form is then determined.

To pass from the absorptive part of this amplitude to the complete function, one may rewrite (63) with the aid of the previous unitarity manipulations:

$$\langle \bar{R}^{(j)}_{x,y_1\cdots y_n}\rangle_0 = \theta(xy_1)[\langle \bar{R}^{(j)}_{x,y_1\cdots y_n}\rangle_0 - \langle \bar{R}^{(j)}_{y_1,x\cdots y_n}\rangle_0]$$
$$- \theta(xy_1)\int \langle \bar{R}^{(0)}_{x,x'}\rangle_0\langle \bar{R}^{(0)}_{y_1,y_1'}\rangle_0\langle \bar{R}^{(0)}_{y_2',y_2}\rangle_0\cdots\langle \bar{R}^{(0)}_{y_n',y_n}\rangle_0$$
$$\times [\langle \hat{R}^{(j)}_{x',y_1'\cdots y_n'}\rangle_0 - \langle \hat{R}^{(j)}_{y_1',x'\cdots y_n'}\rangle_0]$$
$$+ \theta(xy_1)Q^{(j)}(x, y_1\cdots y_n),$$

which, with (67), reduces to

$$\langle \bar{R}^{(j)}_{x,y_1\cdots y_n}\rangle_0 = \theta(xy_1)[\langle \bar{R}^{(j)}_{x,y_1\cdots y_n}\rangle_0 - \langle \bar{R}^{(j)}_{y_1,x\cdots y_n}\rangle_0]. \quad (68)$$

Equation (68) is the statement of integral unitarity for the entire amplitude; to be useful, this must be converted into a statement about the S-amputated function.

Consider, for the time sequence $x^0 > y_1^0$, K-G amputation of both sides of (68),

$$\int \langle \bar{R}^{(0)}_{x,x'}\rangle_0\langle \bar{R}^{(0)}_{y_1',y_1}\rangle_0\cdots\langle \bar{R}^{(0)}_{y_n',y_n}\rangle_0\langle \hat{R}^{(j)}_{x',y_1'\cdots y_n'}\rangle_0$$
$$= \int \langle \bar{R}^{(0)}_{y_2',y_2}\rangle_0\cdots\langle \bar{R}^{(0)}_{y_n',y_n}\rangle_0$$
$$\times [\langle \bar{R}^{(0)}_{x,x'}\rangle_0\langle \bar{R}^{(0)}_{y_1',y_1}\rangle_0\langle \hat{R}^{(j)}_{x',y_1'\cdots y_n'}\rangle_0$$
$$- \langle \bar{R}^{(0)}_{y_1,y_1'}\rangle_0\langle \bar{R}^{(0)}_{x',x}\rangle_0\langle \hat{R}^{(j)}_{y_1',x'\cdots y_n'}\rangle_0], \quad x^0 > y_1^0. \quad (69)$$

Since $\langle \bar{R}^{(0)}_{x,x'}\rangle_0 = \delta(x-x') + \langle \bar{R}^{(0)}_{x,x'}\rangle_0$, where

$\langle \bar{R}^{(0)\prime}_{\mathbf{x},x'} \rangle_0$ represents the nonsingular part (containing the dynamics) of the K–G amputated propagator, there will be one term in each of the multiple integrals of (69) with every propagator replaced by a δ function. The remaining terms will have k δ functions and $n + 1 - k$ $\langle \bar{R}^{(0)\prime} \rangle$ factors in all possible combinations except $k = n + 1$. We denote the total of such terms by

$$r(x, x'; y_1', y_1; \cdots ; y_n', y_n) = \langle \bar{R}^{(0)}_{\mathbf{x},x'} \rangle_0 \langle \bar{R}^{(0)}_{y_1',y_1} \rangle_0 \cdots \langle \bar{R}^{(0)}_{y_n',y_n} \rangle_0 - \delta(x - x')\delta(y_1' - y_1) \cdots \delta(y_n' - y_n),$$

and write (69) in the form

$$\langle \hat{R}^{(j)}_{\mathbf{x},y_1 \cdots y_n} \rangle_0 = \langle \hat{R}^{(j)}_{\mathbf{x},y_1 \cdots y_n} \rangle_0 - \langle \bar{R}^{(j)}_{y_1,\mathbf{x} \cdots y_n} \rangle_0 - \int r(y_1, y_1'; x'x; \cdots ; y_n', y_n) \langle \hat{R}^{(j)}_{\mathbf{y_1'},x' \cdots y_n'} \rangle_0, \quad (70)$$

where a term

$$+ \int r(x, x'; y_1', y_1; \cdots ; y_n', y_n) \langle \hat{R}^{(j)}_{\mathbf{x}',y_1' \cdots y_n'} \rangle_0$$

has been subtracted from each side of (70). Because the integral $\int r \langle \hat{R}^{(j)} \rangle$ maintains its proper time sequence, and (70) has been written for $x^0 > y_1^0$, the last term on the right-hand side of (70) must vanish if the S amplitude is to be retarded. It therefore follows that

$$\langle \hat{R}^{(j)}_{\mathbf{x},y_1 \cdots y_n} \rangle_0 = \langle \hat{R}^{(j)}_{\mathbf{x},y_1 \cdots y_n} \rangle_0 - \langle \hat{R}^{(j)}_{\mathbf{y_1},\mathbf{x} \cdots y_n} \rangle_0, \quad x^0 > y_1^0$$

$$\langle \hat{R}^{(j)}_{\mathbf{x},y_1 \cdots y_n} \rangle_0 = 0, \quad x^0 < y_1^0,$$

or

$$\langle \hat{R}^{(j)}_{\mathbf{x},y_1 \cdots y_n} \rangle_0 = \theta(xy_1)[\langle \hat{R}^{(j)}_{\mathbf{x},y_1 \cdots y_n} \rangle_0 - \langle \hat{R}^{(j)}_{\mathbf{y_1},\mathbf{x} \cdots y_n} \rangle_0] + \theta(x - y_1) f^{(j)}(x, y_2 \cdots y_n), \quad (71)$$

where, as in I, $f^{(j)}$ is determined by imposing the requirement of symmetry of the retarded coordinates. If no contact terms more singular than δ functions have been included in any of the retarded functions comprising $Q^{(j)}$, then none can appear on the right-hand side of (71). Because of the form of (66), one may assert that (71) will be crossing symmetric, which, with the proviso of relativistic invariance, completes the construction.

It is instructive to see the correspondence between a specific solution of the form of (71) and the corresponding special solution obtained for the lower-index functions in the previous sections. The simplest, nontrivial example is furnished by the scalar vertex of index one, and for this case it is clear from a comparison of (46) and (71) that only the previously omitted self-energy legs need be considered here. The necessary $Q^{(1)}$ terms may be read off from (39), and a simple calculation shows that the terms which should be added to (46) are

$$\frac{g}{\pi} \delta(x - z) \int d\kappa^2 \, \rho^{(1)}(\kappa^2) \Delta_R(z - y, \kappa^2) + \frac{g}{\pi} \delta(y - z) \int d\kappa^2 \, \rho^{(1)}(\kappa^2) \Delta_R(x - z, \kappa^2) + \delta(x - y) f^{(1)}(y, z), \quad (72)$$

where

$$\rho^{(1)}(s) = |\bar{r}^{(0)}(s)|^{-2} \operatorname{Im} [\bar{r}^{(1)}(s) \bar{r}^{(0)*}(s)].$$

Evidently, the appropriate choice for $f^{(1)}$ is

$$f^{(1)}(y, z) = \frac{g}{\pi} \int d\kappa^2 \, \rho^{(1)}(\kappa^2) \Delta_R(y - z, \kappa^2).$$

These extra contributions vanish if the phase of $\bar{r}^{(1)}(s)$ is required to be the same as that of $\bar{r}^{(0)}(s)$, or if $\bar{r}^{(1)}(s)$ is itself zero; these conditions are just the content of Eqs. (42).

ACKNOWLEDGMENTS

Much of this work was performed while the author was a summer vistor at Brookhaven National Laboratory. It is a pleasure to thank Professor G. C. Wick for his kind hospitality, and to thank Dr. P. DeCelles for many interesting discussions. A valuable remark concerning the interpretation of the formalism was made by Dr. J. G. Taylor.

Group Properties of Free-Field Equations

V. AMAR AND M. LUCCHINI
Istituto di Fisica dell'Università, Milano, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Milano, Milano, Italy
P. G. SONA
Laboratori C. I. S. E. and Istituto di Scienze Fisiche dell'Università, Milano, Italy
(Received 25 June 1962)

We show that a set of n free-field equations and suitable commutation relations relevant to spin 0 or ½ fields, is invariant with respect to a continuous group of transformations isomorphic to the n-dimensional real orthogonal group. We then find that such a set can always be derived starting from a given particle multiplet. The continuous group found in this way, includes as subgroups the isotopic rotations and the gauge groups related to baryon number and strangeness. Hence this method enables us to reobtain the results reached by other authors in some particular cases, and to extend them to all particle multiplets.

1. INTRODUCTION

In recent papers, some authors[1,2] have tried to geometrize the conservation laws related to the quantum numbers characteristic of the strong interactions, i.e. g^2, g_3, g_N, g_S. They tried to derive them from the invariance of the field equations and commutation rules, with respect to some subgroups of only one continuous group of transformations. These attempts essentially concern some cases of free fields describing the particles of an isotopic spin multiplet with its charge conjugate.

The aim of this work is to show how it is possible to derive for each multiplet the largest continuous group of transformations, not affecting space–time coordinates, conserving free-field equations and the relevant commutation rules. In this group one may select some subgroups whose infinitesimal generators correspond to the usual quantum numbers g_3, g_N, g_S.

However, in our method this result has been reached in a quite formal way, because on searching for these subgroups we had not any *a priori* guiding principle common to all multiplets. We have then followed the *a posteriori* criterion of reobtaining the well-known quantum number assignments.

At any rate, it is remarkable that, for each multiplet, it is always possible to set up a continuous group including as subgroups, the isotopic rotations and the gauge groups related to baryon number and strangeness. Obviously, the infinitesimal generators connected to these quantum numbers are not independent because each multiplet is characterized by a fixed value of baryon number and strangeness.

In Secs. 2 and 3 we derive the continuous group, leaving invariant the free-field equations and commutation rules of an abstract set of independent fields.

In Sec. 4 our method is applied to the physical particle multiplets, and there the subgroups related to quantum numbers of physical interest are presented. As particular cases, the nucleon, K meson, and Σ multiplets are examined.

2. SPIN 0 FIELDS

Consider n independent fields $\varphi_1(x), \cdots \varphi_n(x)$, Hermitean solutions of the Klein–Gordon equation and obeying the following commutation relations:

$$[\varphi_i(x), \varphi_j(y)] = -i\delta_{ij}\Delta(x - y). \tag{1}$$

Consider now the system of equations

$$(\Box - m^2)\varphi_i(x) = 0 \qquad i = 1, 2, \cdots n. \tag{2}$$

From the linearity of the Klein–Gordon operator, it immediately follows that the system (2) is invariant with respect to the group of transformations

$$\varphi_i'(x) = a_{ij}\varphi_j(x), \qquad a_{ij} \text{ real numbers}, \quad \det|a_{ij}| \neq 0. \tag{3}$$

Requiring that the transformations (3) also preserve the commutation relations (1), one finds that the matrix $|a_{ij}| \equiv A$ should be orthogonal, i.e.,

$$AA^T = A^TA = 1. \tag{4}$$

Hence the system (2) and the commutation rules (1), are invariant with respect to the n-dimensional real orthogonal group.

3. SPIN 1/2 FIELDS

Consider $2n$ independent fields $\chi_1(x), \cdots \chi_{2n}(x)$, solutions of the Dirac equation, with the following properties:

[1] F. Gürsey, Nuovo Cimento **7**, 411 (1958).
[2] J. Lukierski, Nuovo Cimento **13**, 410 (1959).

$$\chi_i^c = \overline{C\chi_i^T} = \begin{cases} +\chi_i & i = 1, 2, \cdots n \\ -\chi_i & i = n+1, \cdots 2n, \end{cases} \tag{5}$$

and obeying the following anticommutation relations:

$$\{\bar{\chi}_{i\alpha}(x), \chi_{j\beta}(y)\} = -i\delta_{ij}S_{\beta\alpha}(y-x), \tag{6}$$

$$\{\chi_{i\alpha}(x), \chi_{j\beta}(y)\} = i\delta_{ij}\Sigma_\sigma S_{\alpha\sigma}(x-y)C_{\sigma\beta}. \tag{6a}$$

Consider now the system of equations

$$(\gamma_\mu\partial_\mu + m)\chi_i(x) \equiv D\chi_i(x) = 0 \quad i = 1, 2, \cdots 2n. \tag{7}$$

From the linearity of the Dirac operator D, it follows that the system (7) is invariant with respect to the group of transformations

$$\chi_i' = a_{ij}\chi_j, \qquad a_{ij} \text{ complex numbers}, \quad \det|a_{ij}| \neq 0. \tag{8}$$

Requiring that (6) remains invariant with respect to (8), we reach the following conditions on a_{ij}:

$$\sum_1^{2n} {}_j a_{ij}a_{\rho j}^* = \delta_{i\rho}. \tag{9}$$

Equation (9) reduces the $2n$-dimensional full linear complex group (8) to the $2n$-dimensional unitary group. To have (6a) invariant too, we obtain

$$\sum_1^{2n} {}_j a_{ij}a_{\rho j} = \delta_{i\rho}, \tag{10}$$

i.e., the $2n$-dimensional orthogonal complex group. It then follows that the system (7) and the anticommutation rules (6) and (6a), are invariant with respect to the $2n$-dimensional real orthogonal group.[3]

4. APPLICATIONS TO PARTICLE MULTIPLETS

It is easily seen from the free-field equations and the usual commutation (or anticommutation) relations describing the particles of a multiplet with their charge conjugates, it is possible, by means of a unitary transformation U, to define a set of fields fulfilling the conditions stated in Secs. 2 and 3. Indeed, the commutation rules (1), (6) and (6a) have been obtained considering for every field Φ and its charge conjugate Φ^c, the two independent linear combinations

$$(\Phi + \Phi^c)/\sqrt{2}, \qquad (\Phi - \Phi^c)/i\sqrt{2}.$$

Then, if we have invariance under the real orthogonal group O for these combinations, we have as a consequence, the invariance under the group $G = U^+OU$, isomorphic to O, for the usual fields describing physical particles.

In what follows, we are interested in considering some subgroups of G which will enable us to define the physically important quantum numbers within each multiplet.

A. Nucleon field

Consider the following equation:[4]

$$(D \times I_4)\Psi = 0 \qquad \Psi \equiv \begin{bmatrix} \psi_p \\ \psi_n \\ \psi_p^c \\ \psi_n^c \end{bmatrix}, \tag{11}$$

and the usual anticommutation relations

$$\begin{aligned} \{\bar{\psi}_{\gamma\alpha}(x), \psi_{\gamma'\beta}(y)\} &= -i\delta_{\gamma\gamma'}S_{\beta\alpha}(y-x), \\ \{\psi_{\gamma\alpha}(x), \psi_{\gamma'\beta}(y)\} &= 0 \qquad \gamma, \gamma' = p, n, \\ \{\psi_{\gamma\alpha}^c(x), \psi_{\gamma'\beta}(y)\} &= i\delta_{\gamma\gamma'}\Sigma_\sigma S_{\alpha\sigma}(x-y)C_{\sigma\beta}. \end{aligned} \tag{12}$$

We carry out the unitary transformation

$$\Psi \Rightarrow \Psi' = U\Psi,$$

where

$$U = \frac{1}{\sqrt{2}}\begin{vmatrix} I_2 & I_2 \\ -iI_2 & iI_2 \end{vmatrix},$$

and we obtain

$$\Psi' = \begin{bmatrix} (\psi_p + \psi_p^c)/\sqrt{2} \\ (\psi_n + \psi_n^c)/\sqrt{2} \\ (\psi_p - \psi_p^c)/(\sqrt{2}i \\ (\psi_n - \psi_n^c)/\sqrt{2}i \end{bmatrix} \equiv \begin{bmatrix} \chi_1 \\ \chi_2 \\ \chi_3 \\ \chi_4 \end{bmatrix}.$$

Hence, it follows that (11) and (12) are invariant with respect to a group G_4 isomorphic to the four-dimensional real orthogonal group. Particularly, we have invariance with respect to a group R_4, isomorphic to the four-dimensional real rotations. As it is well known, R_4 is isomorphic[5] to the direct product

[3] We should point out that the limitation to consider in spin ½-case fields with the properties (5), is the same as the one of requiring Hermitean fields in spin-0 case. In fact if we choose the Majorana representation of Dirac matrices [see e.g. P. Roman, *Theory of Elementary Particles* (North-Holland Publishing Company, Amsterdam, 1960), p. 126] the χ_i are Hermitean. Then, the structure of the continuous group is found at a certain extent to be independent from the spin of fields, but it only depends on the number of fields being considered. The fact that with this method we can reach, as shown in Sec. 4, the same results of Gürsey without the formalism used by him, strongly connected with the field spin properties, demonstrates that this connection is not essential.

[4] We use the symbol I_n to indicate the n-dimensional unity matrix.

[5] See, e.g., M. Hamermesh, *Group Theory* (Addison-Wesley Publishing Company, Reading, Massachusetts, 1962).

of two unitary unimodular bidimensional groups U' and U'', i.e.:

$$R_4 \simeq U' \times U''.$$

Consider the following group of matrices acting on Ψ^6.

$$R(a, \mathbf{l}) = \begin{vmatrix} \exp\left(\frac{1}{2}ia\mathbf{l} \times \boldsymbol{\vartheta}\right) & 0 \\ 0 & \exp\left(-\frac{1}{2}ia\mathbf{l} \times \boldsymbol{\vartheta}^*\right) \end{vmatrix}.$$

This group, which obviously corresponds to the usual isotopic rotations, is a subgroup of R_4. Consider also the group of matrices, also acting on Ψ:

$$\bar{R}(\alpha, \beta) = \begin{vmatrix} \alpha I_2 & \beta\sigma_2 \\ (-\beta\sigma_2)^* & \alpha^* I_2 \end{vmatrix},$$

where $|\alpha|^2 - |\beta|^2 = 1$. $\bar{R}(\alpha, \beta)$ is a subgroup of R_4 commuting with $R(a, \mathbf{1})$ and isomorphic to a unitary unimodular bidimensional group, as it is easily shown by the following mapping:

$$\alpha I_2 \Leftrightarrow \alpha$$

$$\beta\sigma_2 \Leftrightarrow i\beta.$$

We obviously have

$$R_4 = R\bar{R}.$$

(One can check that $R' = URU^+$ and $\bar{R}' = U\bar{R}U^+$ are subgroups of the four-dimensional real orthogonal group. Hence R and $\bar{R}$ are subgroups of R_4). $\bar{R}(e^{i\varphi}, 0)$ is a subgroup of $\bar{R}(\alpha, \beta)$ which clearly corresponds to the baryon number.

Let us call ϑ_N the infinitesimal operator corresponding to these bidimensional rotations. The group R_4 is of rank two; we have then only two independent commuting infinitesimal operators which we can assume as ϑ_N and the infinitesimal operator ϑ_3 which induces isotopic rotations around the third axis. It is easily seen that the rotations induced by the infinitesimal operator

$$Q = \vartheta_3 + \tfrac{1}{2}\vartheta_N \tag{13}$$

correspond to the charge.

We then obtain the results of Gürsey,[1] in the group theoretical form illustrated by Lukierski.[2] From our method, it is obvious that we can treat the spin 0 particles in the same way. That is, if we had considered the field Φ relevant to K particles;

$$\Phi = \begin{bmatrix} \varphi_{k^+} \\ \varphi_{k^\circ} \\ \varphi_{k^-} \\ \varphi_{\bar{k}^\circ} \end{bmatrix},$$

we should obtain the same results and, if we now consider the previous bidimensional rotations induced by the infinitesimal operator ϑ_N as corresponding now to strangeness number ϑ_S, the relation (13) is the well-known formula for the charge operator for K particles.

B. Σ field

Consider the following system of equations:

$$(D \times I_6)\Psi = 0 \qquad \Psi = \begin{bmatrix} \psi_{\Sigma^+} \\ \psi_{\Sigma^\circ} \\ \psi_{\Sigma^-} \\ \psi^c_{\Sigma^+} \\ \psi^c_{\Sigma^\circ} \\ \psi^c_{\Sigma^-} \end{bmatrix}, \tag{14}$$

and the anticommutation relations similar to (12). By the unitary transformation $\Psi \to X = U\Psi$ where

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} I_3 & I_3 \\ -iI_3 & iI_3 \end{bmatrix},$$

we comply with the conditions of Sec. 3, for the case $2n = 6$. The system (14) and the relevant commutation rules are invariant with respect to a group G_6 isomorphic to the 6-dimensional orthogonal real group. The group of matrices

$$H(\epsilon_i) = \begin{vmatrix} \exp\left(i \sum_{1}^{3} {}_s \epsilon_s \vartheta_s\right) & 0 \\ 0 & \exp\left(i \sum_{1}^{3} {}_s \epsilon_s \vartheta_s\right)^* \end{vmatrix},$$

where

$$[\vartheta_1, \vartheta_2] = i\vartheta_3 \quad \text{cyclic},$$

$$\vartheta_3 = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{vmatrix},$$

which clearly corresponds to isotopic rotations, is a subgroup of G_6.

G_6 is a simple group of rank three; we have then three infinitesimal independent operators commuting one another. We can select them as

$$\Theta_3, \vartheta_N, \vartheta',$$

where Θ_3 induces the isotopic rotations around the third axis and ϑ_N and ϑ' are the infinitesimal opera-

[1] G. R. Allcock, Nucl. Phys. 27, 204 (1961).

tors, respectively, inducing the following plane rotations:

$$\mathcal{G}_N \to \begin{vmatrix} e^{i\alpha}I_3 & 0 \\ 0 & e^{-i\alpha}I_3 \end{vmatrix},$$

$$\mathcal{G}' \to \begin{vmatrix} e^{i\alpha} & & & & & & 0 \\ & e^{-i\alpha} & & & & & \\ & & e^{i\alpha} & & & & \\ & & & e^{-i\alpha} & & & \\ & & & & e^{i\alpha} & & \\ 0 & & & & & \ddots & \\ & & & & & & e^{-i\alpha} \end{vmatrix}.$$

$\mathcal{G}_N$ can be designated as a baryon number, but we see no physical meaning for $\mathcal{G}'$.

As G_6 is simple, H cannot be an invariant subgroup. Nevertheless Θ_3, $\mathcal{G}_N$, $\mathcal{G}'$, and $\Theta^2 = \Theta_1^2 + \Theta_2^2 + \Theta_3^2$, are commuting one another,[7] as one can easily check.[8] Note that Θ^2 is not an invariant of the group G_6 as it happens in the nucleon case where the isotopic rotations R are an invariant subgroup of R_4.

[7] Θ_1 and Θ_2 are the infinitesimal operators corresponding to the rotations around the first and second isotopic axis.

5. FINAL REMARKS

Obviously, no new physically interesting result can be obtained here, because we have considered only the free-field problem. However, our method allows a general treatment of the group theoretical properties of a set of free-field equations. It enables us to reobtain, as particular cases, the results obtained by other authors and to reach similar results for all particle multiplets.

ACKNOWLEDGMENT

We would like to thank Professor A. Loinger for helpful discussions and comments.

[8] For a group G of n-dimensional real rotations of rank l it is possible to show that one can select $l + 2$ infinitesimal operators $a_1, a_2 \cdots a_l, a_1', a_1''$ so that

$$[a_i, a_j] = 0 \quad \text{and} \quad [a_i, a_1^2 + (a_1')^2 + (a_1'')^2] = 0$$

and that a_1, a_1' and a_1'' are the infinitesimal generators of a subgroup of G isomorphic to three-dimensional real rotations.

Symmetry in Quantum Theory*

GÉRARD EMCH
Institut de Physique Théorique de l'Université de Genève, Switzerland
AND
CONSTANTIN PIRON
Universités de Genève et Lausanne, Switzerland

The aim of this note is to give a precise definition of symmetry in quantum theory in order to generalize Wigner's representation theorem, in the framework of lattice theory and projective geometry. We do not require the concept of physical state; the results are valid for any field (division ring) used in the realization of the lattice of propositions. Physical systems with the most general superselection rules are included in the theory.

I. THE LATTICE OF PROPOSITIONS

EVERY measurement of a physical quantity can in principle be reduced to a series of "yes–no experiments," that is, measurements which have as result only one of two alternatives. We shall refer to the statement relative to a yes–no experiment as a *proposition*. The structure of a physical system as revealed in the set of all observables can therefore be completely described by giving the structure of the set τ of all propositions on the physical system.

We say that a proposition is *true* if the answer to the related yes–no experiment is yes with a probability equal to one; for two elements x and y of τ we write $x \subseteq y$ if "x is true" implies "y is true". This relation satisfies all the properties of a partial ordering relation on τ.

Following to some extent Birkhoff and von Neumann,[1] we shall postulate some further axioms on τ, which define a so-called proposition system.[2]

1. τ is a complete lattice, i.e., given any set of propositions x_i of τ, there exist always in τ an upper and a lower bound, respectively, written $\bigcup x_i$ and $\bigcap x_i$. From that follows the existence of two propositions $\varnothing$ and u such that

$$\varnothing \subseteq x \subseteq u \quad \text{for all} \quad x \quad \text{in} \quad \tau.$$

The symbol $x \cup y$ denotes the proposition "x or y," while $x \cap y$ denotes the proposition "x and y."

2. τ is orthocomplemented, i.e. a mapping $\mathcal{O}$ of τ onto itself is given such that $x \to \mathcal{O}x$ satisfies the following relations:

(i) $x \cup \mathcal{O}x = u, \qquad x \cap \mathcal{O}x = \varnothing;$

(ii) $\mathcal{O}\mathcal{O}x = x;$

(iii) $x \subseteq y \quad \text{implies} \quad \mathcal{O}y \subseteq \mathcal{O}x.$

This operation corresponds to the logical negation.

3. τ is atomic, i.e. for every proposition $x \neq \varnothing$ there exists a proposition $P \subseteq x$, different from $\varnothing$ such that $\varnothing \subseteq y \subseteq P$ implies either $y = \varnothing$ or $y = P$; such a proposition P is called a *point*.

4. Given two propositions $a \subseteq b$, the *segment* $[a, b]$ is defined by $[a, b] = \{x \in \tau \mid a \subseteq x \subseteq b\}$.

Our fourth axiom expresses that every segment inherits the first three axioms; more precisely:

(i) the mapping $\mathcal{O}_r$ defined on $[a, b]$ by:

$$\mathcal{O}_r x \equiv (a \cup \mathcal{O}x) \cap b$$

is an orthocomplementation on $[a, b]$.

(ii) For every point P (not contained in a) such that $a \cup P \subseteq b$, $a \cup P$ is a point for the segment $[a, b]$.

As a consequence we have that $[a, b]$ also satisfies our fourth axiom.

Two propositions x and y which satisfy the relation

$$(x \cap \mathcal{O}y) \cup y = (y \cap \mathcal{O}x) \cup x$$

are said to be *compatible* and we note this fact by $x \leftrightarrow y$.

We say a lattice τ is a *direct union* of a set of lattices τ_i if any element $x \in \tau$ can be written in a unique way in the form $x = \bigcup x_i$ with $x_i \in \tau_i$. A lattice is said to be irreducible, if it cannot be represented as a direct union of two lattices each with more than one element. In the other cases it is said to be *reducible*.

The following results can be proved from the axioms:

* Work supported in part by the Swiss National Research Fund.

[1] G. Birkhoff and J. von Neumann, Ann. Math. 37, 823 (1936).

[2] C. Piron, Thesis presented to the University of Lausanne, Switzerland (unpublished).

(1) Irreducibility of τ is equivalent to any of the following conditions:

(i) $\varnothing$ and u are the only propositions which are compatible with every proposition of τ.

(ii) If P and Q are two arbitrary points in τ, there exists always a third point R in τ such that

$$P \cup Q = P \cup R = R \cup Q.$$

The second of these conditions expresses the *principle of superposition* and a irreducible proposition system is then said to be *coherent.*

(2) Every proposition-system τ can be written in a unique way as a direct union of irreducible proposition systems. A reducible proposition system is said to exhibit *superselection rules.*

A lattice is called a *projective geometry* if it is complete, modular, complemented, atomic and satisfies the $\bigcap$ continuity[3].

(3) It is always possible to embed a proposition system (even if it is not modular) in a projective geometry so that:

(i) this embedding, restricted to the points, is one-to-one.

(ii) the partial ordering relation is preserved.

(iii) the image of a finite union of points is the union of the images of these points.

(iv) for any family of propositions, the image of the intersection of the propositions of the family is the intersection of the images of these propositions. [For the proof, see reference (2).]

It follows that if τ is irreducible, so also is the related projective geometry. Finally, it is known that every irreducible projective geometry, whose projective dimension $(n - 1)$ is at least three, can be realized as the partially ordered set of the linear manifolds of an n-dimensional vector space on a field (division ring).

For instance, the linear manifolds of an infinite Hilbert space form a projective geometry whose partial ordering relation is the inclusion of sets. Let us now consider the family of the *closed* linear manifolds (in the topological sense). With the above partial ordering relation, this family is a lattice which satisfies our four axioms, i.e. a proposition system. Let us denote by $P(x)$, the orthogonal projection which maps the whole space onto the closed linear manifold x; it is well known that

$$P(\mathcal{O}x) = I - P(x) \quad \text{and} \quad P(x \cap y) = \lim_{n\to\infty} [P(x)P(y)]^n.$$

In this case, two propositions x and y are compatible if and only if the projections $P(x)$ and $P(y)$ commute.

A very general proposition system (with any kind of superselection rule) is obtained if we consider an arbitrary family of Hilbert spaces H_i; any proposition x of this system is a family $\{x_i\}$ of closed linear manifolds $x_i \subseteq H_i$. The system in question is then defined as the direct union of the irreducibles systems τ_i consisting of all closed linear manifolds x_i.

II. DEFINITION OF A SYMMETRY. GENERALIZED WIGNER'S THEOREM

Let τ_1 and τ_2 be two proposition systems and μ a one-to-one mapping of τ_1 onto τ_2 with the following properties:

(i) $x \subseteq y$ implies $\mu x \subseteq \mu y$, and conversely;

(ii) $\mu\mathcal{O}_1 x = \mathcal{O}_2 \mu x$,

for every x and y of τ_1.

Such a mapping we call a *morphism* of τ_1 onto τ_2; a morphism of a proposition system τ onto itself (i.e. an automorphism), will be called a *symmetry* of τ.

Lemma 1. Let μ be a morphism of τ_1 onto τ_2, then

$$\mu(\bigcup x_i) = \bigcup \mu x_i,$$

$$\mu(\bigcap x_i) = \bigcap \mu x_i.$$

Proof: From the first property of a morphism we conclude

$$\mu x_i \subseteq \mu(\bigcup x_i) \quad \text{for every} \quad x_i.$$

Then

$$\mu x_i \subseteq \bigcup \mu x_i \subseteq \mu(\bigcup x_i)$$

follows because of the definition of the union. μ^{-1} exists and we thus have

$$x_i \subseteq \mu^{-1}(\bigcup \mu x_i) \subseteq \bigcup x_i.$$

Again from the definition of the union,

$$\mu^{-1}(\bigcup \mu x_i) = \bigcup x_i,$$

or

$$\bigcup \mu x_i = \mu(\bigcup x_i).$$

By analogous arguments we prove the second part of the lemma.

From this lemma we conclude:

(i) $\mu u_1 = u_2$ and $\mu\varnothing_1 = \varnothing_2$;

(ii) $x \leftrightarrow y$ implies $\mu x \leftrightarrow \mu y$, and conversely.

This last assertion says: Every morphism preserves compatibility.

[3] M. L. Dubreil-Jacotin, L. Lesieur, and R. Croisot; *Leçons sur la théorie des treillis,* (Gauthier-Villars, Paris, 1953).

Lemma 2. If P_1 is a point of τ_1, then μP_1 is a point of τ_2 and conversely.

Proof: Let $x_2 \in \tau_2$ such that $\varnothing_2 \subseteq x_2 \subseteq \mu P_1$; then $\varnothing_1 \subseteq \mu^{-1}(x_2) \subseteq P_1$. But P_1 is a point, so that either $\mu^{-1}(x_2) = P_1$, or $\mu^{-1}(x_2) = \varnothing_1$; i.e., either $x_2 = \mu P_1$ or $x_2 = \varnothing_2$.

Q.E.D.

Lemma 3. A morphism is completely determined by its restriction to the points; more precisely,

$$\mu x = \bigcup \mu P_i.$$

This union is understood on the set of the images of all the points P_i of τ_1 which are contained in x. By definition of a morphism, this implies that it is on the set of all the points of τ_2 contained in μx. This third lemma is an immediate consequence of the first two lemmas if we observe that a proposition is the union of the points contained in it.

Lemma 4. Given a morphism, its restriction to any segment is a morphism of this segment; the image of an irreducible segment is again an irreducible segment.

Proof: Let $[a, b]$ be a segment of τ_1, its image is the segment $[\mu a, \mu b]$ of τ_2; the restriction of μ to $[a, b]$ satisfies trivially the first property of a morphism. Let $x_2 \in [a, b]$ and let us construct $\mu \mathcal{O}_r^{(1)} x$. Using the first lemma we find

$$\mu \mathcal{O}_r^{(1)} x = \mu\{(a \cup \mathcal{O}_1 x) \cap b\} = (\mu a \cup \mathcal{O}_2 \mu x) \cap \mu b,$$

which is $\mathcal{O}_r^{(2)} \mu x$. This means that the second property of a morphism is also satisfied by the restriction of μ on $[a, b]$. The first part of this lemma is thus proved. The second part follows if we use, instead of the original definition of irreducibility, the condition of coherence (see first part) which is equivalent. From the first two lemmas, we see that this condition is preserved by a morphism.

It follows from this lemma that a symmetry on a reducible proposition system is expressed as a morphism of every coherent subsystem onto another coherent subsystem which may or may not be different from the first one.

The next theorem is a slight generalization of Wigner's theorem.

Theorem: When τ is a direct union of coherent subsystems, each of them containing at least four independent points, then every symmetry σ on τ maps independently every coherent subsystem τ_α onto a coherent subsystem $\tau_{\sigma(\alpha)}$, by a semilinear transformation.

Proof: From the fourth lemma, we saw that each τ_α is mapped onto a $\tau_{\sigma(\alpha)}$ by a morphism; it is then sufficient to prove that every morphism μ of an irreducible τ_α on an irreducible τ_β is induced by a semilinear transformation. We already know from the first part that it is possible to embed τ_α in the set of all the linear manifolds of a vector space V_α; let i_α be this embedding (here the assumption that τ_α contains at least four independent points is essential); the restriction of i_α to the points of τ_α is a one-to-one mapping on the rays of V_α, so that i_α^{-1} is defined for every ray of V_α. The same arguments also apply to τ_β. Let us denote by φ the one-to-one mapping $i_\beta \mu i_\alpha^{-1}$ acting on the set of all rays Ψ of V_α, and whose range is the set of all the rays of V_β. Using the properties of the i and μ we deduce

$$\varphi(\Psi_1 \cup \Psi_2) = i_\beta \mu i_\alpha^{-1}(\Psi_1 \cup \Psi_2) = \varphi\Psi_1 \cup \varphi\Psi_2,$$

so that for any $\Psi_3 \subseteq \Psi_1 \cup \Psi_2$, we have $\varphi\Psi_3 \subseteq \varphi\Psi_1 \cup \varphi\Psi_2$, and conversely.

At this point we refer to the first fundamental theorem of projective geometry[4] and conclude that φ is induced by a semilinear transformation S of V_α onto V_β. We remind the reader that if V_α (resp. V_β) is a vector space on a field F_α (resp. F_β), a *semilinear transformation* S of V_α onto V_β is a one-to-one mapping, preserving the vector space structure and such that

$$S(fv) = f^s S v \quad \text{for every} \quad f \in F_\alpha \quad \text{and} \quad v \in V_\alpha,$$

where f^s is the image of f under an isomorphism of F_α onto F_β.

There remains to prove that the restriction to $i_\alpha(\tau_\alpha)$ of the mapping induced by S coincides with $i_\beta \mu i_\alpha^{-1}$ on this domain. This is, however, a direct consequence of our third lemma. Our principal theorem is thus completely proved.

Let us now consider the set of all the projectors of a Hilbert space H; we already know that this is a realization of an irreducible proposition system. From our theorem, we conclude that if H is constructed on the field of real numbers,[5] every morphism of H is induced by a linear transformation, because the only automorphism of R is the identity; if H is constructed on the field C of complex numbers, there are two types of morphisms: those which are induced by linear transformations and those which are induced by antilinear transformations.[6] The first type of semilinear transformations correspond to the

---

[4] See for instance E. Artin, *Geometric Algebra* (Interscience Publishers, Inc., New York, 1957).

[5] E. C. G. Stueckelberg, Helv. Phys. Acta. **33**, 727 (1960).

[6] In this case the semilinear transformations, having to satisfy $\mu \mathcal{O}_1 P_1 = \mathcal{O}_2 \mu P_1$ must be continuous, so that the only allowed automorphisms of C are identity and complex conjugation; see for instance, N. Bourbaki, *Espaces vectoriels topologiques* (Hermann, Paris, 1955), Chap. IV, Sec. 5, Exercise 20.

identity on C, and the second to the automorphism of C realized by complex conjugation. This is a new proof of Wigner's representation theorem.[7,8] Finally if H is constructed on the field Q of real quaternions,[9] every morphism is physically equivalent to a linear transformation, and the reason for this is that every automorphism on Q is inner. Let S be a semilinear transformation, then f^S is necessarily of the form $qfq^{-1} = f^S$ for every f in Q and with fixed q in Q. The transformation $S' = q^{-1}S$ is linear and induces the same morphism.

The results so far obtained are independent of the notion of "physical state". In the next section we shall examine the relation of symmetry transformation to the notion of state of a physical system.

III. THE CONCEPT OF STATE

A state[2] of a proposition-system τ is by definition a mapping E of τ into the real numbers R such that

(i) $E(x) \geqq 0$ for every $x \in \tau$,
$E(u) = 1$;

(ii) $x \leftrightarrow y$ implies $E(x) + E(y) = E(x \cup y) + E(x \cap y)$;

(iii) $E(x) = E(y) = 1$ implies $E(x \cap y) = 1$.

The second property is also supposed to be true for any countable set of compatibles x_i. The concept of state is thus a generalization of the notion of probability in the following sense: if C is a family of compatible propositions a state on τ generates a probability on C.

If for a given state E there exists a point P such that $E(P) = 1$, we say that E is a *pure state*; its value $E(Q)$ for another point Q is called the transition probability from P to Q.

Let μ be a symmetry and let us define a new mapping E^μ of τ into R by $E^\mu(a) \equiv E(\mu^{-1}a)$. This is also a state on τ. Moreover, if E is a pure state, so is E^μ and conversely. The transition probability is preserved. These properties are often taken as the definition of a symmetry when τ is realized as the set of all projections of a Hilbert space. The following theorem shows that this definition is, in this case, equivalent to ours.

Theorem. Let τ be the lattice of projections in a Hilbert space and let μ be a one-to-one mapping of τ onto itself, such that for every state E, $E^\mu(a) \equiv E(\mu^{-1}a)$ and $E^{\mu^{-1}}(a) \equiv E(\mu a)$ are also states on τ, then μ is a symmetry in our sense.

Proof: First of all, μ is one-to-one by definition. Since the meaning of $a \subseteq b$ is "if $E(a) = 1$ then $E(b) = 1$" we conclude that $\mu a \subseteq \mu b$ if $a \subseteq b$ and conversely. For every E we have $E^\mu(\mathcal{O}\mu a) = 1 - E^\mu(\mu a) = 1 - E(a) = E(\mathcal{O}a) = E^\mu(\mu\mathcal{O}a)$ and we can conclude $\mathcal{O}\mu a = \mu\mathcal{O}a$. This proves the theorem.

ACKNOWLEDGMENT

We wish to express our very deep gratitude to Professor J. M. Jauch for his interest and invaluable aid during the development of this investigation.

APPENDIX

We saw that every morphism of an irreducible proposition system (with $n \geqq 4$) is induced by a semilinear transformation. The converse is also true if we add the conditions:

(i) $Si_1\mathcal{O}_1P_1 = i_2\mathcal{O}_2i_2^{-1}Si_1P_1$ for every $P_1 \in \tau_1$;

(ii) $S^{-1}i_2\mathcal{O}_2P_2 = i_1\mathcal{O}_1i_1^{-1}S^{-1}i_2P_2$ for every $P_2 \in \tau_2$,

where i_1 and i_2 are the embedding of τ_1 and τ_2 in their respective vector spaces V_1 and V_2.

We have to prove that any semilinear mapping of V_1 onto V_2 such that (i) and (ii) are verified, induces a one-to-one mapping of τ_1 onto τ_2 which satisfies the two properties of a morphism. $\mu \equiv i_2^{-1}Si_1$ is clearly a one-to-one mapping of the points of τ_1 onto those of τ_2. The only apparent difficulty is to show that for any x in τ_1, Si_1x is an element of $i_2(\tau_2)$. For any x in τ_1 we have: $x = \bigcap \mathcal{O}_1P_j$, where the intersection is to be understood on the set of all the P_j such that $x \subseteq \mathcal{O}_1P_j$. With this notation, $Si_1x = \bigcap Si_1\mathcal{O}_1P_j$.

However, from (i) this is $\bigcap i_2\mathcal{O}_2i_2^{-1}Si_1P_j$; we know that i_2 preserves the intersection so that we can draw out i_2 in this expression. This suffices to prove that Si_1x belongs to $i_2(\tau_2)$, and we can write

$$i_2^{-1}Si_1x = \bigcap \mathcal{O}_2i_2^{-1}Si_1P_j,$$

so that $\mu \equiv i_2^{-1}Si_1$ is a mapping defined uniquely on every proposition of τ_1. We would prove by analogous arguments that $\mu^{-1} \equiv i_1^{-1}S^{-1}i_2$ is a mapping defined uniquely on every proposition of τ_2. μ is then a one-to-one mapping of τ_1 onto τ_2.

From the explicit form of μ given above it follows directly that $a \subseteq b$ implies $\mu a \subseteq \mu b$ and conversely.

Moreover, this expression of μx is equivalent to

[7] E. P. Wigner, *Group Theory* (Academic Press Inc., New York, 1959), Appendix to Chapter 20.
[8] R. Hagedorn, Nuovo Cimento Suppl. **12**, 73 (1959).
[9] J. S. Lomont and P. Mendelson, "The Wigner Unitarity–Antiunitarity Theorem." (unpublished); D. Finkelstein, J. M. Jauch, S. Schiminovich, and D. Speiser, J. Math. Phys. **3**, 207 (1962).

$\mathcal{O}_2 \bigcup i_2^{-1} S i_1 P_i$ where the union is to be understood on every P_i such that $P_i \subseteq \mathcal{O}_1 x$, i.e., $\mu P_i \subseteq \mu \mathcal{O}_1 x$; then $\mu x = \mathcal{O}_2 \mu \mathcal{O}_1 x$.

Q.E.D.

If the dimensions of τ_1 and τ_2 are finite, i_1 and i_2 are one-to-one mappings and the conditions (i) and (ii) become equivalent. This is not surprising if we observe that these conditions are $SS^+ = S^+S = I$, when we take the realizations of τ_1 and τ_2 by Hilbert spaces.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 4 APRIL 1963

N-Body Bose System with a Finite Number of States. I. Irreducible Representations

JOSEPH B. AVILES, JR.
U. S. Naval Research Laboratory, Washington, D. C.
(Received 31 August 1962)

For the purpose of investigating the N-Body Bose System with a finite number of states, we analyze $X_{ij} = A_i^\dagger A_j$, $i, j \leq K$. (A and $A^\dagger$ are the usual annihilation and creation operators of the second quantization formalism). The Hamiltonian for a fixed number of particles may be expressed in terms of the (finite-dimensional) irreducible representations of the X_{ij}. A set of fundamental equations is defined for the irreducible representations $X_{ij}(N, K)$, $i, j \leq K$ and is solved for arbitrary N and K. The analysis of the structure of the $X_{ij}(N, K)$ yields a simple, systematic method for listing all possible ways in which $n_1 + \ldots n_i + \ldots + n_K = N$ may be satisfied for n_i positive integral or zero. This leads to a particularly simple method for constructing the $X_{ij}(N, K)$, $i, j \leq K$.

INTRODUCTION

THE Hamiltonian for a system of identical Bosons interacting through two-body forces is:

$$H = \sum H_{ik}^{(1)} A_i^\dagger A_k + \tfrac{1}{2} \sum V_{ijkl}^{(2)} A_i^\dagger A_j^\dagger A_k A_l, \quad (1)$$

where

$$[A_i, A_k^\dagger] = \delta_{ik},$$
$$[A_i^\dagger, A_k^\dagger] = [A_i, A_k] = 0. \quad (2)$$

The sums in (1) are over a complete orthonormal set of single-particle (s.p.) states which are, of course, infinite in number.

The problem of the general diagonalization of (1) is most likely insuperable. The problem may be considered solved for the cases of any number of non-interacting particles in an external field and two interacting particles not in an external field. In both these cases, the problem reduces to finding the eigenvalues of a single particle in an external field. Solution has also been achieved[1] for the case of a gas at low density and temperature (degenerate Bose gas). For such a situation, (1) reduces effectively to a bilinear form. For other cases, (e.g. the Three-Body Problem, liquid He^4) one naturally expects the biquadratic nature of the interaction term to enter in an essential way.

If we now consider the fictitious case in which the sums in (1) are finite, a considerable simplification is obtained; for then the Hamiltonian is representable by finite matrices. We might hope that by studying this simpler situation, we may learn how to handle the biquadratic term more adroitly, and consequently, be able to introduce an approximation which does not destroy its essential nature.

In this communication, we initiate such a study by analyzing the matrix structure of

$$X_{ij} = A_i^\dagger A_j. \quad (3)$$

In particular, the analysis will yield a systematic and rather simple method by which we can construct (i.e. display in matrix form) the irreducible representations of X_{ij} for a system of N particles and K s.p. states.

The reason for taking X_{ij} for the basis of our study is that it commutes with the total number operator defined by

$$\mathfrak{N} = \sum_{i=1}^{K} X_{ii}. \quad (4)$$

In a representation, in which $\mathfrak{N}$ is diagonal, com-

[1] N. N. Bogoliubov, J. Phys. (USSR) 11, 23 (1947).

$\mathcal{O}_2 \bigcup i_2^{-1} S i_1 P_i$ where the union is to be understood on every P_i such that $P_i \subseteq \mathcal{O}_1 x$, i.e., $\mu P_i \subseteq \mu \mathcal{O}_1 x$; then $\mu x = \mathcal{O}_2 \mu \mathcal{O}_1 x$.

Q.E.D.

If the dimensions of τ_1 and τ_2 are finite, i_1 and i_2 are one-to-one mappings and the conditions (i) and (ii) become equivalent. This is not surprising if we observe that these conditions are $SS^+ = S^+S = I$, when we take the realizations of τ_1 and τ_2 by Hilbert spaces.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 4 APRIL 1963

N-Body Bose System with a Finite Number of States. I. Irreducible Representations

JOSEPH B. AVILES, JR.
U. S. Naval Research Laboratory, Washington, D. C.
(Received 31 August 1962)

For the purpose of investigating the N-Body Bose System with a finite number of states, we analyze $X_{ij} = A_i^\dagger A_j$, $i, j \leq K$. (A and $A^\dagger$ are the usual annihilation and creation operators of the second quantization formalism). The Hamiltonian for a fixed number of particles may be expressed in terms of the (finite-dimensional) irreducible representations of the X_{ij}. A set of fundamental equations is defined for the irreducible representations $X_{ij}(N, K)$, $i, j \leq K$ and is solved for arbitrary N and K. The analysis of the structure of the $X_{ij}(N, K)$ yields a simple, systematic method for listing all possible ways in which $n_1 + \ldots n_i + \ldots + n_K = N$ may be satisfied for n_i positive integral or zero. This leads to a particularly simple method for constructing the $X_{ij}(N, K)$, $i, j \leq K$.

INTRODUCTION

THE Hamiltonian for a system of identical Bosons interacting through two-body forces is:

$$H = \sum H_{ik}^{(1)} A_i^\dagger A_k + \tfrac{1}{2} \sum V_{ijkl}^{(2)} A_i^\dagger A_j^\dagger A_k A_l, \quad (1)$$

where

$$[A_i, A_k^\dagger] = \delta_{ik},$$
$$[A_i^\dagger, A_k^\dagger] = [A_i, A_k] = 0. \quad (2)$$

The sums in (1) are over a complete orthonormal set of single-particle (s.p.) states which are, of course, infinite in number.

The problem of the general diagonalization of (1) is most likely insuperable. The problem may be considered solved for the cases of any number of non-interacting particles in an external field and two interacting particles not in an external field. In both these cases, the problem reduces to finding the eigenvalues of a single particle in an external field. Solution has also been achieved[1] for the case of a gas at low density and temperature (degenerate Bose gas). For such a situation, (1) reduces effectively to a bilinear form. For other cases, (e.g. the Three-Body Problem, liquid He^4) one naturally expects the biquadratic nature of the interaction term to enter in an essential way.

If we now consider the fictitious case in which the sums in (1) are finite, a considerable simplification is obtained; for then the Hamiltonian is representable by finite matrices. We might hope that by studying this simpler situation, we may learn how to handle the biquadratic term more adroitly, and consequently, be able to introduce an approximation which does not destroy its essential nature.

In this communication, we initiate such a study by analyzing the matrix structure of

$$X_{ij} = A_i^\dagger A_j. \quad (3)$$

In particular, the analysis will yield a systematic and rather simple method by which we can construct (i.e. display in matrix form) the irreducible representations of X_{ij} for a system of N particles and K s.p. states.

The reason for taking X_{ij} for the basis of our study is that it commutes with the total number operator defined by

$$\mathfrak{N} = \sum_{i=1}^{K} X_{ii}. \quad (4)$$

In a representation, in which $\mathfrak{N}$ is diagonal, com-

[1] N. N. Bogoliubov, J. Phys. (USSR) 11, 23 (1947).

mutation of $\mathfrak{N}$ and X implies that

$$\langle N' | X | N \rangle = X_{N'N}\delta_{N'N},$$

i.e., the matrix for X has the following reduced structure:

$$X = \begin{pmatrix} \ddots & & \\ & X(N_1, K) & \\ & & X(N_2, K) \\ & & & \ddots \end{pmatrix}, \tag{5}$$

in which $X(N_i, K)$ is a matrix depending on a single eigenvalue of $\mathfrak{N}$. When both N and K are finite, the dimensions of the irreducible representation $X(N_i, K)$ of X will also be finite. (The fact that the structure of these representations of X depend also on the number of s.p. states has been indicated in the notation). On the other hand, the operators A and $A^\dagger$ do not commute with $\mathfrak{N}$. Since the reduced structure (5) does not exist for them, an infinite representation is required whether or not K is finite.

The Hamiltonian (1) commutes with $\mathfrak{N}$ and therefore has the structure (5). By using (2), we may write

$$H = \sum H'_{ik}A_i^\dagger A_k + \tfrac{1}{2}\sum V^{(2)}_{ijkl}A_i^\dagger A_k A_j^\dagger A_l,$$

where

$$H'_{ik} = H^{(1)}_{ik} - \tfrac{1}{2}\sum_{m,n} V^{(2)}_{imnk}. \tag{6}$$

The representation of H for given N and K can now be expressed in terms of the representation of the X_{ij} as follows:

$$H(N, K) = \sum H'_{ik}X_{ik}(N, K) + \tfrac{1}{2}\sum V^{(2)}_{ijkl}X_{ik}(N, K)X_{jl}(N, K). \tag{7}$$

We thus see that the irreducible representations of X_{ij} are natural elements for the study of systems with a definite number of particles and a finite number of states.

I. FUNDAMENTAL RELATIONS

By using the well-known properties of $A_i^\dagger$ and A_i, we can obtain the matrix elements for $X_{ij}(=A_i^\dagger A_j)$; these are

$$\langle n'_1, n'_2 \cdots n'_K | A_i^\dagger A_j | n_1, n_2, \cdots n_K \rangle = \begin{cases} [(n_i+1)n_j]^{\frac{1}{2}}\delta^{n_1'}_{n_1}\cdots\delta^{n_i'}_{n_i+1}\quad\delta^{n_j'}_{n_j-1}\cdots\delta^{n_K'}_{n_K} & i \neq j \\ n_i\delta^{n_1'}_{n_1}\cdots\delta^{n_j'}_{n_j}\cdots\delta^{n_K'}_{n_K} & i = j \end{cases} \tag{8}$$

where n_i, the eigenvalues of $N_i(=A_i^\dagger A_i)$, are positive integers or zero, and are restricted by the following condition:

$$\sum_{i=1}^{K} n_i = N. \tag{9}$$

The actual construction of a set of matrices for the X_{ij} for given N and K from (8) and (9) is a complex task for which one must introduce a high degree of systemization. A very natural systemization presents itself once we have analyzed the structure and interrelations of the X_{ij}.

The analysis of the X_{ij} proceeds most effectively from a set of fundamental relations which (a) involves only the X_{ij}, $i, j, \leq K$, and (b) defines them uniquely (to within a unitary transformation). This set of relations is:

$$[X_{ij}, X_{kl}] = X_{il}\delta_{kj} - X_{kj}\delta_{il}, \tag{10'}$$

$$X_{ij}X_{ji} = X_{ii}(X_{jj} + 1), \quad (i \neq j), \tag{10''}$$

$$X_{ij}^\dagger = X_{ji}. \tag{10'''}$$

The main virtue of basing the analysis on (10) is that, in view of (5), these relations are satisfied by the representations of X_{ij} for arbitrary N and K. Our problem will then be to find those representations which, for given N and K, have the smallest number of dimensions, i.e. the irreducible representations. In the future, we shall always reserve the notation $X_{ij}(N, K)$ for irreducible representations.

We shall not assume prior knowledge of the properties of X_{ij}; instead we shall base the analysis solely on the fundamental relations (10). This manner of proceeding, in addition to its methodological interest, lends a strong coherence to the development.

By stating three relations which the X_{ij} must satisfy, we of course imply that the commutation relations (10') are insufficient to define a set of X_{ij}. This becomes apparent if we consider the analagous relations for the fermion operators defined by

$$Y_{ij} = C_i^\dagger C_j, \tag{11}$$

where

$$\left.\begin{aligned} C_i^\dagger C_j + C_j C_i^\dagger &= \delta_{ij} \\ C_i C_j + C_j C_i &= C_i^\dagger C_j^\dagger + C_j^\dagger C_i^\dagger = 0 \end{aligned}\right\}. \tag{12}$$

It follows that

$$[Y_{ij}, Y_{kl}] = Y_{il}\delta_{kj} - Y_{kj}\delta_{il}, \tag{13}$$

which is identical to the boson relation (10')! It is clear that if there is to be a difference between bosons and fermions, then the commutation relations (10') cannot stand alone. The necessity of (10'') will be demonstrated in the next section. Analagous to

(10″), we have for fermions:

$$Y_{ij}Y_{jl} = Y_{il}(1 - Y_{jl}) \qquad i \neq j, \tag{14}$$

which differs essentially from (10″). (A necessary additional relation for fermions is $Y_{ii}^2 = Y_{ii}$. We shall not analyze fermions any further in the present paper.)

The relation (10‴) was introduced in order that the Hamiltonian (7) be Hermitian. The relations (10′) and (10″) alone actually define more general solutions for X_{ij}; however, we have no need for these.

II. OPERATORS FOR A PAIR OF S. P. STATES

We shall now focus our attention on the operators associated with a pair of arbitrary s.p. states, which for convenience we label 1 and 2. These are

$$N_1, N_2, X_{12}, X_{21}. \tag{15}$$

Their fundamental relations are obtained from (10) by limiting the indices to $i, j \leq 2$:

$$\left.\begin{aligned} [N_1, X_{12}] &= X_{12} & [N_1, X_{21}] &= -X_{21} \\ [N_2, X_{12}] &= -X_{12} & [N_2, X_{21}] &= X_{21} \\ [X_{12}, X_{21}] &= N_1 - N_2 \end{aligned}\right\}. \tag{10′a}$$

$$X_{12}X_{21} = N_1(N_2 + 1), \tag{10″a}$$

$$X_{12} = X_{21}^{\dagger}. \tag{10‴a}$$

The properties of the set (15) are most easily deduced by considering the set of Hermitian operators

$$L_0, L_i, \qquad i = x, y, z, \tag{16}$$

which are defined by the linear transformation

$$\left.\begin{aligned} L_0 &= \tfrac{1}{2}(N_1 + N_2) \\ L_x &= \tfrac{1}{2}(N_1 - N_2) \\ L_y &= \tfrac{1}{2}(X_{12} + X_{21}) \\ L_z &= -\tfrac{1}{2}i(X_{12} - X_{21}) \end{aligned}\right\}. \tag{17}$$

The set (15) is then expressed in terms of (16) by

$$\left.\begin{aligned} N_1 &= L_0 + L_x \\ N_2 &= L_0 - L_x \\ X_{12} &= L_y + iL_z \\ X_{21} &= L_y - iL_z \end{aligned}\right\}. \tag{18}$$

The commutation relations of the set (16) follow from (10′a); they are

$$[L_0, L_i] = 0 \qquad i = x, y, z, \tag{19}$$

$$\left.\begin{aligned} [L_x, L_y] &= iL_z \\ [L_y, L_z] &= iL_x \\ [L_z, L_y] &= iL_y \end{aligned}\right\} \tag{20}$$

As the notation has anticipated, (20) is just the commutation relations for components of angular momentum. The properties of the L_i, $i = x, y, z$ are therefore well known. In particular, we know that the irreducible representations of the L_i may be deduced from (20), and that they are associated with a particular (integral or half-integral) eigenvalue l of the magnitude of the total angular momentum

$$\mathfrak{L}^2 = L_x^2 + L_y^2 + L_z^2. \tag{21}$$

The irreducible representations of the L_i have the dimensions $2l + 1$. We shall denote a matrix with these dimensions by $X^{(l)}$.

The operator L_0 is not wholly defined by (19). All that we may say is that for an *irreducible* set of $L_i^{(l)}$, (19) is satisfied only by a matrix $L_0^{(l)}$ of the form

$$\lambda_0 1^{(l)},$$

where $1^{(l)}$ is the unit matrix and λ_0 is arbitrary.

This lack of complete definition of the set (16) [and thus the set (15)] is in accordance with the fact that we have so far only used (10′a) and (10‴a) which are the same for bosons and fermions. Definition of L_0 is completed by first expressing $\mathfrak{L}^2$ in terms of the set (15), viz:

$$\mathfrak{L}^2 = \tfrac{1}{4}[N_1^2 - 2N_1N_2 + N_2^2 + 2(X_{12}X_{21} + X_{21}X_{12})], \tag{22}$$

and then using (10′a) to obtain finally

$$\mathfrak{L}^2 = L_0(L_0 + 1). \tag{23}$$

This equation implicitly defines L_0 in terms of $\mathfrak{L}^2$. [For fermions, we use (14) and obtain $\mathfrak{L}^2 = L_0(L_0 - 1)$.] Since the eigenvalues of $\mathfrak{L}^2$ are $l(l + 1)$, we have

$$l(l + 1) = \lambda_0(\lambda_0 + 1), \tag{24}$$

which has the solutions

$$\lambda_0 = l, -(l + 1). \tag{25}$$

We shall retain only the solution $\lambda_0 = l$ since it implies that the eigenvalues of N_1 and N_2 are positive integers or zero, while the second solution does not.[2]

We are now in a position to find the representations of the set (15) which correspond to the lth irreducible representations of set (16). It is convenient to introduce the integral quantity $\sigma = 2l$ which is an eigenvalue of $N_1 + N_2$.

If we want N_1 and N_2 to be diagonal, we must

[2] The second solution could be used if occupation numbers are defined as $-(n_i + 1)$.

take L_z diagonal in addition to L_0. The eigenvalues of $L_z^{(l)}$ are:

$$m = \tfrac{1}{2}\sigma, \tfrac{1}{2}\sigma - 1, \cdots -\tfrac{1}{2}\sigma + 1, -\tfrac{1}{2}\sigma.$$

If we use these eigenvalues to label the rows and columns of the representation of the L_i, we have the standard formulas

$$\left.\begin{aligned}
&(L_0^{(l)})_{m'm} = \tfrac{1}{2}\sigma\delta_{m'm} \qquad (L_z^{(l)})_{m'm} = m\delta_{m'm} \\
&(L_y^{(l)} + lL_Z^{(l)})_{m'm} \\
&\quad = [(\tfrac{1}{2}\sigma + m + 1)(\tfrac{1}{2}\sigma - m)]^{\frac{1}{2}}\delta_{m',m+1} \\
&(L_y^{(l)} - iL_Z^{(l)})_{m'm} \\
&\quad = [(\tfrac{1}{2}\sigma + m' + 1)(\tfrac{1}{2}\sigma - m')]^{\frac{1}{2}}\delta_{m,m'+1}
\end{aligned}\right\}. \tag{26}$$

From (18), the eigenvalues of $N_1^{(l)}$ and $N_2^{(l)}$ are

$$\begin{aligned}
n_1 &= \tfrac{1}{2}\sigma + m = \sigma, \sigma - 1, \cdots 1, 0, \\
n_2 &= \tfrac{1}{2}\sigma - m = 0, 1, \cdots \sigma - 1, \sigma.
\end{aligned} \tag{27}$$

They are positive integers or zero. The rows and columns of the representations of the set (15) are conveniently labeled by n_2. A simple transformation yields

$$\left.\begin{aligned}
(N_1^{(l)})_{n_2'n_2} &= (\sigma - n_2)\delta_{n_2'n_2} \\
(N_2^{(l)})_{n_2'n_2} &= n_2\delta_{n_2'n_2} \\
(X_{12}^{(l)})_{n_2'n_2} &= [(\sigma - n_2 + 1)n_2]^{\frac{1}{2}}\delta_{n_2',n_2-1} \\
(X_{21}^{(l)})_{n_2'n_2} &= [(\sigma - n_2' + 1)n_2']^{\frac{1}{2}}\delta_{n_2,n_2'-1}
\end{aligned}\right\}. \tag{28}$$

Owing to the linear relation between the sets (15) and (16), the representation (28) is irreducible.

As yet, we have not mentioned any limitation on the number of s.p. states. What we have done is to see which consequences followed solely from considering the relations existing among the four operators associated with an arbitrary pair of s.p. states. However, if we consider a system which has only two states, there is little more to be said since we have found the irreducible representations of all possible operators. For such a system $\mathfrak{N} = N_1 + N_2$ and we therefore identify σ with N. Since we have agreed to use the notation $X_{ij}(N, K)$ for irreducible representations of X_{ij}, we have the following identity:

$$X_{ij}^{(l)} = X_{ij}(\sigma, 2) \qquad i, j \leq 2. \tag{29}$$

We shall now look at the operators

$$N_1, N_2, X_{12}, X_{21} \tag{15}$$

from the standpoint that they are a subset of the irreducible representation

$$X_{ij}(N, K) \qquad i, j \leq K. \tag{30}$$

Since the set (15) and the irreducible set (29) for two s.p. states satisfy the same fundamental relations, we can write

$$X_{ij}(N, K) \quad \begin{vmatrix} \ddots & & \\ & X_{ij}(\sigma_m, 2) & \\ & & X_{ij}(\sigma_q, 2) \\ & & \ddots \end{vmatrix}, \tag{31}$$
$$i, j \leq 2$$

where $\cdots \sigma_m, \sigma_q \cdots$ are particular values of σ. In other words, $X_{ij}(N, K)$, $i, j \leq 2$, is now looked upon as a *reducible* representation of X_{ij}, $i, j \leq 2$. Just how many times a particular value of σ appears in (31) will be determined in Sec. VI.

III. GENERAL PROPERTIES

At this point, we shall deduce certain general properties of the set (30). First, we shall emphasize two simple but important properties: (1) All the various irreducible representations of $X_{ij}(N, K)$ for given N and K, are connected by unitary similarity (i.e. canonical) transformations; (2) Given one irreducible representation, others may be obtained by interchanging the s.p. indices of the operators.

Suppose now we have a representation A. According to the second property, we can obtain new representations $A', A'', \cdots$ in which a given number operator $N_i(N, K)$ has any of the forms of the number operators in representation A. From the first property, we may then infer that all the number operators of a given irreducible representation are connected by unitary similarity transformations. Since a similarity transformation does not alter the eigenvalues of a matrix, we can conclude that all the number operators $N_i(N, K)$, $i \leq K$, have the same eigenvalues with the same multiplicity.

By the same reasoning, we can infer that all $X_{ij}(N, K)$, $i \neq j$, are unitarily similar. However, these operators may not be transformed into number operators. To see this, it suffices to note that $X_{12}(N, K)$, according to (31) and (28), is non-Hermitian, while the number operators are Hermitian. These properties are unaltered by a unitary transformation. We thus see that the set (30) is naturally comprised of two similarity classes.

We can now draw the conclusion that the eigenvalues of $N_i(N, K)$ are positive integers or zero. Furthermore, the restriction

$$\sum_{i=1}^{K} n_i = N \tag{32}$$

implies that N is also a positive integer or zero and

that $n_i \leq N$. It also follows that $n_1 + n_2 = \sigma \leq N$; i.e. the values of σ in (31) are bounded.

IV. REPRESENTATION OF THE NUMBER OPERATORS

The set of commuting Hermitian number operators $N_i(N, K)$, $i \leq K$, will be constructed according to the familiar principle that (32) is satisfied in all possible ways just once. Since we shall always take $N_i(N, K)$, $i \leq K$, diagonal, it will be convenient to list the eigenvalues rather than to display them in matrix form. We therefore define the occupation scheme $O(N, K)$ as a listing of all possible (different) ways in which (32) can be satisfied. Figure 1 shows the general form for such a scheme. For any column in $O(N, K)$ the condition (32) is satisfied and no column is repeated. The construction of the matrices $N_i(N, K)$ from $O(N, K)$ is obvious. The dimensions of these matrices is clearly the same as the number of ways in which N identical particles can be distributed among K states. The answer to this familiar problem is

$$B(N, K) = (N + K - 1)!/N!(K - 1)!. \tag{33}$$

The reason for constructing the $N_i(N, K)$ in the above described fashion is physically obvious if we consider a system of free bosons. We then expect that all possible physical states which can be realized correspond to all possible ways in which N bosons may be found distributed among K free-particle states. In our development, the necessity for this choice will appear when we prove that the set $X_{ij}(N, K)$, $i, j \leq K$, in which $N_i(N, K)$ are constructed as defined above, is irreducible. The irreducibility will be seen (in Sec. IX) to be connected with the similarity of the $N_i(N, K)$, $i \leq K$. We saw in the last section that this is a necessary property.

It will now be shown that all the $N_i(N, K)$ constructed from $O(N, K)$ have the same eigenvalues with the same multiplicity; i.e. they are similar. Let us find the number of times a particular eigenvalue n^ρ appears in (say) the ith *row* of $O(N, K)$. Now since all possible ways of satisfying (32) are included in $O(N, K)$, this amounts to fixing n^ρ particles in the ith s.p. state and asking in how many different ways can the remaining $N - n^\rho$ particles be distributed among the remaining $K - 1$ s.p. states. The answer is

$$B(N - n^\rho, K - 1), \tag{34}$$

which is therefore the multiplicity of the eigenvalue n^ρ. It is clearly independent of which s.p. state we consider.

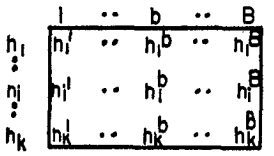


FIG. 1. General form of occupation scheme $O(N, K)$.

We shall return to the actual construction of the occupation scheme for given N and K in Sec. VII.

V. OPERATORS FOR K′ SINGLE-PARTICLE STATES (K′ < K)

We shall now see what conclusions may be drawn concerning the structure of $X_{ij}(N, K)$ by considering only the subset of operators associated with K' s.p. states when $K' < K$. For convenience, we shall label the subset of K' s.p. states in the order 1, 2, $\cdots$ K'. That is, we investigate

$$X_{ij}(N, K), \qquad i, j \leq K' < K, \tag{35}$$

which is a subset of the irreducible set

$$X_{ij}(N, K), \qquad i, j \leq K. \tag{30}$$

The fundamental relations satisfied by (35) are clearly the same as those satisfied by the irreducible representations

$$X_{ij}(\sigma, K'), \qquad i, j \leq K', \tag{36}$$

for which

$$\sum_{i=1}^{K'} N_i(\sigma, K') = \sigma 1(\sigma, K'). \tag{37}$$

Consequently, there exists a representation in which the set (35) has the form

$$X_{ij}(N, K) \quad i, j \leq K' \quad \begin{bmatrix} \ddots & & \\ & X_{ij}(\sigma_m, K') & \\ & & X_{ij}(\sigma_n, K') \\ & & \ddots \end{bmatrix}, \tag{38}$$

in which $0 \leq \sigma \leq N$. That is, we now look upon $X_{ij}(N, K)$, $i, j \leq K'$ as a reducible representation of the operators X_{ij}, $i, j \leq K'$.

We shall now find the number of times in which $X_{ij}(\sigma, K')$, for given σ, appears in (38). Since we need only decide this for number operators, we can make use of the occupation scheme for our analysis. The $N_i(\sigma, K')$ must be constructed in accord with $O(\sigma, K')$ and $N_i(N, K)$ in accord with $O(N, K)$; therefore (38) implies that $O(\sigma, K')$ is contained in $O(N, K)$ possibly more than once. Since a particular arrangement of integers for which $n_1 + \cdots n_{K'} = \sigma$

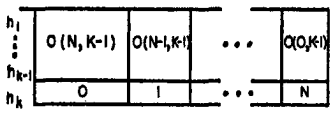


FIG. 2. Standard occupation scheme $O(N, K)$ corresponding to Eqs. (41) and (45).

is satisfied appears in $O(\sigma, K')$ just once, the number of times in which this same arrangement appears in $O(N, K)$ (for the first K' s.p. states) is the multiplicity of $O(\sigma, K')$. Thus we fix σ particles in the first K' s.p. states and ask in how many ways can the remaining $N - \sigma$ particles be distributed among the remaining $K - K'$ s.p. states. The answer is clearly

$$B(N - \sigma, K - K'). \tag{39}$$

This is the multiplicity of $X_{ij}(\sigma, K')$ for given N and K.

We now make the following fundamental observation: The number of times in which $X_{ij}(\sigma, K - 1)$ appears in $X_{ij}(N, K)$ $i, j \leq K - 1$ is simply unity; for we have $K - K' = 1$ and $B(N - \sigma, 1) = 1$ independently of σ and N. This allows us to write:

$$X_{ij}(N, K) \quad i, j \leq K - 1 \begin{bmatrix} X_{ij}(N, K-1) & & & \\ & X_{ij}(N-1, K-1) & & \\ & & \ddots & \\ & & & X_{ij}(1, K-1) \\ & & & & X_{ij}(0, K-1) \end{bmatrix}. \tag{40}$$

This particularly simple result will be quite useful in further considerations.

VI. REDUCTION HIERARCHY; STANDARD REPRESENTATION

If we consider a representation in which the subset of operators $X_{ij}(N, K)$, $i, j \leq K - 1$ have the reduced form shown in (40), it is clear that all of the remaining operators (i.e. those associated with the Kth s.p. state) cannot assume this form; for then we would have a reduced form for the set $X_{ij}(N, K)$, $i, j \leq K$, which is contrary to the assumed irreducibility.

The process of reduction can, however, be continued for operators associated with fewer and fewer s.p. states. The next step would be to express $X_{ij}(N, K)$, $i, j \leq K - 2$, in terms of $X_{ij}(\sigma, K - 2)$, $i, j \leq K - 2$; these latter appear $B(N - \sigma, 2)$ times. By reasoning along the same lines as in the preceding paragraph, we can conclude that once this is done, the remaining operators (associated with s.p. states K, $K - 1$) cannot be similarly reduced. Continuing this process, we eventually arrive at $X_{ij}(N, K)$, $i, j \leq 2$ whose reduced form contains only $X_{ij}(\sigma, 2)$, $i, j \leq 2$; these latter appear $B(N - \sigma, K - 2)$ times.

We now take the representation obtained by the above process as the standard representation. Thus the operators associated with the s.p. states 1 and 2 will exhibit maximum reduction while those associated sucessively with indices no greater than 3, 4, $\cdots K$ will show a decreasing amount of reduction. Furthermore, whenever we express $X_{ij}(\nu, K')$ in terms of $X_{ij}(\sigma, K' - 1)$, we shall arrange these latter matrices, which appear once for each value of σ, such that they appear down the diagonal in the order $\sigma = \nu, \nu - 1, \cdots 1, 0$.

VII. THE STANDARD OCCUPATION SCHEME

The occupation scheme is of course, just a listing of the eigenvalues of the $N_i(N, K)$, $i \leq K$. We shall do this according to the standard representation defined in the preceding section. Let us first consider the number operators of the first $K - 1$ s.p. states; according to (40) we have

$$N_i(N, K) \quad i \leq K - 1 \begin{bmatrix} N_i(N, K-1) & & & \\ & N_i(N-1, K-1) & & \\ & & \ddots & \\ & & & N_i(1, K-1) \\ & & & & N_i(0, K-1) \end{bmatrix}. \tag{41}$$

The matrix $N_K(N, K)$ can be obtained by noting that since

$$\sum_{i=1}^{K-1} N_i(\sigma, K - 1) = \sigma 1(\sigma, K - 1), \tag{42}$$

the region in $N_K(N, K)$ corresponding to $N_i(\sigma, K - 1)$, $i \leq K - 1$, must have the form

$$(N - \sigma)1(\sigma, K - 1) \tag{43}$$

For only then can we satisfy:

$$\sum_{i=1}^{K} n_i = \sigma + n_K = N. \tag{44}$$

We may therefore write

$$N_K = \begin{bmatrix} 0 \cdot 1(N, K-1) & & & \\ & 1 \cdot 1(N-1, K-1) & & \\ & & \ddots & \\ & & & N \cdot 1(0, K-1) \end{bmatrix}. \tag{45}$$

The occupation scheme corresponding to the above structure is given in Fig. 2, in which $O(\sigma, K - 1)$

is the occupation scheme corresponding to the set $N_i(\sigma, K-1)$, $i \leq K-1$,

In view of our discussion on the reduction hierarchy, the occupation scheme shown in Fig. 2 can be used iteratively (with $N, K \to \sigma, K-1 \to \cdots$) for the purpose of obtaining an explicit numerical occupation scheme. We would eventually have a scheme in which the values of $n_K, \cdots n_3$ are explicit while the remainder of the scheme (for n_1 and n_2) is in terms of $O(\sigma', 2)$ which appears $B(N-\sigma', 2)$ times. We have previously found $O(\sigma', 2)$ [see (27)]; it is given in Fig. 3.

We shall now present a convenient method by which we can construct a standard occupation scheme: We start with $O(N, 2)$. By adding on $O(N-1, 2), \cdots O(0, 2)$ and the scheme for n_3 according to Fig. 2 with $K=3$, we obtain $O(N, 3)$. We then add on $O(N-1, 3), \cdots O(0, 3)$ and the scheme for n_4 again according to Fig. 2, but for $K=4$ and thus obtain $O(N, 4)$. By continuing in this fashion, we can build up the standard occupation scheme for arbitrary K.

The general scheme shown in Fig. 4 will illustrate a point which facilitates construction even more: Let us assume that we have completed construction of $O(N, K-1)$, which in Fig. 4 is the region subtended by $n_K = 0$, and now wish to make the extension for $O(N, K)$. We notice that the scheme for $n_1, \cdots n_{K-2}$ for the successive regions subtended by $n_K = 1, 2, \cdots$ is obtained from the first region ($n_K = 0$) by successively dropping off $O(N, K-2)$, $O(N-1, K-2), \cdots$. Also, the corresponding n_{K-1} are obtained from the first region ($n_K = 0$) by successively dropping off $N, N-1, \cdots$. With this final observation, we have reduced the construction of the occupation scheme to a series of simple, automatic operations for which, according to the whole of the preceding analysis, we are guaranteed that we have included all possible ways of satisfying $n_1 + \cdots + n_K = N$ just once. The occupation scheme for $N=2$, $K=4$ is shown at the bottom of Fig. 5.

VIII. EIGENFUNCTIONS OF $N_i(N, K)$; LABELING

The Hermitian matrices $N_i(N, K)$ have $B(N, K)$ independent eigenfunctions. Owing to the degeneracy of each $N_i(N, K)$, any one of them is incapable of defining a unique basis for the $B(N, K)$ dimensional space in which we represent the set X_{ij}, $i, j \leq K$. However, eigenfunctions Ψ which are common to

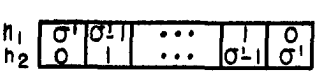


FIG. 3. Standard occupation scheme $O(\sigma, 2)$.

