JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 4, NUMBER 5 MAY 1963

Structure of the Two-Point Function

O. STEINMANN
Institute for Advanced Study, Princeton, New Jersey
(Received 26 December 1962)

The two-point functions of relativistic quantum field theory are discussed in the framework of distribution theory. A derivation in this framework of the Källén–Lehmann representation is given. The singularities in x space are studied for the example of the D_F function. It is in general not possible to extract a singularity with support on the light cone. It is shown however that D_F is at best a measure in any neighborhood of the light cone. Thus Lehmann's statement that D_F is at least as singular as the corresponding free-field function is confirmed.

1. INTRODUCTION

THE structure of the two-point functions in relativistic field theory has been discussed by G. Källen[1] and H. Lehmann.[2] Their main result is a representation of the various two-point functions (see below for their definition) in terms of the corresponding free-field functions. It is the purpose of the present paper to give a more detailed derivation of this representation in the framework of distribution theory and to discuss with its help the nature of the two point function in x space, in particular on the light cone.

We consider the theory of a scalar local field $A(x)$ subjected to the well-known Wightman axioms.[3] The two-point functions of this theory are defined by

$$D^+(\xi) = i\langle A(x)A(y)\rangle_0, \qquad \xi = x - y,$$
$$D^-(\xi) = -D^+(-\xi), \quad D(\xi) = D^+(\xi) + D^-(\xi), \quad (1)$$
$$D^1(\xi) = D^+(\xi) - D^-(\xi),$$

and

$$D_R(\xi) = \theta(\xi)\, D(\xi), \qquad D_A(\xi) = D_R(-\xi),$$
$$D_F(\xi) = \tfrac{1}{2}[D_R(\xi) + D_A(\xi)], \quad (2)$$
$$D_F(\xi) = \theta(\xi)\, D^+(\xi) - \theta(-\xi)\, D^-(\xi),$$

where

$$\theta(\xi) = \begin{cases} 1 & \text{if } \xi^0 > 0, \\ 0 & \text{if } \xi^0 < 0. \end{cases}$$

The functions of the first group are called homogeneous; those of the second group inhomogeneous.

The corresponding free-field functions to mass m will be denoted $\Delta^+(\xi, m^2)$, $\Delta^-(\xi, m^2)$, etc.

In Secs. 2–4 we shall assume the existence of a nonvanishing lowest mass $\mu > 0$ in the theory. The case $\mu = 0$ will be discussed in Sec. 5.

2. THE HOMOGENEOUS FUNCTIONS

As is well known,[3] the Fourier transform

$$\tilde{D}^+(p) = -i(2\pi)^{-1} \int d^4p\, e^{i(p,\xi)}\, D^+(\xi) \quad (3)$$

of the lowest Wightman function is a Lorentz-invariant positive tempered measure with support in

$$V_+^\mu = \{p : p^2 \geq \mu^2, p_0 > 0\}.$$

The set of measures with these properties can be mapped, one-to-one, onto the set of one-dimensional positive tempered measures $\rho(u)$ with support in $u \geq \mu^2$. This mapping is defined as follows[4,5]:

[1] G. Källén, Helv. Phys. Acta 25, 416 (1952).
[2] H. Lehmann, Nuovo Cimento 11, 342 (1956).
[3] A. S. Wightman, Phys. Rev. 101, 860 (1956).
[4] P. D. Methée, Commun. Math. Helv. 28, 225 (1954).
[5] L. Garding, Nuovo Cimento Suppl. 14, 45 (1959).

Let $\psi(p) \in S_4$, where S_4 is the space of infinitely differentiable functions of the 4-vector p with fast decrease at infinity.[6] Then

$$\tilde{\psi}(u) = \int_{V_+} \delta(p^2 - u)\psi(p)\, d^4p, \qquad u \geq \mu^2 \tag{4}$$

is in $u \geq \mu^2$ an infinitely differentiable function with fast decrease at infinity, and can be extended to a function in S_1.

Let $\gamma(p) \in S_4$ be a nonnegative function with the property

$$\tilde{\gamma}(u) \equiv 1. \tag{5}$$

To every function $\tilde{\psi}(u) \in S_1$ corresponds then a function $\psi(p) \in S_4$:

$$\psi(p) = \gamma(p)\tilde{\psi}(p^2). \tag{6}$$

We define the mapping $\tilde{D}^+ \to \rho$ by

$$(\rho, \tilde{\psi}) \equiv (\tilde{D}^+, \gamma\tilde{\psi}) \quad \text{for all} \quad \tilde{\psi} \in S_1. \tag{7}$$

This definition is independent of the special choice of the arbitrary function γ. The measure ρ defined by (7) is obviously positive if $\tilde{D}^+$ is positive.

Conversely, for every positive tempered measure $\rho(u)$ in $u \geq \mu^2$ there exists a $\tilde{D}^+(p)$, defined by

$$(\tilde{D}^+, \psi) \equiv (\rho, \tilde{\psi}). \tag{8}$$

We write this mapping $\rho \to \tilde{D}^+$ symbolically:

$$\tilde{D}^+(p) = \int du\rho(u)\theta(p_0)\ \delta(p^2 - u). \tag{9}$$

This is the Källen-Lehmann representation of $\tilde{D}^+$.

Fourier transformation of the test functions ψ is defined by

$$\phi(\xi) = (2\pi)^{-1} \int d^4p \cdot e^{i(p,\xi)}\psi(p).$$

With this definition, Eq. (4) can be transformed into

$$\tilde{\psi}(u) = i \int d^4\xi\ \Delta^+(\xi, u)\phi(\xi). \tag{10}$$

The representation (9) becomes in ξ space

$$D^+(\xi) = \int du\rho(u)\ \Delta^+(\xi, u). \tag{11}$$

This is to be understood as

$$(D^+, \phi) = \int du\rho(u) \int d^4\xi\ \Delta^+(\xi, u)\phi(\xi), \quad \phi \in S_4. \tag{12}$$

The other homogeneous functions D^-, D, and D^1 can be derived from D^+ by simple algebraic operations. Representations of the form (11) hold for all of them.

3. THE INHOMOGENEOUS FUNCTIONS

Let us consider the function

$$D_R(\xi) = \theta(\xi)\ D(\xi). \tag{13}$$

The other inhomogeneous functions can be obtained from D_R by adding suitable homogeneous functions.

In terms of distribution theory, Eq. (13) means

$$\begin{aligned} &(D_R, \phi) = 0 \quad \text{if} \quad \phi(\xi) \equiv 0 \quad \text{in} \quad \xi^0 \geq 0, \\ &(D_R, \phi) = (D, \phi) \quad \text{if} \quad \phi(\xi) \equiv 0 \quad \text{in} \quad \xi^0 \leq 0. \end{aligned} \tag{14}$$

In order to get a representation for D_R we have to classify $\rho(u)$ with respect to its behavior at infinity: ρ will be said to be of degree N (N integer) if $\int du\ \rho(u)\ u^{-n}$ exists for every integer $n > N$ but diverges for $n \leq N$.

Let $\psi \in S_4$ be a test function. The function $\psi'(u)$ defined by

$$\psi'(u) = \int d^4p\, \frac{\psi(p)}{p^2 - u + i\epsilon p_0}$$

is then continuous and decreases for large u at least as u^{-1}.

Thus, if ρ is of degree $N \leq 0$, the integral $\int du\ \rho(u)\psi'(u)$ exists, and the expression

$$\tilde{D}_R(p) = \int du\rho(u)\, \frac{1}{p^2 - u + i\epsilon p_0}, \tag{15}$$

which we get from (13) by treating the right-hand side as a product in the ordinary sense is a well defined distribution:

$$(\tilde{D}_R, \psi) = \int du\rho(u)\psi'(u).$$

In the case $N > 0$, the expression (15) ceases to make sense and has to be replaced by

$$\tilde{D}_R(p) = Q_N(p^2) \times \int du\rho(u)[Q_N(u)]^{-1} \frac{1}{p^2 - u + i\epsilon p_0}. \tag{16}$$

Here $Q_N(u)$ stands for a polynomial of Nth degree which assumes positive values for all $u \geq \mu^2$. It is easy to see that (16) complies with the requirements (14).

From (15) or (16), respectively, we get the ξ-space representations

$$D_R(\xi) = \int du\rho(u)\ \Delta_R(\xi, u), \tag{17}$$

[6] L. Schwartz, *Théorie des distributions* (Hermann & Cie., Paris, 1959), 2nd ed., Vol. 2.

or

$$D_R(\xi) = Q_N(-\Box) \int du \frac{\rho(u)}{Q_N(u)} \Delta_R(\xi, u). \qquad (18)$$

Equation (18) is to be read as

$$(D_R, \phi) = \int du \frac{\rho(u)}{Q_N(u)} \int d^4\xi \, \Delta_R(\xi, u)\{Q_N(-\Box)\phi(\xi)\}.$$

The freedom in the choice of Q_N leads to the expected ambiguity[7]

$$\sum_{n=0}^{N-1} c_n \Box^n \delta^4(\xi).$$

Indeed, be Q^1, Q^2 two possible polynomials Q_N:

$$Q^i(u) = u^N + \sum_{n=0}^{N-1} a_n^i u^n.$$

The difference between the corresponding D_R^i is

$$\tilde{\Delta}(p) = \tilde{D}_R^1(p) - \tilde{D}_R^2(p)$$

$$= \int \frac{du}{p^2 - u} \rho(u) \frac{Q^1(p^2)Q^2(u) - Q^2(p^2)Q^1(u)}{Q^1(u)Q^2(u)}.$$

The numerator occurring in the integrand is of the form

$$N(p^2, u) = (p^2 - u)\bar{Q}(p^2, u),$$

$$\bar{Q}(p^2, u) = \sum_{n=0}^{N-1} (p^2)^n \bar{Q}_n(u), \qquad \bar{Q}_n \text{ of degree} \leq N - 1.$$

Therefore,

$$\tilde{\Delta}(p) = \sum_0^{N-1} c_n(-1)^n(p^2)^n,$$

$$c_n = (-1)^n \int du\rho(u) \frac{\bar{Q}_n(u)}{Q^1(u)Q^2(u)},$$

which is of the form claimed above.

4. THE STRUCTURE IN ξ SPACE

In this section we will examine the distribution character of the D functions in ξ space for the example of the D_P function.

We start with the representation

$$D_P(\xi) = \int du\rho(u) \, \Delta_P(\xi, u) \qquad (19)$$

or

$$D_P(\xi) = Q_N(-\Box) \int du \frac{\rho(u)}{Q_N(u)} \Delta_P(\xi, u) \qquad (20)$$

by introducing the explicit expression for $\Delta_P(\xi, u)$:

$$\Delta_P(\xi, u) = (4\pi)^{-1} \delta(\xi^2) - \theta(\xi^2) \frac{u}{8\pi} \cdot J_1(u^{\frac{1}{2}}\lambda)/u^{\frac{1}{2}}\lambda,$$

$$\lambda = (\xi^2)^{\frac{1}{2}}. \qquad (21)$$

[7] O. Steinmann, Helv. Phys. Acta. (to be published).

The function $I(t) = t^{-\frac{1}{2}} J_1(t^{\frac{1}{2}})$ is an entire function of t with the Taylor expansion

$$I(t) = \tfrac{1}{2} \sum_{m=0}^{\infty} \frac{(-1)^m}{2^{2m} m! \, (m+1)!} t^m. \qquad (22)$$

$I(t)$ decreases for $t \to \infty$ like $t^{-\frac{1}{2}}$.

According to the degree N of ρ we must distinguish several cases.

Case 1. $N \leq -2$

The integral (19) can be carried out by integrating the two terms in (21) separately. We obtain

$$D_P(\xi) = C \, \delta(\xi^2) - \theta(\xi^2)F(\xi^2), \qquad (23)$$

with

$$C = (4\pi)^{-1} \int du\rho(u) > 0, \qquad (24)$$

$$F(\xi^2) = (8\pi)^{-1} \int du\rho(u) \cdot uI(u\xi^2). \qquad (25)$$

$F(\xi^2)$ exists for all values of ξ^2 and is a bounded continuous function. Thus D_P has a δ singularity on the light cone and is a continuous function elsewhere.This is the same behavior as that of the free-field function Δ_P.

Case 2. $N = -1$

In this case the integral (25) diverges certainly in $\xi^2 = 0$ and may also diverge in any other point. In order to get a convergent integral, we note that

$$(\Box + u) \, \Delta_P(\xi, u) = \delta^4(\xi). \qquad (26)$$

Equation (19) can therefore be written in the form

$$D_P(\xi) = -\Box \int du\rho_1(u) \, \Delta_P(\xi, u) + \int du\rho_1(u) \cdot \delta^4(\xi),$$

where $\rho_1(u) = u^{-1} \cdot \rho(u)$ is of degree -2.

The integral occurring here is of the type that has been considered in Case 1. Thus, making use of

$$\Box \, \delta(\xi^2) = 4\pi \, \delta^4(\xi),$$

we obtain

$$D_P(\xi) = -\int du\rho_1(u) \cdot \delta^4(\xi) + \Box[\theta(\xi^2)F_1(\xi^2)]$$

$$+ \int du\rho_1(u) \cdot \delta^4(\xi) = \Box[\theta(\xi^2)F_1(\xi^2)],$$

with

$$F_1(\xi^2) = (8\pi)^{-1} \int du\rho_1(u)uI(u\xi^2). \qquad (27)$$

From Eq. (22) we have

$$F_1(0) = (16\pi)^{-1} \int du\rho(u) = \tfrac{1}{4}C,$$

C being defined by Eq. (24).

Therefore,

$$G_1(\xi^2) = \theta(\xi^2)[F_1(\xi^2) - \tfrac{1}{4}C] \tag{28}$$

is a bounded continuous function with support in $\xi^2 \geq 0$.

Since $\Box\theta(\xi^2) = 4\,\delta(\xi^2)$, we obtain finally

$$D_P(\xi) = C\,\delta(\xi^2) + \Box G_1(\xi^2). \tag{29}$$

For the discussion of $\Box G_1$ we use again Garding's mapping[5] of invariant four-dimensional distributions on distributions in one variable only. This mapping is defined in our case (all distributions concerned being even under time reversal) by the following procedure: We introduce the space H of the functions

$$h(\tau) = h_1(\tau) + h_2(\tau) \log |\tau|,$$

where

$$h_1, h_2 \in S_1, \qquad h_2(0) = 0.$$

Then, if we topologize H in a suitable way (see reference 5), the formula

$$\bar{\phi}(\tau) = \int \delta(\xi^2 - \tau)\phi(\xi)\, d^4\xi \tag{30}$$

defines a continuous mapping of S_4 onto H. The dual space of H will be called H'.

To every even distribution $T \in S_4'$ corresponds a distribution $\bar{T} \in H'$, defined by

$$(\bar{T}, \bar{\phi}) = (T, \phi), \qquad \bar{\phi} \in H. \tag{31}$$

(To every function $h \in H$ we can find a function $\phi \in S_4$ with $\bar{\phi} \equiv h$.)

We have

$$\begin{aligned} \overline{\Box\phi}(\tau) &= \Gamma\bar{\phi}(\tau), \qquad \Gamma = 4\tau\, d^2/d\tau^2, \\ \overline{\Box T} &= \Gamma'\bar{T}, \qquad \Gamma' = 4(d^2/d\tau^2)\cdot\tau. \end{aligned} \tag{32}$$

Note that the operation Γ' is defined on all of H' in the usual way [i.e. by $(\Gamma'\bar{T}, h) \equiv (\bar{T}, \Gamma h)$], while $(d/d\tau)\bar{T}$ is in general not defined since the functions h are in the origin not continuously differentiable.

We will now apply this procedure to $\Box G_1$. Since G_1 is a continuous function we have simply

$$\overline{G_1}(\tau) \equiv G_1(\xi^2), \qquad \tau = \xi^2.$$

The bar in $\overline{G_1}$ will be omitted in the future. We have

$$\overline{\Box G_1}(\tau) = \Gamma' G_1(\tau) = 4\tau \frac{d^2}{d\tau^2} G_1(\tau) + 8 \frac{d}{d\tau} G_1(\tau).$$

The two terms in this expression make sense individually in spite of the warning given above. This is so because continuous functions are defined as distributions on a much wider class of test functions than H, namely on integrable functions with sufficiently fast decrease at infinity.

$(d/d\tau)G_1(\tau)$ is a measure which does not contain a discrete part. In particular there is no term of the form $c\,\delta(\tau)$. Therefore $(d^2/d\tau^2)G_1$ does not contain any δ' terms. Terms of the form $\delta(\tau - a)$ may be present but will show up in $\Gamma' G_1$ only if $a \neq 0$. Thus the $\delta(\xi^2)$ singularity in (29) cannot be cancelled by the $\Box G_1$ term, i.e. D_P has always a δ singularity on the light cone.

$\Gamma' G_1$ may be continuous away from the light cone. However, it cannot be bounded in a neighborhood of the light cone. In order to see this we consider

$$\tau^{-1}G_1(\tau) = (8\pi)^{-1}\theta(\tau) \times \int du\rho_1(u)\cdot u^2\cdot(u\tau)^{-1}\{I(u\tau) - I(0)\}.$$

According to Eq. (22) the expression $(u\tau)^{-1}\{I(u\tau) - I(0)\}$ converges towards a nonvanishing constant if $\tau \to 0$. Since $\int du\rho_1(u)\cdot u^2$ diverges, $\tau^{-1}G_1$ cannot be bounded in a neighborhood of $\tau = 0$. Therefore $(d^2/d\tau^2)\tau G_1$ cannot exist and be bounded in such a neighborhood, because from this the boundedness of $\tau^{-1}G_1$ would follow.

Thus $\Box G_1$ is certainly no bounded continuous function in the vicinity of the light cone though it may still be an integrable function. In general it will be, however, a distribution of second order.

Case 3. $N = 0$

By using Eq. (26) twice we obtain

$$D_P(\xi) = \Box^2 \int du\rho_2(u)\, \Delta_P(\xi, u) - \int du\rho_2(u)\cdot\Box\,\delta^4(\xi) + \int du\rho_1(u)\cdot\delta^4(\xi),$$

with $\rho_2(u) = \rho(u)\cdot u^{-2}$ being of order -2.

We define analogously to Case 2:

$$F_2(\xi^2) = (8\pi)^{-1} \int du\rho_2(u) u I(u\xi^2),$$

$$G_2(\xi^2) = \theta(\xi^2)\{F_2(\xi^2) - F_2(0)\}.$$

G_2 is again a continuous function with support in

$\xi^2 \geq 0$. A simple calculation yields

$$D_P(\xi) = (4\pi)^{-1} \int du\rho_2(u) \cdot \Box^2\, \delta(\xi^2) - \int du\rho_2(u) \cdot \Box\, \delta^4(\xi) + \int du\rho_1(u) \cdot \delta^4(\xi) - \Box^2\{\theta(\xi^2)F_2(\xi^2)\}$$

$$= \int du\rho_1(u) \cdot \delta^4(\xi) - (16\pi)^{-1} \times \int du\rho_1(u) \cdot \Box^2\theta(\xi^2) - \Box^2 G_2(\xi^2)$$

$$= -\Box^2 G_2(\xi^2). \tag{33}$$

Mapping into H' renders

$$\tilde{D}_P(\tau) = -\Gamma'^2 G_2(\tau).$$

It is in general impossible to extract from this form a singularity with support on the light cone. We know, however, from the discussion of Case 2, that $\Gamma' G_2$ cannot be a bounded function in the vicinity of the light cone. Therefore $\Gamma'^2 G_2$ cannot be an integrable function in this vicinity, since from its integrability the boundedness of $\Gamma' G_2$ would follow; i.e. $\tilde{D}_P$ is on the light cone at least as singular as a measure. In general, it will of course behave worse than a measure. Away from the light cone $\tilde{D}_P$ is, in general, a distribution of fourth order, though in special cases it may be much nicer and may even be a function in some regions.

Case 4. $N > 0$

In this case we must use the representation (18):

$$D_P(\xi) = Q_N(-\Box)\, D'_P(\xi),$$

$$D'_P(\xi) = \int du\rho(u)[Q_N(u)]^{-1}\, \Delta_P(\xi, u).$$

$\rho(u)/Q_N(u)$ is of order 0, i.e. D'_P is of the form discussed in Case 3. In H' we get

$$\tilde{D}_P(\tau) = Q_N(-\Gamma')\, \tilde{D}'_P(\tau).$$

We know that $\tilde{D}'_P$ is at least as singular as a measure in a neighborhood of the light cone. Therefore $\tilde{D}_P$ is in this neighborhood of higher order than that of a measure. Nothing can be said about the singularities off the light cone except that they are of order $\leq 4 + 2N$.

5. THE CASE OF ZERO MASS

In the case $\mu = 0$, the Wightman distribution $\tilde{D}^+(\rho)$ is an invariant positive tempered measure with support in the forward cone V_+. The fact that the singular point $p = 0$ is now contained in the support of $\tilde{D}^+$ requires some extra care in the discussion of the mapping $S_4 \to S_1$ introduced in Sec. 2.

Let $c(p)$ be a function from S_4 with $c(0) = 1$, and $\gamma(p)$ be the function γ defined in Sec. 2. Note that $\gamma(0) = 0$ as a consequence of Eq. (5).

Then we can write, for any $\psi \in S_4$:

$$(\tilde{D}^+, \psi) = (\tilde{D}^+, \psi_0 c + (\psi - \psi_0 c)), \qquad \psi_0 = \psi(0)$$

$$= \psi_0(\tilde{D}^+, c) + (\tilde{D}^+, \gamma\overline{(\psi - \psi_0 c)})$$

$$= \psi_0(\tilde{D}_+, c - \gamma\bar{c}) + (\tilde{D}^+, \gamma\bar{\psi}),$$

where again $(\tilde{D}^+, \gamma\bar{\psi})$ is independent of the special choice of γ and therefore $\alpha = (\tilde{D}^+, c - \gamma\bar{c})$ is independent of γ as well as c.

We get thus the representation

$$\tilde{D}^+(p) = \alpha\, \delta^4(p) + \int du\rho(u)\theta(p_0)\, \delta(p^2 - u),$$

where ρ is defined as in Eq. (7).

The δ term is obviously the contribution coming from the vacuum as an intermediate state and will be assumed to vanish. The second term in the representation (33) is exactly of the form (9) and can be treated in the same way. Thus the representations (11), (17), (18) still hold.

Since ρ/u is no longer defined, we must slightly change the arguments used in Sec. 4.

The degree of ρ is to be defined with the help of the integrals $\int du\, \rho(u)\, (u + a^2)^{-n}$, $a \neq 0$. The degree N obviously does not depend on the value of a. In the case $\mu > 0$, this definition is equivalent to the one given in Sec. 3. In the discussion of Cases 2 and 3 we must replace Eq. (26) by

$$(\Box + u + a^2)\, \Delta_P(\xi, u) = \delta^4(\xi) + a^2\, \Delta_P(\xi, u).$$

From this we get, in Case 2,

$$D_P(\xi) = a^2 \int du\rho_1(u)\, \Delta_P(\xi, u) - \Box \int du\rho_1(u)\, \Delta_P(\xi, u) + \int du\rho_1(u) \cdot \delta^4(\xi),$$

$$\rho_1(u) = \rho(u)(u + a^2)^{-1}.$$

The first term is of the form treated in case 1, the second in the form treated in case 2. The result of the discussion is

$$D_P(\xi) = (4\pi)^{-1} \int du\rho(u) \cdot \delta(\xi^2) - a^2\theta(\xi^2)F_1(\xi^2) + \Box G_1(\xi^2), \tag{34}$$

with

$$F_1(\xi^2) = \int du \rho_1(u) u I(u\xi^2),$$

$$G_1(\xi^2) = \theta(\xi^2)\{F_1(\xi^2) - F_1(0)\}.$$

F_1 and G_1 are again continuous functions.

$D_P(\xi)$ has therefore exactly the same structure as in the case of nonvanishing μ. The same holds also in Case 3, and therefore in Case 4.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor J. R. Oppenheimer for the hospitality extended to me at the Institute for Advanced Study. I also wish to thank the Swiss National Fund for financial support.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 5 MAY 1963

Possible Schemes for Global Symmetry*

DAVID R. SPEISER† AND JAN TARSKI‡
The Institute for Advanced Study, Princeton, New Jersey
(Received 5 September 1962)

An attempt was made to determine all possible schemes for global symmetry as representations of groups. All possible schemes which correspond to connected groups were found, but the conclusions about nonconnected groups are incomplete. Physical interpretation of the schemes is discussed, and a detailed summary of group-theoretic methods is included.

1. INTRODUCTION

THE concepts of isospin and of strangeness, or hypercharge, have proven very useful in systematizing the elementary particle interactions. As we know, the corresponding conservation laws greatly limit the possible structure of strong interactions. These laws, on the other hand, still leave much arbitrariness. It is therefore natural that one should look for additional, or global, symmetries, which might circumscribe the theory in a more definite way.

In the last few years various schemes were proposed by a number of authors, and various approaches were introduced. In particular, Lee and Yang[1] recently examined global symmetry from the point of view of its group structure, and formulated a set of requirements for a global symmetry group. This suggested the following question: *What groups can be used as global symmetry groups?*

It is the purpose of this paper to construct a catalogue of global symmetry groups, for the particular case where the baryons form an octet. We determined, up to equivalence, all symmetry groups which are connected. We also found the possible neutral components (i.e. the connected components of the respective identities) for nonconnected groups. However, it appears that a further investigation of the nonconnected groups would be tedious, and of little significance. We therefore did not pursue this subject further. On the other hand, we discuss some properties of the symmetry groups, such as inclusion relations, and we show, furthermore, some of the physical consequences of the symmetry groups.

This study, we feel, serves a twofold purpose; first, it helps to correlate the previous studies on global symmetry, and extends slightly the results formerly obtained; second, it shows how a problem in interaction symmetry can be approached in a systematic way. We note that our methods could still be used, at least in principle, if the assumptions which we adopt should be modified. However, one should not forget that the relevance of global symmetry, such as considered here, to any higher symmetries which may be realized in nature, is as yet a poorly understood matter. We shall not try, therefore, to correlate our results with the current experimental situation. Only in Sec. 8 will there be a short discussion of these questions.

Our methods depend primarily on the classification of simple Lie groups and on the construction of

* Supported in part by the U. S. Air Force, Office of Scientific Research, Air Research and Development Command.
† Present address: Institut de Physique Théorique de l'Université de Genève, Genève, Switzerland.
‡ Present address: Courant Institute of Mathematical Sciences, New York University, New York, New York.
[1] T. D. Lee and C. N. Yang, Phys. Rev. 122, 1954 (1961).

with

$$F_1(\xi^2) = \int du \rho_1(u) u I(u\xi^2),$$

$$G_1(\xi^2) = \theta(\xi^2)\{F_1(\xi^2) - F_1(0)\}.$$

F_1 and G_1 are again continuous functions.

$D_P(\xi)$ has therefore exactly the same structure as in the case of nonvanishing μ. The same holds also in Case 3, and therefore in Case 4.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor J. R. Oppenheimer for the hospitality extended to me at the Institute for Advanced Study. I also wish to thank the Swiss National Fund for financial support.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 5 MAY 1963

Possible Schemes for Global Symmetry*

DAVID R. SPEISER† AND JAN TARSKI‡
The Institute for Advanced Study, Princeton, New Jersey
(Received 5 September 1962)

An attempt was made to determine all possible schemes for global symmetry as representations of groups. All possible schemes which correspond to connected groups were found, but the conclusions about nonconnected groups are incomplete. Physical interpretation of the schemes is discussed, and a detailed summary of group-theoretic methods is included.

1. INTRODUCTION

THE concepts of isospin and of strangeness, or hypercharge, have proven very useful in systematizing the elementary particle interactions. As we know, the corresponding conservation laws greatly limit the possible structure of strong interactions. These laws, on the other hand, still leave much arbitrariness. It is therefore natural that one should look for additional, or global, symmetries, which might circumscribe the theory in a more definite way.

In the last few years various schemes were proposed by a number of authors, and various approaches were introduced. In particular, Lee and Yang[1] recently examined global symmetry from the point of view of its group structure, and formulated a set of requirements for a global symmetry group. This suggested the following question: *What groups can be used as global symmetry groups?*

It is the purpose of this paper to construct a catalogue of global symmetry groups, for the particular case where the baryons form an octet. We determined, up to equivalence, all symmetry groups which are connected. We also found the possible neutral components (i.e. the connected components of the respective identities) for nonconnected groups. However, it appears that a further investigation of the nonconnected groups would be tedious, and of little significance. We therefore did not pursue this subject further. On the other hand, we discuss some properties of the symmetry groups, such as inclusion relations, and we show, furthermore, some of the physical consequences of the symmetry groups.

This study, we feel, serves a twofold purpose; first, it helps to correlate the previous studies on global symmetry, and extends slightly the results formerly obtained; second, it shows how a problem in interaction symmetry can be approached in a systematic way. We note that our methods could still be used, at least in principle, if the assumptions which we adopt should be modified. However, one should not forget that the relevance of global symmetry, such as considered here, to any higher symmetries which may be realized in nature, is as yet a poorly understood matter. We shall not try, therefore, to correlate our results with the current experimental situation. Only in Sec. 8 will there be a short discussion of these questions.

Our methods depend primarily on the classification of simple Lie groups and on the construction of

* Supported in part by the U. S. Air Force, Office of Scientific Research, Air Research and Development Command.
† Present address: Institut de Physique Théorique de l'Université de Genève, Genève, Switzerland.
‡ Present address: Courant Institute of Mathematical Sciences, New York University, New York, New York.
[1] T. D. Lee and C. N. Yang, Phys. Rev. 122, 1954 (1961).

their representations. These mathematical problems were solved by Killing, Cartan, and Weyl.[2,3] A few other standard theorems were also used. However, we did not need to solve any involved group-theoretic problems ourselves.

A brief summary of our paper follows.

In Sec. 2 we discuss various approaches to global symmetry, especially the approach of Lee and Yang,[1] which we adopt. We also discuss there the ramifications of this approach. The physical assumptions, which are the basis of our work, are summarized in Sec. 3, and we define there various kinds of global symmetry groups. In Sec. 4 we outline the mathematical approach, and list the groups which are candidates for global symmetry groups. In Sec. 5 we present our catalogue of global symmetry groups, and relate our results with the previous studies of global symmetry.

The next two sections illustrate the significance of the different global symmetry groups. This significance is brought out by showing how the symmetries should be extended to the antibaryons and to the mesons (Sec. 6), and by a discussion of the consequences of the symmetry groups (Sec. 7). Finally, in Sec. 8 we make a few observations on the possible relevance of our investigation to experimental tests, and to further theoretical developments.

The appendices deal with mathematical details. In Appendices A and B we summarize various properties of Lie groups and of their representations, respectively. Appendix C contains a discussion of nonconnected groups, and Appendix D contains the proofs of inclusion relations among the groups. We aimed to make the text intelligible without a study of Appendices C and D, but, we should emphasize, this does not apply to Appendices A and B. Much of the material in these two appendices is prerequisite to the text.

2. THE CONCEPT OF GLOBAL SYMMETRY

In this section we discuss and compare some of the approaches which have been introduced in dealing with global symmetry, and in particular, the approach of Lee and Yang.[1]

To illustrate one possible approach, let us consider a scheme which was suggested some time ago by Gell-Mann.[4] This scheme is based on Yukawa couplings, and the equality of some of the coupling constants is postulated. This approach was also used in other works. It has a drawback in that it is not always easy to see the group structure, or the consequences, of a scheme constructed in this manner. However, there is also an advantage, since this approach restricts our considerations to the observed particles. This is in contrast to some of the other approaches, which may require the introduction of hypothetical particles.

Some of the schemes were constructed in close analogy with the I-spin. One assumed a group larger than the I-spin, and the baryons and the mesons are assigned into supermultiplets which are associated with the irreducible representations of this group. One can then try to find some exact or approximate consequences of the assumed group structure. Ultimately one would hope to derive this structure from a deeper fundament, as is the case with the I-spin. The group SO_4 in particular has been used for a number of schemes; see e.g. the work of Pais.[5]

The approach of Lee and Yang, which is also the basis of this work, can be considered as the opposite to the one we just described. Lee and Yang take as the starting point, the representation associated with the eight (long-lived) baryons, rather than the underlying, or abstract, group. A global symmetry group then appears as a linear group, of matrices, which acts on the octet of the baryons and also satisfies certain requirements. This approach is basically phenomenological, and is perhaps an appropriate one to take, since we have no real understanding of the origin of symmetries. (The assumption of an octet of baryons is quite incidental to this approach; see also below.)

The most essential requirement which will be imposed on these linear groups is that they contain the I-spin group and the strangeness group as subgroups. Explicitly, we require that the following set of matrices, taken with respect to the indicated basis, be included:

$$\begin{pmatrix} SU_2 e^{-2i\phi} & & \\ & \begin{pmatrix} R(SO_3)R^{-1} & \\ & 1 \end{pmatrix} e^{-i\phi} & \\ & & SU_2 \end{pmatrix} \begin{pmatrix} \Xi \\ \Sigma \\ \Lambda \\ N \end{pmatrix}. \tag{2.1}$$

The elements of SO_3 are to be related to those of SU_2 in the usual way, and R is a suitably chosen fixed matrix. One may also say: The linear group must reduce in the indicated way if all parameters

[2] E. Cartan, *Sur la structure des groupes de transformations finis et continus*, Thesis, Paris (1894).

[3] H. Weyl, Math. Z. **23**, 271; **24**, 328, 377 (1925). [Reprinted in H. Weyl, *Selecta* (Birkhäuser, Basel, 1956), p. 262 ff.]

[4] M. Gell-Mann, Phys. Rev. **106**, 1296 (1957).

[5] A. Pais, Proc. Nat. Acad. Sci. U. S. **40**, 484 (1954).

except four are kept fixed at zero (provided they are suitably chosen). We shall discuss the requirements more fully in Sec. 3. The groups satisfying these requirements will be called *global symmetry groups*, or simply, *global symmetries*.

We will now make a few comments which should clarify our approach.

(1) One has to assume that some or all of the mass differences, depending on the supermultiplet structure, can be neglected. The assumption is of course crucial in all of the group-theoretic approaches attempted so far, and is their fundamental weakness.

(2) Given any scheme, one should be able to extend it to the antibaryons and to the mesons, and to derive some experimental consequences. As we shall see in the sequel, this does not lead to any difficulties. In particular, we shall assume that the mesons have the symmetries of baryon–antibaryon systems. The possible meson schemes are then determined. Moreover, the partition of meson supermultiplets into isomultiplets is likewise determined. It is the foregoing assumption, we may note, which makes the baryons rather than the mesons more convenient for the starting point. If we had started with schemes for the mesons, then the extensions to the baryons and the antibaryons would be more involved.

(3) Our program presupposes eight fundamental baryonic fields. This may be considered arbitrary, but an assumption of this kind has to be made in order to get concrete results and to understand the methods. Should this assumption prove unsatisfactory, the program could be suitably modified. But we ought to point out that we cannot rely very strongly on the current experimental situation. In the first place, many speculations have been made concerning the possible experimental effects of unstable particles.[6] Furthermore, some of our schemes lead to more mesons than the seven familiar ones, and one is forced to consider the additional mesons as short-lived. We also note that the possibility of having more than eight baryons has likewise been suggested,[7] and many corresponding schemes have been considered.[8]

(4) We have emphasized that in our approach we start with the representations rather than with the underlying groups. The distinction between these becomes particularly relevant in discussing some of the experimental consequences, and also when discussing the mesons, since then we have to deal with inequivalent representations of an underlying group. It is due to this circumstance that, even though we start with a representation, the underlying group nevertheless will play a major role in our investigation, and will yield some useful insights into the significance of the (linear) global symmetry group.

(5) Finally, we mention the possibility of correlating the observed conservation laws (i.e. of I-spin and of strangeness) with discrete groups.[9] We note here that if we were to base our analysis on such discrete groups, rather than on the usual continuous groups, then a systematic study of the possible global symmetries apparently would be much more difficult. We shall not consider this matter further.

3. SUMMARY OF THE ASSUMPTIONS

We will now summarize the assumptions, or requirements, which form the basis of our work. We will also introduce a few definitions. (We note that our requirements differ from those of Lee and Yang[1] in a few minor respects.)

(1) We require that a global symmetry group contain as subgroups the I-spin group and the strangeness group, in the sense explained in Sec. 2.

However, our formulation of the requirement, that the strangeness group be included, will be modified sometimes for mathematical convenience. In order to deal with unimodular matrices we shall replace strangeness by hypercharge, or by a modified hypercharge ($Y' = Y + \alpha N$); see Sec. 4 for details.

We shall refer to the strangeness group, to the hypercharge gauge group, and to the modified hypercharge gauge groups, as the *strangeness-like groups*.

(2) We shall assume that a global symmetry group is a compact Lie group, and that it is a subgroup of the group U_8 of 8×8 unitary matrices.

This restriction ensures that the requirement of conservation of probability is fulfilled, and also excludes certain rather artificial examples; see for instance the examples given by Pontriagin.[10] However, schemes have also been constructed on the basis of noncompact groups; see the example of Higgs,[11] who found it necessary to introduce indefinite metric into the theory. We shall not consider such schemes further.

We remark that this assumption can be derived

[6] See, e.g., J. J. Sakurai, Ann. Phys. N. Y. **11**, 1 (1960).
[7] M. Gell-Mann, Nuovo Cimento Suppl. **4**, 848 (1956).
[8] R. E. Behrends, J. Dreitlein, C. Fronsdal, and B. W. Lee, Rev. Mod. Phys. **34**, 1 (1962).
[9] K. M. Case, R. Karplus, and C. N. Yang, Phys. Rev. **101**, 874 (1956).
[10] L. Pontriagin, *Topological groups* (Princeton University Press, Princeton, New Jersey, 1939), pp. 264–5.
[11] P. W. Higgs, Nucl. Phys. **4**, 221 (1957).

from somewhat weaker premises, since unitarity and compactness are closely related. But we shall not elaborate here.

(3) A global symmetry group is a linear group, or a representation, and therefore we may speak of *reducible* and of *irreducible global symmetry groups*, depending on whether the four isomultiplets form one octet, or two or three supermultiplets.

We shall also consider symmetry groups which relate only to some of the baryons. These will be called *partial global symmetry groups*. We shall assume for a partial global symmetry group the requirements which are analogous to those numbered 1 and 2. In particular, the I-spin and the strangeness groups will be assumed to act on the relevant baryons only. The remaining baryons can be simply ignored, or, one cay say, can be thought of as having quantum numbers zero.

The partial global symmetries will be very helpful to us in finding the global symmetries, both the reducible and the irreducible ones. The partial global symmetries may also be of interest in themselves. However, this latter point of view leads to questions as, e.g., how should the I-spin be interpreted. We shall not pursue this point of view further, but we shall look at these partial symmetries only as useful constructs. For brevity we consider only those partial global symmetries which are irreducible.

(4) We shall say that a symmetry group is *minimal*, if it does not have any (proper) subgroups which define the same supermultiplet structure.

In other words, a global, or a partial global symmetry group is minimal if it does not contain as subgroups any other symmetries which reduce into the same number (and not a larger number) of blocks.

For brevity, our discussion of the nonconnected symmetry groups will be confined largely to the minimal ones, but the connected groups will be discussed fully. (We should note, however, that our understanding of the minimal nonconnected groups is likewise incomplete, and the previous remarks still apply.)

(5) In our approach, a symmetry group is a linear group, and the following question arises: Given two equivalent symmetry groups, do they have the same physical content? We proceed on the assumption that they do. (See also Sec. 7.) However, we do not attempt to define *physical content* precisely, and so cannot answer the question in definite terms.

(6) We now consider the assumptions which will be used in order to extend the global symmetries to the antibaryons and to the mesons. We refer here to the underlying group $\mathcal{G}$ rather than to a specific representation, since in general more than one representation of $\mathcal{G}$ will be involved.

To extend a symmetry to the antibaryons, we shall assume that a system consisting of a baryon and of the corresponding antibaryon is invariant under the action of $\mathcal{G}$. Similarly, we shall assume that there is a system which consists of a baryon, an antibaryon, and a meson, and which is invariant under $\mathcal{G}$. We note that this last condition is fulfilled if we have a Yukawa Lagrangian invariant under $\mathcal{G}$, but our assumption is more general.

Let us now formulate the foregoing assumptions concisely. Let D_B, $D_{\bar{B}}$, and D_M denote the representations which are to be associated with the baryons, the antibaryons, and the mesons, respectively; and let D^E denote the identity representation $[D^E(g) \equiv 1]$. We recall that, if we are given two particles described by vectors v_1 and v_2, then the system which consists of these two particles is described by the tensor product $v_1 \otimes v_2$. Further, the action of a group on a tensor product is given by the Kronecker product representation. Our assumptions now require that

$$D_B \otimes D_{\bar{B}} \supseteq D^E, \tag{3.1a}$$

$$D_B \otimes D_{\bar{B}} \otimes D_M \supseteq D^E. \tag{3.1b}$$

The requirement (3.1a) as stated is sufficient for irreducible symmetries. Otherwise we must assume such a relation for each supermultiplet of the baryons.

Interacting systems can be discussed without introducing any special assumptions. It is only necessary to interpret group invariance as the invariance of n-particle amplitudes (which behave like tensors) with respect to the action of the group. This is done in Sec. 7.

The preceding discussion was given in general terms, since we wanted to avoid any particular field-theoretic framework. However, if we accept the framework of conventional field theory, then we can be more specific. In this case we are interested in the invariance under $\mathcal{G}$ of the following functions:

$$\langle T(\bar{\psi}_a \psi_a)\rangle_0, \quad \langle T(\bar{\psi}_a \psi_b \varphi_c)\rangle_0, \quad \langle T(\bar{\psi}_a \bar{\psi}_b \psi_c \psi_d)\rangle_0, \tag{3.2}$$

and the foregoing discussion applies. See also the article of Behrends *et al.*[8] The last expression in (3.2) refers to baryon–baryon and baryon–antibaryon scattering, as typical of interacting systems.

4. ENUMERATION OF LINEAR GROUPS

In the last section we formulated the mathematical problem of determining the linear groups, or repre-

TABLE I (a). Representations of noncommutative simple Lie groups and algebras, of degree 8 or less, arranged according to the rank.

Rank	Degree 2	3	4	5	6	7	8
1	A_1^2	A_1^3	A_1^4	A_1^5	A_1^6	A_1^7	A_1^8
2		$^{(2)}A_2^3$	B_2^{Spin}	B_2^v	A_2^6	G_2^7	A_2^8
3			$^{(2)}A_3^4$		$\begin{cases} D_3^v \\ C_3^6 \end{cases}$	B_3^v	B_3^{Spin}
4				$^{(2)}A_4^5$			$\begin{cases} D_4^{v,\text{Spin}(2)} \\ C_4^8 \end{cases}$
5					$^{(2)}A_5^6$		
6						$^{(2)}A_6^7$	
7							$^{(2)}A_7^8$

sentations, which are of degree 8 or less, and which satisfy certain requirements. We will now enumerate the linear groups as candidates for global symmetries, but a detailed discussion of the inclusions is postponed to the next section. Likewise, the mathematical details are to a large extent removed to the appendices.

The way which we found the most convenient to approach the problem at hand, i.e. the enumeration of linear groups, is the following: We first enumerate the simple compact Lie groups and their representations. Then we construct the Kronecker products of these representations; the resulting groups are semisimple. This exhausts the connected irreducible linear groups, up to a phase (see below). Out of these irreducible groups, we construct the reducible groups, and then the nonconnected ones.

It may seem that we are making a detour by bringing the abstract Lie groups into the picture, but this seems to be the only feasible approach. Moreover, the underlying group forms an integral part of the theory, and this detour will prove later to be very convenient.

Before we go further, let us observe that a symmetry group as defined above is the Kronecker product of the circle group ($e^{-ir\phi}$) and of a group of unimodular matrices. (This decomposition was implied in Sec. 3.) In the case of a global symmetry, this corresponds to writing the strangeness S as

$$S = (-N) + (S + N). \tag{4.1}$$

Here N is the baryonic charge. We see that, in the group of unimodular matrices, the strangeness group is in effect replaced by the hypercharge gauge group. Clearly, it is sufficient to examine only the groups of unimodular matrices, since the circle group can be adjoined whenever it is needed or desired. The circle group corresponds to the term $(-N)$ in Eq. (4.1), and is of no consequence if just the octet of the baryons is considered.

In case of partial global symmetries an analogous argument applies, but we may have to use a modified hypercharge. E.g., for the system N–Λ–Σ we have to use $S + \frac{2}{3}N$, rather than $S + N$. We also see that special care is needed with the hypercharge when two partial symmetries are combined to form a global symmetry. It is with the hope of avoiding confusion here that we formulated our problem in terms of strangeness rather than hypercharge.

In our tables of representations and symmetries we shall give only the unimodular groups, since a circle group can always be adjoined trivially.

We now proceed to enumerate the compact Lie groups which consist of unimodular matrices. We start with the irreducible representations of the simple noncommutative groups. These (abstract) groups are given by the Cartan–Killing classification,[2] which is summarized in Appendix A. The irreducible representations of these groups have been constructed by Weyl.[3] (See also the Appendix to the work of Dynkin.[12]) In particular, there is a simple formula for the degree of a representation. We shall not try to explain the results or the methods of Weyl in this paper, but we merely list the irreducible representations of simple groups and of degree 8 or less in Tables I(a) and (b).

We should make a few comments concerning these tables. In Table I(a) we refer to the rank of a group. The rank is the dimension of a maximal abelian subgroup, and determines how many conserved quantum numbers are implied. (See Sec. 7.) Next, for each of the rotation groups, except for SO_4 which is not simple, there are listed one vector and one or two spinor representations. The group SU_2 has an irreducible representation of every degree, as is well known. The group SU_3 is the only other group for which are listed representations of more than one degree: two of degree 3, two of degree 6, and one of degree 8. For the unitary unimodular groups SU_n, most of the representations occur in complex conjugate pairs if $n \geq 3$.

The group SO_8 has three inequivalent irreducible representations of degree 8. One can construct them so that they will involve the same matrices, and we shall call the representations *analogous*, for lack of a better term. These representations are not merely isomorphic, as any two representations of a simple

[12] E. B. Dynkin, Am. Math. Soc. Transl. Ser. 2, **6**, 245 (1957; published in Russian in 1952).

TABLE I(b). This table describes the types of the representations of Table I(a). The superscript as in A_2^8 or in B_3^{Spin} shows the degree or the type of representation. The superscript (2) in ${}^{(2)}A_n^k$ indicates that there are two inequivalent representations of the given degree, one the complex conjugate of the other.

Degree 2	3	4	5	6	7	8	Type of associated linear groups
A_1^2	${}^{(2)}A_2^3$	${}^{(2)}A_3^4$	${}^{(2)}A_4^5$	${}^{(2)}A_5^6$	${}^{(2)}A_6^7$	${}^{(2)}A_7^8$	unitary unimodular
$[A_1^2]$	A_1^3	A_1^4	A_1^5	A_1^6	A_1^7	A_1^8	representations of SU_2 (and of SO_3)
	$[{}^{(2)}A_2^3]$			${}^{(2)}A_2^6$		A_2^8	representations of SU_3
	$[B_1^v \simeq A_1^3]$	a	B_2^v	$[D_3^v \simeq A_3^6]$	B_3^v	D_4^v [b]	proper orthogonal (or, vector representations)
$[B_1^{\text{Spin}} \simeq A_1^2]$		B_2^{Spin}				B_3^{Spin}	spinor (for B_n)
		${}^{(2)}[D_3^{\text{Spin}} \simeq A_3^4]$				$D_4^{\text{Spin}(2)}$ [b]	semispinor (for D_n)
$[C_1^2 \simeq A_1^2]$		$[C_2^4 \simeq B_2^{\text{Spin}}]$		C_3^6		C_4^8	symplectic
					G_2^7		automorphism group of octonions

[a] The group SO_4 is not simple; see Table II.
[b] The group D_4 has three inequivalent but analogous representations of degree 8: the vector representation and two spinor representations. (The latter two are not complex conjugates.)

Lie group are isomorphic (at least locally). This situation in fact resembles that of the complex conjugate representations ${}^{(2)}A_n^k$, where the two representations clearly have analogous properties. We shall therefore select only one of such analogous representations to be a candidate for a symmetry group. The chosen representations will be denoted simply by A_n^k in case of SU_{n+1}, and by D_4^8 in case of SO_8.

We next determine the linear groups which are irreducible, connected, and compact, and which consist of unimodular matrices. The basic tool here is the following lemma, which we establish in Appendices A and B.

Lemma 4.1. *Every linear Lie group G which is irreducible, connected, and compact, and which consists of unimodular matrices, is the Kronecker product of irreducible representations of simple (therefore connected), compact, noncommutative Lie groups.*

We can now use Tables I(a) and (b) to construct the linear groups in question. These groups are given in Table II. We note that the factor representations have to refer to *different* groups (which may be isomorphic), since otherwise the representation would reduce.

Ordinarily we shall use primes to denote groups which are isomorphic but independent, as e.g. SU_2 and SU_2'.

Let us now turn to the connected reducible groups of unimodular matrices. Such a group is necessarily of the form

$$R\begin{pmatrix} H & \\ & I \end{pmatrix}R^{-1}, \quad \text{or} \quad R\begin{pmatrix} H & & \\ & I & \\ & & J \end{pmatrix}R^{-1}, \qquad (4.2)$$

where H, I, and J are irreducible, but not necessarily unimodular, and where R is a constant matrix. In other words, such a group might reduce with respect to some basis other than the usual one. The elements of one block, H, I, or J, may not vary independently of the elements of another block. The structure of such matrix groups is clarified by the following two lemmas (proved in Appendix B).

Lemma 4.2. *Write each block as the Kronecker product of simple linear groups: l_i^H, etc. If two of these factors, say l_i^H and l_k^I, are locally isomorphic, then there are two possibilities: The respective matrices may de-*

TABLE II. Kronecker product groups. Primes indicate that the elements in the component groups vary independently. The group $A_1^2 \otimes (A_1')^2$ is equivalent to SO_4.

Rank	Degree 4	6	8
2	$A_1^2 \otimes (A_1')^2$	$A_1^2 \otimes (A_1')^3$	$A_1^2 \otimes (A_1')^4$
3		${}^{(2)}(A_1^2 \otimes A_2^3)$	$A_1^2 \otimes (A_1')^2 \otimes (A_1'')^2$ $A_1^2 \otimes B_2^{\text{Spin}}$
4			${}^{(2)}(A_1^2 \otimes A_3^4)$

TABLE III. Possibilities for the neutral components of irreducible nonconnected groups, for the case where this component breaks into two blocks. The group G^* is the complex conjugate representation to G. (Therefore the elements of the two groups are not independent.) The factors $e^{\pm i\alpha}$ are indicated if they are needed to make the blocks inequivalent.

Form of the neutral component	
$\begin{pmatrix} Ge^{-i\alpha} & \\ & Ge^{i\alpha} \end{pmatrix}$, $\begin{pmatrix} G & \\ & G' \end{pmatrix} Q$	$A_1^2,\ A_1^3,\ A_1^4,\ B_2^{\text{Spin}}$ } Possibilities for G
$\begin{pmatrix} Ge^{-i\alpha} & \\ & Ge^{i\alpha} \end{pmatrix}$, $\begin{pmatrix} G & \\ & G^* \end{pmatrix} Q$, $\begin{pmatrix} G & \\ & G' \end{pmatrix} Q$	${}^{(2)}A_2^3,\ {}^{(2)}A_3^4$ } Possibilities for G
$\begin{pmatrix} A_1^2 \otimes (A_1')^2\, e^{-i\alpha} & \\ & A_1^2 \otimes (A_1')^2\, e^{i\alpha} \end{pmatrix}$, $\begin{pmatrix} A_1^2 \otimes (A_1')^2 & \\ & A_1^2 \otimes (A_1'')^2 \end{pmatrix} Q$, $\begin{pmatrix} A_1^2 \otimes (A_1')^2 & \\ & (A_1'')^2 \otimes (A_1''')^2 \end{pmatrix} Q$	
Possibilities for Q: $\begin{pmatrix} E & \\ & E \end{pmatrix}$, $\begin{pmatrix} Ee^{-i\alpha} & \\ & Ee^{i\alpha} \end{pmatrix}$	

pend on the same or on an independent set of parameters. (In the first case the matrices represent elements of the same underlying group. In the second case we say that the matrices vary independently.) If two factors are not locally isomorphic, then their elements vary independently.

Lemma 4.3. *Let a linear group G be given in one of the two forms of* (4.2), *and let G include the I-spin group and one of the strangenesslike groups as subgroups. Then G is equivalent to a group which reduces with respect to the usual basis* (p, n, $\cdots$, Ξ^-), *and which also fulfills the inclusion requirements.*

Lemma 4.2 follows from the fact that we are dealing with simple groups. Lemma 4.3 is an easy consequence of the inequivalence of the blocks in (2.1). To appreciate this lemma, note the following. If we were to ignore strangeness, then we could construct symmetry groups which would define, e.g., two supermultiplets N'–Λ–Σ and Ξ', where N' and Ξ' are linear combinations of N and Ξ. Lemma 4.3 states that this is no longer possible if strangeness is taken into account.

For brevity we shall not enumerate here the reducible groups. However, examples are given elsewhere in this section, and in the next.

Finally we come to the nonconnected groups. As we mentioned before, we are able to be systematic only with regard to their neutral components. We present a detailed discussion of the nonconnected groups in Appendix C, and we reach there the following conclusion: The neutral component H_0 can include the I-spin group as a subgroup only if H_0 reduces to at most two blocks. If H_0 reduces to two blocks, then they must be of the same degree and must contain the same elements, up to equivalence, but they must not define equivalent representations of H_0. The foregoing discussion of the connected reducible groups can be applied here, and the possible forms of unimodular groups H_0 which reduce into two blocks are shown in Table III. On the other hand, if H_0 is irreducible, then the possible forms of H_0 are already known to us.

If H_0 has two components, then we shall assume for definiteness that the nonconnected group is generated by H_0 and by the matrix

$$R_{2n} = \begin{pmatrix} 0 & (-1)^n E_n \\ E_n & 0 \end{pmatrix}. \tag{4.3}$$

(Such an element sometimes has been improperly called a *discrete element.*) The factor $(-1)^n$ gives the following features: If $n = 2$, then we shall obtain for a partial global symmetry a group which is equivalent to $SU_2 \otimes O_2^+$. For $n = 3$, we preserve the unimodularity in the nonconnected group. If $n = 4$, then we shall obtain the global symmetry group G_0 introduced by Lee and Yang.[1]

We also mention the possibility of generating nonconnected groups by a group H_0 of unimodular matrices and by a scalar matrix ($e^{2\pi i/k}$ where k is integral). This is an example of a reducible group, in which one or more components may be nonconnected. Other groups of this category are possible, and examples will be given in the next section. The construction of such groups presents no new difficulties. In particular, Lemma 4.2 can be easily adapted to nonconnected groups.

5. GLOBAL AND PARTIAL GLOBAL SYMMETRIES

In this section we will state which groups fulfill the requirements for a global or for a partial global symmetry, which requirements were formulated in Secs. 2 and 3. We also discuss the inclusions among the symmetry groups.

The proofs concerning the inclusion of I-spin and of a strangeness-like group, and the proofs of other inclusion relations, are removed to Appendix D. On the other hand, the requirement that a symmetry group be unitary does not lead to complications, as the following lemma shows.

Lemma 5.1. *Let a compact linear group contain the I-spin group and one of the strangeness-like groups as subgroups. Then there exists an equivalent group of unitary matrices which also contains these groups as subgroups.*

The existence of an equivalent group of unitary matrices is of course a well-known fact. But our assertion states more, and a simple proof is given in Appendix B.

We now come to the inclusion relations. In Table IV we state whether each of the simple groups listed in Tables I(a) and (b) contains the I-spin group, the hypercharge (or modified hypercharge) gauge group, and both groups simultaneously. This is, of course, to be understood in the following sense: Is there a representation of the type in question for which the specified inclusions hold? We should point out that there may be a representation which includes the I-spin group, and an equivalent one which includes the hypercharge gauge group, but no equivalent representation which includes both groups simultaneously. Therefore the information in the last column is not an immediate consequence of the information in the two preceding columns. An example which illustrates this point can be found in Appendix D.

Table V is analogous to Table IV, but refers to the Kronecker products.

We will now diverge from the order of the last section, and we will consider the nonconnected groups before the reducible ones. Table VI shows those nonconnected groups (called K, L, and M) which satisfy the inclusion requirements, and which are minimal. It would be very repetitious to give an analogue of Tables IV and V for the nonconnected groups, and such information could be easily derived. In particular, we are leaving out of discussion those

TABLE IV. The inclusion of the I-spin group and of the hypercharge gauge group (perhaps modified) in the representations of Tables I(a) and (b).

Degree and linear group		Particles	Inclusion of I-spin	Y or Y'	both
deg. 2	A_1^2	N; Ξ	yes		
deg. 3	A_1^3	Σ	yes		
	A_2^3		yes		
deg. 3	A_1^3	$N\Lambda$	no	no	no
	A_2^3	$\Xi\Lambda$	yes	yes	yes
deg. 4	A_1^4	$\Sigma\Lambda$	no		
	B_2^{Spin}		no		
	A_3^4		yes		
deg. 4	A_1^4	$N\Xi$	no	no	no
	B_2^{Spin}		yes	yes	yes
	A_3^4		yes	yes	yes
deg. 5	A_1^5	$N\Sigma$	no	no	no
	B_2^v	$\Xi\Sigma$	no	no	no
	A_4^5		yes	yes	yes
deg. 5	A_1^5	$N\Lambda\Xi$	no	no	no
	B_2^v		yes	yes	yes
	A_4^5		yes	yes	yes
deg. 6	A_1^6	$N\Lambda\Sigma$	no	no	no
	A_2^6	$\Xi\Lambda\Sigma$	yes	no	no
	D_3^v		no	no	no
	C_3^6		no	no	no
	A_5^6		yes	yes	yes
deg. 7	A_1^7	$N\Sigma\Xi$	no	no	no
	G_2^7		yes	yes	yes
	B_3^v		yes	yes	yes
	A_6^7		yes	yes	yes
deg. 8	A_1^8	$N\Lambda\Sigma\Xi$	no	no	no
	A_2^8		yes	yes	yes
	B_3^{Spin}		yes	yes	yes
	D_4^8		yes	yes	yes
	C_4^8		no	yes	no
	A_7^8		yes	yes	yes

TABLE V. The inclusion of the I-spin group and of the hypercharge gauge group (perhaps modified) in the representations of Table II.

Degree and linear group	Particles	Inclusion of I-spin	Y or Y'	both
deg. 4 $A_1^2 \otimes (A_1')^2$	$\Sigma\Lambda$	yes		
deg. 4 $A_1^2 \otimes (A_1')^2$	$N\Xi$	yes	yes	yes
deg. 6 $A_1^2 \otimes (A_1')^3$	$N\Lambda\Sigma$	no	no	no
$A_1^2 \otimes A_2^3$	$\Xi\Lambda\Sigma$	yes	yes	yes
deg. 8 $A_1^2 \otimes (A_1')^4$	$N\Lambda\Sigma\Xi$	no	yes	no
$A_1^2 \otimes (A_1')^2 \otimes (A_1'')^2$		no	yes	no
$A_1^2 \otimes B_2^{\text{Spin}}$		yes	yes	yes
$A_1^2 \otimes A_3^4$		yes	yes	yes

TABLE VI. The generators of the minimal nonconnected symmetry groups.

Degree and particles	Neutral component	Discrete generator	Notation
deg. 4— $N\Xi$	$\begin{bmatrix} A_1^2 e^{-i\alpha} & \\ & A_1^2 e^{i\alpha} \end{bmatrix}$	$\begin{bmatrix} O & E_2 \\ E_2 & O \end{bmatrix}$	K
deg. 6— $N\Lambda\Sigma$ / $\Xi\Lambda\Sigma$	$\begin{bmatrix} A_2^3 e^{-i\alpha} & \\ & (A_2')^3 e^{i\alpha} \end{bmatrix}$	$\begin{bmatrix} O & -E_3 \\ E_3 & O \end{bmatrix}$	L
deg. 8— $N\Lambda\Sigma\Xi$	$\begin{bmatrix} A_1^2 \otimes (A_1')^2 & \\ & A_1^2 \otimes (A_1'')^2 \end{bmatrix}$	$\begin{bmatrix} O & E_4 \\ E_4 & O \end{bmatrix}$	M

TABLE VII. Summary of the irreducible global and partial global symmetry groups. The circle groups are omitted, but they can be readily adjoined.

Degree and particles									
2	3		4		5		6	7	8
N Ξ	Σ	$N\Lambda$ $\Xi\Lambda$	$\Sigma\Lambda$	$N\Xi$	$N\Sigma$ $\Xi\Sigma$	$N\Lambda\Xi$	$N\Lambda\Sigma$ $\Xi\Lambda\Sigma$	$N\Sigma\Xi$	$N\Lambda\Sigma\Xi$
A_1^2	A_1^3 A_2^3	A_2^3	A_3^4 $A_1^2 \otimes (A_1')^2$	B_2^{Spin} A_3^4 $A_1^2 \otimes (A_1')^2$ K	A_4^5	B_2^v A_4^5	A_5^6 $A_1^2 \otimes A_2^3$ L	G_2^7 B_2^v A_6^7	A_2^8 B_3^{Spin} D_4^8 A_7^8 $A_1^2 \otimes B_2^{\text{Spin}}$ $A_1^2 \otimes A_3^4$ M

nonconnected groups for which the connected components are irreducible.

The information of Tables IV–VI is summarized in Table VII, where we list the irreducible global and partial global symmetries. Table VIII shows the inclusions among the groups of Table VII.

There remain the reducible global symmetries. Their forms are restricted by Lemmas 4.2 and 4.3. One simple way to construct such groups is to let one partial global symmetry act on some of the baryons, and to let another act independently on the remaining baryons. The following examples illustrate this construction:

B_2^v acts on the N–Λ–Ξ system,

A_1^3 acts on the Σ's; (5.1a)

L acts on the N–Λ–Σ system,

A_1^2 acts on the Ξ's. (5.1b)

One must adjoin the circle group to the unimodular groups appropriately, in order that the strangeness group, or the hypercharge gauge group, be included. This is needed, in particular, in the second example.

More interesting examples arise if the groups which act on different supermultiplets are not independent. One well-known example of this is the doublet approximation group G_{DA}. We define this group in Table IX, where the other groups of this kind are also listed. It is easy to check that no other groups of this kind are possible.

TABLE VIII. Inclusions among the groups of Table VII. The inclusion of M in B_3^{Spin} has not been proved or disproved.

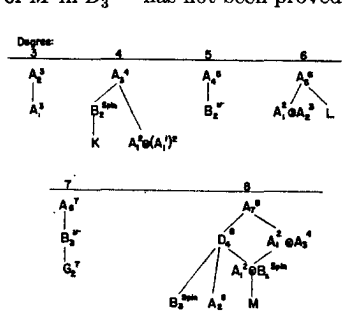


TABLE IX. The reducible global symmetry groups which do not contain independent blocks.

Symmetry groups	Reference bases
$G_{DA} = \begin{bmatrix} A_1^2 & & \\ & A_1^3 & \\ & & A_1^2 \otimes (A_1')^2 \end{bmatrix}$ $G_{HR} = \begin{bmatrix} A_1^2 \otimes (A_1')^2 & & \\ & A_1^3 & \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} N \\ \Xi \\ \Sigma \\ \Lambda \end{bmatrix}$
$K_{DA} = \begin{bmatrix} A_1^2 \otimes O_2^- & & \\ & & \\ & & A_1^2 \otimes (A_1')^2 \end{bmatrix}$, $\bar{K} = \begin{bmatrix} A_1^2 \otimes O_2^- & & \\ & A_1^3 & \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} N \\ \Xi \\ \Sigma \\ \Lambda \end{bmatrix}$
$M_0 = \begin{bmatrix} A_1^2 \otimes (A_1')^2 & & \\ & & \\ & & A_1^2 \otimes (A_1'')^2 \end{bmatrix}$	$\begin{bmatrix} N \\ \Xi \\ \Sigma \\ \Lambda \end{bmatrix}$
$J_\Xi = \begin{bmatrix} A_1^2 e^{-3i\alpha} & & \\ & & \\ & & A_1^2 \otimes A_2^3 e^{i\alpha} \end{bmatrix}$, $J_N = \begin{bmatrix} A_1^2 \otimes A_2^3 e^{-i\alpha} & & \\ & & \\ & & A_1^2 e^{3i\alpha} \end{bmatrix}$	$\begin{bmatrix} \Xi \\ \Sigma \\ \Lambda \\ N \end{bmatrix}$
$H_G = \begin{bmatrix} G_2^7 & \\ & 1 \end{bmatrix}$, $H_B = \begin{bmatrix} B_3^v & \\ & 1 \end{bmatrix}$, $H_A = \begin{bmatrix} A_6^7 & \\ & 1 \end{bmatrix}$	$\begin{bmatrix} N \\ \Xi \\ \Sigma \\ \Lambda \end{bmatrix}$

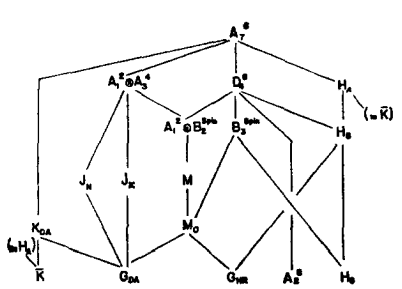
In Table X we show the inclusions for the global symmetry groups, including the reducible groups which are displayed in Table IX. We are not including the groups which have independent blocks, because these groups are too numerous to make such a table interesting.

In Table XI we define two new symmetry groups $\bar{K}'$ and K'_{DA}, and we compare the inclusions for these groups with the inclusions for $\bar{K}$ and for K_{DA}. This comparison shows how sensitive these inclusions are to the detailed structure of the discrete generators in nonconnected groups.

Let us now comment briefly on the symmetry groups. Table VII lists seven irreducible global symmetries, of which four are simple groups, two are Kronecker products, and one is nonconnected. We recall that one can construct other groups of the last category. Next, we see from Table VIII that there are three minimal global symmetries, A_2^8, M, and B_3^{Spin}. The group A_2^8 involves 8 parameters, and is the basis of Gell-Mann's eightfold way[13] and of the work of Ne'eman.[14] The group M involves 9 parameters and is nonconnected. It was discussed by Lee and Yang,[1] who called it G_0. The group B_3^{Spin} involves 21 parameters, and has been discussed by Tiomno[15] and others.[16,17] Further, the group $A_1^2 \otimes A_3^4$, which is not minimal, corresponds to the global symmetry of Gell-Mann.[4] The remaining groups, $A_1^2 \otimes B_2^{\text{Spin}}$, D_4^8, and A_7^8, apparently have not been considered before in the context of global symmetry.

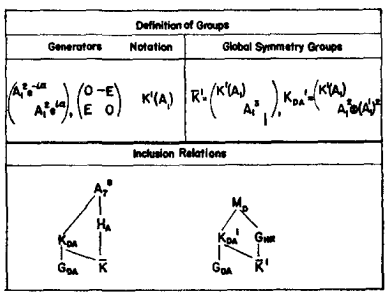
In Table IX, there are three reducible global symmetries which do not include other symmetries as subgroups, namely $\bar{K}$, G_{HR}, and G_{DA}. [We do not consider here (I-spin) $\times$ (hypercharge gauge) as a

TABLE X. Inclusions among the global symmetry groups, which do not contain independent blocks.



[13] M. Gell-Mann, Phys. Rev. 125, 1067 (1962), and Caltech Rept. CTSL-20 (1961).
[14] Y. Ne'eman, Nucl. Phys. 26, 222 (1961).
[15] J. Tiomno, Nuovo Cimento 6, 69 (1957).
[16] G. Feinberg and F. Gürsey, Phys. Rev. 114, 1153 (1959).
[17] R. E. Behrends and A. Sirlin, Phys. Rev. 121, 324 (1961).

TABLE XI. Definition of the groups K', $\bar{K}'$, and K_{DA}', and a comparison of inclusions for $\bar{K}'$ and K_{DA}' with those for $\bar{K}$ and K_{DA}'.



global symmetry.] The group $\bar{K}$, which can be appropriately called hypercharge reflection, involves four parameters and is nonconnected. This group is related to the group $\bar{K}'$ of Table XI, and the latter was discussed by Feinberg and Behrends.[18] The group G_{HR}, which has been called the hypercharge rotation group, has been discussed by Salam and Polkinghorne,[19] Schwinger,[20] and others. (We remark, however, that the use of the term *hypercharge rotation group* has not been consistent.) The group G_{DA}, the doublet approximation, has been discussed by Pais.[21] Of the other groups in Table IX, the group M_0 is the neutral component of M, and has been discussed by Souriau and Kastler.[22] The group H_G has been discussed by Behrends and Sirlin.[17] The remaining groups, it seems, have not been mentioned explicitly in the previous studies.

The foregoing list of references is by no means complete. In particular, various other attempts were made to understand the structure of elementary particle symmetries in terms of the group O_4^* and its subgroups (e.g. reference 5). We also mention that some of the groups that we encountered have been utilized for global symmetry from a different point of view. An example is the group SU_3, or A_2, which is the basis of the models of Sakata[23] and of Thirring.[24] Models based on representations of B_2 have been constructed by Salam and Ward[25] and by Behrends *et al.*[8]

[18] G. Feinberg and R. E. Behrends, Phys. Rev. **115**, 745 (1959).
[19] A. Salam and J. C. Polkinghorne, Nuovo Cimento **2**, 685 (1955).
[20] J. Schwinger, Ann. Phys. N. Y. **2**, 407 (1957).
[21] A. Pais, Phys. Rev. **110**, 574 (1958).
[22] J. M. Souriau and D. Kastler, Conférence internationale d'Aix-en-Provence sur les particules élémentaires, 1961; Vol. I, p. 169.
[23] S. Sakata, Progr. Theoret. Phys. (Kyoto) **16**, 686 (1956).
[24] W. E. Thirring, Nucl. Phys. **10**, 97, and **14**, 565 (1959).
[25] A. Salam and J. C. Ward, Nuovo Cimento **20**, 1228 (1961).

6. ANTIBARYONS AND MESONS

We will now extend the symmetries, which were enumerated in the last section, to the antibaryons and to the mesons. We recall that these extensions ordinarily require representations of the underlying group other than the representation adopted for the baryons, and therefore the distinction between a group and its representations is particularly significant here.

In this section it will be important to distinguish between the classes of representations. These classes, orthogonal, symplectic, and complex, are defined and discussed in Appendix B. We remark that this classification is also fundamental in the proofs of the inclusion relations, which were stated in the previous section. However, there we were able to present the results easily without referring to this classification.

Let us first consider the antibaryons. Their symmetry is defined by the relation (3.1a):

$$D_{\bar{B}} \otimes D_B \supseteq D^E.$$

We conclude that (see Appendix B),

$$D_{\bar{B}} = D_B^*. \tag{6.1}$$

Next we turn to the mesons. We assume that here the symmetries satisfy the relation (3.1b):

$$D_{\bar{B}} \otimes D_B \otimes D_M \supseteq D^E,$$

or equivalently,

$$D_{\bar{B}} \otimes D_B \supseteq D_M^*. \tag{6.2}$$

One possibility is, in view of (3.1a),

$$D_M = D^E. \tag{6.3}$$

This gives one meson, which will be called *globoscalar*. The other possibilities are of course the more interesting ones, and the case (6.3) will be largely ignored from now on.

To obtain the other possibilities for meson schemes, we have to reduce the Kronecker product $D_B^* \otimes D_B$. This product is an irreducible representation. The reduction into irreducible components may therefore yield, besides orthogonal representations, also complex representations in complex conjugate pairs, and symplectic representations in pairs of equivalent representations. Let us illustrate such reductions, before discussing the significance of the classes of representations for meson schemes.

We will consider the symmetry group D_4^8; we may suppose that this is D_4^+. This representation is of course orthogonal, and the same holds for any sub-

group of D_4^v. Consider first the I-spin group. Then the blocks acting on Σ and on Λ form orthogonal representations, while the two blocks which act on N and on Ξ form a pair of equivalent symplectic representations. If we take the hypercharge gauge group, then the action on N and on Ξ gives complex conjugate pairs of representations. (This argument also shows that a global symmetry group cannot be symplectic.)

Let us return to the mesons. It turns out that the reduction of the products $D_B^* \otimes D_B$ into irreducible components yields mostly orthogonal representations. If we associate the mesons with an orthogonal representation, then one can introduce Hermitian fields by taking suitable linear combinations, and the supermultiplet contains, for each meson, its antiparticle (see Sec. 7). This is familiar for the pion triplet. On the other hand, in the complex or in the symplectic case, one can no longer use Hermitian fields.

The pairs of complex conjugate representations for mesons cannot be excluded, however. An example is given by the groups $A_1^2 e^{*i\phi}$ acting on the K-meson doublets, if there are no higher symmetries. In this case, charge conjugation interchanges the mesons in the two complex conjugate representations. It is easy to see that in the case of other complex representations, the effect of charge conjugation would be analogous. We conclude that if charge conjugation is to be defined, then both representations of a complex conjugate pair, or none, must be used in a theory.

The last possibility is a symplectic representation for mesons. In this case there seems to be no compelling reason to utilize both of the equivalent symplectic representations that occur in the product $D_B^* \otimes D_B$. One can also say that there seems to be no way by which the two supermultiplets of mesons could be distinguished. However, this possibility arises for only one symmetry, which is rather artificial.

We will now give a few examples of meson symmetries. For decomposition of Kronecker products we refer to Appendix B.

(a) For the symmetry B_3^{Spin} we have the following decomposition:

$$B_3^{\text{Spin}} \otimes B_3^{\text{Spin}} \simeq B_3^E \oplus B_3^v \oplus B_3^{21} \oplus B_3^{\bar{5}v}.$$

The representations on the right-hand side (r.h.s.) are all orthogonal. This symmetry allows systems of 7, 21, or 35 mesons, besides a globo-scalar.

(b) For the symmetry A_2^8 we have a more varied behavior:

$$A_2^8 \otimes A_2^8 \simeq A_2^E \oplus A_2^{8a} \oplus A_2^{8s} \oplus A_2^{10} \oplus A_2^{10*} \oplus A_2^{27}.$$

The two representations A_2^{10} and A_2^{10} are complex conjugates, and the other representations are all orthogonal. We may have systems of 8, 20, or 27 mesons. A system of 8 mesons may be coupled to the baryons in two linearly independent ways. The antisymmetric and the symmetric couplings, which we indicated, are the most natural, but linear combinations of these are also allowed.

(c) The symmetry $A_1^2 \otimes B_2^{\text{Spin}}$ is typical of Kronecker products:

$$[A_1^2 \otimes B_2^{\text{Spin}}] \otimes [A_1^2 \otimes B_2^{\text{Spin}}] \simeq D^E \oplus A_1^3 \oplus B_2^v \oplus B_2^{10} \oplus (A_1^3 \otimes B_2^v) \oplus (A_1^3 \otimes B_2^{10}).$$

One can easily show (see also Sec. 7) that the representation A_1^3 gives a triplet of mesons with strangeness $S = 0$, i.e. the pions. For mesons with $S \neq 0$ we have to select one of the other representations. All of these representations are orthogonal.

(d) For the group M see reference 1. Here a number of systems of mesons are possible, and we may easily find representations which, when taken together, describe the usual seven mesons.

(e) For a reducible symmetry, as e.g. H_G, we may proceed in a similar way:

$$[G_2^7 \oplus G_2^E] \otimes [G_2^7 \oplus G_2^E] \simeq 2G_2^E \oplus 3G_2^7 \oplus G_2^{14} \oplus G_2^{27}.$$

Let B' denote the seven baryons other than Λ. Then, we must have 7 mesons for the B'–Λ coupling, and for the B'–B' coupling we may have systems of 7, 14, or 27 mesons. Here a globo-scalar meson could provide a direct B'–B' coupling, and also a direct Λ–Λ coupling.

(f) Let us now consider an example of independent blocks. Let A_1^3 act on the Σ's and let B_2^v act on the N–Λ–Ξ system. Then

$$[A_1^3 \oplus B_2^v] \otimes [A_1^3 \oplus B_2^v] \simeq 2D^E \oplus A_1^3 \oplus A_1^5 \oplus B_2^{10} \oplus B_2^{14} \oplus 2(A_1^3 \otimes B_2^v).$$

There is only one possibility for mesons which would couple the Σ's to the N–Λ–Ξ system, namely the last-mentioned representation, which specifies 15 mesons.

(g) Let us consider independent blocks again, with $A_1^2 \otimes (A_1')^2$ acting on the Σ–Λ system, and B_2^{Spin}, on the N–Ξ system. In this case the mesons which would couple the two systems would have to belong to the representation $[A_1^2 \otimes (A_1')^2] \otimes B_2^{\text{Spin}}$. This is the only example of a symplectic representation for mesons.

TABLE XII. The minimal number of mesons required by each symmetry group which does not contain independent blocks. A sum indicates two or three meson supermultiplets, which would couple different pairs of baryons. [Cf. example (c).] In the case of groups in the column headed by G_{DA}, one could also construct various schemes involving less than seven mesons.

Group	Number of mesons	Group	Number of mesons	Group	Number of mesons
A_2^8	8	H_G	7	G_{DA}	3 + 4
B_2^{Spin}	7	H_B	7 + 21	G_{HR}	
D_4^8	28	H_A	14 + 48	$\bar{K}$	
A_7^8	63	J_Z	3 + 6 + 8	K_{DA}	
$A_1^2 \otimes B_2^{\text{Spin}}$	3 + 5	J_N	3 + 6 + 8	M_0	
$A_1^2 \otimes A_3^4$	3 + 15			M	

Table XII shows the minimal number of mesons that each global symmetry group specifies, for groups which do not contain independent blocks. We note again that the I-spin and the strangeness of these mesons are completely determined.

We conclude this section with a few remarks.

(1) As an alternative to deriving the relation $D_B = D_B^*$ as we did, one can consider the baryons and the antibaryons as independent particles, and the possible symmetries of the 16 particles can be examined by modifying our approach. This point of view is indicated in the work of Lee and Yang.[1] We also note that continuous groups which include charge conjugation were constructed from another point of view by Pauli,[26] and were further discussed by others.[27,28]

(2) In the case of complex representations for the mesons, one can satisfy group invariance without allowing antiparticles, by using only one representation of the complex conjugate pair. Models of local relativistic field theory without antiparticles have in fact been constructed, but an extensive use of indefinite metric was required.[29]

(3) If the baryon and the meson symmetries are irreducible, then there is only one coupling constant for the system. Otherwise more coupling constants are possible, depending on the supermultiplet structure.

(4) In some schemes a large number of mesons is specified. There the mesons do not, as a rule, split into few isomultiplets with large $|\mathbf{T}|$, but, rather, into many isomultiplets. These are distinguished by the strangeness and by some new quantum numbers. (See Sec. 7.)

(5) The last remark concerns the Yang–Mills fields,[6,30] or the vector mesons. A general procedure of incorporating such fields was worked out by Utiyama.[31] It is seen that they are associated with the adjoint representation D_{ad} of the underlying symmetry group.[32] (See Appendix B.) While the coupling to the vector mesons is not of the Yukawa type, our assumption (3.1b) still applies, and the vector mesons can be coupled to the baryons if

$$D_B^* \otimes D_B \supseteq D_{ad}.$$

This relation is indeed valid for all connected symmetry groups mentioned in this paper. The work of Utiyama shows that the foregoing relation should be rather general.

7. QUANTUM NUMBERS AND INVARIANT AMPLITUDES

The concept of global symmetry refers primarily to interacting systems, and we will now discuss some consequences of the schemes. While the consequences can be tested by experiments in principle, one should not forget that such tests would be subject to all limitations and ambiguities of global symmetry. (See also the next section.) Our aim here is to show the physical interpretation of the schemes by discussing some direct consequences, and therefore we do not attempt to give an exhaustive tabulation.

Throughout this section we will denote the generators which correspond to the third component of I-spin and to the hypercharge gauge group by T_3 and by Y, respectively.

In the study of representations of simple Lie algebras and Lie groups, a basic concept is that of a weight vector; this is a simultaneous eigenvector of a maximal commuting set of operators. In Appendix B we explain this concept in a little more detail, and we prove the following lemma. This lemma depends on the fact that, except for Σ^0 and Λ, there is no degeneracy of eigenvalues when T_3 and Y are considered simultaneously.

Lemma 7.1. *Let a symmetry group, or its Lie algebra, be given. Then one can choose a maximal commutative subgroup, or subalgebra, in such a way, that each baryon (except as we note) is a weight vector. An exception to this may occur if* Σ^0 *and* Λ *both appear*

[26] W. Pauli, Nuovo Cimento 6, 204 (1957).
[27] F. Gürsey, Nuovo Cimento 7, 411 (1958).
[28] N. Kemmer, J. C. Polkinghorne, and D. L. Pursey, Rept. Progr. Phys. 22, 390 (1959).
[29] M. Günther, Phys. Rev. 125, 1061 (1962).
[30] C. N. Yang and R. L. Mills, Phys. Rev. 96, 191 (1954).
[31] R. Utiyama, Phys. Rev. 101, 1597 (1956).
[32] S. L. Glashow and M. Gell-Mann, Ann. Phys. N. Y. 15, 437 (1961).

in the same supermultiplet. In this case two of the weight vectors may be linear combinations of Σ^0 *and* Λ.

This lemma is illustrated in Fig. 1, in which we relate the baryons with the eigenvalues of the commuting elements, for three representations. The lemma now states the following. Let us consider three representations, which are equivalent to those which we first assumed, and which still satisfy the inclusion requirements. The arrangement of eigenvalues of the baryons would then remain essentially unchanged. E.g., for the group B_3^v, the new symmetry could not have a nonzero generator which would yield zero eigenvalue for each baryon. This conclusion is not obvious *a priori*, since we refer the new representation to the original basis vectors, rather than to a new set. The configurations, such as in Fig. 1, are determined by the representation alone, and can be readily constructed for the familiar representations. We remark that Wigner[33] was one of the first to make extensive use of such diagrams.

Lemma 7.1 can therefore be thought of as a uniqueness lemma. In the sequel we shall assume such diagrams when convenient, and we shall exploit their properties, especially their symmetries and the implied assignments of eigenvalues.

Our justification for the assumption (5) of Sec. 3 consists largely of the foregoing lemma. We may add that this lemma, like Lemma 4.3, would not be valid, if we were to ignore the hypercharge (or strangeness).

The case of meson symmetries is analogous. As with the baryons, a representation completely determines the isomultiplet structure and other quantum numbers. We illustrate this in Fig. 2 for the baryonic symmetry A_2^8, i.e., we show the familiar decomposition of $A_2^8 \otimes A_2^8$. One could easily carry out an analysis of this kind for any of the other symmetry groups, and our assertions in Sec. 6 about the quantum numbers for the mesons could be verified. We shall not do this, however.

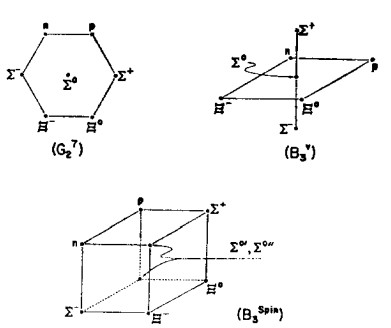


FIG. 1. Weight diagrams for G_2^7, B_3^v, and B_3^{Spin}, together with the assignments of the baryons. The directions of T_3 and of Y are implied by each diagram. In each of the latter two diagrams, T_3 is skew to the figure. The linear combinations $\Sigma^{0\prime}$ and $\Sigma^{0\prime\prime}$ may be identified with the baryons Y and Z.

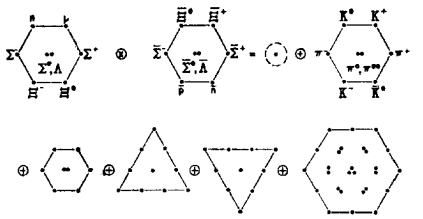


Fig. 2. Weight diagrams showing the decomposition of the product $A_2^8 \otimes A_2^8$. Assignments of the baryons, of the antibaryons, and of the mesons are shown. The meson π^∞ must be pseudoscalar and isoscalar. The diagrams without assignments provide other possibilities for meson schemes.

Let us now turn to the various types of consequences of the symmetry schemes.

Commuting group generators

The Lie algebra of a symmetry group always contains iT_3 and iY. These two operators are diagonal with respect to the baryons, and this implies the familiar conservation laws, or selection rules. However, the symmetry groups, which are of rank 3 or higher, have other linearly independent generators which commute with T_3 and with Y. These other generators are diagonal with respect to the baryons (or almost diagonal, cf. Lemma 7.1), and this leads to additional observables, and additional selection rules. Let us give two examples.

For the first example we will take the reducible symmetry H_B. This corresponds to the group B_3^v, with Λ, a globo-scalar that can be ignored. The weight diagram for B_3^v is shown in Fig. 1. Besides T_3 and Y we have a third observable. We may choose, e.g., a quantity X which is measured in this diagram along an axis parallel to the n–p edge. Then

$$X_p = X_{\Xi^0} = +1,$$
$$X_n = X_{\Xi^-} = -1,$$
$$X_{\Sigma^+} = X_{\Sigma^0} = X_{\Sigma^-} = 0.$$

Therefore the process $\Xi^0 + p \rightarrow \Sigma^+ + \Sigma^0$ is forbidden, since $1 + 1 \neq 0 + 0$.

For the second example (a rather unphysical one) we take the group $SU_3 = A_2^3$ as the symmetry group for the Σ's. This could be, e.g., an independent block in a reducible symmetry. The group A_2 has rank 2, and there exists besides T_3 a second observable, and a new selection rule. The process $\Sigma^+ + \Sigma^- \rightarrow \Sigma^0 + \Sigma^0$ is now forbidden.

[33] E. Wigner, Phys. Rev. 51, 106 (1937).

In the case of B_3^{Spin} the weight vectors correspond to the baryons Y and Z rather than to Σ^0 and Λ, and the foregoing analysis is not immediately applicable. However, new selection rules appear when the mesons are considered, as was shown by Pais.[21]

This argument of Pais, we may add, has sometimes been interpreted as proving the nonexistence of symmetries higher than (I-spin) $\times$ (hypercharge gauge). We should like to point out that Pais' argument establishes contradiction with experiment for symmetries of rank 3 or higher, but not for those of rank 2. There are four groups in our list, namely A_2^8, H_G, G_{HR}, and $\bar{K}$, to which this argument does not apply.

Symmetry Properties of Weight Diagrams

These properties lead to some immediate consesequences. For example, the symmetry of the diagram for G_2^7 (Fig. 1), implies the following:

$$\sigma(\Sigma^+ + n \to p + \Sigma^0) = \sigma(\Xi^- + n \to \Sigma^- + \Sigma^0).$$

This equality can be contrasted with the following for the group A_2^8:

$$\sigma(\Sigma^+ + n \to p + \Sigma^0) = \sigma(\Xi^- + n \to \Sigma^- + \Sigma^{0\prime}),$$

where $\Sigma^{0\prime}$ is a linear combination of Σ^0 and Λ. These considerations also suggest branching ratios, which may be more useful than the absolute cross sections, in testing a theory. For instance, the group A_2^8 implies the relation

$$\frac{\sigma(\Sigma^+ + n \to \Sigma^+ + n)}{\sigma(\Sigma^+ + n \to p + \Sigma^0)} = \frac{\sigma(\Xi^- + n \to \Xi^- + n)}{\sigma(\Xi^- + n \to \Sigma^- + \Sigma^{0\prime})}.$$

The symmetries that are displayed by the weight diagrams, are roughly in the same relation to the invariance under the whole group, as charge symmetry is to charge independence. Various attempts were made to exploit these symmetry properties in applications of global symmetry.[34,35]

Invariant Amplitudes

A general procedure to study the invariance under a symmetry group depends on decomposing the various reaction amplitudes into sums of invariant amplitudes. In particular, the two foregoing types of consequences can be deduced in this way as special cases. For definiteness let us consider the reaction

$$B_1 + B_2 \to B_3 + B_4,$$

where all of the baryons are associated with the representation D. Invariance under the symmetry group then implies that the invariant amplitudes are associated with the identity blocks in the Kronecker product

$$D^* \otimes D^* \otimes D \otimes D.$$

If the representations D^i in the decomposition of $D \otimes D$ should be pairwise inequivalent, then each representation D^i gives D^E once, namely, when combined with D^{i*} (see Appendix B), and therefore defines one invariant amplitude for the process in question. (For further details see reference 8 and the work of Freund *et al.*[36])

The decompositions for some of the global symmetries are as follows:

$$\text{for } B_3^{\text{Spin}}, \quad \mathfrak{I} = a_1T_1 + a_7T_7 + a_{21}T_{21} + a_{35}T_{35},$$

$$\text{for } D_4^8, \quad \mathfrak{I} = a_1T_1 + a_{28}T_{28} + a_{35}T_{35},$$

$$\text{for } A_7^8, \quad \mathfrak{I} = a_{28}T_{28} + a_{36}T_{36}.$$

The invariant amplitudes T_i depend only on the kinematical invariants of the process. The numbers a_i are related to the Clebsch–Gordan coefficients for the group in question, and are determined by the baryons, as Lemma 7.1 shows. The subscripts in a_i and T_i refer to the dimension of the corresponding irreducible representation D^i in $D \otimes D$.

The foregoing examples show that, as a rule, when the symmetry is increased, then some of the invariant amplitudes combine, and the number of independent invariant amplitudes is decreased.

Resonances

The foregoing analysis can be readily applied to the study of resonances. If a resonance is present, then it must be associated with one of the invariant amplitudes, and therefore it should be also observed in some other process as well. This approach to resonances is discussed further in references 1 and 8.

Casimir and Racah Operators

The invariance under the I-spin group implies not only the conservation of T_3 but also of $\mathbf{T}^2$, as is well known. In case of other Lie groups, one can construct analogous operators, the Casimir and the Racah operators. (See Appendix B.) In this way one can define a set of conserved quantum numbers. Two examples are given to show how this relates to the other consequences. We note that the Casimir

[34] L. Radicati and S. Speiser, Nuovo Cimento **24**, 386 (1962).

[35] C. A. Levinson, H. J. Lipkin, and S. Meshkov, Nuovo Cimento **23**, 236 (1962).

[36] P. G. O. Freund, A. Morales, H. Ruegg, and D. Speiser, Nuovo Cimento **25**, 307 (1962).

and the Racah operators can be used for the symmetry group, or for any of its subgroups.

For the first example let us consider the symmetry (5.1a), where B_2^n acts on the N–Λ–Ξ system, and A_1^3 acts independently on the Σ's. We may define $\mathbf{T}_\Sigma$ with reference to the group A_1^3 only, and then $\mathbf{T}_\Sigma$ gives zero for the N–Λ–Ξ system. Similarly, there exists a Casimir operator $\mathbf{T}_{N\Lambda\Xi}$, which gives zero on the Σ's. We conclude that the following processes are forbidden:

$$\Xi^- + p \to \Sigma^0 + \Lambda,$$
$$\Sigma^+ + \Sigma^0 \to \Lambda + \Sigma^+.$$

In these two processes, the quantum numbers which correspond to commuting group generators are conserved. One could also use the invariant amplitudes to show that these processes are forbidden, but the foregoing argument is much simpler.

For the second example let us consider the A_2^8 symmetry and the 20 mesons which correspond to the representations A_2^{10} and $A_2^{10^*}$; i.e., we have 10 charge-conjugate pairs. The isomultiplet structure shows that we have a meson σ^0 and its charge conjugate $\bar{\sigma}^0$, both with $T_3 = 0$, $Y = 0$, and $|\mathbf{T}| = 1$. The second-order Casimir operator for A_2 also gives equal values for σ^0 and $\bar{\sigma}^0$. However, there is a third-order Racah operator for A_2, and this operator gives different values for the two mesons σ^0 and $\bar{\sigma}^0$. Therefore the transition $\sigma^0 \leftrightarrow \bar{\sigma}^0$ is forbidden. Of course, this is also expected from other considerations.