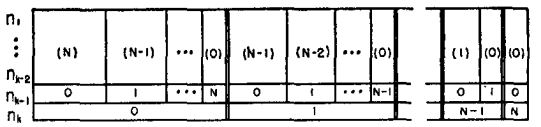


FIG. 4. General form of standard occupation scheme $O(N, K)$. The region occupied by $O(\nu, K\text{-}2)$ is indicated by (ν).

the whole set $N_i(N, K)$, $i \leq K$, form an orthogonal basis for the space. This may be shown as follows: The Hermitian property of $N_i(N, K)$ implies that its eigenfunctions satisfy the condition:

$$0 = (\Psi', N_i(N, K)\Psi) - (\Psi, N_i(N, K)\Psi'). \quad (46)$$

Since hermiticity also implies $(\Psi, \Psi') = (\Psi', \Psi)$, we have

$$0 = (n_i - n_i')(\Psi', \Psi). \quad (47)$$

Now since the set $N_i(N, K)$, $i \leq K$, has been constructed such that the sequence $n_1', \cdots n_K'$ is not identical to the sequence $n_1, \cdots n_K$, Eq. (47) can be true for *all* i only if

$$0 = (\Psi, \Psi'); \quad (48)$$

i.e. The set of *common* eigenfunction must be an orthogonal set.

Actually, we needed to consider only $K-1$ number operators to demonstrate the necessity of (48) since members of the set $N_i(N, K)$, $i \leq K$, are related by the condition

$$\sum_{i=1}^{K} N_i(N, K) = N1(N, K). \quad (49)$$

However, we cannot consider a smaller number than $K-1$. For then, the sequences $n_1', \cdots n_{K'}'$, $K' \leq K-2$ and $n_1, \cdots n_{K'}$, are not in general distinct. This arises from the fact that the multiplicity of $N_i(\sigma, K')$, $i \leq K' \leq K-2$, in $N_i(N, K)$, $i \leq K' \leq K-2$ is in general greater than unity. The smaller set cannot therefore define a unique basis. We can summarize these results in the statement that either the set $N_i(N, K)$, $i \leq K$, or $\mathfrak{N}$ and any $K-1$ members of the previous set are a complete set of commuting operators.

We shall indicate the common eigenfunctions of the set $N_i(N, K)$ $i \leq K$ by the following notation:

$$|N; n_K, \cdots n_2(n_1)\rangle. \quad (50)$$

We have included N to denote that (50) is also an eigenfunction of $\mathfrak{N}$ belonging to the eigenvalue N. Once N is given, we need only specify the eigenvalues of $K-1$ of the number operators; we therefore place n_1 in parenthesis to indicate its superfluity. The reason for choosing n_1 is that in the standard

occupation scheme, the values of n_1 in each $O(\sigma, 2)$ appear in descending order, while the values of n_2 in each $O(\sigma, 2)$, of n_3 in each $O(\sigma', 3)$, etc. appear in ascending order. We retain n_1 for notational convenience.

Evidently, rows and columns of matrices can now be labeled according to the eigenvalues of the set $N_i(N, K)$ $i \leq K$; i.e. by (the columns of) the occupation scheme $O(N, K)$.

IX. SOLUTION FOR X_{ij} (N, K), i ≠ j

The solution for $X_{ij}(N, K)$, $i \neq j$, follows simply from the preceding analysis. We shall first consider $X_{12}(N, K)$. In the standard representation, it is explicitly composed of $X_{12}(\sigma, 2)$, $0 \leq \sigma \leq N$, which appear along the diagonal in a prescribed fashion. We may therefore write

$$\langle N; n_K' \cdots (n_1')| X_{12}(N, K) |N; n_K \cdots (n_1)\rangle$$
$$= \delta_{n_K}^{n_K'} \cdots \delta_{n_3}^{n_3'} \langle \sigma; n_2'(n_1')| X_{12}(\sigma, 2) |\sigma, n_2(n_1)\rangle. \quad (51)$$

If we now substitute the third formula from (28), with the substitution $\sigma - n_2 = n_1$, we obtain

$$= [(n_1 + 1)n_2]^{\frac{1}{2}} \delta_{n_K}^{n_K'} \cdots \delta_{n_3}^{n_3'} \delta_{n_2-1}^{n_3'}(\delta_{n_1+1}^{n_1'}). \quad (52)$$

For convenience, we have introduced a superfluous Kronecker symbol which is in accordance with the restriction $\sum n_i' = \sum n_i = N$.

For arbitrary i and j, $X_{ij}(N, K)$, $i \neq j$, is obtained as follows: We have noted in Sec. III that an interchange of s.p.-state labels is tantamount to a unitary transformation of the set $X_{ij}(N, K)$, $i, j \leq K$. Consider the interchange of 1 with i and 2 with j; that is, the set

$$N_\lambda = N_1, N_2, \cdots N_m, \cdots N_i, N_j, \cdots N_K \quad (53)$$

is transformed into the set

$$\bar{N}_\lambda = N_i, N_j \cdots N_m, \cdots N_1, N_2 \cdots N_K \quad (54)$$

and

$$X_{12} \rightarrow \bar{X}_{12} = X_{ij}, \quad \text{etc.}$$

Since X_{12} is explicitly a function of the eigenvalues of the number operators, its transformation is determined by their transformation. We thus have

$$\bar{X}_{12} = [(\bar{n}_1 + 1)\bar{n}_2]^{\frac{1}{2}} \delta_{\bar{n}_K}^{\bar{n}_K'} \cdots \delta_{\bar{n}_3}^{\bar{n}_3'} \delta_{\bar{n}_2-1}^{\bar{n}_3'}(\delta_{\bar{n}_1+1}^{\bar{n}_1'}) \quad (55)$$

or

$$X_{ij} = [(n_i + 1)n_j]^{\frac{1}{2}}$$
$$\times \delta_{n_K}^{n_K'} \cdots \delta_{n_j-1}^{n_j'} \cdots \delta_{n_i+1}^{n_i'} \cdots \delta_{n_2}^{n_2'}(\delta_{n_1}^{n_1'}), \quad (56)$$

which is the desired expression. This last result agrees with (8).

We have thus completed the task of obtaining a set of $X_{ij}(N, K)$, $i, j \leq K$, solely from the fundamental relations (10). One may verify directly that (56), along with

$$\langle N; n_K' \cdots (n_1')| N_i |N; n_K \cdots (n_1)\rangle$$
$$= n_i \delta_{n_K}^{n_K'} \cdots (\delta_{n_1}^{n_1'}), \quad (57)$$

satisfy the fundamental relations. In this verification, a subtle point arises: It is apparently not necessary to use the principle that the number operators are constructed such that *all* (and not just some) possible arrangements are included for which the condition $n_1 + \cdots + n_K = N$ is satisfied. The point is that (56) is not defined except when all possible arrangements are included. This may be verified directly in simple cases; however, the general necessity of the principle may be inferred from the irreducibility of the set $X_{ij}(N, K)$, $i, j \leq K$, which has been constructed according to that principle. We now demonstrate the irreducibility. In Sec. XI we give a practical method for constructing the matrices $X_{ij}(N, K)$, $i \neq j$.

X. IRREDUCIBILITY OF X_{ij} (N, K), i, j ≤ K

If the set $X_{ij}(N, K)$, $i, j \leq K$, were reducible, we could put it into the form

$$\begin{pmatrix} X_{ij}^a & \\ & X_{ij}^b \end{pmatrix}. \quad (58)$$

Both sets X_{ij}^a and X_{ij}^b, $i, j \leq K$ would then satisfy the fundamental relations (10). In particular, it would necessarily follow that all N_i^a, $i \leq K$, must have the same eigenvalues with the same multiplicity (see Sec. III). The same would also be true for N_i^b. We shall show that it is impossible to find such N_i^a and N_i^b, $i \leq K$.

Let us consider the subset $X_{ij}(N, K)$, $i, j \leq K - 1$. This subset, and consequently X_{ij}^a, $i, j \leq K - 1$ and X_{ij}^b, $i \leq K - 1$, satisfy the fundamental relations (10) for $K - 1$ s.p. states. This implies that we can find a transformation which reduces X_{ij}^a and X_{ij}^b, $i, j \leq K - 1$, to terms of $X_{ij}(\sigma, K - 1)$, $i, j \leq K - 1$, which appear along the diagonal. This transformation may of course be chosen with the reduced structure indicated in (58); it will therefore not disturb the assumed reduced structure of $X_{ij}(N, K)$. After the transformation, $X_{ij}(N, K)$, $i, j \leq K - 1$ is expressed completely in terms of $X_{ij}(\sigma, K - 1)$, which according to our analysis in Sec. V occur only once for each value of $\sigma \leq N$. In fact, the structure of $X_{ij}(N, K)$, $i, j \leq K - 1$, just found differs from the standard form (40) only

in the sequence in which the values of σ appear. We shall denote this sequence by $\sigma_N, \cdots \sigma_\rho, \cdots \sigma_0$. The important point to note is that a particular $X_{ij}(\sigma_\rho, K-1)$ can occur in either X^a_{ij}, $i, j \leq K-1$ or X^b_{ij}, $i, j \leq K-1$, but not in both.

By the same reasoning which led to (45), we can deduce the form which N_K, (N, K) must now assume; for convenience we give N^a_K, and N^b_K:

$$N^a_K = \begin{vmatrix} (N-\sigma_N)1(\sigma_N, K-1) & & \\ & \ddots & \\ & & (N-\sigma_m)1(\sigma_m, K-1) \end{vmatrix}, \quad (59)$$

$$N^b_K = \begin{vmatrix} (N-\sigma_{m+1})1(\sigma_{m+1}, K-1) & & \\ & \ddots & \\ & & (N-\sigma_0)1(\sigma_0, K-1) \end{vmatrix}.$$

It is now observed that a particular value of $n_K = N - \sigma_\rho$ appears in either N^a_K or N^b_K but not in both. Stated otherwise: neither N^a_K nor N^b_K assume all the values 0, 1, $\cdots N$. Let us now look at N^a_i, and N^b_i, $i \leq K-1$. One of these sets, say N^a_i, $\leq K-1$, contains $N_i(N, K-1)$, $i \leq K-1$, whose eigenvalues *do* assume all the values 0, 1, $\cdots N$. We must conclude that N^a_K is not similar to N^a_i, $i \leq K-1$, and therefore, the assumed reduction does not exist.

The above demonstration is, of course, meaningless for the case $K = 2$. The truth for that case follows from the relation between $X_{ij}(N, 2)$, $i, j \leq 2$, and angular momentum which was established in Sec. II.

XI. CONSTRUCTION OF THE MATRICES $X_{ij}(N, K)$, $i \neq j$

We shall now show that the matrices $X_{ij}(N, K)$, $i \neq j$ are easily constructed in the standard representation.

Let us consider the expression

$$\langle N; n'_K \cdots (n'_1)| X_{ij}(N, K) |N; n_K \cdots (n_1)\rangle$$
$$= [(n_i+1)n_j]^{\frac{1}{2}} \delta^{n_K}_{n'_K} \cdots \delta^{n_j}_{n'_{j-1}} \cdots \delta^{n_i}_{n'_{i+1}} (\delta^{n_i}_{n'_i}). \quad (60)$$

The problem of construction is to locate the nonzero matrix elements and then to assign to them the value $[(n_i+1)n_j]^{\frac{1}{2}}$. A nonzero element whose row is given by

$$n'_K, \cdots n'_j, \cdots n'_i, \cdots n'_1 \quad (61)$$

occurs only when its column

$$n_K, \cdots n_j, \cdots n_i, \cdots n_1 \quad (62)$$

is given by

$$n'_K, \cdots n'_i + 1, \cdots n'_j - 1, \cdots n'_1. \quad (63)$$

For a given sequence (61), a unique sequence is defined by (63) (they are linearly related). The only case in which the row given by (61) does not define a column through (63) is when $n'_i = 0$, for then $n_i = -1$. Another possibility apparently occurs when $n'_j = N$, for then $n_j = N+1$; however, since $n'_1 + \cdots + n'_K = N$, this implies that $n'_i = 0$. By virtue of the factor $[(n_i+1)n_j]^{\frac{1}{2}} = [n'_i n_j]^{\frac{1}{2}}$, these singular rows contain only zeros. By a similar analysis, in which we interchange the roles of row and column, we find that columns for which $n_j = 0$ are singular. Let us summarize the above findings: *In the matrix $X_{ij}(N, K)$ $i \neq j$, rows with $n'_i = 0$ have only zeros; all other rows have one nonzero element. Columns with $n_j = 0$ have only zeros; all other columns have one nonzero element.*

These statements are obviously true in any representation in which $N_i(N, K)$, $i \leq K$, is diagonal. The question now is how do we pair off nonzero rows with nonzero columns such that we obtain the correct location of the nonzero matrix elements. The answer, of course, depends on the representation we use. In the standard representation, it is particularly simple: *In the standard representation, the nonzero elements of $X_{ij}(N, K)$, $i \neq j$ are located by pairing off the nonzero rows with the nonzero columns in succession; i.e., the first nonzero row with the first nonzero column, etc.*

The validity of this prescription becomes rather obvious by considering the equivalent statement: When the states defined by (61)—except those with $n'_1 = 0$—are arranged in the sequence in which they appear in the standard occupation scheme, the corresponding states defined by (63) also appear in the sequence given by the standard occupation scheme. This stems from the fact that according to the correspondence between (61) and (63), the sequence of the groups of states with $n'_i = 0, 1, 2 \cdots$ (for fixed $n'_K, \cdots n'_{i+1}$) corresponds to the sequence $n_i = n'_i + 1 = 1, 2, \cdots$ (for fixed $n_K = n'_K, \cdots$, $n_{i+1} = n'_{i+1}$). Similarly the sequence of groups of states with $n'_i = 1, 2 \cdots$ (for fixed $n'_K \cdots n'_i \cdots n'_{i+1}$) corresponds to the sequence $n_i = n'_i - 1 = 0, 1 \cdots$ (for fixed $n_K = n'_K, \cdots, n_j = n'_j + 1, \cdots$, $n_{i+1} = n'_{i+1}$). The important point is that the sequential arrangement within the standard occupation scheme is not altered by the correspondence. Thus the correct corresponding sequence for determining the nonzero columns is obtained from the standard occupation scheme by simply deleting those states for which $n_i = 0$.

In Fig. 5, we give as an example the case for $N = 2$, $K = 4$. We include only $X_{ij}(2, 4)$ for $i < j$,

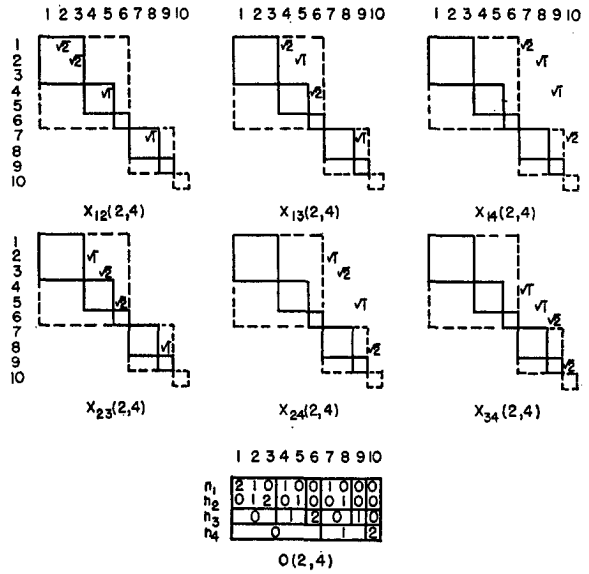


FIG. 5. Representation of the set $X_{ij}(2, 4)$, $i, j \leq 4$.

since the matrices for $i > j$ may be obtained by transposing (they are real). The matrices $N_i(2, 4)$ $i \leq 4$, are easily obtained from the (standard) occupation scheme $O(2, 4)$ which is found at the bottom of Fig. 5. We have also shown the hierarchical properties of the standard representation. The regions indicated by the broken lines in Fig. 5 correspond to $O(\sigma, K - 1 = 3)$, $\sigma = 0, 1, 2$; while the solid-line regions correspond (for given σ) to $O(\nu, K - 2 = 2)$, $\nu = 0, 1, \cdots \sigma$.

According to Fig. 5, all the nonzero elements of $X_{ij}(2, 4)$ $i < j$ lie to the right of the diagonal [all nonzero elements of $X_{ij}(2, 4)$ $i > j$ lie to the left of the diagonal], i.e., the matrices are triangular. This is a general property of the standard representation. For it is clear from the correspondence between (61) and (63), that when $j > i$, the state defined by (63) lies to the right of the state defined by (61) in the standard occupation scheme. This is just the condition that all the nonzero elements lie to the right of the diagonal.

SUMMARY

The objective of this investigation was to give a fairly exhaustive account of the irreducible representations of the operators X_{ij}, $i, j \leq K$ which satisfy the fundamental relations (10).

The fact that each X_{ij} commutes with the total number operator $\mathfrak{N}$ allowed us to classify the representations according to the eigenvalues N.

In Sec. II, we demonstrated the connection between X_{ij}, $i, j \leq 2$ and angular momentum operators. This allowed us to show the necessity of (10″) for definition of the X_{ij}. (The relation (10″) distinguishes bosons from fermions while (10′) and (10″) do not.) We were also able to show that the eigenvalues of the number operators were positive integers or zero.

We then went on to show (in Sec. III) that the irreducible representations $X_{ij}(N, K)$, $i, j \leq K$, could be divided into two sets of similar matrices: The set $N_i(N, K)$, $i \leq K$, and the set $X_{ij}(N, K)$ $i, j \leq K$; $i \neq j$. A set of similar number operators was defined (in Sec. IV) according to the usual principle that $n_1 + \cdots + n_K = N$ was satisfied in all possible ways just once. The fact that all (and not just some) arrangements must be included was later shown to be necessary, for in Sec. X it was found that the set $X_{ij}(N, K)$, $i, j \leq K$, which was constructed according to this principle, was irreducible.

In Sec. V, we studied $X_{ij}(N, K)$, $i, j \leq K' < K$ from the standpoint that the set is a *reducible* representation of X_{ij}, $i, j \leq K'$. We then found how many times the irreducible representations $X_{ij}(\sigma, K')$, $i, j \leq K'$; $0 \leq \sigma \leq N$ appeared in $X_{ij}(N, K)$, $i, j \leq K' < K$. This was found to be $B(N - \sigma, K - K')$. In particular, $X_{ij}(\sigma, K - 1)$ is contained once for each value of σ. This result was used iteratively to define the standard representation (Sec. VI). This same result, along with the fact that all the number operators $N_i(N, K)$, $i \leq K$ are similar to one another was the essential point needed to show that the representation $X_{ij}(N, K)$, $i, j \leq K$, as defined above, was irreducible.

In Sec. VIII, we saw that the common eigenfunctions of the set $N_i(N, K)$, $i \leq K$, formed a complete orthogonal basis for the $B(N, K)$-dimensional space in which the operators are represented. This justified labeling the rows and columns according to these eigenfunctions (i.e. according to arrangements of integers which satisfy $n_1 + \cdots + n_K = N$).

The specific form of the $X_{ij}(N, K)$, $i \neq j$, was obtained in Sec. IX by a transformation of $X_{12}(N, K)$. The transformation was simple the one which (simultaneously) took N_1 into N_i and N_2 into N_j, while leaving the remaining number operators undisturbed.

In Sec. VII, a simple systematic method was given for writing down the occupation scheme for given N and K in the standard representation. This scheme gives the matrices $N_i(N, K)$, $i \leq K$ directly.

In Sec. XI, we saw how this same standard scheme enabled us to construct the matrices $X_{ij}(N, K)$, $i \neq j$, by an equally simple method.

In conclusion, we should mention that this study

is a desirable prelude to the main problem which is the diagonalizing of the Hamiltonian. The difficulties in diagonalizing a (Hermitian) matrix of $B(N, K)$ dimensions are well known. However, the Hamiltonian (7) is *not* the most general Hermitian matrix with these dimensions (when $N > 2$). More general matrices would be obtained by including terms with higher powers of X_{ij}. (These would correspond to 3, 4 $\cdots$ n-body forces). In future investigations, we shall attempt to answer the question: In just what way does the specialization to Hamiltonians of the form (7) simplify the problem of diagonalization? In particular, we shall present the solution to the problem of N bosons interacting through harmonic forces. The solution to this problem is quite simple for "Boltzmann" particles, but is not at all obvious for bosons. We shall show that after removal of the center of mass motion, the second quantized Hamiltonian can be made to depend on the finite representations of the X_{ij}. The Hamiltonian can then be diagonalized by elementary means to obtain the energy levels of the system.

Generalization of a Theorem by Goldberg and Sachs*

I. ROBINSON†
Cornell University, Ithaca, New York and Syracuse University, Syracuse. New York
AND
A. SCHILD
The University of Texas, Austin, Texas
(Received 28 November 1962)

A connection is established between algebraic degeneracy of the Weyl tensor, the existence of a null geodesic shear-free congruence, and certain restrictions on the Ricci tensor which are weaker than the gravitational equations for empty space. The result is roughly, with some important qualifications, that any two of these conditions imply the third. These restrictions on the metric are shown to be invariant under conformal transformations.

1. INTRODUCTION

THE following theorem has been proven by J. N. Goldberg and R. K. Sachs[1,2]:

In the 4-space of general relativity which satisfies Einstein's field equations for the vacuum

$$R_{mn} = 0, \tag{1.1}$$

the space is algebraically degenerate in the sense of the Petrov classification if and only if it contains a shear-free null geodesic congruence.[3]

The last two statements are conform invariant.[4] It is therefore clear that the theorem can be sharpened by replacing the vacuum condition (1.1) by weaker conform invariant conditions. In this work the weakest possible such conditions are obtained.[5]

2. THE TETRAD FORMALISM

We shall often find it convenient to use a notation which suppresses suffixes. Vectors k^a, m^a are denoted by k, m, tensors g^{ab}, V^{ab}, U^{ab}, P^{abc}, C^{abcd} by g, V, U, P, C. Writing two factors next to one another denotes the Cartesian product, e.g., km denotes $k^a m^b$, UV denotes $U^{ab}V^{cd}$. A dot between the factors indicates contraction over the suppressed suffixes adjacent to the dot, two dots indicate double contraction and multiplication by $\frac{1}{2}$. Thus $k\cdot m$ denotes $k_a m^a$, $U\cdot V$ denotes $U^{ac}V_c{}^b$, $U{:}V$ denotes $\frac{1}{2}U^{ba}V_{ab}$, $C\cdot V$ denotes $C^{abce}V_e{}^d$, $V\cdot P{:}U$ denotes $\frac{1}{2}V^a{}_d P^{dbe}U_{eb}$.

With the metric tensor g^{ab} of space-time (signature +2) may be associated a quasiorthogonal tetrad of null vectors k^a, m^a, l^a, $\bar{l}^a$, k and m being real and $\bar{l}$ the complex conjugate of l. These satisfy the equivalent sets of relations

$$g = km + mk + \bar{l}l + l\bar{l}, \tag{2.1}$$

$$k\cdot m = l\cdot \bar{l} = 1, \quad \text{other scalar products zero.} \tag{2.2}$$

In terms of this tetrad, three bivectors U^{ab}, V^{ab}, M^{ab} can be defined:

$$U = ml - lm, \qquad V = k\bar{l} - \bar{l}k,$$
$$M = km - mk + \bar{l}l - l\bar{l}. \tag{2.3}$$

For a suitable choice of orientation of the tetrad, these bivectors are self-dual in the sense

$$U^* = -iU, \quad V^* = -iV, \quad M^* = -iM, \tag{2.4}$$

where duality with respect to a pair of antisymmetric suffixes is defined by

$$W_{..ab..}^{\ast} = \tfrac{1}{2}(-|g|)^{\frac{1}{2}}\epsilon_{abcd}W_{..}{}^{cd}{}_{..}, \tag{2.5}$$

ϵ_{abcd} being completely antisymmetric with $\epsilon_{1234} = 1$ and $|g|$ being the determinant of g_{ab}. The three bivectors satisfy the completeness relation $X = (\frac{1}{2}MM - UV - VU){:}X$ for any self-dual bivector X. Thus any self-dual bivector is a linear combination of U, V and M.

The bivectors (2.3) and their complex conjugates satisfy the relations

$$M : M = \bar{M} : \bar{M} = 2, \quad V : U = \bar{V} : \bar{U} = -1,$$
$$\text{other such products zero.} \tag{2.6}$$

Since the gradient of one of the bivectors (2.3) is self-dual, it must be a linear combination of the three bivectors. Using Eqs. (2.6), we have

$$U_{;a} = \lambda_a U + \mu_a M,$$

* This research has been sponsored by the Office of Aerospace Research, and by the Office of Scientific Research, of the United States Air Force.

† Present address: Southwest Center for Advanced Studies, Dallas, Texas.

[1] J. N. Goldberg and R. K. Sachs, Acta Physica Polonica 22, 13 (1962).

[2] E. Newman and R. Penrose, J. Math. Phys. 3, 566 (1962).

[3] R. K. Sachs, Proc. Roy. Soc. (London) A264, 309 (1961).

[4] F. A. E. Pirani and A. Schild, Bull. Acad. Polon. Sci. Ser. Sci. Math. Astron. Phys. 9, 543 (1961).

[5] A short account of our results, without proofs, was given at the International Conference on Relativistic Theories of Gravitation, Warsaw, July 1962, and will appear in the proceedings of the conference.

$$V_{;a} = -\lambda_a V + \nu_a M,$$
$$M_{;a} = 2\nu_a U + 2\mu_a V, \tag{2.7}$$

where the semicolon denotes covariant differentiation and, by Eqs. (2.3), the vectors μ, ν, λ are given by

$$\mu_a = -m\cdot t_{;a}, \qquad \nu_a = -\bar{l}\cdot k_{;a},$$
$$\lambda_a = \bar{l}\cdot t_{;a} - m\cdot k_{;a}. \tag{2.8}$$

The following products are useful:

$$U\cdot U = 0, \quad U\cdot V = -\tfrac{1}{2}g + \tfrac{1}{2}M, \quad U\cdot M = U,$$
$$V\cdot U = -\tfrac{1}{2}g - \tfrac{1}{2}M, \quad V\cdot V = 0, \quad V\cdot M = -V,$$
$$M\cdot U = -U, \quad M\cdot V = V, \quad M\cdot M = g. \tag{2.9}$$

3. THE CONFORMAL CURVATURE TENSOR

Weyl's conformal curvature tensor[6] is defined by

$$C^a{}_{bcd} = R^a{}_{bcd} + g_{b[d}R^a{}_{c]} + R_{b[d}\delta^a{}_{c]} - \tfrac{1}{3}g_{b[d}\delta^a{}_{c]}R, \tag{3.1}$$

where $R^a{}_{bcd}$ is the Riemann curvature tensor, defined for example by the commutation rule for covariant differentiation, $T_{b;cd} - T_{b;dc} = R^a{}_{bcd}T_a$, $R_{bc} = R^a{}_{bca}$ is the Ricci tensor, and $R = R^a{}_a$ the curvature invariant.

The conformal curvature tensor has the algebraic symmetries

$$C_{abcd} = -C_{bacd} = C_{cdab},$$
$$C_{a[bcd]} = 0, \quad C^a{}_{bca} = 0, \quad {}^*C^* = -C, \tag{3.2}$$

where ${}^*C$ denotes the dual C^*_{abcd} with respect to the first pair of suffixes, C^* the dual $C_{abcd}{}^*$ with respect to the second pair, and thus ${}^*C^*$ denotes the double dual C^{**}_{abcd}.

The Bianchi identities for the curvature tensor imply

$$C_a{}^m{}_{bc;m} = P_{abc}, \tag{3.3}$$

where

$$P_{abc} = -P_{acb} = -R_{a[b;c]} + \tfrac{1}{6}g_{a[b}R_{,c]}. \tag{3.4}$$

We define $C^-{}_{abcd}$ by

$$C^- = C + iC^*. \tag{3.5}$$

It follows from the last of the algebraic properties (3.2) that C^- is self-dual with respect to each pair of suffixes:

$${}^*C^- = C^{-*} = -iC^-. \tag{3.6}$$

Hence C^- can be expanded in terms of Cartesian products of the three bivectors U, V, M. If the algebraic symmetries (3.2) are taken into account, this gives

$$C^- = C^{(5)}UU + C^{(4)}(UM + MU) + C^{(3)}(MM + UV + VU) + C^{(2)}(VM + MV) + C^{(1)}VV. \tag{3.7}$$

Using Eqs. (2.6) and (2.9), we obtain

$$V : C^- : V = C^{(5)},$$
$$V\cdot C^- : V = \tfrac{1}{2}(g + M)C^{(5)} + VC^{(4)},$$
$$C^- : V = -UC^{(5)} - MC^{(4)} - VC^{(3)},$$
$$C^-\cdot V = -\tfrac{1}{2}U(g - M)C^{(5)} + (UV - \tfrac{1}{2}M(g - M))C^{(4)} + (MV - \tfrac{1}{2}V(g - M))C^{(3)} + VVC^{(2)}. \tag{3.8}$$

Let P^* denote the dual $P_{abc}{}^*$ with respect to the last pair of suffixes, and define

$$P^- = P + iP^*. \tag{3.9}$$

Then the differential identities (3.3) become

$$C^-{}_a{}^m{}_{bc;m} = P^-{}_{abc}. \tag{3.10}$$

Relations corresponding to Eqs. (3.5) to (3.10) hold for the complex conjugates

$$C^+ = \bar{C}^- = C - iC^*$$
$$P^+ = \bar{P}^- = P - iP^*. \tag{3.11}$$

4. THE IDENTITIES

Differentiation and use of Eqs. (2.7) immediately gives

$$(V\cdot C^- : V)_{;m} = V\cdot C^-{}_{;m} : V - 2\lambda_m V\cdot C^- : V + C^{(5)}\nu_m U - C^{(4)}\nu_m(2g + M) - 3C^{(3)}\nu_m V,$$
$$(C^- : V)_{;m} = C^-{}_{;m} : V + C^{(5)}\lambda_m U + C^{(4)}(\lambda_m M + 2\nu_m U) + C^{(3)}(\lambda_m V + 2\nu_m M) + 2C^{(2)}\nu_m V,$$
$$(C^-\cdot V)_{;m} = C^-{}_{;m}\cdot V - C^{(1)}\nu_m VV + \text{terms which contain } C^{(2)}, C^{(3)}, C^{(4)} \text{ or } C^{(5)} \text{ as factor.} \tag{4.1}$$

By contraction, we obtain from these the identities which are of central importance in this work:

[6] L. P. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, New Jersey, 1960), Sec. 28.

$$\tfrac{1}{2}(V^{ra}C^{-}{}_{a}{}^{m}{}_{bc}V^{cb})_{;m} = \tfrac{1}{2}V^{ra}P^{-}{}_{abc}V^{cb} - \lambda_m V^{ra}C^{-}{}_{a}{}^{m}{}_{bc}V^{cb} + C^{(5)}\nu_m U^{rm} - C^{(4)}\nu_m(2g^{rm} + M^{rm}) - 3C^{(3)}\nu_m V^{rm}, \qquad (\mathrm{i}_{(1)})$$

$$\tfrac{1}{2}(C^{-}{}_{a}{}^{m}{}_{bc}V^{cb})_{;m} = \tfrac{1}{2}P^{-}{}_{abc}V^{cb} + 2C^{(2)}\nu_m V_a{}^m + \text{terms which contain } C^{(3)},\ C^{(4)} \text{ or } C^{(5)} \text{ as factor}, \qquad (\mathrm{i}_{(2)})$$

$$(C^{-}{}_{a}{}^{m}{}_{bn}V^{nc})_{;m} = P^{-}{}_{abn}V^{nc} - C^{(1)}\nu_m V_a{}^m V_b{}^c + \text{terms which contain } C^{(2)},\ C^{(3)},\ C^{(4)} \text{ or } C^{(5)} \text{ as factor}. \qquad (\mathrm{i}_{(3)})$$

5. DEGENERACY

With the conformal curvature tensor are associated four real null vectors which are called Debever-Penrose vectors. The distinctness or coincidence of the directions of these vectors determines the algebraic structure of the conformal curvature tensor and leads to the classification of Petrov. For details of this theory and for literature references, the reader is referred to the paper by Sachs.[3] Here we shall use the fact that k is a Debever–Penrose vector if and only if $C^{(5)} = 0$ in Eq. (3.7), that k is a double, triple, quadruple Debever–Penrose vector if and only if $C^{(5)} = C^{(4)} = 0$, $C^{(5)} = C^{(4)} = C^{(3)} = 0$, $C^{(5)} = C^{(4)} = C^{(3)} = C^{(2)} = 0$, respectively. In the latter cases, the conformal curvature tensor is said to be degenerate or algebraically special.

The following definitions of degeneracy, and their immediate consequences by Eqs. (3.8), are in each case assumed to hold throughout space-time:

$d_{(0)} \Leftrightarrow k$ is a Debever–Penrose vector
$\Leftrightarrow C^{(5)} = 0$
$\Leftrightarrow V : C^{-} : V = 0$
$\Leftrightarrow V\cdot C^{-} : V = VC^{(4)}$,

$d_{(1)} \Leftrightarrow k$ is (at least) a double Debever–Penrose vector
$\Leftrightarrow C^{(5)} = C^{(4)} = 0$
$\Leftrightarrow V\cdot C^{-} : V = 0$
$\Leftrightarrow C^{-} : V = -VC^{(3)}$,

$d_{(2)} \Leftrightarrow k$ is (at least) a triple Debever–Penrose vector
$\Leftrightarrow C^{(5)} = C^{(4)} = C^{(3)} = 0$
$\Leftrightarrow C^{-} : V = 0$
$\Leftrightarrow C^{-}\cdot V = VVC^{(2)}$

$d_{(3)} \Leftrightarrow k$ is (at least) a quadruple Debever–Penrose vector
$\Leftrightarrow C^{(5)} = C^{(4)} = C^{(3)} = C^{(2)} = 0$
$\Leftrightarrow C^{-}\cdot V = 0$
$\Leftrightarrow C^{-} = VVC^{(1)}$,

$d_{(4)} \Leftrightarrow$ space-time is conformally flat
$\Leftrightarrow C^{(5)} = C^{(4)} = C^{(3)} = C^{(2)} = C^{(1)} = 0$
$\Leftrightarrow C^{-} = 0$.

The following definitions provide a more specific classification of degeneracy:

$D_{(1)} \Leftrightarrow d_{(1)}$, but not $d_{(2)}$
$\Leftrightarrow C^{(5)} = C^{(4)} = 0,\ C^{(3)} \neq 0$,

$D_{(2)} \Leftrightarrow d_{(2)}$, but not $d_{(3)}$
$\Leftrightarrow C^{(5)} = C^{(4)} = C^{(3)} = 0,\ C^{(2)} \neq 0$,

$D_{(3)} \Leftrightarrow d_{(3)}$, but not $d_{(4)}$
$\Leftrightarrow C^{(5)} = C^{(4)} = C^{(3)} = C^{(2)} = 0,\ C^{(1)} \neq 0$.

6. THE FIELD EQUATIONS

We shall consider certain equations which follow directly from Einstein's equations (1.1) or, indeed, from the vacuum field equations with a cosmological constant. Our weaker "field equations" are

$$f_{(1)} \Leftrightarrow V\cdot P^{-} : V = 0,$$
$$f_{(2)} \Leftrightarrow P^{-} : V = 0,$$
$$f_{(3)} \Leftrightarrow P^{-}\cdot V = 0.$$

In symbols:

$$R_{mn} = 0 \Rightarrow f_{(3)} \Rightarrow f_{(2)} \Rightarrow f_{(1)}. \qquad (6.1)$$

The definitions of this section and the one preceeding lead to

Theorem I: Strong degeneracy implies weak field equations:

$$d_{(2)} \Rightarrow f_{(1)}, \qquad (\mathrm{I}_{(1)})$$
$$d_{(3)} \Rightarrow f_{(2)}, \qquad (\mathrm{I}_{(2)})$$
$$d_{(4)} \Rightarrow f_{(3)}. \qquad (\mathrm{I}_{(3)})$$

Theorem $\mathrm{I}_{(1)}$ follows immediately from identity $i_{(1)}$; theorem $\mathrm{I}_{(2)}$ from identity $i_{(2)}$; theorem $\mathrm{I}_{(3)}$ is obvious since $d_{(4)}$ implies the stronger condition $C^{-} = 0$.

7. THE SHEAR-FREE, NULL GEODESIC CONGRUENCE

The field of real null vectors k determines a congruence of null curves which, at each point of

space-time, have k as a tangent. We shall say that k is geodesic if the null curves of the congruence are null geodesics. The condition is that $k_{a;b}k^b$ be proportional to k_a and this is equivalent to

$$l^a k_{a;b}k^b = 0, \tag{7.1}$$

since $k^a k_{a;b}k^b$ is an automatic consequence of $k^a k_a = 0$, and since (7.1) implies its complex conjugate $\bar{l}^a k_{a;b}k^b = 0$.

The null field k is said to be geodesic and shear-free if, in addition to Eq. (7.1),

$$\bar{l}^a k_{a;b}\bar{l}^b = 0 \tag{7.2}$$

is satisfied. This condition has the following geometric interpretation[3,4]: At any point P consider a 2-space S, orthogonal, but not tangent to, the null geodesic l of the congruence which passes through P (Fig. 1). At any other point P' of l, consider any 2-space S', orthogonal to l, but not tangent to it. In S, draw the infinitesimal circle C with center at P. Then the null geodesics $\bar{l}$ of the congruence which pass through the points $\bar{P}$ of the circle C will meet S' in the points $\bar{P}'$ of a curve C'. The shear-free condition ensures that C' is again an infinitesimal circle with center P'.

Differentiation of V, as given by Eq. (2.3), immediately shows that Eqs. (7.1) and (7.2) are equivalent to the single condition $V^a{}_b V^{bc}{}_{;c} = 0$. Substitution for $V^{bc}{}_{;c}$ from Eq. (2.7) shows that the condition is in turn equivalent to $V^{ab}\nu_b = 0$. We shall denote by (gs) the property that k is geodesic and shear-free, so that

$$\begin{aligned} gs &\Leftrightarrow k \text{ geodesic and shear-free} \\ &\Leftrightarrow V^a{}_b V^{bc}{}_{;c} = 0 \\ &\Leftrightarrow V\cdot\nu = 0. \end{aligned}$$

The following lemma is proved in the references cited[1,4]:

Lemma I: If k is geodesic and shear-free, then it is a Debever–Penrose vector:

$$gs \Rightarrow d_{(0)}. \tag{7.3}$$

8. THE MAIN THEOREMS

Theorem II. Degeneracy of the conformal tensor of a specific type implies that the degenerate Debever–Penrose vector is geodesic and shear-free if and only if field equations of the corresponding type are satisfied:

$$\begin{aligned} &(D_{(1)} \quad \text{and} \quad f_{(1)}) \Rightarrow gs, \\ &(D_{(1)} \quad \text{and} \quad gs) \Rightarrow f_{(1)}; \end{aligned} \tag{II$_{(1)}$}$$

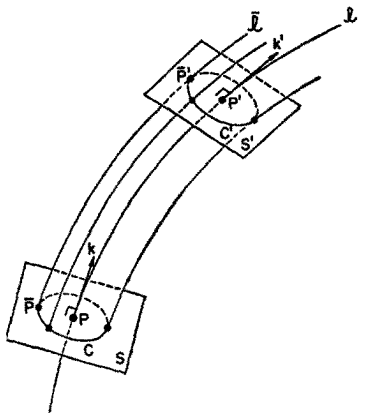


FIG. 1

$$\begin{aligned} &(D_{(2)} \quad \text{and} \quad f_{(2)}) \Rightarrow gs, \\ &(D_{(2)} \quad \text{and} \quad gs) \Rightarrow f_{(2)}; \end{aligned} \tag{II$_{(2)}$}$$

$$\begin{aligned} &(D_{(3)} \quad \text{and} \quad f_{(3)}) \Rightarrow gs, \\ &(D_{(3)} \quad \text{and} \quad gs) \Rightarrow f_{(3)}. \end{aligned} \tag{II$_{(3)}$}$$

The fact that degeneracy and the geodesic and shear-free condition imply corresponding field equations shows that these field equations are the weakest possible ones for the Goldberg–Sachs theorem.

The proof of theorem II is simple. Since $D_{(1)}$ states that $V\cdot C^- : V = 0$, $C^{(5)} = C^{(4)} = 0$, $C^{(3)} \neq 0$, identity $i_{(1)}$ becomes

$$V\cdot P^- : V = 3C^{(3)} V\cdot\nu, \tag{8.1}$$

and II$_{(1)}$ follows. Similarly II$_{(2)}$ and II$_{(3)}$ follow from the identities $i_{(2)}$ and $i_{(3)}$.

Theorem III. If k is geodesic and shear-free and if the weakest field equations $f_{(1)}$ hold, then k is a degenerate Debever-Penrose vector:

$$(gs \quad \text{and} \quad f_{(1)}) \Rightarrow d_{(1)}. \tag{III}$$

Theorems II and III constitute the generalization of the theorem of Goldberg and Sachs. In the next section we shall show explicitly that these theorems relate purely conformal properties.

The lack of symmetry in theorem III suggests that it may be possible to generalize the main theorems by replacing the condition gs by stronger conditions $gs_{(1)}$, $gs_{(2)}$, $gs_{(3)}$, so that any two of the properties $D_{(\alpha)}$, $f_{(\alpha)}$, $gs_{(\alpha)}$ imply the third for each α in the range 1, 2, 3.

The proof of theorem III requires several lemmas. One of us has shown in a previous publication[7] that if $V^a{}_b V^{bc}{}_{;c} = 0$, then a scalar field α can be found so that $(\alpha V^{bc})_{;c} = 0$. The scalar field $\alpha = ae^{ib}$,

[7] I. Robinson, J. Math. Phys. 2, 290 (1961).

a and b real, can be absorbed in V by renormalizing the quasiorthogonal tetrad, k absorbing a and t absorbing e^{ib}. Thus we have

Lemma II. Given gs, k and t can be normalized so that throughout space-time

$$V^{ab}{}_{;b} = 0. \tag{8.2}$$

With this normalization, Eq. (2.7) gives

$$-V\cdot\lambda + M\cdot\nu = 0, \tag{8.3}$$

and, multiplying by $M\cdot$ from the left,

$$\begin{aligned}\nu &= V\cdot\lambda \\ &= M\cdot\nu.\end{aligned} \tag{8.4}$$

Hence, since M and V are skew-symmetric,

$$\nu\cdot\nu = 0, \tag{8.5}$$

$$\nu\cdot\lambda = 0. \tag{8.6}$$

Lemma III. Given gs, and adopting the normalization of lemma II, we have

$$\nu^{a}{}_{;a} = -C^{(4)}. \tag{8.7}$$

To prove the lemma, take the complex conjugate of Eq. (8.2); since V is self-dual [Eq. (8.2)], this gives

$$V^{*ab}{}_{;b} = 0, \tag{8.8}$$

so that V satisfies Maxwell's equations without charges and currents. In flat space-time, each component of V satisfies d'Alambert's equation. In curved space-time, the commutation of covariant derivatives introduces the curvature tensor and, as is well known, the components of a Maxwell field satisfy the equations

$$V_{ab;c}{}^{c} = -R_{ar}V^{r}{}_{b} - V_{ar}R^{r}{}_{b} - V^{c}{}_{r}R^{r}{}_{cba}. \tag{8.9}$$

Multiplying by V^{da} and introducing the conformal curvature tensor from Eq. (3.1), we obtain

$$\begin{aligned}V\cdot V_{;c}{}^{c} &= 2V : C\cdot V \\ &= V : C^{-}\cdot V + V : C^{+}\cdot V \\ &= V : C^{-}\cdot V.\end{aligned} \tag{8.10}$$

The last line follows from the complex conjugate of Eq. (3.7) and the fact that the double dot products of V with $\bar{U}$, $\bar{V}$ and $\bar{M}$ are zero. Remembering that gs is assumed and that therefore, by lemma I, $C^{(5)} = 0$, we have

$$V\cdot V_{;c}{}^{c} = C^{(4)}V. \tag{8.11}$$

Equations (2.7), upon differentiation, give

$$\begin{aligned}V_{;c}{}^{c} = &-V\lambda_{c;}{}^{c} + V\lambda\cdot\lambda - M\nu\cdot\lambda \\ &+ M\nu_{c;}{}^{c} + 2U\nu\cdot\nu + 2V\nu\cdot\mu.\end{aligned} \tag{8.12}$$

By Eqs. (8.5), (8.6), this gives

$$V\cdot V_{;c}{}^{c} = -V\nu_{c;}{}^{c} \tag{8.13}$$

Comparison with Eq. (8.11) completes the proof of lemma III.

We are now ready to prove theorem III. Since, by hypothesis and lemma I, $C^{(5)} = 0$ and $V\cdot P^{-}:V = 0$, the identity $i_{(1)}$ simplifies to

$$\begin{aligned}(V^{ab}C^{(4)})_{;b} = &-2V^{ab}\lambda_{b}C^{(4)} \\ &- 2\nu^{a}C^{(4)} - M^{ab}\nu_{b}C^{(4)}.\end{aligned} \tag{8.14}$$

Adopting the normalization of lemma II, this simplifies further to

$$V^{ab}C^{(4)}{}_{,b} = -5\nu^{a}C^{(4)}. \tag{8.15}$$

Multiplying by λ_{a}, and using Eqs. (8.4) and (8.6), it is seen that

$$C^{(4)}{}_{,a}\nu^{a} = 0. \tag{8.16}$$

We now differentiate Eq. (8.15) covariantly with respect to x^{a}, and obtain

$$-5\nu^{a}{}_{;a}C^{(4)} = V^{ab}C^{(4)}{}_{;ba} = 0, \tag{8.17}$$

since $C^{(4)}{}_{;ba}$ is symmetric in its suffixes. Lemma III then gives

$$(C^{(4)})^{2} = 0, \tag{8.18}$$

and theorem III follows.

9. CONFORMAL TRANSFORMATIONS

Under a conformal change of metric,

$$g'_{ab} = e^{2\psi}g_{ab}, \tag{9.1}$$

where ψ is an arbitrary real function of position in space-time; the mixed conformal curvature tensor C is invariant[6]:

$$C'_{a}{}^{b}{}_{rs} = C_{a}{}^{b}{}_{rs}; \tag{9.2}$$

hence its name. In four dimensions, $(-|g|)^{\frac{1}{2}}\,g^{ab}g^{cd}$ is also conform invariant, and therefore so are $C^{*}{}_{a}{}^{b}{}_{rs}$ and $C^{-}{}_{a}{}^{b}{}_{rs}$:

$$C'^{-}{}_{a}{}^{b}{}_{rs} = C^{-}{}_{a}{}^{b}{}_{rs}. \tag{9.3}$$

The tensor P_{abc} transforms as follows[6]:

$$P'_{abc} = P_{abc} + C_{a}{}^{m}{}_{bc}\psi_{,m}, \tag{9.4}$$

and therefore

$$P'^{-}{}_{abc} = P^{-}{}_{abc} + C^{-}{}_{a}{}^{m}{}_{bc}\psi_{,m}. \tag{9.5}$$

Since a null vector remains null under a conformal change of metric, and orthogonality is preserved, we may let the quasiorthogonal tetrad and the

associated bivectors transform according to:

$$\begin{aligned} k'_a &= k_a, & m'_a &= e^{2\psi} m_a, & t'_a &= e^{\psi} t_a, \\ k'^a &= e^{-2\psi} k^a, & m'^a &= m^a, & t'^a &= e^{-\psi} t^a, \end{aligned} \tag{9.6}$$

$$\begin{aligned} U'_a{}^b &= e^{\psi} U_a{}^b, & V'_a{}^b &= e^{-\psi} V_a{}^b, & M'_a{}^b &= M_a{}^b, \\ U'_{ab} &= e^{3\psi} U_{ab}, & V'_{ab} &= e^{\psi} V_{ab}, & M'_{ab} &= e^{2\psi} M_{ab}, \\ U'^{ab} &= e^{-\psi} U^{ab}, & V'^{ab} &= e^{-3\psi} V^{ab}, & M'^{ab} &= e^{-2\psi} M^{ab}. \end{aligned} \tag{9.7}$$

Writing $C^-{}_a{}^m{}_{bc}$ in the form of Eq. (3.7); we then obtain the transformation properties:

$$\begin{aligned} C'^{(5)} &= e^{-4\psi} C^{(5)}, \quad C'^{(4)} = e^{-3\psi} C^{(4)}, \quad C'^{(3)} = e^{-2\psi} C^{(3)}, \\ C'^{(2)} &= e^{-\psi} C^{(2)}, \quad C'^{(1)} = C^{(1)}. \end{aligned} \tag{9.8}$$

This shows that algebraic degeneracy of the conformal curvature tensor is conform invariant, i.e.,

$$\begin{aligned} &d_{(0)} \Leftrightarrow d'_{(0)}, \quad d_{(1)} \Leftrightarrow d'_{(1)}, \cdots, d_{(4)} \Leftrightarrow d'_{(4)}, \\ &D_{(1)} \Leftrightarrow D'_{(1)}, \quad D_{(2)} \Leftrightarrow D'_{(2)}, \quad D_{(3)} \Leftrightarrow D'_{(3)}. \end{aligned} \tag{9.9}$$

We might also see this directly by considering the definition of degeneracy in terms of the products of C^- and V.

The property of k being geodesic and shear-free, can be shown to be conform invariant,

$$gs \Leftrightarrow gs', \tag{9.10}$$

by giving purely conformal definitions of the terms null geodesic and shear-free.[4] It also follows from

$$\begin{aligned} V'^a{}_b V'^{bc}{}_{:c} &= (-|g'|)^{-\frac{1}{2}} V'^a{}_b ((-|g'|)^{\frac{1}{2}} V'^{bc})_{,c} \\ &= e^{-5\psi} (-|g|)^{-\frac{1}{2}} V^a{}_b [e^{\psi} (-|g|)^{\frac{1}{2}} V^{bc}]_{,c} \\ &= e^{-4\psi} V^a{}_b V^{bc}{}_{:c}, \end{aligned} \tag{9.11}$$

where $_{,c}$ denotes partial differentiation $\partial/\partial x^c$.

From Eq. (9.5) and the definitions of Secs. 5 and 6, it immediately follows that

Theorem IV: When space-time is suitably degenerate, the field equations are conform invariant:

$$d_{(1)} \Rightarrow (f_{(1)} \Leftrightarrow f'_{(1)}), \tag{IV$_{(1)}$}$$

$$d_{(2)} \Rightarrow (f_{(2)} \Leftrightarrow f'_{(2)}), \tag{IV$_{(2)}$}$$

$$d_{(3)} \Rightarrow (f_{(3)} \Leftrightarrow f'_{(3)}). \tag{IV$_{(3)}$}$$

Since degeneracy is either a hypothesis or a conclusion in theorems I, II, and III, it is now clear that these theorems are invariant under a conformal change of the metric.

Kinematics of the Relativistic Two-Particle System*

ALAN J. MACFARLANE
Department of Physics and Astronomy, University of Rochester, Rochester, New York
(Received 28 September 1962)

Knowing the (canonical) forms that the generators of the infinitesimal transformations of P (the restricted Poincaré Group) take when operating within the Hilbert space of states of an arbitrary irreducible representation of P, such as that which affords a kinematic description of a single relativistic particle, we develop the forms in which they appear when operating within the Hilbert space of states of the (reducible) direct product representation of P which describes a system of two noninteracting relativistic particles. On introducing the Clebsch–Gordan (C–G) series of P which expresses the resolution of this latter Hilbert space into a direct integral over the Hilbert spaces wherein operate the irreducible constituents of the direct product representation, we prove that the generators operate in canonical form within each of these subspaces. If we adopt the alternative viewpoint that the C–G series of P must be written exactly so as to achieve this, our analysis may be regarded as providing a fully explicit and mathematically complete derivation of the formula for the C–G coefficients of P that appear in the C–G series. This formula has previously been suggested only on the basis of heuristic physical argument, and its proof is the principal accomplishment of the present work. One important fact which receives further clarification from the explicit nature of our analysis is the following: in order to see how the intrinsic angular momentum of the two-particle system, in a state of given linear momentum, is compounded from the relative angular momentum and intrinsic spins of the particles, one must view the state from a frame of reference wherein it appears to have zero momentum.

1. INTRODUCTION

WE here undertake a detailed investigation of some kinematic properties—chiefly concerning angular momentum—of the relativistic system of two noninteracting particles of nonzero rest masses and arbitrary intrinsic spins.[1] Our starting point is the theory of the representations of the Poincaré or inhomogeneous Lorentz group P in the form originally discovered by Wigner.[2] The first step of our analysis is the derivation therefrom of the forms which the generators of infinitesimal translations and homogeneous Lorentz transformations take when operating within the Hilbert space $H([M, j])$ of states of an arbitrary irreducible representation $[M, j]$ of P, such as the representation $[m, s]$ which describes the kinematics of a single relativistic particle of mass m and spin s. Such results have appeared before in the literature[3] in a variety of equivalent forms: we refer to the precise manner [see Eqs. (3.18) and (3.19) below] in which we present our results as being canonical. We then go on to their application to the treatment of the two-particle system as follows.

If[4] $P^{(1)\alpha}$, $P^{(2)\alpha}$ and $M^{(1)\alpha\beta}$, $M^{(2)\alpha\beta}$ are the generators of infinitesimal translations and homogeneous Lorentz transformations in the Hilbert spaces $H([m_1, s_1])$, $H([m_2, s_2])$ associated with relativistic particles of masses m_1, m_2 and spins s_1, s_2 respectively, then the corresponding generators in the Hilbert space

$$H[(m_1, s,] \otimes [m_2, s_2]) \tag{1.1a}$$

of states of the direct product representation

* Research supported in part by the U. S. Atomic Energy Commission.

[1] This problem has previously been studied by several authors whose work does not overlap the present work to a marked extent; L. Michel, Report on 1953 meeting of IUPAP, p. 272; E. P. Wigner, Nuovo Cimento **3**, 517 (1956); J. S. Lomont, J. Math. Phys. **1**, 237 (1960); G. C. Wick, Ann. Phys. **18**, 65 (1962). On the other hand, the authors of the following references pursue aims at least in part similar to those of the present work; Iu. M. Shirokov, Soviet Phys.—JETP **8**, 703 (1954); H. Joos, *Bemerkungen zur Phase Shift Analysis auf Grund der Darstellungstheorie der inhomogenen Lorentz gruppe*, Oberwolfach, Germany, 1959, (unpublished); H. Epstein, G. Luzzatto, and A. S. Wightman (to be published); B. Barsella, E. Fabri, Phys. Rev. **126** 1561 (1962). See also L. Pukánsky, J. Math. and Mech. **10**, 475 (1961). [I am grateful to the referee to bringing my attention to this paper, which belongs properly to the first group of references.]

[2] E. P. Wigner, Ann. Math. **40**, 149 (1939); V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. (Washington) **34**, 211 (1948). The subject has also been treated extensively in a series of papers by Iu. M. Shirokov, Soviet Phys.—JETP **6**, 664, 919, 929 (1958), *ibid.* **7**, 493 (1958), *ibid.* **9**, 620 (1959). Alternatively, the reader may consult the lectures by A. S. Wightman at Varenna, Nuovo Cimento Suppl. **14**, 81 (1959), and Les Houches, *Dispersion Relations and Elementary Particles*, edited by C. de Witt and R. Omnes, (John Wiley & Sons, Inc., New York, 1961), p. 159.

[3] Iu. M. Shirokov, Dokl. Akad. Nauk SSSR **94**, 857 (1954), and *op. cit.*, footnote 1; L. L. Foldy, Phys. Rev. **102**, 568 (1956); C. Fronsdal, Phys. Rev. **113**, 1568 (1959); H. Joos, *op. cit.*, footnote 1; V. I. Ritus, Soviet Phys.—JETP **13**, 240 (1961).

[4] We use only the letters α, β, γ and δ to denote tensorial indices in Minkowski space. This leaves us free to use the letters κ, λ, μ, ν and ρ to denote the spherical components of a spatial vector or as magnetic quantum numbers; such letters appear only as subscripts and are not subject to the summation convention. Cartesian tensorial indices in ordinary space are denoted by Latin letters and appear only as superscripts.

$$[m_1, s_1] \otimes [m_2, s_2] \tag{1.1b}$$

are related to them by

$$P^\alpha = P^{(1)\alpha} \otimes 1^{(2)} + 1^{(1)} \otimes P^{(2)\alpha}, \tag{1.2}$$

$$M^{\alpha\beta} = M^{(1)\alpha\beta} \otimes 1^{(2)} + 1^{(1)} \otimes M^{(2)\alpha\beta}, \tag{1.3}$$

where $1^{(1)}$ and $1^{(2)}$ are unit operators in the indicated spaces. Now it is well known[5] that the representation (1.1b) of P can be uniquely expressed as a direct integral[6] of irreducible representations of P. Or, put otherwise, the Hilbert space (1.1a) can be expressed as a direct integral over the Hilbert spaces $H_\eta([M, j])$ in which the irreducible constituents $[M, j]$ of (1.1b) are defined, according to the symbolic statement

$$H([m_1, s_1] \otimes [m_2, s_2]) = \sum_{j\eta} \int_{m_1+m_2}^{\infty} dM\, H_\eta([M, j]). \tag{1.4}$$

The notation used here[7] reflects the fact that P is not simply reducible in the sense of reference 7, so that multiple occurrence of equivalent irreducible representations in the direct product (1.1b) is possible. It does indeed occur unless $s_1 = s_2 = 0$ (see below) and the label η serves to distinguish in some manner the equivalent $[M, j]$. We now wish to investigate, using Eqs. (1.2) and (1.3), the manner in which the generators P^α, $M^{\alpha\beta}$ operate within each $H_\eta([M, j])$, given the canonical forms of the one-particle generators. The case of P^α, of course, is trivial. We may also state our objective in a different, more specific manner: to express the states of the direct product Hilbert space (1.1a) in terms of the states of the $H_\eta([M, j])$ in such a way that

(a) P^α and $M^{\alpha\beta}$ operate on the latter in canonical form,

(b) the scalar product in (1.1a) can be expressed as a direct integral (with respect to M) and direct sum (with respect to j and η) of the scalar product in $H_\eta([M, j])$ defined in some canonical manner [see Eqs. (2.6) and (5.27) below]. The latter result in a sense realizes in an explicit form the symbolic statement of Eq. (1.4).

Let us name the required expression for the states of (1.1a) the Clebsch–Gordan (C–G) series of P for (1.1b), and the coefficients of the states of each $H_\eta([M, j])$ which appear in it the C–G coefficients of P for (1.1b). An explicit formula for such C–G coefficients of P has been given before in the work of Shirokov[1] and Joos[1], and in previous work of the author,[8] but so far no complete proof of its correctness (rather only heuristic derivations) has emerged. Our procedure is to assume, in the writing of the C–G series of P for (1.1b), the formula of Shirokov and Joos for the C–G coefficients of P. Then by proving that the expression so obtained for the states of (1.1a) satisfies (a) and (b) as above stated, we indeed prove the formula correct. We first take the special case of particles of zero spin, and thereafter attend to the inevitable complication which a general treatment of spin introduces into the proof.

Notable features of our methods are the following. Firstly we study both $\mathbf{J}$ and $\mathbf{N}$;

$$2J^k = \epsilon^{klm} M^{lm}, \qquad N^k = M^{0k}. \tag{1.5}$$

Previous work has entirely neglected $\mathbf{N}$, which is one of its major shortcomings [see, especially, the discussion following Eq. (5.18) below]. Secondly, the separation of the contributions from the center of momentum (external) and relative motion (internal) of the two-particle system to the $M^{\alpha\beta}$ [see Eqs. (4.24) and (4.25) below] is entirely new. Thirdly, we demonstrate that one must view a two-particle state of given momentum from a reference frame in which its momentum appears to be zero, in order to see how its intrinsic angular momentum is defined in terms of the relative angular momentum and intrinsic spins of the individual particles. Fourthly, we draw attention to the remarkable identities (5.11) and (5.19) which play an important part in obtaining this demonstration.

The order of presentation of material is as follows. Section 2 contains a review of the necessary matter regarding the representations of P. In Sec. 3 we discuss the definition[9] of intrinsic angular momentum, and derive the canonical forms that $\mathbf{J}$ and $\mathbf{N}$ take within an $H([m, s])$. In Secs. 4 and 5, we carry through the program outlined above for the respective cases of the system of two relativistic particles without and with spins. Algebraic proofs

[5] See J. S. Lomont, *op. cit.*, footnote 1.

[6] This concept was developed by J. von Neumann, Ann. Math. **50**, 401 (1949), and R. Godemont, Ann. Math. **53**, 68 (1951). The reader may also refer to M. A. Naimark and S. V. Fomin, Amer. Math. Soc. Translations (Ser. 2) **20**, 55 (1962); J. Dixmier, *Les Algebres d' Operateurs dans L' Espace Hilbertien* (Gauthier-Villars, Paris, France, 1957); J. M. Jauch, *Lectures on the Representations of the Lorentz Group, Part IV* (CERN, Geneva, Switzerland, 1959).

[7] E. P. Wigner, Amer. Jour. Math. **63**, 57 (1941).

[8] A. J. Macfarlane, Revs. Mod. Phys. **34**, 41 (1962); A. J. Macfarlane, Ph.D. Thesis, University of London, 1961 (unpublished).

[9] Our approach is close to that employed in the references listed in footnote 3. See also Chou Kuang Chao and M. I. Shirokov, Soviet Phys.—JETP **7**, 851 (1958). A rather different discussion is given by L. Michel and A. S. Wightman, Phys. Rev. **98**, 1190 (1955); L. Michel, Nuovo Cimento Suppl. **14**, 95 (1959). See also the review article, D. M. Fradkin and R. H. Good, Rev. Mod. Phys. **33**, 343 (1961).

of certain results needed in Sec. 5 are subordinated into appendices A and B.

2. THE REPRESENTATIONS OF P

The Poincaré group P consists of transformations of the type

$$x^\alpha \to x'^\alpha = L^\alpha{}_\beta x^\beta + a^\alpha \tag{2.1}$$

in Minkowski space, where $L^\alpha{}_\beta$ describes a transformation of the restricted homogeneous Lorentz group.[10] If we denote the element (2.1) of P by (a, L), we can state the group multiplication law of P in the form

$$\begin{aligned}(a_1, L_1)(a_2, L_2) &= (a_3, L_3),\\ a_3 &= a_1 + L_1 a_2, \qquad L_3 = L_1 L_2.\end{aligned} \tag{2.2}$$

The representations of P are defined in terms of unitary operators

$$\begin{aligned}U(a, L) &= U(a, 1)U(0, L)\\ &\equiv U(a)U(L),\end{aligned}$$

satisfying

$$U(a_1, L_1)U(a_2, L_2) = U(a_3, L_3), \tag{2.3}$$

with a_3 and L_3 as given by (2.2). In the case of the unitary irreducible representation $[m, s]$ of P describing a system of mass m $(m > 0)$ and intrinsic angular momentum s, they act according to[2]

$$\begin{aligned}[U(a, L)\phi]_\nu(p) &= e^{ip\cdot a}[U(L)\phi]_\nu(p)\\ &= e^{ip\cdot a}\sum_\mu Q_{\nu\mu}(p, L)\phi_\mu(L^{-1}p),\end{aligned} \tag{2.4}$$

where the unitary matrices $Q(p, L)$ satisfy

$$Q(p, L_1)Q(L_1^{-1}p, L_2) = Q(p, L_3), \tag{2.5}$$

on the states[11]

$$\begin{aligned}\phi_s(m) = \{\phi_\nu(p) \mid p^2 = m^2, \omega = (\mathbf{p}^2 + m^2)^{\frac{1}{2}}\\ \geqslant m > 0, -s \leqslant \nu \leqslant s\}\end{aligned} \tag{2.6}$$

of the Hilbert space $H([m, s])$, wherein the scalar product is given by

$$(\phi_s(m), \psi_s(m)) = \sum_\nu \int \frac{d^3p}{(2\omega)}\phi_\nu(p)^*\psi_\nu(p). \tag{2.6}$$

Following Wigner,[2] we define a pure rotation

$$R(p, L) = L(p)\cdot L\cdot L(L^{-1}p)^{-1}, \tag{2.7}$$

where $L(p)$ is given by

$$\begin{aligned}&p^\alpha = (m, \mathbf{0}) = L(p)^\alpha{}_\beta p^\beta,\\ &L(p)^{00} = \omega/m, \quad L(p)^{0k} = -L(p)^{k0} = p^k/m,\\ &-L(p)^{kl} = \delta^{kl} + p^k p^l/[m(m + \omega)].\end{aligned} \tag{2.8}$$

Since $R(p, L)$ satisfies the law

$$R(p, L_1)R(L_1^{-1}p, L_2) = R(p, L_3) \tag{2.9}$$

of multiplication, it follows from (2.9) and (2.5) that we may set

$$Q(p, L) = D^s[R(p, L)], \tag{2.10}$$

where D^s is the usual $(2s + 1)$-dimensional matrix representation of the rotation group.[12]

We introduce the infinitesimal generators P^α, $M^{\alpha\beta}$ of translations and homogeneous Lorentz transformations by means of

$$U(a) = \exp\,[iP^\alpha a_\alpha], \tag{2.11}$$

and

$$U(L) = 1 - \tfrac{1}{2}i\omega_{\alpha\beta}M^{\alpha\beta}, \tag{2.12}$$

where

$$L^{\alpha\beta} = g^{\alpha\beta} + \omega^{\alpha\beta}, \qquad \omega^{\alpha\beta} + \omega^{\beta\alpha} = 0.$$

Then, from (2.3), we obtain the results

$$U(L)P^\alpha U(L)^{-1} = (L^{-1})^\alpha{}_\beta P^\beta, \tag{2.13}$$

$$U(L)M^{\alpha\beta}U(L)^{-1} = (L^{-1})^\alpha{}_\gamma(L^{-1})^\beta{}_\delta M^{\gamma\delta}, \tag{2.14}$$

which exhibit the tensorial character of the generators, and hence the familiar commutation relations

$$[P^\alpha, P^\beta] = 0, \tag{2.15}$$

$$[M^{\alpha\beta}, P^\gamma] = i(P^\alpha g^{\beta\gamma} - P^\beta g^{\alpha\gamma}), \tag{2.16}$$

$$\begin{aligned}[M^{\alpha\beta}, M^{\gamma\delta}] = -i(&g^{\alpha\gamma}M^{\beta\delta} + g^{\beta\delta}M^{\alpha\gamma}\\ &- g^{\alpha\delta}M^{\beta\gamma} - g^{\beta\gamma}M^{\alpha\delta}).\end{aligned} \tag{2.17}$$

In terms of $\mathbf{J}$ and $\mathbf{N}$, the generators of rotations and pure Lorentz transformation defined by (1.5), these become

$$\begin{aligned}&[P^\alpha, P^\beta] = 0,\\ &[J^k, P^0] = 0, \quad [J^k, P^l] = i\epsilon^{klm}P^m,\\ &[N^k, P^0] = -iP^k, \quad [N^k, P^l] = -i\delta^{kl}P^0,\\ &[J^k, J^l] = i\epsilon^{klm}J^m,\\ &[J^k, N^l] = i\epsilon^{klm}N^m \quad [N^k, N^l] = -i\epsilon^{klm}J^m.\end{aligned} \tag{2.18}$$

[10] Information regarding homogeneous Lorentz transformations may be sought in A. J. Macfarlane, J. Math. Phys. 3, 1116 (1962).

[11] The set-theoretic notation $S = \{x \mid P\}$ reads: S is the set of all elements x which satisfy the proposition(s) P.

[12] For a discussion of the properties of quantities associated with the rotation group, see A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1957); M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, Inc., New York, 1957).

We also require the pseudovector W^α defined by

$$W^\alpha = \tfrac{1}{2}\epsilon^{\alpha\beta\gamma\delta}P_\beta M_{\gamma\delta}, \tag{2.19}$$

and note the following properties it possesses:

$$W^0 = \mathbf{P}\cdot\mathbf{J}, \qquad \mathbf{W} = P^0\mathbf{J} + \mathbf{P}\times\mathbf{N}, \tag{2.20}$$

$$[P^\alpha, W^\beta] = 0, \tag{2.21}$$

$$P^\alpha W_\alpha = 0, \tag{2.22}$$

$$[W^\alpha, W^\beta] = -i\epsilon^{\alpha\beta\gamma\delta}P_\gamma W_\delta, \tag{2.23}$$

$$[M^{\alpha\beta}, W^\gamma] = i(W^\alpha g^{\beta\gamma} - W^\beta g^{\alpha\gamma}), \tag{2.24}$$

$$U(L)W^\alpha U(L)^{-1} = (\det L)(L^{-1})^\alpha{}_\beta W^\beta. \tag{2.25}$$[13]

Either the set (2.15)–(2.17) or the set (2.18) of commutation relations specifies the Lie algebra[14] of P. This algebra has for its invariants,[15]

$$I_1 = P^\alpha P_\alpha, \qquad I_2 = W^\alpha W_\alpha, \tag{2.26}$$

and the set Σ of six elements

$$I_1, I_2, P^k, O, \tag{2.27}$$

where O is one of the W^α or else a single linear function of the W^α constitutes a complete commuting set of elements (maximal Abelian subalgebra) within the Lie algebra. The irreducible representations of P are labelled by their eigenvalues of I_1 and I_2, m^2 and $-m^2s(s+1)$ in the case of $[m, s]$, and we may introduce a basis labeled by eigenvalues of P^k, O in the Hilbert space of states of these representations. We have used a basis of exactly this type in Eq. (2.4), as can be seen, in the case of P, by using Eqs. (2.4) and (2.11) to give

$$[P^k\phi]_\nu(p) = p^k\phi_\nu(p). \tag{2.27}$$

The precise definition of O which allows

$$[O\phi]_\nu(p) = \nu\phi_\nu(p) \tag{2.28}$$

to be written is investigated in the next section.

3. DEFINITION OF SPIN

To define the intrinsic spin of the relativistic particle, whose kinematics are described by the representation $[m, s]$ of P, we note that, in the rest frame, spin constitutes the entire angular momentum, and write

$$\begin{aligned}(\mathbf{J}^2\phi)_\nu(\tilde{p}) &= s(s+1)\phi_\nu(\tilde{p}),\\ (J_z\phi)_\nu(\tilde{p}) &= \nu\phi_\nu(\tilde{p}),\\ (J_-\phi)_\nu(\tilde{p}) &= [(s\mp\nu)(s\pm\nu+1)]^{\frac{1}{2}}\phi_{\nu\pm1}(\tilde{p}),\end{aligned} \tag{3.1}$$

where

$$\tilde{p} = (m, \mathbf{0}), \qquad J_- = J^1 \pm iJ^2, \qquad J_z = J^3.$$

By introducing the spherical components[16] J_ρ of $\mathbf{J}$, we can present (3.1) as a single equation.

$$(J_\rho\phi)_\nu(\tilde{p}) = \sum_\mu (S_s{}_\rho)_{\nu\mu}\phi_\mu(\tilde{p}), \tag{3.2}$$

where the spherical components $S_{s\rho}$ of the matrix vector $\mathbf{S}_s$ are given by[17]

$$(S_s{}_\rho)_{\nu\mu} = [s(s+1)]^{\frac{1}{2}}C(s\ 1\ s\ \mu\rho\nu). \tag{3.3}$$

Using (2.20), we may write (3.2) in the form

$$(W_\rho\phi)_\nu(\tilde{p}) = m\sum_\mu (S_s{}_\rho)_{\nu\mu}\phi_\mu(\tilde{p}),$$

and proceed to construct a spin operator for a state of arbitrary p by means of the following development:

$$\begin{aligned}&[U(L(p)^{-1})W_\rho U(L(p))\phi]_\nu(p)\\ &\quad= [W_\rho U(L(p))\phi]_\nu(\tilde{p})\\ &\quad= m\sum_\mu (S_s{}_\rho)_{\nu\mu}[U(L(p))\phi]_\mu(\tilde{p})\\ &\quad= m\sum_\mu (S_s{}_\rho)_{\nu\mu}\phi_\mu(p).\end{aligned} \tag{3.4}$$

In other words, to define the spin components of a state of arbitrary p, we preform a Lorentz transformation to that frame in which it appears to have zero momentum, in agreement with the procedure laid down by Wigner.[18] We may use (3.4) to define the spin operator $\mathbf{S}(p)$:

$$mS(p)^k = U(L(p)^{-1})W^kU(L(p)), \tag{3.5}$$

and with the aid of (2.25) and (2.8) write

$$mS(p)^k = L(p)^k{}_\alpha W^\alpha, \tag{3.6}$$

and[19]

$$mS(p)^k = W^k - (m+\omega)^{-1}p^kW^0. \tag{3.7}$$

We may use Eqs. (3.5) and (2.23) to prove that the components of $\mathbf{S}(p)$ satisfy

$$[S(p)^k, S(p)^l] = i\epsilon^{klm}S(p)^m, \tag{3.8}$$

[13] For an element L of the restricted Lorentz group, $\det L = 1$.

[14] See W. Pauli, *Lectures on Continuous Groups* (CERN, Geneva, Switzerland, 1956); E. C. G. Sudarshan, in "Structure of Dynamical Theories," *Brandeis Summer Institute Lectures on Theoretical Physics*, (Benjamin, New York, 1962), Volume 2, p. 144.

[15] We use the term for those scalar functions of the elements of the Lie algebra which commute with each element of the Lie algebra.

[16] The subscript ρ takes on the values $+1$, 0, -1 with $J_{-1} = \mp 2^{-1/2}J_-$, $J_0 = J_z$.

[17] L. C. Biedenharn, Ann. Phys. 4, 104 (1958).

[18] E. P. Wigner, Rev. Mod. Phys. 29, 255 (1957).

[19] Equation (3.7) is equivalent to one of the forms of relativistic spin operator used by M. H. L. Pryce, Proc. Roy. Soc. (London) **A150**, 166 (1935), *ibid.* **A195**, 62 (1948). See also A. Papapetrou, Prakt. Acad. Athenon 14, 540 (1939); *ibid.* 15, 404 (1940).

and also that

$$\mathbf{S}(p)^2 = -W^2/m^2 = s(s+1). \tag{3.9}$$

It follows from Eqs. (3.4), (3.8) and (3.9) that $\mathbf{S}(p)$ may be regarded as the spin operator of a state of momentum p. Hence, by replacing ω and $\mathbf{p}$ in (3.7) by the translation operators P^0 and $\mathbf{P}$, we obtain an operator $\mathbf{S}(P, W)$ given by

$$m\mathbf{S}(P, W) = \mathbf{W} - (m + P^0)^{-1}\mathbf{P}W^0, \tag{3.10}$$

which may be identified with the operator O of Sec. 2. For the z component of $\mathbf{S}(P, W)$, we have

$$\begin{aligned}[S(P, W)_z\phi]_\nu(p) &= (W_z\phi)_\nu(p) - (m+\omega)^{-1}p_z(W^0\phi)_\nu(p) \\ &= [S(p)_z\phi]_\nu(p) = \nu\phi_\nu(p).\end{aligned} \tag{3.11}$$

Of course $\mathbf{S}(P, W)$ shares with $\mathbf{S}(p)$ the essential properties (3.8) and (3.9).

We next consider the behavior of the operators $\mathbf{S}(p)$ and $\mathbf{S}(P, W)$ under homogeneous Lorentz transformations. Two points of view are available, one considering different states in the same reference frame, the other considering the same state as viewed from different reference frames. Within the first, we ask the question: if $\mathbf{S}(p)$ is the spin operator for a state of momentum p in a given frame, what relationship does it bear to the spin operator $\mathbf{S}(p')$ of a state of momentum p' ($p' = L^{-1}p$) in the same frame of reference? To answer it we write

$$\begin{aligned}&mU(L)S(L^{-1}p)^kU(L)^{-1} \\ &\quad = U(L\cdot L(L^{-1}p)^{-1})W^kU(L\cdot L(L^{-1}p)^{-1})^{-1} \\ &\quad = m[R(p, L)^{-1}]^{kl}S(p)^l,\end{aligned} \tag{3.12}$$

with $R(p, L)$ as given by (2.7). We have here used

$$R(L^{-1}p, L^{-1}) = R(p, L)^{-1}. \tag{3.13}$$

We may evaluate (3.12) for a state of momentum p obtaining

$$\begin{aligned}\sum_{\nu\lambda} D^s_{\mu\nu}[R(p, L)](S^k_s)_{\nu\lambda}D^s_{\lambda\kappa}[R(p, L)^{-1}] \\ = [R(p, L)^{-1}]^{kl}(S^l_s)_{\mu\kappa},\end{aligned} \tag{3.14}$$

which is exactly the $(2s+1)$-dimensional matrix representation of the fundamental formula[20] of the quaternion theory of rotations

$$R(A)^{ij}\tau^i = A\tau^jA^{-1}.$$

If one replaces W_ρ in (3.4) or J_ρ in (3.2) by $S(\vec{p})_\rho$, one can see that the definition of $S(p)_\rho$ agrees with (3.12) by noting that $R(p, L(p)^{-1}) = 1$.

[20] See Eq. (95) of reference 10.

In the second point of view, we see that if $\mathbf{S}(p)$ is the spin operator of a state of momentum p in a given frame of reference, then the spin operator of the same state, as viewed from a reference frame reached from the given one by application of the Lorentz transformation L, is related to it by

$$\begin{aligned}mS'(p')^k &= mS'(L^{-1}p)^k \\ &= L(L^{-1}p)^k{}_\alpha(L^{-1}W)^\alpha \\ &= m[R(p, L)^{-1}]^{kl}S(p)^l.\end{aligned} \tag{3.15}$$

Both of the results (3.12) and (3.15) are consistent with

$$U(L)\mathbf{S}(P, W)U(L)^{-1} = \mathbf{S}(L^{-1}P, L^{-1}W). \tag{3.16}$$

We next turn to the important task of developing the canonical forms of the operation of $\mathbf{J}$ and $\mathbf{N}$ within $H([m, s])$. The required result for $\mathbf{J}$ is obtained by evaluating (2.4) for the infinitesimal spatial rotation

$$L^{\alpha\beta} = g^{\alpha\beta} + \omega^{\alpha\beta}, \qquad \omega^{k0} = 0, \qquad \omega^{ln} = \epsilon^{lnk}n^k\theta,$$

for which[21]

$$U(L) = 1 - i\theta\mathbf{n}\cdot\mathbf{J},$$

and

$$D^s_{\nu\mu}[R(p, L)] = \delta_{\nu\mu} - i\theta\mathbf{n}\cdot(\mathbf{S}_s)_{\nu\mu}.$$

A simple calculation yields the result

$$(J_\rho\phi)_\nu(p) = -i(\mathbf{p}\times\partial)_\rho\phi_\nu(p) + \sum_\mu (S_{s\rho})_{\nu\mu}\phi_\mu(p), \tag{3.17}$$

where $\partial^k = \partial/\partial p^k$, or using (3.4),

$$(\mathbf{J}\phi)_\nu(p) = -i\mathbf{p}\times\partial\phi_\nu(p) + (\mathbf{S}(p)\phi)_\nu(p). \tag{3.18}$$

To treat the case of $\mathbf{N}$, we consider the pure Lorentz transformation

$$L^{\alpha\beta} = g^{\alpha\beta} + \omega^{\alpha\beta}, \qquad \omega^{kl} = 0, \qquad \omega^{k0} = n^k\chi,$$

for which

$$U(L) = 1 - i\chi\mathbf{n}\cdot\mathbf{N}.$$

Using the formalism of Sec. 5 of reference 10, we calculate the two-dimensional matrix of $SU(2C)$ that corresponds to the rotation $R(p, L)$, obtaining

$$A(p, L) = 1 + \tfrac{1}{2}(m+\omega)^{-1}i\boldsymbol{\tau}\cdot\mathbf{p}\times\mathbf{n}\chi,$$

and hence

$$D^s_{\nu\mu}[R(p, L)] = \delta_{\nu\mu} - (m_\circlearrowleft+\omega)^{-1}i\chi\mathbf{n}\cdot\mathbf{p}\times(\mathbf{S}_s)_{\nu\mu}.$$

[21] This agrees with Eq. (82) of reference 10, when one translates from a passive to an active viewpoint. See also the discussion of S. S. Schweber, *Introduction to Relativistic Quantum Field Theory*, (Row-Peterson, Evanston, Illinois, 1961), p. 165.

A direct calculation now leads to the result

$$(\mathbf{N}\phi)_\nu(p) = -i\omega\partial\phi_\nu(p) - (m+\omega)^{-1}\mathbf{p}\times(\mathbf{S}(p)\phi)_\nu(p). \quad (3.19)$$

Equations (3.18) and (3.19) give the canonical forms taken by $\mathbf{J}$ and $\mathbf{N}$[22] when they operate within $H([m, s])$. We note that insertion of (3.18) and (3.19) into (2.20) gives a canonical form for W^α, which can be inserted into (3.7) to provide a consistency check on our calculations.

4. THE RELATIVISTIC SYSTEM OF TWO SPINLESS PARTICLES

We here consider the physical system consisting of two noninteracting relativistic particles of masses m_1, and m_2 and zero intrinsic spins. This is described by the states

$$\phi(m_1, m_2) = \{\phi(p_1, p_2) \mid p_1^2 = m_1^2, \omega_1 \geq m_1 > 0, \; p_2^2 = m_2^2, \omega_2 \geq m_2 > 0\} \quad (4.1)$$

of the Hilbert space $H([m_1, 0] \otimes [m_2, 0])$, wherein scalar product is defined by

$$(\phi(m_1, m_2), \psi(m_1, m_2)) = \int d^3 p_1/(2\omega_1) \times \int d^3 p_2/(2\omega_2)\phi(p_1, p_2)^*\psi(p_1, p_2). \quad (4.2)$$

The transformation law under P is given by

$$[U(a, L)\phi](p_1, p_2) = [U^{(1)}(a, L) \otimes U^{(2)}(a, L)\phi](p_1, p_2) = \exp[i(p_1 + p_2)\cdot a]\phi(L^{-1}p_1, L^{-1}p_2). \quad (4.3)$$

Using (2.13) and (2.14), it follows that

$$P^\alpha = P^{(1)\alpha} \otimes 1^{(2)} + 1^{(1)} \otimes P^{(2)\alpha}, \quad (4.4)$$

$$M^{\alpha\beta} = M^{(1)\alpha\beta} \otimes 1^{(2)} + 1^{(1)} \otimes M^{(2)\alpha\beta}, \quad (4.5)$$

and hence, using (3.18) and (3.19), that

$$(\mathbf{J}\phi)(p_1, p_2) = (-i\mathbf{p}_1\times\partial_1 - i\mathbf{p}_2\times\partial_2)\phi(p_1, p_2), \quad (4.6)$$

$$(\mathbf{N}\phi)(p_1, p_2) = (-i\omega_1\partial_1 - i\omega_2\partial_2)\phi(p_1, p_2), \quad (4.7)$$

where $\partial_1 = \partial/\partial\mathbf{p}_1$, $\partial_2 = \partial/\partial\mathbf{p}_2$. Now (cf. Sec. 1), we know that

$$H([m_1, 0] \otimes [m_2, 0]) = \int_{m_1+m_2}^{\infty} dM \sum_{l=0}^{\infty} H([M, l]), \quad (4.8)$$

where M is seen from (4.4) to be given by

$$M^2 = (p_1 + p_2)^2, \quad (4.9)$$

and l is an intrinsic angular momentum quantum number whose relationship to the internal motion of the two-particle system will emerge in the course of the ensuing discussion. As a first step towards the determination of how $\mathbf{J}$ and $\mathbf{N}$ operate within the $H([M, l])$, it is evidently necessary that we introduce in place of p_1 and p_2, total and relative momentum variables. For the former, the choice

$$K = p_1 + p_2 = (\Omega(M) = (\mathbf{K}^2 + M^2)^{\frac{1}{2}}, \mathbf{K}) \quad (4.10)$$

is natural, and for the latter, the Wightman–Gording four-vector

$$q = M\lambda^{-\frac{1}{2}}(M)\{p_1 - p_2 - [(m_1^2 - m_2^2)/M^2](p_1 + p_2)\} \quad (4.11)$$

is used. This four-vector expression has been constructed so as to satisfy

$$K\cdot q = q^2 + 1 = 0, \quad (4.12)$$

and $\lambda(M)$ is an abbreviation

$$\lambda(M) = \lambda(M^2, m_1^2, m_2^2) = M^4 - 2M^2(m_1^2 + m_2^2) + (m_1^2 - m_2^2)^2. \quad (4.13)$$

In terms of K and q, (4.2) becomes[23]

$$(\phi(m_1, m_2), \psi(m_1, m_2)) = \int dM\, \lambda^{\frac{1}{2}}(M) \int d^3K/[2\Omega(M)] \times \int d^4q\, \delta(K\cdot q)\delta(q^2 + 1) \times \phi(M, \mathbf{K}, q)^*\psi(M, \mathbf{K}, q), \quad (4.14)$$

where $\phi(M, \mathbf{K}, q)$ has arisen from $\phi(p_1, p_2)$ by the replacement of p_1 and p_2 by their expressions in terms of M, $\mathbf{K}$, q as obtained by solving (4.10) and (4.11). Operating on such an object, we easily show[24] that $\mathbf{J}$ and $\mathbf{N}$ take the respective forms

$$(\mathbf{J}\phi)(M, \mathbf{K}, q) = [-i\mathbf{K}\times\partial/\partial\mathbf{K} - i\mathbf{q}\times\partial/\partial\mathbf{q}]\phi(M, \mathbf{K}, \mathbf{q}); \quad (4.15)$$

$$(\mathbf{N}\phi)(M, \mathbf{K}, q) = [-i\Omega(M)\partial/\partial\mathbf{K} - iq^0\partial/\partial\mathbf{q}]\phi(M, \mathbf{K}, q). \quad (4.16)$$

Since the results (4.14) to (4.16) are not yet in the

[22] We note that our results agree with those of Shirokov[1,2] provided that $\mathbf{N}$ (which he equates to M^{k0}) is replaced by $-\mathbf{N}$ throughout.

[23] See the last named reference of footnote 2 and those of footnote 8.

[24] R. Stora, "Multipole Expansions in Elementary Particle Physics," University of Maryland, Tech. Rept. No. 250 (unpublished).

forms we seek, we proceed to a second change of variables,[8] wherein

$$q^\alpha \to e^\alpha, \qquad e^\alpha = L(K)^\alpha_\beta q^\beta. \tag{4.17}$$

Using (2.8) and (4.12), we find

$$e^\alpha = (0, \mathbf{e}), \quad \mathbf{e} = \mathbf{q} - [M + \Omega(M)]^{-1} q^0 \mathbf{K}, \quad \mathbf{e}^2 = 1, \tag{4.18}$$

and

$$q^0 = \mathbf{K}\cdot\mathbf{e}/M, \tag{4.19}$$

$$\mathbf{q} = \mathbf{e} + \mathbf{K}\mathbf{K}\cdot\mathbf{e}/[M(M + \Omega(M))]. \tag{4.20}$$

The physical meaning of $\mathbf{e}$ is obtained as follows. From the definitions of $L(K)$ and q, we see that

$$e^k = M\lambda^{-\frac{1}{2}}(M) L(K)^k{}_\alpha [(p_1 - p_2)^\alpha + cK^\alpha]$$

regardless of c. If $M^2 c = -(m_1^2 - m_2^2)$, we have (4.17) exactly and also $e^0 = 0$. If $c = 1$, we see that $\mathbf{e}$ is a unit vector parallel to the spatial part of $L(K)p_1$, i.e. to the momentum of particle 1 as seen from a frame wherein the total momentum of the two-particle system is zero. Our choice of c is made so that the change of variables $q^\alpha \to e^\alpha$ in the integral (4.14) is easily accomplished.

To handle the change (4.17) of variables, we employ a temporary notation

$$\phi'(M, \mathbf{K}, \mathbf{e}) = \phi'(M, \mathbf{K}, \mathbf{e}(M, \mathbf{K}, q)) \equiv \phi(M, \mathbf{K}, q). \tag{4.21}$$

Partial differentiation with respect to $\mathbf{K}$ and $\mathbf{e}$ now obeys[25]

$$\frac{\partial\phi'}{\partial K^l} = \frac{\partial\phi}{\partial K^l} + \frac{\partial\phi}{\partial q^m}\frac{\partial q^m}{\partial K^l}, \tag{4.22}$$

$$\frac{\partial\phi'}{\partial e^l} = \frac{\partial\phi}{\partial q^m}\frac{\partial q^m}{\partial e^l}, \tag{4.23}$$

and we may employ Eqs. (4.18) to (4.23) to establish the results

$$[\mathbf{K}\times\partial/\partial\mathbf{K} + \mathbf{e}\times\partial/\partial\mathbf{e}]\phi' = [\mathbf{K}\times\partial/\partial\mathbf{K} + \mathbf{q}\times\partial/\partial\mathbf{q}]\phi,$$

$$[\Omega(M)\,\partial/\partial\mathbf{K} - (M + \Omega(M))^{-1}\mathbf{K}\times(\mathbf{e}\times\partial/\partial\mathbf{e})]\phi' = [\Omega(M)\,\partial/\partial\mathbf{K} + q^0\,\partial/\partial\mathbf{q}]\phi.$$

Hence we see that

$$(\mathbf{J}\phi)'(M, \mathbf{K}, \mathbf{e}) = [-i\mathbf{K}\times\partial/\partial\mathbf{K} - i\mathbf{e}\times\partial/\partial\mathbf{e}]\phi'(M, \mathbf{K}, \mathbf{e}), \tag{4.24}$$

$$(\mathbf{N}\phi)'(M, \mathbf{K}, \mathbf{e}) = [-i\Omega(M)\,\partial/\partial\mathbf{K} - (M + \Omega(M))^{-1} \times \mathbf{K}\times(-i\mathbf{e}\times\partial/\partial\mathbf{e})]\phi'(M, \mathbf{K}, \mathbf{e}). \tag{4.25}$$

We have evidently converted $\mathbf{J}$ and $\mathbf{N}$ to the required forms. We also note that the right side of (4.14) can be expressed[7] as

$$\int dM\,\lambda^{\frac{1}{2}}(M)/(4M) \int d^3K/[2\Omega(M)] \int d\mathbf{e}\,\phi'(M, \mathbf{K}, \mathbf{e})^*\psi'(M, \mathbf{K}, \mathbf{e}). \tag{4.26}$$

where $\int d\mathbf{e}$ means integration over the polar angles of the unit vector, $\mathbf{e}$. Having reached Eqs. (4.24) to (4.26), we have no further need for the variable q, and hereafter may omit the primes on ϕ, ψ.

We are now in a position to show explicitly how objectives (a) and (b) of Sec. 1 are achieved in the special case $s_1 = s_2 = 0$ of (1.1b). We define

$$\phi(p_1, p_2) \equiv \phi(M, \mathbf{K}, \mathbf{e}) = 2M^{\frac{1}{2}}\lambda^{-\frac{1}{2}}(M) \sum_{lm} Y_{lm}(\mathbf{e})\phi_{lm}(M, \mathbf{K}). \tag{4.27}$$

Then using (4.24) and

$$(-i\mathbf{e}\times\partial/\partial\mathbf{e})_\rho Y_{lm}(\mathbf{e}) = [l(l+1)]^{\frac{1}{2}} C(l\ 1\ lm\rho m') Y_{lm'}(\mathbf{e}), \tag{4.28}$$

we find

$$(J_\rho\phi)(p_1, p_2) = 2M^{\frac{1}{2}}\lambda^{-\frac{1}{2}}(M) \sum_{lm} Y_{lm}(\mathbf{e})(J_\rho\phi)_{lm}(M, \mathbf{K})$$

$$= 2M^{\frac{1}{2}}\lambda^{-\frac{1}{2}}(M)[\sum_{lm} Y_{lm}(\mathbf{e})(-i\mathbf{K}\times\partial/\partial\mathbf{K})_\rho\phi_{lm}(M, \mathbf{K}) + \sum_{lmm'} Y_{lm}(\mathbf{e})(S_{l\rho})_{mm'}\phi_{lm'}(M, \mathbf{K}),$$

so that, by orthogonality

$$(J_\rho\phi)_{lm}(M, \mathbf{K}) = \sum_{m'} [(-i\mathbf{K}\times\partial/\partial\mathbf{K})_\rho\delta_{mm'} + (S_{l\rho})_{mm'}]\phi_{lm'}(M, \mathbf{K}). \tag{4.29}$$

Hence, operating within $H([M, l])$, we have

$$(\mathbf{J}\phi)_{lm}(M, \mathbf{K}) = (-i\mathbf{K}\times\partial/\partial\mathbf{K})\phi_{lm}(M, \mathbf{K}) + (\mathbf{S}(K)\phi)_{lm}(M, \mathbf{K}), \tag{4.30}$$

where we have set

$$(S(K)_\rho\phi)_{lm}(M, \mathbf{K}) = \sum_{m'} (S_{l\rho})_{mm'}\phi_{lm'}(M, \mathbf{K}). \tag{4.31}$$

Similarly one proves that the operation of $\mathbf{N}$ within $H([M, l])$ obeys

[25] The prime in (4.21) has been introduced to remove the ambiguity in the usual notation for partial differentiation, which is reflected in Eq. (4.22).

$$(\mathbf{N}\phi)_{lm}(M, \mathbf{K}) = -i\Omega(M)\,\partial/\partial\mathbf{K}\phi_{lm}(M, \mathbf{K}) - [M + \Omega(M)]^{-1}\mathbf{K}\times(\mathbf{S}(K)\phi)_{lm}(M, \mathbf{K}). \quad (4.32)$$

Furthermore, from (4.26) and (4.27) and orthogonality, we get

$$(\phi(m_1, m_2), \psi(m_1, m_2)) = \int dM \sum_l (\phi_l(M), \psi_l(M)), \quad (4.33)$$

where we have defined the scalar product in $H([M, l])$ by

$$(\phi_l(M), \psi_l(M)) = \int d^3K/[2\Omega(M)] \times \sum_m \phi_{lm}(M, \mathbf{K})^*\psi_{lm}(M, \mathbf{K}). \quad (4.34)$$

Summarizing, we may define the states $\phi_{lm}(M, \mathbf{K})$ of each $H([M, l])$ that occurs in $H([m_1, 0] \otimes [m_2, 0])$ in such a way [Eq. (4.27)], that

(a) $\mathbf{J}$ and $\mathbf{N}$ operate on them in canonical form, [compare Eqs. (4.30) and (4.32) with Eqs. (3.18) and (3.19)];

(b) the scalar product in $H([m_1, 0] \otimes [m_2, 0])$ is expressed [Eq. (4.33)] as a direct integral with respect to M and sum with respect to l over the scalar product in $H([M, l])$ defined in canonical form, [compare Eq. (4.34) with Eq. (2.6)].

Equation (4.27) is what we name the C-G series of P for $[m_1, 0] \otimes [m_2, 0]$. If we write it in the form

$$\phi(p_1, p_2) = \int dM' \int d^3K'/(2K'^0) \times \sum_{lm} (p_1 0 p_2 0 \mid K'm[M'l])\phi_{lm}(M', \mathbf{K}'), \quad (4.35)$$

with

$$(p_1 0 p_2 0 \mid K'm[M'l]) = 2K^0\delta(\mathbf{K} - \mathbf{K}')\delta(M - M') \times 2M^{\frac{1}{2}}\lambda^{-\frac{1}{2}}(M)\,Y_{lm}(\mathbf{e}), \quad (4.36)$$

we can see that Eq. (4.27) agrees with the corresponding statements in a previous paper.[8] [See, in particular, Eqs. (2.10) and (3.17).]

The section closes with some comment on the definition (4.31) of $\mathbf{S}(K)$. We recall, firstly, that the notation $\mathbf{S}(K)$ implies transformation properties like (3.12) and (3.15) wherein K appears in a fundamental role, and secondly, that the contributions from $\mathbf{S}(K)$ to $\mathbf{J}$ and $\mathbf{N}$ in Eqs. (4.30) and (4.32) have arisen directly from the respective contributions to $\mathbf{J}$ and $\mathbf{N}$ in Eqs. (4.24) and (4.25) from the internal motion of the two-particle system. Thus, to exhibit the consistency of using the notation $\mathbf{S}(K)$, we must show that $\mathbf{e}\times\partial/\partial\mathbf{e}$ has the correct transformation law. Since $\mathbf{e}$ has been built by a well-defined procedure [Eqs. (4.11) and (4.17)] from p_1 and p_2, let us write $\mathbf{e}(p_1, p_2)$ as a temporary measure. We then easily show that

$$e(L^{-1}p_1, L^{-1}p_2)^k = [R(K, L)^{-1}]^{kl}e(p_1, p_2)^l,$$

and hence that, under L,

$$(\mathbf{e}\times\partial/\partial\mathbf{e})^k \to (\mathbf{e}'\times\partial/\partial\mathbf{e}')^k = [R(K, L)^{-1}]^{kl}(\mathbf{e}\times\partial/\partial\mathbf{e})^l, \quad (4.37)$$

where $\mathbf{e} \equiv \mathbf{e}(p_1, p_2)$, $\mathbf{e}' \equiv \mathbf{e}(L^{-1}p_1, L^{-1}p_2)$. Eq. (4.37) is sufficient to give the required assurance of consistency.

5. THE RELATIVISTIC SYSTEM OF TWO PARTICLES OF ARBITRARY SPINS

In this section, we extend the discussion of the preceding one to the case of particles of arbitrary spins s_1 and s_2. We now deal with the Hilbert space (1.1a) of states

$$\phi_{s_1s_2}(m_1, m_2) = \{\phi_{\nu_1\nu_2}(m_1, m_2) \mid p_i^2 = m_i^2, \omega_i \geq m_i > 0, -s_i \leq \nu_i \leq s_i, i = 1, 2\}, \quad (5.1)$$

with scalar product

$$(\phi_{s_1s_2}(m_1, m_2), \psi_{s_1s_2}(m_1, m_2)) = \int d^3p_1/(2\omega_1) \int d^3p_2/(2\omega_2) \times \sum_{\nu_1\nu_2} \phi_{\nu_1\nu_2}(p_1, p_2)^*\psi_{\nu_1\nu_2}(p_1, p_2). \quad (5.2)$$

As in Sec. 4, $\mathbf{J}$ and $\mathbf{N}$ are found to operate in (1.1a) according to

$$(\mathbf{J}\phi)_{\nu_1\nu_2}(p_1, p_2) = (-i\mathbf{p}_1\times\partial_1 - i\mathbf{p}_2\times\partial_2)\phi_{\nu_1\nu_2}(p_1, p_2) + [(\mathbf{S}_1(p_1) + \mathbf{S}_2(p_2))\phi]_{\nu_1\nu_2}(p_1, p_2), \quad (5.3)$$

$$(\mathbf{N}\phi)_{\nu_1\nu_2}(p_1, p_2) = (-i\omega_1\partial_1 - i\omega_2\partial_2)\phi_{\nu_1\nu_2}(p_1, p_2) - (m_1 + \omega_1)^{-1}\mathbf{p}_1\times(\mathbf{S}_1(p_1)\phi)_{\nu_1\nu_2}(p_1, p_2) - (m_2 + \omega_2)^{-1}\mathbf{p}_2\times(\mathbf{S}_2(p_2)\phi)_{\nu_1\nu_2}(p_1, p_2). \quad (5.4)$$

Here also the change of variables $\mathbf{p}_1$, $\mathbf{p}_2$ to M, $\mathbf{K}$, $\mathbf{e}$ may be effected[26]; we then have

$$\mathbf{p}_1\times\partial_1 + \mathbf{p}_2\times\partial_2 = \mathbf{K}\times\partial/\partial\mathbf{K} + \mathbf{e}\times\partial/\partial\mathbf{e}, \quad (5.5)$$

$$\omega_1\partial_1 + \omega_2\partial_2 = \Omega(M)\,\partial/\partial\mathbf{K} - [M + \Omega(M)]^{-1}\mathbf{K}\times(\mathbf{e}\times\partial/\partial\mathbf{e}), \quad (5.6)$$

[26] We regard this as being effected at a single step. When we consider the operation of $\partial/\partial\mathbf{K}$ and $\partial/\partial\mathbf{e}$ on $\phi_{\nu_1\nu_2}(p_1, p_2)$, we shall assume that p_1 and p_2 have been replaced therein by their expressions in terms of M, $\mathbf{K}$, $\mathbf{e}$.

$$\int \frac{d^3p_1}{2\omega_1} \int \frac{d^3p_2}{2\omega_2} = \int dM \frac{\lambda^{\frac{1}{2}}(M)}{4M} \int \frac{d^3K}{2\Omega(M)} \int d\mathbf{e}. \quad (5.7)$$

It is evident from the form of Eq. (5.4) for **N**, that we cannot proceed from here to a solution of our problem as directly as we did in Sec. 4. Specifically, what we cannot yet see is how to compound the contributions from the relative angular momentum and intrinsic spins of particles 1 and 2 to form the intrinsic angular momentum of the two-particle system in such a way that both **J** and **N** are obtained in canonical form. In accordance with the discussion of Sec. 3, the suggested procedure is to view the operation of **J** and **N** on $\phi_{\nu_1\nu_2}(p_1, p_2)$ from a reference frame wherein K appears as $\mathring{K} = (M, \mathbf{0})$. Therefore, using Eqs. (2.4) and (2.10), we introduce[27]

$$\begin{aligned}\phi'_{\nu_1\nu_2}(p_1, p_2) &= [U(L(K))\phi]_{\nu_1\nu_2}(L(K)p_1, L(K)p_2), \\ &= \sum_{\mu_1\mu_2} D^{s_1}_{\nu_1\mu_1}(R_1^{-1}) D^{s_2}_{\nu_2\mu_2}(R_2^{-1}) \phi_{\mu_1\mu_2}(p_1, p_2), \quad (5.9)\end{aligned}$$

where[28]

$$R_1 \equiv R(p_1, K) = R(p_1, L(K)^{-1}),$$
$$R_2 \equiv R(p_2, K) = R(p_2, L(K)^{-1}).$$

We now use

$$\begin{aligned}&\phi_{\nu_1\nu_2}(p_1, p_2) \\ &\quad = \sum_{\mu_1\mu_2} D^{s_1}_{\nu_1\mu_1}(R_1) D^{s_2}_{\nu_2\mu_2}(R_2) \phi'_{\mu_1\mu_2}(p_1, p_2), \quad (5.10)\end{aligned}$$

and Eqs. (5.3) and (5.4), to evaluate $(\mathbf{J}\phi'_{\mu_1\mu_2})(p_1, p_2)$ and $(\mathbf{N}\phi)'_{\nu_1\nu_2}(p_1, p_2)$. The first thing that one notices is that the differential operators in (5.3) and (5.4) effect not only $\phi'_{\mu_1\mu_2}(p_1, p_2)$, but also the D matrices in (5.10). Far from being an embarrassment, this is exactly the circumstance that enables us to obtain canonical forms for **J** and **N**. In the case of **J**, we have at our disposal the remarkable result (proved in Appendix A)

$$\begin{aligned}&(-i\mathbf{p}_1 \times \partial_1 - i\mathbf{p}_2 \times \partial_2)_\rho D^{s_1}_{\nu_1\mu_1}(R_1) \\ &\quad = \sum_{\lambda_1} D^{s_1}_{\nu_1\lambda_1}(R_1)(S_{s_1\rho})_{\lambda_1\mu_1} \\ &\qquad - \sum_{\kappa_1} (S_{s_1\rho})_{\nu_1\kappa_1} D^{s_1}_{\kappa_1\mu_1}(R_1). \quad (5.11)\end{aligned}$$

With the aid of (5.11) and its analog for $D^{s_2}(R_2)$, we obtain[29]

$$\begin{aligned}&(J_\rho\phi)_{\nu_1\nu_2}(p_1, p_2) \\ &\quad = \sum_{\mu_1\mu_2} D^{s_1}_{\nu_1\mu_1}(R_1) D^{s_2}_{\nu_2\mu_2}(R_2)(J_\rho\phi)'_{\mu_1\mu_2}(p_1, p_2) \quad (5.12) \\ &\quad = \sum_{\mu_1\mu_2\lambda_1\lambda_2} D^{s_1}_{\nu_1\lambda_1}(R_1) D^{s_2}_{\nu_2\lambda_2}(R_2) \\ &\qquad \times [(-i\mathbf{p}_1 \times \partial_1 - i\mathbf{p}_2 \times \partial_2)_\rho \delta_{\lambda_1\mu_1}\delta_{\lambda_2\mu_2} \\ &\qquad + (S_{s_1\rho})_{\lambda_1\mu_1}\delta_{\lambda_2\mu_2} \\ &\qquad + \delta_{\lambda_1\mu_1}(S_{s_2\rho})_{\lambda_2\mu_2}]\phi'_{\mu_1\mu_2}(p_1, p_2), \quad (5.13)\end{aligned}$$

and hence, by virtue of the unitary property of D matrices,

$$\begin{aligned}&(J_\rho\phi)'_{\lambda_1\lambda_2}(p_1, p_2) \\ &\quad = \sum_{\mu_1\mu_2} [(-i\mathbf{p}_1 \times \partial_2 - i\mathbf{p}_2 \times \partial_2)_\rho \delta_{\lambda_1\mu_1}\delta_{\lambda_2\mu_2} \\ &\qquad + (S_{s_1\rho})_{\lambda_1\mu_1}\delta_{\lambda_2\mu_2} \\ &\qquad + \delta_{\lambda_1\mu_1}(S_{s_2\rho})_{\lambda_2\mu_2}]\phi'_{\mu_1\mu_2}(p_1, p_2). \quad (5.14)\end{aligned}$$

To put this into a form comparable with (5.3), we introduce $\mathbf{S}_1(p_1)'$ by means of

$$\begin{aligned}&[S_1(p_1)'_\rho\phi]'_{\lambda_1\lambda_2}(p_1, p_2) \\ &\quad = \sum_{\mu_1} (S_{s_1\rho})_{\lambda_1\mu_1}\phi'_{\mu_1\lambda_2}(p_1, p_2), \quad (5.15)\end{aligned}$$

and $\mathbf{S}_2(p_2)'$ by a similar equation. From (5.15), (5.8) and (3.12), it is seen that

$$\begin{aligned}S_1(p_1)'^k &= U(L(K)^{-1})S_1(L(K)p_1)^k U(L(K)) \quad (5.16) \\ &= (R_1^{-1})^{kl} S_1(p_1)^l. \quad (5.17)\end{aligned}$$

The form of (5.17), certainly not a surprising one, is illustrative of a well known feature of relativistic kinematics[30]: the apparent effect of our change of viewpoint on the spin operator of particle 1 is that it is rotated. Naturally a similar discussion applies to particle 2. Using (5.15) and its analog, (5.14) can be put into the form[26]

$$\begin{aligned}(\mathbf{J}\phi)'_{\mu_1\mu_2}(p_1, p_2) &= -i\mathbf{K} \times \partial/\partial\mathbf{K}\phi'_{\mu_1\mu_2}(p_1, p_2) \\ &\quad + \{-i\mathbf{e} \times \partial/\partial\mathbf{e}\phi'_{\mu_1\mu_2}(p_1, p_2) \\ &\quad + [(\mathbf{S}_1(p_1)' + \mathbf{S}_2(p_2)')\phi]'_{\mu_1\mu_2}(p_1, p_2)\}. \quad (5.18)\end{aligned}$$

We observe that both (5.3) and (5.18) for **J** are apparently canonical in form; until we look at **N**, we have no reason to prefer one to the other. However, we already know that equation (5.4) for **N** is not in canonical form. We now show that $(\mathbf{N}\phi)'_{\mu_1\mu_2}(p_1, p_2)$ is.

To take account of the effect of the differential operators in (5.4) on the D matrices of (5.10), we

[27] In order that this be a consistent notation, we must (and, as is easily proved, do) have $[\mathbf{P}^{(1)}\phi]'_{\nu_1\nu_2}(p_1, p_2) = p_1\phi'_{\nu_1\nu_2}(p_1, p_2)$.

[28] The analogous equation in a previous paper (Cited in footnote 8) contains a misprint; the $L(p)$ that appears in it should be replaced by $L(p)^{-1}$.

[29] The contribution to the right-hand side of (5.12) from the second term of (5.11) exactly cancels the contribution from the term of (5.3) involving $\mathbf{S}_1(p_1)$; similarly for particle 2.

[30] H. P. Stapp, Phys. Rev. **103**, 425 (1956). See also V. I. Ritus, footnote 3.

employ a result (proved in Appendix A), which is no less remarkable than Eq. (5.11), namely

$$(-i\omega_1\partial_1 - i\omega_2\partial_2)D^{s_1}_{\nu_1\mu_1}(R_1)$$
$$= (m_1 + \omega_1)^{-1}\sum_{\lambda_1}(\mathbf{p}_1 \times \mathbf{S}_{s_1})_{\nu_1\lambda_1}D^{s_1}_{\lambda_1\mu_1}(R_1)$$
$$- (M + \Omega(M))^{-1}\sum_{\kappa_1} D^{s_1}_{\nu_1\kappa_1}(R_1)(\mathbf{K} \times \mathbf{S}_{s_1})_{\kappa_1\mu_1}. \quad (5.19)$$

Using (5.19) and its analog for $D^{s_2}(R_2)$, we prove the result[31]

$$(\mathbf{N}\phi)'_{\mu_1\mu_2}(p_1, p_2) = -i\Omega(M)\,\partial/\partial\mathbf{K}\phi'_{\mu_1\mu_2}(p_1, p_2)$$
$$- [M + \Omega(M)]^{-1}\mathbf{K} \times \{-i\mathbf{e} \times \partial/\partial\mathbf{e}\phi'_{\mu_1\mu_2}(p_1, p_2)$$
$$+ [(\mathbf{S}_1(p_1)' + \mathbf{S}_2(\phi_2)')\phi]'_{\mu_1\mu_2}(p_1, p_2)\}. \quad (5.20)$$

Equations (5.18) and (5.20) are the required canonical forms for $\mathbf{J}$ and $\mathbf{N}$, with the intrinsic angular momentum contributions to be formed (as we shall see, in particular, in appendix B) by the ordinary rules of vector coupling. This agrees with the discussion of Sec. 3, where to define the spin of a particle in a state of momentum $\mathbf{p}$ in a given frame, we had to view the state from a frame of reference wherein its momentum is $\tilde{\mathbf{p}} = 0$. Here the point is, that to see how to build up the intrinsic angular momentum of the state of the two-particle system of total momentum $\mathbf{K}$ in a given frame—from the relative angular momentum and intrinsic spins of the individual particles—one must view the state from a reference frame wherein its momentum is $\tilde{\mathbf{K}} = 0$. With the aid of (5.18) and (5.20), we may proceed directly to the implementation of the program laid down in the introduction.

In analogy to (4.27), we set

$$\phi'_{\mu_1\mu_2}(p_1, p_2) = 2M^{\frac{1}{2}}\lambda^{-\frac{1}{2}}(M)\sum_{\mu\lambda} C(s_1s_2s\mu_1\mu_2\mu)$$
$$\times C(lsj\lambda\mu m)Y_{l\lambda}(\mathbf{e})\phi'_{j(ls)m}(M, \mathbf{K}), \quad (5.21)$$

and obtain [see appendix B], from Eqs. (5.18) and (5.20), the results

$$(\mathbf{J}\phi)'_{j(ls)m}(M, \mathbf{K}) = \sum_{m'}[-i\mathbf{K} \times \partial/\partial\mathbf{K}\delta_{mm'}$$
$$+ (\mathbf{S}_j)_{mm'}]\phi'_{j(ls)m'}(M, \mathbf{K}), \quad (5.22)$$
$$(\mathbf{N}\phi)'_{j(ls)m}(M, \mathbf{K}) = \sum_{m'}[-i\Omega(M)\,\partial/\partial\mathbf{K}\delta_{mm'}$$
$$- (M + \Omega(M))^{-1}(\mathbf{K} \times \mathbf{S}_j)_{mm'}]\phi'_{j(ls)m'}(M, \mathbf{K}). \quad (5.23)$$

The formation of the intrinsic angular momentum of the two-particle system having been effected, we may now revert to our original view point [that of Eq. (5.8)], obtaining without difficulty[32]

$$(\mathbf{J}\phi)_{j(ls)m}(M, \mathbf{K}) = \sum_{m'}[-i\mathbf{K} \times \partial/\partial\mathbf{K}\delta_{mm'}$$
$$+ (\mathbf{S}_j)_{mm'}]\phi_{j(ls)m'}(M, \mathbf{K}), \quad (5.24)$$
$$(\mathbf{N}\phi)_{j(ls)m}(M, \mathbf{K}) = \sum_{m'}[-i\Omega(M)\,\partial/\partial\mathbf{K}\delta_{mm'}$$
$$- (M + \Omega(M))^{-1}(\mathbf{K} \times \mathbf{S}_j)_{mm'}]\phi_{j(ls)m'}(M, \mathbf{K}), \quad (5.25)$$

in manifestly canonical form. One may also prove, using Eqs. (5.2) and (5.7), Eqs. (5.10) and (5.21), and familiar properties[11] of D matrices, spherical harmonics, and vector coupling coefficients, that[33]

$$(\phi_{s_1s_2}(m_1, m_2), \psi_{s_1s_2}(m_1, m_2))$$
$$= \int_{m_1+m_2}^{\infty} dM \sum_{l=0}^{\infty}\sum_{s}\sum_{j}(\phi_{j(ls)}(M), \psi_{j(ls)}(M)), \quad (5.26)$$

where

$$(\phi_{j(ls)}(M), \psi_{j(ls)}(M))$$
$$= \int d^3K/[2\Omega(M)]$$
$$\times \sum_{m}\phi_{j(ls)m}(M, \mathbf{K})^*\psi_{j(ls)m}(M, \mathbf{K}). \quad (5.27)$$

Quite evidently, the functions $\phi_{j(ls)m}(M, \mathbf{K})$ are the states of a family of Hilbert spaces $H_{ls}([M, j])$, with canonically defined scalar products (5.27), within which $\mathbf{J}$ and $\mathbf{N}$ operate in canonical form [Eqs. (5.24) and (5.25)]. Further, Eq. (5.26) is an explicit realization of the symbolic reduction formula (1.4). It is also evident from Eqs. (5.24), (5.25), and (5.27), that the Hilbert spaces $H_{ls}([M, j])$ for different values of l and s but the same values of M and j are all equivalent to each other. In other words, l and s together constitute the label η of Sec. 1, which enumerates the multiplicity of equivalent $[M, j]$ that occur in the direct product (1.1b). For given j, all positive integral values of l allowed by the vector-coupling lsj occur for each s allowed by the vector coupling s_1s_2s.

As in Sec. 4, we note agreement with the results of a previous paper,[8] [see Eqs. (2.10) and (3.17) there]. To see this, we may combine Eqs. (5.10) and (5.21) to obtain.[32]

---

[31] In this case, the contribution from the first term in (5.19) cancels the unwanted contribution from the term of (5.4) involving $\mathbf{S}_1(p_1)$.

[32] We have here used the result $\phi'_{j(ls)m}(M, \mathbf{K}) = [U(L(K))\phi]_{j(ls)m}(M, \tilde{\mathbf{K}}) = \phi_{j(ls)m}(M, \mathbf{K})$.

[33] The sum over s involves all values consistent with the vector coupling s_1s_2s. That over j involves all values consistent with the vector coupling lsj for each such s and each positive integral value of l.

$$\phi_{\nu_1\nu_2}(p_1, p_2) = \int dM' \int d^3K'/(2K'^0)$$

$$\times \sum_{jlsm} (p_1\nu_1 p_2\nu_2 \,|\, K'm[M', j], l_s)\phi_{j(ls)m}(M', \mathbf{K}'), \quad (5.28)$$

where

$$(p_1\nu_1 p_2\nu_2 \mid K'm[M'j], ls)$$
$$= 2K^0\delta(\mathbf{K} - \mathbf{K}')\delta(M - M')2M^{\frac{1}{2}}\lambda^{-\frac{1}{4}}(M)$$
$$\times \sum_{\mu_1\mu_2\lambda} D^{s_1}_{\nu_1\mu_1}(R_1)D^{s_2}_{\nu_2\mu_2}(R_2)C(s_1s_2s\mu_1\mu_2\mu)$$
$$\times C(ls j\lambda\mu m)Y_{l\lambda}(\mathbf{e}). \quad (5.29)$$

Equation (5.29) is the formula for the C–G coefficient of P which occurs in the reduction of the representation (1.1b) of P. Its insertion into (5.28) yields explicitly the C–G series of P for that representation.

In Eq. (5.21) and the subsequent development, we might equally well have used alternative coupling schemes.

APPENDIX A

We here give proofs of the identities (5.11) and (5.19). We first consider the special case of (5.11) for $s_1 = \frac{1}{2}$. As is well known, $\mathbf{S}_{\frac{1}{2}}$ can be expressed in terms of Pauli matrices [cf Eq. (3.3)] according to

$$\mathbf{S}_{\frac{1}{2}} = \tfrac{1}{2}\boldsymbol{\tau}.$$

Thus the equation we seek to prove can be written in matrix notation:

$$2(-i\mathbf{p}_1 \times \boldsymbol{\partial}_1 - i\mathbf{p}_2 \times \boldsymbol{\partial}_2)D^{\frac{1}{2}}(R_1)$$
$$= D^{\frac{1}{2}}(R_1)\boldsymbol{\tau} - \boldsymbol{\tau}D^{\frac{1}{2}}(R_1). \quad (A1)$$

It is obvious that we need an explicit representation of $D^{\frac{1}{2}}(R_1)$. This may be obtained as follows:

The formalism of Sec. 5 of a previous paper[10] allows us to calculate the matrix A_1 of $SU(2C)$, which is the image of R_1 in the homomorphism of $SU(2C)$ to the rotation group. We obtain

$$A_1 = \frac{[(m_1+\omega_1)(M+\Omega(M)) - \mathbf{p}_1\cdot\mathbf{K} - i\boldsymbol{\tau}\cdot\mathbf{p}_1\times\mathbf{K}]}{[2(m_1+\omega_1)(M+\Omega(M))(m_1M+\mathbf{p}_1\cdot\mathbf{K})]^{\frac{1}{2}}}, \quad (A2)$$

a result first obtained by Shirokov.[34] Since, by definition almost,

$$D^{\frac{1}{2}}(R_1) = A_1, \quad (A3)$$

we have the required explicit form for $D^{\frac{1}{2}}(R_1)$. The proof of (A1) now follows after a simple use of the usual methods of vector algebra and the commutation relations of Pauli matrices.

Having proved (5.11) true for $s_1 = \frac{1}{2}$, we assume it true for some fixed s_1, and seek on this basis to prove it true for $(s_1 + \frac{1}{2})$, and hence, by induction, for all relevant s_1. If we use the definition (3.3) of the matrix vectors $\mathbf{S}_{s_1}$, and the identity[35] which relates $D^{s_1+\frac{1}{2}}(R_1)$ to products of $D^{s_1}(R_1)$ and $D^{\frac{1}{2}}(R_1)$, proof of the inductive step becomes an exercise in the standard techniques of Racah algebra. We omit the details.

Equation (5.19) is proved in like manner with only a little more effort.

We note, for the sake of completeness, that Eq. (A2) may be used along with Eq. (96) of reference 10 to give an explicit representation of the rotation R_1:

$$R_1^{jl} = \delta^{jl} + \frac{K^jp_1^l}{m_1M + \mathbf{p}_1\cdot\mathbf{K}}$$
$$- p_1^jp_1^l\frac{\Omega(M) - M}{(m_1+\omega_1)(m_1M+\mathbf{p}_1\cdot\mathbf{K})}$$
$$- K^jK^l\frac{\omega_1 - m_1}{(M+\Omega(M))(m_1M+\mathbf{p}_1\cdot\mathbf{K})}$$
$$+ p_1^jK^l\frac{2\mathbf{p}_1\cdot\mathbf{K} - (m_1+\omega_1)(M+\Omega(M))}{(m_1+\omega_1)(M+\Omega(M))(m_1M+\mathbf{p}_1\cdot\mathbf{K})}. \quad (A4)$$

This result also is due to Shirokov.[36]

APPENDIX B

We seek to prove that Eqs. (5.22) and (5.23) follow from (5.21), (5.18), and (5.20).

It is readily discovered that this reduces (in each case) to proving that

$$(S_{l\rho})_{\lambda\lambda'}C(s_1s_2s\mu_1\mu_2\mu)C(ls j\lambda'\mu m')$$
$$+ (S_{s_1\rho})_{\mu_1\lambda_1}C(s_1s_2s\lambda_1\mu_2\mu')C(ls j\lambda\mu' m')$$
$$+ (S_{s_2\rho})_{\mu_2\lambda_2}C(s_1s_2s\mu_1\lambda_2\mu')C(ls j\lambda\mu' m')$$
$$= (S_{j\rho})_{mm'}C(s_1s_2s\mu_1\mu_2\mu)C(ls j\lambda\mu m). \quad (B1)$$

The required proof consists of an application of the recurrence relation[37]

$$(S_{j_1\rho})_{m_1n_1}C(j_1j_2j_3n_1m_2m_3)$$
$$+ (S_{j_2\rho})_{m_2n_2}C(j_1j_2j_3m_1n_2m_3)$$
$$= (S_{j_3\rho})_{n_3m_3}C(j_1j_2j_3m_1m_2n_3), \quad (B2)$$

first to the s_1s_2s vector coupling, and then to the lsj vector coupling.

---

[34] Iu. M. Shirokov, Dokl. Akad. Nauk SSSR 99, 737 (1954).

[35] Equation (4.34), A. R. Edmonds, *op. cit.*, footnote 12.

[36] The formula which Shirokov gives [Eq. (9) of the paper cited in footnote 1) contains misprints. However a subsequent use of it [Eq. (41), there] agrees with Eq. (A4).

[37] This is equivalent to the recursion relation, satisfied by vector-coupling coefficients, which Rose (*op. cit.*, footnote 12) uses in his discussion of the Wigner–Eckart theorem.

Derivation of the Gell-Mann–Okubo Mass Formula

H. GOLDBERG AND Y. LEHRER-ILAMED
Israel Atomic Energy Commission Laboratories, Rehovoth, Israel
(Received 8 November 1962)

A short proof is given for a mathematical identity which is used to derive the Gell-Mann–Okubo mass formula.

IT has been proved by S. Okubo[1] that if T^{μ}_{ν} is a tensor operator [under $U(3)$], then any of its irreducible representations may be written in the form

$$T^{\mu}_{\nu} = a\delta^{\mu}_{\nu} + bA^{\mu}_{\nu} + c\sum_{\lambda} A^{\mu}_{\lambda}A^{\lambda}_{\nu}. \tag{1}$$

A^{μ}_{ν} are the representations of the infinitesimal operators of $U(3)$, whereas a, b, and c are functions of the representation. This formula, together with the assumption concerning the mass splitting, namely:

$$\Delta M = \langle i|\, T^3_3 \,|i\rangle, \tag{2}$$

give a mass formula for $U(3)$ symmetries both in the Sakata[2] and the Gell-Mann–Ne'eman[3] models.

It is to be mentioned that Gell-Mann[3] derived the mass formula for the baryons before Okubo's work by assuming (2) for the special case of an octet.

The proof of (1), as given by Okubo, involves somewhat lengthy calculations which obscure the generality of the result. We should like to expose here a proof which is applicable—with obvious modifications—to any semisimple Lie Algebra.

Let $A = (A^{\mu}_{\nu})$ $(1 \leq \mu, \nu \leq 3)$ be a matrix whose elements A^{μ}_{ν} belong to a commutative ring. In this case we have the Cayley–Hamilton identity: If

$$f(x) = \det(xI - A) = a + bx + cx^2 + x^3, \tag{3}$$

then

$$f(A) = aI + bA + cA^2 + A^3 \equiv 0. \tag{4}$$

The identity (4) may be formulated in the following way:[4] Define

$$C^{\mu}_{\nu} = \sum_{\alpha,\beta} A^{\mu}_{\alpha}A^{\alpha}_{\beta}A^{\beta}_{\nu} - \sum_{\alpha,\beta} A^{\alpha}_{\alpha}A^{\mu}_{\beta}A^{\beta}_{\nu} + \sum_{\alpha,\beta}(\delta^{12}_{\alpha\beta}A^{\alpha}_{1}A^{\beta}_{2}A^{\mu}_{\nu} + \delta^{13}_{\alpha\beta}A^{\alpha}_{1}A^{\beta}_{3}A^{\mu}_{\nu} + \delta^{23}_{\alpha\beta}A^{\alpha}_{2}A^{\beta}_{3}A^{\mu}_{\nu}) - \sum_{\alpha,\beta,\gamma}\varepsilon_{\alpha\beta\gamma}A^{\alpha}_{1}A^{\beta}_{2}A^{\gamma}_{3}\delta^{\mu}_{\nu}, \tag{5}$$

where

$$\delta^{\rho\sigma}_{\tau\omega} = \begin{cases} 1 & \text{if } \rho = \tau, \ \sigma = \omega \\ -1 & \text{if } \rho = \omega, \ \sigma = \tau \end{cases} \Bigg\} \ \rho \neq \sigma, \tag{6}$$

0 in all other cases.

δ^{μ}_{ν} is the Kroneker delta function, and $\varepsilon_{\alpha\beta\gamma}$ is the basic antisymmetric tensor.

The Cayley–Hamilton identity is equivalent to the identities

$$C^{\mu}_{\nu} \equiv 0. \tag{7}$$

When the elements A^{μ}_{ν} belong to an associative (noncommutative) free algebra, the identities (7) are generally not valid. However, we may obtain similar identities by symmetrizing the monomials appearing in C^{μ}_{ν}. Namely, if

$$P(abc) = abc + acb + bac + bca + cab + cba, \tag{8}$$

and if

$$6H^{\mu}_{\nu} = \sum_{\alpha,\beta} P(A^{\mu}_{\alpha}A^{\alpha}_{\beta}A^{\beta}_{\nu}) - \sum_{\alpha,\beta} P(A^{\alpha}_{\alpha}A^{\mu}_{\beta}A^{\beta}_{\nu}) - \sum_{\alpha,\beta,\gamma}\varepsilon_{\alpha\beta\gamma}P(A^{\alpha}_{1}A^{\beta}_{2}A^{\gamma}_{3})\delta^{\mu}_{\nu} + \sum_{\alpha,\beta}\delta^{12}_{\alpha\beta}P(A^{\alpha}_{1}A^{\beta}_{2}A^{\mu}_{\nu}) + \sum_{\alpha,\beta}\delta^{13}_{\alpha\beta}P(A^{\alpha}_{1}A^{\beta}_{3}A^{\mu}_{\nu}) + \sum_{\alpha,\beta}\delta^{23}_{\alpha\beta}P(A^{\alpha}_{2}A^{\beta}_{3}A^{\mu}_{\nu}), \tag{9}$$

then[4]

$$H^{\mu}_{\nu} \equiv 0. \tag{10}$$

When A^{μ}_{ν} are operators of a Lie Algebra, the difference

$$C^{\mu}_{\nu} - H^{\mu}_{\nu} \tag{11}$$

is a polynomial of the second degree in A^{μ}_{ν}. Therefore,

$$\sum_{\alpha,\beta} A^{\mu}_{\alpha}A^{\alpha}_{\beta}A^{\beta}_{\nu} = a\delta^{\mu}_{\nu} + bA^{\mu}_{\nu} + c\sum_{\alpha} A^{\mu}_{\alpha}A^{\alpha}_{\nu}, \tag{12}$$

[1] S. Okubo, Prog. Theoret. Phys. (Kyoto) **27**, 949 (1962).
[2] See, for example, M. Ikeda, S. Ogawa, and Y. Ohnuki, Prog. Theoret. Phys. (Kyoto) **22**, 715 (1959) and **23**, 1073 (1960).
[3] M. Gell-Mann, Rept. CTSL-20, California Institute of Technology; Y. Ne'eman, Nuclear Phys., **26**, 222 (1961).
[4] Y. Lehrer, Bull. Res. Council Israel **5A**, 197 (1956).

where a, b and c are scalars under $U(3)$ (and hence functions of the Casimir operators).

In a way similar to the commutative case, Eq. (12) implies

$$\sum_{\alpha\beta\cdots\tau} A^{\mu}_{\alpha}A^{\alpha}_{\beta}\cdots A^{\tau}_{\nu} = a_{(n)}\,\delta^{\mu}_{\nu} + b_{(n)}A^{\mu}_{\nu} + c_{(n)}\sum_{\alpha} A^{\mu}_{\alpha}A^{\alpha}_{\nu}, \qquad (13)$$

where n is the degree of the polynomial in A^{γ}_{δ} on the left, and $n > 3$.

Let $B_1, \cdots, B_r$ be r irreducible matrices of order k. It follows from the theorem of Burnside[5] that any matrix of the same order may be described as a polynomial in these matrices. Hence, T^{μ}_{ν} may be described as a polynomial in the A^{μ}_{ν}. However, the tensorial character of T^{μ}_{ν} implies that the above polynomial is a combination of polynomials of the form

$$\sum_{\alpha\beta\cdots\tau} A^{\mu}_{\alpha}A^{\alpha}_{\beta}\cdots A^{\tau}_{\nu}. \qquad (14)$$

Using (13), Eq. (1) is obtained.

[5] See, for example, V. D. Waerden, *Modern Algebra II*, Translation from the 3rd German edition (Frederick Ungar Publishing Company, New York), 2nd printing, 1950, p. 194.

ACKNOWLEDGMENTS

We would like to thank Y. Dotan, C. A. Levinson, and Y. Ne'eman for helpful discussions.

High-Energy Behavior in Perturbation Theory*

J. C. POLKINGHORNE
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England
(Received 26 October 1962)

The dominant-high energy behavior of a wide class of Feynman diagrams is investigated. When the leading contributions are summed they are shown to give a behavior consistent with the Regge-pole hypothesis. Series expansions for the trajectory and residue of the dominant Regge pole are obtained in this approximation.

1. INTRODUCTION

THE existence of Regge[1] poles in relativistic scattering amplitudes has been the subject of intense study and speculation. One of the most important ways in which their existence would manifest itself is through the high-energy behavior of scattering amplitudes.[2] If the dominant (i.e. the rightmost) Regge pole in the s channel of the scattering amplitude $f(s, t)$ has a trajectory $\alpha(s)$ then

$$f \sim b(s)(t/m^2)^{\alpha(s)}, \qquad t \to \infty, \tag{1}$$

where m is some standard mass, and $b(s)$ is related to the residue $\beta(s)$ of the Regge pole by

$$b(s) = \beta(s)[2\alpha(s) + 1][1 \pm e^{-i\pi\alpha(s)}]/\sin \pi\alpha(s), \tag{2}$$

the sign in (2) being chosen according to the J parity of the trajectory.

Recently Gell-Mann and Goldberger,[3] and Lévy[4] have suggested that the same behavior may be discovered in perturbation theory. An individual Feynman diagram gives an asymptotic behavior involving log t, but when contributions from an infinite number of diagrams are added together, the logarithm may appear in an exponential yielding a power law of type (1). The purpose of this paper is to show that this does indeed occur when the leading asymptotic contributions from a wide class of Feynman diagrams are added together.

The Feynman diagrams considered represent the scattering of two spinless particles by the exchange of a virtual spinless particle (which will be called a meson). The interaction is of Yukawa type and the diagrams considered are ones in which the virtual mesons interact only with the external particles. The diagrams include many in which meson lines cross so that the class is wider than that of ladder diagrams. It is more precisely characterized below (Sec. 4).

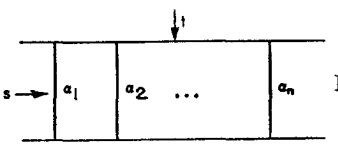


FIG. 1. A ladder diagram.

2. LADDER DIAGRAMS

The method of analysis employed will be illustrated in this section by considering the simple case of ladder diagrams (Fig. 1). The work of Lee and Sawyer[5] already guarantees that a power-law behavior should be found in this case.

For convenience all masses will be set equal to unity. The Feynman parameters associated with meson lines are denoted by $\alpha_1, \cdots, \alpha_n$ and the remaining Feynman parameters by β_i. Then the diagram Fig. 1 gives a contribution to the amplitude

$$I_n(s, t) = g^2\left(\frac{g^2}{16\pi^2}\right)^{n-1}\Gamma(n) \times \int_0^1 d\alpha\, d\beta\, \delta(\Sigma\alpha + \Sigma\beta - 1)\frac{[C(\alpha, \beta)]^{n-2}}{[D(s, t, \alpha, \beta)]^n}. \tag{3}$$

C and D are the familiar numerator and denominator functions occurring in Feynman integrals after symmetric integration has been performed. It is only necessary to notice at this stage that the structure of D is

$$D = \alpha_1 \cdots \alpha_n t + \delta(s, \alpha, \beta). \tag{4}$$

The fact that D is linear in t might seem to suggest that $I_n \sim t^{-n}$ as $t \to \infty$. This would indeed be the case if the lower limits of the α integrations were not zero. However the fact that the coefficient of t

* The research reported in this paper has been sponsored in part by the Air Force Office of Scientific Research, OAR, through the European Office, Aerospace Research, United States Air Force.

[1] T. Regge, Nuovo Cimento 14, 951 (1959); 18, 947 (1960).
[2] G. F. Chew, S. C. Frautschi, and S. Mandelstam, Phys. Rev. 126, 1202 (1962).
[3] M. Gell-Mann and M. L. Goldberger, Phys. Rev. Letters 9, 275 (1962).
[4] M. Lévy, Phys. Rev. Letters 9, 235 (1962).
[5] B. W. Lee and R. F. Sawyer, Phys. Rev. 127, 2266 (1962).

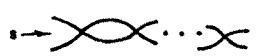


FIG. 2. A contracted ladder diagram.

vanishes on part of the boundary of the region of integration permits I_n to have a less rapidly convergent asymptotic form. This dominant asymptotic part of I_n arises solely from that part of the region of integration in the neighborhood of $\alpha_i = 0$. It is therefore the same as the dominant asymptotic part of

$$I_n' = g^2\left(\frac{g^2}{16\pi^2}\right)^{n-1}\Gamma(n)\int_0^\epsilon d\alpha\int_0^1 d\beta\ \delta(\Sigma\beta - 1) \times \frac{[c(\beta)]^{n-2}}{[\alpha_1\cdots\alpha_n t + d(s,\beta)]^n}, \quad (5)$$

where

$$c(\beta) = C(0,\beta), \quad (6)$$

$$d(s,\beta) = D(s,0,\beta). \quad (7)$$

This dominant contribution will, of course, prove to be independent of ϵ.

It is now necessary to perform some of the α integrations:

$$\int_0^\epsilon \frac{d\alpha_1\cdots d\alpha_n}{[\alpha_1\cdots\alpha_n t + d]^n} = \frac{1}{(n-1)d^{n-1}}\int_0^\epsilon \frac{d\alpha_1\cdots d\alpha_{n-1}\epsilon}{[\alpha_1\cdots\alpha_{n-1}\epsilon t + d]} + \cdots = \frac{1}{t(n-1)d^{n-1}}\int_0^\epsilon \frac{d\alpha_1\cdots d\alpha_{n-2}}{\alpha_1\cdots\alpha_{n-2}} \times \log\left(1 + \frac{\alpha_1\cdots\alpha_{n-1}\epsilon^2 t}{d}\right) + \cdots. \quad (8)$$

In these equations, contributions with a manifestly more rapidly convergent asymptotic behavior as $t \to \infty$ have been omitted. In the Appendix it is shown that the final integral in (8) $\sim(\log t)^{n-1}/\Gamma(n)$ as $t \to \infty$. Thus

$$I_n \sim g^2\left(\frac{g^2}{16\pi^2}\right)^{n-1}\frac{(\log t)^{n-1}}{n-1}\int_0^1 d\beta\frac{\delta(\Sigma\beta - 1)[c(\beta)]^{n-2}}{[d(\beta,s)]^{n-1}} \quad (9)$$

as $t \to \infty$.

The functions c and d are the numerator and denominator functions associated with the contracted Feynman diagram (Fig. 2.) Elementary manipulations show that

$$\left(\frac{1}{16\pi^2}\right)^{n-1}\int_0^1 d\beta\frac{\delta(\Sigma\beta - 1)[c(\beta)]^{n-2}}{[d(\beta,s)]^{n-1}} = \frac{1}{\Gamma(n-1)}[K(s)]^{n-1}, \quad (10)$$

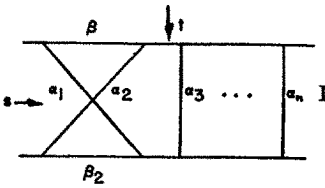


FIG. 3. A Feynman diagram.

where

$$K(s) = \frac{1}{16\pi^2}\int_0^1 \frac{d\beta_1\, d\beta_2\ \delta(\beta_1 + \beta_2 - 1)}{[\beta_1\beta_2 s - (\beta_1 + \beta_2)^2]}. \quad (11)$$

$K(s)$ is just the Feynman function associated with the simple self-energy loop diagram evaluated for two-dimensional energy–momentum vectors, for which it is, of course, convergent. Thus the final expression for the asymptotic form is

$$I_n \sim g^2[g^2K(s)\log t]^{n-1}/\Gamma(n). \quad (12)$$

Therefore,

$$\sum_{n=1}^{\infty} I_n \sim g^2(t)^{-1+g^2K(s)}, \quad (13)$$

which is of the form (1)[5,6] with

$$\alpha(s) = -1 + g^2K(s), \quad (14)$$

$$b(s) = g^2. \quad (15)$$

3. FURTHER DIAGRAMS

The techniques developed in Sec. 2 may be used to analyze other diagrams. In this section we shall consider two typical examples and then state in Sec. 4, the general rules that these illustrate.

The first diagram is Fig. 3. Its contribution is of the form (3) but D is given by

$$D = (\alpha_1\alpha_2 - \beta_1\beta_2)\alpha_3\cdots\alpha_n t + \delta(s,\alpha,\beta). \quad (16)$$

Because the end points $\alpha_1 = 0$, $\alpha_2 = 0$ no longer give a coefficient of t which automatically vanishes, the α_1 and α_2 integrations do not increase the asymptotic behavior of the function. The fact that the coefficient of t vanishes inside the region of multiple integrations is not important, as can be seen by distorting contours appropriately.[7] The

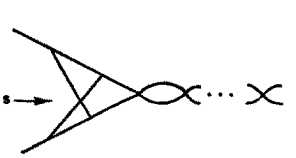


FIG. 4. The contracted diagram associated with Fig. 3.

[6] The same formula has been derived using the renormalization group; B. A. Arbusov, A. A. Lugunov, A. N. Tavkhelidze and R. N. Faustov, Physics Letters **2**, 150 (1962).

[7] *Note added in proof.* This statement requires modification. What is here calculated is the contribution to asymptotic behavior from the edge of the hypercontour. Other contributions will be discussed elsewhere.

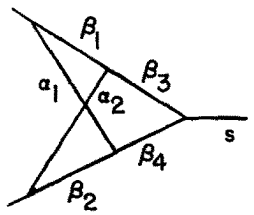


FIG. 5. The vertex part associated with Fig. 3.

asymptotic form can be obtained most easily by writing

$$D = (\alpha_1\alpha_2 - \beta_1\beta_2)\tilde{D}, \tag{17}$$

where

$$\tilde{D} = \alpha_3 \cdots \alpha_n t + \bar{\delta}(s, \alpha, \beta), \tag{18}$$

and

$$\bar{\delta} = \delta(s, \alpha, \beta)/(\alpha_1\alpha_2 - \beta_1\beta_2). \tag{19}$$

The analysis of Sec. 2 applied to $\tilde{D}$ gives

$$I_n \sim g^2\left(\frac{g^2}{16\pi^2}\right)^{n-1} \frac{\Gamma(n)}{t(n-1)\Gamma(n-2)} \times \int d\alpha_1\, d\alpha_2\, d\beta_i\; \delta(\alpha_1 + \alpha_2 + \Sigma\beta - 1) \times \frac{[c(\alpha_1, \alpha_2, \beta)]^{n-2}}{(\alpha_1\alpha_2 - \beta_1\beta_2)[d(s, \alpha_1, \alpha_2, \beta)]^{n-1}} (\log t)^{n-3}, \tag{20}$$

where c and d are the numerator and denominator functions associated with the contracted diagram—Fig. 4. These manipulations are justified because the resulting coefficient of $(\log t)^{n-3}$ is a convergent integral.

The integral in (20) can be simplified to give

$$I_n \sim g^2 V(s) \frac{(g^2 K(s) \log t)^{n-3}}{\Gamma(n-2)}, \tag{21}$$

where

$$V(s) = \left(\frac{g^2}{16\pi^2}\right)^2 \int_0^1 d\alpha_1\, d\alpha_2\, d\beta_i\; \delta(\alpha_1 + \alpha_2 + \Sigma\beta - 1) \times \frac{c'(\alpha, \beta)}{(\alpha_1\alpha_2 - \beta_1\beta_2)[d'(\alpha, \beta, s)]^2}, \tag{22}$$

c' and d' being the numerator and denominator functions associated with the vertex part—Fig. 5. The sum of dominant asymptotic contributions from

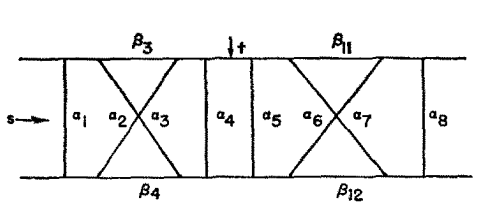


FIG. 6. A Feynman diagram.

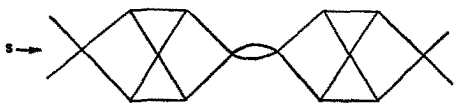


FIG. 7. The contracted diagram associated with Fig. 6.

diagrams of type shown in Fig. 3 just provides an extra term in the series expansion of $b(s)$.

The second example we shall consider is provided by the diagram of Fig. 6. Its denominator function is

$$D = (\alpha_2\alpha_3 - \beta_3\beta_4)(\alpha_6\alpha_7 - \beta_{11}\beta_{12}) \times \alpha_1\alpha_4\alpha_5\alpha_8 t + \delta(s, \alpha, \beta). \tag{23}$$

The resulting asymptotic behavior is found to be

$$I \sim g^2\left(\frac{g^2}{16\pi^2}\right)^7 \frac{\Gamma(8)}{t\cdot 7\cdot\Gamma(4)} \int_0^1 d\alpha_2\, d\alpha_3\, d\alpha_6\, d\alpha_7\, d\beta_i \times \frac{\delta(\Sigma\alpha + \Sigma\beta - 1)[c(\alpha, \beta)]^6}{(\alpha_2\alpha_3 - \beta_3\beta_4)(\alpha_6\alpha_7 - \beta_{11}\beta_{12})[d(s, \alpha, \beta)]^7} (\log t)^3, \tag{24}$$

where c and d are associated with the contracted diagram in Fig. 7. Again this may be reduced to a simpler form, yielding

$$I \sim g^8 \frac{K(s)[K'(s)]^2}{\Gamma(4)} \frac{(\log t)^3}{t}, \tag{25}$$

where

$$K'(s) = 2\cdot\frac{1}{16\pi^2}\left(\frac{g^2}{16\pi^2}\right)^2 \times \int_0^1 d\alpha\, d\beta \frac{[c''(\alpha, \beta)]^2\delta(\Sigma\alpha + \Sigma\beta - 1)}{(\alpha_1\alpha_2 - \beta_3\beta_4)[d''(s, \alpha, \beta)]^3}, \tag{26}$$

c'' and d'' being associated with the self-energy diagram—Fig. 8. Diagrams of this type are thus seen to be associated with the occurrence of higher terms in the expansion of $\alpha(s)$.

4. GENERAL RULES

The class of diagrams we consider is formed in the following way: The single meson exchange diagram is called the unit diagram and denoted by U. Any diagram which cannot be formed by joining the external particle lines of a series of subdiagrams, at least one of which is a U, is called unit irreducible and denoted by the generic symbol D_i. The class of diagrams considered is constructed by joining the external particle lines of a succession of U and D_i

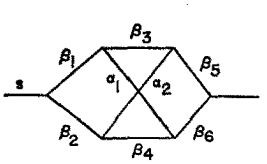


FIG. 8. The self-energy part associated with Fig. 6.

such that there is at least one U and each pair of successive D_i are separated by at least one U.

Associated with each D_i there is an s-channel self-energy diagram S_i formed by joining together both pairs of external lines, and two s-channel vertex parts, V_i and V_i', formed by joining together one or other pair of external lines. The number of meson lines in D_i is denoted by n_i and the coefficient of t in its denominator when D_i is considered as an isolated diagram is denoted by $g_i(\alpha)$, where α represents the set of all the Feynman parameters of D_i.

The series expansion for $\alpha(s)$ is

$$\alpha(s) = -1 + g^2K(s) + g^2 \sum_i K_i(s). \qquad (27)$$

The sum is taken over all unit irreducible diagrams D_i and

$$K_i(s) = \frac{1}{16\pi^2}\left(\frac{g^2}{16\pi^2}\right)^{n_i} \Gamma(n_i + 1) \times \int_0^1 \frac{d\alpha\ \delta(\Sigma\alpha - 1)[C_i(\alpha)]^{n_i}}{g_i(\alpha)[D_i(s, \alpha)]^{n_i+1}}, \qquad (28)$$

where the α's and the forms of C_i and D_i correspond to the self-energy diagram S_i.

The series expansion for $b(s)$ is

$$b(s) = g^2(1 + \sum_i V_i(s))(1 + \sum_i V_i'(s)), \qquad (29)$$

where

$$V_i(s) = \left(\frac{g^2}{16\pi^2}\right)^{n_i} \Gamma(n_i) \times \int_0^1 \frac{d\alpha\ \delta(\Sigma\alpha - 1)[c_i(\alpha)]^{n_i-1}}{g_i(\alpha)[d_i(s, \alpha)]^{n_i}}, \qquad (30)$$

the α's and the forms of c_i and d_i corresponding to the vertex diagram V_i; and $V_i'(s)$ is similarly derived from the vertex diagram V_i'. The occurrence of a product in (29) illustrates the factorizability of Regge pole residues.[8]

5. DISCUSSION

The series expansions obtained for $\alpha(s)$ and $b(s)$ possess many expected properties. The separate terms are real analytic functions with s-channel normal thresholds. As $s \to \infty$, $\alpha(s) \to -1$. However there are also some unphysical properties in this approximation. As Gell-Mann and Goldberger have pointed out, the first term in the expansion for $\alpha(s)$ is unbounded at its normal threshold. A similar result holds in potential theory.[9] Inspection of the asymptotic form in s of the separate terms of the series suggests that they may be an asymptotic series valid for large s.

Many contributions to the asymptotic behavior of the scattering amplitude have been neglected in this analysis. These include all the more rapidly convergent parts of the diagrams considered which were discarded in forming Eqs. (8). There are also contributions from diagrams not considered at all, including isolated D_i diagrams. Inspection of these latter shows that they are expected to contribute only to Regge poles which tend to negative integers less than -1 as $s \to \infty$.

ACKNOWLEDGMENTS

I am indebted to Professor M. Gell-Mann and Professor M. L. Goldberger for a discussion which provoked my interest in this problem.

APPENDIX

In this appendix we show that

$$\int_0^{\epsilon} \frac{d\alpha_1 \cdots d\alpha_m}{\alpha_1 \cdots \alpha_m} \log(1 + a\alpha_1 \cdots \alpha_m t) \sim \Gamma \frac{(\log t)^{m+1}}{(m+1)} \quad \text{as} \quad t \to \infty. \qquad (A1)$$

The region of integration may be divided into two parts: R_1, in which $a\alpha_1 \cdots \alpha_m t < N$, where $N > 1$; R_2, in which $a\alpha_1 \cdots \alpha_m t > N$. The hypervolume of $R_1 \sim t^{-1}(\log t)^{m-1}$ as $t \to \infty$, and so the contribution to the integral from $R_1 \sim (\log t)^{m-1}$ at most. In R_2 we may replace $\log(1 + a\alpha_1 \cdots \alpha_m t)$ by $\log(a\alpha_1 \cdots \alpha_m t)$ without changing the leading asymptotic behavior in t. The result then follows simply.

[8] M. Gell-Mann, Phys. Rev. Letters **8**, 263 (1962); V. N. Gribov and Ya Pomeranchuk, Phys. Rev. Letters **8**, 343 (1962).

[9] R. Blanckenbecler and M. L. Goldberger, Phys. Rev. **126**, 766 (1962).

Transition Matrix for Nucleon–Nucleon Scattering. II*

K. L. KOWALSKI† AND D. FELDMAN
Department of Physics, Brown University, Providence 12, Rhode Island
(Received 28 August 1962)

The Fredholm reduction of the singular integral equation satisfied by the reactance matrix, which was developed in a previous paper, is extended so as to constitute a complete and unified Fredholm formalism for the various integral equations which occur in the momentum-space formulation of two-body (potential) scattering problems. The essential simplification which permits this unified treatment is the demonstration of the formal similarity of the scattering integral equations, whether or not part of the interaction includes a hard core. The principal result consists in an apparatus which is useful in performing those two-body calculations which occur in investigations of high-energy nucleon–nucleus scattering; in particular, the technique is used to obtain the solutions of integral equations satisfied by transition matrices which appear typically in multiple-scattering theories. The relation between the uniqueness of solutions of the scattering integral equations and the validity of the Fredholm formalism is also discussed. Finally, some methods which were previously considered for obtaining approximate solutions of the Fredholm equations are generalized.

INTRODUCTION

IN a previous paper,[1] a procedure was developed for solving the integral equations satisfied by the partial-wave amplitudes of the two-body transition (t) and reactance (K) operators. This study was motivated by the appearance of matrix elements of t between states of unequal energy in problems concerning the scattering of nucleons by nuclei. It was pointed out in I that in such problems, it would be useful to be able to construct approximate solutions of the integral equations satisfied by the t matrices which are automatically exact on the energy shell. The attainment of such a method was the primary objective and the principal result of I.

The emphasis of the present study differs from that of I in that we are concerned primarily with the formal properties of the solutions of the two-body (potential) scattering integral equations without explicit regard to specific applications. The essential results consists in an improved and generalized apparatus which is useful in performing those two-body calculations which occur in investigations of high-energy nucleon–nucleus scattering.

No assumptions are made in this work as to the nature of the two-particle potentials other than that their partial-wave amplitudes are suitably well behaved[2] and that they satisfy the usual Hermiticity and invariance properties associated with the nucleon–nucleon interaction. We include, however, the possibility that the interaction contains a hard core (h.c.), since the use of the h.c. has proved to be such a useful phenomenological tool.

The integral equations satisfied by the t and K operators in the case when the interaction includes a h.c. were obtained in I. An alternative formulation of these equations is carried out in the first section of the present paper, which has the advantage of exhibiting the formal similarity of the different scattering integral equations, whether or not part of the interaction consists of a h.c. The general properties of these equations are discussed in detail in Sec. II.

It was noted in I that K satisfies a manifestly singular integral equation, and it was found that a convenient way of dealing with this type of equation was to reduce it to a Fredholm form. In Sec. III we elaborate upon this Fredholm reduction and, by doing so, obtain a *complete and unified* Fredholm formalism for the various scattering integral equations which occur in the momentum-space formulation of two-body scattering problems. The uniqueness and singular behavior of the solutions obtained with this method are investigated in Sec. IV.

The Fredholm technique is applied in Sec. V to find the solution of equations which appear typically in multiple-scattering theories.[3] Section VI is then devoted to a generalization of some approximation methods which were presented in I for solving the Fredholm equations. Finally, in Sec. VII we comment upon the extension of the Fredholm technique to the integral equations satisfied

* This work was supported, in part, by the U. S. Atomic Energy Commission.
† Present address: Department of Physics, Case Institute of Technology, Cleveland, Ohio.
[1] K. L. Kowalski and D. Feldman, J. Math. Phys. 2, 499 (1961), henceforth referred to as I.
[2] We specifically exclude the Coulomb potential.
[3] K. M. Watson, Phys. Rev. 89, 575 (1953); see also K. L. Kowalski and D. Feldman, Phys. Rev. 130, 276 (1963).

by the momentum-space representatives of the scattering operators, rather than by their partial-wave amplitudes.

I. SCATTERING THEORY WITH A HARD CORE

The Fredholm theory which is to be presented in this paper depends in an important way on the particular form assumed by the scattering integral equations [see Eqs. (2.1) below]. That these equations are valid when the interaction does not contain a h.c. is well known. This section is devoted to the demonstration that, even with a h.c., one can construct useful scattering operators which take into account all the effects of that part of the interaction exclusive of the h.c. and which, in addition, satisfy Eqs. (2.1).

It was shown in I that the appropriate integral equations for the ordinary transition and reactance operators for an interaction with a h.c. are given by

$$t_i^{(+)} = U_i^{(+)} + \Gamma_i^{(+)} + U_i^{(+)}G_i^{(+)}t_i^{(+)}, \tag{1.1a}$$

$$t_i^{(-)} = U_i^{(-)\dagger} + \Gamma_i^{(-)\dagger} + t_i^{(-)}G_i^{(+)}U_i^{(-)\dagger}, \tag{1.1b}$$

and

$$K_i = \bar{U}_i + \bar{\Gamma}_i + \bar{U}_i\bar{G}_iK_i. \tag{1.1c}$$

Here,

$$U_i^{(*)} = [1 + \omega_i^{(*)}]V, \tag{1.2a}$$

and

$$\bar{U}_i = (1 + \bar{\omega}_i)V, \tag{1.2b}$$

where V is the two-particle potential apart from the h.c. The two-body Green's functions are defined by

$$G_i^{(+)} = (E_i - H_0 + i\epsilon)^{-1}, \quad (\epsilon \to +0), \tag{1.3a}$$

and

$$\bar{G}_i = P(E_i - H_0)^{-1}, \tag{1.3b}$$

where H_0 is the relative two-body kinetic energy operator, E_i is an eigenvalue of H_0, and P denotes integrations carried out in the sense of the Cauchy principal value. The subscript i is employed as an energy index. Finally, the definitions and properties of the h.c. operators, $\omega_i^{(*)}$, $\bar{\omega}_i$, $\Gamma_i^{(*)}$, and $\bar{\Gamma}_i$, are discussed in the Appendix.[4]

[4] Equations (1.1) were derived in I under the assumptions that the coordinate representatives of V vanish for distances less than the h.c. radius, and that $t_i{}^{(+)}$ and $K_i[t_i{}^{(-)}]$ operate only to the left (right) of eigenstates of H_0 corresponding to the energy E_i. The last restriction will be relaxed at times in the present investigation (see, for example, Sec. V) and in these instances $\Gamma_i{}^{(+)}$ in (1.1a) must be replaced by the operator $\gamma_i{}^{(+)}$ which is defined in the Appendix (cf. also O. Brander, to be published). The appearance of $\gamma_i{}^{(+)}$ instead of $\Gamma_i{}^{(+)}$ in the general form of (1.1a) arises from the fact that in this case one has to apply the h.c. boundary conditions to the wave operator rather than to the wavefunction as in I. Similar considerations apply to Eqs. (1.1b) and (1.1c). It is easily seen that the various γ's are indistinguishable in their properties from the Γ's with respect to the derivation and general validity of all equations contained in this work which do not explicitly involve the Γ's.