Nonconnected Groups

The consequences which are peculiar to nonconnected groups can be understood most easily with the help of an example, and we therefore note reference 1, where the group M is discussed. Here we only make three comments: (a) The quantum numbers which are associated with the discrete generators are multiplicative, rather than additive. (b) The decomposition into invariant amplitudes is in no way restricted to connected groups, and applies to nonconnected groups equally well. (c) Consequences which depend on properties of the Lie algebra clearly are not affected, if the connected group is enlarged with the help of discrete generators.

8. A FEW FINAL REMARKS

We can summarize our results as follows: We have presented all possible connected symmetry groups and a few nonconnected symmetry groups. We also showed how the symmetries should be extended to the antibaryons, and how they can be extended to the mesons. Finally, we discussed some direct consequences of the symmetry groups. Our work was based on the general and on the specific assumptions which we listed in Secs. 2 and 3. We should now like to make a few concluding remarks.

(1) We have considered global symmetry as referring primarily to interacting particles. We cannot, of course, escape the ambiguities which arise when we take mass differences into account. Still, one may hope that the consequences, which we examined in Sec. 7, can be subjected to some meaningful tests. This problem has been discussed extensively by various authors, and therefore we confine ourselves to four comments.

(a) The predictions of global symmetry may be valid in the limit of very high energies. The most sensitive test of the symmetries would then depend on the invariant amplitudes.

(b) It may be possible to remove these ambiguities, at least partially, by comparing branching ratios rather than the absolute cross sections.

(c) Qualitative tests of global symmetry may be more successful than the quantitative ones, and the resonances immediately suggest themselves. In particular, Lee and Yang mention the excited Y^* state[37] as one of the motivations of their paper.

(d) Future theories may teach us more about the bare particle masses and bare coupling constants. An investigation in this direction was recently attempted by Gell-Mann and Zachariasen.[38]

(2) The concept of global symmetry may be more useful in the study of electromagnetic and weak interactions than for strong interactions. The relevance to weak interactions has been discussed by many authors, and especially by Pais.[39] The relevance to electromagnetic interactions, in particular to the electromagnetic structure of the baryons, has also been considered (e.g. in references 18 and 35).

(3) If our framework of eight baryons is a definite one, then we may say that the groups A_2^8 and H_G are the most attractive possibilities. These groups provide a maximum of formal unification, and require a minimum of new assumptions. For example, we do not have the trouble with additional selection rules, and the resulting meson schemes are reasonably satisfactory. On the other hand, while the groups G_{HR} and $\bar{K}$ likewise do not imply additional selec-

[37] M. Alston, L. W. Alvarez, P. Eberhard, M. L. Good, W. Graziano, H. K. Ticho, and S. G. Wojcicki, Phys. Rev. Letters 5, 520 (1960).

[38] M. Gell-Mann and F. Zachariasen, Phys. Rev. **123**, 1065 (1961).

[39] A. Pais, Rev. Mod. Phys. **33**, 493 (1961), and references given there.

tion rules, these two groups are rather trivial extensions of the group (I-spin) $\times$ (hypercharge gauge). (One other somewhat attractive possibility is the group B_2^{10}, which presupposes 10 baryons.[8])

(4) Since global symmetry in the usual forms is not valid, various interactions were suggested to explain the observed deviations from the postulated symmetry schemes. We refer to the papers mentioned in Sec. 5 for examples, and we confine ourselves here to one comment, which is particularly relevant from our point of view. An examination of our global symmetry groups shows that many of them exhibit a complete symmetry between N and Ξ. In fact, the only exceptions are J_N, J_Ξ, and the symmetries which contain independent blocks. However, these exceptions require a conservation law for the nucleons and also for the cascades, and so cannot be considered satisfactory. It seems therefore difficult to incorporate the N–Ξ asymetry into a global symmetry scheme in a natural way.

(5) A basic question in the study of symmetry is the following: How is a symmetry group to be understood, if it should be found experimentally? We will make only a brief comment here. If in the framework of a dynamical theory one wants to derive symmetries, it seems that one should try to derive the underlying group, rather than specific representations. This is suggested especially by Secs. 6 and 7. It is suggested also by the recent studies of the vector mesons, and, one can say, by the aim for generality and universality in physical theories.

ACKNOWLEDGMENT

We would like to thank Professor C. N. Yang (who suggested this problem) for many stimulating discussions. We also welcome this opportunity to acknowledge fruitful discussions with the following of our colleagues: Drs. and Profs. T. D. Lee, A. Pais, D. L. Pursey, L. Radicati, and E. P. Wigner. One of us (D. R. S.) would further like to acknowledge fruitful discussions with Drs. and Profs. R. Behrends, C. Fronsdal, N. Iwahori, J. Ward, and J. A. Wolf. Finally, we should like to express our gratitude to Professor J. R. Oppenheimer for the generous hospitality offered to us by the Institute for Advanced Study.

One of us (J. T.) would like to acknowledge gratefully the support of an IBM fellowship during the completion of this work.

APPENDIX A. SOME PROPERTIES OF LIE GROUPS

In this appendix we summarize some of the properties of Lie groups. We do not attempt to attain full generality in our discussion, and in various places we emphasize the intuitive idea, rather than the complete rigor. For a detailed presentation of the theory we refer to the texts of Chevalley,[40,41] Pontriagin,[10] Weyl,[42] and Wigner.[43]

This appendix deals with abstract groups. The following appendix is devoted to group representations.

1. *Concept of a Lie group.* A *Lie group* is, roughly speaking, a parameter group, in which the group operations define differentiable functions of the parameters. To a Lie group one can therefore associate topological, differential, and algebraic properties. We shall consider these properties in turn.

2. *Topological properties.* Some of the useful topological concepts are compactness, connectedness, connectivity, dimension, and invariant integration. For definitions of these concepts and their fundamental properties we refer to the book of Pontriagin.[10] One should not confuse connectedness with connectivity. Connectedness is referred to when we ask, whether a given set is connected, and connectivity, when we ask whether it is simply connected.

3. *Differential properties.* The most important of the differential properties is the existence of a Lie algebra. This is the space (a linear space) of the infinitesimal generators of the group. One may say that the Lie algebra serves to investigate a Lie group in a similar way as the derivative serves to investigate a function. The Lie algebra admits a bracket operation, which is anticommutative, and which satisfies the Jacobi identity, i.e., which is analogous to the commutator bracket. A Lie group defines a unique Lie algebra. Similarly, a Lie algebra defines a unique Lie group (up to isomorphism) if we require that the group be connected and simply connected, or if we impose other suitable restrictions.

Two Lie groups which have isomorphic Lie algebras but are not necessarily isomorphic, are said to be *locally isomorphic*. A well known example of such Lie groups is given by SU_2 and SO_3 (the 2×2 unitary unimodular group and the 3×3 proper orthogonal group, respectively).

4. *Algebraic properties.* Group-theoretic concepts include direct products, subgroups, invariant subgroups, etc. In the case of a Lie group, such properties have their analogues in the Lie algebra, and this

[40] C. Chevalley, *Theory of Lie groups* (Princeton University Press, Princeton, New Jersey, 1946), Vol. I.
[41] C. Chevalley, *Theorie des groupes de Lie* (Hermann & Cie., Paris, 1951 and 1955) Vol. II and III.
[42] H. Weyl, *The classical groups* (Princeton University Press, Princeton, New Jersey, 1946), 2nd Ed.
[43] E. P. Wigner, *Group Theory and its Applications to the Quantum Mechanics of Atomic Spectra* (Academic Press Inc., New York, 1959).

greatly simplifies the study of structure of Lie groups.

The concept of a simple group is fundamental. We remark here that in discussing Lie groups, the definition of a simple group differs from the one used otherwise in group theory. We say that a Lie group is simple, if it has no proper invariant subgroups which are Lie groups (but it may have invariant discrete subgroups). This definition leads to a correspondence between simple Lie groups and simple Lie algebras.

In the sequel we shall refer to the following two lemmas (see Chevalley,[40] p. 35, and Pontriagin,[10] p. 77).

Lemma. In a Lie group, the connected component of the identity is an invariant subgroup.

We refer to the connected component of the identity as the *neutral* component. The foregoing lemma implies, in particular, that a simple Lie group is necessarily connected.

Lemma. In a connected Lie group, every discrete invariant subgroup is central.

5. *The classical groups.* These groups are the best known examples of Lie groups.[40,42] The following are some of the classical groups. All of these but the first-mentioned are compact. The subscript n refers to $n \times n$ matrices.

$GL_n(C)$—the full linear group over the complexes,
U_n—the unitary group,
SU_n—the unitary unimodular group,
SO_n—the proper orthogonal group,
O_n^*—the orthogonal group including reflections,
Sp_n—the unitary restricted symplectic group, provided n is even.

The symplectic groups are perhaps the least familiar of these. The group Sp_n leaves invariant a nondegenerate skew-symmetric bilinear form; such a form can exist only in a space of even dimension. One can define Sp_n explicitly as the group of $n \times n$ unitary matrices which satisfy

$$A^T J A = J,$$

where $J^T = -J$, $J^2 = E$. The two following choices (or representations) for J are commonly used:

$$J = \begin{bmatrix} & E_{\frac{1}{2}n} \\ -E_{\frac{1}{2}n} & \end{bmatrix},$$

$$J = \begin{bmatrix} J_2 & & \\ & \ddots & \\ & & J_2 \end{bmatrix}, \quad \text{where} \quad J_2 = \begin{bmatrix} & 1 \\ -1 & \end{bmatrix}.$$

One can also say that Sp_n is associated with rotations in a quarterionic space of dimension $\frac{1}{2}n$, in much the same way as U_n and SO_n are associated with rotations in complex and in real spaces, respectively.

6. *Commutative subgroups and subalgebras, and rank.* A commutative Lie algebra is one in which the bracket operation yields identically zero. Such a Lie algebra corresponds to a commutative Lie group.

Basic concepts in the study of Lie algebras and Lie groups include those of a maximal commutative subalgebra, or, of a maximal commutative connected subgroup. It is a theorem that if the group is simple, then all such subalgebras (and also subgroups) are conjugate to one another within the group, and are *a fortiori* isomorphic. The common dimension of these sets is called the *rank* of the group or algebra. In the case of a simple noncommutative Lie algebra, or a direct sum of such, a maximal commuting subalgebra is called a *Cartan subalgebra.*

7. *The classification of Cartan and Killing.* The classification of simple complex Lie algebras was established by Cartan and Killing.[2] (A Lie algebra is complex if it forms a vector space over the complexes. A complex Lie group is one whose Lie algebra is complex.) This classification also applies to compact Lie groups by virtue of a theorem of Weyl.[3] (See also Pontriagin,[10] p. 271.)

We now present this classification, and we introduce the notation that is generally used. The simple, therefore connected, compact Lie groups consist of the (commutative) circle group, and of the following noncommutative groups. We consider the latter as abstract groups. They have the dimension and local isomorphisms indicated. The subscript n here indicates the rank of the group.

A_n—loc. isom. to SU_{n+1} $(n \geq 1)$, dim. $n^2 + 2n$,
B_n— SO_{2n+1} $(n \geq 1)$, $2n^2 + n$,
C_n— Sp_{2n} $(n \geq 1)$, $2n^2 + n$,
D_n— SO_{2n} $(n \geq 3)$, $2n^2 - n$,
also E_6, E_7, E_8, F_4, G_2.

The five last-mentioned groups have, respectively, the dimensions 78, 133, 248, 52, and 14. Their representations of least degrees are, respectively, those of degrees 27, 56, 248, 26, and 7. We see that of these only the group G_2 can be relevant to global symmetry, such as we consider.

Among the foregoing groups the following local isomorphisms occur, and no others:

$A_1 \cong B_1 \cong C_1$, $B_2 \cong C_2$, and $A_3 \cong D_3$.

Finally, we note that in discussing the classification one often refers to *semi-simple* Lie groups and algebras. These algebras are the direct sums of simple noncommutative Lie algebras. (Equivalently, these are the algebras which have no invariant commutative subalgebras.) Many statements concerning simple Lie groups and algebras have an immediate generalization to the semi-simple ones.

8. *Connectivity.* Given any connected Lie group G, one can construct a locally isomorphic, connected, simply connected Lie group G' (the universal covering group of G). If G is semisimple and compact, the same will be true of G'. The simply connected groups are particularly significant, since every connected group which is locally isomorphic to G and to G', is isomorphic to a group G'/D, where D is a discrete (therefore central) subgroup of G'.

Of the classical groups which are simple and compact, the groups SU_n and Sp_n are simply connected, but the SO_n are not. The construction of Spin (n), a simply connected group which is locally isomorphic to SO_n, is described by Chevalley,[40] p. 61 ff.

We shall use the symbols A_n, B_n, etc., interchangeably for the simply connected compact Lie groups and for their Lie algebras.

9. *The general compact Lie group.* The foregoing classification enables us to describe concisely how an arbitrary connected compact Lie group can be constructed, up to isomorphism. (See Pontriagin,[10] p. 282.)

Theorem. Every connected compact Lie group G can be constructed in the following way. Form the direct product

$$\tilde{G} = H_1 \times H_2 \times \cdots \times H_n \times K_1 \times \cdots K_m,$$

where the H_i are appropriate simple, noncommutative, compact Lie groups which are connected and simply connected; i.e., each H_i is one of the groups $A_n, \cdots, G_2$ in the foregoing classification. The K_i are circle groups. Then G is isomorphic to $\tilde{G}/D$, where D is a discrete (therefore central) subgroup of $\tilde{G}$.

APPENDIX B. SOME PROPERTIES OF GROUP REPRESENTATIONS

In this appendix we summarize some of the properties of group representations which are relevant to this paper, and we give references for the discussions of various special topics. We also prove the lemmas which are stated in the text without proof. General references here are the same as for Appendix A.

1. *Concept of a representation.* In general, a representation is a mapping of a set S of elements into a set of matrices, or of linear transformations on a vector space. If we associate with the set S certain properties like those related to algebraic operations or to continuity, then we also require that these properties be preserved under the mapping. This general definition applies to the representations of groups, of Lie algebras, and of other systems. In particular, in the case of Lie algebras, one requires that the Lie algebra bracket correspond to the commutator bracket of matrices or of transformations.

A representation by $n \times n$ matrices, or by transformations on an n-dimensional vector space, is said to be of *degree n*.

Throughout this paper the equivalence of representations $(D \rightarrow RDR^{-1})$ is denoted by the symbol $\simeq$.

2. *Unitarity and reducibility.* A fundamental and also well known property of group representations is the following: Every representation of a compact group is equivalent to a unitary representation, i.e., to a representation involving only unitary matrices or transformations. This is proved easily with the help of invariant integration. It follows that every representation of a compact group is completely reducible, since the orthogonal complement to an invariant subspace is likewise invariant. In particular, the symmetry groups are completely reducible.

3. *Multivaluedness and faithfulness.* The connectivity of groups is related to what are sometimes called multivalued representations. A well-known example is the correspondence $SO_3 \rightarrow SU_2$. Such a correspondence is a local isomorphism. Moreover, it always becomes single-valued when we replace the first group by its simply connected covering group. Note also that a multivalued representation of a circle group forms again an isomorphic circle group, and consequently multivaluedness can be easily eliminated here.

We recall that a representation of a group is faithful, if only the group identity maps into the unit matrix. It follows that if a linear group is represented in a multivalued manner, as in the foregoing example, then the inverse correspondence is not faithful.

4. *The representations of degree 8 or less.* We will now give a few references, where the representations of Tables I(a) and (b) are discussed in further detail. For a general theory see references 3 and 12 (Appendix).

The representations of A_1 are familiar, and are discussed by Wigner.[43]

The linear groups SU_n, SO_n, and Sp_n, which correspond to A^n_{n-1}, $B^n_{\frac{1}{2}(n-1)}$ or $D^n_{\frac{1}{2}n}$, and $C^n_{\frac{1}{2}n}$, respectively, are among the classical groups, and are discussed thoroughly by Chevalley[40] and by Weyl.[42]

The representations of groups of rank 2 are discussed by Behrends *et al.*[8] The representations of A_2 are also described by Wigner[33] and by Wess.[44] The automorphism group of octonions, G^7_2, is discussed by Chevalley.[45] (See also a note by Pais.[46])

The spinor and the semispinor representations are discussed by Brauer and Weyl[47] and by Cartan.[48] The analogy of the three representations, D^v_4, D^{Spin}_4, and $D^{(\text{Spin}')}_4$, is known as the principle of triality; see references 45 and 48.

For completeness we remark that every representation of the circle group is again the circle group. The circle group is not listed in Tables I(a) and (b).

5. *Kronecker products and Lemma* 4.1. The following assertion can be readily verified with the help of characters (see Wigner,[43] Chap. 16): Let D' and D'' be irreducible representations of the compact groups $\mathcal{G}'$ and $\mathcal{G}''$ respectively; then the Kronecker product $D' \otimes D''$ is an irreducible representation of the direct product $\mathcal{G}' \times \mathcal{G}''$. It can also be shown, with the help of the Peter–Weyl theorem (Pontriagin,[10] Chap. 4) that every irreducible representation of $\mathcal{G}' \times \mathcal{G}''$ is of this form.

We emphasize that the concepts of direct product and of Kronecker product must be clearly distinguished. Note, e.g., that the Kronecker product $SU_2 \otimes SU'_2$ is not a faithful representation of the direct product $SU_2 \times SU'_2$, since $(\pm E_2, \pm E'_2) \to E_4$. Note also that in an irreducible representation of direct products, only one commutative factor can be faithfully represented.

Lemma 4.1 follows at once from the foregoing considerations and from our knowledge of the most general connected compact Lie group; one can readily see that the commutative factors do not enter into the group $\mathcal{G}$.

6. *Proof of Lemma* 4.2. Let us assume for definiteness that the group G is of the form $\binom{H}{I}$, where H and I are simple groups; however, our argument is quite general. Consider the subgroup G which consists of matrices of the form $\binom{E}{q}$ where $q \in I$. It is immediate that such matrices define an invariant subgroup of I, which necessarily equals I, or E_I, or is a proper invariant subgroup which is discrete. In the first case the blocks are independent. In the second case, to each $p \in H$ there corresponds a unique $q \in I$ such that $\binom{p}{q} \in G$. This defines a homomorphism $H \to I$. Since H is simple, this is an isomorphism, or a local isomorphism. Similarly, the last case implies a local isomorphism. The lemma follows. (See the foregoing discussions of connectivity and of multivaluedness.)

7. *Proof of Lemma* 4.3. This lemma depends on the fact that, in a reducible representation, the blocks which are associated with inequivalent representations are uniquely determined. We can state this more explicitly. Let us suppose that a reducible representation breaks into blocks of irreducible representations which act on the respective invariant subspaces

$$V_{11}, \cdots, V_{1,n(1)}, V_{21}, \cdots, V_{m,n(m)}.$$

The representations on $V_{i1}, \cdots, V_{i,n(i)}$ are to be equivalent, but the representations on V_{ij} and V_{kl} are to be inequivalent if $i \neq k$. Then the spaces

$$V_i = V_{i1} \oplus \cdots \oplus V_{i,n(i)}$$

are uniquely determined, but further decomposition of V_i is not unique. A simple proof is given by Wigner[43] (Chap. 12) for representations of compact groups, but our assertion has a much more general validity.

Now, the blocks which a symmetry group must contain are all inequivalent, and if

$$R^{-1}\begin{bmatrix} A_1^2 e^{-i\phi} & & & \\ & A_1^3 & & \\ & & 1 & \\ & & & A_1^2 e^{i\phi} \end{bmatrix} R \subseteq \begin{bmatrix} H & \\ & I \end{bmatrix},$$

then it follows, e.g., that the two vectors $R^{-1}(\Xi^-)$ and $R^{-1}(\Xi^0)$ must both belong to one of the two subspaces defined by the matrices $\binom{H}{I}$. The same applies to the other isomultiplets. We conclude that $\binom{H}{I}$ is equivalent to a linear group, which has a similar block structure, and which, moreover, satisfies the inclusion requirements. The proof is complete.

8. *Proof of Lemma* 5.1. Suppose we are given a compact linear group which fulfills the inclusion requirements; then we can find an equivalent group G of unitary matrices. In particular, the matrices which belong to the I-spin group and to the strangeness-like group in the original linear group map onto a group G^0 of unitary matrices. Now, the

[44] J. E. Wess, Nuovo Cimento **15**, 52 (1960).
[45] C. Chevalley, *The Algebraic Theory of Spinors* (Columbia University Press, New York, 1954).
[46] A. Pais, Phys. Rev. Letters **7**, 291 (1961).
[47] R. Brauer and H. Weyl, Am. J. Math. **57**, 425 (1935). [Reprinted in *Selecta*, reference 3, p. 431 ff.]
[48] E. Cartan, *Leçons sur la théorie des spineurs* (Hermann & Cie., Paris, 1938) vol. I and II.

equivalence between G^0 and its canonical representation can be expressed by a unitary matrix u (Wigner,[43] p. 78, Theorem 1A). We see that the group uGu^{-1} fulfills the specified requirements, and the proof is complete.

9. *Classes of representations.* The results that we will describe here are due to Frobenius and Schur[49] and to Malcev.[50] [See also Wigner[43] (Chap. 24), and Dynkin[12] (Appendix).]

If we are given a representation D, then the complex conjugate matrices define a representation D^*, which may or may not be equivalent to D. In the former case, moreover, the equivalence may be expressed by symmetric or by skew-symmetric matrices; the two possibilities are mutually exclusive if D is irreducible.

This classification can also be described as follows. A representation D may or may not leave invariant a nondegenerate bilinear form (x, y) [i.e. $(x, y) = 0$ for all x (or for all y) implies $y = 0$ (or $x = 0$)]. In the former case the form may be symmetric, and then $D \subseteq SO_n$ (up to equivalence), or the form may be skew-symmetric, and then $D \subseteq Sp_n$. If D is irreducible, then an invariant, bilinear, nondegenerate form is unique, up to a scalar multiple. We therefore define three classes of representations, which can be conveniently designated as classes ± 1 and 0:

(a) D is *orthogonal* if $D^* = CDC^{-1}$, where $C = C^T$; then $D \simeq D' \subseteq SO_n$; class $+1$;

(b) D is *symplectic* if $D^* = CDC^{-1}$, where $C = -C^T$; then $D \simeq D' \subseteq Sp_n$; class -1;

(c) D is *complex* if D^* is not equivalent to D; class 0.

One should distinguish between an orthogonal group, i.e. SO_n, and an orthogonal representation, as defined in (a); this also applies to symplectic groups and representations. In particular, an orthogonal representation need not be real, but it is equivalent to a real one.

If D is reducible, it is still meaningful to say that D does or does not leave invariant a bilinear form, and hence, that it belongs to a class ± 1 or 0. However, D may now be simultaneously orthogonal and symplectic. For example, the linear group $({}^{SU_2}{}_{SU_2})$ is symplectic, since SU_2 is symplectic, and is orthogonal, since it is equivalent to a subgroup of SO_4.

The following properties of this classification are relevant to us:

(1) The representation $D \otimes D$ is orthogonal and contains D^E. Moreover, $D \otimes D' \supseteq D^E$ implies $D' = D^*$ if D and D' are assumed to be irreducible.

(2) Let an orthogonal (or symplectic) representation be reduced to irreducible components. Then the complex representations occur only in complex conjugate pairs, and the symplectic (or orthogonal) representations, in pairs of equivalent representations. Cf. the foregoing example $({}^{SU_2}{}_{SU_2})$.

(3) Let D' and D'' be representations of distinct groups (which may be isomorphic), of classes λ' and λ'', respectively. Then the representation $D' \otimes D''$ is of class $\lambda'\lambda''$.

(4) Let D' and D'' be representations of the same group, of classes λ', $\lambda'' \neq 0$, respectively. Then the representation $D' \otimes D''$ is of class $\lambda'\lambda''$.

Statement (1) depends on the fact that a unitary representation leaves invariant a Hermitian form; this is the content of the assertion that $D \otimes D^* \supseteq D^E$. Statements (3) and (4) express the behavior of symmetry properties of the invariant form, when one makes the extension to the tensor spaces.

The foregoing properties are valid for representations of arbitrary compact groups, and are in no way restricted to Lie groups. One can also make generalizations to noncompact groups.

Of the representations in Tables I(a) and (b), the complex ones are A_n^{n+1} (for $n \geq 2$) and A_2^6. The representations A_1^{2m} and C_n^{2n} are symplectic; this includes B_1^{Spin} and B_2^{Spin}. All others are orthogonal. We also note that the nonidentity irreducible representations of the circle group are complex.

10. *Decomposition of Kronecker product.* The Kronecker product of two representations of the same group is in general reducible, and various examples of this are well known. The following decompositions are relevant to us:

(1) $B^v_{\frac{1}{2}(n-1)} \otimes B^v_{\frac{1}{2}(n-1)}$
$\simeq B^E_{\frac{1}{2}(n-1)} \oplus B^{\frac{1}{2}n(n-1)}_{\frac{1}{2}(n-1)} \oplus B^{\frac{1}{2}n(n+1)-1}_{\frac{1}{2}(n-1)}$,

(2) $D^v_{\frac{1}{2}n} \otimes D^v_{\frac{1}{2}n}$
$\simeq D^E_{\frac{1}{2}n} \oplus D^{\frac{1}{2}n(n-1)}_{\frac{1}{2}n} \oplus D^{\frac{1}{2}n(n+1)-1}_{\frac{1}{2}n}$,

(3) $B_2^{\text{Spin}} \otimes B_2^{\text{Spin}}$
$\simeq B_2^E \oplus B_2^v \oplus B_2^{10}$,

(4) $B_3^{\text{Spin}} \otimes B_3^{\text{Spin}}$
$\simeq B_3^E \oplus B_3^v \oplus B_3^{21} \oplus B_3^{35}$,

(5) $A^n_{n-1} \otimes A^n_{n-1}$
$\simeq A^{\frac{1}{2}n(n-1)}_{n-1} \oplus A^{\frac{1}{2}n(n+1)}_{n-1}$,

[49] G. Frobenius and I. Schur, Preuss. Akad. Wiss. Berlin, 1906, p. 186.

[50] A. Malcev, Am. Math. Soc. Transl. No. 33 (1950; published in Russian, 1944).

(6) $A_{n-1}^{n} \otimes A_{n-1}^{n*}$
$\simeq A_{n-1}^{E} \oplus A_{n-1}^{n^2-1}$,

(7) $A_2^8 \otimes A_2^8$
$\simeq A_2^E \oplus A_2^{8a} \oplus A_2^{8s} \oplus A_2^{10} \oplus A_2^{10*} \oplus A_2^{27}$,

(8) $G_2^7 \otimes G_2^7$
$\simeq G_2^E \oplus G_2^7 \oplus G_2^{14} \oplus G_2^{27}$.

For further examples and for proofs see references 8 and 12 (Appendix), and 42. The first five decompositions correspond to tensors satisfying the familiar symmetry and trace properties. In the decomposition (6), $A_{n-1}^{n^2-1}$ can be interpreted as a representation on traceless tensors with components $v_i \otimes v_j^*$. The decomposition (7) is illustrated in Fig. 2. We shall not describe the decompositions (7) and (8) in detail, but we note the following. The two representations A_2^{8a} and A_2^{8s}, which are associated with antisymmetric and with symmetric tensors respectively, are equivalent. Consequently the direct sum $A_2^{8a} \oplus A_2^{8s}$ can also be decomposed into irreducible subspaces in other ways.

The three representations $D_4^v \otimes D_4^v$, $D_4^{\text{Spin}} \otimes D_4^{\text{Spin}}$ and $D_4^{(\text{Spin}')} \otimes D_4^{(\text{Spin}')}$ are analogous (in the sense explained in Sec. 4). They are not analogous to representations like $D_4^{\text{Spin}} \otimes D_4^{(\text{Spin}')}$, but we have no direct interest in the latter representations.

In the case of direct-product groups, we may proceed as follows. Let two representations of $\mathcal{G}' \times \mathcal{G}''$ be given: $D_i' \otimes D_i''$, where $i = 1, 2$, Then

$$(D_1' \otimes D_1'') \otimes (D_2' \otimes D_2'')$$
$$\simeq (D_1' \otimes D_2') \otimes (D_1'' \otimes D_2'')$$
$$\simeq (D_k' \oplus \cdots \oplus D_l') \otimes (D_m'' \oplus \cdots \oplus D_n'').$$

Here the decomposition of $D_1' \otimes D_2'$, or of $D_1'' \otimes D_2''$, is with reference to $\mathcal{G}'$, or to $\mathcal{G}''$, respectively.

For the case of nonconnected groups see Appendix C and reference 1.

A formula describing the decomposition in the general case involving semisimple Lie groups has been given by Brauer.[51]

11. *The adjoint representation.* The adjoint representation of a Lie algebra is defined as follows. For X and Y in the Lie algebra,

$$X \to \text{ad } X, \quad \text{where} \quad (\text{ad } X)Y = [X, Y].$$

The relation [ad X, ad Y] = ad $[X, Y]$ follows from the Jacobi identity. We remark that this representation is fundamental in the study of structure and representations of Lie groups.

[51] R. Brauer, Compt. Rend. **204**, 1784 (1957).

The degree of the adjoint representation equals the dimension of the Lie group, or of the Lie algebra. For B_n and D_n, the adjoint representation is equivalent to that on second rank skew-symmetric tensors; for A_{n-1}, the adjoint representation is equivalent to $A_{n-1}^{n^2-1}$, mentioned in the foregoing; an example is A_2^8.

The adjoint representation is always orthogonal, since the symmetric form $(X, Y) = \text{tr (ad } X \text{ ad } Y)$ is invariant under the action of the group.

12. *Weights and Lemma* 7.1. Let a semisimple Lie algebra be given, and let us select a maximal commuting set of elements, i.e. a Cartan subalgebra (Appendix A). These elements can be simultaneously diagonalized in a representation. The simultaneous eigenvectors are called the *weight vectors*, and the set of those eigenvalues, which are associated with a weight vector, is called a *weight*. The weights and the weight vectors completely describe a representation, and conversely, a representation determines uniquely the structure of the weights. Examples are shown in Figs. 1 and 2. For a fuller discussion, see references 3, 8, 12 (Appendix), and Racah's notes.[52]

Lemma 7.1 follows trivially from the following facts: (a) The set of operators $\{T_3, Y\}$ may be extended to a Cartan subalgebra. (b) An operator $\alpha T_3 + \beta Y$ will in general have all eigenvalues distinct except for Σ^0 and Λ, and consequently can be diagonal only with respect to the vectors p, n, $\Sigma^{+,-}$, $\Xi^{0,-}$ and $\Sigma^{0'}$, $\Sigma^{0''}$, where the last two vectors are linear combinations of Σ^0 and Λ.

We make one further remark. If to a certain representation D there corresponds a given weight diagram, then to the complex conjugate representation D^*, corresponds the diagram obtained by reflection in the origin. (See reference 50.) In particular, a representation is orthogonal or symplectic if and only if it is symmetric with respect to such reflections.

13. *Casimir and Racah operators.* Given any semisimple Lie algebra, there exists a second-order operator, the Casimir operator, which commutes with every element of the Lie algebra, and so provides a natural generalization of the operator $\mathbf{T}^2$ (or of $\mathbf{J}^2$). The Casimir operator can be defined as $\sum g_{ij} X^i X^j$, where g_{ij} is the tensor which is determined by the bilinear form $(X, Y) = \text{tr (ad } X \text{ ad } Y)$ of the adjoint representation.

If a semisimple Lie algebra is of rank n, then one

[52] G. Racah, "Group theory and spectroscopy," The Institute for Advanced Study, lecture notes, 1951. [Reprinted by CERN, 1961.]

can construct a total of n independent operators of orders 2 or higher, which commute with every element of the Lie algebra. A representation of the Lie algebra is then determined by the values which these operators take. These operators, except for the Casimir operator, were first introduced by Racah.[52,53]

APPENDIX C. NONCONNECTED GROUPS

We will now discuss the structure of nonconnected symmetry groups. Our approach to this subject is based on a theorem of A. H. Clifford. This theorem relates the representation of a group with the restriction to an invariant subgroup, and allows us to determine the possibilities for the nonconnected components of the identities of nonconnected symmetry groups. We remark, however, that relatively little is known about nonconnected Lie groups. For completeness we note a recent article of de Siebenthal.[54]

We first recall that in a topological group the neutral component is an invariant subgroup (Appendix A). The other components are cosets, not subgroups. Next, the following expresses the essentials of Clifford's theorem (reference 42, pp. 159–163).

Theorem. Let a group G and an irreducible representation $D(G)$ be given, and let $\tilde{G}$ be an invariant subgroup of G. Then the restriction $D(\tilde{G})$ [of $D(G)$ to $\tilde{G}$] reduces into irreducible representations $d_i(\tilde{G})$ of $\tilde{G}$ of equal degrees. Any two of these representations, say d_1 and d_2, are conjugate in the sense that

$$d_2(t) = d_1(sts^{-1}), \tag{C.1}$$

for all $t \in \tilde{G}$ and for a suitable $s \notin \tilde{G}$. The two members of this equation are to be interpreted as matrices, taken with reference to appropriate bases. Further, the number of irreducible components of $D(\tilde{G})$ cannot be greater than the index $G/\tilde{G}$ (if this index is finite).

The spaces which are associated with the representations $d_i(\tilde{G})$ are known as systems of imprimitivity.[55]

Equation (C.1) shows that the elements of $d_1(\tilde{G})$ and $d_2(\tilde{G})$ range over the same set of matrices, if the bases are appropriately chosen. The meaning of this equation can be best illustrated by an example. Consider the group generated by the elements

$$t_\phi = \begin{pmatrix} e^{i\phi} & \\ & e^{-i\phi} \end{pmatrix}, \qquad s = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}.$$

This group is equivalent to O_2^*. The elements t_ϕ form an invariant subgroup, and if

$$d_1(t_\phi) = e^{i\phi}, \qquad d_2(t_\phi) = e^{-i\phi},$$

then

$$d_2(t_\phi) = d_1(st_\phi s^{-1}).$$

Here the two representations d_1 and d_2 of (t_ϕ) are not equivalent, and this is ordinarily the case.

It follows from Clifford's theorem that $\dim(d_k)$ is a divisor of $\dim(D)$. However, if the group A_1^3 acting on the Σ's is to be included, then $\dim(d_k) \geq 3$. We see that we have these possibilities for a symmetry group G, having the component of the identity $\tilde{G}$ [we may identify here G with $D(G)$]:

(1) $\tilde{G}$ is irreducible;

(2) $\dim G = 6$ or 8, and G breaks into two blocks $d_1(\tilde{G})$ and $d_2(\tilde{G})$, each of degree 3 or 4, respectively;

(3) $\dim G = 4$, $\tilde{G}$ breaks into two blocks $d_1(\tilde{G})$ and $d_2(\tilde{G})$, each of degree 2, and G is a partial symmetry for N and Ξ (so that the Σ's are not involved).

Case (1) will not be discussed further. For cases (2) and (3) we may use groups from Tables I(a), (b) and II, and also the circle group. Lemma 4.2 is applicable here, but we have the additional restriction that the blocks $d_1(\tilde{G})$ and $d_2(\tilde{G})$ involve the same elements. We are therefore led to these alternatives:

(a) The elements in the two blocks $d_1(\tilde{G})$ and $d_2(\tilde{G})$ vary independently.

(b) An element of one block determines the element in the other block. The matrices in G are therefore of the form $\begin{pmatrix} \tau & \\ & f(\tau) \end{pmatrix}$, with the two sets $\{\tau\}$ and $\{f(\tau)\}$ identical. An examination of the linear groups of degree 4 or less now shows that we must have, up to equivalence,

$$f(\tau) = \tau \quad \text{or} \quad f(\tau) = \tau^*.$$

(c) The blocks $d_i(G)$ are Kronecker product groups, with (a) applicable to some of the factors, and (b) to others. For groups of degree 4 or less, the only Kronecker products are those in which one of the factors is a circle group, and also $A_1^2 \otimes (A_1')^2$. The last alternative corresponds to the group M.

We now see that all possibilities for the connected components of the identities of nonconnected groups are included in Table III.

In the remainder of this appendix we will consider only the three groups K, L, and M, defined in Table VI. We recall that

$$K \simeq SU_2 \otimes O_2^*.$$

Let us discuss now the classes to which these linear groups belong. We easily see that K is of class -1,

[53] G. Racah, Rend. Lincei 8, 108 (1950).
[54] J. de Siebenthal, Commun. Math. Helv. 31, 41 (1956–57).
[55] A. Speiser, *Theorie der Gruppen von endlicher Ordnung* (Springer-Verlag, Berlin, 1937), 2nd Ed., p. 194 ff.

i.e. is symplectic, since SU_2 and O_2^* are of class -1 and $+1$, respectively. It is also trivial that L is complex. With regard to M, we first observe that the connected component M_0 is orthogonal, and that $R_8 = (_E{}^E) \in SO_8$. We can show that M is orthogonal by finding a matrix q such that, simultaneously,

$$qM_0q^{-1} \subseteq SO_8 \quad \text{and} \quad qR_8q^{-1} \in SO_8.$$

But the structure of M_0 allows us to satisfy the first of these conditions by a matrix of the form $(^Q{}_Q)$, and the second condition then also is fulfilled. We summarize:

K is symplectic, L is complex, M is orthogonal.

Finally, let us consider the decomposition of Kronecker product representations. (We can assume here the underlying group to be isomorphic to the linear group in question.) Clifford's theorem applies also to the irreducible representations of the non-connected groups. In particular, the connected component of the identity may reduce into two blocks, or be irreducible. We can illustrate this easily for the group O_2^*:

$$O_2^* \otimes O_2^* = D^E \oplus O_2^* \oplus D^{ps},$$

where D^{ps} denotes the pseudoscalar representation. The connected component of the identity reduces into two blocks in the representation O_2^*, and is irreducible in D^E and D^{ps}.

From the foregoing decomposition we may obtain the decompositions of products of representations of K. We will not discuss the decompositions for L and M, but these two groups do not present any special difficulties.

APPENDIX D. PROOFS OF INCLUSION RELATIONS

We will now prove the relations of inclusion (and of noninclusion) which were asserted in Sec. 5. We will first establish that the groups, which we asserted to be symmetry groups, satisfy the perscribed requirements. Next we shall show that the remaining irreducible groups do not satisfy the requirements. Then we shall establish the inclusions among the symmetry groups. We emphasize once more that a knowledge of the classes of representations is basic for much of this appendix; however, these classes are little used in the proofs we present first.

We now want to establish that each of the groups in Table VII contains (I-spin) $\times$ (hypercharge gauge, perhaps modified) as a subgroup. We first recall that the two groups G_{HR} and G_{DA} fulfill this requirement,[19-21] and that they are defined as follows:

$$G_{HR} = \begin{pmatrix} A_1^2 \otimes (A_1')^2 & & \\ & A_1^3 & \\ & & A_1^E \end{pmatrix} \text{ w.r.t. } \begin{pmatrix} N \\ \Xi \\ \Sigma \\ \Lambda \end{pmatrix},$$

$$G_{DA} = \begin{pmatrix} A_1^2 e^{-i\phi} & & \\ & A_1^2 \otimes (A_1')^2 & \\ & & A_1^2 e^{i\phi} \end{pmatrix} \text{ w.r.t. } \begin{pmatrix} \Xi \\ \Sigma \\ \Lambda \\ N \end{pmatrix},$$

and moreover,

$$A_1^2 \otimes (A_1')^2 \simeq SO_4.$$

In a number of cases the required inclusion can be easily established by comparison with G_{HR} and G_{DA}, if the two latter symmetries are taken with reference to a subset of the baryons. We will now consider the groups by degrees. The inclusions for $A_n^{n+1} = SU_{n+1}$ are trivial, and these groups will not be mentioned.

Degree 4: (a) For K the assertion is trivial; (b) for B_2^{Spin} note that $K \simeq A_1^2 \otimes O_2^*$ is of class $(-1)(+1) = (-1)$, therefore $K \subseteq Sp_4 = B_2^{\text{Spin}}$; (c) for $A_1^2 \otimes (A_1')^2$, cf. the groups G_{HR} and G_{DA}.

Degree 5: Note that B_2^v contains SO_4 as a subgroup, and compare with G_{HR}.

Degree 6: (a) For L the assertion is trivial; (b) for $A_1^2 \otimes A_2^3$, we have as a subgroup

$$A_1^2 \otimes \begin{pmatrix} (A_1')^2 & \\ & e^{i\phi}E_2 \end{pmatrix} e^{-i\phi/3}.$$

The factor $e^{-i\phi/3}$ can be ignored, and compare with G_{DA}.

Degree 7: (a) For G_2^7 see Behrends *et al.*[8,17] (The inclusion also follows from the weight diagram, Fig. 1, which we do not establish.) (b) For B_3^v, compare with G_{HR}. Alternately, G_2^7 is of class $+1$, and therefore $G_2^7 \subseteq B_3^v$.

Degree 8: (a) The group A_2^8 has been discussed elsewhere,[8,13,14] but we shall indicate a simple proof. We first note that $A_2^3 \otimes A_2^{3*} \simeq A_2^8 \oplus A_2^E$. Now, A_2^3 includes as a subgroup $D = (A_1^2 e^{i\phi/3} \oplus e^{-2i\phi/3})$. We examine $D \otimes D^*$ and obtain the desired conclusion. (b) For M, compare with G_{HR} or with G_{DA}. (See also reference 1.) (c) We shall establish below that all of the groups of degree 8 listed in Table VII, except A_2^8 and B_3^{Spin}, contain M as a subgroup, and therefore are symmetry groups. The group B_3^{Spin} contains as a subgroup M_0, the connected component of M, and the same conclusion is valid.

The inclusions for reducible symmetries can be easily established by similar considerations.

Let us now prove noninclusions for those groups of Tables I(a), (b), and II which are not listed in Table VII. These proofs are based on two kinds of argument. The first consists of a direct examination of the diagonal elements of a group, and of showing in this way that the diagonal elements which we require are not contained in the group. This argument can be applied to A_2^6 and to $A_1^2 \otimes (A_1')^4$. The diagonal elements of A_2^6 can be found with the help of the relation $A_2^3 \otimes A_2^3 \simeq A_2^6 \oplus A_2^{3*}$.

The second type of argument depends on the fact that a representation of class -1 cannot decompose in such a way, that each of A_1^3 and A_1^E occurs an odd number of times. This argument is applicable to B_2^{Spin} for Σ and Λ, to C_3^6, to $A_1^2 \otimes (A_1')^2 \otimes (A_1'')^2$, and to C_4^8.

If one of the two groups, I-spin or modified hypercharge gauge, is not included, then for brevity we omit the proof of the inclusion or noninclusion of the other group.

We shall not examine in detail the nonconnected groups, but this would present no difficulties. However, we give an example of a group which contains as subgroups the I-spin group and a modified gauge group, but not both groups simultaneously. Indeed, consider the group, of degree 6, which splits into two independent blocks SU_3 and SU_3'. The I-spin group is contained if we arrange the baryons as N–Λ and Σ, and the modified hypercharge gauge group, if we arrange them as, e.g., p–Λ–Σ^+ and n–Σ^-–Σ^0.

Let us now turn to the inclusions among the symmetry groups. (An inclusion $D \subseteq D'$ is to be understood, of course, as the inclusion $D_0 \subseteq D_0'$ for *some* representations D_0 and D_0' equivalent to D and D' respectively.) We prove first the inclusions which are asserted in Table VIII, i.e. the inclusions among the irreducible symmetry groups. For degrees 3, 5, and 6, the assertions are trivial. For degrees 4 and 7, the assertions follow at once from our knowledge of the classes to which the representations belong. The groups of degree 8, however, require a more detailed examination.

We first use Dynkin's[12] analysis of the inclusions among simple irreducible linear groups. He shows that, for groups of class $+1$, the relation $D' \subseteq D'' \subseteq SO_n$ holds only in special cases, which he tabulates. His work shows that A_2^8 is not included in B_3^{Spin}. Next, the noninclusion of A_2^8, B_3^{Spin} and D_4^8 in $A_1^2 \otimes A_3^4$ follows from the fact that an inclusion would imply a Kronecker product decomposition of a simple group. (See also Dynkin,[12] Theorem 2.2.) On the other hand, the inclusion of $A_1^2 \otimes B_2^{\text{Spin}}$ is trivial. The noninclusion of $A_1^2 \otimes A_3^4$ (and of A_7^8) in D_4^8, and the inclusion of the other groups, follows by considering the classes of the representations. Let us show next that $M \subseteq A_1^2 \otimes B_2^{\text{Spin}}$. We define $B_2^{\text{Spin}} = Sp_4$ by the matrix $J = \left(\begin{smallmatrix} & J_2 \\ -J_2 & \end{smallmatrix}\right)$. The relation $A_1^2 = Sp_2$ shows that the connected component M_0 of M is a subgroup of $A_1^2 \otimes B_2^{\text{Spin}}$. For the discrete generator, we note that $R_8 = E \otimes \left(\begin{smallmatrix} & E \\ E & \end{smallmatrix}\right) \in A_1^2 \otimes B_2^{\text{Spin}}$, and this proves the inclusion of M. Finally, let us consider the group B_3^{Spin}. It is easy to show directly, by using generators, that B_3^{Spin} contains the connected component M_0 of M. Further, B_3^{Spin} does not contain $A_1^2 \otimes B_2^{\text{Spin}}$, since otherwise we would have a locally isomorphic representation of $A_1^2 \otimes B_2^{\text{Spin}}$ of degree 7, and of class $+1$, and this would be a contradiction. However, it seems difficult to determine whether the discrete generator R_8 can belong to B_3^{Spin} if a suitable choice for this (linear) group is made.

The inclusions which are asserted in Table X can be studied along similar lines. We make these observations: (a) A reducible group cannot contain an irreducible group as a subgroup; (b) The group $A_1^2 \otimes A_3^4$ cannot contain H_G as a subgroup, for reasons stated in the foregoing. This kind of argument applies also to a few other cases. (c) The inclusion of H_G in B_3^{Spin} follows from our knowledge or representations of G_2 of degrees 8 or less, or from the fact that $G_2^7 \subseteq B_3^v$. (See also reference 17.) The inclusions which relate the groups H_A, H_B, H_G, J_N, and J_Ξ can now be established easily.

There remain the groups $\bar{K}$, K_{DA}, G_{HR}, G_{DA}, and M_0. The noninclusions for $\bar{K}$ and K_{DA} follow from the fact that these groups contain one (and only one) symplectic block, the group K. These groups consequently cannot be contained in an orthogonal representation. Moreover, these groups are not contained in $A_1^2 \otimes A_3^4$, since they can be expressed by matrices $m_{2\times2} \otimes m_{4\times4}$ only if $\det(m_{4\times4}) = -1$.

Let us now consider the groups G_{HR}, G_{DA}, and M_0. We recall that $M_0 \subseteq B_3^{\text{Spin}}$. Further, G_{HR} is not contained in H_G nor in A_2^8. This may not seem obvious, but can be easily established by a comparison of the weight structures of the Lie algebras. The other inclusions for these groups are trivial.

With regard to the group K', we note that the matrices $\left(\begin{smallmatrix} e^{-i\phi} & \\ & e^{i\phi} \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} & -1 \\ 1 & \end{smallmatrix}\right) \in (A_1')^2$, and therefore $K' \subseteq A_1^2 \otimes (A_1')^2$. The assertions which we made in Table XI follow easily.

Quantum Field-Theory Model whose Truncated Vacuum Expectation Values Vanished beyond Some Order

O. W. GREENBERG*
Department of Physics and Astronomy, University of Maryland, College Park, Maryland
A. L. LICHT†
Department of Physics and Astronomy, University of Maryland, College Park, Maryland, and Naval Ordnance Laboratory, White Oak, Maryland
(Received 7 January 1963)

A local quantum field which has a complete asymptotic field of mass m, and whose truncated vacuum expectation values vanish beyond some order, can lead to no scattering or reactions.

INTRODUCTION AND PROOF OF THEOREM

THERE have been continuing efforts on the part of field theorists to find a model which satisfies the general requirements of local quantum field theory,[1] and leads to scattering and reactions. Up to now these efforts have not been crowned with success. However, some progress has been made, mainly in four directions: (a) models which show what can exist mathematically but do not lead to interaction; (b) models which lead to interaction, but fail to satisfy some of the general requirements; (c) models which shed light on the relations, if any, between different requirements; and (d) studies which show that certain simplifying assumptions do not allow interactions to occur. We will refrain from discussing and referring to the sizable and growing literature on this subject; such a discussion might be done more appropriately in a systematic review article.

The present article falls in category (d) above. We prove:

Theorem. A local quantum field which has a complete asymptotic field of mass m, and whose truncated vacuum expectation values vanish beyond some order N, can lead to no scattering or reactions.

We first give our proof, and then remind the reader of the definition and heuristic significance of truncated vacuum expectation values. For simplicity, we consider the case of a single neutral scalar field; however, the result is true more generally.

Proof of theorem. Glaser, Lehmann, and Zimmermann[2] have shown that if the in field of a local quantum field theory is complete, then the Heisenberg field has the expansion

$$A(x) = A_{\text{in}}(x) + \sum_{n=2}^{\infty} \frac{1}{n!} \times \int dx_1 \cdots dx_n K_{x_1} \cdots K_{x_n} r(x; x_1, \cdots x_n) \times :A_{\text{in}}(x_1) \cdots A_{\text{in}}(x_n): \tag{1}$$

where $\kappa_x = -(\Box_x + m^2)$, and

$$r(x; x_1, \cdots x_n) = (-i)^n \sum_{\substack{\text{perm. of} \\ x_1, \cdots x_n}} \theta(x - x_1) \cdots \times \theta(x_{n-1} - x_n) \langle [\cdots [A(x), A(x_1)], \cdots A(x_n)] \rangle_0$$

is the retarded multiple-commutator function.

We use a result due to Araki, Ruelle, and Steinmann[3] which states that the r functions have contributions only from their truncated parts, i.e.

$$r(x; x_1, \cdots x_n) = (-i)^n \sum_{\substack{\text{perm. of} \\ x_1, \cdots x_n}} \theta(x - x_1) \cdots \times \theta(x_{n-1} - x_n) \langle [\cdots [A(x), A(x_1)], \cdots A(x_n)] \rangle_T ,$$

where $\langle\ \rangle_T$ stands for truncated vacuum expectation value.

On the basis of this result, the vanishing of the truncated vacuum expectation values beyond order N implies that the expansion of the Heisenberg field [Eq. (1)] terminates at $n = N - 1$. Then $A_{\text{out}}(x)$ has the following finite degree expansion in $A_{\text{in}}(x)$:

$$A_{\text{out}}(x) = A_{\text{in}}(x) + \sum_{n=2}^{N-1} \frac{1}{n!} \int dx' \, dx_1 \cdots dx_n \, \Delta(x - x') K_{x'} K_{x_1} \cdots \times K_{x_n} r(x'; x_1, \cdots x_n) :A_{\text{in}}(x_1) \cdots A_{\text{in}}(x_n):,$$

* Partially supported by a grant from the General Research Board of the University of Maryland.
† National Science Foundation Predoctoral Fellow, 1960–1961.
[1] A. S. Wightman, Phys. Rev. **101**, 860 (1956); "Problèmes mathématiques de la théorie quantique des champs," University of Paris lecture notes, 1957 (unpublished).
[2] V. Glaser, H. Lehmann, and W. Zimmermann, Nuovo Cimento **6**, 1122 (1957).
[3] H. Araki, J. Math. Phys. **2**, 163 (1961); D. Ruelle, Nuovo Cimento **19**, 356 (1961); O. Steinmann, Helv. Phys. Acta **33**, 257 (1960).

and according to a folk lemma[4] $S = \mathbb{1}$, $A_{out}(x) = A_{in}(x)$, which means that there is no scattering or reactions.

DEFINITION AND HEURISTIC SIGNIFICANCE OF TRUNCATION

Truncated vacuum expectation values are defined recursively by the implicit equations

$$\langle A(1) \cdots A(n)\rangle_0 = \langle A(1) \cdots A(n)\rangle_T + \sum \langle A(1) \cdots A(j)\rangle_T \cdots \langle A(l) \cdots A(n)\rangle_T,$$

where the sum runs over all divisions of $1, \cdots n$ into two or more groups, in such a way that the numbers increase to the right. Divisions, such as $(1, 3, \cdots)(2, 4, \cdots)$, in which the numbers in different groups interleave, must be included. The definition of truncation has been chosen so that the truncated vacuum expectation values differ from the ordinary ones by having had the contributions from vacuum intermediate states subtracted in a symmetric way. Truncated vacuum expectation values correspond to the connected graphs in perturbation theory. Thus the theorem under discussion in this paper shows that a theory with scattering or reactions must have nonvanishing connected graphs with arbitrarily many external lines.

[4] O. W. Greenberg, J. Math. Phys. 3, 31 (1962).

ACKNOWLEDGMENTS

We thank Professor John S. Toll for several helpful conversations. It is a pleasure to acknowledge the support of the General Research Board of the University of Maryland (O.W.G), and of the National Science Foundation (A.L.L).

Note on Uniqueness of Canonical Commutation Relations*

R. ARNOWITT AND S. DESER
Department of Physics, Brandeis University, Waltham, Massachusetts
(Received 28 May 1962)

It has been pointed out by Wigner that the consistency requirement between the Lagrange and Heisenberg equations of motion does not uniquely determine the canonical commutation relations, at least for one-dimensional systems. It is shown here that this ambiguity does not arise in local field theory whose basic equal-time commutators commute with the translation operator.

I. INTRODUCTION

IN the course of investigating[1] the nature of the axioms underlying the Schwinger Action Principle,[2] it was found that the equal time (anti) commutator between a field and its conjugate momentum, $[\phi(\mathbf{r}), \pi(\mathbf{r}')]_-$, could only be reduced to the form

$$[\phi(\mathbf{r}), \pi(\mathbf{r}')]_- = i\,\delta^3(\mathbf{r}-\mathbf{r}')[1+F], \qquad (1)$$

where F is not necessarily a c number. This reduction was performed on the basis of quite general assumptions, as described in reference 1.

In attempting to see whether F could be forced to zero purely on consistency grounds (rather than by an additional postulate), the freedom of performing arbitrary unitary transformations in the theory was employed. Thus, consider the change generated by the infinitesimal operator $U \simeq 1 + iG$ ($G^+ = G$), $G = \int d^3r[f_1(\phi, \pi)\,\delta\phi + f_2(\phi, \pi)\,\delta\pi]$. The fundamental commutation relations then enable one to evaluate the change ΔR of any function $R(\phi, \pi)$, according to the definition $\Delta R \equiv -i[G, R]$. In general, ΔR will involve the operator F. For the special case in which R is taken to be the Hamiltonian H of the system, this change may be explicitly evaluated in another way using the Heisenberg equations of motion. Thus, in

$$\Delta H = i[H, G] = dG/dt, \qquad (2)$$

the time derivative dG/dt depends on $\dot{\phi}$ and $\dot{\pi}$. The latter derivatives may be evaluated by using the (by assumption) identical Lagrange equations in which F does not enter. The consistency of the two evaluations provides a stringent requirement on F. Indeed, if F is to be a c number, the above condition requires it to vanish (as will be shown below); however, q-number F's could not be so eliminated. An example of this situation had, in fact, been given some time ago by Wigner[3] for the one-dimensional harmonic oscillator. Here, consistency between commutation relations and equations of motion still allowed F to be any one of a one-parameter class of operators,

$$F = (2E_0/\hbar\omega - 1)\exp[i\pi(H - E_0)/\hbar\omega],$$

(E_0 arbitrary[4]) anticommuting with both p and q. However, we show here that, for local field theory whose basic equal-time commutators commute with the translation operator, F is necessarily zero in order that the Lagrange and Heisenberg equations agree.

II. DERIVATION

Consider the general field theoretical system in first-order form.[1] For convenience, the field variables (which may be taken to be Hermitian) may be arranged in a column symbol array with entries $\chi_a(x)$, $(a = 1 \cdots N)$. In the first-order description, the χ_a include both the field amplitudes and their conjugate momenta. Thus, the equal-time commutator (or anticommutator) $[\chi_a(\mathbf{r}), \chi_b(\mathbf{r}')]_-$ includes all the usual commutation (or anticommutation) relations. In general, however, not all field variables are independent, some being functions of the rest if some of the Lagrange equations are constraint equations. The independent variables in the general case will be denoted by ϕ_a, $a = 1 \cdots M$, $M \leq N$. The commutators $[\phi_a(\mathbf{r}), \phi_b(\mathbf{r}')]_-$ are then a complete set of relations; any commutators involving the dependent variables are determined by them. In general, one may write

$$[\phi_a(\mathbf{r}), \phi_b(\mathbf{r}')]_- = if_{ab}\,\delta^3(\mathbf{r}-\mathbf{r}'), \qquad (3)$$

* Supported in part by the Air Force Office of Scientific Research and the National Science Foundation.

[1] R. Arnowitt and S. Deser, J. Math. Phys., 3, 637 (1962).

[2] J. Schwinger, Phys. Rev. 82, 914 (1951); and 91, 713 (1953).

[3] E. P. Wigner, Phys. Rev. 77, 711 (1950).

[4] Note here E_0 represents the ground-state energy, which need not be $\frac{1}{2}\hbar\omega$, with commutation relations of the type of Eq. (1). However, the eigenspectrum of $H-E^0$ is still $n\hbar\omega$ and so F is correctly Hermitian.

where f_{ab} may be an operator. The term $\delta^3(\mathbf{r} - \mathbf{r}')$ is a consequence of the locality assumption.[5] It is also assumed that f_{ab} is independent of the center of mass coordinate $\mathbf{r} + \mathbf{r}'$. This assumption of translational invariance[6] is equivalent to requiring that f_{ab} and the translation generator of the theory U commute, so that the commutators at all space points are numerically equal. This form of the translational invariance requirement is stronger than one in which the commutators are just required to be form invariant under translation. (The latter condition could be satisfied by an operator f_{ab}, depending on $\mathbf{r} + \mathbf{r}'$ and translated by U.) For (anti) commutators, the operators f_{ab} must be (skew-) Hermitian in the Hilbert space, and symmetric (antisymmetric) in the matrix indices ab. To prove that f_{ab} is, in fact, a c-number matrix, consider the double "commutator"

$$K \equiv [\phi_a(\mathbf{r}), [\phi_b(\mathbf{r}'), \phi_c(\mathbf{r}'')]_\pm], \tag{4}$$

where the $\pm$ signs refer to the Fermi (Bose) cases. The quantity K vanishes for $r \neq r', r''$ by the locality condition. However, from Eq. (3) one sees that

$$K = [\phi_a(\mathbf{r}), f_{bc}]i\ \delta^3(\mathbf{r}' - \mathbf{r}''). \tag{5}$$

Choosing $\mathbf{r}' = \mathbf{r}'' = \mathbf{a}$ and $\mathbf{r} \neq a$, one sees that f_{bc} commutes with $\phi_a(\mathbf{r})$, except *perhaps* at $\mathbf{r} = a$. However, repeating the argument with $\mathbf{r}' = \mathbf{r}'' = \mathbf{b}$, $\mathbf{r} \neq \mathbf{b}$ and $\mathbf{b} \neq \mathbf{a}$, implies that f_{bc} commutes with $\phi_a(\mathbf{r})$ at $\mathbf{r} = \mathbf{a}$ also, i.e., throughout all three-space. Since $\phi_a(\mathbf{r})$ represents all the independent variables of the theory, this implies that f_{bc} is a c number.[7]

At this stage, then, f_{ab} is a matrix in (ab) space with c-number coefficients. To show that these coefficients are indeed the conventional ones, we compare the Lagrange and Heisenberg field equations, as discussed in Sec. I. Consider the commutator $[H[\phi], G_\phi]$ between the Hamiltonian $H = \int \mathcal{H}(\phi)\ d^3r$ (which is again only a function of the ϕ_a) and the Hermitian operator $G_\phi = \int \phi_a C_{ab}\ \bar{\delta}\phi_b\ d^3r$. Here C_{ab} is any nonsingular (anti) symmetric matrix and $\bar{\delta}\phi_b$ is an arbitrary infinitesimal quantity (anti) commuting with the ϕ_a themselves. The operator G_ϕ may be viewed as the generator of a unitary transformation $U \simeq 1 + iG_\phi$ which sends ϕ_a into $\phi_a + i/2(fC)_{ab}\ \bar{\delta}\phi_b$. That is,

$$U^{-1}\phi_a U \simeq \phi_a - i[G_\phi, \phi_a] = \phi_a + (fC)_{ab}\ \bar{\delta}\phi_b, \tag{6}$$

upon using Eq. (3) to evaluate the commutator. Therefore,

$$U^{-1}H[\phi]U \simeq H[\phi + (fC)\ \bar{\delta}\phi]. \tag{7}$$

To first order, then, Eq. (7) reads

$$-i[G_\phi, H] = H[\phi + fC\ \bar{\delta}\phi] - H[\phi] = \int d^3r(\delta H/\delta\phi)fC\ \bar{\delta}\phi. \tag{8}$$

On the other hand, using the fact that H generates the time motion, we have[8]

$$-i[G_\phi, H] = -i\int [\phi, H]C\ \bar{\delta}\phi = \int d^3r\dot{\phi}C\ \bar{\delta}\phi. \tag{9}$$

Since the Lagrange equation reads, in first-order form,[9]

$$\dot{\phi} = (\delta H/\delta\phi)A^0, \tag{10}$$

where A^0 is a particular nonsingular matrix characteristic of the spin of the system, Eqs. (8) and (10) together imply that $f = A^0$ (since C has an inverse). This value of f is, in fact, the conventional one. For scalar theory, for example,

$$\phi_a = (\phi, \pi);$$

$$A^0 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

so that

$$\begin{aligned} [\phi(\mathbf{r}), \phi(\mathbf{r}')] &= 0 = [\pi(\mathbf{r}), \pi(\mathbf{r}')], \\ [\phi(\mathbf{r}), \pi(\mathbf{r}')] &= i\ \delta^3(\mathbf{r} - \mathbf{r}'). \end{aligned} \tag{11}$$

In particle systems, the above type of derivation does not determine f uniquely, since the locality and translational invariance conditions are not, in general, available. Thus in field theory these conditions imply that the f_{ab} for two different "degrees of freedom," represented by the fields at two different

[5] The derivation that f is a c number does not actually depend on the presence of $\delta^3(\mathbf{r} - \mathbf{r}')$ on the r.h.s. of Eq. (3), but only on the vanishing of the commutator when $\mathbf{r} \neq \mathbf{r}'$. However, unless one has the δ^3 function, the Lagrange and Heisenberg equations will not be identical.

[6] S. Schweber, Phys. Rev. **78**, 613 (1950).

[7] Our definition of completeness of the set of field operators is different from one that has been used in the axiomatic approach [cf. A. L. Licht and J. S. Toll, Nuovo Cimento **21**, 346 (1961)]. There, a set of fields is called complete if any operator commuting with them at *all* space–time points is a c number. In text, a set is complete if an operator commuting with the set on a *spacelike* surface is a c number. From the point of view of formal local field theory, such complete sets should exist (if the theory exists). They represent just the minimal Cauchy data required by the Lagrange equations to uniquely specify the future motion of the system.

[8] Note (as shown in the Appendix of reference 1) that $\bar{\delta}\phi_a$ must commute with H even if it anticommutes with some of the ϕ_b (which is the case for Fermi fields). Briefly, $\bar{\delta}\phi_a$ would fail to commute with an odd power of Fermi fields in H, but if such terms existed they would also give rise to nonlocal parts in the Heisenberg equations of motion, in contradiction with the locality assumption.

[9] For the independent field variables ϕ_a, the Lagrange equations in first-order form take on the simple Hamiltonian structure of Eq. (10) (see reference 1), where the variational derivative in Eq. (10) is identical to the one defined in Eq. (8).

space points, is the same (i.e., f_{ab} is independent of $\mathbf{r}$). In ordinary quantum mechanics, no simple requirement forcing f to be the same for all the (discrete) degrees of freedom is available.[10] Of course, to the extent that ordinary quantum mechanics can be viewed as the limit of local field theory, our results also apply to it.

Finally, it should be noted that the derivation used here did not utilize the spin-statistics connection. From the result $f = A^0$, the connection then follows as a consequence of the symmetry properties of A^0 and certain assumptions on TCP invariance, as in reference 1.

[10] Naturally, if one postulates f to be the same for each degree of freedom in the particle case, the canonical results then follow, *except* in the one-dimensional case (e.g., Wigner's example). The latter system is anomalous since the derivation in text clearly requires more than two independent operators in the complete set [cf. Eq. (4)].

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 5 MAY 1963

High-Energy Behavior of a Certain Class of Scattering Amplitudes in Perturbation Theory

N. H. FUCHS*
Department of Physics and Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received 7 January 1963)

The high-energy behavior for a certain class of Feynman diagrams is determined. The results obtained indicate that Regge-type behavior of scattering amplitude may be associated with the existence of M-particle bound states for arbitrary M.

I. INTRODUCTION

IT has recently been proposed by Gell-Mann and Goldberger[1] that conventional field theory may predict high-energy behavior of scattering amplitudes in agreement with the existence of Regge poles.[2] The work of Lee and Sawyer[3] on the Bethe–Salpeter equation in the ladder approximation, and that of Federbush and Grisaru, and Polkinghorne[4] on high-energy behavior in perturbation theory lends credence to this proposal.

In the present work we investigate the high-energy behavior of a class of scattering amplitudes using methods of field theory developed by FG. The relationship between our results in perturbation theory and the concepts of Regge pology is discussed.

* National Science Foundation Predoctoral Fellow.
[1] M. Gell-Mann and M. L. Goldberger, Phys. Rev. Letters **9**, 275 (1962).
[2] A. Bottino, A. M. Longoni, and T. Regge, Nuovo Cimento **23**, 954 (1962).
[3] B. W. Lee and R. E. Sawyer, Phys. Rev., **127**, 2266 (1962). (Hereafter denoted by LS).
[4] P. G. Federbush and M. T. Grisaru, to be published. (Hereafter denoted by FG); J. C. Polkinghorne, J. Math. Phys. **4**, 503 (1963).

II. ASYMPTOTIC BEHAVIOR

We consider the diagram of Fig. 1 and calculate the behavior of the corresponding amplitude for large s and fixed t, where we define:

$$s = (K_{M0} - K_{M,N+1})^2, \quad t = (K_{10} + K_{M0})^2, \tag{1}$$
$$pq = \mathbf{pq} - p_0 q_0.$$

We take all masses equal to m, and the coupling in the Lagrangian is of the form $g_{(3)}\phi^3 + g_{(4)}\phi^4$.

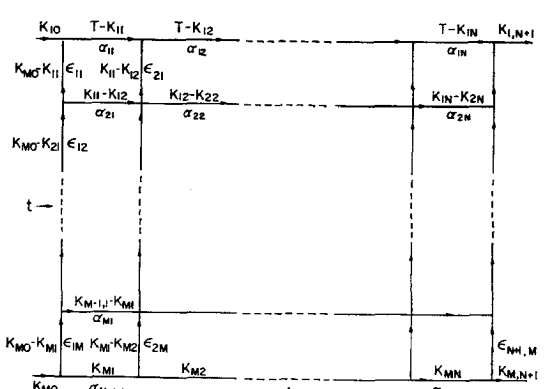


FIG. 1. $T = K_{10} + K_{M0} = K_{M,N+1} + K_{1,N+1}$.

space points, is the same (i.e., f_{ab} is independent of $\mathbf{r}$). In ordinary quantum mechanics, no simple requirement forcing f to be the same for all the (discrete) degrees of freedom is available.[10] Of course, to the extent that ordinary quantum mechanics can be viewed as the limit of local field theory, our results also apply to it.

Finally, it should be noted that the derivation used here did not utilize the spin-statistics connection. From the result $f = A^0$, the connection then follows as a consequence of the symmetry properties of A^0 and certain assumptions on TCP invariance, as in reference 1.

[10] Naturally, if one postulates f to be the same for each degree of freedom in the particle case, the canonical results then follow, *except* in the one-dimensional case (e.g., Wigner's example). The latter system is anomalous since the derivation in text clearly requires more than two independent operators in the complete set [cf. Eq. (4)].

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 5 MAY 1963

High-Energy Behavior of a Certain Class of Scattering Amplitudes in Perturbation Theory

N. H. FUCHS*
Department of Physics and Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received 7 January 1963)

The high-energy behavior for a certain class of Feynman diagrams is determined. The results obtained indicate that Regge-type behavior of scattering amplitude may be associated with the existence of M-particle bound states for arbitrary M.

I. INTRODUCTION

IT has recently been proposed by Gell-Mann and Goldberger[1] that conventional field theory may predict high-energy behavior of scattering amplitudes in agreement with the existence of Regge poles.[2] The work of Lee and Sawyer[3] on the Bethe–Salpeter equation in the ladder approximation, and that of Federbush and Grisaru, and Polkinghorne[4] on high-energy behavior in perturbation theory lends credence to this proposal.

In the present work we investigate the high-energy behavior of a class of scattering amplitudes using methods of field theory developed by FG. The relationship between our results in perturbation theory and the concepts of Regge pology is discussed.

* National Science Foundation Predoctoral Fellow.
[1] M. Gell-Mann and M. L. Goldberger, Phys. Rev. Letters **9**, 275 (1962).
[2] A. Bottino, A. M. Longoni, and T. Regge, Nuovo Cimento **23**, 954 (1962).
[3] B. W. Lee and R. E. Sawyer, Phys. Rev., **127**, 2266 (1962). (Hereafter denoted by LS).
[4] P. G. Federbush and M. T. Grisaru, to be published. (Hereafter denoted by FG); J. C. Polkinghorne, J. Math. Phys. **4**, 503 (1963).

II. ASYMPTOTIC BEHAVIOR

We consider the diagram of Fig. 1 and calculate the behavior of the corresponding amplitude for large s and fixed t, where we define:

$$s = (K_{M0} - K_{M,N+1})^2, \quad t = (K_{10} + K_{M0})^2, \tag{1}$$
$$pq = \mathbf{pq} - p_0 q_0.$$

We take all masses equal to m, and the coupling in the Lagrangian is of the form $g_{(3)}\phi^3 + g_{(4)}\phi^4$.

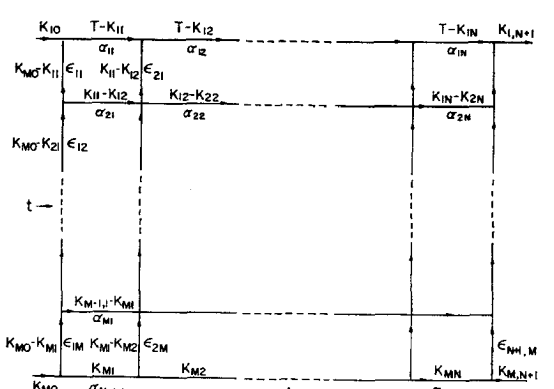


FIG. 1. $T = K_{10} + K_{M0} = K_{M,N+1} + K_{1,N+1}$.

The scattering amplitude F is defined by

$$S = 1 + i(2\pi)^4\delta(K_{10} + K_{M0} - K_{M,N+1} - K_{1,N+1}) \times [(2\pi)^{-\frac{1}{2}}]^4(16E_1E_2E_3E_4)^{-\frac{1}{2}}F. \quad (2)$$

In Fig. 1, the Greek letters in the diagram associated with each internal line correspond to the numbering of the Feynman parameters. The Feynman amplitude F_N^M is easily seen to be

$$F_N^M = g_{(3)}^{2(N+M)}g_{(4)}^{(M-1)(N-1)}\left[\frac{\pi^2}{(2\pi)^4}\right]^{NM}(-1)^{M+1}(N+M-1)! \times \int_0^1 \cdots \int_0^1 \frac{[\mathrm{Disc}_N^M]^{N+M-2}\,\delta[1 - \sum_{i,j}(\alpha_{ij} + \epsilon_{ij})]\,d\alpha_{ij}\,d\epsilon_{ij}}{(D_N^M)^{N+M}}. \quad (3)$$

All the information needed is contained in the matrix displayed in Fig. 2. The quantity D_N^M is the determinant of the matrix and may be written in the form

$$D_N^M = A_N^M s + B_N^M t + C_N^M m^2, \quad (4)$$

where A_N^M, B_N^M, and C_N^M are homogeneous polynomials in the Feynman parameters. The quantity Disc_N^M is the determinant of the matrix with all external momenta set equal to zero.[5]

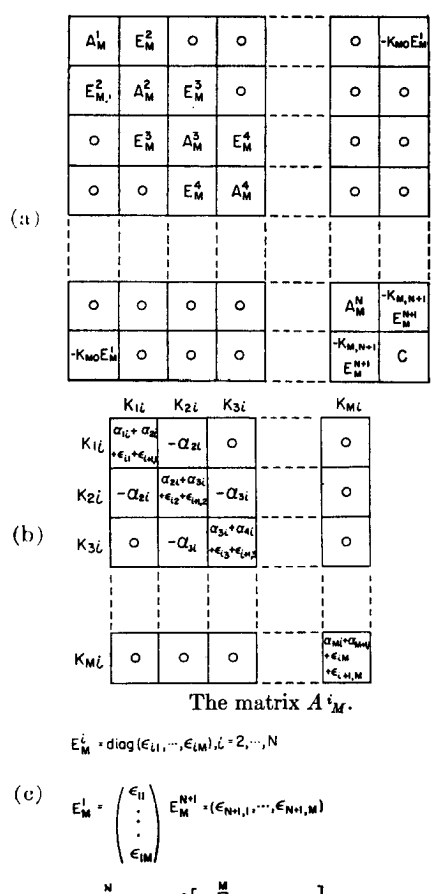


FIG. 2.

We now perform a change of variables:

$$\epsilon_{ij} = r_{ij}\kappa_i \quad 1 \le i \le N+1; \quad 1 \le j \le M, \qquad \sum_{j=1}^{M} r_{ij} = 1 \quad 1 \le i \le N+1. \quad (5)$$

The Jacobian of this transformation is

$$\frac{\partial(\epsilon_{ij})}{\partial(r_{ij},\,\kappa_i)} = \prod_{i=1}^{N+1}\kappa_i^{M-1}. \quad (6)$$

Now, according to Federbush and Grisaru, the high-energy behavior of the scattering amplitude F_N^M is determined by the contribution to the integral (3) of points in $\alpha_{ij}r_{ij}\kappa_i$ space for small κ_i. Hence we set the κ_i equal to zero everywhere except in the coefficient of s. We now require the formula

$$\int_0^1 d\alpha_1 \cdots \int_0^1 d\alpha_{N+1}\,\frac{(\alpha_1 \cdots \alpha_{N+1})^{p-1}}{(\alpha_1 \cdots \alpha_{N+1}s + C)^{q+1}} \underset{s\to\infty}{\to} \frac{(p-1)!\,(q-p)!\,(\ln s)^N}{q!\,C^{q-p+1}N!\,s^p}, \quad (7)$$

which is a generalization of a formula given in the Appendix II to FG. After performing the integration over the κ_i using (7), we find

$$F_N^M \to g_{(3)}^{2(N+M)}g_{(4)}^{(N-1)(M-1)}\left[\frac{\pi^2}{(2\pi)^4}\right]^{NM}(-1)^{M+1} \times \int_0^1 \cdots \int_0^1 d\alpha_{ij}\,dr_{ij}\,\delta(1 - \sum_{i,j}\alpha_{ij}) \times \frac{\prod_{i=1}^{N+1}\delta\left(1 - \sum_{j=1}^{M} r_{ij}\right)(\mathrm{Disc}_N^M)^{N+M-2}}{(\tilde{A}_N^M)^M(B_N^M t + C_N^M m^2)^N} \times (M-1)!\,(N-1)!\,\frac{(\ln s)^N}{N!\,s^M}, \quad (8)$$

where we have introduced $\tilde{A}_N^M$ by

$$A_N^M = \tilde{A}_N^M \times \prod_{i=1}^{N+1}\kappa_i. \quad (9)$$

The quantities B_N^M and C_N^M may be expressed in

[5] J. S. R. Chisholm, Proc. Cambridge Phil. Soc. **48**, 300 (1952).

terms of the determinants of the $M \times M$ matrices $\mathbf{A}_M^i$ of Fig. 2(b) by:

$$C_N^M = \prod_{i=1}^{N} |\mathbf{A}_N^i|, \tag{10}$$

$$B_N^M = C_N^M \times \sum_{j=1}^{N} \left(\prod_{i=1}^{M+1} \alpha_{ij} / |\mathbf{A}_N^j| \right) \tag{11}$$

$$= C_N^M \tilde{B}_N^M.$$

We now turn to the problem of finding a convenient expression for A_N^M. To this end we return to the matrix of Fig. 2(a) and note that we are considering the off-diagonal blocks $\mathbf{E}_M^i$ as having small entries. All $\mathbf{A}_M^i$ and $\mathbf{E}_M^i$ are $M \times M$ matrices and there are no other nonzero entries in $\mathbf{C}$, the submatrix obtained by deleting the last row and column of the matrix of Fig. 2(a). Now let $C_{\alpha\beta}$ be the cofactor of the α, β element of the $NM \times NM$ matrix $\mathbf{C}$. Then, expanding D_N^M in terms of its last row and column,

$$D_N^M = cC_N^M - \sum_{\alpha,\beta} C_{\alpha,\beta} (\mathbf{E}_M^1)_\alpha (\mathbf{E}_M^{N+1})_\beta, \tag{12}$$

where

$$c = \sum_{i=1}^{N} \alpha_{1i} t + m^2 \left[1 - \sum_{j=1}^{M} (\epsilon_{1j} + \epsilon_{N+1,j}) \right]. \tag{13}$$

The quantity A_N^M is given by the terms in the above sum for $1 \leq \alpha \leq M$, $(N-1) \leq \beta \leq NM$, to lowest order in the ϵ_{ij} by

$$A_N^M = \sum_{\alpha,\beta} (\mathbf{E}_M^1 \mathbf{A}_M'^1 \mathbf{E}_M^2 \mathbf{A}_M'^2 \cdots \mathbf{E}_M^N \mathbf{A}_M'^N \mathbf{E}_M^{N+1})_{\alpha\beta}, \tag{14}$$

where $\mathbf{A}_M'^i$ is the adjoint of the matrix $\mathbf{A}_M^i$.

Now we make the change of variables:

$$\alpha_{ij} = s_{ij} \lambda_j \quad 1 \leq j \leq N; \quad 1 \leq i \leq M+1,$$
$$\sum_{i=1}^{M+1} s_{ij} = 1, \quad 1 \leq j \leq N, \tag{15}$$

so that

$$\frac{\partial(\alpha_{ij})}{\partial(s_{ij}, \lambda_j)} = \prod_{j=1}^{N} \lambda_j^M, \tag{16}$$

and we make the definitions

$$\mathbf{E}_M^i = \kappa_i \mathbf{R}_M^i \qquad 1 \leq i \leq N+1,$$
$$\mathbf{A}_M^i = \lambda_i^{M-1} \mathbf{S}_M^i, \quad \mathbf{A}_M'^i = \lambda_i^{M-1} \mathbf{S}_M'^i \quad 1 \leq j \leq N,$$
$$\tilde{A}_N^M = \hat{A}_N^M \prod_{j=1}^{N} \lambda_j^{N-1}, \tag{17}$$
$$C_N^M = \tilde{C}_N^M \prod_{j=1}^{N} \lambda_j^M.$$

Therefore, we now have

$$F_N^M \to g_{(3)}^{2(N+M)} g_{(4)}^{(N-1)(M-1)} (-1)^{M+1} \left[\frac{\pi^2}{(2\pi)^4} \right]^{NM}$$
$$\times \int_0^1 \cdots \int_0^1 dr_{ij}\, ds_{ij} \prod_{j=1}^{N} \delta\left(1 - \sum_{i=1}^{M+1} s_{ij}\right)$$
$$\times \frac{(\ln s)^N}{N!\, s^M} \prod_{i=1}^{N+1} \delta\left(1 - \sum_{j=1}^{M} r_{ij}\right)$$
$$\times \frac{(\tilde{C}_N^M)^{M-2}}{(\hat{A}_N^M)^M} (M-1)!\,(N-1)!$$
$$\times \int_0^1 d\lambda_1 \cdots \int_0^1 d\lambda_N \frac{\delta\left(1 - \sum_{j=1}^{N} \lambda_j\right)}{(\tilde{B}_N^M t + m^2)^N}. \tag{18}$$

But from Eqs. (11) and (15),

$$\tilde{B}_N^M = \sum_{j=1}^{N} \left(\frac{1}{|\mathbf{S}_M^j|} \prod_{i=1}^{M+1} s_{ij} \lambda_j \right). \tag{19}$$

We may therefore apply the Feynman identity to do the λ_i integrals:

$$(N-1)! \int_0^1 d\lambda_1 \cdots \int_0^1 d\lambda_N \frac{\delta(1 - \sum_j \lambda_j)}{(\tilde{B}_N^M t + m^2)^N}$$
$$= \prod_{j=1}^{N} \left(\frac{1}{|\mathbf{S}_M^j|} \prod_{i=1}^{M+1} s_{ij} t + m^2 \right)^{-1}. \tag{20}$$

We now have

$$F_N^M \to g_{(3)}^{2(N+M)} g_{(4)}^{(N-1)(M-1)} \left[\frac{\pi^2}{(2\pi)^4} \right]^{NM} (-1)^{M+1}$$
$$\times \int_0^1 \cdots \int_0^1 ds_{ij} \prod_{j=1}^{N} |\mathbf{S}_M^j|^{M-2}$$
$$\times \prod_{j=1}^{N} \delta\left(1 - \sum_{i=1}^{M+1} s_{ij}\right)$$
$$\times \left[\prod_{j=1}^{N} \left(\frac{1}{|\mathbf{S}_M^j|} \prod_{i=1}^{M+1} s_{ij} t + m^2 \right) \right]^{-1} (M-1)!$$
$$\times \int_0^1 \cdots \int_0^1 dr_{ij} \frac{\prod_{i=1}^{N+1} \delta\left(1 - \sum_{j=1}^{M} r_{ij}\right)}{(\hat{A}_N^M)^M}. \tag{21}$$

III. FACTORING OF THE t DEPENDENCE

We shall now show that the integral over the r_{ij} may be factored into the form

$$\prod_{j=1}^{N} f_M(s_{1j}, \cdots, s_{M+1,j}).$$

Let

$$I_N = (M-1)! \int_0^1 \cdots \int_0^1 dr_{ij} \times \frac{\prod_{i=1}^{N+1} \delta\left(1 - \sum_{j=1}^{M} r_{ij}\right)}{\left[\sum_{\alpha\beta\gamma\delta} (\mathbf{R}_M^1)_{\alpha\beta} (\mathbf{G}^N)_{\beta\gamma} (\mathbf{R}_M^{N+1})_{\gamma\delta}\right]^M}, \tag{22}$$

where

$$\mathbf{G}^N = \mathbf{G}^{N-1} \mathbf{R}^N \mathbf{S}_M'^N, \tag{23}$$

and $\mathbf{G}^0$ is the unit $M \times M$ matrix.

By performing the $r_{N+1,j}$ integrals for $1 \leq j \leq M$, using the Feynman formula, we find

$$I_N = \int_0^1 \cdots \int_0^1 \frac{\prod_{i=1}^{N} \prod_{j=1}^{M} dr_{ij} \prod_{k=1}^{N} \delta\left(1 - \sum_{l=1}^{M} r_{kl}\right)}{\prod_{j=1}^{M} \sum_{i=1}^{M} r_{1i} G_{ij}^N}. \tag{24}$$

By Eq. (23),

$$\sum_{i=1}^{M} r_{1i} G_{ij}^N = \sum_{k=1}^{M} r_{N,k} (\mathbf{S}_M'^N)_{kj} \left(\sum_{i=1}^{M} r_{1i} G_{ik}^{N-1} \right). \tag{25}$$

Now we use the Feynman formula again to write

$$\frac{1}{\prod_{j=1}^{M} \left[\sum_{k=1}^{M} r_{N,k} (\mathbf{S}_M'^N)_{kj} \left(\sum_{i=1}^{M} r_{1i} G_{ik}^{N-1} \right) \right]} = (M-1)! \int_0^1 dx_1 \cdots \int_0^1 dx_M \times \frac{\delta(1 - \sum_i x_i)}{\left[\sum_{ijk} x_i r_{N,k} (\mathbf{S}_M'^N)_{ki} r_{1i} G_{ik}^{N-1} \right]^M}. \tag{26}$$

Finally, therefore,

$$\int_0^1 dr_{N,1} \cdots \int^1 dr_{N,M} \times \frac{\delta(1 - \sum_k r_{N,k})}{\prod_{j=1}^{M} \left[\sum_{k=1}^{M} r_{N,k} (\mathbf{S}_M'^N)_{kj} \left(\sum_{i=1}^{M} r_{1i} G_{ik}^{N-1} \right) \right]}$$

$$= \int_0^1 dx_1 \cdots \int_0^1 dx_M \frac{\delta(1 - \sum_j x_j)}{\prod_{k=1}^{M} \sum_{j=1}^{M} x_j (\mathbf{S}_M'^N)_{kj} \sum_{i=1}^{M} r_{1i} G_{ik}^{N-1}}$$

$$= \int_0^1 dx_1 \cdots \int_0^1 dx_M \frac{\delta(1 - \sum_j x_j)}{\prod_{k=1}^{M} \sum_{j=1}^{M} x_j (\mathbf{S}_M'^N)_{kj}} \times \frac{1}{\prod_{k=1}^{M} \left(\sum_{i=1}^{M} r_{1i} G_{ik}^{N-1} \right)}, \tag{27}$$

which shows that

$$I_N = I_{N-1} \times \int_0^1 dx_1 \cdots \int_0^1 dx_M \frac{\delta(1 - \sum_i x_i)}{\prod_{k=1}^{M} \sum_{j=1}^{M} x_j (\mathbf{S}_M'^N)_{ki}} = I_{N-1} \times f_M(s_{iN}, \cdots, s_{M+1,N}). \tag{28}$$

Now, since $\mathbf{S}_M'^i$ is the same function of the s_{ij} for each i, the factoring is demonstrated by induction.

Before writing down the final result, we will find an explicit expression for $|\mathbf{A}_M^i|$.

Let $\mathfrak{A}^M =$

$$\begin{vmatrix} \alpha_1 + \alpha_2 & \alpha_2 & 0 & \cdots & 0 & 0 \\ \alpha_2 & \alpha_2 + \alpha_3 & \alpha_3 & \cdots & 0 & 0 \\ 0 & \alpha_3 & \alpha_3 + \alpha_4 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & & \alpha_{M+1} + \alpha_M & \alpha_M \\ 0 & 0 & 0 & \cdots & \alpha_M & \alpha_M + \alpha_{M+1} \end{vmatrix}. \tag{29}$$

Then

$$\mathfrak{A}^M = \frac{1}{M!} \sum_P \alpha_{i_1} \alpha_{i_2} \cdots \alpha_{i_M}, \tag{30}$$

where $\sum_P$ denotes the sum over permutations of the indices and $(i_1, \cdots, i_M)$ is a permutation of M of the integers from 1 to $M+1$.

Proof: Proof is by induction on M.

For $M = 1$, $\mathfrak{A}^1 = \alpha_1 + \alpha_2$ and the statement is obviously true. Now, expanding $\mathfrak{A}^M$ in terms of its last column and row,

$$\mathfrak{A}^M = (\alpha_M + \alpha_{M+1}) \mathfrak{A}^{M-1} - \alpha_M^2 \mathfrak{A}^{M-2} = (\alpha_M + \alpha_{M+1}) \frac{1}{M!} \sum_P \alpha_{i_1} \cdots \alpha_{i_{M-1}} - \frac{1}{(M-2)!} \alpha_M^2 \sum_P \alpha_{i_1} \cdots \alpha_{i_{M-2}}. \tag{31}$$

But notice

$$\sum_P \alpha_{i_1} \cdots \alpha_{i_{M-1}} \alpha_{M+1} = \frac{1}{M} \sum_P \alpha_{i_1} \cdots \alpha_{i_M},$$

$$\sum_P \alpha_{i_1} \cdots \alpha_{i_{M-1}} \alpha_M = \frac{1}{M-1} \sum_P \alpha_{i_1} \cdots \alpha_{i_{M-1}}, \quad (32)$$

and the statement is proved.

So we have, at last, an expression for the high-energy behavior of F_N^M:

$$F_N^M \underset{s\to\infty}{\to} (-1)^{M+1} g_{(3)}^{2(N+M)} g_{(4)}^{(N-1)(M-1)} \left[\frac{\pi^2}{(2\pi)^4}\right]^{NM} \times [\varphi_M(t)]^N \frac{(\ln s)^N}{N!\, s^M}, \quad (33)$$

where

$$\varphi_M(t) = \int_0^1 ds_1 \cdots \int_0^1 ds_{M+1}\, \delta(1 - \sum_i s_i) \times \left(\frac{1}{M!} \sum_P s_{i_1} \cdots s_{i_M}\right)^{M-2} f_M(s_1, \cdots, s_{M+1}) \times \left[\frac{s_1 \cdots s_{M+1}}{\frac{1}{M!} \sum_P s_{i_1} \cdots s_{i_M}} t + m^2\right]^{-1}, \quad (34)$$

and f_M is given by Eq. (28).

IV. CONCLUSION

In conclusion, we note that the asymptotic limit of the scattering amplitude corresponding to the diagram of Fig. 1 is a simple generalization of the Regge-type behavior of the ladder diagrams considered by FG. From the work of LS, we know that for the case $M = 1$ (that is, for two particles in the t-channel intermediate state), the asymptotic limit of the sum of ladder diagrams is equal to the sum of the asymptotic limits, at least for the leading term. This leads us to the belief that the same may be true for arbitrary M:

$$\sum_{N=1}^{\infty} F_N^M \underset{s\to\infty}{\to} \text{const} \times s^{-M+\alpha_M(t)}. \quad (35)$$

In fact, our results suggest that (1) in general, M-particle bound states will give rise to Regge poles, (2) the M-particle Regge pole furthest to the right starts at $l = -M + 1$ for infinite t, and (3) in analogy with the results of LS, assuming the validity of an expansion of the position of the pole in powers of the coupling constant, the binding energy of an M-particle state may be calculable. Of course we have not proved the analyticity in the l plane that this implies; it is apparent that our results are more suggestive than substantive.

ACKNOWLEDGMENTS

The author wishes to thank Professor P. Federbush for his suggestions and guidance.

This work is supported in part through funds provided by the U. S. Atomic Energy Commission under Contract AT(30-1) 2098.

Three-Particle Scattering. I. Planar Case*

R. C. WHITTEN
Stanford Research Institute, Menlo Park, California
(Received 6 December 1962)

A formal theory of three-particle scattering in a plane is developed using integral equation methods. Expressions for the scattering amplitude and cross section of elastic and inelastic collisions are derived. The effects of indistinguishability of the colliding particles are discussed.

I. INTRODUCTION

THE theory of three-body scattering in its general case has obvious applications to ionic and chemical reactions; in spite of well-known difficulties it may even be possible to compute the rate constants for such simple cases as three-body electron attachment to hydrogen atoms. Such calculations can be approached via the standard method of collision cross sections, or alternatively, by means of an appropriate representation of the scattering operator.[1] Investigations of some aspects of three-body collisions have been carried out by Delves[2] and by Gallina and coworkers.[3] The latter have obtained the Green's function for three-particle scattering in three-dimensional space, but for S waves only.

Recent developments in the theory of generalized angular momentum representations[4,5] are admirably suited to the treatment of three-particle collisions. For the present we shall restrict our investigation to motion in a plane, emphasizing developments which can be utilized in the extension of our treatment of the problem to three dimensions.