The transition operator which satisfies (1.1a) can be written in the form

$$t_i^{(+)} = U_i^{(+)}\Omega_i^{(+)} + \Gamma_i^{(+)}, \tag{1.4}$$

where the wave operator, $\Omega_i^{(+)}$, is determined by the integral equation

$$\Omega_i^{(+)} = 1 + G_i^{(+)}[U_i^{(+)}\Omega_i^{(+)} + \Gamma_i^{(+)}]. \tag{1.5}$$

Let us introduce the operator

$$t_i^{(+)'} \equiv t_i^{(+)} - \Gamma_i^{(+)} = U_i^{(+)}\Omega_i^{(+)}. \tag{1.6}$$

It follows from Eq. (1.1a), with the aid of (A5a), that $t_i^{(+)'}$ satisfies

$$t_i^{(+)'} = U_i^{(+)'} + U_i^{(+)}G_i^{(+)}t_i^{(+)'}, \tag{1.7}$$

where

$$U_i^{(+)'} = [1 + \omega_i^{(+)}]V[1 + \omega_i^{(-)}]^\dagger. \tag{1.8}$$

Equation (1.7) can be put into a more symmetrical form if we use the relation

$$t_i^{(+)'} = [1 + \omega_i^{(+)}]t_i^{(+)'}, \tag{1.9}$$

which follows immediately from (1.6) and (1.2a) as a consequence of the identity

$$[1 + \omega_i^{(+)}]^2 = 1 + \omega_i^{(+)}. \tag{1.10}$$

Equation (1.10) can be verified with the help of (A4a). We see, then, from Eq. (A3a), that Eq. (1.7) can be rewritten as

$$t_i^{(+)'} = U_i^{(+)'} + U_i^{(+)'}G_i^{(+)}t_i^{(+)'}. \tag{1.11}$$

The most useful feature of the operator $t_i^{(+)'}$ is that it satisfies an equation identical in form to that for $t_i^{(+)}$ in the case without a h.c. [cf. Eqs. (1.1a) and (1.2a) with $\omega_i^{(+)} = \Gamma_i^{(+)} = 0$]. This property will permit a considerable simplification in the work of the later sections. We note that, since $\Gamma_i^{(+)}$ can be regarded as known, $t_i^{(+)'}$ is, for all practical purposes, equivalent to $t_i^{(+)}$.

In I we introduced, instead, of $t_i^{(+)}$, the operator $V\Omega_i^{(+)}$. Although it can be verified that this operator obeys an equation of the same form as (1.11), its connection to $t_i^{(+)}$ is rather indirect and this complicates any practical calculations.

The preceding reformulation of Eq. (1.1a) can be carried out in a completely analogous manner for Eqs. (1.1b) and (1.1c). For example, let

$$K_i' = K_i - \bar{\Gamma}_i. \tag{1.12}$$

One can then show that

$$K_i' = \bar{U}_i' + \bar{U}_i'\bar{G}_iK_i', \tag{1.13}$$

where

$$\bar{U}_i' = (1 + \bar{\omega}_i)V(1 + \bar{\omega}_i)^\dagger. \tag{1.14}$$

The corresponding equation for $t_i^{(-)\prime}$ is stated in Sec. V.

The relationship between $t_i^{(+)}$ and K_i (or, equivalently, $t_i^{(-)}$ and $K_i^\dagger$) for a potential with a h.c. is identical to the case without a h.c., viz.,

$$t_i^{(+)} = K_i - i\pi K_i \delta(E_i - H_0) t_i^{(+)}, \tag{1.15}$$

as was proved in I. The connection between $t_i^{(+)\prime}$ and K_i' can be obtained from (1.15).

II. SCATTERING INTEGRAL EQUATIONS

We consider a class of two-body operators which satisfy the integral equations

$$t_i^{(+)} = V_i^{(+)} + V_i^{(+)} G_i^{(+)} t_i^{(+)}, \tag{2.1a}$$

$$t_i^{(-)} = V_i^{(-)} + t_i^{(-)} G_i^{(+)} V_i^{(-)}, \tag{2.1b}$$

$$K_i = \bar{V}_i + \bar{V}_i \bar{G}_i K_i, \tag{2.1c}$$

where the quantities $V_i^{(*)}$ and $\bar{V}_i$ are related to the two-body potential operator (cf. Sec. I).

The operators satisfying Eqs. (2.1) evidently correspond to the ordinary transition and reactance operators when the interaction does not contain a h.c.[5] When one has a h.c., then, as shown in Sec. I, the scattering problem can nonetheless be reduced to a solution of equations of the form (2.1) provided one carries out a suitable subtraction of the effects of the h.c.

Let us consider Eq. (2.1a), satisfied by $t_i^{(+)}$, for the sake of definiteness. This equation is usually interpreted in one of two ways. First, one can require that $t_i^{(+)}$ operate only to the left of free-particle states, $|i\rangle$, which correspond to the reference energy E_i. In this case the only matrix elements of $t_i^{(+)}$ which appear in (2.1a) are

$$\langle f|t_i^{(+)}|i\rangle = \langle f|V_i^{(+)}|\psi_i^{(+)}\rangle, \tag{2.2a}$$

where $|\psi_i^{(+)}\rangle$ is the state of the system which satisfies the boundary condition of outgoing waves at infinity, and $|f\rangle$ denotes any free-particle state. On the other hand, one can allow $t_i^{(+)}$ to connect states of arbitrary energy. For this last alternative, a representation of the matrix elements of $t_i^{(+)}$ in the form (2.2a) is not possible. These two cases are quite distinct and give rise to integral equations which are different in nature.

When $t_i^{(+)}$ is employed in the first sense given above, it will be referred to as a *physical* operator (because of its simple relation to the wavefunction of the system). In the alternative case, $t_i^{(+)}$ will be called an *unphysical* operator. The same terminology will also be applied to $t_i^{(-)}$ and K_i, except that, for these operators, the appropriate counterparts of (2.2a) are

$$\langle i|t_i^{(-)}|f\rangle = \langle \psi_i^{(-)}|V_i^{(-)}|f\rangle, \tag{2.2b}$$

and

$$\langle f|K_i|i\rangle = \langle f|\bar{V}_i|\psi_i\rangle, \tag{2.2c}$$

where $|\psi_i^{(-)}\rangle$ and $|\psi_i\rangle$ are the incoming- and standing-wave states, respectively.

For the most part, we will confine ourselves to a study of the equations involving physical operators only. It will turn out that the unphysical case can be treated by a straightforward application of the formalism developed for the physical case.

The integral equations for the partial-wave amplitudes corresponding to Eqs. (2.1a) and (2.1c) are[6,7]

$$t_i^{(+)}(k_f|k_i) = V_i^{(+)}(k_f|k_i) + \lambda \int \frac{dk_p k_p^2}{k_p^2 - k_i^2 - i\epsilon} \times V_i^{(+)}(k_f|k_p) t_i^{(+)}(k_p|k_i), \tag{2.3a}$$

and

$$K_i(k_f|k_i) = \bar{V}_i(k_f|k_i) + \lambda P \int \frac{dk_p k_p^2}{k_p^2 - k_i^2} \times \bar{V}_i(k_f|k_p) K_i(k_p|k_i), \tag{2.3b}$$

respectively.[8] Here,

$$\lambda = -(4M/\pi\hbar^2), \qquad k^2 = (2M/\hbar^2)E,$$

and M is the reduced two-particle mass. In Eqs. (2.3), $t_i^{(+)}$, K_i, $V_i^{(+)}$, and $\bar{V}_i$ are finite-dimensional square matrices whose elements (the partial-wave amplitudes) are functions of k_i, k_f, and k_p. [See also the discussion following Eq. (2.5).]

Equations (2.3) are obviously linear, inhomogeneous integral equations of the second kind; however, they are not Fredholm equations in the classical sense.[9] Nevertheless, it can be shown[10-12] that the Fredholm theory can be applied to equations of the form (2.3a) under rather broad conditions on $V_i^{(+)}(k_f|k_p)$, so that these are, in fact, Fredholm equations.

[5] B. Lippmann and J. Schwinger, Phys. Rev. **79**, 469 (1950).

[6] A partial-wave analysis for the nucleon–nucleon case was carried out in I.

[7] The limits on all integrals in this paper may be assumed to run from 0 to ∞.

[8] In most of our subsequent work we will restrict ourselves to the operators $t_i^{(+)}$ and K_i. Completely analogous results hold for $t_i^{(-)}$ and $K_i^\dagger$.

[9] Cf. R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience Publishers, Inc., New York, 1953), Vol. I, Chap. 3, p. 112.

[10] F. Smithies, *Integral Equations* (Cambridge University Press, New York, 1958).

[11] R. Jost and A. Pais, Phys. Rev. **82**, 840 (1951).

[12] A. Salam and P. T. Matthews, Phys. Rev. **90**, 690 (1953).

On the other hand, the type of equation satisfied by $K_i(k_f|k_i)$ is a genuine singular integral equation for which the Fredholm theory is not directly applicable. Unfortunately, there exists no extensive literature which is pertinent to (2.3b).[13] However, it is important to distinguish the singular integral equations to be studied in this paper from those which frequently appear in field-theoretic calculations.[14,15]

It is sometimes convenient to regard both Eq. (2.3a) and Eq. (2.3b) as singular integral equations. Evidently, the first can be transformed into an equation of the second type if one makes use of the identity

$$(z - y \pm i\epsilon)^{-1} = P(z - y)^{-1} \mp i\pi\delta(z - y), \qquad (2.4)$$

and simply regards $t_i^{(+)}(k_i|k_i)$ as part of the inhomogeneous term. Alternatively, one may use (2.4) to express (2.3b) in the form (2.3a). In fact, this is perhaps the simplest reduction of Eq. (2.3b) to a Fredholm form. However, this reduction leads to nothing new since it merely results in the expression of $K_i(k_f|k_i)$ in terms of the Jost–Pais–Khuri[11,16] solution for $t_i^{(+)}(k_f|k_i)$ [cf. Eq. (1.15)], but, unfortunately, the latter solution has not proved to be useful in practical calculations.

The integral equations (2.3) are rather cumbersome to write out repeatedly. So, for the sake of simplicity as well as generality, we study instead a somewhat abstract equation which includes (2.3) as particular examples. Consider the integral equation[17]

$$R(x|y) = f(x|y) + P\int dz \frac{T(x|z)}{z - y} R(z|y), \qquad (2.5)$$

where $R(x|y)$, $f(x|y)$, and $T(x|z)$ are square matrices of finite dimensionality whose elements are complex-valued functions of the real variables x, y, and z.[18] It is assumed that, for all y $(0 \leq y < \infty)$, $f(x|y)$, $T(x|z)$, and $[\partial T(x|z)/\partial z]_{z=y}$ are continuous throughout the two-dimensional domain $0 \leq (x, z) < \infty$, and vanish for $(x, z) \to \infty$. Finally, we suppose that, except for certain discrete y, a continuous solution, $R(x|y)$, of (2.5) exists which vanishes for $x \to \infty$. Most of the results to be derived hold for much weaker conditions on f and T and with R a member of a broader class of functions; the case considered, however, covers many physical situations.

III. FREDHOLM REDUCTION

It was noted in I that a customary technique for studying singular integral equations is to attempt to reduce them to a Fredholm form. For a given equation there is usually a variety of ways in which this reduction can be carried out, where the differences among reduction techniques are ordinarily only a matter of taste. Nevertheless, situations can conceivably arise when the different Fredholm forms of a given singular integral equation are inequivalent, for example, the solution of one of the possible Fredholm equations may not be unique. Also, a Fredholm form and its associated singular integral equation may not be equivalent, that is, the former may have solutions which are not solutions of the latter. Questions concerning the equivalence of different Fredholm forms derived from a singular integral equation, as well as the equivalence of a definite Fredholm form and the singular integral equation itself, are not trivial, but an adequate discussion of these would carry us too far afield. So, except for a somewhat brief discussion in Sec. IV, we shall ignore these problems and proceed in a formal manner.

In the preceding section we indicated a simple method for reducing an equation of the type (2.5) to a Fredholm form. In addition, Goto and Machida[19] have recently shown how to transform (2.5) into a Fredholm equation of the third kind. However, neither of these Fredholm reductions seems to have particularly interesting properties, although the second may be of some value in numerical calculations. In the remainder of this paper we will confine ourselves to the type of reduction which was considered in I; an improved version of this follows.[20]

Let us assume, for the moment, that the determinant of $T(y|y)$ is nonvanishing, viz.,

$$\det T(y|y) \neq 0, \qquad (3.1)$$

so that $T^{-1}(y|y)$ exists. If we multiply Eq. (2.5),

[13] Equations of the form (2.3b) have also appeared in connection with calculations of pion–nucleon scattering in the Tamm–Dancoff approximation; see, e.g., F. J. Dyson, M. Ross, E. E. Salpeter, S. S. Schweber, M. K. Sundaresan, V. M. Visscher, and H. A. Bethe, Phys. Rev. 95, 1644 (1954).

[14] Cf. R. Omnès, Nuovo Cimento 8, 316 (1958).

[15] A comprehensive treatise on the latter type is that of N. I. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff Ltd., Groningen, The Netherlands, 1953).

[16] N. N. Khuri, Phys. Rev. 107, 1148 (1957).

[17] Equations of the form (2.5), and therefore (2.3), apply to a large class of two-body scattering problems, both relativistic and nonrelativistic, for which the interaction Hamiltonian is rotationally invariant [cf. I. F. Gol'fand, Soviet Phys.—JETP 4, 103 (1957)].

[18] In general, $T(x|z)$ also depends upon y. However, we will suppress this dependence in order to simplify the notation.

[19] J. Goto and S. Machida, Prog. Theoret. Phys. Kyoto 25, 64 (1961).

[20] The material in the following two paragraphs follows closely the treatment given in I, but is included for the sake of continuity of the subsequent development.

with x set equal to y, on the left by $\tau(x|y)$, where

$$\tau(x|z) = T(x|z)T^{-1}(y|z), \tag{3.2}$$

and then subtract the resultant expression from (2.5) with x arbitrary, we obtain

$$R(x|y) = f'(x|y) + \tau(x|y)R(y|y) + \int dz\Lambda(x|z)R(z|y), \tag{3.3}$$

where

$$f'(x|y) = [f(x|y) - \tau(x|y)f(y|y)], \tag{3.4}$$

and

$$\Lambda(x|z) = [T(x|z) - \tau(x|y)T(y|z)](z - y)^{-1}. \tag{3.5}$$

The y dependence of $\Lambda(x|z)$ has been suppressed.

The kernel $\Lambda(x|z)$ is nonsingular; thus, Eq. (3.3) is a Fredholm equation of the second kind. The solution of (3.3) can then be written as

$$R(x|y) = f'(x|y) + \tau(x|y)R(y|y) + \int dz\mathfrak{R}_y(x|z)[f'(z|y) + \tau(z|y)R(y|y)], \tag{3.6}$$

where $\mathfrak{R}_y(x|z)$ is the resolvent corresponding to the kernel $\Lambda(x|z)$. The resolvent satisfies

$$\mathfrak{R}_y(x|z) = \Lambda(x|z) + \int ds\Lambda(x|s)\mathfrak{R}_y(s|z) = \Lambda(x|z) + \int ds\mathfrak{R}_y(x|s)\Lambda(s|z). \tag{3.7}$$

It is clear from Eqs. (3.5) and (3.7) that

$$\mathfrak{R}_y(y|z) = 0 \tag{3.8}$$

for all z.

Now, if the solution (3.6) is substituted into the integral of Eq. (2.5), and x is set equal to y, the resultant equation can be solved for $R(y|y)$ to yield

$$R(y|y) = f(y|y) + T(y|y) \times \left[T(y|y) - P\int dzT'(y|z)\zeta(z|y)\right]^{-1} \times P\int dzT'(y|z)\eta(z|y), \tag{3.9}$$

where

$$\eta(x|y) = f(x|y) + \int dz\mathfrak{R}_y(x|z)f(z|y), \tag{3.10a}$$

$$\zeta(x|y) = R(x|y) + \int dz\mathfrak{R}_y(x|z)T(z|y), \tag{3.10b}$$

and

$$T'(x|z) = T(x|z)(z - y)^{-1}. \tag{3.11}$$

When the expression (3.9) for $R(y|y)$ is employed in (3.6), it is found that

$$R(x|y) = \eta(x|y) + \zeta(x|y) \times \left[T(y|y) - P\int dzT'(y|z)\zeta(z|y)\right]^{-1} \times P\int dzT'(y|z)\eta(z|y). \tag{3.12}$$

We have observed earlier that equations of the type (2.3a) can be put into the form (2.3b). This was done in order to include (2.3a) in the representation (2.5). Actually, we see that this is unnecessary, since all of the preceding discussion remains valid if $P(z - y)^{-1}$ is replaced by $(z - y \pm i\epsilon)^{-1}$. Indeed, the kernel, $\Lambda(x|z)$, and therefore the resolvent, $\mathfrak{R}_y(x|z)$, remain unaltered by this change, since

$$[(z - y)\Lambda(x|z)]_{z=y} = 0$$

for all x.

A particularly simple, special case of Eq. (2.5) arises when

$$f'(x|y) = 0, \tag{3.13}$$

whereupon it follows from Eqs. (3.6) and (3.9) that

$$R(x|y) = \zeta(x|y)T^{-1}(y|y)R(y|y), \tag{3.14a}$$

$$R(x|y) = \zeta(x|y)\left[T(y|y) - P\int dzT'(y|z)\zeta(z|y)\right]^{-1}f(y|y). \tag{3.14b}$$

The integral equations of interest [Eqs. (2.1) and (2.3)] satisfy (3.13) as well as

$$f(x|y) = c(y)T(x|y), \tag{3.15}$$

where the factor of proportionality, $c(y)$, has no matrix structure. Henceforth, we will restrict ourselves to the case when (3.15) [and therefore also (3.13)] holds.

The fact that we can limit ourselves to those situations for which Eqs. (3.13) and (3.15) are valid is a direct consequence of the unified treatment of the scattering integral equations, both with and without a h.c., as presented in Sec. I. The result is a considerable simplification of the corresponding discussion given in I.

The preceding formalism was developed under the assumption that $T^{-1}(y|y)$ exists. However, if we refer to Eq. (3.14b), we see that this relation appears to be well defined even if $T^{-1}(y|y)$ does not exist,

and it is not unreasonable to expect that the formalism is valid independently of whether (3.1) is satisfied. We will investigate this point further in the next section.

IV. UNIQUENESS AND SINGULAR BEHAVIOR

Now, given Eq. (3.15), Eq. (2.5) becomes

$$R(x|y) = c(y)T(x|y) + P\int dzT'(x|z)R(z|y), \tag{4.1}$$

which can be reduced to the form (3.14a) where

$$\zeta(x|y) = T(x|y) + \int dz\Lambda(x|z)\zeta(z|y). \tag{3.10c}$$

By construction, any solution of (4.1) satisfies Eqs. (3.14a) and (3.10c); however, the converse is not necessarily true.[21] This situation is characteristic of the general reduction of a singular integral equation to a Fredholm form. In this context, it is interesting to point out that it is impossible to show that $R(x|y)$ satisfies (4.1) if one is given only (3.14a) and (3.10c).

In Sec. III we have implicitly assumed that the solutions of (4.1) and (3.10c) exist, and, moreover, that these solutions are unique. Under these conditions *the* solution of (4.1) is simply

$$R(x|y) = c(y)\zeta(x|y)\Big[T(y|y) - P\int dzT'(y|z)\zeta(z|y)\Big]^{-1}T(y|y). \tag{4.2}$$

If our basic equations of scattering theory are at all physically meaningful, we may expect that the assumptions of existence and uniqueness are usually justified, and therefore, except for definitely singular types of physical phenomena, the solution (4.2) is applicable. In this section we will be concerned with the behavior of (4.2) as one approaches a "singular point," e.g., a point, y, where the solutions of (4.1) or (3.10c) are not unique, or (possibly) where $T^{-1}(y|y)$ does not exist. In other words, we wish to investigate some of the circumstances under which the representation (4.2) may fail.

A necessary and sufficient condition for a solution of Eq. (4.1) to be unique is that the homogeneous equation

$$R_0(x|y) = P\int dzT'(x|z)R_0(z|y) \tag{4.3}$$

have no nontrivial solution. It is clear that (4.3) can be reduced to the form (3.14a),

$$R_0(x|y) = \zeta(x|y)T^{-1}(y|y)R_0(y|y), \tag{4.4}$$

where $\zeta(x|y)$ satisfies Eq. (3.10c). Let us assume, first, that a solution of Eq. (3.10c) exists and that it is unique. Then, a solution, $R_0(x|y)$, of (4.3), not identically zero, exists if and only if

$$T^{-1}(y|y)R_0(y|y) \neq 0. \tag{4.5}$$

We find from Eqs. (4.3) and (4.4) that

$$B(y)R_0(y|y) = 0, \tag{4.6}$$

where

$$B(y) = A(y)T^{-1}(y|y), \tag{4.7a}$$

and

$$A(y) = T(y|y) - P\int dzT'(y|z)\zeta(z|y). \tag{4.7b}$$

It is evident from (4.6) that a nonzero $R_0(y|y)$ exists if and only if $B^{-1}(y)$ does not exist.

Now, suppose a nontrivial solution of (4.3) exists; then from (4.3) and (4.4) we have[22]

$$\zeta(x|y) = P\int dzT'(x|z)\zeta(z|y), \tag{4.8}$$

whence $A(y)$ vanishes, so that, from Eqs. (4.7), $B^{-1}(y)$ will not exist, independently of the nature of $T^{-1}(y|y)$. On the other hand, the nonexistence of $B^{-1}(y)$ implies that $A^{-1}(y)$ does not exist, since $T(y|y)$ has already been assumed to be well defined; hence, by Eqs. (4.6) and (4.7a), there is a nonzero solution for $T^{-1}(y|y)R_0(y|y)$, and so a solution of (4.3) exists. Thus, a necessary and sufficient condition for the existence of a solution of (4.3) is that $A^{-1}(y)$ or $B^{-1}(y)$ not exist.

Let us study next the compatibility of the existence of solutions of (4.3) with the validity of the representation (4.2). We deduce from Eqs. (4.1) and (3.14a) that

$$B(y)R(y|y) = c(y)T(y|y). \tag{4.9}$$

If a solution of (4.3) exists, $A(y)$ vanishes; then, (4.9) [and hence (4.2)] cannot be satisfied for finite $R(y|y)$, unless $T(y|y)$ is equal to zero. But when $T(y|y)$ vanishes, we see from (4.2) that, in general, $R(x|y)$ will not be defined unless $\zeta(x|y)$ also vanishes. Thus, except for the trivial case when $\zeta(x|y) = 0$, the representation (4.2) fails when (4.3) admits of a nontrivial solution.

The form (3.14a) for $R(x|y)$ is meaningful only if the product,

$$T^{-1}(y|y)R(y|y) = c(y)A^{-1}(y)T(y|y), \tag{4.10}$$

[21] For example, the solution of (3.10c) may not be unique.

[22] Note that Eq. (4.8) is compatible with Eq. (3.10c).

is well defined. However, it is evident from (4.10) that, if the solution of Eq. (4.1) is unique [so that $A^{-1}(y)$ exists], the product $T^{-1}(y|y)R(y|y)$, and therefore $R(x|y)$, is defined independently of the existence of $T^{-1}(y|y)$. Also, the nonexistence of $T^{-1}(y|y)$ implies that $R^{-1}(y|y)$ does not exist [unless, of course, $R(y|y)$ does not exist].

We now consider some consequences of the nonexistence of $\zeta(x|y)$. If a solution of Eq. (3.10c) is unique, it can be expressed in the form (3.10b) in terms of the Fredholm resolvent $\mathcal{R}_y(x|z)$, which can, for any particular point y, be written as[12,23]

$$\mathcal{R}_y(x|z) = [1/D(y)]N_y(x|z); \tag{4.11}$$

$D(y)$ does not have a matrix structure and is such that the solution of Eq. (3.10c) is not unique if and only if

$$D(y) = 0. \tag{4.12}$$

As one approaches a point, y, where (4.12) holds, Eq. (4.2) becomes

$$R(x|y) = -c(y)\left[\int dz N_y(x|z)T(z|y)\right] \times \left[\int ds \int dz T'(y|z)N_y(z|s)T(s|y)\right]^{-1} T(y|y), \tag{4.13}$$

which is generally well defined; in particular,

$$R(y|y) = 0, \quad \text{for} \quad D(y) = 0, \tag{4.14}$$

where we have used Eq. (3.8). We see then that any singularities of the solution of (3.10c) as given by (3.10b) do not necessarily give rise to a singular behavior in the solution $R(x|y)$. In fact, the only unusual property of $R(x|y)$ in this case is (4.14).

On the other hand, let us assume that $R(y|y)$ vanishes. Then from (4.9) we have

$$B^{-1}(y)T(y|y) = 0. \tag{4.15}$$

When $T^{-1}(y|y)$ exists, Eq. (4.15) can be satisfied only if $B(y)$ does not exist. Since $T(y|y)$ is assumed to be well defined, the nonexistence of $B(y)$ implies [cf. Eqs. (4.7)] that the integral

$$P\int dz T'(y|z)\zeta(z|y)$$

does not exist. It seems unlikely, with the type of functions $T(y|z)$ which arise in typical physical problems, that the preceding integral will diverge unless $\zeta(z|y)$ does not exist. If this is the case, we conclude that the vanishing of $R(y|y)$ together with the existence of $T^{-1}(y|y)$ implies that the solution of (3.10c) is not unique.

[23] See reference 10, p. 88, and also J. Hamilton, Phys. Rev. **114**, 1170 (1959).

It is interesting to inquire whether or not $\zeta(x|y)$ is defined when $T^{-1}(y|y)$ does not exist. At first, one might think that $\zeta(x|y)$ will be singular in this instance since, then, $\Lambda(x|z)$ does not exist [cf. Eq. (3.5)]. Actually, examples can be found in the theory of Fredholm equations in which the presence of an infinite parameter in the kernel does not necessarily preclude the existence of noninfinite solutions.[24] In the final analysis, to reach any definite conclusion concerning the effect of the nonexistence of $T^{-1}(y|y)$ on $\zeta(x|y)$, it is necessary to consider in detail the explicit Fredholm solution for $\zeta(x|y)$.

It is clear that the entire discussion of this section to this point remains valid if $P(z-y)^{-1}$ is replaced by $(z-y\pm i\epsilon)^{-1}$. Hence, all of our conclusions will apply equally well to the equations satisfied by $t_i^{(+)}(k_f|k_i)$ and $K_i(k_f|k_i)$. With this in mind, it is interesting to transcribe some of our results into a somewhat more physical language. It should be remembered that $T(x|z)$ corresponds essentially to $V_i^{(+)}(k_f|k_p)$ [or $\bar{V}_i(k_f|k_p)$], that is, the potential matrix [cf. Eqs. (2.3) and (2.5)].

When a partial-wave amplitude of the K_i matrix is infinite for real positive k_i, it is well known that this corresponds to a resonance in the angular-momentum state in question. Therefore, the lack of uniqueness of solutions of the singular integral equation for $K_i(k_f|k_i)$ corresponds to a resonance. This connection is by no means obvious, nor is it analogous to what one finds for the $t_i^{(+)}(k_f|k_i)$ equation,[16] since the integral equation for $K_i(k_f|k_i)$ is in no sense a Fredholm equation.

The situation is different for the $t_i^{(+)}$ matrix, since for a wide variety of two-particle potentials (without a h.c.), the poles of $t_i^{(+)}(k_f|k_i)$ appear only when k_i is complex; however, there is little in our formalism which needs to be modified if k_i (i.e., y) is allowed to become complex. It has been shown by Khuri[16] that those singularities of the $t_i^{(+)}$ matrix which occur for k_i on the positive imaginary axis, and hence correspond to bound states, coincide with the nonuniqueness of solutions of the equation for $t_i^{(+)}(k_f|k_i)$. This last, of course, is compatible with our results.

Next, if the inverse of the potential matrix, $\bar{V}_i(k_i|k_i)$ [or $V_i^{(+)}(k_i|k_i)$], does not exist, we know that either $K_i(k_i|k_i)$ [or $t_i^{(+)}(k_i|k_i)$] does not exist, or else $K_i^{-1}(k_i|k_i)$ {or $[t_i^{(+)}(k_i|k_i)]^{-1}$} does not exist. In the former case, we interpret the singularity as being indicative of a phenomenon such as a

[24] For instance, one may consider the case of a separable kernel which contains a parameter as a multiplicative factor. It is easily shown that the resolvent corresponding to such a kernel is a rational function of the parameter, and so need not diverge for infinite values of the latter quantity.

resonance or a bound state. In the latter case the physical implication is not entirely clear. However, if $K_i(k_i|k_i)$, say, has no matrix structure (as in the case in singlet-state scattering), the latter alternative implies that $K_i(k_i|k_i)$ vanishes. Some implications of the nonexistence of $K_i^{-1}(k_i|k_i)$ and $[t_i^{(+)}(k_i|k_i)]^{-1}$ when these quantities have a matrix structure are discussed in the next few paragraphs.

Let us assume that $t_i^{(+)}(k_f|k_i)$ is a partial-wave amplitude of the transition operator $t_i^{(+)}$, where the latter is related to the scattering operator, S_i, by[5]

$$S_i = 1 - 2\pi i\delta(E_i - H_0)t_i^{(+)}. \tag{4.16}$$

In partial-wave notation, Eq. (4.16) becomes

$$S_i(k_i) = 1 + i\pi\lambda k_i t_i^{(+)}(k_i|k_i). \tag{4.17}$$

For the sake of simplicity, we consider only the case when $t_i^{(+)}(k_i|k_i)$ is a 2×2 matrix; this corresponds, of course, to triplet-state nucleon–nucleon scattering. We also assume that the scattering is time-reversal invariant so that $S_i(k_i)$ is a symmetric matrix. Then, since S_i is unitary, we can express $S_i(k_i)$ in the standard manner[25] in terms of the real quantities δ_1, δ_2, and ϵ, where the δ's are the eigenphase shifts and ϵ is the mixing parameter.

The inverse of $t_i^{(+)}(k_i|k_i)$ does not exist if and only if

$$\det t_i^{(+)}(k_i|k_i) = 0. \tag{4.18}$$

In view of Eq. (1.15), Eq. (4.18) implies that the inverse of $K_i(k_i|k_i)$ does not exist and vice versa; a similar relationship exists between $t_i^{(+)'}(k_i|k_i)$ and $K_i'(k_i|k_i)$.[26] Thus, in an investigation of the implications of the nonexistence of the inverse matrices, it is sufficient to consider only $t_i^{(+)}(k_i|k_i)$ and $t_i^{(+)'}(k_i|k_i)$; the former (latter) is of interest with respect to the Fredholm formalism when the interaction does not include (includes) a h.c.

Using the above parametric representation for $S_i(k_i)$, we find, with the aid of (4.17), that Eq. (4.18) leads to

$$(e^{2i\delta_1} - 1)(e^{2i\delta_2} - 1) = 0. \tag{4.19}$$

Equation (4.19) is satisfied only when δ_1 and/or δ_2 is equal to an integral multiple of π.

If $t_i^{(+)'}(k_i|k_i)$ does not possess an inverse, we must have

$$\det [S_i(k_i) - S_i^c(k_i)] = 0, \tag{4.20}$$

where S_i^c is the S_i operator for pure h.c. scattering. We note that $S_i^c(k_i)$ can be expressed in terms of the two real (h.c.) phase shifts, δ_1^c and δ_2^c. Equation (4.20) then yields

$$e^{2i\delta_1'}(e^{2i\delta_2'} - \Delta) = (e^{2i\delta_2'}\Delta^* - 1), \tag{4.21}$$

where

$$\delta_{1,2}' = \delta_{1,2} - \delta_{1,2}^c, \tag{4.22}$$

and

$$\Delta = \cos^2\epsilon + (\sin^2\epsilon)e^{2i(\delta_1^c - \delta_2^c)}. \tag{4.23}$$

In contrast to (4.19), Eq. (4.21) is equivalent to two equations in three unknowns (δ_1', δ_2', and ϵ). Therefore, we expect an entire family of nontrivial solutions whose physical significance, however, is not clear.

We have established, at least in the one- and two-channel cases, what appear to be meaningful relations between the partial-wave amplitudes of the potential and transition matrices. These relations may be of some use as consistency requirements in attempts to reproduce the two-body scattering parameters by means of a (semi) phenomenological potential.

V. UNPHYSICAL OPERATORS

Let us now examine Eqs. (2.1) when no limitation is placed upon the vector space in which $t_i^{(\pm)}$ and K_i operate; these unphysical operators are of particular importance in the theory of multiple scattering.[3] It will suffice to consider Eq. (2.1a) with $V_i^{(+)}$ replaced by $U_i^{(+)'}$, and $t_i^{(+)}$ by $t_i^{(+)'}$, viz.,

$$t_i^{(+)'} = U_i^{(+)'} + U_i^{(+)'}G_i^{(+)}t_i^{(+)'}. \tag{1.11}$$

Equation (1.11) can be solved[27] for $U_i^{(+)'}$ to yield

$$U_i^{(+)'} = t_i^{(+)'}[1 + G_i^{(+)}t_i^{(+)'}]^{-1} = [1 + t_i^{(+)'}G_i^{(+)}]^{-1}t_i^{(+)'}. \tag{5.1}$$

The last equality in Eq. (5.1) implies that $t_i^{(+)'}$ also satisfies

$$t_i^{(+)'} = U_i^{(+)'} + t_i^{(+)'}G_i^{(+)}U_i^{(+)'}. \tag{5.2}$$

Similarly, we note that

$$K_i' = \bar{U}_i' + \bar{U}_i'\bar{G}_iK_i' = \bar{U}_i' + K_i'\bar{G}_i\bar{U}_i'. \tag{5.3}$$

On the other hand, if we define [cf. Eq. (1.6)]

$$t_i^{(-)'} \equiv t_i^{(-)} - \Gamma_i^{(-)\dagger}, \tag{5.4}$$

it can be deduced, in the manner of Sec. I, that

$$t_i^{(-)'} = U_i^{(+)'} + t_i^{(-)'}G_i^{(+)}U_i^{(+)'}. \tag{5.5}$$

[25] J. M. Blatt and L. C. Biedenharn, Rev. Mod. Phys. 24, 258 (1952).

[26] See Sec. I for the definitions of $t_i^{(+)'}$ and K_i'.

[27] This is not possible when $t_i^{(+)'}$ is interpreted as a physical operator.

Comparing Eqs. (5.2) and (5.5) we see that when the transition operators are employed in the unphysical sense there is no distinction between the so-called outgoing and incoming operators.[28]

The characteristic feature of the unphysical scattering operators is that they satisfy *two* integral equations, for example, Eqs. (1.11) and (5.2). Let us study these two equations after an appropriate partial-wave analysis has been carried out. In accordance with the methods of Secs. II–IV, we then need to consider

$$R(x|w) = c(w)T(x|w) + P\int dz \frac{T(x|z)}{z-y} R(z|w), \quad (5.6a)$$

and

$$R(x|w) = c(w)T(x|w) + P\int dz R(x|z) \frac{T(z|w)}{z-y}. \quad (5.6b)$$

All of the notation employed in Eqs. (5.6) has been explained before. However, in the present case it is necessary to emphasize that, besides $T(x|w)$, the quantity $R(x|w)$ depends implicitly upon the parameter y.[29] Finally, all the work to follow will apply with equal validity if $P(z-y)^{-1}$ is replaced by $(z-y \pm i\epsilon)^{-1}$, for reasons which have been elaborated upon previously.

In the following development, we find a representation of $R(x|w)$ which corresponds to a generalization of (4.2). If we use the techniques of Sec. III, the integral equation (5.6a) can be reduced to the Fredholm form

$$R(x|w) = \tau(x|y)R(y|w) + \int dz \Lambda(x|z)R(z|w) + c(w)[T(x|w) - \tau(x|y)T(y|w)]. \quad (5.7)$$

Let us define $\zeta(x|w)$ as the solution of the integral equation

$$\zeta(x|w) = T(x|w) + \int dz \Lambda(x|z)\zeta(z|w), \quad (5.8a)$$

so that

$$\zeta(x|w) = T(x|w) + \int dz \mathcal{R}_y(x|z)T(z|w). \quad (5.8b)$$

Equation (5.7) can then be rewritten as

$$R(x|w) = \zeta(x|y)T^{-1}(y|y)R(y|w) + c(w)[\zeta(x|w) - \zeta(x|y)\tau^T(y|w)], \quad (5.9)$$

where

$$\tau^T(x|w) = T^{-1}(x|y)T(x|w). \quad (5.10)$$

It remains to determine $R(y|w)$.

Now, from (5.6b) we have, for $x = y$,

$$R(y|w) = c(w)T(y|w) + P\int dz R(y|z) \frac{T(z|w)}{z-y}. \quad (5.11)$$

This is precisely the sort of equation one obtains when the integral equations for the physical partial-wave amplitudes of $t_i^{(-)}$ and $K_i^\dagger$ are considered. Equation (5.11) can be reduced to

$$R(y|w) = R(y|y)\tau^T(y|w) + \int dz R(y|z)\Lambda^T(z|w), \quad (5.12)$$

where

$$\Lambda^T(z|w) = [T(z|w) - T(z|y)\tau^T(y|w)](z-y)^{-1}. \quad (5.13)$$

We define a quantity $\zeta^T(x|w)$ as the solution of the integral equation

$$\zeta^T(x|w) = T(x|w) + \int dz \zeta^T(x|z)\Lambda^T(z|w). \quad (5.14a)$$

If $\mathcal{R}_y^T(z|w)$ denotes the resolvent corresponding to the kernel $\Lambda^T(z|w)$, we can then write

$$\zeta^T(x|w) = T(x|w) + \int dz T(x|z)\mathcal{R}_y^T(z|w). \quad (5.14b)$$

The functions $\zeta(x|w)$ and $\zeta^T(x|w)$ are not independent. On comparing Eqs. (3.5) and (5.13), we see that

$$(w-y)\Lambda(x|w) = (x-y)\Lambda^T(x|w). \quad (5.15)$$

Hence,

$$(w-y)\mathcal{R}_y(x|w) = (x-y)\mathcal{R}_y^T(x|w). \quad (5.16)$$

Thus, if $\mathcal{R}_y(x|w)$ [and therefore $\zeta(x|w)$] are known, $\zeta^T(x|w)$ is also known by virtue of Eqs. (5.14b) and (5.16).

Now it is clear that

$$R(y|w) = R(y|y)T^{-1}(y|y)\zeta^T(y|w). \quad (5.17)$$

Inserting the expression (5.17) for $R(y|w)$ into Eq. (5.9) and making use of Eqs. (3.7), we finally obtain

$$R(x|w) = \zeta(x|y)T^{-1}(y|y)R(y|y)T^{-1}(y|y)\zeta^T(y|w) + c(w)(w-y)\mathcal{R}_y(x|w). \quad (5.18)$$

The quantity $R(y|y)$ can be expressed, equivalently, in terms of either $\zeta(x|y)$ [cf. Eq. (4.2)] or $\zeta^T(y|x)$. Equation (5.18) represents the solution of Eqs. (5.6) for $R(x|w)$ in the same sense as does Eq. (4.2) for $R(x|y)$.

[28] In this context it is interesting to note that $\Gamma_i^{(-)\dagger} = \Gamma_i^{(+)}$, and therefore $\gamma_i^{(-)\dagger} = \gamma_i^{(+)}$.

[29] The variable y is related to the reference energy E_i in Eqs. (1.11) and (5.2).

We observe, by comparing Eq. (4.3) (and the ensuing discussion) with Eq. (5.6a) for $c(w) = 0$, that the conditions for a solution of Eq. (5.6a) to be unique are the same as for the $R(x|y)$ defined by Eq. (4.1). Also, we see, on applying (4.10), that

$$T^{-1}(y|y)R(y|y)T^{-1}(y|y) = c(y)A^{-1}(y). \tag{5.19}$$

Thus, the product on the left-hand side of (5.19) is defined unless $R(y|y)$ itself is not uniquely defined.

One significant difference between $R(x|w)$ and $R(x|y)$ is evident when $\zeta(x|y)$, and therefore $\mathfrak{R}_\nu(x|z)$, is not defined. While we noted in Sec. IV that $R(x|y)$ is generally well defined even if $\zeta(x|y)$ does not exist, it is apparent from Eq. (5.18) that this is not the case for $R(x|w)$. In fact, we expect $R(x|w)$ not to be defined when $\zeta(x|y)$ does not exist. Since there seems to be no physical basis for such singularities for real y, we are led to conjecture that the singularities of $\zeta(x|y)$ lie off the real y axis.

VI. CONSTRUCTION OF APPROXIMATE SOLUTIONS

We now consider some approximate methods for solving the integral equation (3.10c) for $\zeta(x|y)$. The object of such methods is to find suitably accurate forms for the resolvent $\mathfrak{R}_\nu(x|z)$. Once we have the latter, it is evident from the work of the preceding section that we can also calculate the functions $\zeta(x|w)$ and $\zeta^T(x|w)$. In this way, we can ultimately determine both $R(x|y)$ and $R(x|w)$.

There exist two basic methods for solving the Fredholm equation (3.10c). First, one may employ the standard Fredholm series solution;[10] however, its application is usually impractical unless the iteration (Neumann) solution is valid. In I, this iteration solution was investigated for the case of a square-well potential and was found to be unreliable, at least for well parameters characteristic of the nucleon–nucleon force.

A second method is to approximate the kernel $\Lambda(x|z)$ by a (finite) series of separable functions, viz.,

$$\Lambda(x|z) \approx \sum_{i=1}^{n} f_i(x)g_i(z). \tag{6.1}$$

With the replacement (6.1), Eq. (3.10c) can be solved exactly in closed form. The remainder of this section will be devoted to the study of the possibility of making the approximation (6.1).

Our principal guide for constructing approximate solutions is the property (3.8) of the resolvent, i.e., we stipulate that any approximate resolvent must satisfy (3.8) in order that $\zeta(x|y)$ be exact for $x = y$. Therefore, if we replace $\Lambda(x|z)$ by a new (approximate) kernel, $\Lambda_a(x|z)$, we demand that: (i) $\Lambda_a(x|z)$ vanish for $x = y$.

Obviously, $\Lambda_a(x|z)$ must be a Fredholm kernel and so we require that: (ii) $\Lambda_a(x|z)$ be well defined for all (x, z) and, in particular, for $z = y$. Finally, we specify that: (iii) $\Lambda_a(x|z)$ have the same asymptotic properties as $\Lambda(x|z)$ for $(x, z) \to \infty$. These three conditions appear to exhaust the possibilities for prescribing the general requirements to be imposed on the kernel $\Lambda_a(x|z)$.

The principal difficulty in obtaining an approximation for $\Lambda(x|z)$ is the fact that the domain of each of the independent variables is infinite. One can always transform to a finite domain, but this sometimes introduces singularities into the kernel. Actually, we will employ such a transformation, but only as a heuristic device.

Let us rewrite $\Lambda(x|z)$ in the form

$$\Lambda(x|z) = [\tau(x|z) - \tau(x|y)]T(y|z)(z - y)^{-1}. \tag{6.2}$$

We then make the change of variables

$$\bar{x} = y(x + y)^{-1}, \qquad \bar{z} = y(z + y)^{-1}, \tag{6.3}$$

and set

$$\tau(x|z) = \bar{\tau}(\bar{x}|\bar{z}). \tag{6.4}$$

If we assume that $\bar{\tau}(\bar{x}|\bar{z})$ can be expanded in a power series about the point $\bar{z} = \frac{1}{2}$ (corresponding to $z = y$) in the range $0 < \bar{z} \leq 1$, and for all $\bar{x}$, we find

$$\Lambda(x|z) = \left(\frac{1}{z - y}\right)\left\{\sum_{n=1}^{\infty} [\partial^{(n)}\bar{\tau}(\bar{x}|\bar{z})/\partial\bar{z}^{(n)}]_{\bar{z}=\frac{1}{2}} \times \frac{(-1)^n}{n!}\left(\frac{z - y}{2(z + y)}\right)^n\right\}T(y|z). \tag{6.5}$$

Since

$$\bar{\tau}(\tfrac{1}{2}|\bar{z}) = 1$$

for all $\bar{z}$, it is clear that in any domain of $\bar{z}$ for which $\bar{\tau}(\bar{x}|\bar{z})$ is analytic,

$$[\partial^{(n)}\bar{\tau}(\bar{x}|\bar{z})/\partial\bar{z}^{(n)}]_{\bar{z}=\frac{1}{2}} = 0 \tag{6.6}$$

for all $n \geq 1$. Therefore, condition (i) is satisfied for each term of the expansion (6.5).

Also, if our expansion is at all valid, it follows that condition (ii) holds for each term in the series (6.5). Moreover, we note that every term except the first vanishes for $z = y$, i.e., on the energy shell.

Condition (iii) cannot be verified unless we specify the nature of $T(x|z)$; however, in view of Eqs. (6.2) and (3.2), it is reasonable to assume that the behavior of $\Lambda(x|z)$ as $z \to \infty$ is determined solely by the factor $T(y|z)(z - y)^{-1}$. If this is the case, then

each term in (6.5) retains this asymptotic property. A similar type of argument leads us to expect that

$$[\partial^{(n)}\bar{\tau}(\bar{x}|\bar{z})/\partial\bar{z}^{(n)}]_{\bar{z}=\frac{1}{2}}$$

vanishes like $T(x|z)$ as $x\to\infty$.

In virtue of the preceding arguments, we conclude that each term of the expansion (6.5) satisfies conditions (i)–(iii). This series may therefore be cut off after any finite number of terms, and one will still be left with a kernel of the form (6.1) which fulfills the requirements (i)–(iii).

It is evident that the series (6.5) is valid provided $\tau(x|z)$ is analytic in that domain of z given by

$$|(z-y)/(z+y)| \leq 1.$$

The question of whether $\tau(x|z)$ is analytic in z in this domain can be answered only if $T(x|z)$ is specified. It is most likely that $\tau(x|z)$ has singularities where $T^{-1}(y|z)$ does not exist. However, in this connection, it is interesting to note that, if one solves (3.10c) with $\Lambda_a(x|z)$ consisting of a finite number of terms of (6.5), one obtains a solution which is generally well defined, even if $T^{-1}(y|y)$ does not exist. This suggests that (6.5) may be useful in providing an asymptotic expansion for $\Lambda(x|z)$, though the complete series may not converge.

A special case of the separable-kernel method is simply to take $\Lambda_a(x|z)$ to be identically zero. The new kernel clearly satisfies conditions (i) and (ii) and possibly (iii). The resultant approximation for $\zeta(x|y)$ was investigated in I for a square-well potential (without a h.c.) and was found to be fairly accurate for x in a rather large neighborhood about y. Also, it can be shown that this approximation, when employed in (4.2) with $x=y$, is equivalent to the choice of a plane or h.c. wavefunction, whichever is appropriate, as a trial function in a variational principle for $R(y|y)$ of the type first derived by Schwinger.[30,31]

The accuracy of the representation of $\zeta(x|y)$, which is obtained by retaining only the first term of (6.5) was also investigated in I, once again for the case of a square-well potential without a h.c. It was found that the agreement between the $\zeta(x|y)$ calculated in this manner and the exact value was excellent.

We have not been able to find any method other than the preceding technique for constructing a separable-kernel approximation which satisfies conditions (i)–(iii). It may well be that other procedures will suggest themselves for particular forms of $T(x|z)$.

VII. EXTENSION OF THE FREDHOLM FORMALISM

All of our work up to now concerning the Fredholm method has been confined to those equations satisfied by the partial-wave amplitudes of the various scattering operators. A question which naturally arises is whether or not this technique can be applied directly to the integral equations for the momentum-space representatives of these operators.

Now, the momentum-space representatives are determined by three-dimensional integral equations in contrast to the one-dimensional equations for the partial-wave amplitudes. However, the singularity in the kernel involves only one variable, namely, the magnitude of the momentum. Thus, a Fredholm reduction which treats all three variables is unnecessary. Also, although one can formally reduce Eqs. (2.1) to Fredholm forms in the manner of Sec. III, one is led to terms containing inverse operators, e.g., the operator generalization of $\tau(x|y)$. But, in actual practice, the only convenient way available for evaluating such terms is by means of a partial-wave analysis. So, in the end, one is led right back to considering the partial-wave amplitudes. This conclusion also implies that the application of the Fredholm technique to the many-body scattering problem is impractical.

APPENDIX: HARD-CORE OPERATORS

The quantities $\omega_i^{(*)}$ and $\Gamma_i^{(*)}$ are defined in terms of their coordinate representatives as

$$\begin{aligned}\langle\mathbf{r}|\omega_i^{(*)}|\mathbf{r}'\rangle &= -\frac{\delta(r-a)}{a^2}\\ &\quad\times\sum_l\left(\frac{2l+1}{4\pi}\right)\frac{g_l^{(*)}(a|r')}{g_l^{(*)}(a|a)}P_l(\mathbf{r}\cdot\mathbf{r}'),\\ \langle\mathbf{r}|\Gamma_i^{(*)}|\mathbf{r}'\rangle &= -\frac{\delta(r-a)\,\delta(r'-a)}{a^4}\\ &\quad\times\sum_l\left(\frac{2l+1}{4\pi}\right)\frac{P_l(\mathbf{r}\cdot\mathbf{r}')}{g_l^{(*)}(a|a)},\end{aligned}\tag{A1}$$

where $\mathbf{r}$ is the relative position vector of the two particles and a is the h.c. radius[32]; $P_l(\mathbf{r}\cdot\mathbf{r}')$ is the Legendre polynomial which is a function of the cosine of the angle between $\mathbf{r}$ and $\mathbf{r}'$. The radial Green's functions, $g_l^{(*)}(r|r')$, are given by

[30] J. Schwinger, Lecture Notes on Nuclear Physics, Harvard, 1947 (unpublished).

[31] A generalization of this variational principle which is valid when one has a h.c. follows from the formal identity of the integral equations for $t_i^{(+)\prime}$ and K'_i to those for $t_i^{(+)}$ and K_i, respectively, when the latter quantities are defined for an interaction without a h.c.

[32] The h.c. radius is, in general, a function of both the total ordinary- and isotopic-spin quantum numbers.

$$g_l^{(+)}(r|r') = [g_l^{(-)}(r|r')]^* = -i\left(\frac{2Mk_i}{\hbar^2}\right) \times \begin{cases} j_l(k_i r')h_l^{(1)}(k_i r), & r > r', \\ j_l(k_i r)h_l^{(1)}(k_i r'), & r < r', \end{cases} \quad \text{(A2)}$$

where j_l is the spherical Bessel function, and $h_l^{(1)}$ is the spherical Hankel function of the first kind.

The h.c. operators $\bar{\omega}_i$ and $\bar{\Gamma}_i$ are obtained from the expressions (A1) for $\omega_i^{(\mp)}$ and $\Gamma_i^{(\mp)}$, respectively, by replacing $g_l^{(\mp)}(r|r')$ by $\bar{g}_l(r|r')$; $\bar{g}_l(r|r')$, in turn, is determined from (A2) by substituting n_l for $(-i)h_l^{(1)}$, where n_l is the spherical Neumann function.

It can be verified with the aid of the coordinate representatives (A1) that the h.c. operators satisfy the following identities[33]:

$$\omega_i^{(\mp)\dagger} G_i^{(\mp)} = G_i^{(\mp)} \omega_i^{(\mp)}, \quad \text{(A3a)}$$

$$\bar{\omega}_i^\dagger \bar{G}_i = \bar{G}_i \bar{\omega}_i, \quad \text{(A3b)}$$

$$\omega_i^{(\mp)2} = -\omega_i^{(\mp)}, \quad \text{(A4a)}$$

$$\bar{\omega}_i^2 = -\bar{\omega}_i, \quad \text{(A4b)}$$

$$\omega_i^{(\mp)\dagger} = G_i^{(\mp)} \Gamma_i^{(\mp)}, \quad \text{(A5a)}$$

$$\bar{\omega}_i^\dagger = \bar{G}_i \bar{\Gamma}_i, \quad \text{(A5b)}$$

[33] We note that $G_i^{(-)} = G_i^{(+)\dagger}$.

$$\Gamma_i^{(\mp)} = \bar{\Gamma}_i \mp i\pi \bar{\Gamma}_i \delta(E_i - H_0)\Gamma_i^{(\mp)}, \quad \text{(A6)}$$

$$\omega_i^{(\mp)} = \bar{\omega}_i \mp i\pi \bar{\Gamma}_i \delta(E_i - H_0)[1 + \omega_i^{(\mp)}]. \quad \text{(A7)}$$

Contrary to the statement made in the Appendix of I, none of the preceding relations is restricted in any way with respect to the states on which it operates.

More generally (see reference 4), one needs also to introduce the operators $\gamma_i^{(\mp)}$ which are defined by

$$\langle \mathbf{r}|1 + G_i^{(\mp)}\gamma_i^{(\mp)}|j\rangle = \langle \mathbf{r}|1 + G_i^{(\mp)}\Gamma_i^{(\mp)}|j\rangle, \quad |\mathbf{r}| \geq a, \qquad = 0, \quad |\mathbf{r}| < a, \quad \text{(A8)}$$

where $|j\rangle$ is any eigenstate of H_0. A corresponding expression for $\bar{\gamma}_i$ is obtained from (A8) by deleting the superscripts $(\pm)$ and placing a bar over all operators. The definition (A8) for the γ's appears quite naturally when the method of I for applying the h.c. boundary conditions is adapted to the representatives, $\langle \mathbf{r}\ |\Omega_i^{(+)}|j\rangle$, of the wave operator $\Omega_i^{(+)}$, for example. The difference between the γ's and the Γ's has been stressed by Brander[4] who, by considering the h.c. as a limiting case of a repulsive square-well potential, has obtained an explicit form for the partial-wave amplitudes of $\gamma_i^{(+)}$.

Unified Variational Formulation of Classical and Quantum Dynamics. II

EDWARD M. CORSON
Georgetown University, Washington 7, D. C.
(Received 22 October 1962)

In a previous article in this series [E. M. Corson, J. Math. Phys. 4, 42(1963); herein referred to as I], we showed that a new variational principle, of Lagrangian structure, includes Hamilton's principle as a unique derived consequent and encompasses both the classical and quantal definitions of "state" *and* the corresponding Lagrangian and Schrödinger equations of motion, respectively. In this theory, the essential difference between classical and quantal domains was shown to arise *solely* from the superadded postulates of determinism vs indeterminism; other than this, the two basic postulates may be reduced to (a) Newton's equations and (b) the new variation principle.

It was shown that in the quantal domain, Schrödinger's equations obtain for the transition amplitudes (propagators) between pure eigenstates of homologous variables in configuration space, at different times, thus clearly implying all the results of conventional Hamiltonian-operator quantum formalism. In the present article we show that our variational formulation, under the familiar *classical* canonical transformation to momentum space, also automatically leads to the corresponding Schrödinger equations for the transition amplitudes between pure eigenstates of homologous variables in momentum space, at different times, the unitarity of the formalism and the reciprocal Fourier integral relations between coordinate and momentum representations. Thus we bridge a long-standing gap in theory and derive, uniquely and apparently for the first time, the basic quantum commutators (P.B.'s) which are postulated in the more familiar Hamiltonian theory.

1. THE COORDINATE-SPACE FORMULATION

FOR convenience of reference, as well as logic and symmetry of presentation of the theoretical development, we start again from the coordinate form of our variational postulate: To any path (more generally, history or process), geometrically considered, between the points P_1 and P_2, we assign the dynamical phase characteristic $\exp(iS/\hbar)$.

We take as our sole variational postulate [cf. Eq. (1′) *infra.*], I:

$$\delta \sum_{\pi \in C} \exp(iS_\pi/\hbar) = 0, \tag{1}$$

where $\pi \in C$ denotes the set of paths contained in the class of paths considered. The two obvious extreme cases are:

(a) *Classical* (deterministic or definite path case):

$$\delta \exp(iS/\hbar) = 0. \tag{2a}$$

Hamilton's principle, $\delta S = 0$, is implied and the usual basic classical concept of state and Lagrange equations of motion follow.

(b) *Quantal* (indeterministic or indefinite path case):

$$\delta \sum_{\pi \in C} \exp(iS/\hbar) = \delta \int \cdots \int \exp\left(\frac{i}{\hbar}\int_{t_1}^{t_2} L[x(\tau)]\, d\tau\right) \times \prod_{t_1}^{t_2} dx(\tau) = 0. \tag{2b}$$

Since all paths, including all possible variations, are already included in $\sum \exp(iS/\hbar)$, it follows that the latter is not a functional of path but merely a function of the end points x_2t_2; x_1t_1 which are arbitrary but fixed under the variation considered, so that the variation is identically zero. Hence,

$$\int \cdots \int \exp(iS[x]/\hbar) \prod_{t_1}^{t_2} dx(\tau) \equiv NK(x_2t_2; x_1t_1) = N\langle x_2t_2 \mid x_1t_1\rangle, \tag{3}$$

where N is a constant, and conventional notation of quantum mechanics is used to facilitate interpretations and discussion. From the known properties of the left-hand side of Eq. (3),[1] the notation $\langle x_2t_2 \mid x_1t_1\rangle$ is appropriate and the usual basic quantal concept of state and Schrödinger equations of motion follow.

We note again that we may also take the alternative, and physically more suggestive, basic variational postulate, II:

$$\delta \mid \sum_{\pi \in C} \exp(iS_\pi/\hbar)\mid^2 \equiv 0. \tag{1′}$$

In the quantal domain this alternate form makes no essential difference but in the classical domain Hamilton's principle appears now as a unique, derived consequent.

2. THE MOMENTUM-SPACE FORMULATION

It is well known that the terminology "coordinate" variable or "momentum" variable is a relative description, with respect to canonical transformations,

[1] E. M. Corson, J. Math. Phys. 4, 42 (1963).

and the dynamical roles of the two sets of variables are readily interchanged. Moreover, on familiar grounds of symmetry, we certainly may expect all the results of the preceding section to reappear in momentum-space form. Indeed, we shall see that the familiar, though rarely used, *classical* canonical transformation to momentum space also automatically leads to the corresponding Schrödinger equations for the transition amplitudes between pure eigenstates of homologous momentum variables, at different times, as well as the reciprocal Fourier integral relations between coordinate and momentum representations. Thus, we derive uniquely, and apparently for the first time, the basic quantum commutators (P. B.'s) which are postulated in the more familiar Hamiltonian operator theory. The general role of operators, eigenstates, etc., now may be viewed as a purely mathematical issue. This is also the case with respect to physical applications where the conventional formalism remains, at present, the technique of choice because the mathematical structure is better known and developed, whereas the theory of function space integration as such, is still in process of development.

We now make the indicated canonical transformation to momentum variables:

$$L(x, \dot{x}, t) = (d/dt)(xp) + K(p, \dot{p}, t), \qquad (4)$$

where, for simplicity of notation, we avoid summations over the number of possible dynamical variables (degrees of freedom). Thus, one must bear in mind that our equations strictly hold for Cartesian variables in one dimension, a matter of importance primarily for the interpretation of certain integrals and powers of h which will appear in certain equations. This however, is not a restriction of any consequence in principle.

Our variational postulate (taking the first statement for simplicity) now involves the form:

$$\sum \exp\left(\frac{i}{\hbar}\int_{t_1}^{t_2} L\, d\tau\right)$$

$$= \sum \exp\left\{\frac{i}{\hbar}\left[(x_2p_2 - x_1p_1) + \int_{t_1}^{t_2} K\, d\tau\right]\right\}, \qquad (5a)$$

or the equivalent converse form

$$\sum \exp\left(\frac{i}{\hbar}\int_{t_1}^{t_2} K\, d\tau\right)$$

$$= \sum \exp\left\{\frac{i}{\hbar}\left](x_1p_1 - x_2p_2) + \int_{t_1}^{t_2} L\, d\tau\right]\right\}, \qquad (5b)$$

where again, the symbolic sums or function space integrals appear only in the quantal domain.

The essential step here is to realize that the symbolic sums require interpretation in terms of different spaces—coordinate or momentum space—and conversely, on the two sides of Eqs. (5a) and (5b), respectively. This is clear from several points of view but most immediately evident from the simple fact that in Eq. (5a), for example, the left-hand side is merely a function of coordinates and times, x_2t_2; x_1t_1; hence, the momentum variables on the right-hand side must integrate out. The same remarks hold, symmetrically, if we start from the equivalent converse form (5b).

On a deeper theoretical level, one realizes (cf. *infra.*) that the symbolic sums over all paths or histories, when coupled with our exponential action form, in the space of the relevant dynamical variables, is actually a new representation of Heisenberg's indeterminacy principle. Thus, in considering Eqs. (5a), (5b), we must keep in mind that (in the quantal domain) the meaning of the symbolic sums is simply that the coordinate and momentum variables are, in effect, released from the usual classical restrictions and treated as independent, in a definite limited sense, even though we are working within a Lagrangian formalism rather than a Hamiltonian formalism wherein this uncoupling, so to speak, is fundamental and natural to the formalism. It is especially to be noted that, apart from the symbolic sums, all our dynamical considerations are classical and this independent status of coordinate and momentum variables does not hold for any single classical path or history. Rather, in the symbolic sums, for *given* initial and final points in the coordinate space picture, all momentum values obtain, and conversely. Therefore, the independent status of the canonically conjugate variables is not generic but quite specifically limited and, indeed, this is the precise equivalent of the relation of the coordinate and momentum diagonal representations of familiar quantum theory.

One sees in the above observations the key points underlying the present theory which, fundamentally, may be said to depend upon asking the appropriate question about transition amplitudes between what turn out to be eigenstates of homologous dynamical variables, rather than the mixed transition amplitudes (or preferably, in this instance, mixed representatives), which are so central to conventional quantum mechanics primarily because of the historical emphasis on determination of eigenvalues.

Returning to Eq. (5a), for example, we have again two obvious extreme cases under the variational postulate:

(a) *Classical:*

$$\delta[\exp(iR[p]/\hbar)] \equiv \delta\left[\exp\left(\frac{i}{\hbar}\int_{t_1}^{t_2} K\,d\tau\right)\right] = 0, \quad (6a)$$

the usual classical concept of state and Lagrange equations of motion, in momentum space, follow

(b) *Quantal:*

$$\delta \sum \exp(iR_\pi/\hbar) \equiv \delta \int_{-\infty}^{\infty}\cdots\int \exp\left[\frac{i}{\hbar}\int_{t_1}^{t_2} K[p(\tau)]\,d\tau\right] \times \prod_{t_1}^{t_2} dp(\tau) = 0. \quad (6b)$$

Here, on grounds of symmetry in the canonically conjugate variables and formal analogy with Eq. (2b), we infer

$$\int\cdots\int \exp(iR[p]/\hbar) \prod_{t_1}^{t_2} dp(\tau) \equiv M\langle p_2 t_2 \mid p_1 t_1\rangle, \quad (7)$$

where again, M is a constant to be determined from the usual normalization. We emphasize that we are using the arguments which are known to substantiate the coordinate-space analog [Eq. (3)], though here in a much more formal way.

THE RECIPROCAL FOURIER INTEGRAL RELATIONS AND THE QUANTUM P.B.'S

We have now the following essential equations: [cf. Eqs. (3), (7), (5a), resp.]

$$\int\cdots\int \exp(iS[x]/\hbar) \prod_{t_1}^{t_2} dx(\tau) \equiv N\langle x_2 t_2 \mid x_1 t_1\rangle, \quad (8a)$$

$$\int\cdots\int \exp(iR[p]/\hbar) \prod_{t_1}^{t_2} dp(\tau) \equiv M\langle p_2 t_2 \mid p_1 t_1\rangle, \quad (8b)$$

$$\sum \exp\left(\frac{i}{\hbar}\int_{t_1}^{t_2} L\,d\tau\right) = \sum \exp\left\{\frac{i}{\hbar}\left[(x_2p_2 - x_1p_1) + \int_{t_1}^{t_2} K\,d\tau\right]\right\}. \quad (8c)$$

It is now easily found that, regardless of the dynamical structure of particular problems, $M = Nh^{-1}$ so that we have [cf. Eqs. (5a), (5b)]

$$\langle x_2 t_2 \mid x_1 t_1\rangle = h^{-1} \iint_{-\infty}^{\infty} \exp\left[\frac{i}{\hbar}(x_2p_2 - x_1p_1)\right] \times \langle p_2 t_2 \mid p_1 t_1\rangle\, dp_2\, dp_1, \quad (9a)$$

and reciprocally,

$$\langle p_2 t_2 \mid p_1 t_1\rangle = h^{-1} \iint_{-\infty}^{\infty} \exp\left[\frac{i}{\hbar}(x_1p_1 - x_2p_2)\right] \times \langle x_2 t_2 \mid x_1 t_1\rangle\, dx_2\, dx_1. \quad (9b)$$

These are, of course, merely a particular form of the familiar relations

$$\langle x|\rangle = h^{-\frac{1}{2}} \int \exp(ixp/\hbar)\, dp\, \langle p|\rangle, \quad (10a)$$

$$\langle p|\rangle = h^{-\frac{1}{2}} \int \exp(-ixp/\hbar)\, dx\, \langle x|\rangle, \quad (10b)$$

and

$$\langle x \mid p\rangle = h^{-\frac{1}{2}} \exp(ixp/\hbar). \quad (11)$$

Since these last equations, in the context of the operator formulation of quantum mechanics, are direct and unique consequences of the fundamental quantum commutators or P. B.'s, it follows that the present theoretical formulation leads uniquely to the conventional formulation, as already anticipated and for the reasons noted.

Eigenfunction Form of the Nonrelativistic Coulomb Green's Function*

KOICHI MANO
Wentworth Institute, Boston, Massachusetts
(Received 22 October 1962)

By treating the contour integrals in the momentum plane, the conventional eigenfunction form is deduced from the characteristic form for the nonrelativistic Coulomb Green's function. A straightforward extension of the argument employed here leads to the most general form of the eigenfunction form.

I. INTRODUCTION

THE eigenfunction form of the nonrelativistic Coulomb Green's function was derived by Mapleton[1] from its characteristic form by dealing with contour integrals in the plane which is physically equivalent to the space of complex energy. However, in view of the fact that both forms of the Green's function are customarily given in terms of momentum, it would be more natural to carry it through by explicitly having the momentum as the variable. Moreover, certain interesting aspects reveal themselves when it is dealt with in this way. In particular it is possible to obtain the most general expression for the eigenfunction form by generalizing the argument developed in connection with the derivation of the conventional form.

II. DECOMPOSITION OF F(α, β; Z)

The nonrelativistic Green's function of a Coulomb field satisfying the diverging or converging wave boundary condition (labeled by + or −) is given by[1,2]

$$G^{(\pm)}(\mathbf{r}, \mathbf{r}'; K) = -\frac{K}{4\pi} \sum_{l=0}^{\infty} (2l+1) P_l(\cos\theta_{rr'}) \times H_l^{(\pm)}(r, K) L_l(r', K) \qquad (r > r'). \tag{1}$$

Restricting the discussion for definiteness to the case of an attractive Coulomb field, the functions L_l are given by

$$L_l(r, K) = e^{\pi/2Ka} \frac{|\Gamma(l+1-i(Ka)^{-1})|}{(2l+1)!} (2Kr)^l e^{iKr} \times F(l+1-i(Ka)^{-1}, 2l+2; -2iKr) \equiv M_l(r, K) F. \tag{2}$$

In Eq. (2), $F = F(l+1-i(Ka)^{-1}, 2l+2; -2iKr)$ stands for the confluent hypergeometric function ${}_1F_1$, $K = \mu v/\hbar$, and $a = \hbar^2/\mu ZZ'e^2$, where Ze and $Z'e$ (Z, $Z' > 0$) are the charges, μ is the reduced mass, and v is the relative velocity. The functions L_l constitute the bounded solutions of the radial equation having the asymptotic form

$$L_l \sim (Kr)^{-1} \sin [Kr - l\pi/2 + \eta(l, K) + (Ka)^{-1} \log 2Kr], \tag{3}$$

where

$$\eta(l, K) = \arg \Gamma(l+1-i(Ka)^{-1}).$$

Corresponding to the decomposition $F = W_1 + W_2$ which will be specified shortly, the function L_l is split into two parts whose asymptotic forms are given by the diverging and converging part of the standing wave represented by the right-hand side of Eq. (3).

In terms of W_1 and W_2 the unbounded solution may be taken as

$$K_l(r, K) = M_l(r, K) i(W_1 - W_2) \tag{4}$$

with the cosine asymptotic form instead of the sine form of Eq. (3).

The functions $H_l^{(\pm)}$ are related to K_l, L_l, and $W_{1,2}$ by[3]

$$H_l^{(\pm)} = K_l \pm iL_l = \pm 2iM_l W_{1,2}.$$

If G_c and G_s are defined by $G^{(\pm)} = G_c \pm iG_s$, they are given by

$$\left.\begin{matrix} G_c \\ G_s \end{matrix}\right\} = \sum \Theta_l(\theta_{rr'}, K) \left\{\begin{matrix} K_l(r, K) \\ L_l(r, K) \end{matrix}\right\} L_l(r', K), \tag{5a, 5b}$$

where

$$\Theta_l(\theta_{rr'}, K) = -(4\pi)^{-1} K(2l+1) P_l(\cos\theta_{rr'}).$$

Equation (5b) foretells that the combination

* Work performed under the contract with the U. S. Air Force, Cambridge Research Laboratories.

[1] R. A. Mapleton, J. Math. Phys. 2, 478 (1961). This paper is referred to hereafter as A.

[2] N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Clarendon Press, Oxford, England, 1949), 2nd Ed., pp. 53 and 113.

[3] The subscripts 1 and 2 on W correspond to (+) and (−) on H_l, respectively.

$2iG_c = G^{(+)} - G^{(-)}$ of $G^{(+)}$ and $G^{(-)}$ is required for the continuum part of the eigenfunction form of the Green's function.

The desired decomposition of F into W_1 and W_2 is shown in A to be given for the case of positive integer β by

$$W_1(\alpha, \beta; z) = \frac{(\beta - 1)!}{2\pi i} \int_{-\infty}^{(0+)} dt\, e^t t^{\alpha-\beta}(t - z)^{-\alpha}, \quad (6a)$$

$$W_2(\alpha, \beta; z) = \frac{(\beta - 1)!}{2\pi i} \int_{-\infty}^{(0+)} dt\, e^t t^{-\alpha}(t + z)^{\alpha-\beta}. \quad (6b)$$

In these equations the initial and the terminal phases of the factors in the integrand are fixed as follows: In Eq. (6a), they are $-\pi$ and $+\pi$ for t, and $+\pi$ and $+\pi$ for $t - z$; in Eq. (6b), $-\pi$ and $+\pi$ for t, and $-\pi$ and $-\pi$ for $t + z$. It is easy to see, by applying the change of variable $t + z \to t$ to Eq. (6b), that the sum of W_1 and W_2 reduces to the integral representation for F when Im $z < 0$ (Im stands for "the imaginary part of"):

$$F(\alpha, \beta; z) = W_1(\alpha, \beta; z) + W_2(\alpha, \beta; z) \quad (\mathrm{Im}\, z < 0). \quad (7)$$

By adjusting the phases of $t + z$ in W_2 so as to coincide with those of $t - z$ in W_1, it is immediately seen that $W_2(\alpha, \beta; z)$ is expressible in terms of $W_1(\beta - \alpha, \beta; -z)$. A similar argument holds for $W_1(\alpha, \beta; z)$. These relations are combined to give

$$W_{1,2}(\alpha, \beta; z) = e^z e^{-2\pi i\alpha} W_{2,1}(\beta - \alpha, \beta; -z). \quad (8)$$

Note that the use of Eq. (8) leads to the possibilities of expressing $F(\alpha, \beta; z)$ (Im $z < 0$) in forms other than that of Eq. (7).

It is to be remembered that $F(\alpha, \beta; z)$ is not equal to the sum of $W_1(\alpha, \beta; z)$ and $W_2(\alpha, \beta; z)$ defined above if Im $z > 0$. In fact, the decomposition of

$$F(\alpha, \beta; z) = \frac{(\beta - 1)!}{2\pi i} \int^{(0+,z+)} dt\, e^t t^{\alpha-\beta}(t - z)^{-\alpha},$$

valid for any z, into two parts, with appropriate initial and terminal phases of the factors in the integrand that are consistent with the condition Im $z > 0$, gives directly

$$F(\alpha, \beta; z) = e^z[W_1(\beta - \alpha, \beta; -z) + W_2(\beta - \alpha, \beta; -z)] \quad (\mathrm{Im}\, z > 0). \quad (9)$$

That Eq. (9) is true may be seen from the fact that it becomes identical with the relation

$$F(\alpha, \beta; z) = e^z F(\beta - \alpha, \beta; -z) \quad (10)$$

when Eq. (7) is applied to the right-hand side of Eq. (9), by noticing that Im $(-z) < 0$. In view of Eq. (8) it is found that[4]

$$F(\alpha, \beta; z) = e^{2\pi i\alpha}[W_1(\alpha, \beta; z) + W_2(\alpha, \beta; z)] \quad (\mathrm{Im}\, z > 0). \quad (11)$$

It should be mentioned here that one could start with the bounded solution with $e^{-iKr} F(l + 1 + i(Ka)^{-1}, 2l + 2; +2iKr)$, rather than the similar expression used in Eq. (2), and decompose this F into diverging and converging wave part U_1 and U_2, respectively.[5] The functions U_1 and U_2 are expressible in terms of W's by

$$U_{1,2}(\alpha, \beta; z) = e^z W_{1,2}(\beta - \alpha, \beta; -z) \quad (\mathrm{Im}\, z > 0) \quad (12)$$

as is seen from Eq. (9). Here again the sum of $U_1(\alpha, \beta; z)$ and $U_2(\alpha, \beta; z)$ is not equal to $F(\alpha, \beta; z)$ when Im $z < 0$. In terms of $U_{1,2}$ the functions L_l, K_l, and $H_l^{(\pm)}$ are given, respectively, by

$$L_l(r, K) = N_l(r, K)F(l + 1 + i(Ka)^{-1}, 2l + 2; +2iKr) = N_l(U_1 + U_2),$$

$$K_l(r, K) = N_l i(U_1 - U_2),$$

and

$$H_l^{(\pm)}(r, K) = \pm 2iN_l U_{1,2},$$

where

$$N_l(r, K) = e^{\pi/2Ka} \frac{|\Gamma[l + 1 + i(Ka)^{-1}]|}{(2l + 1)!} (2Kr)^l e^{-iKr}.$$

Before concluding this section, we demonstrate that certain properties of various functions are displayed when they are considered as functions of a complex variable k. For this purpose define $\mathfrak{F}(k)$ and $\mathfrak{W}_{1,2}(k)$ by

$$\mathfrak{F}(k) = F(\alpha(k), \beta; z(k)),$$

and

$$\mathfrak{W}_{1,2}(k) = W_{1,2}(\alpha(k), \beta; z(k)),$$

where

$$\alpha(k) = l + 1 - i(Ka)^{-1}, \qquad \beta = 2l + 2,$$

[4] Gordon, for instance, gives a similar decomposition which is valid regardless of the sign of Im z. [W. Gordon, Z. Physik, 48, 180 (1928)]. That the same is not the case here may be understood as follows: In the decomposition given by Eq. (7) for Im $z < 0$, specifications on the phases are so rigid that no room is left in which to adjust the phases which would reinstate the validity of Eq. (7) for Im $z > 0$.

[5] Decomposition of this nature is used in R. A. Mapleton, J. Math. Phys. 3, 297 (1962), which refines some of the arguments in his earlier work [J. Math. Phys. 2, 482 (1961)]. It is to be noted that the W functions defined in these two articles are not exactly the same. The $W_{1,2}$ used in the former are what are called $U_{1,2}$ here. It is easily seen, however, that the argument in terms of U is analogous to that in terms of W with the replacement of M_l by N_l.

and $$z(k) = -2ikr.$$

Then the use of Eqs. (8), (9), and (10) results in[6]

$$\begin{aligned} L_l(r, -k) &= (-)^l e^{-\pi/ka} L_l(r, k), \\ H_l^{(\mp)}(r, -k) &= (-)^{l+1} e^{+\pi/ka} H_l^{(\mp)}(r, k), \end{aligned} \tag{13}$$

which in turn leads to

$$G^{(\mp)}(\mathbf{r}, \mathbf{r}'; -k) = G^{(\mp)}(\mathbf{r}, \mathbf{r}'; k). \tag{14}$$

In addition it is seen that

$$L_l(r, k)^* = L_l(r, k). \tag{15}$$

Furthermore, when k is real and positive K, say, the following obtains:

$$\begin{gathered} \mathfrak{F}(K)^* = \mathfrak{F}(-K), \qquad \mathfrak{W}_{1,2}(K)^* = \mathfrak{W}_{1,2}(-K), \\ H_l^{(\mp)}(r, K)^* = e^{2\pi/Ka} H_l^{(\mp)}(r, K), \\ G^{(\mp)}(\mathbf{r}, \mathbf{r}'; K)^* = e^{2\pi/Ka} G^{(\mp)}(\mathbf{r}, \mathbf{r}'; K). \end{gathered}$$

III. DISCRETE PART

The eigenfunction form was obtained in A by considering contour integrals in $\lambda(=k^2)$ plane. In the following, a detailed discussion is developed to accomplish the same in terms of the variable k with the diverging-wave case as an example after having a brief look at the work in A.