Smith's formalism[4,5] for treating three-body collisions is expressed in terms of a generalized angular momentum operator $\mathbf{\Lambda}$, components of which are given by

$$ {}_m\Lambda_{ij}^{kl} = {}_m\xi_i^k \, {}_m\pi_j^l - {}_m\xi_j^l \, {}_m\pi_i^k, \tag{1} $$

where m labels the order in which the particles are paired, e.g., 1 to 2, and 3 to the 1–2 pair, k and l are particle indices, and i and j denote the cartesian components of the appropriate position (momentum) vectors. The symbols ${}_m\xi^i$ and ${}_m\pi^i$ represent the "normalized" position and momentum of the ith and jth particle pairs in the center of mass system with coupling order m:

$$ \begin{aligned} {}_m\xi^1 &= d_m^{-1}(\mathbf{x}^{m+2} - \mathbf{x}^{m+1}), \\ {}_m\xi^2 &= d_m[\mathbf{x}^m - (m_{m+1} + m_{m+2})^{-1} \\ &\qquad \times (m_{m+1}\mathbf{x}^{m+1} + m_{m+2}\mathbf{x}^{m+2})], \\ {}_m\pi^1 &= \frac{\mu}{d_m}\left(-\frac{\mathbf{p}^{m+1}}{m_{m+1}} + \frac{\mathbf{p}^{m+2}}{m_{m+2}}\right), \\ {}_m\pi^2 &= \mu d_m\left(\frac{\mathbf{p}^m}{m_m} - \frac{\mathbf{p}^{m+1} + \mathbf{p}^{m+2}}{m_{m+1} + m_{m+2}}\right), \end{aligned} \tag{2} $$

in which

$$ \begin{aligned} \mu^2 &= m_1 m_2 m_3/M \quad \text{(reduced mass)}, \\ M &= m_1 + m_2 + m_3, \\ d_k^2 &= (m_k/\mu)(1 - m_k/M). \end{aligned} \tag{3} $$

In this representation the center-of-mass position $\mathbf{X}$ and momentum $\mathbf{P}$ are given by

$$ \mathbf{X} = \sum_{k=1}^{3} \frac{m_k}{M}\mathbf{x}^k, \qquad \mathbf{P} = \sum_{k=1}^{3} \mathbf{p}^k. \tag{4} $$

From the variables (2), one can form several independent dynamical variables related to angular momentum of which

$$ \Lambda^2 = \tfrac{1}{2}\sum_{ijkl} |\Lambda_{ij}^{kl}|^2, \tag{5a} $$

$$ \Sigma_t = \sum_i \Lambda_{ii}^{12}, \tag{5b} $$

$$ L_1 = \Lambda_{12}^{11}, \tag{5c} $$

$$ L_2 = \Lambda_{12}^{22}, \tag{5d} $$

$$ L = L_1 + L_2, \tag{5e} $$

$$ Y = L_1 - L_2, \tag{5f} $$

are the most important; they become operators upon making the appropriate quantum-mechanical replacements, e.g., ${}_m\pi_i^1 \rightarrow i\hbar\,(\partial/\partial_m\xi_i^1)$. Since the motion in the center-of-mass system embodies four degrees of freedom, a full description of the assembly

* This research was supported by the National Aeronautics and Space Administration and the National Science Foundation.

[1] F. T. Smith, J. Chem. Phys. **36**, 250 (1962).
[2] L. M. Delves, Nucl. Phys. **9**, 391 (1958–59); **20**, 275 (1960).
[3] V. Gallina, P. Nata, L. Bianchi, and G. Viano, Nuovo Cimento **24**, 835 (1962).
[4] F. T. Smith, Phys. Rev. **120**, 1058 (1960).
[5] F. T. Smith, J. Math. Phys. **3**, 735 (1962).

of particles must contain four and only four independent dynamical variables (exclusive of those associated with internal structure), one of which is conveniently taken as the kinetic energy K. Three useful choices of the other three are (A) Λ^2, L and Σ_t, (B) Λ^2, L and Y, and equivalently (C) Λ^2, L_1 and L_2. With the aid of raising and lowering operators for the appropriate dynamical variables, Smith[5] has constructed specific representations of the unitary transformations connecting representations (A) and (B); representations (B) and (C) differ only by a phase factor. The elements of the transformation are identical with the b coefficients contained in expressions (29) and (30) of reference 5. Representation (A) has been appropriately termed symmetric because all three particles are treated on a completely equal footing. On the other hand, representations (B) and (C) depend upon the order in which the particles are paired; hence they are not treated on an equal basis and we call them asymmetric. This can also be stated in terms of transformation properties under a change in the pairing order; this involves an orthogonal transformation of the ξ's and π's, called by Smith a kinematic rotation. Representation (A) is invariant under such a transformation while (B) and (C) are not.

One can attach the following physical significance to the dynamical variables L, L_1, L_2, Λ^2, and Σ_t. That of L, L_1, and L_2 is immediately apparent: they are respectively the total *ordinary* angular momentum, the ordinary angular momentum of particles 1 and 2 about their center of mass, and the ordinary angular momentum of particle 3 about the center of mass of particles 1 and 2. Actually we should add a prefix to L_1 and L_2 to denote the order of pairing of the particles, e.g., ${}_mL_1$. Smith[4] has related (classically) Λ^2 to the kinetic energy K in the center-of-mass system and a three-body "impact parameter" R by the equation

$$\Lambda^2 = 2\mu K R^2, \tag{6}$$

where R is defined as the minimum value of $\rho = [\sum_{i,j} (\xi_i^j)^2]^{\frac{1}{2}}$ on a straight-line trajectory. Hence Λ^2 is a measure of the tendency of the three particles to simultaneously pass through a given region. We now raise the question, how closely for a given value of R does the system approach the situation in which the particles are equidistant at the instant of time when $\rho = R$? In other words, when would $|\mathbf{x}^1 - \mathbf{x}^2| = |\mathbf{x}^2 - \mathbf{x}^3| = |\mathbf{x}^3 - \mathbf{x}^1|$? This question can best be answered if we first go to the one-dimensional case in which Σ_t and K are the only dynamical variables. The particles approach most closely when Σ_t is a minimum and coincide at some instant of time if $\Sigma_t = 0$. This interpretation of Σ_t can be readily carried over to the planar case; for a given value of Λ^2 the particles approach a three-body collision most closely for $\Sigma_t = 0$ and progressively less closely as Σ_t increases. Obviously, if $\Lambda^2 = 0$, we must have $\Sigma_t = 0$.

As Smith has demonstrated, the classical argument can readily be carried over into the quantum domain. When the eigenvalue of the square of the generalized angular momentum Λ^2 vanishes the particles are coincident at some instant of time and the eigenvalue of Σ_t, σ, must also vanish; this is indeed a consequence of the properties of $\mathbf{\Lambda}$. One can, of course, also characterize the "three-body closeness of approach" by Λ^2, L_1, and L_2, but this description suffers from its asymmetry. It is apparent from the foregoing that for a short-range interaction potential, the three particles will have the greatest probability of undergoing a true "three-body" collision when the quantum numbers λ and σ (which partially characterize the system) are small.

Although the most appropriate representation for treating 3-body scattering is the symmetric one with wavefunctions denoted[5] by $\langle K\Lambda L\Sigma_t \mid \rho\ {}_i\Theta,\ {}_i\Phi,\ \varphi\rangle$, the coordinates ρ, ${}_i\Theta$, ${}_i\Phi$, and φ are related to the "normalized" center-of-mass system cartesian coordinates ξ_i^1, ξ_i^2 by intractable bilinear forms. Instead we shall employ the asymmetric representation with wavefunctions $\langle K\Lambda LY \mid \rho\chi\phi_+\phi_-\rangle$, or rather its equivalent $\langle K\Lambda L_1L_2 \mid \rho\chi\phi_1\phi_2\rangle$, returning to the set $(K, \Lambda, L\Sigma_t)$ later. The transformation equations connecting (ξ^1, ξ^2) and $(\rho, \chi, \phi_1, \phi_2)$ are

$$\begin{aligned} \xi_1^1 &= \rho\cos\chi\cos\phi_1, & \xi_2^1 &= \rho\cos\chi\sin\phi_1, \\ \xi_1^2 &= \rho\sin\chi\cos\phi_2, & \xi_2^2 &= \rho\sin\chi\sin\phi_2, \end{aligned} \tag{7}$$

in which $\rho^2 = \sum_{i,j=1} (\xi_i^j)^2$.

The Schrödinger equation describing the motion of the three particles can be written in terms of the variables ξ_1 and ξ_2 or alternatively in terms of the coordinates $(\rho, \chi, \phi_1, \phi_2)$. The former set yields plane-wave solutions

$$\langle\pi|\xi\rangle = [1/(2\pi)^2]\exp(i\pi\cdot\xi), \tag{8}$$

in which we have chosen our units such that $\hbar = 1$ and the normalization is one particle per unit volume. The scalar product $\pi\cdot\xi = \pi^1\cdot\xi^1 + \pi^2\cdot\xi^2$ can be expressed in terms of the angular coordinates as follows:

$$\begin{aligned} \pi\cdot\xi = k\rho[&\cos\chi\cos\bar{\chi}\cos(\phi_1 - \bar{\phi}_1) \\ &+ \sin\chi\sin\bar{\chi}\cos(\phi_2 - \bar{\phi}_2)], \end{aligned} \tag{9}$$

where $k = |\pi|$ and $\rho = |\xi|$; the barred angles fix the direction of π and the unbarred angles that of ξ. One cannot perform a coordinate rotation to a system in which $\chi = 0$ or $\bar{\chi} = 0$; hence all angles must be retained in the computations.

If we are dealing with a "central" potential, i.e., one in which V is a function of $|\xi|$ only, or one which is "almost" central, the coordinate systems $(\rho\chi\phi_1\phi_2)$ or $\rho\Theta\phi\Phi)$ are appropriate. The solutions of the corresponding Schrödinger equations outside the range of the potential are[5]

$$\langle K\lambda L_1 L_2|\rho\chi\phi_1\phi_2\rangle = (1/k\rho)J_{\lambda+1}(k\rho)\mathcal{J}_{\lambda m_1 m_2}(\chi\phi_1\phi_2), \quad (10a)$$

$$\langle K\lambda L\Sigma_t|\rho\Theta\phi\Phi\rangle = (1/k\rho)J_{\lambda+1}(k\rho)\mathcal{S}_{\lambda m+\sigma}(\Theta\phi\Phi), \quad (10b)$$

where $J_{\lambda+1}(k\rho)$ is the Bessel function of the first kind of order $\lambda + 1$, k is the magnitude of the momentum ($k^2 = 2\mu K$), and

$$\Lambda^2\mathcal{J}_{\lambda m_1 m_2} = \lambda(\lambda + 2)\mathcal{J}_{\lambda m_1 m_2}, \quad (11a)$$

$$L_{1,2}\mathcal{J}_{\lambda m_1 m_2} = m_{1,2}\mathcal{J}_{\lambda m_1 m_2}, \quad (11b)$$

$$\Lambda^2\mathcal{S}_{\lambda m+\sigma} = \lambda(\lambda + 2)\mathcal{S}_{\lambda m+\sigma}, \quad (11c)$$

$$L\mathcal{S}_{\lambda m+\sigma} = m_+\mathcal{S}_{\lambda m+\sigma}, \quad (11d)$$

$$\Sigma_t\mathcal{S}_{\lambda m+\sigma} = \sigma\mathcal{S}_{\lambda m+\sigma}. \quad (11e)$$

$\mathcal{J}_{\lambda m_1 m_2}(\chi\phi_1\phi_2)$ and $\mathcal{S}_{\lambda m+\sigma}(\Theta\phi\Phi)$ are given by

$$\begin{aligned}\mathcal{J}_{\lambda m_1 m_2}(\chi\phi_1\phi_2) = {} & \{[2(\lambda+1)]^{\frac{1}{2}}/2\pi\}e^{im_1(\phi_1-\frac{1}{2}\pi)}e^{im_2(\phi_2-\frac{1}{2}\pi)} \\ & \times \{[\tfrac{1}{2}(\lambda - m_1 + m_2)]!\,[\tfrac{1}{2}(\lambda + m_1 - m_2)]! \\ & \times [\tfrac{1}{2}(\lambda - m_1 - m_2)]!\,[\tfrac{1}{2}(\lambda + m_1 + m_2)]!\}^{\frac{1}{2}} \quad (12a) \\ & \times \sum_{\kappa=0}^{\frac{1}{2}(\lambda-m_1-m_2)} (-1)^\kappa\{\kappa!\,(m_1+\kappa)! \\ & \times [\tfrac{1}{2}(\lambda - m_1 + m_2) - \kappa]! \\ & \times [\tfrac{1}{2}(\lambda - m_1 - m_2) - \kappa]!\}^{-1} \\ & \times \sin^{\lambda-m_1-2\kappa}\chi\cos^{m_1+2\kappa}\chi\end{aligned}$$

$$\mathcal{S}_{\lambda m+\sigma}(\Theta\varphi\Phi) = \{[2(\lambda+1)]^{\frac{1}{2}}/2\pi\}\,\mathfrak{D}^{\frac{1}{2}\lambda}_{\frac{1}{2}m,-\frac{1}{2}\sigma}(2\phi, \pi - 2\Theta, -2\Phi). \quad (12b)$$

The functions $\mathfrak{D}^{\alpha}_{\beta\gamma}$ appearing in (12b) are the elements of the Wigner representation of the three-dimensional rotation group.[6]

In the following section we shall compute the free-particle Green's function, after which the elastic and inelastic three-particle scattering amplitude and collision cross section will be treated. In the final section we shall extend our development to the case of identical particles.

[6] E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (Academic Press Inc., New York, 1959).

2. FREE-PARTICLE GREEN'S FUNCTION

In developing the dynamics of a three-body collision it is convenient to employ the integral equation methods of Lippmann and Schwinger.[7,8,9] In this scheme the state vector $|0\rangle$ of the system is related to the incoming state vector $|i\rangle$ by the equation

$$|0\rangle = |i\rangle + G^{(+)}(E_0)V\,|0\rangle, \quad (13)$$

where V is the interaction potential and $G^{(+)}(E_0)$ is the free-particle Green's function given in operator form by

$$G^{(+)}(E_0) = \lim_{\epsilon\to+0}(E_0 - K + i\epsilon)^{-1}. \quad (14)$$

In Eq. (14) K is the free-particle Hamiltonian operator and ϵ is an adiabatic switching parameter introduced in order to avoid transients in the time dependence of the state function during the scattering process; it has the significance of feeding in the incident wave over a period of time rather than releasing it suddenly.[8] The positive sign adjacent to ϵ in (14) ensures outgoing scattered waves. In position coordinate representation we have

$$\langle\xi|0\rangle = \langle\xi|i\rangle + \int G^{(+)}_{E_0}(\xi, \xi')V(\xi')\langle\xi'|0\rangle\,d\xi', \quad (15)$$

where

$$\begin{aligned}G^{(+)}_{E_0}(\xi, \xi') &= \int \langle\xi|\pi\rangle\langle\pi|G^{(+)}(E_0)|\pi\rangle\langle\pi|\xi'\rangle\,d\pi \\ &= \lim_{\epsilon\to+0}\frac{2\mu}{(2\pi)^4}\int\frac{\exp[i\pi\cdot(\xi - \xi')]}{k_0^2 - k^2 + 2i\mu\epsilon}\,d\pi; \quad k = |\pi|, \quad (16)\end{aligned}$$

using the linear momentum representation of the operator $G^{(+)}_E$; the symbol μ represents the reduced mass and the integral is taken over all momentum space π. The scalar products $\pi\cdot\xi$ are now expressed in terms of the hyperpolar coordinates (χ, ϕ_1, ϕ_2) as in Eq. (9). After carrying out the integration over the azimuthal angles ϕ_1, ϕ_2, we obtain

$$\begin{aligned}G^{(+)}_{k_0}(\xi, \xi') = {} & \lim_{\epsilon\to+0}\frac{2\mu}{(2\pi)^4} \\ & \times \int_0^\infty\int_0^{\frac{1}{2}\pi}\frac{J_0(kr_1\cos\chi)J_0(kr_2\sin\chi)}{k^2 - k^2 + 2i\mu\epsilon} \\ & \times \tfrac{1}{2}\sin 2\chi\,d\chi k^3\,dk, \quad (17)\end{aligned}$$

[7] B. A. Lippmann and J. Schwinger, Phys. Rev. 79, 469 (1950).

[8] M. Gell-Mann and M. L. Goldberger, Phys. Rev. 91, 398 (1953).

[9] G. Gioumousis and C. F. Curtis, J. Chem. Phys. 29, 996 (1958).

where

$$r_1 = [\xi^2 \cos^2 \bar{\chi} + \xi'^2 \cos^2 \bar{\bar{\chi}} - 2\xi\xi' \cos \bar{\chi} \cos \bar{\bar{\chi}} \cos (\bar{\phi}_1 - \bar{\bar{\phi}}_1)]^{\frac{1}{2}},$$

$$r_2 = [\xi^2 \sin^2 \bar{\chi} + \xi'^2 \sin^2 \bar{\bar{\chi}} - 2\xi\xi' \sin \bar{\chi} \sin \bar{\bar{\chi}} \cos (\bar{\phi}_2 - \bar{\bar{\phi}}_2)]^{\frac{1}{2}},$$

and the directions of ξ and ξ' are specified by the sets of angles $(\bar{\chi}\bar{\phi}_1\bar{\phi}_2)$ and $(\bar{\bar{\chi}}\bar{\bar{\phi}}_1\bar{\bar{\phi}}_2)$, respectively.

Using a theorem due to Sonine,[10] one can immediately integrate over χ, obtaining the result

$$G_{k_0}^{(+)}(\xi, \xi') = \lim_{\epsilon \to +0} \frac{2\mu}{(2\pi)^2} \frac{1}{r} \int_0^\infty \frac{J_1(kr)}{k_0^2 - k^2 + 2i\mu\epsilon} k^2 \, dk$$

$$= \lim_{\epsilon \to +0} \frac{2\mu}{(2\pi)^2} \frac{1}{r} \frac{\partial}{\partial r} \left[\int_0^\infty \frac{kJ_0(kr)}{k_0^2 - k^2 + 2i\mu\epsilon} dk \right], \quad (18)$$

in which $r = (r_1^2 + r_2^2)^{\frac{1}{2}}$. The integral in (18) which we shall call I can be evaluated by constructing a contour of integration in the complex plane such that outgoing waves are guaranteed. To do this, one first expresses $J_0(kr)$ as a sum of Hankel functions of the first and second kinds;

$$J_0(\chi) = \tfrac{1}{2}[H_0^{(1)}(\chi) + H_0^{(2)}(\chi)], \quad (19)$$

thus recasting the integral I in the form

$$I = \lim_{\epsilon \to +0} \frac{1}{2} \int_0^\infty \frac{H_0^{(1)}(kr) k \, dk}{k_0^2 - k^2 + 2i\mu\epsilon} + \lim_{\epsilon \to +0} \frac{1}{2} \int_0^\infty \frac{H_0^{(2)}(kr) k \, dk}{k_0^2 - k^2 + 2i\mu\epsilon}. \quad (20)$$

The first integral I_1 in (20) is evaluated by means of Cauchy's integral formula and the appropriate contour of integration shown in Fig. 1(a).

$$\lim_{\epsilon \to +0} \left[\frac{1}{2} \int_0^\infty \frac{H_0^{(1)}(kr)}{k_0^2 - k^2 + 2i\mu\epsilon} k \, dk + \frac{1}{2} \int_C \frac{H_0^{(1)}(kr)}{k_0^2 - k^2 + 2i\mu\epsilon} k \, dk + \frac{1}{2} \int_{i\infty}^0 \frac{H_0^{(1)}(kr)}{k_0^2 - k^2 + 2i\mu\epsilon} k \, dk \right]$$

$$= 2\pi i \, \mathrm{Res}\,(k_0). \quad (21)$$

The second integral in (21) vanishes because $H_0^{(1)}(Z)$ approaches zero in the limit $|Z|$ approaches infinity. Using the contour of integration shown in Fig. 1(b), the second integral I_2 in (20) is computed in a similar manner. Substituting the solutions of I_1 and I_2 into

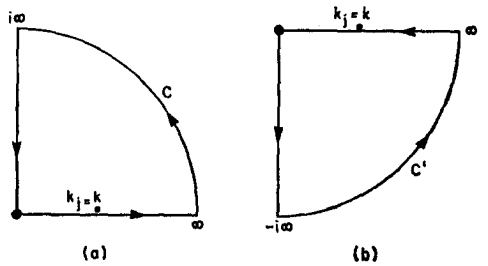


FIG. 1. Contours of integration for integrals I_1 and I_2 [Eq. (20)].

(18), we obtain

$$G_{k_0}^{(+)}(\xi, \xi') = \frac{i\mu}{4\pi} \frac{k}{r} H_1^{(1)}(kr), \quad (22)$$

or, asymptotically for large ξ,

$$G_{k_0}^{(+)}(\xi, \xi') = e^{-i\pi/4} \frac{\mu k^{\frac{1}{2}}}{(2\pi)^{\frac{3}{2}}} \frac{e^{ik\xi}}{\xi^{\frac{1}{2}}} e^{-ik\xi'\alpha}, \quad (23)$$

where

$$\alpha = \cos \chi \cos \chi' \cos (\phi_1 - \phi_1') + \sin \chi \sin \chi' \cos (\phi_2 - \phi_2'). \quad (24)$$

In addition to expressing Eq. (15) in the linear momentum representation we can also cast it in terms of generalized angular momentum eigenfunctions

$\langle \xi | K\lambda m_1 m_2 \rangle$ or $\langle \xi | K\lambda m_+ \sigma \rangle$;

$$G_{k_0}^{(+)}(\xi, \xi') = \sum_{\lambda m_1 m_2} \langle \xi | K\lambda m_1 m_2 \rangle \times \langle K\lambda m_1 m_2 | G_{k_l}^{(+)} | K\lambda m_1 m_2 \rangle \langle K\lambda m_1 m_2 | \xi' \rangle$$

$$= \lim_{\epsilon \to +0} 2\mu \sum_{\lambda m_1 m_2} \mathcal{J}_{\lambda m_1 m_2}(\hat{\xi}) \mathcal{J}^*_{\lambda m_1 m_2}(\hat{\xi}') \frac{1}{\xi\xi'} \times \int_0^\infty \frac{J_{\lambda+1}(k\xi) J_{\lambda+1}(k\xi')}{k_0^2 - k^2 + 2i\mu\epsilon} k \, dk$$

$$= \pi\mu i \sum_{\lambda m_1 m_2} \mathcal{J}_{\lambda m_1 m_2}(\hat{\xi}) \mathcal{J}^*_{\lambda m_1 m_2}(\hat{\xi}') \times \frac{1}{\xi\xi'} H_{\lambda+1}^{(1)}(k_0\xi) J_{\lambda+1}(k_0\xi'); \quad \xi > \xi', \quad (25)$$

and

$$G_{k_0}^{(+)}(\xi, \xi') = \sum_{\lambda m + \sigma} \langle \xi | K\lambda m_+ \sigma \rangle \times \langle K\lambda m_+ \sigma | G_{k_l}^{(+)} | K\lambda m_+ \sigma \rangle \langle K\lambda m_+ \sigma | \xi' \rangle$$

$$= \pi\mu i \sum_{\lambda m + \sigma} \mathcal{J}_{\lambda m_+ \sigma}(\hat{\xi}) \mathcal{J}^*_{\lambda m_+ \sigma}(\hat{\xi}') \times \frac{1}{\xi\xi'} H_{\lambda+1}^{(1)}(k_0\xi) J_{\lambda+1}(k_0\xi'); \quad \xi > \xi'. \quad (26)$$

It is interesting to note that the Green's function of N free particles moving in a plane can be derived

[10] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, London, 1952), 2nd Ed., p. 376.

with the aid of a suitable generalization of the above method. The asymptotic form

$$G_k^{(+)}(\xi, \xi')_N = \frac{e^{-i(\frac{1}{4}+\frac{1}{2}N)\pi}}{(2\pi)^{N-\frac{1}{2}}} \frac{\mu k^{N-5/2}}{|\xi - \xi'|^{N-\frac{1}{2}}} \exp(ik|\xi - \xi'|) \quad (27)$$

is derived in Appendix A. The vectors ξ and π are generalizations of Eq. (2) to N bodies:

$$\xi = \sum_{i=1}^{N-1} \xi^i,$$

$$\pi = \sum_{i=1}^{N-1} \pi^i.$$

3. THREE-PARTICLE ELASTIC SCATTERING

Having derived the Green's function we are now in a position to investigate three-particle scattering dynamics.

Asymptotically the wavefunction of the system takes the form

$$\Psi_{\pi_i}(\xi) \sim \phi_{k_i}(\xi) + e^{-i\pi/4} \frac{k^{\frac{1}{2}}}{\xi^{\frac{1}{2}}} e^{ik\xi} f(\hat{\pi}_i, \hat{\pi}_0), \quad (28)$$

where $f(\hat{\pi}_i, \hat{\pi}_0)$ is the three-body scattering amplitude given by

$$f(\hat{\pi}_i, \hat{\pi}_0) = \frac{\mu}{(2\pi)^{\frac{1}{2}}} \int_0^\infty \int_0^{\frac{1}{2}\pi} \int_0^{2\pi} \int_0^{2\pi} e^{-ik\xi'\alpha} V(\xi'\chi'\phi_1'\phi_2') \times \Psi_{\pi_i}(\xi'\chi'\phi_1'\phi_2') \tfrac{1}{2} \sin 2\chi' \, d\chi' \, d\phi_1' \, d\phi_2' \xi'^3 \, d\xi'. \quad (29)$$

The symbol α has the same significance as in (24) and $\hat{\pi}_i$ and $\hat{\pi}_0$ represent the directions of the incoming and scattered three-particle momenta, respectively. Alternatively, the scattering amplitude can be expanded, with the aid of (25), in a series of generalized angular momentum eigenfunctions

$$f(\hat{\pi}_i, \hat{\pi}_0) = (2\pi)^{\frac{1}{2}} \frac{\mu}{k} \sum_{\lambda m_1 m_2} (-1)^\lambda \mathcal{Y}_{\lambda m_1 m_2}(\hat{\pi}_i) \times \int \frac{J_{\lambda+1}(k\xi')}{\xi'} \mathcal{Y}^*_{\lambda m_1 m_2}(\hat{\xi}') V(\xi') \Psi_{\pi_i}(\xi') \, d\xi'. \quad (30)$$

For many purposes it is more useful to express the scattering amplitude in the generalized angular momentum representations $(\lambda m_1 m_2)$ or $(\lambda m_+ \sigma)$ rather than as an explicit function of the coordinates $(\chi\phi_1\phi_2)$ or $(\Theta\phi\Phi)$. The transformation coefficients are just the generalized angular momentum eigenfunctions

$$f(\chi\phi_1\phi_2; \chi'\phi_1'\phi_2') = \sum_{\substack{\lambda m_1 m_2 \\ \lambda' m_1' m_2'}} \mathcal{Y}^*_{\lambda m_1 m_2}(\chi\phi_1\phi_2) f^{\lambda\lambda'}_{m_1 m_1'; m_2 m_2'} \mathcal{Y}_{\lambda' m_1' m_2'}(\chi'\phi_1'\phi_2'), \quad (31)$$

$$f(\Theta\phi\Phi; \Theta'\phi'\Phi') = \sum_{\substack{\lambda m+\sigma \\ \lambda' m'\sigma'}} \mathcal{Y}^*_{\lambda m+\sigma}(\Theta\phi\Phi) f^{\lambda\lambda'}_{m+m+'; \sigma\sigma'} \mathcal{Y}_{\lambda' m+'\sigma'}(\Theta'\phi'\Phi'). \quad (32)$$

Inverting the expansions (31) and (32) and employing (30), we obtain, for $f^{\lambda\lambda'}_{m_1 m_1'; m_2 m_2'}$ and $f^{\lambda\lambda'}_{m+m+'; \sigma\sigma'}$,

$$f^{\lambda\lambda'}_{m_1 m_1'; m_2 m_2'} = (2\pi)^{\frac{1}{2}} \frac{\mu}{k} (-1)^\lambda \int \frac{J_{\lambda+1}(k\xi')}{\xi'} \mathcal{Y}^*_{\lambda m_1 m_2} \times (\hat{\xi}') V(\xi') \Psi_{\lambda' m_1' m_2'}(\xi') \, d\xi' = f^{\lambda\lambda'}_{m+m+'; m-m-'} e^{i\beta}, \quad (33)$$

$$f^{\lambda\lambda'}_{m+m+; \sigma\sigma'} = (2\pi)^{\frac{1}{2}} \frac{\mu}{k} (-1)^\lambda \times \int \frac{J_{\lambda+1}(k\xi')}{\xi'} \mathcal{Y}^*_{\lambda m+\sigma}(\hat{\xi}') V(\xi') \Psi_{\lambda' m'\sigma'}(\xi') \, d\xi', \quad (34)$$

where β is merely a phase angle. The scattering amplitude is easily transformed from one generalized angular momentum representation to the other by means of the unitary transformation

$$f^{\lambda\lambda'}_{m+m+'\sigma\sigma'} = \sum_{m_- m_-'} b_\lambda(m_-, \sigma) \times f^{\lambda\lambda'}_{m+m+'; m-m-'} b_{\lambda'}(m_-', \sigma'). \quad (35)$$

A variational method for computing the scattering amplitude corresponding to the two-particle collision process has been developed by Schwinger.[11] This procedure can be suitably adapted to the three-particle case by expressing the scattering amplitude as

$$f(\hat{\pi}_0, \hat{\pi}_i) = \frac{\left[\int \psi^*(\xi) U(\xi) \exp(i\pi_i \cdot \xi) \, d\xi\right]\left[\int \exp(-i\pi_0 \cdot \xi') U(\xi') \psi(\xi') \, d\xi'\right]}{\left[\psi^*(\xi) U(\xi) \psi(\xi) \, d\xi - \frac{k^{\frac{1}{2}} e^{-i\pi/4}}{(2\pi)^{\frac{1}{2}}} \iint \psi^*(\xi) U(\xi) \frac{e^{ikR}}{R^{\frac{1}{2}}} U(\xi') \psi(\xi') \, d\xi' \, d\xi\right]}, \quad (36)$$

where $R = |\xi - \xi'|$ and $U(\xi) = \mu V(\xi)$. Variation of f with respect to ψ^* then yields Eq. (29). The utility of this principle lies in computing wavefunctions ψ which make the scattering amplitude on extremum, and therefore presumably give a best fit to it. In practice this is carried out by writing the wave-

[11] D. R. Bates, *Quantum Theory I. Elements* (Academic Press Inc., New York, 1961), p. 350.

function of the system as a function of various parameters, e.g.,

$$\psi = \sum_i \alpha_i \phi_i. \tag{37}$$

Variation of f with respect to the expansion parameters α_i results in the set of simultaneous equations

$$\partial f/\partial \alpha_i = 0, \tag{38}$$

which are then solved for α_i. This procedure, of course, yields only an upper or lower bound to the magnitude of the scattering amplitude depending upon whether the potential is repulsive or attractive; the "goodness" of the approximation depends upon an adroit choice of the basis (ϕ_i). It is hoped that this approach can be exploited in the eventual treatment of actual three-body problems.

In two-particle scattering, the differential scattering cross section $\sigma(\theta, \varphi)$ is related to the scattering amplitude by the equality

$$\sigma(\theta, \varphi) = |f(\theta, \varphi)|^2, \tag{39}$$

where (θ, φ) fix the orientation of the outgoing with respect to the incoming momentum. Derivation of the three-particle analogue of (39) can be approached in two ways which, of course, must yield the same result. The first one is based on the elementary definition of cross section

$$\sigma(\pi_0, \pi_i) = [|\mathbf{j}_{\text{out}}|/|\mathbf{j}_{\text{in}}|]\xi^3 \, d\Omega, \tag{40}$$

where $\mathbf{j}_{\text{in}}$ and $\mathbf{j}_{\text{out}}$ are the probability current densities of the incoming and outgoing particles, respectively, and $\xi^3 \, d\Omega$ is an element of "surface area" of a large hypersphere the center of which is coincident with the center of scattering. The probability current densities are given by

$$\mathbf{j}_{\text{out}} = (1/2i\mu)[\psi_{sc}^* \nabla \psi_{sc} - (\nabla \psi_{sc}^*)\psi_{sc}], \tag{41a}$$

$$\mathbf{j}_{\text{in}} = (1/2i\mu)[\phi_i^* \nabla \phi_i - (\nabla \phi_i^*)\phi_i], \tag{41b}$$

where ψ_{sc} and ϕ_i are the scattered and incident wavefunctions, respectively and μ is the reduced mass. The normalized wavefunctions ϕ_i and ψ_{sc} are

$$\phi_i(\xi) = [1/(2\pi)^2] \exp(i\pi \cdot \xi), \tag{42a}$$

$$\psi_{sc}(\xi) = (k^{\frac{1}{2}}/\xi^{\frac{5}{2}}) e^{-i\pi/4} e^{ik\xi} f(\hat{\pi}_0, \hat{\pi}_i), \tag{42b}$$

from which we obtain, with the aid of (41) and (40),

$$\sigma(\pi_0, \pi_i) = k \, |f(\hat{\pi}_0, \hat{\pi}_i)|^2. \tag{43}$$

The cross section σ has the dimensions of length cubed since $f(\hat{\pi}_0, \hat{\pi}_i)$ can be shown to have dimensions of length squared.

We make the second approach via the appropriate matrix element of the transition rate amplitude R_{π_0, π_i} corresponding to the collision process

$$\sigma(\pi_0, \pi_i) = (2\pi/V_0)\rho(E) \, |R_{\pi_0, \pi_i}|^2, \tag{44}$$

where V_0 is the velocity of the incoming particles given by $V_0 = k/\mu$, and $\rho(E)$ is the number of final states per unit energy:

$$\rho(E) = \frac{k^3}{(2\pi)^4} \frac{dk}{dE} \, d\Omega = \frac{\mu k^2}{(2\pi)^4} \, d\Omega. \tag{45}$$

The matrix element in question is $\langle 0 \, |V| \, i \rangle$ which can be recast in the more useful form[9]

$$R_{\pi_0, \pi_i} = \lim_{\epsilon \to +0} -i\epsilon \langle \pi_0 | \pi_i \rangle, \tag{46}$$

or, explicitly in terms of (42),

$$R_{\pi_0, \pi_i} = \lim_{\epsilon \to +0} (-i\epsilon) \int_0^\infty \int_0^{\frac{1}{2}\pi} \int_0^{2\pi} \int_0^{2\pi} \exp(-i\pi_0 \xi) \times \left[\exp(i\pi_i \cdot \xi) + \frac{k^{\frac{1}{2}}}{\xi^{\frac{5}{2}}} e^{ik\xi} e^{-i\pi/4} e^{-\epsilon\mu\xi/k} f(\hat{\pi}_0, \hat{\pi}_i) \right] \xi^3 \, d\xi \, d\Omega, \tag{47}$$

in which the factor $e^{-\epsilon\mu\xi/k}$ is introduced because of the requirement that we replace the energy E by $E + i\epsilon$ when we move into the first quadrant of the complex plane. The term containing $\exp[i(\pi_i - \pi_0) \cdot \xi]$ contributes a delta function which is of no interest since it indicates *no scattering*. In order to carry out the second integration, one must expand the plane wave (momentum eigenfunction) $\exp(-i\pi_0 \cdot \xi)$ in a series of generalized angular momentum eigenfunctions $\mathcal{J}_{\lambda m_1 m_2}$ (see Appendix B). Making this substitution in (39), and changing the lower limit of the ξ integration to some large but finite value r, we obtain

$$\begin{aligned} R_{\pi_0, \pi_i} = \lim_{\epsilon \to +0} (-i\epsilon) \sum_{\lambda m_1 m_2} (2\pi)^2 (-i)^\lambda \\ \times \int_0^\infty \int_0^{\frac{1}{2}\pi} \int_0^{2\pi} \int_0^{2\pi} \frac{J_{\lambda+1}(k\xi) e^{ik\xi}}{k^{\frac{1}{2}} \xi^{5/2}} \\ \times e^{-\epsilon\mu\xi/k} e^{-i\pi/4} \mathcal{J}_{\lambda m_1 m_2}(\chi'' \phi_1'' \phi_2'') \\ \times \mathcal{J}^*_{\lambda m_1 m_2}(\chi' \phi_1' \phi_2') f(\chi \phi_1 \phi_2; \chi' \phi_1 \phi_2) \\ \times \xi^3 \, d\xi \sin\chi' \cos\chi' \, d\chi' \, d\phi_1' \phi_2'. \end{aligned} \tag{48}$$

The artifice of assuming a large finite value for the lower limit of the integral over ξ is quite legitimate; the integral $\int_0^\infty$ can be split into a sum $\int_r^\infty + \int_0^r$ wherein the second integral $\int_0^r$ is finite and therefore yields no contribution to R_{π_0, π_i} in the limit ϵ approaches $+0$. Since the value of ξ is large in the region of integration, the Bessel function $J_{\lambda+1}(k\xi)$ can be replaced by its asymptotic form. With the aid of (31) we obtain

$$R_{\pi_0, \pi_i} = [(2\pi)^3/\mu] f(\hat{\pi}_0, \hat{\pi}_i). \tag{49}$$

Finally one again arrives at the relation between the scattering amplitude and cross section when the form of R_{π_0,π_i} given by (49) is substituted into (44).

$$\sigma(\pi_0\pi_i) = k\,|f(\hat{\pi}_0, \hat{\pi}_i)|^2, \tag{43}$$

wherein $f(\hat{\pi}_0, \hat{\pi}_i)$ can be represented in $(\chi\phi_1\phi_2)$, $(\chi\phi_-\phi_+)$ or in $({}_i\Theta, {}_i\Phi, \varphi)$ coordinates.

4. PARTIAL-WAVE ANALYSIS

One of the most useful methods of analyzing low-energy two-body scattering is the expansion of the scattering amplitude in "partial waves" each of which is associated with a definite orbital angular momentum. The simplest case is that of a "central" short-range potential, i.e., $V = V(|\xi|)$. Although of very great importance in two-particle scattering, this appears at first glance to be a very unphysical interaction when one extends it to the three-particle case. Nevertheless, consideration of the connection between $|\xi|$ and the moment of inertia of the three particles about an axis through the center of mass and perpendicular to the plane of the particles, i.e., $I = \mu\xi^2$ where μ is the reduced mass and I is the moment of inertia, shows that it is reasonable, at least to first approximation. If the particles must simultaneously be within a certain distance of each other for the potential to act, the moment of inertia I must be less than some value I_0. A "square well" potential would then have the form

$$\begin{aligned} V(|\xi|) &= V_0; \qquad (I_0/\mu) < \xi_0^2 \\ &= 0 \quad (I_0/\mu) > \xi_0^2. \end{aligned} \tag{50}$$

If (50) does represent the interaction potential, the solution to the wave equation can be separated into radial and angular parts; the general solution is of the form

$$\Psi_{\pi_i}(\xi) = \sum_{\lambda m_1 m_2} A_{\lambda m_1 m_2} \frac{R_\lambda(k\xi)}{k\xi} \mathcal{Y}^*_{\lambda m_1 m_2}(\hat{\pi}_i)\mathcal{Y}_{\lambda m_1 m_2}(\hat{\xi}), \tag{51}$$

where $R_\lambda(k\xi)$ approaches

$$(1/k\xi)^{\frac{1}{2}} \sin\,[k\xi - \tfrac{1}{4}(2\lambda + 1)\pi + \eta_\lambda]$$

asymptotically; the η_λ are the *three-body* λ-dependent scattering phase shifts. Now the wavefunction of the system, $\Psi_{\pi_0}(\xi)$, can also be expressed as a function of the scattering amplitude (28). In addition, the asymptotic form of (51) can be separated into a part containing the common factor $e^{ik\xi}$ and another part containing the common factor $e^{-ik\xi}$: After the expansion of $\phi_i(\xi)$ in generalized angular momentum eigenfunctions (see Appendix B), we can equate the coefficients of $e^{ik\xi}$ and $e^{-ik\xi}$ appearing in the two forms of $\Psi_{\pi_i}(\xi)$ and solve for $A_{\lambda m_1 m_2}$:

$$A_{\lambda m_1 m_2} = (2\pi)^2(i)^\lambda e^{i\eta_\lambda}, \tag{52}$$

and

$$\begin{aligned} &f(\hat{\pi}_0, \hat{\pi}_i) \\ &= \frac{(2\pi)^{\frac{5}{2}}}{ik^2} \sum_{\lambda m_1 m_2} (e^{2i\eta_\lambda} - 1)\mathcal{Y}^*_{\lambda m_1 m_2}(\hat{\pi}_0)\mathcal{Y}_{\lambda m_1 m_2}(\hat{\pi}_i). \end{aligned} \tag{53}$$

By comparing (53) with (31) and (33) it becomes evident that

$$\begin{aligned} f^{\lambda\lambda'}_{m_1 m_1',\, m_2 m_2'} &= \frac{(2\pi)^{\frac{5}{2}}}{ik^2}(e^{2i\eta_\lambda} - 1)\,\delta_{\lambda\lambda'}\,\delta_{m_1 m_1'}\,\delta_{m_2 m_2'} \\ &= f^\lambda = f^{\lambda\lambda'}_{m+m+',\,\sigma\sigma'}. \end{aligned} \tag{54}$$

More complicated cases involving "noncentral" potentials will result in more complicated expressions for $f^{\lambda\lambda'}_{m_1 m_1',\, m_2 m_2'}$ or alternatively $f^{\lambda\lambda'}_{m+m+',\,\sigma\sigma'}$, which have off-diagonal terms and involve phase shifts labeled by m_1 and $m_2(m_+$ and $\sigma)$. A general treatment of such potential functions, an example of which is

$$\begin{aligned} V(\xi) &= V_0(|\xi|)(1 + \alpha\hat{\xi}^1\cdot\hat{\xi}^2); \\ &\qquad (\xi)^2 + \alpha\xi^1\cdot\xi^2 < (\xi_0)^2 \\ &= 0; \qquad (\xi)^2 + \alpha\xi^1\cdot\xi^2 > (\xi_0)^2, \end{aligned} \tag{55}$$

will be deferred for a future paper.

INELASTIC SCATTERING

Inelastic scattering is possible if one or more of the particles has internal structure or two of the particles can combine. Typical examples of the latter are three-body attachment of electrons to neutral atoms, three-body electron–ion recombination, and formation of diatomic molecules by means of three-body atomic collisions. The case in which the number of free-particles is conserved will be considered first.

If the three particles all have internal structure, the free-particle wavefunctions are of the form

$$\begin{aligned} \phi_i &= [1/(2\pi)^2]R_m(\rho)S_n(\sigma)T_p(\tau) \\ &\qquad \times \exp\,[i(\pi^1_\alpha\cdot\xi^1 + \pi^2_\alpha\cdot\xi^2)], \end{aligned} \tag{56}$$

where $R_m(\rho)$, $S_n(\sigma)$, and $T_p(\tau)$ represent the internal structures of the particles. The Green's function (14) of the three-body interaction then takes the form

$$G_E^{(+)}(\xi, \xi') = \lim_{\epsilon\to+0} \frac{1}{(2\pi)^4} \sum_{\mu\nu\lambda} \int \frac{\exp\,[i\pi_\alpha\cdot(\xi - \xi')]R_\mu(\rho)S_\nu(\sigma)T_\lambda(\tau)R^*_\mu(\rho')S^*_\nu(\sigma')T^*_\lambda(\tau')}{E - E_\alpha - E_\lambda - E_\nu - E_\mu + i\epsilon}\, d\pi_\alpha. \tag{57}$$

If we now make the substitution

$$(1/2\mu)k^2_{\mu\nu\lambda} = E - E_\lambda - E_\nu - E_\mu, \tag{58}$$

and carry out the integrations indicated in (57) using the methods of Sec. 2, we obtain

$$G_E^{(+)}(\xi\rho\sigma\tau;\ \xi'\rho'\sigma'\tau') = \frac{\sum_{\mu\nu\lambda} i\mu k_{\mu\nu\lambda} H_1^{(1)}(k_{\mu\nu\lambda}\,|\xi - \xi'|)}{4\pi\,|\xi - \xi'|}$$
$$\times R_\mu(\rho)S_\nu(\sigma)T_\lambda(\tau)R^*_\mu(\rho')S^*_\nu(\sigma')T^*_\lambda(\tau') \tag{59}$$

for the Green's function, and thus,

$$\psi(\xi,\ \rho,\ \sigma,\ \tau) = \varphi(\xi,\ \rho,\ \sigma,\ \tau)$$
$$+ \int G_E^{(+)}(\xi,\ \rho\sigma\tau,\ \xi'\rho'\sigma'\tau')V(\xi',\ \rho'\sigma'\tau')$$
$$\times \psi(\xi',\ \rho',\ \sigma',\ \tau')\,d\xi'\,d\rho'\,d\sigma'\,d\tau' \tag{60}$$

for the wavefunction of the system. As in Eq. (56), the wavefunction $\psi(\xi,\ \rho,\ \sigma,\ \tau)$ is separable and can in general be written

$$\psi(\xi,\ \rho,\ \sigma,\ \tau) = \sum_{\mu\nu\lambda} \psi_{\mu\nu\lambda}(\xi)R_\mu(\rho)S_\nu(\sigma)T_\lambda(\tau), \tag{61}$$

which allows the "external" wavefunction to be expressed as

$$\psi_{\mu\nu\lambda}(\xi) = \phi_{\mu'',\nu'',\lambda''}(\xi)\delta_{\mu\mu'',\nu\nu'',\lambda\lambda''}$$
$$+ \sum_{\mu'\nu'\lambda''} \int G^{(+)}_{\mu\mu\lambda}(\xi,\ \xi')\langle\mu\nu\lambda\ |V(\xi')|\ \mu'\nu'\lambda'\rangle$$
$$\times \psi_{\mu'\nu'\lambda'}(\xi')\,d\xi', \tag{62}$$

where

$$\langle\mu\nu\lambda\ |V(\xi')|\ \mu'\nu'\lambda'\rangle$$
$$= \int R_\mu(\rho')S_\nu(\sigma')T_\lambda(\tau')V(\xi',\ \rho',\ \sigma',\ \tau')$$
$$\times R^*_{\mu'}(\rho')S^*_{\nu'}(\sigma')T^*_{\lambda'}(\tau')\,d\rho'\,d\sigma'\,d\tau'. \tag{63}$$

Taking the asymptotic form of $G_E^{(+)}$ [Eq. (23)], we obtain for (63)

$$\psi_{\pi^i\mu\nu\lambda}(\xi) = \phi_{\mu'',\nu'',\lambda''}(\xi)\delta_{\mu\mu'',\nu\nu'',\lambda\lambda''}$$
$$+ [e^{-i\pi/4}k^{\frac{1}{2}}_{\mu\nu\lambda}e^{ik_{\mu\nu\lambda}}/(2\pi)^{\frac{3}{2}}\xi^{\frac{3}{2}}]f(\hat\pi^i_{\mu\nu\lambda},\ \hat\pi^0_{\mu''\nu''\lambda''}). \tag{64}$$

The scattering amplitude is then given by

$$f(\hat\pi^i_{\mu\nu\lambda},\ \hat\pi^0_{\mu'\nu'\lambda'}) = \frac{\mu}{(2\pi)^{\frac{3}{2}}}\int \exp(-ik_{\mu\nu\lambda}\xi'\alpha)$$
$$\times \langle\mu\nu\lambda\ |V(\xi')|\ \mu'\nu'\lambda'\rangle\psi_{\pi^i,\mu'\nu'\lambda'}(\xi')\,d\xi'. \tag{65}$$

If two of the particles, which we shall assume to be structureless, are bound together in the final state, we have a situation in which $|\pi_2|$ is imaginary, or equivalently

$$\chi = \tfrac{1}{2}\pi + i\alpha,$$

where α is real. The Green's function of system then takes the form

$$G_{E'}(\xi,\ \xi') = \lim_{\epsilon\to+0}\frac{2\mu}{(2\pi)^2}$$
$$\times \sum_n \int \frac{\exp[i\pi_2\cdot(\xi_2 - \xi_2')]}{k'^2 - k^2 + 2i\mu\epsilon}\phi_n(\xi_1)\phi^*_n(\xi_1')\,d\pi_2$$
$$= \lim_{\epsilon\to+0}\frac{2\mu}{(2\pi)^2}\sum_n \phi_n(\xi_1)\phi^*_n(\xi_1')$$
$$\times \int \frac{\exp(ik\cosh\alpha_n\,|\xi_2 - \xi_2'|\cos\phi_2)}{k^2 - k^2 + 2i\mu\epsilon}\,k\,dk\,d\phi_2$$
$$= [2\mu/(2\pi)^2]\sum_n \phi_n(\xi_1)\phi^*_n(\xi_1')\tfrac{1}{2}(\pi i)$$
$$\times H_0^{(1)}(k\cosh\alpha_n\,|\xi_2 - \xi_2'|), \tag{66}$$

where $\phi_n(\xi_1)$ are the energy eigenfunctions of the bound (12) pair. Because the energy spectrum of this pair is discrete, the "angles" α_n can assume certain allowed values only, which are related to the internal energy of the pair E_{12} by the equality

$$E^n_{12} = -(1/2\mu)k^2\sinh^2\alpha_n. \tag{67}$$

In the asymptotic region, $G^{(+)}_{E',n}$ takes the form

$$G^{(+)}_{E',n}(\xi,\ \xi') = \frac{\mu e^{i\pi/4}}{(2\pi)^{\frac{3}{2}}}\phi_n(\xi_1)\phi^*_n(\xi_1')$$
$$\times \frac{\exp(ik_n\,|\xi_2 - \xi_2'|)}{k_n^{\frac{1}{2}}\,|\xi_2 - \xi_2'|^{\frac{1}{2}}}, \tag{68}$$

where $k_n \equiv k\cosh\alpha_n$. Using Eqs. (15) and (68) one can immediately write down the wavefunction of the system:

$$\psi_{\pi^i}(\xi) = \sum_n \frac{e^{i\pi/4}e^{ik_n}}{(k_n\xi_2)^{\frac{1}{2}}}\phi_n(\xi_1)f(\hat\pi^0_n,\ \hat\pi^i), \tag{69}$$

where

$$f(\hat\pi^0_n,\ \hat\pi^i) = \frac{\mu}{(2\pi)^{\frac{3}{2}}}$$
$$\times \int \exp(-i\pi_{2n}\cdot\xi_2)\phi_n(\xi_1)V(\xi)\psi_{\pi^i}(\xi)\,d\xi \tag{70}$$

is the three-body scattering amplitude corresponding to the rearrangement process. This treatment can be readily generalized to the case of particles with internal structure.

Computation of the inelastic cross section in terms of the scattering amplitude (65), by the methods of Sec. 4, yields

$$\sigma_{\pi_n^0,\ \pi_{n'}^i}(n' \to n) = (k_n^2/k_{n'})\,|f(\hat\pi^0_n,\ \hat\pi^i_{n'})|^2, \tag{71}$$

where $k_{n'}$ represents the momentum of the incident particles, and n' and n represent the quantum numbers of the internal states of the three incident particles before and after scattering, respectively.

The cross section for scattering into a given exit channel can also be expressed in terms of the three-particle scattering matrix which connects that exit channel with the entrance channel. Normalizing the wavefunction of the system to unit flux in the entrance channel we can write the hyperradial part as[12]

$$R_{\gamma\gamma'}(\xi) = \left(\frac{2\pi}{k_\gamma \xi}\right)^{\frac{1}{2}} \frac{1}{(V_{\gamma'})^{\frac{1}{2}}} \times \left\{\delta_{\gamma\gamma'} \exp\left[-i\, k_{\gamma'}\xi - \frac{2\lambda+1}{4}\pi\right)\right] - S_{\gamma\gamma'} \exp\left[i\left(k_{\gamma'}\xi - \frac{2\lambda+1}{4}\pi\right)\right]\right\}, \quad (72)$$

the angular part as

$$F_{\gamma'}(\bar{\chi}\bar{\phi}_1\bar{\phi}_2;\, \chi\phi_1\phi_2) = \mathcal{J}^*_{\lambda_\gamma, m_{1\gamma}, m_{2\gamma}}(\bar{\chi}\bar{\phi}_1\bar{\phi}_2)\mathcal{J}_{\lambda_{\gamma'}, m_{1\gamma'}, m_{2\gamma'}}(\chi\phi_1\phi_2)$$

in the $(\lambda m_1 m_2)$ representation, and the internal part as $I(\rho, \sigma, \tau)$, where ρ, σ, and τ are the internal coordinates of the three particles. The total flux in the outgoing channel is obtained by squaring the modulus of the complete wavefunction RFI and integrating over the hypersolid angles $\bar{\Omega}$ and the internal coordinates of the particles:

$$\text{flux out} = \int |R_{\gamma\gamma'}(\xi)|^2\, |F_{\gamma'}(\bar{\chi}\bar{\phi}_1\bar{\phi}_2;\, \chi\phi_1\phi_2)|^2 \times |I(\rho\sigma\tau)|^2\, d\bar{\Omega}\, d\Omega v_{\gamma'} \xi^3\, d\rho\, d\sigma\, d\tau = (2\pi/k)^3\, |\delta_{\gamma\gamma'} - S_{\gamma\gamma'}|^2, \quad (73)$$

where $v_{\gamma'}$ is the hypervelocity of the particles in the exit channel γ'. Using the definition of cross section (40) and the fact that the entrance flux is unity, the cross section for exit channel γ' can be expressed as

$$\sigma(\gamma \to \gamma') = (2\pi/k)^3\, |\delta_{\gamma\gamma'} - S_{\gamma\gamma'}|^2 \quad (74)$$

For the case of scattering by a "central" potential (Sec. 4) the phase shifts η_λ can be related to the diagonal elements of the scattering matrix by the familiar relation

$$S_{\lambda\lambda} = e^{2i\eta_\lambda}, \quad (75)$$

where η_λ is in general a complex quantity, the real part of which is associated with elastic scattering and the imaginary part with inelastic scattering.

[12] See, for example, J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (John Wiley & Sons, Inc., New York, 1952), p. 519.

6. IDENTICAL PARTICLES

In the foregoing treatment of three-particle scattering we have considered distinguishable particles only. If we are to extend the case to indistinguishable particles (fermions or bosons), the wavefunctions must be properly symmetrized:

$$\psi = (6)^{-\frac{1}{2}} \sum_{\mathcal{P}} \zeta_{\mathcal{P}} \mathcal{P} \times \{\exp[i(\mathbf{p}^1\cdot\mathbf{x}^1 + \mathbf{p}^2\cdot\mathbf{x}^2 + \mathbf{p}^3\cdot\mathbf{x}^3)]\}, \quad (76)$$

where $\mathbf{p}^i$ and $\mathbf{x}^i$ are the momentum and position of particle i in the laboratory system, $\mathcal{P}$ is the particle permutation operator and $\zeta_{\mathcal{P}}$ is described by

$$\zeta_{\mathcal{P}} = 1 \text{ bosons},$$

$$\zeta_{\mathcal{P}} = \left.\begin{array}{l} -1 \text{ for an odd number of permutations} \\ +1 \text{ for an even number of permutations} \end{array}\right\} \text{fermions.}$$

If we transform to the $\xi - \pi$ coordinate system of Eq. (7), (73) then takes the form

$$\begin{aligned}\psi = \exp(i\mathbf{P}\cdot\mathbf{X})(6)^{-\frac{1}{2}}\{&\exp[i(\pi^1\cdot\xi^1 + \pi^2\cdot\xi^2)] \\ &+ \eta \exp[\tfrac{1}{2}i(\pi^1\cdot\xi^1 - \pi^2\cdot\xi^2 - \sqrt{3}\,\pi^1\cdot\xi^2 - \sqrt{3}\,\pi^2\cdot\xi^1)] \\ &+ \eta \exp[\tfrac{1}{2}i(\pi^1\cdot\xi^1 - \pi^2\cdot\xi^2 + \sqrt{3}\,\pi^1\cdot\xi^2 + \sqrt{3}\,\pi^2\cdot\xi^1)] \\ &+ \eta \exp[i(-\pi^1\cdot\xi^1 + \pi^2\cdot\xi^2)] \\ &+ \exp[\tfrac{1}{2}i(-\pi^1\cdot\xi^1 - \pi^2\cdot\xi^2 + \sqrt{3}\,\pi^1\cdot\xi^2 - \sqrt{3}\,\pi^2\cdot\xi^1)] \\ &+ \exp[\tfrac{1}{2}i(-\pi^1\cdot\xi^1 - \pi^2\cdot\xi^2 - \sqrt{3}\,\pi^1\cdot\xi^2 + \sqrt{3}\,\pi^2\cdot\xi^1)]\}, \end{aligned} \quad (77)$$

where $\mathbf{P}$ is the momentum and $\mathbf{X}$ the position of the center of mass of the three particles; $\eta = -1$ for fermions and $+1$ for bosons. One can now compute the free-particle Green's function just as in Sec. 3, obtaining 6 terms instead of one:

$$G_k^{(+)}(\xi, \xi') = \frac{i\mu k}{4\pi} \times \left[\frac{H_1^{(1)}(kr_1)}{r_1} + \eta\frac{H_1^{(1)}(kr_2)}{r_2} + \eta\frac{H_1^{(1)}(kr_3)}{r_3} + \eta\frac{H_1^{(1)}(kr_4)}{r_4} + \frac{H_1^{(1)}(kr_5)}{r_5} + \frac{H_1^{(1)}(kr_6)}{r_6}\right], \quad (78)$$

where the r_i are related to ξ^1 and ξ^2 as follows:

$$r_1^2 = \xi^2 + \xi'^2 - 2(\xi^1\cdot\xi'^1 + \xi^2\cdot\xi'^2),$$
$$r_2^2 = \xi^2 + \xi'^2 - (\xi^1\cdot\xi'^1 - \xi^2\cdot\xi'^2 - \sqrt{3}\,\xi^1\cdot\xi'^2 - \sqrt{3}\,\xi^2\cdot\xi'^1),$$
$$r_3^2 = \xi^2 + \xi'^2 - (\xi^1\cdot\xi'^1 - \xi^2\cdot\xi'^2 + \sqrt{3}\,\xi^1\cdot\xi'^2 - \sqrt{3}\,\xi^2\cdot\xi'^1),$$
$$r_4^2 = \xi^2 + \xi'^2 - 2(-\xi^1\cdot\xi'^1 + \xi'\cdot\xi'^2),$$
$$r_5^2 = \xi^2 + \xi'^2 - (-\xi^1\cdot\xi'^1 - \xi^2\cdot\xi'^2 + \sqrt{3}\,\xi^1\cdot\xi'^2 - \sqrt{3}\,\xi^2\cdot\xi'^1),$$
$$r_6^2 = \xi^2 + \xi'^2 - (-\xi^1\cdot\xi'^1 - \xi^2\cdot\xi'^2 - \sqrt{3}\,\xi^1\cdot\xi'^2 + \sqrt{3}\,\xi^2\cdot\xi'^1).$$

Computation of the scattering amplitude for the system leads to

$$f(\bar{\chi}\bar{\phi}_1\bar{\phi}_2, \chi\phi_1\phi_2) = \sum_{i=1}^{6} \eta^{(i)} f^{(i)}(\bar{\chi}\bar{\phi}_1\bar{\phi}_2, \chi\phi_1\phi_2), \qquad (79)$$

in which $\eta^{(i)} = +1$ for $i = 1$, 5, and 6; $\eta^{(i)} = -1$ for $i = 2$, 3, and 4 in the case of fermions; $\eta^{(i)} = 1$ for bosons. The different component scattering amplitudes $f^{(i)}$ are of the same form as (29) with the coefficient α of the product $k\xi$ which appears in the exponential factor $e^{-ik\xi'\alpha}$ assuming the values α_i.

$$\alpha_1 = \cos\chi\cos\bar{\chi}\cos(\phi_1 - \bar{\phi}_1) + \sin\chi\sin\bar{\chi}\cos(\phi_2 - \bar{\phi}_2),$$
$$\alpha_2 = \tfrac{1}{2}\cos\chi\cos\bar{\chi}\cos(\phi_1 - \bar{\phi}_1) - \tfrac{1}{2}\sin\chi\sin\bar{\chi}\cos(\phi_2 - \bar{\phi}_2) - \tfrac{1}{2}\sqrt{3}\cos\chi\sin\bar{\chi}\cos(\phi_1 - \bar{\phi}_2) - \tfrac{1}{2}\sqrt{3}\sin\chi\cos\bar{\chi}\cos(\phi_2 - \bar{\phi}_1),$$
$$\alpha_3 = \tfrac{1}{2}\cos\chi\cos\bar{\chi}\cos(\phi_1 - \bar{\phi}_1) - \tfrac{1}{2}\sin\chi\sin\bar{\chi}\cos(\phi_2 - \bar{\phi}_2) + \tfrac{1}{2}\sqrt{3}\cos\chi\sin\bar{\chi}\cos(\phi_1 - \bar{\phi}_2) + \tfrac{1}{2}\sqrt{3}\sin\chi\cos\bar{\chi}\cos(\phi_2 - \bar{\phi}_1),$$
$$\alpha_4 = -\cos\chi\cos\bar{\chi}\cos(\phi_1 - \bar{\phi}_1) + \sin\chi\sin\bar{\chi}\cos(\phi_2 - \bar{\phi}_2),$$
$$\alpha_5 = -\tfrac{1}{2}\cos\chi\cos\bar{\chi}\cos(\phi_1 - \bar{\phi}_1) - \tfrac{1}{2}\sin\chi\sin\bar{\chi}\cos(\phi_2 - \bar{\phi}_2) + \tfrac{1}{2}\sqrt{3}\cos\chi\sin\bar{\chi}\cos(\phi_1 - \bar{\phi}_2) - \tfrac{1}{2}\sqrt{3}\sin\chi\cos\bar{\chi}\cos(\phi_2 - \bar{\phi}_1),$$

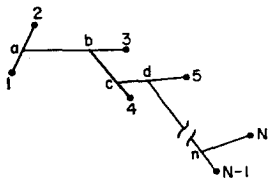


FIG. 2. Coupling of N particles with the aid of the coordinate system given in (A1); a denotes the center of mass of particles 1 and 2, b that of particles 1, 2, and 3. $\cdots$, n that of particles 1 to $N-1$.

$$\alpha_6 = -\tfrac{1}{2}\cos\chi\cos\bar{\chi}\cos(\phi_1 - \bar{\phi}_1) - \tfrac{1}{2}\sin\chi\sin\bar{\chi}\cos(\phi_2 - \bar{\phi}_2) - \tfrac{1}{2}\sqrt{3}\cos\chi\sin\bar{\chi}\cos(\phi_1 - \bar{\phi}_2) + \tfrac{1}{2}\sqrt{3}\sin\chi\cos\bar{\chi}\cos(\phi_2 - \bar{\phi}_1), \qquad (80)$$

corresponding to the various $f^{(i)}$. The relation between the cross section and the scattering amplitude (43) can easily be shown to be correct if we define f by Eq. (79).

If only two of the particles are identical, expression (80) contains only 2 terms in the right-hand member and these are characterized by

$$\alpha_1 = \cos\chi\cos\bar{\chi}\cos(\phi_1 - \bar{\phi}_1) + \sin\chi\sin\bar{\chi}\cos(\phi_2 - \bar{\phi}_2),$$
$$\alpha_2 = -\cos\chi\cos\bar{\chi}\cos(\phi_1 - \bar{\phi}_1) + \sin\chi\sin\bar{\chi}\cos(\phi_2 - \bar{\phi}_2), \qquad (81)$$

ACKNOWLEDGMENTS

The author is indebted to Drs. Felix Smith, Michael Stallybrass, and Joel Brenner for many helpful discussions.

APPENDIX A. N-BODY SCATTERING IN A PLANE

The free-particle Green's function obtained for three bodies in Sec. 3 can easily be generalized to N bodies by using a (normalized) coordinate system corresponding to the coupling scheme shown in Fig. 2.

The coordinate system is a simple extension of Smith's asymmetric one [Eq. (7)]:

$$\begin{aligned}
\xi_1^1 &= \rho\cos\chi_{N-2}\cos\chi_{N-3}\cdots\cos\chi_1\cos\phi_1,\\
\xi_2^1 &= \rho\cos\chi_{N-2}\cdots\cos\chi_1\sin\phi_1,\\
\xi_1^2 &= \rho\cos\chi_{N-2}\cdots\sin\chi_1\cos\phi_2,\\
\xi_2^2 &= \rho\cos\chi_{N-2}\cdots\sin\chi_1\sin\phi_2,\\
\xi_1^3 &= \rho\cos\chi_{N-2}\cdots\sin\chi_2\cos\phi_3,\\
\xi_2^3 &= \rho\cos\chi_{N-2}\cdots\sin\chi_2\sin\phi_3,\\
&\vdots\qquad\qquad\vdots\\
\xi_1^{N-1} &= \rho\sin\chi_{N-2}\cos\phi_{N-1},\\
\xi_2^{N-1} &= \rho\sin\chi_{N-2}\sin\phi_{N-1},
\end{aligned} \qquad (A1)$$

where

$$\rho^2 = \sum_{i=1}^{N-1} [(\xi_1^i)^2 + (\xi_2^i)^2],$$

and

$$0 \leq \chi_i \leq \tfrac{1}{2}\pi, \qquad 0 \leq \phi_i \leq 2\pi.$$

The N-body plane wavefunctions in the center of mass of the assembly are of the form

$$\psi_i(\xi) = [1/(2\pi)^{N-1}] \times \exp [i(\pi^1 \cdot \xi^1 + \pi^2 \cdot \xi^2 + \cdots \pi^{N-1} \cdot \xi^{N-1})] \quad \text{(A2)}$$

and the Green's function is accordingly generalized to

$$G_{k_\alpha}^{(+)}(\xi, \xi') = \lim_{\epsilon \to +0} \frac{2\mu}{(2\pi)^{2N-2}} \times \int \frac{\exp\left[i \sum_{i=1}^{N-1} \pi^i \cdot (\xi^i - \xi'^i) \right]}{k_{\alpha^2} - k^2 + 2i\mu\epsilon} d\pi, \quad \text{(A3)}$$

where

$$k^2 = \sum_{i=1}^{N-1} |\pi^i|^2.$$

Introducing the coordinate system (A1) and integrating over the "azimuthal" angles ϕ_i, we obtain

$$G_{k_\alpha}^{(+)}(\xi, \xi') = \lim_{\epsilon \to +0} \frac{2\mu}{(2\pi)^{(N-1)}} \times \int_0^{\frac{1}{2}\pi} \cdots \int_0^{\frac{1}{2}\pi} \int_0^{\infty} \frac{\prod_{i=1}^{N-2} J_0\left(k a_i \sin \chi_{i-1} \prod_{j=1}^{N-2} \cos \chi_j\right)}{k_\alpha^2 - k^2 + 2i\mu\epsilon} \times \sin \chi_1 \cos \chi_1 \cdots \sin \chi_{N-2} \times \cos^{2N-5} \chi_{N-2}\, d\chi_1 \cdots d\chi_{N-2} k^{2N-5}\, dk, \quad \text{(A4)}$$

where

$$a_i = (|\xi^i|^2 + |\xi'^i|^2 - 2\xi^i \cdot \xi'^i)^{\frac{1}{2}}$$

and $\chi_0 = \frac{1}{2}\pi$.

The integration over the magnitude of the momentum is carried out as in Sec. 2 and

$$G_k^{(+)}(\xi, \xi') = \tfrac{1}{2} i\mu \frac{(-1)^{N-2}}{(2\pi)^{N-2}} \times \left(a^{-1} \frac{\partial}{\partial a}\right)^{N-2} H_0^{(1)}(ka), \quad \text{(A5)}$$

where

$$a = \sum_{i=1}^{N-1} (a^i)^2 = |\xi - \xi'|.$$

Asymptotically, this takes the form

$$G_k^{(+)}(\xi, \xi') \sim e^{-(i\pi/4)(1+2N)} \frac{\mu k^{N-5/2}}{(2\pi)^{N-\frac{1}{2}}} \frac{e^{ika}}{a^{N-\frac{1}{2}}}, \quad \text{(A6)}$$

and the wavefunction of the system can then be expressed as

$$\psi(\xi) = \phi(\xi) + \frac{e^{i\pi(5/4 - \frac{1}{2}N)}}{\xi^{N-\frac{1}{2}}} k^{N-5/2} e^{ik\xi} f(\hat{\pi}_0, \hat{\pi}_i), \quad \text{(A7)}$$

where $f(\hat{\pi}_0, \hat{\pi}_i)$, the scattering amplitude, is written

$$f(\hat{\pi}_0, \hat{\pi}_i) = \frac{\mu}{(2\pi)^{N-\frac{1}{2}}} \int e^{-ik\xi' \alpha} V(\xi') \psi_{\pi_i}(\xi')\, d\xi', \quad \text{(A8)}$$

and

$$\alpha = \sum_{i=0}^{N-2} \sin \chi_i \sin \chi_i' \cos (\phi_{i+1} - \phi_{i+1}') \times \left[\prod_{k=i+1}^{N-2} \cos \chi_k \cos \chi_k' \right]. \quad \text{(A9)}$$

Using the N-body analogue of (40), the cross section for N particles is related to the scattering amplitude by

$$\sigma(\mathbf{k}, \mathbf{k}_0) = (k^{2N-4}/k_0) |f(\hat{\pi}_0, \hat{\pi}_i)|^2. \quad \text{(A10)}$$

APPENDIX B. PLANE WAVE EXPANSION

In Sec. 3, the expansion of the three-particle plane wave $\exp(i\pi \cdot \xi)$ in radial and momentum eigenfunctions proved to be a useful mathematical device:

$$\exp(i\pi \cdot \xi) = \sum_{\lambda m_1 m_2} C_{\lambda m_1 m_2} \frac{J_{\lambda+1}(k\xi)}{k\xi} \mathcal{J}^*_{\lambda m_1 m_2}(\hat{\pi}) \mathcal{J}_{\lambda m_1 m_2}(\hat{\xi}). \quad \text{(B1)}$$

If one multiplies Eq. (5) by $\mathcal{J}^*_{\lambda' m_1' m_2'}(\hat{\xi})\, \mathcal{J}_{\lambda' m_1' m_2'}(\hat{\pi})$ and integrates over all possible directions of $\hat{\pi}$ and $\hat{\xi}$, one obtains, from the orthonormality properties of the $\mathcal{J}$'s,

$$C_{\lambda m_1 m_2} \frac{J_{\lambda+1}(k\xi)}{k\xi} = \iint \exp(i\pi \cdot \xi) \mathcal{J}^*_{\lambda m_1 m_2}(\hat{\xi}) \mathcal{J}_{\lambda m_1 m_2}(\hat{\pi})\, d\Omega_\xi\, d\Omega_\pi. \quad \text{(B2)}$$

We now differentiate Eq. (B2) λ times with respect to $k\xi$ and set $k\xi = 0$; obtaining

$$C_{\lambda m_1 m_2} = (2)^{\lambda+1}(\lambda + 1)(i)^\lambda \iint [\cos \chi \cos \bar{\chi} \times \cos (\phi_1 - \bar{\phi}_1) + \sin \chi \sin \bar{\chi} \cos (\phi_2 - \bar{\phi}_2)]^\lambda \times \mathcal{J}^*_{\lambda m_1 m_2}(\chi \phi_1 \phi_2) \mathcal{J}_{\lambda m_1 m_2}(\bar{\chi} \bar{\phi}_1 \bar{\phi}_2)\, d\bar{\Omega}\, d\Omega, \quad \text{(B3)}$$

where $d\Omega$ is an element of hypersolid angle $d\Omega = \frac{1}{2} \sin 2\chi\, d\chi\, d\phi_1\, d\phi_2$ ($0 \leq \chi \leq \frac{1}{2}\pi$, $0 \leq \phi_1, \phi_2 \leq 2\pi$). Integration of (B3) yields

$$C_{\lambda m_1 m_2} = (2\pi)^2(\lambda+1)^2(i)^\lambda \lambda!$$
$$\times \sum_{\nu=|m_2|}^{\lambda-|m_1|} \left[\frac{[\tfrac{1}{2}(\lambda-m_1+m_2)]!\,[\tfrac{1}{2}(\lambda+m_1-m_2)]!\,[\tfrac{1}{2}(\lambda-m_1 m_2)]!\,[\tfrac{1}{2}(\lambda+m_1+m_2)]!}{[\tfrac{1}{2}(\lambda-m_1-\nu)]!\,[\tfrac{1}{2}(\lambda+m_1-\nu)]!\,[\tfrac{1}{2}(\nu-m_2)]!\,[\tfrac{1}{2}(\nu+m_2)]!}\right]$$
$$\times \left\{\sum_{\kappa=0}^{\frac{1}{2}(\lambda-m_1-m_2)} (-1)^\kappa \{\kappa!\,(m_1+\kappa)!\,[\tfrac{1}{2}(\lambda-m_1+m_2)-\kappa]!\,[\tfrac{1}{2}(\lambda-m_1-m_2)-\kappa]!\}^{-1}\right.$$
$$\left.\times B[\tfrac{1}{2}(\lambda-m_1-2\kappa+\nu+2),\,\tfrac{1}{2}(\nu+m_1+2\kappa-\nu+2)]\right\}^2, \tag{B4}$$

where $B(r, s)$ is the beta function. It was found by inserting various allowed values of λ, m_1, and m_2 into (B4) that, in each case, $C_{\lambda m_1 m_2}$ was independent of m_1 and m_2 and equal to $(2\pi)^2(i)^\lambda$. Proof of this in the general case has so far eluded the author, however. Explicitly,

$$e^{i\pi\cdot\xi} = (2\pi)^2 \sum_\lambda (i)^\lambda$$
$$\times \frac{J_{\lambda+1}(k\xi)}{k\xi} \sum_{m_1 m_2} \mathcal{J}^*_{\lambda m_1 m_2}(\hat{\pi})\mathcal{J}_{\lambda m_1 m_2}(\hat{\xi}). \tag{B5}$$

Note added in proof. Since submission of this paper for publication, Dr. Edward Gerjuoy has pointed out to the author that many years ago Sommerfeld derived the Green's function for a many-dimensional space[13] by a simpler method than that presented here. As applied to many-particle scattering, it involves a spacial partitioning which is different from ours, i.e. Sommerfeld's computation is carried out in the space of *one* of the particles. In spite of its more complicated form, the author, however, feels that the method of computation presented in the foregoing is somewhat preferable for our purposes, particularly for the derivation of Eqs. (25) and (26).

[13] E. Gerjuoy, Ann. Phys. 5, 58 (1958); A. Sommerfeld, *Partial Differential Equations in Physics* (Academic Press Inc., New York, 1949).

Asymptotic Form of the Proper Self-Energy Function in Many-Particle Systems*

DONALD R. FREDKIN
Department of Physics, University of California, La Jolla, California
(Received 11 January 1963)

The asymptotic form at high frequencies of the reciprocal of a two-time thermodynamic Green's function is shown to be a constant. This constant is expressed as the exact expectation value of a second (functional) derivative of the Hamiltonian.

IT is possible to obtain information about the asymptotic behavior of thermodynamic Green's functions from their spectral representations. The coefficients in such asymptotic formulas are given by moments of the spectral density functions and, therefore, are often exactly known from sum rules. In this note, attention is turned to the high-frequency behavior of the reciprocal of (the Fourier transform of) a Green's function. A general formula is derived which, when applied to the single-particle Green's function, shows that the exact self-energy function is the sum of an "exact Hartree-Fock" Hamiltonian, and a function which approaches zero at infinite frequency at least as fast as the reciprocal of the frequency, and which therefore satisfies an unsubtracted dispersion relation. Because the proof is rigorous, it applies to both normal and superfluid systems; in particular, the approximate single-particle Green's function of the BCS theory of superconductivity,[1] as derived by Gorkov,[2] satisfies the general theorem.

If A and B are any two operators, we define the Green's function $G_{AB}(t) = \langle (A(t)\ B(0))_+ \rangle$. Here $(\ )_+$ means time ordering in the sense of Wick,[3] the operators are taken in the Heisenberg picture, and $\langle\ \rangle$ means averaging in a canonical or grand canonical ensemble. (If A and B do not each conserve the number of some type of particle whose number is, in fact, conserved in the temporal evolution of the system, the ensemble must be taken to be grand canonical with respect to that species of particle. It is assumed that all "μN" terms are incorporated in the Hamiltonian for purposes of defining the Heisenberg picture and defining the energy.) The Fourier transform $G_{AB}(\omega)$ is defined by

$$G_{AB}(\omega) = \int dt e^{i\omega t} G_{AB}(t);$$

all integrations are understood to be from $-\infty$ to $+\infty$ unless specifically indicated to the contrary. It is well known[4] that $G_{AB}(\omega)$ has the spectral representation

$$G_{AB}(\omega) = i \int \frac{d\omega'}{2\pi} \rho_{AB}(\omega') \times \left[\frac{\mp f(\omega')}{\omega - \omega' - i\epsilon} + \frac{1 \pm f(\omega')}{\omega - \omega' + i\epsilon} \right], \quad (1)$$

where

$$\rho_{AB}(\omega) = \int dt e^{i\omega t} \langle [A(t), B(0)]_\mp \rangle; \quad (2)$$

the lower sign refers to the case of A and B being both Fermion-type operators, the upper sign refers to all other cases, and $f(\omega) = (e^{\beta\omega} \mp 1)^{-1}$. It is convenient, though illogical, also to denote by $G_{AB}(\omega)$, the function of complex ω defined by

$$G_{AB}(\omega) = i \int \frac{d\omega'}{2\pi} \frac{\rho_{AB}(\omega')}{\omega - \omega'}. \quad (3)$$

It is an elementary mathematical exercise to show that, if $\int \rho(\omega)\, d\omega/2\pi$ exists (as an absolutely convergent Riemann integral or, generally, as a Lebesgue integral), then $\int \omega'(\omega - \omega')^{-1}\rho(\omega')\, d\omega/2\pi$ is of order $o(\omega^{-1})$ as $|\omega| \to \infty$. As a consequence, if the first n moments $M_\nu = \int \omega^\nu \rho_{AB}(\omega)\, d\omega/2\pi$ $(\nu \geq 0)$ exist,

$$G_{AB}(\omega) = i \sum_{\nu=0}^{n-1} \frac{M_\nu}{\omega^{\nu+1}} + o(\omega^{-n}). \quad (4)$$

This result, together with the relations

$$M_\nu = [i(\partial/\partial t)]^\nu \langle [A(t), B(0)]_\mp \rangle|_{t=0}, \quad (5)$$

which follow from (2), are hardly novel.[5] However,

* Research supported by the U. S. Air Force Office of Scientific Research.

[1] J. Bardeen, L. Cooper, and J. Schrieffer, Phys. Rev. 108, 1175 (1957).

[2] L. Gorkov, Soviet Physics—JETP 7, 505 (1958).

[3] G. C. Wick, Phys. Rev. 80, 268 (1950).

[4] L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (W. A. Benjamin, Inc., New York, 1962), and references cited therein.

[5] R. Kubo, in *Lectures in Theoretical Physics* (Interscience Publishers, Inc., New York, 1959), Vol. I, p. 152.

(4) can be rather unenlightening. For example, if A is the absorption operator for a free particle of energy ϵ, and $B = A^*$ is the corresponding emission operator, (4) and (5) yield

$$G_{AB}(\omega) \sim i \sum_{\nu=0}^{\infty} \epsilon^{\nu}/\omega^{\nu+1},$$

which is hardly as convenient as the exact expression $G_{AB}(\omega) = i(\omega - \epsilon)^{-1}$.

Taking the hint from the free-particle example and from experience with diagram summation in perturbation theory,[4,6] consider $iG_{AB}(\omega)^{-1}$. If $M_0 \neq 0$, then from (4),

$$iG_{AB}(\omega)^{-1} = \omega/M_0 - M_1/M_0^2 + O(\omega^{-1}). \quad (6)$$

(The existence of all moments M_ν may be assumed as needed; whether or not this is correct is a question of the finiteness of various renormalization constants,[7] which presumably does not arise in nonrelativistic problems.) Since $G_{AB}(\omega)$ has no complex zeros,[4] the remainder term in (6) is analytic, except on the real axis, and satisfies an unsubtracted dispersion relation. The relation (6) is the principal result of this paper; a simple extension of (6), and the application to single-particle Green's functions, will be given below. The case $M_0 = 0$, omitted above, does not lead to anything new. If M_r is the first nonvanishing moment, $iG_{AB}(\omega)^{-1} = P(\omega) + O(\omega^{-1})$, where $P(\omega)$ is a polynomial of degree $r + 1$ whose coefficients can be expressed in terms of $M_r, \cdots M_{2r+1}$ and, again, the remainder term satisfies an unsubtracted dispersion relation.

If $\{A_\alpha\}$ is a (possibly infinite) set of operators, it is natural to consider the matrix $G(\omega) = [G^{\alpha\beta}(\omega)]$, where $G^{\alpha\beta}(\omega) = G_{A_\alpha A_\beta^*}(\omega)$. The moments M_ν are now matrices $[M_\nu^{\alpha\beta}]$, and (4) is still true as a matrix equation. If M_0 is nonsingular, (6) takes the form

$$iG(\omega)^{-1} = \omega M_0^{-1} - M_0^{-1}M_1M_0^{-1} + O(\omega^{-1}); \quad (6')$$

once again the remainder term satisfies an unsubtracted dispersion relation.

For a many-particle system, the A_α may be taken to be the absorption operators for a complete orthonormal set of one-particle states. Then, because $[A_\alpha, A_\beta^*]_- = \delta_{\alpha\beta}$, M_0 is the unit matrix. It is customary to write the Hamiltonian as

$$\mathcal{H} = \sum_{\alpha,\beta} A_\alpha^* h^{\alpha\beta} A_\beta + (\text{interaction terms}), \quad (7)$$

with h a Hermitian matrix, and to define the proper self-energy part $\sum(\omega) = [\sum^{\alpha\beta}(\omega)]$ by

[6] J. M. Luttinger and J. C. Ward, Phys. Rev. 118, 1417 (1960).

[7] H. Lehman, Nuovo Cimento 11, 342 (1954).

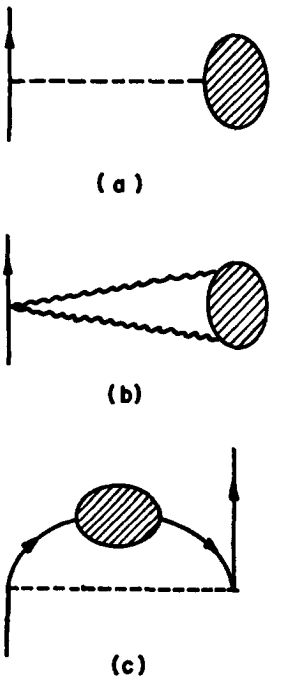


FIG. 1. Three examples of classes of Feynman graphs which contribute to $\langle \partial^2 V/\partial A^* \partial A\rangle$ for the electron–phonon system. Solid lines represent electron propagators, wavy lines phonon propagators, and dashed lines Coulomb interactions. Graphs of type (b) do not occur in the usual theory in which the Hamiltonian is linearized in the phonon "field."

$$iG(\omega)^{-1} = \omega - h - \sum(\omega).$$

From (6′),

$$h + \sum(\omega) = M_1 + O(\omega^{-1}). \quad (8)$$

Using (5), and the Heisenberg equation of motion $i\,\partial A(t)/\partial t = [A(t), \mathcal{H}]$,

$$M_1^{\alpha\beta} = \langle[[A_\alpha, \mathcal{H}], A_\beta^*]_-\rangle = \langle \partial^2\mathcal{H}/\partial A_\alpha^* \partial A_\beta\rangle. \quad (9)$$

The derivative in (8) is to be interpreted as a variational derivative if α is a continuous variable, and appropriate ordering conventions are to be made if fermion operators are involved. Denoting the interaction terms in (7) by V, (7), (8), and (9) imply

$$\sum(\omega) = \langle \partial^2 V/\partial A^* \partial A\rangle + O(\omega^{-1}), \quad (10)$$

or

$$\sum(\omega) = \langle \partial^2 V/\partial A^* \partial A\rangle + \int \frac{d\omega'}{\pi} \frac{\operatorname{Im}\sum(\omega')}{\omega' - \omega}. \quad (10')$$

The first term on the right-hand side of (10′) is formally similar to a Hartree–Fock term; in fact, the unrestricted Hartree–Fock approximation consists of neglecting the dispersion integral and computing the mean value in a manner consistent with this approximation, instead of exactly as in (10′). Figure 1 illustrates some contributions to the mean value for the electron–phonon system. They all have the property that they have no absorptive part according to Langer's "generalized unitarity"

criterion;[8] this is compatible with (4), which, it is easily seen, gives information only about the dispersive part of $G_{AB}(\omega)$.

It is worthy of explicit mention that, if a calculation is made which is more elaborate than the unrestricted Hartree–Fock method, the dispersion integral cannot be neglected (except, of course, in the high-frequency limit). It is essential that the approximations to the two terms in (10′) be made consistently. For this reason, it is not clear what, if any, use can be made of (10) or (10′). In particular, it cannot be concluded from (10) and the success of the nuclear shell model that, at ultra-high energies, the elastic scattering of nucleons by nuclei can be deduced from the shell-model potential. In fact, the shell-model potential contains a substantial contribution from the dispersion integral in (10′) which must approximately make up the difference between the exact value of $\langle \partial^2 V/\partial A^* \, \partial A \rangle$ and the Hartree–Fock value.

Since (10) is alleged to be an exact result, it should hold for the case of a superconductor. Gorkov found,[2] for α labeling an electron momentum and spin state,

$$G^{\alpha\beta}(\omega) = i[(\omega + \epsilon'_\alpha)/(\omega^2 - \epsilon'^2_\alpha - |\Delta_\alpha|^2)]\delta_{\alpha\beta},$$

where $\epsilon'_\alpha = h^{\alpha\alpha} + \langle \partial^2 V/\partial A^*_\alpha \, \partial A_\alpha \rangle$, the average being taken with respect to the BCS density matrix, and $|\Delta_\alpha|^2$ is the energy-gap parameter. Therefore, in this case,

$$\sum^{\alpha\beta}(\omega) = \delta_{\alpha\beta}[\langle \partial^2 V/\partial A^*_\alpha \, \partial A_\alpha \rangle + |\Delta_\alpha|^2/(\omega + \epsilon'_\alpha)],$$

in agreement with (10) and (10′).

[8] J. S. Langer, Phys. Rev. **124**, 997 (1961).

Representations of the Canonical Commutation Relations Describing a Nonrelativistic Infinite Free Bose Gas*†

H. ARAKI‡
Department of Physics, University of Illinois, Urbana, Illinois
AND
E. J. WOODS§
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey, and Theoretical Physics Institute, University of Alberta, Edmonton, Alberta, Canada
(Received 7 January 1963)

The existence of inequivalent representations of the canonical commutation relations which describe a nonrelativistic infinite free Bose gas of uniform density is investigated, with a view to possible applications to statistical mechanics. The functional $E(f, g) = (\Psi, e^{i\phi(f)} e^{i\pi(g)}\Psi)$ is used to describe the inequivalent representations. This functional is calculated for the free Bose gas in a box of volume V, and the limit is then taken as $V \to \infty$. In this way we construct cyclic representations describing an infinite system of particles with a density distribution $\rho(k)$ in momentum space. For a given $\rho(k)$ the operator algebra generated by the $\phi(f)$, $\pi(g)$ is reducible. For the ground-state representation (all particles in the zero-momentum state), the representation is a direct integral of irreducible representations (analogous to BCS theory). For finite temperatures the situation is complicated by the occurrence of representations which are not type I. The physical significance of the reducibility of the representations is discussed.

It is argued that the thermal ensemble for the infinite system is a pure state at zero temperature, although there is some ambiguity as to which operators should belong to the algebra of observables. For finite temperatures, the thermal ensemble seems to be a mixture. The case of an interacting Bose gas is considered briefly.

1. INTRODUCTION

THIS paper is an investigation of inequivalent representations of the canonical commutation relations (hereafter referred to as CCR's) in nonrelativistic field theory which are suitable for describing an infinite free Bose gas of finite density. We are interested in the possibility of an alternative discussion of the thermodynamics of an infinite Bose gas. The conventional approach to the problem is to compute the partition function for $N = \rho V$ particles in a box of volume V, where ρ is the density. The equation of state is obtained from the partition function by differentiation, and one can then take the limit as $N, V \to \infty$ with $N/V = \rho$.

An alternative possibility, and the one to which this paper is devoted, is to begin with an actually infinite system, thereby avoiding the limiting procedure. This approach is a natural one if one seeks to understand the physical significance of the so-called strange representations of the CCR's which describe an infinite number of particles.

Current treatments of the many body problem use the second quantized formalism for the Schrödinger equation. For a system of particles with two-body interactions described by a potential $V(r)$, the Hamiltonian is given by

$$H = (2m)^{-1} \int d\mathbf{x}(\nabla\psi^*)\cdot(\nabla\psi) + \frac{1}{2}\int d\mathbf{x} \int d\mathbf{y}\psi^*(\mathbf{x})\psi^*(\mathbf{y})V(\mathbf{x}-\mathbf{y})\psi(\mathbf{x})\psi(\mathbf{y}), \quad (1.1)$$

where $\psi(\mathbf{x})$, $\psi^*(\mathbf{x})$ are, respectively, the annihilation and creation operators for a particle at the point $\mathbf{x}$.

If one considers a system of uniform density in a box of finite volume V, then the total number of particles is finite. Because one is dealing with finitely many particles, it is possible to realize explicitly the Hilbert space and operators $\psi(\mathbf{x})$, $\psi^*(\mathbf{x})$ by using a configuration space representation (the Fock representation). In the case of an infinite system with infinitely many particles, it is not *a priori* obvious that any suitable representation exists at all. The desired representation would have to be Euclidean invariant, the energy operator for a finite region of space would have to exist, etc.

There are, however, a number of reasons for wanting to consider an actually infinite system. In the first place, thermodynamics is concerned only with the limiting values of physical quantities as the volume of the box becomes infinite. By dealing with an infinite system to begin with, this limiting

* Supported in part by the National Science Foundation and the National Research Council of Canada.
† An earlier version of part of the present work was submitted by one of the authors (E. J. W.) to Princeton University as a Ph.D. thesis.
‡ On leave of absence from the Department of Nuclear Engineering, Kyoto University, Kyoto, Japan.
§ Now at University of Alberta, Edmonton, Alberta, Canada.

procedure could be eliminated from the theory. In fact this limiting process $V \to \infty$ does cause practical difficulties. For example, if one calculates the equation of state by using the expansion

$$\frac{1}{V} \log \mathcal{Q}_V = \sum_{j=1}^{\infty} b_j(V) y^j, \tag{1.2}$$

where $\mathcal{Q}_V$ denotes the grand partition function of the gas in the volume V, and y is the fugacity, then it is not necessarily legitimate to interchange the summation with the limiting process $V \to \infty$. Indeed, Yang and Lee[1] have shown that this accounts for the failure of Mayer's theory[2] to give the equation of state beyond the condensation point.

A second reason for considering a system in an infinite universe rather than a finite box is that one could exploit the Euclidean invariance of the theory from the beginning. In this respect it should be pointed out that in relativistic quantum field theory, it has been very fruitful to exploit the Lorentz invariance of the theory (dispersion relations, PCT theorem, etc.).

It is also worth noting that the BCS model of superconductivity is exactly soluble only in the limit of an infinite volume.

These remarks suggest that one might obtain new insight into the many-body problem by using the idealization of an actually infinite system. There is of course nothing wrong with the box method. Indeed to avoid ambiguity we must impose the requirement that the infinite system is, in some suitable sense, the limit of the finite system.

The first problem is whether any infinite systems exist at all. By this we mean: does there exist a Hilbert space H and dynamical variables in H such that expectation values of physical quantities in H are the limits of the corresponding quantities in the box? More explicitly, does there exist an operator field $\psi(\mathbf{x})$ defined over an infinite Euclidean space, and a unitary representation $U(\mathbf{a}, R)$ of the Euclidean group[3] such that

$$U(\mathbf{a}, R)\psi(\mathbf{x})U^{-1}(\mathbf{a}, R) = \psi(R\mathbf{x} + \mathbf{a}), \tag{1.3}$$

and such that the energy operator for a finite region of space,

$$H_V = (2m)^{-1} \int_V d\mathbf{x} \nabla\psi^* \cdot \nabla\psi + \frac{1}{2} \int_V d\mathbf{x} \int_V d\mathbf{y} \psi^*(\mathbf{x})\psi^*(\mathbf{y}) V(\mathbf{x} - \mathbf{y})\psi(\mathbf{x})\psi(\mathbf{y}), \tag{1.4}$$

is well defined? Also the number operator for a finite region of space,

$$N_V = \int_V d\mathbf{x} \psi^*(\mathbf{x})\psi(\mathbf{x}), \tag{1.5}$$

must exist. In order to discuss the thermodynamics of an infinite Bose gas, we must also require that the limits

$$\rho_{op} = \lim_{V\to\infty} V^{-1} N_V; \qquad h_{op} = \lim_{V\to\infty} V^{-1} H_V \tag{1.6}$$

exist.