By regarding $G^{(+)}(k) \equiv G^{(+)}(\mathbf{r}, \mathbf{r}'; k)$ as a function of $\lambda = k^2$ called $\mathcal{G}^{(+)}(\lambda)$, and assuming $0 < \arg \lambda < 2\pi$ for the principal value of $\arg \lambda$, it is shown in A that $\mathcal{G}^{(+)}(\Lambda_+)$[7] ($\Lambda_+ = K^2 + i\epsilon$, ϵ being a positive real number) consists of two contributions. The first comes from the sum of residues of $-\mathcal{G}^{(+)}(\lambda)/(\lambda - \Lambda_+)$ at the poles of $\mathcal{G}^{(+)}$ on the negative real axis and the second form the integral

$$\begin{aligned} I_C^{(+)} &= -\frac{1}{2\pi i} \int_C \frac{d\lambda\, \mathcal{G}^{(+)}(\lambda)}{\lambda - \Lambda_+} \\ &= \frac{1}{2\pi i} \left[\int_0^\infty \frac{d\lambda\, \mathcal{G}^{(+)}(\lambda)}{\lambda - \Lambda_+} + \int_\infty^0 \frac{d\lambda\, \mathcal{G}^{(+)}(\lambda e^{2\pi i})}{\lambda - \Lambda_+} \right], \end{aligned} \tag{16}$$

taken along the contour C encircling the branch cut on the positive real axis, resulting in the discrete and continuum part, respectively. This is expected to be so because the variable λ is physically equivalent to the complex energy of the system.

In terms of the momentum (i.e., k) plane, the contours to be associated with $G^{(+)}$ and $G^{(-)}$ are those in its upper and lower half-plane, respectively. Evidently, these half-planes correspond to different Riemann sheets of λ plane for which $0 \lesseqgtr \arg \lambda \lesseqgtr \pm 2\pi$.

Now consider the whole upper half-plane from which the neighborhoods of the poles of $G^{(+)}$ on the positive imaginary axis are excluded by surrounding them with infinitesimally small circles. Since there seems to exist no obvious reason why the simplest conceivable choice should not be tried, the function $(2\pi i)^{-1} F_1^{(+)}$, where $F_1^{(\mp)}(k) = G^{(\mp)}(k)/(k \mp K_+)$ [$K_+ = K + i\epsilon'$, ϵ' being an infinitesimal positive real number], is integrated along the boundary of this region to give

$$G^{(+)}(K_+) = \frac{1}{2\pi i} \left[\int_{\mathrm{I}} + \int_{\mathrm{II}} \right] dk\, F_1^{(+)}. \tag{17}$$

In Eq. (17), the contour I consists of the following three parts: the positive real axis, that part of the circle at infinity bounded in the first quadrant; and the positive imaginary axis with indentations to the right of the imaginary axis at each pole of $G^{(+)}$. The contours symmetrical to I with respect to the imaginary axis, the origin, and the real axis are called II, III, and IV, respectively. Now apply the change of variable $k \to -k$ to the second integral of Eq. (17) noticing that the vanishing value of the integral is contributed from an integral along the negative real axis and $-\frac{1}{2}$ times the sum of residues of $F_1^{(+)}$. Then considerations on Eq. (14) suggest that the latter part of the contribution should be equal to $-\frac{1}{2}$ times the sum of residues of $F_1^{(-)}$ which is a part of $(2\pi i)^{-1} \int_{\mathrm{IV}} dk F_1^{(-)}$ that also vanishes. That this is in fact the case may be seen from Eq. (20), to be established below (for $t = 0$ and $s = 1$). For this reason note firstly that the poles of $G^{(\mp)}(k)$ given by Eq. (1) are located at $k_n = \pm i/na$, $n = 1, 2, 3, \cdots$. Secondly it is not difficult to see that[8,9]

$$\begin{aligned} -\Sigma \operatorname{Res} G^{(+)}(k) &= +\Sigma \operatorname{Res} G^{(-)}(k) \\ &= \sum_{n=1}^{\infty} \sum_{l=0}^{n-1} \sum_{m=-l}^{+l} \frac{ina}{2} \phi_{nlm}(\mathbf{r}) \phi_{nlm}(\mathbf{r}')^*, \end{aligned} \tag{18}$$

where

$$\begin{aligned} \phi_{nlm}(\mathbf{r}) = &\left[\left(\frac{2}{na}\right)^3 \frac{(n-l-1)!}{\{(n+l)!\}^3 2n} \right]^{\frac{1}{2}} e^{-r/na} \left(\frac{2r}{na}\right)^l \\ &\times L_{n+l}^{2l+1}\left(\frac{2r}{na}\right) Y_{lm}(\theta, \varphi). \end{aligned} \tag{19}$$

These informations are sufficient to deduce that

$$\begin{aligned} -\Sigma \operatorname{Res} \frac{k' G^{(+)}}{k^s - A} &= (-)^{t+s} \Sigma \operatorname{Res} \frac{k' G^{(-)}}{k^s - A} \\ &= -\frac{1}{2} \left(\frac{i}{na}\right)^{t-s-1} \sum_{n,l,m} \frac{\phi_{nlm}(\mathbf{r}) \phi_{nlm}(\mathbf{r}')^*}{1 - (-ina)^s A}. \end{aligned} \tag{20}$$

[6] To be more exact, $-k$ should appropriately be identified as either $ke^{+i\pi}$ or $ke^{-i\pi}$. However, for most cases, it is permissible and simpler to use $-k$.

[7] For the relationship between the solutions obtained from $G^{(+)}((K^2 + i\epsilon)^{1/2})$ and $G^{(+)}(K)$, see F. M. Odeh, J. Math. Phys. 2, 794 (1961).

[8] See, for example, reference A.

[9] The spherical harmonics Y_{lm} are those defined in H. A. Bethe and E. E. Salpeter, in *Handbuch der Physik*, edited by S. Flügge (Springer-Verlag, Berlin, 1957), Vol. XXXV.

The argument in the above assures one that the second integral of Eq. (17) is transposed by $k \to -k$ to $(2\pi i)^{-1} \int_{IV} dk F_1^{(-)}$. In other words, the original contour integral for $G^{(+)}(K_+)$ taken in the upper half-plane is now transposed to the sum of similar integrals taken in the upper and lower quarter of the right half-plane. Similar transposition of the first integral of Eq. (17) to the one in the third quadrant suggests various representations for $G^{(+)}(K_+)$ as the sum of two integrals taken in appropriate part of the k plane. Thus, for example, it may be expressed as $(2\pi i)^{-1} [\int_{III} + \int_{IV}] dk F_1^{(-)}$ with contours in the lower half-plane. This is a consequence of Eq. (14) which states that $G^{(-)}(-K_+) = G^{(+)}(K_+)$.

Returning to the representation in terms of the integrals in the right half-plane, it reads

$$G^{(+)}(K_+) = \frac{1}{2\pi i} \int_0^\infty dk \left[\frac{G^{(+)}}{k - K_+} - \frac{G^{(-)}}{k + K_+} \right] - \Sigma \operatorname{Res} \frac{G^{(+)}}{k - K_+}. \quad (21)$$

Notice that the integral part of Eq. (21) is not as yet in the form required by the eigenfunction expansion. However, it becomes

$$\frac{1}{2\pi i} \int_0^\infty \frac{dk\, k(G^{(+)} - G^{(-)})}{k^2 - K_+^2} + \frac{1}{2\pi i} \int_0^\infty \frac{dk\, K_+(G^{(+)} + G^{(-)})}{k^2 - K_+^2}, \quad (22)$$

the latter part of which is further reduced as follows: From the integration of $(2\pi i)^{-1} K_+ G^{(\pm)}(k)/(k^2 - K_+^2)$ in the upper or lower half-plane, it is found that

$$\pm \frac{1}{2\pi i} \int_{-\infty}^\infty \frac{dk\, K_+ G^{(\pm)}}{k^2 - K_+^2} = \pm \tfrac{1}{2} G^{(+)}(K_+) + K_+ \Sigma \operatorname{Res} \frac{G^{(\pm)}}{k^2 - K_+^2}, \quad (23\pm)$$

which in turn leads to

$$\frac{1}{2\pi i} \int_{-\infty}^\infty \frac{dk\, K_+(G^{(+)} - G^{(-)})}{k^2 - K_+^2} = G^{(+)}(K_+) + 2K_+ \Sigma \operatorname{Res} \frac{G^{(+)}}{k^2 - K_+^2}. \quad (24)$$

Then Eqs. (21), (22), and (24) result in an expression for $G^{(+)}(K_+)$ given by

$$\frac{1}{2\pi i} \int_0^\infty \frac{2\, dk\, k(G^{(+)} - G^{(-)})}{k^2 - K_+^2} + 2K_+ \times \Sigma \operatorname{Res} \frac{G^{(+)}}{k^2 - K_+^2} - 2 \Sigma \operatorname{Res} \frac{G^{(+)}}{k - K_+}. \quad (25)$$

The sum of the last two terms of Eq. (25) is seen, by the use of Eq. (21), to be equal to

$$-\tfrac{1}{2} \Sigma \operatorname{Res} \frac{2kG^{(+)}}{k^2 - K_+^2} - \tfrac{1}{2} \Sigma \operatorname{Res} \frac{2kG^{(-)}}{k^2 - K_+^2}. \quad (26)$$

It should be noticed here that Eq. (25) with the last two terms replaced by Eq. (26), manifests that the right-hand side of Eq. (17) for $G^{(+)}(K_+)$ is also equal to the integral of $(2\pi i)^{-1} F_2^{(\pm)} = (2\pi i)^{-1} 2kG^{(\pm)}(k)/(k^2 - K_+^2)$, taken in the upper or lower half-plane or the average of these.

Either from the considerations on the symmetry properties of the integral of Eq. (25) and the expression in Eq. (26), or directly from those on the appropriate choice of the integrand and the path, the expression for $G^{(+)}(K_+)$ is finally obtained as

$$G^{(+)}(K_+) = I_2^{(+)} + R_2^{(+)}, \quad (27)$$

$$I_2^{(+)} = \frac{1}{2\pi i} \left[\int_0^\infty \text{ or } \int_{-\infty}^0 \text{ or } \frac{1}{2} \int_{-\infty}^\infty \right] \times \frac{2\, dk\, k(G^{(+)} - G^{(-)})}{k^2 - K_+^2}, \quad (28+, 28-, \text{ or } 28)$$

$$R_2^{(+)} = -\Sigma \operatorname{Res} \frac{2kG^{(+)}}{k^2 - K_+^2} = -\Sigma \operatorname{Res} \frac{2kG^{(-)}}{k^2 - K_+^2} = (na)^{-2} \sum_{n,l,m} \frac{\phi_{nlm}(\mathbf{r}) \phi_{nlm}(\mathbf{r}')^*}{1 + (naK_+)^2}. \quad (29)$$

A closer study on the continuum part $I_2^{(+)}$ will be postponed to the next section. When this is done it will be shown that Eqs. (27), (28+), and (29) give the result identical with that obtained in A.

It is interesting to note that the eigenfunction form given by Eq. (27), in spite of its compact appearance, may not necessarily be the best form to deal with in the actual calculations. The reason for this lies in the very fact that the continuum part requires inherently the simultaneous consideration on the diverging and converging part of the integral of which only one or the other is desired in actual cases. This implies the necessity of appropriately closing the contours in the upper or lower half-plane according to which part of the eigenfunction is being considered. This unnecessary complication which may sometimes lead to some confusion can be avoided if other representations of $I_2^{(+)}$, though not in the eigenfunction form, are used. Such are the integral part obtained by integrating $(2\pi i)^{-1} F_2^{(\pm)}$ in the upper or lower half-plane:

$$I_2^{(+)} = \pm \frac{1}{2\pi i} \int_{-\infty}^\infty \frac{2\, dk\, kG^{(\pm)}}{k^2 - K_+^2}. \quad (30\pm)$$

Still another representation of the similar nature is

$$G^{(+)}(K_+) = I_2^{(+)'} + R_2^{(+)'}, \quad (31)$$

$$I_2^{(+)\prime} = \frac{1}{2\pi i}\int_0^{\infty} \frac{2\,dk\; k(G^{(+)} - G^{(-)})}{k^2 - K_+^2}, \qquad (32\pm)$$

$$R_2^{(+)\prime} = -\tfrac{1}{2}\,\Sigma\,\mathrm{Res}\,\frac{2kG^{(*)}}{k^2 - K_+^2}. \qquad (33\pm)$$

The combinations of Eqs. (32±) and (33±) may be obtained from the integrals of $(2\pi i)^{-1}F_2^{(*)}$ along the paths I and III, respectively.

It may be worthwhile to mention that the expressions corresponding to Eqs. (28−) and (28) may easily be left concealed if the variable λ is used such that only that corresponding to Eq. (28+) is straightforwardly obtained as in A.

Having gone through all the details, it is of some interest to generalize the procedure in the foregoing. Thus, the integral of $(2\pi i)^{-1}F_t^{(*)} = (2\pi i)^{-1}tk^{t-1}G^{(*)}(k)/(k^t - K_+^t)$, where t represents even integers, in the upper or lower half-plane, may be used to express $G^{(+)}(K_+)$. Here the reason for allowing only even values of t is related to the fact that the original differential equation for $G^{(+)}(K)$ involves K in the form of K^2. As a generalization of Eq. (27), it is easily obtained that

$$G^{(+)}(K_+) = I_t^{(+)} + R_t^{(+)} \quad (t = 2, 4, 6, \cdots), \qquad (34)$$

$$I_t^{(+)} = \frac{1}{2\pi i}\left[\int\right]\frac{t\,dk\; k^{t-1}(G^{(+)} - G^{(-)})}{k^t - K_+^t}, \qquad (35)$$

$$R_t^{(+)} = -\,\Sigma\,\mathrm{Res}\,\frac{tk^{t-1}G^{(+)}}{k^t - K_+^t} = -\,\Sigma\,\mathrm{Res}\,\frac{tk^{t-1}G^{(-)}}{k^t - K_+^t}$$

$$= \frac{t}{2}(na)^{-2}\sum_{n,l,m}\frac{\phi_{nlm}(\mathbf{r})\phi_{nlm}(\mathbf{r}')^*}{1 + (-)^{t/2+1}(naK_+)^t}. \qquad (36)$$

In Eq. (35) $[\int]$ stands for either choice of $\int_0^{\infty}$ or $\int_{-\infty}^{0}$ or $\frac{1}{2}\int_{-\infty}^{\infty}$. It may be added that the result for $G^{(+)}(K_+)$ obtained by integrating $(2\pi i)^{-1}F_p^{(*)} = (2\pi i)^{-1}pk^{p-1}G^{(*)}(k)/(k^p \mp K_+^p)$, $p = 1, 3, 5, \cdots$, is also represented by Eqs. (34)–(36) with $t = 2p$. A special example is the case of $p = 1$ which has been discussed in detail.

Finally, a word is added regarding the treatment of $G^{(-)}(K_-)$. It is easily found that the results for this case is given by Eq. (27) or Eq. (34) in which the replacements $(+) \to (-)$ and $K_+ \to K_- = K - i\epsilon'$ are made. On the right-hand sides of Eqs. (28±), (28), (29), (35), and (36), only $K_+ \to K_-$ is required. The results can also be obtained from expressions for $G^{(+)}(K_+)$ by noticing the relation $G^{(-)}(K_-) = G^{(+)}(-K_-)$.

IV. CONTINUUM PART

In this section, only the case of $t = 2$ for $I_t^{(+)}$ is discussed in detail, since it is evidently sufficient to do this and no more.

It is to be noticed that Eq. (28+) for $I_2^{(+)}$ is equivalent to Eq. (16) for $I_G^{(+)}$ of A when the identity $\mathcal{G}^{(+)}(\lambda e^{2\pi i}) = \mathcal{G}^{(-)}(\lambda)$ is utilized. This identity signifies that the value of $\mathcal{G}^{(+)}$ on the first Riemann sheet below the branch cut is equal to that of $\mathcal{G}^{(-)}$ on the second sheet above the cut in the limit, when the both points considered approach a point on the real axis. This association is in fact equivalently contained in the transposition of the contour II to IV by $k \to -k$ which results in the same range of k-integrations for the $G^{(+)}(k)$ and $G^{(-)}(k)$ parts in the form as given by Eq. (28+). Obviously a similar relation exists corresponding to the transposition of I to III.

With the above as a general remark, consider Eq. (28+). When the orthogonality of the spherical harmonics is taken into account to express $P_l(\cos\theta_{rr'})$ as an integral over the angles of the $\mathbf{k}$ space, it is found from Eqs. (5b) and (15) that

$$I_2^{(+)} = \sum_{l=0}^{\infty}\sum_{m=-l}^{+l}\int\frac{d\mathbf{k}\;\phi_{klm}(\mathbf{r})\phi_{klm}(\mathbf{r}')^*}{K_+^2 - k^2}, \qquad (37)$$

where

$$\phi_{klm}(\mathbf{r}) = (2/\pi)^{\frac{1}{2}}L_l(r, k)e^{i\eta(l,k)}Y_{lm}(\hat{r})Y_{lm}(\hat{k})^*, \qquad (38)$$

in which the unit vector $\hat{a} = \mathbf{a}/a$ stands for the angles of the polar coordinates of a vector $\mathbf{a}$. In the following, ϕ_{klm} are also called $\phi_{klm}^{(+)}$ in contrast to $\phi_{klm}^{(-)}$ defined by the right-hand side of Eq. (38), with $\eta(l, k)$ being replaced by $\eta(l, -k) = -\eta(l, k)$. As has been mentioned earlier, Eq. (37) together with Eq. (29) gives the conventional representation of the eigenfunction form of $G^{(+)}(K_+)$. In addition to this, similar representations exist which arise from Eqs. (28−) and (28). However, some care is needed in having the explicit forms for these representations related to the fact that there is involved an integral along the negative real axis. Therefore, they are left untouched until the similar situation is discussed presently when it becomes evident how to express them.

Although the conventional expression for $I_2^{(+)}$ has already been obtained in the above, it is sometimes of certain advantage to use a slightly different expression which is to be found below. With this in mind, notice that for real and positive k, $G^{(*)}(k)$ can be written in the form

$$G^{(*)}(k) = -\frac{k}{(4\pi)^2}\int d\omega_k$$
$$\times\,\Sigma\, i^l(2l+1)H_l^{(*)}(r, k)e^{i\eta(l,k)}P_l(\cos\theta_{rk})$$
$$\times\,[\Sigma\, i^{l'}(2l'+1)L_{l'}(r', k)e^{i\eta(l',k)}P_{l'}(\cos\theta_{r'k})]^*$$
$$= \pm\frac{\pi k}{2i}\int d\omega_k\psi_{1,2}^{(+)}(\mathbf{r}, \mathbf{k})\psi^{(+)}(\mathbf{r}', \mathbf{k})^*. \qquad (39)$$

In Eq. (39),

$$\psi^{(+)}(\mathbf{r},\mathbf{k}) = \sum_{l,m} i^l \phi^{(+)}_{\mathbf{k}lm}(\mathbf{r}), \tag{40}$$

$$= (2\pi)^{-\frac{3}{2}} e^{\pi/2ka}\Gamma(1 - i(ka)^{-1}) e^{i\mathbf{k}\cdot\mathbf{r}} \times F[i(ka)^{-1}, 1; i(kr - \mathbf{k}\cdot\mathbf{r})], \tag{41}$$

in which Gordon's identity[10] is used. Furthermore, $\psi^{(+)}_{1,2}$ corresponds to the $W_{1,2}$ part of L_l such that $\psi^{(+)} = \frac{1}{2}(\psi^{(+)}_1 + \psi^{(+)}_2)$. When $\psi^{(+)}(\mathbf{r},\mathbf{k})$ needs to be considered in the following as a function of k and $\theta = \theta_{rk}$, it shall be called $\psi^{(+)}(\mathbf{r}; k, \theta)$, and similarly for $\psi^{(+)}_{1,2}(\mathbf{r},\mathbf{k})$.

Clearly, $\psi^{(+)}$ in Eq. (39) can be replaced by

$$\psi^{(-)}(\mathbf{r},\mathbf{k}) = \sum_{l,m} i^l \phi^{(-)}_{\mathbf{k}lm}(\mathbf{r}) = (2\pi)^{-\frac{3}{2}} e^{\pi/2ka}\Gamma(1 + i(ka)^{-1}) e^{i\mathbf{k}\cdot\mathbf{r}} \times F[-i(ka)^{-1}, 1; -i(kr + \mathbf{k}\cdot\mathbf{r})]. \tag{42}$$

As is given in the Appendix, the partial wave expansions of $\psi^{(+)}$ and $\psi^{(-)}$ are very nearly the same, implying a close relationship between them. The decomposition of $\psi^{(-)}$, for k = real > 0, into $\psi^{(-)}_{1,2}$ can be carried out in the same way.

By substituting the expression for $G^{(+)} - G^{(-)}$ obtained from Eq. (39), or a similar one with $\psi^{(-)}$ and $\psi^{(-)}_{1,2}$ into Eq. (28+), it is seen, with $\theta' = \theta_{r'k}$, that

$$I_2^{(+)} = \int_0^\infty \frac{dk\, k^2}{K_+^2 - k^2} \times \int d\omega_k\, \psi^{(S)}(\mathbf{r}; k, \theta)\psi^{(S)}(\mathbf{r}'; k, \theta')^*, \tag{43+}$$

where (S) denotes either choice of the signs ($\pm$).

This equation provides an alternative form of the continuum part, since $\psi^{(+)}$ or $\psi^{(-)}$ constitute a complete set of eigenfunctions (together with the discrete eigenfunctions) just as well as the conventional $\phi_{\mathbf{k}lm}$ do.[11] In fact, this form may very well be considered simpler than the conventional expression.

To get similar expression for Eq. (28−), it is only necessary to remember that the range of the integration is the negative real axis. Thus, for example, noting $G^{(+)}(k) - G^{(-)}(k) = -(G^{(+)}(-k) - G^{(-)}(-k))$ from Eq. (14), it is seen that

$$I_2^{(+)} = \int_{-\infty}^0 \frac{dk\, k^2}{K_+^2 - k^2} \times \int d\omega_k\, \psi^{(S)}(\mathbf{r}; -k, \theta)\psi^{(S)}(\mathbf{r}'; -k, \theta')^*. \tag{43−}$$

[10] See Gordon's paper in reference 4.

[11] A. Sommerfeld, *Atombau und Spektrallinien* (Frederick Vieweg und Sohn, Braunschweig, Germany, 1939), Vol. 2, p. 457.

The use of Eq. (A9) gives rise to

$$\int d\omega_k\, \psi^{(S)}(\mathbf{r}; -k, \theta)\psi^{(S)}(\mathbf{r}'; -k, \theta')^* = \int d\omega_k\, e^{-2\pi/ka}\psi^{(S)}(\mathbf{r}; k, \theta)\psi^{(S)}(\mathbf{r}'; k, \theta')^*, \tag{44}$$

so that the right-hand side of Eq. (43−) may also be expressed with that of Eq. (44). Similarly the product $\psi^{(S)}\psi^{(S)*}$ of Eq. (43+) can be replaced by $e^{+2\pi/ka}\psi^{(S)}(\mathbf{r}; -k, \theta)\psi^{(S)}(\mathbf{r}'; -k, \theta')^*$. Of course these representations can be obtained from Eq. (39) and a similar one with $\psi^{(-)}$ functions by noticing

$$\tfrac{1}{2}[\psi_1^{(S)}(\mathbf{r}; k, \theta) + \psi_2^{(S)}(\mathbf{r}; k, \theta)] = \begin{Bmatrix} 1 \\ e^{-2\pi/ka} \end{Bmatrix}\psi^{(S)}(\mathbf{r}; k, \theta) \quad \text{for} \quad \begin{cases} k > 0 \\ k < 0, \end{cases}$$

in connection with Eqs. (7) and (11).

The expression corresponding to Eq. (28) is simply the average of the right-hand sides of Eqs. (43$\pm$).

Furthermore, Eqs. (30$\pm$) lead to the straightforward representations

$$I_2^{(+)} = \int_{-\infty}^\infty \frac{dk\, k^2}{K_+^2 - k^2} \times \int d\omega_k\, \tfrac{1}{2}\psi^{(S)}_{1,2}(\mathbf{r}; k, \theta)\psi^{(S)}(\mathbf{r}'; k, \theta')^*. \tag{45$\pm$}$$

It must be emphasized that contained in Eqs. (30$\pm$) and hence in Eqs. (45$\pm$), are the contours that are to be taken in the upper or lower half-plane according to the choice of ψ_1 or ψ_2 in the integrand of Eqs. (45$\pm$). If the relations

$$\int d\omega_k\, \psi^{(S)}_{1,2}(\mathbf{r}; -k, \theta)\psi^{(S)}(\mathbf{r}'; -k, \theta')^* = \int d\omega_k\, \psi^{(-S)}_{2,1}(\mathbf{r}; k, \theta)\psi^{(-S)}(\mathbf{r}'; k, \theta')^* = \int d\omega_k\, \psi^{(S)}_{2,1}(\mathbf{r}; k, \theta)\psi^{(S)}(\mathbf{r}'; k, \theta')^*$$

are used in Eqs. (45$\pm$), they reduce to Eqs. (43$\pm$) as should be the case.

V. CONCLUDING REMARKS

It has been shown that the most general eigenfunction form for $G^{(+)}(K_+)$ can be given by Eq. (34) in which $I_l^{(+)}$ may be written from Eqs. (43$\pm$) as

$$\pm\frac{l}{2}\int_0^{\pm\infty} \frac{dk\, k^l}{K_+^l - k^l} \times \int d\omega_k\, \psi^{(S)}(\mathbf{r}; \pm k, \theta)\psi^{(S)}(\mathbf{r}'; \pm k, \theta')^*. \tag{46$\pm$}$$

The expressions corresponding to Eqs. (30±) or (32±) are self-evident. Admittedly, these forms may not be of special value except possibly for large values of t where the sum over the discrete part may converge rather rapidly when naK is large.

Perhaps the most important thing to bear in mind when looking at Eq. (34) is to remember the combination of contour integrals through which it is derived. Thus, for example, built into the $\int_0^\infty dk$ form of Eq. (35) is the fact that the diverging and converging part of the integrand demand in connection with $R_l^{(+)}$, the contours I and IV, respectively. The vanishing integral along IV of $G^{(-)}$ may be subtracted from the entire expression [Eq. (34)], leaving only the integral of $G^{(+)}$ along I similar to that given by Eq. (32+) plus (33+). As is already clear, the integral along IV is required only to complete the combination $G^{(+)} - G^{(-)}$ needed for the continuum part.

Another point to make is that the symmetry of the Green's function provides the possibility of decomposing either ψ or ψ^* of Eqs. (46±) into $\psi_{1,2}$ or $\psi^*_{1,2}$.[12]

As a simple exercise to see how Eqs. (43±) and (45±) are to be used, one may take the case of a free particle. This corresponds to the limit $Z \to 0$ so that $\psi^{(+)} = \psi^{(-)} = (2\pi)^{-\frac{3}{2}} e^{i\mathbf{k}\cdot\mathbf{r}}$. Then from the Rayleigh expansion, ψ_1 and ψ_2 are found explicitly. It is a simple matter to see that all of these produce, when treated properly, the result

$$-(4\pi)^{-1}e^{iK_+|\mathbf{r}-\mathbf{r}'|}/|\mathbf{r}-\mathbf{r}'|.$$

It is interesting to note that the correct result is obtainable from Eqs. (43±) by forgetting about the rules of closing the contours and merely evaluating

$$\pm(2\pi)^{-3}\int_0^{+\infty}\frac{dk\,k^2}{K_+^2-k^2}\begin{Bmatrix}\text{D.P.}\\ \text{C.P.}\end{Bmatrix}\int d\omega_k\, e^{i\mathbf{k}\cdot\mathbf{r}}e^{-i\mathbf{k}\cdot\mathbf{r}'}$$

with the contour I or IV alone being associated with the k-integration. In these expressions, D. P. and C. P. of $\Psi(\mathbf{r})$, standing for the diverging and converging part of a function Ψ, designates those parts which satisfy the Ausstrahlungs- and Einstrahlungsbedingung, respectively[13]:

$$\lim_{r\to\infty} r\left(\frac{d\Psi}{dr} \mp ik\Psi\right) = 0.$$

In finding C. P. of the latter, it is to be assumed temporarily that k is a positive number, disregarding the actual sign of k in the subsequent integration.

Finally, it should be remarked that the formal procedure leading to Eqs. (34)–(36) is expected to be applicable to the case of more general potential $U(r)$ provided that the radial equation possesses bounded solutions $L_l(r, K)$ whose diverging and converging part are called $H_l^{(*)}(r, K)$.[14] Then the relation $G^{(*)}(-K) = G^{(*)}(K)$ is still expected to hold. This in turn leads to the validity of Eq. (5b) because of the fact that $G^{(+)} - G^{(-)}$ satisfies the homogeneous differential equation and that $G^{(*)}$ are symmetric with respect to $\mathbf{r}$ and $\mathbf{r}'$. Therefore, the integral of the form of Eq. (35) and the remaining sum over the residues are to give rise to the continuum and discrete part, respectively, of $G^{(+)}(K_+)$.

ACKNOWLEDGMENT

The author is grateful to R. A. Mapleton for helpful conversations.

APPENDIX

The wavefunction of a particle in the positive energy state in a Coulomb field which behaves asymptotically like a planewave plus a diverging spherical wave was obtained by Gordon.[10] It may be represented by

$$\bar{\psi}^{(+)}(\mathbf{r},\mathbf{k}) = \Sigma\, i^l(2l+1)L_l(r,k)e^{i\eta(l,k)}P_l(\cos\theta_{rk})$$
$$= e^{\pi/2ka}\Gamma(1+n)e^{i\mathbf{k}\cdot\mathbf{r}}F(-n,1;i(kr-\mathbf{k}\cdot\mathbf{r})), \quad \text{(A1)}$$

where $n = (ika)^{-1}$. On the other hand, Sommerfeld[15] used along with $\bar{\psi}^{(+)}$, a similar function $\bar{\psi}^{(-)}$ with a converging instead of diverging wave, defined by

$$\bar{\psi}^{(-)}(\mathbf{r},\mathbf{k}) = \bar{\psi}^{(+)}(\mathbf{r},-\mathbf{k})^*. \quad \text{(A2)}$$

The partial wave expansion of $\bar{\psi}^{(-)}$ will be found here from

$$e^{\pi/2ka}\Gamma(1-n)e^{i\mathbf{k}\cdot\mathbf{r}}F(n,1;-i(kr+\mathbf{k}\cdot\mathbf{r})) \quad \text{(A3)}$$

given by Eqs. (A1) and (A2). The procedure used is seen, when applied backward, to give an alternative proof of the Gordon's relation. Consider then $F = F(1-n, 1; i(kr+\mathbf{k}\cdot\mathbf{r}))$ by calling $\rho = +2ikr$ and $\mu = \cos\theta_{rk}$. It may be given by

$$F = \sum_{t=0}^{\infty}\frac{(1-n)_t}{(1)_t t!}\left(\frac{1+\mu}{2}\rho\right)^t,$$

where $(a)_t = a(a+1)\cdots(a+t-1) = \Gamma(a+t)/\Gamma(a)$. The use of an identity

[12] See, for example, the treatment in the first paper in reference 5.

[13] A. Sommerfeld, Ann. Physik 11, 257 (1931), Appendix.

[14] See p. 112 of the book cited in reference 2.

[15] See reference 11. The functions $\bar{\psi}^{(+)}$ and $\bar{\psi}^{(-)}$ are known to be adequate for representing the particle in the initial and final state, respectively. Therefore, it is natural, and in fact simpler, to choose (+) or (−) for (S) in Eq. (43 +), say, when the diverging or converging solution is desired.

$$\frac{1}{(t!)^2}\left(\frac{1 \pm \mu}{2}\right)^t = \sum_{l=0}^{t} \frac{(2l+1)P_l(\pm\mu)}{(t-l)!(t+l+1)!} \tag{A4}$$

that can be proved easily leads to (when $t = l + \nu$)

$$\begin{aligned} F &= \sum_{l=0}^{\infty} (2l+1)\rho^l P_l(\mu) \\ &\quad \times \sum_{\nu=0}^{\infty} \frac{\Gamma(1-n+l+\nu)}{\Gamma(1-n)} \frac{\rho^\nu}{\nu!\Gamma(2l+\nu+2)} \\ &= \sum_{l=0}^{\infty} i^l(2l+1)(-i\rho)^l P_l(\mu) \frac{\Gamma(1-n+l)}{\Gamma(1-n)\Gamma(2l+2)} \\ &\quad \times \sum_{\nu=0}^{\infty} \frac{(1-n+l)_\nu \rho^\nu}{(2l+2)_\nu \nu!} \\ &= \Sigma\, i^l(2l+1)(-i\rho)^l P_l(\mu) \frac{\Gamma(1-n+l)}{\Gamma(1-n)\Gamma(2l+2)} \\ &\quad \times F(1-n+l, 2l+2; \rho). \end{aligned} \tag{A5}$$

From Eqs. (10), (A3), and (A5), it is seen that

$$\begin{aligned} \tilde{\psi}^{(-)}(\mathbf{r}, \mathbf{k}) &= e^{\pi/2ka}\, \Sigma\, i^l(2l+1) \frac{\Gamma(1-n+l)}{(2l+1)!} (2kr)^l \\ &\quad \times e^{ikr} F(1+n+l, 2l+2; -2ikr) P_l(\cos\theta_{rk}) \end{aligned} \tag{A6}$$

$$= (2\pi)^{\frac{3}{2}} \sum_{l,m} i^l \phi_{\mathbf{k}lm}(\mathbf{r}). \tag{A7}$$

As is well known, the difference between the partial wave expansions of $\tilde{\psi}^{(+)}$ and $\tilde{\psi}^{(-)}$ consists of the signs of the phase, $\pm i\eta(l, k)$. This fact may be used to obtain

$$\int d\omega_k\, \psi^{(+)}(\mathbf{r}, \mathbf{k}) \psi^{(+)}(\mathbf{r}', \mathbf{k})^* = \int d\omega_k\, \psi^{(-)}(\mathbf{r}, \mathbf{k}) \psi^{(-)}(\mathbf{r}', \mathbf{k})^*. \tag{A8}$$

Some of the useful relations between $\psi^{(+)}$ and $\psi^{(-)}$ may be derived by combining the basic ones:

$$\left.\begin{aligned} \psi^{(\mp)}(\mathbf{r}; k, \theta)^* &= \psi^{(\mp)}(\mathbf{r}; k, \pi - \theta) = \psi^{(\mp)}(-\mathbf{r}, \mathbf{k}) \\ \psi^{(\pm)}(\mathbf{r}; -k, \theta) &= e^{-\pi/ka} \psi^{(\mp)}(\mathbf{r}; k, \pi - \theta) \end{aligned}\right\}. \tag{A9}$$

Some Analytic Properties of Green Functions and Self-Energy Parts for a System of Interacting Bosons

W. E. PARRY* AND R. E. TURNER†
The Clarendon Laboratory, Oxford, England
(Received 13 November 1962)

Some analytic properties of the single-particle Green functions and proper self-energy parts for a system of interacting bosons at nonzero temperatures are derived. In particular, it is shown how to obtain the analytical continuations of simple functions of the proper self-energy parts, which can be obtained from perturbation theory at an infinite set of points along the imaginary energy axis, to the whole of the complex energy plane. This is useful in determining the poles of the Green functions.

INTRODUCTION

THE single-particle Green functions for a condensed system of interacting bosons can be written down in terms of the proper self-energy parts (see below), namely, the Σ's. These can be evaluated using finite-temperature perturbation theory at an infinite set of points along the imaginary axis in the complex energy plane. The points are at energies $2\pi ni/\beta$, where n is an integer and $\beta = (kT)^{-1}$, k being Boltzmann's constant and T the absolute temperature. We are often interested, however, in the values of the Green functions along the real axis, since the poles of the Green function, for real E, give us the single-particle energies. Unfortunately, these functions are rather complicated functions of the proper self-energy parts, and it is not always easy to obtain the correct analytic continuation of them onto the real axis. This is largely because factors of the form $e^{r\beta E}$ where r is an integer and E the (complex) energy are equal to unity at all the points at which the Σ's have been evaluated, and hence one cannot decide immediately where to include such factors. This paper is largely an attempt to make such decisions easier, and we shall derive some analytic properties of the Green functions and proper self-energy parts on the way. Similar properties for a fermion system have been derived by Luttinger.[1]

In Sec. 1 we briefly review some properties of a function of a complex variable which we shall require. In Sec. 2 we consider the single-particle Green functions, and the last section is devoted to the proper self-energy parts.

1. REQUIRED PROPERTIES OF COMPLEX-VARIABLE FUNCTION

Let $f(z)$, where $z = x + iy$, be a function of a complex variable with the following properties:

(i) $f(z)$ is analytic everywhere, except possibly on the real axis;

(ii) $f(z) \to 0$ as $z \to \infty$ along any straight line in the upper or lower half-plane;

(iii) $f(z)^* = f(z^*)$, which restricts f to real functions.

In general, $f(z)$ will be discontinuous across the real axis. We define:

$$\lim_{\epsilon \to +0} f(x + i\epsilon) = u(x) - iv(x), \tag{1}$$

so that

$$\lim_{\epsilon \to +0} [f(x - i\epsilon) - f(x + i\epsilon)] = 2iv(x), \tag{2}$$

using property (iii). Then it may easily be shown that

$$f(z) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{v(x)}{z - x}\, dx, \tag{3}$$

and, taking real and imaginary parts, we obtain the dispersion relation

$$u(x) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{v(x')}{x - x'}\, dx', \tag{4}$$

where P indicates that the principal part of the integral is to be taken. Finally, using Carleman's theorem, Baym and Mermin[2] have shown that if a function is determined at the infinite set of points $z = 2\pi in/\beta$ (n integral), then properties (i) and (ii) uniquely define an analytic continuation of the function into the whole of the upper and lower half-planes.

---

* At present on leave of absence at University of California, San Diego, California.
† Present address; Department of Physics, The University of Sussex, Brighton, Sussex, England.

[1] J. M. Luttinger, Phys. Rev. 121, 942 (1961).

[2] G. Baym and N. D. Mermin, J. Math. Phys. 2, 232 (1961).

2. SINGLE-PARTICLE GREEN FUNCTIONS

To simplify our calculations, we define our single-boson Green function in the complex E plane by:

$$\mathcal{G}(E,\mathbf{k}) = \frac{1}{Q}\sum_{\nu\gamma} e^{-\beta E_\nu}\left[\frac{\langle\nu|\, a_{\mathbf{k}}\,|\gamma\rangle\langle\gamma|\, a_{\mathbf{k}}^{+}\,|\nu\rangle}{E - E_\gamma + E_\nu} - \frac{\langle\nu|\, a_{\mathbf{k}}^{+}\,|\gamma\rangle\langle\gamma|\, a_{\mathbf{k}}\,|\nu\rangle}{E - E_\nu + E_\gamma}\right], \tag{5}$$

$$Q = \sum_{\nu} e^{-\beta E_\nu}, \tag{6}$$

where $a_{\mathbf{k}}$ and $a_{\mathbf{k}}^{+}$ are the usual creation and annihilation operators for particles with momentum $\mathbf{k}$, and E_ν and $|\nu\rangle$ the eigenvalues and eigenvectors of the generalized Hamiltonian:

$$(H - \mu N)\,|\nu\rangle = E_\nu\,|\nu\rangle. \tag{7}$$

N is the number operators $\Sigma_{\mathbf{k}}\, a_{\mathbf{k}}^{+}a_{\mathbf{k}}$, and μ the chemical potential. If E is $\omega + i\epsilon$, where ω is real and ϵ is vanishingly small, the Fourier transform of Eq. (5) with respect to ω gives the retarded time–temperature-dependent Green function.[3] If $E = -2\pi n i/\beta$ where n is integral, Eq. (5) gives a coefficient of the Fourier series expansion of the causal temperature-dependent Green function.[4] Since the sum in Eq. (5) is a grand canonical average, we assume that it is uniformly convergent.

We note that

$$\langle\nu|\, a_{\mathbf{k}}\,|\gamma\rangle\langle\gamma|\, a_{\mathbf{k}}^{+}\,|\nu\rangle = |\langle\nu|\, a_{\mathbf{k}}\,|\gamma\rangle|^2, \tag{8}$$

and deduce that $\mathcal{G}(E,\mathbf{k})$ is a real function of E. Further,

$$\lim_{|E|\to\infty} \mathcal{G}(E,\mathbf{k}) = \frac{1}{E}. \tag{9}$$

Hence $\mathcal{G}(E,\mathbf{k})$ possesses all the properties (i), (ii), and (iii) of Sec. 1, and the facts listed there lead to standard results for Green functions.

Equation (4) gives us the usual dispersion relation. From a knowledge of the Fourier coefficients $\mathcal{G}(-2\pi n i/\beta,\mathbf{k})$, we can uniquely deduce an analytic continuation $\mathcal{G}(E)$. Writing $\mathcal{G}$ in the Lehmann representation

$$\mathcal{G}(E,\mathbf{k}) = \int_{-\infty}^{\infty} \frac{\rho_{\mathbf{k}}(\omega)}{E - \omega}\, d\omega, \tag{10}$$

we find, using Eq. (3) and the relation

$$\delta(x) = \lim_{\epsilon\to+0} \frac{1}{2\pi i}\left[\frac{1}{x - i\epsilon} - \frac{1}{x + i\epsilon}\right], \tag{11}$$

$$\rho_{\mathbf{k}}(\omega) = Q^{-1}\sum_{\nu\gamma} e^{-\beta E_\nu}[|\langle\nu|\, a_{\mathbf{k}}\,|\gamma\rangle|^2\,\delta(\omega - E_\gamma + E_\nu) - |\langle\gamma|\, a_{\mathbf{k}}\,|\nu\rangle|^2\,\delta(\omega - E_\nu + E_\gamma)]$$
$$= Q^{-1}(e^{\beta\omega} - 1) \times \sum_{\nu,\gamma} e^{-\beta E_\nu}\,|\langle\gamma|\, a_{\mathbf{k}}\,|\nu\rangle|^2\,\delta(\omega - E_\nu + E_\gamma). \tag{12}$$

One further property of $\mathcal{G}(E)$ will be required. If we put $E = \omega + iy$, then

$$\mathcal{G}(E,\mathbf{k}) = \int_{-\infty}^{\infty} \frac{\rho_{\mathbf{k}}(\omega')}{(\omega' - \omega)^2 + y^2}\,(\omega' - \omega)\, d\omega' + iy\int_{-\infty}^{\infty} \frac{\rho_{\mathbf{k}}(\omega')\, d\omega'}{(\omega' - \omega)^2 + y^2}. \tag{13}$$

Suppose that the imaginary part of $\mathcal{G}(E_1,\mathbf{k})$ vanishes ($y_1 \neq 0$); then

$$\mathcal{G}(E_1,\mathbf{k}) = \int_{-\infty}^{\infty} \frac{\omega'\rho_{\mathbf{k}}(\omega')\, d\omega'}{(\omega' - \omega_1)^2 + y_1^2}. \tag{14}$$

Now since the terms under the summation sign in Eq. (12) are always positive, we see that $\rho_{\mathbf{k}}(\omega')$ has the sign of ω'. Hence the integrand in Eq. (14) is never negative, and since ρ is not everywhere zero, we deduce that $\mathcal{G} > 0$. Hence we have

$$\mathcal{G}(E,\mathbf{k}) \neq 0, \qquad \operatorname{Im} E \neq 0. \tag{15}$$

For the case of a degenerate boson gas, one must consider another single-particle Green function.[5,6,7] In the complex E plane it has the representation

$$\tilde{\mathcal{G}}(E,\mathbf{k}) = Q^{-1}\sum_{\nu,\gamma} e^{-\beta E_\nu}\left[\frac{\langle\nu|\, a_{\mathbf{k}}\,|\gamma\rangle\langle\gamma|\, a_{-\mathbf{k}}\,|\nu\rangle}{E - E_\gamma + E_\nu} - \frac{\langle\nu|\, a_{\mathbf{k}}\,|\gamma\rangle\langle\gamma|\, a_{-\mathbf{k}}\,|\nu\rangle}{E - E_\nu + E_\gamma}\right]. \tag{16}$$

Again, since this is a grand canonical average, we assume that the sum converges uniformly. Further, as $|E| \to \infty$,

$$\tilde{\mathcal{G}}(E,\mathbf{k}) \sim 1/E^2.$$

Thus a unique analytic continuation can be obtained from the values at the usual infinite set of points on the imaginary axis.

To show that $\tilde{\mathcal{G}}(E)$ is a real function of E, we have to show that the product of matrix elements $\langle\nu\,|a_{\mathbf{k}}|\,\gamma\rangle\langle\gamma\,|a_{-\mathbf{k}}|\,\nu\rangle$ is real. We introduce a complete set of free-particle states $|\{n\}\rangle$ where $\{n\}$ represents a set of occupation numbers of free-particle levels.

[3] D. N. Zubarev, Usp. Fiz. Nauk. **71**, 71 (1960) [English translation;] Sov. Phys. Usp. **3**, 320 (1960).
[4] T. Matsubara, Progr. Theoret. Phys. **14**, 351 (1955).
[5] N. M. Hugenholtz and D. Pines, Phys. Rev. **116**, 489 (1959).
[6] S. T. Beliaev, Zh. Eksperim. i. Teor. Fiz. **34**, 417 (1958) [English translation;] Sov. Phys.-JETP **6**, 289 (1958).
[7] W. E. Parry and R. E. Turner, Phys. Rev. **128**, 929 (1962).

Then:

$$\langle\nu|\, a_{\mathbf{k}}\, |\gamma\rangle\langle\gamma|\, a_{-\mathbf{k}}\, |\nu\rangle = \sum_{\substack{\{n\}\{n_1\} \\ \{n_2\}\{n_3\}}} \langle\nu \mid \{n\}\rangle\langle\{n\}|\, a_{\mathbf{k}}\, |\{n_1\}\rangle\langle\{n_1\} \mid \gamma\rangle\langle\gamma \mid \{n_2\}\rangle \times \langle\{n_2\}|\, a_{-\mathbf{k}}\, |\{n_3\}\rangle\langle\{n_3\} \mid \nu\rangle. \quad (17)$$

Since the matrix elements of $a_{\mathbf{k}}$ between free-particle states are real, the problem is reduced to showing that products of the form $\langle\nu \mid \{n\}\rangle\langle\{n_3\} \mid \nu\rangle$ are real. To do this, we split the generalized Hamiltonian of Eq. (7) into two parts:

$$H - \mu N = H_0 + H_1, \quad (18)$$

where H_0 is diagonal in the free-particle representation, with eigenvalues ϵ_n corresponding to the eigenstate $|\{n\}\rangle$. Expanding the true eigenstates $|\nu\rangle$ in terms of the free-particle states, substituting in (7), using (18), and taking the product of the resulting equation with an arbitrary free state $\langle\{m\}|$, we have

$$(E_\nu - \epsilon_m)\langle\{m\} \mid \nu\rangle = \sum_{m_1} \langle\{m\}|\, H_1\, |\{m_1\}\rangle\langle\{m_1\} \mid \nu\rangle. \quad (19)$$

This represents a set of simultaneous, linear, homogeneous equations for the $\langle\{m\} \mid \nu\rangle$. The coefficients in these equations are all real, since the matrix elements of H_1 between free-particle states are real. The $\langle\{m\} \mid \nu\rangle$ are determined except for an arbitrary constant, which is to be evaluated by the condition

$$\langle\nu \mid \nu\rangle = \sum_{\{m\}} \langle\nu \mid \{m\}\rangle\langle\{m\} \mid \nu\rangle = 1. \quad (20)$$

This condition does not completely determine the constant, for the introduction of an arbitrary phase factor into the $\langle\{m\} \mid \nu\rangle$ will still satisfy Eq. (20). It is, of course, the same phase factor for all the $\{m\}$. Bearing this in mind, we can write:

$$\langle\{m\} \mid \nu\rangle = \alpha_{m\nu} e^{i\phi_\nu}, \quad (21)$$

where $\alpha_{\nu m}$ is real, on account of the reality of the coefficients in Eqs. (19) and (20), and ϕ_ν is an arbitrary constant. But since

$$\langle\nu \mid \{m\}\rangle = \langle\{m\} \mid \nu\rangle^* = \alpha_{m\nu} e^{-i\phi_\nu}, \quad (22)$$

the reality of Eq. (17) follows immediately. Hence $\tilde{\mathcal{G}}(E, \mathbf{k})$ is a real function of E.

Thus, writing

$$\tilde{\mathcal{G}}(E, \mathbf{k}) = \int_{-\infty}^{\infty} \frac{\tilde{\rho}_{\mathbf{k}}(\omega)\, d\omega}{E - \omega}, \quad (23)$$

we have

$$\tilde{\rho}_{\mathbf{k}}(\omega) = Q^{-1}(e^{\beta\omega} - 1) \times \sum_{\nu,\gamma} \langle\nu|\, a_{\mathbf{k}}\, |\gamma\rangle\langle\gamma|\, a_{-\mathbf{k}}\, |\nu\rangle \delta(\omega - E_\gamma + E_\nu), \quad (24)$$

where we have again used Eqs. (3) and (11).

3. PROPER SELF-ENERGY PARTS

We introduce two self-energy parts in a purely formal manner by writing equations similar to the Dyson equations in electrodynamics:

$$\mathcal{G}(\mathbf{k}, E) = \mathcal{G}_0(\mathbf{k}, E) + \mathcal{G}(\mathbf{k}, E)\Sigma_{11}^{(\beta)}(\mathbf{k}, E)\mathcal{G}_0(\mathbf{k}, E) + \tilde{\mathcal{G}}(\mathbf{k}, E)\Sigma_{02}^{(\beta)}(\mathbf{k}, E)\mathcal{G}_0(\mathbf{k}, E), \quad (25)$$

$$\tilde{\mathcal{G}}(\mathbf{k}, E) = \tilde{\mathcal{G}}(\mathbf{k}, E)\Sigma_{11}^{(\beta)}(\mathbf{k}, -E)\mathcal{G}_0(\mathbf{k}, -E) + \mathcal{G}(\mathbf{k}, E)\Sigma_{02}^{(\beta)}(\mathbf{k}, -E)\mathcal{G}_0(\mathbf{k}, -E), \quad (26)$$

where $\mathcal{G}_0(\mathbf{k}, E)$ is the unperturbed single-particle Green function

$$\mathcal{G}_0(\mathbf{k}, E) = 1/2\pi(E - \eta_{\mathbf{k}}), \quad (27)$$

$$\eta_{\mathbf{k}} = \hbar^2 k^2/2m - \mu. \quad (28)$$

When $E = i\omega_n$, the Σ's play a similar role in thermodynamic perturbation theory[7] to the self-energy parts of the ground state problem.[5,6]

Using the fact that $\tilde{\mathcal{G}}(\mathbf{k}, E)$ is an even function of E [cf. Eq. (16)], we can solve Eqs. (25) and (26) for the Σ's:

$$\Sigma_{11}^{(\beta)}(\mathbf{k}, E) = \frac{1}{\mathcal{G}_0} - \mathcal{G}^-/(\mathcal{G}^-\mathcal{G}^+ - \tilde{\mathcal{G}}^2), \quad (29)$$

$$\Sigma_{02}^{(\beta)}(\mathbf{k}, E) = \tilde{\mathcal{G}}/(\mathcal{G}^-\mathcal{G}^+ - \tilde{\mathcal{G}}^2) \quad (30)$$

where

$$\mathcal{G}^+ = \mathcal{G}(\mathbf{k}, E), \qquad \mathcal{G}^- = \mathcal{G}(\mathbf{k}, -E). \quad (31)$$

Here we encounter a difficulty which one does not meet in the fermion problem. The Σ's behave properly at infinity, and are analytic everywhere except at points at which $\mathcal{G}^-\mathcal{G}^+ = \tilde{\mathcal{G}}^2$. We have been unable to prove that no such points exist. Thus, if we have determined the Σ's by using thermodynamic perturbation theory at a set of points along the imaginary axis, we may not infer the analytic continuation by choosing those Σ's which behave correctly at infinity and are analytic everywhere in the upper or lower half-plane. Rather than deal directly with the Green functions, which, even for fairly low approximations for the Σ's, would be complicated functions of the energy, we can choose a pair of simple combinations of the Σ's which are analytic everywhere, and use these as an intermediate stage in our calculation of the analytic continuations of the Green functions. An example of such a pair of

functions is

$$A(E) = [\Sigma_{11}^{(\beta)}(\mathbf{k}, E) - \mathcal{G}_0^{-1}] = -(\mathcal{G}^-\mathcal{G}^+ - \tilde{\mathcal{G}}^2)/\mathcal{G}^-, \quad (32)$$

$$B(E) = \Sigma_{02}^{(\beta)}(\mathbf{k}, E)A(E) = \tilde{\mathcal{G}}/\mathcal{G}^-. \quad (33)$$

These functions are analytic everywhere off the real axis, since the $\mathcal{G}$'s are analytic, and $\mathcal{G}^-$ cannot vanish [Eq. (16)]. Further, as $|E| \to \infty$, $\mathcal{G}^- \to 1/E$, $\tilde{\mathcal{G}} \to cE^{-2}$ where c is some constant. Thus A and B have the correct behavior at infinity. They will be simpler functions of the energy, and hence it will be easier to choose the correct (and unique) continuation. Having done this, we can determine the Σ's, and substitute them into the Green functions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 4 APRIL 1963

Exact Statistical Mechanics of a One-Dimensional System with Coulomb Forces. III. Statistics of the Electric Field*

ANDREW LENARD
Plasma Physics Laboratory Princeton University, Princeton, New Jersey
(Received 19 November 1962)

The statistics of the electric field in a one-dimensional "sheet" plasma is studied. It is shown that the electric field as a function of the space coordinate is a Markov process, whose basic probabilities are related to the Fourier coefficients of certain functions that play an important role in the previously found development of the grand partition function. Some implications regarding the statistics of particle locations are explored. The one-point probability distribution of the electric field is found to be asymptotically Gaussian in the plasma limit. It is pointed out that the one-dimensional analog of the Holtsmark calculation leads to an incorrect conclusion because electrostatic shielding is not properly taken into account.

1. INTRODUCTION

THE subject of this paper is a further investigation from the point of view of statistical mechanics of a simple one-dimensional system which was studied in two previous papers of this series.[1,2] This is a system of electrically charged sheets capable of motion in one dimension without any inhibition of free crossing over each other. In I, the configurational partition function of the constant pressure ensemble was shown to be an algebraic expression (a rational function of the pressure) which was given as a sum of simple terms, the number of which tends to infinity with the size of the system. The asymptotic form in the infinite system limit was shown to be determined by a characteristic value problem of Sturm–Liouville type (in the case treated, the Mathieu equation). This problem entered in a rather indirect way, and its true origin was only clarified in II. In the latter paper the problem was worked out in the grand canonical ensemble, and use was made of the fact that the grand partition function can be expressed as an average (or functional integral) of a certain functional of a Gaussian random function. The utility of this trick lies in the circumstance that in this model the appropriate Gaussian random function $\phi(x)$ is just "one dimensional Brownian motion,"[3] and is therefore a Markov process. The evolutionary character of this process leads to the expression of the grand partition function in terms of the fundamental solution of a certain partial differential equation. This equation, in turn, may be treated by the method of separation of variables. This way one arrives, in a natural manner, at the characteristic value problem just mentioned.

The Markoffian nature of $\phi(x)$—so essential in this result—is, however, left in the somewhat obscure status of a lucky coincidence. It must be remembered, of course, that $\phi(x)$ plays only a mathematical role, there being no physical significance to the "randomness" implicit in its definition. Nevertheless, the question may be raised whether it is possible to formulate the statistical mechanical problem itself

* This work accomplished under the auspices of the United States Atomic Energy Commission.

[1] A. Lenard, J. Math. Phys. 2, 682 (1961). This paper will be referred to as I.

[2] S. F. Edwards and A. Lenard, J. Math. Phys. 3, 778 (1962). This paper will be referred to as II.

[3] ϕ is displacement, x is time.

functions is

$$A(E) = [\Sigma_{11}^{(\beta)}(\mathbf{k}, E) - \mathcal{G}_0^{-1}] = -(\mathcal{G}^-\mathcal{G}^+ - \tilde{\mathcal{G}}^2)/\mathcal{G}^-, \quad (32)$$

$$B(E) = \Sigma_{02}^{(\beta)}(\mathbf{k}, E)A(E) = \tilde{\mathcal{G}}/\mathcal{G}^-. \quad (33)$$

These functions are analytic everywhere off the real axis, since the $\mathcal{G}$'s are analytic, and $\mathcal{G}^-$ cannot vanish [Eq. (16)]. Further, as $|E| \to \infty$, $\mathcal{G}^- \to 1/E$, $\tilde{\mathcal{G}} \to cE^{-2}$ where c is some constant. Thus A and B have the correct behavior at infinity. They will be simpler functions of the energy, and hence it will be easier to choose the correct (and unique) continuation. Having done this, we can determine the Σ's, and substitute them into the Green functions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 4 APRIL 1963

Exact Statistical Mechanics of a One-Dimensional System with Coulomb Forces. III. Statistics of the Electric Field*

ANDREW LENARD
Plasma Physics Laboratory Princeton University, Princeton, New Jersey
(Received 19 November 1962)

The statistics of the electric field in a one-dimensional "sheet" plasma is studied. It is shown that the electric field as a function of the space coordinate is a Markov process, whose basic probabilities are related to the Fourier coefficients of certain functions that play an important role in the previously found development of the grand partition function. Some implications regarding the statistics of particle locations are explored. The one-point probability distribution of the electric field is found to be asymptotically Gaussian in the plasma limit. It is pointed out that the one-dimensional analog of the Holtsmark calculation leads to an incorrect conclusion because electrostatic shielding is not properly taken into account.

1. INTRODUCTION

THE subject of this paper is a further investigation from the point of view of statistical mechanics of a simple one-dimensional system which was studied in two previous papers of this series.[1,2] This is a system of electrically charged sheets capable of motion in one dimension without any inhibition of free crossing over each other. In I, the configurational partition function of the constant pressure ensemble was shown to be an algebraic expression (a rational function of the pressure) which was given as a sum of simple terms, the number of which tends to infinity with the size of the system. The asymptotic form in the infinite system limit was shown to be determined by a characteristic value problem of Sturm–Liouville type (in the case treated, the Mathieu equation). This problem entered in a rather indirect way, and its true origin was only clarified in II. In the latter paper the problem was worked out in the grand canonical ensemble, and use was made of the fact that the grand partition function can be expressed as an average (or functional integral) of a certain functional of a Gaussian random function. The utility of this trick lies in the circumstance that in this model the appropriate Gaussian random function $\phi(x)$ is just "one dimensional Brownian motion,"[3] and is therefore a Markov process. The evolutionary character of this process leads to the expression of the grand partition function in terms of the fundamental solution of a certain partial differential equation. This equation, in turn, may be treated by the method of separation of variables. This way one arrives, in a natural manner, at the characteristic value problem just mentioned.

The Markoffian nature of $\phi(x)$—so essential in this result—is, however, left in the somewhat obscure status of a lucky coincidence. It must be remembered, of course, that $\phi(x)$ plays only a mathematical role, there being no physical significance to the "randomness" implicit in its definition. Nevertheless, the question may be raised whether it is possible to formulate the statistical mechanical problem itself

* This work accomplished under the auspices of the United States Atomic Energy Commission.

[1] A. Lenard, J. Math. Phys. 2, 682 (1961). This paper will be referred to as I.

[2] S. F. Edwards and A. Lenard, J. Math. Phys. 3, 778 (1962). This paper will be referred to as II.

[3] ϕ is displacement, x is time.

in such a way that a Markoffian property appears in connection with the basic statistics involved in *its* definition. Such a view, if possible, would throw further light on the reasons for the inherent simplicity of the model, a simplicity which is not immediately apparent in its initial formulation. It will be shown that this is indeed possible, and that the Markoffian element in the system is just the electric field regarded as a function of the space coordinate.

That the electric field plays an important role, was already observed in I where it was shown that it is most convenient to consider the total potential energy in the form

$$V = \frac{1}{8\pi}\int_0^L dx\, E^2(x), \tag{1}$$

where $E(x)$ is the electric field. In more complete notation one should indicate also its dependence on the particle locations $x_1, x_2, \cdots, x_N$. The explicit form is

$$E(x; x_1, x_2, \cdots, x_N) = -4\pi \sum_{i=1}^{N} \sigma_i \theta(x_i - x), \tag{2}$$

where the σ_i are the charges, and $\theta(x) = 1$ for $x > 0$, zero otherwise. Thus $E(x)$ is a piecewise constant function subject to the discontinuous increase $4\pi\sigma_i$ as x crosses x_i in the positive direction. The particle positions which enter as parameters in (2), are random variables subject to the basic probability distribution (II.2).[4] Therefore $E(x)$ itself must be considered as a one-parameter family of random variables or, what is the same, a random function.[5]

From a physical point of view it is not difficult to understand why $E(x)$ is Markoffian. Loosely speaking, the reason is that the only information communicated to particles in a region $x > X$ about particles in the complementary region $x < X$ is just the electric field $E(X)$ at the separation point. Therefore, the statistical properties of $E(x)$ for $x > X$ are dependent only on $E(X)$ but not on $E(x')$ with $x' < X$. This is precisely the characteristic property of a Markov process.

To define $E(x)$ completely in a statistical sense we do not need then the joint probability distribution of its values at an arbitrary number of distinct points. Rather the following quantities suffice for a complete description[6]

(a) The probability P_n that $E(x) = E_n$ for any x. E_n is one of the possible values of the electric field labeled by the integer n. The probability is independent of x.[7]

(b) The conditional probability $P_{n,m}(x)$, also referred to as transition probability, that $E(x+x_0) = E_m$ given that $E(x_0) = E_n$. It is sufficient to take $x > 0$, and the probability is independent of x_0.[7]

Between these quantities the following relationships must hold:

$$P_{n,m}(x) = \sum_l P_{n,l}(x - x')P_{l,m}(x') \quad (0 \le x' \le x), \tag{3}$$

and

$$P_n = \sum_l P_l P_{l,n}(x) \qquad (0 \le x). \tag{4}$$

These equations have obvious probability interpretation. It is also convenient to characterize the "evolution" of the Markov process by infinitesimal quantities. These are c_n and p_{nm} $(m \neq n)$, defined as follows. Given that $E(x) = E_n$ the probability that $E(x + dx) \neq E_n$ is $c_n\, dx$. c_n is the probability per unit time[8] that a change from the value E_n occurs. The quantities $p_{n,m}$ are defined by the statement that given a change from the state E_n in the time element dx the final state is E_m. Analytically, we have

$$c_n = -P'_{n,n}(0), \tag{5}$$

and

$$p_{n,m} = \frac{1}{c_n} P'_{n,m}(0) \qquad (m \neq n). \tag{6}$$

where the dash stands for differentiation with respect to x. Conversely, the c_n and the $p_{n,m}$ determine the $P_{n,m}(x)$ via the Kolmogorov differential equations

$$P'_{n,m}(x) = -P_{n,m}(x)c_m + \sum_{l \neq m} P_{n,l}(x)c_l p_{l,m}, \tag{7}$$

together with the initial condition $P_{n,m}(0) = \delta_{n,m}$.[9] The absolute probabilities P_n are then obtained as asymptotic limits

$$P_m = \lim_{x \to \infty} P_{n,m}(x). \tag{8}$$

---

[4] This notation means Eq. (2) of II.

[5] Frequently-used alternative terms are random process or stochastic process. The word "process" is well established by tradition and is intended to suggest that the independent variable is time. This is not the case in the present context, but we shall not make an effort to avoid traditional terminology.

[6] For a general discussion of this topic see W. Feller, *An Introduction to Probability Theory and Its Applications* (John Wiley & Sons, Inc., New York, 1957), Vol. I, Chap. XVII.

[7] In the infinite system limit.

[8] It is customary to think of the independent variable of a Markov process as "time," because this is the case in most applications. Of course, in the present case, this "time" has nothing to do with physical time.

[9] There are some subtle questions regarding the existence and uniqueness of solution for these equations. See reference 6 for a discussion of this topic.

2. CALCULATION

We remind the reader of the nature of the model and the notation used. The system is enclosed in a space of length L. The particles carry charges σ_i which fall into a fixed finite number of groups. We denote these values by σ', σ'', etc. They are assumed to be integral multiples of a common unit. It is convenient to choose this as the unit of charge, and this can be done obviously in such a way that the various charges become integers whose greatest common divisor is unity. It is further assumed that among the various species there is one at least with either sign of charge. In the units chosen, the electric field is capable of values which are integral multiples of 4π. To eliminate superfluous writing we consider then the function

$$\nu(x) = \frac{1}{4\pi} E(x) = -\sum_{i=1}^{N} \sigma_i \theta(x_i - x), \tag{9}$$

which can assume only integer values. The basic probability distribution of particle numbers N', N'', $\cdots$ and positions $x_1, x_2, \cdots, x_N$ ($N = N' + N'' + \cdots$) is given by the grand canonical ensemble (II.2).

Our task consists of obtaining the complete statistical description of $\nu(x)$ as given by (9). This means that, for any finite number h of distinct values $x = y_q$ ($q = 1, 2, \cdots, h$) and corresponding integers n_q, the following probability must be specified:

$$W_h = \text{Prob}\,\{\nu(y_1) = n_1, \nu(y_2) = n_2, \cdots, \nu(y_h) = n_h\}. \tag{10}$$

This is an infinite sequence of functions ($h = 1, 2, \cdots$), each depending on its $2h$ variables y_q and n_q. The y_q are restricted to lie in the interval $(0, L)$, and it is convenient to arrange the indices in such a way that

$$0 < y_1 < y_2 < \cdots < y_h < L. \tag{11}$$

In principle W_h is that fraction[10] of the integral (II.3) which is consistent with the conditions indicated in the parentheses of (10).

We now resort to a device frequently useful in probability theory. We first consider the Fourier series whose coefficients are the values of W_h for different values of the n_q (and fixed y_q). Thus

$$\tilde{W}_h = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \cdots \sum_{n_h=-\infty}^{\infty} W_h \exp\left(i \sum_{q=1}^{h} \xi_q n_q\right). \tag{12}$$

$\tilde{W}_h$ is also a function of $2h$ variables; it is periodic in the ξ_q with period 2π. It may be interpreted as the average value of

$$\exp\left[i \sum_{q=1}^{h} \xi_q \nu(y_q; x_1, x_2, \cdots, x_N)\right], \tag{13}$$

taken over the grand canonical probability distribution (II.2). We have written out explicitly the particle variables in the manner of (2) in order to emphasize that the $2h$ variables ξ_q and y_q are fixed parameters and the averaging goes over the x_i. We now substitute (9) into (13) and interchange the order of the two summations. This produces the quantity

$$\prod_{i=1}^{N} \exp\,[-i\sigma_i R_h(x_i)],$$

where

$$R_h(x) = \sum_{q=1}^{h} \xi_q \theta(x - y_q). \tag{14}$$

We have, therefore,

$$\tilde{W}_h = \Omega^{-1} \sum_{N'=0}^{\infty} \frac{z'^{N'}}{N'!} \sum_{N''=0}^{\infty} \frac{z''^{N''}}{N''!} \cdots \int_0^L dx_1 \cdots \int_0^L dx_N \times \exp\left[-\frac{1}{\theta} V_{N'N''\cdots} - i \sum_{i=1}^{N} \sigma_i R_h(x_i)\right]. \tag{15}$$

Ω is the grand partition function (II.3) and $V_{N'N''\cdots}$ is the total potential energy given, for instance, by (II.13).

The quantity $\Omega\tilde{W}_h$ as given by (15) bears a great resemblance to the expression (II.3) for the grand partition function. We claim that it may be similarly transformed into an average over the Wiener random function $\phi(x)$. This representation is the following

$$\Omega\tilde{W}_h = \left\langle \exp \int_0^L dx\, F_h(\phi(x), x) \right\rangle, \tag{16}$$

where the function F_h is given by

$$F_h(\phi, x) = z' \exp\,\{i\sigma'[\phi - R_h(x)]\} + z'' \exp\,\{i\sigma''[\phi - R_h(x)]\} + \cdots. \tag{17}$$

The meaning of the symbols is the same as in II. The bracket in (16) denotes the Wiener average over the functions $\phi(x)$ normalized by $\phi(0) = 0$. The proof of (16) proceeds along lines entirely parallel to the proof of the analogous formula (II.35), given in Sec. 4 of II, and therefore we do not repeat it here. The dependence of $\tilde{W}_h$ on the $2h$ parameters y_q and ξ_q is contained in F_h via R_h, [Eq. (14) above].

This representation casts $\Omega\tilde{W}_h$ into a general form (II.18) to which the theorem of Kac is applicable.[11]

[10] That is to say, the integral extended over the region in configuration space consistent with the conditions, divided by the complete integral.

[11] For a statement and a proof of this theorem see II. References to original papers on this topic are given there.

Making use of this theorem we obtain then

$$\Omega\tilde{W}_h = \int_{-\pi}^{\pi} d\phi\, \hat{Q}_h(\phi, L; 0, 0), \tag{18}$$

where $\hat{Q}_h = \hat{Q}_h(\phi, x; \phi_0, x_0)$ is the periodic fundamental solution of the partial differential equation

$$\left[\frac{\partial}{\partial x} - \frac{\partial^2}{\partial \phi^2} - F_h(\phi, x)\right]\hat{Q}_h = 0, \tag{19}$$

that is to say, the solution satisfying the initial condition

$$\hat{Q}_h(\phi, x_0; \phi_0, x_0) = \sum_{n=-\infty}^{\infty} \delta(\phi - \phi_0 + 2\pi n). \tag{20}$$

For simplicity we have put the "diffusion constant" $D = 2\pi/\theta = 1$, which merely fixes a particular choice for the unit of length.[12] $\hat{Q}_h$ is periodic[13] in both arguments ϕ and ϕ_0 with period 2π. The coefficient function $F_h(\phi, x)$ in the differential equation (19) is piecewise independent of x as shown by (17) and (14). This gives the clue for a further reduction.

We note, first of all, that $\hat{Q}_h$ as a fundamental solution has the propagation property

$$\hat{Q}_h(\phi, x; \phi_0, x_0) = \int_{-\pi}^{\pi} d\phi'\, \hat{Q}_h(\phi, x; \phi', x')\hat{Q}_h(\phi', x'; \phi_0, x_0), \tag{21}$$

valid for any x' inside the interval (x_0, x). We may apply this relation repeatedly, choosing as the intermediate points $y_1, y_2, \cdots, y_h$ all contained in the interval $(0, L)$. Thus

$$\hat{Q}_h(\phi, L; 0, 0) = \int_{-\pi}^{\pi} d\phi_h \cdots \int_{-\pi}^{\pi} d\phi_1 \times \prod_{q=0}^{h} \hat{Q}_h(\phi_{q+1}, y_{q+1}; \phi_q, y_q), \tag{22}$$

where, by convention, $\phi_0 = 0$, $\phi_{h+1} = \phi$, $y_0 = 0$, and $y_{h+1} = L$. We now inspect the typical factor in this product integrand. In the interval $y_q < x < y_{q+1}$ the function F_h is independent of x, namely

$$F_h = z' \exp\,[i\sigma'(\phi - \xi_1 - \xi_2 - \cdots - \xi_q)] + z'' \exp\,[i\sigma''(\phi - \xi_1 - \xi_2 - \cdots \xi_q)] + \cdots. \tag{23}$$

It is just the function $F(\phi)$ (without subscript) defined by (II.36), with an argument displaced by $\xi_1 + \xi_2 + \cdots + \xi_q$,

$$F_h(\phi, x) = F(\phi - \xi_1 - \xi_2 - \cdots - \xi_q), \tag{24}$$

as long as x is in the interval (y_q, y_{q+1}). Thus F_h is a step function of x, the change at each $x = y_q$ being a displacement by ξ_q in its first argument. The importance of this observation lies in the fact that we have already solved (19), with F replacing F_h, in II, the solution being the function $\hat{Q} = \hat{Q}(\phi, x; \phi_0, x_0)$ (without subscript). Explicitly,

$$\hat{Q}_h(\phi, x; \phi_q, y_q) = \hat{Q}(\phi - \xi_1 - \xi_2 - \cdots - \xi_q, x; \phi_q - \xi_1 - \xi_2 - \cdots - \xi_q, y_q). \tag{25}$$

Thus (22) and thence (18) may be expressed in terms of the function $\hat{Q}$ alone. A further simplification occurs on account of the periodicity of all functions in their ϕ arguments. These occur as integration variables always over the complete period $(-\pi, \pi)$. Hence appropriate substitutions $\phi \to (\phi +$ constant) may be used to simplify matters. In this way the following expression is obtained for $\tilde{W}_h$:

$$\tilde{W}_h = \frac{\int d\phi_{h+1} \int d\phi_h \cdots \int d\phi_1 \prod_{q=0}^{h} \hat{Q}(\phi_{q+1} + \xi_{q+1}, y_{q+1}; \phi_q, y_q)}{\int d\phi\, \hat{Q}(\phi, L; 0, 0)}. \tag{26}$$

Here and in all the following, the integrals over ϕ go from $-\pi$ to π, $\xi_{h+1} = 0$, by definition. The denominator is just the grand partition function given in (II.61). Note that the function $\hat{Q}$ depends only on the difference of its two space arguments.

At this point it is convenient to remove the dependence on the boundaries of the system. This is manifested by the fact that all $h + 1$ variables y_q ($q = 1, 2, \cdots, h + 1$) occur in $\tilde{W}_h$. In contrast, one would expect that if the "observation points" y_q ($q = 1, 2, \cdots, h$) are chosen very far from both boundaries of the large basic interval $(0, L)$, $\tilde{W}_h$ will depend only on the relative positions of these points (with respect to each other, but not with respect to the boundaries). Mathematically, this means the following limit:

$$\begin{cases} y_1 \to \infty \\ y_q - y_{q-1} \quad \text{fixed} \quad (q = 2, 3, \cdots, h) \\ L - y_h \to \infty. \end{cases} \tag{27}$$

[12] It is $2\pi\sigma_0^2/\theta$, where σ_0 is the unit of charge.

[13] In II at first a function was introduced which was not periodic. However, for the case of integer charge numbers, which is of sole interest here, only its periodic conterpart plays a role. (Cf. II, Sec. 5.)

In order to carry this out we need the asymptotic behavior of $\hat{Q}(\phi, x; \phi_0, x_0)$ for $x - x_0 \to \infty$. We assume that this function can be expanded in the series

$$\hat{Q}(\phi, x; \phi_0, x_0) = \sum_{m=0}^{\infty} e^{\gamma_m(x-x_0)} y_m(\phi) y_m(\phi_0), \qquad (28)$$

where $y_m(\phi)$ and γ_m are the characteristic functions and values respectively of the problem (II.62) with periodicity condition (II.63) normalized by (II.64), and assumed to form a complete set of periodic functions.[14] It is further assumed that there is a real positive characteristic value γ_0 such that $\gamma_0 > \text{Re}\ \gamma_m$ for $m \geq 1$. Then in the limit (27) only the first term in the sum (28) survives both in the denominator of (26) and in the two extreme factors of the numerator. The result is

$$\tilde{W}_h = e^{-\gamma_0(y_h - y_1)} \int d\phi_h \cdots \int d\phi_1\, y_0(\phi_h) \times \prod_{q=1}^{h-1} \hat{Q}(\phi_{q+1} + \xi_{q+1}, y_{q+1}; \phi_q, y_q) y_0(\phi_1 + \xi_1). \qquad (29)$$

We shall retain the notation $\tilde{W}_h$ to mean the limit in the sense of (27).

The last step in the calculation is the recovery of the probabilities W_h which are the Fourier coefficients in (12). We have[15]

$$W_h = \frac{1}{(2\pi)^h} \int d\xi_h \cdots \int d\xi_1\, \tilde{W}_h \times \exp\left(-i \sum_{q=1}^{h} \xi_q n_q\right). \qquad (30)$$

This result can be expressed simply in terms of the Fourier coefficients of the functions $\hat{Q}$ and y_0. Let us write

$$\hat{Q}(\phi, x; \phi_0, x_0) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} A_{n,m}(x - x_0) \times \exp\left[i(m\phi - n\phi_0)\right], \qquad (31)$$

and

$$y_0(\phi) = \frac{1}{(2\pi)^{\frac{1}{2}}} \sum_{n=-\infty}^{\infty} A_n \exp(in\phi). \qquad (32)$$

Substitute this into (29) and then into (30). The integrations can then be carried out trivially, and the following final result is obtained:

$$W_h = e^{-\gamma_0(y_h - y_1)} A_{n_1} A_{-n_h} \prod_{q=1}^{h-1} A_{n_q, n_{q+1}}(y_{q+1} - y_q). \qquad (33)$$

Some general properties of the stochastic process with the h-point distribution functions (33) will now be discussed.