There is, of course, one representation satisfying these requirements, namely the Fock representation. Since a total number operator exists in this representation, it describes a system with zero density. Thus the question is whether or not any nontrivial (i.e. finite-density) systems exist.

This approach of considering first the existence question has not always been the traditional one in physics. It might seem desirable to begin with the more physical aspects of the problem, such as finding a suitable replacement for the definition of the partition function via the finite box. However the difficulties that have been encountered in applying the canonical quantization procedure to relativistic quantum field theory suggest that we adopt the more systematic approach.

In Sec. 2 we present a brief mathematical discussion of the representations of the CCR's, designed to make the paper reasonably self-contained. In Secs. 3 and 4 we construct representations of the CCR's which describe an infinite free Bose gas, and discuss their mathematical structure. In Sec. 5 we give a method for constructing the algebra of observables, followed by a discussion of the thermal ensemble for the infinite free Bose gas. In Sec. 6 we give a brief qualitative discussion of the problem for an interacting Bose gas. Section 7 contains a discussion of our results.

2. MATHEMATICAL PRELIMINARIES

We list here a few standard mathematical definitions. Consider a set $\mathbf{A}$ of bounded operators in a Hilbert space H. The set $\mathbf{A}'$ of all bounded operators commuting with every operator in the set $\mathbf{A}$ is called the commutor of $\mathbf{A}$. Clearly $\mathbf{A} \subset \mathbf{A}'' = (\mathbf{A}')'$. The set $\mathbf{A}$ is called self-adjoint if $T \in \mathbf{A}$ implies $T^* \in \mathbf{A}$, i.e. $\mathbf{A}$ is closed under the adjoint operation. The set $\mathbf{A}$ is called a von Neumann algebra if it is self-adjoint and $\mathbf{A} = \mathbf{A}''$. Given an arbitrary self-adjoint

[1] C. N. Yang and T. D. Lee, Phys. Rev. 87, 404 (1952).
[2] J. E. Mayer, J. Chem. Phys. 5, 67 (1937); B. Kahn and G. E. Uhlenbeck, Physica 5, 399 (1938).
[3] The Euclidean group is the group of translations $\mathbf{a}$ and rotations R in 3-dimensional Euclidean space.

set S of operators, S'' is the von Neumann algebra generated by the set S. The algebra of all bounded operators in H is denoted by $\mathbf{B}(H)$.

We say that a self-adjoint set **A** of operators is irreducible if its commutor consists only of multiples of the identity. **A** is called cyclic if there exists a vector $x \in H$ such that the set of all vectors of the form Tx, $T \in \mathbf{A}$ spans H. The vector x is called a cyclic vector for **A**. Clearly **A** is irreducible if and only if every vector is cyclic.

A von Neumann algebra **A** is called a factor if $\mathbf{A} \cap \mathbf{A}'$ consists only of multiples of the identity, i.e. its center is trivial. There exists a classification of factors into types.[4,5] The possible types are I_n, I_∞, II_1, II_∞, and III_∞. A factor of type I_n is isomorphic to the algebra of all $n \times n$ matrices. A factor of type I_∞ is isomorphic to the algebra of all bounded operators in an infinite dimensional (separable) Hilbert space. Factors of type II and III are less simple to describe.

A group representation is called primary if the von Neumann algebra which it generates is a factor. A primary representation is said to be type I (II, III respectively) if this factor is type I (II, III respectively). A primary representation has the property that it cannot be decomposed into disjoint representations. (Two representations are said to be disjoint if no subrepresentation of one is equivalent to a subrepresentation of the other.) A primary representation of type I is a discrete direct sum of copies of one irreducible representation. A representation which can be decomposed into primary representations of type I (II, III, respectively) is said to be of type I (II, III, respectively). An arbitrary representation can be decomposed into representations of type I, II, and III.[6] In the theory of group representations of arbitrary type, the primary representations play a role analogous to that of the irreducible representations in the type I case.

Representations of the CCR's

To make the mathematical discussion rigorous we introduce the smeared operators

$$\phi(f) = \int d\mathbf{x}\phi(\mathbf{x})f(\mathbf{x}), \tag{2.1}$$

$$\pi(f) = \int d\mathbf{x}\pi(\mathbf{x})f(\mathbf{x}), \tag{2.2}$$

where the test functions $f(\mathbf{x})$ must be real in order that $\phi(f)$, $\pi(f)$ be self-adjoint. Further restrictions on the test functions will be discussed below. These operators satisfy the commutation relations

$$[\phi(f), \phi(g)] = [\pi(f), \pi(g)] = 0, \tag{2.3}$$

$$[\phi(f), \pi(g)] = i(f, g), \tag{2.4}$$

where

$$(f, g) = \int d\mathbf{x}f(\mathbf{x})g(\mathbf{x}) \tag{2.5}$$

is the inner product of the real functions $f(\mathbf{x})$, $g(\mathbf{x})$. The annihilation and creation operators $\psi(f)$, $\psi^*(f)$ are related to $\phi(f)$, $\pi(f)$ by

$$\psi(f) = 2^{-\frac{1}{2}}[\phi(f) + i\pi(f)], \tag{2.6}$$

$$\psi^*(f) = 2^{-\frac{1}{2}}[\phi(f) - i\pi(f)]. \tag{2.7}$$

To avoid difficulties due to the fact that the operators $\phi(f)$, $\pi(f)$ are unbounded, we introduce the unitary operators

$$U(f) = e^{i\phi(f)}, \tag{2.8}$$

$$V(f) = e^{i\pi(f)}, \tag{2.9}$$

which satisfy the commutation relations

$$U(f)V(g) = V(g)U(f)e^{-i(f,g)}, \tag{2.10}$$

$$U(f)U(g) = U(f + g), \tag{2.11}$$

$$V(f)V(g) = V(f + g). \tag{2.12}$$

Thus a representation of the CCR's is a map from a real inner product space V of test functions into unitary operators $U(f)$, $V(g)$ on a Hilbert space satisfying Eqs. (2.10–12). We also require that the operators $U(\lambda f)$, $V(\lambda f)$ be weakly continuous in in the real variable λ.[7]

Cyclic Representations

We call a representation of the CCR's cyclic if there exists a cyclic vector Ψ for the von Neumann algebra $\mathbf{R} = \{U(f), V(g)\}''$ generated by the $U(f)$, $V(g)$; i.e. if there exists a vector Ψ such that the set of all vectors of the form $U(f)V(g)\Psi$ spans the Hilbert space H. In this paper we shall consider only cyclic representations.

A detailed treatment of cyclic representations has been given by Araki.[8] For our present purposes

[4] F. J. Murray and J. von Neumann, Ann. Math. **37**, 116 (1936).

[5] G. W. Mackey, Notes on Group Representations (Department of Mathematics, University of Chicago, 1955).

[6] For a more general definition, see reference 5.

[7] For a detailed and rigorous exposition of the representations of the CCR's see I. E. Segal, Trans. Am. Math. Soc. **88**, 12 (1958); J. Lew, thesis, Princeton University, Princeton, New Jersey (1960); H. Araki, thesis, Princeton University (1960); and reference 8.

[8] H. Araki, J. Math. Phys. **1**, 492 (1960).

it is sufficient to note that for a representation with a cyclic vector Ψ, the functional $E(f, g) = (\Psi, U(f)V(g)\Psi)$ determines the representation up to unitary equivalence. We shall also have occasion to use the following result, which is theorem 4.3 of reference 8.

Lemma 2.1: The necessary and sufficient condition that a functional $E(f, g)$, $f, g \in V$ define a representation of the CCR's with a cyclic vector Ψ such that $E(f, g) = (\Psi, U(f)V(g)\Psi)$ is

$$E(f, g)^* = E(-f, -g) \exp [i(f, g)], \qquad E(0, 0) = 1,$$

$$\sum_{i,j=1}^{n} c_i c_j^* E(f_i - f_j, g_i - g_j) \times \exp [i(g_i, f_j) - i(g_j, f_i)] \geq 0,$$

for any integer n, any set of complex numbers c_i, and any collection of test functions f_i, g_i in V; and $E(sf_1 + f_2, tg_1 + g_2)$ is separately continuous in s, t.

The functional $E(f, g)$ will play a central role in this paper. We now turn our attention to the selection of the test functions.

The space V of test functions depends on the case at hand. The Fock representation for nonrelativistic potential scattering is defined for all square integrable functions $f(\mathbf{x})$. In relativistic quantum field theory, one usually takes the space $\mathfrak{D}$ of Schwartz,[9] namely all infinitely differentiable functions vanishing outside a finite region. For the representations considered in this paper, the class of functions for which the representation can be defined depends on the particular representation. It is therefore convenient to choose some space of test functions which is sufficiently restricted to avoid any difficulties, but large enough to generate all the $U(f)$, $V(g)$ by taking strong limits. We shall use the space $\mathfrak{D}$ for this purpose. We now prove a lemma which will be used to extend the definition of the operators $U(f)$, $V(g)$.

Lemma 2.2: Let $U_n \to U$, $V_n \to V$ in the strong operator topology where U_n, V_n are all unitary. Then $U_nV_n \to UV$.

Proof: Clearly U and V are isometric. Let $x \in H$, then we have

$$||(U_n^{-1}U - 1)x||^2 = 2\{||x||^2 - \text{Re}\,(Ux, U_nx)\} \to 0 \quad \text{as} \quad n \to \infty,$$

i.e. $U_n^{-1}U \to 1$ strongly. Now consider

$$||(U_nV_n - UV)x||^2 = 2\{||x||^2 - \text{Re}\,(U_n^{-1}UVx, V_nx)\} \to 0 \quad \text{as} \quad n \to \infty,$$

QED.

Corollary: If U_n and U_n^{-1} both converge strongly, U_n unitary, then $U = \lim_{n\to\infty} U_n$ is unitary.

Lemma 2.3: Given a cyclic representation of the CCR's with functional $E(f, g) = (\Psi_0, U(f)V(g)\Psi_0)$, $f, g \in V$. In general, V is an incomplete inner product space. The definition of $U(f)$, $V(g)$ can be extended to all functions $f, g \in \bar{V}$ (the completion of V) for which there exist sequences $f_n, g_n \in V$ which converge to $f, g \in \bar{V}$ such that

$$\lim_{n,m\to\infty} E(f_n - f_m, 0) = E(0, 0) = 1;$$

$$\lim_{n,m\to\infty} E(0, g_n - g_m) = 1,$$

by defining

$$U(f) = \lim_{n\to\infty} U(f_n);$$

$$V(g) = \lim_{n\to\infty} V(g_n),$$

where the limits are taken in the strong operator topology.

Proof: For any $\Psi \in H$ and $\epsilon > 0$ there exists a

$$\Psi_\epsilon = \sum_{i=1}^{k(\epsilon)} c_i^\epsilon U(f_i^\epsilon) V(g_i^\epsilon)\Psi_0,$$

such that

$$||\Psi - \Psi_\epsilon|| < \epsilon.$$

Let $f_n \in V$ be a sequence such that $E(f_n - f_m, 0) \to 1$. Then

$$\begin{aligned} &||[U(f_n) - U(f_m)]U(g_1)V(g_2)\Psi_0||^2 \\ &= ||\{U(f_n) - \exp [i(g_2, f_m - f_n)]U(f_m)\}\Psi_0||^2 \\ &= 2\{||\Psi_0||^2 - \text{Re} \exp [i(g_2, f_m - f_n)]E(f_n - f_m, 0)\}, \end{aligned}$$

which $\to 0$ as $n, m \to \infty$. Thus there exists an $N(\epsilon)$ such that for $n, m > N(\epsilon)$ and $i = 1, \cdots k(\epsilon)$, we have

$$||[U(f_n) - U(f_m)]c_i^\epsilon U(f_i^\epsilon)V(g_i^\epsilon)\Psi_0|| < \epsilon/k(\epsilon),$$

and hence

$$||[U(f_n) - U(f_m)]\Psi_\epsilon|| < \epsilon.$$

It follows that

$$||[U(f_n) - U(f_m)]\Psi|| < 3\epsilon.$$

[9] L. Schwartz, *Theorie des Distributions* (Hermann & Cie., Paris, 1951).

Thus $U(f_n)$ converges strongly. Since $E(-f_n, 0) = E(f_n, 0)^*$ it follows that $U^{-1}(f_n) = U(-f_n)$ converges strongly also. Hence by the corollary to Lemma 2.2, the operator $U(f) = \lim_{n\to\infty} U(f_n)$ is unitary.

Similarly one can define $V(f)$. It follows from Lemma 2.2 that $U(f)$, $V(g)$ satisfy CCR's. It remains only to show that $U(\lambda f)$, $V(\lambda f)$ are weakly continuous in the real parameter λ. Clearly matrix elements of $U(\lambda f)$, $V(\lambda f)$ must be measurable functions of λ. Weak continuity then follows by applying von Neumann's theorem[10] which states that any representation of the commutation relations for a finite number of degrees of freedom such that matrix elements of $U(\lambda_1, \cdots \lambda_n)$, $V(\lambda_1, \cdots \lambda_n)$ are measurable functions of $(\lambda_1, \cdots \lambda_n)$, is a discrete direct sum of copies of the Schrödinger representation. QED.

The Fock Representation

Our construction of inequivalent representations of the CCR's makes extensive use of the Fock representation (also called the configuration space representation, and the no-particle representation). While the mathematical properties of the Fock representation are well known,[11] they do not seem to be readily available in the literature. We therefore include a brief discussion.

The Fock representation is usually defined as follows: Let $H_F^0 = \mathrm{C}$ (the complex numbers), and let H_F^n be the Hilbert space of symmetric square-integrable functions of n variables (each with domain R^3, and with respect to the measure $d\mathbf{x}_1 \cdots d\mathbf{x}_n$). Then the Hilbert space of state vectors is defined as the direct sum

$$H_F = \sum_{n=0}^{\infty \oplus} H_F^n. \tag{2.13}$$

The creation and annihilation operators $\psi_F^*(f)$, $\psi_F(f)$ are defined by

$$\begin{aligned}
&[\psi_F^*(f)\Psi]^{(n)}(\mathbf{x}_1, \cdots \mathbf{x}_n) \\
&\quad = n^{-\frac{1}{2}} \sum_{i=1}^{n} f(\mathbf{x}_i)\Psi^{(n-1)}(\mathbf{x}_1, \cdots \hat{\mathbf{x}}_i, \cdots \mathbf{x}_n), \\
&[\psi_F(f)\Psi]^{(n)}(\mathbf{x}_1, \cdots \mathbf{x}_n) \\
&\quad = (n+1)^{\frac{1}{2}} \int d\mathbf{x} f(\mathbf{x})\Psi^{(n+1)}(\mathbf{x}, \mathbf{x}_1, \cdots \mathbf{x}_n),
\end{aligned} \tag{2.14}$$

where $\hat{\mathbf{x}}_i$ means omit the ith coordinate. The unitary representation of the Euclidean group is defined by

$$\begin{aligned}
&[U_F(\mathbf{a}, R)\Psi]^{(n)}(\mathbf{x}_1, \cdots \mathbf{x}_n) \\
&\quad = \Psi^{(n)}[R^{-1}(\mathbf{x}_1 - \mathbf{a}), \cdots R^{-1}(\mathbf{x}_n - \mathbf{a})].
\end{aligned} \tag{2.15}$$

[10] J. von Neumann, Math. Ann. **104**, 570 (1931).
[11] J. M. Cook, Trans. Am. Math. Soc. **74**, 222 (1953).

The no-particle or vacuum state Ψ_{F0} is given by

$$\Psi_{F0}^{(0)} = 1, \qquad \Psi_{F0}^{(n)} = 0, \qquad n \neq 0. \tag{2.16}$$

Ψ_{F0} is the unique translationally invariant state, i.e. $U_F(\mathbf{a}, 0)\Psi = \Psi$ implies $\Psi = \alpha\Psi_{F0}$, α complex.

It is easy to verify that the operators

$$\begin{aligned}
\phi_F(f) &= [\psi_F^*(f) + \psi_F(f)]/\sqrt{2}, \\
\pi_F(f) &= i[\psi_F^*(f) - \psi_F(f)]/\sqrt{2},
\end{aligned} \tag{2.17}$$

satisfy at least formally the CCR's. To verify that $\phi_F(f)$, $\pi_F(f)$ as defined above can be extended to self-adjoint operators is not so trivial. We now give an alternative construction of the Fock representation, which is more useful for a rigorous discussion of its mathematical properties.

Von Neumann has defined an infinite tensor product of Hilbert spaces.[12] We shall write the Fock representation as an incomplete infinite tensor product of Schrödinger representations of the commutation relations for one degree of freedom. We assume the reader to be familiar with von Neumann's results.

Let H_i, $i = 1, 2, \cdots$ be the Hilbert space in which the Schrödinger representation is defined with self-adjoint operators Q_i, P_i satisfying $[Q_i, P_i] = i$, and let Ψ_{0i} be the no-particle state. It is well known that the operators Q_i, P_i are irreducible in H_i. Furthermore Ψ_{0i} is a cyclic vector for Q_i alone, and for P_i alone.[13]

The Fock representation is defined as follows: Let

$$H_F = \prod_{i=1}^{\infty \otimes} H_i$$

be the incomplete infinite tensor product containing the vector

$$\Psi_{F0} = \prod_{i=1}^{\infty \otimes} \Psi_{0i}.$$

Define unitary operators $U_i(\lambda)$, $V_i(\lambda)$, λ real, by

$$\begin{aligned}
U_i(\lambda) &= 1 \otimes \cdots \otimes e^{i\lambda Q_i} \otimes 1 \cdots, \\
V_i(\lambda) &= 1 \otimes \cdots \otimes e^{i\lambda P_i} \otimes 1 \cdots.
\end{aligned}$$

To define the $U_F(f)$, $V_F(f)$ choose a complete orthonormal basis (f_i) for V. If

$$f = \sum_{i=1}^{n} \lambda_i f_i,$$

[12] J. von Neumann, Compositio Math. **6**, 1 (1938).
[13] Take H_i to be all square-integrable functions $f(x)$ of the real variable x, Q_i to be multiplication by x. Then $\Psi_{0i} = \pi^{-1/4} \exp(-x^2/2)$. But any function which vanishes only on a set of measure zero is a cyclic vector for Q_i. Similarly, by taking fourier transforms, we see that Ψ_{0i} is a cyclic vector for P_i.

define

$$U_F(f) = \prod_{i=1}^{n} U_i(\lambda_i),$$

$$V_F(f) = \prod_{i=1}^{n} V_i(\lambda_i).$$

Since the $U_i(\lambda)$, $V_i(\lambda)$ are irreducible in each H_i, the $U_F(f)$, $V_F(f)$ are irreducible in H_F.[12] Thus Ψ_{F0} is a cyclic vector, and by an elementary calculation one finds

$$E_F(f, g) = (\Psi_{F0}, U_F(f)V_F(g)\Psi_{F0})$$
$$= \exp\left[-\tfrac{1}{4}||f||^2 - \tfrac{1}{4}||g||^2 - \tfrac{1}{2}i(f, g)\right]. \quad (2.18)$$

It follows from Lemma 2.3 that $U_F(f)$, $V_F(f)$ can be defined for all square-integrable f.

To show that Ψ_{F0} is a cyclic vector for the $U_F(f)$, choose a basis $\Phi_{i,1}, \Phi_{i,2}, \cdots$ for each H_i such that $\Phi_{i,1} = \Psi_{0i}$. Then the set of all vectors of the form

$$\prod_{i=1}^{\infty}{}^{\otimes} \Phi_{i,n(i)},$$

where $n(i) = 1$ except for finitely many i, is a basis for H_F.[12] Since Ψ_{0i} is a cyclic vector for $U_i(\lambda)$ in H_i, any such vector belongs to the linear manifold spanned by all $U(f)\Psi_{F0}$; i.e., Ψ_{F0} is a cyclic vector for all $U_F(f)$. Similarly Ψ_{F0} is a cyclic vector for all $V_F(f)$. These results may be summed up as follows.

Lemma 2.4: Let H_F be the Hilbert space in which the Fock representation $U_F(f)$, $V_F(f)$ is defined with the no-particle state Ψ_{F0}. Then $\{U_F(f), V_F(f), f \in V\}''$ is irreducible, and Ψ_{F0} is a cyclic vector for $\{U_F(f), f \in V\}''$ alone, and also for $\{V_F(f), f \in V\}''$ alone.

In the representation so constructed, Ψ_{F0} is in the domain of any polynomial of $\phi_F(f)$ and $\pi_F(g)$ [the infinitesimal generators of $U_F(\lambda f)$ and $V_F(\lambda g)$]. The vacuum expectation value of products of $\phi_F(f)$ and $\pi_F(g)$ can be calculated easily from $E_F(f, g)$ and coincides with those given by Eqs. (2.13–17). Furthermore,

$$U_F(f)\Psi_{F0} = \sum_{n=0}^{\infty} \frac{1}{n!} \phi_F(f)^n \Psi_{F0}$$

in the strong operator topology, and hence Ψ_{F0} is cyclic with respect to polynomials of $\phi_F(f)$. Therefore the cyclic representation of the CCR's given by $E_F(f, g)$ above is equivalent to that given by Eqs. (2.13–17).

The Fock representation is frequently used to describe the nonrelativistic potential scattering of particles. In the standard treatment, it is presented as being completely equivalent to the n-particle Schrödinger wave equation. In particular, only pure states which have a precise particle number have any physical significance. This means that not only is the number operator N conserved, but it defines a superselection rule.[14,15] Since neither $\phi_F(f)$ nor $\pi_F(f)$ commute with the total number operator N, they cannot be observables. It should be noted that Eq. (2.13) gives the direct sum decomposition of the Hilbert space of state vectors into coherent subspaces (i.e. invariant subspaces on which the algebra of observables is irreducible).

This superselection rule also follows from the Galilean invariance of the theory.[16] Of course this superselection rule does not apply for a photon gas.

Intensive Observables

Our goal is to construct representations of the CCR's which, among other things, have a number density operator ρ_{op} defined formally by[17]

$$\rho_{op} = \lim_{V\to\infty} N_V/V, \quad (2.19)$$

where

$$N_V = \int_V d\mathbf{x}\psi^*(\mathbf{x})\psi(\mathbf{x}). \quad (2.20)$$

While this expression for the number operator as an integral of the formal operator $\psi^*(\mathbf{x})\psi(\mathbf{x})$ is useful for the Fock representation when realized in configuration space, it is not clear what this formula means in general. A rigorous and convenient expression for N_V is obtained by introducing a complete orthonormal set of real functions f_n in the volume V, and using the equations

$$N_V = \sum_n N(f_n) = \sum_n \psi^*(f_n)\psi(f_n),$$
$$= \tfrac{1}{2}\sum_n [\phi^2(f_n) + \pi^2(f_n) - 1]. \quad (2.21)$$

In general, of course, the operator N_V does not exist. In this paper we construct representations in which both N_V and ρ_{op} exist. It is apparent that, in

[14] G. C. Wick, A. S. Wightman, and E. P. Wigner, Phys. Rev. 88, 101 (1952).

[15] J. M. Jauch, Helv. Phys. Acta 33, 711 (1960).

[16] V. Bargmann, Ann. Math. 59, 1 (1954). Bargmann's superselection rule arises in much the same way as the univalence superselection rule in relativistic quantum field theory, which states that one cannot superpose states of integer spin with states of half-integer spin, since they transform differently under rotations. In the above paper, Bargmann shows that the way a state transforms under the Galilean group depends on its mass.

[17] The operators N_V and ρ_{op} always act in the Hilbert space of the infinite volume representations. An alternative possibility, which we shall not use, is to define N_V in the finite volume representation and take the limit of $V^{-1}N_V$ as $V \to \infty$.

order to discuss rigorously the existence of these operators, we must first give a precise definition of the convergence of unbounded operators.

Definition 2.1: Let T_n, T be self-adjoint operators in a Hilbert space H. We shall say that $T_n \to T$ if for all real λ, $e^{i\lambda T_n} \to e^{i\lambda T}$ in the strong operator topology.

Consider an irreducible representation in which ρ_{op} exists. Since $\phi(f)$ and $\pi(g)$ can only change the particle number by a finite integer, they cannot change the density of a system with an infinite volume. Thus we would expect that in an irreducible representation the number density operator should be a constant. We now give a rigorous proof of this fact. In lemmas 2.5 and 2.6, we consider only those representations of the CCR's in which N_V exists.

Lemma 2.5: In an irreducible representation the number density operator is a constant (if it exists).

Proof: The number operator for the volume V is defined by

$$N_V = \lim_{i\to\infty} \sum_{i=1}^{i} \psi^*(f_i)\psi(f_i),$$

where the f_i are a complete orthonormal set of real functions for the volume V, and the limit is taken in the sense of definition 2.1. The number density operator is defined by

$$\rho_{op} = \lim_{V\to\infty} N_V/V,$$

also using definition 2.1. We assume that ρ_{op} exists, and want to show that as $V \to \infty$, $\exp(i\lambda N_V/V)$ commutes with all $U(f)$, $V(g)$ and is therefore a multiple of the identity. Take V sufficiently large that it contains the support of the functions f, g. Since N_V does not depend on the complete set f_i, choose $f_1 = f$. Then all but the first term in N_V commutes with $U(f)$, and a straightforward calculation yields

$$\exp(i\lambda N_V/V)U(f) = U[f\cos(\lambda/V)]V[f\sin(\lambda/V)] \times \exp[i\sin(2\lambda/V)\,||f||^2/4]\exp[i\lambda N_V/V].$$

Similarly, one obtains

$$\exp(i\lambda N_V/V)V(g) = V[g\cos(\lambda/V)]U[-g\sin(\lambda/V)] \times \exp(i\lambda N_V/V)\exp[i\sin(2\lambda/V)\,||g||^2/4].$$

It follows from Lemma 2.2 that

$$\exp(i\lambda\rho_{op}) = \lim_{V\to\infty} \exp(i\lambda N_V/V)$$

commutes with all $U(f)$, $V(g)$. QED.

Lemma 2.6: Given a cyclic representation of the CCR's in which

$$\lim_{V\to\infty} \exp(i\lambda N_V/V)\Psi = \exp(i\lambda\rho)\Psi$$

for some cyclic Ψ and all real λ. Then ρ_{op} exists and is a constant.

Proof: From the proof of Lemma 2.5, we know that $\exp(i\lambda N_V/V)$ commutes with all $U(f)$, $V(g)$ as $V \to \infty$. It follows that

$$\lim_{V\to\infty} \exp(i\lambda N_V/V)U(f)V(g)\Psi = \exp(i\lambda\rho)U(f)V(g)\Psi.$$

But a sequence of unitary operators which converges on a dense set must converge everywhere, i.e.

$$\lim_{V\to\infty} \exp(i\lambda N_V/V) = \exp(i\lambda\rho)$$

in the strong operator topology. QED.

For the free Bose gas, a similar argument shows that $\lim_{V\to\infty} H_V/V$, if it exists, will be a multiple of the identity.

We now consider the distribution of the particles in momentum space (since we are concerned with the free Bose gas). This is not quite so straightforward, since any finite region in momentum space will contain infinitely many particles. This difficulty is avoided by considering the momentum distribution for a finite volume V, and then letting $V \to \infty$. Thus let B be some closed set in momentum space. Let f_n be a complete orthonormal set of functions with periodic boundary conditions whose support in momentum space (a discrete set) is in B. We define the operator

$$N_V(B) = \tfrac{1}{2} \sum_n \psi^*(f_n)\psi(f_n^*).$$

As in the proof of Lemma 2.5 one can show that $\lim_{V\to\infty} N_V(B)/V$, if it exists, commutes with all $U(f)$, $V(g)$. Thus for the infinite free Bose gas, lemmas 2.5 and 2.6 also hold for the particle density distribution in momentum space.

3. THE INFINITE FREE BOSE GAS: GROUND STATE

In this section and Sec. 4 we shall construct representations of the CCR's with the desired properties (number density operator, etc.). Of course in the transition to the idealization of an infinite system there is always the danger of ambiguity. We shall attempt to justify our approach when we consider the density matrix in Sec. 5.

We begin with the simplest possible case, namely the free Bose gas at zero temperature. Thus our

problem is to construct a representation of the CCR's which contains the ground state of the infinite free Bose gas.

It seems reasonable to assume that the ground state for the case of an infinite volume should be cyclic. Now the infinite system must be the limit of the finite system in the sense that physical quantities evaluated for a Bose gas in a box should converge to the corresponding quantities for the infinite Bose gas as the volume $V \to \infty$. This suggests that if we let $E_V(f, g) = (\Psi_V, U_F(f)V_F(g)\Psi_V)$ where Ψ_V is the ground state in a box of volume V and the subscript F refers to the Fock representation, then the functional $E(f, g)$ for the infinite system should be given by $\lim_{V\to\infty} E_V(f, g)$.

Calculation of $E(f, g)$

The Bose gas in a finite box is described by the ordinary Fock representation. The ground state is the state with $n = \rho V$ particles each with wavefunction $f_V(\mathbf{x}) = V^{-\frac{1}{2}}$, $\mathbf{x} \in V$. Introducing the annihilation and creation operators

$$\psi_F(f) = [\phi_F(f) + i\pi_F(f)]/\sqrt{2}, \tag{3.1}$$

$$\psi_F^*(f) = [\phi_F(f) - i\pi_F(f)]/\sqrt{2}, \tag{3.2}$$

we have

$$E_V(f, g) = (n!)^{-1}(\psi_F^*(f_V)^n\Psi_{F0}, U_F(f)V_F(g)\psi_F^*(f_V)^n\Psi_{F0}), \tag{3.3}$$

where Ψ_{F0} is the no-particle state. We use the formulas

$$||\psi_F^*(f)^n\Psi_{F0}||^2 = n! \quad \text{if} \quad ||f|| = 1, \tag{3.4}$$

$$\exp[\psi_F(f)]\psi_F^*(g)\exp[-\psi_F(f)] = \psi_F^*(g) + (f, g), \tag{3.5}$$

$$U_F(f)V_F(g) = \exp\{[i\psi_F^*(f) - \psi_F^*(g)]/\sqrt{2}\} \times \exp\{[i\psi_F(f) + \psi_F(g)]/\sqrt{2}\}E_F(f, g), \tag{3.6}$$

where $E_F(f, g)$, the vacuum functional for the Fock representation, is given by Eq. (2.18). Thus we obtain

$$\begin{aligned} E_V(f, g) &= E_F(f, g)(n!)^{-1} \\ &\times ([\psi_F^*(f_V) - i2^{-\frac{1}{2}}(f, f_V) - 2^{-\frac{1}{2}}(g, f_V)]^n\Psi_{F0}, \\ &\times [\psi_F^*(f_V) + i2^{-\frac{1}{2}}(f, f_V) + 2^{-\frac{1}{2}}(g, f_V)]^n\Psi_{F0}) \\ &= E_F(f, g)\frac{1}{n!}\sum_{r=0}^{n}\left(\frac{n!}{r!\,(n-r)!}\right)^2 \\ &\times [-|(g, f_V) + i(f, f_V)|^2/2]^r(n-r)! \\ &= E_F(f, g)L_n\{\tfrac{1}{2}[(g, f_V)^2 + (f, f_V)^2]\}, \end{aligned} \tag{3.7}$$

where L_n is the nth Laguerre polynomial. Without loss of generality we can assume V sufficiently large that f and g vanish outside it. Then

$$(f, f_V) = \tilde{f}(0)(\rho/n)^{\frac{1}{2}}, \tag{3.8}$$

where

$$\tilde{f}(0) = \int d\mathbf{x} f(\mathbf{x}), \tag{3.9}$$

and similarly for g. Now let $V \to \infty$ with $\rho = n/V$ held constant. Using the formula[18]

$$\lim_{n\to\infty} L_n(z/n) = J_0(2z^{\frac{1}{2}}), \tag{3.10}$$

we obtain

$$E(f, g) = E_F(f, g)J_0\{(2\rho[\tilde{f}(0)^2 + \tilde{g}(0)^2])^{\frac{1}{2}}\}. \tag{3.11}$$

It remains to show that this functional defines a representation of the CCR's with the desired properties. We shall do this by explicitly constructing the representation.

Construction of the Representation

Let H_F denote the Hilbert space in which the Fock representation $U_F(f)$, $V_F(g)$ is defined with vacuum state Ψ_{F0}, and a unitary representation of the Euclidean group $U_F(\mathbf{a}, R)$. Then

$$(\Psi_{F0}, U_F(f)V_F(g)\Psi_{F0}) = E_F(f, g). \tag{3.12}$$

Let M be the Hilbert space of square-integrable functions on the unit circle with respect to the normalized Lebesgue measure $d\theta/2\pi$. Define bounded operators A, B in M as follows

$$(Af)(\theta) = \cos\theta \quad f(\theta), \tag{3.13}$$

$$(Bf)(\theta) = \sin\theta \quad f(\theta). \tag{3.14}$$

Denote the identity function $f(\theta) = 1$ by Φ_0. We then define a representation of the CCR's by

$$H = H_F \otimes M, \tag{3.15}$$

$$\Psi_0 = \Psi_{F0} \otimes \Phi_0, \tag{3.16}$$

$$U(f) = U_F(f) \otimes \exp[i(2\rho)^{\frac{1}{2}}A\tilde{f}(0)], \tag{3.17}$$

$$V(g) = V_F(g) \otimes \exp[i(2\rho)^{\frac{1}{2}}B\tilde{g}(0)], \tag{3.18}$$

$$U(\mathbf{a}, R) = U_F(\mathbf{a}, R) \otimes 1. \tag{3.19}$$

Thus we have

$$\begin{aligned} &(\Psi_0, U(f)V(g)\Psi_0) \\ &= E_F(f, g)\int \exp\{i(2\rho)^{\frac{1}{2}}[\tilde{f}(0)\cos\theta \\ &\quad + \tilde{g}(0)\sin\theta]\}\, d\theta/2\pi \\ &= E_F(f, g)J_0\{2\rho[\tilde{f}(0)^2 + \tilde{g}(0)^2])^{\frac{1}{2}}\}, \end{aligned} \tag{3.20}$$

which is the desired functional.

[18] G. Szegö, *Orthogonal Polynomials* (Am. Math. Soc., New York, 1959), revised ed., Theorem 8.1.3.

We shall prove that Ψ_0 is a cyclic vector for the algebra $\{U(f), V(g)\}''$. Consider the functions

$$f_{s,a}(x) = ase^{-sx}/[4\pi(x^2 + 1)], \qquad s, a > 0. \tag{3.21}$$

We have

$$\lim_{s\to 0} ||f_{s,a}||^2 = 0, \tag{3.22}$$

$$\lim_{s\to 0} \tilde{f}_{s,a}(0) = a. \tag{3.23}$$

It follows from Lemma 2.3 that in the strong operator topology we have

$$\lim_{s\to 0} U(f_{s,a}) = 1 \otimes \exp\,[i(2\rho)^{\frac{1}{2}}aA], \tag{3.24}$$

and similarly,

$$\lim_{s\to 0} V(f_{s,a}) = 1 \otimes \exp\,[i(2\rho)^{\frac{1}{2}}aB]. \tag{3.25}$$

Since Φ_0 is a cyclic vector for $\{A, B\}''$ in the space M, and Ψ_{F0} is a cyclic vector for $\{U_F(f), V_F(g)\}''$ in H_F (see Lemma 2.4) it follows that $\Psi_0 = \Psi_{F0} \otimes \Phi_0$ is cyclic for $\{U(f), V(g)\}''$. Thus we have constructed the desired representation.

In dealing with infinite systems, one expects to encounter nonseparable Hilbert spaces. In particular for the case of the ground state, one expects the particle density ρ to label inequivalent representations of the CCR's (see lemmas 2.5 and 2.6). It might also seem reasonable to expect that, for a given value of ρ, the representation should be irreducible. This is not the case. It is clear from the above construction of the representation that any operator of the form $1 \otimes T$, where T commutes with A and B, commutes with all $U(f)$ and $V(g)$.

We shall now decompose the representation into a direct integral of irreducible representations. Given the above explicit realization of the representation, this is a rather trivial task. We write

$$M = \int_0^{2\pi\oplus} M(\theta)\, d\theta/2\pi, \tag{3.26}$$

where $dim\ M(\theta) = 1$, and apply the theorem[19]

$$H_F \otimes \int^{\oplus} M(\theta)\, d\theta/2\pi = \int^{\oplus} H_F \otimes M(\theta)\, d\theta/2\pi. \tag{3.27}$$

The operators $U(f)$, $V(g)$ are clearly decomposable with respect to this decomposition of H. Thus we have

$$U(f) = \int^{\oplus} U_\theta(f)\, d\theta/2\pi, \tag{3.28}$$

$$V(g) = \int^{\oplus} V_\theta(g)\, d\theta/2\pi, \tag{3.29}$$

where

$$U_\theta(f) = U_F(f) \otimes \exp\,[i(2\rho)^{\frac{1}{2}} \cos\theta\, \tilde{f}(0)], \tag{3.30}$$

$$V_\theta(g) = V_F(g) \otimes \exp\,[i(2\rho)^{\frac{1}{2}} \sin\theta\, \tilde{g}(0)]. \tag{3.31}$$

Since $dim\ M(\theta) = 1$ and the algebra $\{U_F(f), V_F(g)\}''$ is irreducible in H_F, it follows that $\{U_\theta(f), V_\theta(g)\}''$ is irreducible in $H_F \otimes M(\theta)$. Thus we have decomposed the representation defined by Eq. (3.11) into a direct integral of irreducible representations. These results may be summed up as follows.

Theorem 3.1: The functional

$$E(f, g) = E_F(f, g)J_0\{(2\rho[\tilde{f}(0)^2 + \tilde{g}(0)^2])^{\frac{1}{2}}\}$$

defines a cyclic representation of the CCR's. This representation is a direct integral of irreducible representations $U_\theta(f)$, $V_\theta(g)$ where $0 \leq \theta \leq 2\pi$.

Physical Significance of the Reducibility

We now consider the physical significance of the reducibility of the representation. As we remarked earlier, one expects the particle density ρ to label inequivalent representations. Thus we would like to find another quantity which labels the different irreducible representations belonging to a given value of ρ. To do this we shall construct the irreducible representations in a slightly different fashion. As above, let H_F, $\phi_F(f)$, $\pi_F(g)$ refer to the Fock representation. We define a new representation by

$$\phi_\theta(\mathbf{x}) = \phi_F(\mathbf{x}) + (2\rho)^{\frac{1}{2}} \cos\theta, \tag{3.32}$$

$$\pi_\theta(\mathbf{x}) = \pi_F(\mathbf{x}) + (2\rho)^{\frac{1}{2}} \sin\theta. \tag{3.33}$$

Then

$$U_\theta(f) = U_F(f) \exp\,[i(2\rho)^{\frac{1}{2}} \cos\theta\, \tilde{f}(0)], \tag{3.34}$$

$$V_\theta(g) = V_F(g) \exp\,[i(2\rho)^{\frac{1}{2}} \sin\theta\, \tilde{g}(0)], \tag{3.35}$$

which makes the equivalence of the two definitions rather obvious. Thus the irreducible representations describing the ground state of an infinite free Bose gas are simply the Fock representation with a uniform "external" field superimposed. Thus we see that the representations are labeled by the average field strengths

$$\langle\phi_\theta\rangle = \lim_{V\to\infty} V^{-1} \int_V d\mathbf{x}\, \phi_\theta(\mathbf{x}) = (2\rho)^{\frac{1}{2}} \cos\theta, \tag{3.36}$$

$$\langle\pi_\theta\rangle = \lim_{V\to\infty} V^{-1} \int_V d\mathbf{x}\, \pi_\theta(\mathbf{x}) = (2\rho)^{\frac{1}{2}} \sin\theta. \tag{3.37}$$

[19] J. Dixmier, *Les Algebres d'Operateurs dans l'Espace Hilbertien*, (Gauthier-Villars, Paris, 1957), p. 152, proposition 11.

A rigorous proof of these limits is rather trivial. Consider the operator $U_F(\chi_V)$ where $\chi_V(\mathbf{x}) = V^{-1}$ if $\mathbf{x} \in V$ and is zero otherwise. Then $\lim_{V\to\infty} ||\chi_V|| = 0$, and by Lemma 2.3 we have $\lim_{V\to\infty} U_F(\chi_V) = 1$ in the strong operator topology. Hence $\langle\phi_F\rangle = 0$, and similarly $\langle\pi_F\rangle = 0$.

Having provided a label for the irreducible representations, we shall now explain the reducibility. That is, we must understand why there exist operators which are not observables. This is equivalent to saying that there exist states which, although physically equivalent, are mathematically distinguishable.

We shall show that the operators

$$a = (2\rho)^{-\frac{1}{2}}(\langle\phi\rangle + i\langle\pi\rangle) \tag{3.38}$$

and

$$a^* = (2\rho)^{-\frac{1}{2}}(\langle\phi\rangle - i\langle\pi\rangle) \tag{3.39}$$

can be interpreted as annihilation and creation operators for particles in the zero-momentum state. We note that $a = \lim_{V\to\infty} \psi(\chi_V)$ where $\chi_V(\mathbf{x}) = \rho^{-\frac{1}{2}}V^{-1}$, $\mathbf{x} \in V$. If Ψ_V is the normalized ground state for volume V with $n = \rho V$ particles, then $\psi(\chi_V)\Psi_V$ is the normalized ground state with $n - 1$ particles. That is, $\psi(\chi_V)$ is just the usual annihilation operator for zero-momentum particles but without the $n^{\frac{1}{2}}$ factor. Similarly $\psi^*(\chi_V)$ is, except for the $n^{\frac{1}{2}}$ factor, the creation operator for zero-momentum particles. It is therefore plausible to say that a and a^* are annihilation and creation operators for zero-momentum particles in the infinite system.

Since there are infinitely many particles in the zero-momentum state it should, loosely speaking, behave classically. Thus we note that the operators a and a^* act classically in the sense that they commute with each other and with all $U(f)$, $V(g)$. Since the effect of applying a or a^* is to change the number of particles in the zero-momentum state, we see that two states which differ by only a finite number of zero-momentum particles are physically indistinguishable. However two such states are not mathematically identical, unless the operator a is a multiple of the identity. Now if one chooses the ground state, as we did, to have a definite particle number, then one can hardly expect the operator a to be a c number (since it should produce an orthogonal state when applied to the ground state). To sum up, the reason for the reducibility of the representation is that two states which differ by only a finite number of particles in the zero-momentum state are mathematically distinguishable, although physically identical.

It is interesting to note that the operator a is unitary. In our construction of the representation, it is simply multiplication by $e^{i\theta}$, hence $(1 \otimes a)^n\Psi_0 = \Psi_{F0} \otimes \Phi_{-n}$ where $\Phi_n(\theta) = e^{-in\theta}$. The set of all $\Psi_{0n} = \Psi_{F0} \otimes \Phi_n$, $n = 0, \pm 1, \pm 2, \cdots$ form an orthonormal set spanning the manifold of (physically equivalent) degenerate ground states. The state Ψ_{0n} contains $\infty + n$ particles in the zero-momentum state.

The Particle Density Operator

In terms of the field $\psi(\mathbf{x})$ we have

$$\psi_\theta(\mathbf{x}) = \psi_F(\mathbf{x}) + \rho^{\frac{1}{2}}e^{i\theta}. \tag{3.40}$$

Thus the number operator N_V for the finite volume V is given by

$$(N_V)_\theta = (N_V)_F + \rho V + \rho^{\frac{1}{2}} \times \int_V d\mathbf{x}[e^{-i\theta}\psi_F(\mathbf{x}) + e^{i\theta}\psi_F^*(\mathbf{x})]. \tag{3.41}$$

Consider $\rho_{0\theta} = \lim_{V\to\infty} V^{-1}(N_V)_\theta$. From Eqs. (3.36) and (3.37) it follows that

$$\lim_{V\to\infty} V^{-1}\int_V d\mathbf{x}\psi_F(\mathbf{x}) = 2^{-\frac{1}{2}}(\langle\phi_F\rangle + i\langle\pi_F\rangle) = 0. \tag{3.42}$$

One can easily show that $\lim_{V\to\infty} V^{-1}(N_V)_F = 0$. It follows that

$$\rho_{0\theta} = \rho. \tag{3.43}$$

Thus our derivation of the representation has at least led to the correct particle density.

The Hamiltonian

The preceding analysis has been concerned with the field operators at a fixed time, which we shall take to be $t = 0$. Our general principle for determining $\psi(\mathbf{x}, t)$ is that it should be the limit, in some sense, of $\psi(\mathbf{x}, t)$ for the finite volume representation. For the present case, the problem of extending the definition of the field operators to arbitrary times is rather trivial. Let $\psi_F(\mathbf{x}, t)$ denote the time-dependent Fock representation satisfying the free-field equation

$$i\,\partial\psi_F/\partial t = -(\Delta/2m)\psi_F. \tag{3.44}$$

The field

$$\psi_\theta(\mathbf{x}, t) = \psi_F(\mathbf{x}, t) + \rho^{\frac{1}{2}}e^{i\theta} \tag{3.45}$$

clearly satisfies the free-field equation also.

It follows that the Hamiltonian H for $\psi_\theta(\mathbf{x}, t)$ is just the Fock Hamiltonian H_F, i.e.

$$H = H_F = (2m)^{-1}\int d\mathbf{x}(\nabla\psi_F^*)\cdot(\nabla\psi_F). \tag{3.46}$$

Using Eq. (3.40) this can be written

$$H = (2m)^{-1} \int d\mathbf{x}(\nabla\psi_\theta^*)\cdot(\nabla\psi_\theta). \qquad (3.47)$$

Thus the formal expression for the free-particle Hamiltonian remains valid for the case of the strictly infinite system in its ground state. This is not a trivial result. In particular, it should be noted that the operator

$$H' = \int d\mathbf{x}\psi_\theta^*(\mathbf{x})(-\Delta/2m)\psi_\theta(\mathbf{x}) \qquad (3.48)$$

does not exist.

Relation to BCS Theory

We have seen that the ground state of the infinite free Bose gas is degenerate, i.e. there exists more than one state Ψ which satisfies $H\Psi = 0$ where H is the Hamiltonian. In fact, any vector of the form $\Psi = \Psi_{F0} \otimes \Phi$, $\Phi \in M$ has the same energy as the ground state $\Psi_{F0} \otimes \Phi_0$. This situation is analogous to the degeneracy of the ground state of the BCS model of superconductivity.[20]

The degeneracy of the ground state of the BCS model is also related to the fact that the algebra of field operators, which now satisfies anticommutation relations, gives rise to inequivalent irreducible representations. As in the present case, the irreducible representations can be labeled in a natural way by the real parameter θ, $0 \leq \theta \leq 2\pi$. In both cases the degeneracy is related to the invariance of the theory under gauge transformations of the first kind $\psi \to e^{i\alpha}\psi$, α real. This transformation carries the irreducible representation $\psi_\theta(\mathbf{x})$ into $\psi_{\theta+\alpha}(\mathbf{x})$. To prove this for the infinite free Bose gas, we write

$$e^{i\alpha}\psi_\theta(\mathbf{x}) = e^{i\alpha}\psi_F(\mathbf{x}) + \rho^{\frac{1}{2}}e^{i(\alpha+\theta)} \qquad (3.50)$$

and note that $e^{i\alpha}\psi_F(\mathbf{x})$ is unitary equivalent to the Fock representation. It follows that $e^{i\alpha}\psi_\theta(\mathbf{x})$ is unitary equivalent to $\psi_{\theta+\alpha}(\mathbf{x})$.

4. THE INFINITE FREE BOSE GAS: EXCITED STATES

In this section we shall obtain representations of the CCR's which, we hope, will describe an infinite free Bose gas of uniform density at a finite temperature. Thus we shall construct representations which describe a system of particles with a distribution $\rho(\mathbf{k})$ in momentum space. We also consider the case where, in addition to a continuous distribution in momentum space, there is macroscopic occupation of the zero-momentum state (Bose–Einstein condensation). As in the case of the ground-state representations, the functional $E(f, g)$ will be obtained by an heuristic argument. We then construct the cyclic representation with this functional, and show that it has the desired properties.

We first list a number of assumptions about $\rho(\mathbf{k})$.

(i) $\rho(\mathbf{k}) = \rho(k)$,

(ii) $\rho(\mathbf{k})$ is continuous,

(iii) $\rho = \int d\mathbf{k}\rho(\mathbf{k}) < \infty$,

(iv) $\rho(\mathbf{k}) \neq 0$ almost everywhere.

We shall not always make explicit reference to these properties in the following. Much of the discussion is valid under weaker restrictions on $\rho(\mathbf{k})$. However these assumptions simplify the exposition, and they are valid for any case of physical interest.

Calculation of $E(f, g)$

We are interested in the case where the distribution in momentum space is a continuous function $\rho(\mathbf{k})$. We assume that the functional $E_\rho(f, g)$ for this case can be obtained as the limit of functionals describing a discrete distribution of particles in momentum space.

Consider a state Ψ_V (in a box of volume V) with $n_1 = \rho_1 V$ particles each with wavefunction

$$f_1^c(\mathbf{x}) = (2/V)^{\frac{1}{2}} \cos \mathbf{k}_1\cdot\mathbf{x}, \qquad (4.1)$$

another n_1 particles with wavefunction

$$f_1^s(\mathbf{x}) = (2/V)^{\frac{1}{2}} \sin \mathbf{k}_1\cdot\mathbf{x}, \qquad (4.2)$$

and $n_2 = \rho_2 V$ particles with wavefunction

$$f_2^c(\mathbf{x}) = (2/V)^{\frac{1}{2}} \cos \mathbf{k}_2\cdot\mathbf{x}, \qquad (4.3)$$

etc. If we assume that the functional

$$E_V(f, g) = (\Psi_V, U(f)V(g)\Psi_V)$$

has a limit as $V \to \infty$ with $\rho_1, \cdots, \rho_n$ held constant, then a simple modification of the methods of Sec. 3 shows that this limit must be

$$E(f, g) = E_F(f, g) \times \prod_{j=1}^{n} J_0\{2[f_c(\mathbf{k}_j)^2 + g_c(\mathbf{k}_j)^2]^{\frac{1}{2}}\rho_j^{\frac{1}{2}}\} \times J_0\{2[f_s(\mathbf{k}_j)^2 + g_s(\mathbf{k}_j)^2]^{\frac{1}{2}}\rho_j^{\frac{1}{2}}\}, \qquad (4.4)$$

where

$$f_c(\mathbf{k}) = \int d\mathbf{x}f(\mathbf{x}) \cos \mathbf{k}\cdot\mathbf{x},$$

[20] For an analysis of the BCS model from the present point of view, see R. Haag, Nuovo Cimento 25, 287 (1962).

$$f_s(\mathbf{k}) = \int d\mathbf{x} f(\mathbf{x}) \sin \mathbf{k}\cdot\mathbf{x}.$$

We now let the discrete distribution tend to a continuous distribution. Without loss of generality, we shall calculate this limit for the one-dimensional case. Then $\rho(k)$ is approximated by dividing the real line into intervals of length $1/n$, and letting $\rho_j^n = \rho(j/n)/n$. Thus we have

$$E(f, g) = \lim_{n, L\to\infty} E_F(f, g)$$
$$\times \prod_{j=1}^{nL} J_0(2\{\rho(j/n)n^{-1}[f_c(j/n)^2 + g_c(j/n)^2]\}^{\frac{1}{2}})$$
$$\times J_0(2\{\rho(j/n)n^{-1}[f_s(j/n)^2 + g_s(j/n)^2]\}^{\frac{1}{2}}). \quad (4.5)$$

Noting that $\rho(k) = \rho(-k)$, we have

$$\log [E(f, g)/E_F(f, g)]$$
$$= \lim_{n, L\to\infty} \sum_{j=1}^{nL} \{-\rho(j/n)[f_c(j/n)^2 + f_s(j/n)^2$$
$$+ g_c(j/n)^2 + g_s(j/n)^2]n^{-1} + 0(n^{-2})\}$$
$$= -\frac{1}{2}\int_{-\infty}^{\infty} dk\rho(k)[|\tilde{f}(k)|^2 + |\tilde{g}(k)|^2], \quad (4.6)$$

where

$$\tilde{f}(k) = \int dx e^{ikx} f(x). \quad (4.7)$$

The formula for the general case is then given by

$$E_\rho(f, g) = E_F(f, g)$$
$$\times \exp\left\{-\frac{1}{2}\int d\mathbf{k}\rho(\mathbf{k})[|\tilde{f}(\mathbf{k})|^2 + |\tilde{g}(\mathbf{k})|^2]\right\}. \quad (4.8)$$

It should be noted that this equation is valid only for continuous density distributions. If one wants to consider the case where there is macroscopic occupation of the ground state, it is incorrect to insert a delta function in $\rho(\mathbf{k})$. However if $\rho(\mathbf{k}) = \rho_0\,\delta(\mathbf{k}) + \rho_1(\mathbf{k})$ where $\rho_1(\mathbf{k})$ is continuous, then it is obvious from the above arguments that we would obtain

$$E_\rho(f, g) = E_F(f, g)J_0(\{2\rho_0[\tilde{f}(0)^2 + \tilde{g}(0)^2]\}^{\frac{1}{2}})$$
$$\times \exp\left\{-\frac{1}{2}\int d\mathbf{k}\rho_1(\mathbf{k})[|\tilde{f}(\mathbf{k})|^2 + |\tilde{g}(\mathbf{k})|^2]\right\}. \quad (4.9)$$

We now turn to the task of constructing the cyclic representations defined by these functionals.

No Macroscopic Occupation of the Ground State

We construct first the representation defined by Eq. (4.8). We define an operator ρ on the space V of test functions as follows [since ρ is diagonal in momentum space it is defined in terms of the Fourier transform $\tilde{f}(\mathbf{k})$].

$$(\rho f)(\mathbf{k}) = (2\pi)^3\rho(\mathbf{k})\tilde{f}(\mathbf{k}). \quad (4.10)$$

The representation is now defined as follows.

$$H = H_F \otimes H_F, \quad (4.11)$$
$$U(f) = U_F([1 + \rho]^{\frac{1}{2}}f) \otimes U_F(\rho^{\frac{1}{2}}f), \quad (4.12)$$
$$V(g) = V_F([1 + \rho]^{\frac{1}{2}}g) \otimes V_F(-\rho^{\frac{1}{2}}g), \quad (4.13)$$
$$\Psi_0 = \Psi_{F0} \otimes \Psi_{F0}, \quad (4.14)$$
$$U(\mathbf{a}, R) = U_F(\mathbf{a}, R) \otimes U_F(\mathbf{a}, R). \quad (4.15)$$

We have

$$(\Psi_0, U(f)V(g)\Psi_0)$$
$$= E_F(f, g) \exp\{-\tfrac{1}{2}[(f, \rho f) + (g, \rho g)]\}$$
$$= E_\rho(f, g). \quad (4.16)$$

which is the desired functional.

It remains only to prove that Ψ_0 is a cyclic vector for the algebra $\mathbf{R} = \{U(f), V(g)\}''$, i.e. that $\overline{\mathbf{R}\Psi_0} = H$ (the bar denotes the closed linear space spanned by the given set of vectors). Now the annihilation and creation operators are given by

$$\psi(f) = \psi_F([1 + \rho]^{\frac{1}{2}}f) \otimes 1 + 1 \otimes \psi_F^*(\rho^{\frac{1}{2}}f), \quad (4.17)$$
$$\psi^*(f) = \psi_F^*([1 + \rho]^{\frac{1}{2}}f) \otimes 1 + 1 \otimes \psi_F(\rho^{\frac{1}{2}}f). \quad (4.18)$$

Let us call D_F the linear hull of all vectors of the form

$$\psi_F^*(f_1) \cdots \psi_F^*(f_n)\Psi_{F0}.$$

Then $\psi(f)$ on $D_F \otimes D_F$ can be obtained by differentiation of $U(f)$ and $V(g)$. From Eq. (4.17) it follows that

$$\Psi_{F0} \otimes D_F \subset \overline{\mathbf{R}\Psi_0},$$

since $\{\rho^{\frac{1}{2}}f,\ f \in V\}$ is dense in V, and therefore

$$\Psi_{F0} \otimes H_F \subset \overline{\mathbf{R}\Psi_0},$$

since D_F is dense in H_F. It follows that

$$\overline{\{U_F(f)\}''\Psi_{F0}} \otimes H_F \subset \overline{\mathbf{R}\Psi_0}.$$

Since Ψ_{F0} is a cyclic vector for $\{U_F(f)\}''$ (see Lemma 2.4), we have

$$H = H_F \otimes H_F \subset \overline{\mathbf{R}\Psi_0}. \quad (4.19)$$

Thus we have constructed the desired representation. These results may be summed up as follows.

Theorem 4.1: The functional $E_\rho(f, g)$ of Eq. (4.8) defines a cyclic representation of the CCR's.

Reducibility of the Representation

We now investigate the mathematical structure of the operator algebra $\mathbf{R} = \{U(f), V(g)\}''$. By a trivial calculation one can show that the commutor $\mathbf{R}'$ contains the operators

$$\bar{U}(f) = U_F(\rho^{\frac{1}{2}}f) \otimes U_F([1+\rho]^{\frac{1}{2}}f), \tag{4.20}$$

$$\bar{V}(g) = V_F(-\rho^{\frac{1}{2}}g) \otimes V_F([1+\rho]^{\frac{1}{2}}g). \tag{4.21}$$

Thus $\mathbf{R}$ is reducible, as was the case with the ground-state representation. However $\mathbf{R}'$ is now noncommutative. We note that $\mathbf{R} \cup \mathbf{R}'$ contains

$$1 \otimes U_F(f) = U(-\rho^{\frac{1}{2}}f)\bar{U}([1+\rho]^{\frac{1}{2}}f),$$

$$1 \otimes V_F(g) = V(\rho^{\frac{1}{2}}g)\bar{V}([1+\rho]^{\frac{1}{2}}g),$$

$$U_F(f) \otimes 1 = U([1+\rho]^{\frac{1}{2}}f)\bar{U}(-\rho^{\frac{1}{2}}f),$$

$$V_F(g) \otimes 1 = V([1+\rho]^{\frac{1}{2}}g)\bar{V}(\rho^{\frac{1}{2}}g).$$

Thus $(\mathbf{R} \cup \mathbf{R}')''$ is irreducible, and $\mathbf{R} \cap \mathbf{R}' = (\mathbf{R} \cup \mathbf{R}')'$ which contains only multiples of the identity, i.e. $\mathbf{R}$ is a factor.

If the factor $\mathbf{R}$ were type I, then one could write H as a tensor product $H_a \otimes H_b$ such that $\mathbf{R}$ would be the algebra of all operators of the form $T_a \otimes 1$. The representation would then be a discrete direct sum of copies of one irreducible representation. We now show that this is not the case.

The proof that $\mathbf{R}$ is not type I involves showing that under the decomposition $H = H_a \otimes H_b$ the cyclic vector Ψ_0 must be of the form $\Psi_{a0} \otimes \Psi_{b0}$. This is impossible, as one can easily see in a variety of ways. In particular it would imply that $\overline{\mathbf{R}\Psi_0} = H_a \otimes \Psi_{b0}$, in contradiction with the fact that Ψ_0 is a cyclic vector for $\mathbf{R}$.

The following lemmas will be used to show that $\mathbf{R}$ type I implies $\Psi_0 = \Psi_{a0} \otimes \Psi_{b0}$.

Lemma 4.1: Let A and B be unitary operators. If the operator $A \otimes B$ has a discrete spectrum, then both A and B have discrete spectra.

Proof: Let E_A, E_B and E be the spectral measures of A, B and $A \otimes B$ respectively. Then

$$E(X) = E_A \otimes E_B[(x, y); xy \in X],$$

where X is any Borel set in the complex plane (E_A, E_B, and E are zero outside the unit circle). If D is a total set (i.e. a set whose linear hull is dense) in the Hilbert space, then whether or not E has a discrete spectrum is equivalent to whether or not the complex measure $\nu(X) = (\Psi, E(X)\Phi)$ has a discrete spectrum for some $\Psi, \Phi \in D$. We consider vectors of the form $\Psi = \Psi_A \otimes \Psi_B$ and $\Phi = \Phi_A \otimes \Phi_B$, then

$$\nu(X) = \nu_A \otimes \nu_B[(x, y); xy \in X],$$

where

$$\nu_A(X) = (\Psi_A, E_A(X)\Phi_A),$$

and

$$\nu_B(X) = (\Psi_B, E_B(X)\Phi_B).$$

Since ν_A and ν_B are a finite linear combination of nonnegative measures with finite total measure, the desired result can be established by proving that the measure

$$\mu(X) = \mu_A \otimes \mu_B[(x, y); xy \in X],$$

where μ_A and μ_B are nonnegative measures on the unit circle of the complex plane with total measure one, cannot have a discrete spectrum if μ_A has no discrete spectrum.

If μ_A has a continuous spectrum only, then there exists a sequence of points $\theta_0^N = 0, \theta_1^N, \cdots \theta_N^N = 2\pi$, such that

$$\mu_A(\Lambda_i^{AN}) = N^{-1},$$

where

$$\Lambda_i^{AN} = \{e^{i\theta}; \theta \in [\theta_{i-1}^N, \theta_i^N]\} \qquad j = 1, \cdots n.$$

We now prove that for any $e^{i\lambda}$, λ real, $\mu([e^{i\lambda}]) = 0$. Let

$$\Lambda_i^{BN} = \{e^{i\theta}; \theta \in [\lambda - \theta_i^N, \lambda - \theta_{i-1}^N], \text{mod } 2\pi\}.$$

Then

$$\{(x, y); |x| = |y| = 1, xy = e^{i\lambda}\} \subset \bigcup_{i=1}^{N} \Lambda_i^{AN} \times \Lambda_i^{BN}.$$

Hence

$$\mu\{[e^{i\lambda}]\} \leq N^{-1} \sum_{i=1}^{N} \mu_B(\Lambda_i^{BN}) = N^{-1}.$$

Thus

$$\mu\{[e^{i\lambda}]\} = 0,$$

QED.

A similar lemma holds for any normal operators A and B, provided zero is not a discrete eigenvalue of either operator.

Lemma 4.2: Given a Hilbert space $H = H_a \otimes H_b$ and a unitary operator U in H such that for all operators T_a in H_a and T_b in H_b we have

$$U(T_a \otimes 1)U^{-1} = T_a' \otimes 1,$$

$$U(1 \otimes T_b)U^{-1} = 1 \otimes T_b'.$$

Then there exist unitary operators U_a in H_a, and U_b in H_b such that $U = U_a \otimes U_b$.

Proof: By considering projection operators, it follows that $U(x \otimes y)$ where $x \in H_a$, $y \in H_b$, must be of the form $Cx' \otimes C^{-1}y'$ where C is an arbitrary complex number. For some fixed $y \in H_b$, choose some C such that $||C^{-1}y'|| = ||y||$. Then define an operator U_a in H_a by $U_a x = Cx'$.

It remains to define U_b. We have

$$U(x \otimes y) = U_a x \otimes y'(x),$$

where y' may depend on x. Now we have

$$(U[x_1 \otimes y],\ U[x_2 \otimes y]) = (x_1, x_2)(y, y),$$

but also

$$(U[x_1 \otimes y],\ U(x_2 \otimes y]) = (x_1, x_2)(y'(x_1), y'(x_2)).$$

It follows that $y'(x)$ does not depend on x. The operator U_b is now defined by

$$U_b y = y'(x).$$

The operators $U_a \otimes U_b$ and U coincide on all $\sum_i x_i \otimes y_i$ which are dense in H. It follows that $U = U_a \otimes U_b$. QED.

Now assume $\mathbf{R}$ is type I, hence $H = H_a \otimes H_b$. We have shown above that the algebra

$$\bar{\mathbf{R}} = \{\bar{U}(f),\ \bar{V}(g)\}'' \subset \mathbf{R}', \tag{4.22}$$

together with the algebra $\mathbf{R}$, generates all bounded operators in H. It follows that $\mathbf{R}' = \bar{\mathbf{R}}$. Now consider the translation operator $U(\mathbf{a}, 1)$. A trivial calculation shows that $\mathbf{R}$ and $\mathbf{R}'$ are invariant under $U(\mathbf{a}, 1)$, i.e. $U(\mathbf{a}, 1)$ satisfies the condition of lemma 4.2. It follows that $U(\mathbf{a}, 1) = U_a(\mathbf{a}, 1) \otimes U_b(\mathbf{a}, 1)$.

It is known that for any finite $\mathbf{a}$, $U_F(\mathbf{a}, 1)$ has only one discrete eigenvector Ψ_{F0}. It follows from lemma 4.1 that $U(\mathbf{a}, 1) = U_F(\mathbf{a}, 1) \otimes U_F(\mathbf{a}, 1)$ has only one discrete eigenvector $\Psi_0 = \Psi_{F0} \otimes \Psi_{F0}$. Now a second application of lemma 4.1 to the equation $U(\mathbf{a}, 1) = U_a(\mathbf{a}, 1) \otimes U_b(\mathbf{a}, 1)$ yields the result that the unique eigenvector Ψ_0 must be of the form $\Psi_{a0} \otimes \Psi_{b0}$. As we have seen above, this leads to a contradiction. Thus $\mathbf{R}$ is not type I. Since the CCR's are incompatible with the existence of a relative trace function on $\mathbf{R}$, $\mathbf{R}$ is never type II_1. Thus we have proved

Theorem 4.2: For the representation of the CCR's described in theorem 4.1 the operator algebra $\mathbf{R} = \{U(f),\ V(g)\}''$ is a factor either of type II_∞ or III_∞.

The decomposition of the representation into irreducible representations can be obtained as follows. We first construct the representation in a slightly different way. On the space V of test functions we define two operators K, L by

$$K = (1 + 2\rho)^{\frac{1}{2}}, \tag{4.23}$$

$$L = [4\rho(1 + \rho)/(1 + 2\rho)]^{\frac{1}{2}}, \tag{4.24}$$

where ρ is defined by Eq. (4.10). Then we define

$$H = H_F \otimes H_F, \tag{4.25}$$

$$U(f) = U_F(Kf) \otimes 1, \tag{4.26}$$

$$V(g) = V_F(K^{-1}g) \otimes V_F(Lg), \tag{4.27}$$

$$\Psi_0 = \Psi_{F0} \otimes \Psi_{F0}, \tag{4.28}$$

$$U(\mathbf{a}, R) = U_F(\mathbf{a}, R) \otimes U_F(\mathbf{a}, R). \tag{4.29}$$

By noting that $K^2 = K^{-2} + L^2$, we see that

$$E_\rho(f, g) = (\Psi_0,\ U(f)V(g)\Psi_0). \tag{4.30}$$

We omit the proof that Ψ_0 is a cyclic vector.

By lemma 2.4, $\{V_F(f)\}''$ is maximal abelian in H_F. Hence there is a spectral decomposition such that H_F is a L_2 space with measure μ_F over a space Z which is the *algebraic* dual of the space V of test functions, $V_F(g)$ is multiplication by $e^{i(\chi, g)}$, $\chi \in Z$, and the cyclic vector Ψ_0 is the function $\Psi_0(\chi) = 1$.[7,8] The measure μ_F is the normal distribution with unit variance, i.e. the restriction of μ_F to any finite-dimensional subspace of Z is the Gaussian measure

$$\pi^{-n/2} \exp -(x_1^2 + \cdots + x_n^2)\, dx_1 \cdots dx_n.$$

Using this decomposition for the second factor of $H_F \otimes H_F$ we obtain

$$E(f, g) = \int_Z E_\chi(f, g)\, d\mu_F(\chi), \tag{4.31}$$

where

$$E_\chi(f, g) = E_F(Kf, K^{-1}g)e^{i(\chi, Lg)} \tag{4.32}$$

defines an irreducible representation of the CCR's.

Physical Significance of the Reducibility

If $W \in \mathbf{R}'$ is unitary, then all operators in $\mathbf{R}$ have the same expectation value in $W\Psi$ as in Ψ. If $\mathbf{R} = \{U(f),\ V(g)\}''$ contains all observables, then Ψ and $W\Psi$ are physically indistinguishable. We would like to understand, at least intuitively, what the $W\Psi$ are. Recall that for the ground state representations, $\mathbf{R}' = \{a,\ a^*\}''$, where the operators a, a^* could be interpreted as the annihilation and creation operators for zero-momentum particles. In the same way, one can understand the reducibility

of the representations for discrete momentum distributions defined by Eq. (4.4). These representations are a direct integral of inequivalent irreducible representations obtained from the Fock representation by the transformation

$$\psi_F(\mathbf{x}) \to \psi(\mathbf{x}) = \psi_F(\mathbf{x}) + \sum_j e^{i\theta_j} e^{i\mathbf{k}_j\cdot\mathbf{x}} \rho_j^{\frac{1}{2}} . \quad (4.33)$$

However we do not yet understand the intuitive reason for the change in the mathematical structure of R when the momentum distribution becomes continuous.

While the occurence of representations other than type I was hardly anticipated, it is not clear that one can ascribe any particular significance to this fact. It should be noted that, in the general theory of group representations, the primary representations play a role similar to that of the irreducible representations in the type I case. Namely it is known that any type I representation of a separable locally compact group has a unique decomposition into irreducible representations.[21] The generalization of this theorem to arbitrary representations yields a direct integral decomposition into primary representations.[22] However the physical significance of this fact is, to say the least, not very clear.

The mathematical structure of the representation, as we have constructed it, is rather suggestive. Eq. (4.17) indicates that $\psi(f)$ can be interpreted as the sum of an annihilation operator on the first space (or first kind of particle) and a creation operator on the second space (or second kind of particle). This immediately brings to mind the particle–antiparticle description in elementary particle physics. It should also be noted that the operators $\bar{U}(f)$, $\bar{V}(g)$ defined by Eqs. (4.20–21) which commute with $U(f)$, $V(g)$ and, together with $U(f)$, $V(g)$ are irreducible, give a representation of the CCR's unitary equivalent to $U(f)$, $V(g)$. However we do not understand the significance of these remarks.

The Number Density Operator

According to lemma 2.6 it is sufficient to consider the cyclic vector Ψ_0. Let f_n be a complete set of functions for the volume V. Using Eqs. (4.17–18) we have

$$N(f_n)\Psi_0 = \psi_F^*([1+\rho]^{\frac{1}{2}} f_n)\Psi_{F0} \otimes \psi_F^*(\rho^{\frac{1}{2}} f_n)\Psi_{F0} + (f_n, \rho f_n)\Psi_0. \quad (4.34)$$

Now

$$\sum_n (f_n, \rho f_n) = \int d\mathbf{k}\rho(\mathbf{k}) \sum_n |\tilde{f}_n(\mathbf{k})|^2. \quad (4.35)$$

[21] G. W. Mackey, Trans. Amer. Math. Soc. 85, 134, (1957).
[22] J. A. Ernest (preprint).

Consider the function

$$g_{\mathbf{k}}(\mathbf{x}) = \begin{cases} e^{-i\mathbf{k}\cdot\mathbf{x}} & \mathbf{x} \in V \\ 0 & \mathbf{x} \text{ not in } V. \end{cases} \quad (4.36)$$

Then $\tilde{f}_n(\mathbf{k}) = (g_{\mathbf{k}}, f_n)$ and, by Parseval's formula, we have

$$\sum_n |\tilde{f}_n(\mathbf{k})|^2 = ||g_{\mathbf{k}}||^2 = V. \quad (4.37)$$

Thus

$$\sum_n (f_n, \rho f_n) = \rho V, \quad (4.38)$$

where

$$\rho = \int d\mathbf{k}\rho(\mathbf{k}). \quad (4.39)$$

It follows that

$$N_V\Psi_0 = \sum_n N(f_n)\Psi_0 = \chi_V + \rho V\Psi_0, \quad (4.40)$$

where

$$||\chi_V||^2 = \sum_{n,m} (f_n, [1+\rho]f_m)(f_n, \rho f_m) = \sum_n (f_n, [\rho+\rho^2]f_n) = (\rho+\sigma)V, \quad (4.41)$$

and

$$\sigma = \int d\mathbf{k}\rho^2(\mathbf{k}) \quad (4.42)$$

is finite for those $\rho(\mathbf{k})$ of interest. Thus

$$||V^{-1}\chi_V|| = (\rho+\sigma)^{\frac{1}{2}} V^{-\frac{1}{2}}, \quad (4.43)$$

and we obtain

$$\lim_{V\to\infty} V^{-1} N_V\Psi_0 = \rho\Psi_0. \quad (4.44)$$

To obtain the momentum distribution of the particles, assume the region V is rectangular and choose the $f_n(\mathbf{x})$ to be eigenfunctions of the operator $i\nabla$ with periodic boundary conditions. The number operator $N(V, \mathbf{k}_1, \mathbf{k}_2)$ for the number of particles in the volume V with momenta in the interval $(\mathbf{k}_1, \mathbf{k}_2)$ is then given by

$$N(V, \mathbf{k}_1, \mathbf{k}_2) = \sum_n N(f_n),$$

where the sum includes only those values of n such that the eigenvalue to which $f_n(\mathbf{x})$ belongs is in $(\mathbf{k}_1, \mathbf{k}_2)$. We have

$$N(V, \mathbf{k}_1, \mathbf{k}_2)\Psi_0 = \Phi_V + \sum_n (f_n, \rho f_n)\Psi_0. \quad (4.45)$$

Since $V^{-1}\sum_n |\tilde{f}_n(\mathbf{k})|^2$ is uniformly bounded by one

and

$$\lim_{V\to\infty} V^{-1} \sum_n |\tilde{f}_n(\mathbf{k})|^2 = \begin{cases} 1 & \text{if } \mathbf{k} \in (\mathbf{k}_1, \mathbf{k}_2) \\ 0 & \text{if } \mathbf{k} \text{ not in } [\mathbf{k}_1, \mathbf{k}_2], \end{cases} \tag{4.46}$$

we have

$$\lim_{V\to\infty} V^{-1} \sum_n (f_n, \rho f_n) = \rho(\mathbf{k}_1, \mathbf{k}_2) \equiv \int_{\mathbf{k}_1}^{\mathbf{k}_2} d\mathbf{k}\rho(\mathbf{k}). \tag{4.47}$$

The same argument that led to Eq. (4.43) can be used to obtain

$$||V^{-1}\Phi_V|| = O(V^{-\frac{1}{2}}). \tag{4.48}$$

Thus we have

$$\lim_{V\to\infty} V^{-1}N(V, \mathbf{k}_1, \mathbf{k}_2)\Psi_0 = \rho(\mathbf{k}_1, \mathbf{k}_2)\Psi_0, \tag{4.49}$$

which is the desired result.

In view of definition 2.1, the above argument is unsatisfactory in that we have used the unbounded operators N_V, etc. rather than the unitary operators $e^{i\lambda N_V}$, etc. It seems very likely that the unitary operators have the desired values, but we have not done the necessary calculations.

The Hamiltonian

For the free Bose gas in a finite box, the time-dependent field $\psi(f, t)$ satisfies the equation

$$\psi(f, t) = \psi(f_t), \tag{4.50}$$

where

$$i\,\partial f_t/\partial t = -(\Delta/2m)f_t(\mathbf{x}),$$

and

$$f_{t=0} = f.$$

Clearly the operator

$$U(t) = U_F(t) \otimes U_F(t)^* \tag{4.51}$$

satisfies

$$U(t)\psi(f)U^{-1}(t) = \psi(f, t), \tag{4.52}$$

or an equivalent commutation relation with $U(f)V(g)$, and

$$U(t)\Psi_0 = \Psi_0. \tag{4.53}$$

Since Ψ_0 is cyclic for $\mathbf{R}$, Eqs. (4.52–53) uniquely determine $U(t)$.

For the ground-state representation where $H_V\Psi_0 = 0$, we saw that the infinitesimal generator of the time displacements was given by

$$H = \lim_{V\to\infty} H_V,$$

provided one chose the appropriate formal expression for H_V. In the present case the state Ψ_0 obviously has an infinite total energy. This might lead one to to expect that

$$H = \lim_{V\to\infty} [H_V - (\Psi_0, H_V\Psi_0)],$$

i.e. a renormalization. This is not the case. In fact we now prove that the operators $U(t)$ and $U(\mathbf{a}, R)$ are not functions of the field operators.

Lemma 4.3: $U(t)$, $U(\mathbf{a}, R)$ do not belong to $\mathbf{R} = \{U(f), V(g)\}''$.

Proof: $U(t)$, $U(\mathbf{a}, R)$ do not commute with $\tilde{U}(f)$, $\tilde{V}(g)$ which belong to $\mathbf{R}'$ [see Eqs. (4.20–22)]. QED.

Macroscopic Occupation of the Ground State

We now consider the representation of the CCR's defined by Eq. (4.9). The following construction is a straightforward synthesis of Eqs. (3.15–19) and (4.11–15).

$$H = H_F \otimes H_F \otimes M, \tag{4.54}$$

$$U(f) = U_F([1 + \rho]^{\frac{1}{2}}f) \otimes U_F(\rho^{\frac{1}{2}}f) \otimes \exp\,[i(2\rho_0)^{\frac{1}{2}}A\tilde{f}(0)], \tag{4.55}$$

$$V(g) = V_F([1 + \rho]^{\frac{1}{2}}g) \otimes V_F(-\rho^{\frac{1}{2}}g) \otimes \exp\,[i(2\rho_0)^{\frac{1}{2}}B\tilde{g}(0)], \tag{4.56}$$

$$\Psi_0 = \Psi_{F0} \otimes \Psi_{F0} \otimes \Phi_0, \tag{4.57}$$

$$U(\mathbf{a}, R) = U_F(\mathbf{a}, R) \otimes U_F(\mathbf{a}, R) \otimes 1, \tag{4.58}$$

$$U(t) = U_F(t) \otimes U_F(t)^* \otimes 1. \tag{4.59}$$

To prove that Ψ_0 is cyclic for $\mathbf{R} = \{U(f), V(g)\}''$ it suffices to show that $\mathbf{R}$ contains $1 \otimes 1 \otimes A$ and $1 \otimes 1 \otimes B$ (recall proof of cyclicity for the ground-state representation). To do this, consider the functions $f_s(\mathbf{x})$ whose Fourier transforms are

$$\tilde{f}_s(\mathbf{k}) = (2\pi)^{-\frac{1}{2}}\theta(s - k), \tag{4.60}$$

where

$$\theta(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x < 0. \end{cases}$$

Then

$$\lim_{s\to 0} ||\rho^{\frac{1}{2}}f_s|| = 0, \tag{4.61}$$

and

$$\lim_{s\to 0} ||[1 + \rho]^{\frac{1}{2}}f_s|| = 0, \tag{4.62}$$

but

$$\tilde{f}_s(0) = 1. \tag{4.63}$$

Hence by lemma 2.3,

$$\lim_{s\to 0} U(f_s) = 1 \otimes 1 \otimes \exp[i(2\rho_0)^{\frac{1}{2}}A], \tag{4.64}$$

and $\mathbf{R}$ contains $1 \otimes 1 \otimes A$. Similarly one can show that $\mathbf{R}$ contains $1 \otimes 1 \otimes B$.

The decomposition of the ground-state representation into a direct integral of irreducible representations has an obvious analog here. Namely the representation is a direct integral of primary (or factor) representations defined by

$$H_\theta = H_F \otimes H_F, \tag{4.65}$$

$$\Psi_{\theta 0} = \Psi_{F0} \otimes \Psi_{F0}, \tag{4.66}$$

$$U_\theta(f) = U_F([1+\rho]^{\frac{1}{2}}f) \otimes U_F(\rho^{\frac{1}{2}}f) \times \exp[i(2\rho_0)^{\frac{1}{2}}\tilde{f}(0)\sin\theta], \tag{4.67}$$

$$V_\theta(g) = V_F([1+\rho]^{\frac{1}{2}}g) \otimes V_F(-\rho^{\frac{1}{2}}g) \times \exp[i(2\rho_0)^{\frac{1}{2}}\tilde{g}(0)\cos\theta], \tag{4.68}$$

$$U_\theta(\mathbf{a}, R) = U_F(\mathbf{a}, R) \otimes U_F(\mathbf{a}, R), \tag{4.69}$$

$$U_\theta(t) = U_F(t) \otimes U_F(t)^*, \tag{4.70}$$

where $0 \leq \theta \leq 2\pi$. Thus the representation is a direct integral of inequivalent primary representations with multiplicity one.

These primary representations are related to the representation with $\rho_0 = 0$ just as the ground-state representations are related to the Fock representation. That is, if the subscript ρ' refers to the representation with no macroscopic occupation of the ground state, then the above irreducible representations are given by

$$\psi_\theta(\mathbf{x}) = \psi_{\rho'}(\mathbf{x}) + e^{i\theta}\rho_0^{\frac{1}{2}}. \tag{4.71}$$

These representations can be labelled by the average field strengths. Consider the operator $U_\theta(\chi_V)$ where $\chi_V(\mathbf{x}) = V^{-1}$, $\mathbf{x} \in V$. We have

$$\lim_{V\to\infty} ||\rho^{\frac{1}{2}}\chi_V|| = 0, \tag{4.72}$$

and

$$\lim_{V\to\infty} ||[1+\rho]^{\frac{1}{2}}\chi_V|| = 0. \tag{4.73}$$

It then follows from lemma 2.3 that

$$\lim_{V\to\infty} U_\theta(\chi_V) = \exp[i(2\rho_0)^{\frac{1}{2}}\tilde{f}(0)\sin\theta], \tag{4.74}$$

and similarly for $V_\theta(\chi_V)$. Thus we have

$$\langle\psi_\theta\rangle = \lim_{V\to\infty} V^{-1}\int_V d\mathbf{x}\psi(\mathbf{x}) = e^{i\theta}\rho_0^{\frac{1}{2}}. \tag{4.75}$$

Since $\langle\psi_\theta\rangle$ is a unitary invariant, this proves the inequivalence of the representations belonging to different values of θ. These results may be summed up as follows.

Theorem 4.3: The functional $E(f, g)$ of Eq. (4.9) defines a cyclic representation of the CCR's. This representation is a direct integral of inequivalent primary representations $U_\theta(f)$, $V_\theta(g)$ where $0 \leq \theta \leq 2\pi$.

The proof that a number density operator exists, and that the representation does describe a system of particles with the distribution $\rho(\mathbf{k})$ in momentum space, etc. is similar to the discussion for the ground state representations (just replace the Fock representation by the representation of Theorem 4.1). We therefore omit the proofs.

5. DENSITY MATRIX FOR THE INFINITE FREE BOSE GAS

The first step in a theory of the statistical mechanics of an infinite system would seem to be the construction of a suitable density matrix. We give first a plausible argument that to each value of temperature β and chemical potential μ there corresponds an inequivalent representation of the CCR's. A systematic (but not necessarily unambiguous) method for constructing the algebra of observables is presented. The discussion of the density matrix is then based on a consideration of expectation values of observable quantities.

Consider a finite system described by a thermal ensemble $\sigma(\beta, \mu)$. A physical quantity is obtained by averaging the corresponding observable X over the ensemble, i.e. by computing *Trace* $X\sigma$/*Trace* σ. Since the eigenstates of the energy operator can be chosen as products of single-particle states with definite energy and momenta, the calculation involves evaluating the expectation value of the operator X in a state with n_1 particles of momentum $\mathbf{k}_1$, n_2 particles of momentum $\mathbf{k}_2$, etc. If there is no condensation (no macroscopic occupation of the zero-momentum state) then, as the volume of the box gets very large, most of the states in the ensemble will have essentially the same distribution in momentum space. Thus in the limit $V \to \infty$, a single momentum distribution $\rho(\mathbf{k}, \beta, \mu)$ prevails. Since the inequivalent representations of the CCR's which we constructed are labeled by the density distribution $\rho(\mathbf{k})$ in momentum space, this suggests that to each value of the chemical potential μ and temperature β there is associated an inequivalent representation of the

CCR's—namely that representation belonging to $\rho(\mathbf{k}) = \rho(\mathbf{k}, \beta, \mu)$.

These qualitative arguments are confirmed by a calculation of the expectation value of the operator $U(f)V(g)$ over the grand canonical ensemble. The result of this calculation, which is given in appendix 1, is that

$$\lim_{V\to\infty} \frac{\text{Trace } U(f)V(g) \exp(\beta\mu N - \beta H)}{\text{Trace } \exp(\beta\mu N - \beta H)} = E_F(f, g) \exp\left\{-\frac{1}{2}\int \frac{d\mathbf{k}}{(2\pi)^3} \frac{\exp(\beta\mu - \beta E_k)}{1 - \exp(\beta\mu - \beta E_k)} \times [|\tilde{f}(\mathbf{k})|^2 + |\tilde{g}(\mathbf{k})|^2]\right\}, \quad (5.1)$$

where $E_k = k^2/2m$. Thus the limit of the ensemble average of the operator $U(f)V(g)$ is correctly given by its expectation value in the cyclic state Ψ_0 in the representation belonging to

$$\rho(\mathbf{k}, \beta, \mu) = (2\pi)^{-3}[\exp(\beta E_k - \beta\mu) - 1]^{-1}. \quad (5.2)$$

When there is condensation the fluctuations in momentum space do not approach zero as $V \to \infty$. If one uses the grand canonical ensemble to describe the finite system, macroscopic fluctuations will occur only in the zero-momentum state. Thus one would expect the ensemble average of $U(f)V(g)$ to approach an integral over ρ_0 (the density of particles in the zero-momentum state) of the $E(f, g)$ of Eq. (4.9). The calculation of the ensemble average when there is condensation present is complicated by the fact that $\mu = 0$ for the limit of an infinite volume. If one carries out the calculation for zero temperature and lets $\mu \to 0$ as $V \to \infty$ so that the total particle density ρ remains constant, then one obtains

$$E(f, g; \rho, \beta = \infty) = \int dr(r/\rho) \exp(-r^2/2\rho) E_r(f, g), \quad (5.3)$$

where

$$E_r(f, g) = E_F(f, g) J_0\{[2r(\tilde{f}(0)^2 + \tilde{g}(0)^2)]^{\frac{1}{2}}\}. \quad (5.4)$$

It seems likely that if the macrocanonical ensemble were used, the limit of the ensemble average of $U(f)V(g)$ would just be $E_\rho(f, g)$.

Thus the limit of the ensemble average of $U(f)V(g)$ is equal to the expectation value of $U(f)V(g)$ in the cyclic state Ψ_0 in the appropriate representation. However, this does not necessarily imply that the thermal ensemble for the infinite free Bose gas should be a pure state. First, the algebra $\mathbf{R}$ may be reducible (cf. criterion for pure state, below). Second, it is not clear which operators are observables for the infinite system. Third, the fact that the expectation value of $U(f)V(g)$ approaches a limit does not necessarily imply the convergence of different functions of the field operators.

In order to discuss the thermal ensemble for an infinite system, we must adopt a convenient criterion as to whether a given physical state Ψ is a pure state or a mixture. Consider the positive linear functional $(\Psi, X\Psi)$ defined on the algebra of observables $\mathfrak{A}$ as the expectation value of the observable X in the state Ψ. Then the state Ψ is a pure state if and only if it is impossible to write $(\Psi, X\Psi)$ as a convex linear combination of two distinct positive linear functionals. This is equivalent to saying that Ψ is pure if and only if $\mathfrak{A}$ is irreducible on $\overline{\mathfrak{A}\Psi}$ (the subspace spanned by all $X\Psi$, $X \in \mathfrak{A}$).

The Algebra of Observables

In order to apply these remarks, one must first decide which operators are observables. For the finite system the algebra $\mathbf{R} = \{U_F(f), V_F(g)\}''$ is irreducible, but the algebra of observables $\mathfrak{A}$ is not. (In fact, $\mathfrak{A} = \{N\}'$, see discussion of the Fock representation in Sec. 2.) We indicate here a systematic method for constructing the algebra of observables. Although not unambiguous, it is mathematically rigorous. As in the case of the CCR's, we shall use unitary operators in order to avoid any difficulties due to unbounded operators.

We consider a class of Hermitian operators K on the space of L_2 functions as follows. Let n be any positive integer, $f_1 \cdots f_n$ be any orthonormal set of functions in the space $\mathcal{S}$ of Schwartz,[9] and $\lambda_1 \cdots \lambda_n$ any set of real numbers. The operator

$$K = \sum_{i=1}^{n} \lambda_i (f_i \otimes f_i^*) \quad (5.5)$$

is defined by

$$Kg = \sum_{i=1}^{n} \lambda_i (f_i, g) f_i, \qquad g \in L_2. \quad (5.6)$$

The set of all such Hermitian operators K will be denoted here by $\mathfrak{F}$. We then define a unitary operator $\exp[iQ_F(K)]$ on the Hilbert space of the Fock representation by

$$\exp[iQ_F(K)]\, \psi_F^*(f_1) \cdots \psi_F^*(f_n)\Psi_{F0} = \psi_F^*[e^{iK}f_1] \cdots \psi_F^*[e^{iK}f_n]\Psi_{F0}, \quad (5.7)$$

and

$$\exp[iQ_F(K)]\, \Psi_{F0} = \Psi_{F0}. \quad (5.8)$$

The self-adjoint generator is given by[23]

$$Q_F(K) = \sum_{i=1}^{n} \lambda_i \psi_F^*(f_i)\psi_F(f_i^*). \tag{5.9}$$

We now prove that the von Neumann algebra

$$\mathbf{Q}_F = \{\exp [iQ_F(K)]\}'' \tag{5.10}$$

is the algebra of observables for the finite system.

Lemma 5.1: Given the Fock representation for the volume V with total number operator N, then $\mathbf{Q}_F = \{N\}'$.

Proof: Since the $\exp [iQ_F(K)]$ do not change particle number we have $\mathbf{Q}_F \subset \{N\}'$. Hence the subspace H_F^n with fixed particle number n is invariant under $\mathbf{Q}_F$. It remains only to show that $\mathbf{Q}_F$ is irreducible on H_F^n. To prove this we will use the fact that if Ψ is a cyclic vector for a von Neumann algebra $\mathbf{A}$, and if the projection on Ψ is in $\mathbf{A}$, then $\mathbf{A}$ is irreducible.

Let $K_n(\mathbf{x}, \mathbf{y}) = t \sum_{i=1}^{n} f_i(\mathbf{x})f_i(\mathbf{y})$ where f_i is a set of real functions in $\mathcal{S}$ forming a basis of $L_2(V)$. Then

$$\exp (itN) = \lim_{n\to\infty} \exp [iQ_F(K_n)].$$

It follows that the projection operator E_n onto H_F^n is in $\mathbf{Q}_F$. Since $Q_F(K)E_n$ is bounded it is in $\mathbf{Q}_F$. Now the projection operator onto $\psi_F^*(f_1)^n\Psi_{F0}$ is

$$P_n = (n!)^{-1}\psi_F^*(f_1)^n\psi_F(f_1)^nE_n$$

$$= (n!)^{-1}\prod_{j=1}^{n} [Q_F(f_1 \otimes f_1) - (n-j)]E_n,$$

which is in $\mathbf{Q}_F$.

Now any vector Φ in H_F^n can be written as

$$\Phi = \sum_{\Sigma n_i = n} C(n_1, n_2 \cdots)\psi_F^*(f_1)^{n_1}\psi_F^*(f_2)^{n_2} \cdots \Psi_{F0}.$$

By taking linear combinations with different $K(\mathbf{x}, \mathbf{y})$, $Q_F(K)E_n$ with non-Hermitian $K(\mathbf{x}, \mathbf{y})$ is also in $\mathbf{Q}_F$. Since

$$(n!)^{-1}\left[\prod_{i=2}^{\infty} Q_F(f_i \otimes f_1)^{n_i}\right]\psi_F^*(f_1)^n\Psi_{F0}$$

$$= \psi_F^*(f_1)^{n_1}\psi_F^*(f_2)^{n_2} \cdots \Psi_{F0},$$

it follows that $\psi_F^*(f_1)^n\Psi_{F0}$ is a cyclic vector for $\mathbf{Q}_F$ in H_F^n. Hence $\mathbf{Q}_F$ is irreducible on H_F^n. QED.

It is at least plausible to assume that the algebra of observables for the infinite system is the limit, in some sense, of the algebra $\mathbf{Q}_F$ as the volume $V \to \infty$. We now show how the representations of the group G of operators e^{iK} can be studied by the same methods which we used for the representations of the CCR's. Since the discussion is very similar to that given in Sec. 2, we only sketch the details.

We have a map $e^{iK} \to e^{iQ(K)}$ from a group G of operators on a space of complex test functions, into the unitary operators on a Hilbert space. From the Fock representation one obtains the multiplication rule

$$\exp [iQ(K_1)] \exp [iQ(K_2)]$$

$$= \exp [iQ(K_1 \circ K_2)], \tag{5.11}$$

where $\exp (iK_1 \circ K_2)$ is defined by

$$\exp (iK_1 \circ K_2) = \exp (iK_1) \exp (iK_2). \tag{5.12}$$

Eq. (5.11) follows immediately from Eq. (5.7). Although $K_1 \circ K_2$ defined by Eq. (5.12) is not unique, Eq. (5.11) holds for any choice. We now prove that if $K_1, K_2 \in \mathfrak{F}$, then there exists some $K = K_1 \circ K_2$ in $\mathfrak{F}$. Since $\exp (iK_1) \exp (iK_2)$ is unitary, one may write it as $\exp (iK)$. Let

$$K_1 = \sum_{i=1}^{m} \lambda_i f_i \otimes f_i^* \quad \text{and} \quad K_2 = \sum_{i=1}^{n} \mu_i g_i \otimes g_i^*,$$

and let P be the projection on the subspace of L_2 spanned by f_i and g_i, $i = 1, \cdots m$, $j = 1, \cdots n$. We easily see that $(1 - P) \exp (iK) = \exp (iK)(1-P) = (1 - P)$, from which it follows that K can be taken as an operator acting on the finite-dimensional subspace specified by P. Hence K can be taken in $\mathfrak{F}$. Incidentally, we also see by a similar proof that if $f \in L_1$ and $K \in \mathfrak{F}$, then $e^{iK} f \in L_1$, which will be used later.

We also require the commutation relation of $\exp [iQ_F(K)]$ with $U_F(f)V_F(g)$. From the formulas

$$U_F(f)V_F(g) = \exp \{[i\psi_F^*(f) - \psi_F^*(g)]/\sqrt{2}\}$$

$$\times \exp \{[i\psi_F(f) + \psi_F(g)]/\sqrt{2}\}E_F(f, g), \tag{5.13}$$

$$e^{iQ_F(K)}\psi_F^*(f)e^{-iQ_F(K)} = \psi_F^*(e^{iK}f), \tag{5.14}$$

we obtain

$$\exp [iQ_F(K)]U_F(f)V_F(g) \exp [-iQ_F(K)]$$

$$= U_F(f_R - g_I)V_F(g_R + f_I)$$

$$\times \exp \{-i[(f, g) - (f_R - g_I, g_R + f_I)]/2\}, \tag{5.15}$$

[23] If we consider only those $U(f_i)$, $V(f_i)$, $i = 1, \ldots n$ which are relevant for the definition of $Q(K)$ for a given K and forget all other $U(f)$, $V(g)$, then we have a representation of the CCR's for a system of finitely many degrees of freedom. Using the classical result[10] that any unitary representation is a discrete direct sum of copies of one unique irreducible representation, we can easily define $Q(K)$ in each irreducible component just as in the Fock representation. Thus in a representation of the CCR's for an infinite volume, there are two ways of defining $Q(K)$. First, from the $\psi(f)$ and $\psi^*(f)$ for the infinite system as discussed in this footnote, and second, by taking the limit as $V \to \infty$ of $Q_F(K)$ which will be discussed in the main text immediately. In general, these two definitions need not coincide (if they both exist). If they differ, the physically relevant operator seems to be the second one.

where

$$f_R(\mathbf{x}) = \mathrm{Re}\,\{[e^{iK}f](\mathbf{x})\}$$
$$= \mathrm{Re}\left\{f(\mathbf{x}) + \int d\mathbf{y}[e^{iK} - 1](\mathbf{x}, \mathbf{y})f(\mathbf{y})\right\}, \quad (5.16)$$

$$f_I(\mathbf{x}) = \mathrm{Im}\,\{[e^{iK}f](\mathbf{x})\}. \quad (5.17)$$

Note that $e^{iK} - 1$ can be considered as an integral operator and thus as a function of two variables $\mathbf{x}$, $\mathbf{y}$.

Since a representation of the CCR's with a cyclic vector Ψ is uniquely determined by the functional

$$E(f, g) = (\Psi, U(f)V(g)\Psi),$$

this suggests that we consider the functional

$$E(K, f, g) = (\Psi, \exp[iQ(K)]U(f)V(g)\Psi). \quad (5.18)$$

Using the multiplication and commutation rules for the operators $\exp[iQ(K)]$, $U(f)$, $V(g)$ one obtains a straightforward generalization of Lemma 2.1. However we shall not write down the positive definiteness condition on $E(K, f, g)$ here. For our present purposes, it is sufficient to note that any functional $E(K, f, g)$ obtained by a limiting process from functionals satisfying the positive-definiteness condition, is positive definite and therefore defines a representation of the operators $\exp[iQ(K)]$, $U(f)$, $V(g)$. Finally, we remark that Lemma 2.3 can also be extended to the functional $E(K, f, g)$.

Zero-Temperature Case

We give first a heuristic discussion using the unbounded operators $\psi(f)$, in order to provide an intuitive understanding of the more systematic approach involving the unitary operators $e^{iQ(K)}$.

The irreducible representations describing the infinite free Bose gas at zero temperature are related to the Fock representation by

$$\psi_\theta(\mathbf{x}) = \psi_F(\mathbf{x}) + e^{i\theta}\rho^{\frac{1}{2}}. \quad (5.19)$$

In our discussion of the physical significance of the reducibility (see Sec. 3), we noted that the operator

$$a = \int^{\oplus} e^{i\theta}\, d\theta/2\pi \quad (5.20)$$

could be interpreted as an annihilation operator for zero-momentum particles. This suggests that we introduce a "renormalized" number operator N_r in the Hilbert space $H = H_F \otimes M$ as follows. As a basis for M, we choose the functions

$$\Phi_n(\theta) = e^{-in\theta}, \qquad n = 0, \pm 1, \pm 2, \cdots. \quad (5.21)$$

Define an operator N_0 on M by

$$N_0\Phi_n = n\Phi_n. \quad (5.22)$$

Let N_F denote the total number operator for the Fock representation. Then a relative number operator for the infinite system is defined by

$$N_r = N_F \otimes 1 + 1 \otimes N_0, \quad (5.23)$$

that is, N_r just gives the particle number relative to the state $\Psi_0 = \Psi_{F0} \otimes \Phi_0$.

The next step is to separate the Hilbert space $H = H_F \otimes M$ according to the relative particle number. The space M is decomposed as

$$M = \sum_{k=-\infty}^{\infty}{}^{\oplus} M_k, \quad (5.24)$$

where M_k is the one-dimensional space containing Φ_k. H_F is decomposed according to N_F as

$$H_F = \sum_{n=0}^{\infty}{}^{\oplus} H_F^n \quad (5.25)$$

[see Eq. (2.13)]. The desired decomposition of H is then given by

$$H = \sum_{n=-\infty}^{\infty}{}^{\oplus} H_n, \quad (5.26)$$

where

$$H_n = \sum_{k=0}^{\infty}{}^{\oplus} H_F^k \otimes M_{n-k}. \quad (5.27)$$

It is clear that if one takes an observable quantity for the finite free Bose gas and makes the transition to the infinite system by means of the equation

$$Q(K) = \sum_{i=1}^{n} \lambda_i \psi^*(f_i)\psi(f_i^*), \quad (5.28)$$

where

$$K = \sum_{i=1}^{n} \lambda_i f_i \otimes f_i^*,$$

that each H_n is invariant under the resulting operator (if it exists). Furthermore the algebra of observables obtained in this way is irreducible on each H_n (this is perhaps not quite so obvious). We turn now to the systematic construction of the algebra of observables.

We will determine the representation of the operators $e^{iQ(K)}$ for the infinite free Bose gas at zero temperature by computing the functional $E(K, f, g)$. The analog of Eq. (3.3) is

$$E_V(K, f, g) = (\Psi_V, e^{iQ_F(K)}U_F(f)V_F(g)\Psi_V). \quad (5.29)$$

By the same method used to derive Eq. (3.7) one obtains [use Eqs. (5.13–14) as well as (3.4–6)]

$$E_V(K, f, g) = L_n[\tfrac{1}{2}(g + if, f_V) \times (e^{iK}g - ie^{iK}f, f_V)/(f_V, e^{iK}f_V)] \times (f_V, e^{iK}f_V)^n E_F(f, g), \quad (5.30)$$

where $n = \rho V$. Taking the limit $V \to \infty$ with ρ held fixed we obtain

$$E(K, f, g) = J_0\{(2\rho[\tilde{g}(0) + i\tilde{f}(0)]\,[\tilde{g}_K(0) - i\tilde{f}_K(0)])^{\frac{1}{2}}\} \times \exp\left\{\rho \int d\mathbf{x}\, d\mathbf{y}[e^{iK} - 1](\mathbf{x}, \mathbf{y})\right\} E_F(f, g), \quad (5.31)$$

where $g_K = e^{iK}g$.

We now want to construct an operator $e^{iQ(K)}$ in the Hilbert space of the representation of theorem 3.1 such that

$$E(K, f, g) = (\Psi_0, e^{iQ(K)}U(f)V(g)\Psi_0).$$

The method of construction is suggested by Eqs. (5.19) and (5.28). On the basis of a formal calculation, we define the unitary operator $e^{iQ_\theta(K)}$ in H_F as follows

$$e^{iQ_\theta(K)} = e^{iQ_F(K)}U(f)V(g)e^{i\gamma}, \quad (5.32)$$

where the functions f, g are obtained by applying the integral operators

$$-i[(1 - e^{-iK})e^{i\theta} + (e^{i\tilde{K}} - 1)e^{-i\theta}]/\sqrt{2},$$

and

$$[(e^{i\tilde{K}} - 1)e^{-i\theta} - (1 - e^{-iK})e^{i\theta}]/\sqrt{2},$$

to the identity function $\chi(\mathbf{x}) = 1$, respectively, where $\tilde{K}(\mathbf{x}, \mathbf{y})$ is the transposed matrix $K(\mathbf{y}, \mathbf{x})$, and

$$\gamma = \int d\mathbf{x}\, d\mathbf{y}\{[\sin K](\mathbf{x}, \mathbf{y}) + \tfrac{1}{4}e^{-2i\theta}[(e^{iK} - 1)(e^{i\tilde{K}} - 1)](\mathbf{x}, \mathbf{y}) - \tfrac{1}{4}e^{2i\theta}[(1 - e^{-i\tilde{K}})(1 - e^{-iK})](\mathbf{x}, \mathbf{y})\}.$$

Thus we obtain a well-defined unitary operator

$$e^{iQ(K)} = \int^{\oplus} e^{iQ_\theta(K)}\, d\theta/2\pi \quad (5.33)$$

in $H = H_F \otimes M$. To verify that this is the desired operator, one can calculate the expectation value

$$(\Psi_0, e^{iQ(K)}U(f)V(g)\Psi_0) = \int (\Psi_{F0}, e^{iQ_\theta(K)}U_\theta(f)V_\theta(g)\Psi_{F0})\, d\theta/2\pi, \quad (5.34)$$

and show that it agrees with $E(K, f, g)$ given by Eq. (5.31). One can also verify the multiplication and commutation rules for $Q(K)$. That the infinitesimal generator of $e^{iQ(K)}$ is in fact

$$Q(K) = \sum_{i=1}^{n} \lambda_i[\psi_F^*(f_i) \otimes 1 + 1 \otimes e^{-i\theta}\tilde{f}_i(0)\rho^{\frac{1}{2}}] \times [\psi_F(f_i^*) \otimes 1 + 1 \otimes e^{i\theta}\tilde{f}_i(0)^*\rho^{\frac{1}{2}}], \quad (5.35)$$

where

$$K = \sum_{i=1}^{n} \lambda_i(f_i \otimes f_i^*)$$

can be proved by differentiation.

Clearly each subspace H_n occurring in the decomposition of H [see Eq. (5.27)] is invariant under

$$\mathbf{Q} = \{e^{iQ(K)}\}''.$$

It remains to prove that $\mathbf{Q}$ is irreducible when restricted to H_n.

Lemma 5.3: $\{f: f \in L_2 \cap L_1$ and $\tilde{f}(0) = 0\}$ is dense in L_2.

Proof: Any function in the class $\mathcal{S}$ of Schwartz test functions, whose Fourier transform vanishes at the origin, is in the set under consideration. But the set of such functions in $\mathcal{S}$ are obviously dense in L_2. QED.

Lemma 5.4: $\mathbf{Q} = \{e^{iQ(K)}\}''$ is irreducible when restricted to H_n.

Proof: We again use the fact that if Ψ is a cyclic vector for a von Neumann algebra $\mathbf{A}$, and if the projection on Ψ is in $\mathbf{A}$, then $\mathbf{A}$ is irreducible.

If $K(\mathbf{x}, \mathbf{y}) = \sum f_i(\mathbf{x})g_i(\mathbf{y})$ with $\tilde{f}_i(0) = \tilde{g}_i(0) = 0$, then $\exp[iQ(K)] = \exp[iQ_F(K)]$. It follows from Lemma 5.3 that $\mathbf{Q}_F \subset \mathbf{Q}$. Hence $\mathbf{B}(H_F^k) \otimes 1 \subset \mathbf{Q}$. Now consider the restriction $\mathbf{Q}_n$ of $\mathbf{Q}$ to H_n. Then $\mathbf{B}(H_F^k \otimes M_{n-k}) \subset \mathbf{Q}_n$, and in particular the projection on $\Psi_n = \Psi_{F0} \otimes \Phi_n$ is in $\mathbf{Q}_n$. Furthermore, if $\tilde{f}(0) \neq 0$, then $\exp[iQ(f \otimes f)]\ \Psi_n$ has a non-zero component in each $H_F^k \otimes M_{n-k}$, as one can show by direct calculation. [Note that $Q(f \otimes f)$ maps $H_F^k \otimes M_{n-k}$ onto $H_F^{k+i} \otimes M_{n-k-i}$, $j = 0, \pm 1$.] Thus Ψ_n is a cyclic vector for $\mathbf{Q}_n$. It follows that $\mathbf{Q}$ restricted to H_n is irreducible on H_n. QED.

For the operator a of Eq. (5.18) we have $aH_n = H_{n-1}$ and $a \in \mathbf{Q}'$. Thus $H = L \otimes N$ where $L \sim H_n$ and $\mathbf{Q} = \mathbf{B}(L) \otimes 1$. Thus as far as the algebra of observables is concerned, the representation of Theorem 3.1 is a discrete direct sum of copies of a single irreducible representation. Of course the field operators $U(f)$, $V(g)$ mix different H_n.

In constructing the infinite system, it is possible to obtain only L, and not $L \otimes N$. For this one considers only the operators $\exp[iQ(K)]$ instead of the $U(f)V(g)$, and uses the functional $E(K) = (\Psi, \exp[iQ(K)]\ \Psi)$. The space thus constructed is equivalent to the space L because L is generated from Ψ_0 by multiplication by operators in $\mathbf{Q}$. However the mathematical properties of the representation, such

as the irreducibility of $\mathbf{Q}$, are not so obvious from this approach.

However one constructs the infinite system, one obtains at most a direct sum of copies of a single irreducible representation of the algebra of observables. We therefore conclude that the ground state for the infinite free Bose gas is a pure state. This is not a trivial point. It is perhaps worth remarking that if one took $\mathbf{R} = \{U(f), V(g)\}''$ as the algebra of observables, then Eq. (3.27) would be the decomposition of the Hilbert space into coherent subspaces, and the physical state corresponding to Ψ_0 would be the impure state of all $\Psi_{F0} \otimes \Phi$, $\Phi \in M$.

Finite-Temperature Case

The mathematical structure of the representations is not so simple for the finite temperature case. We have not yet completed a systematic investigation of the algebra of observables. In the following we give a brief analysis using the operators $\psi(f)$.

We consider first the case where $T \geq T_c$ (the critical temperature for Bose–Einstein condensation). Then there is no macroscopic occupation of the zero-momentum state, and the representation of the CCR's is defined by Eqs. (4.11–15). Now Eq. (4.17) indicates that $\psi(f)$ can be interpreted as the sum of an annihilation operator for one kind of particle, and a creation operator for a second kind of particle (recall the discussion of the physical significance of the reducibility of the representation in Sec. 4). By analogy with Eqs. (5.19–27), this suggests that we introduce a relative number operator N_r defined by

$$N_r = N_F \otimes 1 - 1 \otimes N_F. \tag{5.36}$$

That is, if one assigns particle number $+1$ to particles of the first kind, and particle number -1 to particles of the second kind, then N_r is the particle number operator (in the sense of elementary particle theory).

The next step is to separate the Hilbert space $H = H_F \otimes H_F$ according to N_r. We define

$$H_n = \sum_{i=0}^{\infty}{}^{\oplus} H_F^{i+n} \otimes H_F^i \quad \text{if} \quad n \geq 0, \tag{5.37}$$

$$H_n = \sum_{i=0}^{\infty}{}^{\oplus} H_F^{i} \otimes H_F^{i-n} \quad \text{if} \quad n < 0. \tag{5.38}$$

Then we have

$$H = \sum_{n=-\infty}^{\infty}{}^{\oplus} H_n, \tag{5.39}$$

and N_r is just multiplication by n on H_n. As we have not calculated $E(K, f, g)$ we use Eq. (5.28) to define $Q(K)$.[23]

Let $\mathbf{Q}$ be the von Neumann algebra associated with the unbounded operators $Q(K)$, i.e. the von Neumann algebra generated by the spectral projections of the $Q(K)$. Clearly each H_n is invariant under $\mathbf{Q}$. One can show that the restriction $\mathbf{Q}_n$ of $\mathbf{Q}$ to H_n coincides with the restriction $\mathbf{R}_n$ of $\mathbf{R}$ to H_n, and that $\mathbf{Q}_n$ is cyclic on H_n. Furthermore $\mathbf{Q}_n$ is clearly reducible on H_n. But whether or not it is a factor is not clear to us. In any case, the reducibility means that the state Ψ_0 is a mixture.

We add a few remarks concerning the rotation and displacement operator $U(\mathbf{a}, R)$ and the time displacement operator $U(t)$. Since $\mathbf{Q} \subset \mathbf{R}$, it follows from Lemma 4.3 that these operators are not observables if we take $\mathbf{Q}$ as the algebra of observables. The operator algebra generated by $\mathbf{R}$ and $U(\mathbf{a}, 0)$ is irreducible because the projection onto the vector Ψ_0 belongs to the algebra generated by $U(\mathbf{a}, 0)$, and $\mathbf{R}$ is cyclic on Ψ_0. Similarly, the operator algebra generated by $\mathbf{Q}_0$ and $U(\mathbf{a}, R)$ restricted to H_0 is irreducible. Thus the conclusion that the thermal ensemble for the infinite free Bose gas is a mixture is only tentative.

For finite temperatures $T < T_c$ the representation of the CCR's is a direct integral of primary representations (i.e. representations in which the algebra R is a factor) labelled by the average field strengths. This case may be analyzed by a straightforward combination of the methods used for $T = 0$ and $T > T_c$. The result, subject to the ambiguities considered above, is that the thermal state is a mixture.

Expectation Values of Observables

The preceding discussion of the algebra of observables is somewhat incomplete. One can, at least formally, construct the observables in the representation defined by the functional $E(f, g)$ by means of Eq. (5.28). However, it is not clear that each observable thus constructed will have the desired expectation value,[23] although our discussion of the functional $E(K, f, g)$ indicates that this is true at least for the zero-temperature free Bose gas. We shall not attempt to settle this issue here. However, it should be pointed out that expectation values of products of the field operators $\phi(\mathbf{x})$ and $\pi(\mathbf{x})$ are at least formally given as functional derivatives of the expectation value of $U(f)V(g)$. This suggests that it is sufficient to consider the functional $E(f, g)$.

We will derive the particle density $\rho(\beta, \mu)$ and the energy density $h(\beta, \mu)$ directly from $E(f, g; \beta, \mu)$

for the free Bose gas. The number operator for the volume V is given by

$$N_V = \tfrac{1}{2} \sum_i [\phi(f_i)^2 + \pi(f_i)^2 - 1], \qquad (5.40)$$

where the f_i form a complete orthonormal set of real functions over the volume V. We have

$$\langle \phi(f_i)^2 + \pi(f_i)^2 - 1 \rangle$$
$$= -\left(\frac{d^2}{dx^2} + \frac{d^2}{dy^2}\right) E(xf_i, yf_i; \beta, \mu)|_{x=y=0} - 1$$
$$= \int d\mathbf{k} \rho(\mathbf{k}, \beta, \mu) |\tilde{f}_i(\mathbf{k})|^2, \qquad (5.41)$$

where $\rho(\mathbf{k}, \beta, \mu)$ is given by Eq. (5.2). As we noted earlier, $\sum_i |\tilde{f}_i(\mathbf{k})|^2 = V$, and thus

$$\rho(\beta, \mu) = \langle N_V/V \rangle = \int d\mathbf{k} \rho(\mathbf{k}, \beta, \mu). \qquad (5.42)$$

The Hamiltonian for the volume V is

$$H_V = \frac{1}{4m} \sum_i \sum_{i=1}^{3} [\phi(\nabla_i f_i)^2 + \pi(\nabla_i f_i)^2 - ||\nabla_i f_i||^2]. \qquad (5.43)$$

Now

$$\langle \phi(\nabla_i f_i)^2 + \pi(\nabla_i f_i)^2 - ||\nabla_i f_i||^2 \rangle$$
$$= -\left(\frac{d^2}{dx^2} + \frac{d^2}{dy^2}\right) E(x\nabla_i f_i, \times y\nabla_i f_i; \beta, \mu)|_{x=y=0} - ||\nabla_i f_i||^2$$
$$= 2 \int d\mathbf{k} \rho(\mathbf{k}, \beta, \mu) k_i^2 |\tilde{f}_i(\mathbf{k})|^2. \qquad (5.44)$$

Hence

$$h(\beta, \mu) = \langle H_V/V \rangle = \int d\mathbf{k} E(\mathbf{k}) \rho(\mathbf{k}, \beta, \mu), \qquad (5.45)$$

where $E(\mathbf{k}) = k^2/2m$. Other thermodynamical quantities can be obtained from $\rho(\beta, \mu)$ and $h(\beta, \mu)$.

6. THE INTERACTING BOSE GAS

This section contains some qualitative remarks on the interacting case. We give a plausible argument that, for suitable Hamiltonians, an infinite system exists. We also indicate a possible method for determining the functional $E(f, g)$.

We restrict ourselves to physical systems which can be described thermodynamically. That is, we assume the Hamiltonian is such that $\beta p = \lim_{V\to\infty} V^{-1} \log \mathcal{Q}_V$ (where $\mathcal{Q}_V$ is the grand partition function for volume V) exists, etc. In classical statistical mechanics van Hove,[24] and Yang and Lee[1] have proved the existence of this limit for certain Hamiltonians. A more rigorous and quite general proof has recently been given by Ruelle,[25] who also considers the quantum mechanical case. In this paper we shall only attempt to make it plausible that certain limits, in which we are interested, exist.

Equation (5.1) suggests that we consider the functional

$$E(f, g; \beta, \mu) = \lim_{V\to\infty} \langle U(f) V(g) \rangle_V, \qquad (6.1)$$

where

$$\langle X \rangle_V = \lim_{V\to\infty} \frac{\text{Trace } X \exp(\beta\mu N - \beta H)}{\text{Trace } \exp(\beta\mu N - \beta H)}$$

is the expectation value of the operator X over the grand canonical ensemble for a system confined to a volume V. For fixed f and g, $U(f)V(g)$ involves the field operators $\phi(\mathbf{x})$, $\pi(\mathbf{x})$ only for $\mathbf{x}$ in a fixed finite region (the support of the functions f and g). Now the physical behavior of a system which can be described thermodynamically should approach a uniform limit as $V \to \infty$. In particular, as we remarked above, the pressure approaches a limiting value. One would expect that the expectation value of any operator referring to some fixed finite region should approach a limit as $V \to \infty$. It therefore seems very plausible on physical grounds that $\lim_{V\to\infty} \langle U(f)V(g) \rangle_V$ exists for suitably restricted Hamiltonians.

It is easy to see that if $E(f, g; \beta, \mu)$ exists, it must satisfy the conditions of Lemma 2.1 and therefore defines a representation of the CCR's. Note that $\langle U(f)V(g) \rangle_V$ is a convex linear combination of terms of the form $(\Psi, U(f)V(g)\Psi)$. Since each term satisfies the conditions of Lemma 2.1, $\langle U(f)V(g) \rangle_V$ does also. It follows from continuity considerations that $E(f, g; \beta, \mu)$ satisfies the conditions of Lemma 2.1.

Equation (6.1) which defines $E(f, g; \beta, \mu)$ is not a practical method for calculating the functional. One possible method is to obtain a differential equation in the parameter β. At infinite temperatures one can neglect the interaction energy, and we have the initial condition

$$E(f, g; \beta, \mu)|_{\beta=0} = E_F(f, g) \exp\left\{-\tfrac{1}{2} \int d\mathbf{x} e^{\alpha} (1 - e^{\alpha})^{-1} [f(\mathbf{x})^2 + g(\mathbf{x})^2]\right\}. \qquad (6.3)$$

From Eq. (6.1) we obtain ($\alpha = \beta\mu$)

$$-(\partial/\partial\beta) E(f, g; \beta, \alpha/\beta) = \lim_{V\to\infty} \langle \{H - \langle H \rangle_V\} U(f) V(g) \rangle_V. \qquad (6.4)$$

[24] L. van Hove, Physica 15, 95 (1949).

[25] D. Ruelle (to be published).

The right-hand side of this equation can be evaluated in terms of $E(f, g; \beta, \alpha/\beta)$ as follows.

Let Φ_i be a complete orthonormal set of real functions for the volume V. The Hamiltonian for the volume V is $H_V = H_{0V} + H_{IV}$. The kinetic energy term H_{0V} is given by Eq. (5.43), and the interaction energy is

$$H_{IV} = \tfrac{1}{2} \sum_{i,j,k,l} V_{ij,kl} \psi^*(\Phi_i)\psi^*(\Phi_j)\psi(\Phi_k)\psi(\Phi_l), \quad (6.5)$$

where $V_{ij,kl} = V_{ji,lk}$. Consider the functions

$$\bar{f} = f + \sum_i \sum_{k=1}^{3} x_{ik} \nabla_k \Phi_i, \quad (6.6)$$

$$\bar{g} = g + \sum_i \sum_{k=1}^{3} y_{ik} \nabla_k \Phi_i, \quad (6.7)$$

where x_{ik}, y_{ik} are real. We have

$$\psi^*(\nabla_k\Phi_i)U(f)V(g) = \frac{1}{\sqrt{2}}\left[-i\frac{\partial}{\partial x_{ik}} - \frac{\partial}{\partial y_{ik}} - i(\bar{f}, \nabla_k\Phi_i)\right]U(f)V(g), \quad (6.8)$$

$$\psi(\nabla_k\Phi_i)U(f)V(g) = \frac{1}{\sqrt{2}}\left[-i\frac{\partial}{\partial x_{ik}} + \frac{\partial}{\partial y_{ik}} + i(\bar{f}, \nabla_k\Phi_i)\right]U(f)V(g), \quad (6.9)$$

evaluated at $x_{ik} = y_{ik} = 0$. Hence

$$\langle H_{0V}U(f)V(g)\rangle = \frac{1}{4m}\sum_i \sum_{k=1}^{3}\left[-i\frac{\partial}{\partial x_{ik}} + \frac{\partial}{\partial y_{ik}} + i(\bar{f}, \nabla_k\Phi_i)\right] \times \left[-i\frac{\partial}{\partial x_{ik}} - \frac{\partial}{\partial y_{ik}} - i(\bar{f}, \nabla_k\Phi_i)\right] \times E(\bar{f}, \bar{g}; \beta, \mu)|_{x_{ik}=y_{ik}=0}. \quad (6.10)$$

Similarly for H_{IV}, consider the functions

$$\bar{f} = f + \sum_i x_i\Phi_i, \quad (6.11)$$

$$\bar{g} = g + \sum_i y_i\Phi_i. \quad (6.12)$$

Then one obtains

$$\langle H_{IV}U(f)V(g)\rangle = \tfrac{1}{4}\sum_{ijkl} V_{ij,kl} \times \left[-i\frac{\partial}{\partial x_l} + \frac{\partial}{\partial y_l} + i(\bar{f}, \Phi_l)\right] \times \left[-i\frac{\partial}{\partial x_k} + \frac{\partial}{\partial y_k} + i(\bar{f}, \Phi_k)\right] \times \left[-i\frac{\partial}{\partial x_i} - \frac{\partial}{\partial y_i} - i(\bar{f}, \Phi_i)\right] \times \left[-i\frac{\partial}{\partial x_i} - \frac{\partial}{\partial y_i} - i(\bar{f}, \Phi_i)\right] \times E(\bar{f}, \bar{g}; \beta, \mu)|_{x_i=y_i=0}. \quad (6.13)$$

Formally we have

$$\langle HU(f)V(g)\rangle = \left\{-(4m)^{-1}\int d\mathbf{x} \times [\nabla\delta_{f(\mathbf{x})} + i\nabla\delta_{g(\mathbf{x})} - \nabla f(\mathbf{x})] \cdot [\nabla\delta_{f(\mathbf{x})} - i\nabla\delta_{g(\mathbf{x})} + \nabla f(\mathbf{x})] + \int d\mathbf{x}\, d\mathbf{y} V(\mathbf{x}-\mathbf{y})F(\mathbf{x}, \mathbf{y})\right\}E(f, g; \beta, \mu), \quad (6.14)$$

where

$$F(\mathbf{x}, \mathbf{y}) = \tfrac{1}{8}[\delta_{f(\mathbf{x})} + i\delta_{g(\mathbf{x})} - f(\mathbf{x})] \times [\delta_{f(\mathbf{y})} + i\delta_{g(\mathbf{y})} - f(\mathbf{y})] \times [\delta_{f(\mathbf{x})} - i\delta_{g(\mathbf{x})} + f(\mathbf{x})] \times [\delta_{f(\mathbf{y})} - i\delta_{g(\mathbf{y})} + f(\mathbf{y})], \quad (6.15)$$

and

$$\delta_{f(\mathbf{x})} \equiv \delta/\delta f(\mathbf{x}).$$

$\langle H_V\rangle$ is obtained by setting $f = g = 0$. Thus Eq. (6.4) is a linear differential equation for $E(f, g; \beta, \mu)$. It is therefore possible that Eqs. (6.3) and (6.4) determine $E(f, g; \beta, \mu)$.

For the free Bose gas $E(f, g; \beta, \mu)$ is given by Eq. (5.1), and one can verify that Eq. (6.4) is satisfied. However the Bose–Einstein condensation presents special problems for a free Bose gas since $\mu = 0$ for $T < T_c$. Thus we are unable to obtain by this method the functional describing the free Bose gas when condensation is present. This difficulty may not occur for an interacting Bose gas, since then $\mu \neq 0$ for all temperatures and densities. On the other hand, it seems probable that in the vicinity of a phase transition $E(f, g; \beta, \mu)$ is two different analytic functionals of β. Thus the differential Eq. (6.4) may not work at a phase transition. To sum up, it is an open question whether or not Eqs. (6.3) and (6.4) determine $E(f, g; \beta, \mu)$.

7. DISCUSSION OF RESULTS

Our approach to the many-body problem may be stated briefly as follows. By considering strictly infinite physical systems, we hope to attain some new insight into the behavior of many-body systems. In particular, we hope to obtain a new way of looking at the theory of condensation. We also hope that

such an investigation might lead to a new method for computing the thermodynamic functions, or to new soluble models of interest.

This paper contains the preliminary results obtained in this general program. It should be emphasized that these results form an extremely small step, and do not shed any light on the ultimate goal. We have chosen to begin the investigation by seeing whether or not any strictly infinite systems exist in the sense of formal quantum theory. The main content of this paper is that such systems do exist, this result being established by constructing representations of the CCR's describing a nonrelativistic infinite free Bose gas. From a strictly mathematical point of view, these representations are of some interest in themselves. From the point of view of physics, the situation is not clear. The existence or nonexistence of strictly infinite systems may be a rather pathological problem, containing nothing relevant to any practical approach to the many body problem.

We shall review briefly our results for the infinite free Bose gas. For the ground state or zero-temperature case, the representation of the field operators is a direct integral of inequivalent irreducible representations (labeled by the average field strengths). The reducibility of the representation is completely analogous to the degeneracy of the ground state in BCS theory. For finite temperatures $T > T_c$ (the critical temperature for Bose-Einstein condensation) the representation is a primary representation. That is, the algebra R of all functions of the field operators is a factor. We proved that R was not a factor of type I (i.e. R is not isomorphic to the algebra of all bounded operators on a Hilbert space). For finite temperatures $T < T_c$, the representation is a direct integral of inequivalent primary representations (labeled by the average field strengths).

We gave a systematic, although not unambiguous, method for constructing the algebra of observables for the infinite system. For the ground-state case, $T = 0$, the algebra of observables obtained in this way from the representation of the CCR's, is a discrete direct sum of copies of one irreducible representation. Thus the thermal ensemble is a pure state for $T = 0$.

For the finite temperature case, the algebra of observables seems to decompose into a discrete direct sum of cyclic representations which are reducible. Thus the thermal ensemble for the infinite free Bose gas seems to be a mixture for $T > 0$. This conclusion is only tentative since there is some ambiguity as to which operators are observables for the infinite system. In particular, our approach leads to the result that the Euclidean transformation operator $U(\mathbf{a}, R)$ and the time displacement operator $U(t)$ do not belong to the algebra of observables for $T > 0$. One would like to understand why the infinitesimal generators of these operators are not observables (or, perhaps, why our method of constructing the algebra of observables is incorrect).

It is interesting to note that the mathematical structure of the representation is quite different when condensation is present. But except for the zero-temperature case (and to some extent for $T < T_c$), we have not understood the physical significance of the mathematical structure of the representations. It seems essential to further progress that this point be cleared up. One can hardly expect to obtain any insight into the many-body problem without completely understanding the free Bose gas. Of course, even if a satisfactory theory of the infinite free Bose gas is achieved, it does not follow that a satisfactory theory for the interacting case exists.

In short, the present situation is that much work remains to be done before the usefulness of this approach can be evaluated.

ACKNOWLEDGMENTS

The first-named author (H. A.) would like to thank Professors R. Jost and M. Fierz for their hospitality in Zürich, and acknowledge the financial support received from the Schweizerischer Nationalfonds (K. A. W.). The second-named author (E. J. W.) would like to express his gratitude to Professor A. S. Wightman for suggesting this problem, and for his invaluable guidance and unfailing encouragement during the course of this work.

APPENDIX 1. CALCULATION OF $E(f, g)$ FOR GRAND CANONICAL ENSEMBLE

In this appendix we compute the functional

$$E(f, g) = \frac{\text{Trace } U(f)V(g) \exp(\beta\mu N - \beta H)}{\text{Trace} \exp(\beta\mu N - \beta H)} \qquad \text{(A1)}$$

for the grand canonical ensemble, which is one of the most important states for considering the equilibrium properties of a system. Of course the canonical ensemble is also interesting, but it is more difficult to compute.

We consider a system in a finite box of volume V. The single-particle energy levels will be denoted by the index j, the corresponding normalized wavefunctions by Φ_j, and the energy by E_j. Then $\Phi_{-j}(\mathbf{x}) = \Phi_j(\mathbf{x})^*$ represents the state with opposite momenta, and Φ_0 is the state with a constant wave-

function. The annihilation and creation operators $\psi(\Phi_i)$ and $\psi(\Phi_i)^*$ will be denoted by a_i and a_i^*. For real $f(\mathbf{x})$, $g(\mathbf{x})$ we have

$$\phi(f) = (2V)^{-\frac{1}{2}} \sum_i (f_i a_i + f_i^* a_i^*), \tag{A2}$$

$$\pi(g) = i(2V)^{-\frac{1}{2}} \sum_i (g_i^* a_i^* - g_i a_i), \tag{A3}$$

where

$$f_i = V^{-\frac{1}{2}}(f, \Phi_i), \tag{A4}$$

$$g_i = V^{-\frac{1}{2}}(g, \Phi_i). \tag{A5}$$

We have

$$U(f) = \prod_i \exp [i(2V)^{-\frac{1}{2}} f_i^* a_i^*] \times \exp [i(2V)^{-\frac{1}{2}} f_i a_i] \exp (-|f_i|^2/4V), \tag{A6}$$

$$V(g) = \prod_i \exp [-(2V)^{-\frac{1}{2}} g_i^* a_i^*] \times \exp [(2V)^{-\frac{1}{2}} g_i a_i] \exp (-|g_i|^2/4V). \tag{A7}$$

Hence

$$U(f)V(g) = \prod_i \exp (-\delta_i^* a_i^*) \times \exp (\delta_i a_i) \exp (-\gamma_i), \tag{A8}$$

where

$$\delta_i = (2V)^{-\frac{1}{2}}(if_i + g_i), \tag{A9}$$

$$\gamma_i = (4V)^{-1}(|f_i|^2 + |g_i|^2 + 2ig_i^* f_i). \tag{A10}$$

[Compare Eq. (3.6)]. We write

$$\beta\mu N - \beta H = \sum \alpha_i N_i, \tag{A11}$$

where $N_i = a_i^* a_i$ and $\alpha_i = \beta(\mu - E_i)$. Thus we have

$$\Xi_V(f, g; \beta, \mu) \equiv \text{Trace } \exp (\beta\mu N - \beta H)U(f)V(g) = \prod_i \text{Trace } [\exp (\alpha_i N_i) \exp (-\delta_i^* a_i^*) \times \exp (\delta_i a_i)] \exp (-\gamma_i), \tag{A12}$$

where the trace is now taken over states of different occupation numbers for the jth energy level only. We now drop the index j, and use the formulas

$$||(a^*)^n \Omega||^2 = n!, \tag{A13}$$

$$e^{\delta a}\Omega = \Omega, \tag{A14}$$

$$e^{\alpha N}(a^*)^n \Omega = e^{\alpha n}(a^*)^n \Omega, \tag{A15}$$

$$e^{\delta a} a^* e^{-\delta a} = a^* + \delta. \tag{A16}$$

where Ω is the no-particle state for one degree of freedom. We have

$$\begin{aligned}
&\text{Trace } (e^{\alpha N} e^{-\delta^* a^*} e^{\delta a}) \\
&= \sum_{n=0}^{\infty} \frac{e^{\alpha n}}{n!} ([a^*]^n \Omega, e^{-\delta^* a^*} e^{\delta a} [a^*]^n \Omega) \\
&= \sum (n!)^{-1} e^{\alpha n} (\Omega, (a - \delta^*)^n (a^* + \delta)^n \Omega) \\
&= \sum_{n=0}^{\infty} (e^{\alpha})^n \sum_{r=0}^{n} \frac{(-|\delta|^2)^r n!}{(r!)^2 (n-r)!} \\
&= \sum_{r=0}^{\infty} \frac{(-|\delta|^2)^r}{(r!)^2} \sum_{n=r}^{\infty} (e^{\alpha})^n n(n-1) \cdots (n-r+1) \\
&= \sum_{r=0}^{\infty} \frac{(-e^{\alpha}|\delta|^2)^r}{(r!)^2} \left(\frac{d}{dx}\right)^r \sum_{m=0}^{\infty} x^m |_{x=e^{\alpha}} \\
&= (1 - e^{\alpha})^{-1} \exp -[e^{\alpha}|\delta|^2 (1 - e^{\alpha})^{-1}].
\end{aligned} \tag{A17}$$

Thus

$$E_V(f, g; \beta, \mu) = \Xi_V(f, g; \beta, \mu)/\Xi_V(0, 0; \beta, \mu) = \prod_i \exp -[e^{\alpha_i}|\delta_i|^2 (1 - e^{\alpha_i})^{-1} + \gamma_i]. \tag{A18}$$

In the limit $V \to \infty$, $V^{-1} \sum_i$ becomes $(2\pi)^{-3} \int d\mathbf{k}$, f_i becomes $\tilde{f}(\mathbf{k}) = \int d\mathbf{k}\, e^{i\mathbf{k}\cdot\mathbf{x}} f(\mathbf{x})$, and E_i becomes $E_k = k^2/2m$. Thus we obtain

$$E(f, g; \beta, \mu) = \lim_{V\to\infty} E_V(f, g; \beta, \mu) = E_F(f, g) \exp \left\{ -\frac{1}{2} \int \frac{d\mathbf{k}}{(2\pi)^3} \frac{\exp (\beta\mu - \beta E_k)}{1 - \exp (\beta\mu - \beta E_k)} \times [|\tilde{f}(k)|^2 + |\tilde{g}(k)|^2] \right\}. \tag{A19}$$

Note added in proof. The case of an infinite Bose gas in an external periodic potential is completely analogous to the free Bose gas, provided one makes the following changes:

$$k^2/2m \to E(\mathbf{k}),$$

$$\tilde{f}(\mathbf{k}) \to f_B(\mathbf{k}) = \int d\mathbf{x} e^{i\mathbf{k}\cdot\mathbf{x}} u_{\mathbf{k}}(\mathbf{x}) f(\mathbf{x}),$$

where the $u_{\mathbf{k}}(\mathbf{x})$ are the Bloch wavefunctions. Thus the mathematical structure of the representations is identical to that for the free Bose gas.

Representation of Annihilation Operators in the Space of Periodic Functions

G. H. DERRICK
Applied Mathematics Department, University of New South Wales, Kensington, New South Wales
(Received 22 October 1962)

A representation is given of fermion annihilation operators a_k which satisfy the usual anticommutation relations $\{a_k, a_{k'}^+\} = \delta_{kk'}$. In this representation the a_k operate on periodic function $f(\zeta)$ of a real variable ζ. The Hilbert space is the set of all square integrable functions $f(\zeta)$ of period 2π, with the scalar product defined by

$$(f, g) = \int_{-\pi}^{\pi} f^* g \, d\zeta.$$

INTRODUCTION

LET us denote by $\psi(\mathbf{r}, s)$ the operator which annihilates a fermion of spin s at coordinate position $\mathbf{r}$.[1] We shall suppose $\mathbf{r}$ ranges continuously over some given domain, finite or infinite, while the spin variable s assumes a discrete set of values, for example, $\pm\frac{1}{2}$. We take the Schrödinger picture, so that $\psi(\mathbf{r}, s)$ does not depend on the time. Then $\psi(\mathbf{r}, s)$ and its Hermitian conjugate $\psi^\dagger(\mathbf{r}, s)$ satisfy the anticommutation relations

$$\begin{aligned}
&\psi(\mathbf{r}, s)\psi(\mathbf{r}', s') + \psi(\mathbf{r}', s')\psi(\mathbf{r}, s) = 0,\\
&\psi(\mathbf{r}, s)\psi^\dagger(\mathbf{r}', s') + \psi^\dagger(\mathbf{r}', s')\psi(\mathbf{r}, s) = \delta(\mathbf{r} - \mathbf{r}')\ \delta_{ss'},\\
&\psi^\dagger(\mathbf{r}, s)\psi^\dagger(\mathbf{r}', s') + \psi^\dagger(\mathbf{r}', s')\psi^\dagger(\mathbf{r}, s) = 0.
\end{aligned}$$

To avoid the difficulties arising from the continuum of values of $\mathbf{r}$, it is usual to introduce a discrete, complete, orthonormal set of space-spin functions $u_k(\mathbf{r}, s)$, and to expand $\psi(\mathbf{r}, s)$ in terms of this set. The functions $u_k(\mathbf{r}, s)$ must satisfy the completeness and orthonormality relations

$$\sum_k u_k(\mathbf{r}, s)u_k^*(\mathbf{r}', s') = \delta(\mathbf{r} - \mathbf{r}')\ \delta_{ss'},$$

$$\sum_s \int d\mathbf{r} u_k(\mathbf{r}, s)u_{k'}^*(\mathbf{r}, s) = \delta_{kk'},$$

but are otherwise arbitrary. Without loss of generality we may suppose that the set $u_k(\mathbf{r}, s)$ be ordered as a simple sequence, with the subscript k taking the values 0, 1, 2, 3, $\cdots$. The annihilation operator a_k for a fermion in "state" k may be defined by

$$a_k = \sum_s \int d\mathbf{r} u_k^*(\mathbf{r}, s)\psi(\mathbf{r}, s),$$

with the inverse relation

$$\psi(\mathbf{r}, s) = \sum_k u_k(\mathbf{r}, s)a_k.$$

Then the set of annihilation operators a_k will satisfy the anticommutation relations

$$\begin{aligned}
&a_k a_{k'} + a_{k'}a_k = 0,\\
&a_k a_{k'}^+ + a_{k'}^+ a_k = \delta_{kk'},\\
&a_k^+ a_{k'}^+ + a_{k'}^+ a_k^+ = 0,\\
&k, k' = 0, 1, 2, \cdots.
\end{aligned} \tag{1}$$

The object of this paper is to give a representation in which the a_k are operators on periodic functions $f(\zeta)$ of a single real variable ζ.

In the usual occupation-number representation, we take the number operators $N_k = a_k^+ a_k$ simultaneously diagonal. The simultaneous eigenstates of these number operators are denoted by a sequence of occupation numbers $|n_0, n_1, n_2, \cdots\rangle$, where each n_k can be 0 or 1. Denoting the vacuum $|0\rangle$, for which $a_k|0\rangle = 0$, we shall order these states so:

$$\begin{aligned}
&|0, 0, 0, 0, \cdots\rangle = |0\rangle,\\
&|1, 0, 0, 0, \cdots\rangle = a_0^+|0\rangle,\\
&|1, 1, 0, 0, \cdots\rangle = a_0^+ a_1^+|0\rangle,\\
&|0, 1, 0, 0, \cdots\rangle = a_1^+|0\rangle,\\
&|0, 1, 1, 0, \cdots\rangle = a_1^+ a_2^+|0\rangle,\\
&|1, 1, 1, 0, \cdots\rangle = a_0^+ a_1^+ a_2^+|0\rangle.
\end{aligned} \tag{2}$$

The criterion used in ordering the states is: To generate the next state, change one occupation number from 0 to 1, or from 1 to 0; the occupation number to be changed is the left-most which can be changed without reproducing a previous state.

With this ordering, the matrices representing N_k and a_k in the occupation number representation are

[1] An excellent introduction to the concept of creation and annihilation operators is given in L. D. Landau and E. M. Lifshitz, *Quantum Mechanics, Nonrelativistic Theory*, English translation (Pergamon Press, New York, 1958), Secs. 62, 63.

$$N_k = \begin{pmatrix} 0_k & & & & & & & \\ & I_k & & & & & & \\ & & I_k & & & & & \\ & & & 0_k & & & & \\ & & & & 0_k & & & \\ & & & & & I_k & & \\ & & & & & & I_k & \\ & & & & & & & \ddots \end{pmatrix},$$

$$a_k = \begin{pmatrix} 0_k & \epsilon_k & & & & & & \\ 0_k & 0_k & & & & & & \\ & & 0_k & 0_k & & & & \\ & & \epsilon_k^\dagger & 0_k & & & & \\ & & & & 0_k & \epsilon_k & & \\ & & & & 0_k & 0_k & & \\ & & & & & & \ddots & \\ & & & & & & & \ddots \end{pmatrix},$$

where 0_k and I_k are respectively the null matrix and the unit matrix with 2^k rows and columns, and

$$\epsilon_0 = 1, \qquad \epsilon_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad \epsilon_{k+1} = \begin{pmatrix} \epsilon_k \\ \epsilon_k \end{pmatrix},$$

$$k = 1, 2, 3, \cdots .$$

2. THE NEW REPRESENTATION

Given two Hilbert spaces H and H', we can induce an isomorphism between the operators on H and the operators on H' as follows[2]: Let $|u_i\rangle$ be a complete denumerable orthonormal set of vectors in H, and $|u_i'\rangle$ such a set in H', where the subscript i takes the same denumerable set of values 0, 1, 2, $\cdots$ in each case. We first define a one–one linear correspondence between the vectors $|\psi\rangle$ of H and the vectors $|\psi'\rangle$ of H' by

$$|\psi'\rangle = \sum_i |u_i'\rangle\,(u_i, \psi). \tag{3}$$

Let F be a linear operator defined on a domain D in H, and let D' be the set of all images $|\varphi'\rangle$ in H' under the transformation (3) of the vectors $|\varphi\rangle$ belonging to D.

Define an operator F' in the domain $D' \subset H'$ by

$$F'\,|\varphi'\rangle = \sum_i |u_i'\rangle\,(u_i, F\varphi). \tag{4}$$

We thus obtain a one–one correspondence between the operators F on H, and the operators F' on H', which is an isomorphism with respect to addition and multiplication; i.e., given any two operators F and G on H, and arbitrary complex numbers λ and μ,

$$\left.\begin{matrix} F \to F' \\ G \to G' \end{matrix}\right\} \text{ imply } \begin{matrix} \lambda F + \mu G \to \lambda F' + \mu G', \\ FG \to F'G', \end{matrix}$$

whenever the domains of F and G enable the above addition and multiplication operations to be defined.

In other words, the mapping $F \to F'$ specified by (3) and (4) gives us a faithful representation in H' of any system of operators in H.

Let us apply (3) and (4) as follows: H is the Hilbert space spanned by the state vectors (2). H' is the Hilbert space of all square integrable complex functions $f(\zeta)$ of a real variable ζ which are periodic with period 2π; the scalar product between $f(\zeta)$ and $g(\zeta)$ is defined as

$$(f, g) = \int_{-\pi}^{\pi} f^* g\, d\zeta.$$

As the complete orthonormal set $|u_i\rangle$ in H we take the states (2) in the order given. For the set $|u_i'\rangle$ in H' we take

$$\frac{1}{(2\pi)^{\frac{1}{2}}}, \frac{e^{i\zeta}}{(2\pi)^{\frac{1}{2}}}, \frac{e^{-i\zeta}}{(2\pi)^{\frac{1}{2}}}, \frac{e^{2i\zeta}}{(2\pi)^{\frac{1}{2}}}, \frac{e^{-2i\zeta}}{(2\pi)^{\frac{1}{2}}}, \cdots$$

in this order.

Then the representation in H' of $N_0 = a_0^\dagger a_0$ given by (3) and (4) is

$$\begin{aligned} N_0' f(\zeta) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} [e^{i(\zeta-\zeta')} + e^{-i(\zeta-\zeta')} + e^{3i(\zeta-\zeta')} \\ &\quad + e^{-3i(\zeta-\zeta')} + \cdots] f(\zeta')\, d\zeta' \\ &= \frac{1}{2}\int_{-\pi}^{\pi} e^{i(\zeta-\zeta')} \sum_s \delta(\zeta - \zeta' - s\pi) f(\zeta')\, d\zeta' \\ &= \tfrac{1}{2}[f(\zeta) - f(\zeta - \pi)]. \end{aligned} \tag{5}$$

The representation of a_0 is

$$a_0' f(\zeta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{m,n} e^{i(m\zeta - n\zeta')} f(\zeta')\, d\zeta',$$

where in the summation (m, n) take the values $(0, 1), (2, -1), (-2, 3), (4, -3), (-4, 5), \cdots$. We have $\sum_{m,n} e^{i(m\zeta - n\zeta')} = \pi e^{-i\zeta'} \sum_s \delta(\zeta + \zeta' - s\pi)$. Hence

$$a_0' f(\zeta) = \tfrac{1}{2} e^{i\zeta}[f(-\zeta) - f(\pi - \zeta)]. \tag{6}$$

In a similar manner we can obtain representations for all the other operators N_k and a_k. The results for $k = 1, 2, 3, \cdots$ are

[2] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Springer-Verlag, Berlin/Vienna, 1932), Chap. II.

$$N'_k f(\zeta) = \tfrac{1}{2} f(\zeta) - \sum_{n=1}^{2^{(k+1)}-1} \left[\frac{2^{-(k+1)} e^{i\zeta'/2} \sin n\pi/2 f(\zeta - \zeta')}{\sin \zeta'/2} \right]_{\zeta' = 2^{-k} n\pi},$$

$$a'_k f(\zeta) = \sum_{n=0}^{2^{(k+1)}-1} \left[\frac{i 2^{-k} e^{i(\zeta - \zeta'/2)} \sin 2^{k-1}(\zeta - \zeta'/2) \sin 2^{k-1}(\zeta - 3\zeta'/2) f(\zeta' - \zeta)}{\sin (\zeta - \zeta'/2)} \right]_{\zeta' = 2^{-k} n\pi}. \tag{7}$$

It should be noted that the results (7) do not hold for $k = 0$.

The results (5), (6) and (7) can be expressed more simply as contour integrals in the complex ζ plane:

$$N'_k f(\zeta) = \frac{1}{2\pi i} \oint_C \varphi_k(\zeta' - \zeta) f(\zeta')\, d\zeta', \qquad a'_k f(\zeta) = \frac{1}{2\pi i} \oint_C \psi_k(\zeta, \zeta') f(\zeta')\, d\zeta', \tag{8}$$

where

$$\begin{aligned} \varphi_k(z) &= 1/2 \sin z && k = 0, \\ &= (e^{iz/2}/4 \sin z/2) \sec (2^{k-1} z) && k = 1, 2, 3, \cdots, \\ \psi_k(\zeta, \zeta') &= e^{i\zeta}/2 \sin (\zeta + \zeta') && k = 0, \\ &= \frac{e^{i(\zeta - \zeta')/2} \sin 2^{(k-2)}(\zeta - \zeta') \sin 2^{(k-2)}(\zeta + 3\zeta') \cot 2^k(\zeta + \zeta')}{i \sin (\zeta - \zeta')/2} && k = 1, 2, 3, \cdots. \end{aligned}$$

The contour C may be any which encloses a length 2π of the real ζ' axis, but excludes all poles of $f(\zeta')$. In general, the elements $f(\zeta)$ of H' are defined only for real ζ, and will not be differentiable. However the subset of functions $f(\zeta)$ which are analytic near the real ζ axis, is everywhere dense in H', so that (8) suffices to define the operators N'_k and a'_k.

The representation of the total number operator $N = \sum_k N_k$ is

$$N' f(\zeta) = \sum_{k=0}^{\infty} \frac{1}{2\pi i} \oint_{C_k} \varphi_k(\zeta' - \zeta) f(\zeta')\, d\zeta'. \tag{9}$$

Let us take for C_k a contour of the type shown in Fig. 1.

We assume that a positive number η exists such that $f(\zeta')$ has no poles within a distance η of the real ζ' axis. The two dotted portions of the contour may be ignored; they make equal and opposite contributions since the integrand is periodic of period 2π.

Hence,

$$\begin{aligned} N' f(\zeta) = \sum_{k=0}^{\infty} \frac{1}{2\pi i} \int_{-\pi}^{\pi} & [\varphi_k(\xi - i\eta - \zeta) f(\xi - i\eta) \\ & - \varphi_k(\xi + i\eta - \zeta) f(\xi + i\eta)]\, d\xi \\ = \frac{1}{2\pi i} \int_{-\pi}^{\pi} & [\varphi(\xi - i\eta - \zeta) f(\xi - i\eta) \\ & - \varphi(\xi + i\eta - \zeta) f(\xi + i\eta)]\, d\xi, \end{aligned}$$

where

$$\varphi(\zeta) = \sum_{k=0}^{\infty} \varphi_k(\zeta).$$

The function $\varphi(\zeta)$ is analytic everywhere except on the real ζ axis and at infinity.

3. DISCUSSION

Since the transformations (3) and (4) induce an isomorphism, the representation of a_k given here is completely equivalent to the usual occupation-number representation. As long as we are dealing with *exact* equations the two representations must yield the same results. In *approximate* work, however, this need not be so. Expressions which are simple in one representation are often quite complicated in the other. Hence what suggests itself as a "natural" approximation scheme in the present representation might well be very involved in the usual representation, and vice versa. The same is true of the introduction of new assumptions into an existing field theory.

A difficulty in application of the representation given here to any actual computation lies in the requirement that the annihilation operators a_k be

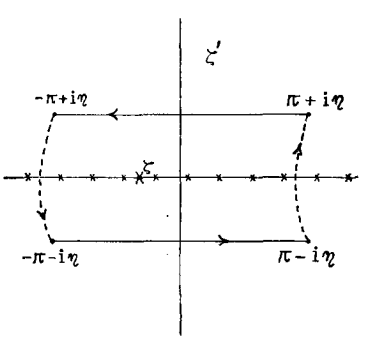


FIG. 1. The contour C_k in the complex ζ' plane used in evaluating (9). The crosses on the real axis represent the poles of the function $\varphi_k(\zeta' - \zeta)$. [Shown for the case $k = 3$ and ζ real.]

ordered in a simple sequence, $k = 0, 1, 2, \cdots$. If, for example, we use momentum eigenfunctions, the annihilation operators would more naturally be labeled by a vector index $\mathbf{k}$, rather than by a single subscript. Be sure, any 3-dimensional array (k_1, k_2, k_3) can be ordered as a simple sequence, but such ordering is highly anisotropic with respect to the three components of $\mathbf{k}$. A further difficulty which rules out numerical computation is the rapidly varying nature of the wavefunctions $f(\zeta)$ corresponding to physical states. For example, the Fermi-sea state $(\prod_{k<k_0} a_k^+)\ |0\rangle$ corresponds to the wavefunction

$$f(\zeta) = e^{-im\zeta}/(2\pi)^{\frac{1}{2}} \quad \text{with} \quad m = \tfrac{1}{2}(2^{k_0} - 1)$$

(taking k_0 even). With k_0 the order of 10^{23}, $f(\zeta)$ is indeed a rapidly varying function.

It should be noted that the particular ordering (2), or the particular choice of base functions $e^{im\zeta}/(2\pi)^{\frac{1}{2}}$ is not essential for the application of the isomorphism (4). For example, we could order the states with occupation numbers n_k according to the natural order of the integers $n_0 + 2n_1 + 4n_2 + \cdots$, or might substitute the base functions $(\sin m\zeta/2)/\pi^{\frac{1}{2}}$ for the exponential form. The representations so obtained are similar in appearance to, but considerably more complicated than, the one given in this paper.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 5 MAY 1963

Note on the One-Dimensional Gas of Impenetrable Point-Particle Bosons

T. D. SCHULTZ
International Business Machines Corporation, Thomas J. Watson Research Center, Yorktown Heights, New York
(Received 10 December 1962)

A one-dimensional gas of impenetrable point-particle bosons is considered. An exact expression for the one-particle density matrix is derived in terms of the Fredholm determinant and resolvent of a certain simple kernel. It is proved that for large r, the density matrix is bounded by const $\times r^{-4/\pi^2}$, and that this implies the absence of a generalized Bose–Einstein condensation, contrary to a recent approximate calculation by Girardeau, who first defined such a condensation.

I. INTRODUCTION

In a recent paper,[1] Girardeau introduced the notion of a generalized Bose–Einstein condensation for a system of $\mathfrak{N}$ particles. In such a condensed state the average number n_0 of particles with momentum $k = 0$ is not of order $\mathfrak{N}$; in the limit of an infinite system at fixed density ρ, however, the average number of particles having arbitrarily small momentum is of order $\mathfrak{N}$. Stated precisely, the criterion for a generalized condensation is

$$\lim_{k_0\to 0} \lim_{\substack{\mathfrak{N}\to\infty \\ \rho \text{ fixed}}} \mathfrak{N}^{-1} \sum_{|k|<k_0} n_k \equiv \lim_{k_0\to 0} P_{k_0} = O(1).$$

It was suggested that such a generalized condensation occurs in the ground state of a one-dimensional gas of $\mathfrak{N}$ impenetrable point-particle bosons satisfying periodic boundary conditions over a fundamental domain of length L. For such a system, Girardeau showed the intimate relation to a system of $\mathfrak{N}$ spinless noninteracting fermions: the energy spectra and, hence, the free energies of the two systems are identical, and the boson wavefunctions are simply related to the corresponding fermion wavefunctions (Slater determinants). Unfortunately, despite the knowledge of the exact ground state wavefunction, it was not possible to calculate n_k exactly, so that a certain linearization approximation in the wavefunction had to be made. Within this approximation, Girardeau found a *complete* generalized Bose–Einstein condensation, i.e. $\lim_{k_0\to 0} P_{k_0} = 1$.

It seemed to us that, because of the simplicity of the system, it should be possible to prove or disprove the existence of such a condensation without making any approximations. In the present note, we disprove it. In Sec. II we formulate the problem to stress the similarity with a certain soluble one-dimensional spin-$\frac{1}{2}$ model. Drawing on results for this spin model[2] we are able to derive an exact

[1] M. Girardeau, J. Math. Phys. 1, 516 (1960).

[2] E. Lieb, T. Schultz, and D. Mattis, Ann. Phys. 16, 407 (1961). Hereinafter, referred to as A.

ordered in a simple sequence, $k = 0, 1, 2, \cdots$. If, for example, we use momentum eigenfunctions, the annihilation operators would more naturally be labeled by a vector index $\mathbf{k}$, rather than by a single subscript. Be sure, any 3-dimensional array (k_1, k_2, k_3) can be ordered as a simple sequence, but such ordering is highly anisotropic with respect to the three components of $\mathbf{k}$. A further difficulty which rules out numerical computation is the rapidly varying nature of the wavefunctions $f(\zeta)$ corresponding to physical states. For example, the Fermi-sea state $(\prod_{k<k_0} a_k^+)\ |0\rangle$ corresponds to the wavefunction

$$f(\zeta) = e^{-im\zeta}/(2\pi)^{\frac{1}{2}} \quad \text{with} \quad m = \tfrac{1}{2}(2^{k_0} - 1)$$

(taking k_0 even). With k_0 the order of 10^{23}, $f(\zeta)$ is indeed a rapidly varying function.

It should be noted that the particular ordering (2), or the particular choice of base functions $e^{im\zeta}/(2\pi)^{\frac{1}{2}}$ is not essential for the application of the isomorphism (4). For example, we could order the states with occupation numbers n_k according to the natural order of the integers $n_0 + 2n_1 + 4n_2 + \cdots$, or might substitute the base functions $(\sin m\zeta/2)/\pi^{\frac{1}{2}}$ for the exponential form. The representations so obtained are similar in appearance to, but considerably more complicated than, the one given in this paper.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 5 MAY 1963

Note on the One-Dimensional Gas of Impenetrable Point-Particle Bosons

T. D. SCHULTZ
International Business Machines Corporation, Thomas J. Watson Research Center, Yorktown Heights, New York
(Received 10 December 1962)

A one-dimensional gas of impenetrable point-particle bosons is considered. An exact expression for the one-particle density matrix is derived in terms of the Fredholm determinant and resolvent of a certain simple kernel. It is proved that for large r, the density matrix is bounded by const $\times r^{-4/\pi^2}$, and that this implies the absence of a generalized Bose–Einstein condensation, contrary to a recent approximate calculation by Girardeau, who first defined such a condensation.

I. INTRODUCTION

In a recent paper,[1] Girardeau introduced the notion of a generalized Bose–Einstein condensation for a system of $\mathfrak{N}$ particles. In such a condensed state the average number n_0 of particles with momentum $k = 0$ is not of order $\mathfrak{N}$; in the limit of an infinite system at fixed density ρ, however, the average number of particles having arbitrarily small momentum is of order $\mathfrak{N}$. Stated precisely, the criterion for a generalized condensation is

$$\lim_{k_0\to 0} \lim_{\substack{\mathfrak{N}\to\infty \\ \rho \text{ fixed}}} \mathfrak{N}^{-1} \sum_{|k|<k_0} n_k \equiv \lim_{k_0\to 0} P_{k_0} = O(1).$$

It was suggested that such a generalized condensation occurs in the ground state of a one-dimensional gas of $\mathfrak{N}$ impenetrable point-particle bosons satisfying periodic boundary conditions over a fundamental domain of length L. For such a system, Girardeau showed the intimate relation to a system of $\mathfrak{N}$ spinless noninteracting fermions: the energy spectra and, hence, the free energies of the two systems are identical, and the boson wavefunctions are simply related to the corresponding fermion wavefunctions (Slater determinants). Unfortunately, despite the knowledge of the exact ground state wavefunction, it was not possible to calculate n_k exactly, so that a certain linearization approximation in the wavefunction had to be made. Within this approximation, Girardeau found a *complete* generalized Bose–Einstein condensation, i.e. $\lim_{k_0\to 0} P_{k_0} = 1$.

It seemed to us that, because of the simplicity of the system, it should be possible to prove or disprove the existence of such a condensation without making any approximations. In the present note, we disprove it. In Sec. II we formulate the problem to stress the similarity with a certain soluble one-dimensional spin-$\frac{1}{2}$ model. Drawing on results for this spin model[2] we are able to derive an exact

[1] M. Girardeau, J. Math. Phys. 1, 516 (1960).

[2] E. Lieb, T. Schultz, and D. Mattis, Ann. Phys. 16, 407 (1961). Hereinafter, referred to as A.

expression for the one-particle density matrix of the hard-sphere boson system in configuration space. In Sec. III we obtain an upper bound to the off-diagonal matrix elements of the density operator between two points at arbitrary distances from one another, and in Sec. IV we use this bound to show that no generalized condensation, whether complete or incomplete, exists.

This result exemplifies the danger of linearization approximations. Whether or not a generalized Bose condensation exists for boson systems with more general interactions and/or more dimensions is, of course, still open.

II. FORMULATION

In the notation of second quantization, the system of impenetrable point-particle bosons is described by the Hamiltonian

$$H = \frac{-\hbar^2}{2m}\int_0^L \psi^\dagger(x)\,\frac{d^2}{dx^2}\,\psi(x)\,dx, \tag{1a}$$

where the impenetrability is expressed through the subsidiary condition on the energy eigenstates Φ_n,

$$\left(\int dx'\psi^\dagger(x)\psi^\dagger(x')\,\delta(x-x')\psi(x')\psi(x)\right)\Phi_n = 0,$$

$$\text{or}\quad (\psi(x))^2\Phi_n = 0,$$

and the field $\psi(x)$ obeys boson commutation rules. It is obviously equivalent to assume that the field $\psi(x)$ is not strictly a boson field, but instead

$$(\psi(x))^2 = (\psi^\dagger(x))^2 = 0, \tag{1b}$$

while retaining

$$[\psi(x),\,\psi(x')] = [\psi^\dagger(x),\,\psi^\dagger(x')]$$
$$= [\psi(x),\,\psi^\dagger(x')] = 0, \qquad x \neq x'. \tag{1c}$$

To stress the analogy with the spin problem already mentioned, the continuum $(0, L)$ is replaced by a set of N points evenly spaced at intervals

$$\epsilon = L/N. \tag{2}$$

The "continuum limit" $N \to \infty$, $\epsilon \to 0$, with L fixed, will be taken at the end of the formal development.

Equations (1) can be expressed in this discrete space if we introduce annihilation and creation operators associated with each point:

$$a_m = \epsilon^{\frac{1}{2}}\psi(m\epsilon); \qquad a_m^\dagger = \epsilon^{\frac{1}{2}}\psi^\dagger(m\epsilon). \tag{3}$$

Then Eq. (1) become

$$H = -\frac{1}{2m\epsilon^2}\sum_1^N a_m^\dagger(a_{m+1} - 2a_m + a_{m-1}), \tag{4a}$$

$$a_m^2 = (a_m^\dagger)^2 = 0, \tag{4b}$$

$$[a_l,\,a_m] = [a_l^\dagger,\,a_m^\dagger] = [a_l,\,a_m^\dagger] = 0, \quad l \neq m. \tag{4c}$$

It is consistent to supplement (4b) with the additional condition

$$\{a_m,\,a_m^\dagger\} = 1. \tag{4d}$$

The total number operator is then

$$\mathfrak{N}_{\text{op}} = \sum_m a_m^\dagger a_m. \tag{4e}$$

The mixed set of commutation/anticommutation rules obeyed by the operators a and $a^\dagger$ and the form of the Hamiltonian (4a) are precisely analogous to the situation with the "isotropic X–Y model" of reference A. The Hamiltonian is diagonalized by first introducing fermion operators

$$C_m = \exp\left(\pi i\sum_1^{m-1} a_l^\dagger a_l\right)a_m;$$
$$C_m^\dagger = a_m^\dagger\exp\left(-\pi i\sum_1^{m-1} a_l^\dagger a_l\right), \tag{5a}$$

in terms of which it becomes

$$H = -\frac{1}{2m\epsilon^2}\sum_1^N C_m^\dagger(C_{m+1} - 2C_m + C_{m-1}), \tag{5b}$$

and then making the canonical transformation

$$C_m = N^{-\frac{1}{2}}\sum_k \exp(ikm\epsilon)\xi_k,$$

$$k = 0,\ \pm 2\pi/L,\ \cdots,\ \pi, \tag{6a}$$

to obtain

$$H = \frac{1}{m\epsilon^2}\sum_k (1 - \cos k)\xi_k^\dagger\xi_k, \tag{6b}$$

and

$$\mathfrak{N}_{\text{op}} = \sum_k \xi_k^\dagger\xi_k. \tag{6c}$$

The ground state of a system of $\mathfrak{N}$ particles[3] is then defined by

$$\xi_k^\dagger\Phi_0 = 0, \qquad |k| < (\mathfrak{N} - 1)\pi/2L,$$
$$\xi_k\Phi_0 = 0, \qquad \text{otherwise.}$$

[3] We assume $\mathfrak{N}$ to be odd. Then the "Fermi sea" is symmetric around $k = 0$. Also, the cyclic boundary condition imposed on the C's and $C^\dagger$'s is then identical to that of the original problem; otherwise, it would be convenient to require in the original problem that the wavefunction change sign when a particle is displaced a distance L.

Equations (6) and (7), when $\epsilon \to 0$, imply the exact results of Girardeau.

Let us now calculate the one-particle density matrix in configuration space, defined by

$$\rho(r', r) = \langle \psi^\dagger(r')\psi(r)\rangle$$
$$\equiv \rho(r - r') \text{ by translational invariance.} \quad (8)$$

Here $\langle \cdots \rangle$ means the expectation value in the state Φ_0. To see how $\rho(r - r')$ is related to n_k we have only to introduce a plane wave expansion of $\psi(r)$:

$$\psi(r) = L^{-\frac{1}{2}} \sum_k \exp(ikr)\eta_k, \quad (9)$$

so that

$$\rho(r - r') = L^{-1} \sum_k \exp[ik(r - r')]n_k, \quad (10a)$$

with

$$n_k = \langle \eta_k^\dagger \eta_k \rangle. \quad (10b)$$

In this formulation, the essential difficulty with bosons, not present with fermions, is that $\langle \xi_k^\dagger \xi_k \rangle$ but not $\langle \eta_k^\dagger \eta_k \rangle$ is easy to compute.

In the discrete space, we define

$$\rho_s = \langle a_m^\dagger a_{m+s} \rangle = \epsilon\rho(s\epsilon), \quad (11)$$

so that

$$n_k = \int_0^L dr \exp(-ikr)\rho(r) = \sum_{s=1}^{N} \exp(-iks\epsilon)\rho_s, \quad (12)$$

the equalities being understood in the continuum limit.

Since n_k must be a real even function of k, ρ_s is real and we can write

$$\rho_s = \tfrac{1}{2}\langle a_m^\dagger a_{m+s} + a_{m+s}^\dagger a_m \rangle$$
$$= \tfrac{1}{2}\langle (a_m^\dagger + a_m)(a_{m+s}^\dagger + a_{m+s})\rangle, \qquad s \neq 0, \quad (13)$$
$$= \epsilon\rho, \qquad s = 0,$$

where $\rho \equiv \mathfrak{N}/L$ is the uniform density of particles.

The evaluation of the expectation value appearing in (13) is carried out in A, Sec. IIC. One finds for this cyclic problem that

$$\rho_s = \tfrac{1}{2} \begin{vmatrix} G_1 & G_2 & \cdots & & G_s \\ G_0 & G_1 & \cdots & & G_{s-1} \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & G_1 & G_2 \\ G_{-(s-2)} & G_{-(s-1)} & \cdots & G_0 & G_1 \end{vmatrix}, \quad (14)$$

where

$$G_m = \langle (C_l^\dagger - C_l)(C_{l+m}^\dagger + C_{l+m})\rangle, \quad l \text{ arbitrary},$$
$$= -\delta_{m0} + \frac{2}{N} \sum_{k \text{ occup.}} \cos km\epsilon$$
$$= \frac{2\epsilon}{L} \frac{\sin \pi\rho m\epsilon}{\sin(\pi m\epsilon/L)} \qquad m \neq 0, \quad (15)$$
$$= -1 + 2\epsilon\rho \qquad m = 0.$$

We are interested in finding ρ_s in the limit $\epsilon \to 0$, $s \to \infty$ with $s\epsilon = r$ arbitrary and fixed. Observe that ρ_s has the structure

$$\rho_s = \tfrac{1}{2} \det \left[\begin{array}{c|c} \boldsymbol{\varepsilon} & \epsilon_s \\ \hline -\mathbf{L} & \boldsymbol{\delta} \end{array}\right], \quad (16)$$

where $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are $(s - 1)$-component vectors with all components proportional to ϵ, while $\mathbf{L}$ is an $(s - 1) \times (s - 1)$ matrix having the form

$$\mathbf{L} = \mathbf{1} - \epsilon\mathbf{K}, \quad (17)$$

where $\mathbf{K}$ has all components of order unity. It is not possible to pass directly to the continuum limit in the matrix of (16) because of the discontinuity going from the first to the second row (and from the $(s - 1)$st to sth column). Also, a direct application of Hadamard's theorem (as in A, p. 464) to (17) gives an upper bound $|\rho_s| \leq O(\epsilon^{\frac{1}{2}})$ which is too weak, in view of the fact that we seek the continuum limit of ρ_s/ϵ. To focus attention on $\mathbf{L}$, which does behave well in the continuum limit, observe that[4]

$$\rho_s = \tfrac{1}{2} \det \left[\begin{array}{c|c} \mathbf{0} & \mathbf{1} \\ \hline -\mathbf{L} & \mathbf{0} \end{array}\right] \left[\begin{array}{c|c} \mathbf{1} & \mathbf{L}^{-1}\boldsymbol{\delta} \\ \hline \boldsymbol{\varepsilon} & \epsilon_s \end{array}\right]. \quad (18)$$

The determinants of the two matrix factors are readily evaluated, so that

$$\rho_s = \tfrac{1}{2}L(\epsilon_s - \boldsymbol{\varepsilon}\mathbf{L}^{-1}\boldsymbol{\delta}), \quad (19)$$

where $L = \det \mathbf{L}$. This equation holds even when $\mathbf{L}$ is singular, if by $L\mathbf{L}^{-1}\,\boldsymbol{\delta}$ we understand the appropriate linear combination of minors.

Equation (19) is the principal formal result. In the continuum limit, L is just the Fredholm determinant $D(\lambda = 1)$ associated with the integral equation

$$\phi(x) = \lambda \int_a^{a+r} K(x - y)\phi(y)\, dy, \quad a \text{ arbitrary}, \quad (20a)$$

with

$$K(x - y) = (\mathbf{K})_{x/\epsilon, y/\epsilon} = \frac{2}{L} \frac{\sin \pi\rho(x - y)}{\sin(\pi(x - y)/L)}. \quad (20b)$$

[4] I am indebted to Dr. E. Lieb and Dr. D. Jepsen for this observation, leading to Eq. (19).

Also in this limit $L(\mathbf{L}^{-1})_{x/\epsilon, y/\epsilon}$ is what is usually called $D(x, y; \lambda = 1)$. In fact, this is the determinant that should stand in (19), which is then valid even if $\mathbf{L}^{-1}$ does not exist. Thus, the determination of $\rho(r)$ is reduced to the evaluation of $D(1)$ and $D(x, y: 1)$ for the kernel $K(x - y)$. Unfortunately, it has not been possible to evaluate these quantities exactly. We now derive an upper bound to $\rho(r)$ which is sufficient for a study of n_k as $k \to 0$.

III. UPPER BOUND TO $\rho(r)$.

We consider separately the magnitudes of the two terms occurring in Eq. (19). We restrict ourselves to $r = s\epsilon < \frac{1}{2}L$, because $\rho(r)$ is periodic with period L, and, being real, it is also even. Thus $\rho(L - r) = \rho(r)$. The methods employed in this section and the associated appendix are due in large part to Jepsen.[4]

In the first term of (19),

$$\tfrac{1}{2}L\epsilon_s \leq \tfrac{1}{2}\,|L|\,|\epsilon_s|. \tag{21a}$$

But

$$|L| = [\prod_j \Lambda_j^2]^{\frac{1}{2}}$$

$$\leq \exp\,[\tfrac{1}{2}\sum_j (\Lambda_j^2 - 1)] = \exp\,[\tfrac{1}{2}\,\mathrm{tr}\,(\mathbf{L}^2 - \mathbf{1})], \tag{21b}$$

where the Λ's are the eigenvalues of $\mathbf{L}$. Also, since $2/\pi \leq |x^{-1}\sin x| \leq 1$,

$$|\epsilon_s| \leq s^{-1}\,|\sin \pi\rho s\epsilon| \leq \pi\rho\epsilon. \tag{21c}$$

In the second term of (19),

$$-\tfrac{1}{2}L\boldsymbol{\varepsilon}\mathbf{L}^{-1}\boldsymbol{\delta} \leq \tfrac{1}{2}\,|L|\,|\sum_j \Lambda_j^{-1}(\boldsymbol{\varepsilon}\cdot\boldsymbol{\phi}_j)(\boldsymbol{\delta}\cdot\boldsymbol{\phi}_j)|, \tag{22}$$

where the $\boldsymbol{\phi}_j$ are the eigenvectors of $\mathbf{L}$.[5] If Λ_0 is the eigenvalue of minimum absolute value, then

$$|\sum_j \Lambda_j^{-1}(\boldsymbol{\varepsilon}\cdot\boldsymbol{\phi}_j)(\boldsymbol{\delta}\cdot\boldsymbol{\phi}_j)|$$

$$\leq |\Lambda_0^{-1}| \sum_j |\boldsymbol{\varepsilon}\cdot\boldsymbol{\phi}_j|\,|\boldsymbol{\delta}\cdot\boldsymbol{\phi}_j| \leq |\Lambda_0^{-1}|\,\boldsymbol{\varepsilon}^2, \tag{23a}$$

where we've used Schwarz's inequality, the completeness of the $\boldsymbol{\phi}$'s, and the fact that $|\boldsymbol{\varepsilon}| = |\boldsymbol{\delta}|$, which is apparent from (14) and (16). But

$$|L|\,|\Lambda_0^{-1}| = (\prod_{j\neq 0} \Lambda_j^2)^{\frac{1}{2}}$$

$$\leq \exp\,[\tfrac{1}{2}\sum_{j\neq 0} (\Lambda_j^2 - 1)]$$

$$= \exp\,[\tfrac{1}{2}\,\mathrm{tr}\,(\mathbf{L}^2 - \mathbf{1})]\exp\,(1 - \Lambda_0^2)$$

$$\leq \exp\,[1 + \tfrac{1}{2}\,\mathrm{tr}\,(\mathbf{L}^2 - \mathbf{1})], \tag{23b}$$

so that

$$|\tfrac{1}{2}L\boldsymbol{\varepsilon}\mathbf{L}^{-1}\boldsymbol{\delta}| \leq \tfrac{1}{2}\boldsymbol{\varepsilon}^2 \exp\,[1 + \tfrac{1}{2}\,\mathrm{tr}\,(\mathbf{L}^2 - \mathbf{1})]. \tag{24}$$

Finally, combining (21) and (24),

$$\rho_s \leq \tfrac{1}{2}(\pi\rho\epsilon + e\boldsymbol{\varepsilon}^2) \exp\,[\tfrac{1}{2}\,\mathrm{tr}\,(\mathbf{L}^2 - \mathbf{1})]. \tag{25}$$

In the appendix it is shown that, in the continuum limit, for arbitrary r_0,

$$\boldsymbol{\varepsilon}^2 \leq \epsilon(2\rho + C_1/L), \tag{26}$$

and

$$\tfrac{1}{2}\,\mathrm{tr}\,(\mathbf{L}^2 - \mathbf{1}) \leq \epsilon\left[-\frac{4}{\pi^2}\ln\left(\frac{r}{r_0}\right)\right.$$

$$\left. + \tfrac{1}{2}(C_2 + C_3) + O\left(\frac{1}{r}\right)\right], \qquad r > r_0. \tag{27}$$

Equation (25) then gives the desired result:

$$\rho(r) = \lim_{\substack{\epsilon\to 0 \\ s = r/\epsilon}} \rho_s/\epsilon \leq Cr^{-4/\pi^2}. \tag{28}$$

Here, C, C_1, C_2, C_3, and C_4 are all constants independent of r, L, and ρ.

IV. NONEXISTENCE OF A GENERALIZED CONDENSATION

We wish to compute P_{k_0} defined by

$$P_{k_0} = \lim_{\substack{\mathfrak{N}\to\infty \\ \rho\ \text{fixed}}} \mathfrak{N}^{-1} \sum_{|k|<k_0} n_k, \tag{29}$$

where

$$n_k = \int_0^L dr\,\exp\,(-ikr)\rho(r). \tag{12}$$

We shall use only the property that $\rho(r) = O(1)$ for $r < r_0$ and $|\rho(r)|^2 < \text{const} \times r^{-2\beta}$, $\beta > 0$, for $r > r_0$, where r_0 is now a large positive constant. For this particular model we have seen that $\beta = 4/\pi^2$.

Then

$$\mathfrak{N}^{-1} \sum_{|k|<k_0} n_k = \frac{1}{\rho L}\int_0^L dr \sum_{|k|<k_0} \exp\,(-ikr)\rho(r)$$

$$= \frac{1}{\rho}\int_0^L dr\,\frac{\sin\,(k_0 + \frac{1}{2}\pi)r}{L\sin\,(\pi r/L)}\,\rho(r). \tag{30}$$

Because $\rho(r) = \rho(L - r)$, we take twice the integral

[5] We have supposed that $\Lambda_j \neq 0$. From (20a) it is known that the eigenvalues λ increase indefinitely. Also, it is readily shown that $K(x - y)$ is positive definite, so that all the λ's are positive. This implies that the Λ's, which are related to the λ's by $\Lambda_j = 1 - \lambda_j^{-1}$, are all <1, and have $\Lambda = 1$ as a limit point. In general, the Λ's being discrete, none of them is precisely $= 0$; only for a discrete set of values of r will some $\Lambda = 0$. We, therefore, exclude this set of measure zero in carrying out the proof. Since $\rho(r)$ is obviously a continuous function of r, the bound thus derived can then be extended to these values of r.

from 0 to $\frac{1}{2}L$. Let us divide the range of integration into two parts, $(0, r_0)$ and $(r_0, \frac{1}{2}L)$. On $(0, r_0)$,

$$\int_0^{r_0} dr \frac{\sin\left(k_0 + \frac{\pi}{L}\right)r}{L \sin(\pi r/L)} \rho(r) \leq \int_0^{r_0} dr \frac{\left(k_0 + \frac{\pi}{L}\right)r}{2r} |\rho(r)| \leq \left(k_0 + \frac{\pi}{L}\right) \times \text{const.} \quad (31)$$

On $(r_0, \frac{1}{2}L)$, we expand $\sin(k_0 + \pi/L)r$, obtaining two terms:

$$M_1 \equiv L^{-1} \int_{r_0}^{\frac{1}{2}L} dr \cos k_0 r \rho(r) \leq L^{-1} \int_{r_0}^{\frac{1}{2}L} dr |\rho(r)| \leq \text{const.} \times L^{-\beta}, \quad (32a)$$

$$M_2 \equiv \int_{r_0}^{\frac{1}{2}L} dr \frac{\sin k_0 r}{L \tan(\pi r/L)} \rho(r). \quad (32b)$$

To estimate M_2, multiply and divide the integrand by r^α, where α is positive and $\frac{1}{2} - \beta < \alpha < \frac{1}{2}$. Then by Schwarz's inequality,

$$\begin{aligned} M_2 &\leq \left[\int_{r_0}^{\frac{1}{2}L} dr \frac{\sin^2 k_0 r}{L^2 r^{-2\alpha} \tan^2(\pi r/L)}\right]^{\frac{1}{2}} \times \left[\text{const} \int_{r_0}^{\frac{1}{2}L} dr\, r^{-2(\beta+\alpha)}\right]^{\frac{1}{2}}, \\ &\leq \left[\frac{1}{\pi^2}\left(\int_0^\infty - \int_0^{r_0}\right) dr \frac{\sin^2 k_0 r}{r^{2(1-\alpha)}}\right]^{\frac{1}{2}} \times \text{const}, \\ &\leq \left[\frac{k_0^{1-2\alpha}}{\pi^2} \int_0^\infty dx \frac{\sin^2 x}{x^{2(1-\alpha)}} - O(k_0^2)\right]^{\frac{1}{2}} \times \text{const}. \end{aligned} \quad (33)$$

Thus, for sufficiently small k_0, $P_{k_0} \leq \text{const} \times k_0$, and so $\lim_{k_0 \to 0} P_{k_0} = 0$; i.e. not even an incomplete generalized condensation exists.

APPENDIX

We determine first an upper bound to ϵ^2, $s = r/\epsilon$, in the continuum limit.

$$\epsilon^2 = 4\epsilon \int_0^r dx \frac{\sin^2 \pi\rho x}{L^2 \sin^2(\pi x/L)} \equiv \epsilon(I_1 + I_2), \quad (A1)$$

where

$$I_1 \equiv 4 \int_0^r dx \frac{\sin^2 \pi\rho x}{\pi^2 x^2} \leq 4 \int_0^\infty dx \frac{\sin^2 \pi\rho x}{\pi^2 x^2} = 2\rho, \quad (A2)$$

and

$$\begin{aligned} I_2 &\equiv 4 \int_0^r dx \sin^2 \pi\rho x [(L^2 \sin^2 \pi x/L)^{-1} - (\pi^2 x^2)^{-1}] \\ &\leq 4 \int_0^{\frac{1}{2}L} dx \sin^2 \pi\rho x \\ &\quad \times [(L^2 \sin^2 \pi x/L)^{-1} - (\pi^2 x^2)^{-1}] = C_1/L. \end{aligned} \quad (A3)$$

C_1 is a positive constant independent of r, L, and ρ. Thus,

$$\epsilon^2 = \epsilon(2\rho + C_1/L). \quad (A4)$$

Now we determine an upper bound for

$$\text{tr}\,(\mathbf{L}^2 - 1) = \epsilon^2 \,\text{tr}\, \mathbf{K}^2 - 2\epsilon \,\text{tr}\, \mathbf{K}.$$

First we have

$$\epsilon \,\text{tr}\, \mathbf{K} = 2\rho r\epsilon. \quad (A5)$$

Second, we consider

$$\epsilon^2 \,\text{tr}\, \mathbf{K}^2 = 4\epsilon \int_0^r dx \int_0^r dy \frac{\sin^2 \pi\rho(x-y)}{L^2 \sin^2 \pi(x-y)/L} \equiv \epsilon(J_1 + J_2), \quad (A6)$$

where

$$J_1 \equiv 4 \int_0^r dx \int_0^r dy \frac{\sin^2 \pi\rho(x-y)}{\pi^2 (x-y)^2}, \quad (A7)$$

and

$$\begin{aligned} J_2 &\equiv 4 \int_0^r dx \int_0^r dy \sin^2 \pi\rho(x-y) \\ &\quad \times [(L^2 \sin^2 \pi(x-y)/L)^{-1} - (\pi^2 (x-y)^2)^{-1}] \\ &\leq 4\pi^{-2} \int_0^{\frac{1}{2}\pi} d\xi \int_0^{\frac{1}{2}\pi} d\eta \\ &\quad \times [(\sin^2(\xi - \eta))^{-1} - (\xi - \eta)^{-2}] \equiv C_2. \end{aligned} \quad (A8)$$

To find an upper bound to J_1 it is convenient to introduce the variables

$$R = \pi\rho(x + y - r) \quad \text{and} \quad \zeta = \pi\rho(x - y),$$

after which J_1 can be transformed to

$$J_1 = 16\pi^{-2} \int_0^\nu dR \int_0^{\nu - R} d\zeta\, \zeta^{-2} \sin^2 \zeta, \quad \nu = \pi\rho r.$$

Integrating by parts on R, and letting $z = \nu - R$, one obtains

$$J_1 = J_{11} + J_{12}, \quad (A9)$$

where

$$J_{11} \equiv 8\pi^{-2}\nu \int_0^\nu dz\, z^{-2} \sin^2 z \leq \frac{4\nu}{\pi}, \quad (A10)$$

and

$$J_{12} \equiv -16\pi^{-2} \int_0^\nu dz\, z^{-1} \sin^2 z.$$

Letting ν_0 be any large positive constant $< \nu$, a bound for J_{12} is obtained:

$$J_{12} \leq -16\pi^{-2} \int_{\nu_0}^{\nu} dz\, z^{-1} \sin^2 z = -8\pi^{-2} \ln (\nu/\nu_0) + 8\pi^{-2} \int_{\nu_0}^{\infty} dz\, z^{-1} \cos 2z - 8\pi^{-2} \int_{\nu}^{\infty} dz\, z^{-1} \cos 2z. \quad \text{(A11)}$$

Because the last term behaves asymptotically like $O(1/\nu)$,

$$J_{12} \leq -8\pi^{-2} \ln (r/r_0) + C_3 + O(1/r) \quad r > r_0. \quad \text{(A12)}$$

Finally, for $r > r_0$,

$$\text{tr } (\mathbf{L}^2 - 1) \leq \epsilon[-8\pi^{-2} \ln (r/r_0) + C_2 + C_3 + O(1/r)]. \quad \text{(A13)}$$

ACKNOWLEDGMENTS

I am very grateful to Dr. D. Jepsen, Dr. E. Lieb, and Dr. D. Mattis for many stimulating discussions and helpful suggestions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 5 MAY 1963

New Method in the Theory of Imperfect Gases and Liquids

ELLIOTT LIEB
International Business Machines Corporation, Thomas J. Watson Research Center, Yorktown Heights, New York
(Received 4 January 1963)

We derive a new expansion method for the thermodynamic properties and correlation functions of imperfect gases. When the interparticle potential is repulsive, we obtain an alternating series of upper and lower bounds for the activity, free energy, pressure, entropy, internal energy, and for the correlation functions. These bounds are valid even if there is a transition. If the potential has an attractive part, we can obtain upper bounds for the activity, free energy, internal energy, and the correlation functions.

INTRODUCTION

IN the standard cluster integral treatment of the imperfect gas,[1] in which one expands thermodynamic quantities in terms of cluster integrals, many questions arise. Firstly, there are the purely mathematical questions such as the convergence of the series; secondly, there are purely practical questions of estimation of error. Even if the density be small, it is difficult to determine the accuracy obtained by retaining the first few terms of the power series. Finally, below the transition temperature, the theory is incapable of estimating any of the thermodynamic quantities. One must then resort to a theory of the liquid state which, at the present time, is even less capable of indicating the errors introduced by the various approximations.

In this paper we introduce an expansion scheme which is similar to, but at the same time different from, the usual cluster integral development. The one important difference is that for a repulsive potential we can estimate the error at each stage of the approximation—and this is true even if there is a transition (as in the case of hard spheres). In other words, we obtain an alternating sequence of upper and lower bounds for the activity, free energy, pressure, entropy, internal energy, and for the correlation functions as well. These bounds allow us to state with confidence where the true results lie with respect to the approximations.

For potentials which have attractive parts, we can obtain upper bounds for the activity, free energy, internal energy and correlation functions.

[1] J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (John Wiley & Sons, Inc., New York, 1940).