3. GENERAL STATISTICAL PROPERTIES

(i) The stochastic process $\nu(x)$ is *stationary*. This means simply that all statistical properties are invariant under an arbitrary displacement, or in other words that all distributions W_h depend only on the differences between the h variables y_q. This is evident from the explicit formula (33) just obtained. It corresponds to the spatially homogeneous nature of thermal equilibrium (when the influence of the boundaries is removed by a limiting process).

(ii) *Independence of distant parts.* This means that random variables relating to distant points or regions are statistically independent. The mathematical expression of this property is the following asymptotic relation

$$W_h(y_1, y_2, \cdots, y_h) \sim W_k(y_1, \cdots, y_k) W_{h-k}(y_{k+1}, \cdots, y_h) \qquad (34)$$

in the limit $y_{k+1} - y_k \to \infty$, all other differences $y_{q+1} - y_q$ being held fixed. This property (the "product decomposition rule") is seen to follow from

$$A_{n,m}(x) \sim A_{-n} A_m \exp(\gamma_0 x) \qquad (x \to \infty) \qquad (35)$$

which itself is a consequence of the fact that in the limit indicated, the $m = 0$ term of (28) gives the asymptotic form of $\hat{Q}$. Other quantities that satisfy the product decomposition rule are the reduced density functions discussed in Section 7 of II.

(iii) $\nu(x)$ *is a Markow process.* This means that the conditional probability

$$\frac{W_{h+1}(n_1, y_1; n_2, y_2; \cdots; n_h, y_h; n_{h+1}, y_{h+1})}{W_h(n_1, y_1; n_2, y_2; \cdots; n_h, y_h)}$$

is, in effect, independent of n_q, y_q $(q = 1, 2, \cdots h-1)$ and of $h (= 1, 2, 3, \cdots)$. This is immediately verified from (33), and one finds for the transition probability

$$\frac{W_2(n, 0; m, x)}{W_1(n)} = P_{n,m}(x) = \frac{e^{-\gamma_0 x} A_{n,m}(x) A_{-m}}{A_{-n}}. \qquad (36)$$

This formula, together with the absolute (i.e., one-point) probability

$$W_1(n) = P_n = A_n A_{-n} \qquad (37)$$

[14] It would be desirable to have a proof of this assumption. The difficulty is that in general the characteristic value problem is not Hermitian. The author has searched in vain for a rigorous argument that would replace the usual one for the Hermitian case. It may be worth pointing out that there is no difficulty in the special case $\sigma' = 1$, $\sigma'' = -1$ studied in I. (Cf. the remarks at the end of Sec. 5 of II.)

[15] It is assumed that the inversion of the order of the operation (30) with the limiting process (27) is legitimate.

gives a complete specification of the Markov process $\nu(x)$.

(iv) *Reversibility.* This property is an expression of the fact that the x axis has no inherently distinguished direction associated with it, or, in other words, that the statistical properties of $\nu(x)$ are identical when viewed either in the direction of increasing or decreasing x. Mathematically, this is expressed by the identities[16]

$$W_h(n_1, y_1; n_2, y_2; \cdots ; n_h, y_h)$$
$$= W_h(-n_h, -y_h; \cdots ; -n_2, -y_2; -n_1, -y_1). \quad (38)$$

For $h = 1$ this is manifestly true as seen from (37). For $h \geq 2$ it is true if and only if

$$A_{n,m}(x) = A_{-m,-n}(x), \quad (39)$$

as one verifies in (33).

(v) The process $\nu(x)$ is *irreducible.* By this we mean that no combination of values for the n_q has zero probability. Mathematically, $W_h > 0$ for any[17] combination of arguments. For this it is sufficient to show that

$$A_{n,m}(x) > 0 \qquad (x > 0), \quad (40)$$

and

$$A_n > 0, \quad (41)$$

for all values of the indices.

Regarding these properties we make the following remarks. The first two are quite intuitive and verified easily. The third, the Markoffian property, is also quite reasonable, as was pointed out in the introduction, and its verification is immediate. On the other hand, the last two of the above mentioned properties, while also quite reasonable, are not so easy to verify. In order to do that, one must have recourse to the definition of the Fourier coefficients $A_{n,m}(x)$ and A_n. The function $\hat{Q}$ satisfies the partial differential equation[18]

$$\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial \phi^2} - \sum_\sigma z_\sigma e^{i\sigma\phi}\right)\hat{Q} = 0, \quad (42)$$

with the initial condition

$$\hat{Q}(\phi, x_0, \phi_0, x_0) = \sum_{n=-\infty}^{\infty} \delta(\phi - \phi_0 + 2\pi n). \quad (43)$$

From this we have for its Fourier coefficients the infinite coupled system of ordinary differential equations

$$\frac{\partial A_{n,m}(x)}{\partial x} + m^2 A_{n,m}(x) - \sum_\sigma z_\sigma A_{n,m-\sigma}(x) = 0, \quad (44)$$

with the initial conditions

$$A_{n,m}(0) = \delta_{n,m}. \quad (45)$$

Similarly, from the characteristic value equation (II.62) for the function $y_0(\phi)$, we get for its Fourier coefficients the infinite number of coupled homogeneous linear equations

$$(\gamma_0 + n^2)A_n - \sum_\sigma z_\sigma A_{n-\sigma} = 0. \quad (46)$$

It is from these equations that (39), (40), and (41) must be derived.

The theory of the system (44) is not quite obvious because an *infinite* number of variables are involved and there is no way to reduce the problem to a successive solution of a finite number of them at a time. A number of rigorous results concerning this system are to be found in the Appendix of this paper, including a proof of the symmetry (39) as well as (40) and (41).

In conclusion we give the quantities c_n and $p_{n,m}$ that characterize the Markov process $\nu(x)$ in a differential sense. From (5) and (36) together with (44) and (45), it follows that

$$c_n = \gamma_0 + n^2. \quad (47)$$

Again, from (6) we get similarly

$$p_{n,m} = z_{m-n}A_{-m}/(\gamma_0 + n^2)A_{-n} \qquad (n \neq m). \quad (48)$$

In view of (46) the sum of (48) over all m is unity, as it must be on account of its probability significance.

4. PARTICLE PROPERTIES

If the electric field $\nu(x)$ changes by an amount σ when x traverses the element dx in the positive direction, there is a particle of charge σ in this element. Therefore, the statistical properties of $\nu(x)$ are intimately associated with the statistical properties of the particles that make up the system. In the following we discuss some of those properties which are easiest obtained by making use of the results derived above.

Let $w_n(x)$ denote the conditional probability that the electric field is constant in the interval $(0, x)$, given that $\nu(0) = n$. Because $\nu(x)$ is Markoffian, we have

$$w_n(x + x') = w_n(x)w_n(x'). \quad (49)$$

Furthermore, by the definition of c_n [cf. Sec. 1],

[16] Originally we defined W_h only for $y_q > 0$. However, in the limit when stationarity holds this restriction can be removed in a trivial way.

[17] We remind the reader that W_h is defined for all y_q distinct. Our statement is not contrary to $W_h \to 0$ with fixed n_q and some y's tending to each other.

[18] We use a notation where the index σ labels the various particle species, $\sigma = \sigma'$, σ'', etc.

$$w_n(x) = 1 - c_n x + o(x) \qquad (x \to 0). \tag{50}$$

Therefore,

$$w_n(x) = \exp(-c_n x) = \exp[-(\gamma_0 + n^2)x]. \tag{51}$$

This is the probability that in the interval $(0, x)$ there are no particles, given that the electric field is n. Since different values for n are mutually exclusive possibilities, the sum

$$\sum_n P_n w_n(x) = \sum_n A_n A_{-n} \exp[-(\gamma_0 + n^2)x] \tag{52}$$

represents the probability that $(0, x)$ contains no particles (regardless of the electric field). We observe that for large x, the asymptotically dominant term is $n = 0$. Thus *a large gap between particles is overwhelmingly*[19] *likely to contain no electric field.*

In an entirely analogous manner we may calculate the probability density that there are particles of charges $\sigma_1, \sigma_2, \cdots$ in the elements $dx_1, dx_2, \cdots$ and no other particles, inside some interval $(0, l)$. This is

$$\sum_n P_n w_n(x_1) c_n p_{n,n+\sigma_1} w_{n+\sigma_1}(x_2 - x_1) \times c_{n+\sigma_1} p_{n+\sigma_1,n+\sigma_1+\sigma_2} \cdots . \tag{53}$$

The terms in the sum are the contributions from the various possible values the electric field may have, at $x = 0$. This expression may be written in a somewhat simpler form by substituting into it (37), (47), and (48). It becomes

$$z_{\sigma_1} z_{\sigma_2} \cdots \sum_n A_n A_{-(n+\sigma_1+\sigma_2+\cdots)} w_n(x_1) \times w_{n+\sigma_1}(x_2 - x_1) w_{n+\sigma_1+\sigma_2}(x_3 - x_2) \cdots . \tag{54}$$

Let us denote this sequence of functions by $e_{\sigma_1}(x_1)$, $e_{\sigma_1\sigma_2}(x_1, x_2) \cdots$. In addition to the indicated variables, they also depend on the size l of the chosen region. Conforming to this notation we may complete this hierarchy of functions by defining e_0 as the quantity (52), the probability that there are no particles at all in this region.

How do these functions depend on l? One sees easily that as $l \to \infty$ they all tend to zero exponentially. However, it is an interesting fact that in this limit their *ratios* tend to finite values, depending only on the relative positions $x_1, x_2, \cdots$. Indeed, making use of (51), one verifies that

$$\xi_{\sigma_1\sigma_2\cdots}(x_1, x_2, \cdots) \equiv \lim_{l\to\infty} \frac{e_{\sigma_1\sigma_2\cdots}(x_1, x_2, \cdots)}{e_0} = z_{\sigma_1} z_{\sigma_2} \cdots \delta_{\sigma_1+\sigma_2+\cdots,0} \times \exp[-\sigma_1^2(x_2 - x_1) - (\sigma_1 + \sigma_2)^2(x_3 - x_2) - \cdots]. \tag{55}$$

It should be stressed that the functions $\xi_{\sigma_1\sigma_2\cdots}$ thus introduced are *not* conditional probabilities because the numerators and the denominator refer to mutually exclusive possibilities. They are, however, non-negative and satisfy the product decomposition rule (cf. above). The argument of the exponential function is just the total potential energy of the charges $\sigma_1, \sigma_2 \cdots$ placed at points $x_1, x_2 \cdots$ divided by the temperature (in the units chosen). These are general properties of the ξ functions, not merely depending on our model. The explicit verification of the formula (55) illustrates a state of affairs recently discussed by Green[20] who has pointed out the interrelations of these functions with the hierarchy $f_{\sigma_1\sigma_2\cdots}$ discussed in Section 7 of II. We wish to call attention in this connection to an interesting aspect of self consistency in statistical mechanics: The distribution of particle positions in a large[21] region, emptied by some fluctuation of all other particles, is just the canonical distribution.[22]

We now raise the question of the probability distribution of the "gap size" between two particles whose charges are known. The quantity

$$\sum_n P_{n-\sigma_1} c_{n-\sigma_1} p_{n-\sigma_1,n} w_n(x_2 - x_1) c_n p_{n,n+\sigma_2}\, dx_1\, dx_2 = \sum_n A_{n-\sigma_1} A_{-n-\sigma_2} z_{\sigma_1} z_{\sigma_2} \times \exp[-(\gamma_0 + n^2)(x_2 - x_1)]\, dx_1\, dx_2$$

represents the probability that there are two particles of charges σ_1 and σ_2 in the elements dx_1 and dx_2 respectively $(x_2 > x_1)$, and that there are none in between. We now take $x_1 < 0$, $x_2 = x_1 + l > 0$ and integrate over all values consistent with these conditions. This gives the probability that the origin (or any other chosen point) is between two particles of charges σ_1 and σ_2 whose distance from each other lies in the interval $(l, l + dl)$:

$$z_{\sigma_1} z_{\sigma_2} \sum_n A_{n-\sigma_1} A_{-n-\sigma_2} \exp[-(\gamma_0 + n^2)l]\, l\, dl. \tag{56}$$

By means of (46) it is easily verified that the sum

[19] In the sense that the probability of the opposite possibility is exponentially small relative to the former as $x \to \infty$.

[20] M. S. Green, "Some Applications of the Generating Functional of the Molecular Distribution Functions," *Lectures in Theoretical Physics* held at the University of Colorado in 1960 (Interscience Publishers, Inc., New York, 1961), Vol. III.

[21] Large, but still small compared to the over-all system size L. Mathematically, we have let $L \to \infty$ first and then $l \to \infty$. It is essential to keep this order of limits in mind.

[22] The number of particles considered must however, remain small compared to the mean number in the region of size l. Thus in (55) the particles are fixed and $l \to \infty$. See also the remarks of Green, reference 20.

of (56) over all σ_1 and σ_2 when integrated over all l in $(0, \infty)$ is just unity, as it must be because this is just the total probability of an exhaustive set of mutually exclusive events. If (56) is summed over σ_1 and σ_2 one obtains the probability distribution of gap size regardless of the species of particles at the two ends.[23] This is

$$\sum_n A_n A_{-n}(\gamma_0 + n^2)^2 \exp\left[-(\gamma_0 + n^2)l\right] l\, dl. \quad (57)$$

An interesting identity emerges from this. The mean *inverse* gap size, formed with the probability distribution (57), is

$$\begin{aligned}\sum_n A_n A_{-n}(\gamma_0 + n^2)^2 \int_0^\infty \exp\left[(\gamma_0 + n^2)l\right] dl \\ = \sum_n A_n A_{-n}(\gamma_0 + n^2) \\ = \sum_n A_n \sum_\sigma z_\sigma A_{-n-\sigma}, \quad (58)\end{aligned}$$

where in the last step (46) was used again. Rewriting this in terms of the definition (32) of the A_n as Fourier coefficients, it becomes

$$\sum_\sigma z_\sigma \int_{-\pi}^{\pi} y_0^2(\phi) \exp(i\sigma\phi) \quad (59)$$

which according to (II.107), is just the total number density.

5. THE PLASMA LIMIT

In this section we shall give an application of the foregoing calculations to a more special problem. What is the one-point probability distribution of the electric field in the "plasma limit"? This limiting condition[24] obtains when the particle kinetic energy is large compared to the potential (or electrostatic) energy, so that in the first approximation particles move independently of each other. The question raised here is thus the analogue for our model of the problem first worked out for a real charged-particle gas by Holtsmark.[25] Nevertheless, it will be seen that the solution is in some respects entirely different. To bring out this difference we shall at first present a calculation patterned closely on the Holtsmark argument, one which in this case leads to an irrelevant conclusion.

In all that follows we shall restrict ourselves to the special case of a two-component gas with $\sigma = \pm 1$.[26] As pointed out in II, it is no loss in generality to set $z_1 = z_{-1}(=z$, for short) in this case. The characteristic equation for $y_0(\phi)$ is then just the Mathieu equation

$$[d^2/d\phi^2 + 2z\cos\phi]y_0(\phi) = \gamma_0 y_0(\phi). \quad (60)$$

The plasma limit occurs when z is very large.[27]

Following Holtsmark, let us assume then that the interaction between particles may be entirely neglected. Thus N particles of charge $\sigma = 1$ and N of charge $\sigma = -1$ are uniformly and independently distributed in the "box" $(0, L)$. Let x be any point inside this interval. The probability that N_1 positive and N_2 negative particles are in $(0, x)$, and the rest in (x, L), is, according to our assumption, just the product of two binomial distributions

$$\binom{N}{N_1}\binom{N}{N_2}\left(\frac{x}{L}\right)^{N_1+N_2}\left(1 - \frac{x}{L}\right)^{2N-N_1-N_2}. \quad (61)$$

In this case $\nu(x) = N_1 - N_2$. Therefore, if n be any integer between $-N$ and N, the probability P_n that $\nu(x) = n$ is just the sum of (61) over all N_1 and N_2 subject to the conditions

$$\begin{cases} 0 \le N_1 \le N, \\ 0 \le N_2 \le N, \\ N_1 - N_2 = n. \end{cases} \quad (62)$$

The characteristic function, or Fourier series, corresponding to the probability distribution P_n is easily shown to be[28]

$$\begin{aligned} Q(\xi) &= \sum_{n=-N}^{N} P_n \exp(in\xi) \\ &= (p^2 + q^2 + 2pq\cos\xi)^N, \quad (63)\end{aligned}$$

where $p = x/L$ and $q = 1 - p$. From the inversion

$$P_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\xi\,(p^2 + q^2 + 2pq\cos\xi)^N, \quad (64)$$

we want to obtain P_n in the infinite system limit. This means $N \to \infty$, $L \to \infty$ with a fixed ratio N/L. The observation point x must also recede from the boundaries so we take p and q as fixed. The integrand in (64) is, for large N, a sharply peaked function around $\xi = 0$; hence in the usual way we approximate the integral by replacing the integrand by the appropriate Gaussian $\exp(-pq\xi^2)$. This gives

$$P_n \sim \text{const.} \exp\left(-\frac{n^2}{4Npq}\right) \quad \binom{N \to \infty}{p,\ q \text{ fixed}}. \quad (65)$$

[23] The probability that some chosen point lies inside a gap whose length is in $(l, l + dl)$. This must be distinguished from the probability that a "randomly chosen" gap has size in $(l, l + dl)$. The two quantities differ by a factor proportional to l.

[24] Discussed in Sec. 8 of I and Sec. 6 of II.

[25] J. Holtsmark, Ann. Physik. **58**, 577 (1919); Physik. Z. **20**, 162 (1919), and **25**, 73 (1924).

[26] This is the case considered in I.

[27] In ordinary units, $z\theta/2\pi\sigma^2$ very large. This implies $\gamma_0 = P/2\pi\sigma^2$ very large, where P is the pressure.

[28] This is just the function $\tilde{W}_1(x, \xi)$, in the present approximation scheme.

Regarding this result we make the following observations: (65) still depends on p which must be interpreted to mean that the influence of the boundaries is not completely eliminated. This is unphysical, but one could argue on grounds of symmetry that the true approximation should have $p = q = \frac{1}{2}$ in this formula. There is, however, a more serious objection. The mean square electric field (or potential energy per unit length) as calculated from (65) is proportional to N. Therefore, the total potential energy is found to be proportional to N^2 instead of N, as all extensive quantities should. This must be regarded as a failure of the "Holtsmark method" to give a workable approximation for the electric field distribution in the one-dimensional plasma.

It is true that the mean square electric field is infinite in the three-dimensional case too. But that is for a different reason, namely because the electric field becomes large too fast as a particle approaches the observation point. In contrast, this effect is absent in the one-dimensional model; here the large electric fields arise from the many distant particles. Our calculation shows that complete independence of particles is in this case insufficient to give the necessary charge cancellation. The lack of proper decrease of the electric field due to a single particle is responsible for this state of affairs.

From a mathematical point of view, we have inadmissibly interchanged two limiting processes. One is the infinite system limit. As always in statistical mechanics, this must be carried out first. The second is the plasma limit which may be interpreted loosely as saying that particles become independent of each other. *The Holtsmark method corresponds to an interchange of the order of these limits.* This is legitimate in the real plasma problem, but not in the one-dimensional case, as we have just seen.

Let us now see how the correct approximation to P_n may be obtained. According to (37) and (32), we need the Fourier coefficients of $y_0(\phi)$ defined in the present case by the Mathieu equation (60). It is known that in the limit of large z, the function $y_0(\phi)$ has sharp maxima at $\phi = 0$ and points displaced by multiples of 2π. In this limit we have[29]

$$\gamma_0 \sim 2z - z^{\frac{1}{2}} \qquad (z \to \infty). \tag{66}$$

Let us then approximate $y_0(\phi)$ by a Gaussian, const. $\exp(-\frac{1}{2}K\phi^2)$, near $\phi = 0$, and determine the constant K by substituting into (60) and making use of (66). The equation is satisfied near $\phi = 0$ if $K = z^{\frac{1}{2}} \sim (\frac{1}{2}\gamma_0)^{\frac{1}{2}}$. Now we calculate the Fourier coefficients A_n, and so the following result emerges:

$$P_n \sim \text{const} \cdot \exp[-n^2/(\tfrac{1}{2}\gamma_0)^{\frac{1}{2}}]. \tag{67}$$

In the plasma limit the electric field distribution is Gaussian.

This result bears some resemblance to the incorrect (65), but we must remember that the "cutoff" in the distribution (67) comes from the physical parameter γ_0[30] which, though large in the plasma limit we are considering, is certainly small compared to N which characterizes the system size. The order of magnitude relations $1 \ll \gamma_0 \ll N$ correspond to the proper order in which the asymptotic limits $N \to \infty$, $\gamma_0 \to \infty$ are carried out.

In conclusion, let us point out that the correct value for the mean squared electric field is obtained from (67). Since γ_0 is large we may approximate sums over n by integrals. If this is done and the proper units are restored for the dimensional quantities, one obtains for the potential energy per unit length, the quantity $\pi\sigma^2(\frac{1}{2}\gamma_0)^{\frac{1}{2}}$ which is consistent with the dominant potential energy term of (I.56).

APPENDIX

We shall present some rigorous results on the coefficients $A_{n,m}(x)$ and A_n. These are needed to establish some of the properties discussed in Sec. 3. Equations (44) together with the initial conditions (45) will be taken as the definition of the $A_{n,m}(x)$. We shall prove the following:

Theorem. *The infinite coupled system of differential equations* (44) *has a solution reducing to the initial values* (45). *If* $A_{n,m}(x)$ *is bounded in* m, *the solution is unique. For this solution the symmetry* (39) *and the positiveness condition* (40) *hold.*

We first convert the differential system (44) and the initial condition (45) into the system of integral equations

$$A_{n,m}(x) = \delta_{n,m} \exp(-m^2 x) + \int_0^x dx' \times \exp[-m^2(x - x')] \sum_\sigma z_\sigma A_{n,m-\sigma}(x'). \tag{68}$$

Consider its formal iteration series

$$\delta_{n,m} \exp(-m^2 x) + \sum_{\sigma_1} z_{\sigma_1} \delta_{n,m-\sigma_1} \int_0^x dx_1 \times \exp[-n^2(x - x_1) - (n - \sigma_1)^2 x_1]$$

[29] These are the first two terms of (I.52). For a derivation see the reference given in I. A not quite rigorous derivation is given in Sec. 6 of II.

[30] The pressure in a suitably chosen unit.

$$+ \sum_{\sigma_1} \sum_{\sigma_2} z_{\sigma_1} z_{\sigma_2} \delta_{n,m-\sigma_1-\sigma_2} \int_0^x dx_1 \int_0^{x_1} dx_2$$
$$\times \exp\left[-n^2(x-x_1)-(n-\sigma_1)^2(x_1-x_2)-(n-\sigma_1-\sigma_2)^2 x_2\right]$$
$$+ \cdots . \tag{69}$$

Let $M \equiv \sum_\sigma z_\sigma$. It is evident that the kth term of the series is less than

$$M^k \int_0^x dx_1 \int_0^{x_1} dx_2 \cdots \int_0^{x_{k-1}} dx_k = \frac{(Mx)^k}{k!}. \tag{70}$$

Therefore the series converges. Moreover, by the Weierstrass "M test," it converges uniformly in the subscripts and x in any finite interval $0 \leq x \leq x_0$. Hence it actually gives a solution to the system (68). Term-by-term differentiation is also legitimate. This proves the first part of the theorem.

Assume now that there are two solutions, both of which are bounded in m and satisfy the same initial conditions. Let the difference be denoted by $\Delta_m(x)$. Then we have

$$\Delta_m(x) = \int_0^x dx' \exp\left[-m^2(x-x')\right] \sum_\sigma z_\sigma \Delta_{m-\sigma}(x'). \tag{71}$$

Let $b(x)$ stand for the least upper bound (in m) of $|\Delta_m(x)|$; it is a Lebesgue measurable function.[31] From (71) follows that

$$|\Delta_m(x)| \leq M \int_0^x dx'\, b(x'), \tag{72}$$

and this implies

$$0 \leq b(x) \leq M \int_0^x dx'\, b(x'). \tag{73}$$

Lemma 1. *The only Lebesgue measurable function that satisfies the inequalities* (73) *for all x on an interval $0 \leq x \leq x_0$ is identically zero there.*

To prove this we note that the right-hand side is a nondecreasing function of x. Hence

$$0 \leq b(x_1) \leq M \int_0^x dx'\, b(x'), \tag{74}$$

for all x_1 and x such that $0 \leq x_1 \leq x \leq x_0$. Let then $\bar{b}(x)$ stand for the least upper bound of $b(x_1)$ for $0 \leq x_1 \leq x$. It is also measurable, and obviously

$$0 \leq \bar{b}(x) \leq M \int_0^x dx'\, b(x')$$
$$\leq M \int_0^x dx'\, \bar{b}(x') \leq Mx\bar{b}(x), \tag{75}$$

because $\bar{b}(x)$ is nondecreasing. From $\bar{b}(x) \leq Mx\bar{b}(x)$ it follows that $\bar{b}(x) = 0$ for $0 \leq x < M^{-1}$. The vanishing of $\bar{b}(x)$ in succeeding intervals $(n-1)M^{-1} \leq x < nM^{-1}$ $(n = 2, 3, \cdots)$ is then easily shown by induction. This proves Lemma 1.

We now conclude that $\Delta_m = 0$ identically in m and x. This proves the uniqueness of the solutions.

It may be worthwhile to note that without the demand of boundedness[32] uniqueness does not hold; in fact, there is then a vast class of solutions satisfying the same initial condition. For instance, in the case when only z_1 and z_{-1} are different from zero, one may take $\Delta_0(x)$ and $\Delta_1(x)$ *any* indefinitely differentiable functions which vanish with all their derivatives at the origin [such as $\exp(-1/x)$]. Then $\Delta_2(x)$, $\Delta_3(x)$, $\cdots$ may be computed in succession by differentiating (71). Hence they become finite linear combinations of $\Delta_0(x)$ and $\Delta_1(x)$ as well as their derivatives, so that they vanish at $x = 0$ without being identically zero. Similarly one may determine $\Delta_{-1}(x)$, $\Delta_{-2}(x)$, $\cdots$ in succession. Our argument shows that for such a sequence of functions, uniform boundedness in any interval $0 \leq x \leq x_0$ is not possible.

The symmetry (39) is verified directly on each term of the series (69).

Finally we come to the positiveness property. All terms in (69) are manifestly nonnegative. Thus it is sufficient to show that for any n and m there is at least one term in the series that is actually positive. This depends on showing that there is an integer k and k choices $\sigma_1, \sigma_2, \cdots, \sigma_k$ (with repetitions allowed) among the specified σ values such that $n = m - \sigma_1 - \sigma_2 - \cdots - \sigma_k$. The truth of this depends on an elementary number theoretical proposition.

Lemma 2. *Let $\sigma_1, \sigma_2, \cdots, \sigma_h$ $(h \geq 2)$ be a set of distinct, nonvanishing integers such that among them there is at least one of each sign and that they have no nontrivial common divisors. Then any integer n may be written in the form*

$$n = k_1\sigma_1 + k_2\sigma_2 + \cdots k_h\sigma_h \tag{76}$$

where the k_i are nonnegative integers.

We remind the reader that our charge numbers σ_i satisfy the hypotheses of the Lemma.

We shall prove this by induction on h. Let $h = 2$. Let the function $r(k_1, k_2)$ be the remainder of $k_1\sigma_1 + k_2\sigma_2$ upon division with $\sigma_1\sigma_2$. Thus r is capable of

[31] It need not be continuous, in principle, even if all $\Delta_m(x)$ are differentiable any number of times.

[32] In view of the fact that our $A_{n,m}(x)$ serve as coefficients in a Fourier series boundedness is, of course, necessary for convergence. In actuality there is a rapid decrease with the subscripts m and n.

values in $0 \leq r \leq |\sigma_1\sigma_2| - 1$. Let us restrict the variables k_1 and k_2 by $0 \leq k_1 \leq |\sigma_2| - 1$ and $0 \leq k_2 \leq |\sigma_1| - 1$, and consider the values of r on these $|\sigma_1\sigma_2|$ ordered pairs (k_1, k_2). Suppose $r(k_1, k_2) = r(k_1', k_2')$. This means that the number $(k_1 - k_1')\sigma_1 + (k_2 - k_2')\sigma_2$ is divisible by $\sigma_1\sigma_2$, and in particular by σ_2. Since the second term is divisible by σ_2, so must the first. But by assumption σ_1 and σ_2 are relatively prime, so that $k_1 - k_1'$ is then divisible by σ_2. Now, $|k_1 - k_1'| \leq |\sigma_2| - 1$, hence this is possible only if $k_1 = k_1'$. Similarly, $k_2 = k_2'$. Thus, for distinct pairs (k_1, k_2) and (k_1', k_2') the function r assumes distinct values. Hence *all possible values for r are actually assumed,* since their total number is equal to the total number $|\sigma_1\sigma_2|$ of distinct pairs (k_1, k_2).

Let now n be an arbitrary integer and write it in the form $n = q\sigma_1\sigma_2 + r$ with q and r integers and $0 \leq r \leq |\sigma_1\sigma_2| - 1$. We have just shown that r may then be always written $k_1\sigma_1 + k_2\sigma_2$ where both k_1 and k_2 are nonnegative. But $q\sigma_1\sigma_2$ itself may be regarded as either a nonnegative multiple of σ_1 or a nonnegative multiple of σ_2, since by assumption, σ_1 and σ_2 have opposite signs. This produces the desired representation of n, and the lemma is proved for $h = 2$.

Let now $h \geq 3$ and suppose the lemma to hold for $h - 1$. Label the σ's so that already among $\sigma_1, \sigma_2, \cdots, \sigma_{h-1}$ there is at least one of each sign; and further let d be their greatest common divisor (which may or may not be equal to 1; but if not, it does not divide σ_h). We now apply the lemma to the $h - 1$ numbers σ_i/d $(i = 1, 2, \cdots, h - 1)$. Thus nonnegative k_i exist such that $-sg(\sigma_h) = \sum_{i=1}^{h-1} k_i\sigma_i/d$. On the other hand the lemma also applies to the two numbers σ_h and $-d \cdot sg(\sigma_h)$. Thus if n be any integer, we have $n = \kappa_1\sigma_h - \kappa_2 \cdot d \cdot sg(\sigma_h)$ with nonnegative κ_1 and κ_2. But then $n = \kappa_1\sigma_h + \kappa_2 \sum_{i=1}^{h-1} k_i\sigma_i$ which gives the desired representation of n. This completes the proof of Lemma 2, and with it the proof of the theorem.

A similar argument can be made to show that all coefficients $A_n > 0$. We first note that from (35) follows[33] the weaker inequality $A_nA_m \geq 0$, so that no two A_m can have different signs. Conventionally, we may take $A_n \geq 0$.[34] Not all A_n vanish, so let for instance $A_k > 0$. It follows from (46) that $A_{k+\sigma} > 0$ for all choices of σ. Continuing this argument, $A_{k+\sigma_1+\sigma_2} > 0$ for all choices of σ_1 and σ_2, and so on. However, from Lemma 2 we see that the subscripts which are reached this way exhaust all integers, and from this (41) follows.

---

[33] Note, however, that we do not have a rigorous general proof of (35). See also footnote 14.

[34] Since a factor of modulus one is arbitrary in $y_0(\phi)$.

Generalized Master Equation and t-Matrix Expansion*†

ROBERT J. SWENSON‡
Lehigh University, Bethlehem, Pennsylvania
(Received 8 December 1962)

Some mathematical aspects of the generalized master equation are discussed. The resolvent operator is expanded in terms of a two-body scattering matrix and this result is used to express the quantities in the generalized master equation in terms of expansions in the scattering matrix. The expansions are utilized in studying the density dependence of these quantities.

ONE of the main problems of nonequilibrium statistical mechanics is the evaluation of the formal correlation function expressions for transport coefficients first obtained by Green.[1] Work in this direction has been done by a number of authors[2] using a variety of techniques. The generalized master equation of Van Hove[3,4] is useful in this respect, and certain of its mathematical aspects will be investigated in this paper. Applications to the problem of transport in gases will be discussed in a subsequent paper.

In order to make practical calculations of transport coefficients using the generalized master equation, it is necessary to expand the quantities which appear in it in powers of a suitable small parameter. The derivation of a generalized master equation given by Van Hove is based on an expansion of the resolvent operator in powers of the interaction, which results in perturbation expansions for the quantities of interest. If the interaction between particles is weak, then these expansions are useful. If, however, the interaction contains an infinite repulsive core, then divergences appear in the above expansions and complicated summations of infinite series must be performed to get expansions in terms of a bounded operator—the two-body scattering matrix or t matrix. In fact, it is only in the lowest order in density that the above summations can easily be carried out.[4] In a recent paper[5] the author derived a generalized master equation without recourse to perturbation theory and these results will be used below to obtain t-matrix expansions for the quantities of interest. In addition to being useful when the interaction contains an infinite repulsive core, these expansions are convenient for discussing the density dependence of the desired quantities. In similar contexts, Resibois[6] has discussed a binary collision expansion for the quantum mechanical three-body scattering problem and Fujita[7] has discussed a binary collision expansion of the quantum statistical pair propagator.

The resolvent operator will be expanded in terms of a t matrix or scattering matrix, an expansion first given by Watson.[8] This is used with the results of I to obtain the desired t-matrix expansions. These expansions are then expressed in terms of diagrams and the density dependence discussed.

The Hamiltonian for a system of N particles in a volume V with density $c = N/V$ is written

$$H = H_0 + H_1, \tag{1}$$

where H_0 is the free-particle Hamiltonian and H_1 is assumed to be of the form

$$H_1 = \sum_{i<j} v(\mathbf{r}_i, \mathbf{r}_j) \equiv \sum_{i<j} v(ij), \tag{2}$$

where $\mathbf{r}_k$ is the position of the kth particle and $v(\mathbf{r}_i, \mathbf{r}_j)$ is the interaction between particles i and j. In order to save writing, define

$$\sum_{i<j} v(ij) = \sum_{\mu} v(\mu);$$

that is, two-particle operators are indicated with only one label. Matrix elements will be computed in the representation furnished by the eigenvectors of H_0,

$$H_0 \, |p\rangle = \epsilon_p \, |p\rangle. \tag{3}$$

* This work was supported by the National Science Foundation.

† Based on part of the author's dissertation submitted to Lehigh University in partial fulfillment of the requirements for the degree of Doctor of Philosophy, October, 1961.

‡ Present address; Chimie Physique II, Université Libre de Bruxelles, Bruxelles, Belgium.

[1] M. S. Green, J. Chem. Phys. **22**, 398 (1954).

[2] S. Fujita and R. Abe, J. Math. Phys. **3**, 350 (1962) and references therein.

[3] L. Van Hove, Physica **23**, 441 (1957). An extension of the work of Van Hove can be found in A. Janner, Helv. Phys. Acta **35**, 47 (1962).

[4] L. Van Hove, in *La Théorie des Gaz Neutres et Ionisés*, edited by C. DeWitt and J. G. Detoeuf (John Wiley & Sons, Inc., New York, 1959).

[5] R. J. Swenson, J. Math. Phys. **3**, 1017 (1962). Future reference to this paper is denoted by I.

[6] P. Resibois, Physica **27**, 33 (1961).

[7] S. Fujita, Physica **27**, 930 (1961).

[8] K. M. Watson, Phys. Rev. **103**, 489 (1956).

Our analysis is based on the properties of the resolvent operator R_l[3,9] defined by

$$R_l = (H - l)^{-1}, \tag{4}$$

where l is a complex number. R_l is related to the unitary time transformation operator according to

$$\exp[-iHt/\hbar] = (2\pi i)^{-1} \int_C dl\, R_l \exp[-ilt/\hbar], \tag{5}$$

with C a counterclockwise contour enclosing a sufficiently large portion of the real axis. Let us define a transition probability $P(t \mid pp_0)$ by[10]

$$P(t \mid pp_0) \equiv \langle p_0| \exp[iHt/\hbar] |p\rangle \times \langle p| \exp[-iHt/\hbar] |p_0\rangle. \tag{6}$$

Clearly $P(0 \mid pp_0) = \delta^{kr}(p-p_0)$, with δ^{kr} a Kronecker delta function. Combining Eqs. (5) and (6), we find

$$P(t \mid pp_0) = -(2\pi)^{-2} \int_C dl\, dl' \times \exp[i(l - l')t/\hbar] X_{ll'}(pp_0), \tag{7}$$

having defined X by

$$X_{ll'}(pp_0) = \langle p_0| R_l |p\rangle\langle p| R_{l'} |p_0\rangle. \tag{8}$$

The generalized master equation is an identity satisfied by the partial transition probability $P_E(t \mid pp_0)$, which in turn is defined as

$$P_E(t \mid pp_0) = (2\pi^2)^{-1} \int_{-\infty}^{\infty} dE' \times \exp[2i(E' - i\eta)t/\hbar] X_{E+E'-i\eta, E-E'+i\eta}(pp_0), \tag{9}$$

with η a small positive number. For $t > 0$ we have the relation

$$P(t \mid pp_0) = \int_{-\infty}^{\infty} dE\, P_E(t \mid pp_0).$$

The identity in question is

$$\frac{dP_E(t \mid pp_0)}{dt} = f_E(t \mid p)\delta^{kr}(p - p_0) + 2\pi \sum_{p'} \int_0^t dt' [w_E(t - t' \mid pp') \times P_E(t' \mid p'p_0) - w_E(t - t' \mid p'p) P_E(t' \mid pp_0)]. \tag{10}$$

This generalized master equation is a consequence of the following identity for X:

$$(l - l')X_{ll'}(pp_0) = F_{ll'}(p)\delta^{kr}(p - p_0) - i \sum_{p'} [\tilde{W}_{ll'}(pp')X_{ll'}(p'p_0) - \tilde{W}_{ll'}(p'p)X_{ll'}(pp_0)], \tag{11}$$

where we have introduced

$$\tilde{W}_{ll'}(pp') \equiv iF_{ll'}(p)W_{ll'}(pp'), \tag{12}$$

and W and F will be defined below [Eqs. (35) and (36)]. The quantities w_E and f_E which appear in Eq. (10) are given in terms of W and F according to

$$f_E(t \mid p) = i(2\pi^2\hbar)^{-1} \int_{-\infty}^{\infty} dE' \times \exp[2i(E' - i\eta)t/\hbar] F_{E+E'-i\eta, E-E'+i\eta}(p),$$

and

$$w_E(t \mid pp') = (2\pi^2\hbar^2)^{-1} \int_{-\infty}^{\infty} dE' \times \exp[2i(E' - i\eta)t/\hbar] \tilde{W}_{E+E'-i\eta, E-E'+i\eta}(pp').$$

It is the quantities F and W with which we are concerned.

First let us use reasoning similar to that of Watson[8] to discuss a t-matrix expansion for F and W. The resolvent operator R_l satisfies the integral equation

$$R_l = d_l - d_l H_1 R_l = d_l - R_l H_1 d_l, \tag{13}$$

where d_l is the unperturbed resolvent

$$d_l = (H_0 - l)^{-1}. \tag{14}$$

Now introduce an operator defined by the integral equation

$$t_l(\mu) = v(\mu) - v(\mu) d_l t_l(\mu). \tag{15}$$

This is the familiar two-body t matrix,[8] which does however depend on N particles through the propagator d. (In the appendix we discuss a second quantized t matrix which is more useful if the effects of quantum statistics are to be studied.) Defining T by

$$T_l = \sum_{\mu} t_l(\mu), \tag{16}$$

Watson[8] has obtained an expansion for R in terms of T;

$$R_l = \sum_{k=0}^{\infty} d_l \{(-T_l d_l)^k\}_{\text{n.r.}}. \tag{17}$$

The subscripts n.r. on the brackets mean that in products of t matrices, consecutive indices are unequal. (Except when needed for clarity, we will not write the argument l or the limits on sums and integrals.) The proof of Eq. (17) is as follows: Solve Eq. (15) for $v(\mu)$,

[9] N. M. Hugenholtz, Physica 23, 481 (1957).

[10] It should be noted that this definition for $P(t \mid pp_0)$ is not coarse-grained. Van Hove's work utilizes a coarse-grained probability; whereas in reference 5 this "fine-grained" probability is used.

$$v(\mu) = t(\mu)[1 - dt(\mu)]^{-1}, \tag{18}$$

and substitute this into Eq. (13) to obtain

$$R = d - d \sum_{\mu} t(\mu)R(\mu), \tag{19}$$

with $R(\mu)$ defined as

$$R(\mu) = [1 - dt(\mu)]^{-1}R. \tag{20}$$

Combining Eqs. (19) and (20), we find

$$R(\mu) = d - d \sum_{\lambda \neq \mu} t(\lambda)R(\lambda). \tag{21}$$

The desired result, Eq. (17), follows from Eqs. (16), (19), and (21).

The t-matrix expansion of R is adopted as a starting point. Using the representation furnished by the eigenstates of H_0, we write the resolvent as the sum of its diagonal and nondiagonal parts

$$R = R_d + R_{nd}, \tag{22}$$

d and nd denoting the diagonal and nondiagonal parts respectively. It is convenient to write Eq. (17) as

$$R = \sum d(-Td)^k = d - dTR, \tag{23}$$

and to leave implicit that when products of t matrices appear consecutive indices are not equal. The diagonal part of Eq. (23) is (defining $D \equiv R_d$)

$$D = d - d(TR)_d = d - dT_dD - d(T_{nd}R_{nd})_d, \tag{24}$$

and the nondiagonal part is

$$R_{nd} = -d(TR)_{nd} = -d(TD)_{nd} - d(TR_{nd})_{nd}. \tag{25}$$

A formal solution for R_{nd} in terms of D is obtained by iterating Eq. (25),

$$R_{nd} = [-dT_{nd} + (dT(dT)_{nd})_{nd} - (dT(dT(dT)_{nd})_{nd})_{nd} + \cdots]D,$$

which can be written more compactly as

$$R_{nd} = \sum_{k=1} \{(-dT)^k\}_{nd}D, \tag{26}$$

with $\{\cdots\}_{nd}$ defined as

$$\{ABCD \cdots\}_{nd} \equiv (A(B(C(D \cdots)_{nd})_{nd})_{nd})_{nd}. \tag{27}$$

Defining an operator G by

$$G = \sum_{k=1} d^{-1}\{(-dT)^k\}_{od} \tag{28}$$

with $\{\cdots\}_{od}$ defined as

$$\{ABCD \cdots\}_{od} = (A(B(C(D \cdots)_{nd})_{nd})_{nd})_d, \tag{29}$$

we obtain a formal solution for D by substituting Eq. (26) into Eq. (24),

$$D = d + dGD.$$

This can be solved for D to yield

$$D = (H_0 - G - l)^{-1}. \tag{30}$$

The operator R can be written

$$R = (I + DU)D \tag{31}$$

with I the unit operator and U defined by

$$DU = \sum_{k=1}^{\infty} \{(-dT)^k\}_{nd}. \tag{32}$$

Notice that G and U depend on the unperturbed resolvent d and the t matrix. The brackets $\{\cdots\}_{od}$ and $\{\cdots\}_{nd}$ have an obvious meaning if we take matrix elements of Eqs. (28) and (32)—they imply that no intermediate state is equal to the initial state, but intermediate states may be equal to other intermediate states. It is useful in some cases to perform sums of infinite series in Eqs. (28) and (32), which results in the further restriction that no intermediate states are equal. The resulting equations are

$$G = \sum_{j=0} \{-T(-DT)^j\}_{\text{i.d.}} = \{-T + TDT - TDTDT + \cdots\}_{\text{i.d.}}, \tag{33}$$

and

$$DU = \sum_{k=1} \{(-DT)^k\}_{\text{i.n.d.}} = D\{-T + TDT - TDTDT + \cdots\}_{\text{i.n.d.}}, \tag{34}$$

where G and U now depend on the t matrix and D, instead of d, and the brackets $\{\cdots\}_{\text{i.d.}}$ and $\{\cdots\}_{\text{i.n.d.}}$ imply that no intermediate state is equal to another intermediate state or the initial state when a matrix element is taken. In the terminology of Van Hove, i.d. stands for irreducible diagonal and i.n.d. for irreducible nondiagonal. These are essentially the results of Van Hove with the difference that here G and U are defined in terms of T rather than H_1. Before proceeding, it should be recalled that we have explicitly omitted the brackets $\{\cdots\}_{\text{n.r.}}$. In the following it is implicit that in products of t matrices with the same arguments, successive indices labeling the particles are kept unequal; i.e. in a product

$$t_l(\mu_1)t_l(\mu_2) \cdots t_l(\mu_m)t_{l'}(\lambda_1)t_{l'}(\lambda_2) \cdots t_{l'}(\lambda_n),$$

we require $\mu_i \neq \mu_{i+1}$ and $\lambda_i \neq \lambda_{i+1}$, but μ_m may equal λ_1.

F and W are defined as follows[5]: F is a difference of D's,

$$F_{ll'}(p) = D_l(p) - D_{l'}(p), \tag{35}$$

with $D_l(p) \equiv \langle p|D_l|\ p\rangle$, and W is defined by the integral equation

$$W_{ll'}(pp_0) = U_l(p_0p)U_{l'}(pp_0) - \sum_{p'} W_{ll'}(pp')D_l(p')D_{l'}(p')U_l(p_0p')U_{l'}(p'p_0), \tag{36}$$

with

$$U_l(p'p) \equiv \langle p'|\ U_l\ |p\rangle.$$

Equation (36) can be iterated to yield

$$W_{ll'}(pp_0) = U_l(p_0p)U_{l'}(pp_0) - \sum_{p'} U_l(p'p)U_{l'}(pp')D_l(p') \times D_{l'}(p')U_l(p_0p')U_{l'}(p'p_0) + \cdots, \tag{37}$$

from which it is evident that

$$W_{ll'}(pp_0) = W_{l'l}(p_0p). \tag{38}$$

In addition to this symmetry relation, some systems possess the more stringent condition called microscopic reversibility, which is

$$W_{ll'}(pp_0) = W_{ll'}(p_0p). \tag{39}$$

We note that any symmetry possessed by W implies the same symmetry for X. An expression for W in terms of D and T is obtained from Eqs. (34) and (37), which, after considerable algebraic simplification, reduces to

$$W_{ll'}(pp_0) = \{\langle p_0|\ (T_l - T_lD_lT_l + T_lD_lT_lD_lT_l - \cdots)\ |p\rangle \times \langle p|\ (T_{l'} - T_{l'}D_{l'}T_{l'} + T_{l'}D_{l'}T_{l'}D_{l'}T_{l'} - \cdots)\ |p_0\rangle\}_{\text{i.d.}}. \tag{40}$$

This is the desired expression for W in terms of T.

It is convenient to introduce a diagrammatic representation of Eq. (40). H_0 is the free-particle Hamiltonian

$$H_0 = \sum_i p_i^2/2m,$$

and we will work in the free-particle representation

$$|p\rangle = |\mathbf{p}_1\mathbf{p}_2 \cdots \mathbf{p}_N\rangle = \prod_i |\mathbf{p}_i\rangle, \tag{41}$$

with $|\mathbf{p}_i\rangle$ given by

$$|\mathbf{p}_i\rangle = V^{-\frac{1}{2}} \exp[i\mathbf{p}_i\cdot\mathbf{r}_i/\hbar], \tag{42}$$

which is normalized in a box of volume V with periodic boundary conditions,

$$\langle \mathbf{p}_i\ |\ \mathbf{p}_i'\rangle = \delta^{kr}(\mathbf{p}_i - \mathbf{p}_i').$$

The effects of quantum statistics will not be treated so we did not symmetrize or antisymmetrize the wavefunction. In the limit of infinite volume, the sums over momenta will become integrals and the Kronecker delta functions will become Dirac delta functions according to[9]

$$\sum_{\mathbf{p}_i} \Rightarrow V(2\pi\hbar)^{-3} \int d\mathbf{p}_i, \tag{43}$$

and

$$\delta^{kr}(\mathbf{a} - \mathbf{b}) \Rightarrow V^{-1}(2\pi\hbar)^3\delta^3(\mathbf{a} - \mathbf{b}). \tag{44}$$

A matrix element of the T matrix in this representation is

$$\langle p'|\ T_l\ |p\rangle = \sum_{i<j} \langle p'|\ t_l(ij)\ |p\rangle, \tag{45}$$

with $t_l(ij)$ given by Eq. (15). From this it is clear that

$$\langle p'|\ T\ |p\rangle = \sum_{i<j} \prod_{r\neq i,j} \delta^{kr}(\mathbf{p}_r' - \mathbf{p}_r) \times \langle \mathbf{p}_i'\mathbf{p}_j'|\ t(ij)\ |\mathbf{p}_i\mathbf{p}_j\rangle, \tag{46}$$

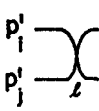
and assuming that $t(\mathbf{r}_i, \mathbf{r}_j)$ depends only on the magnitude of $\mathbf{r}_i - \mathbf{r}_j \equiv \mathbf{r}_{ij}$, we obtain

$$\langle \mathbf{p}_i'\mathbf{p}_j'|\ t(ij)\ |\mathbf{p}_i\mathbf{p}_j\rangle = (2V)^{-1}\delta^{kr}(\mathbf{P}_{ij} - \mathbf{P}_{ij}')t(|\mathbf{p}_{ij} - \mathbf{p}_{ij}'|), \tag{47}$$

where

$$t(|\mathbf{p}_{ij} - \mathbf{p}_{ij}'|) = \int d\mathbf{r}_{ij} \times \exp[i(\mathbf{p}_{ij} - \mathbf{p}_{ij}')\cdot\mathbf{r}_{ij}/2\hbar]t(r_{ij}).$$

It is possible to make the following one-to-one correspondence between matrix elements $\langle p'|\ t_l(ij)|\ p\rangle$ and diagrams:

$\langle p'|\ t_l(ij)|\ p\rangle \quad \Longleftrightarrow$  (48)

associating the l at the vertex with the l in t_l. Since only the states of particles i and j are altered by $t(ij)$, we indicate only these two particles in a diagram; the states of the other particles contribute delta functions of momenta. Associated with the basic diagram (48) is the following contribution to the matrix element of T:

$$(2V)^{-1} \prod_{r\neq i,j} \delta^{kr}(\mathbf{p}_r' - \mathbf{p}_r)\delta^{kr}(\mathbf{P}_{ij} - \mathbf{P}_{ij}') \times t_l(|\mathbf{p}_{ij} - \mathbf{p}_{ij}'|), \tag{49}$$

and Eq. (46) is obtained by summing over all i and j, $i < j$.

Let us write Eq. (40) for W as

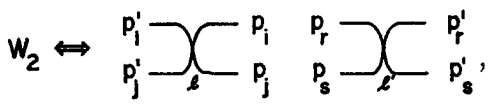
$$W_{ll'}(pp') = W_2 + W_3 + W_4 + \cdots + W_n + \cdots, \quad (50)$$

with

$$W_2 = \{\langle p'| T_l |p\rangle\langle p| T_{l'} |p'\rangle\}_{\text{i.d.}}, \quad (51)$$

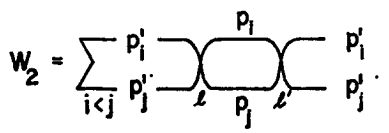
$$W_3 = -\{\langle p'| T_l D_l T_l |p\rangle\langle p| T_{l'} |p'\rangle + \langle p'| T_l |p\rangle\langle p| T_{l'} D_{l'} T_{l'} |p'\rangle\}_{\text{i.d.}}, \text{ etc.}, \quad (52)$$

where the subscript n on W_n indicates that it is the contribution to W from all terms with n T-matrices. The correspondence (48) will now be used to represent W_n by diagrams. For W_2 we obtain diagrams of the form



(53)

where we adopt the usual convention of ordering the vertices in the same order as they appear in the expression represented by the diagram. Since the initial state is equal to the final state, the only diagrams which are nonzero are those for $i = r$ and $j = s$, so we have

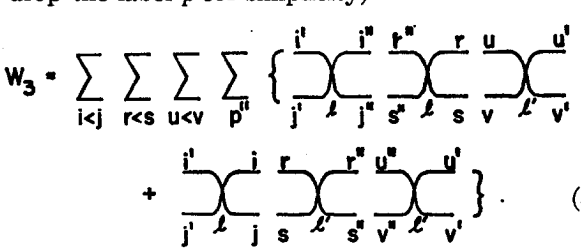


(54)

From Eqs. (48) and (49) we find (factors which will not contribute to the density dependence are omitted),

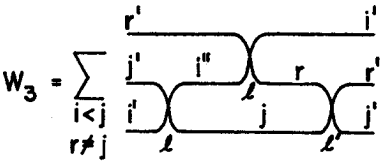
$$W_2 \approx V^{-2} \sum_{i<j} t_l(|\mathbf{p}_{ij} - \mathbf{p}_{ij}'|) t_{l'}(|\mathbf{p}_{ij} - \mathbf{p}_{ij}'|). \quad (55)$$

The diagrammatic representation of W_3 is (we will drop the label p for simplicity)



(56)

For $n > 2$, a large number of diagrams give vanishing contributions to W_n due to the conditions imposed by $\{\ \}_{\text{n.r.}}$, $\{\ \}_{\text{i.d.}}$, and the fact that the initial and final states are the same. The resulting nonzero diagrams for W_3 are

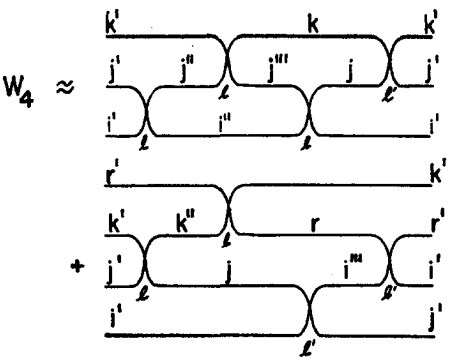


+ a similar diagram for the second term. (57)

This gives for W_3

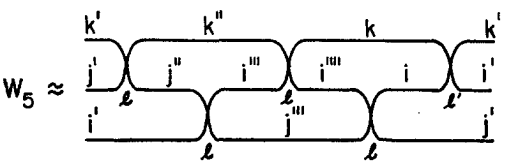
$$W_3 \approx -\sum_{\substack{i<j \\ r\neq j}} V^{-3} t_l(|\mathbf{p}_{ij}' - \mathbf{p}_{ij}''|) D_l(p'') \times t_l(|\mathbf{p}_{ir}'' - \mathbf{p}_{ir}|) t_{l'}(|\mathbf{p}_{rj} - \mathbf{p}_{rj}'|) + \text{similar term.} \quad (58)$$

Let us indicate the types of diagrams which are nonzero in higher order. For W_4 we obtain



+ other connected diagrams with three lines
+ other connected diagrams with four lines, (59)

and for W_5



+ other connected diagrams with three, four, and five lines, (60)

etc.

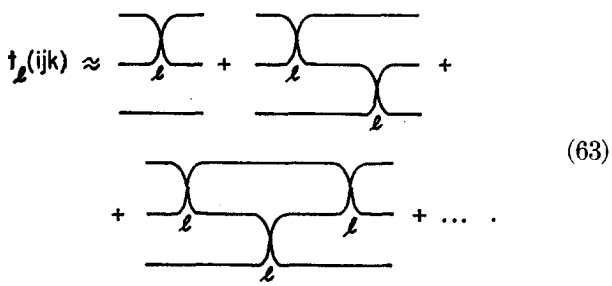
It is observed that there are diagrams with three lines contributing to W_n for all $n > 2$. Let us define ${}^3W$ to be the contribution to W from the sum of all these diagrams. This series for ${}^3W$ can be formally summed. In analogy with Eq. (15) we define a three-body t matrix by

$$t(ijk) = v(ij) + v(ik) + v(jk) - [v(ij) + v(ik) + v(jk)]\, dt(ijk), \quad (61)$$

which is the same as the three-body "collision matrix" defined by Resibois[6] except that d is the many-body free propagator rather than the three-body propagator. If we iterate Eq. (61) and compare it with Eq. (15), we observe that

$$t(ijk) = t(ij) + t(ik) + t(jk) - t(ij)dt(ik) - t(ij)dt(jk) - \cdots + t(ij)dt(jk)dt(ik) + \cdots ; \quad (62)$$

i.e., $t(ijk)$ can be expanded as a series of products of two-body t matrices. Equation (62) can be represented diagrammatically by

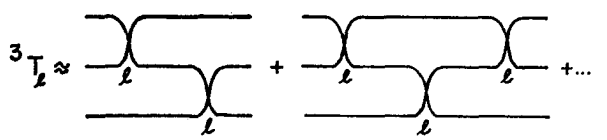


(63)

Let us define a three-body T-matrix by

$$^3T_l = \sum_{i<j<k} t_l(ijk) - \frac{1}{N-2} T_l, \quad (64)$$

where we have subtracted the two-body dependence, $(N-2)^{-1}T_l$. Then the diagram expansion of ^3T is



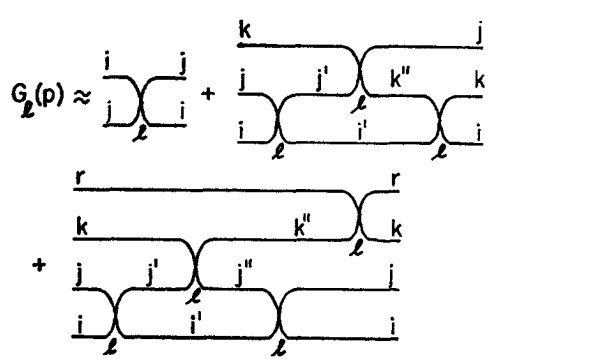
(65)

Comparing this with the diagrams contributing to ^3W, we see that

$$^3W_{ll'}(pp') = \{\langle p'|\, ^3T_l\, |p\rangle\langle p|\, T_{l'}\, |\, \acute{p}\rangle + \langle p'|\, T_l\, |p\rangle\langle p|\, ^3T_{l'}\, |p'\rangle + \langle p'|\, ^3T_l\, |p\rangle\langle p|\, ^3T_{l'}\, |p'\rangle\}_{\text{i.d.}}, \quad (66)$$

a result for the n-body problem which is analogous to that obtained by Resibois[6] for the three-body problem.

In a similar manner, we can express the operator G in terms of diagrams, as is clear from Eq. (33),



+ higher order connected diagrams with three lines

+ higher order connected diagrams with four, five, etc. lines. (67)

The density dependence will now be considered. It is not possible to obtain explicitly the density dependence of F and W, since the t matrix is a many-body operator through its dependence on d, and will in general depend on the density. In order to simplify this analysis, we will ignore this density dependence of t and assume that it can be replaced by a characteristic scattering length a; that is, we write[11]

$$t_l(|\mathbf{p}_{ij} - \mathbf{p}_{ij}'|) \approx a. \quad (68)$$

With this replacement, it follows immediately from Eq. (55) that

$$W_2 \approx c^2a^2, \quad (69)$$

so that W_2 is second order in the scattering length and at least second order in density. From Eq. (58) we obtain

$$W_3 \approx c^3a^3[D_l(p'') + D_{l'}(p'')], \quad (70)$$

thus W_3 is at least of order c^3 and it can contain higher order terms in general from the density dependence of $D(p'')$ and the t matrices. Similarly, one finds

$$W_4 \approx a^4[c^3 + c^4], \quad (71)$$

$$W_5 \approx a^5[c^3 + c^4 + c^5], \quad (72)$$

etc.,

where again we have not treated the density dependence of D and t. With the above simplifications and the definition of ^3W, it is observed that ^3W is the contribution to W from all the terms proportional to c^3.

The density dependence of D can be obtained from that of G [see Eq. (30)] and the density dependence of G can be obtained from Eq. (67). Again we make the same simplifications as above to find

$$G(p) \approx Nca[1 + ca^2 + c^2(a^3 + a^4 + \cdots) + c^3(a^4 + \cdots) + \cdots]. \quad (73)$$

The identity expressed by Eq. (11) and the t-matrix expansions for D and W will be used in a subsequent paper to investigate the evaluation of transport coefficients. The usefulness of this identity

[11] J. M. J. Van Leeuwen and A. S. Reiner, Physica **27**, 99 (1961).

is the following: The transport coefficient is directly related to X, which satisfies Eq. (11). D and W are expanded in terms of a quantity which is small for the problem under consideration, for example an interaction strength or the density, and the expansions are then terminated after a small number of terms. The resulting equation is no longer treated as an identity; it is considered to be an integral equation in which the kernel $\tilde{W}$ and the inhomogenity F are known functions and the equation is to be solved for X.

ACKNOWLEDGMENT

The author wishes to express his grateful thanks to Professor J. A. McLennan, Jr., for his many helpful suggestions and criticisms.

APPENDIX

In order to investigate systems in which quantum statistics are important, it is convenient to work in the second quantized formalism. Only minor changes of the preceding work are needed. The second quantized Hamiltonian is[9]

$$H = \sum_{\alpha} a_{\alpha}^{+} a_{\alpha} p_{\alpha}^{2}/2m + (4V)^{-1} \sum_{\alpha\nu\gamma\sigma} v(\alpha\nu\gamma\sigma) a_{\alpha}^{+} a_{\nu}^{+} a_{\gamma} a_{\sigma}, \tag{A1}$$

where a^{+} and a are creation and annihilation operators and $v(\alpha\nu\gamma\sigma)$ is the appropriate symmetrized potential. Let us define a T operator by

$$T_l = (4V)^{-1} \sum_{\alpha\nu\gamma\sigma} v(\alpha\nu\gamma\sigma) a_{\alpha}^{+} a_{\nu}^{+} a_{\gamma} a_{\sigma} - (4V)^{-2} \sum_{\alpha\nu\gamma\sigma\mu\lambda} v(\alpha\nu\gamma\sigma) v(\gamma\sigma\mu\lambda) a_{\alpha}^{+} a_{\nu}^{+} a_{\gamma} a_{\sigma} d_l a_{\gamma}^{+} a_{\sigma}^{+} a_{\mu} a_{\lambda} + \cdots . \tag{A2}$$

We can demonstrate that

$$R_l = \sum_{k} d_l \{(-T_l d_l)^k\}_{\text{n.r.}}, \tag{A3}$$

and that the analysis of the text remains essentially unaltered in second quantized formalism with the above definition of T.

Note added in proof: It has been objected that dropping the brackets $\{\cdots\}_{\text{n.r.}}$ makes the derivation obscure, and indeed some of the derived equations are incorrect unless properly interpreted. Consequently we add this explanatory note. By ignoring the brackets we count some terms twice; however we see that the error introduced by these "exceptional" terms is of order V^{-1} and thus is unimportant for a large system. These terms arise in a manner very similar to the "exceptional" terms which occur in Van Hove's derivation of a master equation.[4] In order to see this explicitly, let us introduce an operator η which is 0 or 1 and which takes the place of the brackets. We write Eq. (17) as

$$R = d - dT\, d\eta + dT\, d\eta T\, d\eta - \cdots = \sum_{k=0} d(-T\, d\eta)^k, \tag{a}$$

where $\eta = 1$ if it stands between two t matrices with different particle indices and also if it does not stand between two t matrices, and $\eta = 0$ if it stands between two t matrices with the same indices; i.e.,

$$T\eta T\eta = T\eta T = \sum_{\mu} t_{\mu}\eta \sum_{\nu} t_{\nu} = \sum_{\mu\neq\nu} t_{\mu}\eta t_{\nu} + \sum_{\mu} t_{\mu}\eta t_{\mu} = \sum_{\mu\neq\nu} t_{\mu} t_{\nu} + 0.$$

The derivations in the text can be performed in the same manner beginning with Eq. (a). One obtains, for example, in place of Eq. (28),

$$\bar{G} = \sum_{k=1} d^{-1}\{(-dT\eta)^k\}_{\sigma d}, \tag{b}$$

and in place of Eq. (30),

$$\bar{D} = (H_0 - \bar{G} - l)^{-1} = (H_0 - l + \{T - T\, d\eta T + \cdots\}_{\sigma d}\eta)^{-1}. \tag{c}$$

If one now compares (c) to Eq. (30), it is seen that forgetting the brackets in deriving Eq. (30) is equivalent to setting the last η in each term of $\bar{G}$ equal to 1; i.e.,

$$D = (H_0 - G - l)^{-1} = (H_0 - l + \{T - T\, d\eta T + \cdots\}_{\sigma d})^{-1}. \tag{d}$$

The difference between the correct expression $\bar{D}$ and the approximate expression D can be most easily seen by expanding each in a series in t; e.g. the second-order contribution to $\bar{D}$ is

$$\langle p|\, \bar{D}^{(2)}\, |p\rangle = \sum_{p'} \sum_{i<j} \langle p|\, t_{ij}\, |p'\rangle \langle p'|\, t_{ij}\, |p\rangle, \tag{e}$$

and the second-order contribution to D is

$$\langle p|\, D^{(2)}\, |p\rangle = \sum_{p'} \sum_{i<j} \langle p|\, t_{ij}\, |p'\rangle \langle p'|\, t_{ij}\, |p\rangle + \sum_{i<j} \langle p|\, t_{ij}\, |p\rangle \langle p|\, t_{ij}\, |p\rangle. \tag{f}$$

Thus it is seen that the error made by ignoring the bracket is to count the second term in Eq. (f) twice, but it is clear that in the limit of a large system the second term is of order V^{-1} with respect to the first term. If one investigates other derivations in the text, one observes that ignoring the brackets amounts

to counting some terms twice but that the error made is of order V^{-1} and is unimportant in the limit of a large system. Likewise, in going from Eqs. (28) and (32) to Eqs. (33) and (34), some exceptional terms are counted twice, and again this only becomes valid in the limit of a large system.

It should, perhaps, be pointed out that the master equation is meaningless in the limit of a large system, since if one investigates the volume dependence of W in detail it is found to contain arbitrary factors of the volume. It is only when one integrates over all but a few degrees of freedom that these "spurious" volume factors disappear and a meaningful large system limit can be taken. It is with this implicitly in mind that we discuss the master equation for a large system.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 4 APRIL 1963

Nonlinear Coupled Oscillators. I. Perturbation Theory; Ergodic Problem

E. ATLEE JACKSON*
Plasma Physics Laboratory, Princeton University, Princeton, New Jersey
(Received 6 August 1962)

A study is made of some of the problems which arise in determining the long-time behavior of a system of coupled oscillators. Standard perturbation methods are examined in the light of certain classic results due to Poincaré and Whittaker concerning the construction of constants of motion which are analytic in the coupling constant, λ. These considerations lead to the study of perturbation methods which are not ordered in powers of λ. An examination of the various advantages of these methods leads to a method which removes secular terms in such a way as to mitigate the classic problem of small divisors. The significance of studies which attempt to relate the existence of analytic constants of the motion with the ergodic behavior of a system is examined.

I. INTRODUCTION

IN this article we discuss, and attempt to resolve, some of the problems which arise when determining the "long-time" behavior of a system of nonlinearly coupled oscillators. The time scale of such a system can be conveniently measured in terms of the frequencies

$$\omega(m) = \sum_{k=1}^{N} m_k \omega_k,$$

where the m_k's are integers and the ω_k's are the frequencies of the N uncoupled oscillators. The frequencies $\omega(m)$ typically separate into a group of relatively large frequencies, including the set $\{\omega_k\}$, and a group of much smaller frequencies. This latter group contains frequencies which are arbitrarily small (provided the numbers $\{m_k\}$ are sufficiently large), or which vanish if some of the frequencies in $\{\omega_k\}$ are commensurable. These small frequencies commonly arise in standard perturbation methods developed in powers of the coupling constants λ, which thereby relate these frequencies to the actual long-time dynamics of the coupled system. As is well known, such methods suffer from the difficult problem of small divisors for they contain terms proportional to $\omega(m)^{-1}$ and these tend to zero in high orders. Quite apart from this question of convergence, these divisors also cause difficulties even in the lowest orders when one considers the limit of very large systems ($N \to \infty$). In this limit the frequencies $\{\omega_k\}$ can become dense, in which case all of the "small" frequencies $\omega(m)$ tend to zero and all terms containing small divisors become infinite. One way of overcoming this difficulty is to simultaneously take the limit of a vanishing coupling ($\lambda \to 0$) in such a way that $\lambda < \text{min. } \{\omega(m)\}$. As a formal device this may be satisfactory, but from the physical point of view it is quite unsatisfactory, unless, of course, the coupling actually vanishes in this fashion as $N \to \infty$. One of the problems we shall consider is how the behavior of coupled oscillators can be determined when λ is not vanishingly small, but N is sufficiently large for the above inequality to no longer hold.

A second, but much simpler problem which arises in these standard (time-dependent) perturbation

* Present Address; University of Illinois, Urbana, Illinois.

to counting some terms twice but that the error made is of order V^{-1} and is unimportant in the limit of a large system. Likewise, in going from Eqs. (28) and (32) to Eqs. (33) and (34), some exceptional terms are counted twice, and again this only becomes valid in the limit of a large system.

It should, perhaps, be pointed out that the master equation is meaningless in the limit of a large system, since if one investigates the volume dependence of W in detail it is found to contain arbitrary factors of the volume. It is only when one integrates over all but a few degrees of freedom that these "spurious" volume factors disappear and a meaningful large system limit can be taken. It is with this implicitly in mind that we discuss the master equation for a large system.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 4 APRIL 1963

Nonlinear Coupled Oscillators. I. Perturbation Theory; Ergodic Problem

E. ATLEE JACKSON*
Plasma Physics Laboratory, Princeton University, Princeton, New Jersey
(Received 6 August 1962)

A study is made of some of the problems which arise in determining the long-time behavior of a system of coupled oscillators. Standard perturbation methods are examined in the light of certain classic results due to Poincaré and Whittaker concerning the construction of constants of motion which are analytic in the coupling constant, λ. These considerations lead to the study of perturbation methods which are not ordered in powers of λ. An examination of the various advantages of these methods leads to a method which removes secular terms in such a way as to mitigate the classic problem of small divisors. The significance of studies which attempt to relate the existence of analytic constants of the motion with the ergodic behavior of a system is examined.

I. INTRODUCTION

IN this article we discuss, and attempt to resolve, some of the problems which arise when determining the "long-time" behavior of a system of nonlinearly coupled oscillators. The time scale of such a system can be conveniently measured in terms of the frequencies

$$\omega(m) = \sum_{k=1}^{N} m_k \omega_k,$$

where the m_k's are integers and the ω_k's are the frequencies of the N uncoupled oscillators. The frequencies $\omega(m)$ typically separate into a group of relatively large frequencies, including the set $\{\omega_k\}$, and a group of much smaller frequencies. This latter group contains frequencies which are arbitrarily small (provided the numbers $\{m_k\}$ are sufficiently large), or which vanish if some of the frequencies in $\{\omega_k\}$ are commensurable. These small frequencies commonly arise in standard perturbation methods developed in powers of the coupling constants λ, which thereby relate these frequencies to the actual long-time dynamics of the coupled system. As is well known, such methods suffer from the difficult problem of small divisors for they contain terms proportional to $\omega(m)^{-1}$ and these tend to zero in high orders. Quite apart from this question of convergence, these divisors also cause difficulties even in the lowest orders when one considers the limit of very large systems ($N \to \infty$). In this limit the frequencies $\{\omega_k\}$ can become dense, in which case all of the "small" frequencies $\omega(m)$ tend to zero and all terms containing small divisors become infinite. One way of overcoming this difficulty is to simultaneously take the limit of a vanishing coupling ($\lambda \to 0$) in such a way that $\lambda < \text{min. } \{\omega(m)\}$. As a formal device this may be satisfactory, but from the physical point of view it is quite unsatisfactory, unless, of course, the coupling actually vanishes in this fashion as $N \to \infty$. One of the problems we shall consider is how the behavior of coupled oscillators can be determined when λ is not vanishingly small, but N is sufficiently large for the above inequality to no longer hold.

A second, but much simpler problem which arises in these standard (time-dependent) perturbation

* Present Address; University of Illinois, Urbana, Illinois.

methods is the appearance of secular terms (terms proportional to some power of t). Again this problem can be handled by considering limits such that $\lambda \to 0$ as $t \to \infty$, usually in such a way as to make $\lambda^2 t$ constant (see below). In so doing, one can study the "long-time" behavior of a "sufficiently weakly" coupled system. This method is frequently,[1] but not always,[2] used to study the irreversible behavior of large, weakly coupled systems. A second method of treating such secular terms, which does not depend on vanishing coupling, is to use an averaging procedure apparently proposed originally by van der Pol[3] and developed considerably by Kryloff and Bogoliuboff[4] and others. This method removes the secular terms by shifting the frequencies $\{\omega_k\}$ to another set of frequencies $\{\Omega_k\}$. The procedure for determining the frequencies $\{\Omega_k\}$ takes a number of forms depending on the particular equations and solutions of interest[3,4] (e.g., nonautonomous equations, and periodic solutions). In the present study we will present methods which are applicable for the almost-periodic solutions of an autonomous system of equations.

One of the main objectives of the present study is to see if the concept of a frequency shift cannot be related in a simple way to the problem of small divisors. The hope is, therefore, not only to remove the secular terms which arise, but at the same time to do this in such a way as to prevent the denominators from becoming arbitrarily small in high order, or, of more pragmatic importance, in the limit of large N but finite coupling. While a method will be developed which appears to do this, no study will be made here of the convergence of the present scheme. The question of the accuracy of this method will be left to the second paper in this series in which some explicit comparisons will be made with numerical solutions of particular equations of motion.

To motivate our method, we will first discuss some classic results due to Poincaré and Whittaker concerning constants of the motion which are developed in powers of the coupling constant λ. Since all perturbation methods are concerned with the constants of the motion, these results are very relevant to perturbation methods developed in powers of λ. Following this we shall consider two perturbation schemes, neither of which are developed in powers of λ, and re-examine the problem of small divisors in each case. Finally, since systems of coupled oscillators have long been of interest in ergodic studies, we will discuss some aspects of this problem. This turns out to be a natural extension of perturbative studies since these are concerned with the constants of motion, and they in turn have often been erroneously applied to the ergodic problem.

II. CLASSIC RESULTS

The theorem of Poincaré,[5] referred to above, states essentially that if one has a Hamiltonian of the form

$$H = H_0(\mathbf{J}) + \lambda H_1(\mathbf{J}, \boldsymbol{\theta}), \tag{1}$$

which is periodic in all of the variables θ_i, and the Hessian of $H_0(J)$, $|\partial^2 H_0/\partial J_i\, \partial J_k|$, does not identically vanish, then there exists no analytic single-valued constant of the motion (aside from H) of the form

$$\Phi = \sum_{r=0}^{\infty} \lambda^r \phi_r(\mathbf{J}, \boldsymbol{\theta}), \tag{2}$$

where the functions ϕ_r are also periodic in θ_i. A crucial point in the proof of this theorem is that the generalized frequencies $\omega_k(J) = \partial H_0/\partial J_k$ are commensurable on a dense set of points of the energy surface, i.e. they satisfy

$$\omega(m) = \sum_{k=1}^{N} m_k \omega_k = 0 \tag{3}$$

for some nonvanishing sets of integers $\{m_k\}$. While this theorem is not directly applicable to a system of coupled harmonic oscillators (since in that case the Hessian of $H_0(J)$ vanishes), one has a very analagous result for such systems. For a system of coupled harmonic oscillators one can not find constants of the motion of the form (2) which are analytic in the frequencies ω_k (now independent of the J's). However, such constants can be constructed (at least formally) for any given set of ω_k's. A number of years ago Whittaker[6] constructed these constants (the so-called adelphic integrals) which exhibit their nonanalyticity by changing their entire functional form depending on whether or not (3) is satisfied (i.e. for some set $\{m_k\}$). Thus, the adelphic integrals have a very erratic behavior as the set $\{\omega_k\}$ is varied over any finite domain. It is well known that this same effect shows up in many perturbation schemes

[1] See, e.g., R. Brout and I. Prigogine, Physica **22**, 261 (1956).

[2] R. Zwanzig, *Lectures in Theoretical Physics*, edited by W. E. Brittin, B. W. Downs, and J. Downs (Interscience Publishers, Inc., New York, 1961), Vol. 3, p. 106.

[3] I. G. Malkin, *Some Problems in the Theory of Nonlinear Oscillations* (Department of Commerce, Washington, 1959), AEC Translation No. **AEC-tr-3766**, Vols. I and II.

[4] N. Kryloff and N. Bogoliuboff, *Introduction to Nonlinear Mechanics* (Princeton University Press, Princeton, New Jersey, 1947).

[5] H. Poincaré, Acta Math. **13**, 259, (1890); *Méthodes Nouvelles de la Mécanique Céleste* (Dover Publications, Inc., New York, 1957), Vol. 1, p. 233.

[6] E. T. Whittaker, *Analytical Dynamics* (Dover Publications, Inc., New York, 1944), p. 432.

where different methods must be applied depending on whether or not some of the $\{\omega_k\}$ are exactly commensurable. Since one knows that systems with slightly different sets $\{\omega_k\}$ must behave in a similar fashion for times of the order of the difference in these periods, this erratic behavior of the adelphic integrals would appear to be just a formal difficulty arising from the power series in λ. Whether or not this is true is difficult to say, but in any case it is an unwanted feature and so one tries to circumvent it.

In the time-dependent perturbation methods which are developed in powers of λ, the problem of commensurability shows up in the form of secular terms. In this case, one solves the system of equations

$$\ddot{a}_k = -\omega_k^2 a_k - \lambda[\partial H_1(a_1, \cdots, a_N)/\partial a_k] \qquad (k = 1, \cdots, N) \tag{4}$$

by iterating the resulting integral equations

$$a_k(t) = \alpha_k \cos(\omega_k t + \phi_k) - \frac{\lambda}{\omega_k}\int_0^t \sin \omega_k(t-\tau)\frac{\partial H_1(a(\tau))}{\partial a_k}\,d\tau. \tag{5}$$

This yields a series of functions which converge to the true solution at least for times of $O(\lambda^{-1})$, and probably for all times provided the energy surface is finite and H_1 satisfies a Lipschitz condition, etc. However, these solutions are in a form which hides the true long-time behavior of the system. If either (3) is satisfied or H_1 contains a part which is an even function of the a_k's, the first iteration contains secular terms of the form $\lambda t \sin \omega_k t$. In any case the second iteration contains the secular term $\lambda^2 t \sin \omega_k t$, which is the origin of the limiting procedure mentioned above. Such terms indicate that the system is "drifting" from its unperturbed motion (exhibiting an irreversible trend) for times of $O(\lambda^{-1})$ [or $O(\lambda^{-2})$ in the second case]. While this conclusion is correct, it is not of particular importance in determining the behavior of the system for times $t > O(\lambda^{-1})$ for one knows the system is almost periodic over sufficiently long periods.[7] To exhibit this almost periodicity one seeks to determine a set of shifted frequencies $\{\Omega_k(\lambda)\}$, which reduce to the set $\{\omega_k\}$ as $\lambda \to 0$, and which yield the secular terms if expanded in powers of λ [e.g., $\cos(\omega_k + \lambda\mu_k)t \simeq \cos \omega_k t - \lambda t \mu_k \sin \omega_k t$ for $t < O(\lambda^{-1})$]. As mentioned above, this is generally accomplished by a suitable averaging procedure which is relatively simple when applied to a single nonlinear equation. The difficulty presented by a system of equations is that this averaging procedure must be varied depending again on whether or not the set $\{\omega_k\}$ has commensurable members. Thus, one is again confronted with the erratic behavior found in the construction of the adelphic integrals, or degenerate vs nondegenerate perturbation theory.

It would appear that a likely cause of these difficulties stems from the use of the spectrum $\{\omega_k\}$ in (5) since, if the concept of shifted frequencies is taken seriously, the dynamics is determined by the set $\{\Omega_k\}$ rather than $\{\omega_k\}$. In particular, the rate or frequency of the energy exchange between oscillators is not determined by $\omega(m)$, but by

$$\Omega(m) = \sum_{k=1}^{N} m_k \Omega_k. \tag{6}$$

For the "large" frequencies this distinction is of no great importance since it is always assumed that the coupling is small enough for $\omega_k \simeq \Omega_k$ (in order for any perturbation theory to be applicable). However, for the small frequencies in $\{\omega(m)\}$, this distinction is very important, for if some $\omega(m)$ goes to zero in some limit (e.g., $N \to \infty$) the same need not be true of the corresponding $\Omega(m)$, provided λ is finite. Conversely, if N is fixed and λ is increased from zero, the smallest $\omega(m)$ may differ considerably from $\Omega(m)$. Since the long-time behavior of the system is governed by these small frequencies, we may expect the recurrence times of the system to be affected by λ in an important fashion. That this is the case will be shown explicitly in the second paper of this series.

Our objective now is to study methods of first introducing the unknown set $\{\Omega_k\}$ *ab initio* into a perturbation scheme, and subsequently determining this set. These methods lead naturally to a non-power series scheme (in λ), and cast the classical problem of small divisors into quite a different form.

III. PERTURBATION METHODS

In developing a perturbation method which incorporates the above considerations, we will proceed by first considering two methods, each of which have certain desirable features. On the other hand, each method will be found to have certain defects of either a basic or pragmatic nature. Using these results as a foundation we will present a method which includes the best properties of both of these methods and contains none of their defects. The accuracy of this method will be examined in the second paper of this series, where it will be applied to certain special systems whose behavior is determined by computor methods.

We shall first consider a method which is analogous to the iteration of the integral equations (5). To introduce the unknown frequencies $\{\Omega_k\}$ into the

[7] H. Poincaré, Acta Math. **13**, 1890.

integral equations we return to Eqs. (4) and define

$$\Omega_k^2 = \omega_k^2 + \sum_{r=1}^{\infty} \lambda^r \mu_k^{(r)}(\lambda), \tag{7}$$

where the function $\mu_k(\lambda)$ remains to be determined. Substituting (7) into (4), the resulting integral equations are

$$a_k(t) = \alpha_k \cos(\Omega_k t + \phi_k) - \frac{\lambda}{\Omega_k} \int_0^t \sin \Omega_k(t - \tau) F_k(\lambda, \mathbf{a}(\tau))\, d\tau, \tag{8}$$

where

$$F_k(\lambda, \mathbf{a}) = \frac{\partial H_1}{\partial a_k} - \sum_{r=1}^{\infty} \lambda^{r-1} \mu_k^{(r)} a_k, \tag{9}$$

and (α_k, ϕ_k) are determined both by the initial conditions and the unknown Ω_k. The procedure is to iterate (8) in the *explicit* powers of λ, selecting the terms $\mu_k^{(1)}$ so as to remove any secular term arising from $\partial H_1/\partial a_k$. The important difference in the present case is that, in contrast to the case where one iterates with known frequencies $\{\omega_k\}$, the set $\{\Omega_k\}$ is unknown. Consequently, one does not know *a priori*, whether a linear combination $\Omega(m) = \sum_k m_k \Omega_k$ vanishes or not, and therefore, whether a term in the iterated integral leads to a secular term. Because the present method only determines Ω_k *after* the iteration, one must make an assumption about the $\Omega(m)$ which arise in any particular order and justify this assumption *a posteriori*. The assumption we make is:

Assume that any $\Omega(m)$ ($\{m\} \neq \{0\}$) which appears in each iteration does not vanish. (10)

It is clear that this assumption is where we attempt to remove the "erratic problem" discussed in the last section, and therefore, it is of major importance. The perturbation scheme can only be considered valid if this assumption is verified by the resulting equations determining $\Omega(m)$. As we shall see, this is not easy to establish in all rigor, for all Hamiltonians, H_1, but it appears highly probable for certain methods.

With the assumption (10) it is not difficult to see that $\mu_k^{(1)}$ vanishes unless H_1 has an additive part which is even in its variables. These terms can be removed in this order by collecting any quadratic terms of H_1 with the unperturbed Hamiltonian and expressing the remainder of H_1 in a power series in λ [e.g., $H_1(\mathbf{a}) = \sum_{r=0} \lambda^r H_1^{(3+r)}(\mathbf{a})$, where $H_1^{(k)}$ contains the a_i's in kth order]. We can, therefore, quite generally take $\mu_k^{(1)}$ to be zero. Actually, this device which is frequently used is not of fundamental importance one way or the other, for in either case one finds that the $\Omega(m)$ which arises in this order may vanish. That is, Ω_k^2 is determined in this order by the equation

$$\Omega_k^2 = \omega_k^2 + \lambda \mu_k^{(1)}(\boldsymbol{\alpha}),$$

so that the combinations $\Omega(m)$ which appear in this order *may* vanish, depending the values of $\{\omega_k\}$, λ, and $\mu_k^{(1)}$. The fact that the $\Omega(m)$'s do not vanish for certain specific values of these quantities is not sufficient, for (10) is only useful in that it is a general statement, valid for all cases. We shall see, however, that the $\Omega(m)$ which arise in the first iteration are found to be nonzero if we use the next order results in determining the $\{\Omega_k\}$. In general, it will be found that the $\Omega(m)$ arising in any particular order can only be shown to be nonzero by determining these $\Omega(m)$ from the results of the following order.

To see that this is the case, consider the second iteration. The first iteration yields corrected expressions for the functions a_k, which contain coefficients proportional to $\Omega(n)^{-1}$, with certain definite sets of integers $\{n_k\}$ (a special case is illustrated below). Therefore, in the second order, the functions $\mu_k^{(2)}$ will be determined by equations of the form

$$\mu_k^{(2)} = \sum_{\{n\}} h_k(\mathbf{n}, \boldsymbol{\alpha}) \Omega(n)^{-1}, \tag{11}$$

where the integers $\{n\}$ in the sum are *all* those which appear in the first order. It is important to note that none of the $h_k(\mathbf{n}, \boldsymbol{\alpha})$ vanish if $\Omega(\mathbf{n})$ appeared in the first order, as can be shown without much difficulty. In this order the Ω_k's are determined by

$$\Omega_k^2 = \omega_k^2 + \lambda^2 \sum_{\{n\}} \Omega(\mathbf{n})^{-1} h_k(\mathbf{n}, \boldsymbol{\alpha}), \tag{12a}$$

or in other words,

$$\Omega(m) = \sum_{k=1}^{N} m_k [\omega_k^2 + \lambda^2 \sum_{\{n\}} \Omega(\mathbf{n})^{-1} h_k(\mathbf{n}, \boldsymbol{\alpha})]^{\frac{1}{2}}. \tag{12b}$$

It is generally assumed that λ is sufficiently small for the corrections to any ω_k [Eq. (12a)] to be small. Whether or not this is actually the case in the limit of large systems (finite λ) depends on what happens to the $\{\Omega(n)\}$ *and* the $h_k(n, \alpha)$ in this limit. The main advantage of the present method is the form of the Eq. (12b) which determines the frequencies $\Omega(n)$ in terms of an expression containing $\Omega(n)^{-1}$. Thus, if some $\omega(m)$ is very small, it does not follow that the corresponding $\Omega(m)$ is also small, provided λ is finite. An over-simplified example illustrates this point. If

$$\Omega(m) = \omega(m) + \lambda^2 [h/\Omega(m)],$$

then

$$\Omega(m) = \tfrac{1}{2}\omega(m) + \tfrac{1}{2}[\omega(m)^2 + 4\lambda h]^{\frac{1}{2}},$$

so that if $\lambda \to 0$, $\Omega(m) \to \omega(m)$, whereas, if $\omega(m) \to 0$, then $\Omega(m) \sim 0(\lambda^{\frac{1}{2}})$ for finite λ. In the case $\mu_k^{(1)}$ does

not vanish, $\Omega(m) \sim 0(\lambda^{\frac{1}{2}})$ if $\omega(m)$ tends to zero, and the expansion is then in the powers of $\lambda^{\frac{1}{2}}$ rather than λ.

The situation in Eq. (12b) is obviously not as simple as this example. Indeed it appears possible for two (or more) $\Omega(n)$ to approach zero in some limit, provided the corresponding $h_k(\mathbf{n}, \boldsymbol{\alpha})$ behave in an appropriate manner in that limit (for *all* k). Without a more detailed knowledge of the particular limiting procedure, and the functions $h_k(\mathbf{n}, \boldsymbol{\alpha})$, it does not seem possible to exclude this possibility. However, the structure of (12b) puts stringent requirements on *all* the functions $h_k(\mathbf{n}, \boldsymbol{\alpha})$ for $\Omega(n) = 0$ to be a solution. For this reason it seems very unlikely that the assumption (10) can be violated. It should be emphasized that this feature results from the particular structure of (12b), namely that $\Omega(\mathbf{n})$ is (roughly) proportional to $\Omega(\mathbf{n})^{-1}$. As we shall see presently, this is not a necessary feature of perturbation methods—even for those which are not developed in powers of λ.

Before considering other methods, it should be noted that the higher orders proceed in an analogous, if more complicated fashion. Since some of the $\omega(m)$ tend to zero in higher orders (because of the large values of $\{m_k\}$), one expects the corresponding $\Omega(m)$ to differ considerably from $\omega(m)$. It is clear that this considerably alters the standard problem of small divisors, for the divisors $\Omega(m)$ *may* now be bounded away from zero. The complexity of this problem unfortunately makes it extremely difficult to come to any definitive conclusion on this point. One can only say, however, that there is the possibility that all $\Omega(m) \geq \delta > 0$, which is obviously not the case for the frequencies $\omega(m)$.

We now consider a second method which has certain advantages over the previous method, but also has an important disadvantage. Nevertheless, it is useful to discuss this method for further insight into perturbation methods which are not developed in powers of λ. As usual, we shall assume that the functions $a_k(t)$ can be represented by the series

$$a_k(t) = \sum_{\{m\}} \beta_k(\mathbf{m}) \cos(\Omega(\mathbf{m})t + \phi(\mathbf{m})), \qquad (13)$$

where the sum runs over all positive and negative integers. We also assume that $\Omega(m) \neq 0$ unless $\{m_k\} \equiv \{0\}$, which is equivalent to the assumption (10). Under these conditions the functions

$$G_k(t, t') = \sum_{\{m\}} (\omega_k^2 - \Omega^2(m))^{-1} \times \cos(\Omega(m)t + \phi(m)) \cos(\Omega(m)t' + \phi(m))$$

have the well-known property

$$\left(\frac{d^2}{dt^2} + \omega_k^2\right) \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} G_k(t, t')F(t')\, dt' = F(t),$$

provided that $F(t)$ can be represented by a series of the form (13). Equation (4) can, therefore, be written in the form

$$a_k(t) = -\lim_{T\to\infty} \frac{\lambda}{2T} \int_{-T}^{T} G_k(t, t') \frac{\partial H_1(\mathbf{a}(t'))}{\partial a_k}\, dt'.$$

To put this in a form which is suitable for iteration, we define

$$\mathcal{G}_k(t, t') = G_k(t, t') - [2\cos(\Omega_k t + \phi_k)\cos(\Omega_k t' + \phi_k)]/(\omega_k^2 - \Omega_k^2) \qquad (14)$$

and use the fact that

$$-2\lambda \int_{-\infty}^{\infty} \cos(\Omega_k t + \phi_k) \frac{\partial H_1}{\partial a_k}\, dt = 2\int_{-\infty}^{\infty} \cos(\Omega_k t + \phi_k)\left(\frac{d^2}{dt^2} + \omega_k^2\right) a_k\, dt = \alpha_k(\omega_k^2 - \Omega_k^2),$$

where $\alpha_k = \beta_k(m_k = 1, m_i = 0)$, and the integrals are to be understood in the sense of the above limits. Combining these results we have

$$a_k(t) = \alpha_k \cos(\Omega_k t + \phi_k) - \lambda \int_{-\infty}^{\infty} \mathcal{G}_k(t, t') \frac{\partial H_1(\mathbf{a}(t'))}{\partial a_k}\, dt'; \qquad (15a)$$

$$\alpha_k(\omega_k^2 - \Omega_k^2) = -2\lambda \int_{-\infty}^{\infty} \cos(\Omega_k t' + \phi_k) \frac{\partial H_1(\mathbf{a}(t'))}{\partial a_k}\, dt'. \qquad (15b)$$

The present method will be recognized as a classical version of the Brillouin-Wigner method[7] in quantum mechanics. The analogy, however, is not complete, for in the present case the "trial functions" [i.e., $\alpha_k \cos(\Omega_k t + \phi_k)$] contain the *unknown* Ω_k, as does the kernel $\mathcal{G}_k(t, t')$. Furthermore, of course, there are no periodic boundary conditions relating the various Ω_k in the present case. These features could have an important effect on the apparently poor convergence of the usual Brillouin-Wigner method.[8]

Since both of these methods are not developed in powers of λ, but only in their respective explicit powers of λ, it is not surprising to find that they yield different results in any particular "order". In fact, the orders of the second method contain a portion of all orders of the first method, and vice versa. To illustrate this point, consider the perturbative Hamiltonian

[8] K. A. Brueckner, *Theory of Nuclear Structure in the Many Body Problem* (John Wiley & Sons, Inc., New York, 1959), p. 53.

$$H_1 = \tfrac{1}{3} \sum_{k,l,m} V_{klm} a_k a_l a_m \tag{16}$$

where V_{klm} is symmetric in its indices. This Hamiltonian will be considered in more detail in the second paper of this series. Using Eq. (8), one finds to first order

$$\begin{aligned} a_k &= \alpha_k \cos(\Omega_k t + \phi_k) \\ &+ \frac{\lambda}{4\Omega_k} \sum_{l,m} V_{klm} \alpha_l \alpha_m \mathcal{P}(\pm\Omega_l \pm \Omega_m - \Omega_k)^{-1} \\ &\times \{\cos[\pm(\Omega_l t + \phi_l) \pm (\Omega_m t + \phi_m)] \\ &- \cos(\Omega_k t \pm \phi_l \pm \phi_m)\}, \end{aligned} \tag{17}$$

where $\mathcal{P}$ stands for the sum of all permutations of the signs $\pm$ (the coefficients being correlated with the arguments of the trigonometric functions in an obvious manner). The first iteration of the second method, Eq. (15), yields

$$\begin{aligned} a_k &= \beta_k \cos(\Omega_k t + \psi_k) \\ &- \tfrac{1}{2}\lambda \sum_{l,m} V_{klm} \beta_l \beta_m \mathcal{P}[\omega_k^2 - (\Omega_l \pm \Omega_m)^2]^{-1} \\ &\times \cos[\Omega_l t + \psi_l \pm (\Omega_m t + \psi_m)], \end{aligned} \tag{18}$$

where we have replaced (α, ϕ) by (β, ψ) to distinguish them from the constants in (17). In both methods we find that $\omega_k = \Omega_k$ in this order because H_1 [Eq. (16)] is cubic.

If we now consider, by way of example, the case $\dot{a}_k(0) = 0$, $a_k(0) = A_k$, we find from (17) that $\alpha_k = A_k$, $\phi_k = 0$, whereas (β_k, ψ_k) in (18) can only be determined by solving N simultaneous nonlinear algebraic equations. If these equations are in turn solved only to first power in λ, one also obtains $\beta_k = A_k$ $\psi_k = 0$, but in general the results of (18) contain some of the information from all orders of the first method. This means that the second method has a considerable practical advantage over the first method for the iterations of (15a) are much simpler to perform than those of (8), particularly in higher orders. While some of this advantage is offset by the problem of determining the (β_k, ψ_k) in terms of the initial conditions, the latter is an algebraic problem and hence relatively easy.

The disadvantage of the second method is that the denominators in (18) are of the form $\omega_k^2 - (\Omega_l \pm \Omega_m)^2$ rather than $\Omega_k^2 - (\Omega_l \pm \Omega_m)^2$ as in (17) (after collecting some terms). The fact that (18) does not contain the divisors $\Omega(m)$ means that the equations determining the $\{\Omega_k\}$ [Eq. (15b)] will not have the inverse form found in (12). Thus, the argument used for justifying the basic assumption (10) no longer appears to apply. [In special cases, however, this argument may be unnecessary. Thus, for the *linear* case $H = \frac{1}{2}\Sigma\, \dot{a}_k^2 + \omega_k^2 a_k^2 - \lambda a_1 a_2$, the second method yields the correct result in second order, whereas the following method requires all orders. (The author is indebted to Dr. J. Ford for this example.)] In any case we shall see that this feature can be removed, while at the same time retaining the advantage mentioned above.

To accomplish this we note that the disadvantage of the second method comes from the divisors $(\omega_k^2 - \Omega(m)^2)$ in the Green's functions. The only way to alter these divisors is to alter the basic equations (4), as was done in the first method. We are therefore led to the application of the Green's function method to the equations

$$\ddot{a}_k + \Omega_k^2 a_k = -\lambda \frac{\partial H_1}{\partial a_k} + \sum_{n=1} \lambda^n \mu_k^{(n)} a_k. \tag{19}$$

We first require that the μ_k's satisfy

$$\int_{-\infty}^{\infty} \cos(\Omega_k t + \phi_k) \left(\frac{\partial H_1}{\partial a_k} - \sum_{n=1} \lambda^{n-1} \mu_k^{(n)} a_k \right) dt = 0. \tag{20}$$

This is the same as removing the secular terms in the first method, and replaces the condition (15b) in the second method.

Because of the requirement (20), we can now write (19) in the form

$$\begin{aligned} a_k(t) &= \alpha_k \cos(\Omega_k t + \phi_k) - \lambda \int_{-\infty}^{\infty} G_k(t, t') \Big[\frac{\partial H_1(t')}{\partial a_k} \\ &- \sum_{n=1} \lambda^{n-1} \mu_k^{(n)} a_k(t') \Big] dt', \end{aligned} \tag{21}$$

where

$$\begin{aligned} &G_k(t, t') \\ &= \sum_{[m]}{}' \frac{\cos(\Omega(m)t + \phi(m)) \cos(\Omega(m)t' + \phi(m))}{\Omega_k^2 - \Omega^2(m)}, \end{aligned} \tag{22}$$

and the prime signifies that the terms for which $\Omega(m) = \pm\Omega_k$ are excluded in the sum. The kernel (22) now contains no reference to the original frequencies $\{\omega_k\}$, which now only enter through the relationship (7). The first iteration of (16) again yields (18) with ω_k replaced by Ω_k, and all of the previous discussion of the practical advantages and the justification of (10) remain the same. This method will be applied to some specific cases in the second paper of this series.

IV. POINCARÉ'S THEOREM AND THE ERGODIC PROBLEM

Systems of nonlinear coupled oscillators have long been of interest in connection with the ergodic problem of statistical mechanics. Recently some studies have been made in an attempt to generalize a classic study of Fermi's so as to be applicable to such systems. Since these results are obviously

relevant to the long-time behavior of coupled oscillators, we shall examine their significance as regards this problem.

Fermi's[9] study of the ergodic problem was limited to those systems ("kanonische Normalsysteme") whose Hamiltonians satisfied the requirements of Poincare's theorem, discussed in Sec. I, and which therefore do not have constants of the motion of the form (2). By definition, a system is not ergodic if there are at least two invariant domains (under the Hamiltonian equations of motion) on the energy surface, both of finite measure. The boundary between these two domains was shown by Fermi to be determined by a constant of the motion $\Phi(J, \theta, \lambda)$. Now *if* one assumes that all such surfaces must be representable by a constant of the motion of the form (2) (as Fermi did in a footnote), then it follows that a kanonische Normalsysteme is ergodic. It is clear, however, that Fermi's theorem is considerably less general than it is frequently[10] assumed to be, for there are many (namely $2N - 1$) time-independent constants of the motion which separate the energy surface into invariant domains of finite measure.[11] One always acknowledges this fact to a limited extent by excluding in the above definition those domains of the energy surface which are inaccessible in virtue of the classical constants of the motion. The distinction between these constants of the motion and the remaining constants is presumably in our ability to control them ("controlable constants"[11]), that is, the remaining constants assume widely varying values in any small domain of the energy surface.[12] One might expect that the "uncontrollable constants," because of their erratic behavior, define invariant domains of a very filamentary nature, which therefore may still cover (to a large extent) the energy surface, yielding a "course-grained" ergodicity. *If* this is true then the controllable constants are the only ones of importance in determining the ergodic properties of a system. The theorem of Fermi's would then be more significant if, in addition, one could argue that the controllable constants must be of the form (2). It only requires a statement of these assumptions to show how tenuous is the theoretical understanding of the ergodic problem.

[9] E. Fermi, Z. Physik. **24**, 261, (1923).
[10] D. ter Haar, *Elements of Statistical Mechanics* (Rinehart and Company, New York, 1954), p. 358.
[11] A. I. Khinchin, *Statistical Mechanics* (Dover Publications, Inc., New York, 1949), Chap. III.
[12] In particular, Whittaker's adelphic integrals are almost certainly uncontrollable, for they contain arbitrarily small divisors whose coefficients depend on the region of the energy surface.

Recently Balescu[13] studied the problem of generalizing Poincaré's theorem to cover the case of a system of coupled harmonic oscillators. If such a theorem existed, one could then follow Fermi's argument to conjecture about the ergodicity of this system. Balescu's objective, however, appears to be impossible in view of the fact that Whittaker[2] constructed constants of the motion of the form (2) for the present system. In order to generalize Whittaker's discussion and clarify Balescu's result, we consider briefly the standard method[5,6,9,13] of constructing constants of the form (2). Starting with the Hamiltonian

$$H = H_0 + \lambda H_1 = \sum_{k=1}^{N} \omega_k J_k + \lambda \sum_{\{n\}} V_{\{n\}}(J_1, \cdots, J_N) \exp\left(i \sum_k n_k \theta_k\right), \quad (23)$$

we look for constants of the motion of the form

$$\Phi = \sum_{r=0}^{\infty} \lambda^r \Phi_r = \sum_r \sum_{\{n\}} \lambda^r \phi_{\{n\}}^{(r)}(J_1 \cdots J_N) \times \exp\left(i \sum_k n_k \theta_k\right). \quad (24)$$

The necessary and sufficient condition that Φ be a constant is that

$$[\Phi, H] = 0, \quad (25)$$

where [ ] is the usual Poisson bracket. To first order in λ, (25) yields

$$\phi_{\{n\}}^{(0)} \sum_{k=1}^{N} n_k \omega_k = 0,$$

so the only possible nonvanishing terms $\phi_{\{n\}}^{(0)}$ are those for which $\{n\} = \{0\}$ or $\pm\{m\}$, where the set $\{m\}$ satisfies (3). The second-order equations require that

$$[H_0, \Phi_1] + [H_1, \Phi_0] = 0. \quad (26)$$

In order to understand Balescu's result, we write this out in detail:

$$\sum_{\{n\}} \sum_k \omega_k n_k \phi_{\{n\}}^{(1)} \exp\left(i \sum n_k \theta_k\right) + \sum_{\{n\}} \sum_k \frac{\partial V_{\{n\}}}{\partial J_k} \left[m_k \phi_{\{m\}}^{(0)} \exp\left(i \sum (n_k + m_k)\theta_k\right) - m_k \phi_{\{m\}}^{(0)*} \exp\left(i \sum (n_k - m_k)\theta_k\right)\right] - \sum_{\substack{\{n\} \\ \{l\}}} \sum_k \frac{\partial \phi_{\{l\}}^{(0)}}{\partial J_k} n_k V_{\{n\}} \times \exp\left(i \sum (l_k + n_k)\theta_k\right) = 0. \quad (27)$$

Consider the term going as $\exp\left(i \sum m_k \theta_k\right)$:

[13] R. Balescu, Bull. Classe Sci. Acad. Roy. Belg. **42**, 622 (1956).

$$\phi^{(0)}_{\{m\}} \sum_k m_k \frac{\partial V_{\{0\}}}{\partial J_k} - \phi^{(0)*}_{\{m\}} \sum_k m_k \frac{\partial V_{\{2m\}}}{\partial J_k} - V_{\{m\}} \sum_k m_k \frac{\partial \phi^{(0)}_{\{0\}}}{\partial J_k} - V_{\{2m\}} \sum_k 2m_k \frac{\partial \phi^{(0)*}_{\{m\}}}{\partial J_k} = 0. \quad (28)$$

Equation (28) will not be satisfied [and therefore no constant of the motion of the form (24) can be constructed] if the following equations hold:

$$\phi^{(0)}_{\{m\}} = 0, \quad (29a)$$

$$V_{\{m\}} \neq 0, \quad (29b)$$

$$\sum_k m_k \frac{\partial \phi^{(0)}_{\{0\}}}{\partial J_k} \neq 0, \quad (29c)$$

$$\sum_k m_k \omega_k = 0 \qquad (\{m\} \neq \{0\}). \quad (29d)$$

This differs from the theorem due to Balescu only in the requirement (29a) which Balescu claimed is a consequence of his requirement that $[H_1, \Phi_0] \neq 0$. The point we wish to stress, however, is that the conditions (29) are quite different in nature from the type used by Poincaré, in that (29 a; c) are further restrictions on the class of constants of the motion, over and above the restrictions (24). These restrictions are not on the Hamiltonian, as in Poincaré's case, but limit further the allowable class of functions Φ. Because Whittaker does not make these restrictions, he can construct constants of the form (24), and this is the reason for the difference between the results of Balescu and Whittaker.

The considerations of this section show that only limited information about ergodicity can be obtained from a study of constructing constants of the motion of the form (2), for the question remains as to their relevance in this problem. In the case of coupled oscillators, one has, in addition to this, the question as to the effect of the conditions (29a) and (29c). Finally, on the basis of the previous discussions, it seems very unlikely that the precise satisfaction of (29d) should play any role in the course-grained ergodicity of systems with any finite λ.

V. CONCLUSION

In this paper we have discussed in some detail the problems which are found in various perturbation methods as applied to coupled oscillators. Much of the discussion has been on known difficulties and known methods of attack (e.g., shifted frequencies). An attempt has been made here to clarify and unify these points and then to investigate methods which may be capable of treating these problems. The study of these methods showed that it is necessary to make an assumption, (10), about the frequencies $\Omega(m)$ [Eq. (6)], the justification of which seems likely only for certain methods. This together with considerations of practical simplicity in applying the perturbation methods led to our final formulation (20)–(22). To shed further light on the system of coupled oscillators we discussed the relevance and limitations of both Fermi's and Balescu's criterions for the ergodic behavior of interacting systems. In a second paper we will study some specific systems, whose behavior is determined by computer solutions of the equations of motion. This will make it possible to check the present perturbation method for various values of λ, $\{\omega_k\}$, and different forms of V_{jkl} [in Eq. (16)].

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. N. Balazs for many fruitful discussions.

This work was accomplished under the auspices of the U. S. Atomic Energy Commission.

The author would also like to acknowledge the support of the National Science Foundation grant G-19926, at the University of Illinois, for the purpose of revising an earlier report on this research (Matt-95, August 1961, Princeton University).

Lyapunov's Stability Criteria for Plasmas

T. K. FOWLER
Oak Ridge National Laboratory, Oak Ridge, Tennessee
(Received 23 July 1962)

The orbit stability theory of Lyapunov has been adapted to the Vlasov–Boltzmann equation governing plasmas. Both linear and nonlinear stability are considered. The theory is characterized by a search for Lyapunov functions, whose existence implies stability in analogy with particles trapped in a potential well, as in the energy principle. The most important result is an existence theorem for Lyapunov functions quadratic in perturbations in all linearly stable cases in which perturbations damp asymptotically (sufficiently fast). As a corollary, without damping, the existence of a quadratic Lyapunov function is necessary and sufficient to prevent exponential growth of perturbations. A prescription is given for finding Lyapunov functions which are constants of motion. An example is treated. The implication of nonlinear stability from linear stability with damping is discussed, and Dr. C. S. Gardner's direct proof of nonlinear stability of a Maxwellian plasma by Lyapunov's method is reported.

1. INTRODUCTION

IN this paper, Lyapunov's classic theory of stability of finite-dimensional systems[1] is adapted to treat continuous media, in particular plasmas obeying the Vlasov–Boltzmann equation. The general results also apply to moment equations. The theory is characterized by a search for so-called Lyapunov functions, whose existence implies stability in analogy with particles trapped in potential wells. Let vectors $\psi(t)$ describe system perturbations and define stability as boundedness in time of some norm thereof, $||\psi(t)||$. A real, scalar functional, $V[\psi(t), t]$, is a Lyapunov function if, for all ψ and all $t > 0$, $V(\psi)$ is finite and $V(\psi, t) \geq \lambda \ ||\psi||$, some $\lambda > 0$, and if $dV/dt \leq 0$, all $t > 0$. Then $||\psi(t)|| \leq \lambda^{-1} V[\psi(t), t] \leq \lambda^{-1} V[\psi(0), 0]$; hence, stability. We may speak of *linear stability* or *nonlinear stability* depending on whether in computing dV/dt, the time behavior of $\psi(t)$ is determined from approximate, linearized equations, or from the exact, nonlinear equations.

A well known example of Lyapunov's method in the linear stability analysis of plasmas is the energy principle, where the energy perturbation is a Lyapunov function in the stable case.[2,3] Indeed, one interest in Lyapunov's theory is the prospect of finding extended "energy principles". As an example of a nonlinear application, in Appendix A, we report Gardner's proof of nonlinear stability of a Maxwellian plasma in which the free-energy perturbation is the Lyapunov function.

Even in the finite dimensional case, finding Lyapunov functions is largely a matter of luck. However, for finite dimensions there is at least an existence theorem for the linear theory, the existence of a Lyapunov function in the linear approximation being a necessary and sufficient condition for linear stability. The proof of a corresponding theorem for plasmas is perhaps our most important result. We can prove necessity only in the case of asymptotic stability, meaning that perturbations are not only bounded in time but eventually damp away, for example, as a consequence of collisions and particle leakage. In this case, the Lyapunov function damps steadily; $dV/dt < 0$. On the other hand, in the absence of damping, corresponding to frictionless particle motion, or to collisionless plasmas where the phase space distributions do not damp, it is known for finite dimensions that linear stability implies Lyapunov functions which are constants of motion. We discuss this case for plasmas, and obtain criteria for finding constants, but we have not been successful in showing rigorously that a constant must exist in the stable case.

Our most useful result for collisionless plasmas is a corollary of the case with damping. In order to apply the asymptotic stability theorem depending on damping, we simply introduce artificial damping by the transformation $u = \psi \exp(-\mu t)$ and apply our asymptotic stability theorem to the equation satisfied by the u's. Asymptotic stability of the u's guarantees that perturbations ψ of the original

[1] See J. LaSalle and S. Lefschetz, *Stability by Liapunov's Direct Method* (Academic Press Inc., New York, 1961); also F. R. Gantmacher, *The Theory of Matrices* (Chelsea Publishing Company, New York, 1959); Vol. I, pp. 125–129; Vol. II, pp. 117–125; 185–190. Lyapunov's name is transliterated Liapounoff, etc.

[2] I. B. Bernstein, E. A. Frieman, M. D. Kruskal, and R. M. Kulsrud, Proc. Roy. Soc. (London) **A244**, 17 (1958).

[3] M. D. Kruskal and C. R. Oberman, Phys. Fluids **1**, 275 (1958).

system grow slower than $\exp \mu t$. Taking μ small but finite yields, for all practical purposes, a criterion for linear stability of collisionless plasmas against exponential growth of perturbations. Slower growth like powers of t is left open to question.

In addition to existence theorems for Lyapunov functions, another of Lyapunov's results which we have attempted to extend to plasmas is the theorem that, with damping, linear stability implies nonlinear stability. However, we have not achieved rigor, and in any case, the role of damping limits practical usefulness. Lyapunov's nonlinear criterion is derived from the requirement that the greatest rate of growth due to nonlinear terms be less than the slowest rate of damping in the linear approximation. Since the nonlinear growth rate depends on the perturbation itself, the criterion is satisfied only for perturbations of limited magnitude. For slow collision damping and Landau damping, the limit is indeed small. However, the limit is at least finite, and for plasmas, the criterion is surely pessimistic. Further, within the predicted range of stability the plasma is exceedingly stable. Even any temporary growth of nonlinearly stable perturbations is at most comparable to what it would be in the linear approximation.

One aspect of Lyapunov's theory is not touched upon. We restrict ourselves to static equilibria, whereas he studied also the stability of time-dependent equilibria by first transforming to a static problem having the same stability character. For periodic time dependence, the transformation always exists in the finite-dimensional case.[4]

The paper is organized as follows: After defining symbols and terms in Sec. 2, we first prove, in Sec. 3, adaptations of Lyapunov's linear stability theorems in the abstract by methods applicable to partial differentio-integral equations. In Sec. 4, these theorems are applied to plasmas. Nonlinear stability is discussed in Sec. 5. In Appendix A, we record Gardner's proof of nonlinear stability of the Maxwell distribution, with or without collisions.

2. NOTATION AND DEFINITIONS

We wish to study the stability of static solutions of

$$\partial \Psi / \partial t = F(\Psi). \tag{1}$$

In Lyapunov's original problem, Ψ would be a column vector of numbers, say the phase space coordinates of a particle, and $F(\Psi)$ would be an algebraic function of these coordinates. Here, Ψ might be an element of a vector field whose components are functions of a position vector $\mathbf{x}$, and F might be a combination of integral and differential operations in $\mathbf{x}$ space. Whenever necessary, we assume that, for physical problems, acceptable solutions exist and are unique and continuous on $t > 0$.

Let Ψ_0, independent of t, be the equilibrium solution whose stability is under study. Writing a general solution as $\Psi = \Psi_0 + \psi$, the equation for the perturbation, ψ, becomes

$$\partial \psi / \partial t = P\psi + Q(\psi). \tag{2}$$

On the right, we have separated out the linear part of $F(\Psi_0 + \psi)$. Thus P is a linear operator; Q nonlinear. Both P and Q are assumed to be independent of t, implying that F is also. It is sufficient for our purposes to suppose Q to be quadratic in ψ.

We shall also be interested in the linearized equation

$$\partial \varphi / \partial t = P\varphi. \tag{3}$$

For convenience, we have used a special notation φ for solutions of the linearized equation. Throughout the paper, we employ the notation φ whenever the linear approximation is specifically intended.

For easy reference, we list as a group the following standard and special definitions, denoted hereafter by number.

(D.1) *Solutions* $\psi(t)$, $\varphi(t)$: We denote by $\psi(t)$ and $\varphi(t)$ solutions of (2) and (3), respectively, with initial conditions $\psi(0) = \psi$ and $\varphi(0) = \varphi$. For brevity, we write $\psi(t)$ simply as ψ when no confusion is apt to arise.

(D.2) *Scalar product*: (ψ_1, ψ_2) denotes a Hermitian scalar product between vectors ψ_1 and ψ_2. (ψ, ψ) is positive definite, $(\psi_1, \psi_2) = (\psi_2, \psi_1)^*$, etc., where $(^*)$ denotes the complex conjugate.

(D.3) *Norms*: A norm $||\psi||$ of vector ψ is a positive-definite scalar functional satisfying the triangle inequality $||\psi_1 + \psi_2|| \leq ||\psi_1|| + ||\psi_2||$. The norm of operator A on a set S is the least upper bound of $||A(\psi)||/||\psi||$, all $\psi \in S$. The norm of greatest interest is $||\psi|| = (\psi, \psi)^{\frac{1}{2}}$.

(D.4) *Perturbations* S: Let S denote the set of admissible perturbations. If $\psi \in S$, then $c\psi \in S$, any constant c. For the chosen norm, $||\psi||$ is finite if $\psi \in S$. Initial values of both ψ and φ are to be chosen from S. Further, we make the important assumption that Eqs. (2) and (3) map S into itself. As we demonstrate by example in Sec. 4, in practice S is chosen as the largest class of perturbations for which we can prove simultaneously the validity of this assumption, and stability by means of theorems of Sec. 3.

[4] F. R. Gantmacher, reference 1, Vol. II, p. 119.

(D.5) *Stability*: We shall define stability in terms of norms. Given a choice of norm, Ψ_0 is stable if for every $\epsilon > 0$ there exists $\delta(\epsilon) > 0$ such that $||\psi(t)|| \leq \epsilon$, all $t > 0$, all $\psi(0) \in S$ such that $||\psi(0)|| \leq \delta(\epsilon)$.

(D.6) *Lyapunov condition*: See Eq. (27).

(D.7) *Hermiticity*: $A^\dagger$ is the Hermitian conjugate of linear operator A if $(\psi_1, A\psi_2) = (\psi_2, A^\dagger\psi_1)^*$. A is Hermitian if $A = A^\dagger$, anti-Hermitian if $A = -A^\dagger$.

(D.8) *Finite operator*: Linear operator A is finite on S if $(\psi, A\psi)$ is finite, all $\psi \in S$.

We define the following additional properties of a linear operator. Define the ratio $R = (\psi, A\psi)/(\psi, \psi)$. We say that operator A is:

(D.9) *Bounded*, if $\lambda > R > \delta$, some finite λ and δ, all $\psi \in S$;

(D.10) *Positive definite*, if $R \geq \delta$, all $\psi \in S$, some $\delta > 0$; *semi-definite*, if $\delta = 0$; *negative* (semi) definite if $(-A)$ is positive (semi) definite;

(D.11) *Positive nonzero*, if $R > 0$, all $\psi \in S$.

(D.12) *Operators* ξ: Let ξ denote the set of all finite, Hermitian, positive nonzero linear operators on S which are independent of t.

3. LINEAR STABILITY THEOREMS

In this section, we define stability by (D.5) with norm $||\varphi|| = (\varphi, \varphi)^{\frac{1}{2}}$. In the linear approximation, (D.5) is satisfied if (φ, φ) is merely bounded on $t \geq 0$. When normal modes exist, the usual linear criterion that they be bounded is also satisfied. If $\varphi(t) \propto \exp \lambda t$, $(\varphi(t), \varphi(t)) = (\varphi(0), \varphi(0)) \exp (2 \text{ Re } \lambda t)$, whence nonzero (φ, φ) is bounded on positive time if and only if Re $\lambda \leq 0$. We defer to Sec. 4 a discussion of the physical significance of the choice of product for the plasma problem.

The natural Lyapunov functions associated with the scalar product norm are Hermitian forms. Our most important new results are existence theorems (3 and 4) for Lyapunov functions of this type. First, we give sufficient conditions for stability. Note the crucial use of the assumption that (3) maps S into itself. We shall return to this vital point in discussing applications.

Theorem 1. *$(\varphi(t), \varphi(t))$ is bounded on $t \geq 0$, all $\varphi \in S$, if there exists a positive definite $H_+ \in \xi$ such that $[H_+P + P^\dagger H_+]$ is negative semidefinite on S.*

The set ξ was defined by (D.12). We have added the property of definiteness.

We shall show that $(\varphi, H_+\varphi)$ is a Lyapunov function, described in the introduction. It is finite and definite by assumed properties of H_+. Differentiating,

$$\frac{d}{dt}(\varphi, H_+\varphi) = (\varphi, [H_+P + P^\dagger H_+]\varphi). \tag{4}$$

If H_+ satisfies the theorem, and if $\varphi(t) \in S$, all $t \geq 0$, the right side is always negative or zero. Hence, $(\varphi, H_+\varphi)$ is monotone decreasing on $t \geq 0$, and is a Lyapunov function. As in the introduction, by definiteness of H_+, there exists $\delta > 0$ such that $(\varphi, \varphi) \leq \delta^{-1}(\varphi, H_+\varphi)$, whence (φ, φ) is bounded because $(\varphi, H_+\varphi)$ is. Note, by the way, that the existence of H_+ also guarantees uniqueness of solutions of (3) by a standard proof. If φ_1 and φ_2 are solutions, $\varphi_1 - \varphi_2$ is also. Let $\varphi_1 = \varphi_2$ at $t = 0$. Then $(\varphi_1 - \varphi_2, H_+(\varphi_1 - \varphi_2))$, initially zero, remains zero, which, by definiteness of H_+, implies that $\varphi_1 = \varphi_2$, all $t > 0$.

The above theorem provides a convenient test to determine whether $(\varphi, H_+\varphi)$ is a Lyapunov function. The question arises whether the test is ever satisfied. We examine first the case with the right side of (4) equal to zero, whence $(\varphi, H_+\varphi)$ is a constant of motion. By analogy with frictionless particle motion we might expect that constant Lyapunov functions must exist for stable, collisionless plasmas. Though an example is given in Sec. 4, a general existence theorem for this case is lacking.[5] We do show, however, in the next theorem, that the test of Theorem 1 applies if constants exist.

Theorem 2. *If H is Hermitian and independent of t, $(\varphi(t), H\varphi(t))$ is constant if and only if $[HP + P^\dagger H] = 0$ on S.*

Sufficiency follows from (4) if (3) maps S into S. Conversely, if the right side of (4) is zero, all $\varphi \in S$, one can show that $[HP + P^\dagger H] = 0$ acting on S; hence, necessity. Note that H need not be definite. This theorem generalizes the well known quantum mechanical rule for finding constants of motion. Multiplying (3) by i to obtain the Schrödinger equation, the Hamiltonian iP being Hermitian, the theorem reduces to the requirement that H and iP commute. Some variations of theorem 2 were given in reference 8.

Existence of a constant Lyapunov function only shows that perturbations are bounded. We consider now the case in which, measured in the chosen norm, perturbations damp away. For example, even for

---

[5] We can show, in analogy with constructions (5) and (7), there exist constants of the form $(\varphi, H_+\varphi)$, $H_+ \in \xi$, if and only if the time average

$$\lim_{T\to\infty} T^{-1}\int_t^{T+t} dt'(\varphi(t'), B\varphi(t'))$$

exists and is nonzero, some bounded $B \in \xi$. However, though true in finite dimensions, a general proof that the existence of such an average is necessary for stability without damping is lacking.

collisionless plasmas, in a norm $||\varphi|| = (\int d\mathbf{x}\, E^2)^{\frac{1}{2}}$, $\mathbf{E}$ being the electric field perturbation, all finite $||\varphi||$, damp asymptotically if the plasma is stable and exhibits Landau damping. With damping, we obtain an existence theorem for Lyapunov functions.

Suppose $\int_0^\infty dt\ (\varphi, \varphi) < \infty$, all $\varphi \in S$. Let B be bounded and $B \in \xi$, $B = 1$ being an example. Define[6]

$$I(t) = \int_t^\infty dt'\ (\varphi(t'), B\varphi(t')). \tag{5}$$

If $B = 1$, $I(t)$ converges by assumption. By comparison with this case, $I(t)$ converges for all bounded $B \in \xi$. $I(t)$ is positive, since B is positive nonzero, and monotone decreasing, since $dI/dt = -(\varphi, B\varphi)$. Now, write $\varphi(t)$ as $\varphi(t) = T(t)\ \varphi(0)$, $T(t)$ being the linear solution operator for (3) with the property $T(t + \Delta) = T(t)\ T(\Delta)$. Rewrite $I(t)$ as

$$\int_t^\infty dt'\ (T(t' - t)\varphi(t), BT(t' - t)\varphi(t))$$

$$= \left(\varphi(t), \left[\int_0^\infty ds\ T^\dagger(s)BT(s)\right]\varphi(t)\right). \tag{6}$$

In the second step, we made the variable change $s = t' - t$, we brought T on the left over as its Hermitian conjugate, and we interchanged the order of integrating on s and taking the scalar product. We shall denote the operator in brackets by

$$H_+ = \int_0^\infty ds\ T^\dagger(s)BT(s). \tag{7}$$

Then, $I(t) = (\varphi(t), H_+\varphi(t))$.

By its form, H_+ is linear, Hermitian and independent of t. By properties of $I(t)$, it is finite and at least positive nonzero, whence $H_+ \in \xi$, and $(\varphi(t), H_+\varphi(t))$ is monotone decreasing. $(\varphi, H_+\varphi)$ fails in being a Lyapunov function only in that H_+, though positive nonzero, is not necessarily definite. Definiteness, defined by $I(t)/(\varphi(t), \varphi(t)) \geq \delta > 0$, fails if perturbations damp arbitrarily fast, just the case, for example, with Landau damping of plane waves of arbitrarily short wavelength. As we shall see from the next theorem, the difficulty is overcome if we broaden slightly our concept of Lyapunov functions. Alternatively, in practice it may turn out that, for one of the H_+, $||\varphi|| = (\varphi, H_+\varphi)^{\frac{1}{2}}$ is a suitable norm, in which case the norm itself is a Lyapunov function of the usual type.

Theorem 3. *$\varphi(t)$ is bounded on $t \geq 0$ and $\int_0^\infty dt(\varphi(t), \varphi(t)) < \infty$, all $\varphi \in S$, if and only if, for some $H_+ \in \xi$, $[H_+P + P^\dagger H_+]$ is negative definite on S.*

Necessity is proved on identifying H_+ with (7) if we show that $[H_+P + P^\dagger H_+]$ is negative definite. Since $I(t) = (\varphi, H_+\varphi)$, dI/dt is given by (4), but also $dI/dt = -(\varphi, B\varphi)$. Equating these expressions and using boundedness of B yields the negative definiteness condition, $(\varphi, [H_+P + P^\dagger H_+]\varphi) \leq -\delta(\varphi, \varphi)$, some $\delta > 0$. To prove sufficiency, we use negative definiteness in (4) to yield

$$(d/dt)(\varphi, H_+\varphi) \leq -\delta(\varphi, \varphi), \tag{8}$$

valid for all $t \geq 0$ if (3) maps $S \to S$. Integrating (8),[7]

$$(\varphi(t), H_+\varphi(t)) - (\varphi(0), H_+\varphi(0))$$

$$\leq -\delta \int_0^t dt'\ (\varphi(t'), \varphi(t')). \tag{9}$$

Since all quantities are positive, it follows that

$$\int_0^\infty dt\ (\varphi, \varphi) \leq \delta^{-1}(\varphi(0), H_+\varphi(0)),$$

which is finite by finiteness of H_+. $(\varphi(t), \varphi(t))$ is also bounded, all $t \geq 0$. By (4), $(\varphi, H_+\varphi)$ is monotone decreasing, whence (φ, φ) is at least finite because H_+ is nonzero. Since $(\varphi(t), \varphi(t))$ is everywhere finite and integrable, it must be bounded on $t \geq 0$.

We have shown that, with damping, there exists a whole class of H_+ satisfying the theorem, being the class generated by the bounded $B \in \xi$. For each such H_+, $(\varphi, H_+\varphi)$ is a Lyapunov function in a modified sense. Like a Lyapunov function, $(\varphi, H_+\varphi)$ is a monotone, positive test function whose properties imply properties of (φ, φ). In Lyapunov's original method, the property of H_+ is definiteness, implying boundedness of (φ, φ). Here, the property of H_+ is its rate of decrease, expressed by (8), which implies boundedness and also integrability of (φ, φ).

While, as we noted above, in certain norms the preceding theorem can in principle be applied to Landau-damped collisionless plasmas, in norms discussed in the next section, chosen to bound perturbations of the phase space distribution as well as field disturbances, $||\varphi||$ is not expected to damp. To treat this case by theorem 3, we make the transformation $u(t) = \exp(-\mu t)\ \varphi(t)$, $\mu > 0$, where $u \in S$ if φ does. The following useful corollary is obtained by applying theorem 3 to the equation satisfied by the u's, just $(\partial u/\partial t) = (P - \mu)u$.

Theorem 4. *If $\mu > 0$, $\int_0^\infty dt(\varphi, \varphi) \exp(-2\mu t) < \infty$, all $\varphi \in S$, if and only if, for some $H_+ \in \xi$, $[H_+P + P^\dagger H_+ - 2\mu H_+]$ is negative definite on S.*

[6] For this approach, I am indebted to Dr. C. S. Gardner, whose suggestion that constants of motion be constructed from time averages (reference 5) led naturally to construction (5).

[7] This step of the proof was suggested by Dr. C. O. Beasley.

If (u, u) is integrable, (φ, φ) grows slower than $\exp(2\mu t)$. Thus, by choosing μ small but finite, for practical purposes theorem 4 provides a criterion for stability against exponential growth. The theorem always applies, since, if $(\varphi, \varphi) \exp(-2\mu t)$ is at least bounded with a given choice of μ, it is integrable if μ is increased slightly.

4. LINEAR STABILITY OF PLASMAS

We turn now to the application of the preceeding general theorems to plasma problems. We assume plasma dynamics to be governed by the Vlasov-Boltzmann equation, which we now present in convenient notation. Let the time-independent equilibrium state whose stability is to be examined be described by phase space distributions, $f_0^{(i)}(\mathbf{x}, \mathbf{v})$, one for each particle species, i, and by static electric and magnetic fields, $\mathbf{E}_0$ and $\mathbf{B}_0$. The latter are meant to include fields of external origin as well as plasma fields. Let distribution perturbations be f_i; field perturbations, $\mathbf{E}$ and $\mathbf{B}$. The time-evolution of each f_i is governed by

$$\frac{\partial f_i}{\partial t} + L_i f_i + \frac{q_i}{m_i}\frac{\partial f_0^{(i)}}{\partial \mathbf{v}}\cdot(\mathbf{E} + \mathbf{v}\times\mathbf{B}) = \frac{-q_i}{m_i}\frac{\partial f_i}{\partial \mathbf{v}}\cdot(\mathbf{E} + \mathbf{v}\times\mathbf{B}) + \left(\frac{\partial f_i}{\partial t}\right)_{\text{coll}}. \tag{10}$$

Here q_i is the charge and m_i the mass of particles of species i. A collision term has been retained on the right, along with a quadratic term to be discarded on linearization. The operator L_i is

$$L_i = \mathbf{v}\cdot\frac{\partial}{\partial \mathbf{x}} + \frac{1}{m_i}\left(-\frac{\partial \pi_i}{\partial \mathbf{x}} + q_i \mathbf{v}\times\mathbf{B}_0\right)\cdot\frac{\partial}{\partial \mathbf{v}}. \tag{11}$$

The potential $\pi_i(x)$ gives rise to $q_i\mathbf{E}_0$ as well as any externally applied static forces, such as gravitation. L_i governs equilibrium; $L_i f_0^{(i)} = [\partial f_0^{(i)}/\partial t]_{\text{coll}}$. Other embellishments of Eq. (10) might include the addition of source and loss terms.

The field perturbations, $\mathbf{E}$ and $\mathbf{B}$, satisfy Maxwell's equations with charge source $\sum_i q_i \int d\mathbf{v}\ f_i$ and current source $\sum_i q_i \int d\mathbf{v}\ \mathbf{v} f_i$. With properly charge-conserving collision, source, and loss terms, (10) is consistent with these equations in that multiplying (10) by q_i, integrating over $\mathbf{v}$ and summing over species yields the continuity equation.

A simplification arises if $\mathbf{B}$ is neglected or, less drastic, if retardation is neglected in wave propagation. In both cases, instantaneous fields are given as integrals over the instantaneous distributions, and the f_i afford a complete description of the plasma. Then ψ is just a vector with the f_i as components. The plasma analogue of Eq. (2) is obtained by substituting into (10) the solution of Maxwell's equations for $\mathbf{E}$ and $\mathbf{B}$ in terms of the f_i. The left-hand side of (10), together with a linear contribution from the collision term, gives rise to $P\psi$. The remainder gives $Q(\psi)$. The explicit form of the operator P neglecting both collisions and $\mathbf{B}$ was given previously, as was a form retaining $\mathbf{B}$ but neglecting retardation.[8]

Here we shall treat the more general case including retardation effects. The instantaneous f_i no longer constitute a complete plasma description, since with retardation the fields depend on the past history of the particles. The most convenient description seems to be in terms of the f_i plus $\mathbf{E}$ and $\mathbf{B}$. Thus the field components are to be added to the components of ψ, yielding altogether $6 + n$ components if there are n charge species. As dynamical equations, we take Eq. (10) plus the two curl equations of Maxwell,

$$\frac{\partial \mathbf{B}}{\partial t} = -\operatorname{curl}\mathbf{E} \tag{12}$$

$$\frac{\partial \mathbf{E}}{\partial t} = \operatorname{curl}\mathbf{B} - \sum_i q_i \int d\mathbf{v}\ \mathbf{v} f_i. \tag{13}$$

This set of equations contains only first derivatives in time and hence can be put in the form of Eq. (2) to which Lyapunov's theory applies. The description contains two redundant variables in that $\mathbf{E}$ and $\mathbf{B}$ are also subject to Maxwell's divergence equations as constraints,

$$\nabla\cdot\mathbf{E} = \sum_i q_i \int d\mathbf{v}\ f_i; \tag{14}$$

$$\nabla\cdot\mathbf{B} = 0. \tag{15}$$

These constraints may be applied as initial conditions. Because, as we noted, (10) is consistent with the continuity equation, a solution $\psi(f, \mathbf{E}, \mathbf{B})$ of (10)–(13) which satisfies (14) and (15) at $t = 0$, does so for all t. We have not found a more economical description, say in terms of potentials, which can be put in the form of Eq. (2).

We note that, despite the constraints, ψ may represent a perturbation in the fields with no corresponding perturbation of the particle distributions if $\mathbf{E}$ and $\mathbf{B}$ constitute a pure radiation field. For example, ψ might represent the onset of an rf pulse of external origin. Also, ψ may correspond to a net change in particle number.

In discussing linear stability in Sec. 3, we employed, as a norm, the square root of a scalar product. One dimensionless product of interest is

[8] T. K. Fowler, Phys. Fluids 4, 1393 (1961); 5, 249 (1962).

$$(\psi_1, \psi_2) = (N_0 V\theta)^{-1}\Big\{\gamma \sum_{i=1}^{n} \int d\mathbf{x}\, d\mathbf{v}\, (f_i^{(1)})^* f_i^{(2)} + \int d\mathbf{x}\, (\mathbf{E}_1^* \cdot \mathbf{E}_2 + \mathbf{B}_1^* \cdot \mathbf{B}_2)\Big\}. \quad (16)$$

The star denotes the complex conjugate. While only real perturbations are of interest, we have employed the complex product for generality. In addition to a direct attack on Eqs. (10)–(13), on employing the complex product, our methods also apply to possible simpler equations obtained from (10) by Fourier analysis with respect to some ignorable coordinate, such as the azimuthal angle in the case of axially symmetric equilibria.

The range of integration in (16) depends on the boundary conditions of the problem. As the range invariably includes all $\mathbf{v}$ space, the f_i must fall to zero asymptotically in $\mathbf{v}$ fast enough to insure the required finiteness of (ψ, ψ). In idealized problems ignoring the external environment, the range also covers all $\mathbf{x}$, and f_i, $\mathbf{E}$ and $\mathbf{B}$ must suitably vanish at large $\mathbf{x}$ as well.

In (16), V is the plasma volume (or, for infinite plasmas, an arbitrary volume) and

$$N_0 = \sum_i \int d\mathbf{v}\, f_0^{(i)}, \quad (17)$$

$$\theta = N_0^{-1} \sum_i \int d\mathbf{v}\, (\tfrac{1}{2} m_i v^2) f_0^{(i)}, \quad (18)$$

$$\gamma = \tfrac{4}{3}\pi (2n\theta / \sum_i m_i)^{\frac{3}{2}} (\theta / N_0). \quad (19)$$

As normalized, the first term of (ψ, ψ) given by (16) is crudely the square ratio of the number of particles involved in the perturbation to the number of particles in equilibrium. The second term is the ratio of the perturbed field energy to the total equilibrium kinetic energy. While these measures of perturbation magnitudes have some physical appeal and the scalar product (16) is mathematically simple, it is perhaps not possible to say that bounding (16), or any other one macroscopic norm, is for all equilibria a physically satisfactory stability criterion. The problem of chosing meaningful norms has been discussed by Backus.[9] Of course, as was pointed out in Sec. 3, in the linear approximation bounding any quadratic, definite norm at least forbids growing normal modes, $\varphi(t) = g(\mathbf{x}, \mathbf{v}) \exp(\lambda t)$.

We turn now to techniques for applying Lyapunov's linear theory, with examples. We represent the linearization of Eqs. (10)–(13) as follows. If there are n charge species, we first drop collisions and define

$$\psi = \begin{pmatrix} f_1 \\ \vdots \\ f_n \\ \mathbf{E} \\ \mathbf{B} \end{pmatrix} \qquad M = \left(\begin{array}{ccc|cc} -L_1 & & 0 & -\frac{q_1}{m_1}\frac{\partial f_0^{(1)}}{\partial \mathbf{v}}\cdot & \frac{q_1}{m_1}\left(\mathbf{v}\times\frac{\partial f_0^{(1)}}{\partial \mathbf{v}}\right)\cdot \\ & \ddots & & \vdots & \vdots \\ 0 & & -L_n & \frac{-q_n}{m_n}\frac{\partial f_0^{(n)}}{\partial \mathbf{v}}\cdot & \frac{q_n}{m_n}\left(\mathbf{v}\times\frac{\partial f_0^{(n)}}{\partial \mathbf{v}}\right)\cdot \\ \hline -q_1\int d\mathbf{v}\,\mathbf{v} & \cdots & -q_n\int d\mathbf{v}\,\mathbf{v} & 0 & \text{curl} \\ 0 & \cdots & 0 & -\text{curl} & 0 \end{array}\right). \quad (20)$$

The equation $(\partial\psi/\partial t) = M\psi$ should reproduce (12), (13), and n linearized, collisionless equations for the f_i obtained by setting the right-hand side of (10) equal to zero. The shorthand notations employed in (20) are best understood by making this comparison. A vector, say $\mathbf{E}$, denotes its three components in a column, while a vector followed by a dot, such as $[(\partial f_0/\partial \mathbf{v})\cdot]$, denotes its three components in a row, as if the dot product is to be taken when the matrix operates on something. Zeros in the $n \times n$ upper left-hand block of M indicate that this block is diagonal. Curl denotes, of course, an antisymmetric 3×3 matrix of partial derivative operations in space:

$$\text{curl} = \begin{pmatrix} 0 & -\partial_z & \partial_y \\ \partial_z & 0 & -\partial_x \\ -\partial_y & \partial_x & 0 \end{pmatrix}. \quad (21)$$

The integral sign, $\int d\mathbf{v}$, denotes integration on what follows if after M operates on something. The matrix operator M is square, dimension $(6 + n)$.

Neglecting collisions, $M = P$ occurring in Eqs. (2) and (3). To complete P in the general case, a linearized collision operator acting on the first n elements of ψ could be added.

In order to understand the symmetry structure

[9] G. Backus, J. Math. Phys. 1, 178 (1960).

of the operator P, necessary in order to apply Lyapunov's linear theory, we consider as an exampel a Maxwellian plasma. Let $f_0^{(i)} = \eta_i \exp(-\mathcal{E}_i/T_i)$, where η_i is a normalization and $\mathcal{E}_i = \frac{1}{2}m_i v^2 + \pi_i(\mathbf{x})$. Then elements of M containing $\mathbf{v} \times (\partial f_0^{(i)}/\partial \mathbf{v})$ vanish, and $(m_i^{-1}\, \partial f_0^{(i)}/\partial \mathbf{v})$ can be written $[(df_0^{(i)}/d\mathcal{E}_i)\mathbf{v}\cdot]$. Note that, as was discussed in reference 8, such distributions could represent plasmas confined by a static gravitational potential well included in $\pi(\mathbf{x})$.

As a preliminary step to proving stability, we will show that the following diagonal operator, H_+, is definite and belongs to the class ξ, defined by (D.12), and H_+M is anti-Hermitian with respect to the scalar product (16):

$$H_+ = \tfrac{1}{2}N_0 V\theta \begin{bmatrix} \left(-\gamma \dfrac{df_0^{(1)}}{d\mathcal{E}_1}\right)^{-1} & & & 0 \\ & \ddots & & \\ & & \left(-\gamma \dfrac{df_0^{(n)}}{d\mathcal{E}_n}\right)^{-1} & \\ & & & 1 \\ 0 & & & & 1 \end{bmatrix}. \tag{22}$$

H_+ is independent of t. It is Hermitian because it is diagonal with real functions as elements and it is positive definite because $(-df_0^{(i)}/d\mathcal{E}_i)^{-1} = T_i/f_0^{(i)} > 0$. If also $(\varphi, H_+\varphi)$ is finite, all $\varphi \in S$, then $H_+ \in \xi$. From (22) and (16),

$$(\varphi, H_+\varphi) = \frac{1}{2}\sum_i \int d\mathbf{x}\, d\mathbf{v}\, f_i^* f_i \left(\frac{-df_0^{(i)}}{d\mathcal{E}_i}\right)^{-1} + \frac{1}{2}\int d\mathbf{x}\, (\mathbf{E}^*\cdot\mathbf{E} + \mathbf{B}^*\cdot\mathbf{B}). \tag{23}$$

We shall simply restrict S sufficiently to guarantee finiteness of this quadratic form. Then f_i^2 falls off in $|\mathbf{v}|$ at least as fast as $f_0^{(i)}$, and so forth.

Defining the current $\mathbf{j} = \sum_i q_i \int d\mathbf{v}\, \mathbf{v} f_i$, we find

$$(\varphi_1, H_+M\varphi_2) = \frac{1}{2}\sum_i \int d\mathbf{x}\, d\mathbf{v}\, (f_i^{(1)})^* \left(\frac{df_0^{(i)}}{d\mathcal{E}_i}\right)^{-1} L_i f_i^{(2)} + \frac{1}{2}\int d\mathbf{x}\, (\mathbf{j}_1^*\cdot\mathbf{E}_2 - \mathbf{E}_1^*\cdot\mathbf{j}_2) + \frac{1}{2}\int d\mathbf{x}\, (\mathbf{E}_1^*\cdot \operatorname{curl} \mathbf{B}_2 - \mathbf{B}_1^*\cdot \operatorname{curl} \mathbf{E}_2). \tag{24}$$

As an idealization, we take for the range of spatial integration all $\mathbf{x}$ with f_i, $\mathbf{E}$ and $\mathbf{B}$ falling to zero at large $\mathbf{x}$ and $\mathbf{v}$ sufficiently fast to assure the existence of (24). Then the curl operator, Eq. (21), is Hermitian in the scalar product $(\mathbf{a}, \mathbf{b}) = \int d\mathbf{x}\, \mathbf{a}^*\cdot\mathbf{b}$. Similarly, the operator $(df_0^{(i)}/d\mathcal{E}_i)^{-1}L_i$ is anti-Hermitian in the product $(f, g) = \int d\mathbf{x}\, d\mathbf{v}\, f^*g$, which follows from the facts that L_i is anti-Hermitian in this product, the real function $(df_0^{(i)}/d\mathcal{E}_i)^{-1}$ is Hermitian, and they commute, because $L_i(df_0^{(i)}/d\mathcal{E}_i)^{-1} = 0$ and L_i is a first-order differential operator. Making use of these symmetries in (24) yields $(\varphi_1, H_+M\varphi_2) = -(\varphi_2, H_+M\varphi_1)^*$. Thus, acting on the set of perturbations, S_0, satisfying boundary conditions utilized in obtaining this result, H_+M is anti-Hermitian as claimed, and $[H_+M + M^\dagger H_+] = 0$. If we can find a subset, S, of S_0 such that H_+ is finite on S and the Vlasov equation maps S into S, the conditions for linear stability by theorem 1 are satisfied on S. In this case, H_+ would also satisfy theorem 2, whence the Lyapunov function, $(\varphi, H_+\varphi)$, is a constant of motion.

While we shall not attempt a rigorous proof, we suggest that such an S is the set of all φ satisfying (14) and (15) such that $(\varphi, H_+\varphi)$ and $d(\varphi, H_+\varphi)/dt$ exist and φ is analytic in $\mathbf{x}$ and $\mathbf{v}$. Since $M\varphi$ is an analytic functional of φ, we expect solution $\varphi(t)$ to be continuous in t, preserving analyticity in $\mathbf{x}$ and $\mathbf{v}$, and $(\varphi, H_+\varphi)$ to be continuous also, at least on a finite interval $0 \le t \le t_1$. Since also (10)–(13) preserve (14) and (15), $\varphi(t) \in S$ on the interval. Then, by properties of H_+ on S, $(\varphi, H_+\varphi)$ and $d/dt\,(\varphi, H_+\varphi)$ are constant in the interval, the derivative being in fact zero, and the situation at $t = t_1$ is much the same as at $t = 0$. Continuing the argument, interval by interval, we conclude that, if $\varphi(0) \in S$, $\varphi(t) \in S$, all $t > 0$; hence, $S \to S$.

The above argument may also be applied with the boundary condition that the plasma is surrounded by a surface which absorbs particles and energy. With these conditions, integrating by parts in (24) and recalling the Hermiticity definition (D.7), we obtain

$$(\varphi_2, (H_+M)^\dagger\varphi_1)^* = -(\varphi_2, H_+M\varphi_1)^* - \frac{1}{2}\sum_i \int d\mathbf{v}\, d\mathbf{S}\cdot\mathbf{v}\, \beta_p (f_i^{(1)})^* f_i^{(2)} \left(\frac{-df_0^{(i)}}{d\mathcal{E}_i}\right)^{-1} - \frac{1}{2}\int d\mathbf{S}\cdot\beta_F(\mathbf{E}_1^*\times\mathbf{B}_2 + \mathbf{E}_2\times\mathbf{B}_1^*). \tag{25}$$

Here $d\mathbf{S}$ denotes a surface element, and β_p and β_F symbolize the boundary condition instructions to exclude in the integrations regions in which $d\mathbf{S}\cdot\mathbf{v} < 0$, corresponding to returning particles, and, if $\varphi_1 = \varphi_2$, to exclude $d\mathbf{S}\cdot(\mathbf{E}\times\mathbf{B}) < 0$, corresponding to returning radiation, $\mathbf{E}\times\mathbf{B}$ being the Poynting vector. Thus, with $\varphi_1 = \varphi_2$ and Maxwellian $f_0^{(i)}$, the second and third terms on the right side of (25) always are positive or zero. Transposing the first term on the right to the left side, it follows that $[H_+M + M^\dagger H_+]$

is negative semidefinite on S, again compatible with theorem 1.

All conditions of theorem 1 having been satisfied on S, the collisionless Maxwellian plasma is shown to be linearly stable against the set of perturbations, S, which includes, loosely speaking, all perturbations with finite field energy and f_i's localized to the region of phase space occupied by the $f_0^{(i)}$. The proof also holds for any other sets of $f_0^{(i)}$ which damp fast enough at large $|\mathbf{v}|$ and which are everywhere monotonic decreasing functions of the equilibrium particle energies, $\mathcal{E}_i$, since H_+ is definite for any such $f_0^{(i)}$. Thus we have extended our previously reported proof of stability of this class of functions by Lyapunov's method to include effects of retardation in field propagation.[8] Finally, collisions may be included. In Appendix A, we identify expectation values of H_+ in (23) with the lowest-order terms in an expansion of the free energy. Thus we may apply Boltzmann's H Theorem to obtain $d(\varphi, H_+\varphi)/dt \leq 0$, and again $(\varphi, H_+\varphi)$ is a Lyapunov function.

The fact that our linear stability proof has employed the free energy calls to mind the thermodynamic stability criterion requiring free energy to be minimal at equilibrium.[10] Indeed, our requirement that H_+ be definite ensures that any free-energy perturbation is positive, or free energy itself has a minimum. Now, the thermodynamic criterion was proven only near true equilibrium.[11] Theorem 3 in fact assures us that there always exists in the asymptotically stable case a quantity which, like free energy, is minimal at the stationary state whose stability is in question and which in the linear approximation damps monotonically in time. However, the nature of this quantity is not specified, and it may not in any sense be the same quantity in all cases.

5. NONLINEAR STABILITY

As was pointed out in the introduction, one may be lucky enough to show that some functional is a Lyapunov function without linearizing the Vlasov equation, and in Appendix A, we report Gardner's success in proving nonlinear stability of a Maxwellian plasma by this means. In addition to this "direct method", Lyapunov also discussed nonlinear stability as a consequence of linear stability when all modes damp. While it appears to be difficult to apply the technique with rigor to plasma problems, the approach does give insight into the nonlinear extrapolation of linear behavior.

Lyapunov proves nonlinear stability when linear damping competes successfully with growth due to nonlinear terms. We consider first collision damping and present a version of Lyapunov's theorem tailored to the linear stability criteria of theorem 3.

Suppose there exists $H_+ \in \xi$ [ξ is defined by (D.12)] such that $[H_+P + P^\dagger H_+]$ is negative definite on S. Then, nonlinearly, for some $\delta > 0$ and all $\psi \in S$,

$$(d/dt)(\psi, H_+\psi) \leq -\delta(\psi, \psi) + 2\,\mathrm{Re}\,(\psi, H_+Q(\psi)), \tag{26}$$

where $Q(\psi)$, the nonlinear term of (2), is in our case quadratic in ψ. Suppose that, for all $\psi \in S$,

$$|\mathrm{Re}\,(\psi, H_+Q(\psi))| \leq \tfrac{1}{2}\beta(\psi, \psi)(\psi, H_+\psi), \tag{27}$$

which is a kind of boundedness which we shall call a *Lyapunov condition*. Then, by (26), for all $\psi \in S$,

$$(d/dt)(\psi, H_+\psi) \leq [-\delta + \beta(\psi, H_+\psi)](\psi, \psi). \tag{28}$$

Clearly, if (2) maps S into S, $(\psi, H_+\psi)$ is monotone-decreasing for all time if, at $t = 0$,

$$(\psi, H_+\psi) < \beta^{-1}\delta. \tag{29}$$

Nonlinear stability follows in a norm $||\psi|| = (\psi, H_+\psi)^{\frac{1}{2}}$ in accordance with (D.5). If H_+ is positive definite, rather than merely nonzero as was required if $H_+ \in \xi$, the system is also nonlinearly stable in the norm $||\psi|| = (\psi, \psi)^{\frac{1}{2}}$. In either case, at least $(\psi, \psi) \to 0$ as $t \to \infty$. Replacing the bracket in (28) by its maximum value under condition (29), by exact analogy with the sufficiency proof of theorem 3 we see that (ψ, ψ) is bounded and integrable on $t > 0$.

To summarize:

Theorem 5. *If S maps into S and Q satisfies the Lyapunov condition on S, negative definiteness of $[H_+P + P^\dagger H_+]$ on S is a sufficient condition for nonlinear stability in norm $||\psi|| = (\psi, H_+\psi)^{\frac{1}{2}}$ if $H_+ \in \xi$, and in norm $||\psi|| = (\psi, \psi)^{\frac{1}{2}}$ if also H_+ is definite. In either case, $(\psi, \psi) \to 0$ as $t \to \infty$.*

Thus, we have found that, granting the Lyapunov condition, our necessary and sufficient condition (theorem 3) for the linear damping of perturbations ψ is at least sufficient for nonlinear stability. Further, if H_+ is definite, the perturbations are held under as tight control nonlinearly as linearly, since $(\psi, H_+\psi)$ is monotone-decreasing and $(\psi, \psi) \leq \lambda\,(\psi, H_+\psi)$, some $\lambda > 0$. However, in either case stability is assured only for perturbations of limited magnitude

[10] E. A. Gugenheim, *Thermodynamics* (North-Holland Publishing Company, Amsterdam, 1957), p. 97.

[11] For thermodynamic treatments near thermal equilibrium, see references in the paper of P. Glansdorff, I. Prigogine, and D. F. Hays, Phys. Fluids 5, 114 (1962).

at $t = 0$. Formula (29) for the limit is just the condition that an upper bound on the nonlinear growth constant $\beta(\psi, H_+\psi)$ be controlled by a lower bound on the linear damping rate δ. Of course, the theorem loses content if $\beta^{-1}\,\delta \to 0$.

Of the two required bounds, that on β—the Lyapunov condition (27)—is the more difficult to guarantee on an interestingly large set of plasma perturbations, S also known to be mapped into itself by the nonlinear Vlasov equation. For example, consider again the Maxwellian plasma discussed in Sec. 4, with H_+ given by (22). We may assume that, with collisions and an absorbing boundary, H_+ satisfies the conditions of theorem 5 on the set S, discussed in Sec. 4, consisting of all ψ satisfying (14) and (15) with $(\psi, H_+\psi)$ and $d(\psi, H_+\psi)/dt$ finite and ψ analytic in $\mathbf{x}$ and $\mathbf{v}$. By appealing to continuity, as in Sec. 4, one may conclude that S is mapped into itself, at least on an interval $0 \leq t \leq t_1$. Then $(\psi, H_+Q(\psi))$, being one term of $d(\psi, H_+\psi)/dt$, is probably also finite on the interval, and, if the interval is short enough, (27) is satisfied on the interval with some $\beta(t_1)$ depending on t_1. However, without assurance that $\beta(t_1)$ is itself bounded as t_1 is allowed to increase, the Lyapunov condition fails and theorem 5 is inapplicable.

Consider the term $(\partial f/d\mathbf{v})\cdot\mathbf{E}$ contributing to Q. Then, for the Maxwellian equilibrium,

$$(\psi, H_+Q(\psi)) = \int d\mathbf{x}\, d\mathbf{v}\, f\left(\frac{\partial f}{\partial \mathbf{v}}\cdot\mathbf{E}\right)(T/f_0) + \cdots . \quad (30)$$

The difficulty in satisfying (27) is twofold. First, $\partial/\partial\mathbf{v}$ is not a bounded operator in product (16), nor in product $[\psi, \psi] \equiv (\psi, H_+\psi)$. In our special case, by parts integration $\partial/\partial\mathbf{v}$ may be replaced by $\mathbf{v}$, but, as an operator, this quantity is similarly unbounded. Secondly, because (30) is cubic in f and $\mathbf{E}$, a bound on the quadratic form (16) places no limit on (30) if f^2 and E^2 are allowed to accumulate at a point in space.

Now, we may argue that neither of these possible difficulties is real. On an interval such that (27) holds, $(\psi, H_+\psi)$ is monotone-decreasing, whence the field energy is bounded and, by energy conservation, the kinetic energy is also. Thus, it seems unlikely that $\int d\mathbf{x}\, d\mathbf{v}\, \mathbf{v}f^2/f_0$ is steadily increasing (the first difficulty), and we expect Landau damping to suppress accumulation of f^2 and E^2 over distances less than the Debye length. A bound such as (27) might be derived from these plausible limitations, but we have not been able to obtain a more rigorous bound. In any case, collision damping is so slow that the limit on stable perturbations, Eq. (29), is extremely small.