1. THE ACTIVITY, FREE ENERGY, AND PRESSURE

We define the configurational partition function for N particles to be

$$Z_N = V^{-N} \int_V \prod_1^N d^3\mathbf{x}_i I_N(\mathbf{x}_1, \cdots, \mathbf{x}_N), \quad \text{(1.1)}$$

where

$$I_N = \prod_{(i,j)} (1 + f_{ij}) \geq 0. \quad \text{(1.2)}$$

The f_{ij} are the Mayer f functions, i.e.

$$f_{ij} = \exp (-U_{ij}/kT) - 1, \quad \text{(1.3)}$$

where $U(\mathbf{x}_i - \mathbf{x}_j)$ is the two-body potential.

$$J_{12} \leq -16\pi^{-2} \int_{\nu_0}^{\nu} dz\, z^{-1} \sin^2 z = -8\pi^{-2} \ln (\nu/\nu_0) + 8\pi^{-2} \int_{\nu_0}^{\infty} dz\, z^{-1} \cos 2z - 8\pi^{-2} \int_{\nu}^{\infty} dz\, z^{-1} \cos 2z. \quad \text{(A11)}$$

Because the last term behaves asymptotically like $O(1/\nu)$,

$$J_{12} \leq -8\pi^{-2} \ln (r/r_0) + C_3 + O(1/r) \quad r > r_0. \quad \text{(A12)}$$

Finally, for $r > r_0$,

$$\text{tr } (\mathbf{L}^2 - 1) \leq \epsilon[-8\pi^{-2} \ln (r/r_0) + C_2 + C_3 + O(1/r)]. \quad \text{(A13)}$$

ACKNOWLEDGMENTS

I am very grateful to Dr. D. Jepsen, Dr. E. Lieb, and Dr. D. Mattis for many stimulating discussions and helpful suggestions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 5 MAY 1963

New Method in the Theory of Imperfect Gases and Liquids

ELLIOTT LIEB
International Business Machines Corporation, Thomas J. Watson Research Center, Yorktown Heights, New York
(Received 4 January 1963)

We derive a new expansion method for the thermodynamic properties and correlation functions of imperfect gases. When the interparticle potential is repulsive, we obtain an alternating series of upper and lower bounds for the activity, free energy, pressure, entropy, internal energy, and for the correlation functions. These bounds are valid even if there is a transition. If the potential has an attractive part, we can obtain upper bounds for the activity, free energy, internal energy, and the correlation functions.

INTRODUCTION

IN the standard cluster integral treatment of the imperfect gas,[1] in which one expands thermodynamic quantities in terms of cluster integrals, many questions arise. Firstly, there are the purely mathematical questions such as the convergence of the series; secondly, there are purely practical questions of estimation of error. Even if the density be small, it is difficult to determine the accuracy obtained by retaining the first few terms of the power series. Finally, below the transition temperature, the theory is incapable of estimating any of the thermodynamic quantities. One must then resort to a theory of the liquid state which, at the present time, is even less capable of indicating the errors introduced by the various approximations.

In this paper we introduce an expansion scheme which is similar to, but at the same time different from, the usual cluster integral development. The one important difference is that for a repulsive potential we can estimate the error at each stage of the approximation—and this is true even if there is a transition (as in the case of hard spheres). In other words, we obtain an alternating sequence of upper and lower bounds for the activity, free energy, pressure, entropy, internal energy, and for the correlation functions as well. These bounds allow us to state with confidence where the true results lie with respect to the approximations.

For potentials which have attractive parts, we can obtain upper bounds for the activity, free energy, internal energy and correlation functions.

[1] J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (John Wiley & Sons, Inc., New York, 1940).

1. THE ACTIVITY, FREE ENERGY, AND PRESSURE

We define the configurational partition function for N particles to be

$$Z_N = V^{-N} \int_V \prod_1^N d^3\mathbf{x}_i I_N(\mathbf{x}_1, \cdots, \mathbf{x}_N), \quad \text{(1.1)}$$

where

$$I_N = \prod_{(i,j)} (1 + f_{ij}) \geq 0. \quad \text{(1.2)}$$

The f_{ij} are the Mayer f functions, i.e.

$$f_{ij} = \exp (-U_{ij}/kT) - 1, \quad \text{(1.3)}$$

where $U(\mathbf{x}_i - \mathbf{x}_j)$ is the two-body potential.

The excess free energy per particle, F, is

$$F = -kTN^{-1} \ln Z_N. \tag{1.4}$$

In terms of F we can define other thermodynamic quantities[2]: The activity,

$$z = \rho(Z_N/Z_{N+1}) = \rho \exp\left\{\frac{1}{kT}\left[F + \rho(\partial F/\partial\rho)\right]\right\}, \tag{1.5}$$

where $\rho = N/V$; the excess chemical potential,

$$\mu = kT \ln (z/\rho); \tag{1.6}$$

the fugacity,

$$f = kTz; \tag{1.7}$$

and the pressure[3]

$$P = \rho kT + \rho^2(\partial F/\partial\rho) = \rho(kT + \mu - F),$$
$$= kT\left\{\rho \ln z - \int_0^\rho \ln [z(\rho')]\, d\rho'\right\}. \tag{1.8}$$

Suppose now that we have an upper (or lower) bound for z for all densities up to a certain maximum, R, (i.e., $\bar{z} \geq z$ for all $0 \leq \rho \leq R$). Then, from Eq. (1.5),

$$(\partial/\partial\rho)\rho F = kT \ln (\bar{z}/\rho) + \xi, \tag{1.9}$$

where $\xi(\rho) \leq 0$ (or ≥ 0). Consequently,

$$\rho F(\rho) = kT \int_0^\rho \ln\left(\frac{\bar{z}(\rho')}{\rho'}\right) d\rho' + \int_0^\rho \xi(\rho')\, d\rho'. \tag{1.10}$$

It is thus clear from Eq. (1.10) that an upper (lower) bound for z implies an upper (lower) bound for F for $0 \leq \rho \leq R$. The converse is not true, of course. Both an upper and a lower bound for z, will together yield both an upper and a lower bound for the pressure by virtue of Eq. (1.8).

Let us now direct our attention to purely repulsive potentials so that

$$0 \geq f_{ij} \geq -1. \tag{1.11}$$

We shall derive a sequence, $z_1, z_2, \cdots$ for z such that z_n is an upper (lower) bound for z if n is even (odd) and if $0 \leq \rho \leq R_n$. Furthermore, the critical density, R_n, beyond which z_n ceases to be a bound, is infinite for n odd, and for n even generally increases with n and appears to satisfy $\lim_{n\to\infty} R_n = \infty$, although we cannot prove this latter statement. The upper/lower bounds for z generate a sequence of upper/lower bounds, $F_1, F_2, \cdots$ for F. Moreover, one can use z_{n+1} in conjunction with F_n to obtain an upper/lower bound sequence $P_1, P_2, \cdots$ for P. Alternatively, one can use z_n in conjunction with F_{n+1} to obtain a sequence $P_1', P_2', \cdots$.

From Eq. (1.2) we have

$$I_N(\mathbf{x}_1, \cdots, \mathbf{x}_N) = \prod_{i=2}^{N} (1 + f_{1i}) I_{N-1}(\mathbf{x}_2, \cdots, \mathbf{x}_N). \tag{1.12}$$

Since f satisfies (1.11) and since $I_{N-1} \geq 0$, the theorem proved in the Appendix tells us that

$$\begin{aligned} I_N &< I_{N-1} \\ &> (1 + \sum_i f_{1i}) I_{N-1} \\ &< (1 + \sum_i f_{1i} + \sum_{(i,k)} f_{1i} f_{1k}) I_{N-1} \\ &> (1 + \sum_i f_{1i} + \sum_{(i,k)} f_{1i} f_{1k} \\ &\qquad + \sum_{(i,k,l)} f_{1i} f_{1k} f_{1l}) I_{N-1}, \\ &\text{etc.} \end{aligned} \tag{1.13}$$

Using the first of the inequalities (1.13), we can easily do the $\mathbf{x}_1$ integration in Eq. (1.1) with the result that

$$Z_N < Z_{N-1}, \tag{1.14}$$

or

$$z > z_1 = \rho. \tag{1.15}$$

Using z_1 in Eq. (1.10) we have

$$F > F_1 = 0. \tag{1.16}$$

It will be noted that $R_1 = \infty$.

Using the second inequality of (11), we can still do the $\mathbf{x}_1$ integration because f_{1i} is a function of $\mathbf{x}_1 - \mathbf{x}_i$. If we define

$$\int_V f(\mathbf{x})\, d^3\mathbf{x} = \beta_1 < 0 \tag{1.17}$$

(this is the same as the irreducible cluster integral β_1 defined by Mayer[1]), then

$$Z_N > (1 + \rho\beta_1) Z_N^{-1}, \tag{1.18}$$

or

$$z < z_2 = \rho(1 + \rho\beta_1)^{-1}. \tag{1.19}$$

[2] We have followed, with slight modifications, the terminology of H. L. Friedman [*Ionic Solution Theory* (Interscience Publishers, Inc., New York, 1962)] and T. L. Hill [*Statistical Mechanics* (McGraw Hill Book Company, Inc., New York, 1956)]. The *total* free energy per particle and chemical potential are, respectively, $F + kT[\frac{3}{2}\ln(h^2/2\pi mkT) + \ln\rho - 1)$ and $\mu + kT[\ln\rho + \frac{3}{2}\ln(h^2/2\pi mkT)]$, where m is the particle mass and h is Planck's constant.

[3] In the following we shall take it for granted that z, F, μ, and P are intensive variables, i.e., that they depend on N and V only in the combination $\rho = N/V$. If one does not wish to accept this fact as proved, then one can still carry through the following analysis, although with a great deal of further complication. We mention this because if one were willing to cope with the further complication, one could go a long way towards proving that the above thermodynamic quantities are indeed intensive.

In this case $R_2 = -\beta_1$. The corresponding bound for F is

$$F < F_2 = -kT \times \{[(1 + \rho\beta_1)/\rho\beta_1] \ln (1 + \rho\beta_1) - 1\}. \quad (1.20)$$

For the pressure we obtain

$$P < P_1 = \rho kT \times [1 - \ln (1 + \rho\beta_1)] \sim \rho kT[1 - \rho\beta_1], \quad (1.21a)$$

$$P > P_1' = \rho kT\{[(1 + \rho\beta_1)/\rho\beta_1] \ln (1 + \rho\beta_1)\} \sim \rho kT[1 + \tfrac{1}{2}\rho\beta_1]. \quad (1.21b)$$

Thus P_1 and P_1' give the first virial coefficient correctly. The second virial coefficient is actually $-\frac{1}{2}\beta_1$ as will be clear when we consider P_2 and P_2'.

We next consider the third inequality in (1.13). The first two terms can be evaluated as before in Eq. (1.18). The third term,

$$\frac{N(N-1)}{2} \int_V \prod_1^N d^3\mathbf{x}_i f_{12} f_{13} I_{N-1}(2, \cdots, N), \quad (1.22)$$

is, unfortunately, not a factorable integral. We can, however, retain the sense of the inequality by replacing I_{N-1} by the first inequality of (1.13), viz. $I_{N-1}(2, \cdots, N) < I_{N-2}(3, \cdots, N)$. We can now do the integral in (1.22) and we obtain

$$\rho < (1 + \rho\beta_1)z + \tfrac{1}{2}\rho\beta_1^2 z^2. \quad (1.23)$$

We now define the polynomial

$$h(z) = \tfrac{1}{2}\rho\beta_1^2 z^2 + (1 + \rho\beta_1)z - \rho. \quad (1.24)$$

Since $h(0) < 0$ and $h(\pm\infty) = +\infty$, we see that h has two real roots: $\alpha(\rho) > 0$ and $\beta(\rho) < 0$. Eq. (1.23) tells us that either $z > \alpha$ or $z < \beta$. The second possibility can be excluded because of Eq. (1.15) and thus

$$z > z_3 = \alpha(\rho) \quad (1.25)$$

for all ρ.

It is well known[4] that for any system in thermodynamic equilibrium, the pressure is a continuous, monotonically increasing function of the density for a fixed temperature. Equation (1.8) tells us that $z(\rho)$ has the same property. A good upper or lower bound for z should have this same property and we shall now show that $\alpha(\rho)$ in fact does (z_1 and z_2 clearly have this property). Since

$$h(z) = \tfrac{1}{2}\rho\beta_1^2(z - \alpha)(z - \beta), \quad (1.26)$$

$$dh/dz|_{z=\alpha} = \tfrac{1}{2}\rho\beta_1^2(\alpha - \beta) > 0. \quad (1.27)$$

But if we differentiate the equation $h(z) = 0$ with respect to ρ, we find

$$\frac{d\alpha}{d\rho} = \frac{\alpha}{\rho}\left(\frac{dh}{dz}\bigg|_\alpha\right)^{-1} > 0. \quad (1.28)$$

The next problem is to find F_3. To do the integral in Eq. (1.10) we write

$$\int_0^\rho d\rho' \ln \alpha(\rho') = \rho \ln \alpha(\rho) - \int_0^\rho \frac{d\alpha}{d\rho'}\frac{\rho'}{\alpha} d\rho' = \rho \ln \alpha(\rho) - \int_0^{\alpha(\rho)} d\alpha \frac{\rho(\alpha)}{\alpha} = \rho \ln \alpha(\rho) - \int_0^{\alpha(\rho)} \frac{d\alpha}{1 - \beta_1\alpha - \frac{1}{2}\beta_1^2\alpha^2}. \quad (1.29)$$

The last integral is easily computed and, collecting terms, we find

$$F > F_3 = kT\left(1 + \ln \frac{[(x+1)^2 + 2x^2]^{\frac{1}{2}} - 1 - x}{x^2} - \frac{1}{\sqrt{3}x} \ln\left(\frac{\sqrt{3}+1}{\sqrt{3}-1}\right) \times \left\{\frac{[(1+x)^2 + 2x^2]^{\frac{1}{2}} - \sqrt{3}x - 1}{-[(1+x)^2 + 2x^2]^{\frac{1}{2}} - \sqrt{3}x + 1}\right\}\right), \quad (1.30)$$

where

$$x = \rho\beta_1 < 0. \quad (1.31)$$

The corresponding bounds for the pressure are

$$P > P_2 = \rho\left(kT + kT \ln\left(\frac{z_3}{\rho}\right) - F_2\right) = \rho kT\left\{\ln \frac{[(1+x)^2 + 2x^2]^{\frac{1}{2}} - 1 - x}{x^2} + \frac{1+x}{x} \ln (1+x)\right\} \sim -\rho kT(1 - \tfrac{1}{2}x - \tfrac{19}{24}x^2), \quad (1.32)$$

and

$$P < P_2' = \rho(kT + kT \ln (z_2/\rho) - F_3) = -\rho kT\left(\ln (1+x) + \ln \frac{[(1+x)^2 + 2x^2]^{\frac{1}{2}} - 1 - x}{x^2} + \frac{1}{\sqrt{3}x} \ln\left(\frac{\sqrt{3}+1}{\sqrt{3}-1}\right) \times \left\{\frac{[(1+x)^2 + 2x^2]^{\frac{1}{2}} - \sqrt{3}x - 1}{-[(1+x)^2 + 2x^2]^{\frac{1}{2}} - \sqrt{3}x + 1}\right\}\right) \sim \rho kT(1 - \tfrac{1}{2}x + \tfrac{17}{24}x^2). \quad (1.33)$$

[4] T. L. Hill, reference 2, Sec. 28.

Equations (1.32) and (1.33) imply that the second virial coefficient is $-\frac{1}{2}\beta_1$, a well-known result of the Mayer theory.

We should inquire whether the lower bound z_3 is necessarily better than z_1. We see easily from Eq. (1.23) that $z > z_3 > z_1 = \rho$ if $\rho < -2/\beta_1$, and $z > z_1 = \rho > z_3$ if $\rho > -2/\beta_1$. We should not be surprised at this result for, as we remarked in the Appendix, including more terms in (1.13) does not guarantee a better result. Nevertheless, z_3 is still a lower bound for all ρ and Eqs. (1.25)–(1.33) are always valid. We notice, incidentally, that the density beyond which z_1 is the better bound, is twice as large as the density R_2 beyond which the upper bound z_2 ceases to exist.

To find the second upper bound z_4 we use the fourth inequality of (1.13). This time, to retain the sense of the inequality, the term $f_{12}f_{13}I_{N-1}$ must be expanded by using the second inequality of (1.13), that is,

$$\begin{aligned}
&\int f_{12}f_{13}I_{N-1}(2 \cdots N) \\
&> \int f_{12}f_{13}\Big(1 + f_{23} + \sum_{i=4}^{N} f_{2i}\Big)I_{N-2}(3 \cdots N) \\
&> \int f_{12}f_{13}(1 + f_{23})I_{N-2}(3 \cdots N) \\
&\quad + (N-3)\int f_{12}f_{13}f_{24}I_{N-3}(4 \cdots N). \qquad (1.34)
\end{aligned}$$

The integrals in the last line can now be done. For the last term in (1.13) [i.e., $f_{12}f_{13}f_{14}$], we can use the first inequality for I_{N-1} twice [i.e., $I_{N-1}(2 \cdots N) < I_{N-3}(4 \cdots N)$]. Collecting results, one finds

$$\rho > z + \rho[\beta_1 z + \tfrac{1}{2}(\beta_1^2 + 2\beta_2)z^2 + \tfrac{2}{3}\beta_1^3 z^3], \qquad (1.35)$$

where

$$\beta_2 = \frac{1}{2}\int f_{12}f_{13}f_{23}\, d^3\mathbf{x}_2\, d^3\mathbf{x}_3 \qquad (1.36)$$

is the three-particle irreducible cluster integral.

We shall not discuss Eq. (1.35) in any detail; instead we shall discuss the general features of (1.35) and all higher-order equations obtained by taking more and more terms in the series of inequalities (1.13). As we have already seen in connection with z_3 and in the derivation of (1.35), to obtain an inequality involving z [such as (1.35)] one must use (1.13) many times, applying it over and over again until one obtains factorable integrals which can be evaluated in terms of Z_{N-i} and certain cluster integrals. One will always end up with either of the following types of inequalities:

$$\rho < z + \rho \sum_{i=1}^{L} a_i z^i \quad (L \text{ even}), \qquad (1.37a)$$

or

$$\rho > z + \rho \sum_{i=1}^{L} a_i z^i \quad (L \text{ odd}), \qquad (1.37b)$$

The a_n's (which will in general depend on L) will be given in terms of cluster integrals. The former inequality, (1.37a), will hold when L is even and a_L will be positive. For the case of odd L, (1.37b) will hold and a_L will be negative. It is important to note, however, that for a given L, the coefficients a_n are not uniquely determined. Beginning with $L = 2$, one can easily see that in applying the inequality (1.13) many times, several choices will present themselves. When L is even, for example, the term $f_{12}f_{13}I_{N-1}(2 \cdots N)$ can be replaced simply by $f_{12}f_{13}I_{N-2}(3 \cdots N)$, as we did for the z_3 case, or we can replace it by $f_{12}f_{13}(1 + f_{23})I_{N-2}(3 \cdots N)$, or we can replace it by $f_{12}f_{13}(1 + f_{23})[1 + (N-2)f_{24} + \frac{1}{2}(N-2)(N-3)f_{24}f_{25}]I_{N-2}(3 \cdots N)$, which can then be further decomposed. There are many possibilities, and we shall return later to a discussion of some of the possible choices.

We first discuss how (1.37a) can yield a lower bound, and (1.37b) an upper bound, for z. The two cases must be considered separately.

L even: Define

$$H(z) = 1 - \sum_{1}^{L} a_i z^i, \qquad (1.38)$$

whence

$$\rho(z) < \bar{\rho}(z) = z/H(z), \qquad (1.39)$$

if $H(z) > 0$. Since $H(0) = 1$ and $H(\pm\infty) = -\infty$, $H(z)$ has at least one positive root—call the smallest one Q. Thus for $0 \leq z \leq Q$, (1.39) gives an upper bound, $\bar{\rho}$, for ρ as a function of z. Furthermore, $\bar{\rho} \to +\infty$ as $z \to Q$. We shall now prove that $\bar{\rho}(z)$ is a monotonically increasing function of z. Consequently, we are permitted to invert (1.39) and obtain a lower bound, $\bar{z}$, for z which satisfies the equation

$$0 = \bar{z} - \rho H(\bar{z}) \equiv h(\bar{z}). \qquad (1.40)$$

This lower bound is a monotonically increasing function of ρ, and as ρ goes from zero to infinity, $\bar{z}$ goes from 0 to Q. Since the true $z(\rho)$ is unbounded as $\rho \to \infty$ (we have already proved $z > \rho$), $\bar{z}$ will not be a very good bound for large ρ. On the other hand, there is reason to believe that the value of Q generally

tends to increase with L. Before turning to the proof of monotonicity, we remark that even if $\bar{p}(z)$ were not monotonic we could always construct a monotonic function by "rectifying" the $\bar{p}(z)$ curve, because we have already proved that the true $p(z)$ is a monotonic function. The lower bound $\bar{z}$, would then be a discontinuous function of ρ.

Proof of Monotonicity: Let us regard ρ as a positive parameter in the polynomial Eq. (1.40). It is sufficient to prove that $d\bar{z}/d\rho > 0$. Equation (1.40) has L roots; $\bar{z}$ is the root which vanishes when $\rho = 0$. Let ρ be very small. The L roots are approximately $\bar{z} = \rho$ and the $L - 1$ roots of $-a_L x^{L-1} = 1/\rho$. Since $a_L > 0$, none of these latter roots will be positive real, so that when ρ is small, $\bar{z}$ will be the only positive real root. For the general ρ, let the real roots be $r_1 > r_2 > \cdots$. Since real roots always appear and disappear in pairs at a common point, $\bar{z}$ will always be either r_1 or r_3 or r_5, etc. and never r_2 or r_4, etc. As we showed in the discussion leading to Eq. (1.28),

$$\frac{d\bar{z}}{d\rho} = \frac{\bar{z}}{\rho}\left(\frac{\partial h}{\partial z}\bigg|_{z=\bar{z}}\right)^{-1}$$

$$= \frac{\bar{z}}{\rho^2 a_L}\prod_{j\neq i}(\bar{z} - r_j)^{-1}\prod_K(\bar{z} - \alpha_K)^{-1}, \qquad (1.41)$$

where $\bar{z} = r_i$, and α_K are the complex roots. Since i is odd, the two products in Eq. (1.41) are positive. Thus $d\bar{z}/d\rho$ is positive if $\bar{z}$ is positive. But we have already shown that $\bar{z}$ is positive for small ρ. Hence $\bar{z}$ is always positive, Q.E.D.

To recapitulate, we have shown that (1.37a) yields a lower bound for z as a function of ρ, valid for all ρ. We turn now to (1.37b).

$$p(z) > \bar{p}(z) = z/H(z). \qquad (1.42)$$

If $H(z)$ has real positive roots, the smallest of which is Q, then (1.42) will imply that $z < Q$ for all ρ. But this is patently absurd since $z(\rho)$ is unbounded. Therefore, $R(z)$ can have no real positive roots and $\bar{p}(z)$ is positive for all $z > 0$. The function $\bar{p}(z)$ has a positive slope for small z and goes to zero for large z. There is at least one value of z for which $d\bar{p}/dz = 0$, but there may be more if Eq. (1.40) has more than two real positive roots for some $\rho > 0$. In any event, let Q be the smallest value of z for which $d\bar{p}/dz = 0$, and let $R = \bar{p}(Q)$. Then for $0 \leq \rho \leq R$ we may invert (1.42) and thereby obtain a monotonically increasing upper bound, $\bar{z}(\rho)$ for z. As we warned at the beginning, the lower bound situation is more favorable than the upper bound. The latter exists only for $\rho < R$ while the former exists for all ρ. But R depends on L and there is reason to believe that it tends to increase with L.

For either L odd or even we can immediately find the corresponding bound $\bar{F}$ for F with the help of Eq. (1.39) or Eq. (1.42). From Eq. (1.10),

$$\bar{F}(\rho) = \frac{kT}{\rho}\int_0^\rho \ln\left(\frac{\bar{z}(\rho')}{\rho'}\right)d\rho'$$

$$= kT\left[1 + \ln\left(\frac{\bar{z}(\rho)}{\rho}\right) - \frac{1}{\rho}\int_0^\rho \frac{d\bar{z}(\rho')}{d\rho'}\frac{\rho'}{\bar{z}(\rho')}\,d\rho'\right]$$

$$= kT\left[1 + \ln\left(\frac{\bar{z}(\rho)}{\rho}\right) - \frac{1}{\rho}\int_0^{\bar{z}(\rho)}\frac{dz}{H(z)}\right]. \qquad (1.43)$$

The last integral is elementary.

We turn now to consider the relationship of our procedure to the standard cluster integral methods of the theory of imperfect gases.[2] In the standard theory one has the following equations:

$$\rho = \sum_1 j b_j z^j = z[1 + \sum_1 (j+1)b_{j+1}z^j], \qquad (1.44)$$

$$P = kT\sum_1 b_j z^j \qquad (1.45a)$$

$$= \rho kT\left[1 - \sum_1 \frac{j}{j+1}\beta_j \rho^j\right], \qquad (1.45b)$$

$$F = kT\left[1 - \ln(1 + \sum_1 (j+1)b_{j+1}z^j) - \left[\frac{1 + \sum_1 b_{j+1}z^j}{1 + \sum_1 (j+1)b_{j+1}z^j}\right]\right], \qquad (1.46a)$$

$$= -kT\sum_1 \frac{1}{j+1}\beta_j \rho^j, \qquad (1.46b)$$

where the b_j are the cluster integrals for j particles ($b_1 = 1$), and the β_j are the irreducible cluster integrals for $j + 1$ particles. There exists a simple algebraic relation between the two.[5] Consider the function

$$\bar{H}(z) = \frac{z}{\rho(z)} = 1 - \sum_1^\infty \alpha_j z^j, \qquad (1.47)$$

where we have supposed that $\bar{H}$ is analytic, and hence has a power series, for sufficiently small $|z|$. Equating the two power series in Eqs. (1.44) and (1.47), and defining $C_j \equiv (j+1)b_{j+1}$, we have

$$\sum_1^\infty (C_j - \alpha_j)z^j = \left(\sum_1^\infty \alpha_j z^j\right)\left(\sum_1^\infty C_j z^j\right), \qquad (1.48)$$

whence

$$\alpha_j = C_j - \sum_{s=1}^{j-1}\alpha_s C_{j-s}. \qquad (1.49)$$

---

[5] T. L. Hill, reference 2, pp. 140, 144.

The α's are thus simply related to the C's; the first few relations are

$$\alpha_1 = C_1 = 2b_2,$$
$$\alpha_2 = C_2 - C_1^2 = 3b_3 - 4b_2^2,$$
$$\alpha_3 = C_3 - 2C_2C_1 + C_1^3 = 4b_4 - 12b_3b_2 + 8b_2^3,$$
$$\alpha_4 = C_4 - 2C_3C_1 - C_2^2 + 3C_1^2C_2 - C_1^4$$
$$= 5b_5 - 16b_4b_2 - 9b_3^2 + 36b_2^2b_3 - 16b_2^4, \quad (1.50)$$

and

$$2b_2 = \alpha_1, \quad 3b_3 = \alpha_2 + \alpha_1^2, \quad 4b_4 = \alpha_3 + 2\alpha_2\alpha_1 + \alpha_1^3,$$
$$5b_5 = \alpha_4 + 2\alpha_3\alpha_1 + \alpha_2^2 + 3\alpha_1^2\alpha_2 + \alpha_1^4. \quad (1.51)$$

Whether or not the power series in (1.44) and (1.47) converge, Eq. (1.49) is a formal relation between the α's and the C's. Furthermore, if $\tilde{H}(z)$ is a meromorphic function of z in the complex plane, then the series in (1.47) will converge up to the first zero of $\rho(z)$, and vice-versa. Thus the series in Eq. (1.47) may very well converge when the conventional series, (1.44) diverges.

It will be seen that the α_j which are sums of products of cluster integrals are the ideal a_j to be used in (1.37a) and (1.37b). It will also be seen that if we look at any a_j in (1.37), there will exist a λ (depending on j), such that if $L > \lambda$, a_j *can be* made equal to α_j. When we say "can be", we have in mind that it is possible to iterate the inequalities, (1.13), so that a_j is never equal to α_j. This brings us to the question of choices for the a_j.

As was remarked above, the b_j (and hence the α_j) can be written in terms of irreducible cluster integrals. Another way to write the b_j is

$$b_j = \sum_{n=j-1}^{\frac{1}{2}j(j-1)} b_j^n, \quad (1.52)$$

where b_j^n is that part of the cluster integral for j particles which contains exactly n bonds. Now, in deriving Eqs. (1.37) up to $L = 3$ (i.e., the equations for z_1 to z_4), it may be observed that we were always able to arrange matters so as to include all terms involving L or fewer f bonds, and no terms involving more than L bonds. In other words, the α_j in Eq. (1.50) can themselves be thought of as sums of integrals involving various numbers of bonds. A term such as $b_2b_3 = (b_2^1)(b_3^2 + b_3^3)$ is a sum of two terms with three and four bonds, respectively. If we admit in α_j only terms with up to L bonds, we have $\alpha_j = 0$ for $j > L$, and obtain the appropriate a_j coefficients in (1.37). [This procedure, it should be remarked, is not the same thing as ignoring terms with $n > L$ in Eq. (1.51)]. Having obtained this set of a_j's, Eq. (1.51) then defines a new set of b_j's. It is these latter coefficients which, when inserted into Eq. (1.46a) gives an upper/lower bound for F as explained before.

Another choice which suggests itself is to replace all β_j's by "ring" integrals, as is so frequently done. This would give rise to a set of b_j's and, by Eq. (1.50) to a set of a_j's. It seems to be true that these a_j's are a legitimate set to insert in Eq. (1.37a) (L would then be infinite), although we have not proved this.

II. CORRELATION FUNCTIONS AND INTERNAL ENERGY

In this section we show how the procedures developed above may be applied to find upper/lower bounds for correlation functions, potentials of average force, internal energy, and entropy.

We define the r-particle correlation function by

$$g(\mathbf{x}_1, \cdots, \mathbf{x}_r) = \frac{V^r}{V^N Z_N} \int I_N(\mathbf{x}_1, \cdots, \mathbf{x}_N) \prod_{r+1}^{N} d^3\mathbf{x}_i. \quad (2.1)$$

From these we define[6] the potentials of average force by

$$w(\mathbf{x}_1, \cdots, \mathbf{x}_r) = -kT \ln g(\mathbf{x}_1, \cdots, \mathbf{x}_r). \quad (2.2)$$

The correlation functions satisfy

$$\int_V g(\mathbf{x}_1, \cdots, \mathbf{x}_r)\, d^3\mathbf{x}_r = Vg(\mathbf{x}_1, \cdots, \mathbf{x}_{r-1}), \quad (2.3)$$

from which it follows that for an infinite system, and for a suitably behaved potential,

$$\lim_{\mathbf{x}_r \to \infty} g(\mathbf{x}_1, \cdots, \mathbf{x}_r) = g(\mathbf{x}_1, \cdots, \mathbf{x}_{r-1}). \quad (2.4)$$

We also define the excess internal energy per particle by

$$E = -T^2 (\partial/\partial T)F/T, \quad (2.5)$$

and the excess entropy per particle by[7]

$$S = -\partial F/\partial T = (E - F)/T. \quad (2.6)$$

The internal energy is also given in terms of the two-particle correlation function by

$$E = \tfrac{1}{2}\rho \int G(\mathbf{x})U(\mathbf{x})\, d^3\mathbf{x}, \quad (2.7)$$

where

$$g(\mathbf{x}_1, \mathbf{x}_2) = G(\mathbf{x}_1 - \mathbf{x}_2). \quad (2.8)$$

[6] This agrees with T. L. Hill, reference 2, p. 193.

[7] The *total* internal energy and entropy per particle are $E + \frac{3}{2}kT$ and $S - k[\frac{3}{2} \ln (h^2/2\pi mkT) + \ln \rho - \frac{5}{2}]$, respectively.

Finally, we can obtain F in terms of E by

$$F = T\int_T^\infty \frac{E(T')}{T'^2}\,dT', \tag{2.9}$$

which latter expression, when inserted into Eq. (2.6), gives S entirely in terms of E.

In order to obtain upper/lower bounds for E, we certainly are not permitted to differentiate the previously derived bounds for F. But from the fact that $U \leq 0$, Eq. (2.7) tells us that upper/lower bounds for G will yield lower/upper bounds for E. These bounds will in turn yield new lower/upper bounds for F with the help of Eq. (2.9), and these new bounds will turn out to be different from the ones we obtained before. Finally, both an upper and a lower bound for E will yield both an upper and a lower bound for S with the help of Eqs. (2.6) and (2.9). Alternatively, one does not have to use Eq. (2.9) in (2.6); one can use one of the bounds for F found previously. Bounds for the correlation functions will also yield bounds for the potentials of average force.

Let us turn now to the question of finding bounds for the correlation functions. We demonstrate the analysis only for the two-particle function, the extension to higher functions being obvious. Using the first inequality of (1.13), together with the fact that $1 + f_{12} = \exp(-U/kT) \geq 0$, we have

$$\begin{aligned} G(\mathbf{x}_1 - \mathbf{x}_2) &= (1 + f_{12})V^{2-N}Z_N^{-1} \\ &\times \int \prod_{j=3}^{N}(1 + f_{1j})I_{N-1}(2, \cdots, N)\prod_3^N d^3\mathbf{x}_i \\ &< (1 + f_{12})V^{2-N}Z_N^{-1}\int I_{N-1}(2, \cdots, N)\prod_3^N d^3\mathbf{x}_i \\ &= (1 + f_{12})(z/\rho). \end{aligned} \tag{2.10}$$

To obtain a lower bound, we use the second inequality of (1.13) on the factor $\prod(1 + f_{1i})$, and the first inequality of (1.13) on the term $f_{13}I_{N-1}(2, \cdots N)$ which arises from the former inequality. The result is

$$G(\mathbf{x}_1 - \mathbf{x}_2) > (1 + f_{12})(z/\rho)[1 + \beta_1 z]. \tag{2.11}$$

The next bound, obtained by using the third inequality of (1.13) on $\prod(1 + f_{1i})$ and the second inequality on $\prod(1 + f_{2i})$, gives an improved dependence on G on $\mathbf{x}$, viz.

$$\begin{aligned} G(\mathbf{x}_1 - \mathbf{x}_2) &< (1 + f_{12})(z/\rho) \\ &\times [1 + \beta_1 z + \tfrac{3}{2}(\beta_1 z)^2 + zK(\mathbf{x})], \end{aligned} \tag{2.12}$$

where

$$K(\mathbf{x}_1 - \mathbf{x}_2) = \int f_{13}f_{23}\,d^3\mathbf{x}_3. \tag{2.13}$$

As before, one can proceed to obtain better inequalities on G by using (1.13) repeatedly. The only unknown quantity in the inequalities is z, but the bounds on z obtained before can be used.

It is to be noticed that at each stage we explicitly take out the factor $(1 + f_{12})$. [We would take out the factor $(1 + f_{12})(1 + f_{13})(1 + f_{23})$ for three particles, etc.]. This means that even for a very large potential, U, the corresponding bound on E [Eq. (2.7)], will be quite sensible. On the other hand, while our bounds on G may be good for small $|\mathbf{x}_1 - \mathbf{x}_2|$, they are incorrect asymptotically. From Eq. (2.4) we see that as $\mathbf{x} \to \infty$, $G(\mathbf{x}) \to 1$, but our bounds do not have this property. What happens, presumably, is that as we use higher inequalities in (1.13), $G(\mathbf{x})$ gets better for larger and larger $\mathbf{x}$, but is never correct at $\mathbf{x} = \infty$. Yet there is some advantage in this for, setting $|\mathbf{x}_1 - \mathbf{x}_2| = \infty$ in Eqs. (2.10) to (2.12), we obtain, respectively,

$$\begin{aligned} &\rho < z, \qquad \rho > z[1 + \beta_1 z], \\ &\rho < z[1 + \beta_1 z + \tfrac{3}{2}\beta_1^2 z^2], \end{aligned} \tag{2.14}$$

and so on for higher inequalities. These latter bounds on z are similar to, but qualitatively different from the bounds [Eq. (1.37)] we obtained before. They are also not quite as good as the former bounds, but they have the advantage of making the integral in Eq. (1.43) easier to do. It is also clear that the bounds in Eq. (2.14) are more similar to the familiar standard Eq. (1.44).

We conclude with the observation that one can obtain some integral inequalities for G. Unfortunately, these cannot be inverted and still keep the sense of inequality. As an example, consider

$$\begin{aligned} G(\mathbf{x}_1 - \mathbf{x}_2) &> (1 + f_{12})V^{2-N}Z_N^{-1} \\ &\times \int (1 + (N-3)f_{13})I_{N-1}(2 \cdots N)\prod_3^N d^3\mathbf{x}_i \\ &= (1 + f_{12})\left(\frac{z}{\rho}\right)\left[1 + \rho\int f_{13}G(\mathbf{x}_3 - \mathbf{x}_2)\,d^3\mathbf{x}_3\right] \\ &> (1 + f_{12})\left(\frac{z}{\rho}\right) + z\int f_{13}G(\mathbf{x}_3 - \mathbf{x}_2)\,d^3\mathbf{x}_3. \end{aligned} \tag{2.15}$$

The last inequality was introduced merely to make the equation tractable by Fourier transforms. It is indeed a great pity that we are not permitted to invert Eq. (2.15). If we could do so [or if we ignore the question of bounds, and regard (2.15) simply as a good approximation to G], then we could obtain straightaway the Debye–Hückel law for an electron gas. There would be no need for the usual tenuous assumptions about analytic continuation and con-

vergence that plague the standard cluster integral treatment of the electron gas.[8] Equation (2.15) does have the advantage, however, that if we insert Eq. (2.10) into its right-hand side, we obtain

$$G(\mathbf{x}_1 - \mathbf{x}_2) > (1 + f_{12})(z/\rho)[1 + \beta_1 z + zK(\mathbf{x})], \quad (2.16)$$

an equation that complements Eq. (2.12) above.

III. MORE GENERAL POTENTIALS

The methods we have presented fail unless $U > 0$. If, on the other hand, $U < 0$ then all the inequalities in (1.13) become lower bounds, viz.

$$I_N > I_{N-1} > (1 + \sum_i f_{1i})I_{N-1} \qquad (U < 0), \quad (3.1)$$

etc.

One could then carry through the same analysis as before and obtain a series of upper bounds for z, F, E, and the correlation functions (but not for P and S). This case is a little different from the previous one in two important respects.

(i) Each bound is valid for all ρ (i.e. $R_n = \infty$ for all n).

(ii) Since all terms are now positive, there are many more possible choices in the application of (1.13). For example, we can see without further ado, that if we truncate the series in Eq. (1.44) we obtain an upper bound for z, viz.,

$$\rho < \sum_{j=1}^{L} j b_j z^j \quad (3.2)$$

for any L.

The purely attractive potential, unfortunately, is not very interesting. For the mixed potential we can use the following decomposition:

$$U(\mathbf{x}) = U^+(\mathbf{x}) + U^-(\mathbf{x}), \quad (3.3)$$

$$U^+(\mathbf{x}) = U(\mathbf{x}) \quad \text{if} \quad U(\mathbf{x}) > 0,$$
$$= 0 \quad \text{if} \quad U(\mathbf{x}) < 0, \quad (3.4)$$

$$U^-(\mathbf{x}) = U(\mathbf{x}) \quad \text{if} \quad U(\mathbf{x}) < 0,$$
$$= 0 \quad \text{if} \quad U(\mathbf{x}) > 0. \quad (3.5)$$

[8] H. L. Friedman, reference 2, Chap. 12.

Consequently,

$$1 + f = \exp(-U/kT) = (1 + f^+)(1 + f^-), \quad (3.6)$$

where

$$f^- > 0, \qquad -1 < f^+ < 0. \quad (3.7)$$

One can now use (1.13) on $(1 + f^+)$ and $(1 + f^-)$ separately. The expansion of $(1 + f^-)$ always gives an upper bound for z; therefore, one can only use combinations of (1.13) applied to $(1 + f^+)$ that likewise give upper bounds for z (i.e. the second, fourth, etc. inequalities).

APPENDIX

Let a_j be a set of real numbers defined for $j = 1, 2, \cdots, M$ such that $0 \geq a_j \geq -1$ for all j. Let $f(s)$ be defined as

$$f(s) = \prod_{j=1}^{M} (1 + sa_j) = \sum_{n=0}^{M} s^n K_n(\{a\}), \quad (A1)$$

where

$$K_n(\{a\}) = \sum_{(\alpha_1, \alpha_2, \cdots, \alpha_n)} a_{\alpha_1} \cdots a_{\alpha_n}, \quad K_1 = 1. \quad (A2)$$

The summation in Eq. (A2) is over all distinct n-tuples $\alpha_1, \cdots, \alpha_n$.

Theorem:[9] If $0 \leq s \leq 1$ then

$$f(s) \gtrless \sum_{n=0}^{L} s^n K_n(\{a\}), \quad (A3)$$

where L is any integer from 0 to M, and where the $>$ sign holds if L is odd, the $<$ sign holds if L is even.

Proof: The sum in Eq. (A3) is just the first $L + 1$ terms in a Taylor series for f about $s = 0$. Define $g_n(s) = d^n f/ds^n$. By the remainder theorem for Taylor's series, the theorem will be proved if we can show that for all $0 < s < 1$, $g_{L+1}(s) > 0$ (L odd), $g_{L+1}(s) < 0$ (L even). But a simple calculation shows that $g_n(s) = n!\, f(s) K_n(\{b\})$ where $b_j = a_j(1 + sa_j)^{-1}$. Since $f(s) > 0$ and $b_j < 0$ for $0 < s < 1$, the theorem is proved.

Remark: There is no guarantee that the $(L + 2)$th partial sum gives a better bound for f than the Lth partial sum. It may very well be worse.

[9] *Note Added in Proof.* Professor Pólya has kindly pointed out to us that a slightly generalized version of this theorem is to be found in G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis* (Springer Verlag, Berlin, 1954), 2nd Ed., Vol. II, Chap. 5, Theorem 163, pp. 67 and 255.

Asymptotic Form of the Coefficients of Some Ising-Model Series

B. W. NINHAM
School of Mathematics, The University of New South Wales, Kensington, New South Wales
(Received 1 November 1962)

The exact asymptotic form of the coefficients of some two-dimensional Ising-model series is derived. A preliminary comparison with some three-dimensional series suggests that the asymptotic nature of their coefficients is not inconsistent with the same analytic behavior.

I. INTRODUCTION

PRINCIPALLY through the pioneering work of Domb[1] and co-workers, methods have been developed to compute a number of terms in a power-series expansion of the partition function of the Ising model. With these results established, a great deal of effort has been spent in trying to predict critical points, and the behavior of thermodynamic functions near the critical point, for many lattices. This is usually accomplished by fitting the coefficients of the series to some functional form by various approximation techniques. A particular prediction is generally supposed to be reasonable if the same technique gives close agreement with the known exact results for the two-dimensional model. Unfortunately one can never be certain of the validity of such predictions unless he knows the precise analytic form of the coefficients or, in other words, the solution of the problem. The recent work of Domb and Sykes[2] and Baker[3] would seem to indicate that the nature of the singularity in the magnetic susceptibility for both two- and three-dimensional lattices is of the same mathematical form. One might expect that such similarities ought somehow to be hidden in the as yet unknown functional form of the coefficients of the various series expansions.

We shall derive exact asymptotic expressions for the coefficients of the series expansions of the known two-dimensional partition function, and spontaneous magnetization. The latter are obtainable through a well-known method of Darboux,[4] while the former can be found by the application of some powerful theorems of Hamy,[5] based on Darboux's work. We then show that the coefficients of the series expansion for the three-dimensional simple cubic lattice appear to exhibit a form not inconsistent with the same analytic behavior as that of the two-dimensional square lattice. Without too much work, we extrapolate to a critical point in agreement with an analysis of Domb.[6]

Such a result is highly suggestive. Indeed, if it could be shown that the three-dimensional partition function belongs to the same class of functions, in a sense to be defined, as does that the two-dimensional partition function (a conjecture which seems likely), the asymptotic behavior of the three-dimensional series could be established with certainty. The problem of predicting critical points from series expansions could then be regarded as solved. Further, if such a proof could be constructed, it might even be possible to predict some of the functions from which these expansions arose. Systematic methods for such predictions can be developed using a modification of the theory of converging factors of Dingle[7] and Mueller.[8]

We shall confine our analysis of the exact two-dimensional results to the simple square lattice. The techniques may be applied with equal facility to other lattices.

II. SPONTANEOUS MAGNETIZATION OF THE SQUARE LATTICE

The spontaneous magnetization of the square lattice was found by Onsager and Kaufman,[9] and Yang and Lee[10] to be

$$M = [(1 + x^2)(1 - 6x^2 + x^4)^{\frac{1}{2}}/(1 - x^2)^2]^{\frac{1}{4}}, \qquad (1)$$

where $x = \exp(-2J/kT)$. J is a constant which measures the strength of the interaction between neighboring lattice points, and T is the temperature.

[1] C. Domb, Advan. Phys. 9, 149 (1960).
[2] C. Domb, and M. F. Sykes, Proc. Roy. Soc. (London) A240, 214 (1957).
[3] G. A. Baker, Phys. Rev. 124, 768 (1961).
[4] M. G. Darboux, J. Math. 3, 377 (1878).
[5] M. Hamy, J. Math. 4, 203 (1908).
[6] C. Domb and M. F. Sykes, Phil. Mag. 2, 733 (1957).
[7] R. B. Dingle, "Asymptotic Expansions and Converging Factors," papers I–VI, Proc. Roy. Soc. (London) A (1957).
[8] H. Mueller, thesis, University of Western Australia (1959).
[9] L. Onsager, Nuovo Cimento (Suppl. 2) 6, 216 (1949).
[10] C. N. Yang, Phys. Rev. 85, 808 (1952).

The corresponding series expansion given by Domb[11] is, up to the term in x^{18},

$$M = 1 - 2x^4 - 8x^6 - 34x^8 - 152x^{10} - 714x^{12} - 3472x^{14} - 17318x^{16} - 88048x^{18}. \quad (2)$$

We derive the asymptotic form of the coefficients of this series from Eq. (1) by a theorem of Darboux.

Suppose that, on the circle of convergence, a function $f(z)$ has an algebraic singularity so that

$$f(z) = (z - \alpha)^k \phi(z) + \psi(z), \quad (3)$$

where ϕ and ψ are two functions regular at $z = \alpha$. Then Darboux has shown that the asymptotic form of the coefficients in the expansion of $f(z)$ can be obtained by substituting for its expansion that of $\phi(\alpha)(z - \alpha)^k$. Higher approximations may be obtained by replacing $f(z)$ by

$$\left[\sum_{r=0}^{p} \frac{(z-\alpha)^r}{r!} \phi^{(r)}(\alpha)\right](z - \alpha)^k,$$

the error involved in stopping at the nth term being always of the order of $1/n$ times this term.

Applying the theorem to the function M of Eq. (1), we write $x^2 = z$, so that

$$M = (1 - z)^{-\frac{1}{2}}(1 + z)^{\frac{1}{2}}(z - 3 + 2\sqrt{2})^{\frac{1}{2}} \times (z - 3 - 2\sqrt{2})^{\frac{1}{2}}. \quad (4)$$

The nearest algebraic singularity to the origin, which gives the dominant contribution, is at $z = 3 - 2\sqrt{2}$. Thus we take $\alpha = (3 - 2\sqrt{2})$, $k = \frac{1}{2}$ and the theorem shows that, to first approximation, the asymptotic expansion of M is given by the expansion of

$$M \simeq [(z - (3 - 2\sqrt{2}))]^{\frac{1}{2}}[(1 - z)^{-\frac{1}{2}}(1 + z)^{\frac{1}{2}} \times (z - 3 - 2\sqrt{2})^{\frac{1}{2}}]_{z=\alpha}. \quad (5)$$

We write

$$M = \sum_{n=0}^{\infty} a_n z^n, \quad (6)$$

and expand $[z - (3 - 2\sqrt{2})]^{\frac{1}{2}}$ in a power series in z. The coefficients are then given, to first Darboux approximation, as

$$a_n \simeq (10\sqrt{2} - 12)^{\frac{1}{2}} \times [(n - 9/8)!/n!\,(-\tfrac{1}{8})!](3 + 2\sqrt{2})^n. \quad (7)$$

The values found from this first approximation, and the second approximation to the coefficients of z^n are compared in Table I with the exact coefficients of Eq. (2). The maximum possible error in the first approximation is of the order of $1/n$ times the approximant, while that of the second is $O(1/n^2)$. The values tabulated lie well within these limits.

TABLE I. Comparison between actual coefficients of the series expansion for M and Darboux first and second approximation.

Order of coefficient of x^2 n	Coefficient Darboux first approximation	Second approximation	Actual coefficient
6	670	700	714
7	3 300	3 410	3 472
8	16 550	17 020	17 318
9	84 830	86 620	88 048

It is clear then that Darboux's method yields an asymptotic form of considerable numerical accuracy. However, it deals only with contributions from the dominant singularity. For our later comparison with three-dimensional series, it is desirable to try to estimate the form of contributions from less important singularities, lying outside the circle of convergence and vanishingly small in numerical magnitude, with increasing order. The methods of Hamy which we use below throw some light on this problem.

III. ASYMPTOTIC BEHAVIOR OF THE PARTITION FUNCTION OF THE SQUARE LATTICE

The partition function in the absence of a magnetic field has been given by Onsager.[11] In terms of a function $\Lambda(1, x)$ it is

$$\ln \Lambda(1, x) = \ln (1 + x^2) + \frac{1}{2\pi}\int_0^{\pi} \ln \{\tfrac{1}{2}[1 - (1 - k_1^2 \sin^2 \varphi)^{\frac{1}{2}}]\}\, d\varphi, \quad (8)$$

where

$$(1/N) \ln (\text{P.F.}) = -\tfrac{1}{2}C \ln x + \ln \Lambda(1, x). \quad (9)$$

C is the coordination number of the lattice, which is 4 for the square lattice. The variable k_1 is defined by

$$k_1 = 4x(1 - x^2)/(1 + x^2)^2. \quad (10)$$

The series expansion in powers of x^2 up to the term in x^{24} of the function $\Lambda(1, x)$ was found by Domb to be

$$\ln \Lambda(1, x) = x^4 + 2x^6 + \tfrac{9}{2}x^8 + 12x^{10} + \tfrac{112}{3}x^{12} + 130x^{14} + 490\tfrac{1}{4}x^{16} + 1956\tfrac{2}{3}x^{18} + 8174\tfrac{1}{5}x^{20} + 35302x^{22} + 156777\tfrac{1}{2}x^{24} + \cdots. \quad (11)$$

[11] L. Onsager, Phys. Rev. 65, 117 (1944).

We integrate by parts once, and find

$$\ln \Lambda(1, x) = \ln (1 + x^2)$$

$$+ \frac{1}{2\pi} \int_0^{\pi} \frac{\phi \cos \varphi}{\sin \varphi} d\varphi[(1 - k_1^2 \sin^2 \varphi)^{-\frac{1}{2}} - 1]. \quad (12)$$

In order to determine the asymptotic behavior of the series expansion for $\ln \Lambda (1, x)$ we first find the coefficient of $(x^2)^n$ in the integrand of Eq. (12), for large n, and then integrate over ϕ. Consider then the function

$$f(z) = f(x^2) = (1 - k_1^2 \sin^2 \varphi)^{-\frac{1}{2}}$$

$$= (1 + z)^2[(1 + z)^4 - 16z(1 - z)^2 \sin^2 \varphi]^{-\frac{1}{2}}. \quad (13)$$

The coefficients of the expansion of $f(z)$ may be found by the method of Darboux. However this deals only with contributions from the chief singularity. We therefore appeal to some theorems of Hamy.

The coefficient b_n of z^n in $f(z)$ may be written as

$$b_n = \frac{1}{2\pi i} \oint z^{-(n+1)} f(z)\, dz, \quad (14)$$

where the integration is to be conducted along a closed contour around the origin in the complex z plane.

A Theorem of Hamy

Definition. Contours of the first and third kind.

Given a point α in the complex plane, and a contour BCD, the contour is defined as one of the *first* kind with respect to α if:

1. The extremities B, D of the contour are at a greater distance from the origin than the point α.

2. The line 0α intersects the contour at a single point or satisfies this condition after a suitable deformation.

This is illustrated in Fig. 1. Consider the contour αb of Fig. 2. Suppose that all the points of the contour are farther from the origin than α. Such a contour is one of the *third* kind with respect to α. Hamy also defines contours of the second kind, but these will not concern us.

Theorem. Evaluation of $I = 1/2\pi i \int F(z)\, z^{-(m+1)}\, dz$ along a contour of the first kind.

Let α be the singularity of $F(z)$ closest to the origin. Suppose that α is separated from the origin and any other singularities of $F(z)$ by finite distances, and that the contour does not include any singularities of $F(z)$ at a distance from the origin $\leq |\alpha|$. Further suppose that in the neighborhood of α, $F(z)$ can be expanded as

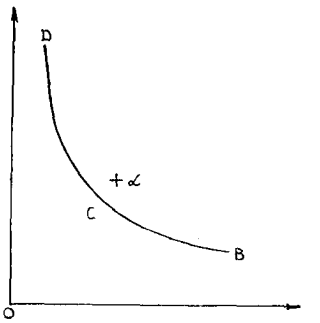


FIG. 1. Contour of the first kind.

$$F(z) = \phi(z) + A_1\left(1 - \frac{z}{\alpha}\right)^{\alpha_1}$$

$$+ A_2\left(1 - \frac{z}{\alpha}\right)^{\alpha_2} + \cdots + A_p\left(1 - \frac{z}{\alpha}\right)^{\alpha_p} \psi_p(z), \quad (15)$$

where $\phi(z)$ is regular, and $\psi_p(z)$ finite in the neighborhood of α. $A_1, A_2, \cdots A_p$ are constants, and $\alpha_i > -1$, $i = 1, 2, 3, \cdots p$, and the α_i satisfy $\alpha_1 < \alpha_2 < \alpha_3 \cdots < \alpha_p$. Under these conditions, the expression

$$\frac{A_1}{\alpha^n} \frac{(n - \alpha_1 - 1)!}{(-\alpha_1 - 1)!\, n!} + \frac{A_2}{\alpha^n} \frac{(n - \alpha_2 - 1)!}{(-\alpha_2 - 1)!\, n!} + \cdots \quad (16)$$

furnishes an asymptotic expansion for the integral I taken along a contour of the first kind with respect to α, in the sense that the ratio of successive terms is of the order of $1/n$.

Corollary 1. Consider the integral I along a closed contour which encircles the origin, and suppose that there are no branch points along the contour. Let α be that singularity of $F(z)$ external to C, which lies closest to the origin. The contour may be deformed arbitrarily. In particular, take a contour DEAB of radius $> |\alpha|$. The contour ABDE, Fig. (3), is clearly of the first kind with respect to α, and the expression (16) is applicable.

Corollary 2. Suppose that the contour ABC is of the first kind with respect to two singular points, α and β, equidistant from the origin. Let E be a point further from the origin than α or β, Fig. 4. Then

$$\int_{ABC} = \int_{ABE} + \int_{EBC}. \quad (17)$$

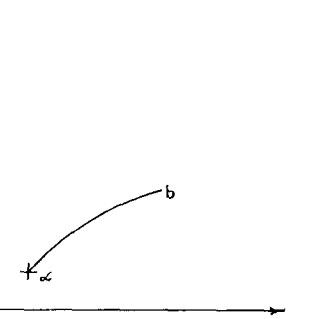


FIG. 2. Contour of the third kind.

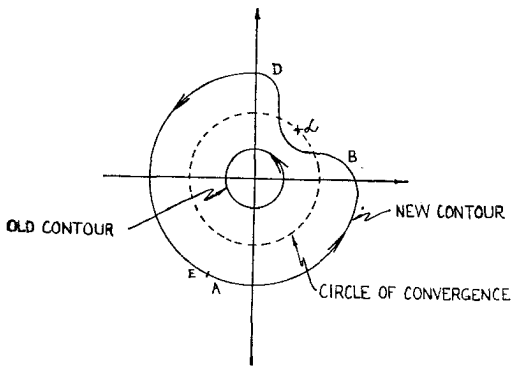


FIG. 3. Illustration of Corollary 1.

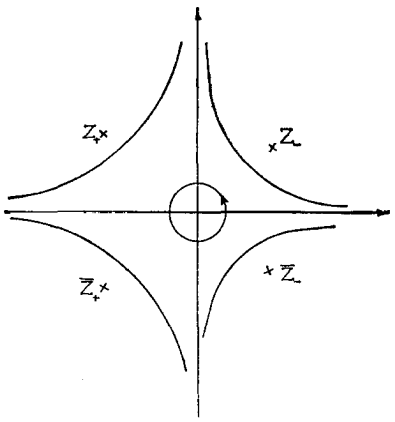


FIG. 5. Contour for the evaluation of b_n.

The paths ABE, EBC are of the first kind with respect to α and β, respectively. The theorem must be applied to both points and the contributions added.

Let us return now to Eqs. (13) and (14) and apply the theorem to the integral b_n. The singularities of $f(z)$ which we denote by z_+, $\bar{z}_+$, z_-, $\bar{z}_-$ where bars indicate complex conjugates, are

$$z_- = (1 - e^{2i\varphi}) \pm [(1 - 2e^{2i\varphi})^2 - 1]^{\frac{1}{2}}; \qquad ph(i) = \tfrac{1}{2}\pi, \tag{18}$$

and since $\phi \geq 0$, $|z_+| \geq |z_-|$. The original contour may be distorted to obtain the new contour shown in Fig. 5. The chief contributions to the integral will be those from the singularities z_-, $\bar{z}_-$. Expanding $f(z)$ about z_- as indicated in Eq. (16), we obtain

$$f(z) = (z - z_-)^{-\frac{1}{2}}\left\{\left[\frac{f(z)}{(z - z_-)^{-\frac{1}{2}}}\right]_{z=z_-} + (z - z_-)\left[\frac{d}{dz} f(z)(z - z_-)^{\frac{1}{2}}\right]_{z=z_-} + \cdots\right\}$$
$$= \left(1 - \frac{z}{z_-}\right)^{-\frac{1}{2}} A_1 + \cdots \tag{19}$$
$$+ \text{higher terms},$$

where

$$A_1 = (-z_-)^{-\frac{1}{2}}[f(z)(z - z_-)^{\frac{1}{2}}]_{z=z_-}. \tag{20}$$

If we compare Eq. (19) with the expansions (14) and (15), we see that for large n the contribution to the integral from z_- is

$$\frac{(n - \frac{1}{2})!}{(-\frac{1}{2})!\, n!} \frac{A_1}{(z_-)^n} + \cdots. \tag{21}$$

Substituting for (z_-) from Eq. (18) we find

$$A_1 = i(\cos^{-\frac{1}{2}}\varphi)(e^{2i\varphi} - 1)^{\frac{1}{2}}. \tag{22}$$

The singularity $\bar{z}_-$ gives an analogous contribution with A_1 replaced in (21) by its conjugate. Then from both singularities, z_- and $\bar{z}_-$, the total contribution to b_n for large n is

$$b_{n-} \simeq 2\frac{(n - \frac{1}{2})!}{(-\frac{1}{2})!\, n!} \operatorname{Re}\left\{\frac{i(e^{2i\varphi} - 1)^{\frac{1}{2}}}{\cos^{\frac{1}{2}}\phi} \frac{1}{(z_-)^n} + \cdots\right\}, \tag{23}$$

with

$$z_- = (1 - 2e^{2i\varphi}) - [(1 - 2e^{2i\varphi})^2 - 1]^{\frac{1}{2}}, \tag{24}$$

and Re denotes the real part of the quantity in brackets.

In precisely the same way, the contributions to b_n from the singularities z_+ and $\bar{z}_+$ are found to be

$$b_{n+} \simeq \frac{2(n - \frac{1}{2})!}{(-\frac{1}{2})!\, n!} \operatorname{Re}\left\{\frac{(e^{2i\varphi} - 1)^{\frac{1}{2}}}{\cos^{\frac{1}{2}}\varphi} \frac{1}{(z_+)^n}\right\}. \tag{25}$$

The term in b_{n+} is vanishingly small compared with b_{n-}. However, we shall retain it for later comparison with a three dimensional series.

We return again to the original integral Eq. (12) whose asymptotic expansion we set out to find. Summarizing briefly, the first step, i.e., the evaluation of the coefficient for large n of (x^{2n}) in the expansion of $(1 - k_1^2 \sin^2\varphi)^{-\frac{1}{2}}$ has been completed, and is given by the sum of b_{n-} and b_{n+}. It remains to carry out the integration over ϕ. Substituting b_{n-} into the integrand of Eq. (12), the coefficient of $(x^2)^n$ for large n becomes

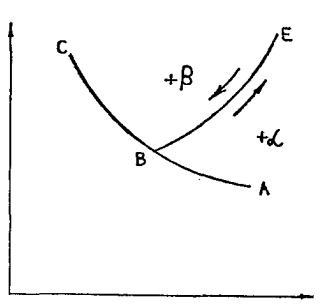


FIG. 4. Illustration of Corollary 2.

$$C_{n-} \simeq \frac{1}{2\pi}\int_0^{\pi} \frac{\phi\cos\varphi}{\sin\varphi}\,d\varphi \times \left(\frac{2(n-\frac{1}{2})!}{(-\frac{1}{2})!\,n!}\right)\mathrm{Re}\left\{\frac{i(e^{2i\varphi}-1)^{\frac{1}{2}}}{\cos^{\frac{1}{2}}\varphi}\frac{1}{(z_-)^n}\right\}. \quad (26)$$

We change variables to $\omega = re^{i\phi}$, so that Eq. (26) becomes

$$C_{n-} \simeq \frac{(n-\frac{1}{2})!}{(-\frac{1}{2})!\,n!}\frac{\sqrt{2}}{\pi} \mathrm{Re}\int_{\epsilon}^{e^{i\pi}} \frac{(\omega^2+1)^{\frac{1}{2}}\ln\omega\,d\omega}{\omega^{\frac{1}{2}}(\omega^2-1)^{\frac{1}{2}}\{1-2\omega^2-[(1-2\omega^2)^2-1]^{\frac{1}{2}}\}^n}.$$

The path of integration is along the unit semicircle in the upper half ω plane. After rationalization of the denominator the integral may be written in the more convenient form

$$C_{n-} \simeq K\,\mathrm{Re}\int_C F(\omega)G^n(\omega)\,d\omega, \quad (27)$$

where

$$K = \frac{\sqrt{2}}{\pi}\frac{(n-\frac{1}{2})!}{(-\frac{1}{2})!\,n!}, \quad (28)$$

$$F(\omega) = (\ln\omega)[\omega^{-\frac{1}{2}}(\omega^2+1)^{\frac{1}{2}}(\omega^2-1)^{\frac{1}{2}}], \quad (29)$$

$$G(\omega) = (1-2\omega^2) + [(1-2\omega^2)^2-1]^{\frac{1}{2}}. \quad (30)$$

A Further Theorem of Hamy

The integral (27) belongs to one of a special class dealt with by Hamy. He is concerned with the asymptotic evaluation of

$$J = \int_C F(\omega)G^n(\omega)\,d\omega, \quad (31)$$

taken along a contour of the third kind, subject to the following restrictions:

1. $|G(\omega)|$ is not a maximum at one of the extremities of the contour.

2. Nowhere along the contour is $G^{(1)}(\omega)$ equal to zero.

3. $|G(\omega)|$ is a maximum at $\omega = \gamma$, and $G(\omega)$ is analytic in the neighborhood of $\omega = \gamma$.

Then, expanding $F(\omega)$ in the vicinity of γ as

$$F(\omega) = (\omega-\gamma)^{\beta}[B_1 + B_2(\omega-\gamma) + B_3(\omega-\gamma)^2 + \cdots], \quad (32)$$

it is shown that J is given by the asymptotic expansion [Eq. (33)] in which the ratio of successive terms is again of the order of $(1/n)$;

$$J \simeq \frac{-2\pi i}{(-\beta-1)!}\left(\frac{G}{G^{(1)}}\right)^{\beta+1}G^n\left\{B_1\frac{(n-\beta)!}{(n+1)!} + (\beta+1)\frac{(n-\beta-1)!}{(n+1)!} \times\left[\frac{B_1}{2}\left[(\beta+2)\frac{GG^{(2)}}{(G^{(1)})^2}-2\beta\right] - \frac{G}{G^{(1)}}B_2\right] + \cdots\right\}. \quad (33)$$

We have used the abbreviation

$$G, G^{(1)}, G^{(2)} \equiv G(\gamma), G^{(1)}(\gamma), G^{(2)}(\gamma). \quad (34)$$

For further remarks on the applicability of the above expression, the original papers of Hamy should be consulated.

Our integral [Eq. (27)] falls into this class, for with $G(\omega) = (1-2\omega^2) + [(1-2\omega^2)^2-1]^{\frac{1}{2}}$, $G(\omega)$ satisfies conditions (1), (2), and (3). Further, $|G(\omega)|$ max $= 3 + 2\sqrt{2}$ at $\omega = \gamma$ where $\gamma = e^{\frac{1}{2}i\pi} = i$. Whence, expanding $F(\omega)$ about $\omega = i$ as indicated in Eq. (32), and applying the theorem, we find, after some algebra that

$$C_{n-} \simeq \mathrm{Re}\,\frac{(3+2\sqrt{2})^n}{(-\frac{1}{2})!}\frac{(n-\frac{1}{2})!}{(-\frac{1}{2})!\,n!}\frac{i}{\sqrt{2}} \times\left(\text{Real term} - \tfrac{3}{2}i\frac{(n-\frac{3}{2})!}{(n+1)!} \times\left\{\frac{\pi}{4}\left[\frac{5}{4}\left(1+\frac{1}{2\sqrt{2}}\right)\right]-\frac{\pi}{2\sqrt{2}}\right\}+\cdots\right). \quad (35)$$

The first term inside the bracket is real and therefore gives no contribution. Finally, we obtain

$$C_{n-} \simeq (3+2\sqrt{2})^n\left[\frac{(n-\frac{1}{2})!}{(-\frac{1}{2})!\,n!}\right]^2 + \frac{1}{(n-\frac{1}{2})(n+1)}(0.571)\left[1+<O\left(\frac{2}{n}\right)\right]. \quad (36)$$

This is the asymptotic value of the coefficient of of $(x^2)^n$ in the expansion of ln $\Lambda(1, x)$ [Eq. (12)]. The coefficients calculated from this first approximation are compared in Table II with the exact values of Eq. (11). The maximum possible error in the calculated coefficients is theoretically of the

TABLE II. Comparison between coefficients of the series expansion for ln $\Lambda(1, x)$ and the asymptotic values.

Order of coefficient	Actual coefficient	First approximation
7	130	110
8	490.25	423
9	1 956.66	1 784
10	8 174.2	7 640
11	35 302	33 600
12	156 777.5	151 700

order of $2/n$ times the coefficient, and we see that these lie well within these limits. Application of precisely the same method yields a contribution from the two unimportant singularities of the original integral [Eq. (12)], of the form

$$C_{n+} \simeq (3 - 2\sqrt{2})^n \left[\frac{(n - \frac{1}{2})!}{(-\frac{1}{2})!\, n!}\right]^2 \frac{1}{(n + 1)} \times \text{const}\left[1 + O\left(\frac{2}{n}\right)\right]. \quad (37)$$

In calculating higher-order coefficients numerically, such a contribution is completely insignificant, but nevertheless we assume that it exists.

IV. SIMILARITY BETWEEN THE SQUARE AND CUBIC LATTICE SERIES FOR THE PARTITION FUNCTION

A. High-Temperature Expansions in the Absence of a Magnetic Field

At high temperatures $x = \exp(-2J/kT) \to 1$, and the series expansions (2), (11) are of no use. However expansion of $\Lambda(1, x)$ with respect to different variables will lead to series valid above the critical point. Thus $\Lambda(1, x)$ may be expanded in terms of the variables $u = (1 - x)/(1 + x)$ or $t = (1 - x^2)$, to obtain series convergent at high temperatures. With a square lattice the high-temperature expansion follows immediately from the low-temperature expansion by noting that the variable $k_1^2 = 16x(1 - x^2)/(1 + x^2)^2$ of Eq. (12) is invariant to the transformation $u = (1-x)/(1+x)$. Thus, above the critical point,

$$\ln \Lambda(1, x) = \ln \tfrac{1}{2}(1 + x)^2 + U_2(u). \quad (38)$$

The series $U_2(u)$ is identical with Eq. (11), with u^2 replacing x^2 as the expansion parameter. For the cubic lattice, Wakefield[12] gives

$$\ln \Lambda(1, x) = \ln \tfrac{1}{4}(1 + x)^3 + U_3(u), \quad (39)$$

where $U_3(u)$ is the series

$$U_3(u) = 3u^4 + 22u^6 + (375/2)u^8 + 2004u^{10} + (48267/2)u^{12} + \cdots . \quad (40)$$

From the relation

$$(1/N) \ln \text{P.F.} = -\tfrac{1}{2}c \ln x + \ln \Lambda(1, x), \quad (41)$$

he derives

$$(1/N) \ln \text{P.F.} = \ln 2 + U_{3A}(u), \quad (42)$$

where

$$U_{3A}(u) = (3/2)u^2 + (15/4)u^4 + (45/2)u^6 + (1503/8)u^8 + (20\,043/10)u^{10} + (96\,537/4)u^{12} + \cdots . \quad (43)$$

Wakefield has carried out a careful graphical analysis of the series $U_{3A}(u)$ which behaves for higher terms as $U_3(u)$, and obtains for the coefficient a_n an asymptotic form

$$a_n \simeq \text{const}\ (20.9)^n/(2n - 1)(2n)(2n + 1). \quad (44)$$

His determination of the critical point of the cubic lattice is based entirely upon this approximation. The approximation (44) may be written in the now familiar form of the two-dimensional series for the high-temperature expansion of the partition function, Eq. (36). For if we write

$$a_n \simeq \text{const}\ (C_1)^n/(n - \tfrac{1}{2})(n + \tfrac{1}{2})n, \quad (45)$$

remembering that for large n

$$(n - \tfrac{1}{2})!/(-\tfrac{1}{2})!\, n! \simeq 1/(\pi n)^{\frac{1}{2}}, \quad (46)$$

it is tempting to conjecture that, by analogy with the two-dimensional series, the a_n's have the precise form

$$a_n \simeq (\text{const})\left[\frac{(n - \frac{1}{2})!}{(-\frac{1}{2})!\, n!}\right]^2 \frac{1}{(n - \frac{1}{2})(n + 1)} (C_1)^n \times \left(1 + \frac{C_2}{n - \frac{3}{2}} + \cdots\right). \quad (47)$$

In fact the form (46) does appear to describe the series $U_3(u)$ with somewhat greater accuracy than that of Wakefield.

Estimation of the Critical Point

We have fitted the coefficients of $U_3(u)$ to a form given by the first approximation of Eq. (46) by the reciprocal-difference formula of Thiele,[13] and find for the constant C_1 the value 20.6. This is to be compared with the value 20.9 of Wakefield. This value for the constant leads to a critical point at $u_c = 1/20.6$, i.e., $x_c = 0.639$. Surprisingly this is in precise agreement with the value determined by Domb[14] using a different approach.

B. Low-Temperature Expansions

For the square lattice, the asymptotic form of the coefficients of $\ln \Lambda(1, x)$ are given according to our analysis by Eqs. (36) and (37);

$$a_n \sim (3 + 2\sqrt{2})^n \left[\frac{(n - \frac{1}{2})!}{(-\frac{1}{2})!\, n!}\right]^2 \frac{C_1}{(n - \frac{1}{2})(n + 1)} \times \left[1 + O\left(\frac{1}{n}\right)\right]$$

[12] A. J. Wakefield, Proc. Cambridge Phil. Soc. **47**, 419, 799 (1951).

[13] L. M. Milne-Thomson, *The Calculus of Finite Difference* (MacMillan and Company, Ltd., London, 1951), 1st Ed., p. 104.

$$+ (3 - 2\sqrt{2})^n \left[\frac{(n - \frac{1}{2})!}{(-\frac{1}{2})!\, n!}\right]^2 \frac{C_2}{(n + 1)} \times \left[1 + O\left(\frac{1}{n}\right)\right]. \quad (48)$$

For the cubic lattice, Wakefield has derived the low-temperature expansion of $\ln \Lambda(1, x)$ as

$$\ln \Lambda(1, x) = Z(x) = x^6 + 3x^{10} - \tfrac{7}{2}x^{12} + 15x^{14} - 33x^{16} + (313/3)x^{18} - (561/2)x^{20} + 849x^{22} - (9847/4)x^{24} + 7485x^{26} - (45\,069/2)x^{28} + \cdots. \quad (49)$$

Again, resorting to graphical analysis, he establishes that the coefficients α_n of Eq. (49) can be best expressed in the form

$$\alpha_n = (-1)^{n-1}\gamma_n + \beta_n. \quad (50)$$

He finds a best fit, by trial and error, for this two-component series, which gives

$$\alpha_n \simeq -\frac{(-3.525)^n(\text{const})}{(n - \frac{1}{2})(n + \frac{1}{2})} + \frac{(2.434)^n(\text{const})}{(n - \frac{1}{2})n(n + \frac{1}{2})}. \quad (51)$$

If we recall Eqs. (46), (36), and (37), the asymptotic form of the corresponding two-dimensional series is

$$a_n \simeq \frac{(3 - 2\sqrt{2})^n(\text{const})}{n(n + 1)} + \frac{(3 + 2\sqrt{2})^n(\text{const})}{n(n - \frac{1}{2})(n + 1)}. \quad (52)$$

The similarity between the two cases seems at least remarkable, especially since Wakefield appears to obtain his results without any knowledge of the form of the two-dimensional expansion for $\ln \Lambda(1, x)$.

V. CONCLUSION

We make the remark that most numerical methods for the determination of critical phenomena from three-dimensional lattice series are first tested by comparison with approximate estimations carried out on two-dimensional lattices with the same method. The behavior of thermodynamic functions of the two-dimensional lattices in the neighborhood of the critical points, and the points themselves, are known exactly. However as mentioned in the introduction, it would seem that from a test of a given method using two-dimensional series, it is not really possible to draw any conclusions regarding the behavior of the three-dimensional series, unless in fact these do have the same form.

From his recent numerical analysis of Ising model series, Baher[3] draws the conclusion that the spontaneous magnetization for the single cubic, body-centered cubic, and face-centered cubic lattices behave near the critical point as

$$M \sim (x^2 - x_c^2)^{3/10}; \qquad x^2 = e^{-2J/kT}. \quad (53)$$

If his conclusions are correct, they most strongly support our conjectures. For the two-dimensional square lattice we see [Eq. (5)] that the spontaneous magnetization behaves near the critical point as $(x^2 - x_c^2)^{\frac{1}{8}}$, and that *this information alone is sufficient to determine the asymptotic form of the coefficients of the series expansion for M*.

We are aware that conjectures regarding the similarity in behavior of two- and three-dimensional Ising model series must be considered as tentative and very dangerous. However, keeping this in mind, the remarkable similarity which appears without any "forcing" of the answer between the square- and cubic-lattice series surely warrants further investigation. Such an investigation would necessarily require:

1. A proof that the three-dimensional partition functions fall into the class of functions discussed by Hamy and Darboux. It could then be shown that the coefficients of the series expansions for the three-dimensional model must be of Hamy–Darboux form. The critical point predictions of Domb could certainly then be regarded as highly accurate.

2. One might try to develop a systematic method for fitting coefficients to the assumed Hamy–Darboux form via reciprocal differences (appropriate to series of the assumed form).

3. A thorough investigation of all the known series to determine whether the coefficients can be fitted more satisfactorily with the Hamy–Darboux form than with techniques used so far. This is almost certainly so, but any conclusions without the precondition (1) are not really justified.

4. If these requirements can be satisfied, the question of predicting the functions from which the series arose might then be tackled. Knowledge of the exact asymptotic form for the two-dimensional series would provide a check on the reliability of such predictions. As mentioned in the introduction a modification of the theory of converging factors developed by Dingle[7] may show some light on this problem.

ACKNOWLEDGMENTS

The author is indebted to Dr. R. B. Dingle for drawing to his attention to both the problem, and the work of Darboux and Hamy, and for his constant advice and criticism. The helpful comments of Professor John Blatt, Dr. T. D. Schultz, and Dr. D. W. Jepson are gratefully acknowledged. The work was supported in most part by a C.S.I.R.O. postgraduate studentship held at the University of Western Australia during 1957.

Nonlinear Coupled Oscillators. II. Comparison of Theory with Computer Solutions

E. ATLEE JACKSON
Departments of Physics and Mechanical Engineering, University of Illinois, Urbana, Illinois
(Received 28 November 1962)

A study is made of a particular system of coupled oscillators to determine how the amount of energy exchange and the recurrence time depend on the parameters of the system, and the accuracy with which these are predicted by the perturbation theory developed in part I of this series. The system consists of $N + 1$ particles, connected by springs which have a quadratic force term, the strength of which, λ_i, can vary between different particles ("defects"). The equations of motion for the perfect chain ($\lambda_i = \lambda$) were solved on a computer for the cases $N = 4$, $\lambda = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$; and $N = 9$, $\lambda = \frac{1}{2}$; and supplemented by the previous calculations of Fermi, Pasta, and Ulam, in which $N = 32$, $\lambda = \frac{1}{4}$, 1. The "energy" in the linear modes, $E_k(t) = \frac{1}{2}(\dot{a}_k^2 + \omega_k^2 a_k^2)$, are determined from these computations and compared with the perturbation theory in the first paper of this series. As was found by Fermi, Pasta, and Ulam, when only E_1 is initially excited, then only the first few modes acquire an appreciable amount of energy, after which nearly all the energy returns to the first mode in a time τ_λ. The second-order perturbation theory is found to give an accurate estimate (within 15%) of both τ_λ and the amount of energy exchange for all the above cases, except $N = 32$, $\lambda = 1$, which requires higher-order analysis. It is shown that the nonergodic behavior of this system does not result simply from the incommensurability of the uncoupled frequencies $\{\omega_k\}$, but also from the particular form of mode interaction and the initial conditions used in all calculations, both of which affect the coupled frequency spectrum $\{\Omega_k\}$. To alter the mode interaction a preliminary study is made of "imperfect" chains (variable λ_i), which directly couples low and high modes. It is found that either a large, or else systematic, variation of λ_i is necessary to appreciably alter the nonergodic behavior of this system. A theoretical examination of the dependence of the coupled frequencies $\{\Omega_k\}$ on the initial conditions shows that the ergodic behavior will, in general, be strongly dependent upon the initial conditions.

I. INTRODUCTION

IN the first paper of this series[1] (referred to hereafter as NOI), a study was made of the problems which arise in perturbation theories developed in powers of the coupling constant λ. Alternative methods were examined in which the unknown coupled frequencies $\{\Omega_k(\lambda)\}$ were introduced *ab initio*. A perturbation scheme was then developed which determines the $\{\Omega_k\}$ in a way which appears to eliminate the appearance of small divisors. In the present study, this theory is applied to a particular system of coupled oscillators and compared with results obtained from computer solutions of the equations of motion. In addition, the theory will be used to examine the behavior of the system under conditions which are more general than those used for the computer calculations. The ultimate objective of such studies is to understand the role of coupling strengths, initial conditions, uncoupled frequencies $\{\omega_k\}$, and form of mode interaction on the behavior of this important class of systems.

In the present study we will examine the behavior of a one-dimensional chain of particles, connected by nonlinear springs. A numerical analysis of such systems has been made previously by Fermi, Pasta, and Ulam[2] (FPU), and, in order to make use of their results, we will restrict our study to one of the nonlinear interactions used by them (cubic terms in the Hamiltonian). Using a chain of 32 particles, they studied the case when, at $t = 0$, only the lowest linear mode was excited. Contrary to their expectations, the energy did not become distributed among a large number of modes, but only the first few modes became excited before the original situation was closely reproduced (corresponding to a Poincaré cycle). Their results, using the system of Eqs. (3), with $\lambda_i = \lambda$, are shown in Figs. 1 and 2. Moreover, they found that the qualitative features of these results were largely independent of the type of nonlinear coupling (e.g., cubic, quartic, nonanalytic). FPU did not offer any explanation for this lack of ergodicity in the coupled systems, but recently Ford[3] proposed an explanation of this phenomena. Ford noted that weakly coupled oscillators, whose uncoupled frequencies are ω_k, will exchange energy periodically with a frequency

$$\omega(m) = \sum_k^N m_k \omega_k, \tag{1}$$

[1] E. A. Jackson, J. Math. Phys. **4**, 551 (1963).

[2] E. Fermi, J. Pasta, and S. Ulam, "Studies of Nonlinear Problems I," Los Alamos Scientific Laboratory Report LA-1940 (1955).

[3] J. Ford, J. Math. Phys. **2**, 387 (1961).

where the appropriate integers $\{m_k\}$ depend on the form of the nonlinear interaction and N is the number of particles in the chain. Since none of the frequencies $\omega(m)$ vanish identically in the FPU case, Ford concluded that the system should not be ergodic. However, the frequency $\omega(m)$ is only significant in the case of sufficiently weak coupling (as discussed in NOI). As will be shown in Sec. III, the cases studied by FPU do not belong in the category of weak coupling, so, while Ford's first observation is correct, it is not entirely relevant to the calculations of FPU or to any case where N is sufficiently large. One of the objectives of the present study is to survey the behavior of such systems over a wider range of coupling constants and frequencies $\{\omega_k\}$ by varying N. In this way we will cover the cases of both weak coupling (where Ford's analysis is relevant) and strong coupling (to test the modifications predicted by the present theory). In the latter case we will be primarily interested in comparing the "recurrence time" τ_λ (defined as the time required for the near recurrence of the initial conditions) with that predicted from

$$\Omega(m) = \sum_k^N m_k \Omega_k(\lambda) \tag{2}$$

[i.e., $\tau_\lambda = 2\pi/\Omega(m)$], and to see how it varies as λ increases. If τ_λ increases when λ increases, this would mean that the energy exchange would continue for a longer period of time when the coupling is stronger—a feature which is presumably necessary, even if not sufficient, for ergodic behavior. Actually, what is found to occur for the particular initial condition discussed above, is that τ_λ decreases as λ increases (Sec. IV). This behavior of τ_λ will be shown to be dependent on the initial condition, and therefore is not necessarily a general feature of such systems (Sec. VI).

Another important property of the perfect chain is that energy is only transmitted to the higher modes once the intermediary modes have become excited. In the case when only the lowest mode is initially excited, this property is obviously a dominant factor in how effectively the energy becomes distributed among the modes. To remove this property we consider the effect of a variable coupling constant, λ_i, in the nonlinear force acting between different particles. Such a "defective" chain directly couples various modes of the system, depending on the form of λ_i. A preliminary study of the effectiveness of this coupling is given in Sec. V.

The interesting advantages and difficulties of using the equations of motion of the action-angle variables for such studies is illustrated in Appendix B.

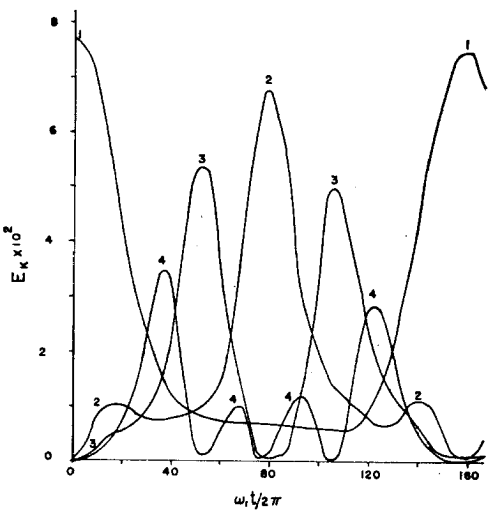


FIG. 1. The energy, E_k, in the first four modes vs periods of basic mode, as computed by Fermi, Pasta, and Ulam. $A_k = 4\delta_{k1}$, $N = 32$, $\lambda = \frac{1}{4}$, $E_5 < 0.015$, $E_k < 5 \times 10^{-3}$ for $k > 5$.

II. LINEAR CHAIN OF COUPLED OSCILLATORS

The theory of NOI will now be applied to a particular class of systems which can be studied in some detail. We consider a linear chain of $N + 1$ particles, all of the same mass, m, coupled by nonlinear springs, with the first and last particle held fixed. If the time is measured in units of $(m/\mu)^{\frac{1}{2}}$, where μ is the linear spring constant, then the equation of motion for the displacement, x_i, of the ith particle from its position of equilibrium is

$$\ddot{x}_i = (x_{i+1} + x_{i-1} - 2x_i) + \lambda_i(x_{i+1} - x_i)^2 - \lambda_{i-1}(x_i - x_{i-1})^2 \quad (i = 1, \cdots N - 1), \tag{3a}$$

$$x_0(t) = x_N(t) = 0. \tag{3b}$$

The quantity $\mu\lambda_i$ is the quadratic spring constant

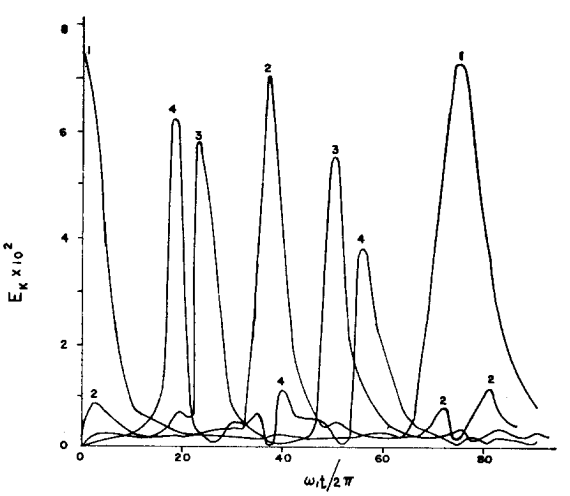


FIG. 2. Same condition as in Fig. 1, except $\lambda = 1$.

for the spring between particle i and $(i + 1)$, and allows for the possibility of "defects" in the nonlinear forces. The Hamiltonian of this system is

$$H = \tfrac{1}{2}\sum_{i=0}^{N-1} \dot{x}_i^2 + (x_{i+1} - x_i)^2 + \tfrac{1}{3}\sum_{i=0}^{N-1} \lambda_i (x_{i+1} - x_i)^3. \quad (4)$$

Before proceeding to analyze this system, we may note that there are two possible limiting processes involving $N \to \infty$. The finite string is obtained by setting μ, m^{-1}, and λ all proportional to N, in which case the frequency spectrum [$(\mu/m)^{\frac{1}{2}}$ times ω_k of (7)] becomes discrete. This case has recently been studied in detail by Zabusky.[4] The other possible limit (with μ, m, and λ constant) yields the infinite string, in which case the frequency spectrum becomes dense. It is only the latter system which is usually used in the study of irreversible processes, and this is the limit which we will be concerned with in the present study.

We now define the normal modes, $a_k(t)$ by

$$a_k(t) = \left(\frac{2}{N}\right)^{\frac{1}{2}} \sum_{i=0}^{N} x_i \sin\frac{ik\pi}{N} \quad (k = 0, \cdots, N), \quad (5a)$$

$$a_k = O(0 \geq k \geq N).$$

We have extended the definition of a_k for $k > N$ and $k < 0$ in order to simplify the range of the summations below. Thus, in all sums involving a_k, we shall leave the range of summation unspecified, always meaning by this, $-\infty \leq k \leq \infty$, unless otherwise specified. Using the relationship

$$\sum_{i=0}^{N} \sin\frac{ik\pi}{N}\sin\frac{il\pi}{N} = \tfrac{1}{2}N\delta_{k-l},$$

Eq. (5a) can be inverted to yield

$$x_i = \left(\frac{2}{N}\right)^{\frac{1}{2}} \sum_k a_k \sin\frac{ik\pi}{N}, \quad (5b)$$

which automatically satisfies (3b). Substituting (5b) into (44), one can put the transformed Hamiltonian in the form

$$H \equiv H_0 + \lambda H_1$$

$$= \tfrac{1}{2}\sum_k \dot{a}_k^2 + \omega_k^2 a_k^2 + \frac{\lambda}{3}\sum_{jkl} V_{jkl} a_j a_k a_l, \quad (6)$$

where $\omega_k = 2 \sin (k\pi/2N)$, and V_{jkl} is symmetric in the indices, so that the general equations of motion are

$$\ddot{a}_j = -\omega_j^2 a_j - \sum_{kl} V_{jkl} a_k a_l. \quad (7)$$

[4] N. J. Zabusky, J. Math. Phys. 3, 1028 (1962).

The coefficients of V_{jkl} depend, of course, on the quantities λ_i. In the case of the perfect chain ($\lambda_i = \lambda$) one can show that (see Appendix A)

$$V_{jkl} = \gamma_{jk}\delta_{j+k-l} - \tfrac{1}{3}\gamma_{jk}\delta_{j+k+l-2N} + (l \leftrightarrow k) + (j \leftrightarrow l), \quad (8)$$

where

$$\gamma_{jk} = (2N)^{-\frac{1}{2}}[\omega_{2j} + \omega_{2k} - \omega_{2(j+k)}] \quad (9)$$

(note that $\gamma_{jk} \equiv \gamma_{j,2N-j-k}$), and δ_K is the usual Kronecker function (more commonly written δ_{K0}). The brackets in (8) represent the repetition of the first two terms except for the indicated interchange of indices. Since the perfect chain is one of the most important cases, we shall give the equations of motion in detail:

$$\ddot{a}_j = -\omega_j^2 a_j - 2\lambda \sum_k \gamma_{jk} a_k a_{j+k} - \lambda \sum_k \gamma_{k,j-k} a_k a_{j-k} + \lambda \sum_k \gamma_{jk} a_k a_{2N-j-k}. \quad (10)$$

Aside from some simplifications and notational changes, these are the same equations used by Ford. The last group of terms in (10), the so-called Umklapp terms, do not play a significant role in the dynamics of the chain if only the low modes are excited.

To obtain more coupling between the modes it is necessary to allow for defects in the nonlinear forces. Such defects, hopefully, represent impurities or some other imperfections which produce additional scattering of the linear modes. A particularly simple defect is $\lambda_i = \delta_{i-i}\lambda$, for some specific i; in other words, only one nonlinear spring in the system. One easily finds that, in this case,

$$V_{jkl} = 3\left(\frac{2}{N}\right)^{\frac{1}{2}}\left[\sin\left(\frac{j(i+1)\pi}{N}\right) - \sin\frac{ji\pi}{N}\right] \times [j \leftrightarrow k]\cdot[j \leftrightarrow l],$$

where the last two terms are the same as the first except for the indicated interchange. One can see that any mode j, for which ij is nearly a multiple of N, is only weakly coupled with any other mode. Thus, to have appreciable scattering, over and above the perfect coupling, it is necessary to make the variations in λ_i more systematic. As an example, we shall consider

$$\lambda_i = \lambda \sum_{m=0}^{N-1} \sigma_m \cos\frac{im\pi}{N}, \quad (11)$$

where $\sigma_0 = 1$. To determine V_{jkl} now requires some analysis, which is done in Appendix A. The effects of these imperfections will be studied in Sec. V.

III. SPECIAL SOLUTIONS

In this section we will analyze the behavior of this system, and for simplicity, consider only those cases when the chain is initially at rest;

$$a_j(0) = A_j, \qquad \dot{a}_j(0) = 0. \tag{12}$$

The analysis will be based on the perturbation method developed in NOI, given by

$$a_j = \alpha_j \cos \Omega_j t - \lambda \int_{-\infty}^{\infty} G_j(t, \tau) \times \left[\frac{\partial H_1(\tau)}{\partial a_j} - \sum_{n=1}^{\infty} \lambda^{n-1} \mu_j^{(n)} a_j(\tau) \right] d\tau, \tag{13}$$

and

$$\int_{-\infty}^{\infty} \cos \Omega_j t \left[\frac{\partial H_1}{\partial a_j} - \sum \lambda^{n-1} \mu_j^{(n)} a_j(t) \right] dt = 0, \tag{14}$$

where

$$G_j(t, \tau) = \sum_{\{m\}}{}' \frac{\cos \Omega(m) t \cos \Omega(m) \tau}{\Omega_j^2 - \Omega(m)^2}. \tag{15}$$

The prime signifies that the terms $\Omega(m) = \pm\Omega_j$ are excluded from the sum. The corresponding equations in NOI have been simplified to obtain Eqs. (12–15) by using the second condition in Eq. (12) to drop all phase factors. The method of analysis consists simply in iterating Eqs. (13) and (14) in the explicit powers of λ.