Finally, one may attempt to avoid the above difficulties in satisfying the Lyapunov condition by changing norms. A more convenient form of theorem 5 for general norms is given in Bellman's book.[12] The appropriate Lyapunov condition requires that for every $\epsilon > 0$ there exists $\delta(\epsilon) > 0$ such that $||Q(\psi)|| \leq \epsilon\, ||\psi||$, all $\psi \in S$ such that $||\psi|| \leq \delta$. A norm rigorously satisfying this requirement is (one-dimensional, one-species plasma, electrostatic perturbations only);

$$||\psi(f, E)|| = \limsup_{(\text{all } x)} \left\{ |E(x)| + \sum_{n=0}^{\infty} \nu^n \int dv \left| \frac{\partial^n f}{\partial v^n} \right| \right\}. \quad (31)$$

The series takes care of the velocity derivative difficulty, and the absence of spatial integrals avoids the point accumulation problem. The generalization to three dimensions and magnetic perturbations is straightforward. Now the problem is to prove, by any means, linear stability in such a norm on a set mapped into itself. Bellman's proof is developed from the integral equation equivalent to (2),

$$\psi(t) = T(t)\psi(0) + \int_0^t dt'\, T(t - t')Q[\psi(t')], \quad (32)$$

where $T(t)$ is again the solution operator for the linearized equation (3). Specifically, it is required of linear behavior that, in norm (31), $||T(t)||$ is bounded and integrable on $t \geq 0$.

Theorem 5 is not useful in treating Landau-damped collisionless plasmas because the f_i's do not damp. Since condition (27) requires knowledge of the f_i's, one cannot, as in the linear theory, resort to a scalar product involving only damping field quantities, say $\int d\mathbf{x}\, E^2$. However, the second version of the nonlinear stability theorem, employing norms like (31), can be adapted to prove nonlinear stability as a consequence of linear Landau damping. In fact, one shows also that Landau damping of the fields persists nonlinearly, a problem discussed by Montgomery and Gorman.[13] Returning to the one-dimensional plasma $\mathbf{B} = 0$, let $U = R\psi$, where the transformation simply multiplies f by $\exp(-\lambda t)$ and E by $\exp(\lambda t)$. Then $U(0) = \psi(0)$ and, dropping collisions, $Q(\psi) = Q(U)$. Operating on (32) by $R(t)$ gives

[12] R. Bellman, *Stability Theory of Differential Equations* (McGraw-Hill Book Company, Inc., New York, 1953), pp. 78–85.

[13] D. Montgomery and D. Gorman, Phys. Rev. **124**, 1309 (1961).

$$U(t) = R(t)T(t)U(0) + \int_0^t dt'\, R(t)T(t-t')Q[U(t')]. \quad (33)$$

Bellman's version of the nonlinear stability proof may be applied to (33). One concludes that, if, in norm (31), $||R(t)T(t)||$ is bounded on $t > 0$ and $\int_0^t dt'\ ||R(t)T(t-t')||$ is bounded as $t \to \infty$, $||U||$ is bounded nonlinearly, whence $|E|$ damps at least as fast as $\exp(-\lambda t)$. It is known from Fourier–Laplace transform analysis that the linear behavior of a Maxwellian plasma satisfies this requirement with boundary conditions eliminating oscillations of arbitrarily large wavelength. We choose λ a little less than the minimum linear Landau damping rate. Though $\partial^n f/\partial v^n \propto t^n$, the factor $\exp(-\lambda t)$ introduced by the transformation controls this growth, any $\lambda > 0$.

In conclusion, we note that, to the extent that a Lyapunov condition of form (27) holds even without damping, theorem 5 may be employed to set bounds on nonlinear growth of linearly stable systems for finite times. One simply transforms to $u = \psi \exp(-\mu t)$, as in obtaining theorem 4. By nonlinear stability of the u's, the ψ's grow no faster than $\exp(\mu t)$ if initially ψ satisfies limit (29). Since in this case $\delta \propto \mu$, the magnitude limit is inversely proportional to the length of time for which the ψ's are effectively bounded; just μ^{-1}.

ACKNOWLEDGMENTS

I would like to express appreciation to Professors H. Grad and J. Berkowitz for the hospitality of Courant Institute of Mathematical Sciences, New York University, during a portion of this work. I am indebted to H. Weitzner and particularly to C. S. Gardner of that institution for helpful suggestions and comments. I would like to thank Dr. Gardner for permission to report his proof in Appendix A.

APPENDIX A

We wish to show the relation of $(\varphi, H_+\varphi)$, given by (23), and free energy. In doing so, we also give Gardner's proof of nonlinear stability of a Maxwellian plasma, which is an extension of Newcomb's linear proof.[14,15] The proof bypasses linear analysis by exhibiting a Lyapunov function for the nonlinear equation.

Consider for simplicity a plasma of one-particle species with a Maxwell distribution in equilibrium, $f_0 = N \exp(-\mathcal{E}/T)$, where $\mathcal{E} = \frac{1}{2}mv^2 + \pi(\mathbf{x})$ and π was defined following Eq. (11). Let $\mathbf{B}_0 = 0$. The following quantity is essentially the free energy corresponding to a perturbation $\psi(f, \mathbf{E}, \mathbf{B})$:

$$F(\psi) = \int d\mathbf{x}\, d\mathbf{v}\, [T(G(f+f_0) - G(f_0)) + (f+f_0)\mathcal{E}] + \frac{1}{2}\int d\mathbf{x}\, (E^2 + B^2), \quad (A1)$$

where $G(x) = (x \ln N^{-1}x - x)$. Aside from the subtraction of a linear term constant by particle conservation, the first term is T times negative entropy normalized to zero at equilibrium, and the second and third terms are the internal energy. As particle and energy conserving boundary conditions, we let $\mathbf{x}$ and $\mathbf{v}$ integrations cover all space and f, $\mathbf{E}$ and $\mathbf{B}$ damp to zero at large $|\mathbf{x}|$ and $|\mathbf{v}|$. Presuming that these conditions are preserved by continuous solutions of (10)–(13), F is forever constant without collisions and, by Boltzmann's H theorem, monotone-decreasing with collisions. Since $F(0)$ is constant if f_0 is, the change in free energy due to the perturbation $[\Delta F = F(\psi) - F(0)]$, has similar time behavior.

We expand

$$G(f+f_0) = G(f_0) + fG'(f_0) + \tfrac{1}{2}f^2G''(f_0) + \cdots. \quad (A2)$$

Here $G'(f_0) = -\mathcal{E}T^{-1}$, and $G''(f_0) = f_0^{-1}$. For the moment we drop higher-order terms. Introducing (A2) into (A1), we see that in this approximation the free energy perturbation ΔF is just the constant $(\varphi, H_+\varphi)$ in (23).[16] Thus the first term of (23) is seen to arise from an expansion of the entropy function, as in reference 14.

The existence of positive definite ΔF quadratic in perturbations was employed by Newcomb to prove stability of a Maxwellian plasma.[14] Kruskal and Oberman[3] adapted the method to obtain in the small m/e limit an energy principle for all functions of $\mathcal{E}$, and possibly other variables, such that $\partial f_0/\partial \mathcal{E} < 0$. It is their development of Newcomb's proof which we have followed above.

Gardner has observed that, if rather than throw away higher-order terms in (A2), we apply the mean value theorem, we may terminate the series by replacing $G''(f_0)$ by $G''(f_0 + f_1)$, some $f_1(t)$ lying be-

[14] See the Appendix to the paper of I. B. Bernstein, Phys. Rev. **109**, 10 (1958).

[15] For similar methods in magnetohydrodynamics, see J. Berkowitz, H. Grad, and H. Rubin, *Proceedings of the Second International Conference on the Peaceful uses of Atomic Energy* (United Nations, Geneva, 1958), Vol. 31, p. 177.

[16] Free energy is what Buneman has called "*r.f.* energy" in his article appearing in *Radiation and Waves in Plasmas* [Edited by M. Mitchner (Stanford University Press, Stanford, California, 1961)].

tween zero and f_0. Now, the nonlinear solution is required to be positive, whence a median value thereof, $f_0 + f_1$, is also positive. Then for all time $G''(f_0 + f_1) = T/(f_0 + f_1) > 0$, and again ΔF is composed of two positive definite quantities, one being the perturbed field energy, $\frac{1}{2}\int dx\ (E^2 + B^2)$. For all ψ such that ΔF is initially finite, it follows from the constancy of ΔF with respect to the nonlinear Eq. (10) without collisions that each term, in particular the field energy, is bounded nonlinearly, and from the damping of ΔF with collisions that the field energy then damps nonlinearly; hence, stability in either case. ΔF is an example of a Lyapunov function.[1]

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 4 APRIL 1963

Partition Function for Certain Simple Lie Algebras*

JAN TARSKI†
Department of Physics, University of Wisconsin, Madison, Wisconsin
(Received 17 August 1962)

The partition function 'which yields the multiplicities of weights in representations' is computed for the following Lie algebras: A_2, B_2, G_2, and A_3.

1. INTRODUCTION

IN this paper we obtain explicit expressions for the partition function for certain simple Lie algebras. This function allows us to compute directly the multiplicity of a weight in a representation of a Lie algebra or group.[1] We determine this function through a study of the combinatorial properties of the weights.

The representations of the Lie algebras in question have been considered in connection with various physical problems. As examples, we cite studies of supermultiplets[2] and of global symmetry.[3,4]

The concepts of weights and of their multiplicities are fundamental in representation theory, and are also very intuitive. However, this terminology is seldom found in the physical literature, and therefore we shall give a brief explanation. Given a (semisimple) Lie algebra, we consider a maximal set of linearly independent, mutually commuting elements. In a representation, the matrices associated with such a set can be simultaneously diagonalized, and the eigenvalues which are associated with a simultaneous eigenvector form a *weight*. The dimension of the space associated with a given weight is the *multiplicity* of that weight.

We see that a knowledge of weights and of their multiplicities provides a very detailed description of representations. To give an example of these concepts, let us consider the familiar case of irreducible representations of the three-dimensional rotation group. Here a single eigenvalue forms a weight, and every weight has multiplicity one. (All representations are assumed finite-dimensional. We also remark that Behrends *et al.*[3] use the word *multiplicity* in a different sense.)

Our emphasis is on the simple Lie algebras of rank 2, i.e. on A_2, B_2, and G_2, and we also discuss A_3. The case of rank 1 is trivial, while for Lie algebras of rank 3 or greater (except for A_3) the explicit forms of the partition function are, apparently, very involved. Since the representations of the four Lie algebras in question have been analyzed extensively by more direct methods,[2,3] there seems to be little immediate application for our results. Nevertheless, formulas which provide a closed-form solution to a general problem usually turn out to be of value. Our results also illustrate a recursion

* Most of the results presented here are included in a technical report which was distributed by the Department of Mathematics, University of California, Berkeley, California, in 1958, but have not been published previously. The preparation of this report was supported by U.S.A.F., Office of Scientific Research, Air Research and Development Command. The completion of this work was supported by the Atomic Energy Commission and by the Research Committee of the Graduate School from funds supplied by the Wisconsin Alumni Research Foundation.

† Present Adress: Courant Institute of Mathematical Sciences, New York University, New York, New York.

[1] B. Kostant, Trans. Am. Math. Soc. **93**, 53 (1959).

[2] E. Wigner, Phys. Rev. **51**, 106 (1937).

[3] R. E. Behrends, J. Dreitlein, C. Fronsdal, and B. W. Lee, Rev. Mod. Phys. **34**, 1 (1962).

[4] D. R. Speiser and J. Tarski, J. Math. Phys. (to be published).

tween zero and f_0. Now, the nonlinear solution is required to be positive, whence a median value thereof, $f_0 + f_1$, is also positive. Then for all time $G''(f_0 + f_1) = T/(f_0 + f_1) > 0$, and again ΔF is composed of two positive definite quantities, one being the perturbed field energy, $\frac{1}{2}\int dx\ (E^2 + B^2)$. For all ψ such that ΔF is initially finite, it follows from the constancy of ΔF with respect to the nonlinear Eq. (10) without collisions that each term, in particular the field energy, is bounded nonlinearly, and from the damping of ΔF with collisions that the field energy then damps nonlinearly; hence, stability in either case. ΔF is an example of a Lyapunov function.[1]

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 4 APRIL 1963

Partition Function for Certain Simple Lie Algebras*

JAN TARSKI†
Department of Physics, University of Wisconsin, Madison, Wisconsin
(Received 17 August 1962)

The partition function 'which yields the multiplicities of weights in representations' is computed for the following Lie algebras: A_2, B_2, G_2, and A_3.

1. INTRODUCTION

IN this paper we obtain explicit expressions for the partition function for certain simple Lie algebras. This function allows us to compute directly the multiplicity of a weight in a representation of a Lie algebra or group.[1] We determine this function through a study of the combinatorial properties of the weights.

The representations of the Lie algebras in question have been considered in connection with various physical problems. As examples, we cite studies of supermultiplets[2] and of global symmetry.[3,4]

The concepts of weights and of their multiplicities are fundamental in representation theory, and are also very intuitive. However, this terminology is seldom found in the physical literature, and therefore we shall give a brief explanation. Given a (semisimple) Lie algebra, we consider a maximal set of linearly independent, mutually commuting elements. In a representation, the matrices associated with such a set can be simultaneously diagonalized, and the eigenvalues which are associated with a simultaneous eigenvector form a *weight*. The dimension of the space associated with a given weight is the *multiplicity* of that weight.

We see that a knowledge of weights and of their multiplicities provides a very detailed description of representations. To give an example of these concepts, let us consider the familiar case of irreducible representations of the three-dimensional rotation group. Here a single eigenvalue forms a weight, and every weight has multiplicity one. (All representations are assumed finite-dimensional. We also remark that Behrends *et al.*[3] use the word *multiplicity* in a different sense.)

Our emphasis is on the simple Lie algebras of rank 2, i.e. on A_2, B_2, and G_2, and we also discuss A_3. The case of rank 1 is trivial, while for Lie algebras of rank 3 or greater (except for A_3) the explicit forms of the partition function are, apparently, very involved. Since the representations of the four Lie algebras in question have been analyzed extensively by more direct methods,[2,3] there seems to be little immediate application for our results. Nevertheless, formulas which provide a closed-form solution to a general problem usually turn out to be of value. Our results also illustrate a recursion

* Most of the results presented here are included in a technical report which was distributed by the Department of Mathematics, University of California, Berkeley, California, in 1958, but have not been published previously. The preparation of this report was supported by U.S.A.F., Office of Scientific Research, Air Research and Development Command. The completion of this work was supported by the Atomic Energy Commission and by the Research Committee of the Graduate School from funds supplied by the Wisconsin Alumni Research Foundation.

† Present Adress: Courant Institute of Mathematical Sciences, New York University, New York, New York.

[1] B. Kostant, Trans. Am. Math. Soc. **93**, 53 (1959).

[2] E. Wigner, Phys. Rev. **51**, 106 (1937).

[3] R. E. Behrends, J. Dreitlein, C. Fronsdal, and B. W. Lee, Rev. Mod. Phys. **34**, 1 (1962).

[4] D. R. Speiser and J. Tarski, J. Math. Phys. (to be published).

relation which follows from the multiplicity formula.

Let us review now some concepts[3,5] relevant to a semisimple Lie algebra $\mathfrak{g}$. Let $\alpha_1, \cdots, \alpha_k$ be the simple positive roots of $\mathfrak{g}$; these are linearly independent. Ordinarily we shall think of the roots α_i as basis vectors:

$$\alpha_1 = (1, 0, \cdots, 0), \cdots, \alpha_k = (0, \cdots, 0, 1).$$

Certain linear combinations of the α_i constitute the complete set of positive roots of $\mathfrak{g}$:

$$\varphi_i = \sum c_{ij}\alpha_j, \quad \text{for} \quad i = 1, \cdots, M. \tag{1.1}$$

Here the c_{ij} are nonnegative integers, and $\varphi_i \neq 0$. We can define the partition function P for $\mathfrak{g}$ as follows:

Definition 1.1. Let $n_1, \cdots, n_k$ be k nonnegative integers. Then $P(n_1, \cdots, n_k)$ equals the number of ways in which $\sum n_i\alpha_i$ can be partitioned into a sum of roots φ_i, i.e., the number of ways we can write

$$(n_1, \cdots, n_k) \equiv \sum n_i\alpha_i = \sum m_i\varphi_i,$$

where the m_i are nonnegative integers. If some $n_i < 0$, then $P(n_1, \cdots, n_k) = 0$.

In terms of the function P, we have the following multiplicity formula [reference 1, Eq. (1.1.5)]:

$$m_\lambda(\nu) = \sum_{\sigma \in W} (\operatorname{sgn} \sigma) P[\sigma(g + \lambda) - (g + \nu)]. \tag{1.2}$$

Here $m_\lambda(\nu)$ is the multiplicity of the weight ν in an irreducible representation of $\mathfrak{g}$ having λ as its highest weight, $g = \frac{1}{2}\sum \varphi_i$, and W is the Weyl group.

Let us set $\lambda = 0$ in Eq. (1.2). Then $m_\lambda(\mu) = 0$ for $\mu \neq 0$, and

$$P(\mu) = -\sum_{\sigma \in W, \sigma \neq e} (\operatorname{sgn} \sigma) P[\mu - (g - \sigma g)], \tag{1.3}$$

for $\mu \neq 0$ [reference 1, Eq. (1.1.6)]. This relation determines P recursively if we are also given that $P(0) = 1$ and that $P(\mu) = 0$ whenever some $n_i < 0$.

The foregoing definition of P can also be used if we have a set of vectors $\{\varphi_i\}$ which is not related to a Lie algebra, provided each φ_i has integral nonnegative components and is nonzero. However, Eq. (1.3) does not apply in such cases.

In Sec. 2 we deduce a few combinatorial relations. In Sec. 3 the explicit forms of P are tabulated, and the proofs are outlined. In Sec. 4 we relate some of the properties of P to the recursion relation (1.3). Finally, an appendix contains the expressions $\sigma(\mu)$ for $\sigma \in W$. These are given in order to facilitate the use of the multiplicity formula.

[5] H. Weyl, Math. Z. **23**, 271; **24**, 328, 377 (1925). Reprinted in H. Weyl, *Selecta* (Birkhäuser, Basel, 1956), pp. 262 ff.

2. SOME COMBINATORIAL RELATIONS

In this section we present a few combinatorial lemmas that will be useful later. Our approach depends primarily on relating $P(n_1, \cdots, n_k)$ to the partition function q for one-dimensional problems.

We make two remarks before defining the function q. First, we always have $P \geq 0$, and

$$P(0) = 1.$$

If each $n_i \geq 0$ and if the α_i are included among the φ_i, then $P > 0$. The conditions $n_i \geq 0$ will always be understood, unless an explicit statement to the contrary is made.

Second, an M-tuple $(m_1, \cdots, m_M)$ consisting of nonnegative integers and satisfying $\sum m_i\varphi_i = (n_1, \cdots, n_k)$ will be called an *M-tuple for* $(n_1, \cdots, n_k)$.

We now define for positive integers $j_1, \cdots, j_l$, some of which may be equal,

$$q(n; j_1, \cdots, j_l) = P(n), \tag{2.1}$$

where the function $P(n)$ refers to the set

$$\{\varphi_i = j_i\}_{1 \leq i \leq l}.$$

In other words, $q(n; j_1, \cdots, j_l)$ equals the number of ways of writing

$$n = \sum m_i j_i,$$

where each m_i is a nonnegative integer.

We note a few properties of q. We may have $m_l = 1, 2, \cdots, [n/j_l]$, and therefore

$$q(n; j_1, \cdots, j_l) = \sum_{a=0}^{[n/j_l]} q(n - aj_l; j_1, \cdots, j_{l-1}). \tag{2.2}$$

This equation can be used for an explicit evaluation of q. This equation also yields

$$q(n; j_1, \cdots, j_{l-1}, 1) - q(n - 1; j_1, \cdots, j_{l-1}, 1) = q(n; j_1, \cdots, j_{l-1}). \tag{2.3}$$

Next, let us adopt the notation

$$q(n; \overbrace{1, \cdots, 1}^{l+1}) = q_l(n). \tag{2.4}$$

The function $q_l(n)$ can be evaluated by elementary considerations. One gets

$$q_l(n) = \binom{n + l}{l} = \frac{1}{l!}\prod_{a=1}^{l}(n + a). \tag{2.5}$$

We now come to the lemmas which relate $P(n_1, \cdots, n_k)$ to q.

Lemma 2.1. Let P be the partition function for the following set of vectors:

$\theta_0 = (1, 0), \quad \theta_i = (j_i, k_i) \quad for \quad i = 1, \cdots, N.$

The vector (0, 1) *may, but need not, be present among the* θ_i. *Let* max $(j_i/k_i) \leq n_1/n_2$. *Then*

$$P(n_1, n_2) = q(n_2; k_1, \cdots, k_N).$$

Proof. In order that $(m_0, \cdots, m_N)$ may be an $(N + 1)$-tuple for (n_1, n_2), we must have

$$n_1 = m_0 + \sum_{i=1}^{N} m_i j_i, \qquad n_2 = \sum_{i=1}^{N} m_i k_i. \qquad (2.6a, b)$$

The assumed inequality implies that if Eq. (2.6b) is satisfied by nonnegative m_i, for $1 \leq i \leq N$, then Eq. (2.6a) is satisfied by a (unique) $m_0 \geq 0$. Therefore P equals the number of ways to satisfy Eq. (2.6b), and the lemma follows.

The same type of argument can be used to prove the following lemma.

Lemma 2.2. Let P *be the partition function for the following set of vectors*:

$$\xi_1 = (1, 0, 0), \quad \xi_2 = (0, 0, 1), \quad \xi_i = (j_i, k_i, l_i) \quad for \quad i = 3, \cdots, N.$$

Let $j_i, l_i \leq k_i$ *and let* $n_2 \leq n_1, n_3$. *Then*

$$P(n_1, n_2, n_3) = q(n_2; k_3, \cdots, k_N).$$

In Lemma 2.3 we find a more involved expression for P.

Lemma 2.3. Let P *be the partition function for the following set of vectors*:

$$\eta_1 = (0, 1, 0), \quad \eta_2 = (1, 1, 1),$$
$$\eta_i = (j_i, k_i, 0) \quad for \quad i = 3, \cdots, r,$$
$$\eta_i = (0, k_i, l_i) \quad for \quad i = r + 1, \cdots, N.$$

For $i = 3, \cdots, N$, *let* $k_i \leq j_i$ *or let* $k_i \leq l_i$, *as appropriate, and let* $n_1 + n_3 \leq n_2$. *Then*

$$P(n_1, n_2, n_3) = \sum_{a=0}^{\min(n_1, n_3)} q(n_1 - a; j_3, \cdots, j_r) \times q(n_3 - a; l_{r+1}, \cdots, l_N).$$

Proof. Let $(m_1, \cdots, m_N)$ be an N-tuple for (n_1, n_2, n_3). Necessarily $0 \leq m_2 \equiv a \leq \min(n_1, n_3)$. Consider the equations for n_1 and for n_3:

$$n_1 - a = \sum_{i=3}^{r} m_i j_i, \qquad n_3 - a = \sum_{i=r+1}^{N} m_i l_i.$$

These equations are independent, and if they are satisfied by nonnegative m_i, for $3 \leq i \leq N$, then the corresponding equation for n_2 can be satisfied by an $m_1 \geq 0$. The conclusion follows, as before.

Sometimes one can determine P by considering first a difference function defined by P. This, in fact, is our technique for those cases where the foregoing lemmas are not directly applicable, and we shall give an example in the next section. Here we state the basic lemma.

Lemma 2.4. Let $P(n_1, \cdots, n_k)$ *be the partition function for the set of vectors* $\{\zeta_1, \cdots, \zeta_M\}$, *and let* $\zeta_1 = (1, 0, \cdots, 0)$. *Then the difference function*

$$C(n_1, n_2, \cdots, n_k) = P(n_1, n_2 \cdots, n_k) - P(n_1 - 1, n_2, \cdots, n_k)$$

is the partition function for the set of vectors $\{\zeta_2, \cdots, \zeta_M\}$. *(An analogous conclusion clearly holds if we replace* 1 *by any other* $j \leq M$.)

Proof. The correspondence

$$(m_1, m_2, \cdots, m_M) \leftrightarrow (m_1 - 1, m_2, \cdots, m_M)$$

is a one-to-one correspondence between those M-tuples for $(n_1, n_2, \cdots, n_k)$ in which $m_1 > 0$, and all M-tuples for $(n_1 - 1, n_2, \cdots, n_k)$. There remain the M-tuples for $(n_1, n_2, \cdots, n_k)$ in which $m_1 = 0$, and the lemma follows.

3. THE EVALUATION OF P

The roots of a semisimple Lie algebra $\mathfrak{g}$ can be considered[3,5] as vectors in the Euclidean k-space R^k, where k is the rank of $\mathfrak{g}$. There is a natural choice for the angles between the root vectors and for their lengths. (However, we shall be concerned only with relative lengths.) In the case of Lie algebras of rank 2, in particular, we can represent the roots as vectors in a plane. These vectors extend into rays which divide the plane into a number of sectors. This is shown in Fig. 1 for the positive roots.

The function P has a different form in each sector. The details of the geometric configurations will be important for us in Sec. 4 and in the Appendix. However, our evaluation of P is algebraic, and the value of the diagrams in the present section is largely heuristic.

For the semisimple Lie algebra of rank 1, there is only one positive root vector, and therefore $P(n_1) = 1$.

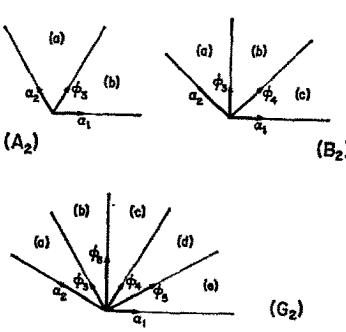


FIG. 1. Positive root vectors and the cases for the Lie algebras of rank 2.

TABLE I. Partition function for the four Lie algebras.

Algebra	Case	Value of $P(n_1, n_2)$ or of $P(n_1, n_2, n_3)$
A_2	Positive root vectors: $\alpha_1 = \varphi_1 = (1, 0);\quad \alpha_2 = \varphi_2 = (0, 1);\quad \varphi_3 = (1, 1)$; $\lvert\alpha_2\rvert = \lvert\alpha_1\rvert;\quad \sphericalangle(\alpha_1, \alpha_2) = 120°$	
	(a) $n_1 \leq n_2$	$n_1 + 1$
	(b) $n_2 \leq n_1$	$n_2 + 1$
B_2	Positive root vectors: $\alpha_1 = \varphi_1 = (1, 0);\quad \alpha_2 = \varphi_2 = (0, 1);\quad \varphi_3 = (1, 1);\quad \varphi_4 = (2, 1)$; $\lvert\alpha_2\rvert = \sqrt{2}\,\lvert\alpha_1\rvert;\quad \sphericalangle(\alpha_1, \alpha_2) = 135°$	
	(a) $n_1 \leq n_2$	$b(n_1)$
	(b) $n_2 \leq n_1 \leq 2n_2$	$b(n_1) - q_2(n_1 - n_2 - 1)$ $= q_2(n_2) - b(2n_2 - n_1 - 1)$
	(c) $2n_2 \leq n_1$	$q_2(n_2)$
	$b(n) \equiv q(n; 2, 1, 1)$ $= \frac{1}{4}(n + 2)^2$ for n even, $= \frac{1}{4}(n + 1)(n + 3)$ for n odd;	$q_2(n) = \frac{1}{2}(n + 1)(n + 2)$.
G_2	Positive root vectors: $\alpha_1 = \varphi_1 = (1, 0);\quad \alpha_2 = \varphi_2 = (0, 1);\quad \varphi_3 = (1, 1)$; $\varphi_4 = (2, 1);\quad \varphi_5 = (3, 1);\quad \varphi_6 = (3, 2)$; $\lvert\alpha_2\rvert = \sqrt{3}\,\lvert\alpha_1\rvert;\quad \sphericalangle(\alpha_1, \alpha_2) = 150°$	
	(a) $n_1 \leq n_2$	$g(n_1)$
	(b) $n_2 \leq n_1 \leq \frac{3}{2}n_2$	$g(n_1) - h(n_1 - n_2 - 1)$
	(c) $\frac{3}{2}n_2 \leq n_1 \leq 2n_2$	$h(n_2) - g(3n_2 - n_1 - 1)$ $+ h(2n_2 - n_1 - 2)$
	(d) $2n_2 \leq n_1 \leq 3n_2$	$h(n_2) - g(3n_2 - n_1 - 1)$
	(e) $3n_2 \leq n_1$	$h(n_2)$
	$g(n) \equiv q(n; 3, 3, 2, 1, 1)$ $= (1/432)(n + 6)(n^3 + 14n^2 + 54n + 72)$ for $n \equiv 0 \bmod 6$, $= (1/432)(n + 5)^2(n^2 + 10n + 13)$ for $n \equiv 1$, $= (1/432)(n + 4)(n^3 + 16n^2 + 74n + 68)$ for $n \equiv 2$, $= (1/432)(n + 3)^2(n + 5)(n + 9)$ for $n \equiv 3$,	$g(n) = (1/432)(n + 2)(n + 8)(n^2 + 10n + 22)$ for $n \equiv 4$, $= (1/432)(n + 1)(n + 5)(n + 7)^2$ for $n \equiv 5$;
		$h(n) \equiv q(n; 2, 1, 1, 1, 1)$ $= (1/48)(n + 2)(n + 4)(n^2 + 6n + 6)$ for n even, $= (1/48)(n + 1)(n + 3)^2(n + 5)$ for n odd.
A_3	Positive root vectors: $\alpha_1 = \varphi_1 = (1, 0, 0);\quad \alpha_2 = \varphi_2 = (0, 1, 0);\quad \alpha_3 = \varphi_3 = (0, 0, 1)$; $\varphi_4 = (1, 1, 0);\quad \varphi_5 = (0, 1, 1);\quad \varphi_6 = (1, 1, 1)$; $\lvert\alpha_1\rvert = \lvert\alpha_2\rvert = \lvert\alpha_3\rvert;\quad \sphericalangle(\alpha_1, \alpha_2) = \sphericalangle(\alpha_2, \alpha_3) = 120°;\quad \sphericalangle(\alpha_1, \alpha_3) = 90°$	
	(a) $n_2 \leq n_1, n_3$	$q_3(n_2)$
	(b) $n_1 \leq n_2 \leq n_3$	$q_3(n_2) + r(n_2 - n_1 - 1, -1, n_2 + 1)$
	(c) $n_1 \leq n_3 \leq n_2$ and $n_1 + n_3 \leq n_2$	$r(n_1, n_3 + 1, n_1 + 1)$
	(d) $n_1 \leq n_3 \leq n_2$ and $n_2 \leq n_1 + n_3$	$r(n_1, n_3 + 1, n_1 + 1) - q_3(n_1 + n_3) + q_3(n_2)$
	Interchanging n_1 and n_3 in cases (b), (c), and (d) leads to analogous cases (b'), (c'), and (d'), respectively.	
	$q_3(n) = \frac{1}{6}(n + 1)(n + 2)(n + 3)$;	
	$r(n, k, m) = \sum_{j=0}^{n}(k - j)(m - j) = (n + 1)[km - \frac{1}{2}n(k + m) + \frac{1}{6}n(2n + 1)]$.	

A single ray takes the place of sectors of Fig. 1.

Our results for A_2, B_2, G_2, and A_3 are summarized in Table I, and the root vectors are also described there. Let us first discuss the Lie algebras of rank 2. For those cases which correspond to the outside regions of Fig. 1, the expressions for P follow from Lemma 2.1. For the remaining cases we obtain P by finding first a difference function, except for case (c) of G_2, where we have to start with a second difference function.

Let us illustrate this in detail for case (b) of B_2. We consider, for $0 \leq k \leq n_2$,

$$C(k, n_2) = P(2n_2 - k, n_2) - P(2n_2 - k - 1, n_2). \quad (3.1)$$

By Lemma 2.4, $C(k, n_2)$ is the partition function for the vectors (0, 1), (1, 1), (2, 1), and we conclude [cf. Lemma 2.1, Eq. (2.3), and Table I],

$$\begin{aligned} C(k, n_2) &= q(k; 2, 1) \\ &= b(k) - b(k - 1). \end{aligned} \quad (3.2)$$

For $k = 0$ we know that $P(2n_2, n_2) = q_2(n_2)$ [case (a)], and Eq. (3.2) remains valid since $b(-1) = 0$. Equations (3.1) and (3.2) now yield

$$P(2n_2 - k, n_2) = q_2(n_2) - b(k-1), \qquad (3.3)$$

as we assert in Table I. In the foregoing, the roles of n_1 and n_2 can be interchanged, and then the alternate expression for P results.

The last remark suggests a certain symmetry between n_1 and n_2, which seems to be lacking for case (c) of G_2. Indeed, one can construct an alternate expression for P for this case, to restore the symmetry. But this expression would be quite involved, and so is omitted.

The functions q in Table I can be evaluated by a repeated use of Eq. (2.2). For a numerical computation of $g(n)$ and $h(n)$ one can conveniently use the second difference functions, defined by

$$f^{(0)}(n) = f(n) \quad \text{and} \quad f^{(i+1)}(n) = f^{(i)}(n+1) - f^{(i)}(n).$$

We find [see Eq. (2.3)],

$$h^{(2)}(n) = b(n+2) = \tfrac{1}{4}(n+4)^2 \quad \text{for } n \text{ even},$$
$$= \tfrac{1}{4}(n+3)(n+5) \quad \text{for } n \text{ odd};$$
$$g^{(2)}(6n-2) = g^{(2)}(6n) = g^{(2)}(6n+2) = (n+1)^2,$$
$$g^{(2)}(6n-1) = n(n+1),$$
$$g^{(2)}(6n+1) = g^{(2)}(6n+3) = (n+1)(n+2). \qquad (3.4)$$

The Lie algebra A_3 remains to be discussed. Cases (a), (c), and (c′) follow directly from Lemmas 2.2 and 2.3. To determine P in the remaining cases, we have to consider appropriate difference functions, and the arguments are much the same as before. However, we shall not construct any diagrams here.

4. DISCUSSION OF THE PROPERTIES OF P

The form of the function P, especially in the case of B_2 and G_2, is rather striking. We will now show how this form can be related to the recursion formula (1.3).

This formula has the following meaning: We construct a polygon with vertices defined by the vectors $\sigma(g)$, and we label its vertices with alternating signs. [See Fig. 2(α).] We imagine it to be movable over the plane, with the orientation of its sides preserved. Then, if we place the polygon with its center anywhere except at $-g$, the alternating sum indicated by the vertices vanishes. [The exception corresponds to the condition $\mu \neq 0$ in Eq. (1.3).]

For definiteness, let us consider B_2, and let us refer to Fig. 2. We place the octagon at the position (β). We see that we must have $P = $ (const.) along the outside line. and, more generally, that $P(n_1, n_2)$ depends on only one of the arguments in each of the outside regions (a) and (c). Recall that $P(n_1, \cdots, n_k) = 0$ if some $n_i < 0$.

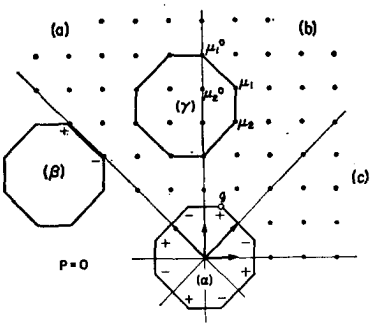


FIG. 2. Illustration of the recursion relation. The figure refers to B_2.

Now imagine the octagon inside the region (a), and compare with the position (γ). In region (a), P depends only on n_2, and we conclude that

$$P(\mu_1^0) - P(\mu_1) = P(\mu_2^0) - P(\mu_2); \qquad (4.1)$$

this difference remains constant along the vertical lines. This corresponds to the fact, that the term $b(k-1)$ in Eq. (3.3) depends only on k, and is independent of n_2.

It seems remarkable that the difference in Eq. (4.1) is that function which also expresses P in region (c), i.e. $b(n)$. There may be a deeper reason for this, or a more direct reason, but we have not been able to find one.

ACKNOWLEDGMENTS

I am greatly indebted to Professor B. Kostant for much help when I was making my acquaintance with the theory of Lie groups, and for suggesting this problem. I would like to thank Professor R. G. Sachs for the hospitality of the Physics Department of the University of Wisconsin during the completion of this work.

APPENDIX: EXPRESSIONS FOR REFLECTED VECTORS

To facilitate the computation of multiplicities by Kostant's formula (1.2), we give explicit expressions for the permuted vectors $\sigma(\mu)$, for the Lie algebras considered in this paper. The following discussion is based on references 1, 3, and 5.

Let us suppose that we have a Lie algebra $\mathfrak{g}$ of rank k, and that we have constructed in R^k the vectors α_i which correspond to the simple positive roots. (These vectors must form the specified angles.) Then the principal Weyl chamber consists of vectors $\beta \in R^k$ such that $\sphericalangle(\alpha_i, \beta) \leq 90°$ for all α_i. Successive reflections of the principal chamber in its walls

TABLE II. Reflected vectors. We denote (n_1, n_2) or (n_1, n_2, n_3) by ρ.

A_2	$\sigma_1\rho = (n_1, n_1 - n_2)$	$\sigma_5\rho = (n_2 - n_1, n_2)$ $\sigma_{j\pm 3}\rho = -\sigma_j(n_2, n_1)$
B_2	$\sigma_1\rho = (n_1, n_1 - n_2)$ $\sigma_2\rho = (2n_2 - n_1, n_2 - n_1)$	$\sigma_7\rho = (2n_2 - n_1, n_2)$ $\sigma_{j\pm 4}\rho = -\sigma_j\rho$
G_2	$\sigma_1\rho = (n_1, n_1 - n_2)$ $\sigma_2\rho = (3n_2 - n_1, 2n_2 - n_1)$ $\sigma_3\rho = (2n_1 - 3n_2, n_1 - 2n_2)$	$\sigma_{10}\rho = (2n_1 - 3n_2, n_1 - n_2)$ $\sigma_{11}\rho = (3n_2 - n_1, n_2)$ $\sigma_{j\pm 6}\rho = -\sigma_j\rho$
A_3	$\sigma_1\rho = (n_2 - n_1, n_2, n_3)$ $\sigma_2\rho = (n_2 - n_1, n_2, n_2 - n_3)$ $\sigma_3\rho = (n_1, n_2, n_2 - n_3)$ $\sigma_4\rho = (n_1, n_1 - n_3, n_2 - n_3)$ $\sigma_5\rho = (n_3 - n_2, n_3 - n_1, n_3)$ $\sigma_6\rho = (n_2 - n_3, n_3 - n_1, n_3)$	$\sigma_7\rho = (n_1, n_1 - n_3, n_1 - n_2)$ $\sigma_8\rho = (n_1, n_1 + n_3 - n_2, n_1 - n_2)$ $\sigma_9\rho = (n_1, n_1 + n_3 - n_2, n_3)$ $\sigma_{10}\rho = (n_3 - n_2, n_1 + n_3 - n_2, n_3)$ $\sigma_{11}\rho = (n_3 - n_2, n_1 + n_3 - n_2, n_1 - n_2)$ $\sigma_{j+12}\rho = -\sigma_j(n_3, n_2, n_1)$

yield all the other chambers. The group W is generated by such reflections. Given two chambers D^i and D^j, there is a unique $\sigma \in W$ which maps D^i on to D^j.

For B_2 and for G_2, some of the chambers correspond to the cases of Sec. 3, but this does not apply to A_2 nor to A_3 (and probably would not apply to any other simple Lie algebra). The chambers for A_2, B_2, and G_2 are shown in Fig. 3. They are so numbered that, for $\sigma \in W$,

$$\operatorname{sgn} \sigma = (-1)^{i+j} \quad \text{if} \quad \sigma : D^i \to D^j.$$

In addition, D^1 is the principal chamber. We shall denote by σ_{i+1} the element of W defined by

$$\sigma_{i+1} : D^1 \to D^i.$$

Therefore,

$$\operatorname{sgn} \sigma_i = (-1)^i. \tag{A1}$$

In Table II we summarize the action of the σ_i on vectors $\rho \in D^1$, for the four Lie algebras. If $\rho \in D^i \neq D^1$, then we can find the action of σ_i by considering $\sigma_i\sigma_{i+1}^{-1}$. For A_3 we do not picture the chambers, but the relation (A1) still applies. In case of the Lie algebras of rank 2, the assertions in this table can be obtained directly from the diagrams. In case of A_3, however, this method is not practical, and we proceed as follows.

We define a one-to-one correspondence between triples (n, n_2, n_3) and 4-tuples $(k_1, \cdots, k_4)$ satisfying

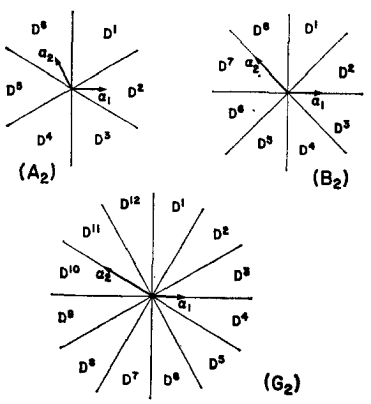


FIG. 3. Weyl chambers for the Lie algebras of rank 2.

$\sum k_i = 0$, by identifying k_i as the coefficient of the unit e_i in the sum $\sum n_i(e_i - e_{i+1})$. Explicitly, we have

$$(n_1, n_2, n_3) \leftrightarrow (n_1, n_2 - n_1, n_3 - n_2, -n_3).$$

A triple belongs to the principal chamber D^1 if and only if the numbers in the corresponding 4-tuple are decreasing. Equivalently, $(n_1, n_2, n_3) \in D^1$ if and only if

$$\tfrac{1}{2}n_2 \leq n_1, n_3 \quad \text{and} \quad n_1 + n_3 \leq 2n_2.$$

Now, the elements $\sigma \in W$ (together with their signatures $\operatorname{sgn} \sigma$) are in a one-to-one correspondence with the permutations of the 4-tuple. The latter can be easily examined, and the assertions in Table II can thus be obtained.

Lie Algebraic Solution of Linear Differential Equations

JAMES WEI AND EDWARD NORMAN
Socony Mobil Oil Company, Inc., Research Department, Paulsboro, New Jersey
(Received 13 November 1962)

The solution $U(t)$ to the linear differential equation $dU/dt = h(t)U$ can be represented by a finite product of exponential operators; In many interesting cases the representation is global. $U(t) = \exp[g_1(t)H_1] \exp[g_2(t)H_2] \ldots \exp[g_n(t)H_n]$ where $g_i(t)$ are scalar functions and H_i are constant operators. The number, n, of terms in this expansion is equal to the dimension of the Lie algebra generated by $H(t)$. Each term in this product has time-independent eigenvectors. Some applications of this solution to physical problems are given.

I. INTRODUCTION

CONSIDER a linear differential equation of the first order

$$dU(t)/dt = H(t)U(t), \qquad U(0) = I, \tag{1}$$

where H and U are time-dependent linear operators in a Banach space or a finite-dimensional space. The operator $H(t)$ is a given function of time. The initial value of U is the identity operator, and one seeks the solution U as a function of time.

The operator H often appears in physics as the Hamiltonian in an equation of motion, or as a transition probability matrix in a Master equation. It usually has a finite basis, and can be written as a finite sum

$$H(t) = \sum_{i=1}^{m} a_i(t)H_i, \tag{2}$$

where $a_i(t)$ are a set of linearly independent complex valued functions of time, and H_i are constant operators. The set of operators $\{H_1 \cdots H_m\}$ may be enlarged by repeated commutation to a Lie algebra L. The number of linearly independent operators H_i in L (the dimension of L) is often a finite number $n \geq m$; this is always the case if $H(t)$ is a finite matrix operator.

It will be shown that if L is finite dimensional, the system of Eq. (1) with a time-dependent operator can be uncoupled into a set of independent systems

$$U(t) = U_1(t)U_2(t) \cdots U_n(t), \tag{3}$$

where each component $U_i(t)$ is an operator satisfying

$$\frac{d}{dt} U_i(t) = \dot{g}_i(t)H_iU_i(t), \qquad U_i(0) = I, \tag{4}$$

where $\dot{g}_i(t)$ is a complex valued function of time. The solutions to (4) are $U_i(t) = \exp[g_i(t)H_i]$, and $U(t)$ becomes a product of exponentials

$$U(t) = \prod_{i=1}^{n} \exp[g_i(t)H_i]. \tag{5}$$

This is an exact solution.

The scalar functions $g_i(t)$ are the solutions to a set of nonlinear differential equations

$$\frac{d}{dt} g_i(t) = \sum_{i=1}^{n} \eta_{ik}a_k(t) \qquad g_i(0) = 0, \tag{6}$$

where η_{ik} are nonlinear functions of the g's. Thus we have reduced the linear differential equations of (1) to the nonlinear differential equations of (6), reminiscent of the reduction of the linear Master equation to the nonlinear Boltzmann equation. For the important case where L is a solvable Lie algebra, the η_{ik} form a triangular matrix, and the system (6) can be solved by quadrature.

Approximate and formal solutions of Eq. (1) in terms of a single exponential have been given by Birkhoff,[1] Feynman,[2] Kubo,[3] and Fer.[4] Magnus,[5,6] showed rigorously that the solution $U(t)$ can be expressed *locally* as

$$U(t) = \exp\left[\sum_{i=1}^{n} f_i(t)H_i\right], \tag{7}$$

where the $f_i(t)$ are a set of scalar functions of time satisfying a set of nonlinear differential equations. Magnus also showed that if two sets of operators have the same abstract Lie algebra and expansion functions $a_k(t)$, they will have solutions in the same form with the same functions $f_i(t)$; other properties of $H(t)$ become irrelevant to the problem. It will be shown that this often results in the reduction of many separate problems to one problem. If the dimension n of the Lie algebra is smaller than the dimension of the space on which $H(t)$ operates (this

[1] G. Birkhoff, J. Math. Phys. **16**, 104 (1937).
[2] R. Feynman, Phys. Rev. **84**, 108 (1951).
[3] R. Kubo, J. Chem. Phys. **20**, 770 (1952).
[4] F. Fer, Bull. Classe Sci. Acad. Roy. Belg. **44**, 5, 818 (1958).
[5] W. Magnus, Commun. Pure Appl. Math. **7**, 649 (1954).
[6] J. Mariani and W. Magnus, Research rept. no. BR-37, New York University, New York, 1961.

is certainly the case if the operand space is a Banach space), the difficulty of the problem is reduced.

These properties of Magnus' form of the solution of Eq. (1) are retained by the form given by Eq. (5). One of the advantages of (5) is that it is globally valid in a number of interesting cases. Another advantage of (5) is its relative ease in computation and added physical insight. As Eq. (7) is the exponential of a sum, $U(t)$ must be reevaluated for each value of t by a power series or an eigenfunction expansion; in Eq. (5), the eigenvalues and eigenfunctions of each constant operator H_i can be evaluated once for all times, and may possess some physical interpretation. The method of Magnus as outlined by Wichmann[7] requires a finite-dimensional representation of the Lie algebra—not necessarily a trivial problem unless the algebra is semisimple. No such representation will be required in the method of this paper.

A brief description of the Lie algebraic terms relevant to this paper is given in this section.[8,9]

The solution $U(t)$ lies in a subset of all linear operators GL; this subset is called the associative algebra R generated by $H(t)$. The algebra R consists of all operators in $H(t)$ plus all products

$$H_{\alpha_1}^{\beta_1} H_{\alpha_2}^{\beta_2} H_{\alpha_3}^{\beta_3} \cdots H_{\alpha_r}^{\beta_r}, \qquad \alpha_i = 1 \text{ to } m, \quad \beta_i = 1, 2, 3 \cdots,$$

and all linear combinations of such products.

Every associative algebra gives rise to a Lie or commutator algebra L. We shall consider the Lie algebra generated by $H(t)$. The Lieproduct of elements H_i and H_k is $[H_i, H_k]$ or $H_iH_k - H_kH_i$. The Lie algebra L consists of all operators in $H(t)$ plus all Lieproducts

$$[H_{\alpha_1}[H_{\alpha_2}[H_{\alpha_3} \cdots [H_{\alpha_{r-1}}, H_{\alpha_r}] \cdots]]] \quad \alpha_i = 1 \text{ to } m,$$

and all linear combinations of such products.

Every element of L generated by $H(t)$ is also an element of R generated by $H(t)$. The reverse is not true; $H_1H_2 - H_2H_1$ is a Lie element, but H_1H_2, in general, is not. The Lie algebra L is thus a subset of R, and is defined by the commutation *structural constants*

$$[H_i, H_k] = \sum_{r=1}^{n} \gamma_{ik}^r H_r.$$

Two Lie algebras of operators have the same abstract Lie algebra if they have corresponding bases which give rise to the same structural constants.

An example of a Lie algebra is the set $\{Q, P, I\}$ with the commutation rules

$$[Q, P] = i\hbar I, \qquad [Q, I] = 0, \qquad [P, I] = 0.$$

A subset S of L is called a *subalgebra* if it is closed under the operations of addition, multiplication by a scalar and commutation. A subalgebra S is called an *ideal* if the commutator $[X, Y]$ of $X \in S$ and $Y \in L$ is in S.

The set of those elements of L which are the result of commutation of two Lie elements forms the *derived algebra* L'. L' is an ideal of L. The derived algebra of L' is denoted by L''. By induction, one forms the *derived series*

$$L \supset L' \supset L'' \supset \cdots.$$

A Lie algebra L is said to be *solvable* if $L^{(h)} = \{0\}$ for some h; this clearly means the rest of the terms in the derived series are also zero.

The union of two solvable ideals is again a solvable ideal; the union of all solvable ideals is called the *radical*. An algebra is said to be *semisimple* if its radical is $\{0\}$. An algebra is said to be *simple* if it has no ideal other than L and $\{0\}$, and if $L' \neq \{0\}$.

The *lower central series* is constructed by relabeling $L^2 = L'$, and defining L^{n+1} to be all the elements that result from commutation of elements from L with elements from L^n, i.e. $L^{n+1} = [L, L^n]$. $L \supset L^2 \supset L^3 \supset \cdots$. A Lie algebra L is said to be *nilpotent* if $L^h = \{0\}$ for some h.

Examples of these terms will be given in Sec. III.

II, THE UNCOUPLING THEOREM

The results of this section are contained in a paper by the present authors.[10] It is, however, worthwhile to present here the proofs of Eqs. (5) and (6) so that the calculations in Sec. III of this paper will be meaningful.

The main theorem is based upon two lemmas.

(i) (Baker–Hausdorff) If $X, Y \in L$, then $e^X Y e^{-X} \in L$, and $e^X Y e^{-X} = Y + [X, Y] + [X, [X, Y]]/2! + [X, [X, [X, Y]]]/3! + \cdots$. Let us define the linear operator adX, $X \in L$ by the equation

$$(adX)Y = [X, Y], \qquad Y \in L.$$

Powers of the operator adX are defined in the usual way and thus

$$(adX)^2 Y = [X, [X, Y]], \quad \text{etc.}$$

[7] E. H. Wichmann, J. Math. Phys. 2, 876 (1961).

[8] N. Jacobson, *Lie Algebras* (Interscience Publishers, Inc., New York, 1962).

[9] N. Bourbaki, *Elements de Mathematique* (Hermann & Cie, Paris, France, 1961), Vol. 26, Chap. on "Algebres de Lie."

[10] J. Wei and E. Norman, "On Global Representations of the Solutions of Linear Differential Equations as a Product of Exponentials" (to be published in Proc. Am. Math. Soc.).

In terms of adX the Baker–Hausdorff formula can be rewritten as

$$e^X Y e^{-X} = (e^{adX})Y. \tag{8}$$

(ii) Let $H_1, H_2, \cdots, H_n$ be a basis for L. Then

$$\left[\prod_{j=1}^{r} \exp(g_j H_j)\right] H_i \left[\prod_{j=r}^{1} \exp(-g_j H_j)\right] = \sum_{k=1}^{n} \xi_{ki} H_k, \qquad r = 1, \cdots, n, \tag{9}$$

where each $\xi_{ki} = \xi_{ki}(g_1, \cdots, g_r)$ is an analytic function of its arguments.

The proof of lemma (i) is accomplished by direct computation. Equation (9) of lemma (ii) follows by repeated application of lemma (i), and the analyticity of the ξ_{ki} is easily established.

Theorem. Suppose the linear operator $H(t)$ can be expressed in the form

$$H(t) = \sum_{i=1}^{m} a_i(t) H_i \quad m \text{ finite}, \tag{10}$$

where the $a_i(t)$ are scalar functions of time, and the H_i are time-independent operators. Let the Lie algebra L generated by $H(t)$ be of finite dimension n. Then there exists a neighborhood of $t = 0$, in which the solution of the equation

$$dU/dt = H(t)U, \qquad U(0) = I \tag{11}$$

may be expressed in the form

$$U(t) = \exp[g_1(t)H_1] \times \exp[g_2(t)H_2] \cdots \exp[g_n(t)H_n], \tag{12}$$

where $H_1, \cdots, H_n$ is a basis for L, and the $g_i(t)$ are scalar functions of time. Moreover, the $g_i(t)$ depend only on the Lie algebra L and the $a_i(t)$.

Proof: First note that we might just as well write $A(t) = \sum_{i=1}^{n} a_i(t)H_i$ instead of $A(t) = \sum_{i=1}^{m} a_i(t)H_i$, by simply setting $a_i(t) \equiv 0$ for $i > m$. Note also that at time $t = 0$, $U(0) = I$ is in the form (12) with all $g_i(t) = 0$.

Now let U be of the form (12). Since

$$\frac{dU}{dt} = \sum_{i=1}^{n} \dot{g}_i(t) \left[\prod_{j=1}^{i-1} \exp(g_j H_j)\right] \times X_i \left[\prod_{j=i}^{n} \exp(g_j H_j)\right], \tag{13}$$

$$AU = \sum_{i=1}^{n} a_i(t) H_i \cdot U.$$

We obtain, upon substitution of (13) into (12) and post-multiplication by the inverse operator U^{-1},

$$\sum_{i=1}^{n} a_i(t) H_i = \sum_{i=1}^{n} \dot{g}_i(t) \left[\prod_{j=1}^{i-1} \exp(g_j H_j)\right] \times H_i \left[\prod_{j=i-1}^{1} \exp(-g_j H_j)\right] = \sum_{i=1}^{n} \dot{g}_i(t) \left[\prod_{j=1}^{i-1} \exp\{g_j ad H_j\}\right] H_i. \tag{14}$$

Application of lemma (ii) to the terms on the right of Eq. (14) yields

$$\sum_{k=1}^{n} a_k(t) H_k = \sum_{i=1}^{n} \sum_{k=1}^{n} \dot{g}_i(t) \xi_{ki} H_k. \tag{15}$$

Since the operators H_k are linearly independent, we have a linear relation between the $a_k(t)$ and the $\dot{g}_i(t)$. The elements ξ_{ki} of the transform matrix ξ are analytic functions of the g_i's;

$$\begin{matrix} a \\ \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \end{matrix} = \begin{matrix} \xi \\ \begin{Vmatrix} \xi_{11} & \cdots & \xi_{1n}^{i} \\ & & \\ \xi_{n1} & \cdots & \xi_{nn} \end{Vmatrix} \end{matrix} \begin{matrix} \dot{g} \\ \begin{bmatrix} \dot{g}_1 \\ \dot{g}_2 \\ \vdots \\ \dot{g}_n \end{bmatrix} \end{matrix}, \qquad g(0) = 0. \tag{16}$$

Since the ξ_{ki} are analytic functions of g, we have that the determinant Δ of ξ is an analytic function of g. We also know that at $t = 0$ $\xi = I$, and hence $\Delta(0) \neq 0$. These two facts show that there must exist a neighborhood N_0 of $t = 0$ in which $\Delta \neq 0$, i.e. in which ξ is invertable. We can thus write (16) in the form

$$dg/dt = f(a, g) = \xi^{-1}a, \qquad g(0) = 0. \tag{17}$$

Since ξ^{-1} is analytic in N_0, we are assured of a neighborhood of $t = 0$ in which the solution of (17) exists and is unique. This completes the proof of the theorem.

The authors have shown[11] that the uncoupling theorem is global for *all* solvable Lie algebras, and for the real "split 3-dimensional" simple Lie algebra.

III. SOME PHYSICAL EXAMPLES

(a) The method of uncoupled exponentials will be illustrated in this section by examples from low-dimensional Lie algebras. Many physical systems give rise to low-dimensional Lie algebras.

There is only one one-dimensional Lie algebra, which is of course Abelian. For *any* Abelian Lie algebra, $e^X e^Y = e^{X+Y}$, and the solution of our problem is trivial.

[11] Reference 10, Theorem 2.

There are two distinct two-dimensional Lie algebras, one of which is Abelian. The non-Abelian algebra was discussed by Sack[12] as the "Quantum Mechanical Shift Operator." This algebra is spanned by the operators X and Y with the commutation rules

$$[X, Y] = Y. \tag{18}$$

The Landau–Teller transition probabilities of a system of simple harmonic oscillators[13] is a realization of the above algebra in an infinite-dimensional space. Defining the raising and lowering operators as

$$R = \sum_{j=0}^{\infty} (j+1)(-E_{j,j} + E_{j+1,j}),$$

$$S = \sum_{j=1}^{\infty} j(-E_{j,j} + E_{j-1,j}),$$

we obtain the commutator

$$[R, S] = R + S = \sum_{j=0}^{\infty} \{-(2j+1)E_{j,j} + (j+1)(E_{j,j+1} + E_{j+1,j})\},$$

where E_{jk} is the infinite matrix in which all elements are zero except in the (j, k) position where the entry is unity. This algebra is abstractly identical to the algebra of (18) where R is isomorphic to X, and $R + S$ is isomorphic to Y.

A second realization is provided by the kinetics of the deuterium exchange reaction.[14] It is again a Master equation problem with only nearest-neighbor transition probabilities, but now the operand space is finite dimensional. The operators are given by

$$R' = \sum_{j=0}^{s-1} (s-j)(-E_{j,j} + E_{j+1,j}),$$

$$S' = \sum_{j=1}^{s} j(-E_{j,j} + E_{j-1,j}),$$

which possess the commutator

$$[R', S'] = R' - S' = \sum_{j=0}^{s} (2j-s)E_{j,j} - \sum_{j=0}^{s-1} (j+1)E_{j,j+1} + \sum_{j=0}^{s-1} (s-j)E_{j+1,j},$$

where $s + 1$ is the dimension of the operand space, and E_{jk} is an $(s+1) \times (s+1)$ matrix. This algebra is again isomorphic to the algebra of (18) where $-S'$ is isomorphic to X, and $R' - S'$ is isomorphic to Y. These operators also arise in the problem of the servicing of machines in queueing theory.[15]

This algebra is solvable since the derived series is

$$L = \{X, Y\}, \quad L' = \{Y\}, \quad L'' = \{0\}.$$

The solution of the equation

$$dU/dt = H(t)U = (a(t)X + b(t)Y)U \qquad U(0) = I$$

can be written as

$$U(t) = \exp [f(t)X \cdot \exp g(t)Y].$$

Equation (14) gives $aX + bY = \dot{f}X + \dot{g}(e^{adfX})Y$, and Eq. (8) gives $(e^{adfX})Y = e^{f}Y$. One obtains then

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{Vmatrix} 1 & 0 \\ 0 & e^{f} \end{Vmatrix} \begin{bmatrix} \dot{f} \\ \dot{g} \end{bmatrix},$$

or

$$\dot{f} = a$$

$$\dot{g} = be^{-f},$$

which can easily be solved by quadrature. Thus we have reduced many different problems to the same quadrature problem.

(b) The simplest non-Abelian three-dimensional Lie algebra is spanned by the operators $\{Q, P, I\}$ with the commutation rules

$$[Q, P] = i\hbar I, \quad [Q, I] = 0, \quad [P, I] = 0. \tag{19}$$

This algebra is solvable, and it is well known that

$$e^{Q+P} = e^{Q}e^{P}e^{i\hbar I/2}.$$

The operators for the scattering of x rays by crystal lattices[16] form a 4-dimensional Lie algebra which is an easy extension of the algebra of (19). It is spanned by the operators $\{Q^2 + P^2, Q, P, I\}$ with the additional commutation rules

$$[P^2 + Q^2, Q] = -2i\hbar P, \quad [P^2 + Q^2, P] = 2i\hbar Q, \quad [P^2 + Q^2, I] = 0. \tag{20}$$

These two algebras are solvable since

$$L = \{P^2 + Q^2, Q, P, I\}, \quad L' = \{Q, P, I\},$$

$$L'' = \{I\}, \quad L''' = \{0\},$$

The algebra of (19), (20) can be expressed more conveniently in terms of the bases

$$\{W \equiv P^2 + Q^2, X \equiv Q - iP, Y \equiv Q + iP, I\};$$

the commutation rules become

$$[W, X] = 2\hbar X, \quad [W, Y] = -2\hbar Y, \quad [X, Y] = -2\hbar I.$$

If an equation of motion is written as

[12] R. A. Sack, Phil. Mag. **3**, 497 (1958).
[13] E. W. Montroll and K. E. Shuler, Advan. Chem. Phys. **1**, 371 (1959).
[14] J. R. Anderson and C. Kemball, Proc. Roy. Soc. (London) **A226**, 472 (1954).
[15] W. Feller, *An Introduction to Probability Theory and Applications* (John Wiley & Sons, Inc., New York, 1957), p. 416.
[16] G. H. Weiss and A. A. Maradudin, J. Math. Phys. **3**, 771 (1962).

$$dU/dt = \{a_1(t)W + a_2(t)X + a_3(t)Y + a_4(t)I\}U,$$

and the solution is

$$U = \exp[g_1(t)W]\cdot\exp[g_2(t)X]\cdot\exp[g_3(t)Y]\cdot\exp[g_4(t)I],$$

we have from Eqs. (8) and (14):

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{2\hbar g_1} & 0 & 0 \\ 0 & 0 & e^{-2\hbar g_1} & 0 \\ 0 & 0 & -2\hbar g_2 & 1 \end{Vmatrix} \begin{bmatrix} \dot{g}_1 \\ \dot{g}_2 \\ \dot{g}_3 \\ \dot{g}_4 \end{bmatrix},$$

or

$$\dot{g}_1 = a_1, \qquad \dot{g}_2 = e^{-2\hbar g_1}a_2,$$
$$\dot{g}_3 = e^{2\hbar g_1}a_3, \qquad \dot{g}_4 = a_4 + 2\hbar g_2 e^{2\hbar g_1}a_3 .$$

These equations are easily solved by quadrature.

All of the algebras discussed in examples (a) and (b) are solvable; hence the given solutions are global.

(c) A celebrated three-dimensional Lie algebra is the "split 3-dimensional simple algebra" characterized by

$$[E, F] = H, \quad [E, H] = 2E, \quad [F, H] = -2F. \tag{21}$$

These rules of commutation are satisfied in a finite-dimensional operand space by the three components of angular momentum for a rotator[17] where

$$E = (m_x + im_y)/i\hbar, \qquad F = (m_x - im_y)/i\hbar,$$
$$H = -2m_z/\hbar.$$

In an infinite-dimensional operand space, this is realized by the simple harmonic oscillator[18] where

$$E = Q^2/2i\hbar, \quad F = P^2/2i\hbar, \quad H = (QP + PQ)/2i\hbar.$$

Given this Lie algebra and the dimension of the representation, it is known that these are the only irreducible representations in the sense of isomorphism.

For a simple harmonic oscillator where the restoring force is time-dependent, the wave equation may be written as

$$dU/dt = (a(t)H + b(t)E + c(t)F)U,$$

and the operator U is expanded as

$$U(t) = \exp[h(t)H]\cdot\exp[g(t)E]\cdot\exp[f(t)F];$$

then Eqs. (8) and (14) give

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{Vmatrix} 1 & 0 & g \\ 0 & e^{-2h} & g^2e^{-2h} \\ 0 & 0 & e^{+2h} \end{Vmatrix} \begin{bmatrix} \dot{h} \\ \dot{g} \\ \dot{f} \end{bmatrix},$$

or

$$\dot{h} = a - ge^{-2h}c, \quad \dot{g} = be^{2h} - g^2e^{-2h}c, \quad \dot{f} = e^{-2h}c. \tag{22}$$

This algebra is not solvable, and the scalar functions h, g, and f cannot be found by quadrature. The solution of the system (22) is reduced to quadrature after solving the Riccati equation

$$\dot{u} - u^2 + p(t)u + q(t) = 0,$$
$$p(t) = -\dot{c}/c, \quad q(t) = bc - a^2 - \dot{a} + (\dot{c}/c)a,$$

satisfied by $\dot{h}(t)$.

The general continuous-time Markov process in three-dimensional space consists of six independent operators which form a Lie algebra $\{E_{21} - E_{11}; E_{31} - E_{11}; E_{12} - E_{22}; E_{32} - E_{22}; E_{31} - E_{33}; E_{32} - E_{33}\}$. This algebra is not solvable. A general reduction principle and method will be given in the next section which can be used to decompose any Lie algebra into the solvable radical and a number of simple subalgebras. We shall be content to state the results of the decomposition here.

The Lie algebra L is decomposed to the sum

$$L = R + S,$$

where R is the solvable radical spanned by

$$A = \begin{Vmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 1 & 0 \end{Vmatrix}, \qquad B = \begin{Vmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{Vmatrix},$$

$$C = \begin{Vmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{Vmatrix},$$

and S is the simple subalgebra spanned by

$$E = \begin{Vmatrix} 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{Vmatrix}, \qquad F = \begin{Vmatrix} -1 & -\frac{1}{2} & 0 \\ 2 & 1 & 0 \\ -1 & -\frac{1}{2} & 0 \end{Vmatrix},$$

$$H = \begin{Vmatrix} -1 & -1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{Vmatrix}.$$

S, as a subalgebra, has the commutation rule of Eq. (21). The rest of the multiplication table is

[17] P. A. M. Dirac, *Quantum Mechanics* (Oxford University Press, New York, 1957).
[18] Reference 17.

	E	F	H	A	B	C
A	0	0	0	0	$-B$	$-C$
B	C	0	$-B$	B	0	0
C	0	$-B$	C	C	0	0.

In this example and in example (c), the solution is again global.

Another complicated algebra is that of the proper orthochronous inhomogeneous Lorentz group.[19] The Lie algebra is 11-dimensional with the following elements and commutation rules:

$$\begin{aligned}
&\{I, H, P_1, P_2, P_3, J_1, J_2, J_3, g_1, g_2, g_3\};\\
&[H, P_i] = [H, J_i] = 0; \quad [H, g_i] = -iP_i;\\
&[P_i, P_j] = 0; \quad [P_i, J_j] = i\epsilon_{ijk}P_k; \qquad (23)\\
&[P_i, g_j] = -i\delta_{ij}I;\\
&[J_i, J_j] = i\epsilon_{ijk}J_k; \quad [J_i, g_j] = i\epsilon_{ijk}g_k;\\
&[g_i, g_j] = -i\epsilon_{ijk}J_k,
\end{aligned}$$

where

$$\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1,$$

$$\epsilon_{321} = \epsilon_{132} = \epsilon_{213} = -1;$$

all other ϵ_{ijk} are zero.

It will be shown that this algebra can be decomposed into a five-dimensional radical and two three-dimensional simple subalgebras.

IV. REDUCTION PRINCIPLE AND METHOD

If, in the equation $dU/dt = H(t)U$, the Lie algebra L generated by $H(t)$ is solvable; then Eq. (17) can be solved by quadrature. The task for high-dimensional algebras is more tedious—but no more difficult—than for low-dimensional algebras. On the other hand, if L is not solvable, then the difficulties in solving (17) mount rapidly as the dimension of L increases. In this section, we present a method for reducing a Lie algebra to its component parts. If the algebra is solvable, no reduction is needed; if it is simple, no reduction is possible. However, as indicated in the preceding section, there do exist physically interesting algebras which are neither solvable nor simple.

First we shall present the relevant theoretical results, and show how they are used in the reduction of our problem. Proofs of these results may be found in the literature,[20,21] but these proofs are of the nature of existence proofs and thus are not very useful in practice.

[19] J. S. Lomont and H. E. Moses, J. Math. Phys. 3, 405 (1962).
[20] Reference 8.
[21] Reference 9.

Definition 1. A Lie algebra L is said to be the *semidirect* sum of the subalgebra L_1 and ideal L_2, $L = L_1 \oplus L_2$, if $L_1 \cap L_2 = 0$, and if every element $x \in L$ can be written uniquely as a sum $x = x_1 + x_2$ with $x_1 \in L_1$ and $x_2 \in L_2$. If L_1 and L_2 are both ideals of L then the sum is said to be *direct*.

The definition can be extended in an obvious way to any number of summands.

Structure theorem. A finite-dimensional semisimple Lie algebra L may be decomposed into the *direct* sum $L = L_1 \oplus L_2 \oplus \cdots \oplus L_r$, where the L_i are ideals which are simple algebras.

Levi's theorem. If L is a finite-dimensional Lie algebra with radical R, then there exists a semisimple subalgebra S of L such that L is the semidirect sum $L = S \oplus R$.

In the equation $dU/dt = H(t)U$, where $H(t)$ generates L, the decomposition $L = S \oplus R$ gives rise to the corresponding decomposition, $H = H_S + H_R$ of H, where $H_R \in R$ and $H_S \in S$. Let

$$dU/dt = HU = (H_S + H_R)U. \qquad (24)$$

Define U_R and U_S by

$$\frac{dU_S}{dt} = H_S U_S, \quad \frac{dU_R}{dt} = (U_S^{-1} H_R U_S) U_R.$$

Since R is an ideal in L, we see [by lemma (i)] that $U_S^{-1} H_R U_S$ is in R. It is easy to verify that $U = U_S U_R$ satisfies (24). The fact that R is solvable makes it easy to find U_R once U_S has been found. To reduce the task of finding U_S we make use of the structure theorem. Write $S = S_1 \oplus S_2 \oplus \cdots \oplus S_r$. Then $H = H_1 + \cdots + H_r$ where $H_i \in S_i$. Making use of the fact that $[U_i, H_j] = 0$ for $i \neq j$, it is easy to see that $U_S = U_1 U_2 \cdots U_r$, where U_i satisfies the equation $dU_i/dt = H_i U_i$.

We shall now describe a method by which one can carry out in practice, the reduction embodied in the structure theorem and Levi's theorem.

Definition 2. A subalgebra $\mathfrak{C}$ of a Lie algebra L is called a *Cartan subalgebra* if (1) $\mathfrak{C}$ is nilpotent and (2) the set of all $x \in L$ such that $[x, h] \in \mathfrak{C}$ for every $h \in \mathfrak{C}$ (the *normalizer* of $\mathfrak{C}$) is simply $\mathfrak{C}$ itself.

The construction of a Cartan subalgebra is the first step in carrying out the reduction. Frequently a Cartan subalgebra can be found by inspection. Failing this, one may be constructed by first finding a regular element.

Definition 3. An element $h \in L$ is called *regular* if the number of independent elements $x \in L$ so that $(\mathrm{ad}h)^k x = 0$ for some k, is minimal.

If h is a regular element, then the set $\mathfrak{C}$ defined

by $\mathfrak{K} = \{x \in L \mid (\mathrm{ad}h)^k x = 0 \text{ for some } k\}$ is a Cartan subalgebra.

There exists a straightforward method for finding a regular element.[22]

We will now illustrate the way in which a Cartan subalgebra is used to effect the decomposition of L. The general method is easily inferred from the example.

A Cartan subalgebra of the proper orthochronous inhomogeneous Lorentz group can be determined by inspection to be $\mathfrak{K} = \{I, H, g_1, J_1, P_1\}$. Next, we determine all the eigenvalues of the linear operators ad$I$, ad$H$, etc. The matrices corresponding to these linear transforms are quite simple, and the eigenvalues are readily found and given in Table I.

Next we define an element x of L to be a generalized eigenvector of h with eigenvalue λ if $(\mathrm{ad}h - \lambda)^k x = 0$ for some positive integer k. The generalized eigenvectors are found and tabulated in Table II.

Next, by a suitable linear combination, one obtains a set of simultaneous eigenvectors to these five operators, that span L. The simultaneous generalized eigenvectors and the corresponding eigenvalues are tabulated in Table III.

At this stage, it is convenient to remove the elements belonging to the radical from Table III. The method of finding the radical of a Lie algebra is quite straightforward, and is fully described by Jacobson.[23] The radical of this algebra is found to be $R = \{I, H, P_1, P_2, P_3\}$. After deletion of the radical, Table III becomes Table IV.

The contents of Table IV forms a semisimple algebra of six dimensions. Thus we have accomplished the decomposition given by Levi's theorem, $L = S \oplus R$. The next step is to decompose S into simple subalgebras according to the structure theorem. By a linear combination of the adjoint operators, we replace Table IV by Table V. By noting that $[B, A] = 2(J_1 - ig_1)$ and $[D, C] = 2(J_1 + ig_1)$, the decomposition is now complete:

$$L = L_1 \oplus L_2 \oplus R,$$

where

$$R = \{I, H, P_1, P_2, P_3\},$$

$$L_1 = \{J_1 - ig_1,\ (g_2 - J_3) - i(g_3 + J_2),\ (g_2 + J_3) + i(g_3 - J_2)\},$$

$$L_2 = \{J_1 + ig_1,\ (g_2 + J_3) - i(g_3 - J_2),\ (g_2 - J_3) + i(g_3 + J_2)\}.$$

[22] Reference 8, p. 60.
[23] Reference 8, p. 73.

TABLE I.

operator	eigenvalues	multiplicity
ad I	0	11
ad H	0	11
ad P_1	0	11
ad J_1	0, 1, −1	5, 3, 3
ad g_1	0, i, −1	7, 2, 2

TABLE II.

operator	eigenvalues	generalized eigenvectors
ad I	0	all elements of L
ad H	0	all elements of L
ad P_1	0	all elements of L
ad g_1	0	$I, H, g_1, J_1, P_1, P_2, P_3$
	i	$g_2 - J_3,\ g_3 + J_2$
	$-i$	$g_2 + J_3,\ g_3 - J_2$
ad J_1	0	I, H, g_1, J_1, P_1
	1	$g_2 - ig_3,\ J_2 - iJ_3,\ P_2 - iP_3$
	−1	$g_2 + ig_3,\ J_2 + iJ_3,\ P_2 + iP_3$

TABLE III.

eigenvector	ad I	ad H	ad P_1	ad g_1	ad J_1
$P_2 + iP_3$	0	0	0	0	−1
$P_2 - iP_3$	0	0	0	0	1
$(g_2 - J_3) + i(g_3 + J_2)$	0	0	0	i	−1
$(g_2 + J_3) + i(g_3 - J_2)$	0	0	0	$-i$	−1
$(g_2 + J_3) - i(g_3 - J_2)$	0	0	0	$-i$	1
$(g_2 - J_3) - i(g_3 + J_2)$	0	0	0	i	1

TABLE IV.

eigenvector	ad g_1	ad J_1
$D = (g_2 - J_3) + i(g_3 + J_2)$	i	−1
$B = (g_2 + J_3) + i(g_3 - J_2)$	$-i$	−1
$C = (g_2 + J_3) - i(g_3 - J_2)$	$-i$	1
$A = (g_2 - J_3) - i(g_3 + J_2)$	i	1

TABLE V.

eigenvector	ad$(J_1 - ig_1)$	ad$(J_1 + ig_1)$
A	2	0
B	−2	0
C	0	2
D	0	−2

ACKNOWLEDGMENT

The authors are indebted to Professor Wilhelm Magnus for his stimulating discussions and encouragements.

Erratum: Zero-Point Energy of an Electron Lattice

[J. Math. Phys. 1, 395 (1960)]

ROSEMARY A. COLDWELL-HORSFALL
Department of Physics, The University, Sheffield 10, England
AND
ALEXEI A. MARADUDIN
Westinghouse Research Laboratories, Pittsburgh 35, Pennsylvania

An error was made in the calculation of the eighth moment of the frequency spectrum of the electron lattice. Instead of the values quoted in Eq. (4.15) read

$$u_8 = 0.085125, \qquad v_8 = 0.424824.$$

The subsequent equations now become

$$E_0/N = \tfrac{3}{2}\hbar\omega_p(0.553467), \tag{4.16}$$

$$c_2 = 23.548, \qquad c_4 = -52.206, \tag{4.18}$$

with the extrapolated value

$$E_0/N = \tfrac{3}{2}\hbar\omega_p(0.5143). \tag{4.19}$$

Using the more accurate value

$$c_2 = 29.984, \tag{4.20}$$

$$c_4 = -107.99, \quad \text{and} \quad c_6 = 83.670. \tag{4.21}$$

The extrapolated value of the zero-point energy using these three c_{2n} is

$$E_0/N = \tfrac{3}{2}\hbar\omega_p(0.5116) = 2.658/r_s^{\frac{1}{2}}\ \text{Ry}.$$

Thus we see that the result obtained by the moment-trace method is in very good agreement with Carr's result, $2.66/r_s^{\frac{1}{2}}$ Ry, and also with the value $2.661/r_s^{\frac{1}{2}}$ Ry obtained by Laster[1] who has computed the frequency histogram for 16 000 wave vectors in the first Brillouin zone. We would like to thank G. Bambakidis for finding the error in u_8, and S. J. Laster for communicating his results to us.

[1] S. J. Laster, Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science, Southern Methodist University (1962).