As was discussed in NOI, the first iteration yields

$$a_j(t) = \alpha_j \cos \Omega_j t - \tfrac{1}{2}\lambda \sum_{kl} V_{jkl} \alpha_k \alpha_l \mathcal{P} \times [\Omega_j^2 - (\Omega_k \pm \Omega_l)^2] \cos (\Omega_k \pm \Omega_l) t, \tag{16}$$

where $\mathcal{P}$ stands for the sum of all permutations in the signs ($\pm$). Because H_1 [Eq. (6)] is cubic, $\mu_j^{(1)} = 0$, so that $\Omega_j = \omega_j$ in this order.

Before proceeding to higher orders it is instructive to consider some of the features of Eq. (16). Following FPU, it is useful to consider the "energy of the jth mode,"[5] defined by

$$E_j = \tfrac{1}{2}(\dot{a}_j^2 + \omega_j^2 a_j^2). \tag{17}$$

FPU found that if only A_1 is initially nonzero (so that initially only E_1 is nonzero), then only the first few E_j's become appreciable in magnitude before E_1 returns to essentially its initial value (Figs. 1 and 2). Ford's explanation of this nonergodic behavior is based on an analysis which is apparently equivalent to Eq. (16), with $\Omega_j = \omega_j$. As Ford noted, all of the relevant frequencies $\omega(m)$ which appear in Eq. (16) are nearly integer multiples of $2\omega_1 - \omega_2$ for sufficiently large N. Hence, the variations of E_j are predicted, on the basis of this theory, to recur periodically with a period

$$\tau_0 = 2\pi/(2\omega_1 - \omega_2). \tag{18}$$

In the study made by FPU, N and λ are sufficiently large ($N = 32$, $\lambda = \frac{1}{4}; 1$) for τ_0 to be a poor estimate of the recurrence time. Specifically, one finds

$$\tau_0 = 2.5 \times 10^4, \quad \tau_{\frac{1}{4}} = 10^4, \quad \tau_1 = 5 \times 10^3, \tag{19}$$

where the first is obtained from Eq. (18) when $N = 32$, and the remaining two from Figs. 1 and 2 ($\lambda = \frac{1}{4}; 1$, respectively). The subscript on τ_0 signifies the theoretical recurrence time as $\lambda \to 0$, N fixed (as discussed in NOI). One can see that τ_0 differs by 150%–400% from the recurrence times found by FPU, so the FPU cases are not in the limit of weak coupling. For this reason, this low-order theory does not give an adequate explanation of the FPU results. In particular, it does not predict whether τ_λ increases or decreases as λ is increased from zero. Presumably the exchange of energy between various modes will improve, and hence the system will become more "ergodic," if τ_λ increases with λ. Thus it is not possible to decide how well the linear modes will exchange energy if λ is "large" (in the sense $\lambda > \min \{\omega(m)\}$, as discussed in NOI), using only the present results. It should be emphasized that this is not necessarily just a "qualitative" question for if all $\omega(m) = 0$ the recurrence time is entirely determined by higher orders (i.e., the "qualitative" difference is between an infinite and a finite period). Also, as can be seen from Ford's results, the predicted behavior of $E_2(t)$ and $E_3(t)$ does not even agree qualitatively with the FPU results (e.g., E_3 is predicted to have three maxima within a time τ_λ, whereas it has only two maxima in Fig. 1). In the following section we shall consider systems with various values of N and λ, some of which are small enough for this lowest-order theory to be applicable, and others (including the FPU cases) which require higher orders. In following sections, we shall attempt to use the insight gained from the analysis of these cases to study more general initial conditions.

We now consider the second-order theory. Using Eq. (16) one finds from Eq. (14) that

$$\mu_j^{(2)} = -\sum_{kl} \tfrac{1}{2}\alpha_k^2 (2 - \delta_{j-k}) \{ V_{jjl} V_{lkk} \Omega_l^{-2} + V_{jkl}^2 \mathcal{P} [\Omega_l^2 - (\Omega_j \pm \Omega_k)^2]^{-1} \}, \tag{20}$$

[5] Since we shall be largely concerned only with the behavior of the energies E_j, one might expect that the use of action-angle variables are most natural. Unfortunately, the equations of motion for the angle variables present certain formal difficulties, discussed in Appendix B.

and $\Omega_j^2 = \omega_j^2 + \lambda^2\mu_j^{(2)}$ in this order. The additive correction to Eq. (16) is obtained from Eq. (13) and is given by

$$a_j^{(2)} = \tfrac{1}{2}\lambda^2 \sum_{klmn} V_{jkl}V_{lmn}\alpha_k\alpha_m\alpha_n \times \mathcal{P}'[\Omega_l^2 - (\Omega_m \pm \Omega_n)^2]^{-1}[\Omega_j^2 - (\Omega_m \pm \Omega_n \pm \Omega_k)^2]^{-1} \cos(\Omega_m \pm \Omega_n \pm \Omega_k)t, \quad (21)$$

where the prime excludes those terms for which $(\Omega_m \pm \Omega_n \pm \Omega_k)^2 \equiv \Omega_j^2$.

While Eqs. (16), (20), and (21), together with the initial conditions of Eq. (12), complete the analytical analysis, there remains a considerable amount of numerical analysis in order to compare this theory with the computer solutions of the equations of motion. The task which remains is to determine the (α_j, Ω_j) from the initial conditions and the equations $\Omega_j^2 = \omega_j^2 + \lambda^2\mu_j^{(2)}$. Without reverting to further computer calculations, this requires the use of certain approximations to the theory, to be discussed shortly. The analysis is considerably simpler in the case of a perfect chain, characterized by Eq. (8), because of the limited number of terms in the series of Eqs. (16), (20), and (21). We shall therefore consider this case first, and examine the imperfect chain in Sec. V.

IV. THE PERFECT CHAIN

In this section we will study the behavior of perfect chains subject to the initial conditions

$$x_i(0) = \sin(i\pi/N),$$

which is equivalent to

$$A_j = \tfrac{1}{2}N\delta_{j-1} \quad (22)$$

in Eq. (12). This also appears to be the initial conditions used by FPU,[6] and from their results we know that only the first few modes will be excited. As a matter of practical expediency we will therefore examine the behavior of only the first three modes and take $\alpha_j = 0$ for $j \geq 4$ (except in certain cases). This approximation is not particularly good in the FPU case, $\lambda = \frac{1}{4}$, and fails entirely when $\lambda = 1$, as can be judged from Figs. (1) and (2). Nonetheless, it turns out to be a sufficiently good approximation for our purposes. (Actually it is of some interest to see how well such approximations of the theory agree with the true behavior of the system.)

A second simplification results from the fact that the frequencies $\Omega(m)$ in the denominators of Eqs. (16) and (21) are either "large" or "small" and hence, if one is interested only in the average behavior of $a_j(t)$, it is only necessary to keep the terms with small divisors. Having made these simplifications, one finds from Eqs. (16) and (21),

$$a_1 = \alpha_1 \cos\Omega_1 t - \lambda V_{112}\alpha_1\alpha_2 \frac{\cos(\Omega_2 - \Omega_1)t}{\Omega_1^2 - (\Omega_2 - \Omega_1)^2} - \lambda V_{123}\alpha_2\alpha_3 \frac{\cos(\Omega_3 - \Omega_2)t}{\Omega_1^2 - (\Omega_3 - \Omega_2)^2},$$

$$a_2 = \alpha_2 \cos\Omega_2 t - \tfrac{1}{2}\lambda V_{112}\alpha_1^2 \frac{\cos 2\Omega_1 t}{\Omega_2^2 - (2\Omega_1)^2} - \lambda V_{123}\alpha_1\alpha_3 \frac{\cos(\Omega_3 - \Omega_1)t}{\Omega_2^2 - (\Omega_3 - \Omega_1)^2},$$

$$a_3 = \alpha_3 \cos\Omega_3 t - \lambda V_{123}\alpha_1\alpha_2 \frac{\cos(\Omega_2 + \Omega_1)t}{\Omega_3^2 - (\Omega_2 + \Omega_1)^2} + \tfrac{1}{2}\lambda^2 V_{112}V_{123}\alpha_1^3[\Omega_2^2 - (2\Omega_1)^2]^{-1} \times [\Omega_3^2 - (3\Omega_1)^2]^{-1}\cos(3\Omega_1 t) + \lambda^2\alpha_1\alpha_2^2\{[\Omega_1^2 - (\Omega_2 - \Omega_1)^2]^{-1} \times [\Omega_3^2 - (2\Omega_2 - \Omega_1)^2]^{-1}V_{112}V_{123} - [(2\Omega_2)^2 - \Omega_4^2]^{-1}[\Omega_3^2 - (2\Omega_2 - \Omega_1)^2]^{-1}V_{134}V_{224}\}\cos(2\Omega_2 - \Omega_1)t. \quad (23)$$

We have not included any of the terms from Eq. (21) in a_1 or a_2 because of their relative magnitude (another considerable simplification).

We shall apply this relatively simple theory to a number of systems with different values of N and λ. In this way we will not only be able to check the theory over the range of weak to strong coupling, but also see what terms in the theory are necessary to explain the behavior of the system.

Case $N = 4$

In this case the third mode is never appreciably excited for any value of λ (the energy surface is only bounded if $\lambda \lesssim 0.85$ with the present initial conditions). We therefore take $\alpha_3 = 0$ in Eq. (23) and find that the energies, Eq. (17), are given by

$$\langle E_1(t)\rangle = \tfrac{1}{2}\Omega_1^2 A_1^2 - \langle E_2(t)\rangle,$$

$$\langle E_2(t)\rangle = \frac{\lambda^2\alpha_1^4 V_{112}^2}{8\Omega_{11}^2}\sin^2\tfrac{1}{2}\Omega_{11}t, \quad (24)$$

where $\Omega_{ij} = \Omega_i + \Omega_j - \Omega_{i+j}$ and we have taken $\Omega_j \simeq j\Omega_1 \simeq j\omega_1$ in all the frequencies $\Omega(m)$ which are

[6] They did not specify their initial conditions analytically, only verbally. Equation (22) differs from Ford's conclusion that $A_j = 250\,\delta_{j-1}$, but is believed to be correct.

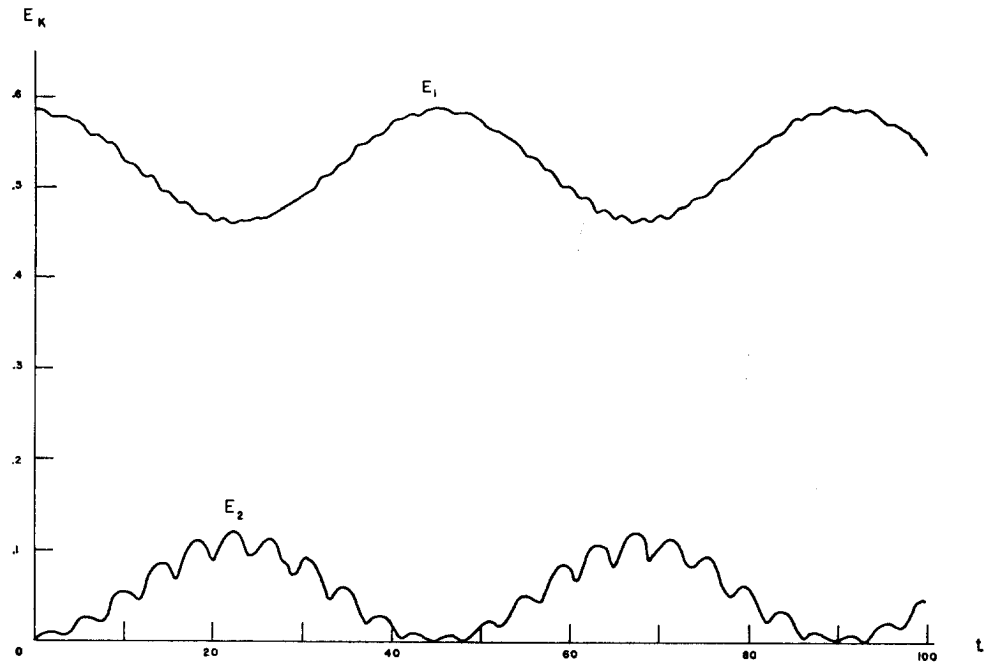


FIG. 3. Energy in the first two modes for the case $N = 4$, $A_k = \sqrt{2}\,\delta_{k1}$, $\lambda = \frac{1}{4}$, $E_3(t) \leq 0.01$ $(2\pi/\omega_1 = 8.2)$.

not small. The brackets $\langle\ \rangle$ indicate that these are the values of E_i average over a short time $(\sim 2\pi/\Omega_1)$. The constant α_1 is determined by the initial conditions, Eq. (22) and Eq. (23). One finds

$$(\tfrac{1}{2}N)^{\frac{1}{2}} = \alpha_1 + \frac{1}{4}\left(\frac{\lambda}{\Omega_{11}}\right)^2 \frac{V_{112}^2}{\Omega_2^2}\,\alpha_1^3. \tag{25}$$

Once Ω_{11} is known, Eq. (25) gives α_1, which in turn determines the magnitude of the energy exchange between modes 1 and 2 [Eq. (24)]. The equations determining $\Omega_i \simeq \omega_i + (\lambda^2/2\omega_i)\mu_i^{(2)}$ also contain α_1 so that these should be solved together with Eq. (25). Fortunately, Ω_i is less sensitive to the value of α_1 than vice versa so we shall first solve

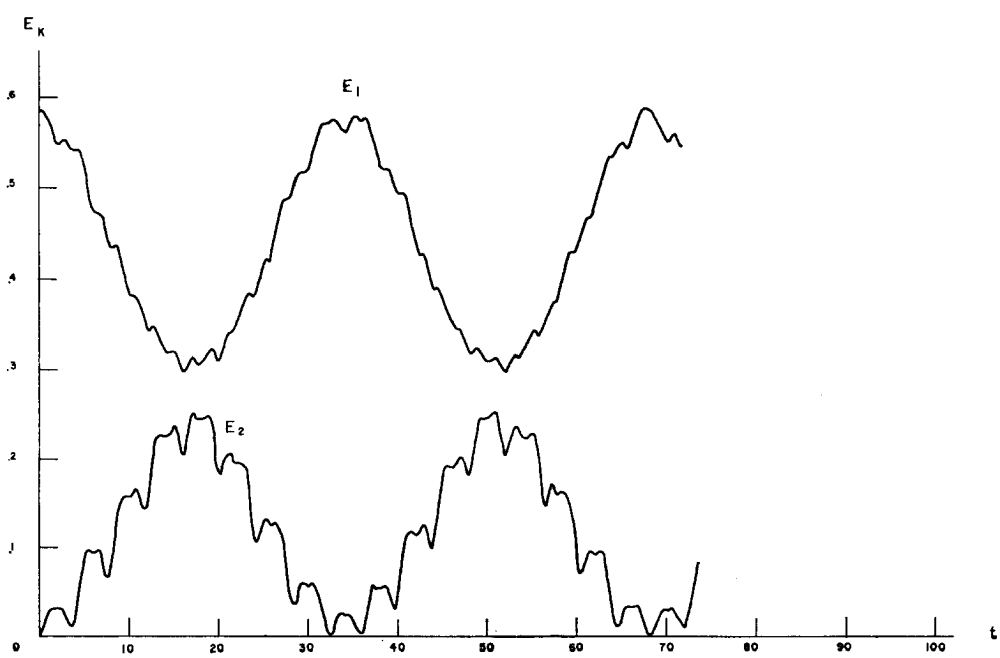


FIG. 4. Energy in the first two modes for the case $N = 4$, $A_k = \sqrt{2}\,\delta_{k1}$, $\lambda = \frac{1}{2}$, $E_3(t) \leq 0.04$ $(2\pi/\omega_1 = 8.2)$.

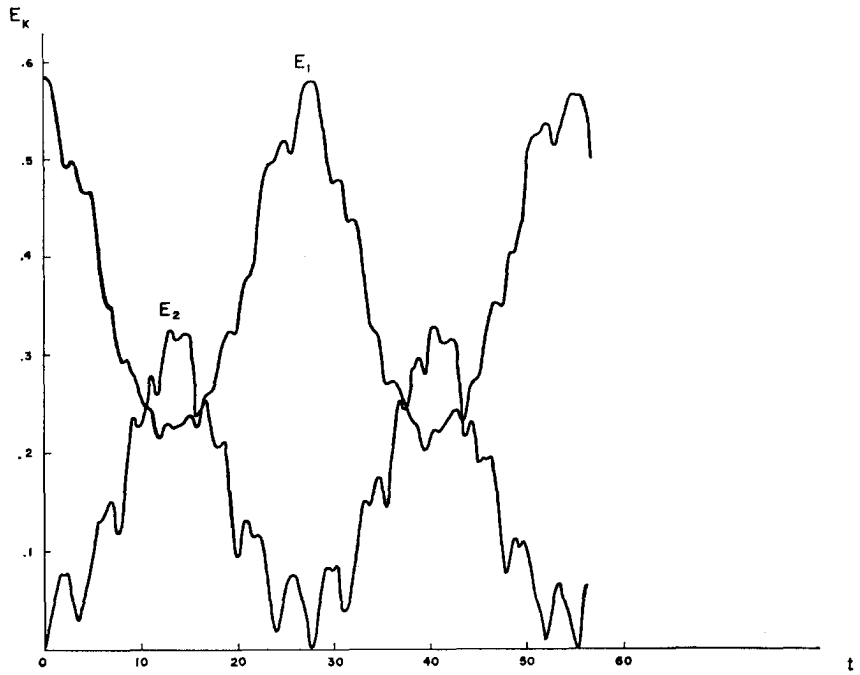


FIG. 5. Energy in the first two modes for the case $N = 4$, $A_k = \sqrt{2}\ \delta_{k1}$, $\lambda = \frac{3}{4}$, $E_3 \leq 0.1$ $(2\pi/\omega_1 = 8.2)$.

for Ω_{ij}, taking $\alpha_1 = (N/2)^{\frac{1}{2}}$, and then correct α_1 using Eq. (25). This approximation could obviously be improved, but it turns out to be unnecessary. Using Eq. (20) and the values of V_{ikl}, ω_i, and $\alpha_1 = (N/2)^{\frac{1}{2}}$, one obtains the following equations for Ω_{ij}:

$$\Omega_{11} = 0.1166 + 0.0301\lambda^2 + \frac{0.079\lambda^2}{\Omega_{11}} - \frac{0.096\lambda^2}{\Omega_{12}},$$

$$\Omega_{12} = 0.3318 - 0.380\lambda^2 - \frac{0.020\lambda^2}{\Omega_{11}} + \frac{0.1916\lambda^2}{\Omega_{12}}. \quad (26)$$

For the present case, the equations of motion (3a) were solved on a computer for $\lambda = \frac{1}{4}, \frac{1}{2}$, and $\frac{3}{4}$. The results are shown in Figs. (3–5). The two main properties of E_1 and E_2 are the recurrence time τ_λ and the maximum amount of energy exchange (i.e., $\langle E_2 \rangle_{\max.}$). The oscillations which are observed on $E_2(t)$ in these figures comes from the fact that N is small. In that case the "small" and "large" divisors in Eq. (16) only differ by a factor of about 10, so their neglection in Eq. (23) is only justified on the average. In Table I we compare the recurrence time $\tau_\lambda = 2\pi/\Omega_{11}$ and $\langle E_2 \rangle_{\max.}$ as obtained from Eqs. (26) and (24), with the results shown in Figs. 3–5.

TABLE I. Comparison of theory with computed recurrence times and maximum energy exchange.

λ	τ_λ	Theoretical τ_λ	$\langle E_2 \rangle_{\max.}$	Theoretical $\langle E_2 \rangle_{\max.}$
0	...	54.1	...	0
1/4	45.0	45.8	0.11	0.113
1/2	34.0	36.7	0.23	0.220
3/4	27.5	30.8	0.30	0.285

The accuracy of the present theory can be seen to be very good even for $\lambda = \frac{3}{4}$. For $\lambda \geq 0.85$, the energy surface is unbounded, so $\lambda = \frac{3}{4}$ represents a large distortion [the case $\lambda = 1$ was run and $E_1(t)$ indeed becomes unbounded shortly after one recurrence time]. The error in the theoretical values of τ_λ are in the direction one would expect from the numerical approximations used in determining Ω_{11} [i.e., $\alpha_1 = (\frac{1}{2}N)^{\frac{1}{2}}$].

While the present case is of limited interest, because N is small, it illustrates certain features which are common to systems with large N. The incommensurability of the frequencies Ω_i (in particular, Ω_{11}) is found to be enhanced by the coupling. This can be pictured as having two effects on the process of energy exchange between modes. First of all, the recurrence time τ_λ is decreased as λ increases, so that the time during which the second mode could transmit energy to the third is decreased. Moreover, from Eqs. (24) and (25), one sees that if Ω_{11} increases with λ, then $\langle E_2 \rangle_{\max.}$ does not increase as λ^2, so that the *amount* of energy exchanged is reduced. Both of these effects reduce the possibility of exchanging energy with the third mode, and thus make the coupled system very nonergodic. Of course, since

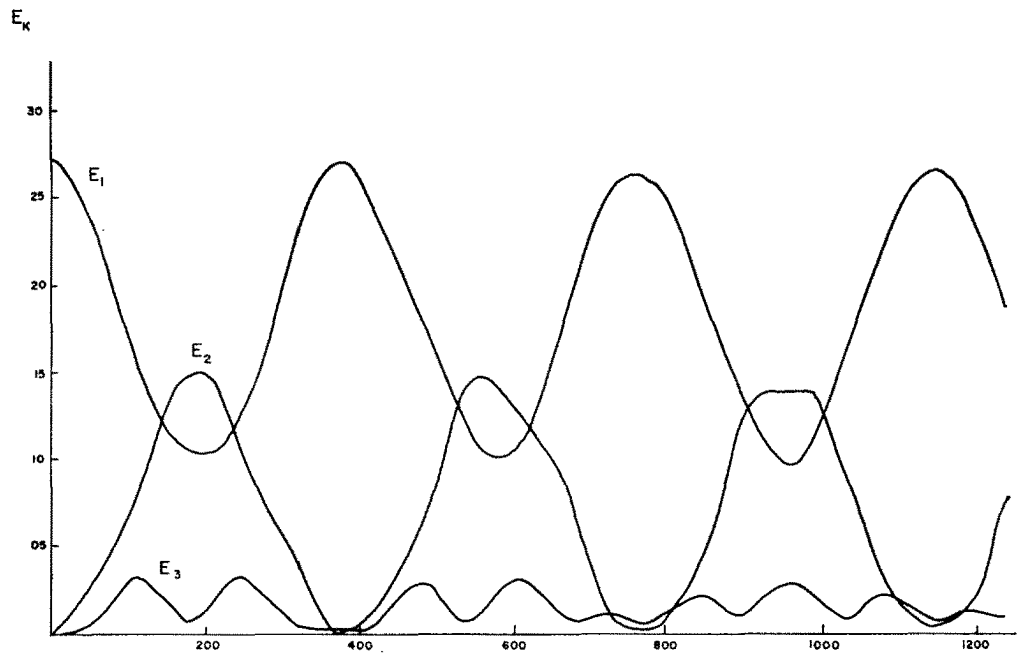


FIG. 6. Energy in modes for the case $N = 9$, $\lambda = \frac{1}{4}$.

N is small, the frequency $2\omega_1 - \omega_2$ is not very small, so, on the basis of Ford's work, one would not expect an ergodic behavior. The point of importance here is that, not only is Ford's observation correct, but that the incommensurability becomes even more pronounced as λ becomes large.

Case $N = 9$

The case $N = 9$ was run on a computer for $\lambda = \frac{1}{4}$, the results of which are shown in Fig. 6. The frequencies are sufficiently commensurable in this case (because N is larger) for $E_k(t)$ to be much more regular. Thus the expressions in Eq. (23) are not only accurate in the time-averaged sense, but also at each instant. To analyze this case we shall still approximate E_1 and E_2 by Eq. (24), but now examine the behavior of $E_3(t)$ as predicted by Eq. (23). In the present case we will neglect the last term of a_3 because $\alpha_2 \simeq 0.4 \ll \alpha_1$ and the denominators of this term are relatively large compared to the other second-order term. One then finds from Eq. (23) the approximate result:

$$\begin{aligned} E_3(t) = {} & [32\Omega_2^2]^{-1}\lambda^4 V_{112}^2 V_{123}^2 \alpha_1^6 \Omega_{11}^{-2}[\Omega_{12}^{-2}\sin^2 \tfrac{1}{2}\Omega_{12}t \\ & + (\Omega_{11} + \Omega_{12})^{-2}\sin^2 \tfrac{1}{2}(\Omega_{11} + \Omega_{12})t \\ & - 2\Omega_{12}^{-1}(\Omega_{11} + \Omega_{12})^{-1}\sin \tfrac{1}{2}\Omega_{12}t \\ & \times \sin \tfrac{1}{2}(\Omega_{11} + \Omega_{12})t \cos \tfrac{1}{2}\Omega_{11}t]. \end{aligned} \tag{27}$$

The frequencies $(\Omega_{11}, \Omega_{12})$ which are needed are determined by Eq. (20), which can now be approximated by

$$\Omega_j \simeq \omega_j - \frac{\lambda^2}{4\omega_j}\sum_{k,l}\alpha_k^2(2 - \delta_{j-k}) \times V_{jkl}^2 \mathcal{P}[\Omega_l^2 - (\Omega_j \pm \Omega_k)^2]^{-1} \tag{28}$$

and all large divisors can now be ignored. To actually solve these equations we shall again make the rather crude approximation $\alpha_k = (\frac{1}{2}N)^{\frac{1}{2}}\delta_{k-1}$, and then determine $(\Omega_{11}, \Omega_{12}, \Omega_{13}, \cdots)$ from the resulting coupled equations. When this is done one finds

$$\Omega_{11} \simeq 0.0157, \quad \Omega_{12} \simeq 0.0362, \quad \Omega_{13} \simeq 0.0655. \tag{29}$$

Using this result one obtains a corrected value of α_1 from Eq. (25), which is $\alpha_1 = 1.88$. Substituting these results into Eqs. (27) and (24), one obtains the predicted behavior of E_1, E_2, and E_3 as shown in Fig. 7. The theoretical recurrence time of 400 compares very favorably with the actual value of 380. The maximum of E_2 varies between 0.15 and 0.14, compared with the theoretical value of 0.15. This variation in $(E_2)_{\max.}$ probably results from the third term of a_2 in Eq. (23), which has been ignored. The asymmetry in the theoretical curve for E_3 comes from the fact that $\Omega_{12} \neq 2\Omega_{11}$ in Eq. (29). On the other hand, this inequality is what causes the "tail" on E_3 (near $t = 400$), which can be seen in Fig. 6. It should be noted also that if one replaces Ω_{ij} by $\omega_{ij} = \omega_i + \omega_j - \omega_{i+j}$, then, since $\omega_{12} \simeq 3\omega_{11}$, one would expect E_3 to have three maxima, as was found by Ford. The fact that there are only two

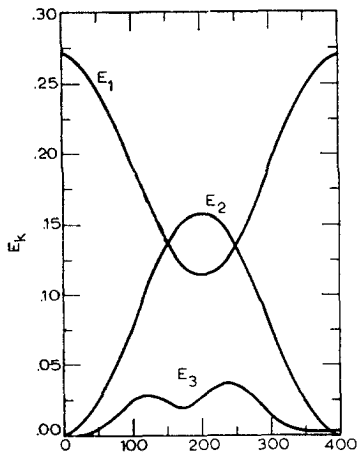


FIG. 7. Theoretical prediction of energy in modes when $N = 9$, $\lambda = \frac{1}{4}$.

maxima is due to the shift in the frequencies which make $\Omega_{12} \simeq 2\Omega_{11}$. Thus a qualitative, as well as quantitative difference in the behavior of the energies results from the shift in frequencies.

From these results it is clear that the second-order theory is quite adequate to explain the behavior of a "reasonable-sized" system. In passing it should be noted that the study of systems for which N is from 5 to 10 should be sufficient to demonstrate all the essential properties of large systems.

Case $N = 32$

As N becomes larger, for a fixed value of λ, the interaction between the normal modes becomes stronger. The cases calculated by FPU, which will be studied in this section, have sufficient interaction between the modes to require higher-order analysis in order to obtain a thoroughly accurate description of their calculations. This is particularly true of the case $\lambda = 1$. We shall be contented here to use only the second-order theory and the numerical approximations developed above because, rather remarkably, they turn out to be reasonably good when $\lambda = \frac{1}{4}$. In the case $\lambda = 1$, no attempt will be made to reproduce the results shown in Fig. 2, except for an estimate of the recurrence time.

Using the approximation $\alpha_j = (\frac{1}{2}N)^{\frac{1}{2}}\,\delta_{j-1}$, one obtains the following approximate solution of Eq. (20) when $\lambda = \frac{1}{4}$:

$$\begin{aligned} \Omega_{11} &\simeq 6.87 \times 10^{-4}, \\ \Omega_{12} &\simeq 1.31 \times 10^{-3}, \\ \Omega_{13} &\simeq 2 \times 10^{-3}. \end{aligned} \tag{30}$$

This indicates a recurrence time of $2\pi/\Omega_{11} = 0.91 \times 10^4$, compared with FPU's value of 10^4. This agreement is better than might be expected considering the above approximation. Using the values (30) in (23), the initial conditions (22) yield the corrected values of α_j:

$$\alpha_1 \simeq 2.90, \qquad \alpha_2 \simeq -0.77, \qquad \alpha_3 \simeq 0.22. \tag{31}$$

These should be understood to be only approximate solutions (accurate to about 2%, 10%, and 30%, respectively). To obtain these values it is necessary to consider also α_4 and α_5, so that the labor required to improve these values is considerable. Because of this lack in accuracy one can only expect a qualitative agreement between the theoretical behavior of the E_k's and FPU's results.

To obtain even a reasonable reproduction of the curves in Fig. 1, it is necessary to improve on the expressions in Eqs. (24) and (27). Returning to Eq. (23), one obtains

$$\begin{aligned} E_1 \simeq{}& \tfrac{1}{2}\Omega_1^2\{A_1^2 + (B_1^2 + 2A_1B_1)\sin^2 \tfrac{1}{2}\Omega_{11}t \\ &+ (C_1^2 + 2A_1C_1)\sin^2 \tfrac{1}{2}\Omega_{12}t \\ &+ 2B_1C_1 \cos \tfrac{1}{2}(\Omega_{12} - \Omega_{11})t \sin \tfrac{1}{2}\Omega_{11}t \sin \tfrac{1}{2}\Omega_{12}t\}, \end{aligned} \tag{32}$$

$$\begin{aligned} E_2 \simeq{}& \tfrac{1}{2}\Omega_2^2\{B_2^2 \sin^2 \tfrac{1}{2}\Omega_{11}t + C_2^2 \sin^2 \tfrac{1}{2}\Omega_{12}t + 2B_2C_2 \\ &\times \cos \tfrac{1}{2}(\Omega_{11} + \Omega_{12})t \sin \tfrac{1}{2}\Omega_{11}t \times \sin \tfrac{1}{2}\Omega_{12}t\}, \end{aligned}$$

where

$$B_1 = \frac{\lambda V_{112}\alpha_1\alpha_2}{\Omega_1\Omega_{11}}, \qquad C_1 = \frac{\lambda V_{123}\alpha_2\alpha_3}{\Omega_1\Omega_{12}},$$

$$B_2 = \frac{\lambda V_{112}\alpha_1^2}{2\Omega_2\Omega_{11}}, \qquad C_2 = \frac{\lambda V_{123}\alpha_1\alpha_3}{\Omega_2\Omega_{12}}.$$

The corresponding expression for E_3 is rather lengthy and will not be reproduced here. It should be noted, however, that the last term of a_3 [Eq. (23)] now becomes significant. When Eqs. (30) and (31) are substituted into Eq. (32) one obtains the results shown in Fig. 8. Comparing this figure with Fig. 1, it is seen that there is excellent agreement between the maxima of the E_k's and recurrence times, fair agreement between the minima, and qualitative agreement as to the all-over shapes of the E_k's. Considering the numerical mutilations which have been perpetrated on the theory, the agreement is on the whole quite satisfactory. Nonetheless, these results do not constitute a very critical check of the theory, which would require a more accurate computation.

As a point of curiosity (since the approximations of the α_j's is now extremely crude), the recurrence time for the case $\lambda = 1$ was determined from the second-order theory. It is found that the theory yields $\Omega_{11} = 1.44 \times 10^{-3}$, corresponding to a recurrence time of $\tau_1 = 4.36 \times 10^3$. This compares very favorably with the actual value of $\tau_1 = 5 \times 10^3$

obtained by FPU. This result indicates that improving the estimates of the α_i's does not materially effect the values of Ω_{ik}, apparently because of compensating errors.

The results of this section testify to the correctness of the basic arguments in NOI which led to the present theory. In particular, the present results show that the frequencies $\omega(m)$ [Eq. (1)] are not relevant in determining the periodic behavior of coupled oscillators. Thus, the nonergodic behavior of this system does *not* arise from any singular property of $\omega(m)$ but from at least two other properties. The first is the fact that the relevant frequencies $\Omega(m)$, λ) [Eq. (2)] tend to *increase* as λ is *increased.* Thus, while the amount of energy exchange is increased as λ increases, the duration of this exchange decreases (compare Figs. 1, 2; and Figs. 3, 4, 5). This feature, which will be shown in Sec. VI to be dependent on the initial conditions, would not preclude an ergodic behavior, except for the second property of this system. This property, which is inherent in all perfect chains, is that the energy is transmitted to the higher modes only via the intermediary modes. This is the case quite independent of the form of the nonlinearity (cubic, quadratic, and probably even for nonanalytic interactions). The only way to remove this property is to introduce the imperfections discussed above. We shall therefore consider, in a preliminary fashion, the effects arising from such imperfections.

V. THE IMPERFECT CHAIN

In the present section we shall present a preliminary investigation of the effects due to imperfect coupling. To study these effects and make a comparison with the results of the last section, the equations of motion were solved on a computer using the initial conditions of Eq. (12) and $N = 4$. While admittedly the case $N = 4$ is of limited interest, the small amount of computer time required to run a case makes it an economical starting point.

The first computation was made for the case where the energy was initially all in the first mode. In a perfect chain, the third mode would become appreciably excited only once the second mode had acquired considerable energy, for $V_{113} = 0$. To produce a direct coupling between modes 1 and 3, the coupling constants

$$\lambda_i = \lambda \sum_{m=0}^{3} \sigma_m \cos \frac{im\pi}{4} \qquad (\sigma_0 = 1)$$

must contain nonvanishing components σ_1 or σ_3. This follows from the general expression for V_{113}

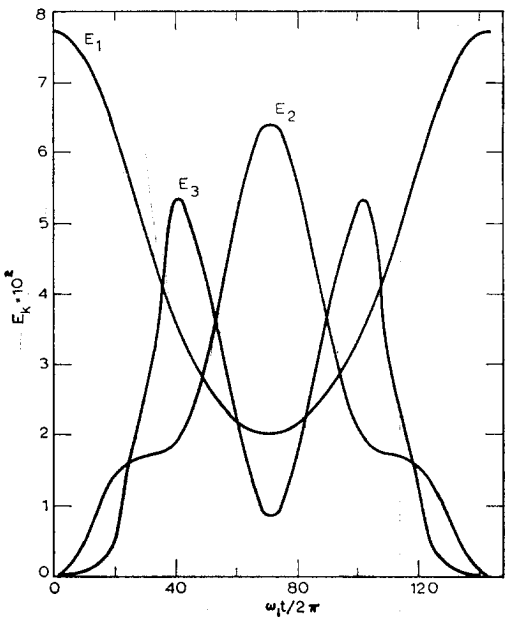


FIG. 8. Theoretical prediction of energy in modes for the FPU case. $N = 32$, $\lambda = \frac{1}{4}$.

obtained in Appendix A. The values λ_i were originally selected by the relationship

$$\lambda_i = 0.575 + 0.3 \cos [\tfrac{1}{4}(i + 1)\pi],$$

which in terms of the present notation yields

$$\lambda = 0.425, \quad \sigma_1 = 0.602, \quad \sigma_2 = 0.146, \quad \sigma_3 = 0.104.$$

The average value of the coupling constant is

$$\bar{\lambda} = \tfrac{1}{4} \sum_{i=0}^{3} \lambda_i = \tfrac{1}{2},$$

so the present case should be compared with the case $N = 4$, $\lambda = \frac{1}{2}$ of the last section. The results of the computations are shown in Fig. 9. It can be seen that this coupling has increased the energy exchange between modes 1 and 3 (compare with Fig. 4), and also the energy exchange to mode 2. While the periodicity of $E_k(t)$ has been somewhat disrupted, neither this effect nor the enhanced energy exchange produces anything that could be termed ergodic.

A second case which was studied was for the initial condition

$$A_k = \delta_{K-2}, \qquad (33)$$

which, in a perfect chain, would not appreciably excite any other mode, because $V_{221} = 0$. To couple mode 2 to mode 1 and 3, the defective coupling

$$\lambda = 0.5, \qquad \sigma_1 = -0.088 = -\sigma_3, \qquad \sigma_2 = 0$$

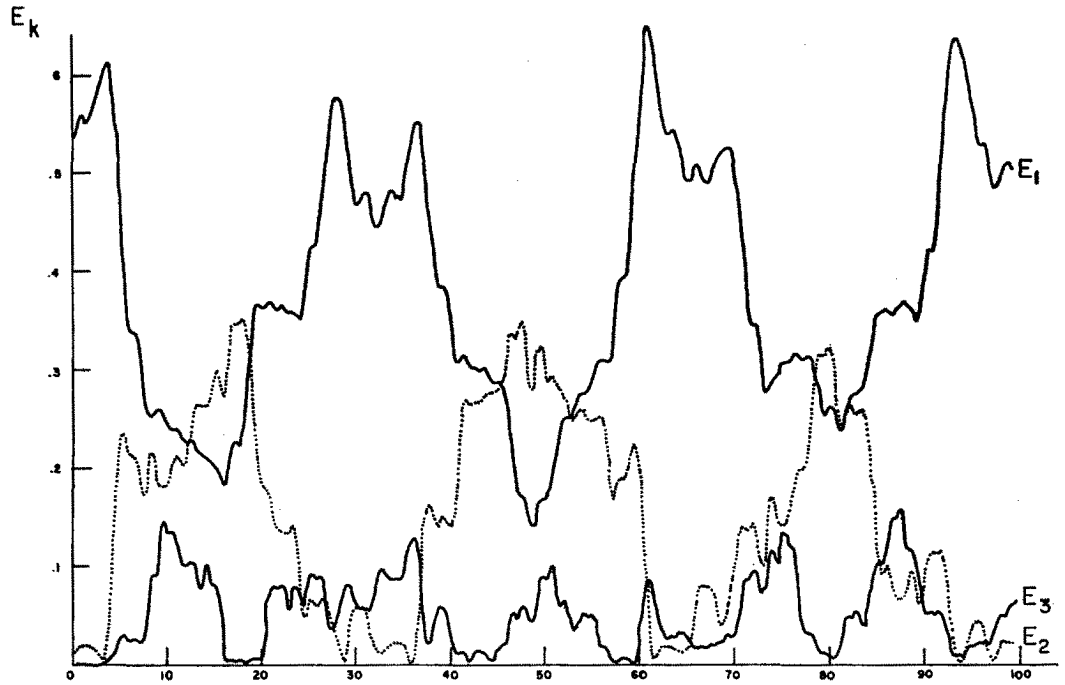


FIG. 9. Energy of modes in an imperfect chain which couples modes 1 and 3. $\lambda_i = 0.575 + 0.3 \cos [\frac{1}{2}(i + 1)\pi]$, $N = 4$.

(i.e., $\lambda_i = 0.5 + 0.25 \cos [\frac{1}{2}(i + 1)\pi]$) was used. The result of this computation is shown in Fig. 10. While the energy exchange between modes 2 and 1 is quite complete, mode 3 does not become appreciably excited. Moreover, the periodicity is very pronounced, even though the recurrence time is relatively long.

These two samplings show that defective coupling, *per se*, need not produce an ergodic behavior even when it has an appropriate regularity. Apparently

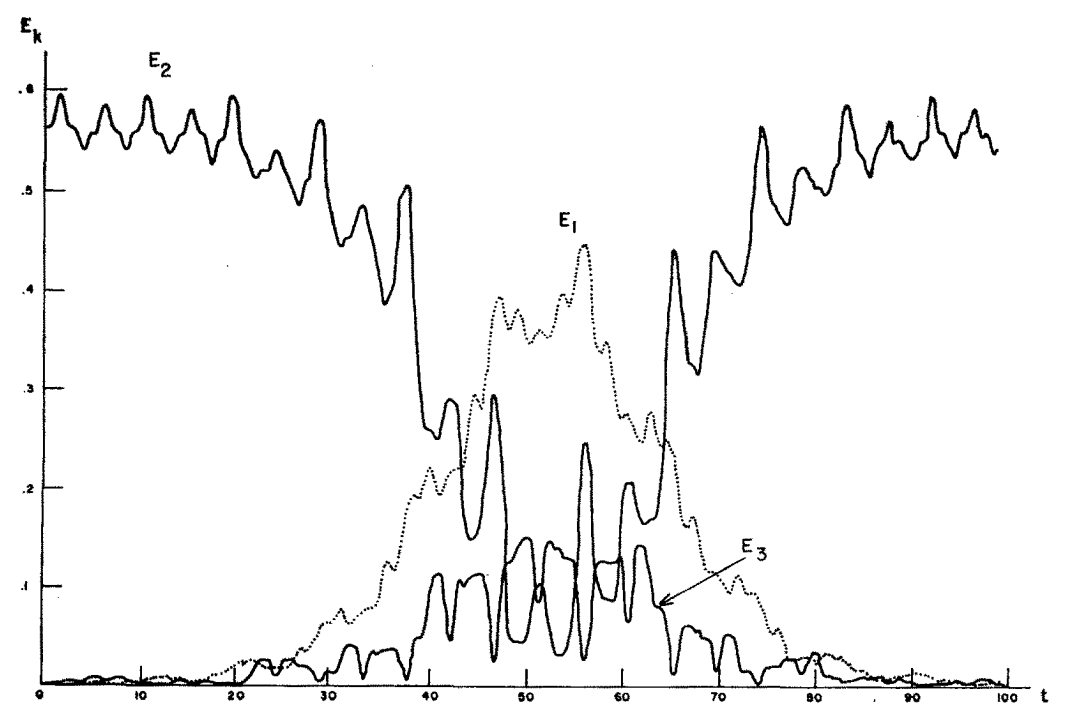


FIG. 10. Energy of modes in an imperfect chain which couples mode 2 to mode 1 and 3. $\lambda_i = 0.5 + 0.25 \cos [\frac{1}{2}(i + 1)\pi]$, $N = 4$.

the defects must also be sufficiently large. To estimate the required magnitude of the defects, we shall consider systems with a large number of particles and fixed total energy (as $N \to \infty$). In keeping with the above calculations, we consider here the case where only a few modes are initially excited [of $O(1)$]. Under these conditions ω is proportional to N^{-1}, so a_k must be of $O(N)$ if H_0 [Eq. (6)] is to remain constant as $N \to \infty$. If the system is to behave ergodically then presumably all $a_k^{(1)}$ must be of the same order [Eq. (16)]. For this to be true, all coefficients $V_{jkl}[\Omega_j \pm \Omega_k \pm \Omega_l]^{-1}$ must be of the same order. If λ is sufficiently small, and no $(\omega_j \pm \omega_k \pm \omega_l)$ vanishes, then these divisors are either of $O(N^{-1})$ or $O(N^{-3})$. Those coupling coefficients V_{jkl} which are *only* connected to large divisors (V_{LD}) must therefore be of $O(N^2)$ times as large as V_{SD} (those with small divisors). The V_{LD} are those $\{V_{jkl}\}$ for which $j \pm k \pm l \neq 0$, whereas V_{SD} are those for which *some* $j \pm k \pm l$ vanishes. Using the general expression (A5), one finds that V_{SD} is of $O(\sigma_0 N^{-7/2})$, whereas V_{LD} is of $O(\sigma_m N^{-7/2})$ for some integer m. We conclude that if all $a_k^{(1)}$ are to be of the same order, then

$$\sigma_m = \sigma_0 O(N^2) = O(N^2), \tag{34}$$

since $\sigma_0 = 1$. Moreover, Eq. (34) apparently must hold for "nearly all" integers m because the number of V_{LD} is of $O(N)$ times the number of V_{SD}. In the light of these results it is not surprising that the above computer results showed no appreciable enhancement of energy exchange. In those cases $|\sigma_m| \leq \frac{1}{2}$, whereas Eq. (34) shows that σ_m should be around 10 for the various a_k's to be of the same order. It should be noted that such large values may lead to negative values of λ_i, and even so, represents only a necessary but not sufficient condition for effective energy exchange among the modes.

VI. MORE GENERAL EXCITATIONS

All of the above computations were made for cases in which only one mode was initially excited. The fact that the system did not behave ergodically for this particular initial condition does not preclude the possibility of an ergodic behavior for other initial conditions (or at least a "nearly ergodic" behavior). Conversely, if the energy becomes distributed over the modes in one case, it may not do so in another case. To examine this problem we shall distinguish between two extreme cases, one of weak coupling ($\lambda \to 0$), and the case of strong coupling ($\lambda \to \infty$). Moreover we shall be interested in the case when all modes are initially excited. From the statistical viewpoint this is of greatest interest since it corresponds to the largest portion of the energy surface, in contrast to the above computations which examine only a small portion of this surface.

To examine the weak coupling case, we consider a perfect chain for which N is very large but bounded. In that case $\omega_j \simeq j\pi/N$, except for $j \simeq N$, and V_{jkl} can be approximated by the expression (A5). The frequencies $\{\Omega_j\}$ are given by Eq. (28), which can now be written in the form

$$\Omega_j \simeq \omega_j + \frac{\lambda^2 \alpha_j^2 V_{jj,2j}^2}{16\omega_j^2 \Omega_{jj}} + \sum_{k \neq j} \frac{\lambda^2 \alpha_k^2}{4\omega_j}\left[\frac{V_{j,k,j+k}^2}{\omega_{j+k}\Omega_{jk}} - \frac{V_{j,k,|j-k|}^2}{\omega_{|j-k|}\Omega_{k|j-k|}}\right], \tag{35}$$

where the large divisors have been approximated, using $\Omega_j \simeq \omega_j$ (because λ is small) and $\omega_{ij} \simeq 2\omega_{i+j}$ (because N is large). While it is clear that the frequencies $\{\Omega_j\}$ depend on the initial conditions, through the α's it is not clear from Eq. (35) what this dependency might be. However, if λ is small, the α_j's may be approximated by the corresponding A_j's. Also if none of the ω_{jk} vanish then $\Omega_{jk} \simeq \omega_{jk}$ for small λ. The question of how small λ must be for this to be valid will be answered shortly. Using the approximation (A5),

$$V_{jkl} \simeq [8/(2N)^{\frac{1}{2}}]\omega_j\omega_k\omega_l \mathcal{P}(\delta_{j=k=l}).$$

Using these results, the bracket [ ] in Eq. (35) equals

$$\frac{32}{N}\omega_j^2\omega_k^2\left(\frac{\omega_{j+k}}{\omega_{jk}} - \frac{\omega_{|j-k|}}{\omega_{k|j-k|}}\right).$$

Because $\omega_{jk} \simeq \frac{1}{8}\omega_j\omega_k\omega_{j+k}$, this bracket vanishes in this limit. While the approximations made here are somewhat crude, it seems very likely that the last term of Eq. (35) is less important than the second term if λ is sufficiently small. Thus one obtains

$$\Omega_j \simeq \omega_j + (32\lambda^2/N)A_j^2\omega_j.$$

This frequency shift, which can be represented as a decrease in the effective mass, is due to the exchange of energy from the mode j to the mode $2j$, and then back to mode j. From this one finds

$$\Omega_{ij} = \omega_{ij} + (32\lambda^2/N)(A_i^2\omega_i + A_j^2\omega_j - A_{i+j}^2\omega_{i+j}). \tag{36}$$

One of the main assumptions has been that $\Omega_{ij} \simeq \omega_{ij}$ (i.e., weak coupling), for which Eq. (36) now gives a criteria. In the FPU case, one finds that weak coupling corresponds to $\lambda \ll 0.1$, compared with their values of 0.25 and 1.

Presumably, the system will behave in a more nearly ergodic fashion if the frequencies $\{\Omega_{ij}\}$ are smaller than the $\{\omega_{ij}\}$. Whether this is the case or not depends on the sign of the bracket ( ) in Eq. (36). If the low modes are predominately excited then Eq. (36) predicts that $\Omega_{ij} > \omega_{ij}$ (i.e., less commensurable), as was found in the above computations. If initially the modes all have nearly the same energy ($A_j \sim \omega_j^{-1}$), then again $\Omega_{ij} > \omega_{ij}$. On the other hand, if more energy is initially concentrated in the higher modes (e.g., $A_j \sim j$), then Eq. (36) predicts that $\Omega_{ij} < \omega_{ij}$. This would indicate that there is more tendency for the energy to be concentrated in the lower modes in an ensemble of such systems (or presumably on a time average). It would be of considerable interest to see whether or not these systems actually have this preference when λ is small.

The case of strong coupling is, as would be expected, much more difficult to examine. If one starts from the basic physical model [Eq. (3)], and lets $N \to \infty$, with λ fixed, then the frequencies ω_k become commensurable and one obtains strong coupling. One also obtains an infinite number of equations determining the (Ω_{ij}, α_j). A more tractable approach would be to take Eq. (6) as the basic physical model, and not assume any relationship between ω_k, V_{jkl} and N (i.e., allow these to be selected arbitrarily). In that case the ω_k's can be taken to be commensurable, while the V_{jkl} and N can be chosen in some convenient manner. By taking N small, the number of equations is reduced and there is some hope of analyzing the situation. We therefore consider the case when

$$\omega_k = k\omega, \tag{37}$$

and V_{jkl} is given by Eq. (8); but now Eq. (9) is not assumed to hold (i.e., the α_{jk}'s are arbitrary).[7] It will also be assumed that λ is sufficiently small for $\Omega_j \simeq \omega_j \gg \Omega_{kl}$ (but not $\Omega_{ij} = \omega_{ij} \equiv 0$). In this case Eqs. (35) and (16) determine the (α_j, Ω_{ij}). In the present case it does not appear to be justified to set $\alpha_j = A_j$, and hence they must be determined. As a particular example we take $N = 4$,[7] in which case Eqs. (35) and (16) reduce to:

$$\begin{aligned}
A_1 &= \alpha_1 - W_{11}\alpha_1\alpha_2 - W_{12}\alpha_2\alpha_3, \\
A_2 &= \alpha_2 - \tfrac{1}{2}W_{12}\alpha_1\alpha_2 + \tfrac{1}{4}W_{12}\alpha_1^2 - \tfrac{7}{32}\tilde{W}_{23}\alpha_3^2, \\
A_3 &= \alpha_3 + \tfrac{1}{3}W_{12}\alpha_1\alpha_2 - \tfrac{1}{8}\tilde{W}_{23}\alpha_2\alpha_3, \\
\Omega_{11} &= \tfrac{1}{2}W_{11}^2\Omega_{11}\alpha_1^2 - W_{12}^2\Omega_{12}(\tfrac{1}{6}\alpha_1^2 - \tfrac{2}{3}\alpha_2^2 + \tfrac{1}{2}\alpha_3^2), \\
\Omega_{12} &= \tfrac{1}{4}W_{11}^2\Omega_{11}\alpha_1^2 + W_{12}^2\Omega_{12}(\tfrac{1}{3}\alpha_1^2 + \tfrac{2}{3}\alpha_2^2 - \alpha_3^2),
\end{aligned} \tag{38}$$

where

$$W_{ij} = \frac{\lambda V_{i,j,i+j}}{2\omega\Omega_{ij}}, \qquad \tilde{W}_{ij} = \frac{\Omega_{ij}}{\omega} W_{ij},$$

where ω is given by Eq. (37). These equations are still sufficiently complicated to require considerable analysis. However, one can see that, except for the last terms of the second and third equations, these equations depend on the ratio of λ/ω, rather than on λ or ω alone. The exceptional terms arise from the so-called Umklapp terms, and are important only when the high modes are excited. Another feature which can be seen from Eq. (38) is that Ω_{ij} should be nearly proportional to λ, and α_j should be nearly independent of λ. Once again the Umklapp terms are the exception to this result. It would be of some interest to see whether these predictions are borne out by computations, and to further check Eq. (38) against numerical solutions.

VII. CONCLUSION

The present study has shown the importance of the shifted frequencies in the determination of the behavior of coupled oscillators. The perturbation method developed in NOI was also found to be quite adequate, even in low order, for determining these frequencies. The confirmation of this perturbative method was, however, rather limited and further studies with different initial conditions would be of considerable interest. In addition to the agreement between the theory and computer solutions, the theory indicates certain general features. One of the most important results is that the "near-ergodic" behavior of such systems can be expected to be strongly dependent on initial conditions. In particular, the present analysis suggests that, for weak coupling, an ensemble composed of perfect coupled oscillators will tend to have their lower modes preferentially excited. Also the effects of imperfections on the gross properties of these systems appear to be small unless *the imperfections* are quite large. Thus the "speck of dust" should not substantially effect the recurrence properties of this system unless the recurrence time is sufficiently long. Further, judiciously selected computations should shed valuable light on these conclusions.

APPENDIX A

In this section we will derive the general expression for V_{jkl} in the case of the imperfect chain, characterized by the coupling constants of Eq. (11). Substituting Eq. (5b) into the Hamiltonian Eq. (4), one

[7] Calculations, based on this model, have been recently performed by Dr. J. Ford (private communication).

obtains

$$H = \tfrac{1}{2}\sum_k \dot{a}_k^2 + \omega_k^2 a_k^2 + \frac{1}{24}\left(\frac{2}{N}\right)^{\frac{1}{2}} \sum_{i=0}^{N-1} \lambda_i \times \left[\sum_k a_k\left(-\omega_k^2 \sin\frac{ik\pi}{N} + \omega_{2k}\cos\frac{ik\pi}{N}\right)\right]^3, \quad \text{(A1)}$$

where $\omega_k = 2\sin(K\pi/2N)$ has been used.

Since λ_i is in the form of Eq. (11), only the terms of the cubic term in Eq. (A1) which will survive after the summation on i are those proportional to $\cos(is\pi/N)$. Collecting only these terms, the cubic term of Eq. (A1) yields

$$\sum_{j,k,l} a_j a_k a_l \left[P\left\{\omega_j^2\omega_k^2\omega_{2l}\sin\frac{ij\pi}{N}\sin\frac{ik\pi}{N}\cos\frac{il\pi}{N}\right\} + \omega_{2j}\omega_{2k}\omega_{2l}\cos\frac{ij\pi}{N}\cos\frac{ik\pi}{N}\cos\frac{il\pi}{N}\right], \quad \text{(A2)}$$

where P stands for the sum of the three distinct permutations of (j, k, l) in the brackets $\{\ \}$. Multiplying Eq. (A2) by λ_i, Eq. (11), and using trigonometric identities, one finds

$$\frac{\lambda}{8}\sum_{j,k,l}\sum_{m=0}^{N-1} a_j a_k a_l \sigma_m(1+\delta_m)\mathcal{P} \times \left[P\left\{\omega_j^2\omega_k^2\omega_{2l}\left[\cos\frac{i\pi}{N}(j-k\pm l\pm m) - \cos\frac{i\pi}{N}(j+k\pm l\pm m)\right]\right\} + \omega_{2j}\omega_{2k}\omega_{2l}\cos\frac{i\pi}{N}(l\pm k\pm j\pm m)\right],$$

where $\mathcal{P}$ represents the sum of all (four or eight) permutations of the signs $(\pm)$. The term δ_m can be dropped if one takes ± 0 (when $m = 0$) as two terms. Substituting this expression into Eq. (A1) and summing on i, one finds

$$H = \tfrac{1}{2}\sum_k a_k^2 + \omega_k^2 a_k^2 + \tfrac{1}{3}\lambda \sum_{j,k,l} V_{jkl} a_j a_k a_l, \quad \text{(A3)}$$

where

$$V_{jkl} = \frac{1}{32}\left(\frac{2}{N}\right)^{\frac{1}{2}}\sum_{m=0}^{N-1}\sigma_m(1+\delta_m)\mathcal{P}[P\{\omega_j^2\omega_k^2\omega_{2l}(\delta_{j-k\pm l\pm m} + \delta_{j-k\pm l\pm m\pm 2N} - \delta_{j+k\pm l\pm m} - \delta_{j+k\pm l\pm m-2N})\} + \omega_{2j}\omega_{2k}\omega_{2l}(\delta_{l\pm k\pm j\pm m} + \delta_{l\pm k\pm j\pm m\pm 2N})]. \quad \text{(A4)}$$

Since V_{jkl} is symmetric in the indices (j, k, l), Eqs. (A3) and (A4) are of the desired form [Eq. (6)]. The expression [Eq. (A4)] for V_{jkl} can be simplified considerably. One useful simplification can be obtained in the limit of large N, for, since $\omega_k \sim N^{-1}$ in this limit, Eq. (A4) reduces to

$$V_{jkl} \simeq \frac{1}{32}\left(\frac{2}{N}\right)^{\frac{1}{2}}\sum_{m=0}^{N-1}\sigma_m(1+\delta_m)\omega_{2j}\omega_{2k}\omega_{2l}\mathcal{P} \times (\delta_{l\pm k\pm j\pm m} + \delta_{l\pm k\pm j\pm m\pm 2N}). \quad \text{(A5)}$$

We will now indicate the method of simplifying Eq. (A4) for the case of a perfect chain (and very similarly for the imperfect chain). The bracket [ ] of Eq. (A4), when rearranged, can be put in the form

$$P[-\omega_j^2\omega_k^2\omega_{2l} + \omega_j^2\omega_l^2\omega_{2k} + \omega_l^2\omega_k^2\omega_{2j} + \omega_{2j}\omega_{2k}\omega_{2l}]\delta_{j+k-l} - [P(\omega_j^2\omega_k^2\omega_{2l}) - \omega_{2j}\omega_{2k}\omega_{2l}]\delta_{j+k+l-2N}. \quad \text{(A6)}$$

Using the fact that $\omega_k = 2\sin(k\pi/2N)$, together with simple trigonometric identities, the two brackets of Eq. (A6) reduce to

$$4[\omega_{2j} + \omega_{2k} - \omega_{2l} + \omega_{2(l-j)} + \omega_{2(l-k)} - \omega_{2(j+k)} + \omega_{2(j+k-l)}],$$

and

$$-4[\omega_{2j} + \omega_{2k} + \omega_{2l} - \omega_{2(j+k+l)} - \omega_{2(j+k)} - \omega_{2(j+l)} - \omega_{2(k+l)}],$$

respectively. Using the Kronecker functions in Eq. (A6), one obtains for Eq. (A6),

$$8(2N)^{\frac{1}{2}}P\{\gamma_{jk}\delta_{j+k-l}\} - 8(2N)^{\frac{1}{2}}\gamma_{jk}\delta_{j+k+l-2N}, \quad \text{(A7)}$$

where γ_{jk} is defined in Eq. (9). Substituting Eq. (A7) into Eq. (A4), and setting $\sigma_0 = 1$, one obtains

$$V_{jkl} = P\{\gamma_{jk}\delta_{j+k-l} - \tfrac{1}{3}\gamma_{jk}\delta_{j+k+l-2N}\}, \quad \text{(A8)}$$

which is identical with Eq. (8).

APPENDIX B

There appear to be very few studies in which the equations of motion for the action-angle variables have been explicitly integrated to determine the behavior of coupled nonlinear oscillators. It is perhaps of some interest therefore to show, by means of a simple example, some of the disadvantages and advantages of this approach.

In many of the examples studied above, the energy was largely confined to the first two modes, so we may expect that the Hamiltonian

$$H = \sum_{k=1}^{2}\Omega_k J_k + \lambda V_{112}\left(\frac{8J_1^2J_2}{\Omega_1^2\Omega_2}\right)^{\frac{1}{2}}\sin^2\theta_1\sin\theta_2 \quad \text{(B1)}$$

should yield comparable results for the energies $E_k = \Omega_k J_k$. We are interested in the slow (and large-scale) variations of J_k, so we shall approximate H by its time averaged value

$$\bar{H} = \sum_{k=1}^{2}\Omega_k J_k + \tfrac{1}{2}\lambda V_{112}\frac{J_1}{\Omega_1}\left(\frac{2J_2}{\Omega_2}\right)^{\frac{1}{2}}\sin(2\theta_1 - \theta_2).$$

The equations of motion are therefore

$$\dot{J}_1 = (\lambda V_{112}/\Omega_1)(2/\Omega_2)^{\frac{1}{2}} J_1 J_2^{\frac{1}{2}} \cos(2\theta_1 - \theta_2),$$
$$\dot{J}_2 = (\lambda V_{112}/\Omega_1)(2\Omega_2)^{-\frac{1}{2}} J_1 J_2^{\frac{1}{2}} \cos(2\theta_1 - \theta_2), \quad \text{(B2)}$$

and

$$2\dot{\theta}_1 - \dot{\theta}_2 = 2\,\partial\bar{H}/\partial J_1 - \partial\bar{H}/\partial J_2$$
$$= 2\Omega_1 - \Omega_2 + (\lambda V_{112}/\Omega_1)(2/\Omega_2)^{\frac{1}{2}}$$
$$\times (J_2^{\frac{1}{2}} - \tfrac{1}{4} J_2^{-\frac{1}{2}} J_1)\sin(2\theta_1 - \theta_2). \quad \text{(B3)}$$

Using the initial conditions [Eq. (33)]:

$$a_1(0) = (\tfrac{1}{2}N)^{\frac{1}{2}}, \quad a_2(0) = 0, \quad \dot{a}_1(0) = \dot{a}_2(0) = 0,$$

and the relationships

$$a_k = (2J_k/\Omega_k)^{\frac{1}{2}} \sin\theta_k, \quad \dot{a}_k = (2J_k\Omega_k)^{\frac{1}{2}} \cos\theta_k,$$

one finds for the initial conditions

$$J_1(0) = \tfrac{1}{4}N\Omega_1, \quad \theta_1(0) = \tfrac{1}{2}\pi, \quad J_2(0) = 0, \quad \text{(B4)}$$

but $\theta_2(0)$ is undetermined. To determine $\theta_2(0)$, one notes that $a_2/\dot{a}_2 \to +0$ as $t \to 0$, a_2 is negative near $t = 0$, $\dot{\theta}_2 > 0$, and assuming continuity, one concludes that

$$\theta_2(0) = \pi. \quad \text{(B5)}$$

Using these initial conditions, it might seem reasonable to approximate $2\theta_1 - \theta_2$ by $(2\Omega_1 - \Omega_2)t$ (the phase factors cancel) and use only the Eqs. (B2), or equivalently setting $\lambda = 0$, in Eq. (B3). However, because of the initial condition $J_2(0) = 0$, the situation must be treated with more care. We shall assume instead that

$$2\theta_1 - \theta_2 = \omega t, \quad \text{(B6)}$$

and determine ω in some low order of approximation. The reason this initial condition requires special care is that neither Eq. (B2) nor Eq. (B3) satisfy a Lipschitz condition at $J_2 = 0$, and, in fact, Eq. (B2) has two solutions satisfying Eq. (B4). These two solutions are:

$$J_1(t) = \tfrac{1}{4}N\Omega_1; \quad J_2(t) = 0,$$

and, using Eq. (B6),

$$J_1(t) = \tfrac{1}{4}N\Omega_1 \operatorname{sech}^2[\alpha \sin\omega t],$$
$$J_2(t) = \tfrac{1}{8}N\Omega_1 \tanh^2[\alpha \sin\omega t], \quad \text{(B7)}$$

where

$$\alpha = \lambda V_{112} N^{\frac{1}{2}}/4\omega(\Omega_1\Omega_2)^{\frac{1}{2}}.$$

We see that J_1 and J_2 have periods π/ω, so if we had taken $\omega = 2\Omega_1 - \Omega_2$, the predicted periods of E_1 and E_2 would have been one-half of the actual periods. To determine ω we insert Eq. (B7) into Eq. (B3), expand and retain only the lowest-order terms in λ (i.e., λ^0). Because of the term $J_2^{-\frac{1}{2}} \sim \lambda^{-1}$, this is not equivalent to setting $\lambda = 0$ in Eq. (B3). One finds by this method

$$\omega = \tfrac{1}{2}(2\Omega_1 - \Omega_2), \quad \text{(B8)}$$

which now yields the correct periods for $E_k(t) = \Omega_k J_k$. If one expands the solutions of Eq. (B7), they are found to agree with our previous solutions [Eq. (24)], for the energies $E_k(t)$.

The principal advantage of this method is that one obtains the energies directly. Moreover, in the present simple case, Eqs. (B2) can be integrated exactly [however, not Eq. (B3)], giving a better idea of how the energy is exchanged when the coupling is strong. The principal disadvantages come from the singularity in Eq. (B3) at $J_2 = 0$, and the determination of Ω_k (i.e., the shift away from ω_k).

ACKNOWLEDGMENTS

The computer solutions and part of the analysis of this study were performed while the author was at the Plasma Physics Laboratory, Princeton, New Jersey, supported by the Atomic Energy Commission. The author is particularly indebted to Mrs. Joan Peskin for her considerable effort in performing the computer calculations despite a number of handicaps.

The author would like to express his appreciation to Dr. Nandor Balaz, Dr. John Green, and Dr. Norman Zabusky for interesting correspondences during the course of this research. The author is also indebted to Dr. Joseph Ford for a number of interesting correspondences concerning his computations of similar systems.

A large part of the present analysis has been supported by The National Science Foundation.

Statistical Theory of the Energy Levels of Complex Systems. IV

FREEMAN J. DYSON AND MADAN LAL MEHTA*
Institute for Advanced Study, Princeton, New Jersey
(Received 21 January 1963)

In previous papers of this series, a theory was constructed for the description of the statistical properties of the eigenvalues of a random matrix of high order. This paper deduces from the same theoretical model the statistical behavior to be expected for a finite stretch of observed eigenvalues chosen out of a much longer stretch of unobserved ones. In comparing the model with experimental data, we have always to deal with such a finite stretch of observed levels. Three "statistics" are investigated; these are quantities which can be computed directly from the observed data, and for which the expectation values and statistical variances can be calculated from the theory. One statistic gives only a precise way of measuring the average level spacing. One provides a test of the model by measuring the degree of long-range order in the level series. The third provides an independent test of the model by measuring the degree of short-range order. In Sec. V these methods are applied to a preliminary analysis of some experimental data on neutron capture levels in heavy nuclei. The results are inconclusive. Discrepancies between theory and observation are large, but the discrepancies might be produced either by incompleteness of the data or by incorrectness of the theoretical model.

I. INTRODUCTION

SUPPOSE that the energy levels of a complex system are observed in an interval of energy of length $2L$, and are found at the positions $[E_1, E_2, \cdots, E_n]$. The statistical theory developed in a series of earlier papers[1] suggests the following hypothesis for the statistical behavior of the E_j. There exists a mean-level-spacing D and a very large integer N such that

$$E_j = (ND/2\pi)\theta_j, \qquad j = 1, \cdots, n. \tag{1}$$

The angles $[\theta_1, \cdots, \theta_n]$ are a section of a longer series $[\theta_1, \cdots, \theta_N]$. The $[\theta_1, \cdots, \theta_N]$ are distributed on the whole circle $0 \leq \theta_j \leq 2\pi$ with the probability distribution

$$Q_N(\theta_1, \cdots, \theta_N) = C_N \prod_{i<j} |e^{i\theta_i} - e^{i\theta_j}|. \tag{2}$$

The $[\theta_1, \cdots, \theta_n]$ are the subset of $[\theta_1, \cdots, \theta_N]$ contained in a randomly chosen arc of angular length

$$2\alpha = (4\pi L/ND), \tag{3}$$

the remaining θ_j which lie outside this arc being unobserved.

The purpose of the present paper is to calculate various properties of the observed level-series which follow from this hypothesis. In I, II, and III, a number of consequences were deduced from Eq. (2) concerning the statistical properties of the complete series of angles $[\theta_1, \cdots, \theta_N]$. Experimental data is always restricted to a partial series $[\theta_1, \cdots, \theta_n]$. For the analysis of experiments, it is essential to study the properties of partial series. The calculations will be made under the conditions

$$1 \ll n \ll N. \tag{4}$$

That is to say, the number n of observed levels is large enough to be described in statistical terms, but is still very small compared with the number of unobserved levels.

The method of procedure is to search for convenient statistics of the observed level series. The word "statistic" is here used in the technical sense customary among statisticians. A "statistic" is a number W which can be computed from an observed sequence of levels alone without other information, and for which the average value $\langle W \rangle$ and the variance

$$V_W = \langle (W - \langle W \rangle)^2 \rangle \tag{5}$$

are known from the theoretical model. A convenient statistic should satisfy two conditions: (i) the computation of W from the observed data should be simple, and (ii) the theoretical figure of merit

$$\Phi_W = [V_W/\langle W \rangle^2] \tag{6}$$

should be as small as possible. This Φ_W is the mean-square fractional deviation which is to be expected between the observed W and the theoretical $\langle W \rangle$, if the theoretical model is a valid one.

A statistic may serve either of two purposes. If $\langle W \rangle$ involves some unknown parameter contained in the theoretical model, then W provides a measure-

* On leave of absence from Tata Institute of Fundamental Research, Bombay, India.

[1] F. J. Dyson, J. Math. Phys. 3, 140, 157, 166 (1962); hereafter referred to as I, II, III, respectively.

ment of this parameter with fractional error proportional to $\Phi_W^{\frac{1}{2}}$. If $\langle W \rangle$ is independent of parameters, then W provides a test of the theory. In the second case, the theory has clearly failed if $(W - \langle W \rangle)^2$ is found to be much larger than V_W.

In the theoretical model described above, there is only one parameter, namely the "ideal level-spacing" D. The integer N is not an effective parameter, since all properties of the model become independent of N as $N \to \infty$. In the analysis of observations, one needs only one statistic to measure D, and any additional statistics will provide tests of the model.

In practice one often has level series which are mixtures of two or more superimposed series with different values of angular momentum and parity. For the analysis of such series one must consider a generalized theoretical model. The generalized model consists of m uncorrelated level series superimposed on the same interval of length $2L$, their theoretical level spacings being $[D_1, \cdots, D_m]$. This generalized model contains m free parameters D_μ. It is more convenient to take for the parameters the overall mean level-spacing

$$D = [\sum D_\mu^{-1}]^{-1} \tag{7}$$

and the fractions

$$f_\mu = (D/D_\mu), \qquad \sum f_\mu = 1, \tag{8}$$

f_μ being the proportion of levels belonging to the μth series.

When one is dealing with a double series ($m = 2$), it is possible with two statistics to measure D and $f_1 = 1 - f_2$, and with a third statistic to obtain a meaningful test of the model. When the series is multiple ($m > 2$), the information that can be obtained in this way is necessarily more fragmentary.

The following three sections (II–IV) are devoted to a detailed examination of the theoretical properties of three types of statistic. Finally, in Sec. V, the practical use of this analysis will be illustrated by a preliminary comparison of theory with observation in the case of series of neutron capture levels in heavy nuclei.

II. LINEAR STATISTICS

The simplest statistic for the measurement of D is

$$W_1 = n, \qquad \langle W_1 \rangle = (2L/D). \tag{9}$$

This is one of a general class of linear statistics

$$W = \sum_{i=1}^{n} f(E_i), \tag{10}$$

$$\langle W \rangle = D^{-1} \int_{-L}^{L} f(x)\, dx, \tag{11}$$

where $f(x)$ is any function defined on the interval $[-L, +L]$, the zero of energy being for convenience chosen at the center of the observed interval. The special function $f(x) = 1$ gives the statistic (9). The mean values (9) and (11) hold both for simple and multiple level series.

The variance of W for a single level series is

$$V_W = \iint_{-(L/D)}^{(L/D)} f(Du)f(Dv) \times [\delta(u - v) - Y_2(u - v)]\, du\, dv, \tag{12}$$

where Y_2 is the two-level cluster function given by Eq. (III, 51). Let $\phi(t)$ be the Fourier transform of $f(x)$ defined by

$$f(x) = \int_{-\infty}^{\infty} \phi(t) \exp [2\pi i x t]\, dt, \tag{13}$$

$$\phi(t) = \int_{-\infty}^{\infty} f(x) \exp [-2\pi i x t]\, dx. \tag{14}$$

Then Eq. (III, 14) gives

$$V_W = D^{-1} \int_{-\infty}^{\infty} \phi(t)\phi(-t)[1 - b(Dt)]\, dt, \tag{15}$$

with $b(k)$ given by Eq. (III, 56).

The function $\phi(t)$ is large only for values of t of the order L^{-1}. Provided that $f(x)$ is a smooth function, the whole of the integral (15) will come from values of t of this order, and so the approximation

$$b(k) = 1 - 2\,|k| \tag{16}$$

may be used with an error of order $(D/2L) = n^{-1}$. Terms of order n^{-1} will always be neglected. Thus for smooth $f(x)$, Eq. (15) becomes

$$V_W = 2 \int_{-\infty}^{\infty} \phi(t)\phi(-t)\,|t|\, dt \tag{17}$$

$$= -\frac{1}{\pi^2} \iint_{-L}^{L} f'(x)f'(y) \log |x - y|\, dx\, dy. \tag{18}$$

Observe that these expressions are now independent of D. Equation (17) may be used whenever the integral converges, and Eq. (18) whenever $f(x)$ is differentiable. However, neither expression applies when $f(x) = 1$, because of the discontinuities at the end points.

When $f(x) = 1$, Eq. (14) gives

$$\phi(t) = [\sin (2\pi L t)/(\pi t)], \tag{19}$$

and Eq. (17) diverges. In this case it is necessary to go back to the exact Eq. (15). The integral (15) may be split into the two parts $|t| < \eta$ and $|t| > \eta$, where

$$L^{-1} \ll \eta \ll D^{-1}. \tag{20}$$

In the part $|t| < \eta$ we use the exact Eq. (19) for $\phi(t)$ and the approximation (16) for $b(k)$. In the part $|t| > \eta$ we use Eq. (III, 74). Eq. (15) then gives

$$V_n = (2/\pi^2)[\ln (2\pi n) + 1 + \gamma - (\pi^2/8)], \tag{21}$$

where the n in the logarithm means the expectation-value $\langle n \rangle = (2L/D)$. The figure of merit for the statistic n is

$$\Phi_n = (2/\pi^2 n^2)[\ln (2\pi n) + 1 + \gamma - (\pi^2/8)]. \tag{22}$$

Equation (21) shows that the statistical fluctuation of n is in practice very small. For example, when $n = 100$, $V_n = 1.4$, and so D is measured to 1% accuracy by the simple statistic $(2L/n)$. Eq. (21) confirms and makes more precise the result obtained in Eq. (II, 112).

Although the variance of n is in practice small, it is clear from Eq. (18) that n is not the best statistic for measuring D. If we write

$$f(x) = g(x/L), \tag{23}$$

where $g(u)$ is a smooth function defined for $(-1 < u < +1)$, then Eq. (18) is a pure number, independent of D and L. Hence $V_W < V_n$ for large n. In other words, *any* smooth function $g(u)$ gives a more precise measurement of D than n does, provided n is large enough. Quantitatively, the figure of merit for the statistic W is

$$\Phi_W = \frac{4}{\pi^2 n^2}\left[\frac{\iint_{-1}^{1} g'(x)g'(y) \log |x - y|\, dx\, dy}{\left\{\int_{-1}^{1} g'(x)x\, dx\right\}^2}\right]. \tag{24}$$

It is of some interest to determine from Eq. (24) the best possible function $g(u)$. By taking a variation of Eq. (24), we find that the optimum $g'(u)$ satisfies the integral equation

$$\int_{-1}^{1} g'(y) \log |x - y|\, dy = x, \tag{25}$$

which has the solution

$$g(x) = (1/\pi)(1 - x^2)^{\frac{1}{2}}. \tag{26}$$

The optimum statistic is

$$W = \sum_{i=1}^{n} [1 - (E_i/L)^2]^{\frac{1}{2}}, \qquad \langle W \rangle = (\pi L/2D), \tag{27}$$

which has the variance

$$V_W = \tfrac{1}{2}, \tag{28}$$

and the figure of merit

$$\Phi_W = (8/\pi^2 n^2). \tag{29}$$

Eq. (28) shows that the average number of levels in an interval of length $2L$, namely

$$\langle n \rangle = (2L/D) = (4/\pi)\langle W \rangle, \tag{30}$$

can be measured with an absolute mean-square error less than one unit, using the statistic $(4W/\pi)$ with W given by Eq. (27). If we used for $\langle n \rangle$ the statistic n, we would have a larger error given by the square root of Eq. (21). The fact that $\langle n \rangle$ is measurable to within one unit in an interval of arbitrary length is an indication of the essentially crystalline character of the level series.

In the case of a multiple level series, the variance of a linear statistic is the sum of the variances computed for the individual series. Thus Eq. (21) becomes

$$V_n = \frac{2m}{\pi^2}\left[\log (2\pi n) + 1 + \gamma - \frac{\pi^2}{8}\right] + \frac{2}{\pi^2}\sum_{1}^{m} \log f_\mu, \tag{31}$$

while Eq. (28) becomes simply

$$V_W = \tfrac{1}{2}m, \tag{32}$$

m being the number of independent series. Thus a rare series contributes as much as a dense series to the error in measurement of D. This effect of a rare series may be of some practical importance, for example, when a few p-wave levels are mixed into a series of neutron-capture levels which are predominantly s wave. The effect will be even worse than is indicated by Eq. (32), because in practice not all the levels in the p-wave series will be observed.

III. LEAST-SQUARE STATISTICS

It is customary[2] to represent the experimental level distributions graphically by a plot of $N(E)$ against E, where $N(E)$ is the number of levels having energy between zero and E. The resulting graphs are in appearance like staircases, and the staircase gives a good visual impression of the over-

[2] J. Rainwater, *Handbuch der Physik*, edited by S. Flügge (Springer-Verlag, Berlin, 1957), Vol. 40; T. Ericson, Phil. Mag. Suppl. 9, No. 36, 425 (1960); see also references 4 and 6 below.

all regularity of the level series. It is customary to measure the average level spacing by drawing a line having the same average slope as the staircase. Fluctuations in the level series then appear as deviations of the staircase from the straight line.

The aim of this section is to make a quantitative study of this "staircase" method of analysis of data, and to find out precisely what can be learned from it. It should be possible to construct, from the deviations between the staircase and the straight line, a statistic which will serve to test whether the overall irregularity of the observed level series is in agreement with the theoretical model or not.

It is convenient to take the observed energy interval as $[-L, +L]$, with zero in the center. For negative E, $N(E)$ is defined as *minus* the number of levels between E and zero. The problem is to analyze the deviation of the staircase graph

$$y = N(E) \tag{33}$$

from a suitably chosen straight line

$$y = AE + B. \tag{34}$$

A suitable method of analysis is to fit Eq. (34) to (33) by a least-squares criterion. The mean-square deviation of Eq. (33) from the best fit (34) is

$$\Delta = \operatorname*{Min}_{A,B} \left\{ \frac{1}{2L} \int_{-L}^{L} [N(E) - AE - B]^2 \, dE \right\}. \tag{35}$$

We choose this Δ as the statistic which measures the irregularity of the level series.

We consider three possible ways of fitting the straight line to the staircase, which give rise to three different statistics Δ_1, Δ_2, Δ_3. For Δ_1 we assume $B = 0$, so that the line is constrained to pass through the fixed point $y = N(E) = E = 0$ at the center of the interval, only the slope A being fitted to the data. Eq. (35) then gives

$$\Delta_1 = \frac{1}{2L} \int_{-L}^{L} [n(E)]^2 \, dE - \frac{3}{4L^4} \left[\int_{-L}^{L} n(E) E \, dE \right]^2, \tag{36}$$

$$n(E) = N(E) - (E/D). \tag{37}$$

For Δ_2 we assume that the line is constrained to pass through the fixed point $y = N(-L)$, $E = -L$ at the lower end of the observed interval. Then

$$\Delta_2 = \frac{1}{2L} \int_{-L}^{L} [n'(E)]^2 \, dE - \frac{3}{16L^4} \left[\int_{-L}^{L} n'(E)(E + L) \, dE \right]^2, \tag{38}$$

$$n'(E) = N(E) - N(-L) - [(E + L)/D]. \tag{39}$$

For Δ_3 we leave the line unconstrained so that the parameters A, B may vary independently. This gives

$$\Delta_3 = \Delta_1 - \frac{1}{4L^2} \left[\int_{-L}^{L} n(E) \, dE \right]^2, \tag{40}$$

or equivalently

$$\Delta_3 = \Delta_2 - \frac{9}{16L^4} \left[\int_{-L}^{L} n'(E)(E - \tfrac{1}{3}L) \, dE \right]^2. \tag{41}$$

The reason for considering the unsymmetrical statistic Δ_2 is that the fitting of data in the past[2] has usually been done in this way, with the line forced to pass through the lower end point of the staircase. Δ_1 was considered in order to see whether a symmetric fitting would give a more useful statistic. Unexpectedly it turns out that Δ_1 is statistically much more unreliable than Δ_2. On general grounds one would expect Δ_3 to be more informative than either Δ_1 or Δ_2. In fact, the superiority of Δ_3 is greater than one would naively suppose. The additional labor of making a 2-parameter fit to determine Δ_3 is very much worthwhile.

The first step in the theoretical analysis is to calculate expectation values for Δ_1, Δ_2, Δ_3. If k is any nonzero integer, we write

$$\rho_k = \sum_{j=1}^{N} \exp(ik\theta_j), \tag{42}$$

and then by Eq. (III, 14),

$$\langle \rho_k \rho_{-l} \rangle = N \, \delta_{kl}(1 - b(k/N)). \tag{43}$$

When k is small compared with N, the approximation (16) holds, and so

$$\langle \rho_k \rho_{-l} \rangle = 2 \, |k| \, \delta_{kl}, \qquad |k| \ll N. \tag{44}$$

We can express Δ_1, Δ_2, Δ_3 in terms of the ρ_k as follows, using Eqs. (36), (40), and (41):

$$\Delta_3 = V_3 - (3/4L^4)(R_3)^2, \tag{45}$$

$$\Delta_1 = \Delta_3 + (1/4L^2)(R_1)^2, \tag{46}$$

$$\Delta_2 = \Delta_3 + (9/16L^4)(R_2)^2, \tag{47}$$

$$V_3 = \frac{1}{16\pi^2 L^2} \sum \sum \frac{\rho_k \rho_{-l}}{kl} \times [2L\psi(k - l) - \psi(k)\psi(-l)], \tag{48}$$

$$R_3 = \frac{1}{4\pi^2} \sum \frac{\rho_k}{k} \Delta\psi(k), \tag{49}$$

$$R_1 = \frac{1}{2\pi i} \sum \frac{\rho_k}{k} (2L - \psi(k)), \tag{50}$$

$$R_2 = -\frac{1}{4\pi^2}\sum \frac{\rho_k}{k}\left[\Delta\psi(k) + \frac{2\pi i L}{3} \times \left(\psi(k) - 2L\exp\left(\frac{2\pi i k L}{ND}\right)\right)\right]. \quad (51)$$

We have used the notations

$$\psi(k) = \phi(k/ND), \qquad \Delta\psi(k) = \tfrac{1}{2}ND(\psi(k+1) - \psi(k-1)), \quad (52)$$

with $\phi(t)$ defined by Eq. (19). The sums over k and l always extend over integer values excluding zero. The expectation values of Δ_1, Δ_2 and Δ_3 can now be evaluated by means of Eqs. (43), (44) and (III, 74). The results are

$$\langle\Delta_1\rangle = \frac{2}{\pi^2}\left[\ln(\pi n) + \gamma - \frac{\pi^2}{8} - \frac{3}{8}\right], \quad (53)$$

$$\langle\Delta_2\rangle = \frac{5}{4\pi^2}\left[\ln(2\pi n) + \gamma - \frac{\pi^2}{8} - \frac{21}{20}\right], \quad (54)$$

$$\langle\Delta_3\rangle = \frac{1}{\pi^2}\left[\ln(2\pi n) + \gamma - \frac{\pi^2}{8} - \frac{5}{4}\right]. \quad (55)$$

Putting in the numerical values of the constants, we find

$$\langle\Delta_1\rangle = (2/\pi^2)[\ln n + 0.1132], \quad (56)$$

$$\langle\Delta_2\rangle = (5/4\pi^2)[\ln n + 0.1313], \quad (57)$$

$$\langle\Delta_3\rangle = (1/\pi^2)[\ln n - 0.0687]. \quad (58)$$

The mean-square deviations (56)–(58) are in all practical cases extraordinarily small. For example, in a series of 100 observed levels we have $\langle\Delta_2\rangle < 0.6$, $\langle\Delta_3\rangle < \frac{1}{2}$. This means that the deviation of the staircase graph from a straight line is almost always less than one unit. The theory predicts that level series should be spaced with an extremely high degree of regularity.

To emphasize the strong tendency toward regular spacing of energy levels, it is interesting to contrast the results (56)–(58) with the corresponding mean-square deviations which would be expected for a completely random sequence of numbers. For n numbers distributed at random in the interval $[-L, +L]$, the expected mean-square deviations are given by

$$\langle\Delta_1\rangle_R = (3/20)n, \quad (59)$$

$$\langle\Delta_2\rangle_R = (1/10)n, \quad (60)$$

$$\langle\Delta_3\rangle_R = (1/15)n. \quad (61)$$

In going from a sequence of random numbers to a sequence of energy levels, the effect on the mean-square deviations is roughly to replace n by $(\ln n)$. This illustrates how dramatically nonrandom an energy-level series really is.

In order to make practical use of the statistics Δ_1, Δ_2, or Δ_3, we need to know their theoretical variances. To calculate

$$V_{\Delta 3} = \langle(\Delta_3)^2\rangle - \langle\Delta_3\rangle^2, \quad (62)$$

it is necessary to multiply out the expressions (48) and (49) and to evaluate expectation values of products of four ρ_k. It happens very fortunately that in this problem the approximation

$$\langle\rho_k\rho_{-l}\rho_m\rho_{-p}\rangle = 4\,|km|\,\delta_{kl}\delta_{mp} + 4\,|kl|\,\delta_{kp}\delta_{lm} + 4\,|kp|\,\delta_{k,-m}\delta_{l,-p} \quad (63)$$

can be used, the error being of the relative order $(1/n)$. After some heavy algebra, we find the result

$$V_{\Delta 3} = \frac{1}{\pi^4}\left[\frac{4\pi^2}{45} + \frac{7}{24}\right] = \frac{1.1690}{\pi^4}, \quad (64)$$

a pure number independent of n and D. According to Eq. (6), the figure of merit of the statistic Δ_3 is then

$$\Phi_{\Delta 3} = 1.17(\ln n)^{-2}. \quad (65)$$

The statistical fluctuations of Δ_3 are expected to be of the order of $(\ln n)^{-1}\langle\Delta_3\rangle$. This means that Δ_3 is a good enough statistic to provide a meaningful test of the theory.

The behavior of Δ_1 and Δ_2 is rather different. Consider for example the variance of Δ_1. Eq. (46) gives for this quantity

$$V_{\Delta 1} = V_{\Delta 3} + (1/2L^2)[\langle\Delta_3 R_1^2\rangle - \langle\Delta_3\rangle\langle R_1^2\rangle] + (1/16L^4)[\langle R_1^4\rangle - \langle R_1^2\rangle^2]. \quad (66)$$

Now Eq. (50) shows R_1 to be a linear combination of the ρ_k, and Eq. (63) shows that the ρ_k behave like independent Gaussian variables so far as their second and fourth moments are concerned. Therefore R_1 behaves like a Gaussian variable, and

$$\langle R_1^4\rangle = 3\langle R_1^2\rangle^2 = (48L^4/\pi^4)(\ln n)^2. \quad (67)$$

It can be shown by a detailed examination that the use of Eq. (63) is justified in this context so long as terms of relative order $(\ln n)^{-1}$ are neglected. Thus Eq. (67) holds with an error of order $(\ln n)$. Since $V_{\Delta 3}$ is of order unity, the second term in Eq. (66) can be at most of order $(\ln n)$. Therefore Eqs. (66) and (67) give to the leading order in $(\ln n)$,

$$V_{\Delta 1} = (2/\pi^4)(\ln n)^2, \quad (68)$$

and the figure of merit for Δ_1 is

$$\Phi_{\Delta 1} = \tfrac{1}{2}. \tag{69}$$

By exactly the same argument, since R_2 in Eq. (47) is also a linear combination of the ρ_k, we find

$$\langle R_2^4 \rangle = (16L^8/27\pi^4)(\ln n)^2, \tag{70}$$

$$V_{\Delta 2} = (1/8\pi^4)(\ln n)^2, \tag{71}$$

$$\Phi_{\Delta 2} = (2/25) = 0.08, \tag{72}$$

again neglecting terms of relative order $(\ln n)^{-1}$. In this case, because the numerical coefficient of Eq. (71) is so small, the neglected terms may be comparable with those retained when $(\ln n)$ is of the order of 4 or 5. Thus Eqs. (71) and (72) are rather unreliable in contrast to Eqs. (64) and (65) which have errors controlled to order $(1/n)$.

Summarizing the foregoing results, we may say that Δ_1 is useless as a practical statistic, since its statistical scatter is of the same order as its mean value. With Δ_2 the situation is much improved, since the figure of merit is of the order of 0.1, and the scatter is only of the order of $\frac{1}{3}$ of the mean value. However, the theoretical calculation of the variance of Δ_2 is quite inaccurate. The statistic Δ_3 is to be preferred to Δ_1 and Δ_2 for two reasons: first because it has a somewhat smaller variance than Δ_2, and second because the variance of Δ_3 is much more accurately known.

The statistical properties of Δ_1, Δ_2 and Δ_3 may be understood more easily in terms of a simple mathematical model. Let $z_0, z_1, \cdots, z_n$ be $(n+1)$ independent Gaussian random variables with mean zero and variance $(1/\pi^2)$. Then Δ_3 has statistical behavior similar to that of the quantity

$$\Gamma_3 = \sum_{j=1}^{n} j^{-1} z_j^2. \tag{73}$$

The mean of Γ_3 is

$$(1/\pi^2) \sum j^{-1} \sim (1/\pi^2)(\ln n), \tag{74}$$

while the variance of Γ_3 is

$$(2/\pi^4) \sum j^{-2} \sim (2/\pi^4)(\pi^2/6), \tag{75}$$

a number independent of n. The statistic Δ_3 is "good" because it is effectively a sum of a large number of independent contributions $(j^{-1}z_j^2)$. On the other hand, by Eqs. (46) and (47), the quantities Δ_1 and Δ_2 behave like

$$\Gamma_1 = \Gamma_3 + (\ln n) z_0^2, \tag{76}$$

$$\Gamma_2 = \Gamma_3 + \tfrac{1}{4}(\ln n) z_0^2, \tag{77}$$

respectively. For large n, the behavior of Γ_1 and Γ_2 is dominated by the single term in z_0^2, and their large statistical scatter is mainly due to this term.

The foregoing calculations have all referred to a single level series. In the case of a mixed series, with m components having fractional densities $f_1, \cdots, f_m$, the results are changed as follows. The expressions (45)–(51) for Δ_1, Δ_2 and Δ_3 are still valid, but ρ_k is now a sum of m independent contributions. Eqs. (43) and (44) are replaced by

$$\langle \rho_k \rho_{-l} \rangle = N \delta_{kl} [1 - \sum_{\mu} f_\mu b(k/f_\mu N)], \tag{78}$$

$$\langle \rho_k \rho_{-l} \rangle = 2\,|k|\, m \delta_{kl}, \qquad |k| \ll N. \tag{79}$$

The right-hand side of Eq. (63) is multiplied by m^2. The effect of these changes is that in Eqs. (53)–(55) all the right-hand sides are multiplied by m, and at the same time $(\ln n)$ is replaced by

$$\ln \bar{n} = \ln n + m^{-1} \sum_{\mu} (\ln f_\mu). \tag{80}$$

This $\bar{n}$ is the geometric mean of the numbers of observed levels in each of the m series. In Eqs. (64), (68), and (71), the right-hand sides are multiplied by m^2, and $(\ln n)$ is again replaced by $(\ln \bar{n})$. For these results to hold it is assumed that each of the single series populations nf_μ is large compared with unity.

Consider in particular the statistic Δ_3 for a multiple series. Eqs. (58) and (64) now become

$$\langle \Delta_3 \rangle = (m/\pi^2)[\ln \bar{n} - 0.0687], \tag{81}$$

$$V_{\Delta 3} = 1.1690(m^2/\pi^4). \tag{82}$$

If n is of the order of 100, a measurement of Δ_3 should be able to distinguish reliably between an unmixed series $(m = 1)$ and a double series $(m = 2)$. The difference between the values of $\langle \Delta_3 \rangle$ for $m = 1$ and $m = 2$ is well outside the standard deviation of either. However, a decision between $m = 2$ and $m = 3$ will not be so unambiguous.

If a series is known to be double $(m = 2)$, then $\langle \Delta_3 \rangle$, according to Eq. (81), depends on the mixing ratio $(f_1 : f_2)$. However, the variance of Δ_3 is just large enough to make a significant measurement of the mixing ratio by means of Δ_3 impossible.

IV. ENERGY STATISTIC

In Sec. VI of III it was shown that the quantity

$$W = -\sum_{i<j} \ln |\exp(i\theta_i) - \exp(i\theta_j)| + \tfrac{1}{2}N(\ln N), \tag{83}$$

representing the total energy of a Coulomb gas, is at the same time a very sensitive statistic. By means

of W one could test accurately whether a given set of angles $[\theta_1, \cdots, \theta_N]$ distributed around the unit circle was or was not in agreement with the theoretical probability distribution (2).

We now wish to construct a statistic Q, analogous to W, but requiring knowledge of a finite stretch of observed levels $[E_1, \cdots, E_n]$ only. The simplest statistic of this kind would be

$$Q_1 = -\sum_{i<j} f(E_i, E_j) \ln |2\pi(E_i - E_j)|, \tag{84}$$

where $f(E, E')$ is some weight function which vanishes except when both E and E' are within the observed energy interval $[-L, +L]$. Without loss of generality we may assume $f(x, y) = f(y, x)$. It turns out that Q_1 is not a satisfactory statistic, for the following reason. Suppose that one energy level E_i varies slightly while the others are fixed. Then

$$(\partial Q_1/\partial E_i) = -(\partial W_i(E_i)/\partial E_i), \tag{85}$$

$$W_i(E) = \sum_{j \neq i} f(E, E_j) \ln |2\pi(E - E_j)|. \tag{86}$$

The expectation value $\langle W_i(E)\rangle$ is a strongly varying function of E, and so Q_1 varies strongly when individual levels E_i are moved. Consequently the statistical scatter of Q_1 cannot be expected to be small. A "good" statistic is one with a small statistical scatter. Therefore Q_1 is not likely to be good, and in looking for a good statistic we should try to construct a quantity which is stationary under variations of the E_i.

We wish then to add counter terms to Q_1 in order to cancel the derivative (85). The cancellation cannot be exact, but at least it should be possible to cancel the major part of Eq. (85).

Let E_i be held fixed and expectation values taken with respect to the other levels E_j. Then Eq. (86) with Eq. (III, 9) gives

$$\langle W_i(E)\rangle = D^{-1}[V(E) - G(E_i, E)], \tag{87}$$

$$V(E) = \int f(E, E') \ln |2\pi(E - E')| \, dE', \tag{88}$$

$$G(E_i, E) = \int f(E, E') \ln |2\pi(E - E')| \times Y_2[(E_i - E')/D] \, dE'. \tag{89}$$

In Eq. (89) we are interested only in values of E and E' which are in the neighborhood of E_i. We assume for the purposes of the present heuristic argument that $f(E, E')$ is a smoothly varying function. Thus for E and E' in the neighborhood of E_i, we may write

$$f(E, E') = F(E_i) + \tfrac{1}{2}F'(E_i)[(E - E_i) + (E' - E_i)], \tag{90}$$

with

$$F(E) = f(E, E). \tag{91}$$

Therefore, to the first order in $(E - E_i)$,

$$D^{-1}G(E_i, E) = F(E_i)(\ln D + 2U) + \tfrac{1}{2}F'(E_i)(E - E_i)(\ln D + 2U - 1), \tag{92}$$

where, by Eq. (III, 89),

$$U = \frac{1}{2}\int \ln |2\pi x| \; Y_2(x) \, dx = 1 - \tfrac{1}{2}(\gamma + \ln 2) = 0.365. \tag{93}$$

Substituting from Eq. (92) into (84) and (85), we find

$$\langle Q_1\rangle = -\tfrac{1}{2}D^{-2}\int V(E) \, dE + (U + \tfrac{1}{2}\ln D)D^{-1}\int F(E) \, dE, \tag{94}$$

$$\langle \partial Q_1/\partial E_i\rangle = -D^{-1}V'(E_i) + F'(E_i)(U + \tfrac{1}{2}\ln D - \tfrac{1}{2}). \tag{95}$$

In order to cancel the first term on the right-hand side of Eq. (95), we add to Q_1 the expressions

$$Q_2 = P \sum V(E_i), \tag{96}$$

$$Q_3 = -\tfrac{1}{2}P^2 \int V(E) \, dE, \tag{97}$$

with

$$P = \Big[\sum F(E_i)\Big] \Big/ \Big[\int F(E) \, dE\Big]. \tag{98}$$

This quantity P must be used instead of its expectation value D^{-1} in Eq. (96), (97), since the statistic Q must not contain explicitly the theoretical parameter D. Eq. (96), (97) then give

$$\langle Q_2\rangle = -2\langle Q_3\rangle = D^{-2}\int V(E) \, dE, \tag{99}$$

$$\langle \partial Q_2/\partial E_i\rangle = D^{-1}V'(E_i) + \Big[\int V \, dE \Big/ \int F \, dE\Big]D^{-1}F'(E_i), \tag{100}$$

$$\langle \partial Q_3/\partial E_i\rangle = -\Big[\int V \, dE \Big/ \int F \, dE\Big]D^{-1}F'(E_i). \tag{101}$$

Therefore $(Q_2 + Q_3)$ precisely cancels the first

terms on the right of both Eqs. (94) and (95). We are still left with the undesirable terms in $(\ln D)$. To cancel these we add a fourth term,

$$Q_4 = \tfrac{1}{2}(\ln P) \sum F(E_i), \tag{102}$$

which gives

$$\langle Q_4 \rangle = -\tfrac{1}{2}D^{-1}(\ln D) \int F(E)\, dE, \tag{103}$$

$$\langle \partial Q_4/\partial E_i \rangle = -\tfrac{1}{2}(\ln D)F'(E_i) + \tfrac{1}{2}F'(E_i). \tag{104}$$

So we write

$$Q_T = Q_1 + Q_2 + Q_3 + Q_4, \tag{105}$$

and then

$$\langle Q_T \rangle = UD^{-1} \int F(E)\, dE, \tag{106}$$

$$\langle \partial Q_T/\partial E_i \rangle = UF'(E_i). \tag{107}$$

Equations (106) and (107) show that Q_T is stationary under all variations of the E_i for which $[\sum F(E_i)]$ is constant.

The expression Q_T still has one undesirable feature. The logarithms in Eqs. (84) and (88) depend on the scale or unit in terms of which the energies E_i are expressed. According to Eq. (106), the scale dependence cancels out from the expectation value $\langle Q_T \rangle$. However, Q_T itself is not scale-free. To remove the ambiguity we introduce a quantity R with the dimensions of energy. The magnitude of R will be chosen later; typically it will be of the order of a few times the mean-level-spacing D. For the present we assume only the inequalities

$$D < R < L. \tag{108}$$

The scale-free statistic Q is obtained from Q_T by changing E_i into $(E_i/2\pi R)$ in all the logarithms. We have then explicitly

$$\begin{aligned} Q = &-\sum_{i<j} f(E_i, E_j) \ln |(E_i - E_j)/R| \\ &+ \sum_{ij} \sum F(E_i)U(E_j) - \tfrac{1}{2}U_0 \sum_{ij} \sum F(E_i)F(E_j) \\ &+ \tfrac{1}{2}[\sum_i F(E_i)] \ln \left(2\pi R \sum_i F(E_i) \Big/ \int F(E)\, dE\right), \end{aligned} \tag{109}$$

where

$$U(E) = \left(\int f(E, E') \times \ln |(E - E')/R|\, dE'\right) \Big/ \left(\int F(E')\, dE'\right), \tag{110}$$

$$U_0 = \left(\iint f(E, E') \times \ln |(E - E')/R|\, dE\, dE'\right) \Big/ \left(\int F(E')\, dE'\right)^2. \tag{111}$$

Although we have been led to the statistic Q by valid heuristic arguments, the main justification for the use of Q is the fact that its variance is small and calculable. The computation of the variance of Q requires that we multiply out the square of Eq. (109) and then make extensive use of the identities (III, 12). After tremendous cancellations of terms (which of course are not accidental but arise from the stationary character of Q) the variance of Q reduces to the expression

$$\begin{aligned} V_Q = &-\tfrac{1}{4} \iiiint Y_4(x, y, z, w) \\ &\times L(x, y)L(z, w)\, dx\, dy\, dz\, dw \\ &+ \iiint Y_3(x, y, z) \\ &\times L(x, y)L(x, z)\, dx\, dy\, dz \\ &- \frac{1}{2} \iint Y_2(x, y)[L(x, y)]^2\, dx\, dy \\ &+ \frac{1}{2} \iiiint [\delta(x - z) - Y_2(x, z)] \\ &\times [\delta(y - w) - Y_2(y, w)] \\ &\times L(x, y)L(z, w)\, dx\, dy\, dz\, dw, \end{aligned} \tag{112}$$

where

$$L(x, y) = f(xD, yD) \ln |(x - y)D/R|. \tag{113}$$

Each term in Eq. (112) corresponds precisely to a term in Eq. (III, 91). A similar expression for the expectation value of Q is easily obtained, namely

$$\begin{aligned} \langle Q \rangle = &\frac{1}{2} \iint Y_2(x, y)L(x, y)\, dx\, dy \\ &+ \tfrac{1}{2} \ln (2\pi R/D) \int F(xD)\, dx. \end{aligned} \tag{114}$$

The previous expression for $\langle Q \rangle$ given by Eq. (106) was obtained from the approximation (90), which is only valid when $f(E, E')$ is everywhere smooth. Eq. (114) is a more accurate expression which can be used with any choice of $f(E, E')$.

We now finally specify the statistic Q to be used for the analysis of data, by making the choice

$$f(x, y) = 1 \quad \text{for} \quad |x - y| < R, \quad |x| < L, \quad |y| < L, \tag{115}$$

$$f(x, y) = 0 \quad \text{otherwise},$$

for the weight function. This choice is guided by two considerations. (1) It will be seen later that the variance V_Q is increased by any discontinuities in $L(x, y)$. It is therefore advantageous to place discontinuities of $f(xD, yD)$ along the lines $|x-y|\,D = R$ where the logarithmic factor in Eq. (113) vanishes. (2) The labor of computing Q would become prohibitive if any but the simplest possible $f(x, y)$ were chosen.

This choice of $f(x, y)$ implies

$$F(x) = 1, \qquad |x| < L, \tag{116}$$

$$U(x) = -(R/L), \qquad |x| < L - R, \tag{117}$$

$$U(x) = -(2L)^{-1}\{R + (L - |x|) \times (1 - \ln [(L - |x|)/R])\}, \qquad L - R < |x| < L, \tag{118}$$

$$U_0 = -(R/L) + \tfrac{1}{3}(R/L)^2. \tag{119}$$

The number of terms which have to be computed in obtaining Q is approximately Mn, where

$$n = 2L/D \tag{120}$$

is roughly the number of observed levels, while

$$M = R/D \tag{121}$$

is a parameter still at our disposal.

In calculating $\langle Q\rangle$ and V_Q we shall neglect terms of order unity compared with the leading terms which are of order n. Then Eq. (114) gives, by virtue of Eqs. (III, 12) and (93),

$$\langle Q\rangle = nU - nM \int_1^\infty Y_2(Mu)(\ln u)\, du. \tag{122}$$

For $M \geq 1$, the asymptotic expansion Eq. (III, 55) of $Y_2(Mu)$ is very accurate. Taking only the first term of the expansion, we find

$$\langle Q\rangle = n[U - (\pi^2 M)^{-1}]. \tag{123}$$

This shows that the effect on $\langle Q\rangle$ of taking M even as small as 1 or 2 is quite unimportant and is easily calculable.

We come next to the evaluation of V_Q. If we neglect end effects of relative order n^{-1}, we may simplify Eq. (112) in the following ways: (1) Replace $L(x, y)$ by $K((x - y)/M)$, where

$$K(u) = \text{Min}\,(0, \ln |u|). \tag{124}$$

(2) Suppress in each term the integration over x, replacing it by a factor $(2L/D) = n$. After these changes, the integrands in V_Q are functions only of the coordinate differences $(x - y)$, etc., and the integrands vanish only when one of these differences exceeds M. Let now V_Q^0 be the value of V_Q when $M = \infty$. We obtain V_Q^0 from V_Q if we replace $K(u)$ by

$$K^0(u) = \ln |u|. \tag{125}$$

The four integrals constituting V_Q^0 are identical with the terms of Eq. (III, 91), except that the factor N multiplying each term is replaced by n. But the value of Eq. (III, 91) is known and is given by Eq. (III, 88). Therefore,

$$V_Q^0 = nC, \qquad C = \tfrac{3}{2} - \tfrac{1}{8}\pi^2 = 0.2663. \tag{126}$$

It remains to estimate the difference between V_Q and V_Q^0. The first three terms of $(V_Q - V_Q^0)$ are integrals including factors such as

$$\left[\int_{x+M}^\infty + \int_{-\infty}^{x-M}\right] \times Y_4(x, y, z, w) \ln |(x - y)/M|\, dy, \tag{127}$$

which are of order $(\pi^2 M)^{-1}$ like the second term of Eq. (122). We shall not calculate these terms, which are all of the form $n(\pi^2 M)^{-1}$ multiplied by factors of the order of unity. The fourth term of $(V_Q - V_Q^0)$ is of a different kind. It may be written

$$T_4 = \tfrac{1}{2}n \iiint [\delta(x - z) - Y_2(x - z)] \times [\delta(y - w) - Y_2(y - w)] \{K[(x - y)/M]K[(z - w)/M] - K_0[(x - y)/M]K_0[(z - w)/M]\}\, dy\, dz\, dw. \tag{128}$$

In this term there is no Y function which decreases as the separations $(x - y)$ and $(z - w)$ become large. But it is easy to evaluate Eq. (128) explicitly. We write

$$\tilde{K}(p) = \int_{-\infty}^\infty e^{2\pi i p u} K(u)\, du, \tag{129}$$

$$\tilde{K}_0(p) = \int_{-\infty}^\infty e^{2\pi i p u} K_0(u)\, du, \tag{130}$$

for the Fourier transforms of $K(u)$ and $K_0(u)$. Then

$$T_4 = \tfrac{1}{2}nM^2 \int_{-\infty}^\infty [1 - b(p)]^2 \times [(\tilde{K}(Mp))^2 - (\tilde{K}_0(Mp))^2]\, dp, \tag{131}$$

by virtue of Eq. (III, 14). Now we have, for large values of q,

$$\tilde{K}(q) = -[2\,|q|]^{-1} + [2\pi^2 q^2]^{-1} \cos(2\pi q) + \cdots, \tag{132}$$

while

$$\tilde{K}_0(q) = -[2\ |q|]^{-1} \tag{133}$$

precisely. Thus the difference $[(\tilde{K})^2 - (\tilde{K}_0)^2]$ in Eq. (131) is of order $(Mp)^{-3}$ for large values of Mp. Hence the important part of the integral comes from small p, for which we may use the approximation

$$1 - b(p) = 2\ |p|. \tag{134}$$

An elementary calculation then gives the result

$$T_4 = n(\pi^2 M)^{-1}. \tag{135}$$

So this term is of the same order as the other three contributions to $(V_Q - V_Q^0)$. Note that if we had chosen $f(x, y)$ in such a way that the product $L(x, y)$ had discontinuities, the term T_4 would have been proportional to

$$n \int [1 - b(p)]^2 p^{-2}\, dp, \tag{136}$$

with no factor $(\pi^2 M)^{-1}$. Just by making $L(x, y)$ continuous at $|x - y| = M$, we have avoided the serious additional statistical scatter in Q which would arise from "edge effects" if the discontinuities in $f(x, y)$ had not been cancelled by the vanishing of the logarithmic factor.

To summarize the results so far, the statistic Q defined by Eqs. (109) and (115)–(119) has the expectation value (123) and the variance

$$V_Q = n[C + \theta(\pi^2 M)^{-1}], \tag{137}$$

where θ is a coefficient of the order of unity.

The choice of the parameter M must be a compromise between convenience and accuracy. Since the labor of computing Q is proportional to M, it is clearly desirable to keep M as small as possible. Eq. (137) indicates the price that must be paid in additional statistical scatter if M is made too small. In most practical circumstances it would seem that $M = 2$ would be a reasonable choice.

We discuss lastly the effects of applying the statistic Q to a multiple level series. The effects are exactly the same as those which were seen for the statistic W in Sec. VI of III. Equations (123) and (137) become, for a multiple series,

$$\langle Q \rangle = n[U - (\pi^2 M)^{-1} - \tfrac{1}{2} \sum_\mu f_\mu (\ln f_\mu)], \tag{138}$$

$$V_Q = n[C + \theta(\pi^2 M)^{-1} - \sum_\mu f_\mu (\ln f_\mu)]. \tag{139}$$

Again Eq. (138) is exact, whereas the correction term in Eq. (139) is an approximation similar to Eq. (III, 96). If Q is computed for a given level series and the result compared with Eq. (138), we obtain a measurement of the quantity $[\sum_\mu f_\mu \ln f_\mu]$ with an error of the order of $n^{-\frac{1}{2}}$. This information concerning the mixing ratios in the series is much more precise than that provided by the statistic Δ_3 in Sec. III. The statistic Q gives also a much more sensitive test to decide whether our basic theoretical hypothesis has any relevance to real nuclei.

V. COMPARISON WITH OBSERVED SERIES OF NEUTRON-CAPTURE LEVELS

We have discussed three types of statistic which might be useful in the analysis of observational data. They were denoted by W, Δ, and Q, and their theoretical properties were described in Secs. II, III, and IV, respectively. The statistic W plays a minor role, since its function is only to give an accurate estimate of the mean level density. The other two statistics, Δ and Q, are designed to test the validity of our theoretical model. In particular, Δ (in any of its three variant forms Δ_1, Δ_2, or Δ_3) measures the extent to which the large-scale regularity or long-range order of the level series agrees with the predictions of the model, while Q measures the extent to which the local fluctuation or short-range disorder agrees with the theory. The fact that Δ emphasizes the long-range order is obvious from its definition. The fact that Q emphasizes the short-range disorder can be understood by examining Eqs. (109) and (115); the important contributions to Q come from the first term on the right-hand side of Eq. (109), and this term is precisely a sum over pairs of neighboring or nearly neighboring levels.

In the present section we make a very preliminary analysis[3] of a small fraction of the available data on neutron-capture levels, using the statistics Δ and Q. The analysis is insufficient, both in quantity and in depth, for us to reach any reliable conclusions about the validity of our theory. Superficially, the observed values of Δ and Q disagree badly with the calculated values. However, the disagreements are of a kind that could reasonably be explained as effects of imperfections in the data combined with normal statistical fluctuations. We do not feel our-

[3] C. E. Porter and N. Rosenzweig, Suomalaisen Tiedeakat. Toimituksia AVI, No. 44 (1960); Phys. Rev. **120**, 1698 (1960), have made statistical analyses of neutron-capture levels and also of levels in complex atomic spectra. They found the observed distributions to be in agreement with the random-matrix model. Our analysis in this paper does not differ in principle from that of Porter and Rosenzweig, except that we aim to be more quantitative and precise. By being more precise we give ourselves a better chance of disproving the theoretical model.

selves competent to pass judgment on the quality of the experimental data. We therefore present the results of this preliminary and inconclusive analysis, just as an example of what can be done. We hope that others who are more familiar with the practical details of the experiments will carry the analysis further. If all the existing data were to be analyzed, it would be a simple matter to mechanize the process and carry out the computations electronically. Our calculations were all done by hand, and we can therefore give no guarantee that the results are free from computational error.

The level series which we have analyzed are those produced by neutron capture in U^{238}, Th^{232}, and Ta^{181}. These are cases in which accurate and recent data exists, and in which the capturing element is effectively monoisotopic. The capture levels in U^{238} and Th^{232} should form a single series, while those in Ta^{181} should form a double series with the spin-values 3 and 4.

In general the quality and completeness of the data will be highest when the kinetic energy of the captured neutron is lowest. Above a certain neutron energy, the resolution of the time-of-flight spectrometer becomes so poor that the weaker levels will not be recorded. We therefore analyze the data only over a restricted range of energy within which there seems to be a good chance of detecting all the existing levels. In practice the upper limit of the "good" energy range is chosen to be that point at which the staircase graph of $N(E)$ against E, as described in Sec. III, stops being approximately linear and begins to curve systematically downward. Of course this choice of the upper limit in no way guarantees that some levels are not being missed within the analyzed range.

For U^{238}, we analyzed 55 levels observed by the Columbia group[4] between 0 and 1 keV, supplemented with two additional levels reported at Harwell[5] to exist with energies of 755 and 809 eV. This series represents some of the cleanest and most accurate of all existing data. The fact that even here the existence of two levels is doubtful indicates how cautious one must be in assuming any observed level series to be complete.

For Ta^{181}, we analyzed 68 levels observed[6] between 0 and 334 eV. The small mean level spacing gives a good opportunity for analyzing a large number of levels in a range of energies where the instrumental resolution is high. The comparison between U^{238} and Ta^{181} also enables us to check the predictions of the theory concerning the difference in statistical behavior between single and double series.

TABLE I. Summary of results of the analysis of observed level series.

Nucleus	U 238	Ta 181	Th 232
Number of levels	57	68	154
Energy range	0–1 keV	0–334 eV	1–4 keV
Δ_1 Observed	1.778	3.411	3.265
Δ_1 Theory	0.84 ± 0.58	1.53 ± 1.24	1.04 ± 0.72
Δ_2 Observed	1.299	1.443	8.717
Δ_2 Theory	0.53 ± 0.15	0.93 ± 0.31	0.66 ± 0.18
Δ_3 Observed	1.278	1.437	3.123
Δ_3 Theory	0.40 ± 0.11	0.87 ± 0.22	0.50 ± 0.11
Q Observed	14.62	74.1	61.17
Q Theory	18.2 ± 3.9	51.4 ± 8.6	52.4 ± 6.4

For Th^{232}, we analyzed 154 levels observed[7] between 1 and 4 keV. The data in this energy range is inevitably of poorer quality than the U^{238} data which we have analyzed. For example, there are 80 observed levels between 1 and 2.5 keV, and 74 between 2.5 and 4 keV; according to the theory of Sec. II, Eq. (21), these two numbers should differ by 1 or 2 units at most if all existing levels were observed. The analysis of the Th^{232} data thus provides an experimental check on the sensitivity of the statistics Δ and Q to imperfections in the data. Presumably the underlying theory would be equally applicable to U^{238} and to Th^{232} if the data were perfect.

The results of the analysis of the three series are summarized in Table I. The limits of error quoted for the theoretical numbers are root-mean-square deviations. The theoretical numbers for Ta^{181} are for a double series with mixing ratio $f = \frac{1}{2}$; the numbers would change very little for any value of f between $\frac{1}{4}$ and $\frac{3}{4}$. The values of Q are computed with the parameter R chosen to be 40 eV, 10 eV, and 80 eV, for U^{238}, Ta^{181}, and Th^{232}, respectively. The theoretical value of Q for Ta^{181} considered as a single series would have been 24.7 ± 4.5.

The reader is free to draw whatever conclusions he pleases from these numbers. We make only the following explanatory remarks:

(a) Even in "good" data such as the U^{238} series, there remain two important possible sources of experimental error. First, some of the weakest ob-

[4] J. L. Rosen, J. S. Desjardins, J. Rainwater, and W. W. Havens, Jr., Phys. Rev. 118, 687 (1960).
[5] F. W. K. Firk, J. E. Lynn, and M. C. Moxon, Rept. Intern. Conf. Nucl. Structure, Harwell, England (1960).
[6] J. S. Desjardins, J. L. Rosen, W. W. Havens, Jr., and J. Rainwater, Phys. Rev. 120, 2214 (1960).
[7] W. W. Havens, Jr. (private communication).

served levels may be formed by capture of p-wave neutrons; these levels belong to separate series uncorrelated with the main s-wave series. Second, some of the real s-wave levels may be unobserved because they happen to have exceptionally small widths.[8]

(b) Both the kinds of errors described in (a) will have a drastic effect in increasing the values of Δ, but will have a much smaller effect in increasing Q. The point here is that a single missing level or a single spurious level can double Δ by throwing the long-range order of the series into disarray. On the other hand, Q is a measure of local or short-range disorder, and a fraction x of false data will increase Q only by a fraction roughly equal to x.

(c) The discrepancies between observed and calculated values are very large for Δ, but moderately small for Q. Further, the discrepancies in Δ are much the worst in the case of Th^{232} where the quality of the data is poorest.

(d) As a result of remarks (a), (b), and (c), we can say that the discrepancies which have been found are qualitatively consistent with what one would expect to arise from imperfections in the data, if the underlying theoretical model were correct. On the other hand, the discrepancies are certainly consistent with the theoretical model being incorrect.

[8] Very informative theoretical discussions of the difficulties of identifying weak levels as due to s- or p-wave neutrons are to be found in references 4 and 6.

VI. CONCLUSION

We would be very happy if we could report that our theoretical model had been strikingly confirmed by the statistical analysis of neutron-capture levels. We would be even happier if we could report that our theoretical model had been decisively contradicted. In the second case, the observations would have demonstrated that a purely statistical description is inadequate for these highly excited nuclear states; we should then be faced with the pleasant task of discovering the unknown symmetries or quantum numbers in terms of which a nonstatistical description should be defined.

Unfortunately, our model is as yet neither proved nor disproved. We can only end by appealing to the experimenters to persevere and improve the already excellent quality of their data. Until we have some level series which one can, with confidence, say are free from a single missing or spurious level, a decisive test of the statistical model will hardly be possible.

ACKNOWLEDGMENTS

We are grateful to Professor W. W. Havens, Jr., and his group at the Columbia University Nevis Laboratory for valuable discussions and for communication of unpublished results. One of us (M.L.M.) would like to thank Professor J. R. Oppenheimer for hospitality at the Institute for Advanced Study and for financial support.

Statistical Theory of the Energy Levels of Complex Systems. V

MADAN LAL MEHTA* AND FREEMAN J. DYSON
Institute for Advanced Study, Princeton, New Jersey
(Received 21 January 1963)

This paper is divided into three disconnected parts. (i) An identity is proved which establishes an intimate connection between the statistical behavior of eigenvalues of random unitary matrices over the real field and over the quaternion field. It is proved that the joint distribution function of *all* the eigenvalues of a random unitary self-dual quaternion matrix of order N is identical with the joint distribution function of a set of N *alternate* eigenvalues of a random unitary symmetric matrix of order $2N$. A corollary of this result is the following: the distribution of spacings between next-nearest-neighbor eigenvalues in a real symmetric matrix of large order is identical with the distribution of nearest-neighbor spacings in a self-dual Hermitian quaternion matrix of large order. (ii) A conjecture is made which gives an exact analytic formula for the partition function of a finite gas of N point charges free to move on an infinite straight line under the influence of an external harmonic potential. This conjecture is at the same time a statement about the statistical properties of the eigenvalues of Hermitian matrices whose elements are Gaussian random variables. (iii) A list is made of several other problems which remain unsolved in the statistical theory of eigenvalues of random matrices.

I. INTRODUCTION

THIS is the last in a series of papers[1] concerned with the statistical behavior of the energy levels of complex systems. Our desire to tidy up and finish our work has resulted in a paper whose contents are rather miscellaneous. We have to report one definitive theorem (Sec. II), one conjecture which we are unable to prove (Sec. III), and finally a list of unsolved problems (Sec. V) on which we have worked long enough to ascertain that they are interesting and nontrivial. Since we decided to leave this field, we considered it important not to give the false impression that we have answered all the outstanding questions. It is our hope that the list of unsolved problems will stimulate others to begin where we have left off, so that the study of this singularly beautiful branch of mathematical physics may continue without interruption.

In paper I, three types of ensembles of random unitary matrices were defined. These were named orthogonal, unitary, and symplectic, because of their relation to the corresponding classical groups; they are characterized by the three values 1, 2, and 4 of a parameter β. In paper TW, it was shown how each of these three ensembles may be relevant to the description of a complex system, depending on the symmetry group of the system and on its behavior under the operation of time reversal. The orthogonal ensemble ($\beta = 1$) is most frequently relevant in practice and was discussed in detail in papers II, III, and IV. The unitary ensemble ($\beta = 2$) is almost trivial from a mathematical point of view and is perhaps never relevant in practice. Section II of this paper is devoted to a brief study of properties of the symplectic ensemble ($\beta = 4$). Our main theorem states that the series of all eigenvalues of a matrix in the symplectic ensemble of order N is statistically equivalent to the series of alternate eigenvalues of a matrix in the orthogonal ensemble of order $2N$. Thus many statistical properties of the symplectic ensemble are easily reducible to properties of the orthogonal ensemble which were computed in paper III.

Sections III–IV of this paper are concerned with the eigenvalues of Gaussian random matrices and with the corresponding Coulomb gas analogs. A number of properties of the Gaussian matrix ensembles are proved rigorously, and others are conjectured. The main conjecture is a formula for the partition function of a finite Coulomb gas on a straight conducting wire with a harmonic potential. A similar conjecture was previously made for the finite Coulomb gas on a circle with zero potential (conjecture A of paper I); in that case the conjecture was proved by Wilson[2] and by J. Gunson (unpublished). We hope that the new conjecture will yield equally fast to our colleagues' mathematical skill.

The final Sec. V of this paper is self-explanatory, consisting of a list of topics and questions which we

* On leave of absence from Tata Institute of Fundamental Research, Bombay, India.

[1] F. J. Dyson, J. Math. Phys. **3**, 140, 157, 166, 1191, and 1199 (1962), quoted in what follows as I, II, III, BMM and TW respectively; F. J. Dyson and M. L. Mehta, J. Math. Phys. **4**, 701 (1963) is quoted as IV.

[2] K. Wilson, J. Math. Phys. **3**, 1040 (1962).

did not succeed in elucidating and which we think will repay further development.

II. THE SYMPLECTIC ENSEMBLE

Let the eigenvalues of a random unitary matrix of rank N be given by $\exp[i\theta_j]$, $j = 1, \cdots, N$. The joint probability distribution of the angles θ_i was calculated in Sec. V of I, and is

$$Q_{N\beta}(\theta_1, \cdots, \theta_N) = C_{N\beta} \prod_{i<k} |\exp(i\theta_i) - \exp(i\theta_k)|^\beta, \quad (1)$$

with

$$C_{N\beta} = (2\pi)^{-N}\{\Gamma(1 + \tfrac{1}{2}\beta)\}^N\{\Gamma(1 + \tfrac{1}{2}N\beta)\}^{-1}. \quad (2)$$

Here β is the parameter which takes the values 1, 2, or 4, according as the random unitary matrix is chosen from the orthogonal, the unitary, or the symplectic ensemble.

When $\beta = 4$, the product on the right of Eq. (1) can be expressed as a determinant,[3]

$$Q_{N4} = C_{N4} \det[\exp(ip\theta_i), p\exp(ip\theta_i)], \quad (3)$$

with $2N$ rows and columns. The index j labels a pair of columns and takes values from 1 to N, while the row index p takes the $2N$ half-integer values

$$p = -N + \tfrac{1}{2}, \quad -N + \tfrac{3}{2}, \cdots, N - \tfrac{3}{2}, \quad N - \tfrac{1}{2}. \quad (4)$$

The expression (3) is symmetric in the angles θ_i. There is no loss of generality in supposing the θ_i to be labeled in ascending order, so that

$$0 < \theta_1 < \theta_2 < \cdots < \theta_N < 2\pi. \quad (5)$$

Let $R_{N1}(\theta_1, \cdots, \theta_N)$ be the probability density for the $[\exp(i\theta_i)]$ to appear as a set of alternate eigenvalues of a random unitary matrix of rank $2N$ chosen from the orthogonal ensemble. Then

$$R_{N1}(\theta_1, \cdots, \theta_N) = [(2N)!/2\cdot N!] \times \int_{\theta_1}^{\theta_2} d\phi_1 \int_{\theta_2}^{\theta_3} d\phi_2 \cdots \int_{\theta_N}^{\theta_1+2\pi} d\phi_N \times Q_{2N,1}(\theta_1, \phi_1, \theta_2, \phi_2, \cdots, \theta_N, \phi_N). \quad (6)$$

The angles appear in Eq. (6) in increasing order, and therefore the value of $Q_{2N,1}$ obtained from Eq. (1) is a simple alternant determinant. Thus Eq. (6) becomes

$$R_{N1} = [(2N)!/2\cdot N!] \int \cdots \int d\phi_1 \cdots d\phi_N \times C_{2N,1}(-i)^N \det[\exp(ip\theta_i), \exp(ip\phi_i)]. \quad (7)$$

The integration over the ϕ_i can be explicitly performed, and gives the result

$$R_{N1} = [(2N)!/2\cdot N!]C_{2N,1}(-i)^N \det[\exp(ip\theta_i), (-ip^{-1})\{\exp(ip\theta_{i+1}) - \exp(ip\theta_i)\}], \quad (8)$$

with the convention

$$\theta_{N+1} = \theta_1 + 2\pi. \quad (9)$$

With the help of a little algebra, Eq. (8) is easily reduced to identity with Eq. (3). The following result has then been proved.

Theorem:

$$Q_{N4}(\theta_1, \cdots, \theta_N) \equiv R_{N1}(\theta_1, \cdots, \theta_N). \quad (10)$$

Or in words, the probability distribution of a set of N alternate eigenvalues of a matrix in the orthogonal ensemble of order $2N$ is identical with the probability distribution of the set of all eigenvalues of a matrix in the symplectic ensemble of order N.

This theorem has the consequence that all statistical properties of eigenvalues in the symplectic ensemble are deducible from properties of eigenvalues in the orthogonal ensemble. A similar theorem, relating the properties of the unitary and orthogonal ensembles, has been conjectured in Sec. VII of III and subsequently proved by Gunson.[4] The two theorems together imply that statistical properties of all three types of ensemble are reducible to properties of the orthogonal ensemble alone. It is curious that two such similar theorems should exist, without any obvious or logical connection between them. The proofs of both are mere verifications, giving no hint of deeper causes or physical principles which might underlie them.

We next summarize the facts concerning level-spacing distributions in the three types of ensemble. For simplicity, consider an infinitely long level series with mean spacing $D = 1$. The probability distribution for a spacing x between nearest-neighbor levels is given by

$$S_\beta(x) = [d^2/dx^2]E_\beta(x), \quad (11)$$

where $E_\beta(x)$ is the probability that a randomly chosen interval of length x is empty of levels. As usual, $\beta = 1$, 2, or 4 for the three types of matrix ensemble. The functions $E_\beta(x)$ are calculable in terms of the two quantities

[3] The determinant (3) is called a "confluent alternant," since it is obtained as a limiting form of the simple alternant Det $[\exp(ip\phi_q)]$ when the ϕ_q become equal in pairs. For a proof that the confluent alternant is equal to Eq. (1) with $\beta = 4$, see H. W. Segar, Messenger of Mathematics 22, 57 (1893).

[4] J. Gunson, J. Math. Phys. 3, 752 (1962).

$$E(x) = \prod_i (1 - x\lambda_i^2), \tag{12}$$

$$E'(x) = \prod_i (1 - x\mu_i^2), \tag{13}$$

where the λ_i and μ_i are, respectively, the eigenvalues of the integral equations

$$\lambda F(y) = \int_0^1 \cos(\tfrac{1}{2}\pi xyz) F(z)\, dz, \tag{14}$$

$$\mu F(y) = \int_0^1 \sin(\tfrac{1}{2}\pi xyz) F(z)\, dz. \tag{15}$$

Namely, we have

$$E_1(x) = E(x), \tag{16}$$

$$E_2(x) = E(x)E'(x), \tag{17}$$

$$E_4(x) = \tfrac{1}{2}[E(2x) + E'(2x)]. \tag{18}$$

Equation (16) was discovered by Gaudin.[5] Equation (17) is a consequence of Gunson's Theorem.[4] Equation (18) is a consequence of the theorem stated above [Eq. (10)].

Gaudin[5] has carried through a numerical calculation of the functions $E(x)$ and $S_1(x)$. More recently Kahn[6] has adapted Gaudin's method to compute $E'(x)$, $S_2(x)$, and $S_4(x)$. Thus all three spacing distributions are now accurately known.

The two-level correlation function for the symplectic ensemble can also be obtained immediately from Eq. (10), using the results obtained for the orthogonal ensemble in Sec. III of III. Let the probability density for finding eigenvalues x_1 and x_2 in a symplectic level series, irrespective of the positions of other levels, be

$$1 - Y_{2s}(|x_1 - x_2|). \tag{19}$$

Then the function Y_{2s} is given by

$$Y_{2s}(r) = \{q(r)\}^2 - \{(d/dr)q(r)\}\left\{\int_0^r q(x)\, dx\right\}, \tag{20}$$

with

$$q(r) = [\sin(2\pi r)]/(2\pi r). \tag{21}$$

It is easy now to verify the correctness of the remarks made at the end of Sec. V of III.

III. COULOMB GAS WITH HARMONIC POTENTIAL

Consider a gas of N point charges with positions $[x_1, \cdots, x_N]$, free to move on the infinite straight line $[-\infty < x < +\infty]$. Suppose that the energy of the gas is given by

$$W = -\sum_{j<k} \ln|x_j - x_k| + \tfrac{1}{2}\sum_j x_j^2. \tag{22}$$

The first term in W represents an electrostatic repulsion between each pair of charges, while the second term represents a harmonic potential attracting each charge independently toward the point $x = 0$. In Sec. Vb of II it was mentioned that this Coulomb gas provides an exact model for the statistical behavior of the eigenvalues of a matrix of order N chosen from a Gaussian ensemble. Many properties of the Coulomb gas are implicitly contained in earlier papers[7] in which the eigenvalue distributions of Gaussian ensembles were studied.

The distribution of eigenvalues of a matrix chosen from a suitably normalized Gaussian ensemble is given by

$$P_{N\beta}(x_1, \cdots, x_N) = C_{N\beta}[\exp(-\beta W)], \tag{23}$$

$$C_{N\beta} = (2\pi)^{-\frac{1}{2}N}[\Psi_N(\beta)]^{-1}, \tag{24}$$

$$\Psi_N(\beta) = (2\pi)^{-\frac{1}{2}N} \times \int \cdots \int_{-\infty}^{\infty} [\exp(-\beta W)]\, dx_1 \cdots dx_N. \tag{25}$$

Here $\beta = 1$, 2 or 4 for the three types of ensemble, the case $\beta = 1$ being the usual one which is applicable to real symmetric matrices. The same equations (23)–(25) describe the distribution of the charges in the Coulomb gas in thermal equilibrium at a temperature given by $kT = \beta^{-1}$.

Many of the statistical properties of the eigenvalue distributions will be determined by the partition function $\Psi_N(\beta)$ and its derivatives with respect to β. It is therefore of interest to find an exact analytical expression for $\Psi_N(\beta)$.

Conjecture D: For every integer N and real or complex β, we have identically

$$\Psi_N(\beta) = \beta^{-\frac{1}{2}N - \frac{1}{4}N(N-1)\beta}\{\Gamma(1 + \tfrac{1}{2}\beta)\}^{-N} \times \left\{\prod_{g=1}^{N} \Gamma(1 + \tfrac{1}{2}\beta g)\right\}. \tag{26}$$

This conjecture appears very similar to Conjecture A of I, which was subsequently proved by Gunson and by Wilson.[2] However, we have not succeeded in adapting Wilson's or Gunson's method to yield a proof of Eq. (26). Some new trick seems to be needed.

Again the evidence in favor of the truth of Conjecture D is overwhelmingly strong. First, it is possible to verify Eq. (26) directly for the physical values 1, 2, and 4 of β. The case $\beta = 1$ has been

[5] M. Gaudin, Nucl. Phys. 25, 447 (1960).
[6] Peter B. Kahn, "Energy-Level Spacing Distributions," Brookhaven preprint, No. 6392.
[7] M. L. Mehta, Nuclear Phys. 18, 395 (1960); M. L. Mehta and M. Gaudin, *ibid.* 18, 420 (1960).

verified more than once.[8] The case $\beta = 2$ is the easiest; one needs only to introduce Hermite polynomials and exploit their orthogonality properties. In the case $\beta = 4$, the product $\prod (x_i - x_k)^4$ can be written as a confluent alternant, similar to Eq. (3), but involving Hermite polynomials instead of exponentials, and again the properties of Hermite polynomials lead to a proof of Eq. (26).

Conjecture D has also been verified for general β in the cases $N = 1, 2, 3$. The calculation is trivial for $N = 1$ or 2. The verification for $N = 3$ is by no means trivial, but it has been carried through by Gaudin.[9] Gaudin's method is to transform the right-hand side of Eq. (25) into an integral over the symmetric variables

$$y_1 = x_1 + x_2 + x_3,$$
$$y_2 = x_2x_3 + x_3x_1 + x_1x_2,$$
$$y_3 = x_1x_2x_3,$$

and to use the properties of the successive discriminants of an algebraic equation.

Finally we are able to prove, as in Sec. VIII of I, that Conjecture D must hold for all complex β if it holds for $\beta = 2k$ with k a positive integer. The argument rests on the fact that the energy W given by Eq. (22) is bounded below. More precisely,[10]

$$W \geq W_0 = \tfrac{1}{4}N(N-1)[1 + \ln 2] - \tfrac{1}{2}\sum_{g=1}^{N} g \ln g, \qquad (27)$$

the minimum being attained when the x_i are at the zeros of the Nth Hermite polynomial $H_N(x)$. Thus Eq. (25) may be written

$$\Psi_N(\beta) = \int_0^Y P(y)y^\beta\, dy, \qquad Y = \exp(-W_0), \qquad (28)$$

where $P(y)$ is a positive weight function. It is also easy to verify that the right-hand side of Eq. (26) is bounded by $C\,|Y^\beta|$ for any complex β in the half-plane (Re $\beta > 0$). Therefore Carlson's Theorem can be applied as in Sec. VIII of I. We conclude that Conjecture D is equivalent to the following apparently weaker statement,

$$(2\pi)^{-\frac{1}{2}N}\int \cdots \int_{-\infty}^{\infty} dx_1 \cdots dx_N \times \{\prod_{i<j}(x_i - x_j)^{2k}\} \exp(-k\sum x_i^2) = (2k)^{-\frac{1}{2}N(1+(N-1)k)}\{k!\}^{-N}\left\{\prod_{g=1}^{N}(kg)!\right\}. \qquad (29)$$

It is possible to go further than this and reduce Conjecture D to a finite algebraic identity. For this purpose we introduce the notation

$$D_i = [(\partial/\partial x_i)]_{x_i=0}, \qquad (30)$$

with the understanding that all differentiations are to be carried out before the variables x_i are set equal to zero. The identity

$$(k\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty} x_i^n \exp(-kx_i^2)\, dx_i = \{i\, D_i/(2k)^{\frac{1}{2}}\}^n \exp(-\tfrac{1}{2}x_i^2) \qquad (31)$$

enables us to replace all integrations in Eq. (29) by differentiations at the point $x_i = 0$. Eq. (29) then takes the form

$$\{\prod_{i<j}(D_i - D_j)^{2k}\} \exp(-\tfrac{1}{2}\sum x_i^2) = (-1)^{\frac{1}{2}N(N-1)k}\{k!\}^{-N}\left\{\prod_{g=1}^{N}(kg)!\right\}. \qquad (32)$$

Since only one term in the exponential series contributes, Eq. (32) is a finite combinatorial identity. A somewhat neater form of the identity is obtained by interchanging the roles of D_i and x_i. We display this form of Eq. (32) as

Conjecture E: The identity

$$\left\{\frac{1}{2}\sum_{i=1}^{N}(\partial^2/\partial x_i^2)\right\}^{\frac{1}{2}N(N-1)k}\{\prod_{i<j}(x_i - x_j)^{2k}\} = \{\tfrac{1}{2}N(N-1)k\}!\,\{k!\}^{-N}\left\{\prod_{g=1}^{N}(kg)!\right\} \qquad (33)$$

holds for positive integer N and k.

Conjecture E is equivalent to Conjecture D, and may perhaps be easier to prove. In particular, Conjecture E may be more amenable to generalization. It turned out, in the case of Conjecture A in I, that the key to the proof[2] lay in the structure of the generalized Conjecture C. We have not been able to find the analog of Conjecture C for the present problem.

IV. GAUSSIAN ENSEMBLES WITH $\beta = 2$ AND $\beta = 4$

In this section we collect some results, supplementing the calculations in our earlier papers[7] which referred exclusively to the case $\beta = 1$.

Consider the ensemble of Hermitian matrices M, invariant under unitary transformations and having

[8] M. L. Mehta, Nucl. Phys. **18**, 395 (1960), and other references given there. See also N. G. de Bruijn, J. Ind. Math. Soc. **19**, 133 (1955); we are very sorry that we came to know of this important reference only recently.

[9] M. Gaudin (private communication).

[10] T. J. Stieltjes, *Sur Quelques Théorèmes d'Algèbre, Oeuvres Complètes* 1, 440–441 (P. Noordhoff Ltd., Groningen, The Netherlands, 1914).

all elements, apart from Hermiticity, independent random complex variables. The joint distribution function of the eigenvalues of such a matrix is given by Eq. (23) with $\beta = 2$;

$$P_{N2}(x_1, \cdots, x_N) = C_{N2} \exp(-\sum x_i^2) \prod_{i<k} (x_i - x_k)^2. \quad (34)$$

The density of eigenvalues is given by[11]

$$N \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} P_{N2}(x, x_2, \cdots, x_N)\, dx_2 \cdots dx_N = \sum_0^{N-1} \phi_i^2(x) = N\phi_N^2(x) - 2[N(N-1)]^{\frac{1}{2}}\phi_{N-1}(x)\phi_{N+1}(x), \quad (35)$$

where $\phi_j(x)$ are the normalized harmonic oscillator wavefunctions

$$\phi_j(x) = (2^j j!\, \pi^{\frac{1}{2}})^{-\frac{1}{2}} e^{\frac{1}{2}x^2}\left(-\frac{d}{dx}\right)^j e^{-x^2}. \quad (36)$$

The n-level correlation function is

$$R_n(x_1, \cdots, x_n) = \frac{N!}{(N-n)!} \int_{-\infty}^{\infty} \cdots \times \int_{-\infty}^{\infty} P(x_1, \cdots, x_N)\, dx_{n+1} \cdots dx_N = \det [K_N(x_j, x_k)]_{j,k=1,2,\cdots,n}, \quad (37)$$

where

$$K_N(x, y) = \sum_0^{N-1} \phi_i(x)\phi_i(y), \quad (38)$$

with ϕ_i given by Eq. (36).

The probability that there are no eigenvalues in the interval $(-\theta, \theta)$ is

$$R^{(2)}(2\theta) = \det\left[\delta_{pq} - \int_{-\theta}^{\theta} \phi_p(x)\phi_q(x)\, dx\right]_{p,q=0,1,\cdots,N-1}, \quad (39)$$

which, in the limit $N \to \infty$, yields precisely the value of the function $E_2(x)$ as given by Eqs. (17), (12), and (13).

In the symplectic Gaussian ensemble ($\beta = 4$), one may apply similar methods. Thus the density of eigenvalues at the origin is

$$N \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\exp(-2 \sum x_i^2) \times \prod_{i<k} (x_i - x_k)^4]_{x_1=0}\, dx_2 \cdots dx_N = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \sum_{j=0}^{[\frac{1}{2}(N-1)]} \frac{(2N-2-4j)!}{2^{N-1-2j}\{(N-1-2j)!\}^2}, \quad (40)$$

where $[\frac{1}{2}(N-1)]$ is $\frac{1}{2}(N-1)$ or $(\frac{1}{2}N - 1)$ according to whether N is odd or even. As $N \to \infty$, the level density at the origin becomes

$$D^{-1} \sim \pi^{-1}(2N)^{\frac{1}{2}}, \qquad N \gg 1. \quad (41)$$

For the probability that there is no eigenvalue in the interval $(-\theta, \theta)$, one obtains

$$R^{(4)}(2\theta) = \chi_1(\theta\sqrt{2}) + \chi_2(\theta\sqrt{2}), \quad (42)$$

where

$$\chi_1(\tau) = \det\left[\delta_{pq} - \int_{-\tau}^{\tau} \phi_{2p-1}(x)\phi_{2q-1}(x)\, dx\right] \qquad p, q = 1, 2, \cdots, N, \quad (43)$$

and

$$\chi_2(\tau) = -\int_{\tau}^{\infty} dx \times \det\begin{bmatrix} 0 & \phi_{2q-1}(x) \\ \phi_{2p-1}(x) & \delta_{pq} - \int_{-\tau}^{\tau} \phi_{2p-1}(y)\phi_{2q-1}(y)\, dy \end{bmatrix} \qquad p, q = 1, 2, \cdots, N. \quad (44)$$

The function $\chi_1(\tau)$ is the Fredholm determinant of an integral equation with the symmetric kernel

$$\sum_{p=1}^{N} \phi_{2p-1}(x)\phi_{2p-1}(y); \qquad |x|, |y| < \tau,$$

and therefore in the limit $N \to \infty$ becomes[5] identical with $E'(2x)$ defined by Eq. (13). The parameter x is here

$$x = (S/D) = (2\theta/D) = (2\tau N^{\frac{1}{2}}/\pi), \quad (45)$$

by virtue of Eq. (41). To make Eq. (42) reduce to Eq. (18) in the limit $N \to \infty$, we should have

$$\chi_2(\tau) \to \tfrac{1}{2}[E(2x) - E'(2x)], \qquad N \to \infty. \quad (46)$$

We have not succeeded in verifying Eq. (46) directly.

V. OPEN QUESTIONS

In this section we shall make a list of nine unsolved problems which we consider interesting.

a. *Asymptotic form of the level-spacing distributions for large spacings.* The analytical expressions given by Eqs. (11)–(18) for the spacing distributions are due essentially to Gaudin.[5] These elegant expressions are well suited to numerical calculation of $S_\beta(x)$ when x is of the order of unity. However, there remains the problem of obtaining an accurate evalua-

[11] This result is contained in an unpublished manuscript of Professor E. P. Wigner. We are grateful to Professor Wigner for permission to quote from his unpublished work.

tion of $S_\beta(x)$ for large x. This problem was discussed in greater detail in paper II.

b. *The $L - I$ problem*: This problem also arises out of the work of Gaudin.[5] The integral equations (14) and (15) may be combined into the form

$$\gamma f = If, \tag{47}$$

where I is the integral operator defined by

$$If(y) = \int_{-1}^{1} \exp(ityz) f(z)\, dz, \tag{48}$$

and γ is one of the eigenvalues of I. Gaudin introduced the differential operator

$$L = (z^2 - 1)(d/dz)^2 + 2z(d/dz) + t^2 z^2, \tag{49}$$

operating on a function $f(z)$ defined on the interval $-1 \leq z \leq 1$, with suitable boundary conditions at the endpoints. Gaudin observed that the two operators L and I commute. They therefore have a complete set of eigenfunctions in common.

Now the operator L is well-known, and its eigenvalues and eigenfunctions have been studied and tabulated.[5] However, according to Eqs. (12) and (13), for the level-spacing distributions we require the eigenvalues of I and not those of L. The fact that I and L commute means that they have the same eigenfunctions, but this fact does not imply any particular connection between their eigenvalues.

Since L has a nondegenerate spectrum, and I commutes with L, I must be a function of L,

$$I = \phi(L), \tag{50}$$

and the eigenvalues γ_j of I are then given immediately by

$$\gamma_j = \phi(L_j), \tag{51}$$

where the L_j are the known eigenvalues of L. The unsolved problem is to determine, from the definitions (48) and (49), *which* function ϕ makes Eq. (50) hold as an operator identity. If this problem were solved, it would probably also throw new light on problem a.

c. *Identities Between Cluster Functions.* In Eq. (12) of III, the following relation between cluster functions was stated as true:

$$\int_{-\infty}^{\infty} Y_n(x_1, \cdots, x_n)\, dx_n = (n - 1) Y_{n-1}(x_1, \cdots, x_{n-1}). \tag{52}$$

These cluster functions are similar to the Ursell–Mayer functions which are commonly used in the kinetic theory of gases.[12] Now Dr. A. Lenard has pointed out to us that Eq. (52) certainly does *not* hold for a normal gas whose molecules interact only with short-range forces. In particular, for a perfect gas we have $Y_2 = 0$, $Y_1 \neq 0$, and Eq. (52) already fails for $n = 2$. Our "proof" of Eq. (52) involved an illegal interchange of limits and is quite worthless.

In the case of the one-dimensional Coulomb gas which was studied in paper III, we have checked Eq. (52) for several values of n by computing the cluster functions explicitly. In each case which we examined, Eq. (52) holds. Thus the deductions which were made from Eq. (12) in paper III are presumably correct. However, a rigorous proof of Eq. (52) even for the Coulomb gas is still lacking. We make a conjecture that the relations (52) will hold for a gas in any number of dimensions, if and only if the gas is incompressible. By incompressible, we mean simply that the velocity of propagation of sound waves in the gas becomes infinite as the frequency tends to zero. A Coulomb gas with charges all of the same sign is, in this sense, always incompressible.

Problem c is to prove this conjecture, or more generally, to find necessary and sufficient conditions for Eq. (52) to hold. It seems likely that the solution of this problem may give rise to a new branch of kinetic theory, namely the statistical mechanics of incompressible fluids.

d. *Proof of the Gaussian Conjecture.* This is the proof of Conjecture D or Conjecture E, as discussed in Sec. III of this paper.

e. *Ensembles of Asymmetric Matrices.* Suppose that M is an $(N \times N)$ matrix in which each element is an independent, real, Gaussian random variable, without any condition of symmetry. The eigenvalues of M will be N complex numbers Z_j. What can be said about the distribution of the Z_j in the complex plane when N is large?

The particular interest of this problem lies in the fact that it is completely untouched by all the methods that have been successfully applied to the study of symmetric matrices. For example, Wigner's original method[13] of determining the distribution of eigenvalues was to calculate the moments of the distribution function. When the eigenvalues are real, the moments determine the distribution uniquely. However, when the eigenvalues are complex, the knowledge of the moments gives no useful information about the distribution. In particular, any distribution with circular symmetry about the origin in the complex plane will have all its moments zero.

A more general way of formulating the asymmetric

[12] J. E. Mayer and M. G. Mayer, *Statistical Mechanics*, (John Wiley & Sons, Inc., New York, 1940).

[13] E. P. Wigner, Ann. Math. **62**, 548 (1955).

matrix problem is the following: Is it possible to define any reasonable ensemble of asymmetric matrices for which the distribution of eigenvalues in the complex plane can be calculated?

f. *The Five Exceptional Lie Groups.* In paper I we have defined matrix ensembles which were called orthogonal, unitary, and symplectic. These ensembles are in precise one-to-one correspondence with the classical Lie groups belonging to the three principal series.[14] Now it is well-known that there exist, besides the principal series, 5 and only 5 simple Lie groups. These are known as the exceptional Lie groups, and they have many beautiful properties.[15] We pose the problem of defining 5 exceptional matrix ensembles, related to the exceptional Lie groups in the same way as the ensembles of paper I are related to the groups of the principal series.

This problem is closely related to the question, raised in footnote 10 of TW, of fitting octonions appropriately into physics. The exceptional Lie groups are all closely connected with structures based on the octonion algebra.[15]

g. *Physical Application of the Unitary and Symplectic Ensembles.* It was proved in TW that the appropriate ensemble for describing states of a complex system is orthogonal, unitary, or symplectic, according as the corepresentation of the symmetry-group of the system to which these states belong is of Wigner type I, III, or II. Unfortunately, in any system with the full symmetry of the 3-dimensional rotation group, all co-representations are of type I, and so all states must be described by orthogonal ensembles. In particular, in the application of the theory to neutron-capture levels of heavy nuclei or to complex atomic spectra, the nuclei or atoms have rotational symmetry and only the orthogonal ensemble is relevant.

For a more searching test of the theory, it would be highly desirable to find a practical situation to which the unitary or symplectic ensembles could be applied. For this purpose one must find a physical system of sufficient complexity, having a large number of accurately observed levels, and not possessing rotational symmetry. It seems to be a difficult problem to find a system satisfying these conditions. Until this problem is solved, the study of the unitary and symplectic ensembles will remain somewhat academic.

h. *Statistical Effects of Missing and Spurious Levels.* In paper IV it was pointed out that the statistical analysis of observed series of energy levels can easily be vitiated by the effects of missing and spurious levels. Even one missing level, or one spurious level, can have a drastic effect on certain types of statistical assessment.

It would be desirable to make the results of paper IV more precise by calculating quantitatively the effects of missing and spurious levels. To carry through such calculations would not be difficult, only rather laborious. Problem h is not really an unsolved problem; it is just a job that remains to be done. When calculations of effects of missing and spurious levels have been made, it will be possible to judge with greater confidence whether the discrepancies between theory and experiment recorded in paper IV can be explained away as due to imperfections in the experimental data.

i. *Better Analysis of Experimental Data.* In many ways the weakest part of this entire investigation is the analysis in paper IV of experimental data on neutron-capture levels. The analysis is complicated, and the results are disappointingly crude. To make firmer contact between theory and experiment, we urgently need to discover better and more discriminating methods for the numerical analysis of data.

ACKNOWLEDGMENT

One of us (M.L.M.) would like to thank Professor J. R. Oppenheimer for hospitality at the Institute for Advanced Study and for financial support.

[14] H. Weyl, *The Classical Groups, Their Invariants and Representations* (Princeton University Press, Princeton, New Jersey, 1939).

[15] B. A. Rozenfel'd, Doklady Akad. Nauk S.S.S.R. **106**, 600 (1956).

Wigner–Eckart Theorem and Simple Lie Groups

JEAN GINIBRE
Laboratoire de Physique Théorique et Hautes Energies, Orsay, (Seine-et-Oise), France.
(Received 26 November 1962)

For simple Lie groups, matrix elements of vector operators (i.e. operators which transform according to the adjoint representation of the group) within an irreducible (finite-dimensional) representation, are studied. By use of the Wigner–Eckart theorem, they are shown to be linear combinations of the matrix elements of a finite number of operators. The number of linearly independent terms is calculated and shown to be at most equal to the rank l of the group. l vector operators are constructed explicitly, among which a basis can be chosen for this decomposition. Okubo's mass formula arises as a consequence.

INTRODUCTION

THE study of elementary particle interaction symmetries has been renewed in the last few months by systematic use of simple Lie groups.[1-7] One may make the following hypotheses: the interactions can be split into two parts:

A first part I_0 which is invariant under the transformations of a group G which entails the direct product of the isospin group SU_2 and the strangeness gauge group as a subgroup.

A second part I_1 which is "small" with respect to the first one and may entail a part of strong, electromagnetic and weak interactions.

If I_1 is treated as a perturbation, then:

To zero order, the elementary particles can be classified in supermultiplets which form bases for irreducible representations of G.

To first order, many interesting quantities are given by matrix elements of operators, in a given irreducible representation, the transformation properties of which are related to those of I_1. Such are the mass differences (respectively the magnetic moments) of particles of the same supermultiplet, if one keeps the non-G-invariant part of the strong (respectively electromagnetic) interactions.

If the transformation properties of I_1 are arbitrary, nothing more can be deduced from the previous hypotheses for these matrix elements. However one can use an analogy: in the similar case where I_0 is replaced by strong SU_2-invariant interactions, and I_1 by minimal electromagnetic non-SU_2-invariant interactions, we know from the relation $Q = T_3 + \frac{1}{2}(N + S)$ that the noninvariant part transforms like the third component T_3 of isospin, i.e. like a component of a tensor operator belonging to the adjoint (or regular) representation of the symmetry group.

One can generalize to the case of higher symmetry by supposing that the transformation properties of I_1 are those of some component of a tensor operator belonging to the adjoint representation of G,[1,3] hereafter called vector operators by analogy with the SU_2 case. The mass differences are then matrix elements within a given supermultiplet of some component of a vector operator. (For instance, in the usual theories with rank-2 simple Lie groups, they must be given by an operator commuting with isospin and strangeness, thus transforming like the infinitesimal generator associated with strangeness.) Therefore, they are restricted by the Wigner–Eckart theorem, which imposes relations between them. (This is similar to the case of the SU_2 group, where the matrix elements of a vector operator within an irreducible representation with given j are proportional to those of J and therefore are all expressed in terms of one single parameter.) Mass relations will appear as a consequence of the application of the Wigner–Eckart theorem to the study of matrix elements $\langle|M\sigma|\rangle$ of vector operators $M\sigma$ within an irreducible representation of G.

Here we shall restrict ourselves to the case where G is one of the simple Lie groups.

In Sec. 1, we show that the $\langle|M\sigma|\rangle$'s are linear combinations of the matrix elements of some operators. We then calculate by means of group characters, the number of linearly independent terms (Sec. II). In Sec. III, we give a general method for obtaining a basis for these operators. We apply this method

[1] M. Gell-Mann, "The Eightfold Way," Cal. Tech. Rept. 20 (March 1961); Phys. Rev. 125, 1067 (1962).
[2] R. E. Behrends, J. Dreitlein, C. Fronsdal, and W. Lee, Rev. Mod. Phys. 34, 1 (1962).
[3] S. Okubo, Progr. Theoret. Phys. (Kyoto) 27, 949 (1962).
[4] S. Okubo, Progr. Theoret. Phys. (Kyoto) 28, 24 (1962).
[5] D. Speiser and J. Tarski, J. Math. Phys. 4, 588 (1963).
[6] L. A. Radicati and D. Speiser, Nuovo Cimento 24, 385 (1962).
[7] B. d'Espagnat and J. Prentki, Nuovo Cimento 24, 497 (1962).

to the classical groups A_l, B_l, C_l, D_l in Sec. IV, and to the exceptional G_2 group in Sec. V.

1. APPLICATION OF THE WIGNER–ECKART THEOREM

The theorem may be stated as follows:

Let $\mathfrak{H}$ be a Hilbert space, and $x \to U(x)$ a unitary representation of a group G in $\mathfrak{H}$. In the space of the linear operators on $\mathfrak{H}$, it induces a representation, which to x associates the transformation $M \to U(x)MU^{-1}(x)$, (i.e. such that the matrix elements $\langle\psi_1 |M| \psi_2\rangle$ are invariant). We suppose that every representation of G is completely reducible. Suppose that ψ_1, M, and ψ_2 belong to irreducible subspaces E_1, F, and E_2 of the respective representations. The tensor product $F \otimes E_2$ may be split up into the direct sum of irreducible subspaces, H_α, some of which, H_i, are equivalent to E_1.

$$F \otimes E_2 = \sum_\alpha H_\alpha. \tag{1}$$

Let P_α be the projection of $F \otimes E_2$ on H_α, and Q_i the isomorphism of H_i onto E_1. Then

$$\langle\psi_1 |M| \psi_2\rangle = \sum_i \lambda_i \langle\psi_1 \mid Q_i P_i M \otimes \psi_2\rangle, \tag{2}$$

where the λ_i's are constants depending on E_1, F, and E_2, but not on the particular ψ_1, M, and ψ_2 chosen in E_1, F, and E_2.

Remark: The decomposition (1) is not unique if some of the H_α's are equivalent, but this obviously does not disturb the result.

We are interested in the case where $E_1 = E_2$ and F is equivalent to the adjoint representation (ρ) of G. Therefore the problem is reduced to:

decompose the direct product of (ρ) and of a given representation E into a direct sum of irreducible representations, and determine the number of terms equivalent to E,

find explicit expressions for the $Q_i P_i$, or equivalently for a sufficient number of vector operators yielding a generating system for the $\langle|M|\rangle$'s.

II. NUMBER OF TERMS IN THE DECOMPOSITION, EQ. (2)

The root diagram[8] of a simple Lie group of rank l is a set of vectors in l-dimensional Euclidian space. The planes π, perpendicular to the roots, divide that space in domains D_i. The symmetries with respect to the planes π generate a group W under which the diagram is invariant. The W-transformed v_i of a vector v belonging to no π build a finite set of vectors, and there is exactly one of them in each D_i. The dominant[8] v's of all the systems of v_i's lie in one of the D's; let it be D_1.

An irreducible representation is defined by its highest weight L or by its character[8,9]:

$$\chi(L) = \xi(L)/\xi(0) \text{ where } \xi(L) = \sum_{S \in W} \delta_S e^{i(SK)_j \varphi^j}, \tag{3}$$

with

$$K = L + R,$$

$$R = \tfrac{1}{2} \sum \alpha^+ \qquad (\alpha^+ \text{ are the positive roots}), \tag{4}$$

j labels the coordinates in l-dimensional space, φ^j are parameters which define the classes of conjugate elements in G. $\delta_S = +1$ if S is the product of an even number of reflexions, and $\delta_S = -1$ for an odd number.

We note $\chi(\rho)$ the character of the adjoint representation.

We want to decompose $E \otimes (\rho)$ into a direct sum. In terms of characters, this reads

$$\chi(L) \cdot \chi(\rho) = \sum_{L'} \chi(L'),$$

or equivalently

$$\xi(L) \cdot \chi(\rho) = \sum_{L'} \xi(L'). \tag{5}$$

We use the geometrical method given by Behrends *et al.*[2] and Speiser.[10] $\xi(L)$ is represented by the set of points SK with multiplicities δ_S, $\chi(\rho)$ by the weight diagram of (ρ), i.e. the root diagram of G, with multiplicity l at the origin, and $+1$ everywhere else. Therefore, the first member in (5) is represented by the points SK with multiplicities $l\delta_S$ and $SK + \alpha$ with δ_S, where α is any nonzero root. Each point in D_1 with multiplicity m yields m terms in the second member of (5). Therefore, in the general case where L is such that the root diagram translated by K is entirely contained in D_1, one finds:

l times the initial representation E,

once any representation with highest weight $L + \alpha$, α being any nonzero root.

$$\chi(L) \cdot \chi(\rho) = l\chi(L) + \sum_\alpha \chi(L + \alpha). \tag{6}$$

Thus the number of terms in (2) is l in general.

The argument fails if the root diagram translated by K is crossed by one of the π's which limit D_1. We then find in D_1, additional points of the type

[8] G. Racah, *Group Theory and Spectroscopy* CERN 61-8 (1961).

[9] H. Weyl, Z. Math. 23, 271 (1925); *ibid.* 24, 328, 377 (1926).

[10] D. Speiser, Lecture notes, Istambul Summer School (1962).

$SK + \alpha$ ($S \in W$, α nonzero root) with multiplicities δ_S (other points SK are certainly not in D_1). The number of terms will thus be increased (respectively decreased) by the number of pairs (S, α) such that $SK + \alpha = K$ with $\delta_S = +1$ (respectively -1), or equivalently by the number of roots α such that $K + \alpha$ be an SK with $\delta_S = +1$ (respectively -1). The number of such terms is determined in the appendix, where we prove the following result:

E being any irreducible representation with highest weight L, and (ρ) the adjoint representation, the number of terms equivalent to E in the decomposition of $E \otimes (\rho)$ into a direct sum, and therefore the number of linearly independent terms in (2) is

$$l - q \tag{7}$$

where l is the rank of the group; q is the number of vanishing coefficients in the decomposition of L on the fundamental weights.

Application to the Rank-2 Groups

The representations are defined by the two integers κ_1 and κ_2 which are the above-mentioned coefficients. One then finds two terms if $\kappa_1\kappa_2 \neq 0$ and only one if $\kappa_1\kappa_2 = 0$ (the case $\kappa_1 = \kappa_2 = 0$ is trivial). In particular, the lowest-order representations where two terms occur correspond to $\kappa_1 = \kappa_2 = 1$ i.e. (8) for A_2, (16) for B_2, and (64) for G_2.[2]

III. EXPLICIT DETERMINATION OF A BASIS

We now want to determine a set of vector operators $M^{(p)}$ providing a generating system in (2), i.e. such that the matrix elements of any vector operator M within an irreducible representation E be linear combinations of those of the $M^{(p)}$'s, with coefficients depending only on M and on E.

$$\langle \psi_1 \,|M|\, \psi_2\rangle = \sum \lambda_p(M, E)\langle \psi_1 \,|M^{(p)}|\, \psi_2\rangle, \tag{8}$$

for $\psi_1, \psi_2 \in E$.

Remark: Clearly such $M^{(p)}$'s provide a basis for the matrix elements of M between two equivalent irreducible representations E_1 and E_2, by just inserting in their matrix elements the isomorphism Q of E_2 unto E_1:

$$\langle \psi_1 \,|M|\, \psi_2\rangle = \sum \lambda_p(M, E)\langle \psi_1 \,|QM^{(p)}|\, \psi_2\rangle, \tag{9}$$

for $\psi_1 \in E_1$, $\psi_2 \in E_2$; $E_1 \simeq E_2$.

On the other hand, we discard the matrix elements of M between two inequivalent irreducible representations (which of course can be $\neq 0$).

We now closely follow Racah's method.[8] We note $\Lambda(G)$ the Lie algebra of G, g its dimension, $\{X_\sigma\}$ ($\sigma = 1, \cdots g$) a basis in $\Lambda(G)$ satisfying $[X_\rho, X_\sigma] = C^\tau_{\rho\sigma}X_\tau$, $X = a^\sigma X_\sigma$ any element of $\Lambda(G)$. In an irreducible representation of $\Lambda(G)$, X_σ goes to A_σ and X to $A = a^\sigma A_\sigma$. In particular, in the adjoint representation, X_σ goes to $E_\sigma = (C^\tau_{\rho\sigma}a^\rho)\,\partial/\partial a^\tau$. In an irreducible representation, we may look for the $M^{(p)}$'s in the form of noncommutative polynomials of the infinitesimal generators of the group:

Lemma 1. The associative algebra $\mathfrak{A}$ generated by the infinitesimal generators within an irreducible representation is the whole matrix algebra over the representation space.

Proof: $\mathfrak{A}$ is irreducible in the complex field. By Schur's lemma, its commutator algebra consists of the multiples of the unit matrix, whence the result follows.[11]

Of course we do not need all these polynomials, for clearly $[X_\rho, X_\sigma]$ and $C^\tau_{\rho\sigma}X_\tau$ yield the same operator. The noncommutative polynomials of the X's, if one ignores the commutation relations in $\Lambda(G)$, clearly build the tensor algebra over $\Lambda(G)$ and we need the quotient algebra of $\mathfrak{G}$. By the ideal I generated by the $(X_\rho \otimes X_\sigma - X_\sigma \otimes X_\rho - C^\tau_{\rho\sigma}X_\tau)$. We now prove

Lemma 2. In each coset, there is exactly one completely symmetric tensor, and it has the lowest order in its coset.

Proof: There is one means that every tensor in $\mathfrak{G}$ can be reduced to a symmetric one by use of the commutation relations: one symmetrizes the homogeneous terms of highest-order r; one is left with a symmetric part of order r (eventually zero) and terms containing commutators, therefore reducing to degree $\leq r - 1$. One then applies the same operation to the homogeneous parts of decreasing order, until one reaches $r = 1$ which is clearly symmetric. This procedure does not increase the order. There is but one means that there should be no completely symmetric tensor in I. If T is such a tensor, its highest-order homogeneous part T_0 is invariant by complete symmetrization. But $T \in I$ means that T_0 is of the form

$$\sum_{\rho,\sigma} P_{\rho\sigma}(X)(X_\rho X_\sigma - X_\sigma X_\rho)Q_{\rho\sigma}(X),$$

and it vanishes by complete symmetrization; thus $T = 0$.

Therefore we can restrict ourselves to the symmetrized polynomial operators. To such an operator

[11] H. Weyl, *Classical Groups* (Princeton University Press, Princeton, New Jersey, 1946).

$M(X)$ we can associate a polynomial $M(a)$ on $\Lambda(G)$ defined by substitution of a_σ to X_σ in $M(X)$. This correspondence Φ is clearly one to one. We are now looking for vector operators $M(X) = \{M_\sigma(X), \sigma = 1, \cdots g\}$. We call vector covariant of (ρ)[12] any set of polynomials of the a's, $M(a) = \{M_\sigma(a), \sigma = 1, \cdots g\}$, which transform according to (ρ). We prove

Lemma 3. ϕ carries a vector operator into a vector covariant (and ϕ^{-1} does the converse).

Proof. $M(X)$ is a vector operator means

$$[X_\rho, M_\sigma(X)] = C^\tau_{\rho\sigma} M_\tau(X), \tag{10}$$

or

$$[X_\rho, X_\mu]\, \partial M_\sigma/\partial X_\mu = C^\tau_{\rho\sigma} M_\tau(X), \tag{11}$$

or

$$(C^\lambda_{\rho\mu} X_\lambda)\, \partial M_\sigma/\partial X_\mu = C^\tau_{\rho\sigma} M_\tau(X), \tag{12}$$

where the first factor on the left is supposed to occupy the place of the suppressed X_μ in the second one. If X_σ is now replaced everywhere by a_σ, $M_\sigma(X)$ goes to $M_\sigma(a)$, and (12) goes to

$$E_\rho M_\sigma(a) = C^\tau_{\rho\sigma} M_\tau(a). \tag{13}$$

Therefore $M(a)$ is a vector covariant.

Conversely, if $M_\sigma(a)$ satisfies (13), the symmetrized $M_\sigma(X)$ satisfies (10). The method applies not only to scalar[8,13] and vector operators, but also to any type of tensor operators.

The problem is now to find the vector covariants of (ρ). If $M_\sigma(a)$ is one of them, and $Y = b^\sigma X_\sigma \in \Lambda(G)$, then $N = b^\sigma M_\sigma(a)$ is an invariant of (ρ), which depends on two vectors X and Y and is linear in Y. The correspondence between $M_\sigma(a)$ and N is one to one. We are thus led to consider the vector invariants of (ρ).

We already know the invariants depending on one vector.[10,14] The group G being identified with one of its faithful irreducible representations, the adjoint representation (ρ) can be defined as representing $x \in G$ by the automorphism of $\Lambda(G)$:

$$A \to xAx^{-1} \quad \text{for any} \quad A \in \Lambda(G).$$

This immediately yields the invariants

(I) $\operatorname{Tr} A^p$ (p positive integer),

(II) $\det(A - \omega)$ identically with respect to ω, and therefore all the coefficients of the various powers of ω. The two types of invariants are connected by the Newton formulas[11] which express the sums of the pth power of n variables in terms of their symmetric functions and conversely. This would be obvious for diagonalizable A's, the n variables x_i being taken as the eigenvalues of A. Then $\operatorname{Tr} A^p = \sum_i (x_i)^p$, and the symmetric functions are the coefficients of $\det(A - \omega)$. In the general case, we slightly modify the classical proof[11] of Newton's formulas. Let

$$\Delta(\lambda) = \det(1 - \lambda A), \quad \text{and} \quad C = (1 - \lambda A)^{-1}.$$

Then

$$d\Delta/d\lambda = \Delta'(\lambda) = \Delta(\lambda) \operatorname{Tr} A \cdot C;$$

but

$$C = 1 + \lambda A + \cdots + (\lambda A)^n + \cdots,$$

$$\Delta'(\lambda) = \Delta(\lambda) \times [\lambda \operatorname{Tr} A^2 + \lambda^2 \operatorname{Tr} A^3 + \cdots + \lambda^n \operatorname{Tr} A^{n+1} \cdots],$$

which yields the desired formulas by identification of the formal series of λ. (ρ) is known to have l algebraically independent invariants[14] and Racah has shown that this method yields a basis for them.[13] Let $F^{(p)}(a)$ be such a basis.

From these invariants, polarization D_{ba}[11] gives invariants of two vectors linear in the second one:

$$D_{ba} F^{(p)}(a) = (b^\sigma)\, \partial F^{(p)}(a)/\partial a^\sigma,$$

and derivation with respect to a yields a family of vector covariants

$$M^{(p)}_\sigma(a) = \partial F^{(p)}(a)/\partial a^\sigma. \tag{14}$$

Remark: $p = 1$ in (I) and the term ω^{n-1} in (II) both yield $\operatorname{Tr} A$ which is zero for any simple Lie group, for then (ρ) is irreducible and has no linear invariant.

$p = 2$ in (I) yields the invariant quadratic form $a^\sigma a_\sigma$, from which we deduce the already-known vector covariant $M^{(1)}_\sigma(a) = a_\sigma$.

It should be investigated whether or not polarization of one-vector invariants yields a basis for two-vector invariants.[15] We failed to prove it and shall admit this result, and in particular that (14) yields a basis for our vector covariants.

IV. EXPLICIT FORM OF THE $M^{(p)}$'S FOR THE CLASSICAL GROUPS A_l, B_l, C_l, D_l

A_l is the real (or unitary complex) unimodular group in $n = l + 1$ dimensions. $\Delta(\omega) = \det(A - \omega)$

[12] We use the word in a more restrictive sense than Weyl does,[11] but no confusion can occur in this paper.

[13] G. Racah, Rend. Lincei. 8, 108 (1950).

[14] E. Cartan, *Thèse* (Gauthier-Villars, Paris, 1894).

[15] For instance, in SU_2, $x_1 y_2 - x_2 y_1$ is an invariant of two vectors and is not obtained by polarization of any invariant of one vector. The same thing occurs for any skew invariant.

has degree $l+1$ and yields l invariants, namely the coefficients of the powers of ω from 0 to $l-1 = n-2$. One may take equivalently the Tr A^p for $p = 2, \cdots l+1$. We give here the result for the full linear group $Gl(n)$. Transition to the unimodular group $Sl(n)$ can be made by suitable extraction of traces. If one uses the X^λ_μ, represented by the A^λ_μ with 1 at the place $\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}$ and 0 everywhere else, as a basis for $\Lambda(G)$, and the coefficients a^μ_λ of A as corresponding coordinates, then from the Tr A^p, one obtains

$$M^{(p)\lambda}_{\mu} = SX^\lambda_\rho X^\rho_\sigma \cdots X^\tau_\mu \quad (p-1 \text{ factors})$$

$$p = 2, \cdots l+1, \qquad (15)$$

where S means symmetrization with respect to the order of factors.

In particular, for $l = 2$, we get Okubo's formula (A-8 in reference 3), from which the mass formula can be deduced.

B_l is the rotation group in $n = 2l+1$ dimensions. As a basis in $\Lambda(G)$, one may take the X_{ij} represented by the A_{ij} with $+1$ and -1 at the places ij and ji, and zero everywhere else. $A = a^\sigma A_\sigma$ becomes $A = \frac{1}{2}a^{ij}A_{ij}$ with $a^{ij} = -a^{ji}$, or $A = -A^T$. $\Delta(\omega) = \det(A-\omega)$ is an odd polynomial, and the coefficients of $\omega, \cdots \omega^{2l-1}$ yield l invariants. One may equivalently take the Tr A^{2p} for $p = 1, \cdots l$ (Tr A^{2p+1} is obviously zero).

From $F^{(p)} = \text{Tr } A^{2p}$, one gets

$$M^{(p)}_{ij} = SX_{ik}X_{kl}\cdots X_{nj} \quad (2p-1 \text{ factors}), \qquad (16)$$

(contractions over intermediate indices are to be taken with the conserved metric g^{ij}).

D_l is the rotation group in $n = 2l$ dimensions and all that has been said for B can be repeated here, except that here, $\Delta(\omega)$ is an even polynomial; the invariant of degree $2l$, which may be taken as $\det A$, being the determinant of a $2l$-dimensional skew matrix, is a perfect square, and must be replaced by one of its square roots.

C_l is the symplectic group in $n = 2l$ dimensions. It leaves the nondegenerate skew form (x, Hy) invariant, where $H = -H^T$ $\det H = 1$, $H^2 = -1$. The elements A of $\Lambda(C_l)$ are defined by $A^TH + HA = 0$, or HA symmetric. Then $\Delta(\omega) = \det(A-\omega) = \det(HA-\omega H) = \det[(HA)^T - \omega H^T] = \Delta(-\omega)$. Δ is an even polynomial and the coefficients of ω^{2p-2} $(p = 1, \cdots l)$ yield l invariants. One may equivalently take the $F^{(p)} = \text{Tr } A^{2p}$, $(p = 1, \cdots l)$. The Tr A^q for q odd are zero, for

$$\text{Tr } A^q = (-)^q \text{ Tr } [H(HA)]^q$$

$$= (-)^q \text{ Tr } [(HA)^T H^T]^q$$

$$= (-)^q \text{ Tr } [-(HA)H]^q = (-)^q \text{ Tr } A^q,$$

by carrying the last H from right to left and using again $H^2 = -1$. Therefore Tr $A^q = 0$ for q odd.

As a basis in $\Lambda(C_l)$, we may take the X^i_i represented by the $A^i_i = h_{ik}A^{ki}$, where A^{ij} has $+1$ at the places ij and ji and 0 everywhere else. From $F^{(p)} = \text{Tr } A^{2p}$ we then get

$$M^{(p)i}_{i} = SX^k_i X^l_k \cdots X^i_m \qquad (17)$$

$(2p-1$ factors, $p = 1, \cdots l)$.

V. THE EXCEPTIONAL GROUP G_2

G_2 has rank 2, therefore it has only two independent invariants or vector covariants. If one considers it is a subgroup of 0_7, one gets three such terms. We shall here derive directly the algebraic relation which reduces their number from 3 to 2. (See also reference 13). For that purpose we need the definition of G_2 given in reference 2 in a slightly different form. The seven eight-dimensional γ matrices γ_i $(i = 1, \cdots 7)$ are defined (up to an equivalence) by

$$[\gamma_i, \gamma_j]_+ = -2g_{ij} \quad (g_{ij} \text{ is the metric of } 0_7), \qquad (18)$$

and the matrix C by

$$\gamma^T_i = -C\gamma_i C^{-1};$$

C satisfies $C^T = C$.

C can be cast into the form

$$C = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & g_{ij} \end{pmatrix} (C = 1 \text{ if } g_{ij} = \delta_{ij})$$

by choosing a basis formed by a spinor η such that $\eta^T C\eta = 1$, and by the seven spinors $\eta_i = \gamma_i\eta$. The spin representation of 0_7 satisfies $U^TCU = C$, thus is a subgroup of the complex orthogonal group 0_8. The γ_i can be taken antihermitian, so that it is also a subgroup of real 0_8. The η_i provide a basis for the ordinary representation of real 0_7, which leaves g_{ij} invariant, and is a second subgroup of real 0_8. G_2 is the intersection of these two subgroups. To any 7×7 matrix $A \in \Lambda(0_7)$ ($A = \frac{1}{2}a^{ij}A_{ij}$ with the notations of Sec. IV), the spin representation associates the 8×8 matrix $B = -\frac{1}{2}a^{ij}G_{ij}$[16] where $G_{ij} = \frac{1}{4}[\gamma_i, \gamma_j]$ corresponds to $-A_{ij}$ in the representation. B is entirely defined from A by

$$[\gamma^k, B] = a^{kl}\gamma_l. \qquad (19)$$

In the preceding basis, γ_i has the form

[16] The minus sign comes from the one in Eq. (18).

0	0	1	0
0			
−1	Γ_{ijk}		
0			

where the ± 1 stand at the ith place in the corresponding subrows. The condition that B belongs to $\Lambda(G_2)$ can be expressed by

$$B = \begin{bmatrix} 0 & 0 \\ 0 & B' \end{bmatrix},$$

where B' is a 7×7 matrix. Identification in (19) yields $B' = A$. If we identify the 7×7 matrix A with the 8×8 matrix

$$\begin{bmatrix} 0 & 0 \\ 0 & A \end{bmatrix},$$

the $A \in \Lambda(G_2)$ can be defined by

$$[\gamma^k, A] = a^{kl}\gamma_l. \tag{20}$$

Remark: Expressing the γ's in terms of the Γ's gives

$$\Gamma^{ij}{}_l\, a^{lk} + \Gamma^{i}{}_l{}^{k} a^{lj} + \Gamma^{jk}_{l} a^{li} = 0, \tag{21}$$

which expresses the fact that G_2 leaves the trilinear totally skew form $\Gamma_{ijk}x^iy^jz^k$ invariant.
Iteration of (20) yields

$$(A^p)^{kl}\gamma_l = [\cdots [\gamma^k, A], \cdots, A], \tag{22}$$

where $(A^p)^{kl}$ is the element kl of A^p, and with p factors A on the right hand side. It can be proved by complete induction that

$$(A^p)^{kl}\gamma_l = \sum_{q=0}^{q=p} C_p^q (-)^q A^q \gamma^k A^{p-q}, \tag{23}$$

where $C_p^q = p!/q!\,(p-q)!$

Multiplying by γ_m and taking the trace, one gets from (23),

$$(A^p)^{km} = -\tfrac{1}{8} \sum_q C_p^q (-)^q \operatorname{Tr}(\gamma^m A^q \gamma^k A^{p-q}). \tag{24}$$

Contraction with g_{mk} gives

$$\operatorname{Tr} A^p = -\tfrac{1}{8} \sum_q C_p^q (-)^q \operatorname{Tr}(A^q \gamma_k A^{p-q} \gamma^k). \tag{25}$$

We now consider the two invariants $\operatorname{Tr} A^2$ and $\operatorname{Tr} A^4$.

Multiplying (23) on the left by γ_k and summing over k gives remarkable identities for $p = 1$ and $p = 2$, respectively:

$$\gamma_k\gamma^k A - \gamma_k A \gamma^k = a^{kl}\gamma_k\gamma_l = -4A,$$

or

$$\gamma_k A \gamma^k = -3A, \tag{26}$$

and

$$\gamma_k A^2 \gamma^k = A^2 - (\operatorname{Tr} A^2)\mathbb{1} \tag{27}$$

where $\mathbb{1}$ is the 8×8 unit matrix. (25) for $p = 4$ is easily transformed by use of (26) and (27) into

$$-3 \operatorname{tr} A^4 = \operatorname{Tr} A^2(A^2 - (\operatorname{Tr} A^2)\mathbb{1})$$

or

$$\operatorname{Tr} A^4 = \tfrac{1}{4}(\operatorname{Tr} A^2)^2. \tag{28}$$

This allows us to keep only the invariants of order 2 and 6, and, therefore, the vector covariants of order 1 and 5.

VI. CONCLUSION

We have seen that for simple Lie groups, the matrix elements $\langle \psi_1 | M_\sigma | \psi_2 \rangle$ of a vector operator M_σ within an irreducible representation E are linear combination of those of a finite number of operators, with coefficients depending only on M_σ and E; the number of linearly independent terms is $l - q$, where l is the rank of the group and q is defined in Sec. II.

We found l vector operators, in the form of polynomials of the infinitesimal generators of the group, and admitted a step in the proof showing that they yield a generating system for the preceding decomposition. (Of course only $l - q$ of them are linearly independent in a given E). Further study of the vector invariants of the adjoint representation (ρ) of the group is needed to complete the proof. We conclude by noting that the method applies to any type of tensor operator, i.e. to any set of operators which transforms according to a tensor power of (ρ).

APPENDIX

We prove here the result stated in Sec. II. We need the following lemmas:

Lemma 1. For any particular α^+, the sum in R can be split into 3 partial sums:

$$R = R_0 + R_b + \tfrac{1}{2}\alpha^+,$$

where

$$\begin{cases} R_0 \in \pi_\alpha \ (\pi_\alpha \text{ being the } l-1 \text{ plane orthogonal to } \alpha^+ \text{ and } \alpha^-), \\ R_b \text{ is on the same side of } \pi_\alpha \text{ as } \alpha^+. \end{cases}$$

Proof: There exists an $(l-1)$ plane U and an $(l-1)$ plane V such that

(a) U entails no root and every > 0 root is on the same side of U as α^+.

(b) V entails α^+ and α^- but no other root.
(c) $V \cap U$ is contained in π_α.

As the defining equations for π, U and V, one can take

$$\pi: \alpha_i x^i = 0,$$

$$U: h(x) = x^1 + \epsilon_1 x^2 + \cdots + \epsilon_{l-1} x^l = 0 \quad (\epsilon\text{'s} > 0),$$

$$V: \alpha_i x^i + \lambda h(x) = 0, \quad \text{with}$$

$$\lambda = -\alpha^2(\alpha_1 + \alpha_2\epsilon_1 + \cdots + \alpha_l\epsilon_{l-1})^{-1}.$$

There are ∞^{l-1} such families depending on the $\epsilon_1, \cdots, \epsilon_{l-1}$. Condition (a) can be satisfied by choosing each ϵ small enough with respect to the preceding one so that the sign of $h(x)$ be determined for any root by the sign of the first nonzero coordinate. Condition (c) is satisfied by construction and (b) can be too, for the only restriction imposed on the preceding V is that it contains α^+ (and α^-).

U and V divide the space into four regions: (*a*) and (*b*), which are on the same side of U as α^+, (*a*) containing the perpendicular to U, and (*c*) and (*d*) which are on the other side, (*c*) containing the perpendicular to U. We note R_h, the sum of the roots contained in the region (*h*). Then

$$R = \tfrac{1}{2}(R_a + R_b + \alpha^+) \qquad R_b = -R_d, \qquad R_a = -R_c.$$

$R_a + R_d$ is contained in π_α, for the root diagram is invariant by π symmetry. Therefore,

$$R = \tfrac{1}{2}(R_a + R_d) + R_b + \tfrac{1}{2}\alpha^+,$$

whence lemma 1 follows with

$$R_0 = \tfrac{1}{2}(R_a + R_d).$$

R_b clearly has the required property.

Lemma 2. $R - \frac{1}{2}\alpha^+$ is orthogonal to α^+ if and only if there is an $(l-1)$ plane V containing α^+ and such that the set of positive roots $\neq\alpha^+$ be the set of roots on the one side of V (property $\mathcal{P}$).

Proof: By lemma 1, $R - \frac{1}{2}\alpha^+ = R_0 + R_b$. If $R - \frac{1}{2}\alpha^+$ lies in π_α, R_b lies in π_α too. But $\pi_\alpha \cap (b)$ is contained in U, so R_b is in U. Now R_b is a sum of roots none of which is in U and which all are on the same side of U. Therefore there cannot be any root in (*b*), and $R_b = 0$.

The positive roots $\neq\alpha^+$ are those in (*a*) and (*b*). There is none in (*b*), so they lie in (*a*). As there is no root in (*b*), there is none in (*d*) either. So V has the required property.

Conversely, if there is a V with property $\mathcal{P}$, invariance of the root diagram by π_α symmetry (which leaves invariant the two sides of V) yields the result.

R is dominant, therefore lies in D_1. Lemmas 1 and 2 then show that D_1 is the domain which for each pair (α^+, α^-) is on the same side of π_α as α^+. We next prove:

Lemma 3. π_α effectively limits D_1 if and only if there is an $(l-1)$ plane with property $\mathcal{P}$.

Proof: Let α^+ be a positive root and u^+ any vector in π_α on the same side of some U than the positive roots. Let T be the plane orthogonal to u^+. α^+ is in T. "π_α limits D_1" is equivalent to "there is some u^+ which is on the + side of π_β for every $\beta^+ \neq \alpha^+$", which is equivalent to "there is some T containing α^+ such that every β^+ is on the u^+ side of T", whence lemma 3.

We come back to our problem. Special cases occur when for some α and $S \in W$ $SK + \alpha = K$. SK and K have the same length, therefore

$$\begin{cases} K = J + \frac{1}{2}\alpha \quad \text{where} \quad J \text{ lies in } \pi_\alpha, \\ SK = J - \frac{1}{2}\alpha, \end{cases}$$

and S must be symmetry with respect to π_α, so $\delta_S = -1$.

It follows from the preceding lemmas that this can occur only for positive roots α^+, the π's of which effectively limit D_1. It then occurs once for each root α^+ such that $L = K - R$ be parallel to π_α.

Remembering that the relevant α's are in number l and that the l fundamental weights are parallel to the l combinations of $l-1$ of their π's, we come to the result stated in Sec. II.

ACKNOWLEDGMENTS

The author is grateful to Professor B. d'Espagnat, Professor Ph. Meyer, and Professor M. Gourdin for the interest which they have taken in this work.